NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Computer Engineering

CHANNEL NOISE EQUALIZATION

Graduation Project

COM- 400

Student: Ahmad Fathi Al-hajali

Supervisor: Assoc. Prof. Dr. Rahib Abiyev

Nicosia – 2003

ACKNOWLEDGEMENTS

"First, I would like to thank my supervisor Assoc. Prof. Dr. Rahib Abiyv for his invaluabl advice and belief in my work and myself over the course of this graduation project.

Second, Iwould like to all of the references that provided me with several very helpful suggestions to accomplish this project.

Third, it is a pleasure to thank my family for their outstanding patience, and love.

Finally, thanks to all my friends for their advices and support. In particular, thanks for Hamza özkarsli for his special friendship and support."

ABSTRACT

Standard equalization techniques start by modeling a communication channel as a filter, with a specific transfer function. The equalizer, which is part of the receiver, then estimates the parameters of this (unknown) transfer function, and attempts to undo the effects of this transfer function. Thus the equalizer is also a filter of a special kind. The simplest implementation, if the channel transfer function were perfectly known to the receiver, would be to filter the received signal by a filter whose transfer function is the inverse of the channel transfer function. However, this process is imperfect because the receiver generally knows or can estimate the channel only imperfectly, and most channels change with time, which calls for an adaptive estimation process. Moreover, even if the receiver could perfectly estimate the channel, the process of filtering with an inverse channel filter enhances the noise, which is invariably present along with the signal.

The linear equalizers, which are so-called because they perform a linear operation (filtering) on the received signal. Some of the issues with linear equalization can be addressed by implementing instead a decision feedback equalizer (DFE), which is a non-linear equalizer which works by recognizing that ISI effects on a symbol depend on previously transmitted symbols as well as symbols that are yet to come. This equalizer structure includes a feedback portion that subtracts the ISI effect of previously detected symbols from a current symbol. The only unknown is then the effect of symbols that have not yet been detected, which is minimized by appropriate (feedforward) filtering. The problem with DFE's is that the feedback loop introduces instabilities in the presence of noise. Thus both linear equalizers and DFE's have limitations on their usefulness in the presence of noise.

TABLE OF CONTENTS

INTRODUCTION

Every day, in our work and in our leisure time, we come in contact with and we use a variety of modern communication system and communication media, the most common being the telphone, radio and television. Through these media we are able to communicate (nearly) instantaneously with people on different continents, transact our daily business, and receive information about various developments and events of note that occur all over the world. Electronic mail and facsimile transmission have made it possible to rapidly communicate written messages across great distance.

Can you imagine a world without telephones radio, and television? Yet, when you think about it , most of these modern-day communication systems werw invented and developed during the past century. As we move towards a more information centric world, and the appetite for data rates increases, communication system designers face ever-increasing challenges in utilizing available bandwidth more efficiently. In the wireless world, this means sending more bits in every available Hertz of spectrum, while keeping the quality of the reception the same as before. In the wired world, spectrum limitations arise because of the transmission characteristics of media such as copper wire, and even, as distance increases, optical fiber. As more bits per Hertz are sent, the bits overlap and interfere with each other, and reduce the received signal quality unless special algorithms, called channel equalization, are used. Equalization theory is well understood, but practical implementations are an area of active research and development. Signal representation via chromatic derivatives, which accurately describe the local behavior of a signal, enables realizable implementations of equalization for complex channels.

In chapter one we begin with the structure of a communication system, digital communication system, channels, charactristics of a communication channels, and the mathematical representation of channels. Chapter two deals with carrier modulation methods for transmitting digital information over a communication channels, and we describe (ASM), (PSK), and (QAM).Chapter three is concerned with designing of transmitting and receiving filters, and channel equalization, we concentrate our discussion with adaptive equalizer and its design for mitigating the effects of amplitude and phase distortion that is encountered in signal transmission.

CHAPTER ONE

ELEMENTS OF AN ELECTRICAL COMMUNICATION SYSTEM

1.1 Structure of Communication System

Electrical communication system is designed to send messages or information from a source that generates the messages to one or one destinations. In general, a communiction system can be represented by the functional block diagram shown in Figure 1.1.

The information generated by the source may be of the form of voice (speech source), a picture (image source), or plain text in some particular language, such as English, Japanese, German, French, etc. an essential feature of any source that generates information is that its output is described in probabilistic terms; that is, the output of a source is not deterministic. Otherwise, there would be no need to transmit the message.

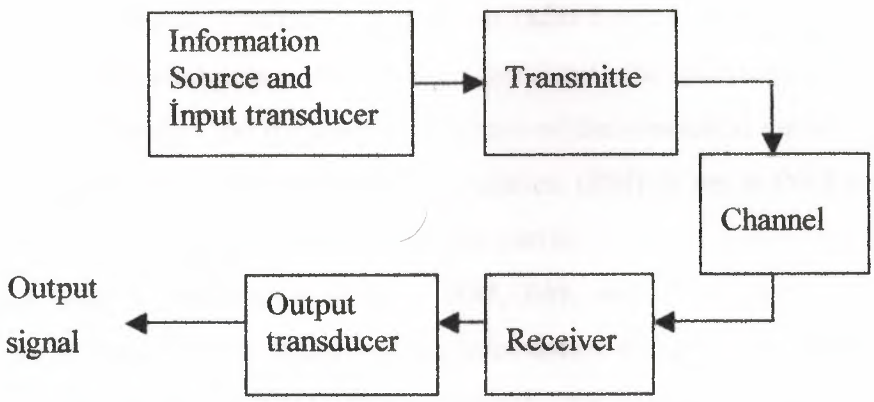


FIGURE 1.1. Functional block diagram of a communication system.

A transducer is usually required to convert the output of a source into an electrical signal that is suitable for transmission. For example, a microphone serves as the transducer that converts an acoustic speech signal into an electrical signal, and a video camera converts an image into an electrical signal. At the destination, a similar transducer is required to convert the electrical signals that are received into a form that is suitable for the user; for example, acoustic signals, images, etc.

The heart of the communication system consists of three basic parts, namely, the transmitter, the channel, and the receiver. The functions performed by these three elements are described below.

1.1.1 The Transmitter

The transmitter converts the electrical signal into a form that is suitable for transmission through the physical channel or transmission medium. For example, in broadcast, the Federal Communications Commission (FCC) specifies the frequency range for each transmitting station. Hence, the transmitter must translate the information signal to be transmitted into the appropriate frequency range that matches the frequency allocation assigned to the transmitter. Thus, signals transmitted by multiple radio stations do not interfere with one another. Similar functions are performed in telephone communication systems, where the electrical speech signals from many users are transmitted over the same wire.

In general, the transmitter performs the matching of the message signal to the channel by a process called modulation. Usually, modulation involves the use of the information signal to systematically vary the amplitude, frequency, or phase of asinusoidal carrier. For example, in AM radio broadcast, the information signal that is transmitted is contained in the amplitude variations of the sinusoidal carrier, which is the center frequency in the frequency band allocated to the radio transmitting station. This is an example of amplitude modulation. In FM radio broadcast, the information signal that is transmitted is contained in the frequency variations of the sinusoidal carrier. This is an example of frequency modulation. Phase modulation (PM) is yet a third method for impressing the information signal on a sinusoidal carrier.

In general, carrier modulation such as AM, FM, and PM is performed at the transmitter, as indicated above, to convert the information signal to a form that matches the characteristic of the channel. Thus, through the process of modulation, the information signal is translated in frequency to match the allocation of the channel. The choice of the type of modulation is based on several factors, such as the amount of bandwidth allocated, the types of noise ana interference that the signal encounters in transmission over the channel, and the electronic devices that are available for signal amplification prior to transmission. In any case, the modulation process makes it possible to accommodate the transmission of multiple messages from many users over the same physical channel.

In addition to modulation, other functions that are usually performed at the transmitter are filtering of the information-bearing signal, amplification of the modulated signal, and in the case of wireless transmission, radiation of the signal by means of a transmitting antenna.

1.1.2 The Channel

The communications channel is the physical medium that is used to send the signal from the transmitter to the receiver. In wireless transmission, the channel is usually the atmosphere (free space). On the other hand, telephone channels usually employ a variety of physical media, including wirelines, optical fiber cables, and wireless (microwave radio). Whatever the physical medium for signal transmission, the essential feature is that the transmitted signal is corrupted in a random manner by a variety of possible mechanisms. The most common form of signal degradation comes in the form of additive noise, which is generated at the front end of the receiver, where signal amplification is performed. This noise is often called thermal noise. In wireless transmission, additional additive disturbances are man-made noise and atmospheric noise picked up by a receiving antenna. Automobile ignition noise is an example of man-made noise, and electrical lightning discharges from thunderstorms is an example of atmospheric noise. Interference from other users of the channel is another form of additive noise that often arises in both wireless and wireline communication systems.

In some radio communication channels, such as the ionospheric channel that is used for long-range, short-wave radio transmission, another form of signal degradation is multipath propagation. Such signal distortion is characterized as a nonadditive signal disturbance, which manifests itself as time variations in the signal amplitude, usually called fading. This phenomenon is described in more detail in Section 1.3.

Both additive and nonadditive signal distortions are usually characterized as random phenomena and described in statistical terms. The effect of these signal distortions must be taken into account in the design of the communication system.

In the design of a communication system, the system designer works with mathematical models that statistically characterize the signal distortion encountered on physical channels. Often, the statistical description that is used in a mathematical model is a result of actual empirical measurements obtained from experiments involving signal transmission over such channels. In such case, there is a physical justification for the mathematical model used in the design of communication systems. On the other hand, in some communication system designs, the statistical characteristics of the channel may vary significantly with time. In such cases, the system designer may design a communication system that is robust to the variety of signal distortions. This can be accomplished by having the system adapt some of its parameters to the channel distortion encountered.

1.1.3 The Receiver

The function of the receiver is to recover the message signal contained in the received signal. If the message signal is transmitted by carrier modulation, the receiver performs carrier demodulation to extract the message from the sinusoidal carrier. Since the signal demodulation is performed in the presence of additive noise and possibly other signal distortion, the demodulated message signal is generally degraded to some extent by the presence of these distortions in the received signal. As we shall see, the fidelity of the received message signal is a function of the type of modulation, the strength of the additive noise, the type and strength of any other additive interference, and the type of any nonadditive interference.

Besides performing the primary function of signal demodulation, the receiver also performs a number of peripheral functions, including signal filtering and noise suppression.

1.2 Digital Communication System

Up to this point, we have described an electrical communication system in rather broad terms based on the implicit assumption that the message signal is a continuous time-varying waveform. We refer to such continuous-time signal waveforms as analog signals and to the corresponding information sources that produce such signals as analog sources. Analog signals can be transmitted directly via carrier modulation over the communication channel and demodulated accordingly at the receiver. We call such a communication system an analog communication system.

Alternatively, an analog source output may be converted into a digital form and the message can be transmitted via digital modulation and demodulated as a digital signal at the receiver. There are some potential advantages to transmitting an analog signal by means of digital modulation. The most important reason is that signal fidelity is better controlled through digital transmission than analog transmission. In particular, digital transmission allows us to regenerate the digital signal in long-distance transmission, thus eliminating effects of noise at each regeneration point. In contrast, the noise added in analog transmission is amplified analog with the signal when amplifiers are used periodically to boost the signal level in long-distance transmission. Another reason for choosing digital transmission over analog is that the analog message signal may be highly redundant. With digital processing, redundancy may be removed prior to modulation, thus conserving channel bandwidth. Yet a third reason may be that digital communication systems are often cheaper to implement.

In some applications, the information to be transmitted is inherently digital, e.g., in the form of English text, computer data, etc. In such cases, the information source that generates the data is called adiscrete (digital) source.

In a digital communication system, the functional operations performed at the transmitter and receiver must be expanded to include message signal discretization at the transmitter and message signal synthesis or interpolation at the receiver. Additional functions include redundancy removal, and channel coding and decoding.

Figure 1.2 illustrates the functional diagram and the basic elements of a digital communication system. The source output may be either an analog signal, such as audio or video signal, or a digital signal, such as the output of a Teletype machine, which is discrete in time and has a finite number of output characters. In

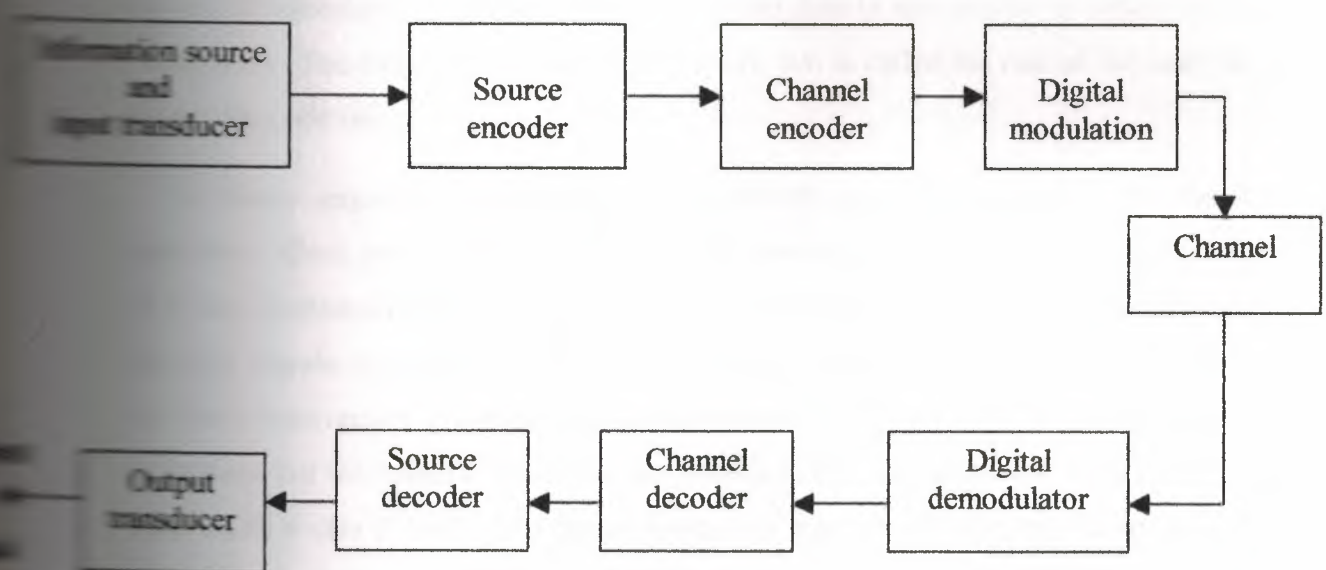


FIGURE 1.2. Basic elements of digital communication system

a digital communication system, the messages produced by the source is usually converted into a sequence of binary digits. Ideally, we would like to represent the source output (message) by as few binary digits as possible. In other words, we seek an efficient representation of the source output that results in little or no redundancy. The process of efficiently converting the output of either an analog or a digital source into a sequence of binary digits is called source encoder or data compression. The process of efficiently converting the output of either an analog or a digital source into a sequence of binary digits is called source encoder or data compression.

The sequence of binary digits from the source encoder, which we call the information sequence, is passed to the channel encoder. The purpose of the channel encoder is to introduce in a controlled manner some redundancy in the binary information sequence, which can be used at the receiver to overcome the effects of noise and interference encountered in the transmission of the signal through the channel. Thus, the added redundancy serves to increase the reliability of the received data and improves the fidelity of the received signal. In effect, redundancy in the information sequence aids the receiver in decoding the desired information sequence. For example, a (trivial) form of encoding of the binary information sequence is simply to repeat each binary digit m times, where m is some positive integer. More sophisticated (nontrivial) encoding involves taking k information bits at a time and mapping each k-bit sequence into a unique n-bit sequence, called acode word. The amount of redundancy introduced by encoding the data in this manner is measured by the ratio n/k. The reciprocal of this ratio, namely, k/n is called the rate of the code or, simply, the code rate.

The binary sequence at the output of the channel encoder is passed to the digital modulator, which serves as the interface to the communications channel. Since nearly all of the communication channels encountered in practice are capable of transmitting electrical signals (waveforms), the primary purpose of the digital modulator is to map the binary information sequence into signal waveforms. To elaborate on this point, let us suppose that the coded information sequence is to be transmitted one bit at a time at some uniform rate R bits/s The digital modulator may simply map the binary digit 0 into a waveform$s_0(t)$ and the binary digit 1 into a waveform $s_1(t)$. In this manner, each bit from the channel encoder is transmitted separately. We call this binary modulation. Alternatively, the modulator may transmit b coded information bits at a time by using M =2b distinct waveformsi(t), I = 0, 1, ..., m-1, one waveform for each of the 2b possible b-bits equences. We call this M-ary modulation (M >2). Note that a new b-bit sequence enters the modulator every b/R seconds. Hence, when the channel bit rateR is fixed, the amount of time available to transmit one of the M waveforms corresponding to a b-bit sequence is b times the period in a system that uses binary modulation.

At the receiving end of a digital communications system, the digital demodulator processes the channel-corrupted transmitted waveform and reduces each waveform to a single number that represents an estimate of the transmitted data symbol (binary or M-ary). For example, when binary modulation is used, the demodulator may process the received waveform and decide on whether the transmitted bit is a 0 or 1. In such a case, we say the demodulator has made a binary decision. As one alternative, the demodulator may make a ternary decision; that is, it decides that the transmitted bit is either a 0 or 1 or it makes no decision at all, depending on the apparent quality of the received signal. When no decision is made on a particular bit, we say that the demodulator has inserted an erasure in the demodulated data. Using the redundancy in the transmitted data, the decoder attempts to fill in the positions where erasures occurred. Viewing the decision process performed by the demodulator as a form of quantization, we observe that binary and ternary decisions are special cases of a demodulator that quantizes to Q levels, where $Q \geq 2$ In general, if the digital communications system employs M-ary modulation, where $m=0,1, \ldots , M$ represent the M possible transmitted symbols, each corresponding to $k = \log_2 M$ bits, the demodulator may make A Q-ary decision, where $Q \geq M$ .In the extreme case where no quantization is performed, $Q = \infty$.

When there is no redundancy in the transmitted information, the demodulator must decide which of the M waveforms was transmitted in any given time interval. Consequently, $Q \quad M$, and since there is no redundancy in the transmitted information, no discrete channel decoder is used following the demodulator. On the other hand, when there is redundancy introduced by a discrete channel encoder at the transmitter, the Q-ary output from the demodulator occurring every k/R seconds is fed to the decoder, which attempts to reconstruct the original information sequence from knowledge of the code used by the channel encoder and the redundancy contained in the received data.

A measure of how well the demodulator and encoder perform is the frequency with which errors occur in the decoded sequence. More precisely, the average probability of a bit-error at the output of the decoder is a measure of the performance of the demodulator-decoder combination. In general, the probability of error is a function of the code characteristics, the types of waveforms used to transmit the information over the channel, the transmitter power, the characteristics of the channel (i.e., the amount of noise), the nature of the interference, etc., and the method of demodulation and decoding. These items and their effect on performance will be discussed in detail in subsequent chapters.

As a final step, when an analog output is desired, the source decoder accepts the output sequence from the channel decoder, and from knowledge of the source encoding method used, attempts to reconstruct the original signal from the source. Due to channel decoding errors and possible distortion introduced by the source encoder and, perhaps, the source decoder, the signal at the output of the source decoder is an approximation to the original source output. The difference or some function of the difference between the original signal and the reconstructed signal is a measure of the distortion introduced by the digital communications system.

1.2.1 Early Work in Digital Communications

Although Morse is responsible for the development of the first electrical digital communication system (telegraphy), the beginnings of what we now regard as modem digital communications stem from the work of nyquist (1924), who investigated the problem of determining the maximum signaling rate that can be used over a telegraph channel of a given bandwidth without intersymbol interference. He formulated a model of a telegraph system in which a transmitted signal has the general form

$$S(t) = \sum_{n} a_n g(t-nT)$$

where g(t) represents a basic pulse shape and {an} is the binary data sequence of {±1} transmitted at a rate of 1/T bits per second. Nyquist set out to determine the optimum pulse shape that was bandlimited to W Hz and maximized the bit rate 1/T under the constraint that the pulse caused no intersymbol interference at the sampling times k/T, k = 0, ±1, ±2,.... His studies led him to conclude that the maximum pulse rate1/T is 2Wpulses per second. This rate is now called the yquist rate.Moreover, this pulse rate can be achieved by using the pulses $g(t) = (\sin 2\pi Wt)/2\pi Wt$.This pulse shape allows the recovery of the data without intersymbol interference at the sampling instants nyquist's result is equivalent to a version of the sampling theorem for bandlimited signals, which was later stated precisely by Shannon (1948). The sampling theorem states that a signal of bandwidth W can be reconstructed from samples taken at the Nyquist rate of 2W samples per second using the interpolation formula

$$S(t) = \sum_n s\left(\frac{n}{2W}\right) \frac{\sin 2\pi W(t - n/2W)}{2\pi W(t - n/2W)}$$

In light of Nyquist's work. Hartley (1928) considered the issue of the amount of data that can be transmitted reliably over a bandlimited channel when multiple amplitude levels are used. Due to the presence of noise and other interference, Hart-ley postulated that the receiver could reliably estimate the received signal amplitude to some accuracy, say A_δ. This investigation led Hartley to conclude that there is maximum data rate that can be communicated reliably over a bandlimited channel when the maximum signal amplitude is limited to Amax (fixed power constraint) and the amplitude resolution is A_δ.

Another significant advance in the development of communications was the work of Wiener (1942) who considered the problem of estimating a desired signal waveform s(t)in the presence of additive noise n(t), based on observation of the received signal r(t) = s(t) + n(t). This problem arises in signal demodulation. Wiener determined the linear filter whose output is the best mean-square approximation to the desired signal s(t). The resulting filter is called the optimum linear (Wiener) filter. Hartley's and Nyquist results on the maximum transmission rate of digital information were precursors to the work of Shannon (1948 a, b) who established the mathematical foundations for information theory and derived the fundamental limits for digital communication systems. In his pioneering work. Shannon formulated the basic problem of reliable transmission of information in statistical terms, using probabilistic models for information sources and communication channels. Based on such a statistical formulation, he adopted a logarithmic measure for the information content of a source. He also demonstrated that the effect of a transmitter power constraint, a bandwidth constraint, and additive noise can be associated with the channel and incorporated into a single parameter, called the channel capacity For example, in the case of an additive white (spectrally flat) Gaussian noise interference, an ideal bandlimited channel of bandwidth W has a capacity C given by

$$C = W \log_2\left(1 + \frac{P}{WN_0}\right) \text{ bits/s}$$

where P is the average transmitted power and n0 is the power spectral density of the additive noise. The significance of the channel capacity is as follows: If the information rate R from the source is less than C (R < C), then it is theoretically possible to achieve reliable (error-free) transmission through the channel by appropriate coding. On the other hand, if R > C, reliable transmission is not possible regardless of the amount of signal processing performed at the transmitter and receiver. Thus, Shannon established basic limits on communication of information and gave birth to a new field that is now called information theory.

Initially the fundamental work of Shannon had a relatively small impact on the design and development of new digital communications systems. In part, this was due to the small demand for digital information transmission during the 1950's. Another reason was the relatively large complexity and, hence, the high cost of digital hardware required to achieve the high efficiency and high reliability predicted by Shannon's theory.

Another important contribution to the field of digital communications is the work of Kotelnikov (1947), which provided a coherent analysis of the various digitalcommunication systems based on a geometrical approach. Kotelnikov approach was later expanded by Wozencraft and Jacobs (1965).

The increase in the demand for data transmission during the last three decades, coupled with the development of more sophisticated integrated circuits, has led to the development of very efficient and more reliable digital communications systems. In the course of these developments Shannon's original results and the generalization of his results on maximum transmission limits over a channel and on bounds on the performance achieved have served as benchmarks for any given communications system design. The theoretical limits derived by Shannon and other researchers that contributed to the development of information theory serve as an ultimate goal in the continuing efforts to design and develop more efficient digital communications systems.

Following Shannon's publications came the classic work of Hamming (1950) on error detecting and error-correcting codes to combat the detrimental effects of channel noise. Hamming's work stimulated many researchers in the years that followed, and a variety of new and powerful codes were discovered, many of which are used today in the implementation of modem communication systems.

1.3 Communication Channels and Their Characteristics

As indicated in the preceding discussion, the communication channel provides the connection between the transmitter and the receiver. The physical channel may be a pair of wires that carry the electrical signal, or an optical fiber that carries the information on a modulated light beam, or an underwater ocean channel in which the information is transmitted acoustically, or free space over which the information-bearing signal is radiated by use of an antenna. Other media that can be characterized as communication channels are data storage media, such as magnetic tape, magnetic disks, and optical disks.

One common problem in signal transmission through any channel is additive noise. In general, additive noise is generated internally by components such as resistors and solid-state devices used to implement the communication system. This is sometimes called thermal noise. Other sources of noise and interference may arise externally to the system, such as interference from other users of the channel. When such noise and interference occupy the same frequency band as the desired signal, its effect can be minimized by proper design of the transmitted signal and its demodulator at the receiver. Other types of signal degradations that may be encountered in transmission over the channel are signal attenuation, amplitude and phase distortion, and multipath distortion.

Increasing the power in the transmitted signal may minimize the effects of noise. However, equipment and other practical constraints limit the power level in the transmitted signal. Another basic limitation is the available channel bandwidth. A bandwidth constraint is usually due to the physical limitations of the medium and the electronic components used to implement the transmitter and the receiver. These two limitations result in constraining the amount of data that can be transmitted reliably over any communications channel. Shannon's basic results relate the channel capacity to the available transmitted power and channel bandwidth.

Below, we describe some of the important characteristics of several communication channels.

1.3.1 Wireline Channels

The telephone network makes extensive use of wire lines for voice signal transmission, as well as data and video transmission. Twistedpair wire lines and coaxial cable are basically guided electromagnetic channels, which provide relatively modest bandwidths Telephone wire generally used to connect a customer to a central office has a bandwidth of several hundred kilo-hertz (kHz). On the other hand, coaxial cable has a usable bandwidth of several megahertz (MHz). Figure 1.3 illustrates the frequency range of guided electromagnetic channels, which includes waveguides and optical fibers.

Signals transmitted through such channels are distorted in both amplitude and phase and further corrupted by additive noise. Twisted-pair wireline channels are also prone to crosstalk interference from physically adjacent channels. Because wireline channels carry a large percentage of our daily communications around the country and the world, much research has been performed on the characterization of their transmission properties and on methods for mitigating the amplitude and phase distortion encountered in signal transmission.

1.3.2 Fiber Optic Channels

Optical fibers offer the communications system designer a channel bandwidth that is several orders of magnitude larger than coaxial cable channels. During the past decade, optical fiber cables have been developed that have a relatively low signal attenuation, and highly reliable photonic devices have been developed for signal generation and signal detection. These technological advances have resulted in a rapid deployment of optical fiber channels, both in domestic telecommunication systems as well as for trans-atlantic and trans-pacific communications. With the large bandwidth available on fiber optic channels it is possible for telephone companies to offer subscribers a wide array of telecommunication services, including voice, data, facsimile, and video.

The transmitter or modulator in a fiber optic communication system is a light source, either a light-emitting diode (LED) or a laser. Information is transmitted by varying (modulating) the intensity of the light source with the message signal. The light propagates through the fiber as a light wave and is amplified periodically (in the case of digital transmission, it is detected and regenerated by repeaters) along the transmission path to compensate for signal attenuation. At the receiver, the light intensity is detected by a photodiode, whose output is an electrical signal that varies in direct proportion to the power of the light impinging on the photodiode.

It is envisioned that optical fiber channels will replace nearly all wireline channels in the telephone network by the turn of the century.

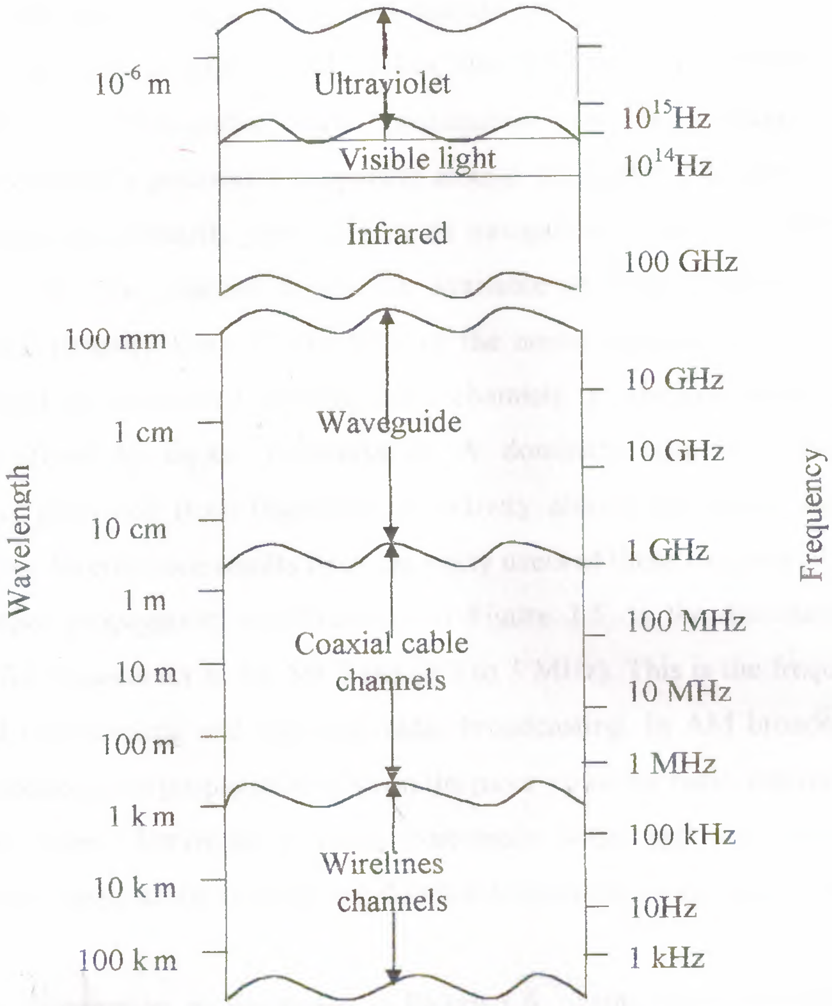


FIGURE 1.3. Frequency range for guided wire channel.

1.3.3 Wireless Electromagnatic Channels

In wireless communication systems, electromagnatic energy is coupled to the propagation medium by an antenna, which serves as the radiator. The physical size and the configuration of the antenna depend primarily on the frequency of operation. To obtain efficient radiation of electromagnatic energy the antenna must be longer than 1/10 of the wavelength. Consequently, a radio transmitting in the AM frequency band, say at 1 MHz (corresponding to a wavelength of $\lambda = c/fc = 300m$), requires an antenna of at least 30 meters.

Figure 1.4 illustrates the various frequency bands of the electromagnetic spectrum. The mode of propagation of electromagnetic waves in the atmosphere and in free space may be subdivided into three categories, namely, ground-wave propagation, sky-wave propagation, and light-of-sight (LOS) propagation. In the VLF and ELF frequency bands, where the wavelengths exceed 10 km, the earth and the ionosphere act as a waveguide for electromagnetic wave propagation. In these frequency ranges, communication signals practically propagate around the globe. For this reason, these frequency bands are primarily used to provide navigational aids from shore to ships around the world. The channel bandwidth available in these frequency bands are relatively small (usually from 1% to 10% of the center frequency), and hence, the information that is transmitted through these channels is relatively slow speed and, generally, confined to digital transmission. A dominant type of noise at these frequencies is generated from thunderstorm activity around the globe, especially in tropical regions. Interference results from the many users of these frequency bands.

Ground-wave propagation, as illustrated in Figure 1.5, is the dominant mode of propagation for frequencies in the MF band (0.3 to 3 MHz). This is the frequency band used for AM broadcasting and maritime radio broadcasting. In AM broadcasting, the range with ground-wave propagation of even the more powerful radio stations is limited to about 100 miles. Atmospheric noise, man-made noise, and thermal noise from electronic components at the receiver are dominant disturbances for signal transmission of MF.

Sky-wave propagation, as illustrated in Figure 1.6, results from transmitted signals being reflected (bent or refracted) from the ionosphere, which consists of several layers of charged particles ranging in altitude from 30 to 250 miles above the surface of the earth. During the daytime hours, the heating of the lower atmosphere by the sun causes the formation of the lower layers at altitudes below 75 miles. These lower layers, especially the D-layer serves to absorb frequencies below 2 MHz, thus severe y limiting sky-wave propagation of AM radio broadcast. However, during the night-time hours, the electron density in the lower layers of the ionosphere drops sharply and the frequency absorption that occurs during the daytime is significantly reduced. As a consequence, powerful AM radio broadcast stations can propagate over large distances via sky wave over the F-layer of the ionosphere, which ranges from 90 miles to 250 miles above the surface of the earth.

A frequently occurring problem with electromagnetic wave propagation via sky wave in the HF frequency range is signal multipath. Signal multipath occurs when the

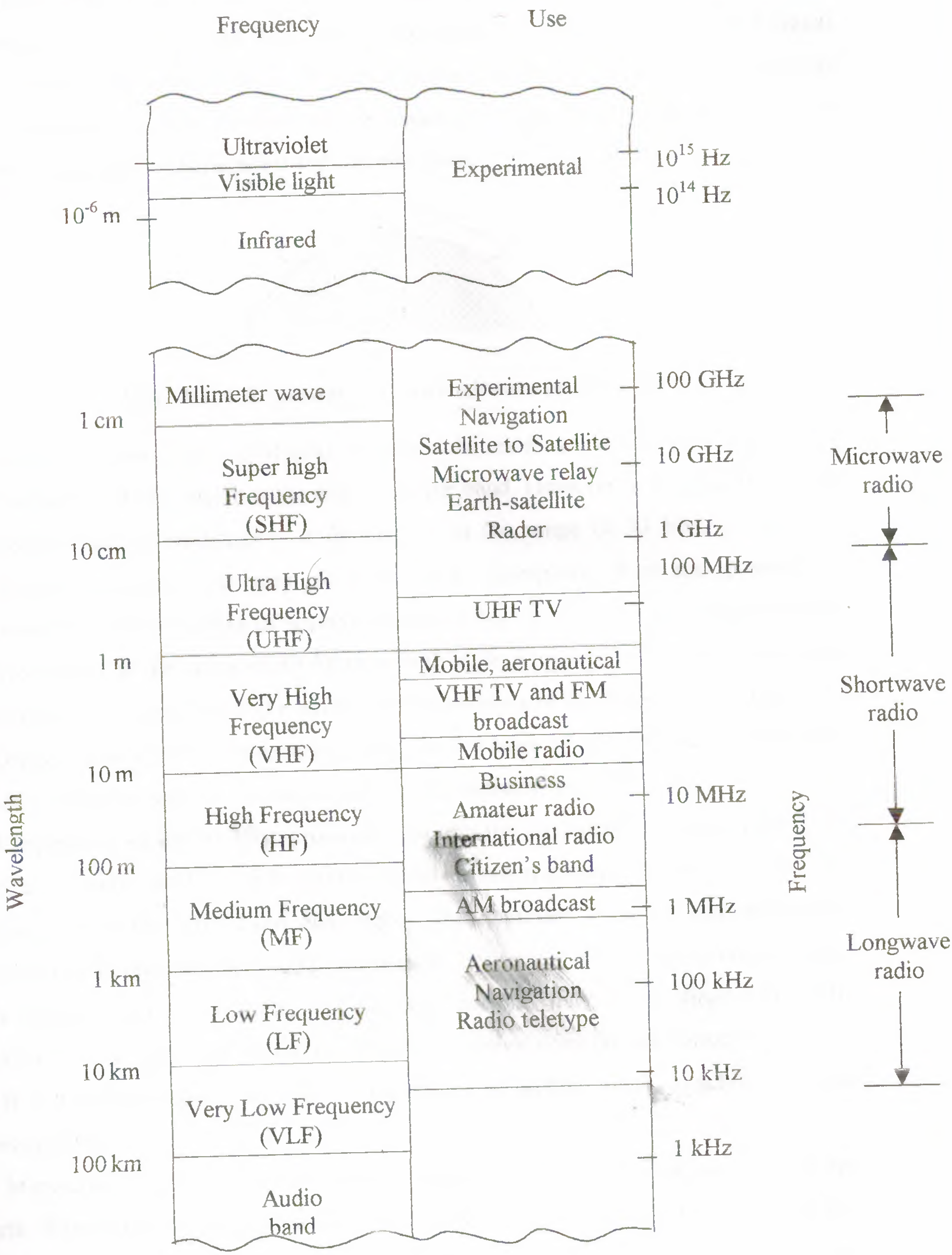


FIGURE 1.4. Frequency range wireless electro magnetic channel.

transmitted signal arrives at the receiver via multiple propagation paths at different delays. Signal multipath generally results in intersymbol interference in a digital communication system. Moreover, the signal components arriving via different propagation paths may add destructively, resulting in a phenomenon called signal fading, which most people have experienced when listening to a distant radio station at night when sky wave is the dominant propagation Mode. Additive noise at HF is a combination of atmospheric noise and thermal voice.

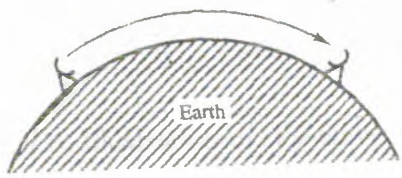


FIGURE 1.5. Illustration of ground-wave propagation.

Sky-wave ionospheric propagation ceases to exist at frequencies above approximately 30 MHz, which is the end of the HF band. However, it is possible to have ionospheric scatter propagation at frequencies in the range of 30 MHz to 60MHz, resulting from signal scattering from the lower ionosphere. It is also possible to communicate over distances of several hundred miles by use of troposphere scattering at frequencies in the range of 40 MHz to 300 MHz. Tropoacatter results from signal scattering due to particles in the atmosphere at altitudes of 10 miles or less. Generally, ionospheric scatter and tropospheric scatter involve large signal propagation losses and require a large amount of transmitter power and relatively large antennas.

Frequencies above 30 MHz propagate through the ionosphere with relatively little loss and make satellite and extraterrestrial communications possible. Hence, at frequencies in the VHP band and higher, the dominant mode of electromagnetic propagation is line-of-sight (LOS) propagation. For terrestrial communication systems, this means that the transmitter and receiver antennas must be in direct LOS with relatively little or no obstruction. For this reason, television stations transmitting in the VHF and UHF frequency bands mount their antennas on high towers to achieve a broad coverage area.

In general, the coverage area for LOS propagation is limited by the curvature of the earth. If the transmitting antenna is mounted at a height h feet above the surface of the earth, the distance to the radio horizon, assuming no physical obstructions such as mountains, is approximately=$\sqrt{2h}$ miles. For example, a TV antenna mounted on a tower of 1000 ft. in height provides coverage of approximately 50 miles. As another example, microwave radio relay systems used extensively for telephone and video transmission at frequencies above 1 GHz have antennas mounted on tall towers or on the top of tall buildings.

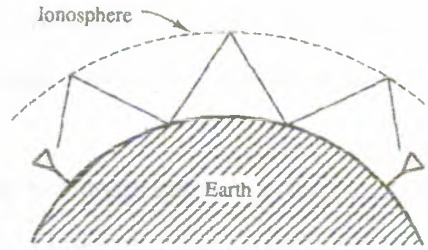


FIGURE 1.6. Illustration of sky-wave propagation.

The dominant noise limiting the performance of communication systems in the VHF and UHF frequency ranges is thermal noise generated in the receiver front end and cosmic noise picked up by the antenna. At frequencies in the SHF band above 10 MHz, atmospheric conditions play a major role in signal propagation. Figure1.7 illustrates the signal attenuation indB/mile due to precipitation for frequencies in there range of 10 to 100GHz. We observe that heavy rain introduces extremely high propagation losses that can result in service outages (total breakdown in the communication system).

At frequencies above the EHF band, we have the infrared and visible light regions of the electromagnetic spectrum, which can be used to provide LOS optical

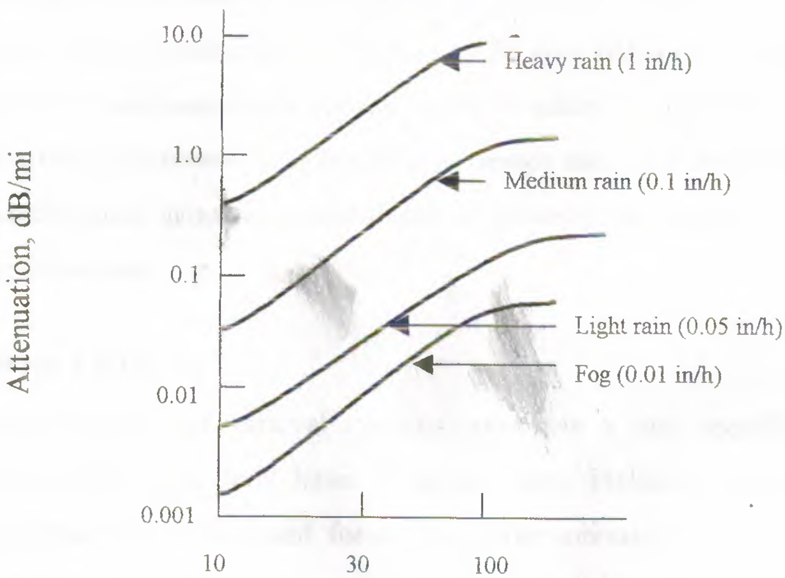


FIGURE 1.7. Illustration of sky-wave propagation.

communication in free space. To date, these frequency bands have been used in experimental communication systems, such as satellite-to-satellite links.

1.3.4 Underwater Acoustic Channels

Over the past few decades, ocean exploration activity has been steadily increasing. Coupled with this increase is the need to transmit data collected by sensors placed under water to the surface of the ocean. From there it is possible to relay the data via a satellite to a data collection center.

Electromagnetic waves do not propagate over long distances under water except at extremely low frequencies. However, the transmission of signals at such low frequencies is prohibitively expensive because of the large and powerful transmitters required. The attenuation of electromagnetic waves in water can be expressed in terms of the skin depth, which is the distance a signal is attenuated by $1/e$. For sea water, the skin depth $\delta = 250/\sqrt{f}$ where f is expressed in Hz and δ is in meters. For example, at 10 kHz, the skin depth is 2.5 meters. In contrast, acoustic signals propagate over distances of tens and even hundreds of kilometers.

An underwater acoustic channel is characterized as amultipath channel due to signal reflections from the surface and the bottom of the sea. Because of wave motion, the signal multipath components undergo time-varying propagation delays, which result in signal fading. In addition, there is frequency-dependent attenuation, which is approximately proportional to the square of the signal frequency.

Ambient ocean acoustic noise is caused by shrimp, fish, and various mammals. Near harbors, there is also man-made acoustic noise in addition to the ambient noise. In spite of this hostile environment, it is possible to design and implement efficient and highly reliable underwater acoustic communication systems for transmitting digital signals over large distances.

1.3.5 Storage Channels

Information storage and retrieval systems constitute a very significant part of data-handling activities on a daily basis. Magnetic tape, including digital audio tape and video tape, magnetic disks used for storing large amounts of computer data, optical disks used for computer data storage, and compact disks are examples of data storage systems that can be characterized as communication channels. The process of storing data on a magnetic tape or a magnetic or optical disk is equivalent to transmitting a signal over a telephone or a radio channel. The readback process and the signal processing involved in storage systems to recover the stored information is equivalent to the functions performed by a receiver in a telephone or radio communication system to recover the transmitted information.

Additive noise generated by the electronic components and interference from adjacent tracks is generally present in the readback signal of a storage system, just as is the case in a telephone or a radio communication system.

The amount of data that can be stored is generally limited by the size of the disk or tape and the density (number of bits stored per square inch) that can be achieved by the write/read electronic systems and heads. For example, a packing density of 10^9 bits per square inch has been recently demonstrated in an experimental magnetic disk storage system. (Current commercial magnetic storage products achieve a much lower density.) The speed at which data can be written on a disk or tape and the speed at which it can be read back is also limited by the associated mechanical and electrical subsystems that constitute an information storage system.

Channel coding and modulation are essential components of a well-designed digital magnetic or optical storage system. In the readback process, the signal is demodulated and the added redundancy introduced by the channel encoder is used to correct errors in the readback signal.

1.4 Mathematical Models For Communication Channels

In the design of communication systems for transmitting information through physical channels, we find it convenient to construct mathematical models that reflect the most important characteristics of the transmission medium. Then, the mathematical model for the channel is used in the design of the channel encoder and modulator at the transmitter and the demodulator and channel decoder at the receiver. Below, we provide a brief description of the channel models that are frequently used to characterize many of the physical channels that we encounter in practice.

1.4.1 The Additive Noise Channel

The simplest mathematical model for a communication channel is the additive noise channel, illustrated in Figure 1.8. In this model, the transmitted signal s(t) is corrupted by an additive random noise process n(t) Physically, the additive noise process may arise from electronic components and amplifiers at the receiver of the communication system, or from interference encountered in transmission as in the case of radio signal transmission.

IF the noise is introduced primarily by electronic components and amplifiers at the receiver, it may be characterized as thermal noise. This type of noise is characterized statistically as a Gaussian noise procass. Hence, the resulting mathematical model for model applies to a broad class of physical communication channels and because of its mathematical tractability this is the predominant channel model used in our the channel is usually called the additive Gaussian noise channel. Because this channel communication system analysis and design. Channel attenuation is easily incorporated into the model. When the signal undergoes

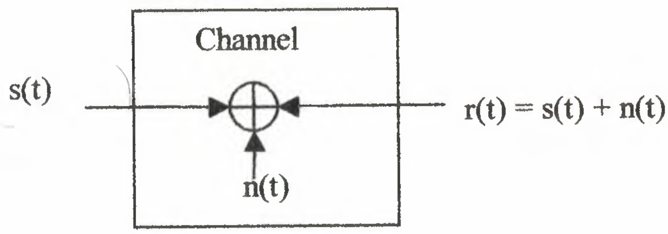


FIGURE 1.8. The additive noise channel.

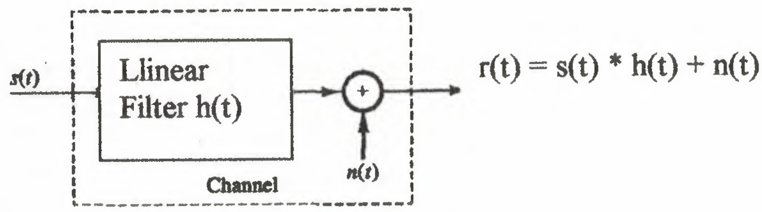


FIGURE 1.9. The linear filter channel with additive noise.

attenuation in transmission through the channel, the received signal is

$$r(t) = \alpha\, s(t) + n(t) \qquad (1.4.1)$$

where α represents the attenuation factor.

1.4.2. The Linear Filter Channel

In some physical channels such as wireline telephone channels, filters are used to ensure that the transmitted signals do not exceed specified bandwidth limitations and thus do not interfere with one another. Such channels are generally characterized mathematically as linear filter Channels with additive noise, as illustrated in Figure 1.9. Hence, if the channel input is the signal s(t) the channel output is the signal

$$r(t) = s(t) * h(t) + n(t)$$

$$\int_{-\infty}^{+\infty} h(\tau)s(t - \tau)\, d\tau + n(t) \qquad (1.4.2)$$

where $h(\tau)$ is the impulse response of the linear filter and denotes convolution.

1.4.3. The Linear Time-Variant Filter Channel

Physical channels such as underwater acoustic channels and ionospheric radio channels which result in time-variant multipath propagation of the transmitted signal may be characterized mathematically as time-variant linear filters. Such linear filters are characterized by a time-variant channel impulse response $h(\tau\,;\,t)$, where $h(\tau\,;\,t)$ is the response of the channel at time t due to an impulse applied at time $t - \tau$. Thus, τ represents the "age"(elapsed-time) variable. The linear time-variant filter channel with additive noise is illustrated Figure 1.10. For an input signal *s(t),* the channel output signal is

$$r(t) = s(t) * h(\tau\,;\,t) + n(t)$$

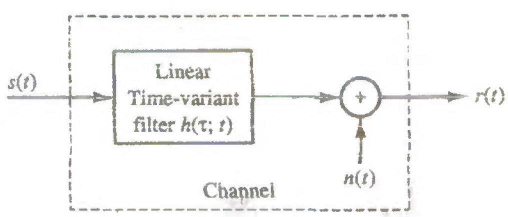


FIGURE 1.10. Linear time-variant filter channel with addaptive noise.

$$\int_{-\infty}^{+\infty} h(\tau\,;\,t)s(t - \tau)\, d\tau + n(t) \qquad (1.4.3)$$

A good model for multipath signal propagation through physical channels, such as the ionosphere (at frequencies below 30 MHz) and mobile cellular radio channels, is a special case of (1.4.3) in which the time-variant impulse response has the form

$$h(\tau\ ;\ t) = \sum_{k=1} a_k(t)\delta(t - \tau_k) \qquad (1.4.4)$$

where the $\{a_k(t)\}$ represents the possibly time-variant attenuation factors for the L multipath propagation paths. If (1.4.4) is substituted into (1.4.3), the received signal has the form

$$h(\tau\ ;\ t) = \sum_{k=1}^{1} a_k(t)s(t - \tau_k) + n(t) \qquad (1.4.5)$$

Hence, the received signal consists of L multipath components, where each component is attenuated by $\{a_k(t)\}$ and delayed by $\{\tau_k\}$.

The three mathematical models described above adequately characterize a large majority of physical channels encountered in practice. These three channel models are used in this text for the analysis and design of communication systems.

CHAPTER TWO

DIGITAL TRANSMISSION VIA CARRIER MODULATION

Digitally modulated signals with lowpass spectral characteristics can be transmitted directly through such channels without the need for frequency translation of the signal.

There are many communication channels, including telephone channels, radio channels, and satellite channels, that pass signals within a band of frequencies that is far removed from dc.Such channels are called bandpass channels, digital information may be transmitted through such channels by using a sinusoidal carrier that is modulated by the information sequence in either amplitude, phase, or frequency, or some combination of amplitude and phase. The effect of impressing the information signal on one or more of the sinusoidal parameters is to shift the frequency content of the transmitted signal to the appropriate frequency band that is passed by the channel. Thus, the signal is transmitted by carrier modulation.

In this chapter we treat several carrier-modulation methods, including amplitude modulation, phase modulation, combined amplitude and phase modulation. We will consider the spectral characteristics of these carrier-modulation signals. We begin our discussion with digital PAM, which is transmitted by modulating the amplitude of the carrier.

2.1 Carrier-Amplitude Modulation

In baseband digital PAM, the signal waveforms are of the form

$$S_m(t) = A_m g_T(t), \qquad m = 1, 2, 3, \ldots, M \qquad (2.1.1)$$

where $g_T(t)$ is the transmitting filter impulse response whose shape determines the spectral characteristics of the transmitted signal and A_m is signl amplitude that takes the discrete values

$$A_m = (2m - 1 - M), \qquad m = 1, 2, 3, \ldots, M \qquad (2.1.2)$$

The spectrum of the baseband signals is contained in the frequency band $|f| \leq W$ where W is the bandwidth of $|G_T(f)|^2$.

To transmit the digital signal waveforms through a bandpass channel by amplitude modulation, the baseband signal waveforms $s_m(t)$, m = 1,2,..., M an multiplied by a sinusoidal carrier of the form $\cos 2\pi f_c t$, as shown inFigure 2.1, where f_c is the carrier frequency ($f_c > w$) and corresponds to the center frequency in the passband of the channel. Thus, the transmitted signal waveformsmay be expressed as

$$U_m(t) = A_m g_T(t) \cos 2\pi f_c t \,, \qquad m = 1, 2, \ldots, M \qquad (2.1.3)$$

Amplitude modulation of the carrier $\cos 2\pi f_c t$ by the baseband signal waveforms $S_m(t) = A_m g_T(t)$ shifts the spectrum of the baseband signal by an amount f_c and, thus, places the signal into the passband of the channel. Recall that the Fourier transform of the carrier is $[\delta(f - f_c) + \delta(f + f_c)]/2$. Because multiplication of two signals in the time domain correspond to the convolution of their spectra in the frequency domain, the spectrum of the amplitude-modulated signal given by (2.1.3) is

$$U_m(f) = \frac{A_m}{2} [G_T(f - f_c) + G_T(f + f_c)] \qquad (2.1.4)$$

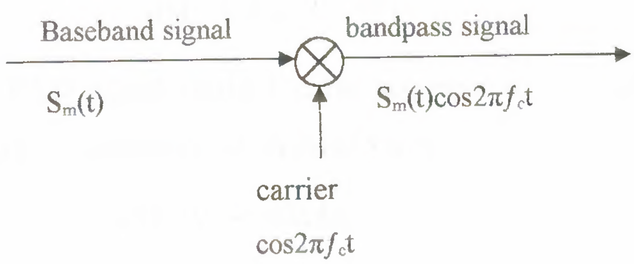
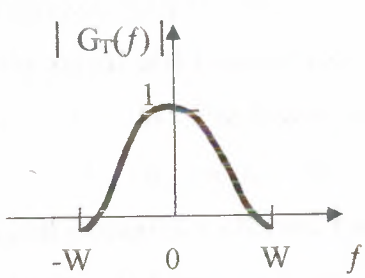
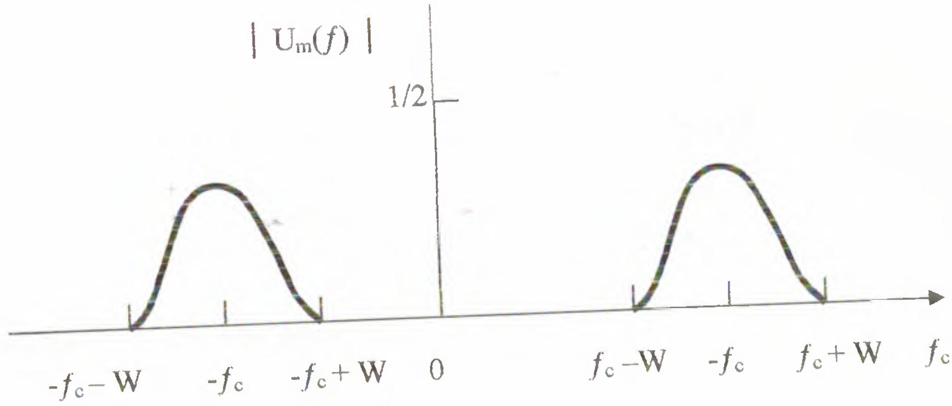


FIGURE 2.1. Amplitude modulation of a sinusoidal carrier by the baseband PAMsignal.



(a)



(b)

FIGURE 2.2. Spectra of (a) baseband and (b) amplitude-modulated signal.

Thus the spectrum of the baseband signal $s_m(t) = A_m g_T(t)$ is shifted in frequency by the carrier frequency f_c. The result is a DSB-AM signal, as illustrated in Figure 2.2.

The above illustration of the spectral characteristics of the amplitude-modulated signal carry over to the power-spectral density of a sequence of transmitted symbols. Thus, if

$$v(t) = \Sigma\, a_n g_T(t - nT) \tag{2.1.5}$$

is the baseband PAM signal and $\{a_n\}$ is the sequence of amplitude values, the power-spectral density of the amplitude-modulated signal

$$u(t) = v(t)\cos 2\pi f_c t \tag{2.1.6}$$

is

$$S_U(f) = ¼\, [Sv(f - f_c) + Sv(f + f_c)] \tag{2.1.7}$$

where $Sv(f)$ is the power-spectral density of the baseband signal, and $S_U(f)$is the power-spectral density of the bandpass signal u(t).

The modulated signal u(t)given by (2.1.6) is a DSB-SC amplitude-modulated signal. The upper sideband of the signal u(t) is comprised of the frequency content of u(t) for $|f| > f_c$ i.e., for $f_c < |f| \le f_c + W$. The lower sideband of u(t) is comprised of the frequency content for $|f| < f_c$, i.e., for $f_c - W \le$ for $|f| < f_c$ Hence, the DSB-SC amplitude-modulated signal occupies a channel bandwidth of 2W, which is twice the bandwidth required to transmit the baseband signal.

The energy of the passband signal waveforms $U_m(t)$, m = 1, 2, …, M, given by (2.1.3) is defined as

$$E_m = \int_{-\infty}^{\infty} u^2_m(t)dt = \int_{-\infty}^{\infty} A^2_m g^2_T(t)\cos^2 2\pi f_c t\, dt$$

$$= \frac{A^2m}{2}\int_{-\infty}^{\infty} g^2_T(t)\, dt + \frac{A^2m}{2}\int_{-\infty}^{\infty} g^2_T(t)\cos 4\pi f_c t\, dt \qquad (2.1.8)$$

We note that $f_c \gg w$, whenthe term

$$\int_{-\infty}^{\infty} g^2_T(t)\cos 4\pi f_c t\, dt \qquad (2.1.9)$$

involves the integration of the product of a slowly varying function, namely $g^2_T(t)$, with a rapidly varying sinusoidal term, namely $\cos 4\pi f_c t$ as shown in Figure 2.3. Because $g_T(t)$ is slowly varying relative to $\cos 4\pi f_c t$, the integral in (2.1.9) over a single cycle of $\cos 4\pi f_c t$ is zero, and, hence, the integral over an arbitrary number of cycles is also zero. Consequently,

$$E_m = \frac{A^2m}{2}\int_{-\infty}^{\infty} g^2_T(t)\, dt = \frac{A^2m}{2} E_g \qquad (2.1.10)$$

Where E_g is the energy in the signal pulse $g_T(t)$ Thus, we have shown that the energy in the passband signal is one-half of the energy in the baseband signal. The scale factor of ½ is due to the carrier component $\cos 2\pi f_c t$, which has an average power of ½.

When the transmitted pulse shape $g_T(t)$ is rectangular, i.e.,

$$g_T(t) = \begin{cases} \sqrt{\dfrac{E_g}{T}} & 0 \le t \le T \\ 0 & \text{Otherwise} \end{cases} \qquad (2.1.11)$$

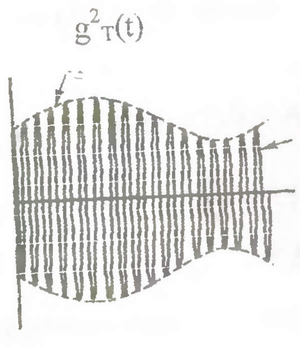


FIGURE 2.3. The signal $g^2_T(t)\cos 4\pi f_c t$

the amplitude-modulated carrier signal is usually called amplitude -shift keying (ASK).

Finally, we note that impressing the baseband signals $S_m(t)$ onto the amplitude of the carrier signal $\cos 2\pi f_c t$ does not change the basic geometric representation of the digital PAM signal waveforms. The signal waveforms $U_m(t)$ may be expressed as

$$U_m(t) = S_m \psi(t) \tag{2.1.12}$$

Where the basic signal waveform $\psi(t)$ is defined as

$$\psi(t) = \sqrt{\frac{2}{E_g}}\, g_T(t) \cos 2\pi f_c t \tag{2.1.13}$$

and

$$S_m = \sqrt{\frac{E_g}{T}}, \qquad A_m, \qquad m = 1, 2. . ., M \tag{2.1.14}$$

Note that the only change in the geometric representation, compared to baseband signals, is the scale factor $\sqrt{2}$, which appears (2.1.13) and (2.1.14).

2.1.1 Amplitude Demodulation And Detection

The demodulation of a bandpass digital PAM signal may be accomplished in one of several ways by means of correlation or matched filtering. For illustrative purposes, we consider a correlation-type demodulator. Suppose the transmitted signal is

$$u_m(t) = A_m g_T(t) \cos 2\pi f_c t, \qquad 0 \leq t \leq T \tag{2.1.15}$$

The received signal may be expressed as

$$r(t) = A_m g_T(t) \cos 2\pi f_c t + n(t), \qquad 0 \leq t \leq T \tag{2.1.16}$$

Where n(t)is a bandpass noise process, which is represented as

$$n(t) = n_c(t) \cos 2\pi f_c t - n_s(t) \cos 2\pi f_c t \tag{2.1.17}$$

By crosscorrelating the received signal r(t) with the basic function $\psi(t)$ as shown in Figure 2.4, we obtain the output

$$\int_{-\infty}^{\infty} r(t)\psi(t)dt = A_m \sqrt{\frac{2}{E_g}} \int_{-\infty}^{\infty} g^2_T(t)\cos^2\pi f_c t\, dt + \int_{-\infty}^{\infty} n(t)\psi(t)dt$$

$$= A_m\sqrt{Eg/2} + n \tag{2.1.18}$$

Where n represents the additive noise component at the output of the correlator.

2.1.2 Carrier Phase Recovery

In the above development we assumed that the function $\psi(t)$ is perfectly synchronized with the signal component in r(t) in both

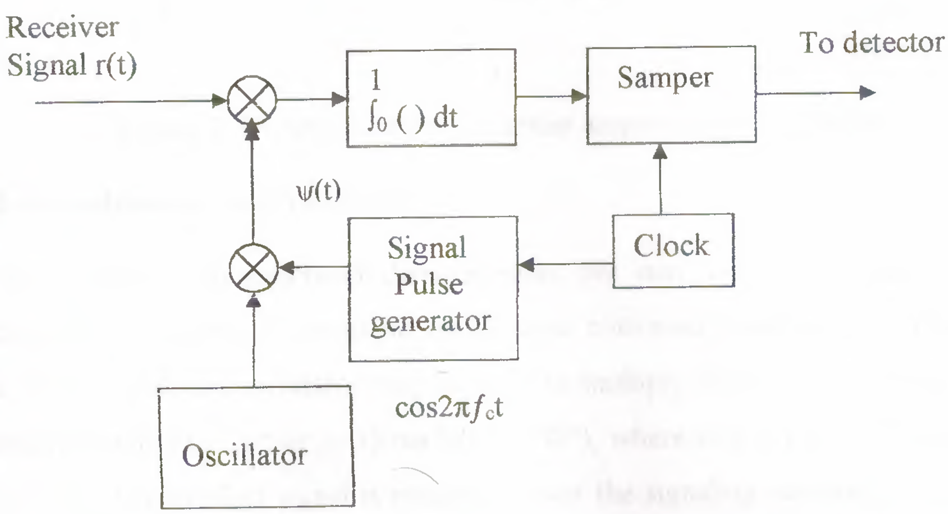


FIGURE.2.4. Demodulation of bandpass digital PAM signal.

time and carrier phase. In practice, however, these ideal conditions do not hold because of the propagation delay encountered in transmitting a signal through the channel and because of frequency and phase drifts that occur in any practical oscillator that generates the carrier signal $\cos 2\pi f_c t$.As a consequence, it is necessary to generate a phase coherent carrier at the receiver to perform the demodulation of the received signal.

Because the message signal, given by (2.1.5) is generally zero mean, the DSB-SC amplitude-modulated signal u(t) given by (2.1.6) has zero-average power at $f = f_c$. Consequently, it is not possible to estimate the carrier phase directly from r(t), However, if we square r(t), we generate a frequency component at $f = 2f_c$ that has nonzero-average power. This component can be filtered out by a narrowband filter tuned to $2f_c$, which can be used to drive a PLL. A functional block diagram of the receiver that employs a PLL for estimating the carrier phase is shown in Figure 2.5.

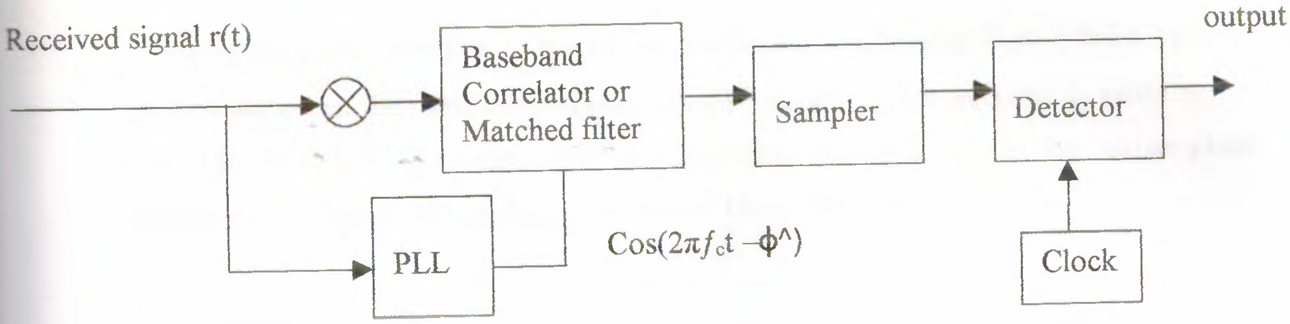
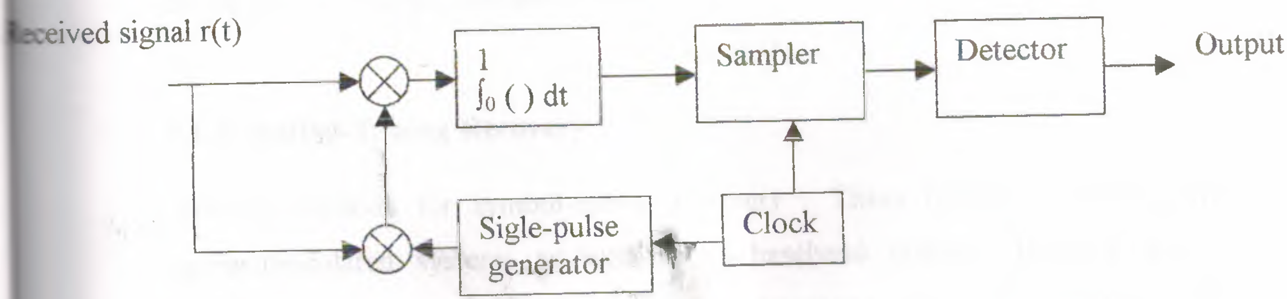


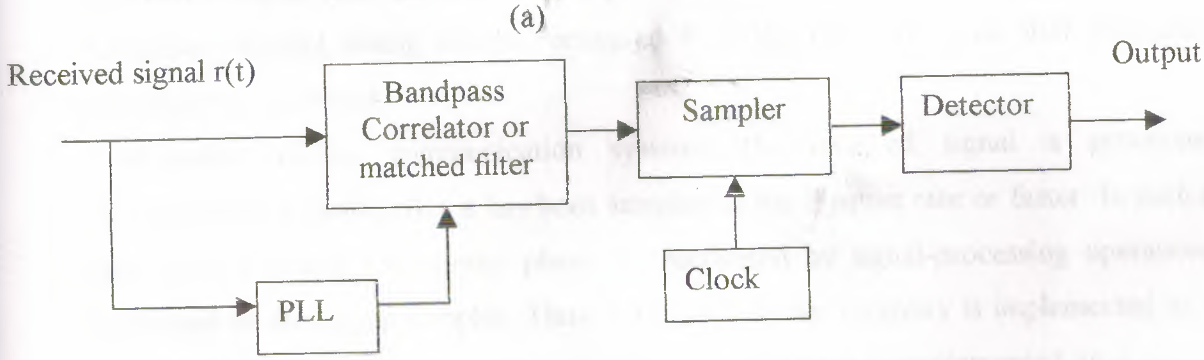
Figure 2.5. Demodulation of carrier amplitude-module signal

2.1.3 Demodulation And Detection

As an alternative to baseband demodulation, we may perform crosscorrelation or matched filtering either at passband or at some convenient intermediate frequency. In particular, a bandpass correlator may be used to multiply the received signal r(t) by the amplitude-modulated carrier $g_T(t)\cos(2\pi f_c t - \Phi\hat{})$, where $\cos(2\pi f_c t - \Phi\hat{})$ is the output of the PLL. The product signal is integrated over the signaling interval T, the output of the integrator is sampled at t = T and the sample is passed to the detector. If a matched filter instead of a correlator is used, the filter impulse response is $g_T(T - t)\cos[2\pi f_c(T - t) - \Phi\hat{}]$, he functional block diagram for these demodulators are shown in Figure 2.6.



(a)



(b)

FIGURE 2.6. Bandpass demodulation of digital PAM signal via (a) bandpass correlation and (b) bandpass matched filtering.

To describe the operation of the optimum detector for the case of an AWGN channel, let us assume that the transmitted signal is $u_m(t) = s_m\ \psi(t)$ where $\psi(t)$ and s_m given by (2.1.13) and (2.1.14), respectively. Furthermore, we assume that the carrier-phase estimate $\Phi\text{^} = \Phi$ where Φ is the carrier phase. Then, the input

to the detector is

$$r = \int_{-\infty}^{\infty} r(t)\psi(t)dt$$

$$= s_m \int_{-\infty}^{\infty} \psi^2(t)dt + \int_{-\infty}^{\infty} n(t)\psi(t)dt$$

$$= A_m\sqrt{}Eg/2 + n \qquad (2.1.19)$$

Where n(t) is a sample function of the bandpass noise process, which is given by (2.1.17) in terms of its quadrature components.

For equiprobable messages, the optimum detector bases its decision on the distance metrics

$$D(r, s_m) = (r - s_m)^2, \qquad m = 1, 2, \ldots, M$$

or, equivalent, on the correlation metrics

$$C(r, s_m) = 2rs_m - s^2_m$$

2.1.4 Symbol-Timing Recovery

Several methods for symbol-timing recovery . These methods generally apply to carrier-modulated systems as well as to baseband systems. Because any carrier-modulated signal can be converted to a baseband signal by a simple frequency translation, symbol timing can be recovered from the received signal after frequency conversion to baseband.

In many modem communication systems, the received signal is processed (demodulated) digitally after it has been sampled at the Nyquist rate or faster. In such a case, symbol timing and carrier phase are recovered by signal-processing operations performed on the signal samples. Thus, a PLL for carrier recovery is implemented as a digital PLL and a clock recovery loop of a type described is implemented as a digital loop. Timing recovery methods based on sampled signals have been described and analyzed by Mueller and Muller (1976).

2.1.5 Signal Demodulation In The Presence Of Channel Distortion

In most practical communication systems, the channel distorts the transmitted signal. The most common type of channel distortion encountered in practice is linear distortion (amplitude and phase distortion). This type of distortion results in intersymbol interference (ISI) that degrades the performance of the communication system.

The ISI may be reduced by proper design of the transmitting and receiving filters. for designing optimum transmitting and receiving filters for a channel with linear distortion may be applied directly by converting the bandpass channel into an equivalent baseband channel, To be specific, suppose that the bandpass channel is characterized by the frequency response $C_{bp}(f_c)$ and let f_c denote the center of the frequency band. Let $C(f)$ denote the frequency response of an equivalent baseband (lowpass) channel, where $C(f)$is defined as

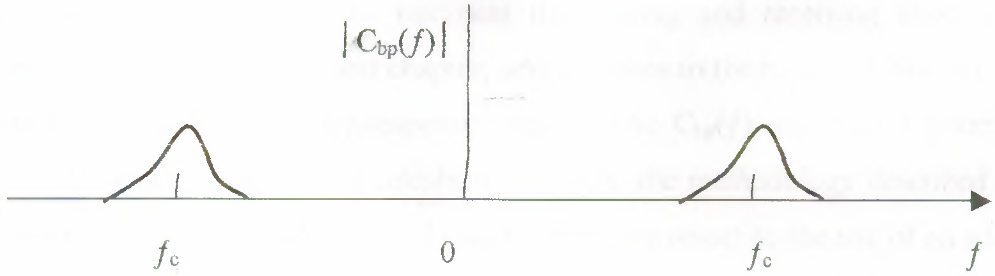
$$C(f - f_c) = \begin{cases} 2\,C_{bp}(f), & f > f_c \\ 0, & f < f_c \end{cases} \qquad (2.1.20)$$

Because $C^*_{bp}(-f) = C_{bp}(f)$, it follows that

$$C^*(-f - f_c) = \begin{cases} 0, & f > -f_c \\ 2C^*_{bp}(-f), & f < -f_c \end{cases} \qquad (2.1.21)$$

Therefore, the frequency response $C_{bp}(f)$ of the bandpass channel may be expressed as

$$C_{bp}(f) = ½\,[C(f - f_c) + C^*(-f - f_c)] \qquad (2.1.22)$$



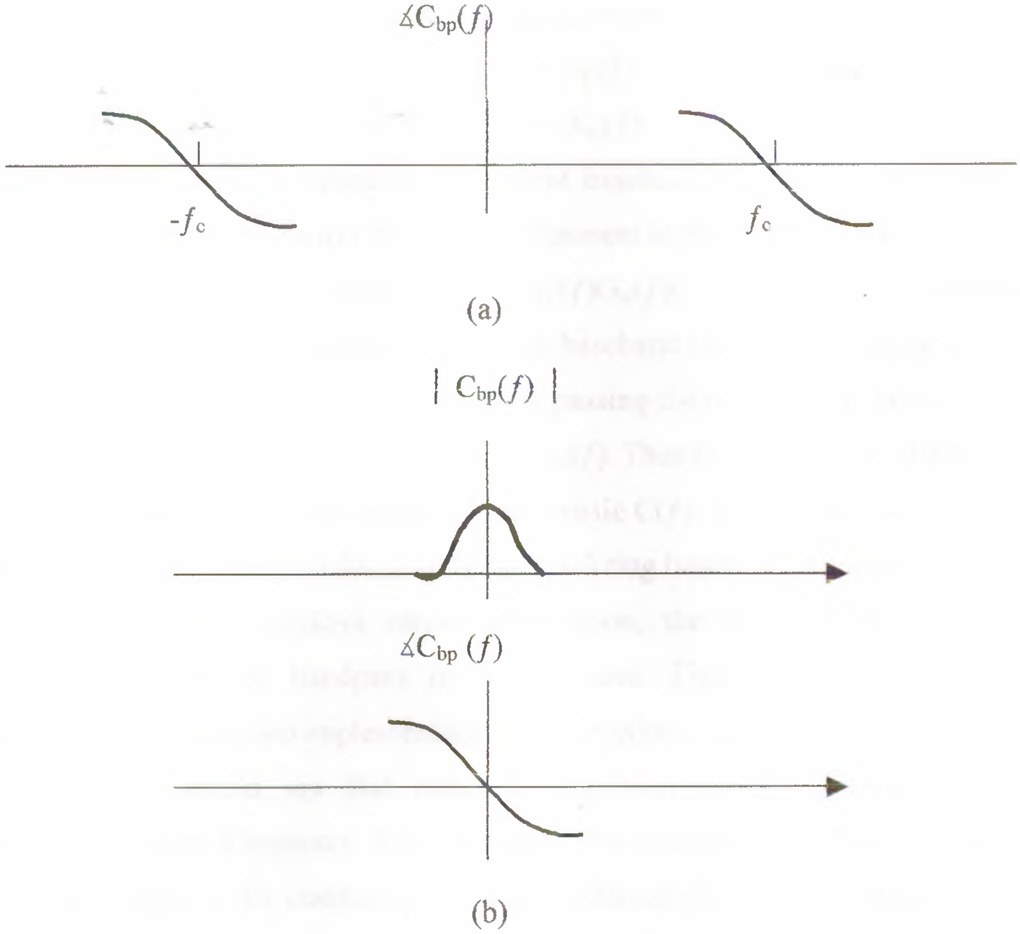


FIGURE 2.7. Magnitude and phase response for a (a) bandpass channel (b) its equivalent baseband channel.

Where $C(f)$ is the frequency response of the equivalent baseband channel. This relationship is illustrated in Figure 2.7. The corresponding relationship between the impulse responses of the bandpass and baseband channels is easily established from (2.1.23). If C_{bp} denotes the impulse response of the bandpass channel and c(t) denotes the impulse response of the baseband channel,then

$$C_{bp} = \mathrm{Re}[c(t)e^{j2\pi fct}] \qquad (2.1.23)$$

Once we have determined the frequency response $C(f)$ of the equivalent baseband channel as described above, the optimum transmitting and receiving filters can be designed as we will see in the next chapter, which applies to the baseband channel.

When the channel frequency-response characteristic $C_{bp}(f)$ unknown a priori, as is the case, for example, in dial-up telephone channels, the methodology described in the next chapter, is no longer applicable. In such a case, we resort to the use of an adaptive equalizer to compensate for the channel distortion. The basil approach in this case is to design the transmitting and receiving fiters$G_T(f)$ and $G_R(f)$either for zero ISI or for controlled ISI (partial-response signals), based on the assumption that the channel is ideal. For example, we might select $G_T(f)$ and $G_R(f)$ to satisfy the condition

$$G_T(f)\ G_R(f) = X_{rc}(f) \tag{2.1.24}$$

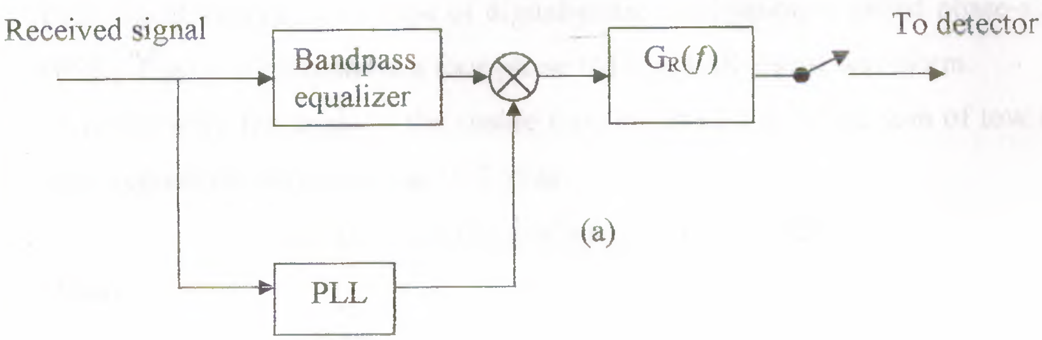
Which results in zero ISI when the equivalent baseband channel is ideal. Then, if we transmit an isolated pulse $g_T(t)$ the signal component at the output of the receiving filter has the frequency-response characteristic $G_T(f)G_R(f)C(f)$ where $C(f)$ represents the frequency response of the nonideal quivalent baseband channel. The IS resulting from the nonideal channel may be compensated by passing the receive signal through a linear adaptive equalizer, with frequency response $G_E(f)$. Thus $G_E(f)$ serves as the system that compensates for the nonideal channel characteristic $C(f)$. This is basically the approach described in the next chapter for adaptively equalizing baseband channels.

In a system that employs carrier modulation, the adaptive equalizer may be implemented either at bandpass or at baseband. Figure 2.8 illustrates the two configurations. These two implementations are functionally equivalent.

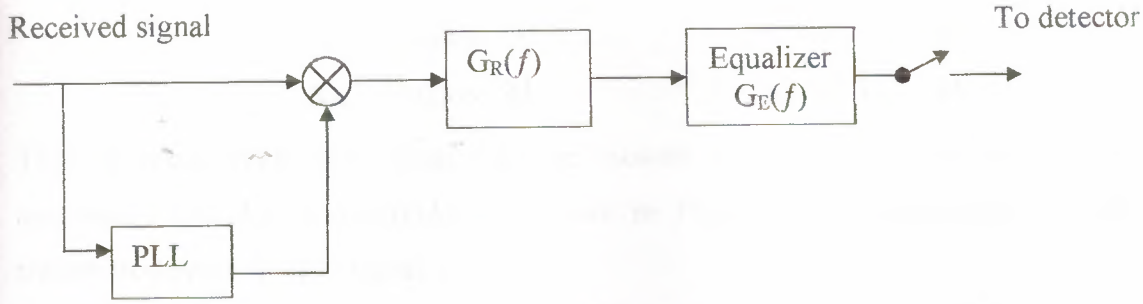
Finally, we should say that either a decision-feedback equalizer (DFE) or a maximum-likelihood sequence detector (viterbi algorithm) may be used as a alternative to a linear equalizer for combating ISI due to channel distortion. These two nonlinear equalizers yield a significantly better performance compared to linear equlizer on *channels with severe distortion.*

2.2 Carrier Phase Modulation

In carrier-phase modulation the information that is transmitted over a communication channel is impressed on the phase of the carrier. Because the range of thecarrier phase is $0 \leq \theta \leq 2\pi$, the range of the carrier phases used to transmit digital information via digital-phase modulation are $\theta_k = 2\pi k / M$, $k = 0, 1, \ldots, M-1$. Thus, for binary phase modulation ($M = 2$) the two carrier phases are $\theta = 0$ and $\theta_1 = \pi$ radians. For M-ary-phase modulation $= 2^k$ where k is the number of information bits per transmitted symbol.



(a)



(b)

FIGURE 2.8. Channel equalization at (a) bandpass (b) baseband.

The general representation of a set of Carrier-phase-modulated signal waveforms is

$$u_m(t) = g_T(t)\cos(2\pi f_c + 2\pi m/M), \qquad m = 0, 1, \ldots, M-1 \qquad (2.2.1)$$

where $g_T(t)$ is the transmitting filter pulse shape, which determines the spectral characteristics of the transmitted signal. These signals have identical energy, i.e.,

$$E_m = \int_{-\infty}^{\infty} u^2_m(t)dt = \int_{-\infty}^{\infty} g^2_T(t)\cos(2\pi f_c + 2\pi m/M)\,dt$$

$$= \tfrac{1}{2}\int_{-\infty}^{\infty} g^2_T(t)dt + \tfrac{1}{2}\int_{-\infty}^{\infty} g^2_T(t)\cos(4\pi f_c + 4\pi m/M)\,dt$$

$$= \tfrac{1}{2} E_g \equiv E_s, \text{ For all } M \qquad (2.2.2)$$

where E_g is the energy of the pulse $g_T(t)$ and E_s denotes the energy per transmitted symbol. The term involving the double-frequency component in (2.2.2) averages out to zero when $f_c \gg W$ where W is the bandwidth of $g_T(t)$. Hence, all waveforms possess the same energy. Note that when $g_T(t)$ is a rectangular pulse, it is defined as

$$g_T(t) = \sqrt{\frac{2E_s}{2}}, \qquad 0 \le t \le T \qquad (2.2.3)$$

When a rectangular pulse is employed, the transmitted signal waveforms

$$u_m(t) = \sqrt{\frac{2E_s}{T}}\cos(2\pi f_c + 2\pi m/M), \qquad m = 0, 1, \ldots, M-1 \quad (2.2.4)$$

have a constant envelope and the carrier phase changes abruptly at the beginning of each signal interval. This type of digital-phase modulation is called phase-shift keying (PSK) Figure 2.9 illustrates a four-phase (M = 4) PSK signal waveform.

By viewing the angle of the cosine function in (2.2.4) as the sum of tow angles, we may express the waveforms in (2.2.1) as

$$u_m(t) = g_T(t)A_{mc}\cos 2\pi f_c t - g_T(t)A_{ms}\sin 2\pi f_c t \qquad (2.2.5)$$

Where

$$A_{mc} = \cos 2\pi m / M, \qquad m = 0, 1, \ldots, M-1$$
$$A_{ms} = \sin 2\pi m / M, \qquad m = 0, 1, \ldots, M-1 \qquad (2.2.6)$$

Thus, a phase-modulated signal may be viewed as two quadrature carriers with amplitudes $g_T(t)A_{mc}$ and $g_T(t)A_{ms}$ as shown in Figure 2.10, which depend on the transmitted phase in each signal interval.

It follows from (2.2.5) that digital-phase-modulated signals can be represents geometrically as two-dimensional vectors with components $\sqrt{E_s}\cos 2\pi m / M$ and $\sqrt{E_s}\sin 2\pi m / M$, i.e.,

$$s_m = (\sqrt{E_s}\cos 2\pi m / M,\ \sqrt{E_s}\sin 2\pi m / M) \qquad (2.2.7)$$

Note that the orthogonal basis functions are $\psi_1(t) = \sqrt{2 / E_g}\, g_T(t)\cos 2\pi f_c t$ and $\psi_2(t) = -\sqrt{2 / E_g}\, g_T(t)\sin 2\pi f_c t$. Signal point constellations for M = 2,4,8 are illustrated in Figure 2.11. We observe that binary-phase modulation is identical to binary PAM.

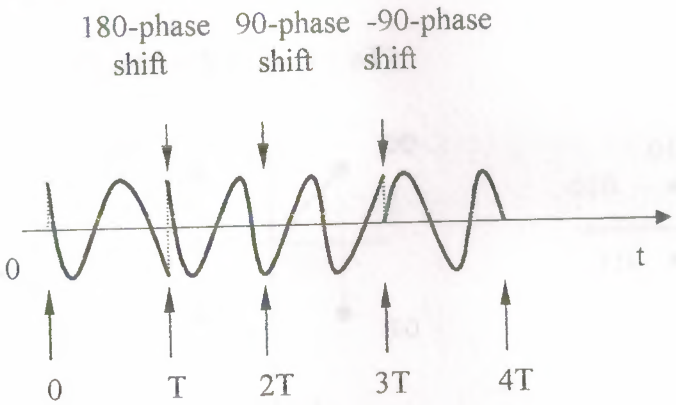


FIGURE 2.9. Example of a four-phase PSK signal.

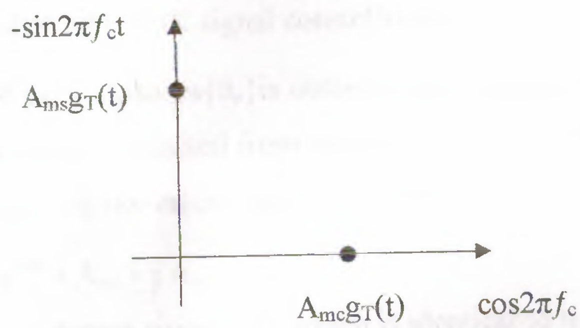


FIGURE 2.10. Digital-phase modulation viewed two amplitude modulated quadrature carriers.

The mapping or assignment of k information bits into the $M = 2^k$ possible phases may be done in a number of ways. The preferred assignment is to use Gray encoding, in which adjacent phases differ by one binary digit as illustrated in Figure 2.11. Because the most likely errors caused by noise involve the erroneous selection of an adjacent phase to the transmitted phase, only a single bit error occurs in the k-bit sequence with Gray encoding.

The spectral characteristics of a digital-phase-modulated signal are similar to the spectral characteristics of a PAM signal. From (2.2.1) we observe that the bandpass signal waveforms $u_m(t)$ in the signaling interval $0 \le t \le T$ may be expressed as

$$u_m(t) = \mathrm{Re}[g_T(t)e^{j2\pi m/M}e^{j2\pi f_c t}], \qquad m = 0, 1, \ldots, M-1 \qquad (2.2.8)$$

The corresponding transmitted bandpass signal for a sequence of information symbols may then be expressed as

$$u(t) = \mathrm{Re}[v(t)\, e^{j2\pi f_c t}] \qquad (2.2.9)$$

Where, by definition,

$$v(t) = \Sigma\, e^{j\theta_n} g_T(t - nT) \qquad (2.2.10)$$

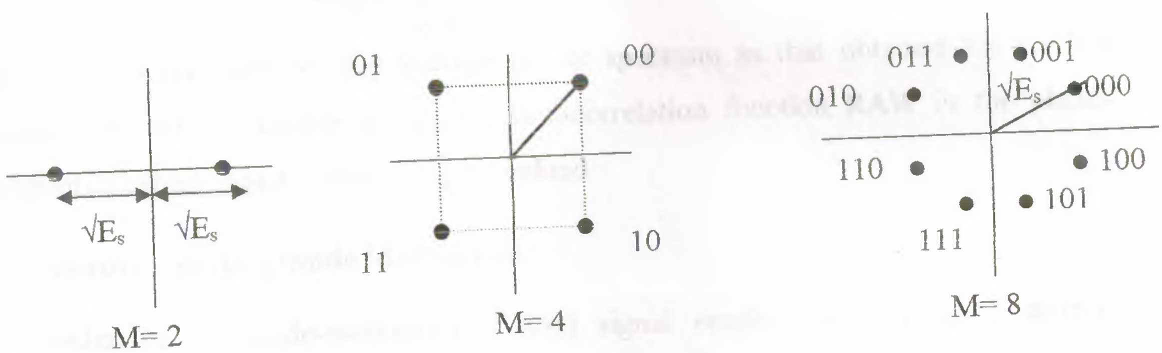


FIGURE 2.11. PSK signal constellation.

The transmitted sequence of carrier phases$\{\theta_n\}$is obtained by mapping the sequence of k-bit symbols into the carrier phases selected from the set of M phases $\{2\pi m / M,\ m = 0, 1, \ldots, M-1\}$. By defining a complex-valued amplitude A_n as

$$A_n = e^{j\theta_n} = A_{nc} + j\, A_{ns} \qquad (2.2.11)$$

We obtain an expression for the lowpass signal v(t), which is identical to the expression for a PAM signal, except that now the sequence$\{A_n\}$ is complex-valued. This difference requires that we define the autocorrelation function for the sequence $\{A_n\}$ as

$$R_A(n) = E\,[A^*_k A_{n+k}]$$

$$= E\,[e^{-j\theta k}\, e^{j\theta n+k}] = E\,[e^{j(\theta n+k - \theta k)}] \qquad (2.2.12)$$

Aside from this minor difference, the derivation for the power-density spectrum of the transmitted signal u(t) given by (2.2.9) follows by its procedure Thus, we obtain the average autocorrelation function of the transmitted signl as

$$R_U(\tau) = \frac{1}{2}\sum_{n=-\infty}^{\infty} R_A(m)R_g(\tau - mT)\cos 2\pi f_c\tau \qquad (2.2.13)$$

Where $R_g(\tau)$ is the time autocorrelation function of the pulse $g_T(f)$. The Fourier transform of $R_U(\tau)$ yields the power-spectral density as

$$S_U(f) = \frac{1}{4}\left[S_v(f - f_c) + S_v(-f - f_c)\right] \qquad (9.2.14)$$

Where the power spectrum of the lowpass signal $\upsilon(t)$ is

$$S_v(f) = 1 / T\ S_A(f)|G_T(f)|^2 \qquad (9.2.15)$$

And

$$S_A(f) = \sum_{n=-\infty}^{\infty} R_A(n)e^{-j2\pi fnT} \qquad (2.2.16)$$

This is the same form for the average power spectrum as that obtained for a PAM signal. The only difference is that the autocorrelation function RAW in the phase-modulated carrier signal is now complex-valued.

2.3. Quadrature Amplitude Modulation

A quadrature amplitude-modulated (QAM) signal employ two qudrature carriers $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$, each of which is modulated by an independent sequence of information bits. We may view this method of signal transmission as a form of quadrature-carrier multiplexing. The transmitted waveforms have the form

$$u_m(t) = A_{mc}\, g_T(t)\cos 2\pi f_c t + A_{ms}\, g_T(t)\sin 2\pi f_c t, \quad m = 1, 2, \ldots, M \qquad (2.3.1)$$

Where $\{A_{mc}\}$ and $\{A_{ms}\}$are the sets of amplitude levels that are obtained by mapping k-bit sequences into signal amplitudes. For example, Figure 2.20 illustrates a 16-QAM signal constellation that is obtained by amplitude modulating each quadrature carrier by M = 4 PAM In general, rectangular signal constellations result when two quadrature carriers are each modulated by PAM.

More generally, QAM may be viewed as a form of combined digital amplitude and digital-phase modulation. Thus, the transmitted QAM signal waveforms may be expressed as

$$u_{mn}(t) = A_m g_T(t)\cos(2\pi f_c t + \theta_n), \qquad m = 1, 2, \ldots, M_1$$
$$n = 1, 2, \ldots, M_2 \qquad (2.3.2)$$

If $M_1 = 2^{k1}$, and $M_2 = 2^{k2}$ the combined amplitude and phase-modulation method results in the simultaneous transmission of $k_1 + k_2 = \log_2 M_1 M_2$ binary digits occurring at a symbol rate $R_b / (k_1 + k_2)$ Figure 2.21 illustrates the functional block diagram of a QAM modulator.

It is clear that the geometric signal representation of the signals given by (2.3.1) and (2.3.2) is in terms of two-dimensional signal vectors of the form

$$S_m = (\sqrt{E_s}A_{mc}, \sqrt{E_s}A_{ms}) \qquad m = 1, 2, \ldots, M \tag{2.3.3}$$

Examples of signal space constellations for QAM are shown in Figure 2.22.

The power-spectral density of a QAM signal is identical in form to that for a digital-phase-modulation signal given by (2.2.14)-(2..2.16). The transmitted bandpass signal u(t) for a sequence of information symbols may be expressed as in (2.2.9), where the baseband information-bearing signal is

$$\upsilon(t) = \sum_{n=-\infty}^{\infty} A_n e^{j\theta n} g_T(t - nT) \tag{2.3.4}$$

In (2.3.4) A_n is the signal amplitude and θ_n is the signal phase of the transmitted carrier. Hence, the autocorrelation of the transmitted sequence of amplitudes and

Phases is

$$\begin{aligned} R_A(n) &= E\,[A_k\, e^{-j\theta k}\, A_{n+k}\, e^{j\theta n+k}] \\ &= E\ \ [A_k\, A_{n+k} e^{j(\theta k+n - \theta k)}\,] \end{aligned} \tag{2.3.5}$$

In the special case in which the symbols in the transmitted sequence are uncorrelated and zero mean, the autocorrelation in (2.3.5) reduces to

$$R_A(n) = \begin{cases} \sigma^2_{A,} & n = 0 \\ 0\ , & n \neq 0 \end{cases} \tag{2.3.6}$$

Where $\sigma^2_A = E\,[A^2_k]$ In this case the power-spectral density of the input sequence white and, hence, the average power-spectral density of the transmitted signal spends solely on the spectral characteristics of the transmitted pulse, i.e., on

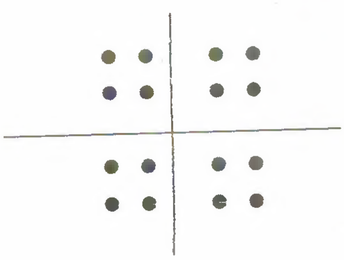


FIGURE 2.12. M = 16-QAM signal constellation.

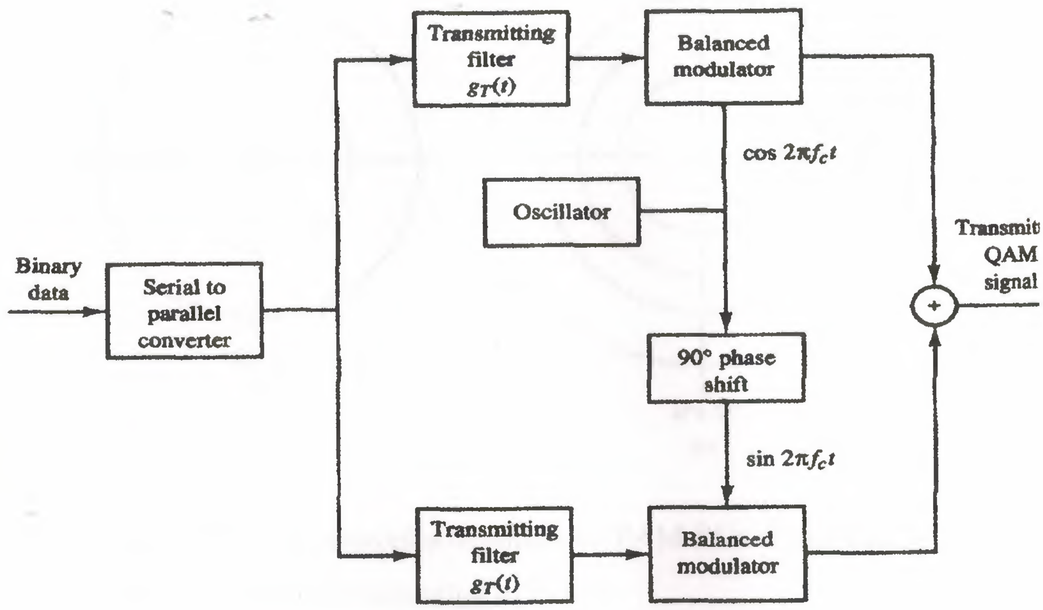


FIGURE 2.13 Functional block diagram of modulator for QAM.

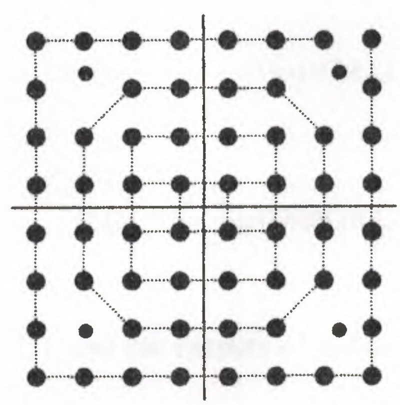


FIGURE 2.14. Rectangular signal-pace constellations for QAM.
M=64, M=32, M=16, M=8, M=4

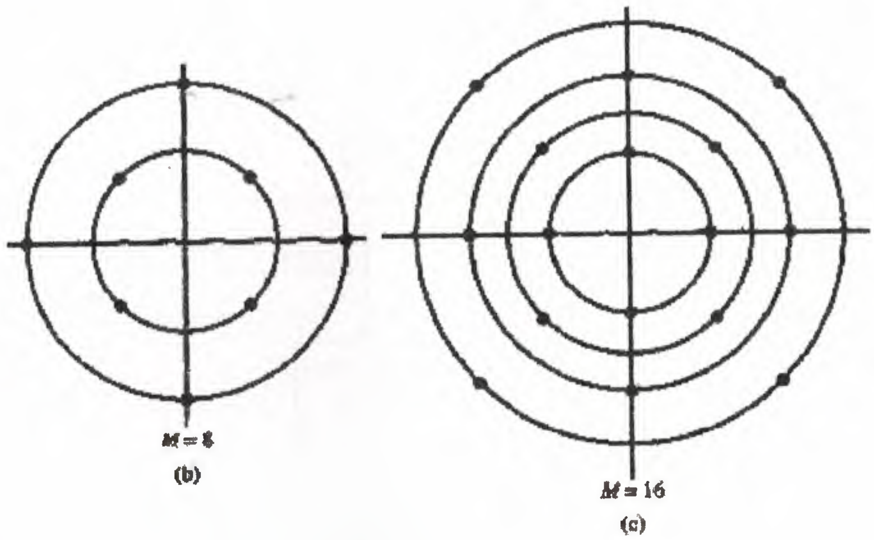


FIGURE 2.15. Examples of combined PAM-PSK signal-diagrams.

2.3.1. Demodulation and Detection of QAM

let us assume that a carrier-phase offset is introduced in the transmission of the signal through the channel. In addition, the received signal is corrupted by additive gaussian noise. Hence, r(t) may be expressed as

$$r(t) = A_{mc}g_T(t)\cos(2\pi f_c t + \emptyset) + A_{ms}g_T(t)\sin(2\pi f_c t + \emptyset) + n(t) \quad (2.3.7)$$

suppose that an estimate $\acute{\emptyset}$ of the carrier phase is available at the demodulator. Then, the received signal may be correlated with the two basis functions

$$\psi_1(t) = \sqrt{\frac{2}{E_g}}\, g_T(t)\cos(2\pi f_c t + \acute{\emptyset})$$

$$\psi_2(t) = \sqrt{\frac{2}{E_g}}\, g_T(t)\sin(2\pi f_c t + \acute{\emptyset}) \quad (2.3.8)$$

as illustrated in Figure (2.23), and the outputs of the correlators are sampled and passed to the detector.

The input to the detector consists of the two sampled component r_1, r_2, where ($E_s \equiv E_g / 2$)

$$r_1 = A_{mc}\sqrt{E_s}\cos(\emptyset - \acute{\emptyset}) + A_{ms}\sqrt{E_s}\sin(\emptyset - \acute{\emptyset}) + n_c\sin\acute{\emptyset} - n_s\cos\acute{\emptyset}$$

$$r_2 = A_{mc}\sqrt{E_s}\sin(\emptyset - \acute{\emptyset}) + A_{ms}\sqrt{E_s}\cos(\emptyset - \acute{\emptyset}) + n_c\sin\acute{\emptyset} - n_s\cos\acute{\emptyset} \quad (2.3.9)$$

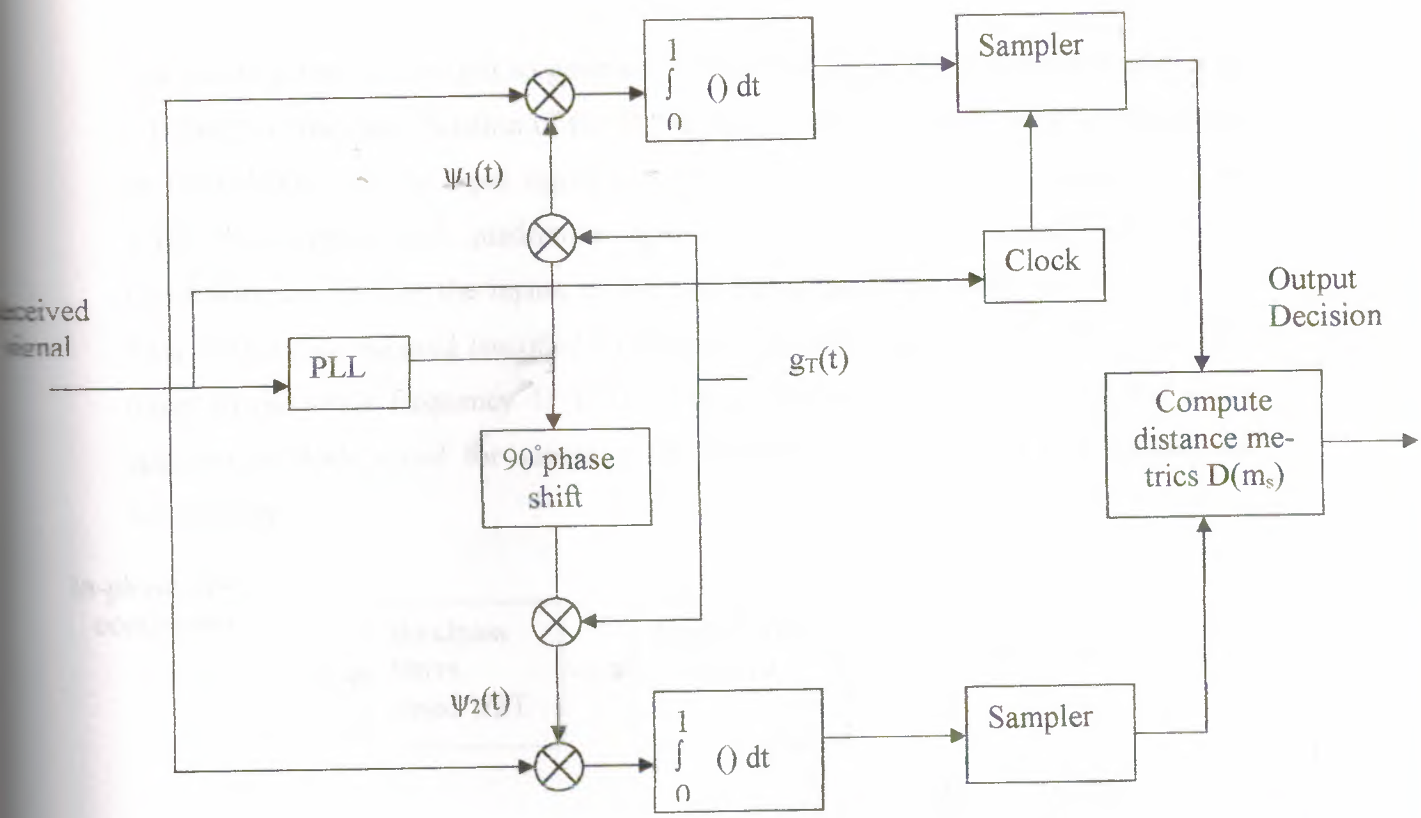


FIGURE 2.16. Demodulation and detection of QAM signal.

We observe that the effect of an imperfect phase estimate is twofold. First, the desired signal components in r_1 and r_2 are reduced in amplitude by the factor cos(Ø – Ǿ). In turn, this reduces the SNR by the factor $\cos^2$ (Ø – Ǿ). Second, there is a leakage of the quadrature signal components into the desired signal. This signal leakage, which is scaled by sin(Ø – Ǿ), causes a significant performance degradation unless Ø – Ǿ is very small. This point serves to emphasize the importance of having an accurate carrier-phase estimate in order to demodulate the QAM signal. The optimum detector computes the distance metrics

$$D(r, s_m) = |r - s_m|^2, \qquad m = 1, 2, \ldots, M \qquad (2.3.10)$$

And selects the signal corresponding to the largest value of $D(r, s_m)$. If a correlation metric is used in place of a distance metric, it is important to recognize that correlation metrics must employ bias correction because the QAM signals are not equal energy signals.

Symbol timing may be extracted from the received signal by using one of the methods for symbol synchronization. For QAM signals, the spectral line methods have proved to be particularly suitable for timing recovery. Figure 2.24 illustrates a spectral line method which is based on filtering out a signal component at the frequency 1 / 2T and squaring the filter output to generate a sinusoidal signal at the desired symbol rate1 / T. Because the demodulation of the QAM signal is accomplished, as described above, by multiplication of the input signal with the two quadrature-carrier signals $\psi_1(t)$ and $\psi_2(t)$, the in-phase and quadrature signal components at the outputs of the two correlators are used as the inputs to the two bandpass filters tuned to 1/2T. The two filter outputs are squared (rectified), summed, and then filtered by a narrowband filter tuned to the clock frequency 1/ T Thus, we generate a sinusoidal signal that is the appropriate clock signal for sampling the outputs of the correlators to recover the information.

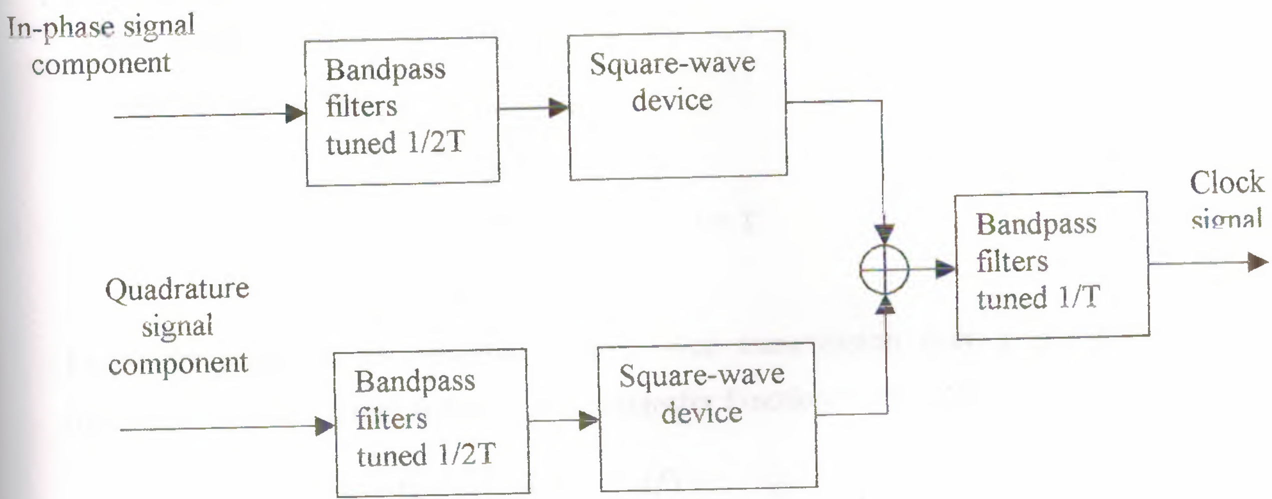


FIGURE 2.17. Block diagram of timing recovery method for QAM.

Finally, we should consider the problem of linear channel distortion that maybe encountered in a practical communication system employing QAM modulation If the channel frequency response characteristic $C_{bp}(f)$ (or the equivalent baseband frequency response) is known, we may use the methodogy described Section 2.1.3 to design optimum transmitting and receiving filters. On the other hand, if the channel frequency response is not known a priori, an adaptive equalizer may be used to compensate for the channel distortion. The equalizer may be implemented either as a bandpass equalizer or as a baseband equalizer.

CHAPTER THREE

SYSTEM DESIGN IN THE PRESENCE OF CHANNEL DISTORTION AND CHANNEL EQUALIZATION

3.1 System Design In The Presence Of Channel Distortion

We described a signal design criterion that results in zero ISI at the output of the receiving filter. Recall that a signal pulse x(t) will satisfy the condition of zero ISI at the sampling instants $t = nT$, $n = \pm1, \pm2, \ldots$, if its spectrum $X(f)$ satisfies the condition given by

$$\sum_{n=-\infty}^{\infty} x\left(f + \frac{m}{T}\right) = T$$

From this condition we conclude that ISI-free transmission over a channel, the transmitter-receiver filters and the channel transfer function must satisfy

$$G_T(f)C(f)G_R(f) = X_{rc}(f) \tag{3.1.1}$$

Where $X_{rc}(f)$ denotes the Fourier transform of an appropriate raised cosine pulse whose parameters depend on the channel bandwidth W and the transmission interval T. Obviously there are an infinite number of transmitter-receiver filter pairs that satisfy the above condition.

In this section we consider the design of the transmitter-receiver filter pairs that maximize the SNR at the output of the receiving filter. Therefore, in this section we are concerned with the design of a digital communication system with zero ISI and minimum error probability for a channel with distortion. We first present a brief coverage of various types of channel distortion and then derive the equations for transmitter-receiver filters.

We distinguish two types of distortion. Amplitude distortion results when the amplitude characteristic $|C(f)|$ is not constant for $|f| \leq W$ The second type of distortion, called phase distortion, results when the phase characteristic $\Theta_c(f)$ is nonlinear in frequency.

Another view of phase distortion is obtained by considering the derivative of $\Theta_c(f)$ Thus, we define the envelope delay characteristic as

$$\tau(t) = -\frac{1}{2\pi}\frac{d\Theta_c(f)}{df} \tag{3.1.2}$$

When $\Theta_c(f)$ is linear in f the envelope delay is constant for all frequencies. In this case, all frequencies in the transmitted signal pass through the channel with the same fixed-time delay. In such a case, there is no phase distortion. However, when $\Theta_c(f)$ is nonlinear, the envelope delay $\tau(t)$ varies with frequency, and the various frequency components in the input signal undergo different delays in passing through the channel. In such a case we say that the transmitted signal has suffered from delay distortion. Both amplitude and delay distortion cause ISI in the received signal. For example, let us assume that we have designed a pulse with a raised cosine spectrum that has zero ISI at the sampling instants. An example of such a pulse is illustrated in Figure below (a). When the pulse is passed through a channel filter with constant amplitude $|C(f)| = 1$ for $|f| < W$ and a quadratic phase characteristic (linear envelope delay), the received pulse at the output of the channel is shown in Figure below (b). Note that the periodic zero crossings have been shifted by the delay distortion so that the resulting pulse suffers from ISI. Consequently, a sequence of successive pulses would be smeared into one another and the peaks of the pulses would no longer be distinguishable due to the ISI.

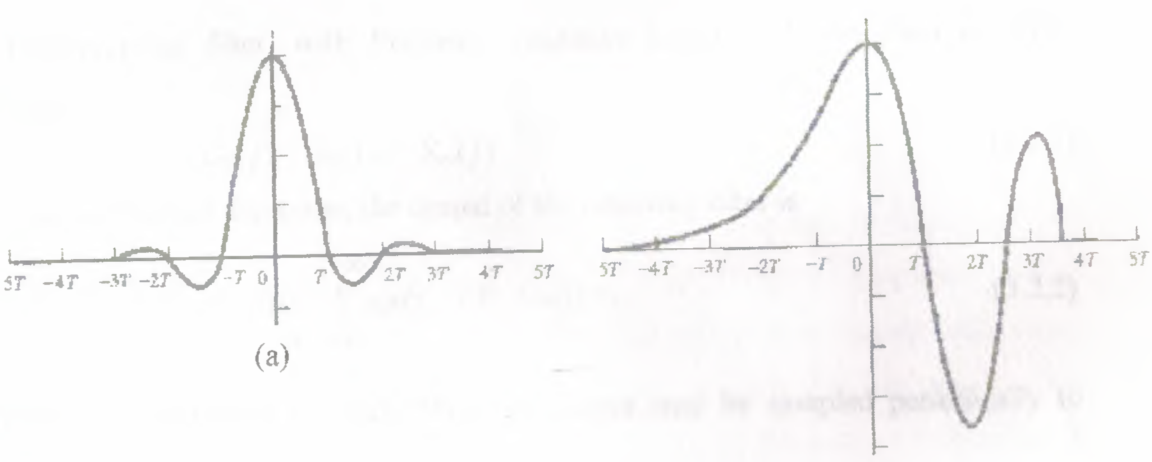


FIGURE 3.1. Effect of channel distortion (a) channel input (b) channel output.

3.2 Channel Equalization

In the preceding section we described the design of transmitting and receiving filters for digital PAM transmission when the frequency response characteristics of the channel are known. Our objectives were to design these filters for zero ISI and maximum SNR at the sampling instants. This design methodology is appropriate when the channel is precisely known and its characteristics do not change with time.

In practice we often encounter channels whose frequency response characteristics are either unknown or change with time. For example, in data transmission over the dial-up telephone network, the communication channel will be different every time we dial a number because the channel route will be different. Once a connection is made, however, the channel will be time-invariant for a relatively long period of time. This is an example of a channel whose characteristics are unknown a priori. Examples of time-varying channels are radio channels, such as ionospheric propagation channels. These channels are characterized by time-varying frequency response characteristics. These types of channels are examples where the optimization of the transmitting and receiving filters, as described in the preceding section, is not possible.

Under these circumstances, we may design the transmitting filter to have a square-root raised cosine frequency response, i.e.,

$$G_T(t) = \begin{cases} \sqrt{X_{rc}(f)}\, e^{-j2\pi f t0}, & |f| \leq W \\ 0, & |f| > W \end{cases}$$

and the receiving filter, with frequency response $G_R(f)$ to be matched to $G_T(f)$ Therefore,

$$|G_T(f)|\,|G_R(f)| = X_{rc}(f) \qquad (3.2.1)$$

Then, due to channel distortion, the output of the receiving filter is

$$Y(t) = \sum_{n=-\infty}^{+\infty} a_n x(t - nT) + \upsilon(t) \qquad (3.2.2)$$

where $x(t) = g_T(t) * c(t) * G_R(t)$. The filter output may be sampled periodically to product the sequence

$$y_m = \sum_{n=-\infty}^{+\infty} a_n x_{m-n} + \upsilon_m$$

$$= x_0a_m + \sum_{n\neq m} a_nx_{m-n} + \upsilon_m \qquad (3.2.3)$$

where $x_n = x(nT)$, $n = 0, \pm1, \pm2, \ldots$ The middle term on the right-hand side of (3.2.3) represents the ISI.

In any practical system it is reasonable to assume that the ISI affects a finite number of symbols. Hence, we may assume that $x_n = 0$ for $n < L_1$ and $n>L_2$ where L_1 and L_2 are finite, positive integers. Consequently, the ISI observed at the output of the receiving filter may be viewed as being generated by passing the data sequence $\{a_m\}$ through an FIR filter with coefficients $\{x_n, -L_1 \leq n \leq L_2\}$, as shown in Figure 3.2. This filter is called the equivalent discrete-time channel filter. Since its input is the discrete information sequence (binary or M-ary), the output of the discrete-time channel filter may be characterized as the output of a finite-state machine corrupted by additive Gaussian noise. Hence, the noise-free output of the filter is described by a trellis having M^L states where $L = L_1 + L_2$.

3.2.1 Maximum Likelihood Sequence Detection

The optimum detector for the information sequence $\{a_m\}$based on the observation of the received sequence $\{y_m\}$ given by (3.2.3), is a ML sequence detector. The detector is akin to the ML sequence detector described in the context of detecting partial response signals that that have controlled ISI. The Viterbi algorithm provides a method for searching through the trellis for the ML signal path. To accomplish this search, the equivalent channel filter coefficients $\{x_n\}$ must be known or measured by some method. At each stage of the trellis search, there are M^L surviving sequences with M^L corresponding Euclidean distance path metrics.

Due to the exponential increase in the computational complexity of the Viterbi algorithm with the span (length L) of the ISI, this type of detection is practical only when M and L are small. For example, in mobile cellular telephone systems, which employ digital transmission of speech signl, M is usually selected to be small, e.g., M = 4 and L 2. In this case, the ML sequence detector may be implemented with reasonable complexity. However, when *M* and *L* are large, the ML sequence detector becomes impractical. In such a case other more practical but supoptimum methods are used to detect the information sequence $\{a_m\}$ in the presence of ISI. Nevertheless, the performance of the ML sequence detector for a channel with ISI serves as a benchmark for comparing its performance with that of suboptimum methods. Two suboptimum methods are described below.

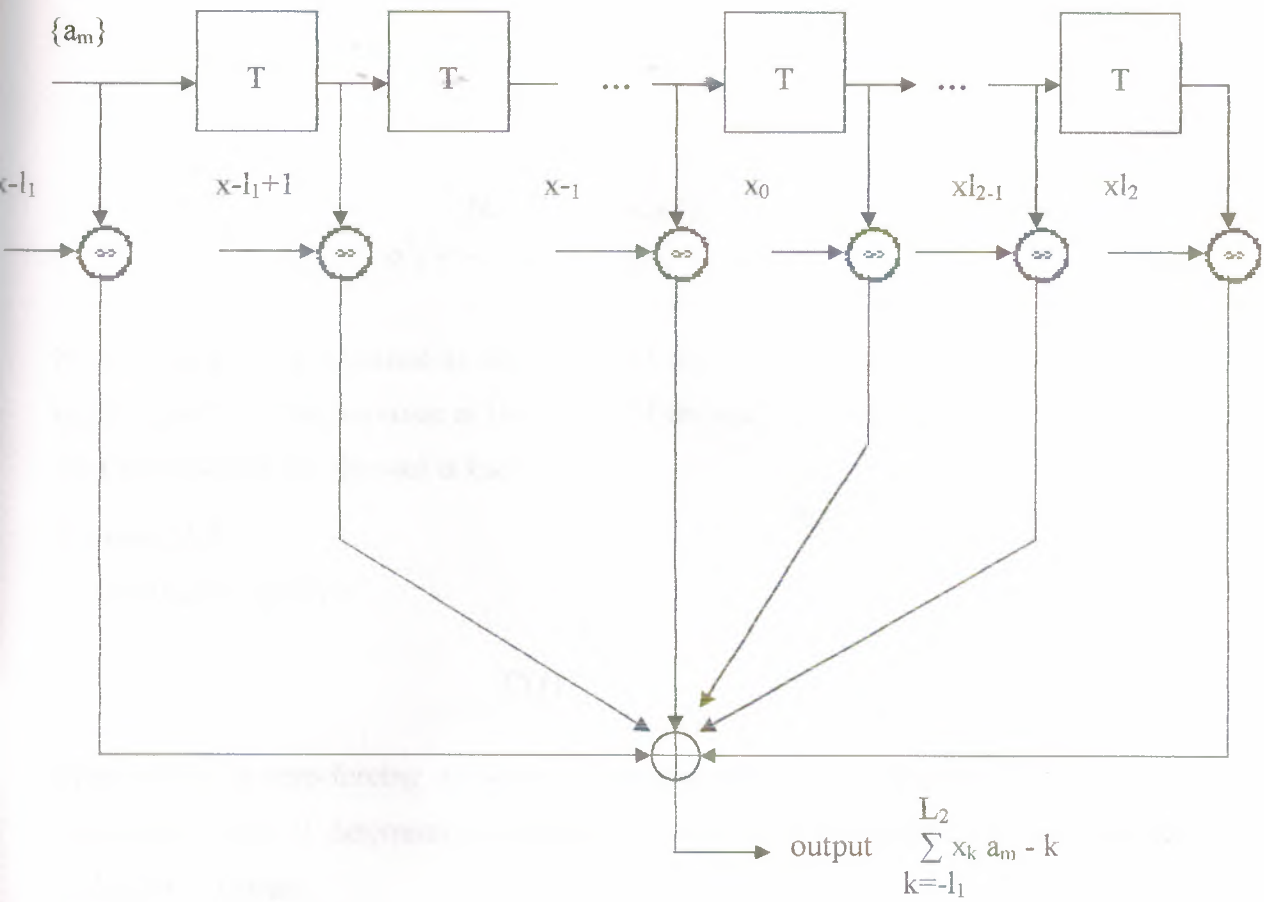


FIGURE 3.2. Equivalent discrete-time channel filter.

3.2.2 Linear Equalizers

To compensate for the channel distortion, we may employ a linear filter with adjustable parameters. The filter parameters are adjusted on the basis of measurements of the channel characteristics. These adjustable filters are called channel equalizer or, simply, equalizer.

On channels whose frequency response characteristics are unknown, but time-invariant, we may measure the channel characteristics, adjust the parameters of the equalizer, and once adjusted, the parameters remain fixed during the transmission of data. Such equalizers are called preset equalizers On the other hand, a daptive equalizers update their parameters on a periodic basis during the transmission of data.

Where $s_n(f)$ is the power-spectral density of the noise. When the noise is $s_n(f)| = N_0 / 2$, and the variance becomes

$$\sigma^2_{\upsilon} = \frac{N_0}{2} \int_{-W}^{W} \frac{|x_{rc}(f)|}{|C(f)|^2} df \tag{3.2.6}$$

Note that the noise variance at the output of the zero-forcing equalizer is, ingeneral higher than the noise variance at the output of the optimum receiving filter $|G_R(f)|$, for the case in which the channel is known.

Example 3.2.1

The channel given in

$$|C(f)| = \frac{1}{\sqrt{1 + (f / W)^2}}$$

equalized by a zero-forcing equalizer. Assuming that the transmitting and receiving filters satisfy (3.2.1), determine the value noise variance at the sampling instants and the probability of error.

Solution

When the noise is white, the variance of the noise at the output of the zero-forcing equalizer (input to the detector) is given by (3.2.6) Hence,

$$\sigma^2_{\upsilon} = \frac{N_0}{2} \int_{-w}^{w} \frac{|x_{rc}(f)|}{|C(f)|^2} df$$

$$= \frac{TN_0}{2} \int_{-w}^{w} [1 + (\frac{f}{W})^2] \cos^2 \frac{\pi |f|}{2W} df$$

$$= N_0 \int_{0}^{1} (1 + x_2) \cos^2 \frac{\pi |f|}{2} dx$$

$$= (\frac{2}{3} - \frac{1}{\pi^2}) N_0$$

we obtain the average transmitted power as

$$p_{av} = \frac{(M^2 - 1)d^2}{3T} \int_{-w}^{w} |G_T(f)|^2 \, df$$

$$= \frac{(M^2 - 1)d^2}{3T} \int_{-w}^{w} |X_{rc}(f)| \, df$$

$$= \frac{(M^2 - 1)d^2}{3T}$$

from general expression for the probability of error, we have

$$p_M = \frac{2(M - 1)}{M} Q\left(\sqrt{\frac{3\, p_{av} T}{(M^2 - 1)(2/3 - 1/\pi^2)N_0}}\right)$$

If the channel were ideal, the argument of the Q-function would be $6p_{av} / (M^2 - 1)N_0$ Hence, the loss in performance due to the equalized nonideal channel is given by the factor $2(2/3 - 1/\pi^2)$ 1.133 or 0.54 dB.

Let us now consider the design of a linear equalizer from a time-domain viewpoint. We noted previously that in real channels, the ISI is limited to a finite number of samples, say L samples. As a consequence, in practice the channel equalizer is approximated by a finite duration impulse response (FIR) filter, or transversal filter, with adjustable tap coefficients $\{c_n\}$, as illustrated in Figure 3.3. The time delay τ between adjacent taps may be selected as large as T, the symbol interval, in which case the FIR equalizer is called a symbol-spaced equalizer. In this case the input to the equalizer is the sampled sequence given by (3.2.10). However, we note that when $1 / T < 2W$, In frequencies in the received signl above the folding frequency $1 / T$ are into frequencies below $1 / T$. In this case, the equalizer compensates for the aliased channel-distorted signal.

On the other hand, when the time delay τ between adjacent taps is select such that $1 / \tau \geq 2W > 1 / T$, no aliasing occurs and, hence, the inverse channel equalizer compensates for the true channel distortion. Since $\tau < T$ the channel equalizer is said to have fractionally spaced equalizer and it is called a fractionally spaced equalizer. In practice, τ is often selected as $\tau = T/2$. Notice that, in this case, the sampling rate at the output of the filter $G_R(f)$ is $2/T$.

The impulse response of the FIR equalizer is

$$g_E(t) = \sum_{n=-N}^{N} c_n \delta(n - n\tau) \qquad (3.2.7)$$

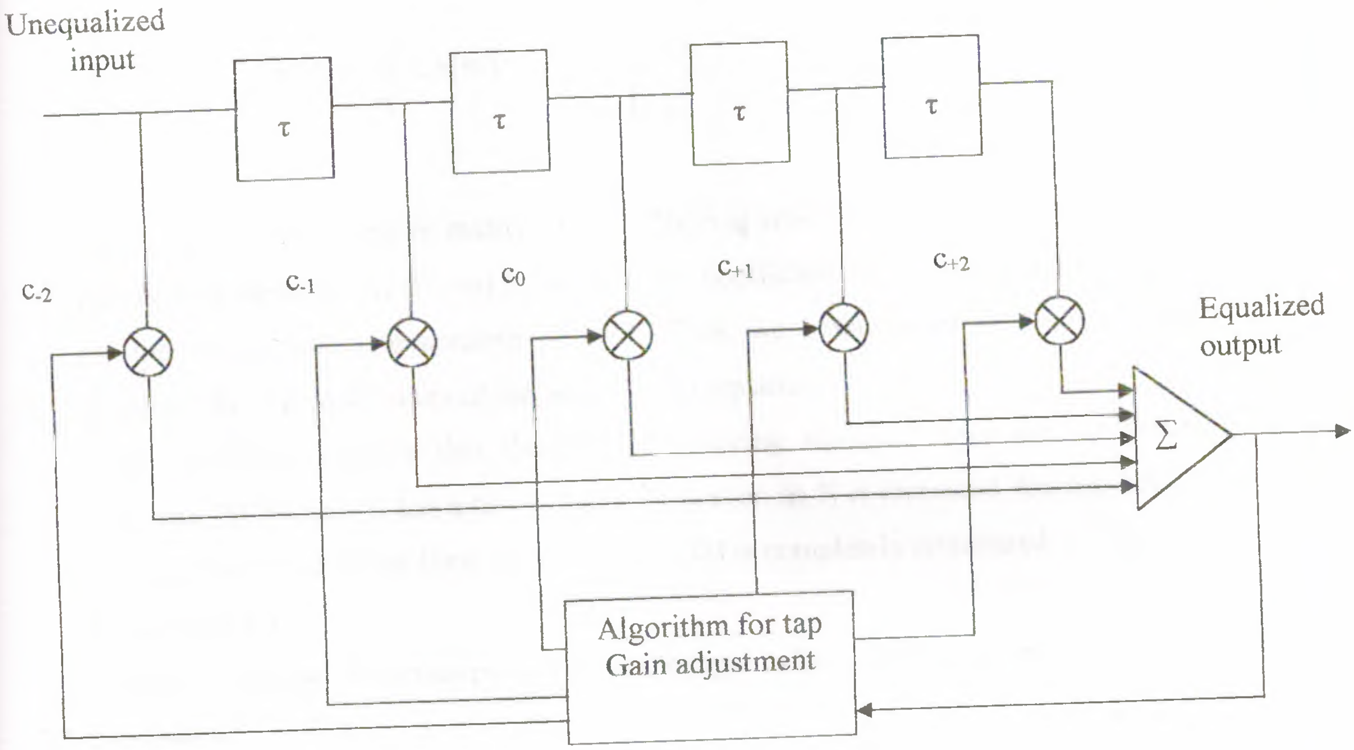


FIGURE 3.4. Linear transversal filter.

and the corresponding frequency response is

$$G_E(f) = \sum_{n=-N}^{N} c_n e^{-j2\pi f t\tau} \qquad (3.2.8)$$

where $\{c_n\}$are the $(2N + 1)$ equalizer coefficients, and N is chosen sufficient large so that the equalizer spans the length of the ISI, i.e., $2N + 1 \geq L$. since $X(f)C(f)G_R(f)$ and $x(t)$ is the signal pulse corresponding to $X(f)$ then the equalized output signal pulse is

$$q(t) = \sum_{n=-N}^{N} c_n x(n - n\tau) \qquad (3.2.9)$$

The zero-forcing condition can now be applied to the samples of q(t) taken times $t =$ mT. These samples are

$$q(mT) = \sum_{n=-N}^{N} c_n x(mT - n\tau), \qquad m = 0, \pm1, \pm2, \ldots, \pm N \quad (3.2.10)$$

Since there are 2N + 1 equalizer coefficients, we can control only 2N + 1 sampled values of q(t) Specifically, we may force the conditions

$$q(mT) = \sum_{n=-N}^{N} c_n x(mT - n\tau) = \begin{cases} 1, & m = 0 \\ 0, & m = \pm1, \pm2, \ldots, \pm N \end{cases} \quad (3.2.11)$$

which may be expressed in matrix form as $\mathbf{Xc} = \mathbf{q}$ where $\mathbf{X}$ *is a* $(2N + 1) \times (2N + 1)$ matrix with elements $x(mT - n\tau)$ is the $(2N + 1)$ coefficient vector and $\mathbf{q}$ is the $(2N + 1)$ column vector with one nonzero element. Thus, we obtain a set of 2N + 1 linear equations for the coefficients of the zero-forcing equalizer.

We should emphasize that the FIR zero-forcing equalizer does not completely eliminate ISI because it has a finite length. However, as N is increased, the residual ISI can be reduced and in the limit as $N \to \infty$, the ISI is completely eliminated.

Example 3.2.2

Consider a channel distortion pulse x(t), at the input to the equalizer, given by the expression

$$X(t) = \frac{1}{1 + (2t / T)^2}$$

where 1 / T is the symbol rate. The pulse is sampled at the rate 2 / T and equalized by a zero-forcing equalizer . Determine the cofficients of a five-tap zero-forcing equalizer.

solution

According to (3.2.8),the zero forcing equalizer must satify the equations

$$q(mT) = \sum_{n=-2}^{2} c_n x(mT - nT / 2) = \begin{cases} 1, & m = 0 \\ 0, & m = \pm1, \pm2, \end{cases}$$

The matrix $\mathbf{X}$ with elements $x(mT - nT / 2)$ is given as

$$
\mathbf{X} = \begin{bmatrix} \frac{1}{5} & \frac{1}{10} & \frac{1}{17} & \frac{1}{26} & \frac{1}{37} \\ 1 & \frac{1}{2} & \frac{1}{5} & \frac{1}{10} & \frac{1}{17} \\ \frac{1}{5} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{17} & \frac{1}{10} & \frac{1}{5} & \frac{1}{2} & 1 \\ \frac{1}{37} & \frac{1}{26} & \frac{1}{17} & \frac{1}{10} & \frac{1}{5} \end{bmatrix} \qquad (3.2.12)
$$

The cofficient vector **C** and the vector **q** are given as

$$
\mathrm{C} = \begin{bmatrix} c_{-2} \\ c_{-1} \\ c_0 \\ c_1 \\ c_2 \end{bmatrix} \qquad \mathbf{q} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad (3.2.13)
$$

Then, the linear equations $\mathbf{Xc} = \mathbf{q}$ can be solved by inverting the matrix **X**. Thus, we obtain

$$
\mathbf{c}_{op} = \mathbf{X}^{-1}\mathbf{q} = \begin{bmatrix} -2.2 \\ 4.9 \\ -3 \\ 4.9 \\ -2.2 \end{bmatrix} \qquad (3.2.14)
$$

One drawback to the zero-forcing equalizer is that it ignores the presence of additive noise. As a consequence, its use may result in significant noise enhancement. This is easily seen by noting that in a frequency range where $C(f)$ is small, the channel equalizer $G_E(f) = 1 / C(f)$ compensates by placing a large gain in that frequency range. Consequently, the noise in that frequency range is greatly enhanced. An alternative is to relax the zero- ISI condition and select the channel equalizer characteristic such that the combined power in the residual ISI and the additive noise at the output of the equalizer that optimized based on the minimum mean-sequare error (MMSE) criterion accomplishes to desired goal.

To elaborate, let us consider the noise corrupted output of the FIR equalizer, which is

$$z(t) = \sum_{n=-N}^{N} c_n y(t - n\tau) \qquad (3.2.15)$$

where y(t) is the input to the equalizer, given by (). The output is sampled at times t = mT. Thus, we obtain

$$z(mT) = \sum_{n=-N}^{N} c_n y(mT - n\tau) \qquad (3.2.16)$$

The desired response samples at the output of the equalizer at t = mT is the transmitted symbol a_m. The error is defined is the difference between a_m and z(mT) . then, the mean-squre error (MSE) between the actual output sample z(mT) and the desired value a_m is

$$MSE = E[z(mT) - a_m]^2$$

$$= E\left[\sum_{n=-N}^{N} c_n y(mT - n\tau) - a_m\right]^2$$

$$= \sum_{n=-N}^{N} \sum_{k=-N}^{N} c_n c_k R_Y(n - k) - 2\sum_{k=-N}^{N} c_k R_{AY}(k) + E[a^2{}_m] \qquad (3.2.17)$$

where the correlations are definrd as

$$R_Y(n - k) = E[y(mT - n\tau)y(mT - k\tau)]$$

$$R_{AY}(k) = E[y(mT - k\tau)a_m] \qquad (3.2.18)$$

and the expectation is taken with respect to the random information sequence $\{a_m\}$ and the additive noise.

The MMSE solution is obtained by the differentiating () with respect to the equalizer cofficients$\{c_n\}$. Thus, we obtain the necessary conditions for the MMSE as

$$\sum_{n=-N}^{N} c_n R_Y(n - k) = R_{AY}(k), \quad k = 0, \pm 1, \pm 2, \ldots, \pm N \qquad (3.2.19)$$

These are (2N + 1) linear equation for the equalizer cofficients. In contrast to the zero-forcing solution described previously, these equations depend on the statistical properties (the autocorrelation) of the noise as well as the ISI through the autocorrelation$R_Y(n)$.

In practice, we would not normally know the autocorrelation $R_Y(n)$ and thecrosscorrelation $R_{AY}(n)$. However, these correlation sequences can be estimated by transmitting a test signal over the channel and using the time-average estimates

$$\check{R} = \frac{1}{K}\sum_{k=1}^{k} y(kT - n\tau)y(kT)$$

$$\check{R}_{AY} = \frac{1}{K}\sum_{k=1}^{k} y(kT - n\tau)y(kT)a_k \tag{3.2.20}$$

in place of the ensemble averages to solve for the equalizer coefficients given by (3.2.19).

3.2.3 Adaptive Equalizers

We have shown that the tap coefficients of a linear equalizer can be determined by solving a set of linear equations. In the zero-forcing optimization criterion, the linear equations are given by (3.2.11). On the other hand, if the optimization criterion is based on minimizing the MSE, the optimum equalizer coefficients are determined by solving the set of linear equations given by (3.2.19).

In both cases, we may express the set of linear equations in the general matrix form

$$\mathbf{Bc} = \mathbf{d} \tag{3.2.21}$$

where $\mathbf{B}$ is a $(2N+1) \times (2N+1)$ matrix, $\mathbf{c}$ is a column vector representing the $2N+1$ equalizer coefficients, and d is a $(2N+1)$ dimensional column vector. The solution of (3.2.21) yields

$$\text{copt} = \mathbf{B}^{-1}\mathbf{d} \tag{3.3.22}$$

In practical implementations of equalizers, the solution of (3.2.21) for the optimum coefficient vector is usually obtained by an iterative procedure that avoids the explicit computation of the inverse of the matrix $\mathbf{B}$. The simplest iterative procedure is the method of steepest descent, in which one begins by choosing arbitrarily the coefficient vector $\mathbf{c}$, say $\mathbf{c}_0$. This initial choice of coefficients corresponds to a point on the criterion function that is being optimized. For example, in the case of the MSE criterion, the initial guess Co corresponds to a point on the quadratic MSB surface in the$(2N + 1)$ dimensional space of coefficients. The gradient vector, defined as g_0, which is the derivative of the MSE with respect to the $2N + 1$ filter coefficients, is then computed at this point on the criterion surface and each tap coefficient is changed in the direction opposite to its corresponding gradient component. The change in the j^{th} tap coefficient is proportional to the size of the j^{th} gradient component.

For example, the gradient vector, denoted as $\mathbf{g}_k$ for the MSE criterion, found by taking the derivatives of the MSE with respect to each of the 2N + 1 coefficients, is

$$\mathbf{g}_k = \mathbf{B}\mathbf{c}_k - \mathbf{d}, \quad k = 0, 1, 2, \ldots \tag{3.2.23}$$

Then the coefficient vector $\mathbf{c}_k$ is updated according to the relation

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \Delta\mathbf{g}_k \tag{3.2.24}$$

where Δ is the step-size parameter for the iterative procedure. To ensure convergence of the iterative procedure, Δ is chosen to be a small, positive number. In such a case, the gradient vector $\mathbf{g}_k$ converges toward zero, i.e., $\mathbf{g}_k \rightarrow \mathbf{0}$ as $k \rightarrow \infty$ and the coefficient vector $\mathbf{c}_k \rightarrow \mathbf{c}_{opt}$ as illustrated in Figure 3.5 based on two dimensional optimization. In general, convergence of the equalizer tap coefficients to $\mathbf{c}_{opt}$ cannot be attained in a finite number of iterations with the steepest-descent method. However, the optimum solution $\mathbf{c}_{opt}$ can be approached as closely as desired in a few hundred iterations. In digital communication systems that employ channel equalizers, each iteration corresponds to a time interval for sending one symbol and, hence, a few hundred iterations to achieve convergence to $\mathbf{c}_{opt}$ corresponds to a fraction of a second.

Adaptive channel equalization is required for channels whose characteristics change with time. In such a case, the ISI varies with time. The channel equalizer must track such time variations in the channel response and adapt its coefficients to reduce the ISI. In the context of the above discussion, the optimum coefficient vector $\mathbf{c}_{opt}$ varies with time due to time variations in the matrix $\mathbf{B}$ and, for the case of the MSE criterion, time variations in the vector $\mathbf{d}$. Under these conditions, the iterative method described above can be modified to use estimates of the gradient components. Thus, the algorithm for adjusting the equalizer tap coefficients may be expressed as

$$\hat{\mathbf{C}}_{k+1} = \hat{\mathbf{C}}_k - \Delta\hat{\mathbf{g}}_k \tag{3.2.25}$$

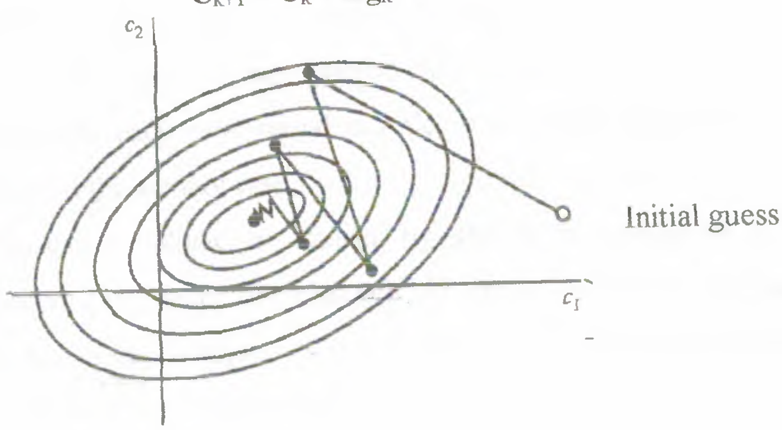


FIGURE 3.5. Example of convergance characteristics of gradient algorithm.

where $\hat{g}_k$ denotes an estimate of the gradient vector $\hat{g}_k$ and $\hat{c}_k$ denotes the estimate of the tap coefficient vector.

In the case of the MSE criterion, the gradient vector $\hat{g}_k$ given by (3.2.23) may also be expressed as

$$g_k = -E\,[e_k \mathbf{y}_k]$$

An estimate $\hat{g}_k$ of the gradient vector at the k^{th} iteration is computed as

$$\hat{g}_k = -\,e_k \mathbf{y}_k \tag{3.2.26}$$

where e_k denotes the difference between the desired output from the equalizer at the k^{th} time instant and the actual output z(kT), and $\mathbf{y}_k$ denotes the column vector of 2N + 1 received signal values contained in the equalizer at time instant k. The error signal e_k is expressed as

$$e_k = e_k - z_k \tag{3.2.27}$$

where z_k = z(kT) is the equalizer output given by (3.2.15), and a_k is the desired symbol. Hence, by substituting (3.2.26) into (3.2.25), we obtain the adaptive algorithm for optimizing the tap coefficients (based on the MSE criterion) as

$$\hat{c}_{k+1} = \hat{c}_k + \Delta e_k \mathbf{y}_k \tag{3.2.28}$$

Since an estimate of the gradient vector is used in (3.2.28), the algorithm is called a stochastic gradient algorithm. It is also known as the LMS algorithm.

A block diagram of an adaptive equalizer that adapts its tap coefficients according to (3.2.28) is illustrated in Figure 3.6. Note that the difference between the desired output a_k and the actual output z_k from the equalizer is used to form the error signal a_k. This error is scaled by the step-size parameter A, and the scaled-error signal Δe_k multiplies the received signal values {y(kT - nτ)} at the 2N + 1 taps. The products Δe_k y(kT - nτ) at the (2N +1) taps are then added to the previous values of the tap coefficients to obtain the updated tap coefficients, according to (3.2.28). This computation is repeated as each new signal sample is received. Thus, the equalizer coefficients are updated at the symbol rate.

Initially, the adaptive equalizer is trained by the transmission of a known pseudorandom sequence {am} over the channel. At the demodulator, the equalizer employs the known sequence to adjust its coefficients. Upon initial adjustment, the adaptive equalizer switches from a training mode to a decision-directed mode, in which case the decisions at the output of the detector are sufficiently reliable so that the error signal is formed by computing the difference between the detector output and the equalizer output, i.e.,

$$e_k = \tilde{a}_k - z_k \tag{3.2.29}$$

where $\tilde{a}_k$ is the output of the detector. In general, decision errors at the output of the detector occur infrequently and, consequently, such errors have little effect on the nce of the tracking algorithm given by (3.2.28).

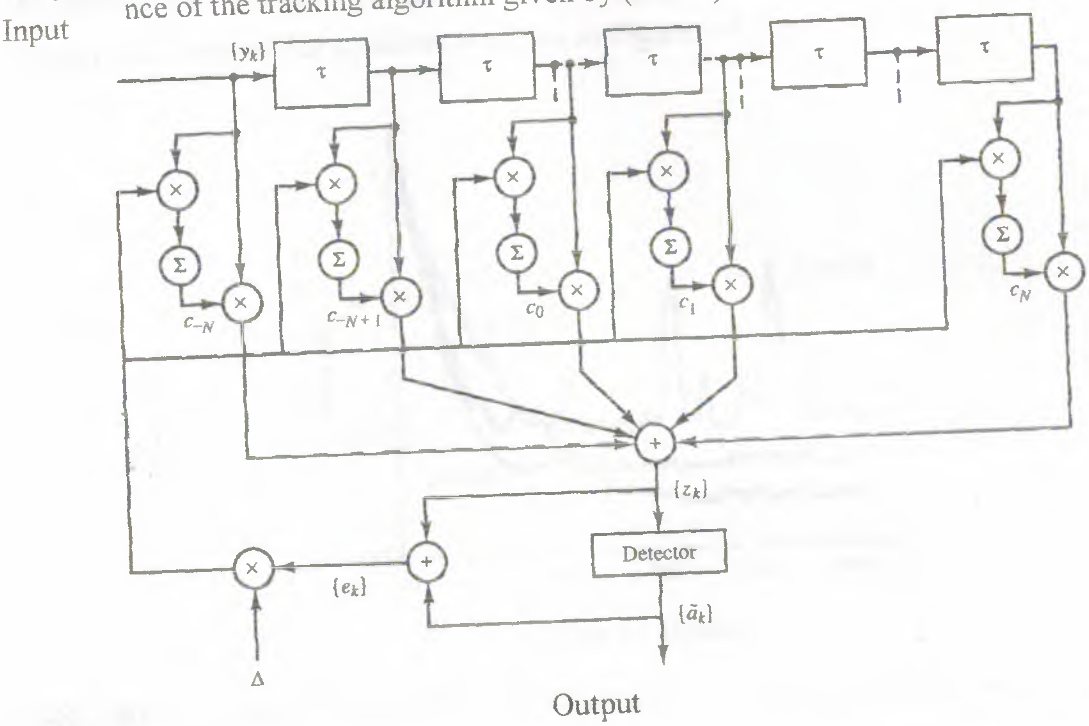


FIGURE 3.6. Linear adaptive equalizer based on the MSE criterion.

A rule of thumb for selecting the step-size parameter so as to ensure convergance and good tracking capabilities in slowly varying channels is

$$\Delta = \frac{1}{5(2N + 1)P_R} \tag{3.2.30}$$

where P_R denotes the received signal-plus-noise power, which can be estimated from the received signal.

The convergence characteristics of the stochastic gradient algorithm in (3.2.28) is illustrated in Figure 3.7. These graphs were obtained from a computer simulation of an 11-tap adaptive equalizer operating a channel with a rather modest amount of ISI. The input signal-plus-noise power P_R was normalized to unity. The rule of thumb given in (3.2.30) for selecting the step size gives A 0.018. The effect of making A too large is illustrated by the large jumps in MSE as shown for A 0.115. As A is decreased, the convergence is slowed somewhat, but a lower MSE is achieved, indicating that the estimated coefficients are closer to $\mathbf{c}_{opt}$.

Although we have described in some detail the operation of an adaptive equalizer which is optimized on the basis of the MSE criterion, the operation of an adaptive equalizer based on the zero-forcing method is very similar. The major difference lies in the method for estimating the gradient vectors $\mathbf{g}_k$ at each iteration. A block diagram of an adaptive zero-forcing equalizer is shown in Figure 3.8. For

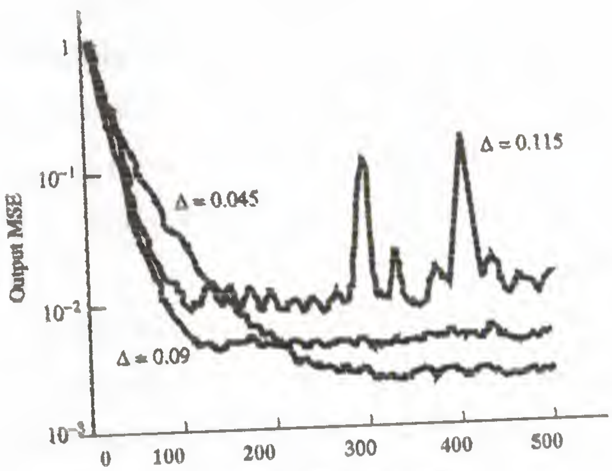


FIGURE 3.7. Initial convergence characteristics of the LMS algorithm with different step sizes.

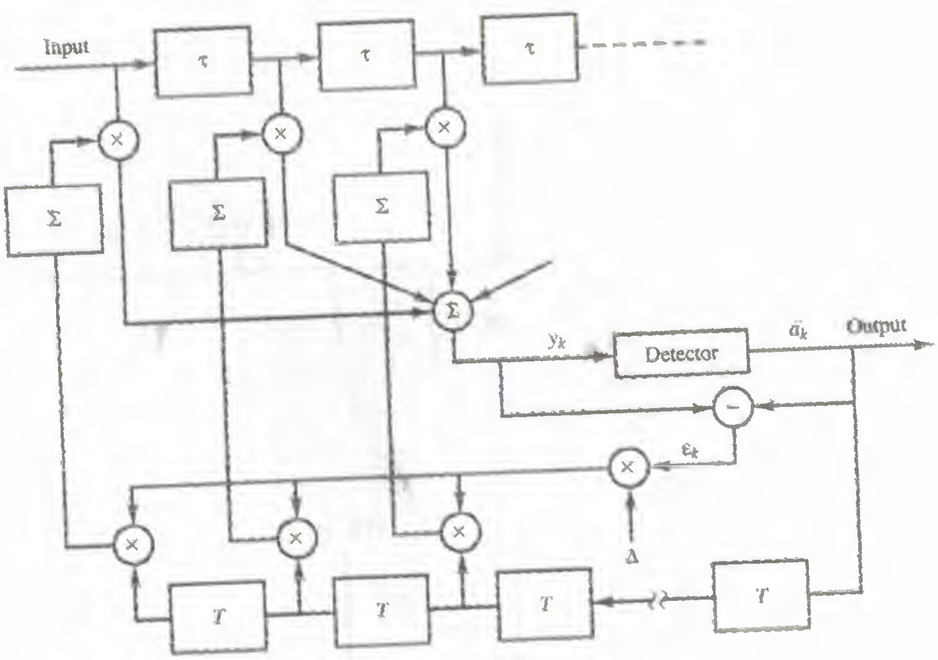


FIGURE 3.8. An adaptive zero-forcing equalizer.

more details on the tap coefficient update method for a zero-forcing equalizer, the reader is referred to the papers by Lucky (1965, 1966), and the texts by Lucky, Slaz, and Weldon (1968) and Proakis (1989).

3.2.4 Decision-Feedback Equalizer

The linear filter equalizers described above are very effective on channels, such as wireline telephone channels, where the ISI is not severe. The severity of the ISI is directly related to the spectral characteristics and not necessarily to the time span of the ISI. For example, consider the ISI resulting from two channels, which are illustrated in Figure 3.9. The time span for the ISI in Channel A is 5 symbol intervals on each side of the desired signal component, which has a value of 0.72. On the other hand, the time span for the ISI in Channel B is one symbol interval on each side of the desired signal component, which has a value of 0.815. The energy of the total response is normalized to unity for both channels.

In spite of the shorter ISI span, Channel B results in more severe ISI. This is evidenced in the frequency response characteristics of these channels, which are shown in Figure 3.10. We observe that Channel B has a spectral null (the frequency response C(0 for some frequencies in the band $|f| \leq W$) at $f = 1 / 2T$, whereas this does not occur in the case of Channel A. Consequently, a linear equalizer will introduce a large gain in its frequency response to compensate for

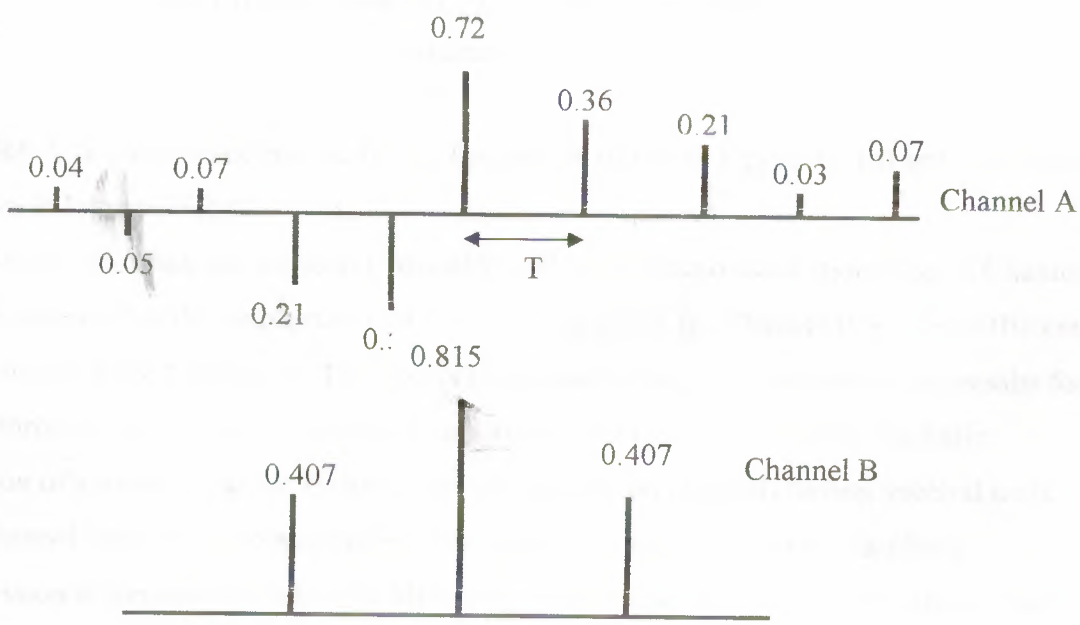
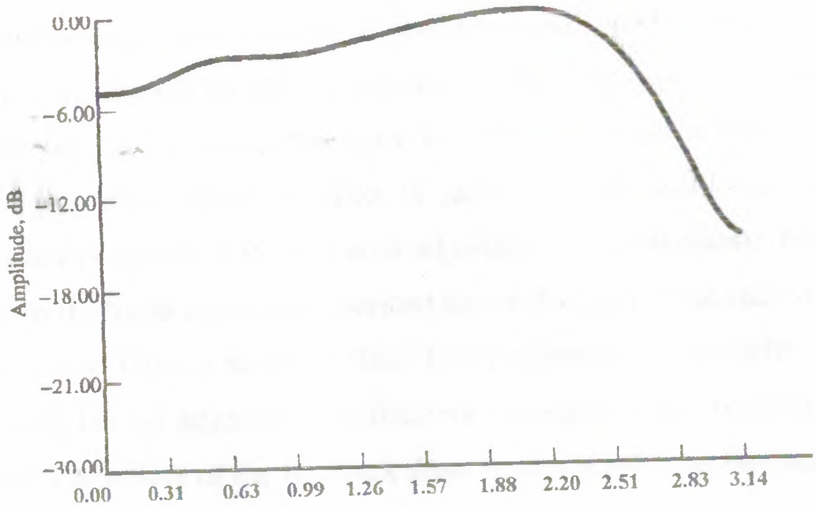
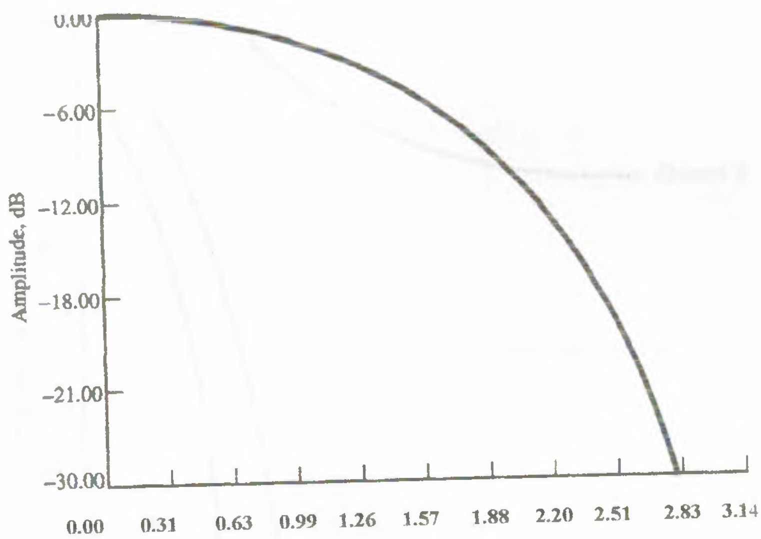


FIGURE 3.9. Two channel with ISI.



ω frequency
(a)



ω frequency
(b)

FIGURE 3.10. Amplitude spectra fpr (a) Channel A shown in Figure 3.9 (a) and (b) Channel B shown in Figure 3.9(b)

the channel null. Thus, the noise in Channel B will be enhanced much more than in Channel A. This implies that the performance of the linear equalizer for Channel B will be sufficiently poorer than that for Channel A. This fact is borne out by the computer simulation results for the performance of that two linear equalizers shown in Figure 3.11. Hence, the basic limitation of a linear equalizer is that it performs poorly on channels having spectral nulls. Such channels are often encountered in radio communications, such as ionospheric transmission at frequencies below 30 MHz and mobile radio channels, such as those used for cellular radio communications.

A decision feedback equalizer (DFE) is a nonlinear equalizer that employs previous decisions to eliminate the ISI caused by previously detected symbols on the current symbol to be detected. A simple block diagram for a DFE is shown in Figure 3.12. The DFE consists of two n ters. The first filter is called a feedforward filter, and it is generally a fractionally spaced FIR filter with adjustable tap coefficients. This filter is identical in form to the linear equalizer described above. Its input is the received filtered signal y(t). The second filter is feedback filter It is implemented as an FIR filter with symbol-spaced taps having adjustable coefficients. Its input is the set of previously detected symbols. The output of the feedback filter is subtracted from the output of the feedforward filter to form the

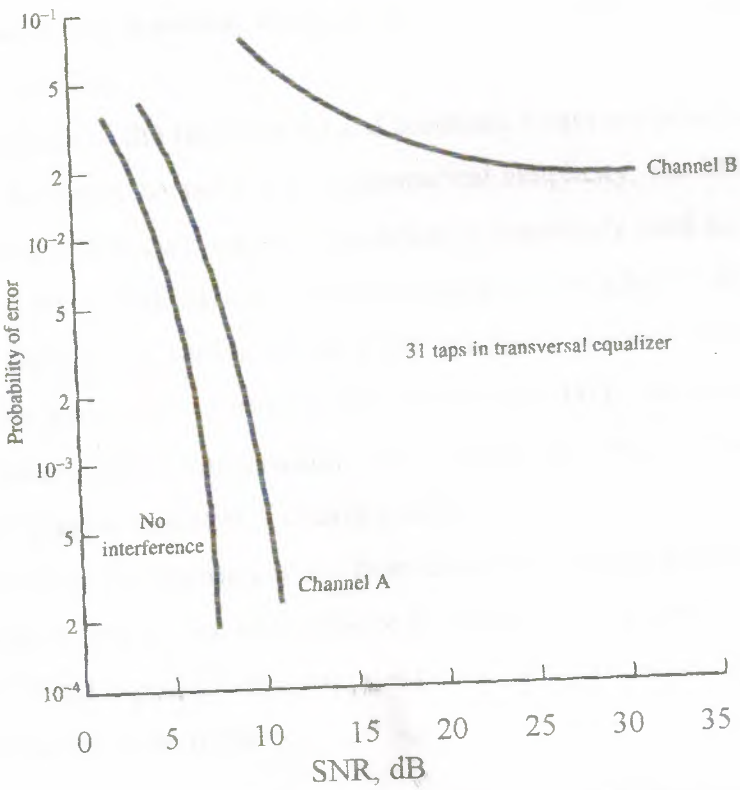


FIGURE 3.11. Error rate performance of linear MSE equalizer.

input to the detector. Thus, we have

$$z_m = \sum_{n=1}^{N_1} c_n y(mT - n\tau) - \sum_{n=1}^{N_2} b_n \tilde{a}_{m-n} \qquad (3.2.31)$$

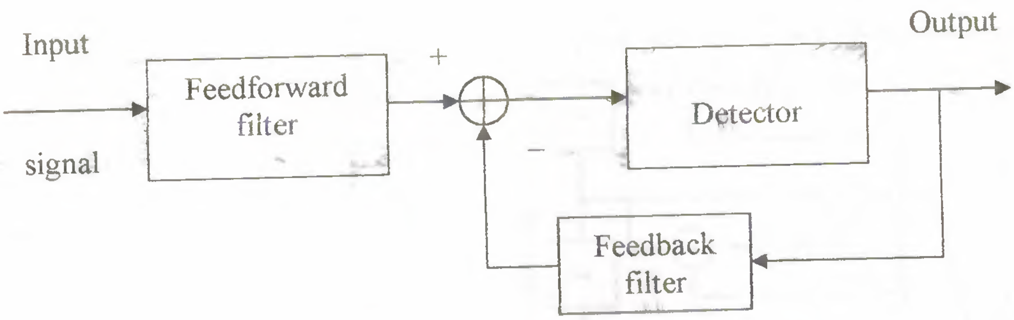


FIGURE 3.12. Block diagram of DFE.

where $\{c_n\}$ and $\{b_n\}$ are the adjustable coefficients of the feedforward and feedback filters, respectively, $\tilde{a}_{m-n} =$ 1, 2, ..., N_2 are the previously detected symbols, N_1 is the length of the feedforward filter, and N_2 is the length of the feedback filter. Based on the input z_m, the detector determines which of the possible transmitted symbols is closest in distance to the input signal z_m. Thus, it makes its decision and outputs $\tilde{a}_m$. What makes the DFE nonlinear is the nonlinear characteristic of the detector, which provides the input to the feedback filter.

The tap coefficients of the feedforward and feedback filters are selected to optimize some desired performance measure. For mathematical simplicity, the MSE criterion is usually applied and a stochastic gradient algorithm is commonly used to implement an adaptive DFE. Figure 3.13 illustrates the block diagram of an adaptive DFE whose tap coefficients are adjusted by means of the LMS stochastic gradient algorithm. Figure 3.14 illustrates the probability of error performance of the DFE, obtained by computer simulation, for binary PAM transmission over Channel B The gain in performance relative to that of a linear equalizer is clearly evident.

We should mention that decision errors from the detector that are fed to the feedback filter have a small effect on the performance of the DFE. In general, a small loss in performance of 1-2 dB is possible at error rates below 10^{-2}, but the decision errors in the feedback filters are not catastrophic.

Although the DFE outperforms a linear equalizer, it is not the optimum equalizer from the viewpoint of minimizing the probability of error. As indicated previously, the optimum detector in a digital communication system in the presence of ISI is a ML symbol sequence detector. It is particularly appropriate for channels with severe ISI, when the ISI spans only a few signals. We observe that the performance of the ML sequence detector is about 4.5 dB better than that of the DFE at an error probability of 10^{-4}.

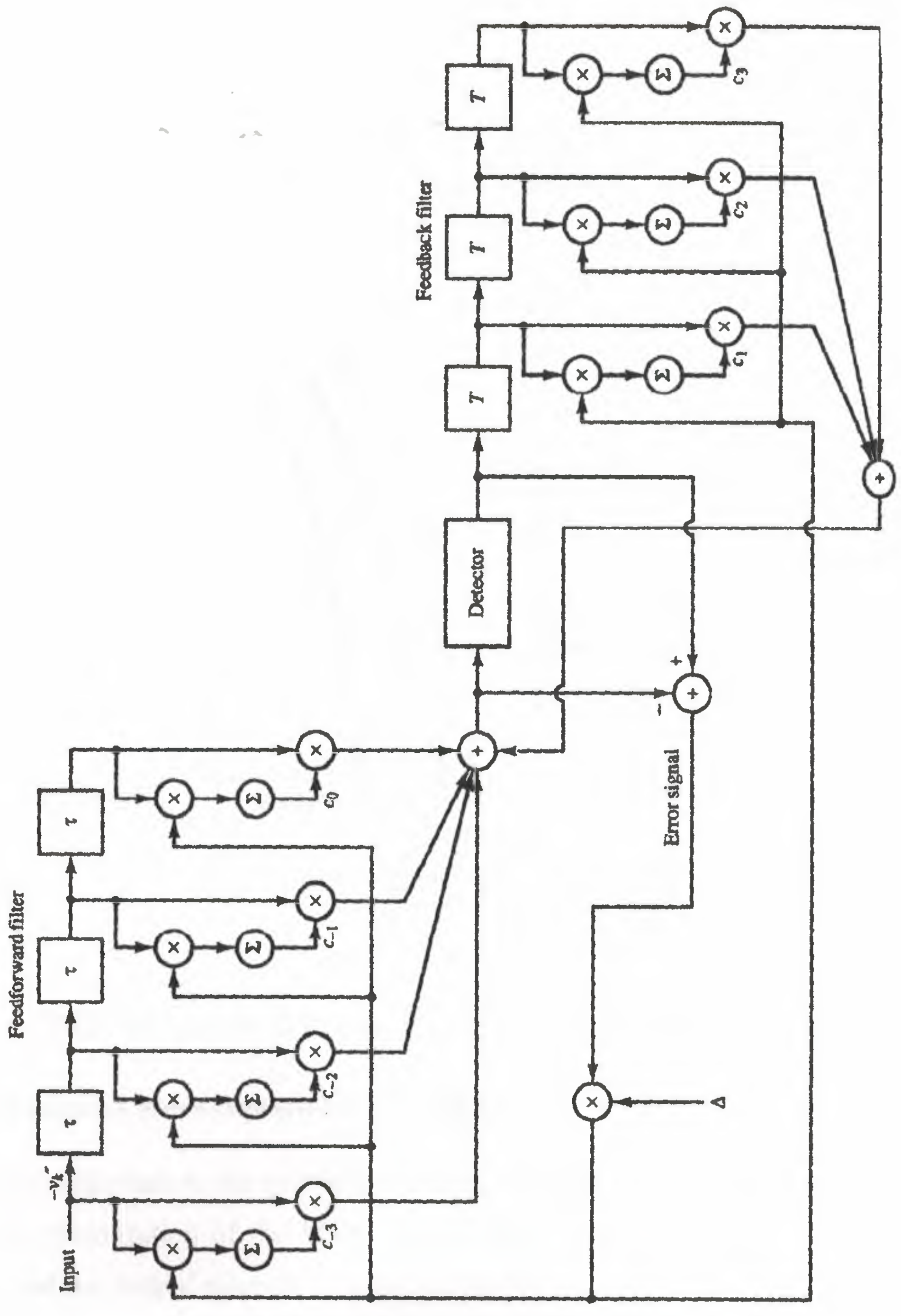


FIGURE 3.13. Adaptive DEF.

Hence, this is one example where the ML sequence detector provides a significant performance gain on a channel with a relatively short ISI span.

In conclusion, we mention that adaptive equalizers are used widely in highspeed digital communication systems for telephone channels. High-speed telephone line modems (at bit rates above 2400 bits/sec) generally include an adaptive equalizer that is implemented as an FIR filter with coefficients that are adjusted based on the minimum mean squared error criterion. Depending on the data speed, the equalizer typically spans between 20 and 70 symbols. The LMS algorithm given by (3.2.28) is usually employed for the adjestment of the equalizer coefficients adaptively.

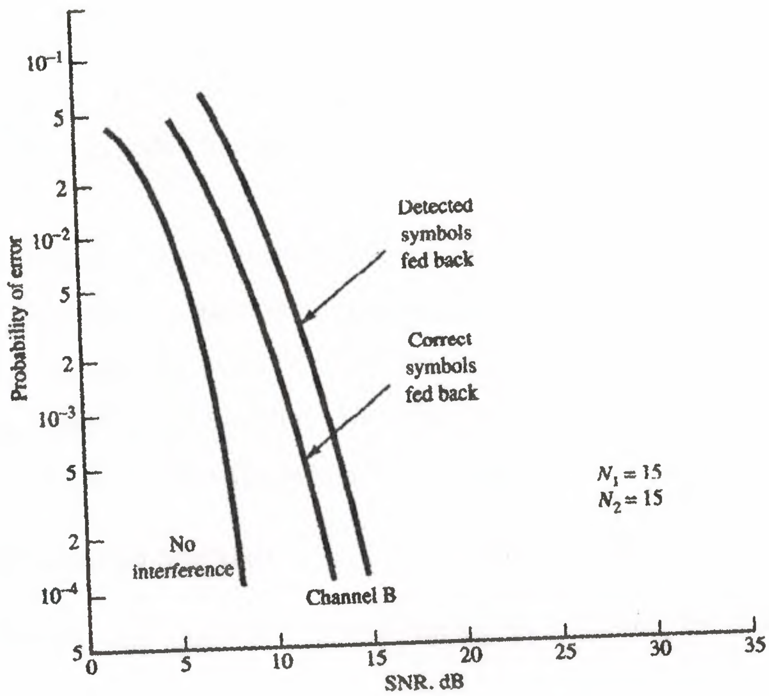


FIGURE 3.14. Performance of DFE with and without error propagation.

3.3 Minimum Mean-Square-Error Method

Another approach to the problem of timing recovery from the received signal is based on the minimization of the MSE between the samples at the output of the receiving filter and the desired symbols. In practice, the MSE

$$\text{MSE} = E\{[y_m(\tau_0) - a_m]^2\} \tag{3.3.1}$$

is approximated by the time-average-square error

$$(\text{MSE}) = \sum_m [y_m(\tau_0) - \hat{a}_m]^2 \tag{3.3.2}$$

Where

$$Y_m(\tau) = \sum a_n x(mT - nT - \tau_0) + \upsilon(mT) \tag{3.3.3}$$

and $\hat{a}_m$ is the output symbol from the detector.

The minimum of (MSE) with respect to the timing phase is τ_0 found by differentiating (3.3.2) with respect to τ_0 .Thus, we obtain the necessary condition

$$\sum_m [y_m(\tau_0) - \hat{a}_m] \frac{dy_m(\tau_0)}{d\tau_0} = 0 \qquad (3.3.4)$$

An interpretation of the necessary condition in (3.3.4) is that the optimum sampling time corresponds to the condition that the error signal $[y_m(\tau_0) - \hat{a}_m]$ is uncorrelated with the derivative $dy_m(\tau_0) / d\tau_0$. Since the detector output is used in the formation of the error signal $[y_m(\tau_0) - \hat{a}_m]$, this timing phase estimation method is said to be decision-directed.

Figure 3.15 illustrates an implementation of the system that is based on the condition given in (3.3.4). Note that the summation operation is implemented as a lowpass filter, which averages a number of symbols. The averaging time is roughly equal to the reciprocal of the bandwidth of the filter. The filter output drives the voltage-controlled oscillator (VCO), which provides the best MSE estimate of the timing phase τ_0.

3.4 Maximum-Likelihood Methods

In the ML criterion, the optimum symbol timing is obtained by maximizing the likelihood function.

$$\Lambda(\tau_0) = \sum_m a_m y_m(\tau_0) \qquad (3.4.1)$$

where $y_m(\tau_0)$is the sampled output of the receiving filter given by (3.3.3). From a mathematical viewpoint, the likelihood function can be shown to be proportional tothe probability of the received signal (vector) conditioned on a known transmitted signal. Physically, $\Lambda(\tau_0)$ is simply the output of the matched filter or correlator the receiver averaged over a number of symbols.

A necessary condition for TO to be the ML estimate is that

$$\frac{d\,\Lambda(\tau_0)}{d\,\tau_0} = \sum_m a_m \frac{dy_m(\tau_0)}{d\,\tau_0} = 0 \qquad (3.4.2)$$

This result suggests the implementation of the tracking loop shown in Figure 3.16 We observe that the product of the detector output $\hat{a}_m$ with the input $dy_m(\tau_0) / d\tau_0$ is averaged by alowpass filter that drives the VCO. Since the detect output is used in the estimation method, the estimate τ ^ is decision-directed.

As an alternative to the use of the output symbols from the detector, we may use a non-decision-directed method that does not require knowledge of the information symbols. This method is based on averaging over the statistics the symbols. For example, we may square the output of the receiving filter and maximize the function

$$\Lambda_2(\tau_0) = \sum_m y^2_m(\tau_0) \qquad (3.4.3)$$

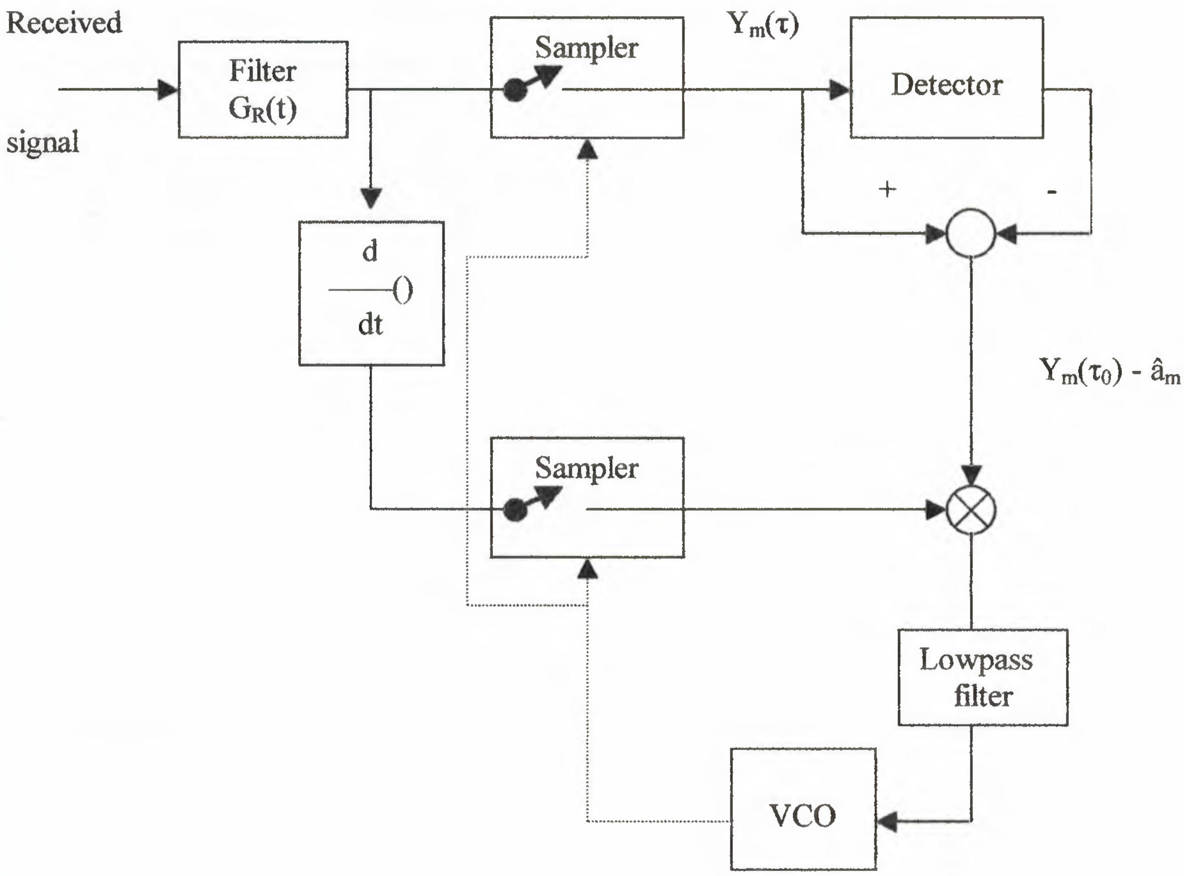


FIGURE 3.15. Timing recovery based on minimization of MSE.

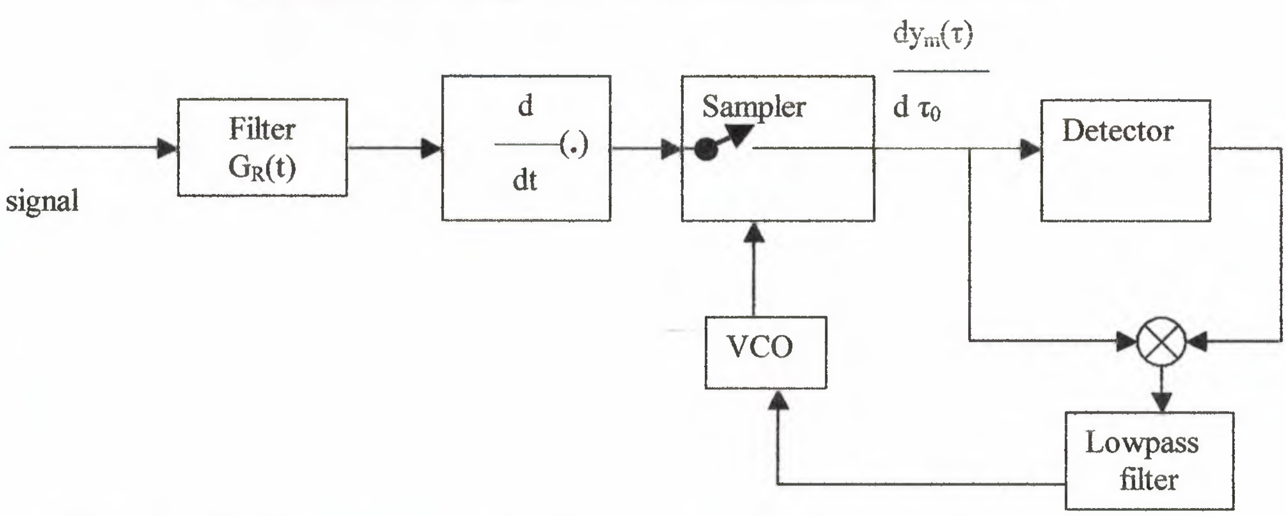


FIGURE 3.16. Decision-directed ML timing recovery method for baseband PAM.

With respect to τ_0. Thus, we obtain

$$\frac{\Lambda_2(\tau_0)}{d\,\tau_0} = 2\sum_m y_m(\tau_0)\frac{dy_m(\tau_0)}{d\tau_0} = 0 \qquad (3.4.4)$$

The condition for the optimum τ_0 givin by () may be satisfied by the implementation shown in Figure 3.17. In this case, there is no need to know data sequence $\{a_m\}$. Hence, the method is non-decision-directed.

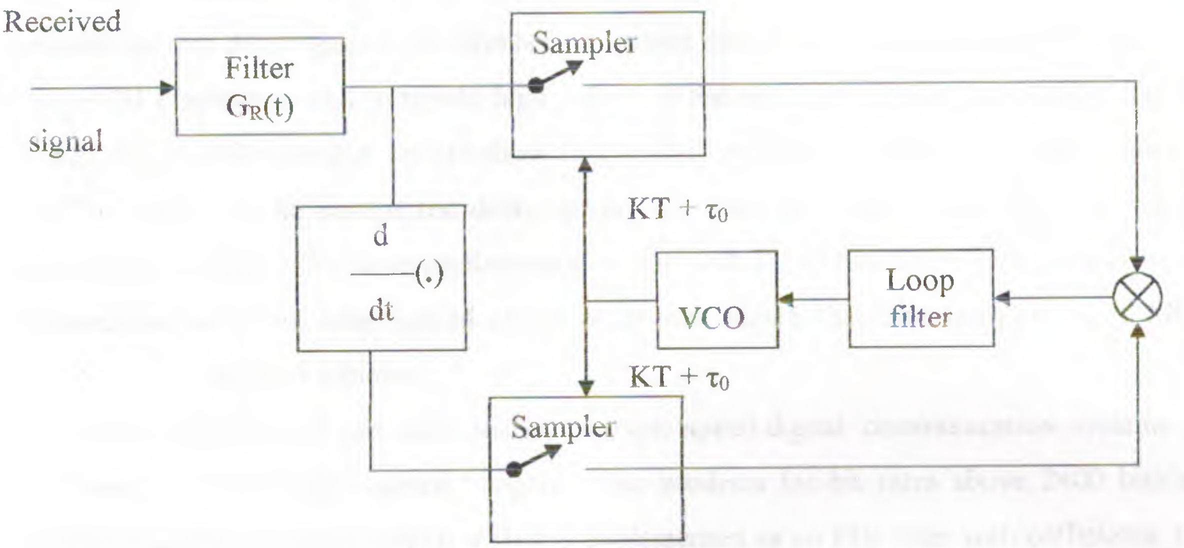


FIGURE 3.17. Non-decision-directed estimation of timing for baseband PAM.

CONCLUSION

A performance analysis with linear and decision feedback equalizers of broadband wireless communications systems with coherent QPSK modulation in channels representative of Rician fading has been carried out. The results for the bit error rate were based on Monte Carlo simulations with the exponential and the uniform power delay profile models. A simplified method was presented and used for generating the received signal. It was shown that the performance of CQPSK with linear or DF equalizers strongly depends on the normalized rms delay spread, the number of echoes and the power delay profile model. For severe ISI conditions with relatively high values of the normalized delay spread and for both power delay profile models, results show that a DFE performs better than a linear equalizer. For low values of the normalized delay spread, for instance = 0:1 , a system with a linear equalizer can achieve the same performance as one with a DF equalizer. For a typical indoor environment with rms delay spread of 7.5 ns, it was shown that data rates up to136 Mbs/s with low error rate can achieved.

Adaptive equalizers are used widely in high speed-digital communication systems for telephone channels. High –speed telephone line modems (at bit rates above 2400 bits/sec) generaly include in adaptive equlizer that is implemented as an FIR filter with cofficients that are adjusted based on the mean minimum sqared error criterion. Depending on the data speed , the equalizer typically spans between 20 and 70 symbols. The LMS algorithm given by (3.28) is usually employed for the adjustment of the equalizer cofficients adaptively.

REFERENCES

[1] Alder, R.L., COPPERSMITH, D., AND HASSNER, M (1983), "Algorithm for Sliding BlockCodes," IEEE Trans. Inform. Theory, vol. IT-29, pp. 5-22. January.

[2] ANDERSON, J. B., AULUN, T., AND SUNDBERG, C. W.(1986), Digital Phase Modulation, Plenum, New York.

[3] AULTUN, T., RYDBECK, N., AND SUNDBERG, C. W. (1981), "Continuous Phase Modulation-Part II: Partial Response Signalling," IEEE Trans. Commun., vol COM-29, pp.210-255, March.

[4] BELLO, P. A., AND NELIN, B. D. (1962a), "Predetection Diversity Combining with Selectively Fading Channels," IRE Trans. Commun. Sys., vol CS-10,pp. 32-44, March.

[5] BELLO, P. A., AND NELIN, B. D. (1962b), "The Influence of Fading Spectrum on the Binary Error Probabilities of Incoherent Matched Filter Receivers," IRE Trans. Commun. Sys., vol CS-11,pp. 170-186, june.

[6] BENEDETIO, S., BIGLIER, E., AND CASTELLANI, V. (1987), Digital Transmission Theory, Prentice-Hall, Englewood Cliffs, N.J.

[7] A.Semmar, H. T. Huynh, and M. Lecours, "Theperformance of coherent QPSK communications over frequeny-selective fading channels of broadband.

[8] M. Kavehrad and P. McLane, "Performance of lowcomplexity channel and diversity for spread spectrum in indoor wireless communication," *Bell. Syst., Tech.*

J. 164, pp. 1927–1965, Oct. 1985.

[9] J. G. Proakis, *Digital communications*. New-York,McGraw-Hill, 1995.

[10] G. L. Turin, "Communications through noisy, randommultipaths channels," *IRE Convention Rec., Part 4*,pp. 154–166, 1956.

[11] F. D. Garber and M. B. Pursely, "Performance ofbinary FSK communications over frequency-selective rayleigh channels," *IEEE Trans. Commun.*, vol. 37, pp. 83–89, Jan. 1989.

[12] S. Haykin, *Adaptive Filter Theory*, 3rd Edition, NJ: Prentice Hall, 1996.

[13] S.U.H. Qureshi, "Adaptive equalization," *Proceedings of IEEE*, Vol. 73, pp. 1349–1387, 1985.

[14] J-C. Feng and C.K. Tse, "On-line adaptive chaotic demodulator based on radial-basis-function neural networks," *Phys. Rev. E*, Vol. 63, pp. 026202:1–10, 2001.

[15] TANG, D. L .R. (1970), "Block Codes for a Class of Constrained Noiseless Channel,"Information and Control, vol. 17, pp .436-461.

[16] ZIEMER, R. E., AND PETERSON, R.L. (1985), Digital Communication and Spread Spectrum Systems, Macmillan, New York.

[17] ZIEMER, R. E., AND TRANTER, W.H. (1990), Principles of Communications Systems, Modulation, and Noise, Houghton Mifflin, Boston.

[18] UNGERBOCK, G.(1974),"Adaptive Maximum Likelihood Receiver for Carrier Modulated Data Transmission Systems,"IEEE Trans. Commun., vol. COM-22, pp.624-636, MAY.

[19] SAKRISON, D. J. (1968), Communication Theory: Transmission of Waveforms and Digital Information, New York, Wiley.

[20] OMURA, J. (1971),"Optimal Receiver Design for Convolutional Codes and Channels With Memory Via Control Theoretical Concepts,"INFO. Sci., vol 3, pp. 243-266.