الرؤية : أن يكون المعهد رائداً ومعتمداً أكاديمياً في مجالات العلوم الهندسية والتكنولوجية على المستوى المحلي والإقليمي.

Ministry of Higher Education and Scientific Research
Higher Institute of Engineering and Technology, Kafrelsheikh

DIGITAL COMMUNICATION SYSTEMS EEE 3213

Electrical Engineering Department

Electronics and Electrical Communications Engineering

Third Year

Dr.Amira Alshoraky

2025-2026

الرسالة: يلتزم برنامج هندسة الإلكترونيات والاتصالات الكهربائية بإعداد خريجين مؤهلين بالكفاءات اللازمة إلى جانب الأخلاقيات المهنية الضرورية لتلبية المتطلبات المتزايدة لسوق العمل، واستراتيجيات التطوير، وإجراء البحوث العلمية المبتكرة، وتقديم الخدمات المجتمعية.

Contents

Preface

Today, communication enters our daily lives in so many different ways that it is very easy to overlook the multitude of its facets. Mobile phones, radio and TV, Internet access points, ….etc., are all capable of providing rapid communications from every corner of the globe. Indeed, the list of applications involving the use of communication in one way or another is almost endless.

Digital communication are becoming increasingly attractive due to its advantages that make it easy to handle and flexible compared to analog communication

This book was written for under-graduate students interested in electrical communication systems, those require a solid understanding and a consolidated comprehensive text of digital communication systems. Content of the book has been selected to be adequate for a one-semester course in digital comm. systems. The clear trend in today's academia is to teach less analog and more digital material. The book focuses on digital communication systems, aiming to provide its basics to the students of communication engineering, and explain them in the simplest and most efficient way. It covers the basic building blocks of pulse and digital communications systems. Our material on base band pulses and carrier modulation has been completely recast in order to achieve a simpler exposition and introduce a stronger engineering flavor.

The book is organized as follows;

Chapter 1, Introduction, introduces the overall structure of the digital communication system followed by its advantages and disadvantages.

Chapter 2, **Digital pulse modulation**, covers the process of converting the analog signal into digital one. It discusses the processes of sampling, quantization, and coding that are fundamental to the digital transmission of analog signals. This chapter may be viewed as the transition from analog to digital communication. The characteristics of pulse amplitude modulation, PAM are discussed in detail, as it is the basic to all forms of pulse modulation, be they of the analog or digital type. PCM signal are discussed and evaluated, followed by its differential version, DPCM, delta modulation, and its adaptive format are also discussed.

Chapter 3, Multiplexing Techniques, differentiate between analog, and digital multiplexing, describing the hierarchy structure of each of them.

Chapter 4, Digital Modulation, discusses *passband data transmission*, where a sinusoidal carrier wave is employed to facilitate the transmission of the digitally modulated wave over a band pass channel. Different digital modulation techniques are discussed such as; ASK, FSK, and PSK with its M-ary techniques as power and spectrum efficient digital modulation techniques. Hybrid amplitude/phase modulation schemes including QAM is also given. This is followed by MSK, and GMSK discussion as methods to overcome FSK disadvantages. The chapter is ended by evaluation of the effect of channel noise on the performance of digital communication systems.

Moving through this book, and studying the numerous examples, the student learns the fundamentals of the subject and acquires the tools and skills necessary to analyze and design elements of modern communication systems. The prerequisite for this book is a course in analog communication.

The text is ended by a sample of the references used to cover the area of the "Digital Communication".

Hopefully this text including something useful, but as every thing in our life still something missed.

Chapter

1

Introduction

Digital communications refers to the field of study concerned with the transmission of digital data in contrast with analogue communications. While analogue communications use a continuously varying signal, a digital transmission can be broken down into discrete messages. Transmitting data in discrete messages allows for greater signal processing capability. The ability to process a communication signal means that errors caused by random processes can be detected and corrected. Digital signals can also be sampled instead of continuously monitored and multiple signals can be multiplexed together to form one signal.

One advantage of digital information is that it tends to be far more resistant to transmitted and interpreted errors than information symbolized in an analogue medium. This accounts for the clarity of digitally-encoded telephone connections, compact audio disks, and for much of the enthusiasm in the engineering community for digital comm.'s technology. Digital communication is so full of concepts, processes, algorithms, and complexity because this is the route to lower cost.

In 1841, the great mathematician Augustin-Louis Cauchy first proposes the sampling theorem. Nearly 80 years later J.R. Carson published a mathematical analysis of time sampling in communications.
In a 1928 lecture at the American Institute of Electrical Engineers Harry Nyquist provides proof of the sampling theorem in "Certain Topics in Telegraph Transmission Theory". In 1937, A. Reeves proposes Pulse Code wave Modulation (PCM). In 1948, John Bordeen, William Shockley, and Walter Brattain invent the bipolar junction transistor at Bell Labs-compact digital circuitry is a reality. Two years later, in 1950 Richard W. Hamming publishes significant work on error correction and detection codes. Voice digitization and transmission first became feasible in the late 1950s when solid state electronics became available. In 1962 Bell System personnel established the first commercial use of digital transmission when they began operating a T1 carrier system for use as a trunk group in a Chicago area exchange. After the Tl system a family of T-carrier systems (Tl, TlC, TlD, T2, T3, T4) were developed, all of which involved Time Division Multiplexing, TDM of digitized voice signals. In 1965, NHK Technical Research Institute publicly demonstrates a PCM digital audio recorder with a 30 kHz sampling rate and 12-bit resolution. The world's first commercially designed digital microwave radio system was established in Japan by Nippon Electric Company (NEC) in 1968.
In the early 1970s digital microwave systems began to appear in the United States for specialized data transmission services. The first digital microwave link in the U.S. public telephone network was

supplied by NEC of Japan for a New York telephone link. Digital microwave systems were subsequently developed and installed by several U.S. manufacturers for use in intermediate-length toll and exchange area circuits.

Bell System's first commercial use of digital fiber optic transmission occurred in September of 1980 on a short-haul route. Three years later the first long-haul system between New York and Washington, D.C., was put into service.

The objective of designing a communication system is for the electrical signal at the transmitting end to be reproduced at the receiving end with minimal distortion. To achieve this, different techniques are used, depending on issues such as type of data, type of communication medium, distance to be covered, and so forth. Figure 1.1, illustrates the typical functional elements of a digital communication system.

The information source generates particular symbols at a particular rate. The source encoder translates these symbols in sequences of 0's and 1's. The channel encoder is oriented towards translating sequences of 0's and 1's to other sequences of 0's and 1's, to realize high transmission reliability and efficiency.

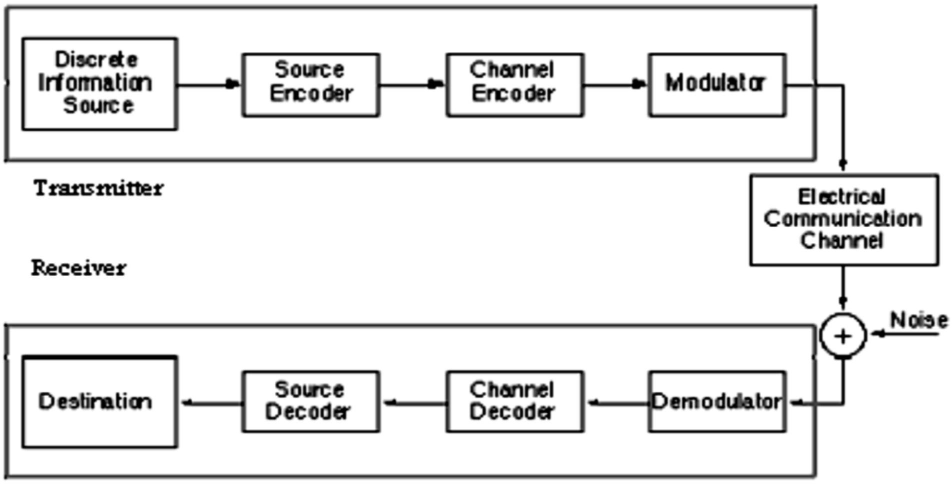


Figure 1.1: Digital Communication System

The modulator accepts streams of 0's and 1's, and converts them to electrical waveforms suitable for transmission.

The communication channel provides the connection between the source and destination. It may be wired or wireless with a finite bandwidth. The transmitted waveform suffers from amplitude and phase distortions in addition to power decrease due to channel attenuation especially with large distance.

Finally, the waveform is corrupted by unwanted signals, referred to as noise. The primary objective of a communication system is to suppress the bad effects of noise as much as possible. Attenuated signal may be recovered by amplifiers, but noise can't be eliminated at all as the amplifiers amplify the added noise as well. Amplification alone does not solve the problem, particularly when the system has to cover large distances.

The inverse process takes place at the destination side.

The demodulator converts the electrical waveforms to sequences of 0's and 1's, the channel decoder translates the sequence of 0's and 1's to the original sequence of 0's and 1's. It also performs error correction and clock recovery. The source decoder finally translates the sequence of 0's and 1's into symbols. Figure 1.2 illustrates a more detailed block diagram of digital communication system with main processes that carried to transmit the base band signal.

If individual time values of the discrete-time signal, instead of being measured precisely (which would require an infinite number of digits), are approximated to a certain precision-which, therefore, only requires a specific number of digits-then the resultant data stream is termed a digital signal. The process of approximating the precise value within a fixed number of digits, or bits, is called quantization.

In summary, a digital signal is a quantized discrete-time signal; a discrete-time signal is a sampled analogue signal. In most applications, digital signals are represented as binary numbers, so their precision of quantization is measured in binary digits or bits.

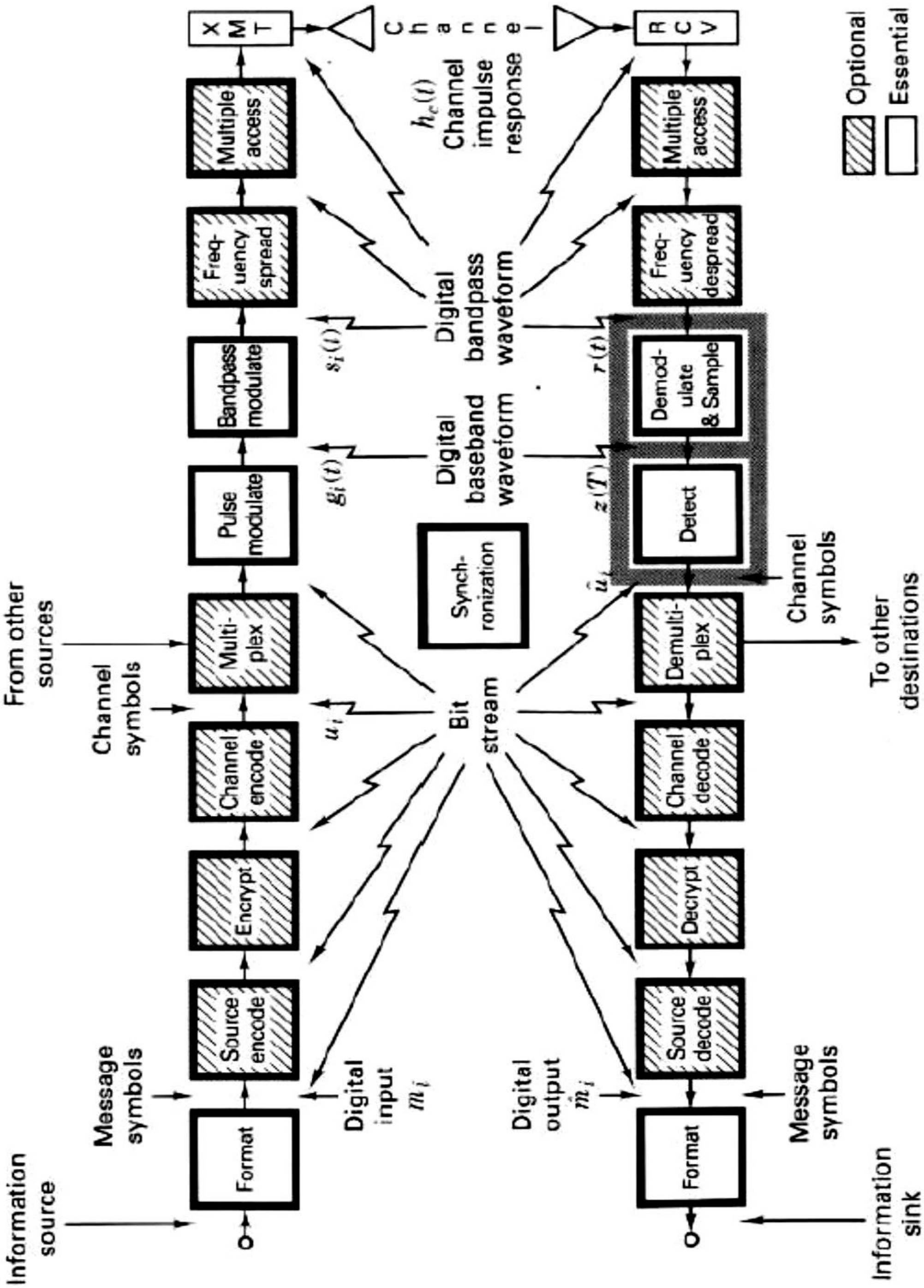


Figure 1.2 Base band transmission

1.1 Digital Network Evolution

The evolution of the analogue telephone network into one that is all digital except for the access lines is summarized in figure 1.3.

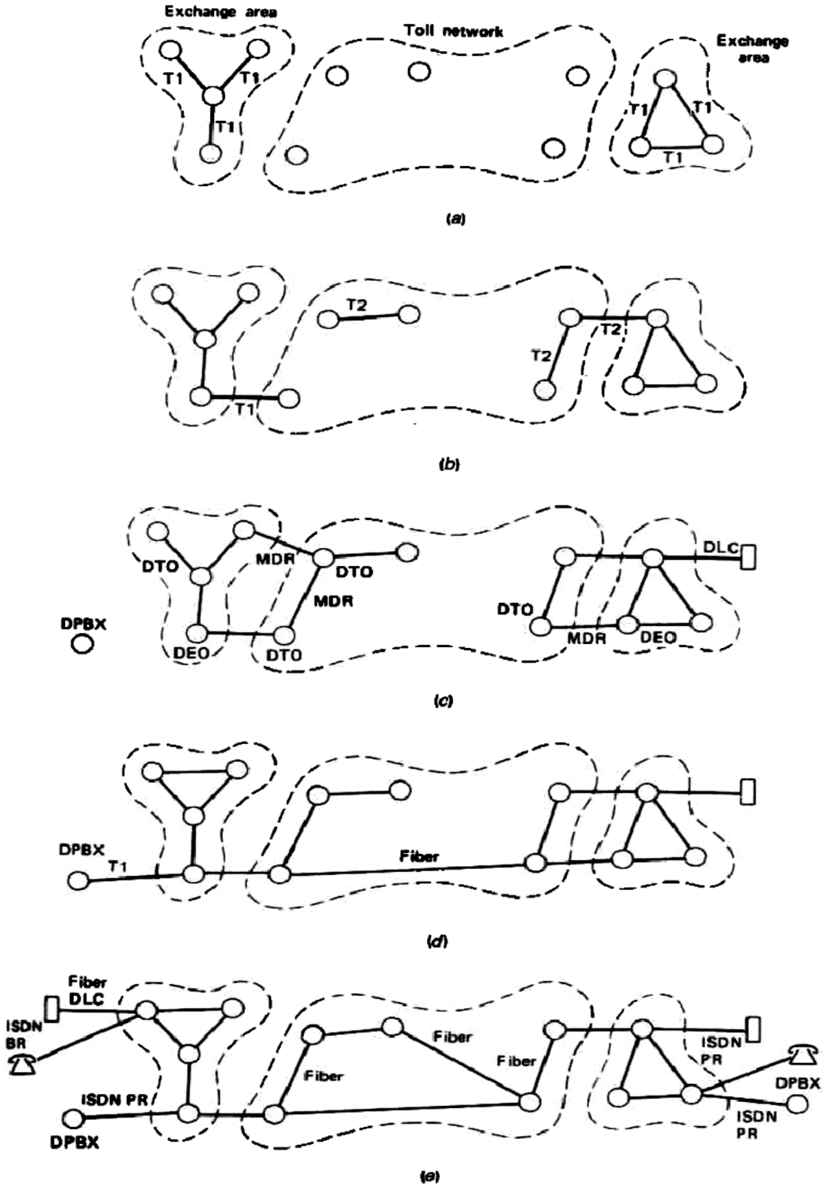


Figure 1.3 Digital network evolution

The process began in the 1960s (a) with T1 systems being installed on relatively short haul interoffice trunks within the exchange areas. Next, in the early 1970s (b), digital transmission was introduced into the toll (trunk) network with T2 systems for relatively short routes between toll offices. It was in the late 1970s (c) that digitization really began to take over.

Tl coverage expanded greatly, Digital Loop Carrier (DLC) systems came into use, Digital subscriber carrier systems were actually introduced in the early 1970s, but these systems utilized a voice digitization technology (delta modulation) that was incompatible with the rest of the network and therefore did not figure into the integrated network. These early digital loop carrier systems are now obsolete), and digital switches became available at all levels of the network: PBXs (DPBXs), End Offices (DEOs), Tandem Offices, and Toll Offices (DTOs). Moreover, Microwave Digital Radios (MDRs) proved to be advantageous to use in both the exchange areas and the shorter toll network routes due to low interface costs to digital switches. Thus, the late 1970s produced a number of integrated islands where digital switches within a region were interconnected by digital transmission links but there was little digital connectivity between the islands. (Data under voice was installed as an over build to analogue routes for limited-capacity data services.). A fully integrated and interconnected digital network became a reality in the early 1980s (d) when fiber optic transmission emerged as the technology of choice for high-density long-haul routes.

Digital connectivity to business customer premises equipment also occurred in this time frame as Tl became the preferred voice trunk interface for large PBXs. End-to-end digital connectivity for voice or data services became a reality in the late 1980s (e) with the introduction of ISDN basic rate (ISDN BR, 28 + D) and ISDN Primary Rate (ISDN PR, 238 + D) digital connections to the customer. In addition, fiber technology became much more ubiquitous as DS3 rate systems eliminated T2 systems in the toll network and fiber-based systems became the preferred technology for loop carrier and feeder systems-even at relatively short distances.

1.2 Advantages of digital Voice Networks

The main advantage of digital information is that it tends to be far more resistant to transmitted and interpreted errors than information symbolized in an analogue medium. This accounts for the clarity of digitally-encoded telephone connections as example. However, digital communication has its own unique pitfalls, and there are multitudes of different and incompatible ways in which it can be sent. The technical advantages of digital communications are;

1. Ease of multiplexing,
2. Ease of signaling,
3. Use of modern technology,
4. Integration of transmission and switching,
5. Signal regeneration,
6. Performance monitor ability,
7. Accommodation of other services,
8. Operation at low signal-to-noise/interference ratios,

9. Ease of encryption.

These features are listed in the order of their relative importance for general telephony. In particular applications, however, certain considerations may be more or less significant. Because of all these advantages, and because recent advances in wideband communication channels and solid-state electronics have allowed scientists to fully realize these advantages, digital communications has grown quickly, and is quickly edging out analogue communication due to the vast demand to transmit computer data and the ability of digital communications to do so. For instance, the last item, ease of encryption, is a dominant feature favoring digital networks for the military. Encryption, for example, is practical and generally useful only if the secure form of the message is established at the source and translated back into the clear form only at the destination. Thus an end-to-end digital system that operates with no knowledge of the nature of the traffic (i.e., provides transparent transmission) is a requirement for digital encryption applications. For similar reasons end-to-end digital transmission is needed for direct transmission of data (no modem). When a network consists of a mixture of analogue and digital equipment, universal use of the network for services such as data transmission dictates conformance to the least common denominator of the network: the analogue channel.

1.2-1 Ease of Multiplexing

The multiplexing function involves nothing more than cyclically sampling many data streams. Such an operation assumes all of the data streams are synchronized to each other. The process of synchronizing the data streams requires logic circuitry that is much more complex. Digital techniques were first applied to general telephony in interoffice T-carrier (TDM) systems. In essence, these systems traded electronics costs at the ends of a transmission path for the cost of multiple pairs of wires between them. Although FDM of analogue signals had also been used to reduce cable costs, FDM equipment is much more expensive than TDM equipment, even when the cost of digitization is included. After voice signals have been digitized, TDM equipment costs are quite small by comparison. Since digitization occurs only at the first level of the TDM hierarchy, high-level digital TDM is even more economical than high-level FDM counterpart. It should be pointed out that TDM of analogue signals is also very simple and does not require digitization of the sample values.

The drawback of analogue TDM lies in the vulnerability of narrow analogue pulses to noise, distortion, crosstalk, and inter-symbol interference. These degradations cannot be removed by regeneration as in a digital system. Hence, analogue TDM is not feasible except for noiseless, distortion -free environments (Analogue TDM has been used in a few telephone applications that use pulse-width-modulated TDM, and some older PBXs.) In essence the ability to regenerate a signal, even at the expense of greater band width, is almost a requirement for TDM transmission.

1.2-2 Ease of Signaling

Control information (e.g., on-hook/off-hook, address digits) is inherently digital and, hence, readily incorporated into a digital transmission system. One means of incorporating control information into a digital transmission link involves time division multiplexing the control as a separate but easily identifiable control channel. Another approach involves inserting special control codes into the message channel and having digital logic in the receiving terminals decode that control information. In either case, as far as the transmission system is concerned, control information is indistinguishable from message traffic.

In contrast, analogue transmission systems required special attention for control signaling.

Many analogue transmission systems presented unique and sometimes difficult environments for inserting control information. The control formats depend on the nature of both the transmission system and its terminal equipment. In some interfaces between network subsystems control information had to be converted from one format to another. Signaling on analogue links therefore represented a significant administrative and financial burden to the operating telephone companies.

The move to common-channel signaling removed most of the signaling costs associated with interoffice trunks but did not change the situation for individual subscriber lines, which must carry signaling on the same facility as the message channel.

The use of Digital Subscriber Lines (DSLs) reduces the signaling costs relative to analogue subscriber lines, which helps offset the higher cost of a DSL and a digital telephone. DSLs are a fundamental aspect of ISDN.

In summary, digital systems allow control information to be inserted into and extracted from a message stream independently of the nature of the transmission medium (e.g., cable, fiber, microwave, satellite). Thus the signaling equipment can (and should) be designed separately from the transmission system. It then follows that control functions and formats can be modified independently of the transmission subsystem. Conversely, digital transmission systems can be upgraded without impacting control functions at either end of the link.

1.2-3 Use of Modern Technology

A multiplexer or switching matrix for time division digital signals is implemented with the same basic circuits used to build digital computers: logic gates and memory. The basic cross point of a digital switch is nothing more than an AND gate with one logic input assigned to the message signal and other inputs used for control (cross point selection). Thus the dramatic developments of digital integrated circuit technology for computer logic circuits and memory are applicable directly to digital transmission and switching systems. In fact, many standard circuits developed for use in computers are directly usable in a switching matrix or multiplexer. Nevertheless, the implementation of TDM is much less expensive than analogue FDM.

Even greater advantages of modern technology have been achieved by using Large Scale Integrated (LSI) circuits designed specifically for telecommunications functions such as voice encoding/decoding, multiplexing/de-multiplexing, switching matrices, and special-purpose and general-purpose Digital Signal Processing (DSP).

The relative low cost and high performance of digital circuits allows digital implementations to be used in some applications that are prohibitively expensive when implemented with comparable analogue components. Completely non blocking switches, for example, are not practical with conventional analogue implementations, except in small sizes. In a modern digital switch the cost of the switching matrix itself is relatively insignificant. Thus, for medium-size applications, the size of the switch matrix can be increased to provide non blocking operations, if desired. The automatic call distributor developed by Collins- Rockwell is an early example of a digital switch operating in an analogue environment. A digital implementation was chosen largely because it could economically provide a non blocking operation, and it has an inherent functional advantage over analogue implementations. This advantage is derived from the relative ease with which digital signals can be multiplexed.

The technological development to have the most significant impact on the telephone network is certainly fiber optic transmission. Although fibers themselves do not favor digital transmission over analogue transmission, the interface electronics to a fiber system function primarily in an on-off (nonlinear) mode of operation. Thus digital transmission dominates fiber applications, although analogue optical technology is commonly used in analogue video distribution.

1.2-4 Integration of Transmission and Switching

Traditionally the analogue transmission and switching systems of telephone networks were designed and administered by functionally independent organizations.

In the operating telephone companies, these two equipment classes are referred to as outside plant and inside plant, respectively.

In fact, the first stages of digital switches generate first-level TDM signals by nature, even when interfaced to analogue transmission links. Thus the multiplexing operations of a transmission system can be easily integrated into the switching equipment. The basic advantage of integrating the two systems is shown in figure 1.4.

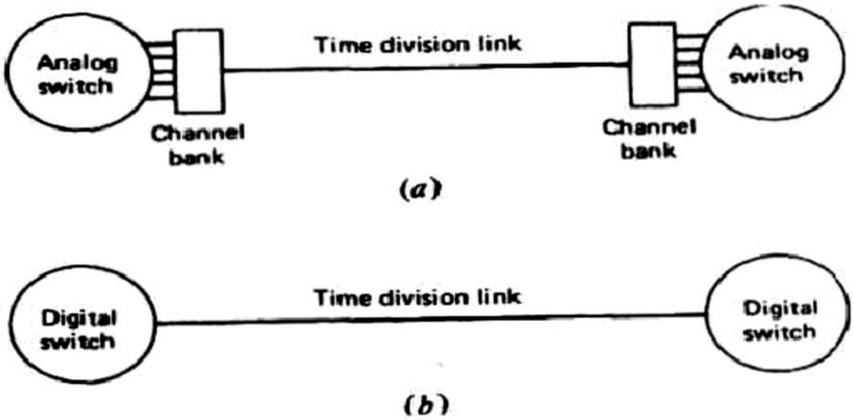


Figure 1.4 Integration of transmission and switching:
(a) non-integrated transmission and switching,
(b) integrated time division switching and transmission

The de-multiplexing equipment (channel banks) at the switching offices is unnecessary, and first-stage switching equipment is eliminated. If both ends of the digital TDM trunks are integrated into a digital switch, the channel banks at both ends of the trunk are eliminated. In a totally integrated network voice signals are digitized at or near the source and remain digitized until delivered to their destination. Furthermore, all interoffice trunks and internal links of a switching system carry TDM signals exclusively.

Thus first-level multiplexing and de-multiplexing are non-existent except at the periphery of the network. Integration of transmission and switching functions not only eliminates much equipment but also greatly improves end-to-end voice quality. By eliminating multiple A/D and D/A conversions and by using low-error-rate transmission links, voice quality is determined only by the encoding / decoding processes.

In summary, the implementation benefits of a fully integrated digital network are:

1. Long-distance voice quality is identical to local voice quality in all aspects of noise, signal level, and distortion.
2. Since digital circuits are inherently four-wire, network-generated echoes are eliminated, and true full-duplex, four-wire digital circuits are available.
3. Cable entrance requirements and mainframe distribution of wire pairs is greatly reduced because all trunks are implemented as sub channels of a TDM signal.

1.2-5 Signal Regeneration

As shown in figure 1.5, the transmission process of the binary data despite the existence of certain imperfections, does not alter the essential nature of the information.

Figure 1.5 signal regeneration in a digital repeater line

Of course, if the imperfections cause sufficient changes in the signal, errors detection occur and the binary data in the receiver does not represent the original data exactly.

A fundamental attribute of a digital system is that the probability of errors transmission can be made arbitrarily small by inserting regenerative repeaters at intermediate points in the transmission link. If spaced close enough together, these intermediate nodes detect and regenerate the digital signals before channel-induced degradations become large enough to cause decision errors. The end-to- end error rate can be made arbitrarily small by inserting a sufficient number of regeneration nodes in the transmission link.

The A/D conversion process inherently introduces possibility of error that can be minimized by establishing enough discrete levels. Increased resolution requires more bits and consequently more bandwidth for transmission. Hence, a digital transmission system readily provides a trade-off between transmission quality and bandwidth.

1.2-6 Performance Monitor ability

An additional benefit of the source-independent signal structure in a digital transmission system is that the quality of the received signal can be ascertained with no knowledge of the nature of the traffic.

The transmission link is designed to produce well-defined pulses with discrete levels. Any deviation in the received signal, represents a degradation in transmission quality.

In general, analogue systems cannot be monitored or tested for quality while in service since the transmitted signal structure is unknown.

FDM signals typically include pilot signals to measure channel continuity and power levels. The power level of a pilot is an effective means of estimating the Signal-to-Noise Ratio, SNR-only in a fixed noise environment.

One common method of measuring the quality of a digital transmission link is to add parity, or Cyclic Redundancy Check (CRC), bits to the message stream. The redundancy introduced to the data stream enables digital logic circuits in a receiver to readily ascertain channel error rates. If the error rate exceeds some nominal value, the transmission link is degraded.

1.2-7 Accommodation of other Services

Any digitally encoded message (whether inherently digital or converted from analogue) presents a common signal format to the transmission system. It accommodates control (signaling) information. In an analogue network the transmission standard is the 4-kHz voice circuit. All special service such as data or facsimile must be transformed "to look like voice." In particular, data signals must be converted to an analogue format through the use of modems.

The standard analogue channel was necessarily optimized for voice quality. Use of an analogue network for non voice services often requires special compensation for various analogue transmission impairments. In contrast, the main parameter of quality in a digital system is the error rate. Low-error-rate channels are readily obtainable. When desired, the effects of channel errors can be effectively eliminated with error control procedures implemented by the user.

An additional benefit of the common transmission format is that traffic from different types of sources can be intermixed in a single transmission medium without mutual interference. The use of a common transmission medium for analogue signals is sometimes complicated because individual services require differing levels of quality.

For example, television signals, which require greater transmission quality than voice signals, were not usually combined with FDM voice channels in a wideband analogue transmission system.

1.2-8 Operation at Low S/N,I Ratios

Noise and interference in an analogue voice network become most apparent during speech pauses when the signal amplitude is low. Relatively small amounts of noise occurring during a speech pause can be quite annoying to a listener. The same levels of noise or interference are virtually unnoticeable when speech is present. Hence it is the absolute noise level of an idle channel that determines analogue speech quality. Subjective evaluations of voice quality led to maximum noise level standards of - 62 dBm0 for short-haul systems and -56 dBm0 for long-haul systems.

For comparison, the power level of an active talker is typically -16 dBm0. Thus representative end-to-end SNRs in analogue networks are 46 and 40 dB for short- and long-haul systems, respectively. SNRs on individual transmission systems are necessarily higher.

In a digital system speech pauses are encoded with a particular data pattern and transmitted at the same power level as active speech. Because signal regeneration virtually eliminates all noise arising in the transmission medium, idle channel noise is determined by the encoding process and not the transmission link. Thus speech pauses do not determine maximum noise levels as they do in an analogue system. Digital transmission links provide virtually error-free performance at SNRs of 15-25 dB, depending on the type of line coding or modulation used. The ability of a digital transmission system to reject crosstalk is sometimes more significant than its ability to operate in relatively high levels of random noise. Low levels of crosstalk are eliminated by the regeneration process in a digital repeater or receiver. Even if the crosstalk is of sufficient amplitude to cause detection errors, the effects appear as random noise and, as such, are unintelligible.

Considering the fact that a digital system typically needs a greater bandwidth than a comparable analogue system and that wider bandwidths imply greater crosstalk and noise levels, the ability to operate at lower SNRs is partly a requirement of a digital system and partly an advantage.

1.2-9 Ease of Encryption

Although most telephone users have little need for voice encryption, the ease with which a digital bit stream can be scrambled and unscrambled means that a digital network (or a digital cellular system) provides an extra bonus for users with sensitive conversations. In contrast, analogue voice is much more difficult to encrypt and is generally not nearly as secure as digitally encrypted voice. Ease of encryption stimulated early use of digital voice systems by the military.

1.3- Digital Signal Processing

It is another significant application of digital technology. Basically, signal processing refers to an operation on a signal to enhance or transform its characteristics. Signal processing can be applied to either analogue or digital waveforms. Amplification, equalization, modulation, and filtering are common examples of signal processing functions.

Digital Signal Processing (DSP) refers to the use of digital logic and arithmetic circuits to implement signal processing functions on digitized signal waveforms. Sometimes analogue signals are converted to digital representations for the express purpose of processing them digitally. Then the digital representations of the processed signals are converted back to analogue. These operations are illustrated in figure 1.6, where

a sine wave corrupted by noise is digitally filtered to remove the noise.

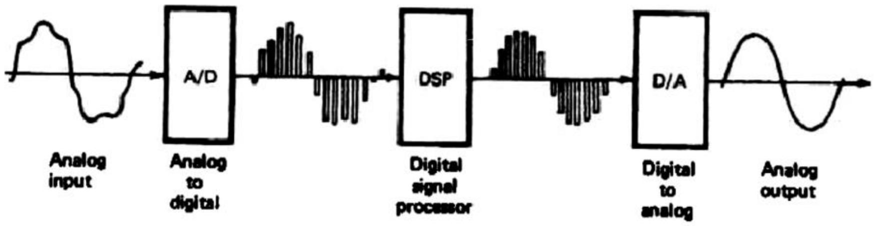


Figure 1.6 Digital signal processing of an analogue signal

1.4 Disadvantages of Digital Voice Networks

The basic technical disadvantages of digital implementations are;

1. Increased bandwidth,
2. Need for time synchronization,

3. Topologically restricted multiplexing,
4. Need for conference/extension bridges,
5. Incompatibilities with analogue facilities.

1.3-1 Increased Bandwidth

The bandwidth expansion of digital voice comes when the samples are encoded into binary codes and transmitted with an individual pulse for each bit in the code. Thus a T_1 system requires approximately eight times as much bandwidth as do 24 analogue voice channel since each sample is represented by an 8-bit code word and each bit is transmitted as a separate discrete pulse. Although more sophisticated digitization algorithms can be used to encode voice at a lower bit rate than that used on Tl systems (64 kbps), even the most sophisticated algorithms cannot provide comparable voice quality without at least a two-to-one bandwidth penalty. This bandwidth penalty imposed by voice digitization is directly dependent on the form of transmission coding or modulation used. With greater sophistication in the modulation / demodulation equipment, greater efficiency in terms of the bit rate in a given bandwidth is achievable. Basically, greater transmission efficiency is achieved by increasing the number of levels in the line code with limited transmit power. However, the distances between discrete signal levels in the receiver are reduced dramatically. Thus, the transmitted signal is no longer as immune to noise and other imperfections as it is with lower information densities. Using a combination of advanced digital modulation, lower rate digitization, and error-correcting codes, point-to-point digital radios could provide voice channel efficiencies comparable to or even better than analogue microwave systems.

Full development along these lines never occurred, however, because the emergence of optical fibre transmission eliminated the incentive to do so.

1.3-2 Need for Time Synchronization

Whenever digital information is transmitted from one place to another, a timing reference, or "clock," is needed to control the transfer. The clock specifies when to sample the incoming signal to decide which data value was transmitted. The optimum sample times usually correspond to the middle of the transmitted pulses. Thus, for optimum detection, the sample clock must be synchronized to the pulse arrival times. In general, the generation of a local timing reference for detecting the digital signal is not difficult. More subtle problems arise, however, when a number of digital transmission links and switches are interconnected to form a network. Not only must the individual elements of the network maintain internal synchronization, but also certain network wide synchronization procedures must be established before the individual subsystems can interoperate properly.

The need for some form of synchronization is not unique to digital networks. Single side band FDM transmission systems present similar requirements for carrier synchronization in analogue networks.

In analogue systems, however, the synchronization requirements are less critical by about two orders of magnitude.

1.3-3 Topologically Restricted Multiplexing

To the general public, the most apparent use of multiplexing is broadcast services for radio and television. In these systems the airspace is shared by using FDM of individual broadcast channels. With this system there are no operational restrictions to the geographic location of transmitters and receivers. As long as the transmitters confine their emissions to their assigned bandwidth and each receiver uses a sufficiently selective filter to pass only the desired channel, the network operates without mutual interference. On the other hand, TDM is not nearly as amenable to applications involving distributed sources and destinations. Since the time of arrival of data in a time slot is dependent on the distance of travel, distributed TDM systems require a guard time between time slots. FDM systems also require guard bands between the channels to achieve adequate channel separation. The width of the FDM guard bands, however, is not dependent on the geographic location of the transmitters. In a TDM system the guard times must be increased as the geographic separation between transmitters increases. Furthermore, each time division source must duplicate the synchronization and time slot recognition logic needed to operate a TDM system. For these reasons, TDM has been used primarily in applications (e.g., interoffice trunks) where all of the

information sources are centrally located and a single multiplexer controls the occurrence and assignment of time slots.
TDMA satellites and cellular systems are examples of applications of TDM for distributed sources. These systems use sophisticated synchronization techniques so that each ground station or mobile unit times its transmission to arrive at the satellite or base station at precisely defined times, allowing the use of small guard times between time slots. Notice that these applications involve only one destination; a satellite or a base station. If an application involves multiple, distributed sources and destinations (with transmission in more than one direction), larger guard times are unavoidable. Figure 1.7 shows such an application but uses FDM instead of TDM.

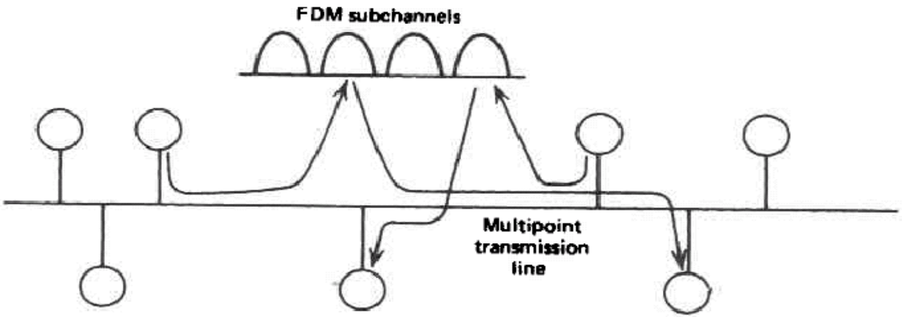


Figure 1.7 FDM on distributed multipoint line

The main engineering consideration for this system is to ensure that the FDM channels have sufficient isolation to allow a high-powered source to be adjacent to a receiver with the worst-case receive level. Obviously, adequate FDM isolation requires a certain amount of bandwidth overhead, but it is usually fairly easy to design filters with adequate isolation for a large range of signal levels so distance considerations are minimized.

1.3-4 Need for Conference/Extension Bridges

The process of combining multiple analogue signals to form a conference call or function as multiple extensions on a single telephone line can be accomplished by merely bridging the wire pairs together to superimpose all signals. Nowhere is this more convenient than when multiple extensions share a single two-wire line, as indicated in figure 1.8. When digitized voice signals are combined to form a conference either the signals must be converted to analogue so they can be combined on two-wire analogue bridges or the digital signals must be routed to a digital conference bridge, as shown in figure 1.9.

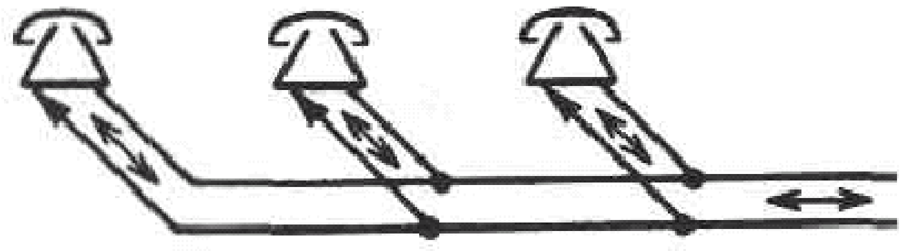


Figure 1.8 Analogue telephones connected to two-wire line

The digital bridge selectively adds the (four-wire) signals together (using DSP) and routes separate sums back to the conferees as shown. When conferencing is implemented in association with a switching system, the need for a digital conference bridge is not much of a disadvantage and in fact can significantly improve the quality of a conference by eliminating echoes and signal loss caused by power division

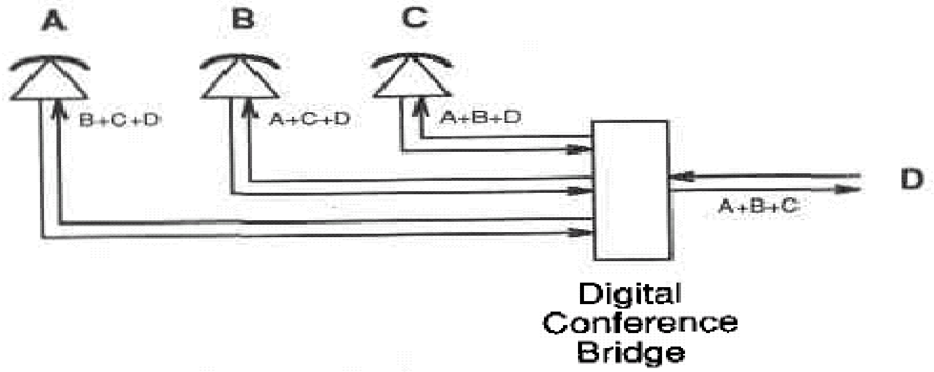


Figure 1.9 Use of conference bridge for digital telephones

However, when digital extensions need to have their signals combined so multiple extensions can be active in a conversation, the need for a centralized bridge can be an onerous problem. Residential telephone wiring typically follows a daisy-chain pattern, as indicated in figure 1.8. Thus the need to rewire all outlets and install a centralized conference box is a significant impediment to the deployment of digital station equipment in residential areas.

1.3-5 Incompatibilities with Analogue Facilities

When digital equipment was first used in private and public telephone networks, it necessarily provided standard analogue interfaces to the rest of the network. Sometimes these interfaces represented a major cost of the digital subsystem. The foremost example of this situation arose in digital end offices. The standard analogue subscriber loop interface is particularly incompatible with electronic switching machines (analogue or digital). Another aspect of digital switching that complicates its use in analogue environments is the artificial delay inserted by a typical digital matrix. One way to eliminate the problems with the analogue interface is to use digital subscriber loops and digital telephones. Unfortunately, the overwhelming investment in the loop plant for analogue telephones complicates a wide spread deployment of digital subscriber equipment. Most notable of the long-established practices that complicate a transition to digital loops are single wire pairs, loading coils, bridged taps, higher resistance (a bridged tap is an unused pair of wires connected at some point to an in-use pair as another extension or for possible future reassignment of a cable pair). Traditionally analogue communication system are rapidly being replaced with more modern digital modulation systems that offer several outstanding advantages over traditional analogue systems such as ease of processing, ease of multiplexing, and noise immunity.

Chapter

2

Digital Pulse Modulation

When an analog message is conveyed over an analog communication system, the full message is typically used at all times. To transmit analog message signals, such as speech signals or video signals, by digital means, the signal has to be converted into digital form. This process is known as analog-to-digital conversion. The sampling process is the first process performed in this conversion, and it converts a continuous - time signal into a discrete-time signal or a sequence of numbers. Digital transmission of analog signals is possible by virtue of the sampling theorem, and the sampling operation is performed in accordance with the sampling theorem.

In this chapter, using the Fourier transform technique, we present this remarkable sampling theorem and discuss the operation of sampling and practical aspects of sampling.

To send the same analog signal over a digital communication system it requires that only its samples be transmitted at periodic intervals. Because the receiver can, therefore, receive only samples of the message, it must attempt to reconstruct exactly the original message at all times from only its samples, even at times in between the samples. At first glance it may seem astonishing that only samples of a message and not the entire waveform can adequately describe all the information in the signal.

So, the aim of pulse modulation is to transfer a narrow band analog information, such as a phone call over a wideband low pass channel as a two-level signal. It involves modulating a carrier that is a train of regularly recurrent pulses using an analogue modulating wave. At the receiving end, the original waveforms may be reconstituted from the information regarding the samples, if these are taken frequently enough. Pulse modulation may be subdivided broadly into two categories; analogue and digital. In the former, pulse analogue modulation, a base band information signal may modulate or vary one of the carrier pulses parameters, while in the latter, a code which indicates the sample amplitude to the nearest predetermined level is sent.

In pulse analogue modulation, the amplitude of a base band information signal S(t) may modulate or vary; the pulse amplitude, to produce a Pulse Amplitude Modulation (PAM); the Pulse Width (duration), to give Pulse Width Modulation (PWM); the time delay between pulses in a sequence, to give Pulse Position Modulation (PPM). Pulse amplitude modulation may be an end in itself and can be used as modulation schemes in their own right for analogue communication systems allowing, for example, many separate information carrying signals to share a single physical channel by interleaving the individual

signal pulses. Such pulse interleaving is called Time Division Multiplexing (TDM).
In pulse digital modulation, the samples of the PAM signal may be quantized and coded to produce Pulse Code Modulation (PCM), Differential Pulse Code Modulation (DPCM), Delta Modulation (DM), or Adaptive Delta Modulation (ADM), …etc.

2.1 Sampling Principles

In signal processing, sampling is the reduction of a continuous (analogue) signal to a discrete signal to convert it to digital one.
A discrete signal is a signal that is not continuous in time. That is, it has values only at disconnected points of the time axis. A common example is the conversion of a sound wave (a continuous-time signal) to a sequence of samples (a discrete-time signal).
A sample refers to a value or set of values at a point in time and / or space. The sampling interval is the interval T_S where sampling is performed by measuring the value of the continuous signal every T_S seconds.

2.1-1 Sampling Theorem for low pass signals

The sampling theorem for low pass signals states that;
a band-limited signal S(t) of finite energy, having no spectral components above f_m Hz, can be completely represented in its samples and recovered back if the sampling frequency is twice of the highest frequency content of the signal. i.e.,

$$f_s \geq 2\, f_m$$

Figure 2.1 illustrates the Fourier transform S(f) of the band-limited signal S(t) which is zero for $f > f_m$, i.e.,

$$S(f) = 0 \qquad \text{for} \quad |f| > f_m \qquad (2.1)$$

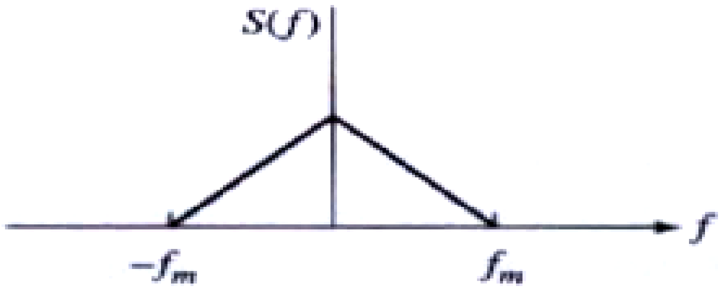


Figure 2.1 Representation of S(f)

A band-limited signal with maximum frequency f_m specified by equation (2.1) is often referred to as a low-pass signal.

A sampler which is sometimes referred to as a continuous-to-discrete (C / D) converter, is a subsystem that extracts samples from a continuous signal. A theoretical ideal sampler (a transistor working as a switch) as shown in figure 2.2 multiplies a continuous signal S(t) which is to be sampled with a Dirac comb (pulse train) P(t).

If the pulse train consists of narrow pulses, the output of the multiplier is a sampled version of the original waveform that depends on the sample values of the input.

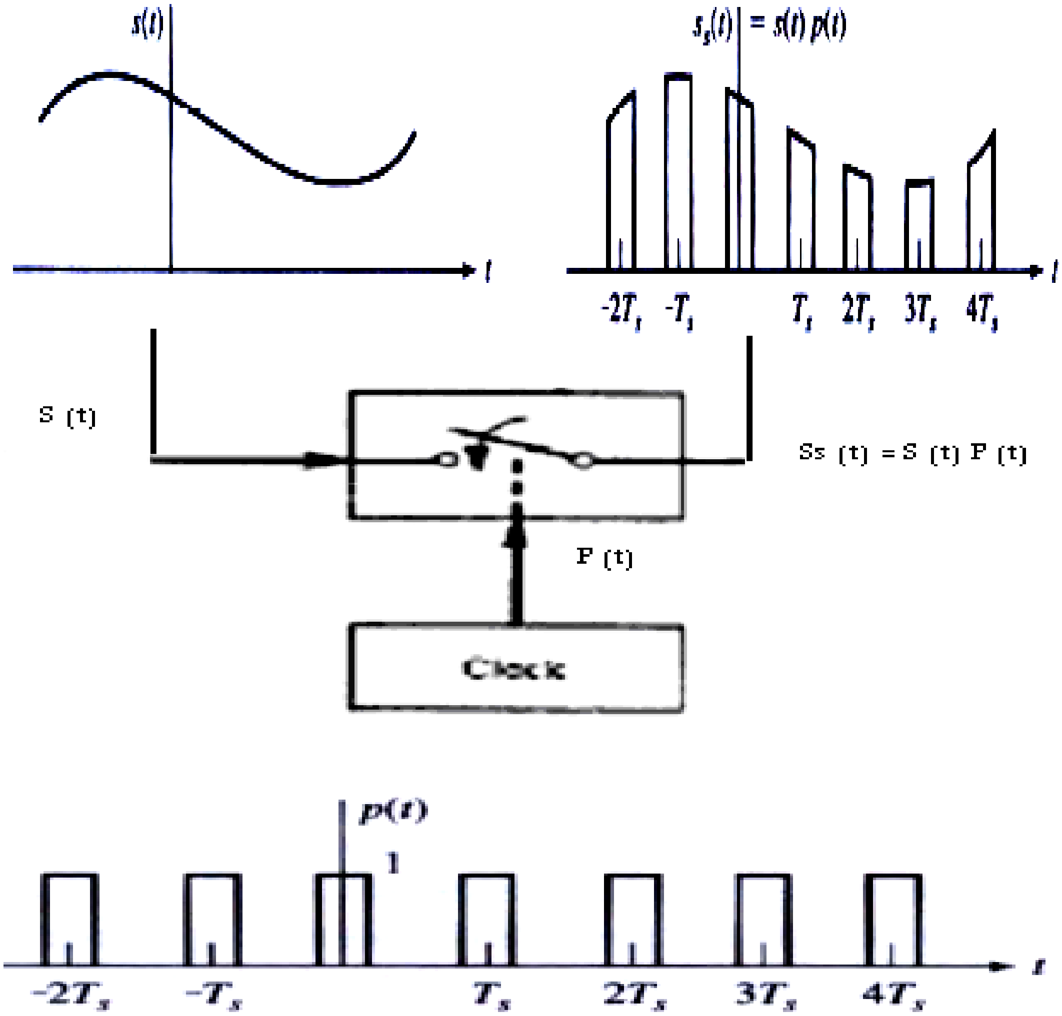


Figure 2.2 Sampling process

2.1-2 Nyquist criteria

A classical result in sampling systems was established in 1933 by Harry Nyquist when he derived the minimum sampling frequency required to extract all information in a continuous, time-varying waveform. This minimum sampling frequency must be greater than twice the highest frequency of the signal being sampled.

This result-the Nyquist criterion-is defined by the relation;

$$f_S \geq 2 f_m \ (= f_N) \qquad (2.2)$$

where, f_N is called the Nyquist frequency, of the sampling system.The samples must be "close enough" to each other to give all of the information. The restriction is that the spacing T_S between samples be less than $1 / 2 f_m$. The upper limit of T_S,

$1/(2 f_m)$ is known as the *Nyquist Sampling Interval*.

For example, a voice signal with maximum frequency of 4 kHz must be sampled at least 8000 times per second (which is the sampling rate used by nearly all telephony systems) to comply with the conditions of the sampling theorem. Note that the higher the frequency f_m, the faster the function varies, and the closer together the sample points should be in order to permit function reconstruction. Since the multiplying pulse train is assumed to be periodic, it can be expanded in a Fourier series. The P(t) shown in figure 2.2 is an even function, so using the trigonometric series;

$$S_S(t) = S(t)\ P(t)$$

$$S_S(t) = S(t)\left[a_o + \sum_{n=1}^{\infty} a_n \, Cos\, n\, \omega_s\, t\right]$$

$$= a_o\, S(t) + \sum_{n=1}^{\infty} a_n\, S(t)\, Cos\, n\, \omega_s\, t \qquad (2.3)$$

The goal is to isolate the first term in the final expression (2.3), which is proportional to the original base band signal S(t). Each of the terms in the summation of equation (2.3) is off the form of S(t) multiplied by a sinusoid. When a time signal is multiplied by a sinusoid, the result is a shift of all frequencies of the signal by an amount equal to the frequency of the sinusoid. The frequency content of each term in equation (2.3) is then centered around the frequency of the multiplying sinusoid (the carrier frequency). The Fourier transform of $S_S(t)$ is sketched in figure 2.3.

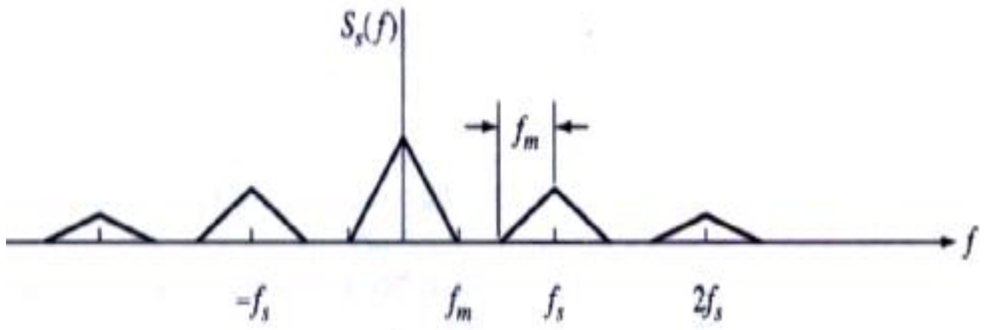


Figure 2.3 Fourier transform of natural sampled wave

The shape centered at the origin is the transform of $a_o S(t)$, and the shifted versions represent the transforms of the various harmonic terms are not overlapped, provided that $f_S > 2\ f_m$. These terms can be separated from each other using linear filters, where the original signal $a_o S(t)$ can be recovered from the sampled waveform S_S (t) by passing it through a low-pass filter with a cutoff frequency f_m.

2.1-3 Aliasing phenomena

The sampling theorem indicates that S(t) can be perfectly recovered from its samples when $f_s > 2 f_m$. If the sampling is attempted in the real world, error results if the sampling rate is not high enough.

If sampling occurs at too slow rate, $f_s < 2 f_m$ (under sampling condition), frequency components that did not exist in the original waveform will be produced in the desired frequency band. The lower side band of the sampling frequency f_S overlaps (appears within) the base band and thought to be part of it that is irretrievably corrupted, and the details of the original analogue waveform is lost, so it can't possibly be recovered intact. This overlap of the spectra is known as aliasing or fold over. Aliasing problems are not confined to speech digitization process. The potential for aliasing is present in any sample data system. Motion picture taking, for example, is another sampling system that can produce aliasing.

To avoid aliasing error as gracefully as possible in practice, the signal is sampled at a rate slightly higher than the Nyquist rate f_N.

If $f_s > 2f_m$, then as shown in figures 2.3, there is a gap between the upper limit f_m and the lower limit $f_s - f_m$ of $S_S(f)$. This range from f_m to $f_s - f_m$ is called a guard band. As an example, speech transmitted via telephone is generally band limited using a low-pass filter with a 3-dB cutoff around $f_m = 3.4$ kHz. The Nyquist rate is, thus, 6.8 kHz. For digital transmission, the speech is normally sampled at the rate $f_s = 8$ kHz. The guard band is then $f_s - 2f_m = 1.2$ kHz. The use of a sampling rate higher than the Nyquist rate also has the desirable effect of making it somewhat easier to design the low-pass reconstruction filter so as to recover the original signal from the sampled signal. Thus the sampled signal is sufficiently attenuated at the overlap frequency of 4 k Hz to adequately reduce the energy level of the fold over spectrum. Also, figure 2.4-a, illustrates a gap for filtering of at least 0.2 f_N = 0.4 f_m.

Actually, sampling at the Nyquist frequency makes the two bands just touch as shown in figure 2.4-b, so that whilst, theoretically, an ideal filter could recover the base band, no such filter exists, of course.

As shown in figure 2.4-c, and figure 2.5 for a base band signal sampled at only 1.6 f_m (0.8 f_N), there is aliasing error causes overlap area of 0.4 f_m. The spectrum in figure 2.5-a indicates that the lower side frequency of f_S - f_m is at 0.6 f_m (i.e., 1.6 f_m - f_m).

The waveform indicates how samples of f_m fit f_s - f_m. Figure 2.5-b, shows that the shorter wavelength sinusoid is the original f_m itself and the samples are positioned at sampling intervals of T_S that is related to the base band period, T_m, by;

$$f_s = 1 / T_S = 1.6 f_m = 1.6 / T_m$$

therefore, $$T_S = T_m / 1.6 = \frac{5}{8}\, T_m$$

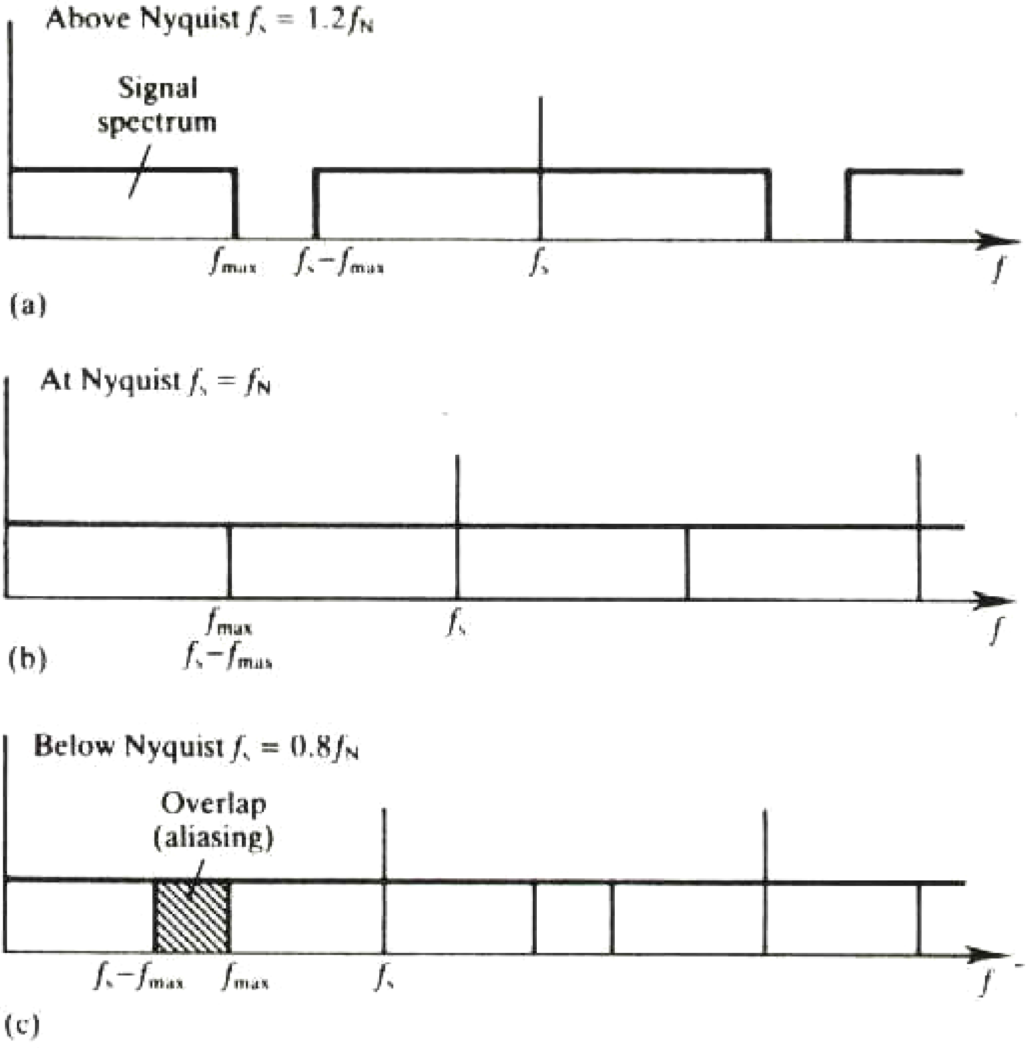


Figure 2.4 The Nyquist criterion

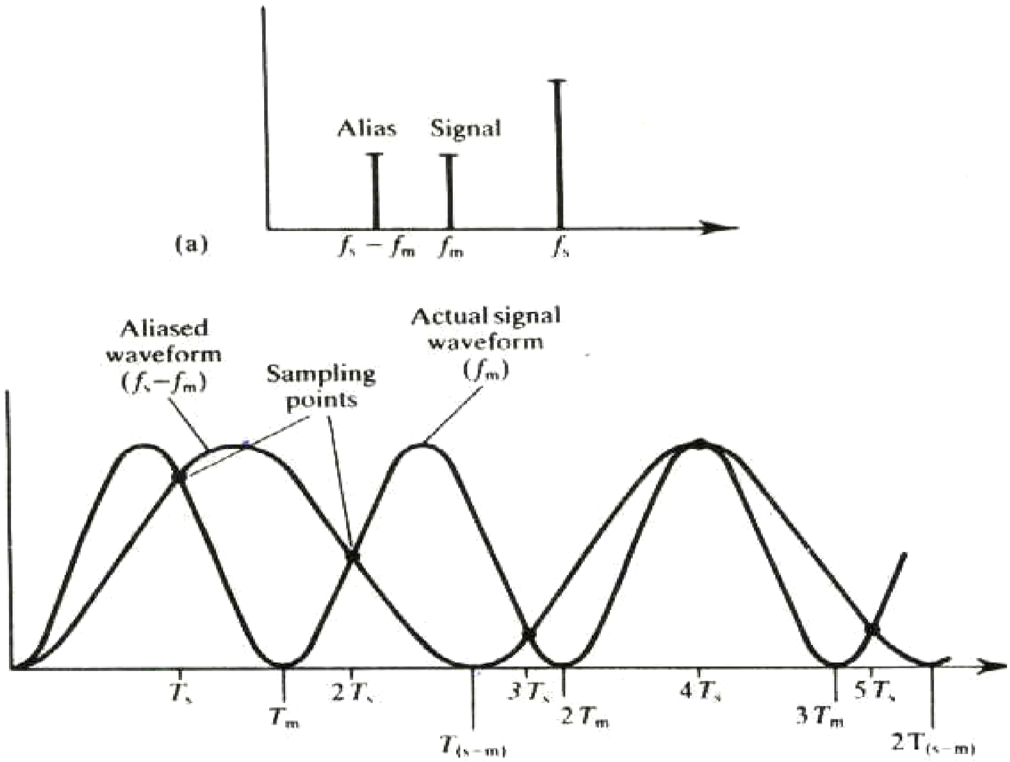


Figure 2.5 Waveform and spectra showing aliasing:

(a) spectrum of signal sampled below f_S (at f_S = 0.8 f_N);

(b) waveform showing how samples of f_m fit $f_S - f_m$ also

(T_S=5 / 8 T_m, $T_{(s - m)}$= 5 / 3 T_m)

A quantitative measure of the distortion introduced by aliasing can be defined as the ratio of un-aliased to aliased power in the reconstructed signal. For ideal reconstructed filter with rectangular amplitude response then, in the absence of an anti-aliasing filter, the signal to distortion ratio (SDR) is given by;

$$\text{SDR} = \frac{\int_0^{fs/2} S(f)\,df}{\int_{fs/2}^{\infty} S(f)\,df} \qquad (2.4)$$

where, S (f) is the two sided power spectral density of the base band information signal S (t). More generally, using a filter with frequency response H (f) for reconstruction then the integral limits are extended to give;

$$\mathrm{SDR} \approx \frac{\int_0^\infty S(f)\left|H(f)\right|^2 df}{\int_0^\infty S(f-f_s)\left|H(f)\right|^2 df} \qquad (2.5)$$

An approximation sign is used as the spectral replicas centered on 2 f_S Hz and above are assumed to be totally suppressed by |H (f)|.

So, unless the Nyquist criterion is met, it is impossible to recover the original base band signal S(t) from the sampled signal $S_S(t)$ by filtering.

Figure 2.6, illustrates another simple example of aliasing in the time domain. A sinusoid at a frequency of 3 Hz is shown. Suppose we sample this sinusoid at four samples per second. The sampling theorem tells us that the minimum sampling rate for unique recovery is six samples per second, so four samples per second is not fast enough. The samples at the slower rate are indicated in the figure. But alas, these are the same samples that would result from a sinusoid at 1 Hz, as shown by the dashed curve. The 3 Hz signal is thus disguising itself (aliasing) as a 1 Hz signal, or a 1 Hz sinusoid is folded back to fall in the 3 Hz sinusoid band.

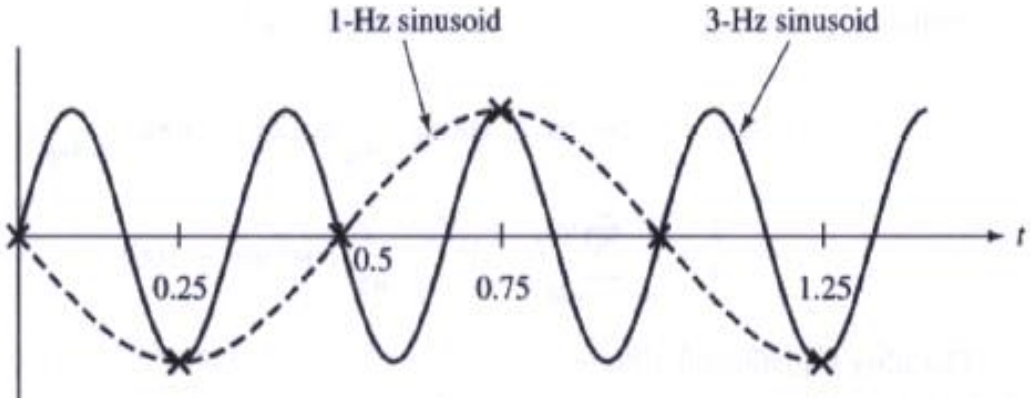


Figure 2.6 Example of aliasing

2.1-4 Sampling Theorem in Frequency Domain

The sampling theorem in time-domain stated that, if the band limited signal is sampled at the rate of $f_S > 2\ f_m$ in time domain, then it can be fully recovered from its samples.

A dual to this time-domain sampling theorem, is the sampling theorem in the frequency domain.

It states that, a time-limited signal S(t) which is zero for $|t| > |T_S|$, specified by;

$$S(t) = 0 \qquad \text{for } |t| > |T_S| \tag{2.6}$$

can be represented and recovered fully from its samples of its frequency spectrum, if the samples are taken at intervals less than $1/2T_S$ Hz apart as shown in figure 2.7.

Then the Fourier transform S(ω) is given by;

$$S(\omega) = \sum_{n=-\infty}^{\infty} S(n\omega_S)\ \frac{Sin\, T_o\left(\omega - n\omega_S\right)}{T_o\,(\omega - n\omega_S)} \tag{2.7}$$

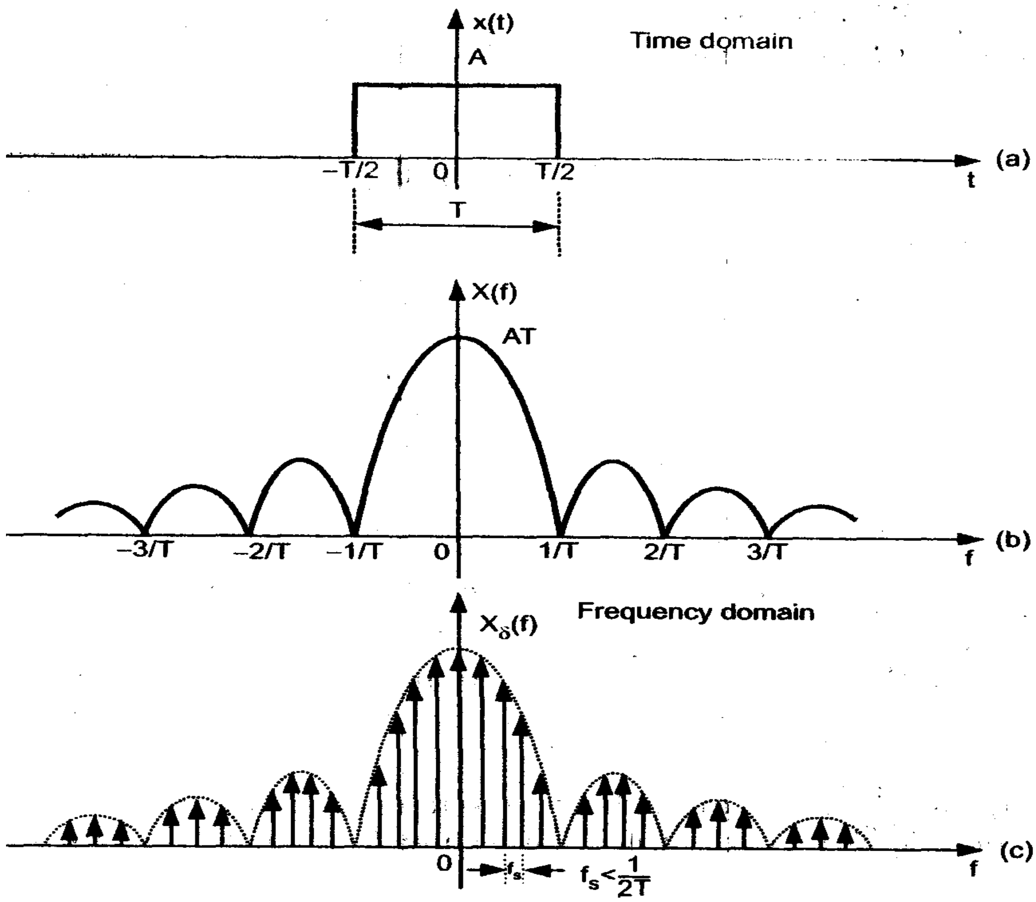


Figure 2.7 Sampling theorem in frequency domain
a- S(t) time limited to ± T/2,
b- Continuous spectrum of S(t),
c- Sampled spectrum $S_\delta(f)$

2.1-5 Sampling Theorem for Band pass signals

The Sampling Theorem for Band pass signals can be written as follows;

The band pass signal S(t) whose maximum bandwidth is 2 f_m can be completely represented into and recovered from its samples if it is sampled at the minimum rate of twice the bandwidth.

Figure 2.8 shows the spectrum of band pass signal with BW = 2 f_m,

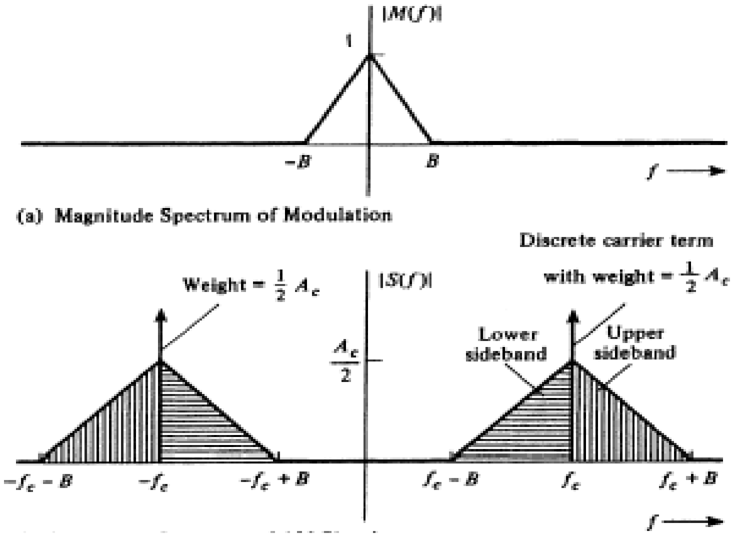


Figure 2.8 Spectrum of band pass signal with bandwidth 2 f_m,

then the minimum sampling rate of the band pass signal should be 4 f_m sample per second.

When the signal is a discrete base band signal, it can be thought of as a list of numbers representing the sample values of an analogue waveform. One way to send such a list through a channel is to send a pulse waveform-one pulse is placed at each sampling point. Each pulse carries information about the corresponding sample values. Each sample value can be conveyed as the amplitude, (PAM), width, (PWM), or position, (PPM).

2.2 Pulse Amplitude Modulation, PAM

Pulse-amplitude modulation, PAM is the simplest and most basic form of analog pulse modulation, where sampling process is equivalent to amplitude modulation of a constant-amplitude pulse train. It is a form of signal modulation where the message information is encoded in the amplitude of a series of signal pulses. Hence the technique represented in figure 2.9 is usually referred to as a PAM, because the successive output intervals can be described as a sequence of pulses with amplitudes change according to amplitude of the modulating signal at the sampling instants.

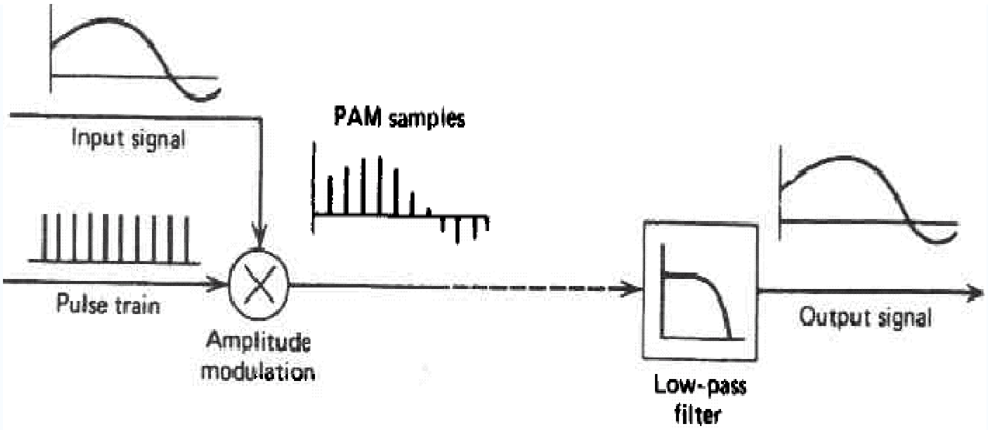


Figure 2.9 Basic concepts of Pulse Amplitude Modulation

Figure 2.10, portrays the spectrum of the input signal and the resulting spectrum of the PAM pulse train.

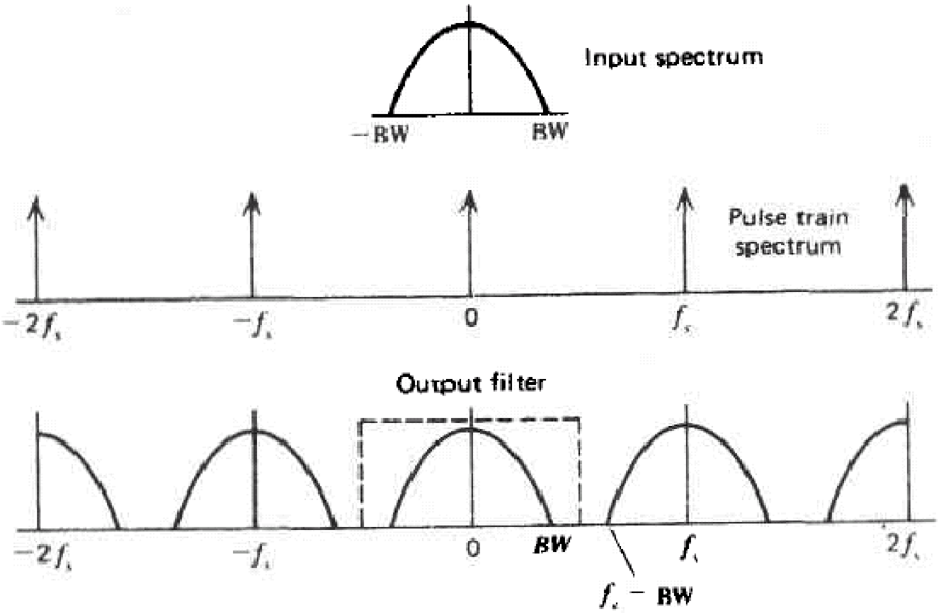


Figure 2.10 Spectrum of PAM signal

The PAM spectrum can be derived by observing that a continuous train of impulses has a frequency spectrum consisting of discrete terms at multiples of the sampling frequency. The input signal amplitude modulates these terms individually.

Thus a double-sideband spectrum is produced about each of the discrete frequency terms in the spectrum of the pulse train. The original signal waveform is recovered by a low-pass filter designed to remove all the higher frequency AM components but the original signal spectrum. As shown in figure 2.10, the reconstructive low-pass filter must have a cutoff frequency that lies between f_m and $f_S - f_m$. Hence, separation is only possible if $f_S - f_m$ is greater than f_m (i.e., if $f_S \geq 2.2\ f_m$). So, practical filters with gradual roll-off in the filter stop band characteristics can be used, otherwise fold over distortion occur.

If the original signal has frequency components above one-half of the sampling rate, this causes that the base band signal will be under-sampled and the base band spectrum of S(t) can't be recovered exactly even with an ideal rectangular low pass filter.

Also most input analog signals are filtered (band limited) with an anti-aliasing filter (usually a low-pass filter with a cutoff frequency near the frequency $f_S/2$) immediately before the sampling circuit as shown in figure 2.11.

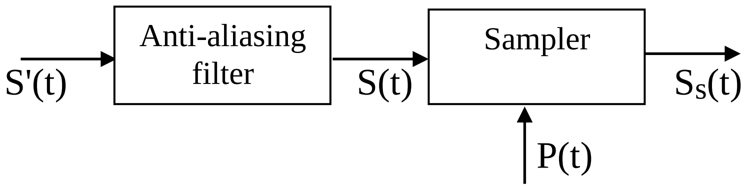


Figure 2.11 Using anti-aliasing filter

Whilst this filter may remove high frequency energy from the information signal the resulting distortion is generally less than that introduced if the same energy is aliased to incorrect frequencies by the sampling process.

Depending upon the shape of the pulse of PAM, there are mainly two types of PAM;

i- Ideally, or instantaneously sampled PAM,

ii- Practically sampled PAM, which divided into two types;

a- Naturally sampled PAM,

b- Flat-top sampled PAM,

2.2-1 Ideal, Instantaneous, Impulse Sampled PAM

Ideal sampler shown in figure 2.12 is used to sample an analog signal S(t) shown in figure 2.13-a, instantaneously at a uniform rate, once every T_s sec.

The result of this sampling process, is an infinite sequence of samples {S(nTs)}, where n takes on all possible integers.

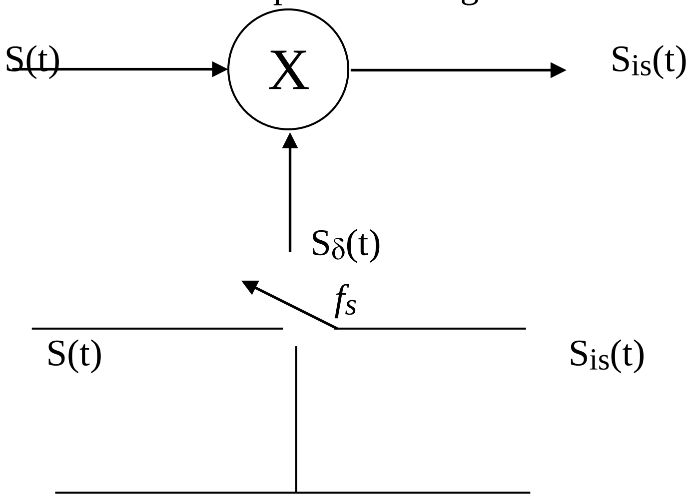


Figure 2.12 Mathematical model and functional diagram of the instantaneous sampler

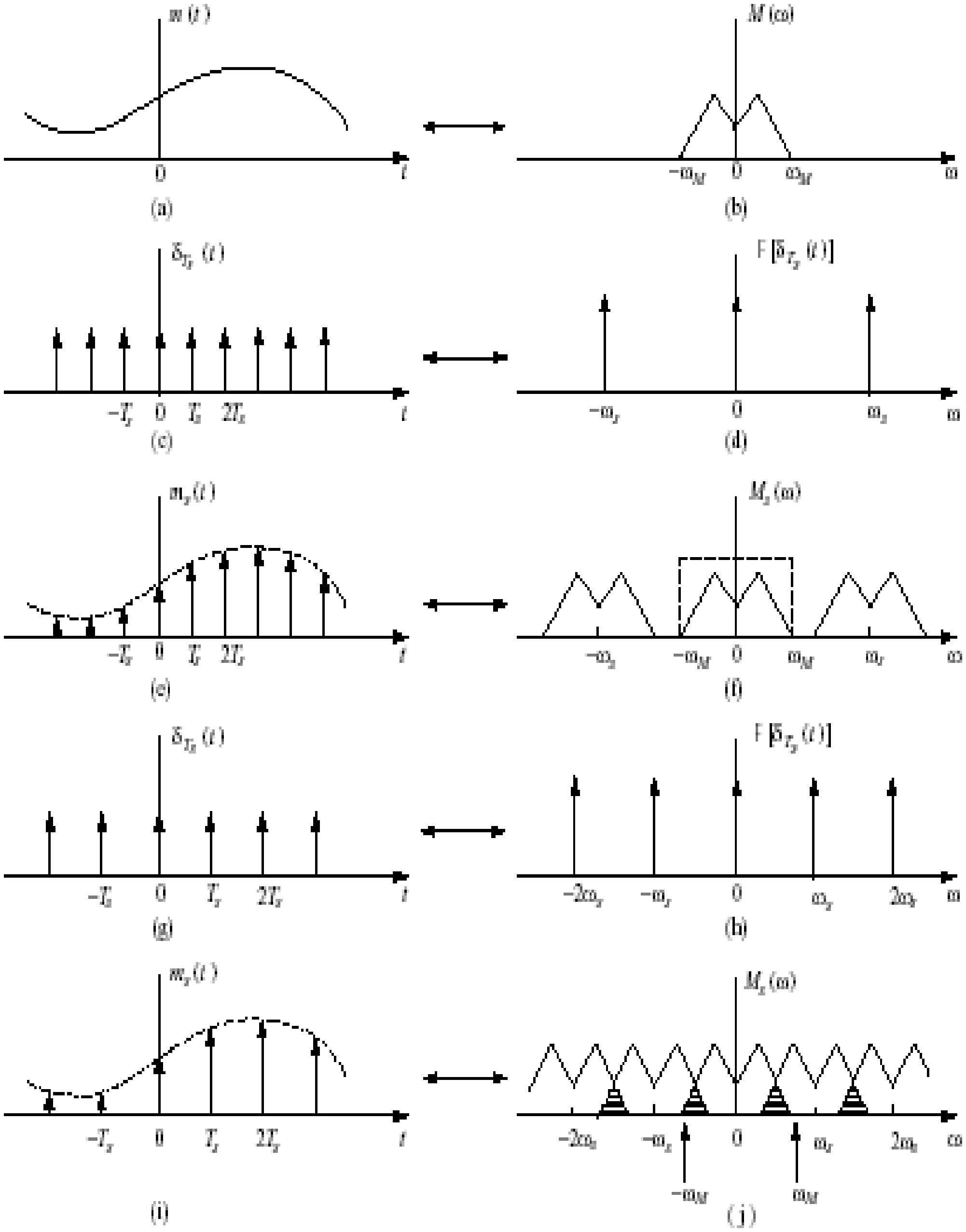


Figure 2.13 Instantaneous sampling

Figure 2.13-e illustrates that $S_{is}(t)$ be obtained by multiplication of S(t) by a periodic train of unit impulse functions $S_{\delta}(t)$ with period T_S, that is;

$$S_{is}(t) = S(t)\ S_\delta(t) = S(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s)$$

$$= \sum_{n=-\infty}^{\infty} S(t)\ \delta(t - nT_s) = \sum_{n=-\infty}^{\infty} S(nT_s)\ \delta(t - nT_s) \qquad (2.8)$$

Because the impulse train $S_\delta(t)$ is a periodic signal of period T_S, it can be expressed as a Fourier series. The trigonometric Fourier series is;

$$S_\delta(t) = \frac{1}{T_s}\ [1 + 2\cos\omega_S + 2\cos 2\omega_S + 2\cos 3\omega_S + \ldots\ldots.]$$

Therefore,

$$S_{is}(t) = S(t)\ S_\delta(t)$$

$$= \frac{1}{Ts}\ [s(t) + 2s(t)\cos\omega_S + 2s(t)\cos 2\omega_S + 2s(t)\cos 3\omega_S + ..] \qquad (2.9)$$

The transform of the sampled waveform can be written as;

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

$$S_{is}(f) = \frac{1}{T_s} \sum_{n=-\infty}^{+\infty} S\ (f - nf_s) \qquad (2.10)$$

The signal $S_{is}(t)$ shown in figure 2.13-e is referred to as the ideal sampled signal.

It is clear that the spectrum of the sampled waveform $S_{is}(f)$ consists of an infinite number of scaled by $1/T_S$ and shifted copies (periodically repeated every f_S) of the spectrum of the original signal S(t).

The original continuous signal S(t) can be completely recovered (demodulated) from sampled signal $S_{is}(t)$ using a LPF with cut off frequency f_m for $f_s > 2 f_m$ as shown in figure 2.13-f.

The sampling process considered so far is known as *ideal sampling* because it involves ideal impulse functions (assuming pulses with duration τ = zero). Obviously, ideal sampling is not practical. It is theoretically only because it is not possible to have a pulse whose width approaches zero.

Practical Sampling of an analog signal is performed by means of high-speed switching circuits, and the sampling process takes the form of **natural sampling** or **flat-top sampling.**

2.2-2 Naturally Sampled PAM

It is performed as the product of two signals; S(t) that forms the input to a gating (sampler) circuit, and the gating control P(t) as shown in figure 2.14. The tops of the pulses of the resulted natural sampled signal $S_{ns}(t)$ follow the variations of the signal being sampled.

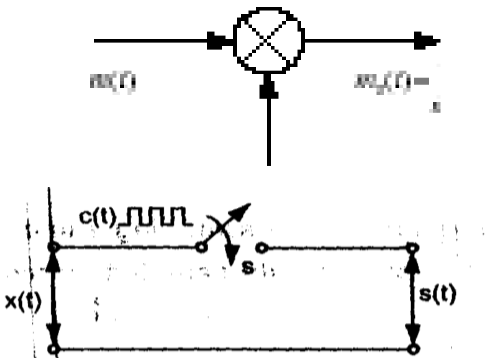


Figure 2.14 Mathematical model and functional diagram of

the natural sampler

Natural sampling of a band-limited signal S(t) is shown in figure 2.15. The sampled signal $S_{ns}(t)$ can be expressed as;

$$S_{ns}(t) = S(t) \times P(t) \qquad (2.11)$$

where P(t) is the periodic train of rectangular pulses with fundamental period T_S, and each rectangular pulse in P(t) has duration τ and unit amplitude.

Recall that the periodic pulse train P(t) can be expressed in a Fourier series as follows:

$$P(t) = \sum_{n=-\infty}^{\infty} C_n e^{j\left(2\pi n f_s t\right)}, \qquad (2.12)$$

where C_n is the Fourier coefficient, given by;

$$C_n = \frac{\tau}{T_s} Sinc\left(\frac{n\tau}{T_s}\right) e^{-jn\pi\tau/T_s} \qquad (2.13)$$

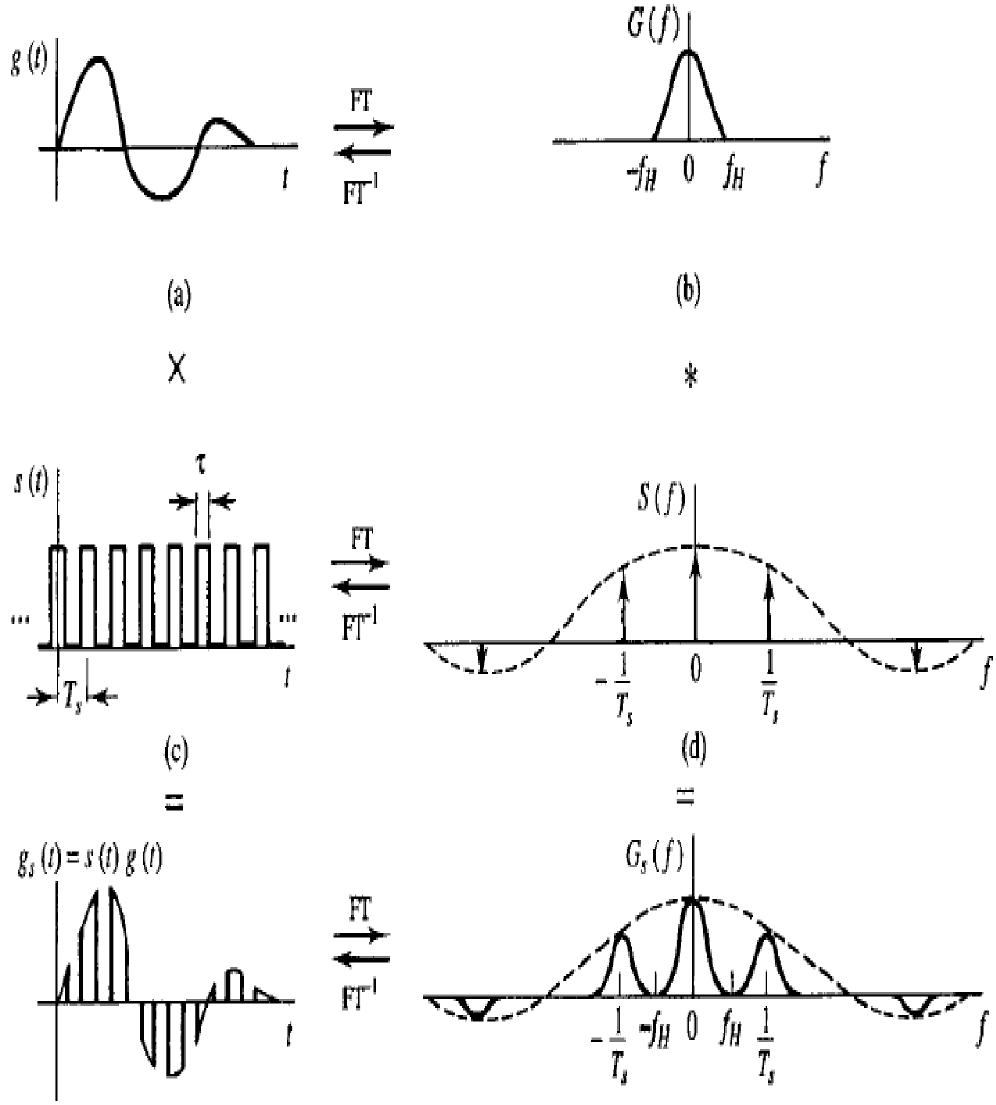


Figure 2.15 Time and frequency domain illustration of natural sampling PAM signal: a- signal S (t); b- signal spectrum; c- sampling function; d- spectrum of sampling function; e- sampled signal; f- spectrum of sampled PAM signal.

The Fourier transform of P(t) is;

$$P(\omega) = \sum_{n=-\infty}^{\infty} C_n \, \delta(\omega - n\omega_s) \qquad \omega_s = 2\pi / T_s \qquad (2.14)$$

Thus the sampled waveform, $s_{ns}(t)$ is;

$S_{ns}(t) = s(t) \sum_{n=-\infty}^{\infty} C_n e^{j(2\pi n f_s t)}$ (2.15)

Then the Fourier transform of the naturally sampled wave $S_{ns}(t)$ is found by periodically shifting and repeating the Fourier transform of the original signal. It is given by;

$$S_{ns}(f) = F[S_{ns}(t)] = \sum_{n=-\infty}^{\infty} C_n F\left[s(t) e^{j(2\pi n f_s t)}\right]$$

$$= \sum_{n=-\infty}^{\infty} C_n S(f - n f_s) \qquad (2.16)$$

It is clear that the effect of the natural sampling is to multiply the n^{th} shifted spectrum $S(f - nf_s)$ by a constant C_n.

Similar to the ideal sampling case, equation (2.16) shows that the natural sampled signal $S_{ns}(f)$ consists of an infinite number of copies of $S(f)$, which are periodically shifted in frequency every f_s hertz. However, here the copies of $S(f)$ are not uniformly weighted (scaled) as in the ideal sampling case, but rather they are weighted by the Fourier series coefficients of the pulse train. The spectrum of the sampled waveform $S_{ns}(t)$ is shown in figure 2.15-e, where S(t) is again a band limited waveform. It can be seen from figure 2.15-f that, despite the above difference, the original signal S(t) can be equally well reconstructed from $S_{ns}(t)$ using a LPF as long as the Nyquist criterion is satisfied. Finally, it should be noted that natural sampling can be considered to be a practical approximation of ideal sampling, where an ideal impulse is approximated by a narrow rectangular pulse. With this perspective, it is not surprising that when the width τ of the pulse train approaches zero, the spectrum in equation (2.16) converges to equation (2.10).

To obtain the frequency components of a PAM waveform in order to determine the channel requirements, assume the base band analog signal is level shifted so that no part of it is negative to make all samples positive, so;

$$S(t) = E_m (1 + \text{Cos } \omega_m t) \qquad (2.17)$$

and the sampling function has constant amplitude spectrum with pulse duration τ, and sampling interval T_S such that;

$\tau << T_S$, so;

$$P(t)=\frac{\tau}{T_S}(1+2 \text{ Cos } \omega_S t +2 \text{ Cos } 2\omega_S t +2 \text{ Cos } 3\omega_S t +...) \qquad (2.18)$$

The sampled function will be;

$$V_{PAM} (t) = E_m (1 + \text{Cos } \omega_m t)\frac{\tau}{T_S} (1 + 2 \text{ Cos } \omega_S t$$

$$+ 2 \text{ Cos } 2\omega_S t + 2 \text{ Cos } 3\omega_S t + \text{........})$$

$$V_{PAM}(t) = \frac{E_m \tau}{T_S}(1+\text{Cos } \omega_m t+2\text{Cos } \omega_S t+2\text{Cos } \omega_S t \text{ Cos } \omega_m t$$

$$+ 2 \text{ Cos } 2\omega_S t + 2 \text{ Cos } 2\omega_S t \text{ Cos } \omega_m t + \text{........})$$

$$V_{PAM}(t)= \frac{E_m \tau}{T_S}[1 + \text{Cos } \omega_m t + \text{Cos } (\omega_S - \omega_m)t + 2 \text{ Cos } \omega_S t$$

$$+ \text{Cos } (\omega_S + \omega_m) t + \text{Cos } (2\omega_S - \omega_m) t + 2 \text{ Cos } 2\omega_S t$$

$$+ \text{Cos } (2\omega_S + \omega_m) t + \ldots\ldots..] \qquad (2.19)$$

which is a dc level plus the original base band plus a set of 100 % amplitude modulation on the sampling frequency and its harmonics. The spectrum of the sampled signal is shown in figure 2.16.

2.2-3 Flat-Top Sampled PAM

Flat-top sampling is the most popular sampling method. This sampling process involves two simple operations:

i- Instantaneous sampling of the analog signal S(t) every T_S seconds.
ii- Maintaining the value of each sample for a duration of τ seconds.

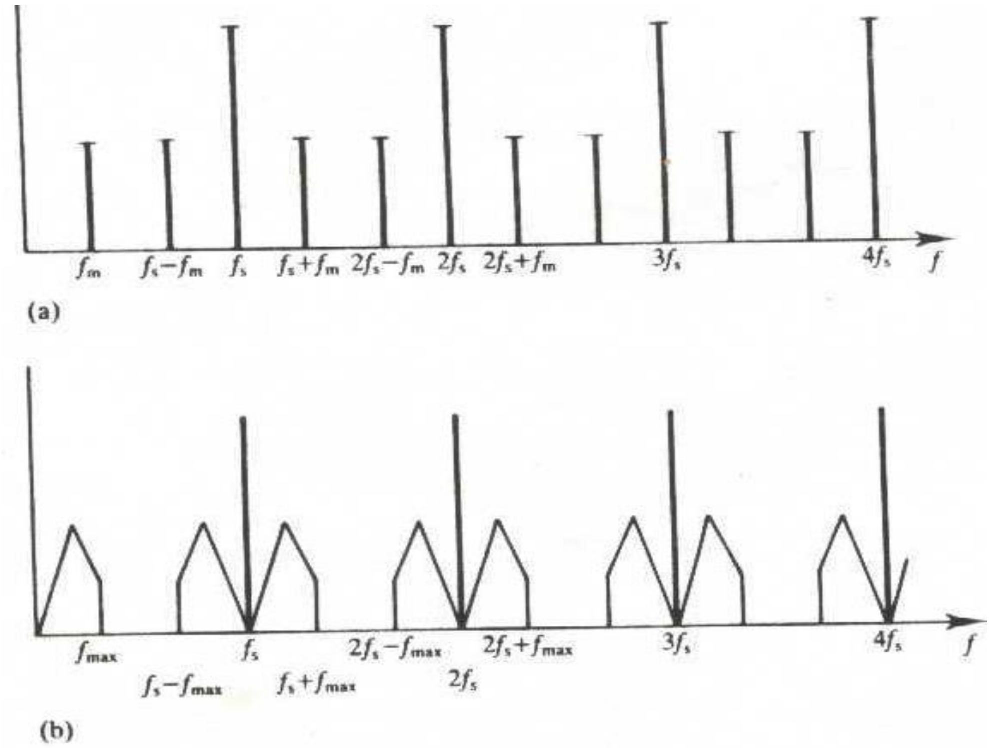


Figure 2.16 Spectra of sampled signals:

a- single frequency base band (f_S = 3.2 f_m = 1.6 f_N);

b- band of signal frequencies ((0 - f_{max}) – f_S = 3.2 f_{max}= 1.6 f_N)

The mathematical model of these two operations are indicated in figure 2.17-a. In circuit technology, these two operations are referred to as sample and hold as illustrated in figure 2.17-b that is used to generate a flat-top sampled PAM waveform shown in figure 2.18-c.

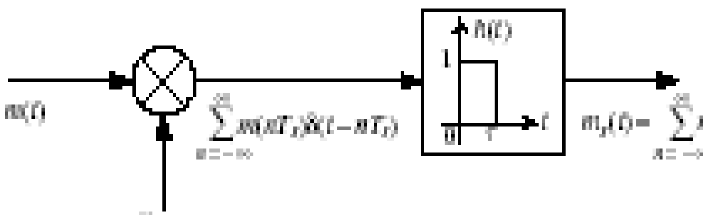


Figure 2.17-a Mathematical model of flat-top sampler

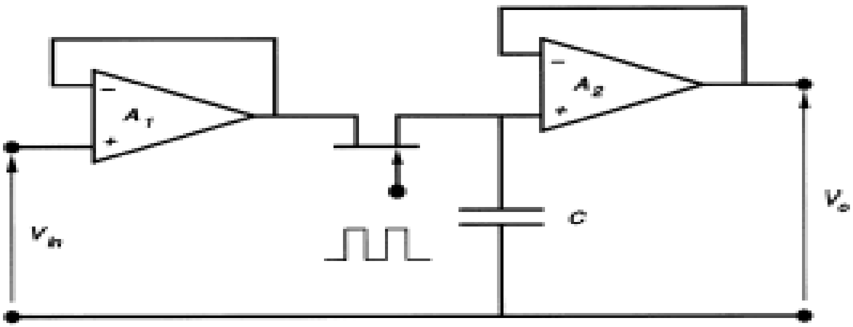


Figure 2.17-b A practical sample-and-hold circuit

Figure 2.17-b illustrates that, the operational amplifiers are connected as voltage followers with a gain of unity.

The output impedance of A_1 is low enough for it to drive the required charge into the capacitor. The N-channel JFET is switched on by a pulse applied to the gate and the capacitor charges up to the value of the input voltage. When the JFET switch is turned off, the high input impedance of A_2 drains minimal current from C.

Figure 2.18 illustrates a periodic pulse train P(t), a portion of a typical analog signal S(t), and the result S_{fs} (t) of controlling the pulse heights with the sample values.

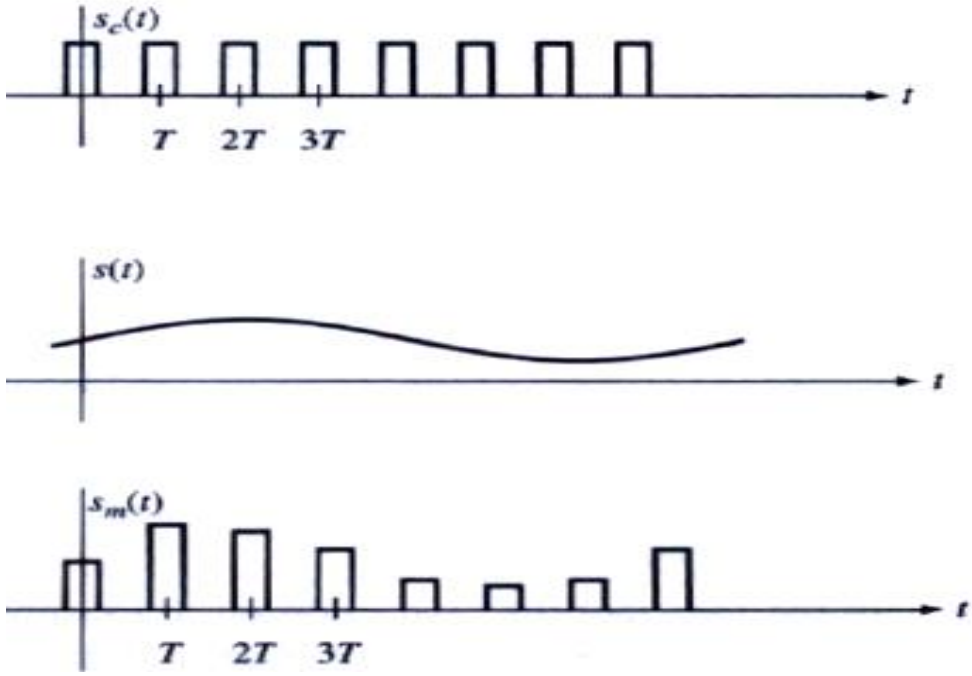


Figure 2.18 Flat top Pulse Amplitude Modulation

Note that since the pulse tops are horizontal, the modulated waveform is not simply the product of the pulse train P(t) and the analog signal S(t). It is called flat-top sampled PAM, as the pulse height depends only upon the value of S(t) at the sampling point, and not on the signal values across the range of the pulse width.

Note that in figure 2.17-a, a filter with impulse response h(t) is added at the end to change impulse samples to pulse samples.

Again, we are interested in the spectrum of the flat-top sampled signal $S_{fs}(t)$, which is related to the original signal S(t) through the following expressions:

Using the ideal sampled signal $S_{is}(t)$ given by equation (2.8) $S_{fs}(t)$ can be expressed as;

$S_{fs}(t) = S_{is}(t) * h(t)$

$$= \left[S(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \right] * h(t)$$

$$= \left[\sum_{n=-\infty}^{\infty} S(nT_s)\, \delta(t - nT_s) \right] * h(t) \qquad (2.20)$$

Using the convolution property of the Fourier transform, the Fourier transform of $S_{fs}(t)$ is determined as follows;

$$S_{fs}(f) = F\left[S(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \right] F[h(t)]$$

$$= \frac{1}{T_s} H(f) \sum_{n=-\infty}^{\infty} S(f - nf_s) \qquad (2.21)$$

where H(f) is the Fourier transform of the rectangular pulse h(t), given by $H(f) = \tau \,\mathrm{sinc}\,(f\tau)\, e^{-j\pi f\tau}$.

Equations (2.20), and (2.21), imply that the spectrum of the signal produced by flat-top sampling is essentially the spectrum of the signal produced by ideal sampling shaped by H(f). Since H(f) has the form of sinc function, each spectral component of the ideal sampled signal is weighted differently, hence causing amplitude distortion. The primary effect is an attenuation of high-frequency components. This effect is known as the *aperture effect.*

As a consequence of this distortion caused by H(f), it is not possible to reconstruct the original signal S(t) using a LPF, even when the Nyquist criteria is satisfied.

Demodulation of flat-top sampled PAM that is performed by detecting the amplitude level of the carrier at every symbol period, T_S requires a little more work. The aperture effect can be compensated by connecting an equalizing filter with a frequency response $H_{eq}(f) = 1/H(f)$ in cascaded with the low pass reconstructed filter, as shown in figure 2.19.

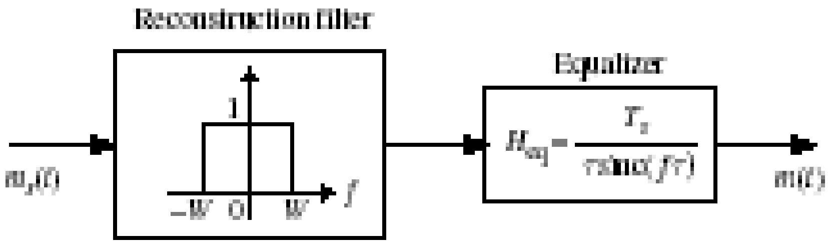


Figure 2.19 Demodulation of flat-top PAM

Finally, it should be noted that for a duty cycle $\tau/T_s \leq 0.1$, the amplitude distortion due to H(f) is less than 0.5%. In this case, equalization may not be necessary in practical applications.

We could use a sample and hold circuit to recover a staircase approximation of the original waveform. The holding time is set equal to the sampling period. The result is shown in figure 2.20 for a representative S(t).

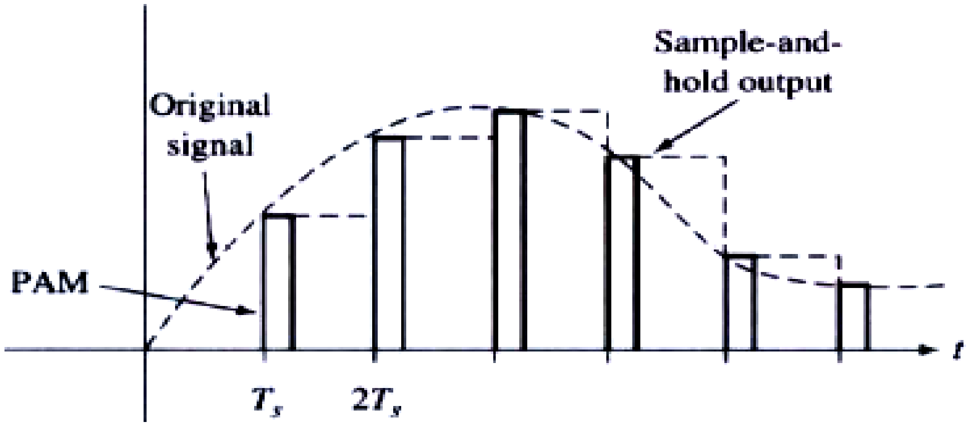


Figure 2.20 Sample and hold for PAM demodulation

The resulting staircase function can be low pass filtered to get a smooth approximation to the original waveform.

The design of the S&H circuit is best illustrated by the following example.

Example 2.1, The S&H circuit shown in figure 2.17-b uses a JFET as a series sampling gate. The voltage follower A_2 takes a current of 500 nA. The JFET has a source-to-drain resistance of 25 Ω when it is in the on state and may be considered to be an open-circuit when it is in the off state. The signal amplitude is 2.0V, the sample-time is 5 μs and the hold time is 500 μs. The capacitor has a value 0.4 μF. Calculate the error in the output at the end of the hold time, assuming the signal source has a negligible resistance.

Solution, When the sampling gate is closed, assume that the current flowing into the capacitor i >> 550 nA (the leakage current taken by A_2):

$\tau = C\,R_S = 0.4 \times 10^{-6} \times 25 = 10\ \mu\text{s}$

For an RC circuit,

$v_C = V_S\,(1 - e^{-t/\tau})$

When $t = 5\ \mu s$, and $V_S = 2.0$ V, $v_c = 1.987$ V.

Charge on C at $t_1 = 5\ \mu s$,

$Q = C \times v_C = 0.4 \times 10^{-6} \times 1.987 = 0.795\ \mu C$

Charge lost by C in time $t_2 = 500\ \mu s$,

$\Delta Q = I \times \Delta t = 500 \times 10^{-9} \times 500 \times 10^{-6} = 25 \times 10^{-11}$ C

Charge on C at time t_2,

$Q - \Delta Q = (0.795 \times 10^{-6} - 25 \times 10^{-11}) = 0.794 \times 10^{-6}$ C,

Voltage across C at time t_2,

$$v_{C2} = \frac{Q - \Delta Q}{c} = 1.986\ V$$

Percentage error in the output is 1.4%.

Figure 2.21 demonstrates a fold over process occurring in speech if a 5.5 kHz signal is sampled at an 8 kHz rate to produced PAM signal. Notice that the sample values are identical to those obtained from a 2.5 kHz input signal. Thus after the sampled signal passes through the 4 kHz output filter, a 2.5 kHz signal arises that did not come from the source. This example illustrates that the input must be band limited, before sampling, to remove frequency terms greater than $f_S/2$, even if these frequency terms are ignored (i.e., are inaudible) at the destination.

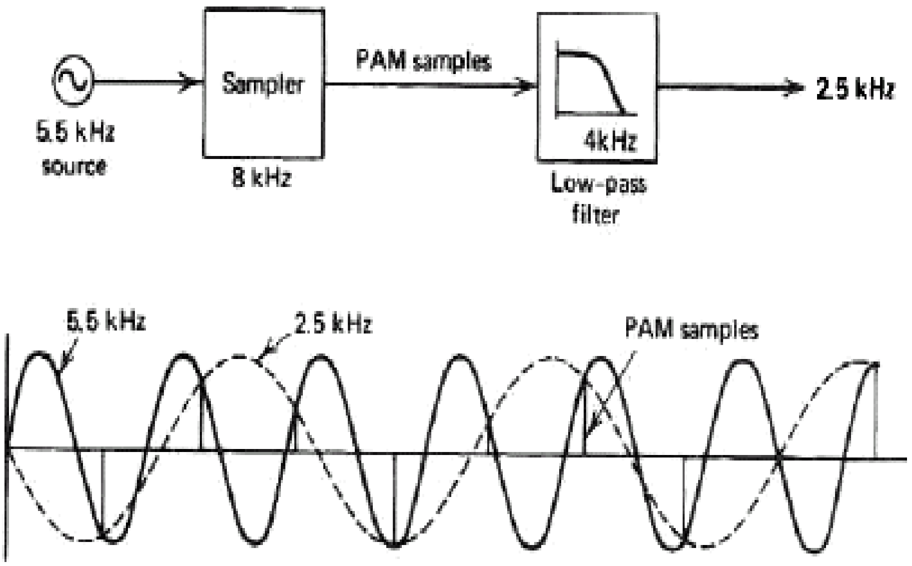


Figure 2.21 Fold over of 5.5-kHz signal into a 2.5-kHz signal

Thus, a complete PAM system, as given in figure 2.22, must include a band limiting filter before sampling to ensure that no spurious or source-related signals get folded back into the desired signal bandwidth. The input filter of a voice codec may also be designed to cut off very low frequencies to remove 60-cycle hum from power lines.

Figure 2.22 shows the signal being recovered by a sample-and-hold circuit that produces a staircase approximation to the sampled waveform.

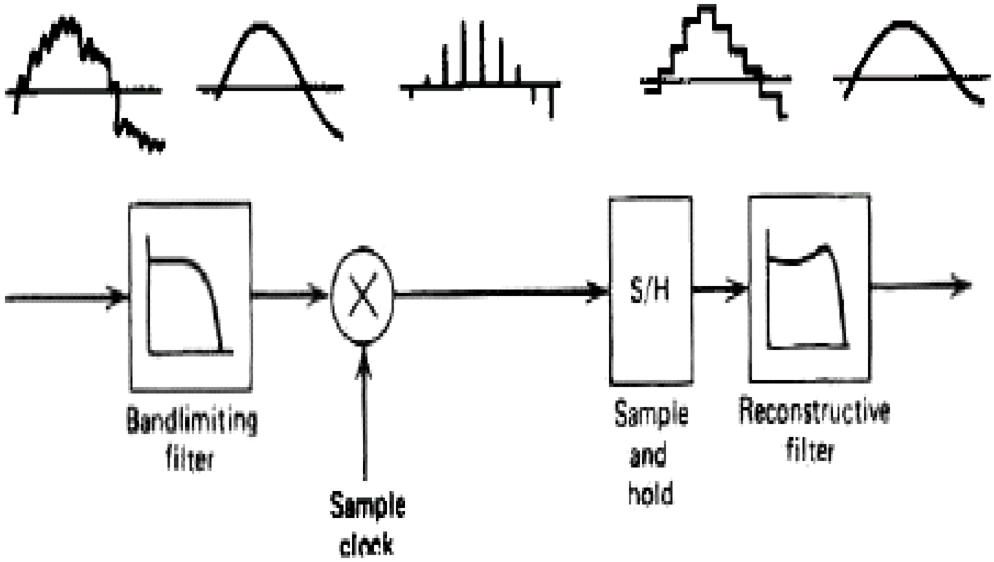


Figure 2.22 End-to-end PAM system

By interleaving the samples from multiple sources, PAM systems can be used to share a transmission facility in a TDM manner which is the main (perhaps the only advantage of PAM). This is because a PAM signal only occurs in slots of time, leaving the idle time for the transmission of other PAM signals. However, this advantage comes at the expense of a larger transmission bandwidth. PAM systems are not generally useful over long distances owing to the vulnerability of the individual pulses to noise, distortion, inter-symbol interference, and crosstalk. Instead, for long-distance transmission the PAM samples are converted into a digital format, thereby allowing the use of regenerative repeaters to remove transmission imperfections before errors result. Pulse-amplitude modulation is now rarely used, having been largely superseded by Pulse Code Modulation, and, more recently, by pulse-position modulation.

In particular, all telephone modems faster than 300 bit/s use quadrature amplitude modulation (QAM) that uses a two-dimensional constellation.

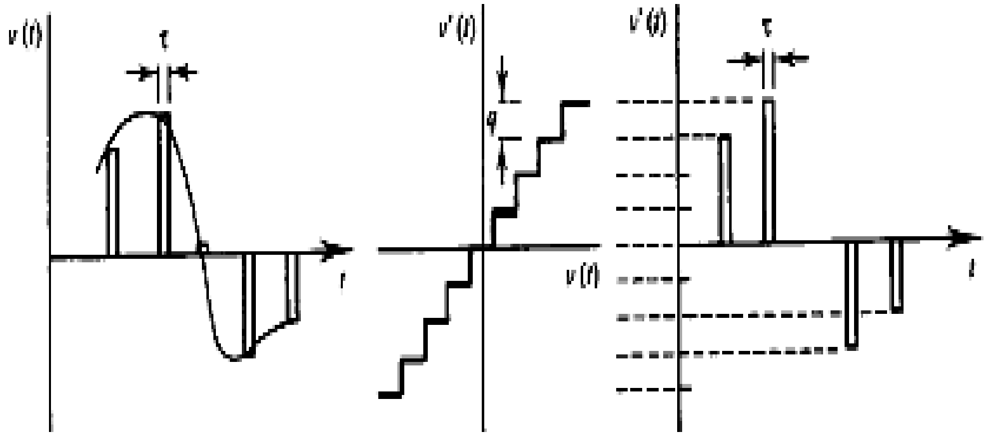
It should be noted, however, that the widely popular Ethernet communication standard is a good example of PAM usage. In particular, 100 BASE-T2 (running at 100 Mb/s) Ethernet medium is utilizing 5 level PAM running at 25 mega pulses / sec over two wire pairs. A special technique is used to reduce inter-symbol interference between the unshielded pairs. Later, 1000 BASE-T medium raised the bar to use 4 pairs of wire running each at 125 mega pulses / sec to achieve 1000 Mb/s data rates, still utilizing PAM-5 for each pair. Several proposals were considered for wire-level modulation, including PAM with 12 discrete levels (PAM-12), 10 levels (PAM-10), or 8 levels (PAM-8).

It should be noted that PAM transmission does not improve the noise performance over base band modulation (which is the transmission of the original continuous signal).

It is well known that in analog FM, bandwidth can be traded for noise performance. PAM signals require a larger transmission bandwidth without any improvement in noise performance. This suggests that there should be better pulse modulations than PAM in terms of noise performance. Two such forms of pulse modulation are; PWM and PPM.

2.3 Pulse Code Modulation, PCM

Although PAM signal becomes discrete (in time) rather than continuous but nevertheless remain analogue in nature since all pulse amplitudes within a specified range are allowed. If a PAM signal is quantized such that each analogue sample is adjusted in amplitude to coincide with the nearest of a finite set of allowed amplitudes (quantizing levels) as shown in figure 2.23, then the resulting signal is no longer analogue but discrete.



a- Analogue i/p signal, b- Quantizer ch/s, c- Quantized o/p signal

Figure 2.23 Quantization of a PAM signal

When the sampled values after quantizing, are encoded (represented by the digital words corresponding to the quantizing level, into which the sample value falls), as shown in figure 2.24, the resulting signal is no longer discrete but digital.

For this purpose, we choose a set of (digital) values. It is customary to choose the number of values that can be represented as power of 2, as 2^n where n is an integer to represent each sample as binary numbers using n binary digits (bits).

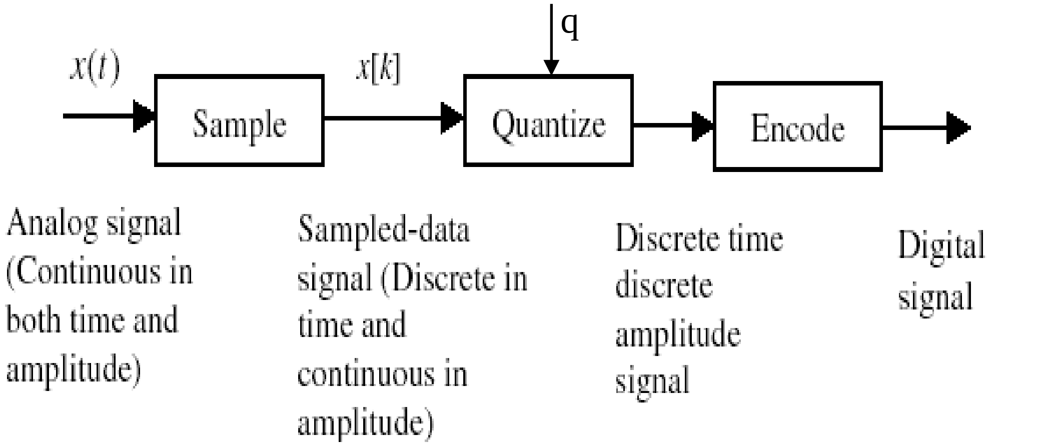


Figure 2.24 Sampling, quantization and encoding process

The number of quantizing levels, M each level represented by n-bits, is;

$$M = 2^{n} \tag{2.22}$$

The steps of the quantization procedure are:

Step 1. If not given, decide how many bits will be used for quantization, for example n bits per sample.

Step 2. Divide the interval between the minimum and maximum signal values V_{fs} into 2^{n} - 1 small intervals, each step will be

$q = V_{fs}/ (2^{n}- 1)$ volts.

This is assuming that all steps are of equal size.

As an example, for V_{fs} = 7 volts, and the number of allowed quantization levels were eight = 2^{3}, then the pulse amplitudes could be represented by the binary numbers from zero (000) to seven (111) using n = 3 bits to represent the binary value of each sample, but for

$n = 4$ bits, we have 2^4 or 16 values for sample magnitude can be represented by 4 bits with binary representation (0000) through (1111). Quantized PAM signals is usually a precursor to generate Pulse Code Modulation (PCM) which has some significant advantages over other base band modulation types.

This is easy to see since the quantized PAM signal no longer exactly represents the original, continuous analogue, signal but a distorted version of it.

Figure 2.25 depicts a typical quantization process in which all sample values falling in a particular quantization interval, are represented by a single discrete value located at the center of the quantization interval, so it is rounded off to send only discrete values.

In this manner the quantization process introduces a certain amount of error ε_q or distortion into the signal samples.

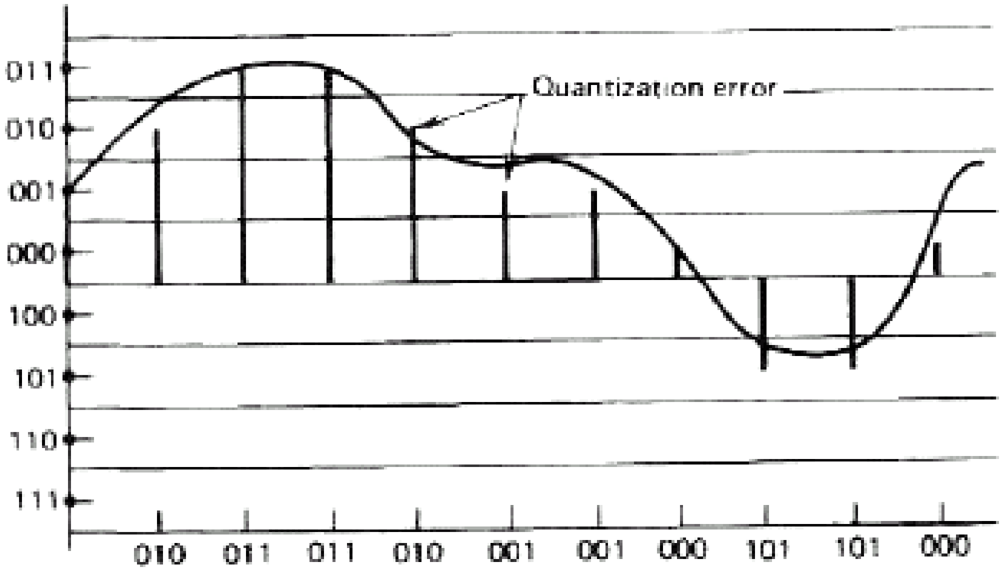


Figure 2.25 Quantization error

This error, known as quantization noise, is minimized by establishing a large number of small quantization intervals.

Of course, as the number of quantization intervals increases, so must the number of bits increases to uniquely identify the quantization intervals.

There are many different types of voice digitization algorithms.

The choice of a particular type is primarily dependent on the implementation cost, the complexity, and the performance requirements implied by the application data rate and quality.

The algorithm chosen for T1 system, (companded PCM) provides excellent quality for all types of input signals (e.g., voice or data) at a moderate data rate (64 kbps) at moderate cost.

The algorithms used in the first-generation DPBXs used lower cost coding techniques (higher rate PCM or delta modulation) because, at the time, a switching application was more sensitive to digital conversion cost and less sensitive to quality or data rate, used un-companded PCM at a data rate of 144 kbps because it was cheaper than companded PCM at the time. Subsequent advantages derived from integrating transmission and switching and a dramatic drop in the cost of Tl-compatible digital voice coders have made obsolete the use of "switching only" voice digitization algorithms. In fact, if the digital network were to be designed today, a more complicated but economically viable codec with a data rate significantly below 64 kbps would probably be utilized.

Transmissions applications with strict bandwidth limits such as high frequency (HF) or digital cellular radio require much more sophisticated voice digitization algorithms to achieve data rates on the order of 8-16 kbps. As a help in reducing the data rate, the performance requirements of these applications are also relaxed as much as the application allows.

Many applications can use very low rate digitization algorithms with significant quality reductions.

Voice digitization techniques can be broadly categorized into two classes: those digitally encoded analog waveforms, (waveform coding) as faithfully as possible and those processing waveforms to encode only the perceptually significant aspects of speech and hearing processes, (voice coding). The first category represents the general problem of analog-to-digital and digital-to-analog conversions and is not restricted to speech digitization.

Except in special cases, telephone equipment designed to transparently reproduce an analog waveform used one of these techniques. Thus, when studying these common waveform encoding techniques, we are, in fact, studying the more general realm of analog-to-digital conversion. The second category of speech digitization is concerned primarily with producing very low data rate speech encoders and decoders for narrowband transmission systems or digital storage devices with limited capacity. A device from this special class of techniques is commonly referred to as a "vocoder" (voice coder). Very low data rate (e.g., 1200-bps) vocoder techniques generally produce unnatural or synthetic sounding speech. As such, low-data-rate vocoders do not provide adequate quality for general telephony. A great deal of effort has been expended to develop medium-rate (e.g., 8-kbps) voice coders with natural speech qualities, primarily for digital cellular applications. These coders are implemented as a combination or hybrid of the low-bit-rate techniques and the waveform coders. Thus, these techniques represent a third class of voice digitization algorithm.

PCM is an extension of PAM wherein after a PAM signal has been quantized the possibility exits of transmitting not the sample itself but a binary number indicating its height as binary digits known as a digital code word. PCM is a standardized method that is used in the telephone network to change an analog signal to a digital one for transmission through the digital communication network. Figure 2.26 shows the sequence of events before a speech signal is completely converted into binary (digital) form.

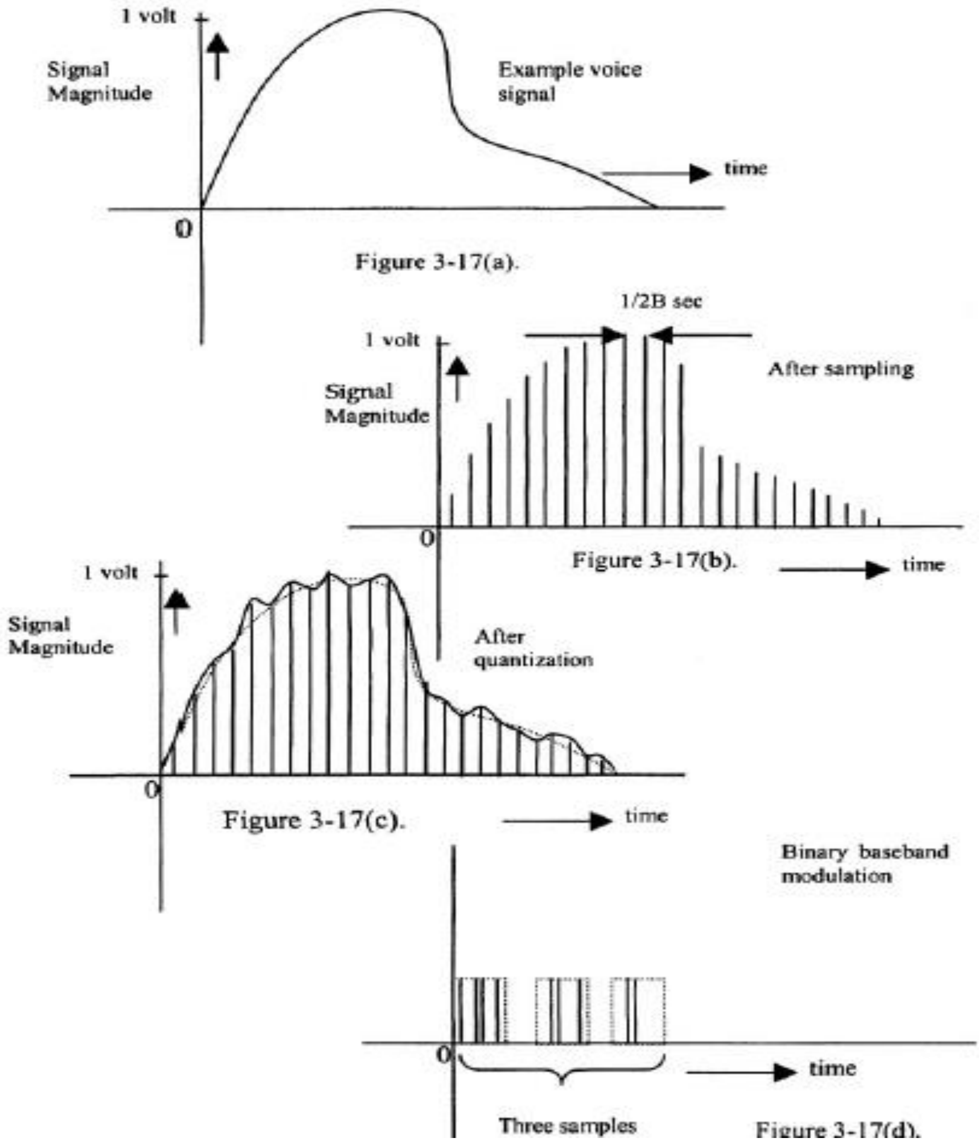


Figure 2.28 Conversion of analogue signal into a bit stream;
a- Analogue signal, b- Sampling process, c-Quantization process,
d- Encoding process.

Assuming the analogue speech signal has a bandwidth of B = 4 kHz, the sampling theorem defines the minimum sampling rate of this signal to be 2B = 8,000 bauds (samples per second) to preserve the information content of the signal. Samples are taken at intervals (intersample time)of $T_s = 1/\ 2B = 125\ \mu s$, then each sample is quantized into one of 256 levels and encoded into digital eight-bit words. Seven bits represent the sample information plus 1 parity bit, then the overall data rate of one PCM speech signal becomes;

$$R_b = f_b = (f_s\ \textit{sample/ sec})\ (\text{n bits/sample})$$

$$R_b = 8{,}000\ \text{sample/sec} \times 8\ \text{bit/sample} = 64\ \text{Kbps.}$$

Because the signaling is binary, the baud rate measures the number of transitions per sec, R_b/n equals the bit rate. The theoretically minimum absolute band width of the PCM signal is;

$$B_{min} = B_{Nyquist} = ½\ \text{Baud rate} = 32\ \text{kHz}$$

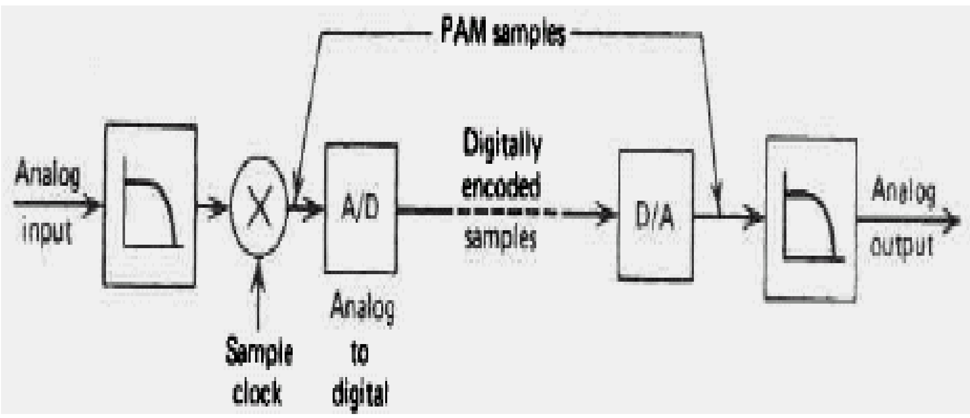
and this is realizable if the PCM waveform consists of (sin x) / x pulse shapes. If rectangular pulse shaping is used, the absolute bandwidth is infinity and the first null band width is, $B_{Null} = R_b = 1 / T_b = 64$ k Hz. This same data rate is available for data transmission through each speech channel in the network. The frequency of voice signal that will be transmitted is chosen to be 300 -to-3,400 Hz, leaving enough guard band for filtering.

Broadcast-quality color television signals have an analog base band bandwidth of somewhat less than 5 MHz. for conventional PCM encoding of these video signals, a sampling rate of;

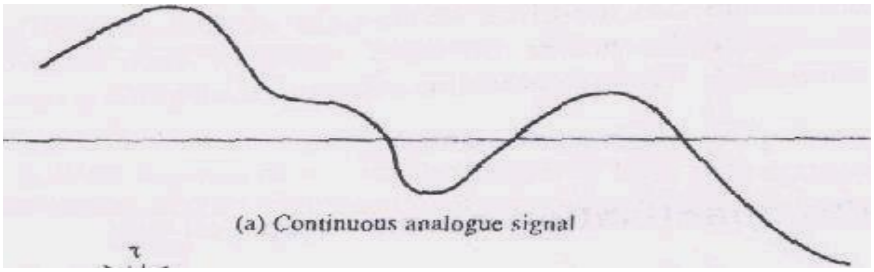
f_s = 10 M sample/sec., and a 9-bit per sample coding scheme is used.

Thus the resulting transmission rate is 90 Mb/s. Most television pictures have a large degree of correlation, which can be exploited to

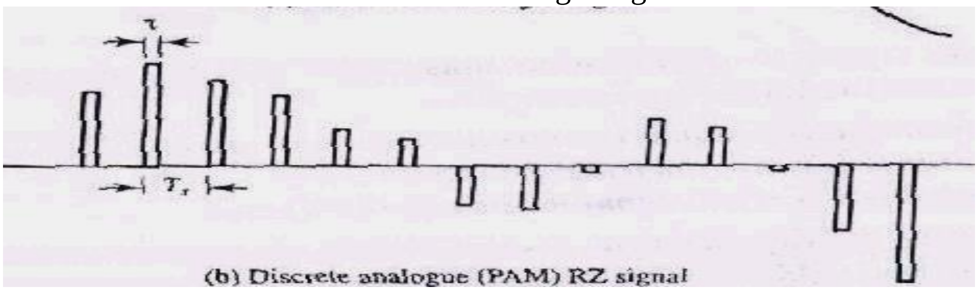
reduce the transmission rate, so digital broadcast-quality color television signals requiring only 20 to 45 Mb/s transmission rates.
This results in the conversion of an analog signal into a string of binary numbers. This binary data can be transmitted digitally using many of the base band or pass band modulation schemes.
The quantization is performed using methods that minimize the quantization error. Also, some signal levels are more sensitive to noise than others. The lower signal levels, for example, are affected by noise more heavily than the higher amplitudes.
Again, we can manipulate the quantization in such a way that finer resolution (step size) can be used to represent lower signal amplitudes so that it is easier to detect them at the receiving side.
Thus, as shown in figure 2.27-a, a PAM system can be converted into a PCM system by adding an analog-to-digital (A/D) converter at the source and a digital-to-analog (D/A) converter at the destination.
Figure 2.27-a -to- f shows the relationship between an information, PAM, quantized PAM and PCM signal.



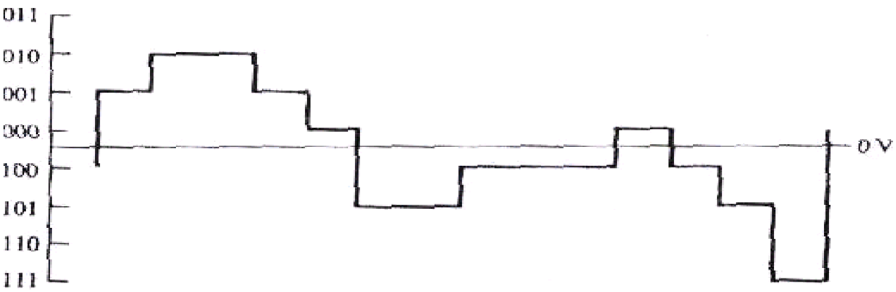
a- Pulse code modulation



b- Continuous analogue signal



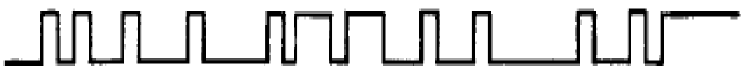
c- Discrete analogue (PAM) RZ signal



d- Discrete digital (quantized PAM) NRZ signal

· 001 · 010 · 010 · 001 · 000 · 101 · 101 · 100 · 100 · 100 · 000 · 100 · 101 · 111

e-Binary coded (quantized) PAM



f- PCM NRZ signal

Figure 2.27 Relationship between PAM, quantized PAM and PCM signal

Figure 2.27-c and d, also illustrates the difference between a pulsed signal with duty cycle (τ / T_S) less than 1.0 and a pulsed signal with duty cycle equal to 1.0. The former are often referred to as return to zero (RZ) signals and the latter as non-return to zero (NRZ) signals.

There is clearly a bandwidth penalty to pay for PCM, if information is to be transmitted in real time, since, in the given example, three/four binary pulses are transmitted instead of one quantized PAM pulse. (The penalty here is a factor of three /four since, for the same pulse duty cycle, each PCM pulse must be one third/fourth the duration of the PAM pulse. PCM signals have greater noise immunity than PAM signals.

One of the measures needed by a voice communication engineer is the quality of speech delivered to the listener. Measurements of speech quality are complicated by subjective attributes of speech as perceived by a typical listener.

One subjective aspect of noise or distortion on a speech signal involves the frequency content, or spectrum, of the disturbance in conjunction with the power level.

Successive quantization errors of a PCM encoder are generally assumed to be distributed randomly and uncorrelated to each other. Thus the commutative effect of quantization errors in a PCM system can be treated as additive noise with a subjective effect that is similar to band limited white noise.

PCM and its variants, such as adaptive PCM (APCM) and adaptive differential PCM (ADPCM) have been extensively used for voice communications on digital links since the 1960s and continue to be popular in today's multimedia networks.

2.3-1 Signal-to-Quantization noise Ratio (SN_qR)

Quantization noise (error) is the difference between the actual value of an analog sample and the assigned digital value due to quantization. For example, if the analog value of a sample is in the range 4/16-5/16, the assigned value is 9/32. For a sample whose value is 0.26 volts the assigned value is 9/32 = 0.281. Thus, there is an error of 0.281- 0.26 = 0.021 volts due to quantization.

The signal-to-quantizing-noise ratio of a quantized signal can be determined as;

$$\mathbf{SN_qR} = \frac{E\left\{S^2(t)\right\}}{E\left\{[y(t)-S(t)]^2\right\}} = \frac{S}{N_q} = \frac{\overline{v^2}}{\overline{\overline{\varepsilon_q^2}}} \qquad (2.23)$$

where, $E\{.\}$: expectation or averaging

S (t) : quantizer input signal (sample) power

y (t) : quantizer output signal decoded signal power

2.3-1.1 SN_qR for Linear PCM

In determining the expected value of the quantization noise, it is convenient to make the following assumptions;

1- Linear quantization (equal increments between quantization levels, i.e., all quantization intervals have equal lengths, q (uniform quantization), the quantization noise is uniformly distributed and independent of the sample values).

2- Zero mean signal (symmetrical pdf about 0 v).

3- Uniform signal pdf (all signal levels (samples) are equally Likely). Figure 2.28 shows that the quantized signal can be decomposed into the sum of the analogue signal and the difference between the quantized and the analogue signals. The difference signal is essentially random and can therefore be thought of as a special type of noise process. The RMS value of this quantization noise can be calculated, this leads to the concept of a Signal-to-Quantization noise Ratio, (SN_qR).

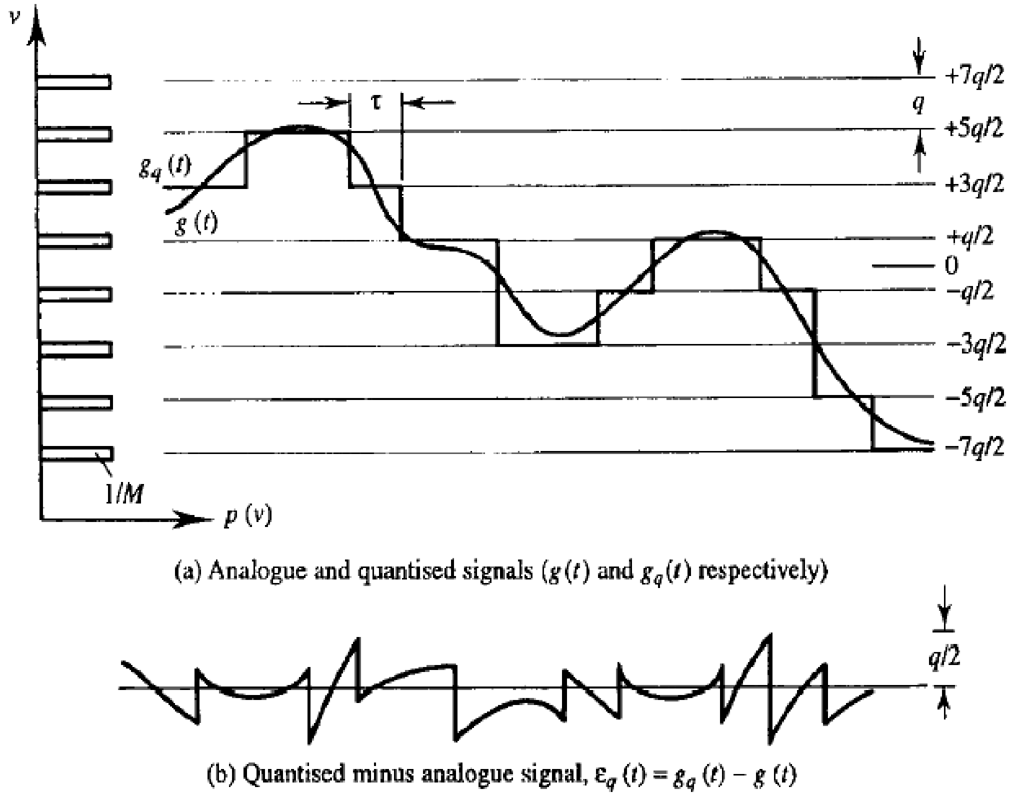


Figure 2.28 Quantization error interpreted as noise

4- The error Y(t) – S(t) is limited in amplitude to q / 2. (Decoded output samples are ideally positioned at the middle of a quantization interval.) The maximum quantization error is the half of the step size q/2. For n-bit PCM, there are 2^n-1 steps.

Therefore, the step size q, is $1/2^n$-1 volts and the maximum quantization error is half of that.

5- A sample value is equally likely to fall anywhere within a quantization interva1, implying a uniform probability density of amplitude l / q.

6- Signal amplitudes are assumed to be confined to the maximum range of the coder. If a sample value exceeds the range of the highest quantization interval, overload distortion (also called peak limiting) occurs.

Example 2.2: For an 8-bit PCM system to be used to transmit an analog signal with range between 0 and 6 volts and bandwidth of 4 kHz, we have the following parameters.

Voltage level range: 0 to 6 volt gives step size, q = 6/ (2^8-1)

= 6/255 volts.

Maximum quantization error = 1/2 × 6/255 = 6/510 volts. Maximum signal size for mid-step resolution = 6 - 6/510 = 6 × 509/510 volts.

Minimum sampling rate = 8000 bauds (symbol/sec)

Minimum bit rate (8 bits/symbol) = 8 × 8000 = 64 kbps

The above is a straightforward description of PCM. Actual systems are much more complex due to certain features of the voice signal, such as the information is carried more in one part of the signal spectrum than the other.

The quantization levels of area 1/ M each are represented by its pdf p(υ), where M is the even number of quantization levels that is assumed to be large enough, and the signal varies from sample to sample.

The distance between adjacent quantizing levels is q v, with the pdf of allowed levels given by:

$$P(v)=\sum_{i=-M}^{M}(1/M)\,\delta(v-iq/2) \qquad (2.24)$$

where, i takes on odd values only. The mean square signal power after quantization is:

$$\overline{\upsilon^2} = \int_{-\infty}^{\infty} \upsilon^2 P(\upsilon) d\upsilon$$

$$= \frac{2}{M}\left[\int_0^{\infty} \upsilon^2 \,\delta(\upsilon - \frac{q}{2}) d\upsilon + \int_0^{\infty} \upsilon^2 \,\delta(\upsilon - 3q/2) d\upsilon + \ldots.\right]$$

$$= \frac{2}{M}\left(\frac{q}{2}\right)^2 \left[1^2 + 3^2 + 5^2 + \ldots\ldots + (M-1)^2\right]$$

$$\therefore \overline{v^2} = \frac{2}{M}\left(\frac{q}{2}\right)^2 \left[\frac{M(M-1)(M+1)}{6}\right]$$

i.e.;

$$\overline{\upsilon^2} = \frac{M^2 - 1}{12} q^2 \qquad volt^2 \tag{2.25}$$

Denoting the quantization error (difference between the un-quantized and quantized signals) as ε_q as showed in figure 2.28, with maximum value q /2, then it follows from assumption 3 that the probability density function, pdf of the quantizing error ε_q is uniform over the range [- q /2, q /2] as shown in figure 2.29, with;

$$P(\varepsilon_q) = \begin{cases} 1/q, & -q/2 \le \varepsilon_q < q/2 \\ 0, & elsewhere \end{cases} \tag{2.26}$$

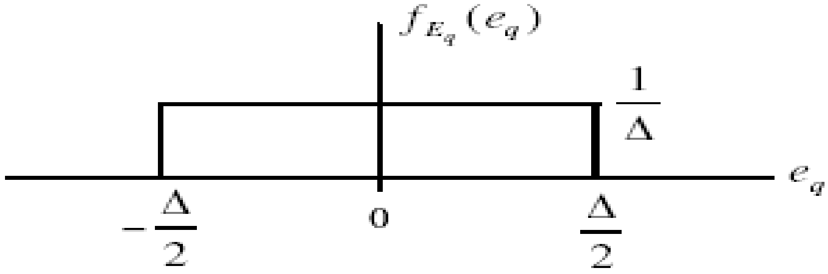


Figure 2.29 PDF of quantizing error

Denoting the quantizing error of the kth sample by $\varepsilon_q[k]$, we have;

$$\overline{\varepsilon_q} = E\{\varepsilon_q[k]\} = \int_{-q/2}^{q/2} \varepsilon_q P(\varepsilon_q) d\varepsilon_q = 0 \qquad (2.27)$$

So the quantizing error has zero mean. The variance (mean square quantization error (noise)) of $\varepsilon_q[k]$ is;

$$\overline{\varepsilon_q^2} = E\{\varepsilon_q^2[k]\} = \int_{-q/2}^{q/2} \varepsilon_q^2 P(\varepsilon) d\varepsilon_q$$

$$= \frac{1}{q}\left[\frac{\varepsilon_q^3}{3}\right]_{-\frac{q}{2}}^{+\frac{q}{2}} = \frac{q^2}{12} \quad V^2 \qquad (2.28)$$

using equation (2.25), then,

$$SN_qR = \frac{\overline{\upsilon^2}}{q^2/12} = M^2 - 1 \qquad (2.29)$$

$$SN_qR|_{dB} = 10 \log_{10}\left(\frac{\overline{v^2}}{q^2/12}\right)$$

$$= 10.8 + 20 \log_{10}\left(\frac{v}{q}\right) \qquad (2.30)$$

For a sine wave input, the SN_qR produced by uniform quantization is;

$$SN_qR|_{dB} = 10\ \log_{10}\left(\frac{E_m^2/2}{q^2/12}\right)$$

$$= 7.78 + 20\ \log_{10}\left(\frac{E_m}{q}\right) \qquad (2.31)$$

where E_m is the peak amplitude of the sine wave.

The average SN_qR given by equation (2.29) is therefore given by:

$$SN_qR = \frac{\overline{v^2}}{\overline{\varepsilon_q^2}} = M^2 - 1 \qquad (2.32)$$

For large SN_qR the approximation $SN_qR = M^2$ is used.

Expressing in dB, $SN_qR = 20 \log M$

Since the peak signal level is $M\,q / 2$ then the peak SN_qR is:

$$(SN_qR)_{peak} = \frac{(Mq/2)^2}{\overline{\varepsilon_q^2}} = 3M^2 \qquad (2.33)$$

Expressing in dB the $(SN_qR)_{peak}$ is:

$$(SN_qR)_{peak} = 10 \log (3\ M^2) = 10 \log 3 + 20 \log M$$

$$(SN_qR)_{peak} = 4.8 + 20 \log M$$

$$= 4.8 + 6\ n \qquad \text{dB} \qquad (2.34)$$

Example 2.3, A sine wave with a l-V minimum amplitude is to be digitized with a minimum SN_qR of 30 dB. How many uniformly spaced quantization intervals are needed, and how many bits are needed to encode each sample?

Using equation (2.31), the maximum size of a quantization interval is determined as;

$$q = (1)\ 10^{-(30 - 7.78)/20} = 0.078 \text{ V}$$

Thus $E/q = 1/q \approx 13$ quantizing intervals are needed for each polarity for a total of 26 intervals in all. The number of bits required to encode each sample is determined as;

$$n = \log_2 (26) = 4.7 = 5 \text{ bits per sample}$$

When measuring quantization noise power, the spectral content is often weighted in the same manner as noise in an analog circuit. Unfortunately, spectrally weighted noise measurements do not always reflect the true perceptual quality of a voice encoder / decoder. If the spectral distribution of the quantization noise more or less follows the spectral content of the speech waveform, the noise is masked by the speech and is much less noticeable than noise uncorrelated to the speech. On the other hand, if the quantization process produces energy at voice band frequencies other than those contained in particular sounds, they are more noticeable.

Whilst it is true that PCM signals are more tolerant of noise than the equivalent quantized PAM signals it is also true that both suffer the same degradation due to quantization noise. Uniform PCM system uses a conventional analog-to-digital converter to generate the binary sample codes. The number of bits required for each sample is determined by the maximum acceptable noise power. Minimum digitized voice quality requires a signal-to-noise ratio in excess of 26 dB. For a uniform PCM system to achieve a SN_qR of 26 dB, equation (2.31) indicates that $q_{max} = 0.123\ E_m$. For equal positive and negative signal excursions (encoding from $- E_m$ to E_m), this result indicates that just over l6 quantization intervals, or 4 bits per sample, are required.

In addition to providing adequate quality for small signals, a telephone system must be capable of transmitting a large range of signal amplitudes, referred to as dynamic range. Dynamic range (DR) is usually expressed in decibels as the ratio of the largest possible amplitude signal to the smallest possible amplitude signal (other than zero) that can be decoded by the D/A converter in the receiver:

$$DR = 10 \log_{10} (P_{max} / P_{min})$$

$$= 20 \log_{10} (V_{max} / V_{min}) \qquad (2.35)$$

V_{min} is the quantum value (resolution)

Example 2.4; Consider a PCM system with the following parameters;

Maximum analog i/p frequency = 4 kHz

Maximum decoded voltage at the receiver = $\pm$ 2.55 v

Minimum dynamic range = 46 dB

Determine; a- minimum sampling rate – b- minimum number of bits used in the PCM code, c- resolution, d- quantization error.

Solution;

a- As $f_s \geq 2\, f_m$, so the sampling rate will be;

$$f_s = 2 \text{ x } 4 \text{ kHz} = 8 \text{ kHz}$$

b- As $DR = 20 \log \ (V_{max} / V_{min})$

$$46 \text{ dB} = 20 \log \ (V_{max} / V_{min})$$

$$2.3 = \log \ (V_{max} / V_{min})$$

$$DR = 10^{2.3} = (V_{max} / V_{min}) = 199.5$$

The minimum number of bits is determined as;

$$2^n - 1 = DR = 199.5, \text{ so, } n \approx 8 \text{ bits}$$

because the input amplitude range is $\pm$ 2.55, one additional bit, the sign bit, is required. Therefore 9 bits are required and the total number of PCM codes is 29 = 512 (there are 255 positive codes, and 255 negative codes, and 2 zero codes). The actual dynamic range is calculated as;

$$DR\,\Big|_{dB} = 20 \log \ (2^n - 1) = 20 \log (256 - 1) = 48.13 \text{ dB}$$

c- The resolution is determined by dividing the maximum positive or maximum negative voltage by the number of positive or negative nonzero codes;

$$\text{resolution} = V_{max} / (2^n - 1) = 2.55 / (2^8 - 1)$$

$$= 2.55 / (256 - 1) = 0.01 \text{ V}$$

d- the maximum quantization error is;

$\epsilon = q / 2 = 0.01 / 2 = 0.005$ V

A typical minimum dynamic range is 30 dB. Thus signal values as large as 3l times E_m must be encoded without exceeding the range of quantization intervals. Assuming equally spaced quantization intervals for uniform coding, the total number of intervals is determined as 496, which requires 9-bit code words, (This result is derived with the assumption of minimum performance requirements). Higher performance objectives (less quantization noise and greater dynamic range) require as many as 13 bits per sample for uniform PCM systems. This coding performance was established when it was likely that multiple conversions would occur in an end-to-end connection. Now that the possibility of multiple A/D and D/A conversions has been eliminated, end-to-end voice quality is much better than it was in the analog network).

The performance of an n-bit uniform PCM system is determined by observing that;

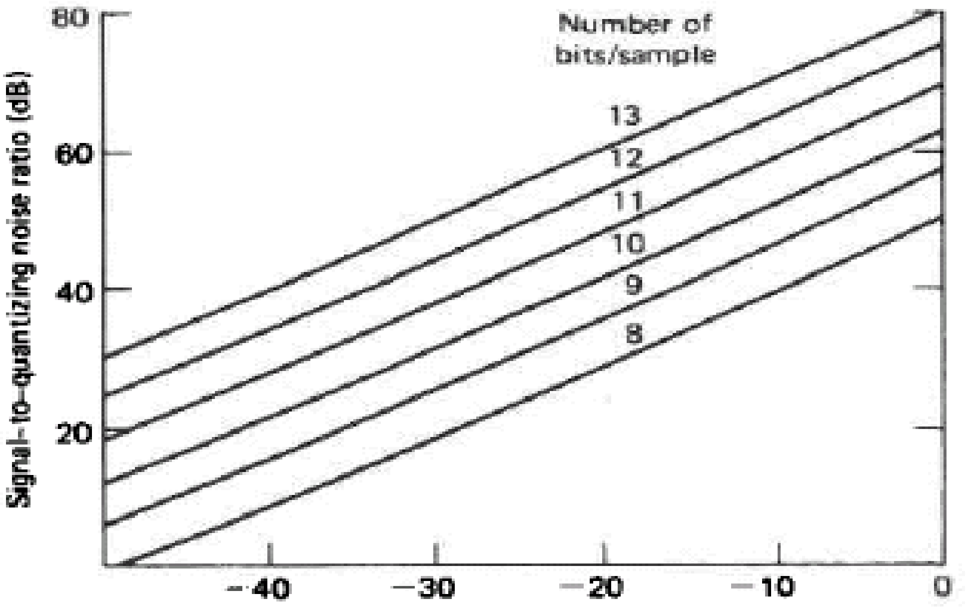
$$q = 2\, E_{m\text{-max}} / 2^n \qquad (2.36)$$

where $E_{m\text{-max}}$ is the maximum (non overloaded) amplitude.

Substituting equation (2.36) into equation (2.31) produces the PCM performance equation for uniform coding;

$$SN_qR = 1.78 + 6\,n + 20 \log_{10} (E_m / E_{m-\max}) \qquad (2.37)$$

The first two terms of equation (2.37) provide the SN_qR when encoding a full-range sine wave. The last term indicates a loss in SN_qR when encoding a lower level signal. These relationships are presented in figure 2.30, which shows the SN_qR of a uniform PCM system as a function of the number of bits per sample and the magnitude of an input sine wave.



$E_m / E_{m\text{-max}}$

Figure 2.30 SN_qR for uniform PCM coding

High-quality PCM encoders produce quantization noise that is evenly distributed across voice frequencies and independent of the encoded waveforms. Thus quantization noise ratios defined in equation (2.29) are good measures of PCM performance.

Example 2.5, What is the minimum bit rate that a uniform PCM encoder must provide to encode a high-fidelity audio signal with a dynamic range of 40 dB? Assume the fidelity requirements dictate passage of a 20-kHz bandwidth with a minimum signal-to-noise ratio of 50 dB. For simplicity, assume sinusoidal input signals.

To prevent fold over distortion, the sampling rate must be at least 40 kHz.

Assuming an excess sampling factor comparable to that used in D-type channel banks (4000/3400), we choose a sampling rate of 48 kHz as a compromise for a practical band limiting filter. By observing that a full-amplitude signal is encoded with an SN_qR of 40 + 50 = 90 dB, we can use equation (2.37) to determine the number of bits n required to encode each sample:

$$n = (90 - 1.78) / 6 = 15 \text{ bit}$$

Thus the required bit rate is;

(15 bit/sample) (48,000 sample/sec) = 720 k bit/ sec.

For a given number of quantization levels, M, the number of binary digits required for each PCM code word is;

$n = \log_2 M$. The PCM peak signal to quantization noise ratio, $(SN_qR)_{peak}$ is therefore;

$$(SN_qR)_{peak} = 3\, M^2 = 3\, (2^n)^2 \qquad (2.38)$$

If the ratio of peak to mean signal power, $\upsilon^2_{peak} / \upsilon^2$, is denoted by α then, the average signal to quantization noise ratio is:

$$SN_qR = 3\, (2^{2n})\, (1/\alpha) \qquad (2.39)$$

Expressing in dB this becomes;

$$SN_qR = 4.8 + 6n - \alpha_{dB} \qquad (2.40)$$

For a sinusoidal signal $\alpha = 2$ (3 dB), for speech $\alpha = 10$ dB. The SN_qR for an n-bit PCM voice system can therefore be estimated using the rule of thumb 6 (n - 1) dB.

Example 2.6, A digital communications system is to carry a single voice signal using linearly quantized PCM. What PCM bit rate will be required if an ideal anti-aliasing filter with a cut-off frequency of 3.4 kHz is used at the transmitter and the signal to quantization noise ratio is to be kept above 50 dB?

$$SN_qR = 4.8 + 6n - \alpha_{dB}$$

For voice signals $\alpha = 10$dB, i.e.:

$$n = (50 + 10 - 4.8) / 6 = 9.2$$

10 bit / sample are therefore required. The sampling rate required as given by Nyquist's rule is;

$$f_S = 2.2 \text{ x } 3.4 \text{ kHz} = 7.4 \text{ kHz (k samples / sec.)}$$

The PCM bit rate (or more strictly binary baud rate) is therefore:

$$R_b = f_S\, n = 7.48 \text{ x } 10^3 \text{ x } 10 \text{ bits / sec.}$$

$$= 74.8 \text{ k bit /s.}$$

2.3-1.2 SN_qR for Decoded PCM

Figure 2.31 Shows the complete block diagram of PCM system.

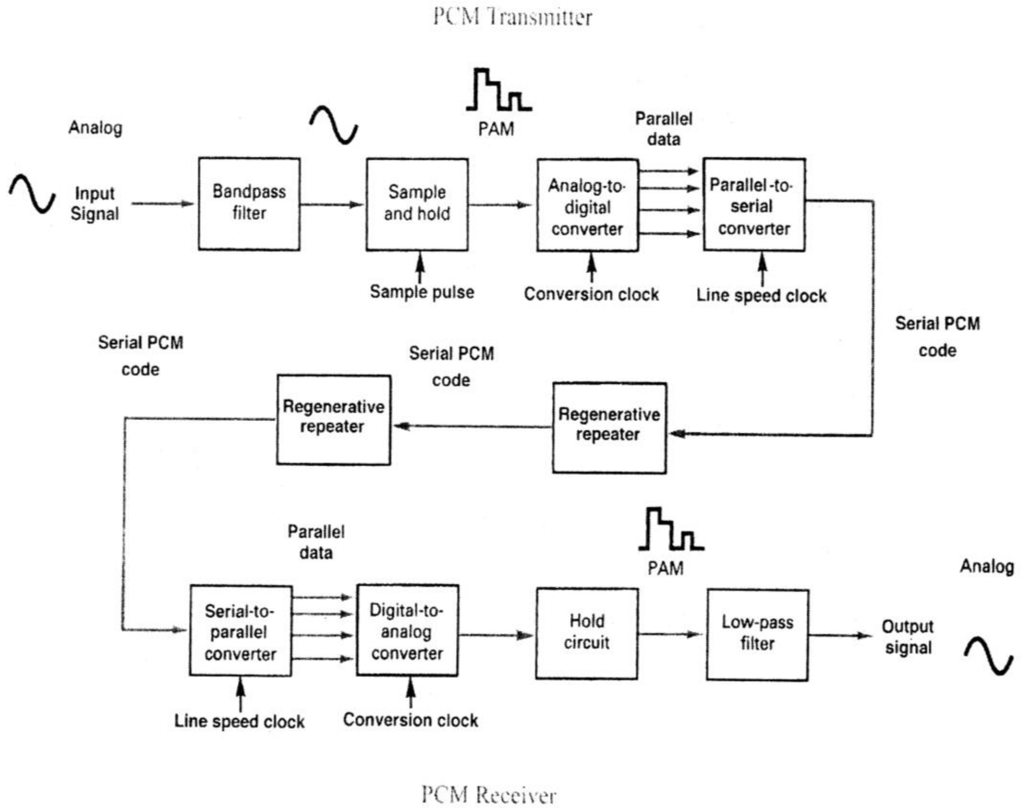


Figure 2.31 PCM system

If all PCM code words are received and decoded without error then the SNR of the decoded signal is essentially equal to the signal to quantization noise ratio, SN_qR, as given in equations (2.29 - 2.33).

In the presence of channel and/or receiver noise, however, it is possible that one or more symbols in a given code word will be changed sufficiently in amplitude to be interpreted in error. For binary PCM this involves a digital 1 being interpreted as a 0 or a digital 0 being interpreted as a 1.The effect that such an error has on the SNR of the decoded signal depends on which symbol is detected in error. The least significant bit in a binary PCM word will introduce an error in the decoded signal equal to one quantization level. The most significant bit would introduce an error of many quantization levels.

The following reasonably simple analysis gives a useful expression for the SNR performance of a PCM system in the presence of noise. We first assume that the probability of more than one error occurring in a single n-bit PCM code word is negligible. We also assume that all bits in the code word have the same probability of being detected in error, (P_e). Using subscripts 1, 2, ….., n to denote the significance of PCM code word bits (1 corresponding to the least significant, n corresponding to the most significant) then the possible errors in the decoded signal are:

$$\varepsilon_1 = q$$
$$\varepsilon_2 = 2q$$
$$\varepsilon_3 = 4q$$
$$\ldots\ldots\ldots \qquad (2.41)$$
$$\varepsilon_n = 2^{n-1} q$$

The mean square decoding error, $\overline{\varepsilon_{de}^2}$ is the mean square of the possible errors multiplied by the probability of an error occurring in a code word, i.e.:

$$\overline{\varepsilon_{de}^2} = n\, P_e(1/n)[(q)^2 + (2q)^2 + \ldots\ldots + (2^{n-1} q)^2]$$

$$= P_e\,(q)^2\,[4^0 + 4^1 + 4^2 + 4^3 + \ldots\ldots + 4^{(n-1)}] \qquad (2.42)$$

The square bracket is the sum of a geometric progression with the form:

$$S_n = a + a\,r + a\,r^2 + \ldots + a\,r^{(n-1)} = \frac{a(r^n - 1)}{r - 1} \qquad (2.43)$$

where a = 1 and r = 4. Thus:

$$\overline{\varepsilon_{de}^2} = P_e\, q^2\, (4^n - 1)/3 \qquad (\text{volt}^2) \qquad (2.44)$$

Since the error or noise which results from incorrectly detected bits is statistically independent of the noise which results from the quantization process we can add them together on a power basis, i.e.:

$$SN_TR = \frac{\overline{\upsilon^2}}{\overline{\varepsilon_q^2} + \overline{\varepsilon_{de}^2}} \qquad (2.45)$$

where $\overline{v^2}$ is the received signal power. Using equation (2.25) for $\overline{v^2}$ and equation (2.44) for $\overline{\varepsilon_{de}^2}$ and remembering that the number of quantization levels M = 2^n we have;

$$SN_TR = \frac{M^2 - 1}{1 + 4(M^2 - 1)P_e} \qquad (2.46)$$

Equation (2.46) allows us to calculate the average SN_TR of the decoded PCM signal including both quantization noise and the decoding noise which occurs due to corruption of individual PCM bits by channel or receiver noise. If we denote the channel SN_TR using the subscript in and the decoded PCM SN_TR using the subscript out then for binary, polar NRZ signaling, using simple center point decisions, we have:

$$SN_TR_{out} = \frac{M^2 - 1}{1 + 4(M^2 - 1)1/2\, erfc((1/2)\, SNR_{in})^{1/2}} \qquad (2.47)$$

The SN_TR_{out} in equation (2.47) is linear ratio (not dB values) and the function erfc (x) is the complementary error function. Equation (2.47) is sketched for various values of n = $\log_2$ M in figure 2.32.

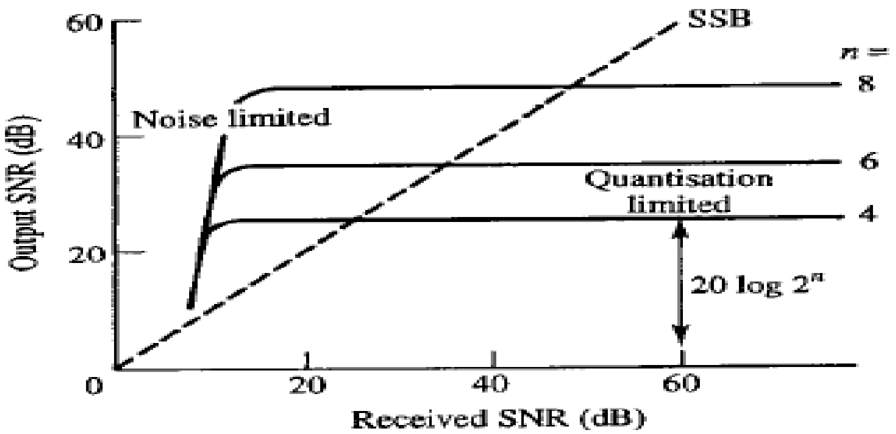


Figure 2.32 Input / output SNR for PCM

The noise immunity advantage of PCM illustrated by this figure is clear. The x-axis is the SNR of the received PCM signal. The y-axis is the SNR of the reconstructed (decoded) information signal. If the SNR of the received PCM signal is very large then the total noise is dominated by the quantization process and the output SNR is limited to SN_qR. In practice, however, PCM systems are operated at lower input SNR values near the knee or threshold of the curves in figure 2.32. The output SNR is then significantly greater than the input SNR. At very low input SNR, when the noise is of comparable amplitude to the PCM pulses, then the interpretation of code words starts to become unreliable. Since even a single error in a PCM code word can change its numerical value by a large amount then the output SNR in this region (i.e. below threshold) decreases very rapidly.

Example 2.7, Find the overall SNR_{out} for the reconstructed analogue voice signal in example (2.5) if receiver noise induces an error rate, on average, of one in every 10^6 PCM bits.

Using equations (2.46), (2.32);

$$SNR_{out} = \frac{SN_qR}{1 + 4\,SN_qR\,P_e}$$

$$SN_qR = 4.8 + 6n - \alpha \text{ dB} = 4.8 + (6 \times 10) - 10$$

$$\therefore \quad SN_qR = 54.8 \text{ dB (or } 3.020 \times 10^5)$$

$$SNR_{out} = \frac{3.020\,x10^5}{1 + 4(3.020\,x10^5)(1x10^{-6})}$$

$$= 1.368 \times 10^5 = 51.4 \text{ dB}$$

The SNR available with PCM systems increases with the square of the number of quantization levels as given by equation (2.29) while the baud rate, and equivalently the bandwidth, increases with the logarithmic of the number of quantization levels which is clear in example 2.3. Thus the bandwidth can be exchanged for SNR. Close to threshold PCM is superior to all analogue forms of pulse modulation at low SNR. However, all practical PCM systems have a performance which is an order of magnitude below their theoretical optimum.

As PCM signals contain no information in their pulse amplitude they can be regenerated using non-linear processing at each repeater in a long haul system.

Such digital regenerative repeaters allow accumulated noise to be removed and essentially noiseless signals to be retransmitted to the next repeater in each section of the link. The probability of error does accumulate from hop to hop however.

We see a three way trade-off among energy, bandwidth, and processing complexity. An increase in any one means the other two can be reduced, more or less, for the same performance. A PCM digitizer, for example, can be replaced by a more complex digitizer, which puts out fewer bits, which consume less transmission bandwidth.

Transmission error may be driven down by more complex coding, instead of more transmission energy. With modern coded modulation, coding complexity can even be exchanged for bandwidth.

The cheapest of energy, bandwidth, and processing is now processor complexity.

For the listener it would be a nuisance if he had to listen to a very low volume signal corrupted by a relatively high quantization error (low SN_qR). To achieve the same SN_qR for a small-amplitude signal as for a large-amplitude signal, a quantizer with a non-uniform step size is required. To achieve this non-uniform step-size quantization, given a uniform step-size quantizer, it is necessary to precede it with a nonlinear input-output device known as a compandor. Note that the compandor followed by the linear quantizer amplifies the low volume signals more than the high-volume signals.

2.3-1.3 Companded PCM

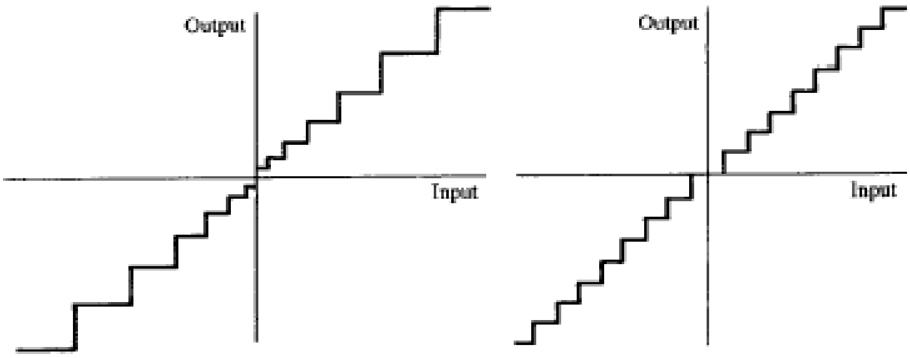
The goal in the coder design is to get as good an average S/N as possible when the sampling rate and the number of bits for each sample are given.

Linear quantizing is not the optimum solution because the derived expressions for SN_qR assumed that the information signal has uniformly quantized samples i.e., that all quantization levels are used equally. At low signal levels the quantizing noise is high and the SNqR is very low. At high signal levels the quantizing noise is the same even though we would tolerate a high noise level. We should define quantizing levels in such a way that performance is acceptable over a wide dynamic range of the voice. This requires that quantum levels are not uniformly spaced and we call this non-uniform quantizing.

As indicated in equation (2.37) and figure 2.30, the SN_qR increases with the signal amplitude E_m.

For example, a 26-dB SN_qR for small signals and a 30-dB dynamic range produces a 56-dB SN_qR for a maximum amplitude signal. In this manner a uniform PCM system provides unneeded quality for large signals. Moreover, the large signals are the least likely to occur. For these reasons the code space in a uniform PCM system is very inefficiently utilized.

One way to arrange to use the PCM system efficiently is to adapt non-linear quantization or equivalently, companding. Non-linear quantization is illustrated in figure 2.33-a.



a- Non-linear quantization b- Linear quantization
Figure 2.33 Quantization characteristics

If the information signal pdf has small amplitude for a large fraction of time and large amplitude for a small fraction of time (as is usually the case) then the step between adjacent quantization levels is made small for low levels and larger for higher levels, (the quantization intervals be directly proportional to the sample value).

One way to understand the companding process is to think of compressing the dynamic range of the analog signal first by compressor circuitry prior to transmission, which amplifies low levels more than higher levels as shown in figure 2.34-a. After this we may use linear quantization, and the signal values after compression and linear quantizing will actually be non-uniformly quantized. In the decoder of the receiver, we use linear quantizing to reproduce the compressed sample values. Then we low pass filter the sample sequence to reproduce the compressed analog signal. We then expand this analog signal by amplifying low levels less than high levels to cancel the distortion that was produced by the compressor in the encoder as shown in figure 2.34-b.

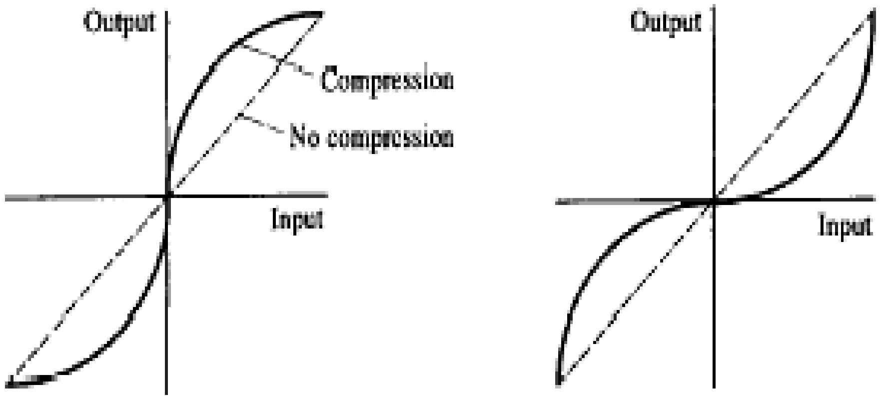


Figure 2.34 Typical compression and expansion (compander) characteristics

After linear decoding in the receiver, the noise level is the same at any sample level. In expansion a low-level signal is reduced to its original value and quantizing noise is attenuated. This makes the noise level lower at low signal levels than at high signal levels and improves the SNqR at low signal levels. This improvement of the average SNqR at low analog signal levels is essential because noise is most disturbing at low signal levels, and the quantizing noise does not disturb the listener very much if the signal level is high as well.

Figure 2.35 shows a complete non-uniform quantizing system.

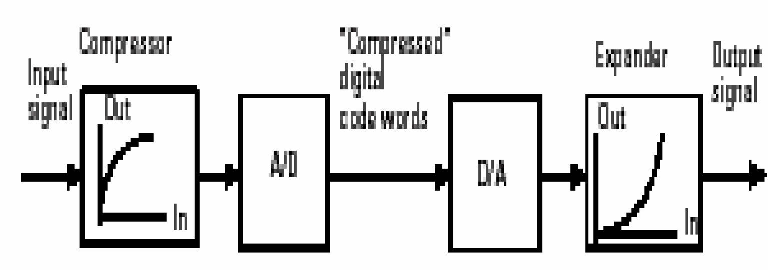


Figure 2.35 Complete non-uniform quantizing system

It is clear that coding performance is improved using non-uniform quantization intervals. The SN_qR will be maintained constant, as nearly as possible for all signal levels. So, companding is a mean of improving the dynamic range of a communication system, (match the difference between men and women speech for example), such that increasing the amount of compression, increases the dynamic range at the expense of the S/N for high signal amplitudes.

With PCM, companding may be accomplished using analog or digital techniques. Early PCM systems used analog companding, whereas more modern systems use digital companding.

Analog Companding

Historically, analog compression was implemented using specially designed diodes inserted in the analog signal path in a PCM transmitter prior to the sample-and-hold circuit. Analog expansion was also implemented with diodes that were placed just after the low pass filter in the PCM receiver. The following figure 2.36 shows the basic process of analog companding.

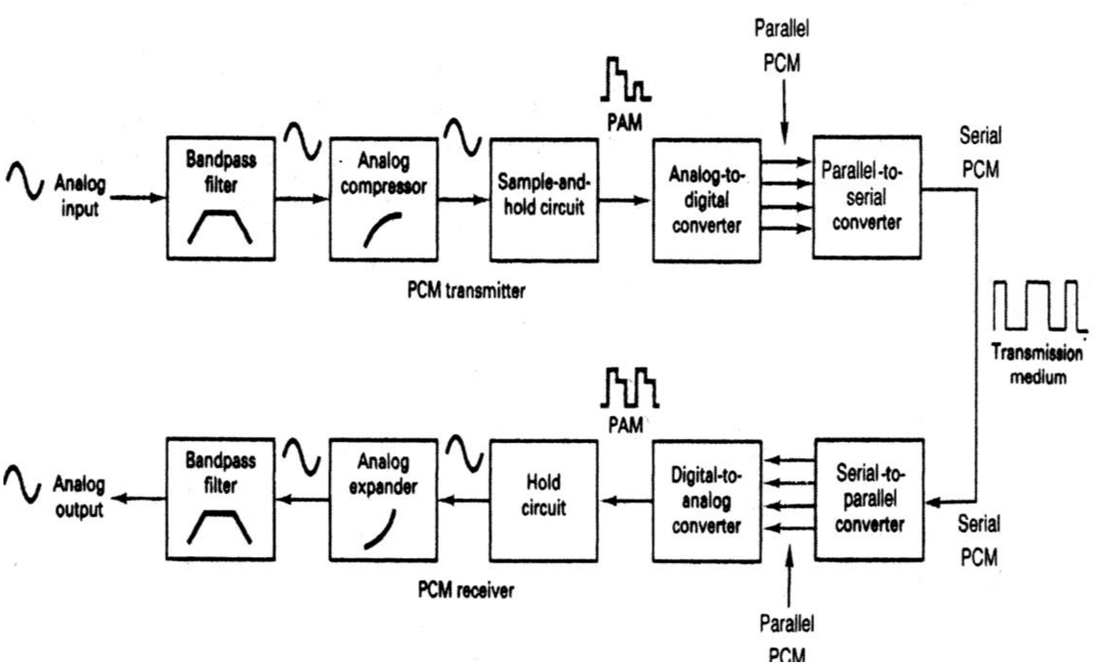


Figure 2.36 PCM system with analog companding

In the transmitter, the dynamic range of the analog signal is compressed, sampled and then converted to a linear PCM code. In the receiver, the PCM code is converted to a PAM signal, filtered and then expanded back to its original dynamic range. Different signal distributions require different companding characteristics. For instance, voice quality telephone signals require a relatively constant SN_qR performance over a wide dynamic range, which means that the distortion must be proportional to signal amplitude for all input signal levels. This requires a logarithmic compression ratio, which requires an infinite dynamic range and an infinite number of PCM codes.

Of course, this is impossible to achieve.

When the quantization intervals are not uniform, a nonlinear relationship exists between code words and the sample values that they represent. However, there are two main different nonlinear coding schemes of analog companding currently being used that closely approximate a logarithmic function and are often called log-PCM codes. They have been standardized internationally for speech by the International Telecommunication Union, ITU; and are known as A-law which is used in European standard countries, and the μ-law, which is used in North America and Japan. Here are some key points about these coding schemes:

1- Companding curves are based on the statistics of human voice and many good solutions can be found.

2- These schemes provide quite the same quality, but they are not compatible.

3- A conversion device, a trans-coder, is needed between countries using different standards.

A-Law Companding

In Europe, the ITU-T has established A-law companding families of compression characteristics (Rec. G. 732) that defines the continuous curves given by the following formulas to be used to approximate true logarithmic companding. The normalized A-law compression characteristic curve is divided into linear and logarithmic sections defined as;

$$F_A(x) = \begin{cases} \operatorname{sgn}(x).\left[\dfrac{A|x|}{1+\ln(A)}\right] & for\ 0 \le |x| \le \dfrac{1}{A} \\ \operatorname{sgn}(x).\left[\dfrac{1+\ln(A|x|)}{1+\ln(A)}\right] & for\ \dfrac{1}{A} \le |x| \le 1 \end{cases} \qquad (2.48\text{-a})$$

or,

$$V_{out} = V_{max}\ \frac{A\,V_{in}/V_{max}}{1+\ln A} \qquad 0 \le \frac{V_{in}}{V_{max}} \le \frac{1}{A}$$

$$= \frac{1+\ln(AV_{in}/V_{max})}{1+\ln A} \qquad \frac{1}{A} \le \frac{V_{in}}{V_{max}} \le 1 \qquad (2.48\text{-b})$$

where, $F_A(x)$ is the normalized output signal from the compressor,

$|x| = \left|\dfrac{V_{in}}{V_{max}}\right|$ is the normalized input signal to the compressor (V_{max} is the maximum uncompressed analog input amplitude, volts, V_{in} is the amplitude of the input signal at a particular instant of time, volts), while, V_{out} is the compressed output signal amplitude, volts, and sgn (x) is the signum function (polarity (sign) of x) which is +1 for x > 0 and -1 for x < 0.

The SN_qR is constant in the logarithmic section and directly proportional to the signal value in the linear section.

The parameter A defines the curvature of the compression characteristic with A = 1 giving a linear law as shown in figure 2.37-a. The commonly adapted value is a standard value of A = 87.6, which gives a 24 dB improvement in SN_qR over linear PCM for small signals ($|x| < 1/A$) and a constant SN_qR of 38 dB for large signals ($|x| > 1/A$).

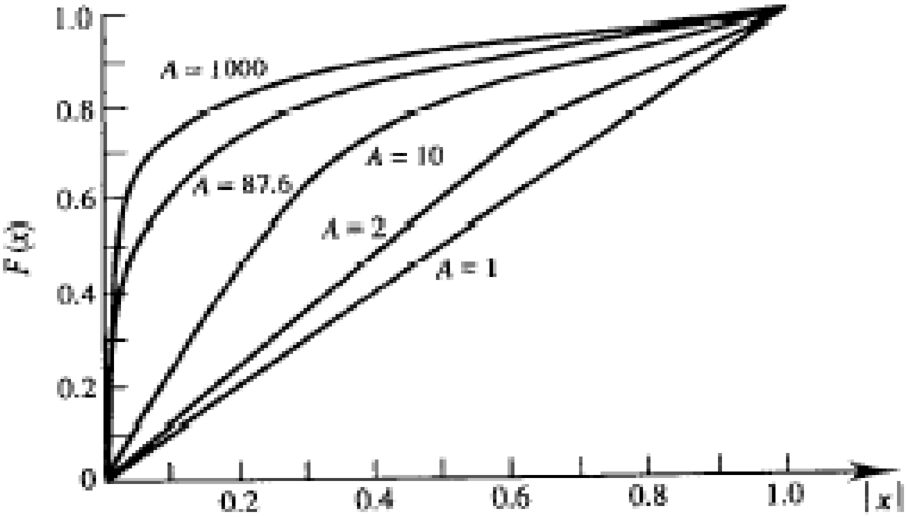


Figure 2.37-a A-law compression ch/s for several values of A

The dynamic range of the logarithmic (constant SN_qR) region of this characteristic is;

$20 \log_{10} [1/(1/A)]=39$ dB. The overall effect is to allow 11 bit (2048 level) linear PCM, which would be required for adequate voice signal quality, to be reduced to 8 bit (256 level) companded PCM. A 4 kHz voice channel sampled at its Nyquist rate (8 kHz) therefore yields a companded PCM bit rate of 64 k bit/sec.

The dashed line in figure 2.38 indicates that ITU-T/CCITT recommendations define the continuous curves given by the preceding formulas but approximate them with a curve with linear segments for easier implementation.

As an example of the performance of a nonlinear coding scheme, figure 2.38 represents the SN_qR dependence on the signal level for A-law companding. The signal level may vary within the range of 40 dB while SN_qR remains nearly unchanged. However, when signal level is high, linear quantizing would give better performance, as the “without companding” dashed line shows.

We see from figure 2.38 that at low levels the SN_qR of A-law companding is more than 20 dB better than linear coding. The curve gives this performance when the signal is a sine wave and the ripple of the curve is a consequence of the approximation of the compression curve with linear segments.

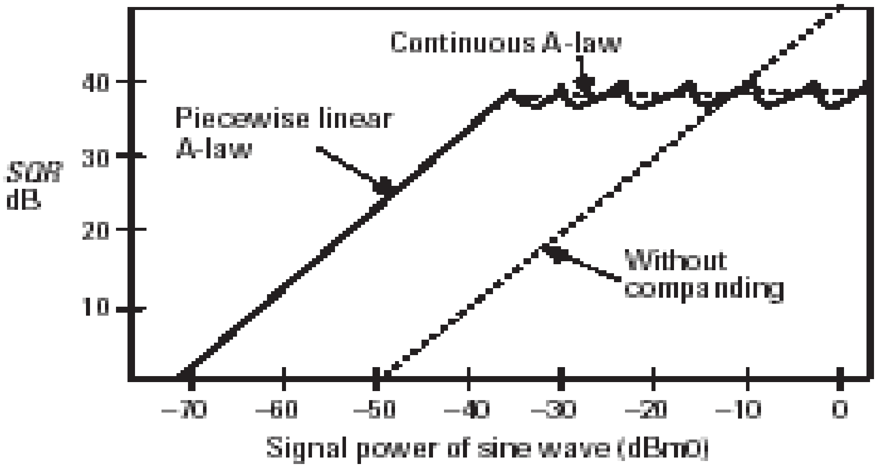


Figure 2.38 Companding performance

The inverse or expansion characteristic is defined as;

$$F^{-1}{}_{A}(y) = \begin{cases} \operatorname{sgn}(y).\left[\dfrac{|y|[1+\ln(A)]}{A}\right] & 0 \le |y| \le \dfrac{1}{1+\ln(A)} \\ \operatorname{sgn}(y).\left[\dfrac{e^{|y|[1+\ln(A)]-1}}{A}\right] & \dfrac{1}{1+\ln(A)} \le |y| \le 1 \end{cases} \qquad (2.49)$$

where, $y = F_A(x)$

μ-law Companding

It is the family of compression characteristics (Rec. G.733) used in North America and Japan, which is defined as follows:

$$F_\mu(x) = \operatorname{sgn}(x) \,.\, \frac{\ln(1+\mu|x|)}{\ln(1+\mu)} \qquad 0 \le |x| \le 1 \qquad (2.50\text{-a})$$

$$V_{out} = \frac{V_{\max}\ \ln(1+\mu V_{in}/V_{\max})}{\ln(1+\mu)} \qquad 0 \le \frac{V_{in}}{V_{\max}} \le 1 \qquad (2.50\text{-b})$$

where, $F_\mu(x)$ represents the normalized output signal from the compressor, while, other parameters are as described with A-low companding.

The companding parameter μ is used to define the amount of compression. It is a dimensionless constant with standard value of 255. Figure 2.39-a shows the compression curves for several values of μ that determines the range of signal power in which the SN_qR is relatively constant. Note that the higher the μ, the more compression.

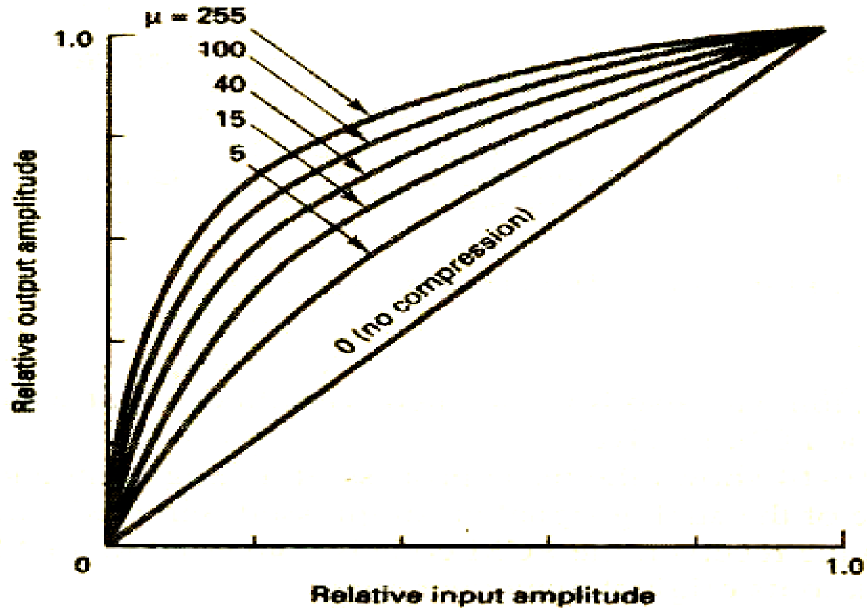


Figure 2.39-a μ-law compression characteristics

Voice transmission requires a minimum dynamic range of 40 dB and a seven-bit PCM code. For a relatively constant SN_qR and a 40-dB dynamic range, a $\mu \geq 100$ is required. The early Bell system PCM systems used a seven-bit code with a $\mu = 100$. However, the most recent PCM systems use an eight-bit code and a $\mu = 255$.

Example 2.7, for a compressor with $\mu = 255$, determine;

a- The voltage gain for the following relative values of V_{in}:

V_{max}, 0.75 V_{max}, 0.5 V_{max}, and 0.25 V_{max}.

b- The compressed output voltage for a maximum input voltage of 4V.

c- Input and output dynamic ranges and compression.

Solution; a- Substituting into equation 2.50-b, the following voltage gains are achieved for the given input amplitudes:

V_{in}	Compressed Voltage Gain
V_{max}	1.00
0.75 V_{max}	1.26
0.5 V_{max}	1.75
0.25 V_{max}	3.00

b- Using the compressed voltage gains determined in step (a), the output voltage is simply the input voltage times the compression gain:

V_{in}	Voltage Gain	V_{out}
V_{max} = 4 V	1.00	4.00
0.75 V_{max} = 3 V	1.26	3.78
0.50 V_{max} = 2 V	1.75	3.50
0.25 V_{max} = 1 V	3.00	3.00

c- Dynamic range is calculated by substituting into equation (2.35):

input dynamic range = 20 log 4/1 = 12 dB

output dynamic range = 20 log 4/3 = 2.5 dB

compression = input dynamic range - output dynamic range

= 12 dB – 2.5 dB = 9.5 dB

To restore the signals to their original proportions in the receiver, the compressed voltages are expanded by passing them through an amplifier with gain characteristics that are the complement of those in the compressor. For the values given in the example, the voltage gains in the receiver are as follows:

V_{in}	**Expanded Voltage Gain**
V_{max}	1.00
0.75 V_{max}	0.79
0.5 V_{max}	0.57
0.25 V_{max}	0.33

The overall circuit gain is simply the product of the compression and expansion factors which equals one for all input voltage levels. For the values given in the example;

V_{in}	**Overall Circuit Gain**
V_{max}	1 x 1 = 1
0.75 V_{max}	1.26 x 0.79 ≈ 1
0.5 V_{max}	1.75 x 0.57 ≈ 1
0.25 V_{max}	3.00 x 0.33 ≈ 1

The inverse or expansion characteristic for a μ-law compandor is defined as;

$$F_{\mu}^{-1}(y) = \text{sgn}(y)\left(\frac{1}{\mu}\right)[(1+\mu)^{|y|} - 1] \qquad (2.51)$$

where y = the compressed value, = $F_{\mu}(x)$ $(-1 \leq y \leq 1)$

sgn (y) = polarity of y

Digital Companding

Digital Companding involves digital compression in the transmitter such as; after the analog input signal is first sampled and converted to a linear PCM code and then the linear code is digitally compressed.
In the receiver the compressed PCM code expanded prior to PCM decoding (converted back to analog). Figure 2.40 shows the block diagram for a digitally companded PCM system.
The most recent digitally compressed PCM systems use a 12-bit linear PCM code and an eight-bit compressed PCM code.

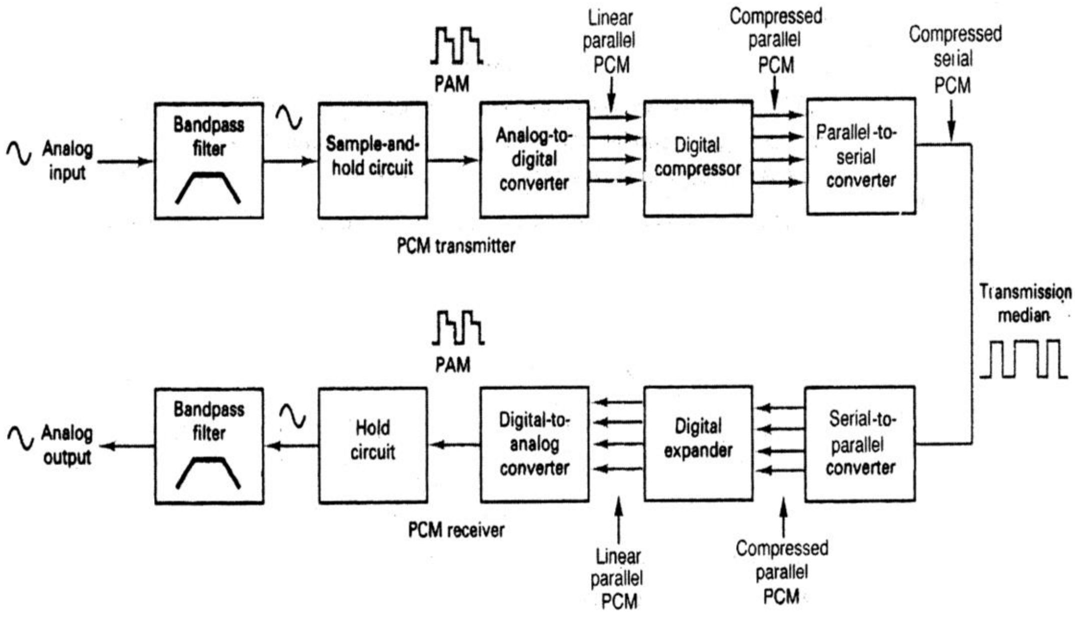


Figure 2.40 Digitally companded PCM system

The compression and expansion curves closely resemble the analog µ-law curves with a $\mu = 255$ by approximating the curve with a set of eight straight-line segments (segments 0 through 7) for the positive half of the analog signal. The slope of each successive segment is exactly one-half that of the previous segment.

Notice that the first portion of the A-law characteristic given by equation (2.48) is linear by definition. The remaining portion of the characteristic ($1/A \leq |x| \leq 1$) can be closely approximated by linear segments for easier implementation. In all, there are eight positive and eight negative segments, 16-segments in all. The first two segments of each polarity (four segments in all near the origin) are co-linear and therefore are sometimes considered as one straight-line segment. The A-law characteristic is normally implemented as a 13-segment piecewise linear approximation to equation (2.48) as illustrated in figure 2.37-b.

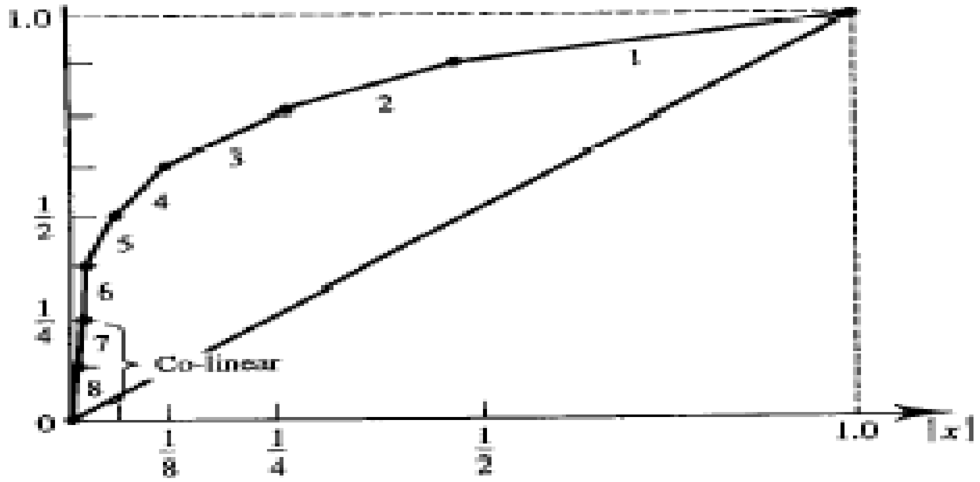


Figure 2.37-b 13-segment compression A-law realized by piecewise linear approximation.

Thus the segmented approximation of the A-law characteristic is sometimes referred to as a "13-segment approximation."

Figure 2.39-b shows the 12-bit-to-8-bit digital compression curve for positive values only.

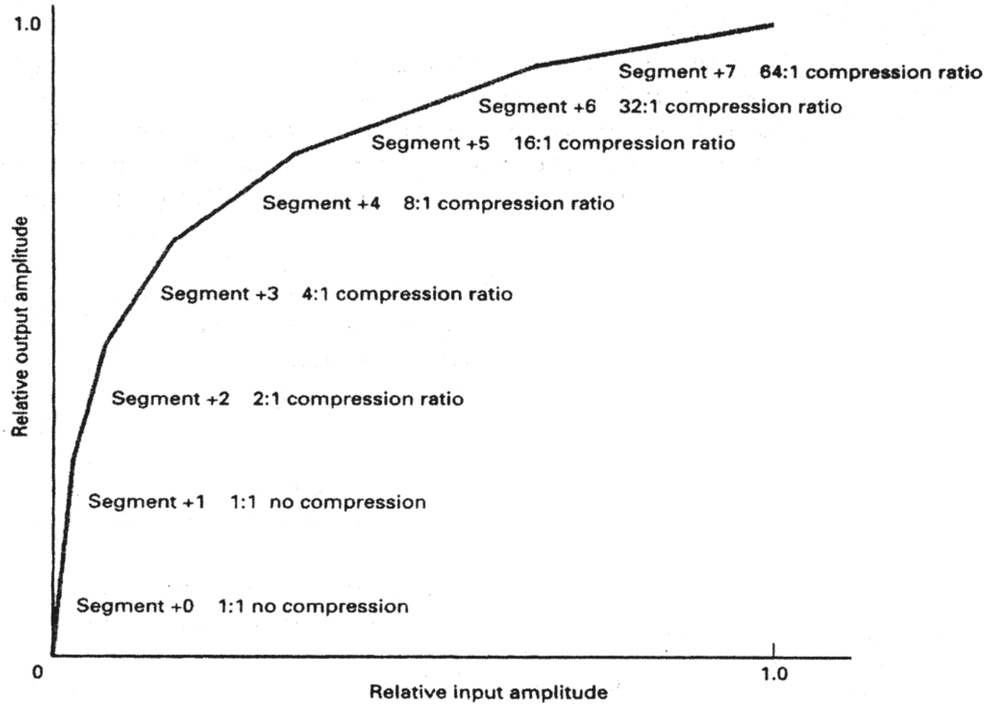


Figure 2.39-b µ255 compression characteristics (positive values only).

The curve for negative values is identical except the inverse. Although there are 16 segments (eight positive and eight negative), this scheme is often called 13-segment compression because the curve of segments +0, +1, -0, and -1 is a straight line with a constant slope and is considered as one segment. The performance of a µ-law approximation is presented in figure 2.39-c.

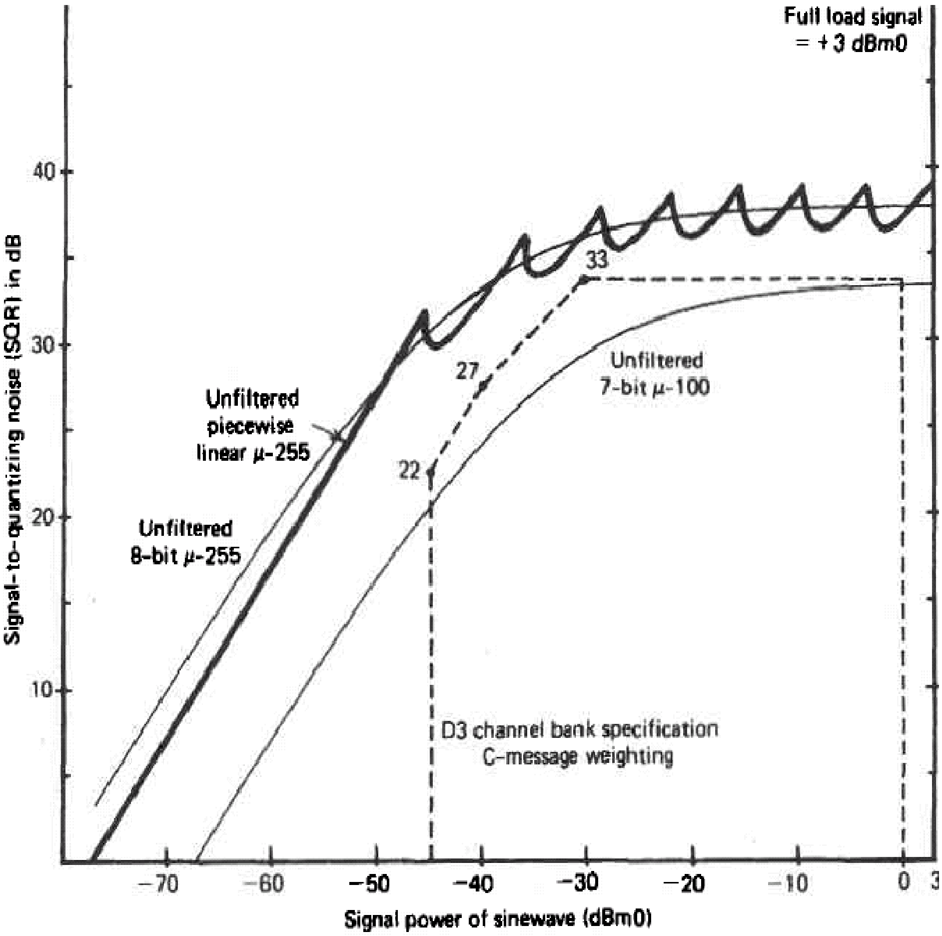


Figure 2.39-c SN_qR of μ-law coding with sine wave inputs.

Figure 2.41 displays the theoretical performance of the A-law approximation that compares it to the performance of a μ-law approximation.

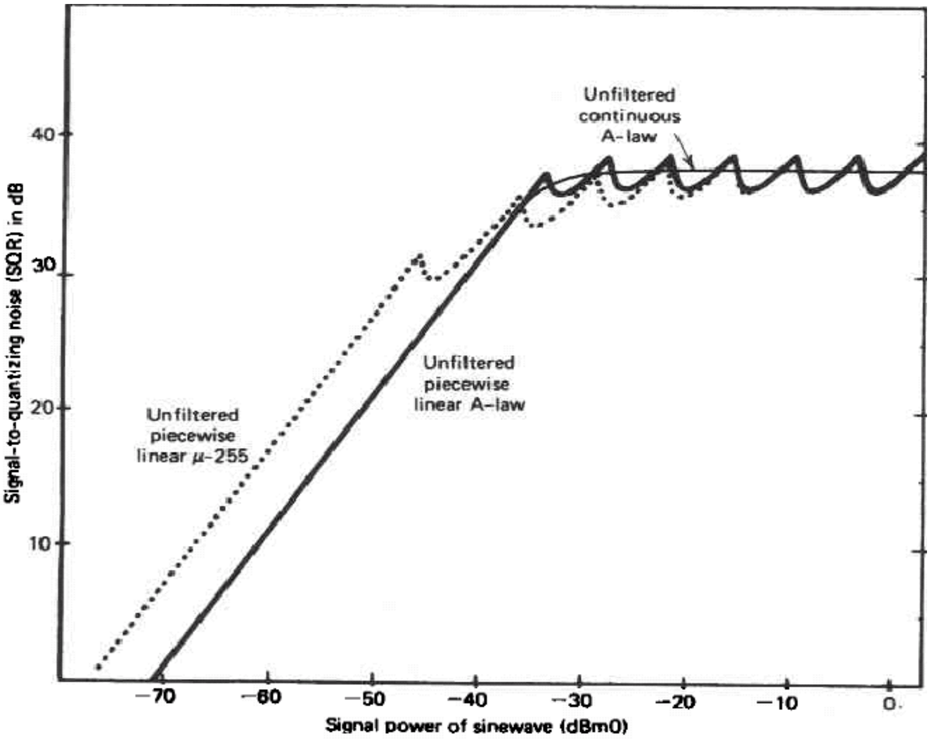


Figure 2.41 SN_qR of A-law PCM coding with sine wave inputs

For an intended dynamic range, the μ-law compander tends to give slightly improved SN_qR for voice signals when compared with the A-law but it has a slightly smaller dynamic range. In practice, like the A-law, the μ-law is usually implemented as a piecewise linear approximation.

A- and μ-law 64 k bit / sec companded PCM have been adapted by ITU-T as international toll quality standards (recommendation G.711) for digital coding of voice frequency signals. The sampling rate is 8 k Hz and the encoding law uses 8 binary digits per sample.

For communication between countries using different companding laws (one using A- and one using μ-) conversion from one to the other is the responsibility of the country using the μ-law.

2.3-2 Channel Bandwidth Requirements for Digital Transmission

Among other things, the above example highlights the relation between digital transmission and the required channel bandwidth.

For analog transmission of a B Hz signal, the channel bandwidth must be greater than or equal to B Hz. However, if the signal is digitized using k bits per sample, then the channel must be able to sustain a bit rate of 2×k×B or higher. For most of the binary modulation schemes discussed, this translates approximately to k×B Hz of channel bandwidth, thus requiring k times as much bandwidth as the analog transmission. This is the cost of digital transmission.

In communications via cables, this requirement is easily met these days by advancements in optical fiber manufacturing. In wireless communications, however, low bit rate digital schemes are sought after because the bandwidth is always at a premium.

Modulation schemes are investigated with high bandwidth efficiency. The bandwidth efficiency may be defined as the number of bits carried per Hz of the channel bandwidth. Combinations of MPSK and ASK usually provide good solutions for such systems.

2.3-3 Tradeoff between Channel Bandwidth and SN_qR

Consider a sinusoidal modulating message signal of bandwidth

W Hz that requires a minimum sampling rate of 2W sample per second, with each sample represented by an n-bits code word. The bit duration time T_b has a maximum value of;

$$T_s = 1 / f_s = 1/ 2W = nT_b,$$

So, $$T_b = 1 / 2 n W \qquad (2.52)$$

The channel bandwidth B required to transmit a pulse of this duration is given by; $B = 2 n W$

Using equations (2.28, 2.29), for SN_qR using sinusoidal modulating signal with amplitude E_m. The quantizer step size will be;

$q = 2 E_m / M$

The average quantizing noise power is;

$$N_q = q^2 / 12 = (2 E_m/M)^2 / 12 = E_m^2 / 3 M^2$$

Thus SN_qR of the PCM will be; $SN_qR = 3 M^2 / 2$

So, SN_qR increases with M^2.

$SN_qR\big|_{dB} = 10 \log_{10} SN_qR = 1.8 + 20 \log_{10} M$

Then SN_qR increases logarithmly with M.

as $M = 2^n$, so;

$$SN_qR = 3 (2^n)^2 / 2 = 3 \times 4^n / 2 = (3 / 2)\, 2^{(B/W)} \qquad (2.53)$$

This relation indicates that the PCM system is capable of improving SN_qR exponentially with bandwidth expansion ratio B/W.

2.4 Differential PCM, DPCM

In conventional PCM, there are often successive samples taken in which there is little difference between the amplitudes of the two samples. This necessitates transmitting several identical PCM codes, which is redundant. Differential Pulse Code Modulation (DPCM) is designed specifically to take advantage of the sample-to-sample redundancies in a typical speech waveform.

With DPCM the difference in the amplitude of two successive samples is transmitted rather than the actual sample. Thus the band limiting filter in the encoder and the smoothing filter in the decoder are basically identical to those used in conventional PCM systems.

The DPCM structure shown in figure 2.49 is more complicated.

Since the range of sample differences is less than the range of individual samples, fewer bits are needed to encode difference samples and thus lower the transmission rate. The sampling rate is often the same as for a comparable PCM system. Thus the band limiting filter in the encoder and the smoothing filter in the decoder are basically identical to those used in conventional PCM systems.

The DPCM structure shown in figure 2.42 is more complicated. The band limiting filter is used to limit the analog input signal frequency to one-half the sampling rate. A conceptual means of generating the difference samples for a DPCM coder is to store the previous input sample directly in a sample-and-hold circuit and use an analog differentiator (subtractor) to measure the change. The change in the signal is then quantized and encoded for transmission. However, because the previous input value is reconstructed by a feed back loop that integrates the encoded sample differences. In essence the feedback signal is an estimate of the input signal as obtained by integrating the encoded sample differences.

Thus the feedback signal is obtained in the same manner used to reconstruct the waveform in the decoder.

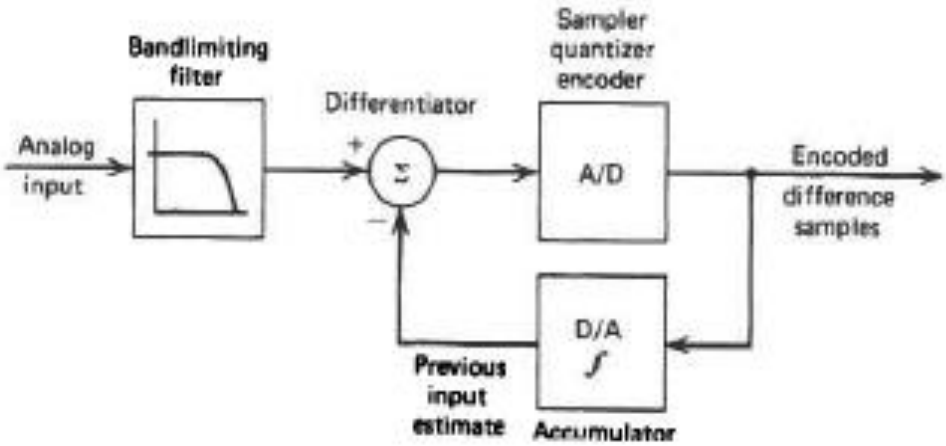


Figure 2.42 Functional block diagram of DPCM

The advantage of the feed back implementation is that quantization errors do not accumulate indefinitely. If the feed back signal drifts from the input signal, as a result of an accumulation of quantization errors, the next encoding of the difference signal automatically compensates for the drift. In a system without feed back the output produced by a decoder at the other end of the connection might accumulate quantization errors without bound.

As in PCM systems the analog-to-digital conversion process can be uniform or companded. Some DPCM systems also use adaptive techniques (syllabic companding) to adjust the quantization step size in accordance with the average power level of the signal.

Example 2.6, speech digitization techniques are sometimes measured for quality by use of an 800-Hz sine wave as a representative test signal. Assuming a uniform PCM system is available to encode the sine wave across a given dynamic range. Determine how many bits per sample can be saved by using a uniform DPCM system.

A basic solution can be obtained by determining how much smaller the dynamic range of the difference signal is in comparison to the dynamic range of the signal amplitude. Assume the maximum amplitude of the sine wave is A, so that;

$$X(t) = A \, \mathrm{Sin}\,(2\pi \,.\, 800\, t)$$

The maximum amplitude of the difference signal can be obtained by differentiating and multiplying by the time interval between samples:

$$\frac{dx}{dt} = A(2\pi)(800)\,Cos\,(2\pi 800 t)$$

$$\left|\Delta x(t)\right| \max = A(2\pi)(800)\left(\frac{1}{8000}\right) = 0.628A$$

The saving in bits per sample can be determined as;

$$\mathrm{Log}_2\,(1 / 0.628) = 0.67 \text{ bits}$$

Example 2.6 demonstrates that a DPCM system can use 2/3 bit per sample less than a PCM system with the same quality. Typically DPCM systems provide a full l-bit reduction in code word size. The larger savings is achieved because on average speech waveforms have a lower slope than an 800-Hz tone.

Figure 2.43 shows a simplified block diagram of a DPCM transmitter. The analog input signal is band limited to one-half the sample rate, then compared with the preceding accumulated signal level in the differentiator. The output of the differentiation is the difference between the two signals. The difference is PCM encoded and transmitted. The ADC operates the same as in a conventional PCM system, except that it typically uses fewer bits per sample.

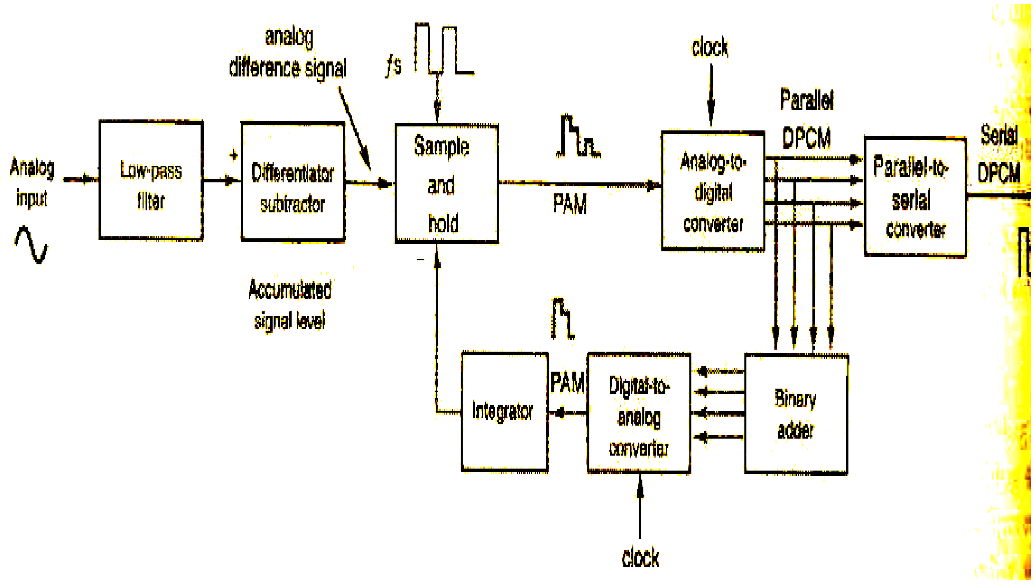


Figure 2.43 a simplified block diagram of a DPCM transmitter

Figure 2.44 shows a simplified block diagram of a DPCM receiver. Each received sample is converted back to analog, stored, and then summed with the next sample received.

In the receiver shown in figure 2.44, the integration is performed on the analog signals, although it could also be performed digitally.

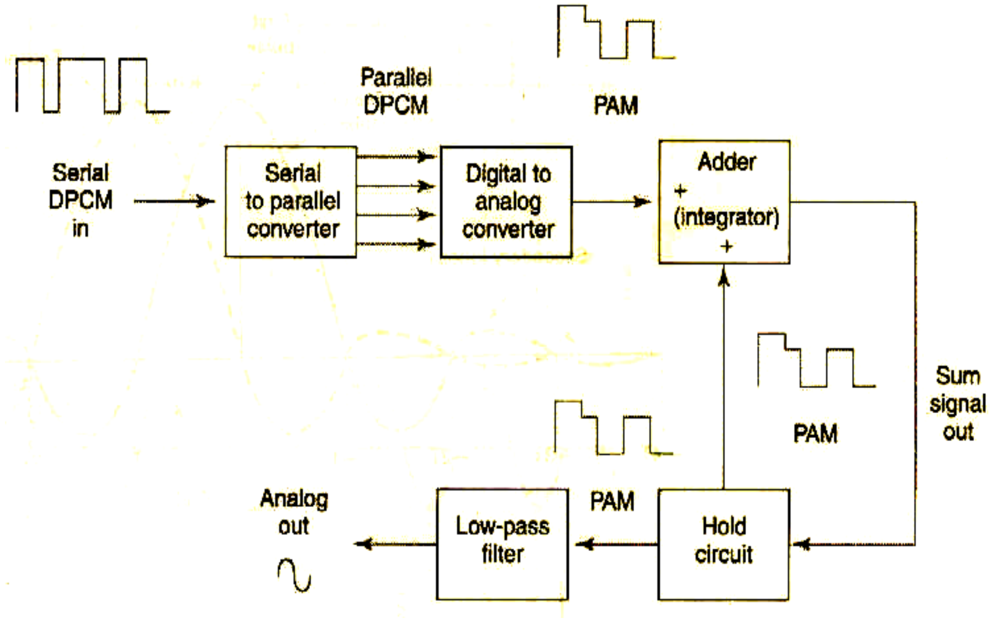


Figure 2.44 a simplified block diagram of a DPCM receiver

2.4-1 DPCM Implementations

Differential PCM encoders and decoders can be implemented in a variety of ways depending on how the signal processing functions are partitioned between analog and digital circuitry. At one extreme the differencing and integration functions can be implemented with analog circuitry, while at the other extreme all signal processing can be implemented digitally using conventional PCM samples as input. Figure 2.45 shows block diagrams of the two extremes of DPCM implementation with differing amounts of digital signal processing. Figure 2.45-a depicts a system using analog differencing and integration. Analog-to- Digital conversion is performed on the difference signal, and D/A conversion for the feed back loop is immediately performed on the limited-range difference code. Analog summation and storage in a sample-and-hold (S/H) circuit is used to provide integration. Notice that these D/A converters convert a limited difference signal. Figure 2.45-b shows a system where all signal processing is performed by digital logic circuits. The A/D converter produces full-amplitude-range sample codes which are compared to digitally generated approximations of the previous amplitude code. Notice that the A/D converter in this case must encode the entire dynamic range of the input where as the A/D converter in the other version operates on only the difference signals.

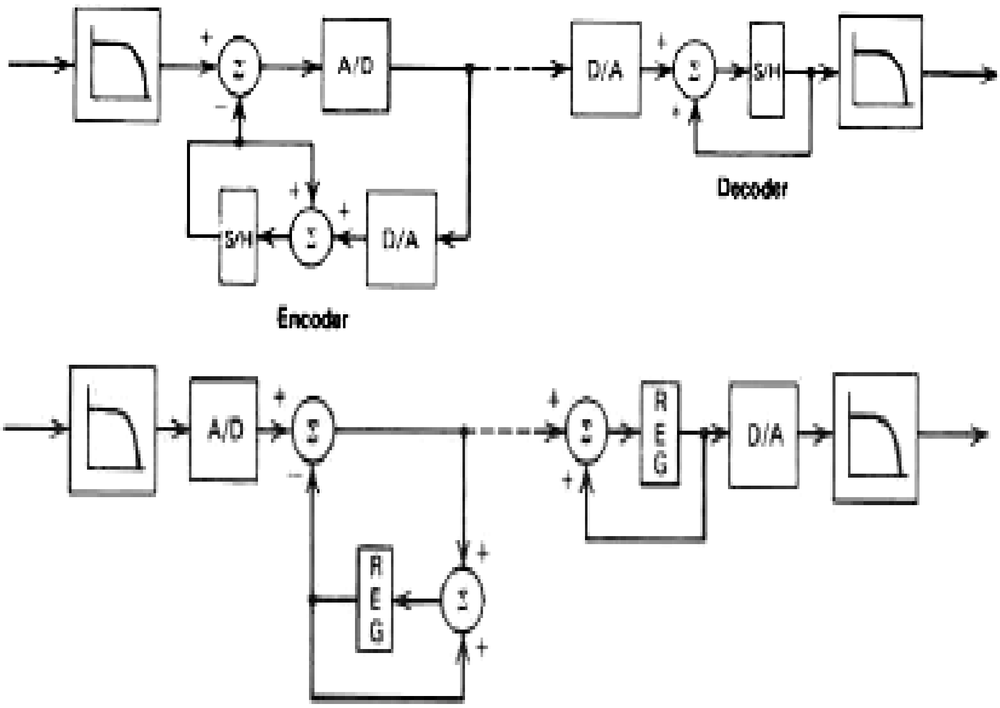


Figure 2.45 DPCM implementations: (a) analog integration; (b) digital differencing

Due to the availability of digital signal processing components some of which contain internal A/D converters digital processing as shown in figure 2.45-b is generally the most effective means of implementing a DPCM algorithm.

In fact, most DPCM applications involve processing speech signals that have already been digitized into analog processing as an aid in processing log-PCM signals some DSP components provide internal μ-law and A-law conversion functions. The decoders in the two implementations shown in figure 2.50 are exactly like the feed back implementations in the corresponding encoder. This reinforces the fact that the feed back loop generates an approximation of the input signal (delayed by one sample).

If no channel errors occur, the decoder output (before filtering) is identical to the feedback signal. Thus the closer the feedback signal matches the input, the closer the decoder output matches the encoder input.

2.4-2 Higher Order Prediction

DPCM encoder is considered as a special case of a linear predictor with encoding and transmission of the prediction error. The feed back signal of a DPCM system represents first-order prediction of the next sample value, and the sample difference is a prediction error. Under this view point the DPCM concept can be extended to incorporate more

than one past sample value into the prediction circuitry.
Thus the additional redundancy available from all previous samples can be weighted and summed to produce a better estimate of the next input sample. With a better estimate the range of the prediction error decreases to allow encoding with fewer bits. For systems with constant predictor coefficients, results have shown that most of the realizable improvement occurs when using only the last three sample values, as shown in figure 2.46. For conceptual purposes this implementation shows analog differencing and integration as in figure 2.45-a.The most effective implementations use digital memory, multiplication, and addition in a DSP component in lieu of the analog processing shown, particularly because most applications involve already digitized (PCM) signals.

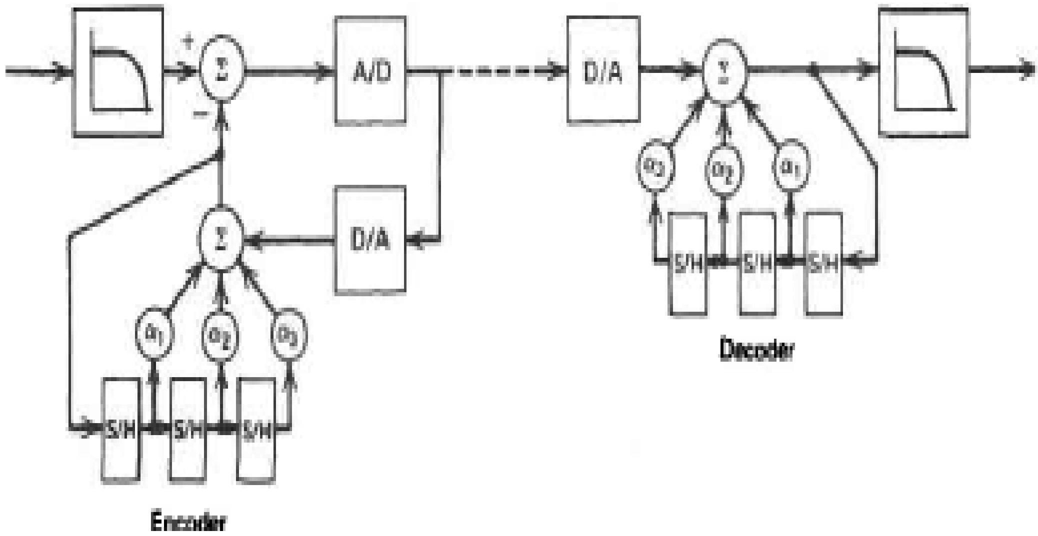


Figure 2.46 Extension of DPCM to third-order prediction.

As mentioned in section 2.4, analysis of differential PCM systems with first-order predication typically provides a 1-bit-per-sample reduction in code length relative to PCM systems with equivalent performance. Extended DPCM systems utilizing third order prediction can provide reductions of 2 bits per sample. Thus a standard DPCM system can provide 64-kbps PCM quality at 56 kbps, and third-order linear Prediction can provide comparable quality at 48 kbps. However, some subjective evaluation have indicated that somewhat higher bit rates are needed to match 64-kbps PCM quality.

2.4-3 Adaptive Differential PCM, ADPCM

Relatively straight forward implementation of DPCM can provide savings of 1.5 - 2 bits per sample with respect to standard PCM encoding. Even greater savings can be achieved by adding adaptation logic to the basic DPCM algorithm to create what is referred to as adaptive differential PCM (ADPCM).

Many forms of ADPCM have been investigated and used in various applications. Two of the most prevalent applications are voice messaging and DCM equipment for increasing the number of voice channels on a Tl line. With respect to the latter application, ITU-T has established a 32-kbps ADPCM standard (Recommendation G.721). This algorithm has been extensively tested and characterized to not significantly degrade toll quality voice circuits when inserted into the internal portions of the network. Design considerations of the standard are:

1. Multiple tandem encoding and decoding between both PCM and analog interfaces.
2. End-to-end signal quality for voice, voice band data, and facsimile
3. Effects of random and bursty channel errors
4. Performance on analog signals degraded by loss, noise, amplitude distortion, phase distortion, and harmonic distortion
5. Easy trans-coding with μ-law and A-law PCM

The 32-kbps rate implies a 2 : 1 savings in channel bandwidth with respect to standard PCM. A significant impairment introduced by implementations of the ADPCM standard is the corruption of modem signals carrying data rates greater than 4800 bps. Voice band data at rates of 4800 bps and below are adequately supported.

The G.721 ADPCM algorithm is conceptually similar to that shown in figure 2.46 but more sophisticated in that it uses an eighth-order predictor, adaptive quantization, and adaptive prediction. Furthermore, the algorithm is designed to recognize the difference between voice or data signals and use a fast quantizer adaptation mode for voice and a slow adaptation mode for data. Subjective evaluation of the G.721 algorithm using the mean opinion score (MOS) method of evaluating speech quality is shown in figure 2.47.

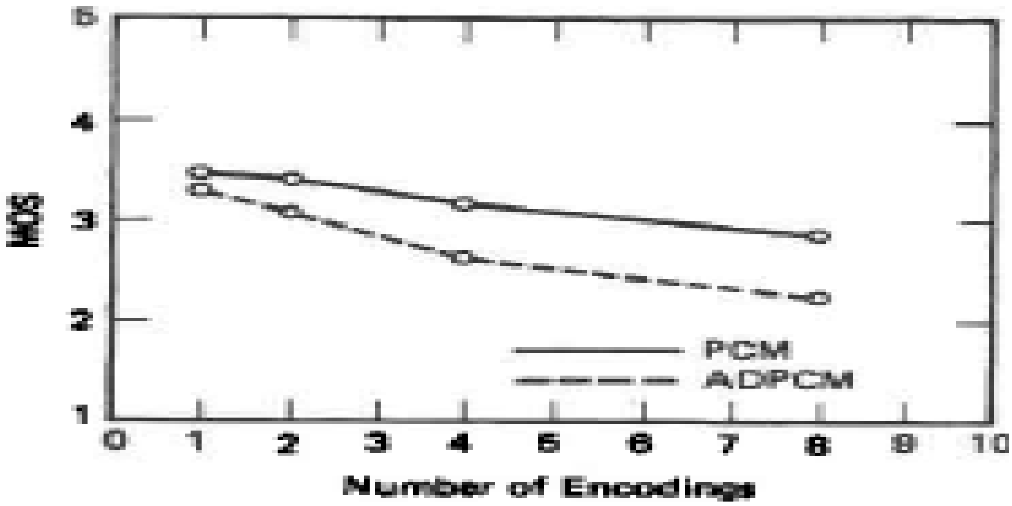


Figure 2.47 Average MOS versus number of encodings for PCM and ADPCM for carbon microphone.

The MOS method uses trained listeners to evaluate the speech quality on a scale of 1:5. Figure 2.47 shows the average scores of 32-kbps ADPCM and 6 4-kbps PCM as a function of the number of analog tandem encodings.

Because ADPCM at 32 kbps provides good quality at a moderate cost and power consumption, it is used in several cordless telephone or low-tier-digital cellular systems those utilize simple, low-power mobile units, and small cells and only supports pedestrian speeds:

Personal Access Comm. System (PACS) (North America)

Second Generation Cordless Telephones (CT2) (Europe)

Digital European Cordless Telephones (DECT) (Europe)

Personal Handy phone System (PHS) (Japan)

2.5 Delta Modulation, DM

Delta modulation (DM) is another digitization technique that specifically exploits the sample-to-sample redundancy in a speech waveform. In fact, DM can be considered as a special case of DPCM using only 1 bit per sample of the difference signal to achieve digital transmission of analog signals. The single bit specifies merely the polarity of the difference sample and thereby indicates whither the signal has increased (the sample is larger than) or decreased (the sample is smaller than) since the last (previous) sample.

An approximation to the input wave form is constructed in the feed back path by stepping up one quantization level when the difference is positive ("one") and stepping down one quantization level when the difference is negative ("zero"). If the current sample value is equal to the previous one, code the first such occurrence opposite to the previous bit and then alternate to ‘0’ and ‘1’ for later occurrences.

In this way the input signal is encoded as a sequence of 'ups" and "downs" in a manner resembling a staircase. Figure 2.48 shows a DM approximation of a typical waveform that does not need k bits per sample as in PCM. One bit per sample will suffice.

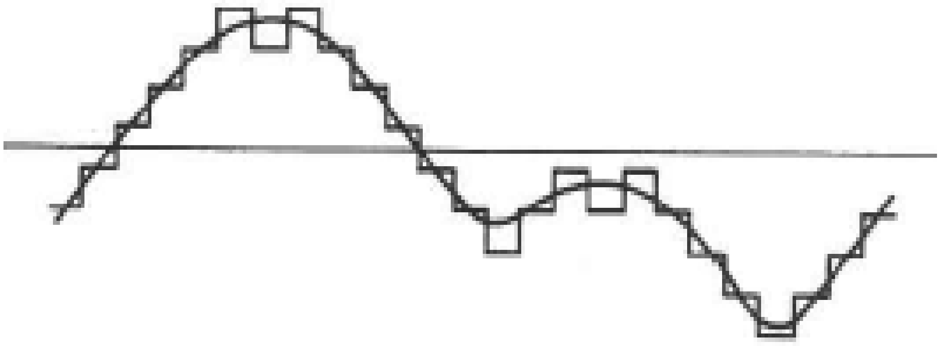


Figure 2.48 Waveform encoding by delta modulation

Notice that the feedback signal continues to step in one direction until it crosses the input, at which time the feedback step reverses direction until the input is crossed again. Thus, when tracking the input signal, the DM output "bounces" back and forth across the input waveform, allowing the input to be accurately reconstructed by a smoothing filter. The main attraction of DM is its simplicity.

Since each encoded sample contains a relatively small amount of information (1 bit), DM systems require a higher sampling rate than PCM or multi-bit DPCM systems for comparable quality of speech. From another viewpoint, "over sampling" is needed to achieve better prediction from one sample to the next. Consequently, the advantage in saving bandwidth in sampling is lost in rendering the quality equal to the PCM.

2.5-1 Delta Modulation Transmitter

Figure 2.49 shows a block diagram of a delta modulation transmitter. The input analog is sampled and converted to a PAM signal, which is compared with the output of the DAC. The output of the DAC is a voltage equal to the regenerated magnitude of the previous sample, which was stored in the up-down counter as a binary number.

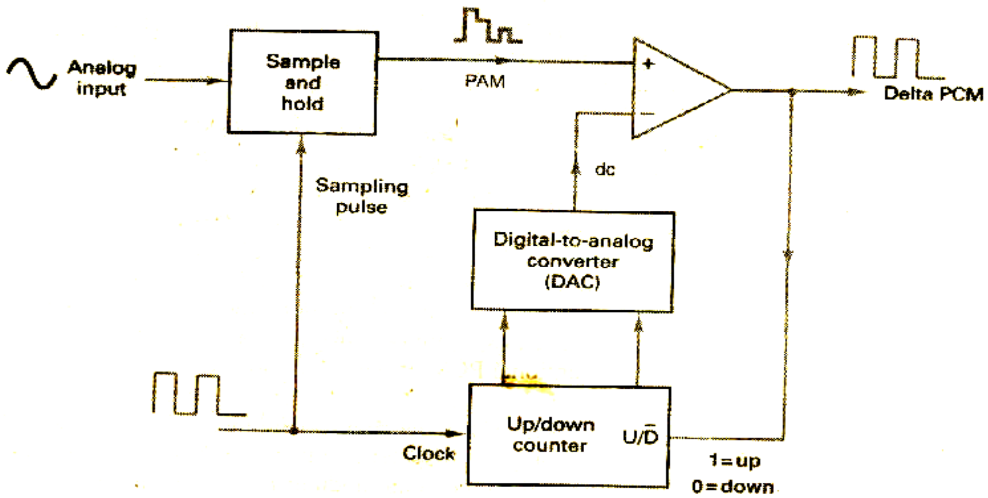


Figure 2.49 Delta modulation transmitter

The up-down counter is incremented or decremented depending on whether the previous sample is larger or smaller than the current sample. The up-down counter is clocked at a rate equal to the sample rate. Therefore, the up-down counter is updated after each comparison. Figure 2.50 shows the ideal operation of a delta modulation encoder. Initially, the up-down counter is zeroed, and the DAC is outputting 0 V. The first sample is taken, converted to a PAM signal, and compared with zero volts. The output of the comparator is a logic 1 condition (+V) indicating that the current sample is larger in amplitude than the previous sample. On the next clock pulse, the up-down counter is incremented to a count of 1.

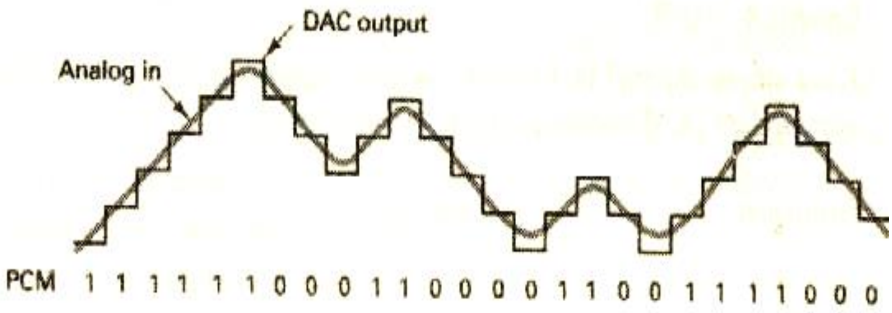


Figure 2.50 Ideal operation of a delta modulation encoder

The DAC now outputs a voltage equal to the magnitude of the minimum step size (resolution), The steps change value at a rate equal to the clock frequency (sample rate). Consequently, with the input signal shown, the up-down counter follows the input analog signal up until the output of the DAC exceeds the analog sample; then the up-down counter will begin counting down until the output of the DAC drops below the sample amplitude. In the idealized situation shown in figure 2.50, the DAC output follows the input signal. Each time the up-down counter is incremented, a logic1 is transmitted, and each time the up-down counter is decremented, a logic 0 is transmitted.

2.5-2 Delta Modulation Receiver

Figure 2.51 shows the block diagram of a delta modulation receiver. As you can see, the receiver is almost identical to the transmitter except for the comparator. As the logic 1s and 0s are received, the up-down counter is incremented or decremented accordingly. Consequently, the output of the DAC in the decoder is identical to the output of the DAC in the transmitter.

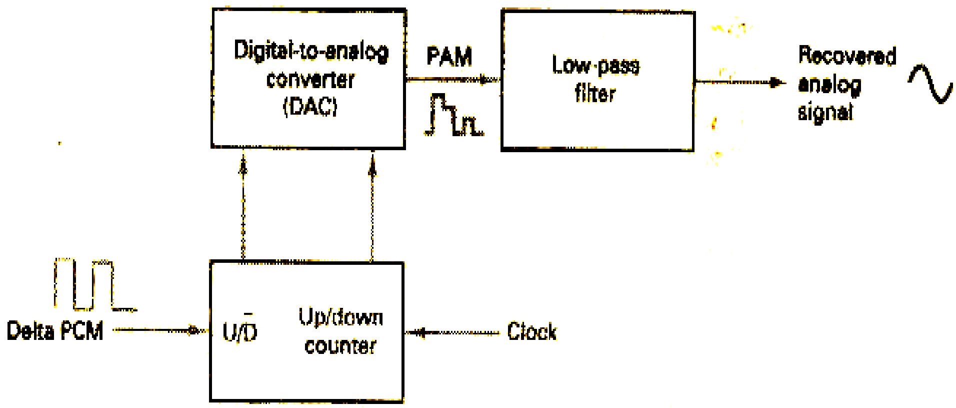


Figure 2.51 Delta modulation receiver

With delta modulation, each sample requires the transmission of only one bit; therefore, the bit rates associated with delta modulation are lower than conventional PCM systems. However, there are two problems associated with delta modulation that do not occur with conventional PCM: slope overload and granular noise.

A. Slope overload

Basically, slope overload distortion occurs for large and fast signal transition, when the slope (rate of change) of the analog input signal exceeds the maximum rate of change that can be generated by the feedback loop. Figure 2.52 shows what happens when the analog input signal changes at a faster rate than the DAC can maintain. Since the maximum rate of change in the feedback loop is merely the step size times the sampling rate, a slope overload condition occurs if;

$$\left|\frac{d\,x(t)}{dt}\right| > q\,f_s \tag{2.54}$$

The relatively high sampling rate of a delta modulator (increasing the clock frequency) produces a wider separation of these spectrums and hence, fold over distortion is prevented with less stringent roll-off requirements for the input filter. It reduces also the probability of slope overload occurring. Another way to prevent slope overload is to increase the magnitude of the minimum step size.

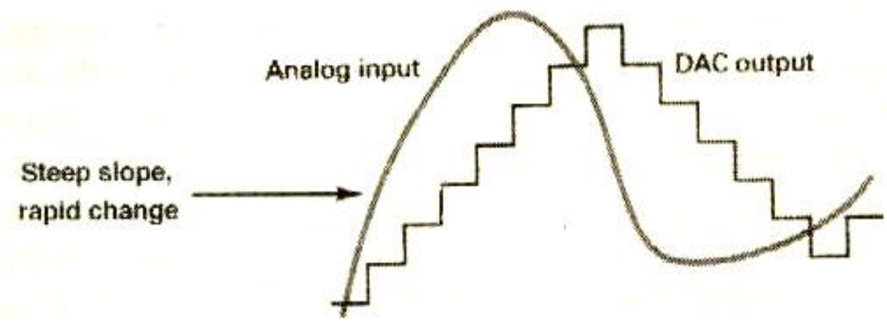


Figure 2.52 Slope overload distortion

Slope overload is not a limitation of just a DM system, but an inherent problem with any system, such as DPCM in general, that encodes the difference in a signal from one sample to the next. A difference system encodes the slope of the input with a finite number of bits and hence a finite range. If the slope exceeds that range, slope overload occurs. In contrast, a conventional PCM system is not limited by the rate of change of the input, only by the maximum encodable amplitude. Notice that a differential system can encode signals with arbitrarily large amplitudes, as long as the large amplitudes are attained gradually.

B. Granular noise

Figure 2.53 contrasts the original and reconstructed signals associated with a granular noise in delta modulation system. It can be seen that when the original analog input signal has relatively constant amplitude, the reconstructed signal has variations that were not present in the original signal. Granular noise in delta modulation is analogous to quantization noise in conventional PCM.

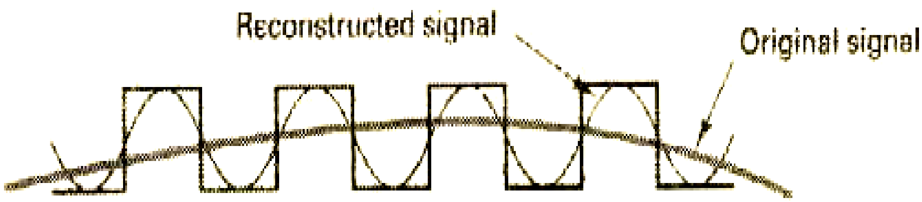


Figure 2.53 Granular noise

Granular noise can be reduced by decreasing the step size. Granular noise is more prevalent in analog signals that have gradual slopes and whose amplitudes vary slowly by small amounts. Slope overload is more prevalent in analog signals that have steep slopes or whose amplitudes vary rapidly.

Therefore, to reduce the granular noise, a small resolution is needed, and to reduce the possibility of slope overload occurring, a large resolution is required. Obviously, a compromise is necessary, where the design of a DM (or DPCM) necessarily involves a trade-off between the two types of distortion. The optimum DM step size in terms of minimizing the total of the granular and the slope overload noise is required. The perceptual effects of slope overload on the quality of a speech signal are significantly different from the perceptual effects produced by granular noise.

As indicated in figure 2.54, the slope overload noise reaches its peaks just before the encoded signal reaches its peaks. Hence, slope overload noise has strong components identical in frequency and approximately in phase with a major component of the input. In fact, overload noise is much less objectionable to a listener than random or granular noise at an equivalent power level. Hence, from the point of view of perceived speech quality, the optimum mix of granular and slope overload noise is difficult to determine.

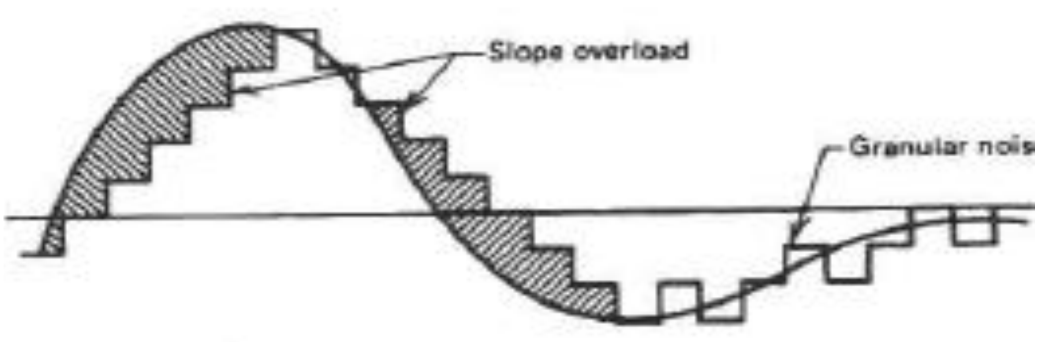


Figure 2.54 Slope overload and granular noise of DM system

Many versions of DM for voice encoding focused on ways of implementing adaptive delta modulation (ADM) to improve the performance at a given bit rate. The intense interest at that time was related to the simplicity, good tolerance of channel errors, and relatively low cost implementation. The cost factor is no longer relevant because even relatively complicated coding algorithms now have insignificant costs compared to most system costs. ADM is still used in some old PBXs, in some military secure voice radio systems, and as a means of encoding the residual error signal of some predictive coders.

2.5-3 Adaptive Delta Modulation, ADM

Adaptive delta modulation is a delta modulation system where the step size of the DAC is automatically varied, depending on the amplitude characteristics of the analog input signal. Figure 2.55 shows how an adaptive delta modulator works. When the output of the transmitter is a string of consecutive 1s or 0s, this indicates that the slope of the DAC output is less than the slope of the analog signal in either the positive or the negative direction, Essentially, the DAC has lost track of exactly where the analog samples are, and the possibility of slope overload occurring is high.

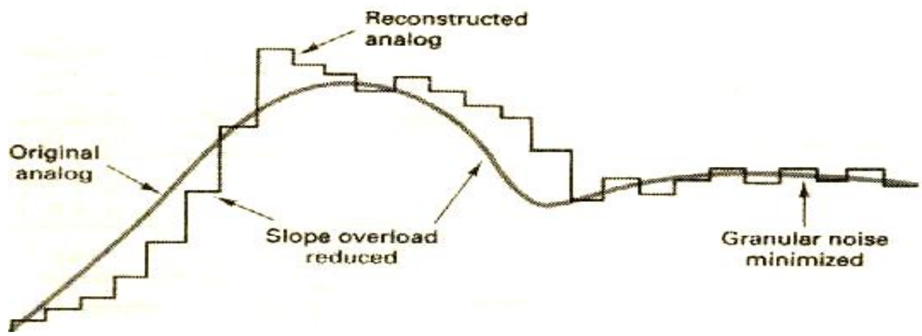


Figure 2.55 Adaptive delta modulation

With an adaptive delta modulator, after a predetermined number of consecutive 1s or 0s, the step size is automatically increased. After the next sample, if the DAC output amplitude is still below the sample amplitude, the next step is increased even further until eventually the DAC catches up with the analog signal. When an alternative sequence of 1s and 0s is occurring, this indicates that the possibility of granular noise occurring is high. Consequently, the DAC will automatically revert to its minimum step size and, thus, reduces the magnitude of the noise error.

A common algorithm for an adaptive delta modulator is when three consecutive 1s or 0s occur, the step size of the DAC is increased or decreased by a factor of 1.5. Various other algorithms may be used for adaptive delta modulators, depending on particular system requirements.

The differential systems described in the previous sections (DPCM, ADPCM, ADM) operate with lower data rates than PCM systems because they encode a difference signal that has lower average power than the raw input signal. The ratio of the input signal power to the power of the difference signal is referred to as the prediction gain. Simple DPCM systems (first-order predictors) provide about 5 dB of prediction gain.

ADPCM provides greater levels of prediction gain depending on the sophistication of the adaptation logic and the number of past samples used to predict the next sample.

The prediction gain of ADPCM is ultimately limited by the fact that only a few past samples are used to predict the input and the adaptation logic only adapts the quantizer, not the prediction weighting coefficients (the α's in figure 2.46).

Chapter

3

Multiplexing Techniques

The installation of communication systems is very costly, and the costliest element is the transmission medium. Quite clearly great economics would result if a single transmission medium can carry several signals all combined together.

Multiplexing is the technique used to combine a number of signals and send them over the medium to make the best use of the transmission medium and ensure that its bandwidth is utilized to its full capacity.

So, it is economically feasible to utilize the available bandwidth of optical fiber or coaxial cable or a radio system in a single high-capacity system shared by multiple users. In order for the signals to be received independently they must be sufficiently separated in some sense.

This quality of separateness is usually called orthogonality. Orthogonal signals can be received independently of each other whilst non-orthogonal signals cannot. There are many ways in which orthogonality between signals can be provided.

Two approaches to multiplexing are analog, or Frequency-Division Multiplexing, (FDM); and digital, or Time-Division Multiplexing, (TDM). The actual equipment that performs the multiplexing is called a channel bank. Analog or A-type channel banks perform analog multiplexing; digital or D-type channel banks perform digital multiplexing. In FDM, the frequency band of the system is divided into several narrowband channels, one for each user all the time. Use of multiplexing technique is possible if the bandwidth of the channel is higher than the bandwidth of the individual data sources to facilitate good utilization of the channel bandwidth. In TDM, the transmission time of the system is divided into several narrow time slot channels, one for each user that uses the total system bandwidth. Use of multiplexing technique is possible if the capacity of the channel is higher than the data rates of the individual data sources to facilitate good utilization of the channel capacity.

Despite the need to convert the voice signals to a digital format at one end of a Tl line and back to analog at the other, the combined conversion and multiplexing cost of a digital TDM terminal was lower than the cost of a comparable analog FDM terminal.

3.1 Analog Multiplexing

The traditional way of providing orthogonality in analogue telephony and audio/video broadcast applications is to transmit different information signals using different carrier frequencies. Such combined signals are disjoint (non-overlapped) in frequency and can be received separately using filters. Using different carriers, to isolate signals from each other in FDM technique, number of signals from different sources are translated into different frequency bands at the transmitting side and sent over the same transmission medium by using them to modulate the carrier signals with different and appropriate frequency so that they do not interfere with each other.

FDM was the original multiplexing technique for analogue communications and is now experiencing a resurgence in fiber optic systems in which different wavelengths are used for simultaneous transmission of many information signals. In FDM telephony, 300 Hz – 3.4 kHz bandwidth telephone base band signals, are stacked in frequency at 4 kHz spacing with small frequency guard bands between them to allow signals separation using practical filters. Figure 3.1 shows an example of a communication system in which the signals from three data sources can be combined (multiplexed) together and sent through a single transmission medium. At the receiving end, the signals are separated (de-multiplexed) using a bank of filters.

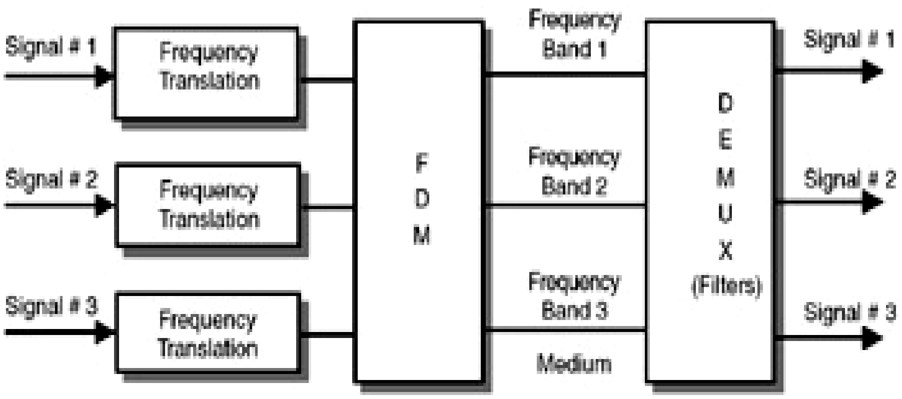


Figure 3.1 Multiplexing and De-multiplexing.

Figure 3.2 shows how an FDM signal can be generated.

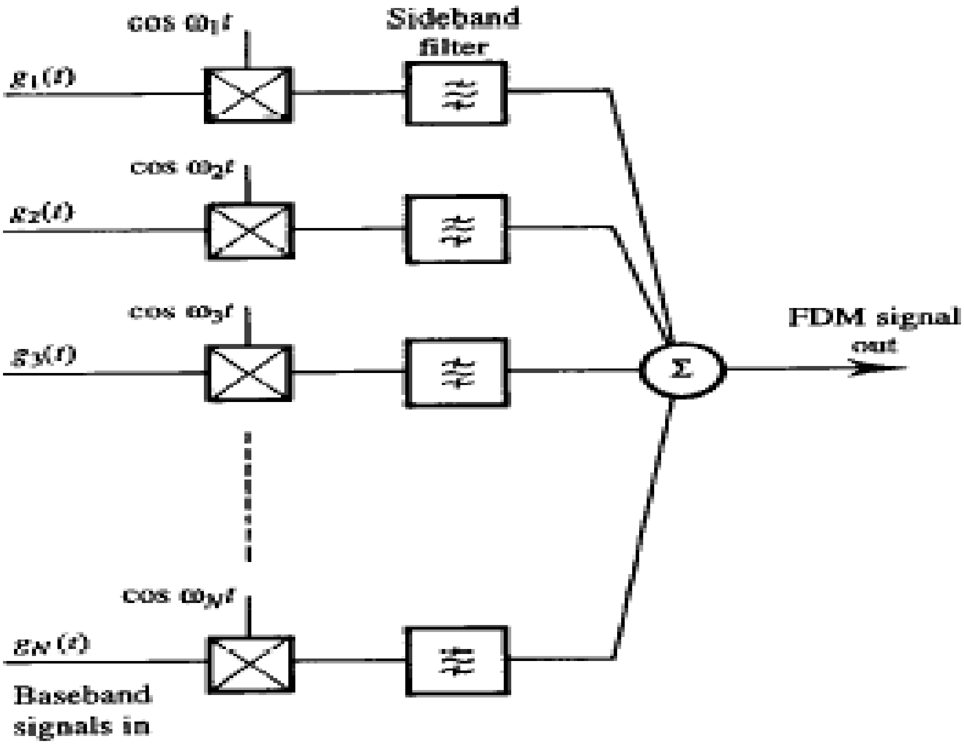


Figure 3.2 Generation of an FDM signal.

3.1-1 FDM Hierarchy

Actual FDM is accomplished as a multilevel process. FDM hierarchy multiplex 12- voice signals each of 4 k Hz together to create a Basic group as one of the fundamental unit for marketing. Five-groups are multiplexed together to create a Super group. Ten Super groups multiplexed together give a Master group. Six Master groups multiplexed together give a Jumbo group. Three Jumbo groups multiplexed together give a Jumbo group multiplex.

3.1-1.1 Formation of a Basic Group

In the trunk or toll system, 12 channels form a Basic group. The Basic group is formed by SSB-SC amplitude modulation of 12 sub-carriers at 64, 68, 72; . . . ; 108 kHz. This modulation technique is used to beat the corrupting influence of noise on the information content of the transmission as we put as much as possible, if not all, of the available power into one of the sidebands. An added advantage to this scheme is that the required bandwidth is reduced to one-half of its original value. Clearly, this would allow twice as many messages to be sent on the same channel as before. The price to be paid for this advantage is that to demodulate an SSB signal, it is necessary to reinstate the carrier at the receiver. The reinstated carrier has to be in synchronism with the original carrier, otherwise demodulation yields an intolerably distorted signal. Providing a synchronized local oscillator requires complex equipment at the transmitter as well as at the receiver. In SSB radio, an attenuated form of the carrier is transmitted with the signal. This is used to synchronize a local oscillator in the receiver. In the telephone system, a centrally generated pilot signal is distributed to all offices for demodulation purposes. In some cases, a local oscillator without synchronization is used. If the frequency error is small (approximately -5 Hz), successful demodulation can be achieved.

The required carriers are generated from a 4 kHz crystal-controlled oscillator and multiplied by the appropriate factor. The upper sidebands

are removed and the lower side bands are added together to form the Basic group. Figure 3.3-a shows a block diagram for channel 1. Figure 3.3-b shows a schematic representation of the spectrum of the Basic group that covers the frequency range from 60 kHz to 108 kHz.
For a small-capacity trunk, the Basic group may be transmitted without further processing. The transmission channel can be a twisted pair or coaxial cable.

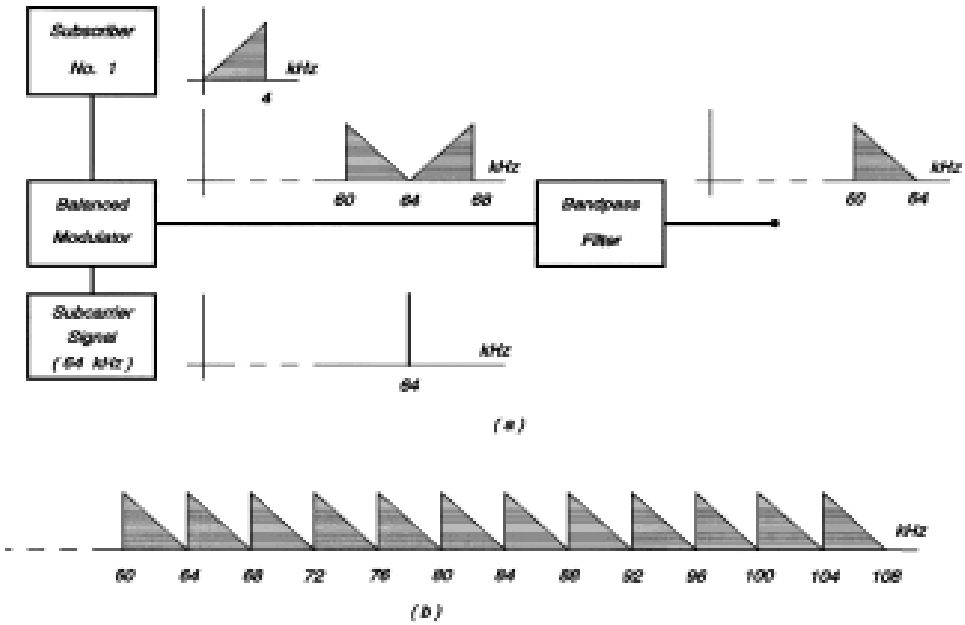


Figure 3.3 Formation of Basic group with its Spectrum

3.1-1.2 Formation of a Super Group

For higher capacity channels, five Basic groups are combined to form a Super group. Figure 3.4-a shows the block diagram of the Super group 1. Note that to make the filtering problem easier, the carrier frequency is chosen to be 420 kHz. Figure 3.4-b shows the frequency spectrum of the Super group that occupying the frequency range from 312 kHz to 552 kHz.

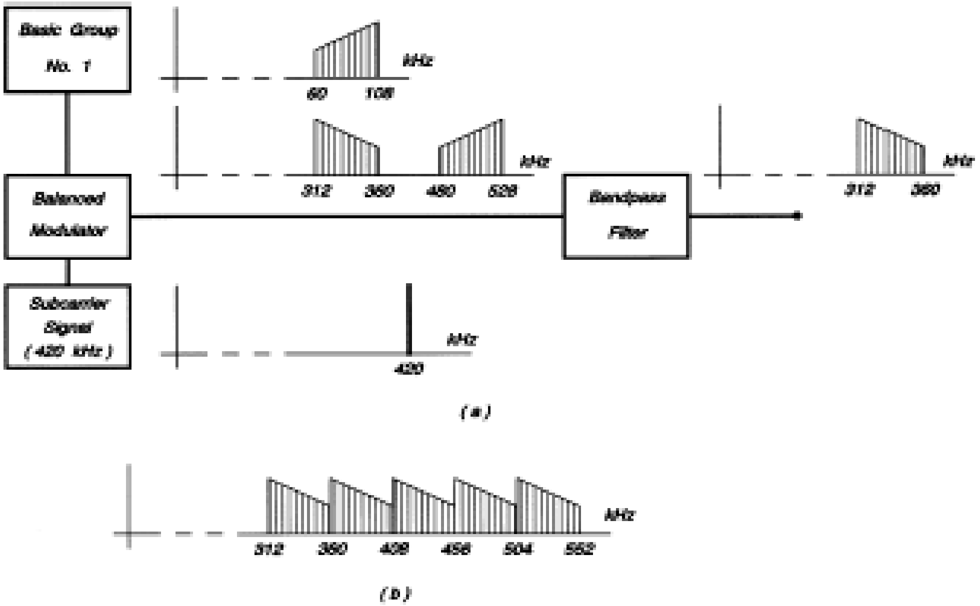


Figure 3.4 Formation of the Super group and its Spectrum

Table 3.1 shows the carrier frequencies and band widths for each Super group. For a 60-channel trunk, the signal can be transmitted in this form. Again a twisted pair with coil loading or amplification and coaxial cable may be the medium of transmission. By organizing the 12 Basic groups into a Super group of 5 it clear that the sub-carrier frequencies, the balanced modulators and band pass filters can all be duplicated five times over.

Table 3.1

Supergroup number	Carrier frequency (kHz)	Bandwidth (kHz)
1	420	312–360
2	468	360–408
3	516	408–456
4	564	456–504
5	612	504–552

3.1-1.3 Formation of a Master Group

To create a 600-channel trunk, 10 Super groups are combined to form a Master group. The frequency spectrum of the Master group is shown in figure 3.5 that occupying the frequency range from 564 kHz to 3,084 kHz and containing a total of 600 voice channels.

Note that there are gaps of 8 kHz between each Super group spectrum. These gaps are designed to make the filtering problem easier.

The carrier frequencies and bandwidths of the 10 Super groups are given in Table 3.2.

The Super group can be transmitted over coaxial cable or it can be used to modulate a 4 GHz carrier for terrestrial microwave transmission or even sent over a satellite link.

Table 3.2

supergroup number	Carrier frequency (kHz)	Bandwidth (kHz)
1	1116	564–804
2	1364	812–1052
3	1612	1060–1300
4	1860	1308–1548
5	2108	1556–1796
6	2356	1804–2044
7	2652	2100–2340
8	2900	2348–2588
9	3148	2596–2836
10	3396	2844–3084

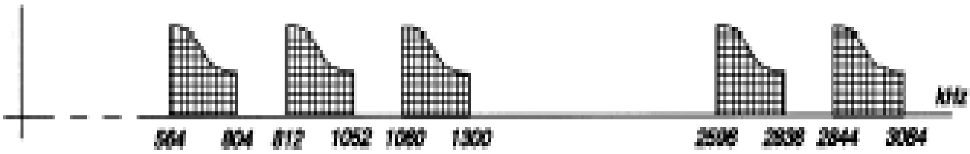


Figure 3.5 Formation of the Maser group and is Spectrum.

Six master groups multiplexed together give a Jumbo group occupying the frequency range from 564 kHz to 17,548 kHz and containing 3,600 channels. Three Jumbo groups multiplexed together give a Jumbo group multiplex containing 10,800 channel.

Analog multiplexing and continual improvements in the technology enabled costly transmission systems to be shared to carry thousands of telephone signals, thereby making long-distance telephone service affordable to us all. But analog multiplexing suffered from noise, distortion, and other impairments, and was costly to maintain as it requires a huge number of modulators / demodulators, oscillators and filters. Analog multiplexing on coaxial cable has been used in the old Bell System since 1946 for long-distance telephone transmission. Two such coaxial make a two-way pair, with each coaxial carrying transmission in one direction. As the signal travels along the coaxial, it becomes weaker and weaker and must there be amplified before it becomes too weak. Amplification is a one-way affair, and thus a coaxial can only carry signals in one direction. A number of coaxial are placed together to form a coaxial cable for use in a transmission system. The multiplexing system used with coaxial cable is called L-carrier.

Various generations of the technology are indicated by a number suffixed after the L. The key factor in the L-carrier system is the distance between the amplifiers, called repeaters that amplify and retransmit the signal to the next section of the cable.

The age of analog multiplexing is now over; digital multiplexing proved superior in nearly all respects and is now quite widespread. In the late 1980s, AT&T replaced nearly all the analog multiplexing in use on its long-distance network with digital multiplexing systems.

3.2 Digital Multiplexing

In FDM, voice signals were ‘‘stacked’’ in the frequency spectrum so that many such signals could be transmitted over the same channel without interference. Each channel is connected to the transmission medium for the whole time but each channel is allocated a different frequency band. In digital multiplexing, the time of channel use is divided among all users. Each voice signal is assigned the use of the complete channel bandwidth (All the channels use exactly the same frequency band) using one of non-overlapping time slots on a periodic basis using a technique known as TDM. Normally, all time slots of a TDM system are of equal length. Also, each sub channel is usually assigned a time slot with a common repetition period called a frame interval. Use of multiplexing technique is possible if the capacity of the channel is higher than the data rates of the individual data sources to facilitate good utilization of the channel bandwidth.

At the transmitter side the multiplexer collects the data from each source, and the combined bit stream is sent over a single medium. Framing information is needed for the switching circuit at the receiver side that separates the data corresponding to the individual sources (time slots) in the de-multiplexer as shown in figure 3.6. When the de-multiplexer detects the frame synchronization word, it knows that this is the start of a new frame and the next time slot contains the information of user channel 1.

Many PAM or PCM signals could be time multiplexed using an electronic switch that is operated by gated pulses. The switch is closed for the duration of the pulse. A TDM system with two inputs PAM signals is shown in figure 3.7. The samplers or commutators are shown here as switches which are driven in synchronism.

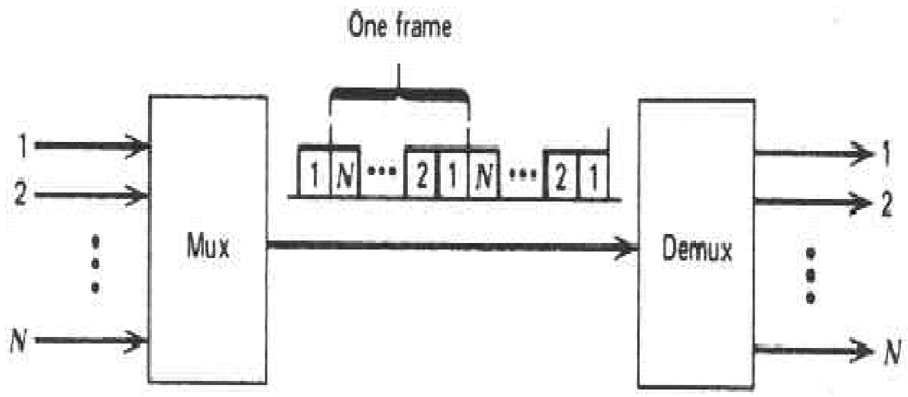


Figure 3.6 Digital Multiplexing and de-multiplexing

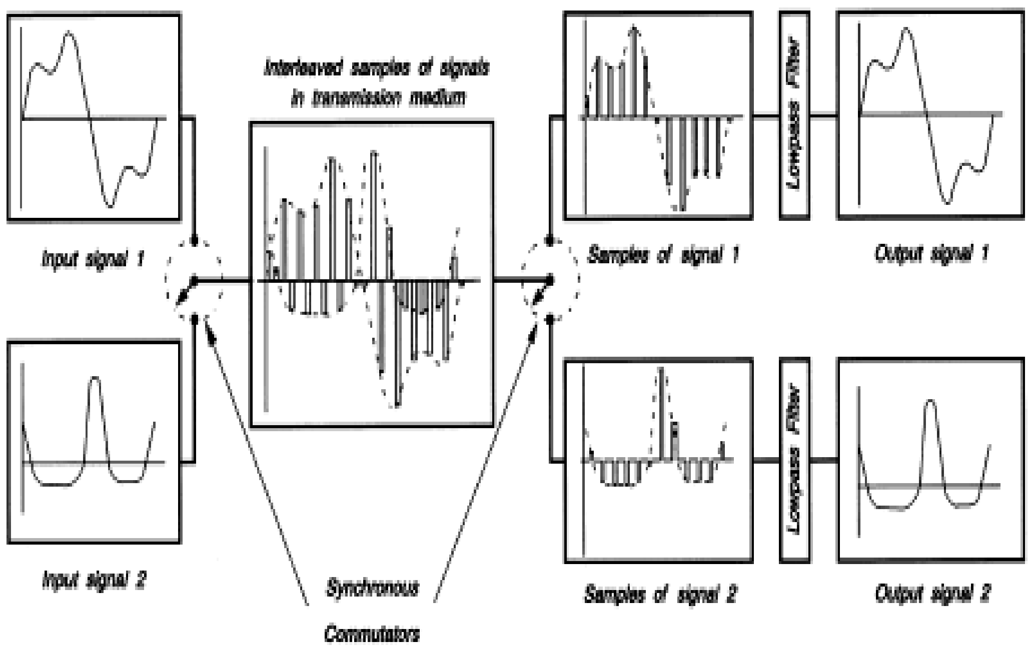


Figure 3.7 TDM Principle

TDM obviously increases the overall sample rate and therefore the required bandwidth for transmission as shown in figure 3.8. It is however, possible to reduce the band width of the TDM signal dramatically by appropriate filtering since, strictly, it is only necessary that the TDM signal provides the correct amplitude information at the sampling instants.

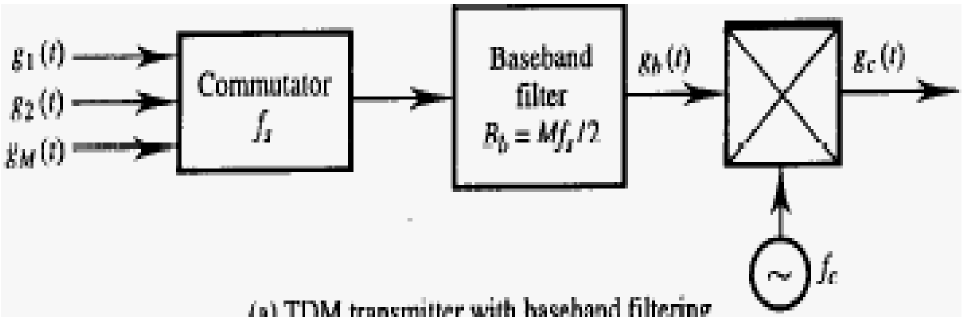


Figure 3.8 TDM transmitter with base band filtering

If a minimum bandwidth TDM signal is formed by filtering then sampling accuracy becomes critical in that samples taken at times other than the correct instant will result in cross-talk between channels.

Cross talk also occurs between the channels of a TDM signal even if filtering is not explicitly applied. This is because the transmission medium itself may band limit the signal. Such band limiting effects may often be at least approximated by RC low-pass filtering. In this case the response of the medium results in pulses with exponential rising and falling edges is shown in figure 3.9.

If the guard time between rectangular pulses (time between the trailing edge of one pulse and the rising edge of the next) is t_g, and the time constant of the transmission channel is RC, then the amplitude of each pulse will decay to a fraction $e^{-tg / RC}$ of its peak value by the time the next pulse starts.

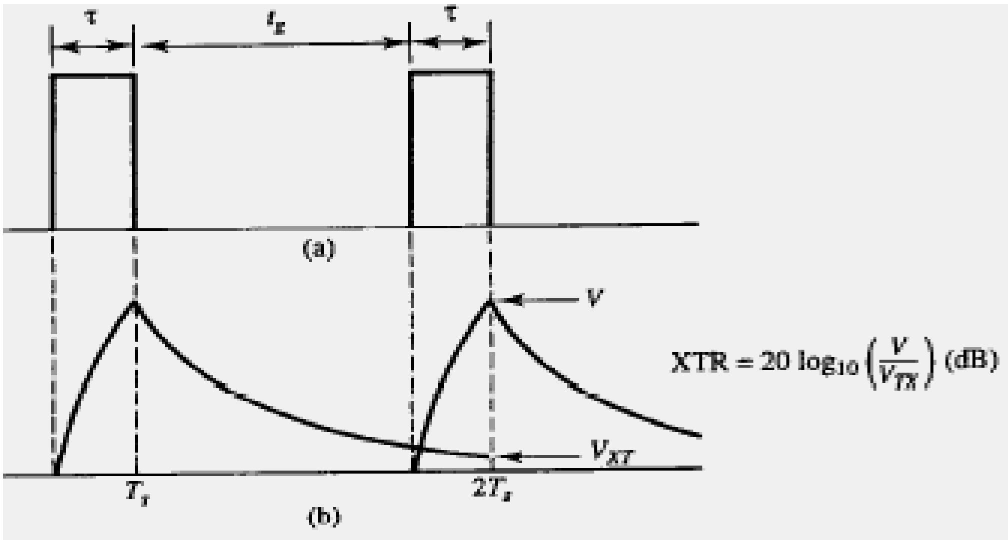


Figure 3.9 Cross-talk between tributary channels of a TDM signal: a- signal at channel input, b-Signal at channel output.

For the RC characteristic the channel bandwidth is;

$f_{3dB} = 1 / (2\pi RC)$, therefore the cross-talk ratio (XTR) in dB at the pulse trailing edge (the optimum XTR sampling instant) is:

$$XTR = 20 \log e^{-2\pi f_{3dB}(tg + \tau)}$$

$$= 54.6\, f_{3dB}\,(tg + \tau) \text{ dB} \qquad (3.1)$$

This is the first order estimate of the required guard time to maintain a desired cross-talk ratio in a band limited channel, for a given rectangular pulse width τ, at the channel input. (If sampling occurs at the center of the τ s nominal pulse slot, rather than at its end, then XTR is reduced by approximately 6 dB when τ << RC.)

Digital TDM forms the basis of the telephone hierarchy for transmitting multiple simultaneous telephone calls over high speed 2, or 140 M bit/s data links.

When TDM signal is developed on a four-channel system as shown in figure 3.10, for example, the sampling frequency for each individual channel is displaced in time relative to the preceding and succeeding channel.

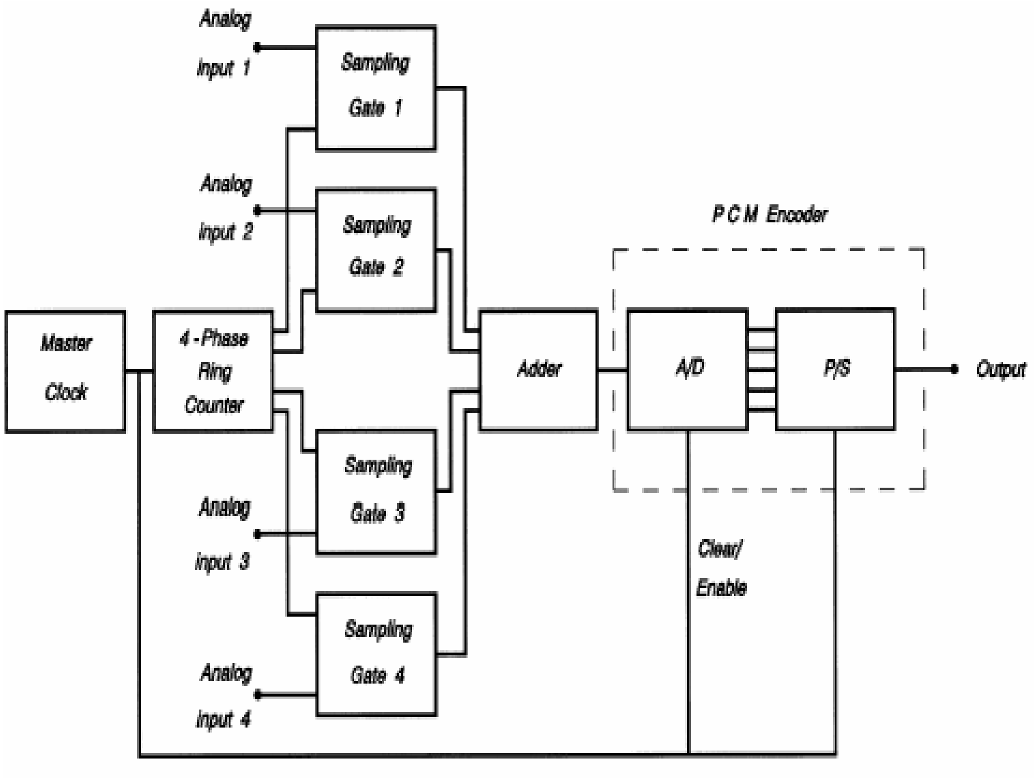


Figure 3.10-a Block diagram of a four-channel PCM System

This means that the sampling takes place successively,

so that channel 1 is sampled first, followed by channels 2, 3 and 4, then channel 1 is sampled again and so on.

Practical TDM systems based on PAM have been built and used in the telephone system (No. 101 ESS-PBX).

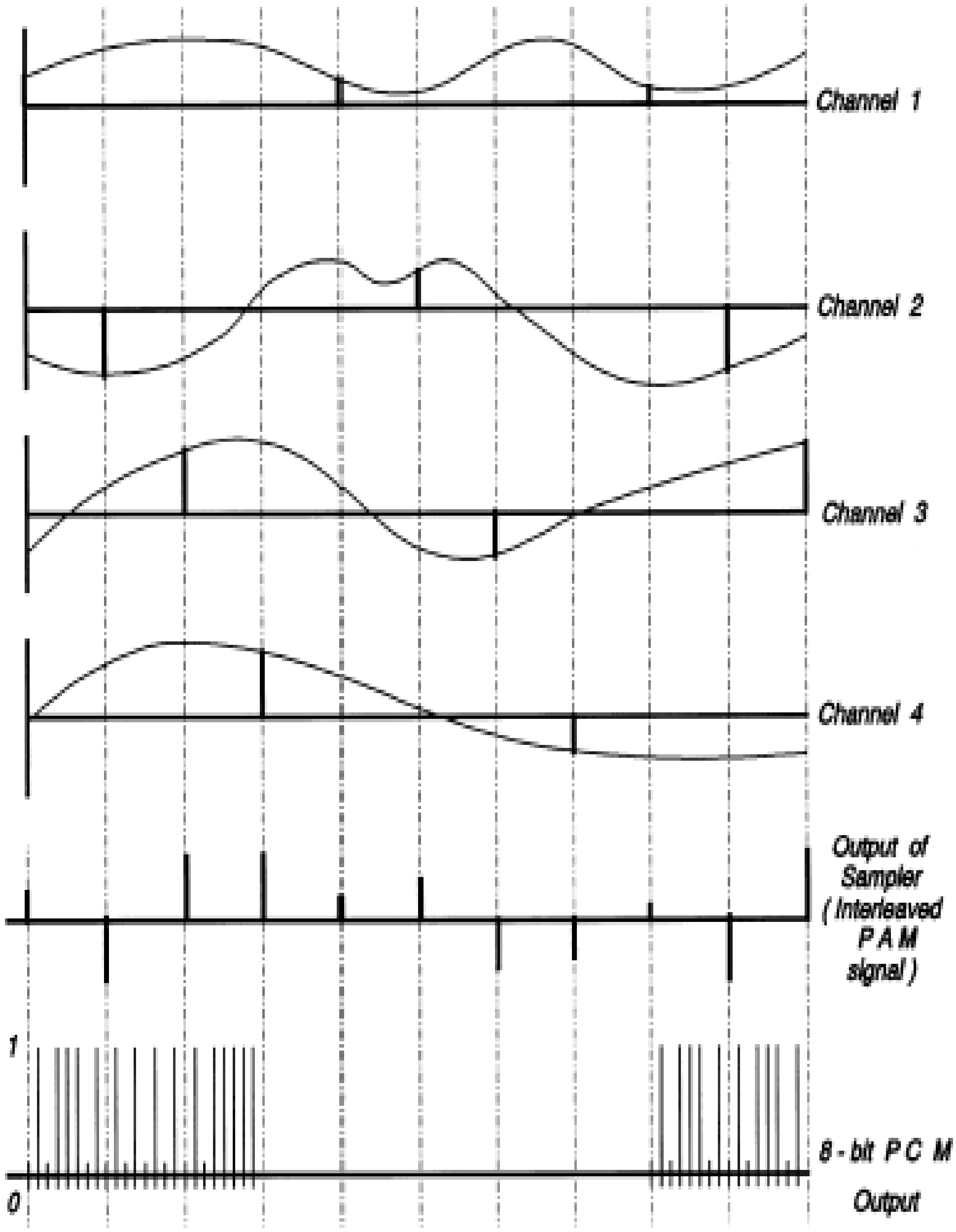


Figure 3.10-b Signals of the four-channels, the interleaved PAM and the 8-bit PCM System

3.2-1 TDM Hierarchy

PCM-coded speech is transmitted as 8-bit samples 8,000 times a second, which makes up a 64-Kbps data rate.

These eight-bit words from different users are interleaved into a frame at a higher data rate. In a manner similar to the FDM hierarchy, there are two recommended standards:

- **North American, AT&T Standard**
- **European, CEPT Standard**

American Telephone and Telegraph, AT&T established a digital TDM hierarchy that has become the standard for North America and Japan.

It uses the μ-law for quantizing, and the system was designed for channel-associated signaling.

All higher levels are implemented as a combination of some number of lower level signals. The designation of the higher level digital multiplexers reflects the respective input and output levels.

For example, an M12 multiplexer combines four DSl signals to form a single DS2 signal, (Because T2 transmission systems have become obsolete, the M l2 function exists only in a functional sense within M l3 multiplexers, which multiplex 28 DS1 signals into 1 DS3 signal).

A similar digital hierarchy has also been established by ITU-T as an international standard. In 1959 the European countries formed a non-political organization "CEPT", The European Committee of Postal and Telecommunication Administrations. This committee recommended standards for compatibility of the telecommunications systems employed throughout Europe. Both European and American PCM frame are repeated at PCM sampling rate that is 8,000 times in a second.

3.2-1.1 AT&T System

It is known as Bell T1 PCM Carrier System uses an eight-bit PCM in 24 voice-channel banks as shown in figure 3.11. The number of bits generated for one scan of the channels (frame) is 24 x 8 = 192. One bit (a frame alignment bit) is required for frame synchronization (S-bit) so the total number of bits per frame is 193. For band-limited 4 kHz analog signal, sampling rate 8 k Hz, the system produces a gross line bit rate of (193 x 8000) = 1.544 M bit/s. The 24 voice channels require 1.536 M bps and an additional 8 k bps are needed for synchronization purposes, thereby giving 1.544 M bps as the overall bit rate, The minimum bandwidth required to transmit the signal is 1.5 MHz. Starting with DS0 which is a 64 k bps digital version of a voice signal as a fundamental building block, DS1 or T1 level frame has 24-PCM channel. It is the most common frame structure in telecommunications networks used in North American standard areas.

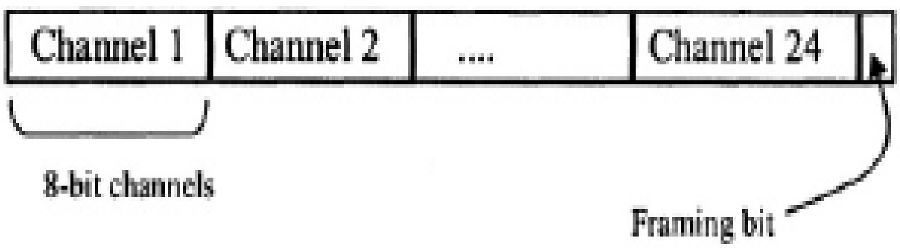


Figure 3.11 The 193-bit, 125 μ sec. DS-1 frame.

The use of the eight-bit code means that the voice signals are quantized at 256 (2^8) levels. Some of the less significant bits may be robbed and used for signaling purposes such as dialing and detection of hook switch ON/OFF.

The quantization error resulting from this is considered to be tolerable, although several compression schemes are used to minimize its effect.

The 1.544-Mbps frame structure is shown in figure 3.12.

A multi-frame is constructed from 12 subsequent frames and their 12 S-bits make up the 6-bit frame and 6-bit multi-frame synchronization words.

In T1, the least significant bit of each channel in every sixth frame is used for signaling. As a consequence, only seven bits in each time slot are transparently carried through the network and the basic user data rate is 56 Kbps instead of the 64 Kbps in the European systems.

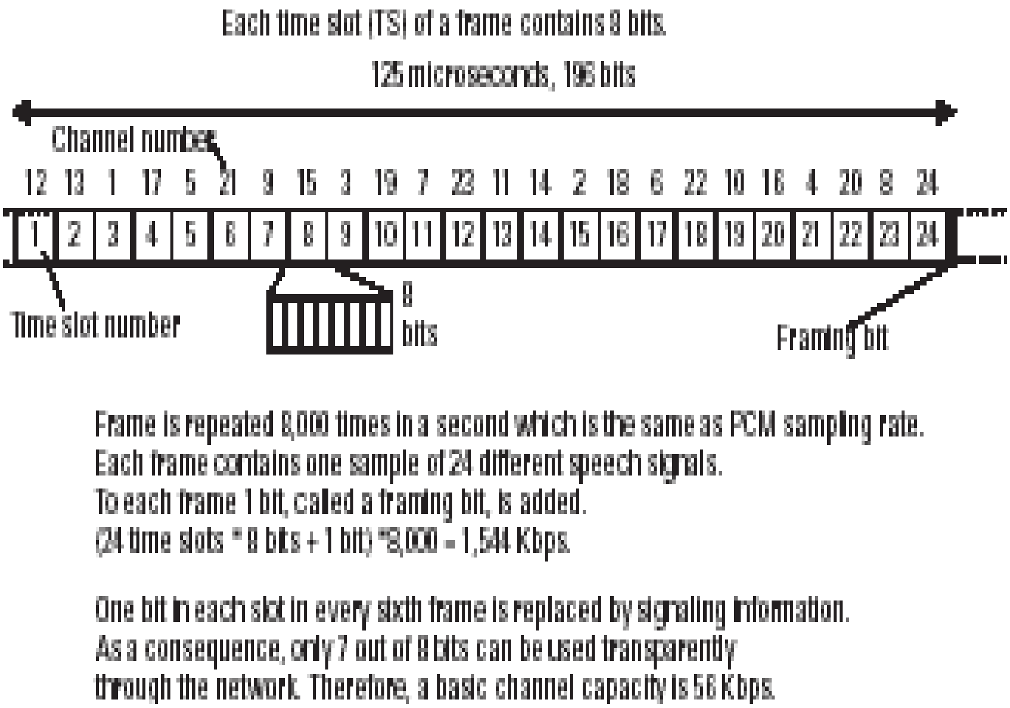


Figure 3.12 The 1.544 Mbps PCM frame

For frame synchronization and for de-multiplexing of signaling information, frames make up a multi-frame structure with two alternative lengths, a *Super frame* (SF) containing 12 frames or an *Extended super frame* (ESF) containing 24 frames. The framing bits of ESF, one in each frame, carry frame synchronization information including CRC code and data channel for network management messages.

In transatlantic connections, E1 frames are adapted to the T1 frame structure and trans-coding between μ-law and A-law PCM is carried out. Each time slot in E1 is transmitted further in one time slot of T1.

Table 3.3 and following figure list the various multiplex levels, their bit rates, and the transmission media used for each. Notice that the bit rate of a high-level multiplex signal is slightly higher than the combined rates of the lower level inputs. The excess bits are included for certain control and synchronization functions. As DS0 with 64 k bps digital voice signal as a fundamental building block, DS1 or T1 level frame has 24-PCM voice channel digitally multiplexed together is sometimes called a digital group, or a digroup for short.

As shown in table 3.3, four DS-1 signals time multiplexed together give a DS-2 signal containing 96 voice channels and requiring a data rate of 6.312 M bps.

Seven DS-2 signals time multiplexed together give a DS-3 signal containing 672 channels and requiring 44.736 M bps.

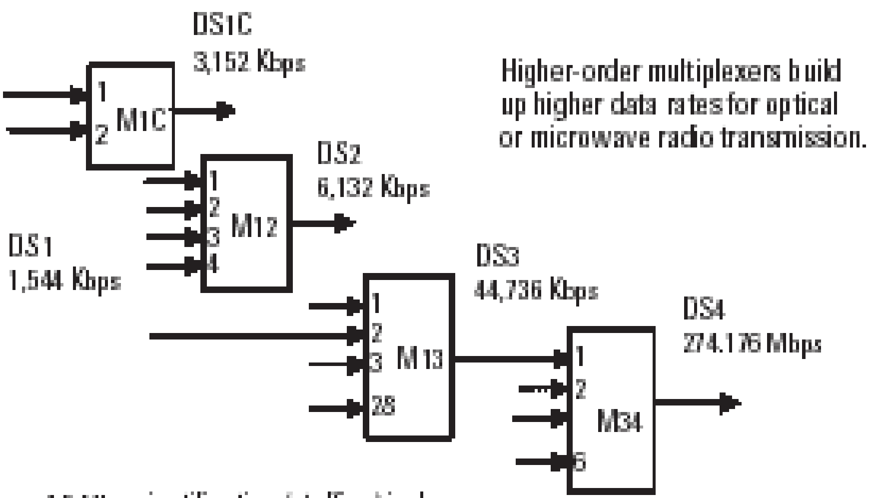
Six DS-3 signals time multiplexed together give a DS-4 signal containing 4032 channels and requiring a data rate of 274.176 M bps.

The extra bits in the higher capacity digital signals are used for timing and synchronization information to assist in the separation and de-multiplexing of the individual channels.

The timing and clocking information is contained within the digital bit stream and thus is self-synchronizing, or asynchronous. Such near syn-chronous timing schemes form a plesiochronous digital hierarchy (PDH).

Table 3.3 Digital TDM signals of North America and Japan

Digital Signal Number	Number of Voice Circuits	Multiplexer Designation	Bit Rate (Mbps)	Transmission Media
DS1	24	D channel bank (24 analog inputs)	1.544	T1 paired cable
DS1C	48	M1C (2 DS1 inputs)	3.152	T1C paired cable
DS2	96	M12 (4 DS1 inputs)	6.312	T2 paired cable
DS3	672	M13 (28 DS1 inputs)	44.736	Radio, Fiber
DS4	4032	M34 (6 DS3 inputs)	274.176	T4M coax, WT4 waveguide, radio



Above 1.5 Mbps, justification (stuffing) is done at each stage because tributary rates are allowed to be plesiochronous. Demultiplexing has to be done step by step because justification bits must be stripped off in order to locate the information content.

When the signal is sent over a twisted-pair telephone wire, it suffers considerable degradation from noise, bandwidth limitation, and phase delay. It is therefore necessary to place repeaters and equalization circuits at intervals of approximately 6000 ft to restore the pulses. 6000 feet (approximately 2000 m) of twisted pair non-loaded wire has a 3 dB bandwidth of approximately 4 kHz.

Signaling Information

There are two types of information carried by a DS-1 frame: the user information and the signaling information. The example of user information is digital voice or data. The signaling information is used by the network to delineate and control user information. In the case of DS-1 system, there may be as many as 24 users using each frame. There is the need for routing and control of this information. This is done by robbed bit signaling in voice communication and common channel signaling in data communications.

In robbed bit signaling one of every 48 bits in a voice channel is stolen by the network to be used for signaling. Recall that each voice channel has 8 bits per frame. Therefore, stealing one bit out of six frames per channel leaves 47 out of the 48 bits used for actual user information. The stolen bit is the least significant bit of a channel of every sixth frame. Thus, user channel contains a signaling channel with a bit rate 1/48 times the bit rate of voice channel. Consequently, the 8-bit PCM does not get a true bit rate of 64 kbps.

Instead, it has a bit rate of $(47/48)^{th}$ of 64 kbps. The signaling bit rate per voice channel is $1/48 \times 64$ kbps.

In summary, when DS-1 is used to carry voice only traffic, the 24 channels are used for 8-bit PCM. Signaling information is added to each voice channel by stealing one bit per channel in every sixth frame. In common channel signaling mechanism, a separate channel is stolen from every frame to carry the signaling information about a group or groups of channels. In this case, twenty-three channels are used for data, and the 24th channel is dedicated to carry the signaling information about all the remaining 23 channels.

The main use of this channel is fast recovery from a framing error.

In the 23 channels, the user information is carried by using 7 bits per channel, thus giving a data rate of ;

$7 \times 8000 = 56$ kbps. The 8th bit of each data channel is used to define a signaling channel for each data channel.

This is done when DS-1 frame is used to carry data from data terminals.

Another type of signaling is the in-channel signaling in which the user and signaling information is carried over the same channel. It is used by allocating one bit of every data channel for this purpose.

When DS-1 is used to carry combined voice and data information, then all the 24 channels are used. Multiple DS-1 frames can be multiplexed to obtain higher data rate systems, such as DS-2, using 4 DS-1 frames, and DS-3 combining 30 DS-1 frames.

3.2-1.2 The 30/32-Channel CEPT PCM System

The 30/32 channel PCM system uses a frame and multi-frame structure is shown in figure 3.12. The TDM frame depicts the number of bits in each channel where out of the 32 slots, 30 slots are used to carry voice and two slots (slot 0 and slot 16) are used to carry synchronization and signaling information. The 2.048 Mbps used in areas that go by European standards. Starting with E0 which is an 64 k bps digital version of a voice signal as a fundamental building block, E1 or T1 level frame has 30-PCM channel. There is reserved time slot for CAS information in the 2-Mbps frame structure.

Signaling Information

In the E-l carrier system, there are 32 channels, each for 64 kbps data rate, providing a total bit rate of 2.048 Mbps. Thirty of these channels are used for user data (voice or non-voice) and the remaining two for various signaling and control functions. The term T-1 is used to describe the raw bit rate of 1.544 Mbps using DS-1 format. There is no such differentiation of terms in E-l: the term E-l conveys the meanings for the signal format as well as the raw bit rate.

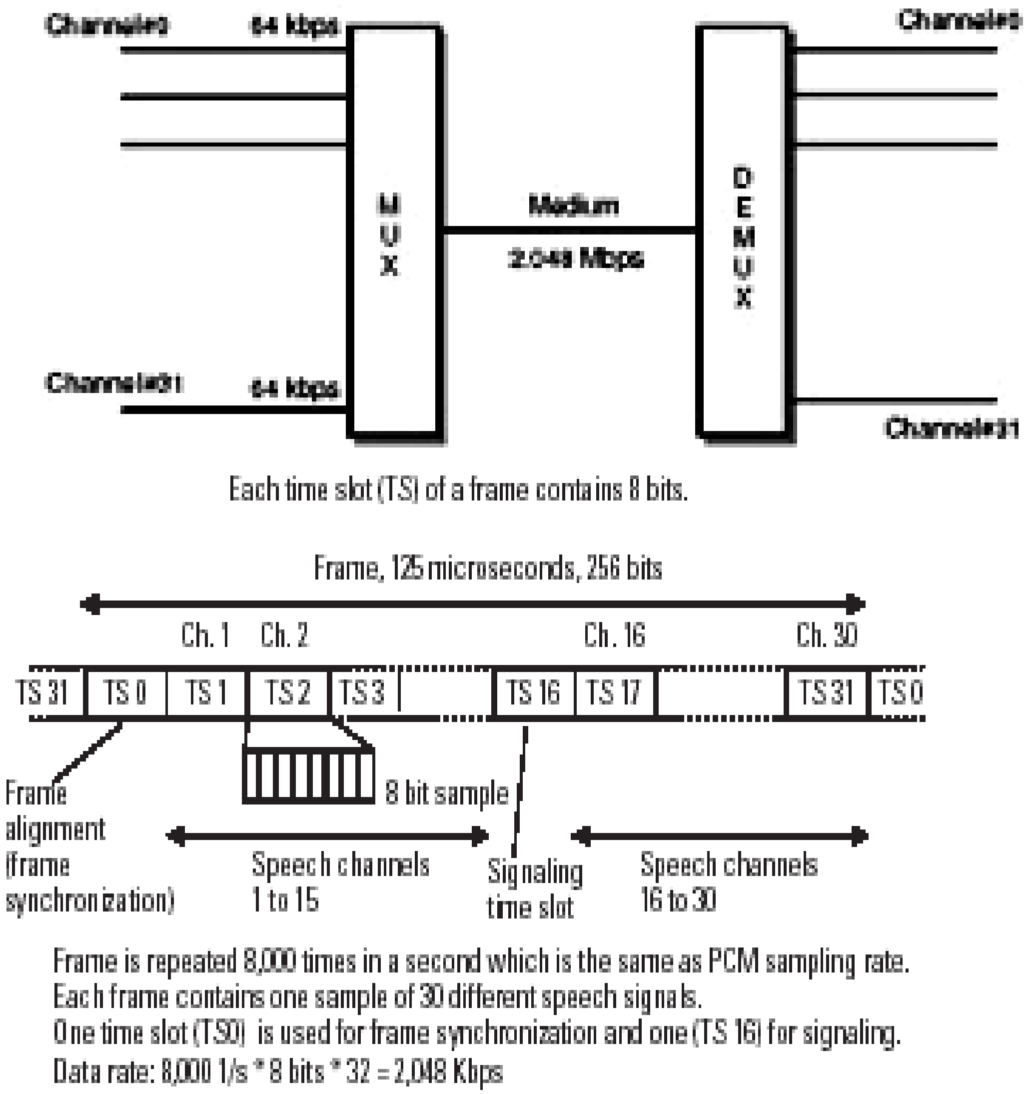


Figure 3.12 30 / 32-Channel CEPT PCM System with 2,048-kbps frame structure

Multiplexing is performed at the transmitter side and the TDM frame has the following channel allocation:

- Time slot 0: frame alignment word (FAW), frame service word (FSW).
- Time slots 1 to 15: digitized speech for channels 1 to 15.

• Time slot 16: multi-frame alignment word (MFAW), multi-frame service word (MFSW) signaling information.

• Time slots 17 to 31; digitized speech for channels 16 to 30.

Common channel signaling can be adapted for this system and if this is used then it would have 31 data channels and a single synchronization channel, or Frame Synchronization Time Slot.

Each time slot contains an eight-bit sample value and each channel produces data at the rate of 64 Kbps. These voice channels or data channels are synchronously multiplexed where the multiplexer takes the 8 bits of each channel and produces a gross line bit stream at the rate of 2.048 Mbps (64 kbps × 32) data stream known as the 2-Mbps Frame, which is often called E1 (first level in European hierarchy).

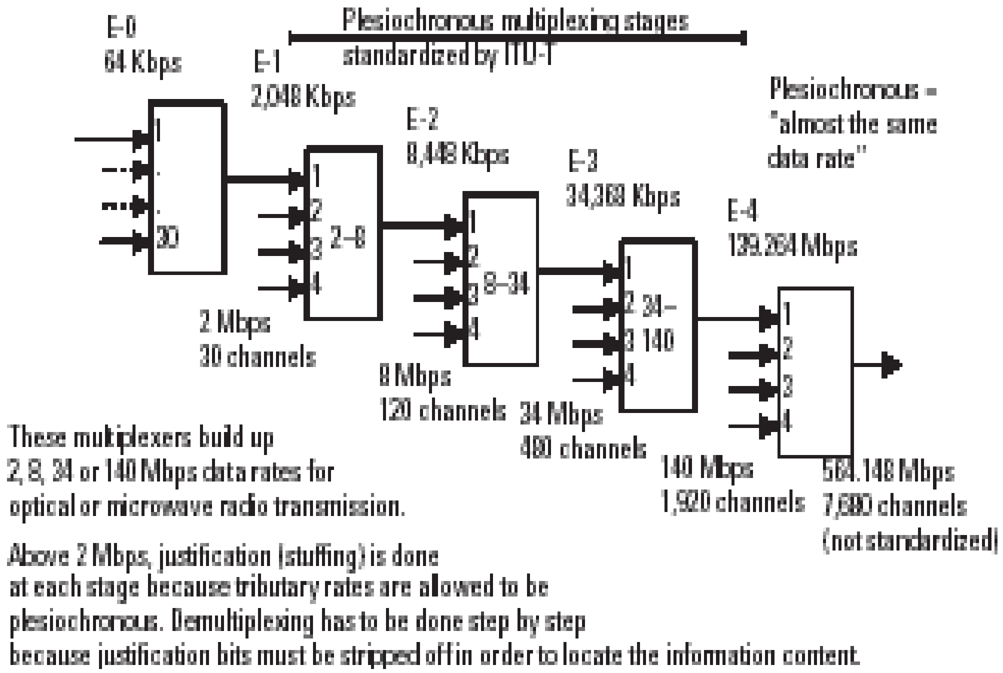
It uses the A-law in the quantization process. The frame is repeated 8,000 times a second, which is the same as the PCM sampling rate. For error-free operation the tributaries (64-Kbps data streams of the users) have to be synchronized with the clock signal of the 2-Mbps multiplexer. The data rate of 2,048 Kbps for the multiplexer is allowed to vary by 50 *parts per million* (ppm), and as a consequence each user of the network has to take timing from the multiplexer in the network and generate data exactly at the data rate of the multiplexer divided by 32.

At the receiving end, the de-multiplexer separates the data corresponding to each channel. As shown in Table 3.4, this hierarchy is similar to the North American standard but involves different numbers of voice circuits at all levels.

Table 3.4 Digital TDM signals of Europe

Level Number	Number of Voice Circuits	Multiplexer Designation	Bit Rate (Mbps)
E1	30		2.048
E2	120	M12	8.448
E3	480	M23	34.368
E4	1920	M34	139.264
E5	7680	M45	565.148

The following figure indicates higher order multiplexing stages for European standard illustrates in table 3.4



A- PCM Frame and Multi-frame Structure

A.1- The Frame Structure

The frame consists of 32 time slots (TS), as shown in figure 3.13. Each TS consists of 8 bits. In these time slots bit 1 is used to indicate the polarity of the sample and bits 2 to 8 indicate the amplitude of the sample.

A.2-The Multi-frame Structure

In order for signaling information (dial pulses) for all 30 channels to be transmitted, the multi-frame consists of 16 frames numbered F0 to F15. This structure is shown in figure 3.13. Signaling for two channels is transmitted in time slot 16 in each frame.

B- Time Durations

CCITT recommendation is that the 8 kHz is used as the sampling frequency for 4 kHz voice signals. The following calculated time durations are all shown in figure 3.13

B.1- Frame duration

The frame duration for f_S = 8 kHz, is determined as;

Frame duration = 1/f_S = 125 μ second

B.2- Time Slot duration

The time slot duration is determined as follows:

$$\text{Time slot duration} = \frac{frame\ duration}{timeslots\ per\ frame}$$

= 125 x 10^{-6} / 32 = 3.906 μ second

B.3- Bit duration

The bit duration is determined as follows:

$$\text{Bit duration} = \frac{timeslot\ duration}{bits\ per\ timeslot}$$

$$\text{Bit duration} = \frac{3.906\,x10^{-6}}{8} = 488\ n\ \sec ond$$

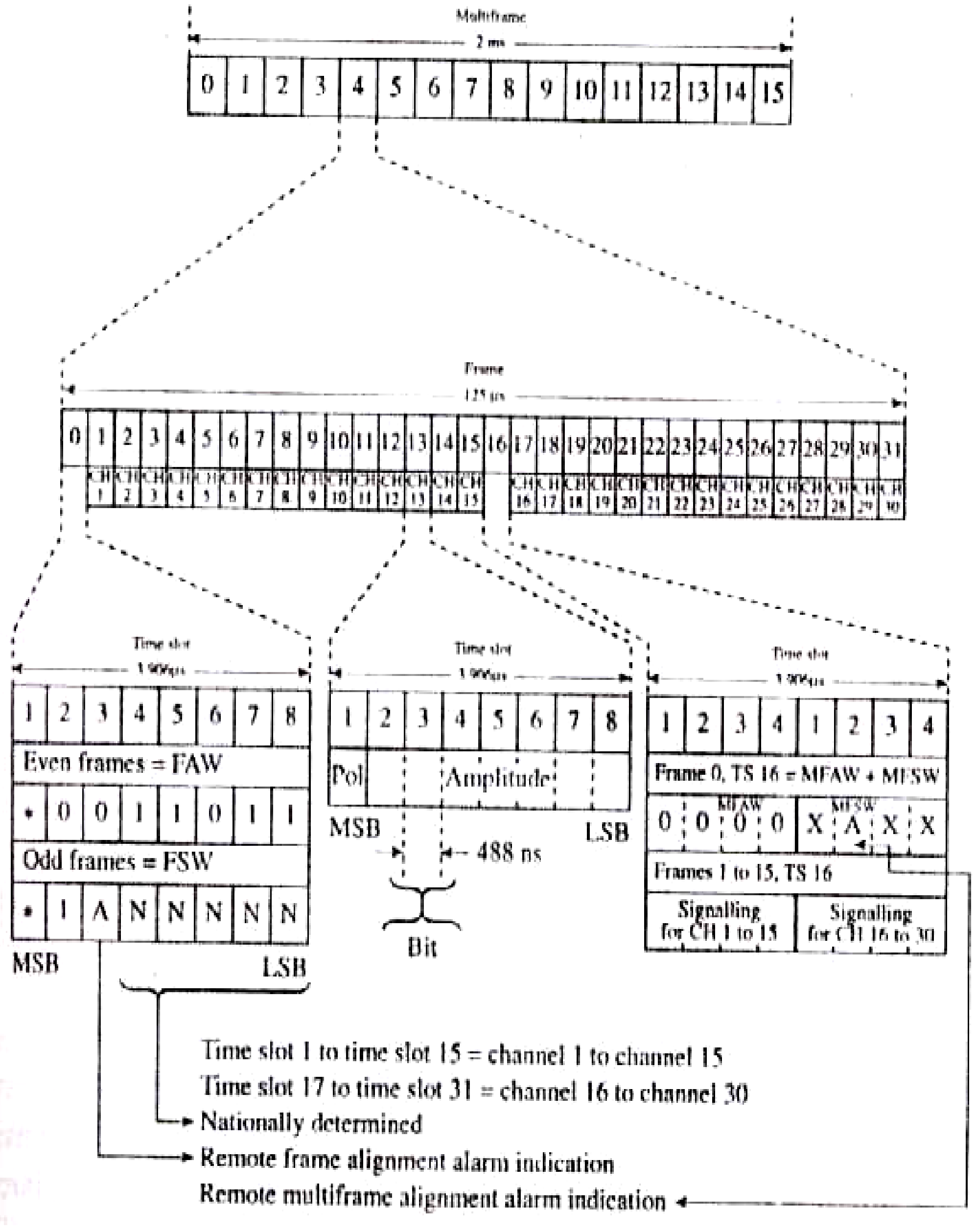


Figure 3.13 Frame and multi-frame structure for a 30 / 32- channel PCM system

B.4- Multi-frame duration

The Multi-frame duration is determined as follows:

Multi-frame duration = frame duration x frames per multi-frame;

$= 125 \times 10^{-6} \times 16 = 2$ m second.

C- Gross line bit rate

The gross line bit rate is determined as:

$$\text{Gross line bit rate} = \frac{1}{bit\ duration} = \frac{1}{488\, x 10^{-9}}$$

$$= 2.048 \text{ M bit/sec.}$$

It is the most common frame structure in telecommunications networks, as the primary rate 2,048-Kbps frame used in the European standard areas. This is the basic data stream that carries speech channels and ISDN-B channels through the network and it is called E-1.This primary rate frame is built up in digital local exchanges that multiplex 30 speech or data channels at bit rate of 64 Kbps into the 2,048-Kbps data rate. ITU-T defines this frame structure in Recommendation G.704.

TDM is normally associated only with digital transmission links and the backbone digital links of the PSTN (T-carrier, digital microwave, and fiber optics) use a synchronous variety of TDM. Although analog TDM transmission can be implemented by interleaving samples from each signal, the individual samples are usually too sensitive to all varieties of transmission impairments. Analog TDM techniques have been used in some PBXs before digital electronics became so inexpensive that the digitization penalty disappeared.

We discriminate between two types of TDMs to deal with the different ways in which time for channel use could be allocated. The form of TDM shown in figure 3.6 is sometimes referred to as synchronous TDM to specifically imply that each sub channel is assigned a certain amount of transmission capacity determined by the time slot duration and the repetition rate. In contrast, another form of TDM referred to as statistical, or asynchronous TDM. With this second form of multiplexing, sub channel rates are allowed to vary according to the individual needs of the sources.

The frame alignment word is needed to inform the de-multiplexer where the words of the channels are located in the received 2-Mbps data stream. The frame synchronization time slot (TS0) includes frame alignment information and it has two different contents that are alternated in subsequent frames as shown in figure 3.14. The de-multiplexer looks for this time slot in the received data stream and, when it is found, locks onto it and starts picking up bytes from the time slots for each receiving user. Each user receives 8 bits in 125-µs periods, which makes 64 Kbps. A fixed alignment word is not reliable enough for frame synchronization because it may happen that a user's data from one channel simulates the synchronization word and the de-multiplexer might lock to this user time slot instead of TS0. This is why there is one alternating bit (D2) in time slot 0 as indicated in figure 3.14 and due to this the de-multiplexer is able to detect the situation where one channel constantly transmits a word that is equal to the *frame alignment word* (FAW).

To make frame alignment even more reliable, the *cyclic redundancy check 4* (CRC-4) procedure was added in the mid-1980s. C-bits are allocated to carry a four-bit error check code that is calculated over all bits of a few frames. The receiver performs error check calculations over all bits of the frames and it is able to detect false frame alignment even if the frame alignment word is simulated by one user that alters bit two.

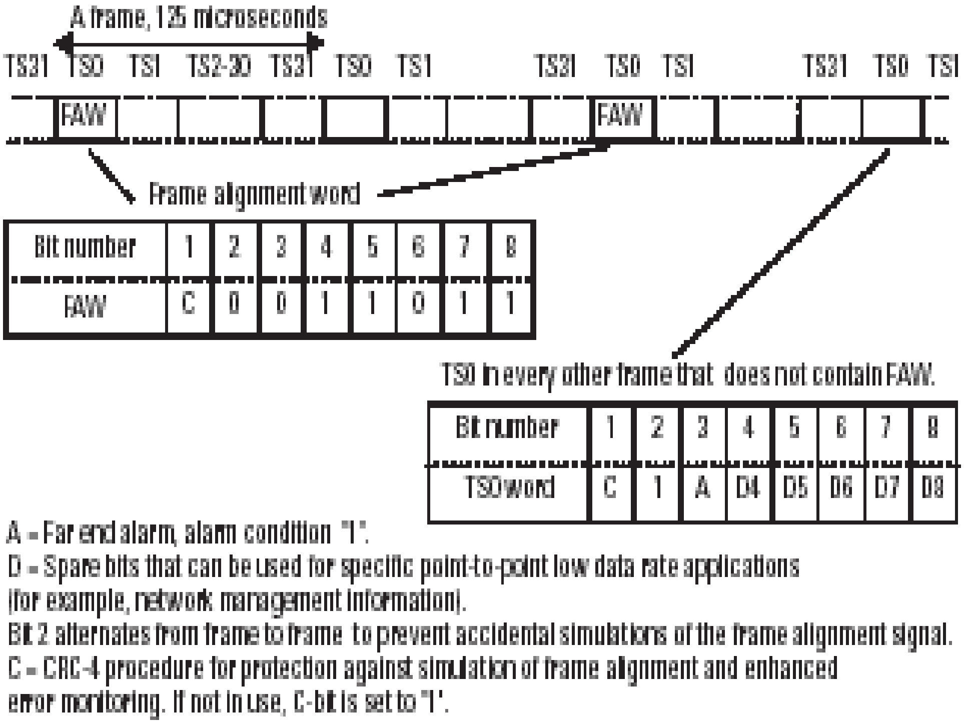


Figure 3.14 the 2,048-kbps frame alignment word in TS0.

Each receiver of the 2,048-Kbps data stream detects errors in order to monitor the quality of the received signal. Error monitoring is mainly based on the detection of errors in the frame alignment word. The receiver compares the received word in every other TS0 with the error-free frame alignment word. In addition to the frame alignment word, the CRC-4 code is used to detect low error rates. Errors in the frame alignment word do not give reliable results when the error rate is very low. It may take a long time before an error is detected in TS0 although many errors may have occurred in other time slots of the frame. The C-bit in figure 3.14 is set to 1 if CRC is not used.

The TS0 in every other frame also contains a far-end alarm information bit A as shown in figure 3.14. This is used (set to 1) to tell the transmitting multiplexer that there is a severe problem in the transmission connection and reception is not successful at the other end of the system. This may caused by, a high error rate, loss of frame alignment, or loss of signal. With the help of the far-end alarm, consequent actions can take place. These actions include rerouting user channels to another operational system. D-bits can be used for transmission of network management information. At international borders they are usually set to 1.

Multi-frame Structure of the Signaling Time Slot

Time slot 16 (TS16) is defined to be used for the channel associated signaling to carry separate signaling information to all user channels of the frame. TS16 is a transparent 64-Kbps data channel like any other time slot in the frame.

Thirty channels share the signaling capacity of TS16.

A frame structure is needed to allocate the bits of this time slot to each of the 30 speech channels.

The information about the location of the signaling data of each speech channel is given to the signaling de-multiplexer with the help of the multi-frame structure containing a multi-frame alignment word for multi-frame synchronization.

The data rate available for each speech channel is 2 Kbps.

For CCS, the signaling information of all users is carried in data packets and any time slot can be used for this. Each packet carries information about the call to which it is related and signaling information. CCS packets can in some cases, for example, in the short message service of GSM, also carry user data.

3.2-2 Synchronous TDM

In synchronous TDM, the sampling rate of the various channels are identical, all transmissions from multiplexed users occur at specified time instants. For example, each user is allowed to transmit for a time beginning at a given instant and ending at another instant. If the user does not have data to transmit at the beginning of the specified interval, the channel remains unused.

Synchronous TDM is performed by defining channel as having a certain data rate. A multiplex cycle or frame is then defined as consisting of a certain number of bits to be repeated. This multiplex cycle is further divided into time slots such that the sum of slots is equal to the multiplex frame.

Example 3.1: Consider 8-bit PCM voice transmission. Let a digital transmission channel have a data rate of 768 kbps or bits per sec.
This channel has 12 times the data rate required by a single voice source. Suppose that a TDM cycle consists of ($12 \times 8 = 96$) bits. Let the cycle be divided into 12 slots, each of 8 bits. Then 12 voice sources, each using 64 kbps 8-PCM can be multiplexed on this line. Note that in 8-bit PCM, 8 bits represent a single sample. Therefore, samples from 12 sources are multiplexed and interleaved on a single channel.
In synchronous TDM, the time slots are allocated to traffic sources without any regard to whether these time slots will be used continuously or not. It is well known that a voice source is active only about 40% of the time. That means that for 60% of the time there will be no data from each voice source and the slots will remain unused for about 60% of time.

3.2-3 Asynchronous, Statistical TDM

In this method of TDM, signals are with different sampling rates, specific slots are not allocated to individual users. Instead, all data sources store their data in a buffer and then spit them out at a fixed rate. The multiplexer visits data buffers one by one. If a buffer contains some data to be transmitted, it is transmitted. Otherwise, the multiplexer goes on to the next data buffer. Therefore, the channel bandwidth is not wasted if any buffer has some data to send all the time.

The difference between synchronous and statistical TDM is shown in figure 3.15.

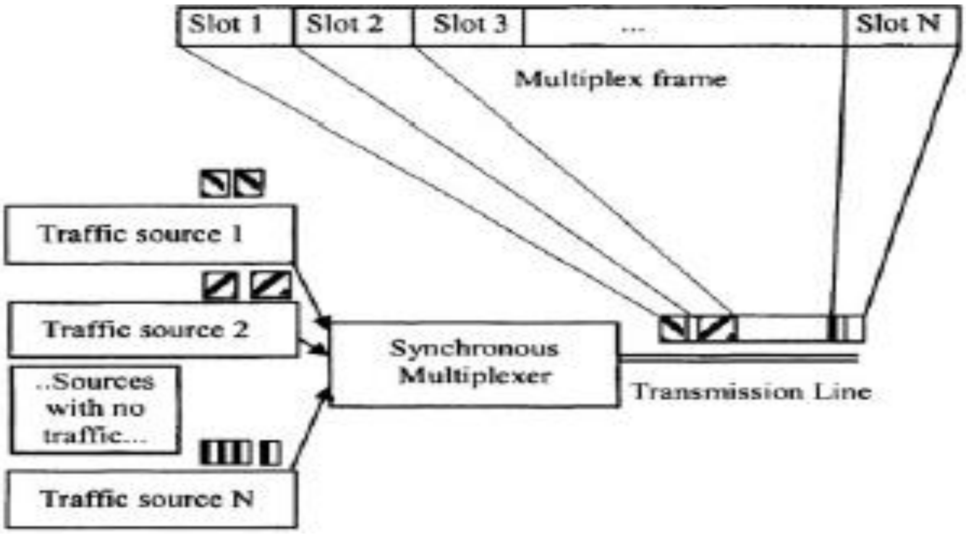


Figure 3.15 a- Synchronous multiplexing. Even if frame is partially filled, the source has to wait for the next frame to send more than one slot of traffic

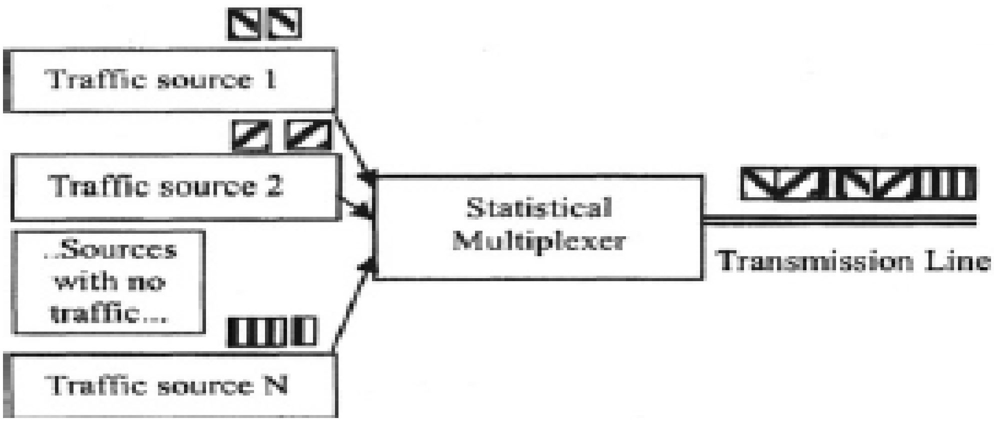


Figure 3.15- b Statistical multiplexing. If frame is partially filled, the source doesn't has to wait for the next frame to send more than one slot of traffic

The figure shows data from N users multiplexed on a single transmission line. A fixed frame length is defined in synchronous TDM in figure 3.15-a. Each of the N data sources is allocated a specified part of the frame. The figure shows that during one particular TDM cycle only three users have data to send. The frame carries data from these three users and remains largely empty. Shown in ‘white’ is the portion of the channel frame for which no one transmits any data, even if some of the sources have data to transmit.

Contrary to this figure 3.15-b shows that there is no multiplex cycle in statistical TDM. The multiplexer visits each source for the same duration of time as in synchronous TDM. However, if a source does not have any data to send, the multiplexer does not stay there for the maximum allocated amount of time. Instead, it moves on to the next source of data. In this way, more data is transmitted for the duration of the same frame. In synchronous TDM even if an attached source does not have any data to send, the channel expects that the source will use the allocated slot. However, in statistical TDM, if a source has nothing to send, the multiplexer goes on to next source.

In synchronous TDM, each source can transmit data only in the designated time slot or slots. Consequently, if the multiplexed data is heading towards a switch, the switch knows which slots are being used by each source.

The data can be switched according to this information. However, the situation is different for statistical TDM. In this case, any slot could be carrying data from any of the sources. In fact, there is no particular need of dividing the frame into slots. The information about who is the source of a slot has to be included with the data. So, the user data blocks for statistical TDM consist of a header that tells the switch about the source or destination (or both) of data.

The frame for statistical TDM does not have to be of a fixed length. Depending on factors, such as traffic from other sources, a source can transmit a variable number of bits each time. In this way, the frame for statistical TDM requires pretty much all functions of synchronous communication via frames, namely, flag, address and other frame control information.

3.2-4 Statistical Versus Synchronous TDM

The slot allocation in synchronous TDM is just like allocating channel in circuit switched network. It does not require addressing and framing information. Therefore, all sources of traffic use their designated time slots as if the slots were separate channels. When traffic load is high enough to keep the TDM channel busy for most time, the synchronous TDM is more efficient than statistical TDM as it does not have any frame headers. Therefore, statistical TDM is not always better than synchronous TDM or vice versa. In fact, they both have strengths and weaknesses. The main strengths of statistical TDM lie in the following facts:

1. It does not require a strict definition of beginning and ending of a slot/frame. This property makes its implementation simple. This resembles packet switching mechanism.
2. It utilizes the channel capacity more efficiently, especially when individual sources have bursty traffic.

Bursty traffic is characterized by repeated patterns of sudden data generation followed by long pauses.

3. It is better suited to traffic sources with varying requirements of channel capacity.

Sources requiring higher capacities might send longer frames, and more often than the other slow speed sources. Even though synchronous TDM can provide this capability by assigning multiple time slots, the smallest unit of data in synchronous TDM is the slot itself and there is always a possibility that a slot remains only partially filled, thus wasting the channel capacity.

In addition to the above benefits of using statistical TDM, there are some drawbacks associated with it as well. Some of these are:

1. The data has to be stored in buffers that requires additional cost of memory.
2. Due to buffer storage, there will be delay distortion introduced. This would deteriorate the quality of real-time data.
3. There is an additional overhead attached to each frame that helps to identify the boundaries, addressee/addresser and other information about data. At higher loads, this header cuts down on the efficiency of the line.

The synchronous TDM has the following benefits over the other.

1. Once slot allocation has occurred, there is a fixed relation between a data source and its slots. This is analogous to circuit allocation in circuit switched network.

2. There is no extra delay distortion, which makes it ideal for voice-like communication.

There are tradeoffs in terms of some disadvantages, such as:

1. In conditions of low traffic volume from all or some data sources, the link utilization is lower than statistical TDM.
2. The multiplexer and traffic sources all have to be synchronized. Usually, the way this is done is by having a central clock providing synchronization to the main multiplexer with clocks from each source synchronized with the main clock. The main clock runs at the line capacity, and is sometimes called as the *master clock.*

This is the real cost of synchronous TDM and offsets the memory cost of statistical TDM.

Local clocks of individual sources generate data locally. Data is multiplexed according to a predetermined order.

The multiplex cycle is divided into frames and slots, and one or more slots are assigned for individual data sources.

Chapter

4

Digital Modulation

Digital modulation is a process that impresses a digital symbol onto a carrier signal suitable for transmission. For short distance transmissions, base band modulation often called line coding is usually used. Figure 4.1 shows several base band modulation waveforms.
The first one is the non-return to zero-level (NRZ-L) modulation which represents a symbol 1 by a positive square pulse with length T and a symbol 0 by a negative square pulse with length T. The second one is the uni-polar return to zero modulation with a positive pulse of T/2 for symbol 1 and nothing for 0. The third is the bi-phase level or Manchester modulation which uses a waveform consisting of a positive first-half T pulse and a negative second-half T pulse for 1 and a reversed waveform for 0.

For long distance and wireless transmissions, band pass modulation is usually used. Band pass modulation is also called carrier modulation. Digital base bands with base band voltage changes only between fixed levels are used to alter the parameters of a high-frequency sinusoidal carrier signal. The most common such base band is a binary digit (bit) stream giving the application its usual name of binary modulation.

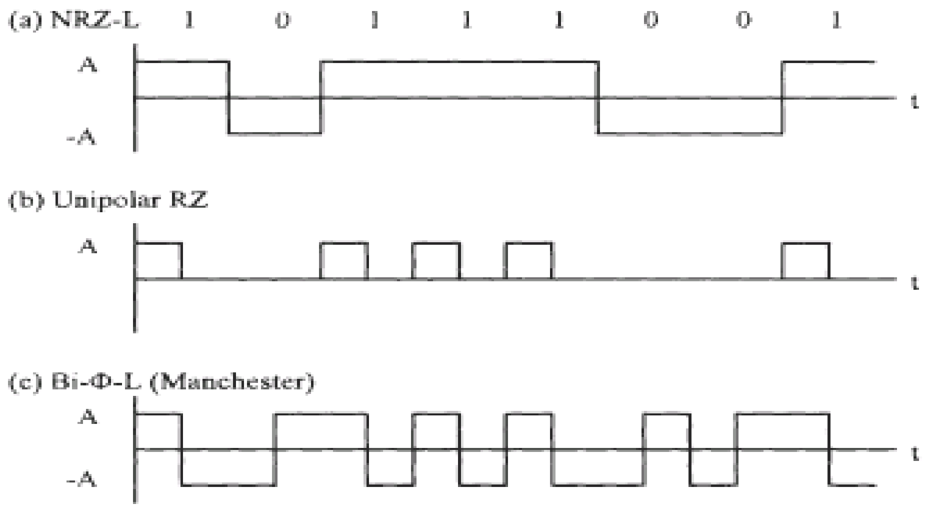


Figure 4.1 Base band digital modulation examples.

The digital modulation schemes are classified into two large categories: constant envelope and non-constant envelope. Under constant envelope class, there are three subclasses: Frequency Shift Keying, FSK, Phase Shift Keying, PSK, and Continuous Phase Modulation, CPM. Under non-constant envelope class, there are three subclasses: Amplitude Shift Keying, ASK, Quadrature Amplitude Modulation, QAM, and other non-constant envelope modulations.

Among the schemes, ASK, FSK, and PSK are basic modulations, and Minimum Shift Keying, MSK, Gaussian Minimum Shift Keying, GMSK, and QAM, etc. are advanced schemes. The advanced schemes are variations and combinations of the basic schemes, where;

where;

$$\mathbf{V_c(t) = E_c \, Sin\,(2\pi f_c\, t + \theta)}$$

ASK **FSK** **PSK**

QAM

The constant envelope class is generally suitable for communication systems whose power amplifiers must operate in the nonlinear region of the input-output characteristic in order to achieve maximum amplifier efficiency as TWTA (travelling wave tube amplifier) in satellite communications. However, the generic FSK schemes in this class are inappropriate for satellite application since they have very low bandwidth efficiency in comparison with PSK schemes. Binary FSK, BFSK was used in the low-rate control channels of AMPS (advanced mobile phone service) cellular systems, and ETACS (European total access communication system). The data rates are 10 Kbps for AMPS and 8 Kbps for ETACS.

The PSK schemes, including BPSK, QPSK, Offset Quadrature PSK, OQPSK, and MSK have been used in satellite communication systems.

QPSK is worth special attention due to its ability to avoid 180^{o} abrupt phase shift and to enable differential demodulation. It has been used in digital cellular mobile systems, such as the United States digital cellular (USDC) system.

The PSK schemes have constant envelope but discontinuous phase transitions from symbol to symbol. The CPM schemes have not only constant envelope, but also continuous phase transitions. Thus they have less side lobe energy in their spectra in comparison with the PSK schemes. MSK and GMSK are two important schemes in CPM class. MSK is a special case of CPFSK, but it also can be derived from OQPSK with extra sinusoidal pulse-shaping. MSK has excellent power and bandwidth efficiency. Its modulator and demodulator are also not too complex. MSK has been used in NASA's Advanced Communication Technology Satellite (ACTS).

GMSK has a Gaussian frequency pulse. Thus, it can achieve even better bandwidth efficiency than MSK. GMSK is used in the US cellular digital packet data (CDPD) system and European GSM (global system for mobile communication) system.

The generic non-constant envelope schemes, such as ASK and QAM, are generally not suitable for systems with nonlinear power amplifiers. However, QAM, with a large signal constellation, can achieve extremely high bandwidth efficiency.

QAM has been widely used in modems used in telephone networks, such as computer modems. QAM can even be considered for satellite systems. In this case, however, back-off in TWTA'S input and output power must be provided to ensure the linearity of the power amplifier.

4.1 Criteria of Choosing Modulation Schemes

The essence of digital modem design is to efficiently transmit digital bits and recover them from noise corruptions and other channel impairments. The choice of modulation technique has a direct impact on the capacity of the digital communication system. It determines the bandwidth efficiency of a single physical channel in terms of the number of bits/ second / hertz. The performance of a modulation scheme can be characterized by the bandwidth of the modulated signal, immunity to added noise and the complexity of the modulator / demodulator hardware. Based on these performance criteria,
a modulation technique has to be chosen for a given application.
How well a modulation scheme performs on a noisy channel is characterized by the Bit Error Rate (BER). The BER is related to the signal-to-noise ratio (SNR). For a given BER, say 10^{-3}, the modulation technique that requires the least SNR is the best.
Consideration must be given to achieve the following performance criteria by which a suitable digital modulation technique can be selected;

● High bandwidth efficiency

● High power efficiency

● Low carrier-to-co-channel interference power ratio (CCI)

● Low out-of-band radiation

● Robust to multi-path effects

● Low cost and ease of implementation

● Constant or none constant envelope

– Constant: only phase is modulated

– Non-constant: phase and amplitude modulated

This section covers the relative spectral efficiencies and some variations of the main modulation types as used in practical systems. Fortunately, there are a limited number of modulation types which form the building blocks of any system. The three primary criteria of choosing modulation schemes are: power efficiency, bandwidth efficiency, and system complexity.

4.1-1 Power Efficiency

It is the ability of a modulation technique to preserve the fidelity of the digital message at low power levels. Designer can increase noise immunity by increasing signal power.

Power efficiency is a measure of how much signal power should be increased to achieve a particular BER for a given modulation scheme. In the past both the modulator and demodulator were implemented completely in hardware, though nowadays, with the advent of digital signal processors modulator and demodulator implementations are software oriented, and hence a lot of software is used.

The bit error rate, or bit error probability of a modulation scheme is inversely related to E_b/N_o, signal energy per bit / noise power spectral density. For example, P_e of ASK in the AWGN channel is given by;

$$P_e = Q\left(\sqrt{\frac{2E_b}{N_o}}\right) \tag{4.1}$$

where Q (y) is the Gaussian integral, sometimes referred to as the Q-function. It is defined as;

$$Q\,(y) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2}\, du \tag{4.2}$$

which is a monotonically decreasing function of y.

The following table 4.1 illustrates the values of Q (y) and erfc (y) functions.

Therefore the power efficiency of a modulation scheme is defined straightforwardly as the required E_b/N_o for a certain bit error probability (P_e) over an AWGN channel. $P_e = 10^{-5}$ is usually used as the reference bit error probability.

Table 4.1 Values of Q (y) and erfc (y) functions

Definition: $Q(y) = \int_y^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$

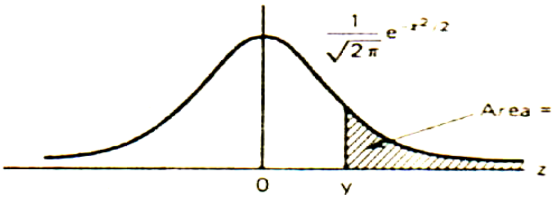
(1) $P(X > \mu_x + y\sigma_x) = Q(y) = \int_y^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$

(2) $Q(0) = \frac{1}{2}$; $\quad Q(-y) = 1 - Q(y)$, when $y \geq 0$

(3) $Q(y) \approx \frac{1}{y\sqrt{2\pi}} e^{-y^2/2}$ when $y \gg 1$ (approximation may be used for $y > 4$)

(4) $\text{erfc}^*(y) \triangleq \frac{2}{\sqrt{\pi}} \int_y^\infty e^{-z^2}\, dz = 2Q(\sqrt{2}y)$, $y > 0$

(5) $\text{erfc}(y) = 1 - \text{erf}(y)$



y	$Q(y)$	y	$Q(y)$	y	$Q(y)$	$Q(y)$	y
0.05	0.4801	1.05	0.1469	2.10	0.0179		
0.10	0.4602	1.10	0.1357	2.20	0.0139		
0.15	0.4405	1.15	0.1251	2.30	0.0107		
0.20	0.4207	1.20	0.1151	2.40	0.0082	10^{-3}	3.10
0.25	0.4013	1.25	0.0156	2.50	0.0062		
0.30	0.3821	1.30	0.0968	2.60	0.0047		
0.35	0.3632	1.35	0.0885	2.70	0.0035		
0.40	0.3446	1.40	0.0808	2.80	0.0026	$\frac{10^{-3}}{2}$	3.28
0.45	0.3264	1.45	0.0735	2.90	0.0019		
0.50	0.3085	1.50	0.0668	3.00	0.0013		
0.55	0.2912	1.55	0.0606	3.10	0.0010	10^{-4}	3.70
0.60	0.2743	1.60	0.0548	3.20	0.00069		
0.65	0.2578	1.65	0.0495	3.30	0.00048		
0.70	0.2420	1.70	0.0446	3.40	0.00034		
0.75	0.2266	1.75	0.0401	3.50	0.00023		
0.80	0.2119	1.80	0.0359	3.60	0.00016	$\frac{10^{-4}}{2}$	3.90
0.85	0.1977	1.85	0.0322	3.70	0.00010		
0.90	0.1841	1.90	0.0287	3.80	0.00007	10^{-5}	4.27
0.95	0.1711	1.95	0.0256	3.90	0.00005		
1.00	0.1587	2.00	0.0228	4.00	0.00003	10^{-6}	4.78

4.1-2 Bandwidth Efficiency

It describes how efficiently the allocated bandwidth is utilized or the ability of a modulation scheme to accommodate data within a limited bandwidth.

It is defined as the throughput, measured for any digital modulation technique as the ratio of the transmission bit rate to the minimum bandwidth required for a particular modulation scheme, R_b/W (f_b/W).

It indicates the number of bits that can be propagated through the transmission medium for each Hertz of system bandwidth. It is sometimes called information density or spectral efficiency, often used to compare the performance of one digital modulation technique to another.

Assuming the system uses Nyquist (ideal rectangular) filtering at base band, which has the minimum bandwidth required for inter-symbol interference-free transmission of digital signals, then the bandwidth at base band is 0.5 f_S, f_S is the symbol rate, and the bandwidth at carrier frequency is $W = f_S$,. Since, $f_S = f_b / \log_2 M$, for M-ary modulation, the bandwidth efficiency is;

$$f_b / W = \log_2 M \tag{4.3}$$

In binary systems, one transmitted symbol contains1bit of information.The theoretical limit (Nyquist rate), is 2 symbol/s/Hz.

For binary signals, this means that we can transmit only 2 b/s/Hz. For f_b = 10 Mbps rate, binary transmission system requires a theoretical minimal BW (Nyquist BW) of 5 MHz, and a practical BW of about 6 MHz. Many system applications require a considerably higher spectral efficiency than is possible with binary signaling. An increased spectral efficiency can be achieved by channel encoding (conversion) of the binary signal into an M-level signal. Binary systems are more power efficient, but less spectrally efficient than multistate M-ary systems.

Two equally bandwidth efficient methods for transmitting PAM are; (1) Single-side band PAM and (2) Quadrature PAM, in which the information sequence is splitted into two parallel sequences that are transmitted via PAM on the quadrature carriers Cos $\omega_c t$ and Sin $\omega_c t$. For both methods the required channel bandwidth W is approximately equal to 1/2T, since $1/T = R_b / n = R_b / \log_2 M$ symbol/sec, we obtain the result,

$$W = R_b \; / \; 2 \log_2 M$$

and the bandwidth efficiency for PAM as measured by the ratio R_b/W in b/s/Hz;

$$\frac{R_b}{W} = 2 \log_2 M \tag{4.4}$$

Example 4.1; to generate an M = 4-level signal, a two bits modulator (4-PAM) will take two consecutive bits at a time to form a symbol, and will map the signal amplitude to one of four possible levels, for example -3 volts, -1 volt, 1 volt, and 3 volts. Thus, the output symbol rate f_S of the M = 4-level signal is calculated as;

$f_S = 1/ T_S = 1/ 2T_b = f_b / 2$. Transmission at the Nyquist rate (2 symbols/s/Hz) leads to a spectral efficiency of 4 b/s/Hz. Similarly, in an 8-PAM system, 3 bits form a symbol; that is, $T_S = 3\ T_b$. For this reason, the theoretical spectral efficiency of an 8-level PAM system is 6 b/s/Hz. Similarly, 16-PAM gives, $T_S = 4\ T_b$, with 8 b/s/Hz.

Example 4.2; a- Find the minimum required bandwidth for the base band transmission of a four level PAM pulse sequence having a data rate of R_b = 2400 bits/sec. If the system has transfer characteristics with 100% excess bandwidth.

b- The same 4-ary PAM sequence is modulated onto a carrier wave, so that the base-band spectrum is shifted and centred at frequency f_c. Find the minimum required DSB bandwidth for transmitting the modulated PAM sequence, using the same transfer function as given in (a).

Solution; a- $M = 4 = 2^2 = 2^n$, symbol or pulse rate is;

$$R_S = \frac{R_b}{n} = \frac{2400}{2} = 1200\ symbol/\sec$$

Minimum bandwidth $W_N = ½ (1 + α) R_S = ½ (2) (1200)$

$$= 1200 \text{ Hz}$$

b- $W_{DSB} = (1 + α) R_S = (2) (1200) = 2400$ Hz

Example 4.3, It is required to transmit a bit rate of f_b = 90 Mb/s with an 8-PAM system, then the minimal Nyquist channel BW is;

$f_N = f_S /2 = (f_b /3)/2 = 90/6 = 15$ MHz. Assuming a roll-off (attenuation) factor of $α = 0.2$. The filter guard-band attenuation reaches about 50 dB at;

$$f_c = f_N\ (1+ α) = 15 \text{ MHz } (1.2) = 18 \text{ MHz}$$

The practical spectral efficiency of this system is;

$$f_b / f_G = 90 \text{ Mb/s} / 18 \text{ MHz} = 5 \text{ b/s/Hz}.$$

This spectral efficiency is 20 % below the ideal spectral efficiency of Nyquist channel.

4.1-3 System Complexity

System complexity refers to the amount of circuits involved and the technical difficulty of the system. Associated with the system complexity is the cost of manufacturing, which is of course a major concern in choosing a modulation technique.

Usually the demodulator is more complex than the modulator.

Coherent demodulator is much more complex than non-coherent demodulator since carrier recovery is required. The carrier frequency and phase of the received modulated signal must be precisely established.

Non-coherent (envelope) detection: requires no reference wave. It does not exploit phase reference information.

Also, it has less complex receiver, but worse performance.

The modulation techniques employed non-coherently are; ASK, FSK, Differential PSK (DPSK), CPM, and hybrids.

4.2 Trading off Simplicity and Bandwidth

When engineers speak of network bandwidth, they're referring to the practical frequency limit of a network medium. In serial communication, bandwidth is a product of data volume (binary bits per transmitted "word") and data speed ("words" per second). The standard measure of network bandwidth is bits per second, or bps. There is a fundamental trade off in communication systems. Simple hardware can be used in transmitters and receivers to communicate information. However, this uses a lot of spectrum which limits the number of users. Alternatively, more complex transmitters and receivers can be used to transmit the same information over less bandwidth. The transition to more and more spectrally efficient transmission techniques requires more and more complex hardware. Complex hardware is difficult to design, test, and build. This trade off exists whether communication is over air or wire, analogue or digital.

Figure 4.2 shows three basic binary carrier modulations.

They are amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK). In ASK, (OOK), the modulator puts out a burst of carrier for every symbol 1, and no signal for every symbol 0. In a general ASK scheme, the amplitude for symbol 0 is not necessarily 0. In FSK, for symbol 1 a higher frequency burst is transmitted and for symbol 0 a lower frequency burst is transmitted, or vice versa. In PSK, a symbol 1 is transmitted as a burst of carrier with 0 initial phase while a symbol 0 is transmitted as a burst of carrier with 180^{o} initial phase.

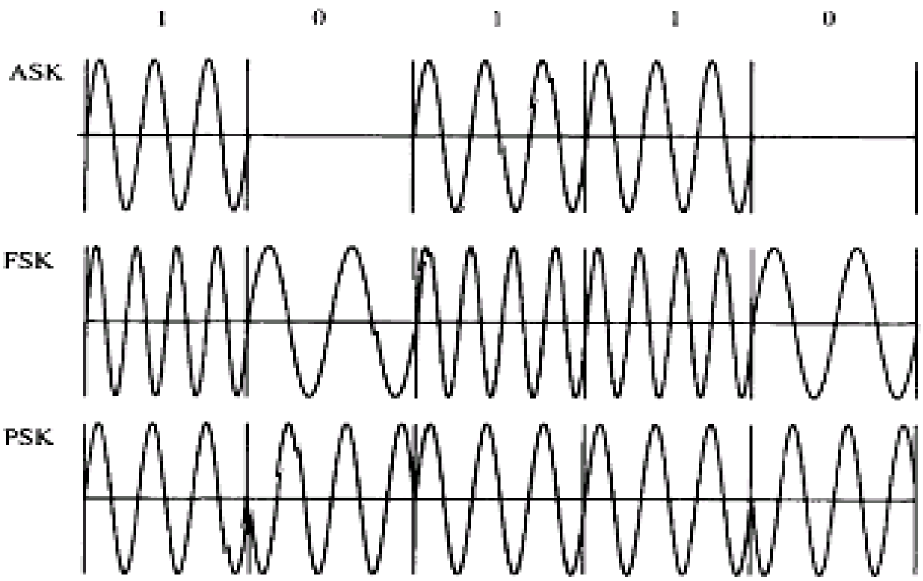


Figure 4.2 Three basic band pass modulation schemes

Based on these three basic schemes, a variety of modulation schemes can be derived from their combinations.

For example, by combining two binary PSK (BPSK) signals with orthogonal carriers a new scheme called quadrature phase shift keying

(QPSK) can be generated. By modulating both amplitude and phase of the carrier, we can obtain a scheme called quadrature amplitude modulation (QAM).

4.3 Digital modulation Applications

The following table 4.2 covers the main digital modulation formats, their main applications in both wireless communications and video.

Table 4.2 Various digital modulation and its application

Modulation Format	Application
MSK, GMSK	GSM, CDPD
BPSK	Deep Space Telemetry
QPSK, π/4 DQPSK	Satellite, CDMA, TETRA
OQPK	CDMA, Satellite
FSK, GFSK	DECT, Paging, AMPS, CT2, Land mobile, Public Safety
8 PSK	Satellite, aircraft
16 QAM	Microwave digital radio Modem
32 QAM	Terrestrial Microwave

4.4 Amplitude Shift Keying, ASK

In ASK, the base band digital signal is used to modulate the amplitude of an analogue carrier. One way to perform ASK would be to transmit the carrier signal with constant amplitude for one level of digitally encoded base band signal (e.g. NRZ) voltage and sending nothing for the other level. This is known as a binary ASK where, a bit stream of ones and zeros representing the modulating signal, is used to obtain ASK signal. The binary digit 1 is represented by the presence of the carrier and digit 0 is represented by the absence of the carrier. Figure 4.3, shows the corresponding ASK modulated signal.

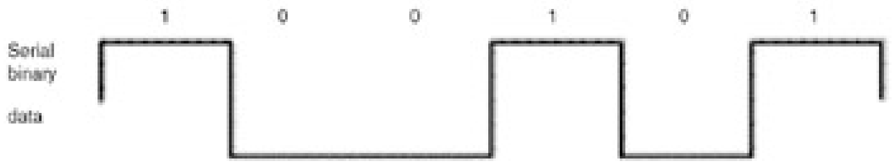


Figure 4.3-a Modulating signal.

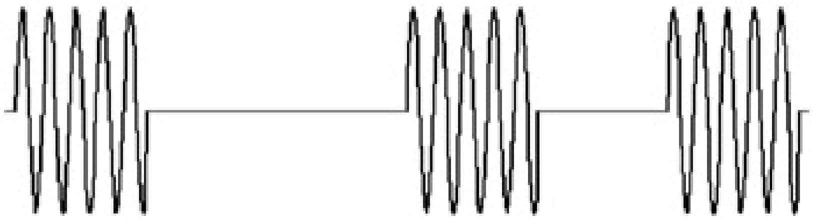


Figure 4.3-b ASK signal.

Accordingly, the ASK signal can be mathematically represented by;

$$S(t) = A \sin \omega_c t \quad \text{for binary } 1 \qquad (4.5\text{-a})$$

$$S(t) = 0 \quad \text{for binary } 0 \qquad (4.5\text{-b})$$

Fourier series is used to analyze the binary base band signal to obtain its components such as;

$$S_1(t) = \frac{1}{2} + \frac{2}{\pi}\left[Cos\,\omega_o t - \frac{1}{3} Cos\,3\omega_o t + \frac{1}{5} Cos\,5\omega_o t - \ldots \right] \qquad (4.6)$$

for uni-polar bit stream, but for bipolar bit stream;

$$S_2(t) = \frac{4}{\pi}\left[Cos\,\omega_o\,t - \frac{1}{3}Cos\,3\,\omega_o t + \frac{1}{5}Cos\,5\,\omega_o\,t - \ldots\right] \qquad (4.7)$$

where ω_o, is the fundamental frequency of the bit stream.

Figure 4.4 illustrates the formation of ASK signal that obtained by multiplying the base band signal by the carrier signal;

$$V_C(t) = \text{Cos}\ \omega_C\ t$$

Then;

$$V_{ASK} = V_c\ (t) \text{ x } S_1\ (t)$$

$V_{ASK} = \text{Cos}\ \omega_C t \text{ x}$

$$\left(\frac{1}{2} + \frac{2}{\pi}\left[Cos\,\omega_o\,t - \frac{1}{3}Cos\,3\,\omega_o t + \frac{1}{5}Cos\,5\,\omega_o\,t - \ldots\right]\right) \qquad (4.8\text{-a})$$

$$V_{ASK} = \frac{1}{2}Cos\,\omega_c\,t + \frac{2}{\pi}Cos\,\omega_c t\,Cos\,\omega_o t - \frac{2}{3\pi}Cos\,\omega_c t\,Cos\,3\,\omega_o t + \ldots.$$

$$\begin{aligned} &= \frac{1}{2}Cos\,\omega_c\,t + \frac{1}{\pi}Cos\,(\omega_c - \omega_o)t + \frac{1}{\pi}Cos\,(\omega_c + \omega_o)\,t \\ &\quad - \frac{1}{3\pi}Cos\,(\omega_c - 3\,\omega_o)t - \frac{1}{3\pi}Cos\,(\omega_c + 3\,\omega_o)t \qquad (4.8\text{-b}) \\ &\quad + \frac{1}{5\pi}Cos\,(\omega_c - 5\,\omega_o)t + \frac{1}{5\pi}Cos\,(\omega_c + 5\,\omega_o)t - \ldots.. \end{aligned}$$

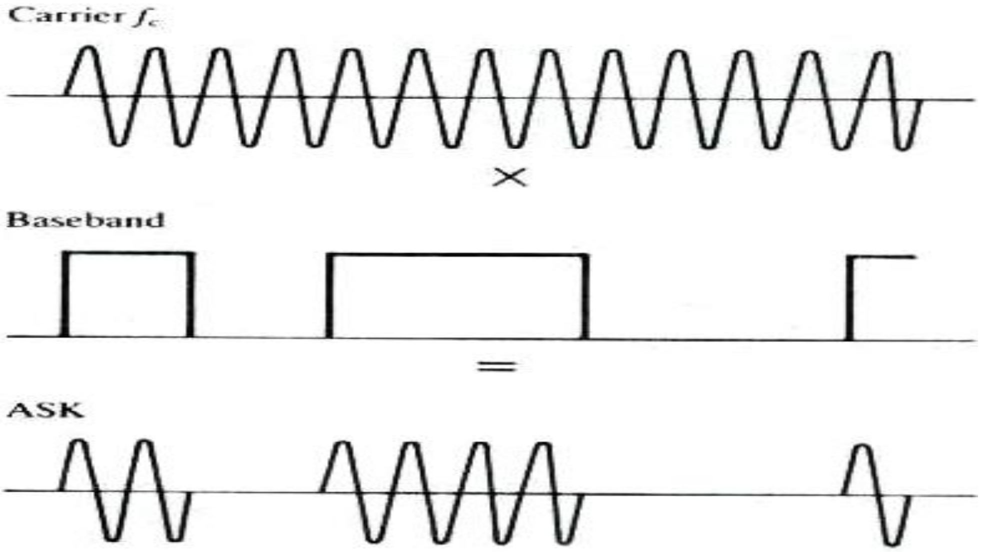


Figure 4.4 Formation of ASK-modulated bit stream

Figure 4.5 illustrates the frequency spectrum of the ASK signal for a bit rate $1 / T_b = f_o / 2 = 1 / 2$ T. It is possible to use more than two values for the amplitude of the modulated signal. The bandwidth of ASK signal which needs to be transmitted depends on the requirement of the application. Using at least the third harmonic sideband pair ($f_c \pm 3 f_o$), then ASK bandwidth is given by the formula;

$$BW_{ASK} = 6 f_o = 3 \text{ (bit rate maximum)} \qquad (4.9)$$

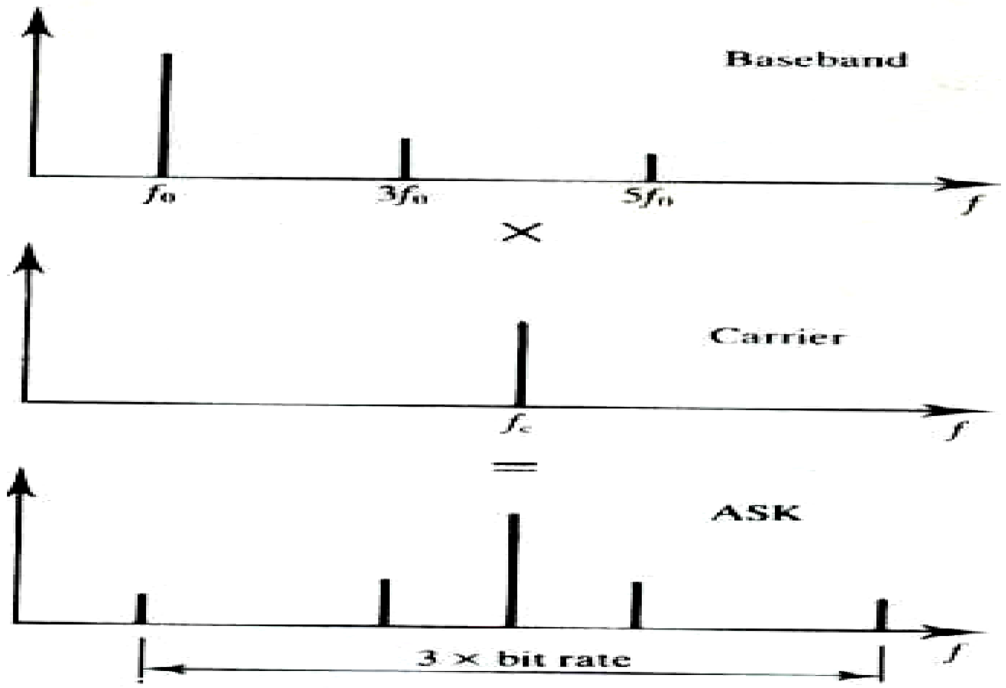


Figure 4.5 ASK spectrum

ASK is rarely used nowadays at audio and radio frequencies as it is susceptible to noise, and is not used on cables.

It is used in optical fibre communication because the noise is less.

Obviously spectral efficiency depends on the requirement of system bandwidth for a certain modulated signal. For example, the one-sided power spectral density of an ASK signal modulated by an equi-probable independent random binary sequence is given by;

$$\psi_s(f)=\frac{E_C^2\,T}{4}Sinc^2\left[T\left(f-f_c\right)\right]+\frac{E_C^2}{4}\delta\left(f-f_c\right) \qquad (4.10)$$

and is shown in figure 4.6, where T is the bit duration, E_C is the carrier amplitude, and f_C is the carrier frequency.

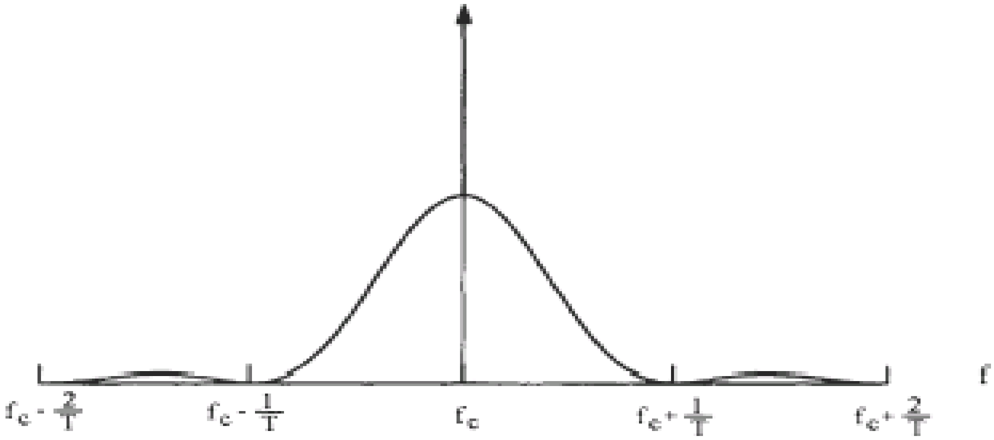


Figure 4.6 Power spectral density of ASK.

From the figure we can see that the signal spectrum stretches from - ∞ to ∞. Thus to perfectly transmit the signal an infinite system bandwidth is required, which is impractical. The practical system bandwidth requirement is finite, which varies depending on different criteria. For example, in figure 4.6, most of the signal energy concentrates in the band between two nulls, thus a null-to-null bandwidth requirement seems adequate.

4.5 Frequency Shift Keying, FSK

It is a form of frequency modulation in which the modulating signal shifts the output frequency between predetermined values. In binary FSK, two signals with different carrier frequencies are used to represent binary data. Since these carriers have different frequencies, they can easily be distinguished at the receiver. The binary digits are represented by two different frequencies close to the carrier frequency as shown in figure 4.7.

A BFSK signal is mathematically represented by;

$$S_1\,(t) = A\ \mathrm{Sin}\,(2\,\pi\, f_1)\,t \quad \text{for binary} \quad 1 \qquad (4.11\text{-a})$$

$$S_2\,(t) = A\ \mathrm{Sin}\,(2\,\pi\, f_2)\,t \quad \text{for binary} \quad 0 \qquad (4.11\text{-b})$$

f_1 can be $f_c + f_m$ and f_2 can be f_c - f_m, where f_c is the carrier frequency and $2f_m$ is the frequency deviation.

The simplest way to analyze the FSK modulation is to treat it as two separate interleaved digit streams- one for the ones and one for the zeros. Then each is ASK modulated at a different frequency and the two added to give the complete FSK waveform. This process is shown in figure 4.8.

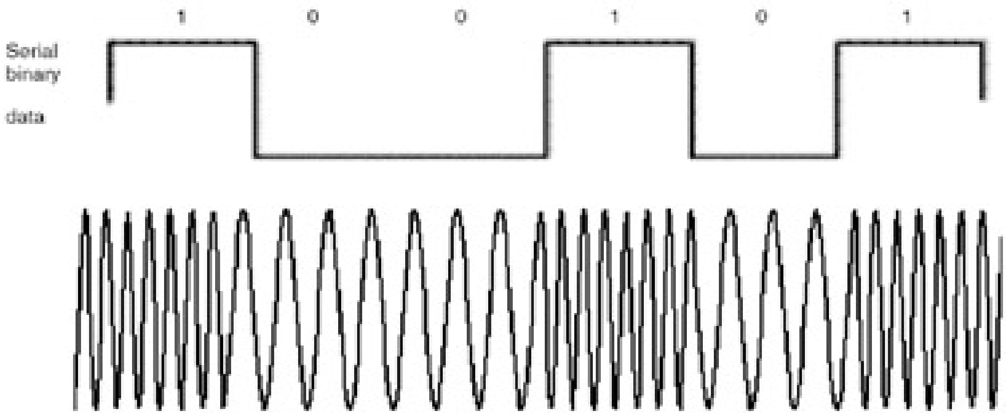


Figure 4.7 FSK signal

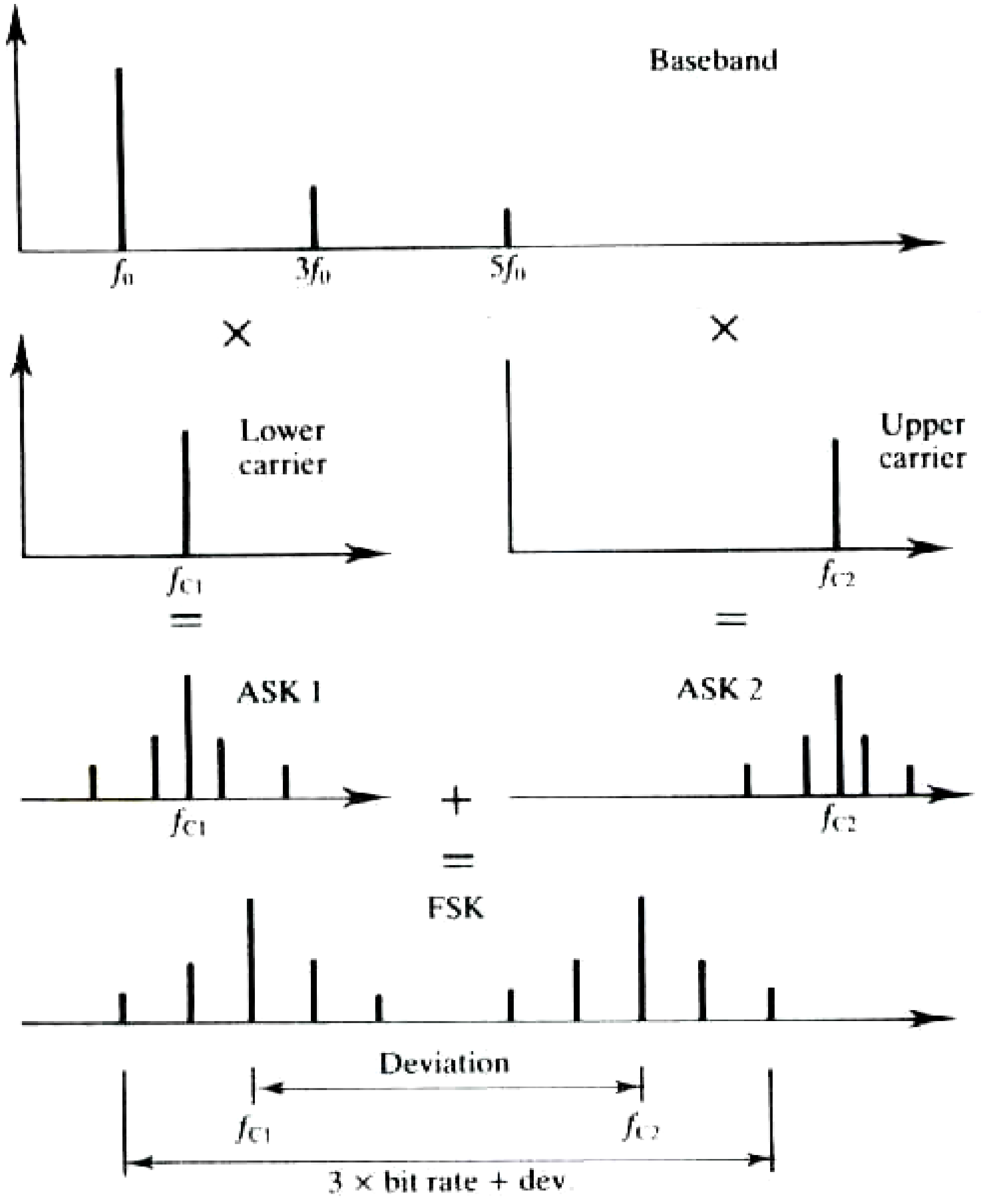


Figure 4.8 Formation of FSK signal as the sum of two ASK signals

So,

$V_{FSK} = \text{Cos}\ \omega_1 t \times S_1(t) + \text{Cos}\ \omega_2 t \times S_2(t)$

where $S_2(t)$ is the complement of $S_1(t)$ so that,

$S_2(t) = 1 - S_1(t)$.

Therefore ;

$V_{FSK} = \text{Cos}\ \omega_1 t$ x

$$\left(\frac{1}{2}+\frac{2}{\pi}\left[Cos\,\omega_o\,t-\frac{1}{3}Cos\,3\,\omega_o t+\frac{1}{5}Cos\,5\,\omega_o\,t-\ldots\right]\right) \text{ + Cos}$$

$$\omega_2 t \text{ x } \left(\frac{1}{2}-\frac{2}{\pi}\left[Cos\,\omega_o\,t-\frac{1}{3}Cos\,3\,\omega_o t+\frac{1}{5}Cos\,5\,\omega_o\,t-\ldots\right]\right)$$

$\therefore$ $V_{FSK} =$

$$\begin{aligned}
&\frac{1}{2}Cos\,\omega_1\,t+\frac{1}{\pi}Cos\,(\omega_1-\omega_o)t+\frac{1}{\pi}Cos\,(\omega_1+\omega_o)\,t\\
&-\frac{1}{3\pi}Cos\,(\omega_1-3\,\omega_o)t-\frac{1}{3\pi}Cos\,(\omega_1+3\,\omega_o)t\\
&+\frac{1}{5\pi}Cos\,(\omega_1-5\,\omega_o)t+\frac{1}{5\pi}Cos\,(\omega_1+5\,\omega_o)t-\ldots..\\
&+\frac{1}{2}Cos\,\omega_2\,t-\frac{1}{\pi}Cos\,(\omega_2-\omega_o)t-\frac{1}{\pi}Cos\,(\omega_2+\omega_o)\,t\\
&+\frac{1}{3\pi}Cos\,(\omega_2-3\,\omega_o)t+\frac{1}{3\pi}Cos\,(\omega_2+3\,\omega_o)t\\
&-\frac{1}{5\pi}Cos\,(\omega_2-5\,\omega_o)t-\frac{1}{5\pi}Cos\,(\omega_2+5\,\omega_o)t+\ldots...
\end{aligned} \tag{4.12}$$

which is as expected, merely two ASK spectra centred around ω_1 and ω_2 and separated by $\Delta\,\omega = \omega_2 - \omega_1$, the frequency shift.

Figure 4.9 shows this spectrum for a bit rate $1/T_b$ of fundamental frequency $f_o = 1 / T = 1 / 2\,T_b$.

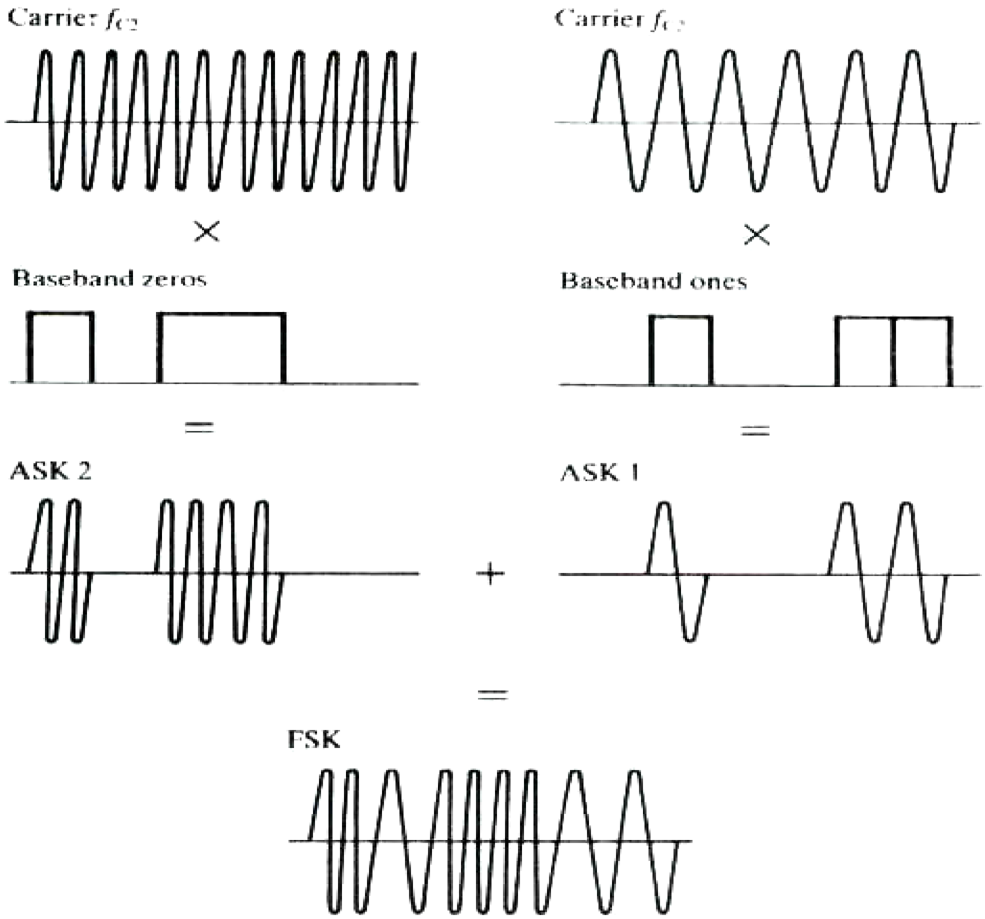


Figure 4.9 Formation of FSK spectrum

The bandwidth requirement of FSK signal using the third harmonic sideband pair is given by;

$$BW_{FSK} = 3\,f_o + 2\,\Delta f + 3\,f_o$$

$$= 6\, f_o + 2\, \Delta f$$

$$= 6\, (1 / 2T_b) + 2\, \Delta f$$

$$BW_{FSK} = 3\, (1 / T_b) + 2\, \Delta f$$

$$= 3\, (\text{bit rate}) + 2\, \Delta f \qquad (4.13)$$

This bandwidth is larger than that of the ASK signal by the amount of the frequency shift, which is an unavoidable disadvantage of FSK. But this is more than compensated by its main advantage which is that it is much less susceptible to corruption by unwanted amplitude modulation - due to noise or transients. FSK has the added benefit over the ASK that if a bit is lost during transmission, it is known that the bit is lost, since there should always be a carrier signal for zero or one.

The baud rate in this case is again equal to the bit rate. As compared to ASK, signal synchronization is easier to maintain in FSK. In ASK, it is possible for the receiver to lose sync with the transmitter when a long string of no-carrier signals occurs.

Spectral efficiency is calculated as R_b / W.

Example 4.4; For a BFSK signal with maximum frequency 51 kHz, and minimum frequency 49 kHz, and an input bit rate of 2 kbps, determine the following;

a- The peak frequency deviation,

b- Minimum bandwidth, and c- Baud rate

d- Spectral efficiency

e- Write down the FSK waveform equation when the carrier amplitude is 5 volt

Solution; a- The peak frequency deviation is determined as;

$$\Delta f = \frac{|f_{\max} - f_{\min}|}{2} = \frac{|51-49|}{2} = 1 kHz$$

b- The minimum band width is determined as;

$BW_{FSK} = 2\,(\Delta f + f_m)$, f_m , f_a is the fundamental frequency

$BW_{FSK} = 2\,(1000 + 1000) = 4$ k Hz

c- The baud rate for binary FSK (N =1) is determined as;

$$\text{baud rate} = \frac{f_b}{N} = \frac{f_b}{1} = f_b = 2\,kbps$$

d- Spectral efficiency = R_b/BW = 2000/4000 = 0.5 b/s/Hz

e- For the carrier signal with 5 volt amplitude, and frequency,

$f_c = f_{\max} - \Delta f$

$f_c = 51 - 1 = 50$ k Hz , then;

$$V_{FSK} = V_c \text{ Cos } 2\pi\,[f_c \pm \Delta f]\,t$$

$$V_{FSK} = 5 \text{ Cos } 2\pi\,[\,50 \times 10^3 \pm 1 \times 10^3\,]\,t$$

Example 4.4'; Calculate the frequency modulation index, deviation ratio, and BW required to transmit an 1200 b/sec. digital signal, when $f_0 = 21000$ Hz, and $f_1 = 13000$ Hz.

Solution; Tone separation = 21000 – 13000 = 8000 Hz = 2 Δf,

the frequency modulation index is calculated as;

$$\beta_f = \frac{\Delta f}{f_m} = \frac{\Delta f}{f_a} = \frac{4000}{600} = 6.67$$

when Δf is at a maximum, and f_m is at a maximum, then the modulation index is referred to as the deviation ratio, D, so for an FSK modulator, the deviation ratio, D is given by;

$$D = \frac{Tone\,separation}{bit\,rate} = \frac{8000}{1200} = 6.67 = \beta_f$$

The required BW is calculated as;

BW_{FSK} = Tone separation + bit rate = 8000 + 1200 = 9200 Hz.

When the BW is limited to the bit rate, then the receiver will still be able to recover the original data.

As the deviation ratio is approximately equal to the energy concentrated in the carrier and the first-order sidebands, it is not necessary to transmit all the sidebands.

4.5-1 Frequency-Shift Keying Modulator (Transmitter)

The FSK modulator of the modem has the block diagram shown in figure 4.10. A crystal controlled oscillator provides an accurate clock frequency which is fed into two divide-by-N circuits. The outputs are passed though narrow band pass filters tuned to the required frequencies f_1 and f_2. The outputs of the band pass filters go to the inputs of the multipliers. The multiplicands are the binary signal to be FSK-coded and its complement. Note that the binary signal and its complement are used to control the flow of signal from the band pass filters to the adder. When the binary input is a 1, a signal of frequency f_1 is fed to the adder; f_2 is cut off. When it is a 0, f_1 is cut off and f_2 appears at the input of the adder.

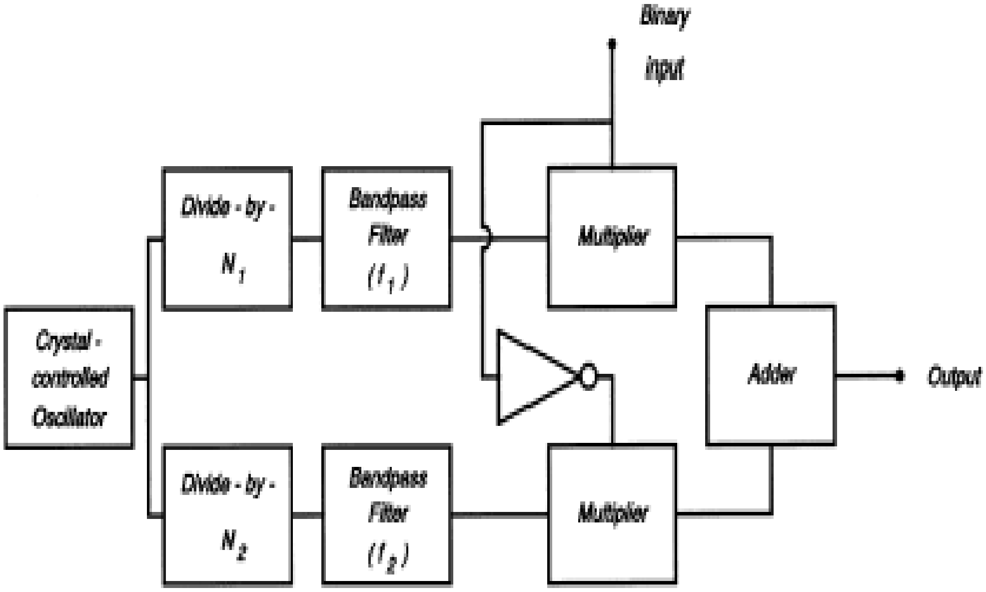


Figure 4.10 FSK modulator with a crystal-controlled oscillator driver

The divide-by-N circuits can be realized by using two 12-stage ripple counters (2^{12} = 4096) with suitable resets to get the required values of N. The band pass filters can be simple LC-tank circuits with modest Q factors tuned to 980 Hz and 1180 Hz, respectively. There are several other schemes which can be used to realize the FSK coder.

4.5-2 Frequency-Shift Keying Demodulator (Receiver)

A simple FSK demodulator of the modem is shown in figure 4.11. Two narrow band pass filters are tuned to the two frequencies present in the FSK signal. Their outputs are used to drive the envelope detectors. Note that the envelopes of the two frequencies f_1 and f_2 are both square waves.

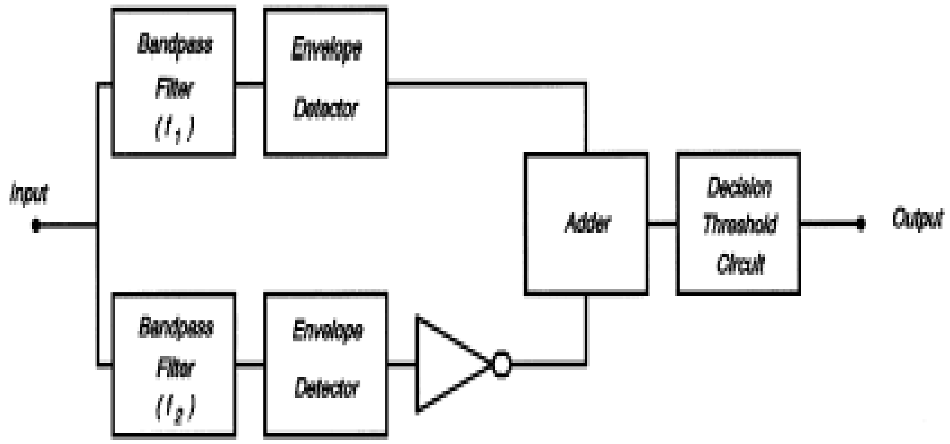


Figure 4.11 FSK Demodulator

The output of one of the envelope detectors is inverted and added to the other to give the required binary output.

The narrow band pass filters can be LC parallel-tuned amplifiers with modest Q factors.

There are several modems on the market which use sophisticated phase-locked loop techniques for both the modulator and demodulator; most use integrated circuits.

As said earlier, ASK is the simplest type of digital pass band modulations. The complexity increases for FSK and further for the PSK.

If M carriers are used for FSK with M > 2, then we have the M-ary FSK with the baud rate equal to 1/ n^{th} of the bit rate. FSK is used widely in cable and radio communication systems.

4.5-3 Multi-level FSK

The data throughput can be increased by using more than two frequencies. Thus using M levels or frequencies, a symbol can represent $\log_2$ M bits, i.e.; 4 frequencies will convey 2 bits per symbol. For 25 kb/s FSK information rate, the data rate required for 4-FSK is 6.25 kb/sec; thus requiring a symbol rate of 3.125 kbps, and use the four signalling frequencies;

$f_{oo} = f_c - 3\, f_b/2$; offset = - 4687.5 Hz

$f_{1o} = f_c - f_b/2$; offset = - 1562.5 Hz

$f_{11} = f_c + f_b/2$; offset = +1562.5 Hz

$f_{o1} = f_c + 3\, f_b/2$; offset = +4687.5 Hz

the data is shaped using a 10^{th} order Bessel filter, with a 3 dB point at 4 kHz, so that the carrier oscillator can move smoothly from one signalling frequency to the next within the symbol period, i.e., 1/3.125 msec.

4.6 Phase Shift Keying, PSK

It is the most frequently used digital modulation technique, where, the data information is imbedded in the phase of the carrier. The two commonly used PSK techniques are binary PSK (BPSK) and quaternary (quadrature) PSK (QPSK). PSK modulations with other alphabets are sometimes used; these are denoted by 8-PSK, 16-PSK, and so on, with the digit representing the number of phases. BPSK and QPSK are quadrature modulations and can be viewed as linear modulations; that is, their transmitted signal can be viewed as a superposition of independent pulses. In this section, we will focus on these two modulations and will explore their spectra, error probabilities, and other properties. Strictly speaking, PSK with $M > 4$ is neither quadrature nor linear.

4.6-1 Binary PSK, BPSK,

In BPSK, binary 1 and 0 are represented by two phases of the carrier, as the phase is measured relative to the previous bit interval, the baud rate and bit rate are equal. This is because the carrier phase (360 or 2π rad) is equally divided into two halves. Thus the information is contained in the phase of the modulated signal, that is mathematically, a BPSK signal is represented by;

$$V_{BPSK}(t) = A \text{ Cos } (\omega_C t + \theta (t)) \qquad (4.14)$$

where, $\theta(t) = 0\ (180)$ for binary 0

or $= 180\ (0)$ for binary 1

The same carrier frequency is used for both types of bits (0 and 1) but the phase is inverted for one or the other as shown in figure 4.12.

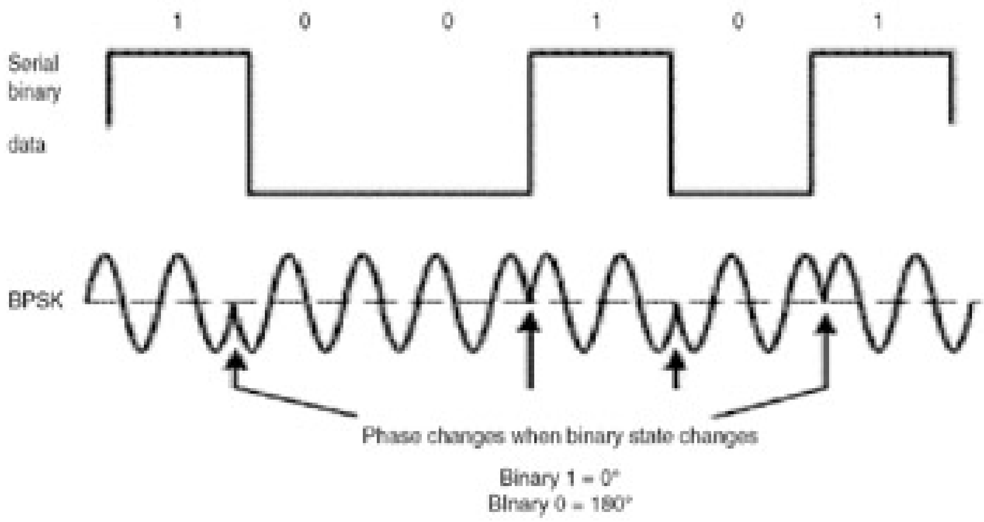


Figure 4.12 BPSK Signal

The bandwidth occupied by BPSK is the same as that of ASK, in addition to it is less susceptible to noise corruption.

Figure 4.13 shows the BPSK modulator. Note that we now need a bipolar base band because one phase is merely minus the other (180° phase change = multiplying by - 1). Multiplies the carrier with it to get BPSK as follows:

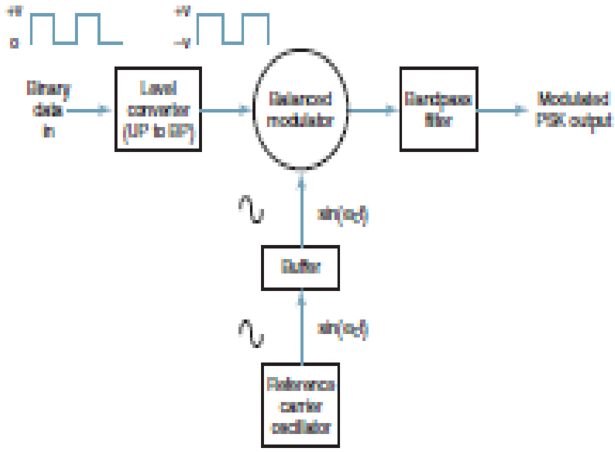


Figure 4.13 BPSK modulator

$$
\begin{aligned}
\upsilon_{BPSK} &= \upsilon_c \,.\, s(t) \\
&= Cos\ \omega_c\ t\ x\ \frac{4}{\pi}\left(Cos\,\omega_o\ t - \frac{1}{3}Cos\,3\omega_o\ t + \frac{1}{5}Cos\,5\omega_o\ t - \dots\right) \\
&= \frac{4}{\pi}\left(Cos\,\omega_c\ t.Cos\,\omega_o\ t - \frac{1}{3}Cos\,\omega_c\ t.Cos\,3\omega_o\ t + \dots\right) \\
&= \frac{2}{\pi}\left(\begin{array}{l} Cos(\omega_c + \omega_o)\ t + Cos(\omega_c - \omega_o)t - \frac{1}{3}Cos(\omega_c + 3\omega_o)t \\ -\frac{1}{3}Cos(\omega_c - 3\omega_o)t + \dots \end{array}\right)
\end{aligned}
\qquad (4.15)
$$

which is a series of diminishing sidebands as in figure 4.14

Obviously this is the same spectrum as ASK but without the carrier (the DSBSC version?) and its bandwidth may be calculated in the same way, so, if the third harmonic pair are included, we get:

$$BW_{BPSK} = 3\ (\text{bit rate}) \qquad (4.16)$$

As such it has become the only method used for digital transmission on microwave radio, although it is used at other frequencies too.

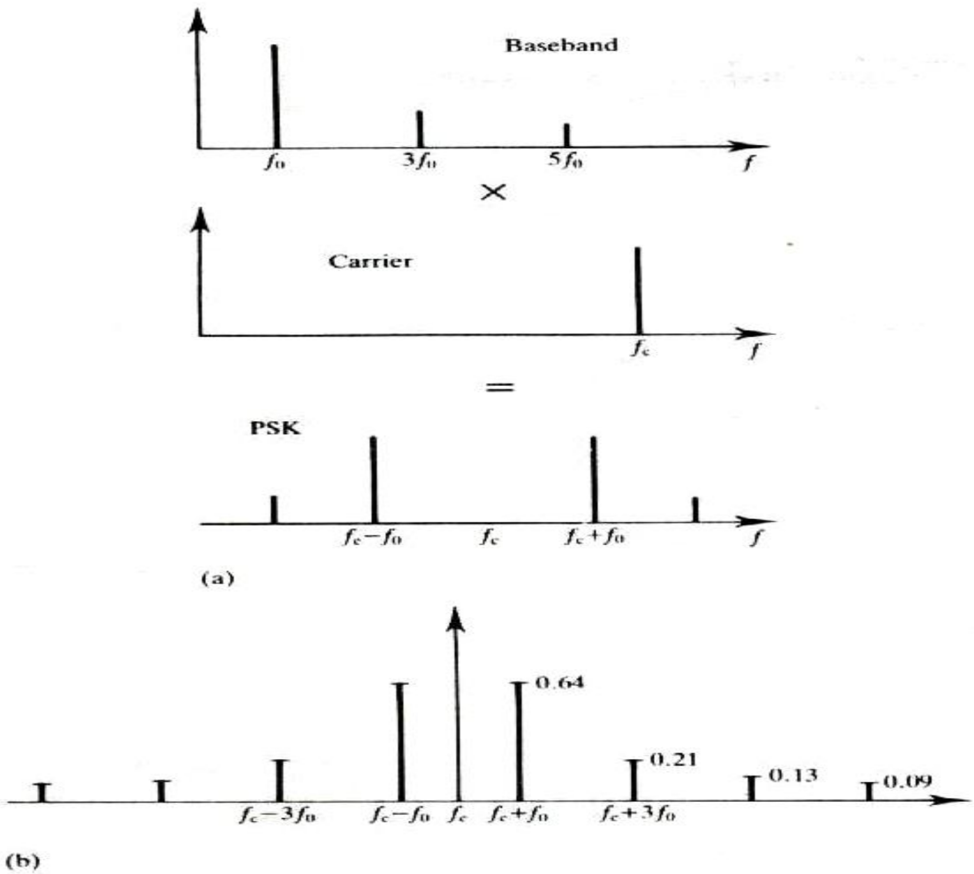


Figure 4.14 BPSK spectrum

Comparing bandwidth efficiency of PSK which equals;

$$\frac{R_b}{W} = \log_2 M \tag{4.17}$$

with that of PAM given by equation (4.4), we observe that PAM is a factor of 2 better by virtue of the fact that the PAM signal can be transmitted by SSB or, alternatively, by use of two independent quadrature carriers (DSB).

Example 4.5; For a BPSK modulator with a carrier frequency of 70 MHz and an input bit rate of 10 Mbps, determine the following;

a- Maximum and minimum upper and lower side frequencies, also draw the output spectrum

b- The minimum Nyquist bandwidth,

c- The baud rate.

d- spectral efficiency

Solution; a- The BPSK waveform has the following equation;

$$V_{BPSK} = \text{Sin}\,(2\,\pi\, f_a\, t) \text{ x } \text{Sin}\,(2\,\pi\, f_c\, t),$$

with f_a is the fundamental frequency of the base band input signal

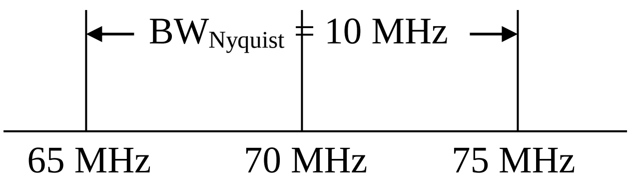
$$V_{BPSK} = \text{Sin}\,[2\,\pi\,(5 \text{ MHz})\, t] \text{ x } \text{Sin}\,[2\,\pi\,(70 \text{ MHz})\, t],$$

$$V_{BPSK} = \frac{1}{2}\text{Cos}\, 2\,\pi\,(70\text{-}5) \text{ MHz } t - \frac{1}{2}\text{Cos}\, 2\,\pi\,(70 + 5) \text{ MHz } t$$

Maximum upper side frequencies, USF = (70 + 5) M Hz

Minimum lower side frequency, LSF = (70 – 5) M Hz

Therefore, the output spectrum will be as shown;



b- The minimum Nyquist bandwidth;

$BW_{Nyquist}$ = 75 MHz – 65 MHz = 10 MH z

c- The baud rate = Symbol rate = baud per sec.

$$\text{baud rate} = \frac{f_b}{N} = \frac{10\,M\ Hz}{1} = 10\,M\ Hz$$

or Symbol rate = 10 M baud

For Binary modulation, baud rate = bit rate, bps. However, in multilevel modulation the bit rate is always greater than the baud rate. For BPSK,

f_b = BW x N = 10 x 10^6x 1 = 10 mega baud = 10^7 bps.

d-The theoretical spectral efficiency of BPSK modems is;

1 b/s/Hz. So BPSK is more efficient than BFSK system.

Using practical filters with approaching a roll-out factor of 0.3,

a spectral efficiency of 1 b/s/Hz /(1.3) = 0.77 b/s/Hz is achieved.

For coherent demodulation of BPSK signal a carrier frequency that is in synchronism with the received modulated wave is required. The carrier recovery (CR) system provides the receive multiplier with a sinusoidal frequency that has exactly the same frequency and the same phase as the transmitted un-modulated carrier wave.

The band limited received IF signal, $V_{BPSK}(t)$, is multiplied by the recovered carrier wave K Cos $\omega_c t$ as shown in figure 4.15. The receiver multiplier, followed by a LPF performs the coherent phase demodulation process. The received signal is described as;

$$V_{BPSK}(t) = r(t) = C_r \cos[\omega_c t + \theta(t)] \quad (4.18\text{-a})$$

Where C_r is the peak amplitude of the modulated carrier.

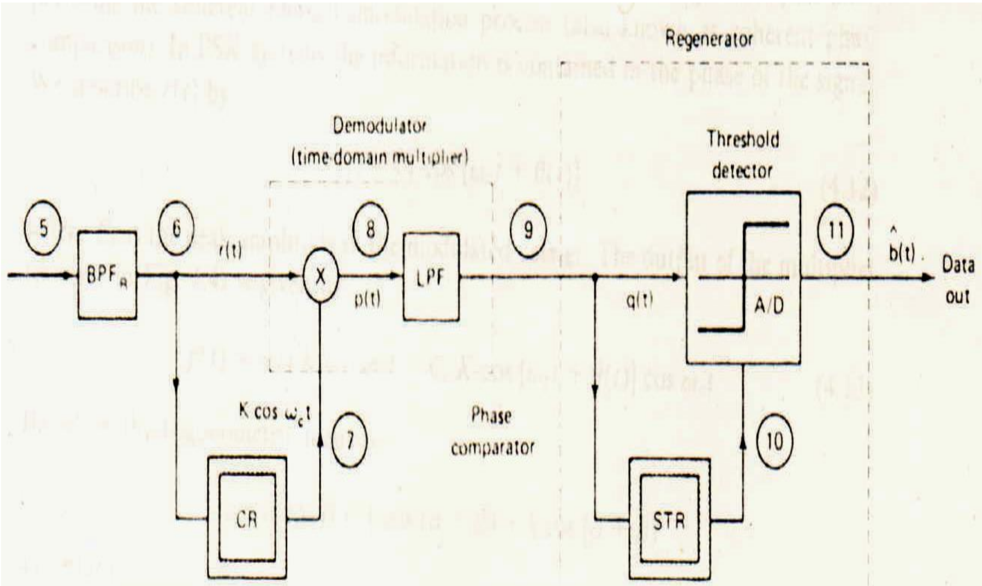


Figure 4.15 BPSK demodulator

The output of the multiplier is given by;

$$P(t) = r(t)\,\mathbf{k} \cos \omega_c t$$

$$= C_r \mathbf{k} \cos[\omega_c t + \theta(t)] \cos \omega_c t \qquad (4.18\text{-b})$$

$$\therefore P(t) = \tfrac{1}{2} C_r \mathbf{k} \cos[\omega_c t + \theta(t) - \omega_c t] +$$

$$\tfrac{1}{2} C_r \mathbf{k} \cos[\omega_c t + \theta(t) + \omega_c t]$$

$$= \tfrac{1}{2} C_r \mathbf{k} \cos[\theta(t)] + \tfrac{1}{2} \cos[2\omega_c t + \theta(t)] \quad (4.18\text{-c})$$

The receiver LPF removes the double-frequency spectral components. At the threshold comparator input, we have;

$$q(t) = \tfrac{1}{2} C_r \mathbf{k} \cos[\theta(t)] \qquad (4.18\text{-d})$$

In equation (4.18-d), $C_r k / 2$ represents a gain constant, while

$\cos[\theta(t)]$ is the time-variable band limited base band signal.

For $\theta(t) = 0^o$ or 180^o this signal equals +1 or -1, respectively. Finally, the one-bit A/D converter (threshold comparator) provides the digital output $\widehat{b}(t)$.

In terms of performance, ASK and FSK require twice as much power to attain the same bit error rate performance as phase shift keying (PSK). Consequently, the vast majority of mobile-satellite systems employ PSK.

In addition to the two advantages of PSK, the Bandwidth is reduced by multilevel schemes. All these make PSK a very desirable method for some applications - high bit rates and high carrier frequencies, for example.

In a multi-phase PSK (called M-ary PSK), the carrier phase may be divided into M equal parts. For M equal divisions of the phase, each of the M resulting signals will have one of the M values of phases between 0°and 360°. In M-ary PSK, the bit rate would be n times the baud rate, $n = \log_2(M)$.

4.6-2 Quadrature PSK, QPSK

It is a special case of MPSK, in which M = 4. Therefore, the carrier phase is divided into four distinct phases with difference of $[360 / M]^o = 360/4 = 90^o$ between adjacent signal phases. In this case, each symbol represents $\log_2 4 = 2$ bits. Two bits in the bit stream are taken, and four phases of the carrier frequency are used to represent the four combinations of the two bits; 11, 10, 00, and 01 such as;

$$S(t) = E_C \sin(\omega_C t + 45^o) \quad \text{for binary} \quad 11$$

$$S(t) = E_C \sin(\omega_C t + 135^o) \quad \text{for binary} \quad 10$$

$$S(t) = E_C \sin(\omega_C t + 225^o) \quad \text{for binary} \quad 00$$

$$S(t) = E_C \sin(\omega_C t + 315^o) \quad \text{for binary} \quad 01 \qquad (4.19)$$

Figure 4.16-a shows the modulator of the QPSK signal. The outputs of both Balanced modulators are BPSK signals. The I multiplier output signal has phase 0^o or 180^o relative to the carrier, and the Q multiplier output signal has phase 90^o or 270^o relative to the carrier. The multiplier outputs are then summed to give a four-phase signal. Thus QPSK can be regarded as two BPSK systems operating in quadrature. Note that either 90^o or 180^o phase transitions are possible. As an example, a 180^o phase transition would occur when the IQ digit

combination changed from 11 to 00. For unfiltered QPSK signal, phase transitions occur instantaneously and the signal has a constant-amplitude envelope. However, phase changes for filtered QPSK signals result in a varying envelope amplitude.

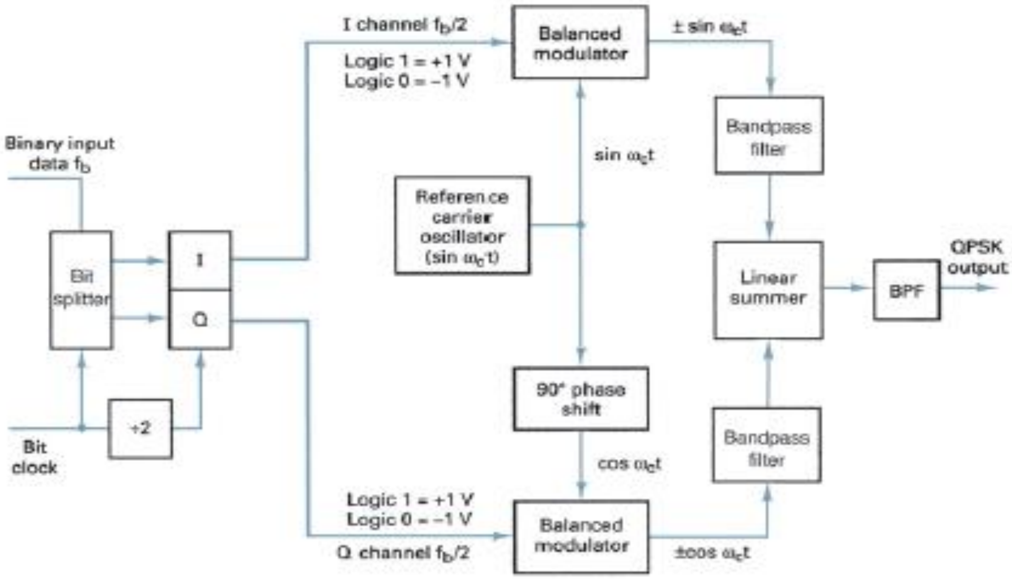


Figure 4.16-a QPSK modulator

The permitted values for the complex envelope are illustrated by the QPSK signal constellation shown in figure 4.16-b. The signal constellation is a plot of the permitted values for the complex envelope. QPSK may be generated by using the quadrature generation technique of figure 4.16-a, where the base band signal processor is a serial-to-parallel converter that reads in two bits of data at a time from the serial binary input stream, and outputs the first of the two bits to I channel and the second bit to Q channel. If the two input bits are both binary ones, (11), then I channel is E_C and Q channel is E_C. This is represented by the top right-hand dot for S(t) in the signal constellation for QPSK signaling in figure 4.16-b

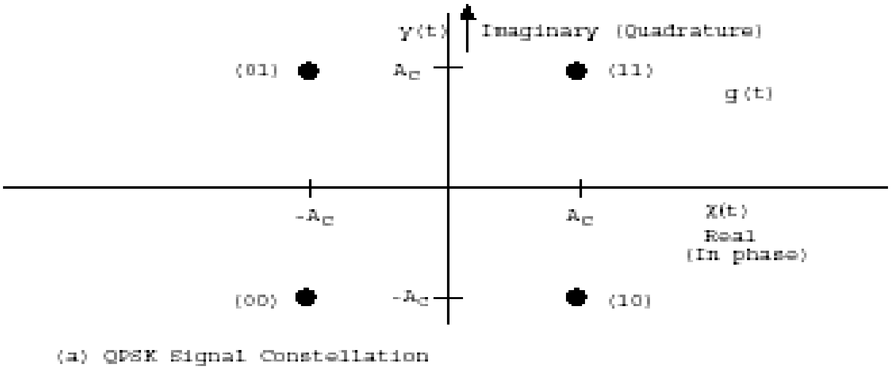


Figure 4.16-b. QPSK signal constellation

Likewise, the three other possible two-bit words, (10), (01), and (00), are also shown. The QPSK signal is also equivalent to a four-phase phase-shift-keyed signal (4PSK) since all the points in the signal constellation fall on a circle where the permitted phases are $\theta(t) =$ 45, 135, 225, and 315. There is no amplitude modulation on the QPSK signal since the distances from the origin to all the signal points on the signal constellation are equal.

In particular, a 180^o phase change results in a momentary change to zero in envelope amplitude. The QPSK signal at the modulator output is normally filtered to limit the radiated spectrum, then transmitted over the transmission channel to the receiver input. Because the I and Q modulated signals are in quadrature (orthogonal), the receiver is able to demodulate and regenerate them independently of each other, operating effectively as two BPSK receivers. The regenerated I and Q streams are then recombined in a parallel-to-serial converter to form the original input data streams; however, this stream is of course subject to error because of the effects of noise and filtering. The following figure 4.16-c illustrates the demodulation of QPSK signal.

The unfiltered (constant envelope) QPSK signal is shown in figure 4.16-d. QPSK provides twice the data throughput in the same bandwidth compared to BPSK. This is because the symbols in each quadrature channel occupy the same spectral space and have half the spectral width of a BPSK signal with the same data rate as the QPSK signal.

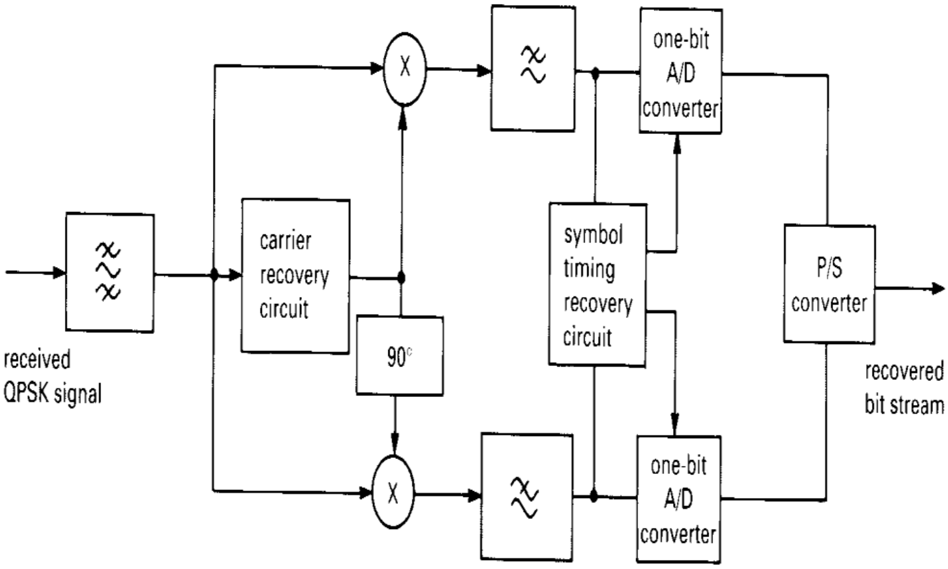


Figure 4.16-c Coherent QPSK demodulator

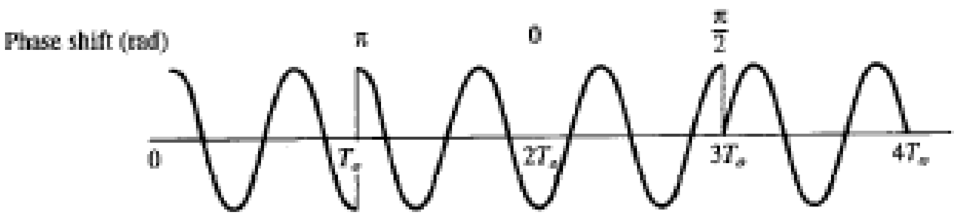


Figure 4.16-d Unfiltered QPSK signal (T_o = 2 T_b)

Example 4.6; For a QPSK modulator with a carrier frequency of 70 MHz and an input bit rate of 10 Mbps, determine the following;

a- The minimum double-sided Nyquist bandwidth, f_N

b- The baud rate.

c- Compare the results with those achieved with the BPSK modulator in example 4.5

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

$$f_{bQ} = f_{bI} = \frac{f_b}{2} = \frac{10\,M\ bps}{2} = 5\text{ M bps.}$$

The highest fundamental frequency ;

$$f_a = \frac{f_{bQ}}{2} = \frac{f_{bI}}{2} = \frac{5\,M\ bps}{2} = 2.5\,M\ bps$$

The output waveform from either balanced modulator is;

$$V_{QPSK} = \text{Sin}\,(2\,\pi\, f_a\, t) \times \text{Sin}\,(2\,\pi\, f_c\, t),$$

$$V_{QPSK} = \text{Sin}\,[2\,\pi\,(5\text{ MHz})\,t] \times \text{Sin}\,[2\,\pi\,(70\text{ MHz})\,t],$$

$$V_{QPSK} = \frac{1}{2}\text{Cos}\,2\,\pi\,(f_c - f_a)\,t - \frac{1}{2}\text{Cos}\,2\,\pi\,(f_c + f_a)\,t$$

$$V_{QPSK} = \frac{1}{2}\text{Cos}\,2\pi\,(70\text{-}2.5)\text{ MHz}\,t - \frac{1}{2}\text{Cos}\,2\pi\,(70+2.5)\text{ MHz}\,t$$

$$V_{QPSK} = \frac{1}{2}\text{Cos}\,2\,\pi\,(67.5)\text{ MHz}\,t - \frac{1}{2}\text{Cos}\,2\,\pi\,(72.5)\text{ MHz}\,t$$

a-The minimum Nyquist bandwidth;

$$BW_{Nyquist} = 72.5\text{ MHz} - 67.5\text{ MHz} = 5\text{ MHz}$$

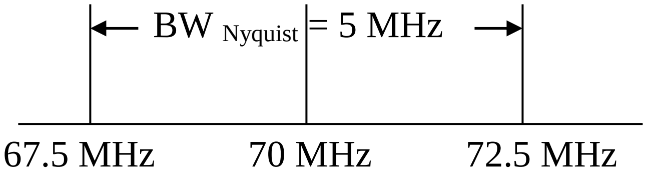
The minimum bandwidth for the QPSK can also be determined by simply using the equation;

$$BW_{QPSK} = \frac{f_b}{N} = \frac{10\,M\ Hz}{2} = 5\,M\ Hz$$

b- the symbol rate equals the bandwidth; thus;

Symbol rate = 5 M baud

Therefore, the output spectrum will be as shown;



It can be seen that for the same input bit rate the minimum bandwidth required to pass the QPSK modulated signal is half that required for the BPSK modulated signal. Also, the baud rate for the QPSK modulator is one-half that of the BPSK modulator. Because of this, it is very popular in systems where bandwidth is at a premium. Phase shift keying (BPSK and QPSK) is extensively used in wireless radio communication systems and telephone network.

4.6-3 Offset-keyed-QPSK, OQPSK

A block diagram of an offset-keyed QPSK system is shown in figure 4.17-a. This block diagram is very similar to that of QPSK. The difference lies in the data transitions between the I and Q streams as they enter the multipliers. The incoming data stream is applied to a serial-to-parallel converter. One of the converter output streams, the Q stream, is then offset with respect to the other by delaying it by an amount equal to the incoming signal bit duration,

$T_b = T_s/2$. Figure 4.17-a shows the serial NRZ data stream $\{S_k\}$ is converted serial-to-parallel (S/P) and becomes two parallel data streams, $\{a_k\}$ and $\{b_k\}$. A differential encoder (diff. enc) may be inserted into the modulator. This encoder, with diff. decoder at the receiver, is required if the carrier recovery circuitry introduced a phase ambiguity. The S/P converter assures that the I and Q data streams are synchronous. The $T_b = T_s/2$ delay element is inserted if OQPSK modulation is required. This delay line does not change the synchronous relationship needed between the I and Q channels. When the VCO is not locked (in synchronism) to an integer multiple of the data rate, the system performance degradation is very small (< 0.1 dB).

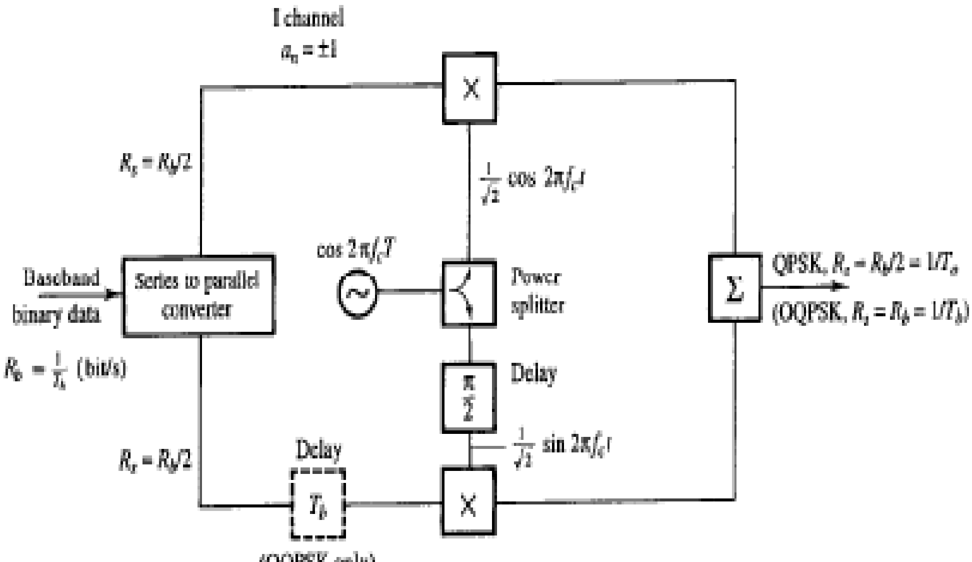


Figure 4.17-a schematic for QPSK (OQPSK) modulator

When the incoming $\{C_k\}$ and $\{d_k\}$ data rates are asynchronous (independent of each other), as shown in figure 4.17-b, and are fluctuating around specified bit rates. It is frequently advantageous to modulate directly the I and Q data streams. In this case the QPSK modulator provides, in effect an asynchronous multiplexing scheme.

If the data streams are band limited with a LPF before modulation, this scheme is known as a two-level QAM modulator.

With advancements in microchip technology, systems using MPSK for M > 4 are easily available these days.

This results in the relationship between the I and Q streams and the input data stream shown in figure 4.17-b.

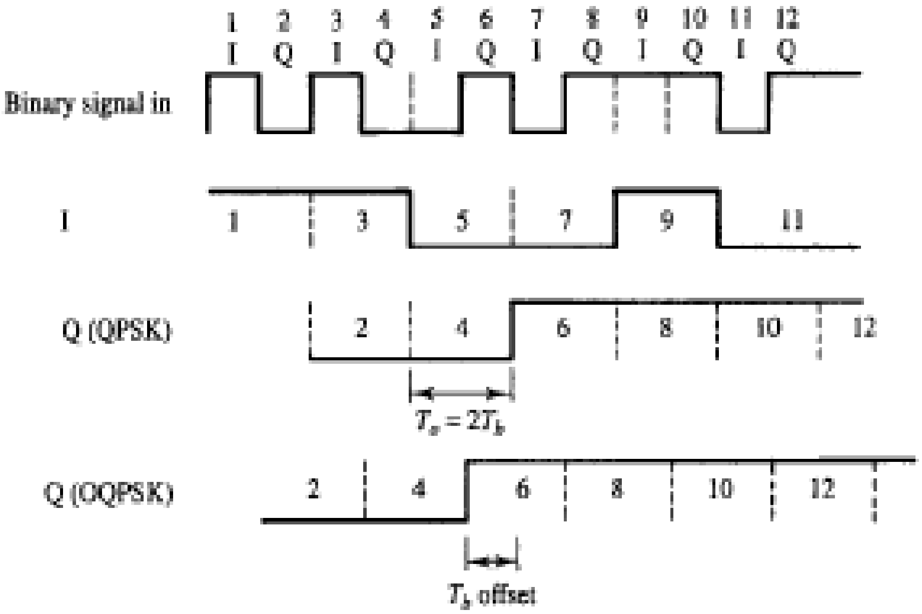


Figure 4.17-b Relationship between the I and Q streams and the input data stream

The resulting instantaneous phase states at the modulator output are the same as for QPSK. However, because both data streams applied to the multipliers can never be in transition simultaneously, only one of the vectors that comprise the offset-keyed quadric-phase modulator output signal can change at any one time. The result is that only 90^{o} phase transitions occur in the modulator output signals.

Like QPSK, an unfiltered OQPSK signal has a constant-amplitude envelope. However, for filtered OQPSK signals, the result is a maximum amplitude envelope variation of 3 dB (70 %) compared to the 100% amplitude envelope variation for conventional QPSK systems. This lower-amplitude envelope variation imparts certain advantages to OQPSK as compared to QPSK in both non-linear satellite and also line-of-sight microwave systems. For example, when a band limited OQPSK signal is transmitted through an amplitude-limiting device, there is only partial regeneration of the spectrum amplitude back to the unfiltered level. For QPSK under the same circumstances, however, there is an almost complete regeneration to the unfiltered level. The receiver is identical to that of QPSK, with the exception that the regenerated I data stream is delayed by a unit bit duration $T_b = T_s/2$, so that when combined with the regenerated Q stream, the original “input data” stream is recreated. Of course, this is subject to error because of the effects of noise and filtering.

Figure 4.17-c shows the Schematic for QPSK (OQPSK) demodulator.

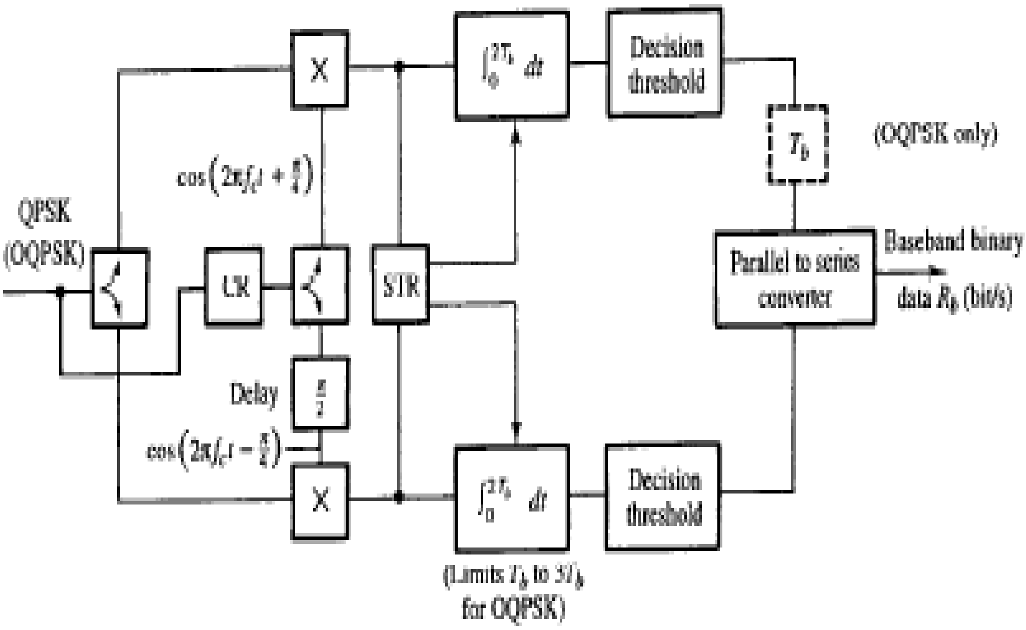


Figure 4.17-c Schematic for QPSK (OQPSK) demodulator.

4.6-4 8- PSK

It is a modulation technique can be viewed as an extension of the QPSK system. In the classical 8-PSK modulator block diagram, shown in figure 4.18, the f_b rate data are split into three binary parallel streams, each having a transmission rate of $f_b/3$. The two-level to four-level converter provides one of the four possible levels of a polar base band signal at a and b.

If the binary symbol A is a logic one (zero), then the output level A has one of the two possible (positive, negative) signal states.

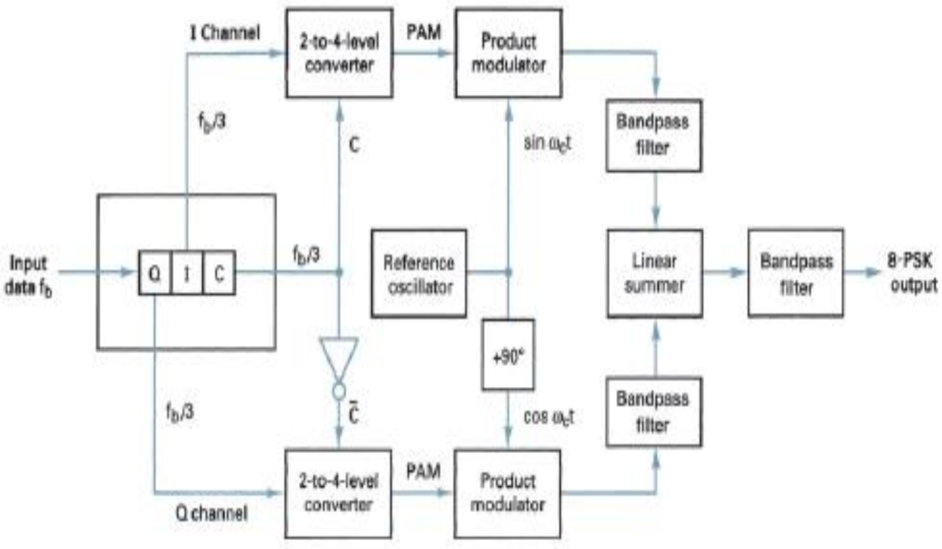


Figure 4.18 8-PSK modulator block diagram

The logic state of the C bit determines whether the higher or lower signal level should be present at A or at B. When C = 1, the amplitude of A is greater than that of B; if C = 0, the converse is true. The four-level polar base band signals at a and b are used to DSB-SC amplitude modulate the two quadrature carriers.

A modern approach to the design of an 8-PSK modulator of high speed (120 Mb/s) transmission, using only digital devices, is illustrated in figure 4.19.

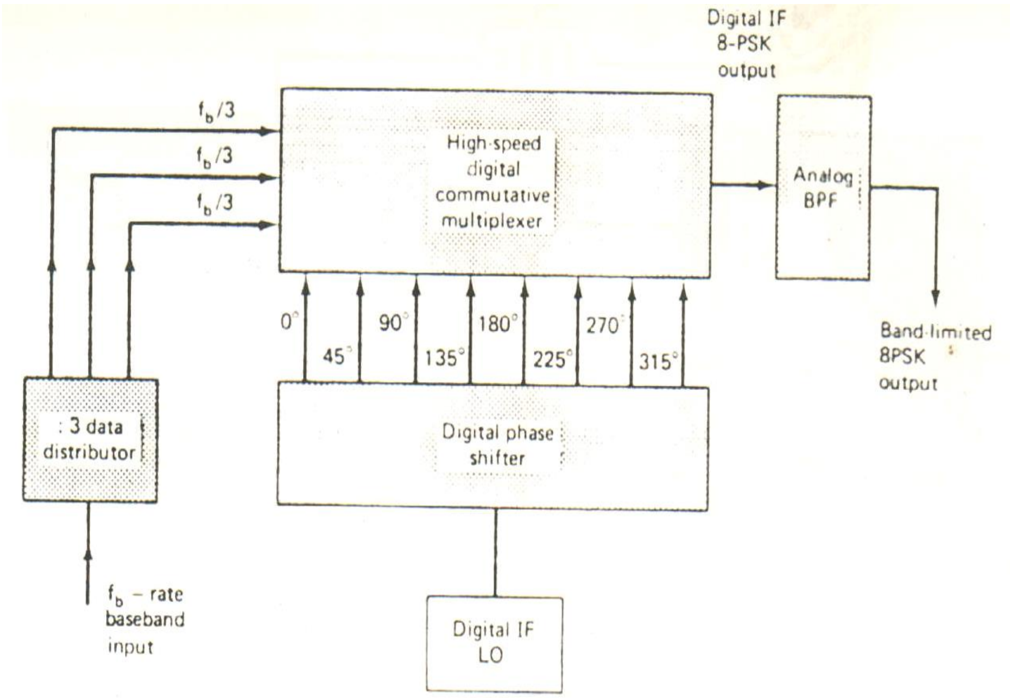


Figure 4.19 High speed 8-PSK modulator using digital subsystems.

The f_b rate input binary base band information is serial-to-parallel converted in the data distributor unit. These parallel $f_b/3$ rate data streams switch on and off the logic gates of the high-speed commutative IF multiplexer.

Depending on the base band logic states, one of the eight digital IF vectors is connected to the digital IF output. This digital phase-shifted 8-PSK carrier is filtered by means of a conventional BPF; thus a band limited 8-PSK signal is obtained.8-level PSK is used at audio frequencies to get 4800 bits per second into a 4 kHz voice frequency channel with a carrier of 1800 Hz occupying most of the useful channel (300 to 3400).

At gigahertz radio frequencies the requirement is for bit rates in the megahertz range for both terrestrial (16-QAM standard) and satellite (QPSK standard) links which mean, respectively, 4 and 2 bits per symbol, so increasing information rates for a given bandwidth.

Example 4.7; For a 8-PSK modulator with a carrier frequency of 70 MHz and an input bit rate of 10 Mbps, determine the following;

a- The minimum double-sided Nyquist bandwidth, f_N

b- The baud rate.

c- Compare the results with those achieved with the BPSK and QPSK modulators in examples 4.5, 4.6.

Solution; a- The bit rate in the I and Q, and c channels is equal to one-third of the input bit rate, or;

$$f_{bc} = f_{bQ} = f_{bI} = \frac{f_b}{3} = \frac{10\,M\,bps}{3} = 3.33 \text{ M bps.}$$

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

$$f_a = \frac{f_{bc}}{2} = \frac{f_{bQ}}{2} = \frac{f_{bI}}{2} = \frac{3.33\,M\,bps}{2} = 1.667\,M\,bps$$

The output waveform from the balanced modulators is;

$$V_{8\text{-PSK}} = \text{Sin}\,(2\,\pi\, f_a\, t) \times \text{Sin}\,(2\,\pi\, f_c\, t),$$

$$V_{8\text{-PSK}} = \text{Sin}\,[2\,\pi\,(1.667 \text{ MHz})\, t] \times \text{Sin}\,[2\,\pi\,(70 \text{ MHz})\, t],$$

$$V_{8\text{-PSK}} = \frac{1}{2}\text{Cos } 2\pi\,(f_c - f_a)\,t - \frac{1}{2}\text{ Cos } 2\pi\,(f_c + f_a)\,t$$

$$V_{8\text{-PSK}} = \frac{1}{2}\text{Cos } 2\pi\,(70\text{–}1.667)\text{ MHzt} - \frac{1}{2}\text{ Cos } 2\pi\,(70+1.667)\text{ MHzt}$$

$$V_{8\text{-PSK}} = \frac{1}{2}\text{Cos } 2\pi\,(68.333)\text{ M Hzt} - \frac{1}{2}\text{ Cos } 2\pi\,(71.667)\text{ M Hzt}$$

b-The min. Nyquist bandwidth

$BW_{Nyquist}$ = (71.667 – 68.333) MHz = 3.333 MHz

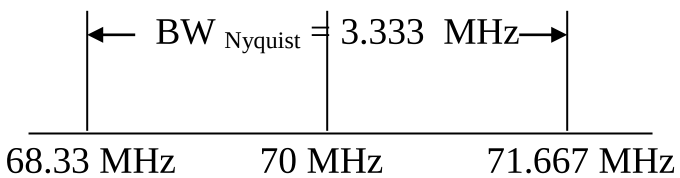
The minimum bandwidth for the 8-PSK can also be determined by simply using the equation;

$$BW_{8\text{-PSK}} = \frac{f_b}{N} = \frac{10\,M\ Hz}{3} = 3.33\,M\ Hz$$

c- The symbol rate equals the bandwidth; thus

Symbol rate = 3.333 M baud

Therefore, the output spectrum will be as shown;



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

4.6-5 16-ary PSK

It is an M-ary encoding technique where M = 16. Four bits (quad bits) are combined, N = 4, producing 16 different output phases. Therefore the minimum bandwidth and baud equal one-fourth the bit rate (f_b / 4). It is clear that as the level of encoding increases (the values of N, and M), more output phases are possible. With 16-PSK, the angular separation between adjacent output phases is only 22.5^o. there, 16-PSK can undergo only a 11.5^o phase shift during transmission and still retain its integrity.

Increasing M levels further causes limitation in the level of encoding and bit rate possible with PSK, as a point is eventually reached where receivers can not discern the phase of the received signalling element. In addition, phase impairments inherent on communications lines have a tendency to shift the phase of the PSK signal, destroying its integrity and producing errors.

4.7 Quadrature-Amplitude Modulation, QAM

QAM is a form of digital modulation similar to PSK except the digital information is contained in both the amplitude and the phase of the transmitted carrier; for this reason, the term amplitude phase keying is also used.

With QAM, amplitude and phase-shift keying are combined in such a way that the positions of the signalling elements on the constellation diagrams are optimized to achieve the greatest distance between elements, thus reducing the likelihood of one element being misinterpreted as another element. Obviously, this reduces the likelihood of errors occurring.

4.7-1 8- QAM

It is an M-ary encoding technique, (M = 8, N = 3). Unlike 8-PSK, the output signal from 8-QAM modulator is not a constant-amplitude signal.

4.7-2 16- QAM

It uses L = 4-level base band streams. Also, both the phase and the amplitude of the transmitted carrier are varied. The block diagram for a QAM-SC modulator is shown in figure 4.20-a.

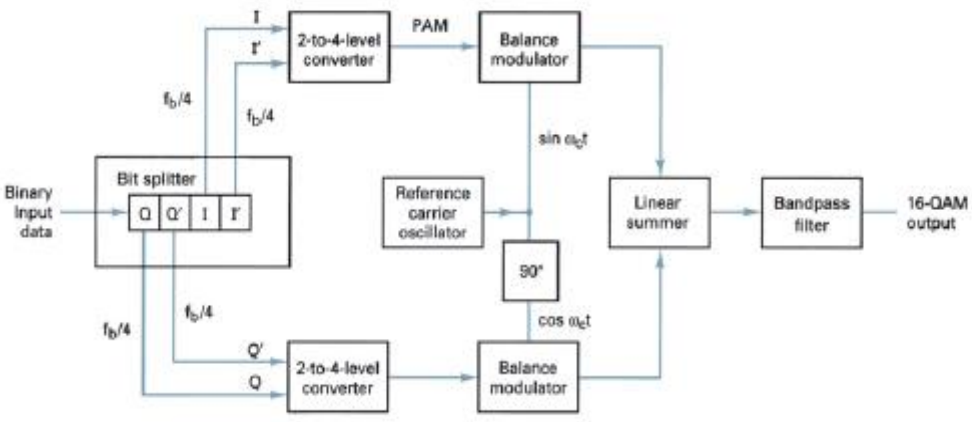


Figure 4.20-a M-ary QAM modulator block diagram

The phase constellation of 16-QAM is illustrated in figure 4.20-b.

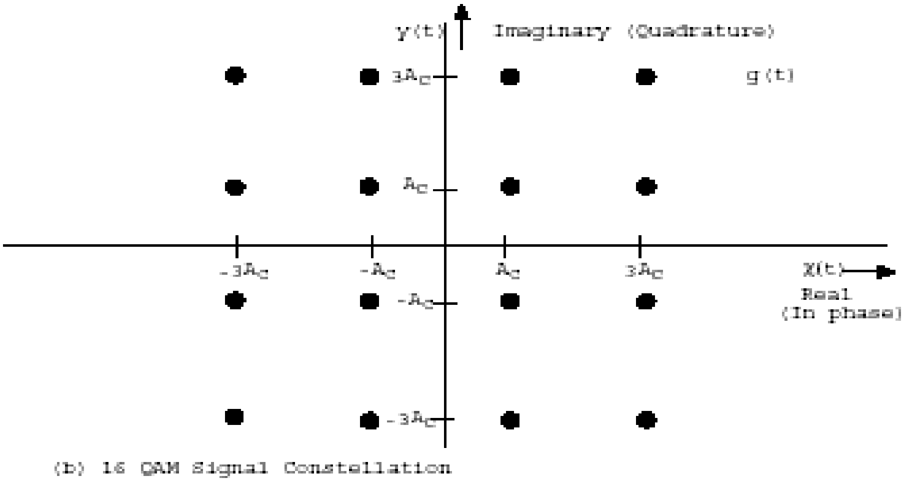


Figure 4.20-b. Phase constellation of 16-QAM

If an M = 16-ary QAM modulated signal having a theoretical efficiency of 4 b/s/Hz is desired, these commuted binary streams converted into L = 4-level base band streams.

For the above example, the resultant 4-level symbol streams of the I and Q channels are (5 Mb/s) / $\log_2 4$, or 2.5 M symbols/s. If a pre modulation LPFs are used, then the minimum bandwidth (Nyquist bandwidth with $\alpha = 0$) of these filters is 1.25 MHz. The minimum IF bandwidth requirements is identical to the double-sided minimum base band bandwidth (i.e., 2.5 MHz). This example illustrates that

a 10 Mb/s, M = 16-ary QAM signal can be transmitted in a theoretical minimum bandwidth of 2.5 MHz; thus, a spectral efficiency of

4 b/s/Hz has been obtained. Practical, high-speed, 400-Mb/s, 16-ary systems have been achieving a bandwidth efficiency of approximately 3.7 b/s/Hz.

The f_b rate binary source is commuted into two binary symbol streams, each having a rate of $f_b/2$. The following 2-to-M-level base band converter converts these $f_b/2$ rate data streams into L-level PAM signals having a symbol rate of;

$$f_s = (f_b/2) / \log_2 L \quad \text{symbols/sec.} \tag{4.20}$$

For example, if the source bit rate is f_b = 10 Mb/sec, then the commuted binary base band streams have an $f_b/2$ = 5 Mb/s rate.

The spectral efficiency and P_e performance of the QAM system is identical with that of QPSK systems. As an example indicates spectral efficiencies that are achieved in some practical radio systems.

The TDMA version of the North American Digital Cellular (NADC) system, achieves a 48 K bits-per-second data rate over a 30 kHz bandwidth or 1.6 bits per second per Hz. It is a π/4 DQPSK based system and transmits two bits per symbol.

Another example is a microwave digital radio using 16-QAM. This kind of signal is more susceptible to noise and distortion than something simpler such as QPSK. This type of signal is usually sent over a direct line-of-sight microwave link or over a wire where there is very little noise and interference. In this microwave-digital-radio example the bit rate is 140 M bits per second over a very wide bandwidth of 52.5 MHz. The spectral efficiency is 2.7 bits per second per Hz. To implement this, it takes a very clear line-of-sight transmission path and a precise and optimized high-power transceiver.

Example 4.8; For a 16-QAM modulator with a carrier frequency of 70 MHz and an input bit rate of 10 Mbps, determine the following;

a- The minimum double-sided Nyquist bandwidth, f_N

b- The baud rate.

c- Compare the results with those achieved with the BPSK, QPSK, and 8-PSK modulators in examples 4.5, 4.6, and 4.7.

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

$$f_{bI} = f_{bI'} = f_{bQ} = f_{bQ'} = \frac{f_b}{4} = \frac{10\,M\ bps}{4} = 2.5 \text{ M bps.}$$

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

$$f_a = \frac{f_{bI}}{2} = \frac{f_{bI'}}{2} = \frac{f_{bQ}}{2} = \frac{f_{bQ'}}{2} = \frac{2.5\,M\ bps}{2} = 1.25\,M\ bps$$

The output waveform from the balanced modulators is;

$V_{16\text{-}QAM} = \text{Sin}\,(2\,\pi\, f_a t) \times \text{Sin}\,(2\,\pi\, f_c t)$,

$$V_{16\text{-QAM}} = \text{Sin}\,[2\pi\,(1.25\text{ MHz})t] \times \text{Sin}\,[2\pi\,(70\text{ MHz})t],$$

$$V_{16\text{-QAM}} = \frac{1}{2}\text{Cos}\,2\pi\,(f_c - f_a)t - \frac{1}{2}\,\text{Cos}\,2\,\pi\,(f_c + f_a)t$$

$$V_{16\text{-QAM}} = \frac{1}{2}\text{Cos}2\pi(70\text{-}1.25)\text{ MHzt} - \frac{1}{2}\text{Cos}2\pi(70\text{+}1.25)\text{ MHzt}$$

$$V_{16\text{-QAM}} = \frac{1}{2}\text{Cos}\,2\pi\,(68.75)\text{ MHzt} - \frac{1}{2}\text{Cos}2\,\pi\,(71.25)\text{ MHzt}$$

b-The min. Nyquist bandwidth;

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

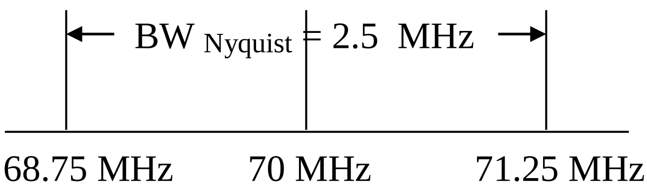
The minimum bandwidth for the 16-QAM can also be determined by simply using the equation;

$$BW_{16\text{-QAM}} = \frac{f_b}{N} = \frac{10\,M\;Hz}{4} = 2.5\,M\;Hz$$

c- The symbol rate equals the bandwidth; thus

Symbol rate = 2.5 M baud

Therefore, the output spectrum will be as shown;



It can be seen that for the same input bit rate the minimum bandwidth required to pass the output of the 16-QAM modulator is equal to one-fourth of that required for the BPSK modulator, one-half that of QPSK, and 25 % less than that required for the 8-PSK modulator. Also, in each case the baud rate has been reduced by the same factors.

4.7-3 64- QAM

It is an M-ary encoding technique of 64 state (8x8) QAM, (M = 8, N = 3). Unlike 8-PSK, the output signal from 8-QAM modulator is not a constant-amplitude signal.

Also, with 256-QAM signal has 256 state (16x16), (M = 16, N = 4). Both the phase and the amplitude of the transmitted carrier are varied.

Example 4.9; for 256-QAM signals with f_b = 1.6 Mbps transmission rate is used. The 256-state (16 x 16) QAM signal, contains 16 base band levels in the in-phase, I, and 16 levels in the quadrature, Q, channels, respectively. Each level is formed by 4 bits. For our example the symbol rate of the I-channel (or Q-channel) demodulator is;

f_s = 1600 kb/s / 2 (number of channels) / 4 (bits / symbol)

= 200 k symbols/sec = 200 k baud, f_N = 100 kHz.

Theoretical bandwidth efficiency limits; The following table 4.3 shows the theoretical bandwidth efficiency limits for the main modulation techniques. These figures cannot actually be achieved in practical radios since they require perfect modulators, demodulators, filter and transmission paths. If the radio had a perfect (rectangular in the frequency domain) filter, then the occupied bandwidth could be made equal to the symbol rate.

Techniques for maximizing spectral efficiency include;

- Relate the data rate to the frequency shift (as in GSM).
- Use pre-modulation filtering to reduce the occupied bandwidth. Raised cosine filters, as used in NADC, PDC, and PHS give the best spectral efficiency.
- Restrict the types of transitions.

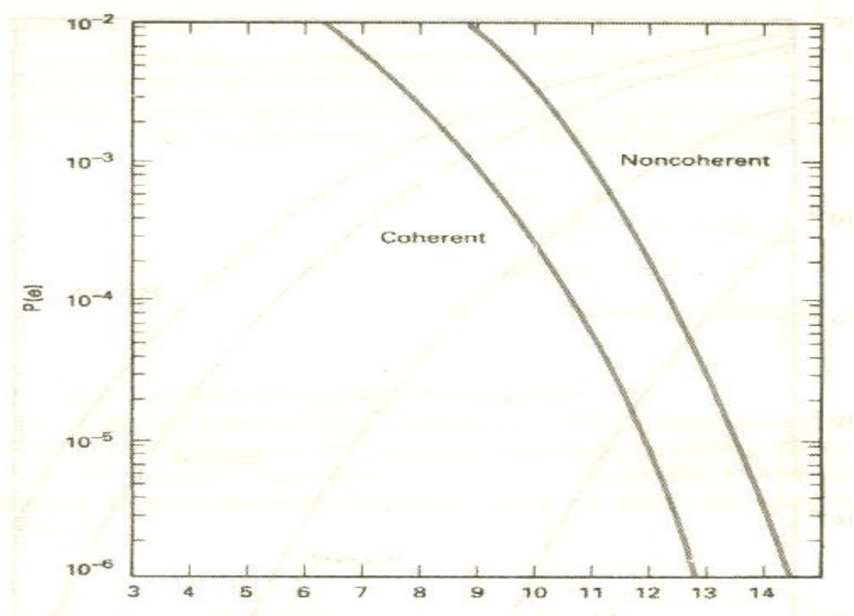
Table 4.3 Theoretical bandwidth efficiency limits for the main modulation techniques.

Modulation	Encoding Scheme N bits	Output possible M level	Min. bandwidth W Hz	Baud rate	BWη
ASK	1	2	f_b	f_b	1
FSK	1	2	f_b	f_b	1
BPSK	1	2	f_b	f_b	1
QPSK	2	4	$f_b/2=6000$	$f_b/2$	2
8-PSK	3	8	$f_b/3=4000$	$f_b/3$	3
8-QAM	3	8	$f_b/3=4000$	$f_b/3$	3
16-QAM	4	16	$f_b/4=3000$	$f_b/4$	4
16-PSK	4	16	$f_b/4=3000$	$f_b/4$	4
32-PSK	5	32	$f_b/5$	$f_b/5$	5
32-QAM	5	32	$f_b/5$	$f_b/5$	5
64-PSK	6	64	$f_b/6$	$f_b/6$	6
64-QAM	6	64	$f_b/6$	$f_b/6$	6
128-PSK	7	128	$f_b/7$	$f_b/7$	7
128-QAM	7	128	$f_b/7$	$f_b/7$	7

Note that for values of N > 1, the number of output conditions increases, and the minimum bandwidth and bauds decreases. Therefore, digital modulation schemes with N > 1 achieves bandwidth compression (i.e., less bandwidth is required to propagate a given bit rate). When data compression is performed, higher data transmission rates are possible for a given bandwidth.

4.8 Minimum Shift keying, MSK

Many data sets operating in the range 50 b/s to 1 Mb/s use non-coherent FSK modems. Non-coherent demodulators are simpler but require a higher E_b/N_o than coherent systems as shown in figure 4.21'.



E_b/N_o

Figure 4.21' Comparison between coherent and non-coherent FSK demodulator

Since any higher E_b/N_o requirement than the minimum possible is not acceptable for power-constrained communication systems, non-coherent FSK is not acceptable. The coherent FSK modulation / demodulation method, known as MSK or fast frequency shift keying, FFSK, has a good performance. MSK is a binary form of continuous phase frequency shift keying (CPFSK). It is a particularly spectrally efficient form of coherent frequency-shift keying, as it has a narrower spectrum than wider deviation forms of FSK. The width of the spectrum is also influenced by the waveforms causing the frequency shift.

The GSM digital cellular phones use MSK, where the peak frequency deviation is selected to produce orthogonal waveforms for binary one and binary zero data. (Digital phones use a speech codec to convert the analog voice source to a digital data source for transmission over the system.) Orthogonality occurs when $\Delta f = 1/4R$, where R is the bit rate. Actually, GSM uses Gaussian shaped MSK (GMSK). That is, the digital data waveform (with rectangular binary one and binary zero pulses) is first filtered by a low-pass filter having a Gaussian shaped frequency response (to attenuate the higher frequencies).

This Gaussian filtered data wave form is then fed into the frequency modulator to generate the GMSK signal. This produces a digitally modulated FM signal with a relatively small bandwidth.

The theoretical spectral efficiency of MSK systems is 2 b/s/Hz. The logic state 1 corresponds to transmit frequency f_2 (f_1), while logic state 0 (-1) to f_1 (f_2). The frequency deviation is given by;

$$\Delta f = \frac{f_2 - f_1}{2} = \frac{1}{4T_b} \tag{4.21}$$

where T_b is the unit bit duration of the input data streams.

Note that a coherent relation between the transmitted frequencies and the bit rate is required such that; in MSK the difference between the higher and lower frequency deviation is equal to half the bit rate.

For this reason, the MSK demodulator is usually a coherent quadrature detector, similar to that for QPSK. In this case, the error rate performance is the same as that of BPSK and QPSK.

An implementation method of the MSK modulator is shown in figure 4.21-a.

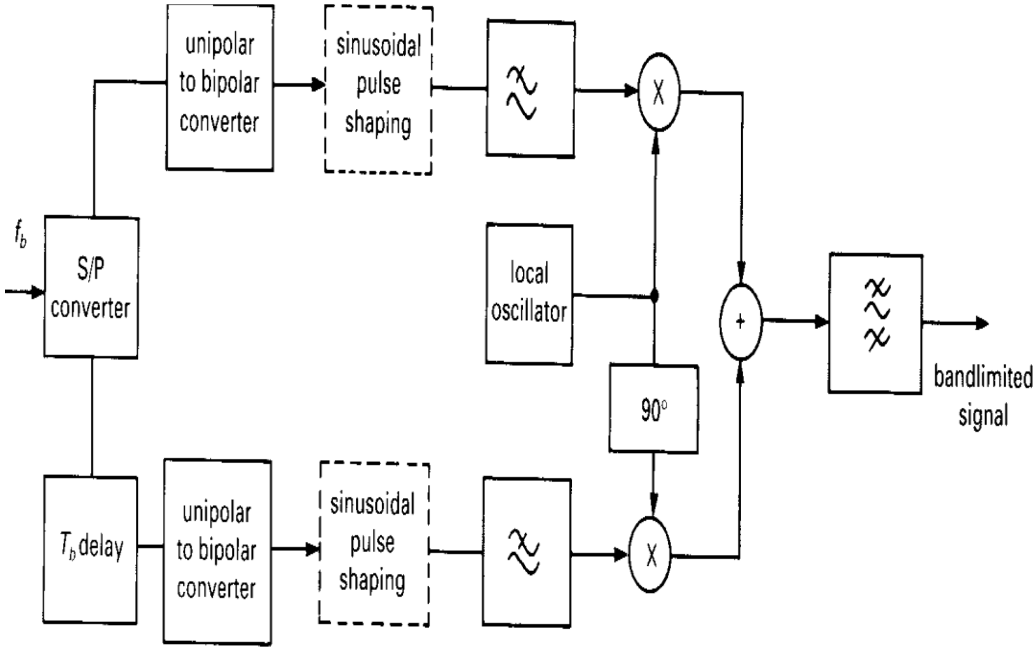


Figure 4.21-a. Block diagram of an MSK modulator

The un-modulated carrier of frequency f_C is multiplied by both the in-phase and quadrature base band signals. The serial-to-parallel converted data are fed into the sinusoidal pulse-shapers. In the base band I channel the pulse shaper generates a Cos [$\pm\ \pi\ t/2T_b$] pulse sequence, while in the Q channel the offset delay T_b in conjunction with the pulse shaper, provides the pulse sequence given by;

$$\text{Cos}\left[\pm\frac{\pi(t-T_b)}{2T_b}\right] = \text{Sin}\left[\pm\frac{\pi\ t}{2T_b}\right]$$

The sinusoidal pulse shapers could be implemented as non-linearly switched filters.

The sinusoidal pulse shaping means that the modulator output has either a positive or negative linear phase-change rate relative to the carrier depending on the input data.

The amplitude and phase of the output signals of the modulating multipliers is such that the sum of these signals (which is the un-filtered MSK modulated signal) has a constant-amplitude envelope carrier signals (i.e., time invariant).

This is a desirable characteristic for improving the power efficiency of transmitters. Amplitude variations can exercise non-linearity in an amplifier's amplitude-transfer function, generating spectral re-growth, a component of adjacent channel power. Therefore, more efficient amplifiers (which tend to be less linear) can be used with constant-envelope signals, reducing power consumption.

The MSK demodulator operates in the same manner as the QPSK receiver described earlier. However, different filtering is required to ensure ISI-free transmission.

A frequency shift keying signal, $V_{FSK}(t)$, can be considered as the transmission of a sinusoid, the frequency of which is shifted between the following two frequencies;

$$f_1 = f_c - \Delta f = f_c - ¼\, T_b$$

$$f_2 = f_c + \Delta f = f_c + ¼\, T_b \qquad (4.22)$$

It is described by;

V_{FSK} (t) = A Cos [$2\pi(f_c \pm \Delta f)t$]

= A Cos [$\pm 2\pi(\Delta f)t$] Cos [$2\pi(f_c)t$]

– A Sin [$\pm 2\pi(\Delta f)t$] Sin [$2\pi(f_c)t$] (4.23)

Thus the MSK signal V_{MSK}(t) is;

V_{MSK} (t) = A Cos [$\pm \pi t/2T_b$] Cos [$2\pi(f_c)t$] –

A Sin [$\pm \pi t/2T_b$] Sin [$2\pi(f_c)t$] (4.24)

As a result, the waveforms used to represent a 0 and a1bit differ by exactly half a carrier period as shown in figure 4.21-b. This is the smallest FSK modulation index that can be chosen such that the waveforms for 0 and 1 are orthogonal.

Figure 4.21-c shows the schematic of MSK demodulator.

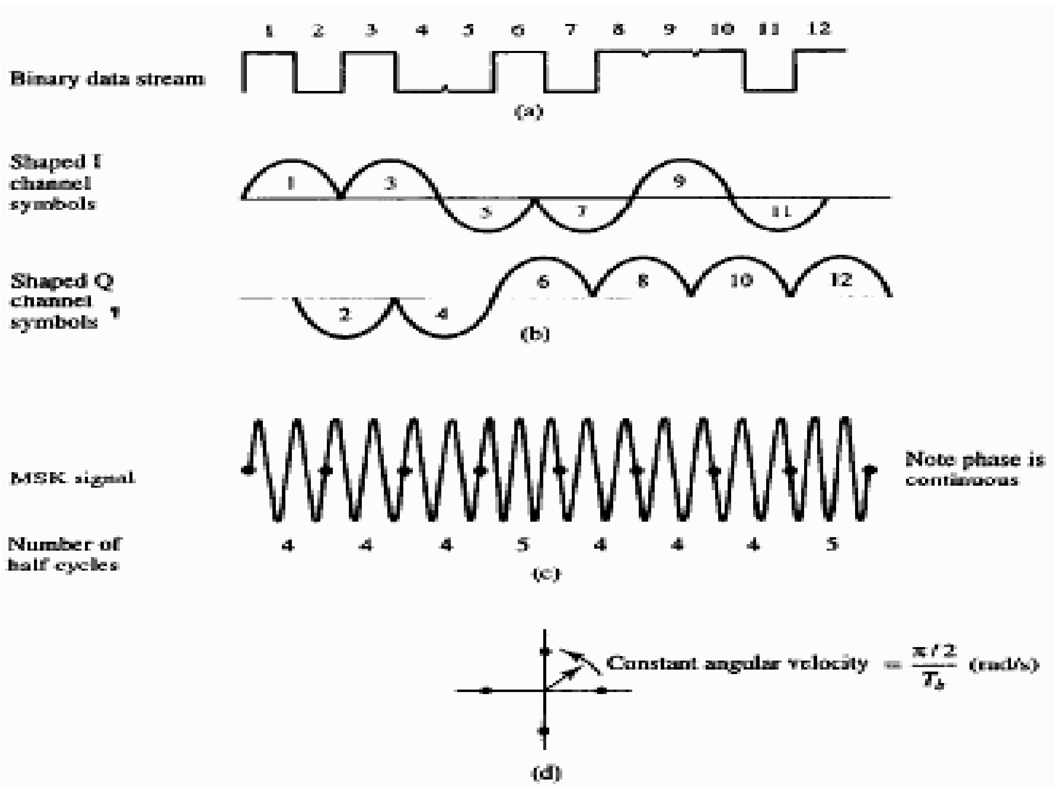


Figure 4.21-b MSK transmitter waveforms and phasor diagrams

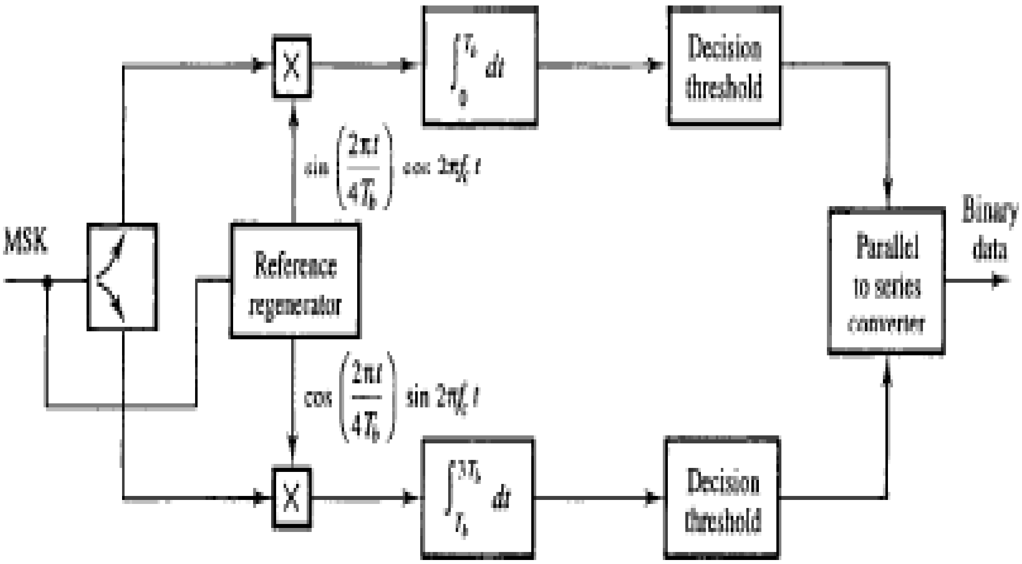


Figure 4.21-c Schematic of MSK demodulator

4.9 Gaussian Minimum Shift Keying

GMSK is a derivative of MSK where the bandwidth required is further reduced by passing the modulating waveform through a Gaussian filter so, the modulation method is referred to as GMSK. The Gaussian filter minimizes the instantaneous frequency variations over time. GMSK has a constant envelope, good BER performance and is self-synchronizing, spectrally efficient modulation scheme. It is particularly useful in GSM mobile radio systems. Generation of GMSK is illustrated in figure 4.22. Figure 4.23 illustrates typical modulated carrier waveforms for three systems; MSK, OQPSK and QPSK. The binary data rate in each case is $1/T_b$.

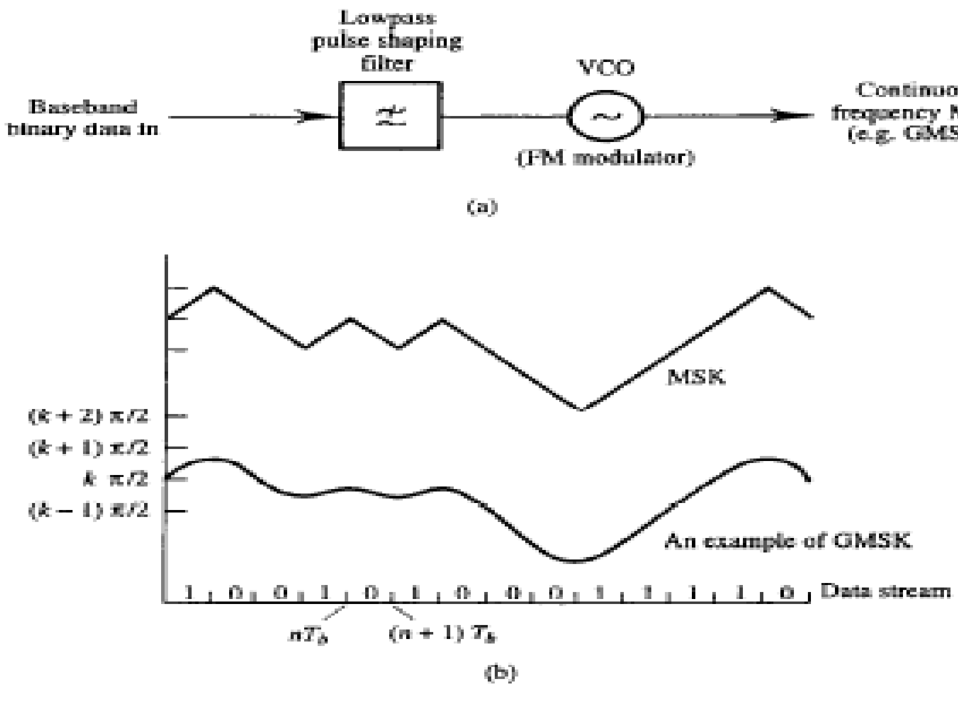


Figure 4.22 (a) Generation of GMSK, (b) MSK and GMSK phase trajectories for typical bit sequences with the MSK

Phase transitions in QPSK occur only every 2 T_b seconds (i.e., a multiple of the symbol duration), whereas for MSK, transitions may occur every T_b seconds.

In QPSK, abrupt $\pm 90^o$ phase transitions and phase reversals of $\pm 180^o$ are possible. In the MSK case the carrier phase varies linearly over T_b seconds in $\pi/2$ increments and is continuous phase waveform.

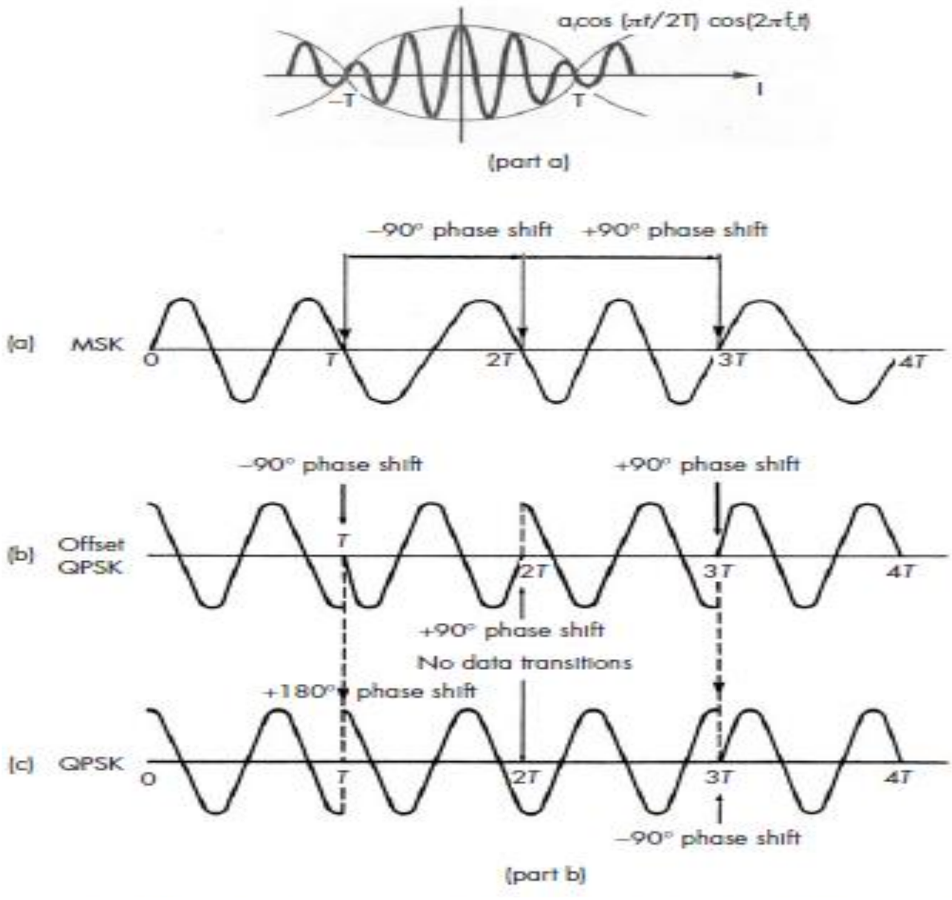


Figure 4.23 MSK modulation: (a) generation of MSK via quadrature

modulation, and (b) typical modulated carrier waveforms for; MSK, OQPSK and conventional QPSK.

4.10 Performance of Digital Communication Systems in the presence of Noise

In digital communication system, E_b/N_o is the more often parameter used to describe the system performance in the presence of noise. E_b is bit energy can be described as signal power S times the bit time T_b. N_o is noise power spectral density, and can be described as noise power N divided by BW W. Since the bit time and bit rate R_b are reciprocal, so;

$$\frac{E_b}{N_o} = \frac{S\, x T_b}{N / W} = \frac{S / R_b}{N / W} = \frac{S}{N}\left(\frac{W}{R_b}\right) \qquad (4.25)$$

E_b/N_o is one of the most important metrics of digital communication systems performance.

It is a standard quality measure for the quality of these systems.

The reason for using the bit energy to describe the digital signal is that;

A power signal (analog waveform) is described as a signal having finite average power and infinite energy, and the energy signal (digital waveform) is that signal having zero average power and finite energy. This classification is useful in comparing analog and digital waveforms. The analog waveform is considered as having infinite duration that corresponding to infinite amount of energy, hence, energy is not a useful way to characterize the analog waveform.

Power (rate of delivering the energy) is a more useful parameter for analog waveforms. However, in a digital communication system a symbol is transmitted (and received) by using a transmission waveform within the symbol time T_S. Focusing on one symbol, the power (averaged over all time) goes to zero. Hence, power is not a useful way to characterize a digital waveform.

The symbol energy (power integrated over T_S) is a more useful parameter for characterizing digital waveforms.

Figure 4.23 shows the variation of the probability P_e versus E_b/N_o, where for $E_b/N_o \geq X_o$, $P_e \leq P_o$.

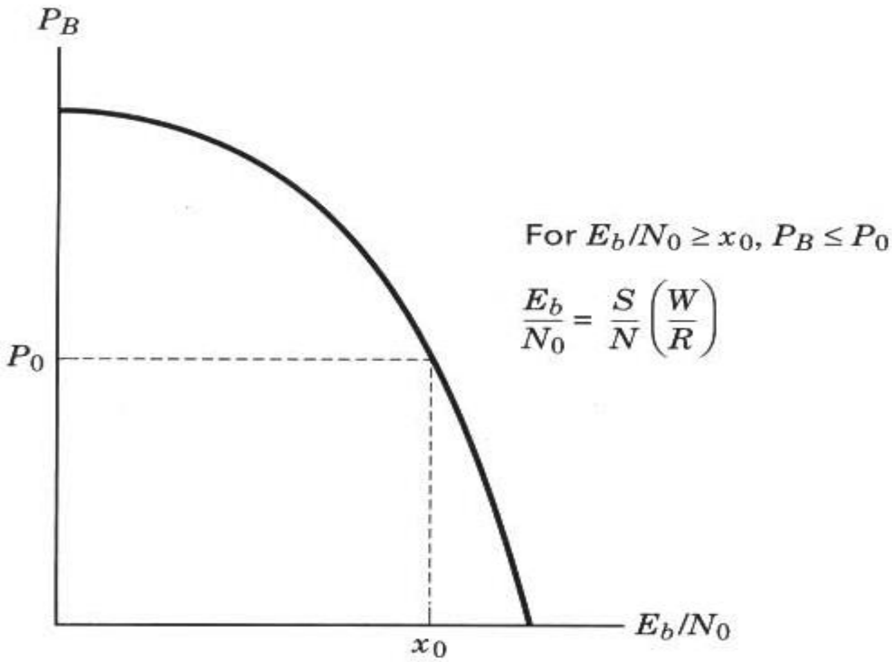


Figure 4.23 P_e versus E_b/N_o

The required E_b/N_o can be considered a metric that characterizes the performance of one system versus another; the smaller the required E_b/N_o, the more efficient is the detection process for a given P_e.

The SNR in a PAM signal depends on the form of the receiver. If the receiver simply samples the received waveform at periodic points in time, the sample values are;

$$r(nT_s) = S(nT_s) + n(nT_s) \qquad (4.26)$$

where n(t) is the additive noise. The SNR is then;

$$\frac{S}{N} = \frac{\overline{s^2}}{\overline{n^2}} \qquad (4.27)$$

The numerator of equation (4.27) is the average signal power, while the denominator is the average noise power.

If the channel can be modeled as an ideal LPF, the noise power is simply N_oW.

For PAM the probability of symbol error is calculated as;

$$P_M = \frac{M-1}{M} erfc\left(\sqrt{\frac{3}{M^2-1}\left(\frac{nE_b}{N_o}\right)}\right) \qquad (4.28)$$

The penalty in increasing the number M of signaling waveforms is loss in average SNR per bit of approximately 4 dB for M = 4, and for large M, the penalty approaches 6 dB for every factor of 2 increase in M.

The Power efficiency of PAM signal is calculated as;

$$\eta_{Power,PAM} = \frac{3\log_2 M}{M^2-1} \qquad (4.29)$$

Also, the spectral efficiency is given by;

$$SE_{PAM} = 2 \log_2 M$$

It is clear that the spectral efficiency improves with M, but this decreases power efficiency.

In an M-level base band transmission system, one or more bit errors are measured by $P_e = f\,(S/N)$. The end result is as follows;

$$P_e = 2\left[1 - \frac{1}{M}\right] Q\left(\sqrt{\frac{3}{M^2 - 1}\frac{S}{N}}\right) \qquad (4.30)$$

where, M = 2, 4, 8, ……

The numerical values of the Q function can easily be obtained from table 4.1 given before for the complementary error functions, denoted by erfc (y).

Note that the practical spectral efficiency of PAM systems, defined at the 35-to-50 dB out-of-band attenuation point, is about 20 % lower than the theoretical Nyquist rate .

The required S/N, in the Nyquist BW for a BER of 10^{-8} is 34 dB, assuming Gaussian Noise. Generally, $T_s = T_b \log_2 M$.

Table 4.4 indicates that a digital transmission system is more spectrally efficient if it has the capability of transmitting a greater number of bps in a given BW. Also it is clear that higher-state systems having a larger number of output levels have an increased spectral efficiency.

However, higher-level systems require a higher SNR. For a binary system (2-level), the practical spectral efficiency is 1.66 b/s/Hz, whereas 8-PAM system has a 5 b/s/Hz practical spectral efficiency.

To achieve this improvement in the spectral efficiency, an increased SNR from 15 dB to 28 dB is required assuming 10^{-8} BER.

Table 4.4, PAM and base band spectral efficiency and SNR requirements.

Number of levels	**Theoretical SE (b/s/Hz)**	**Practical SE (35 dB attenuation)**	**Required SNR** (P_e = BER = 10^{-8})
2-PAM	2	1.66 [- 20 %]	15
4-PAM	4	3.33 [- 20 %]	21.5
8-PAM	6	5 [- 20 %]	28
16-PAM	8	6.66 [- 20 %]	34
32-PAM	10	8.33 [- 20 %]	40

Digital systems have a certain threshold value for the S/N. From table 4.5, we find that if the S/N is worse than 18 dB, errors occur quite frequently. At a few decibels better value for the S/N, the transmission is almost error free.

Table 4.5 Examples of error rates and mean time between errors for a 64 kbps channel.

S/N (dB)	Error Rate	Mean Time Between Errors
10.3	10^{-2}	1.5 ms
14.4	10^{-4}	150 ms
16.6	10^{-6}	15 seconds
18	10^{-8}	26 minutes
19	10^{-10}	2 days
20	10^{-12}	6 months
21	10^{-14}	50 years

Example 4.10; It is required to transmit a bit rate of f_b = 800 kb/s rate binary NRZ data stream that converted to 16-level PAM base band signal. Calculate the power and spectral efficiency of this PAM signal .

Solution; As each 16-PAM symbol contains 4 bits of information, the corresponding 16-PAM symbol rate is ؛

$f_S = f_b / 4 = 800/4 = 200$ k baud (symbol / sec.)

The Nyquist frequency is $f_N = f_S /2 = 100$ kHz.

A theoretical (ideal) α = 0 filter would filter out all spectral components above 100 kHz. With α = 0.2, filter has a practical attenuation of 35 dB at 20% above the 100 kHz Nyquist frequency

So, all spectral components beyond 120 kHz are attenuated by 35 dB.

The Power efficiency of this PAM signal is;

$$\eta_{Power,PAM} = \frac{3n}{M^2 - 1} = \frac{3x4}{(16)^2 - 1} = 4.7\%$$

and its spectral efficiency is;

$SE_{PAM} = f_b / W$ = 800 kbps/100 k Hz

$= 2 \log_2 M = 2 \times 4 = 8$ b/s/Hz.

Example 4.11; Determine the Nyquist BW and the required SNR for an f_b = 15 Mbps rate transmission system if the available base band BW is 3 MHz and the specified $P_e = 10^{-8}$. How many transmission levels are required.

Solution; The practical spectral efficiency is;

$SE = f_b / f_N$ = 15 Mbps / 3 MHz = 5 b/s/Hz

The best possible performance is obtained with the lowest P_e, so we have to select a transmission system that satisfies this spectral efficiency requirement with minimum SNR.

From table 4.4, we note that;

4-PAM has SE of 4 b/s/Hz, and SNR of 21.5 dB, whereas;

8-PAM has SE of 6 b/s/Hz, and SNR of 28.0 dB. Thus we have no choice but to choose an 8-PAM system, which requires an increased SNR.

In 8-PAM, each symbol level is formed by 3 bits; thus $f_s = f_b / 3$ (because $T_s = 3\ T_b$).

Therefore, the corresponding Nyquist BW is;

$f_N = f_s / 2 = (f_b / 3)/2$ = (15 Mbps / 3) / 2 = 2.5 MHz with ideal filter $\alpha = 0$, a 2.5 MHz base band BW would be sufficient. With practical $\alpha = 0.2$, channel filter, the f_b = 15 Mb/s rate 8-level signal is transmitted in a practical BW of, 2.5 x (1 + 0.2) = 3 MHz. This 3 MHz BW includes the out-of-band attenuation of the $\alpha = 0.2$ roll-off channel.

Note that the noise BW of the receiver is identical to the Nyquist BW of the receive LPF ($f_s/2$) assuming that the overall Nyquist channel is equally matched between the transmitter and the receiver.

Example 4.12, The information in an analog waveform, with maximum frequency f_m = 3 kHz, is to be transmitted over an M-ary PAM system, where the number of pulse levels is M =16. The quantization distortion is specified not to exceed ± 1% of the peak-to-peak analog signal.

a- What is the minimum number of bits/sample, or bits/ PCM word that should be used in digitizing the analog waveform ?

b- What is the minimum required sampling rate, and what is resulting bit transmission rate?

c- What is the PAM pulse or symbol transmission rate?

d- If the transmission BW (including filtering) equals 12 kHz, determine the BW efficiency for this system.

Solution, here we are concerned with two types of levels: the number of quantization levels for fulfilling the distortion requirement and the 16 levels of the multilevel PAM pulses,

a- $n \geq \log_2 1 / 0.02 = \log_2 50 = 5.6.$

therefore, use n = 6 bits/sample to meet the distortion requirement.

b- Using the Nyquist sampling criteria, the min. sampling rate

$f_s = 2\, f_m = 6000$ sample/s.

Each sample will give rise to a PCM word composed of 6 bits.

Therefore, the bit transmission rate;

$$f_b = n\, f_s = 6 \times 6000 = 36 \text{ kbps}$$

c- Since multilevel pulses are to be used with;

$M = 2^n = 16$ levels, then $n = \log_2 16 = 4$ b/symbol.

Therefore, the bit stream will be partitioned into groups of 4 bits to form the new 16-level PAM digits, and the resulting symbol transmission rate R_s is;

$$f_s = f_b / n = 36000 / 4 = 9000 \text{ symbols / second.}$$

d- BW efficiency is described by data throughput,

$$f_b / W = 36000 / 12000 = 3 \text{ bits / sec / Hz.}$$

The following table 4.6 gives P_e formulae for ideal coherent detection of base band and IF modulated binary signals.

Table 4.6 P_e formulae for ideal coherent detection of base band and IF modulated binary signals.

Signaling	**Signal**	P_e
Base band	**Uni-polar ASK**	½ erfc $\sqrt{\frac{1}{2}\frac{E_b}{N_o}}$
signaling	**Polar ASK**	½ erfc $\sqrt{\frac{E_b}{N_o}}$
IF/RF	**ASK**	½ erfc $\sqrt{\frac{1}{2}\frac{E_b}{N_o}}$
signaling	**BPSK (orthogonal)**	½ erfc $\sqrt{\frac{1}{2}\frac{E_b}{N_o}}$
	BPSK	½ erfc $\sqrt{\frac{E_b}{N_o}}$

The following figure 4.24-a gives a comparison between the performance of some binary systems.

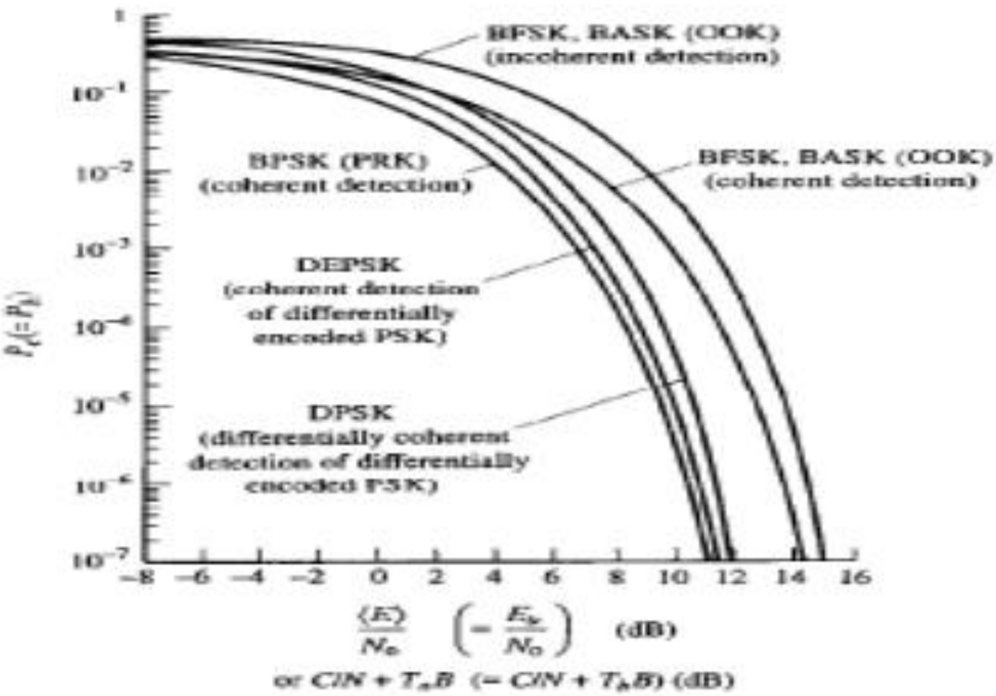


Figure 4.24-a Comparison of binary ASK/PSK/FSK systems performance.

Table 4.7, gives a comparison between the performance of different digital modulation techniques with $P_e = 10^{-6}$. Figure 4.24-b gives a comparison between the performance of different levels of QAM system.

Table 4.7 a comparison between the performance of different digital modulation Techniques with $P_e = 10^{-6}$

Modulation Technique	C/N Ratio (dB)	E_b/N_0 Ratio (dB)
BPSK	10.6	10.6
QPSK	13.6	10.6
4-QAM	13.6	10.6
8-QAM	17.6	10.6
8-PSK	18.5	14
16-PSK	24.3	18.3
16-QAM	20.5	14.5
32-QAM	24.4	17.4
64-QAM	26.6	18.8

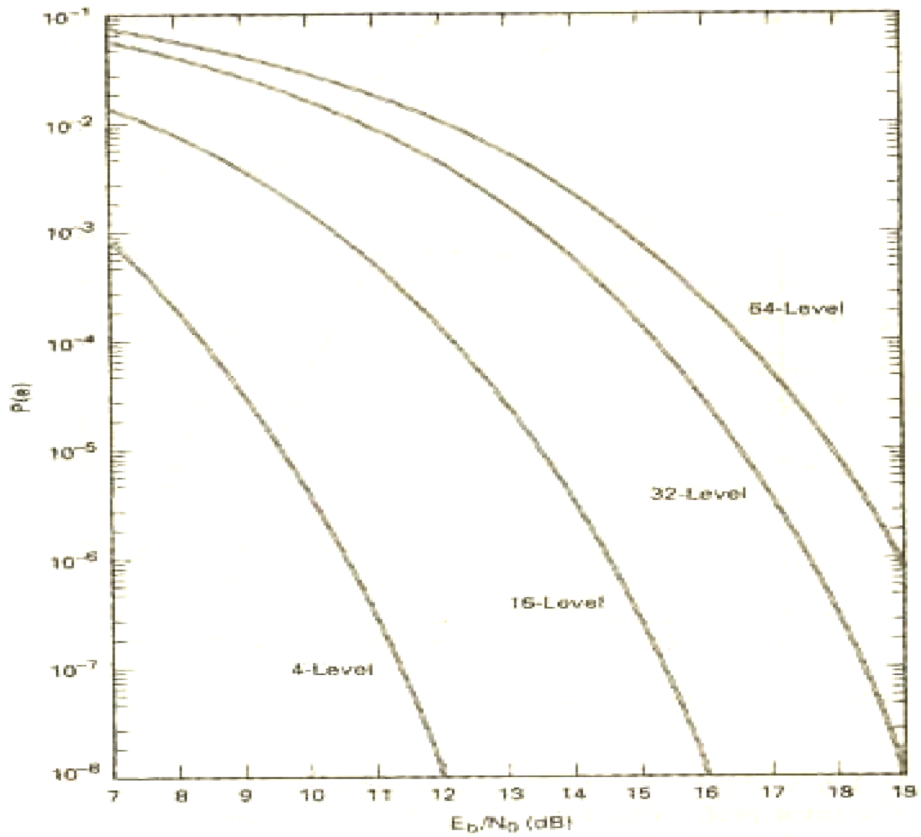


Figure 4.24 Comparison between the performance of different levels of QAM system.

Calculation of E_b/ N_o in digital radio systems.

In digital systems the basic measure of transmission quality is BER. With digital radio links, we employ the ratio E_b/ N_o as a measure of signal quality. Given a certain modulation type, we can derive BER from an E_b/ N_o curve.

In words, E_b/ N_o means energy per bit per noise spectral density ratio.

N_o is simply the thermal noise in 1 Hz of BW or:

The thermal noise level, P_n in 1 Hz of BW at absolute zero Kelvin is;

$P_n = -228.6$ dBW/ Hz of BW for a perfect receiver at absolute zero.

At room temperature 290° K we have:

$P_n = -204$ dBW / Hz of BW for a perfect receiver.

$= -174$ dBm / Hz of BW for a perfect receiver.

A *perfect receiver* is a receiving device that contributes no thermal noise to the communication channel, (cannot occur in real life).

Converting for a real receiver gives;

$P_n = -204$ dBW/ Hz + NF dB + 10 log B, (a)

where B is the BW in Hz of the receiver. NF is the noise figure of the receiver. An example of application of equation (a) might be a receiver with a 3-dB noise figure and a 10-MHz BW. What would be the thermal noise power (level) in dBW of the receiver?

Use Equation (a).

$P_n = -204$ dBW/ Hz + 3 dB + 10 log (10×10^6)

$= -204$ dBW/ Hz + 3 dB + 70 dB = -131 dBW.

$N_o = -204$ dBW/ Hz + NF dB. (b)

Example. For a receiver has a noise figure of 2.1 dB. What is its thermal noise Power density in 1 Hz of BW, N_o?

$N_o = -$ 204 dBW + 2.1 dB = -201.9 dBW/ Hz.

The receive signal level, RSL represents the total power (in dBm or dBW) entering the receiver front end, during, 1-sec duration (energy). We want the power carried by just 1 bit. For example, if the RSL were 1 W, and the signal was at 1000 bps, the energy per bit would be

1/ 1000 or 1 mW per bit. Then we define E_b as:

E_b = RSL dBm or dBW − 10 log (bit rate). (c)

As an example using typical values. The RSL into a certain receiver was − 89 dBW and bit rate was 2.048 Mbps. What is the value of E_b?

E_b = − 89 dBW − 10 log (2.048×10^6)

= − 89 dBW − 63.11 dB = −152.11 dBW.

We can now develop a formula for E_b/N_o:

E_b/N_o = RSL dBW - 10 log (bit rate) - (-204 dBW + NFdB). (d)

E_b/N_o = RSL dBW - 10 log (bit rate) + 204 dBW - NFdB. (d')

Some notes on E_b/N_o and its use. E_b/N_o, for a given BER, will be different for different types of modulation (e.g., FSK, PSK, QAM, etc.). When working with E_b, we divide RSL by the bit rate, not the symbol rate nor the baud rate. There is a theoretical E_b/N_o and a practical E_b/N_o. The practical is always a greater value than the theoretical, greater by the *modulation implementation loss* in dB, which compensates for system imperfections.

Figure 4.25 is an example of where BER is related to E_b/N_o. There are two curves in the figure. The first from the left is for BPSK/ QPSK and the second is for 8-ary PSK. The values are for coherent detection. Coherent detection means that the receiver has a built-in phase reference as a basis to make its binary or higher-level decisions.

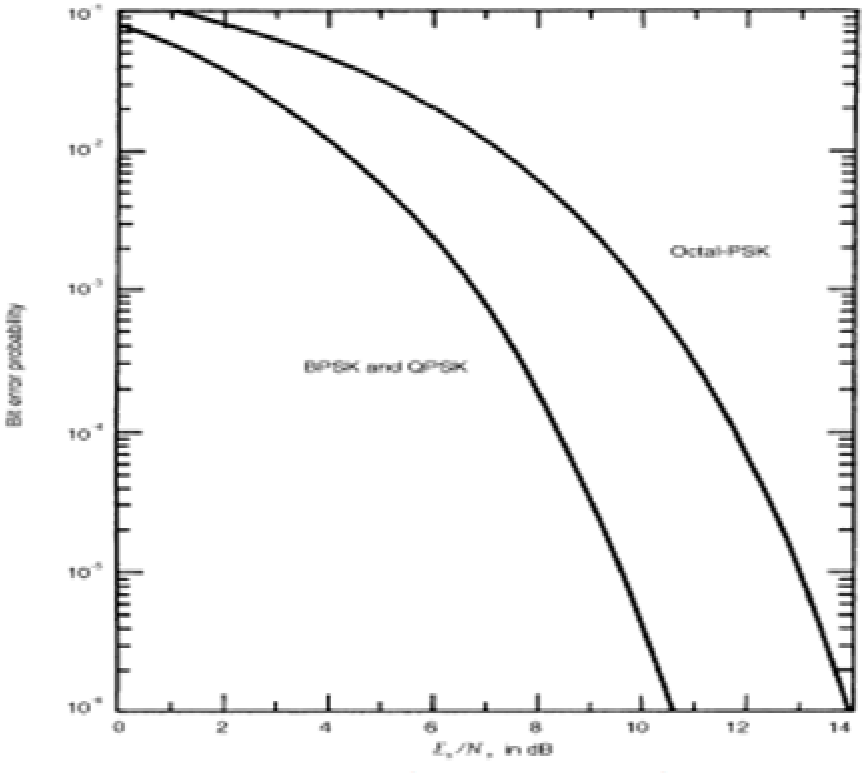


Figure 4.25 BER versus E_b/N_o performance for BPSK/QPSK and 8-ary PSK

References

1- Vijay K. Garg, "Wireless Communications and Networking", Elsevier Inc., 2007.

2- Marvin K. Simon, "Bandwidth-Efficient Digital Modulation with Application to Deep-Space Communications", JPL Publication 00 - 17, June 2001.

3- Simon Haykin, "Communication Systems", 4th Edition, John Wiley & Sons, Inc. 2001.

4- Bateman, Andy, "Digital Communications", 1st Edition, Addison-Wesley, 1998.

5- Glover and Grant, "Digital communications", 1st Edition, Prentice Hall, 1998.

6- R, P. Singh and S.D. Sapre, "Communication systems: analog and Digital", Tata McGraw-Hill book company, 1995.

7- L. W. Couch II, "Digital and analog communication systems", Macmillan Publishing company, 1990.

8- B.P . Lathi , "Modern Digital and analog communications" ,

9- Simon Haykin, "An introduction to analog and digital communication", John Wiley & Sons, 1989.

Ministry of Higher Education
Higher institute of Engineering & Technology, Kafr El-Sheikh

وزارة التعليم العالي
المعهد العالي للهندسة والتكنولوجيا بكفرالشيخ

رسالة المعهد:

يلتزم المعهد بإعداد خريجين متخصصين ومكتسبين للمهارات والجدارات طبقا للمعايير القومية الأكاديمية المرجعية لتلبية احتياجات سوق العمل المحلى والاقليمي وقادرين على إجراء بحوث علمية مبتكره وتقديم خدمات مجتمعية في إطار القيم الأخلاقية