NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Electrical and Electronic Engineering

MULTIPATH FADING REDUCTION USING CONVOLUTIONAL CODES

Graduation Project
EE – 400

Student: Mousab Radwan (20032829)

Supervisor: Dr. Ali Serener

Nicosia – 2007

ACKNOWLEDGMENT

All praise and glory to almighty ALLAH, the lord of the universe, who is the entire source of all knowledge and wisdom endowed to mankind. All thanks are due to him who gave me the ability and patience throughout my studies for completing this task.

I would like to acknowledge to my MOTHER AND MY FATHER (ASSOC.PROF.DR FAE'Q RADWAN) who has brought all of his efforts to support me, without knowing the return and who has patiently encouraged me to be the best everywhere.

I would like to thanks my project supervisor DR. ALI SERENER for his intellectual support, encouragement and enthusiasm which made it possible to accomplish this project. I appreciate his most gracious encouragement and very valued constructive critics throughout my education.

My special thanks goes to NEU education staff especially to electrical and electronic engineering teaching staff for their generosity and special concern of me and all E.E.E STUDENTS.

ABSTRACT

This report discusses convolutional codes, which are used as channel coding to reduce the errors that occur in the channel as much as possible. During transmission in the channel intersymbol interference and multipath fading distorts the signal.

The main goal of this report is to show the effect the convolutional codes on multipath fading.

MATLAB simulations are performed to show that convolutional codes improve the performance of communication systems affected by intersymbol interference and multipath fading.

TABLE OF CONTENTS

INTRODUCTION

Communication channels suffer from the effects of multipath, and intersymbol interference (ISI). Multipath occurs from the reflection in difference phases, and it creates small scale fading. Intersymbol interference occurs when transmitted pulses are distorted in the channel, when they interfere with each other.

This report investigates single carrier communications over various types of channels. It particularly investigates the additive white Gaussian noise (AWGN) channel and Rayleigh fading channel. Using MATLAB simulation on simulink, the performance of convolutional coding over fading channel is analyzed. The results show the superior effect of convolutional codes on multipath fading.

The first chapter describes the elements of the digital communication system and the noise within the system.

Chapter two describes multipath fading in detail and shows how they affect the transmitted signal.

Chapter three describes convolutional codes in detail.

Finally, chapter four includes the results obtained through simulation using simulink.

CHAPTER ONE

COMMINICATION SYSTEM OVERVIEW

1.1 Elements of an Electrical Communication System

Electrical communication systems are designed to send message or information from a source that generates the messages to one or more destinations. In general, a communication system can be represented by the functional block diagram shown in figure 1.1. The information generated by 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 is output is described in probabilistic terms; i.e., the output of a source is not deterministic. Otherwise, there would be no need to transmit the message.

A transducer is usually required to convert the output of source into an electrical signal that is suitable for transmission. For example, a microphone serves 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; e.g., acoustic signals, images, etc.

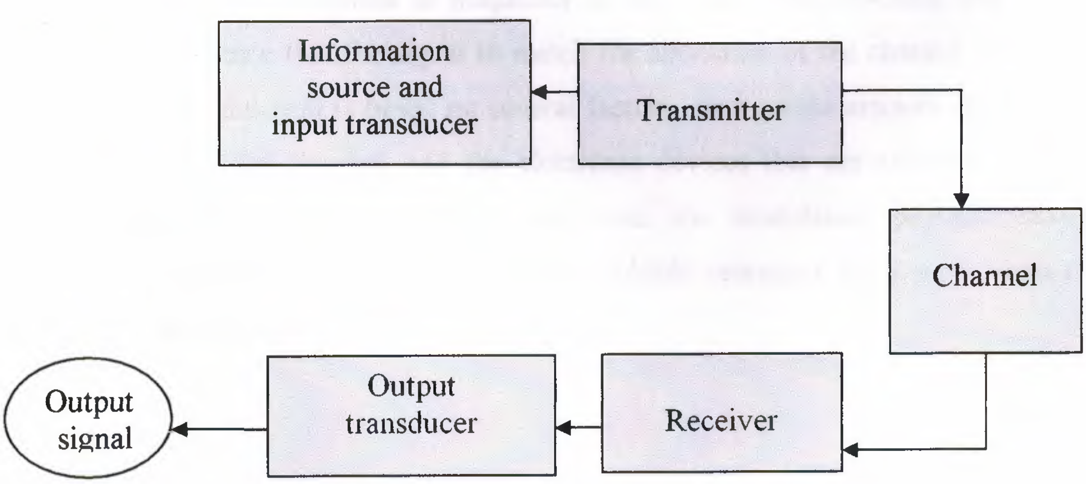

Figure 1.1 Functional block diagram of communication system [1].

1.1.1 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 a radio TV broadcast, the Federal communication 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 buy multiple radio station 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 process called modulation. Usually, modulation involves the use of the information signal to systematically vary either the amplitude, frequency, or phase of a sinusoidal 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. In FM radio broadcast, the information signal that is transmitted is contained in the frequency variations of the sinusoidal carrier. 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 characteristics of the channel. Thus through the process of modulation, the information signal is translated in frequency of the bandwidth allocated, the types of noise and interference that the signal to match the allocation of the channel. The choice of the type of modulation is based on several factors, such as the amount 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.

1.1.2 Channel

The communication channel is the physical medium that is used to send the signal from the transmitter to the receiver. In the wireless transmission, the channel is usually the atmosphere (free space). On the other hand, telephone channel 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 channel, such as ionospheric channel that is used for long range, Short wave radio transmission, another form of signal degradation is multi-path propagation.

Such signal distortion is characterized as a non-additive signal disturbance which manifests itself as time variations in the signal amplitude, usually called fading. This phenomenon is described in more detail in the following section 1.3.

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

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

1.1.3 Receiver

The function of the receiver is to cover the message signal contained in the received signal. If the message signal is transmitted by carrier modulation, the receiver performs carrier demodulation in order to extract the message form 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 non additive 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 along 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 a discrete (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.

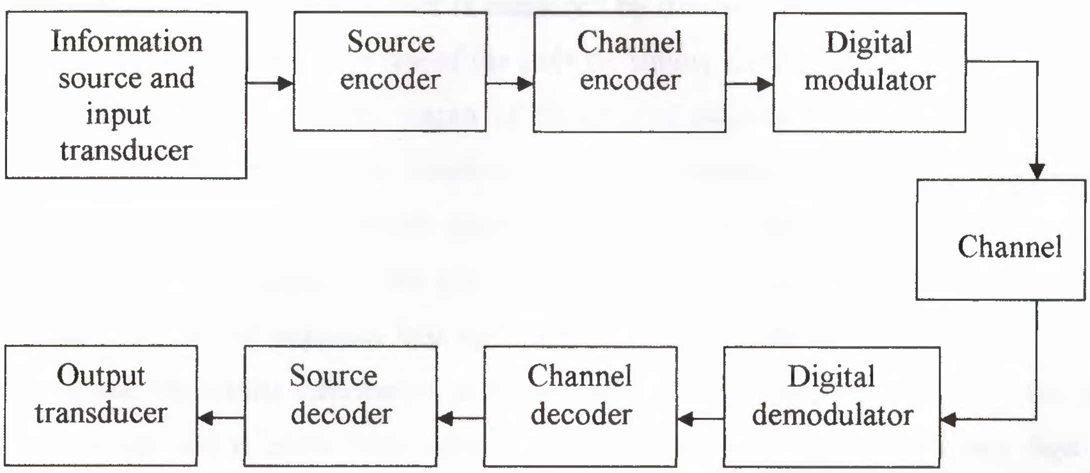

Figure 1.2 Basic elements of digital communication system [1].

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 computer which is discrete in time and has a finite number of output characters. In a digital communication system, the messages produced by the source are 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 encoding 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 a code 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 the 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 $M=2^b$ distinct waveforms $S_i(t)$, $i = 0, 1, \ldots, M-1$, one waveform for each of the 2^b possible b-bit sequences. 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 rate R is fixed, the amount of time available to transmit one of the M waveforms corresponding to a b-bit sequence is b times the time 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 a 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 $m = 0, 1, \ldots, M-1$ represent the M possible transmitted symbols, each corresponding to $b = \log_2 M$ bits, the demodulator may make a Q-ary decision, where $Q > M$. 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 $Q = M$, 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 b/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 decoder 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.

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.

1.3 Communication Channels and Their Characteristics

As indicated in our 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.

The effects of noise may be minimized by increasing the power in the transmitted signal. However, equipment and other practical constraints limit the power level in the transmitted signal. 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.

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. Twisted-pair wirelines 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 kilohertz (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.

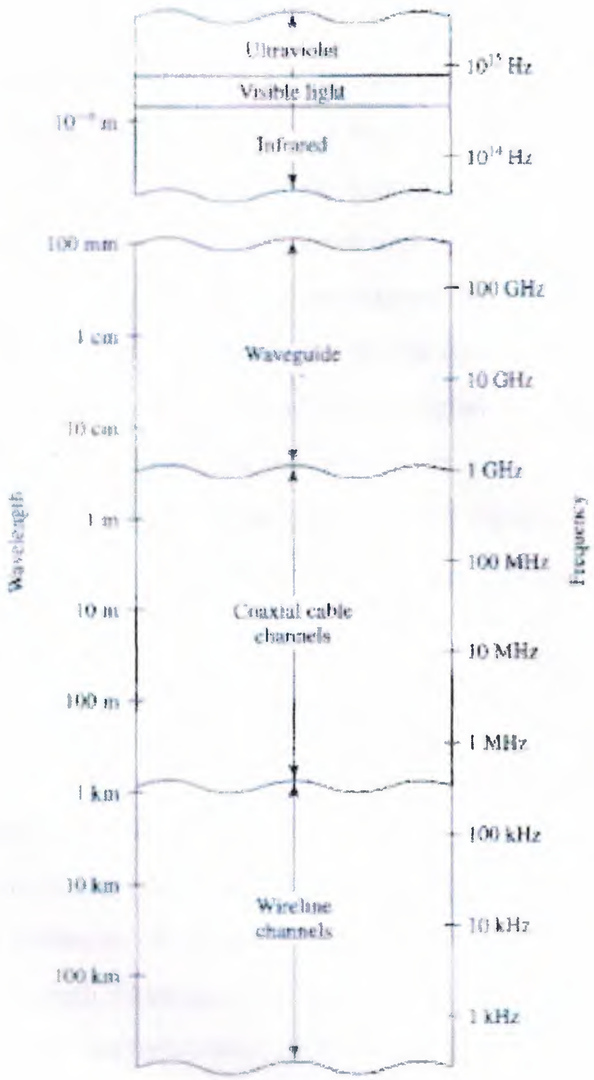

Figure 1.3 Frequency range of guided wireline channels [1].

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 which 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 transatlantic and transpacific communications. With the large bandwidth available on fiber optic channels it is possible for the 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 in the next few years.

1.3.3 Wireless Electromagnetic Channels

In radio communication systems, electromagnetic 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 electromagnetic energy, the antenna must be longer than 1/10 of the wavelength. Consequently, a radio station transmitting in the AM frequency band, say at 1 MHz (corresponding to a wavelength of $\lambda = c/f_c = 300$ m) 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 line-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 bandwidths available in these frequency bands are relatively small (usually from 1-10% 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.

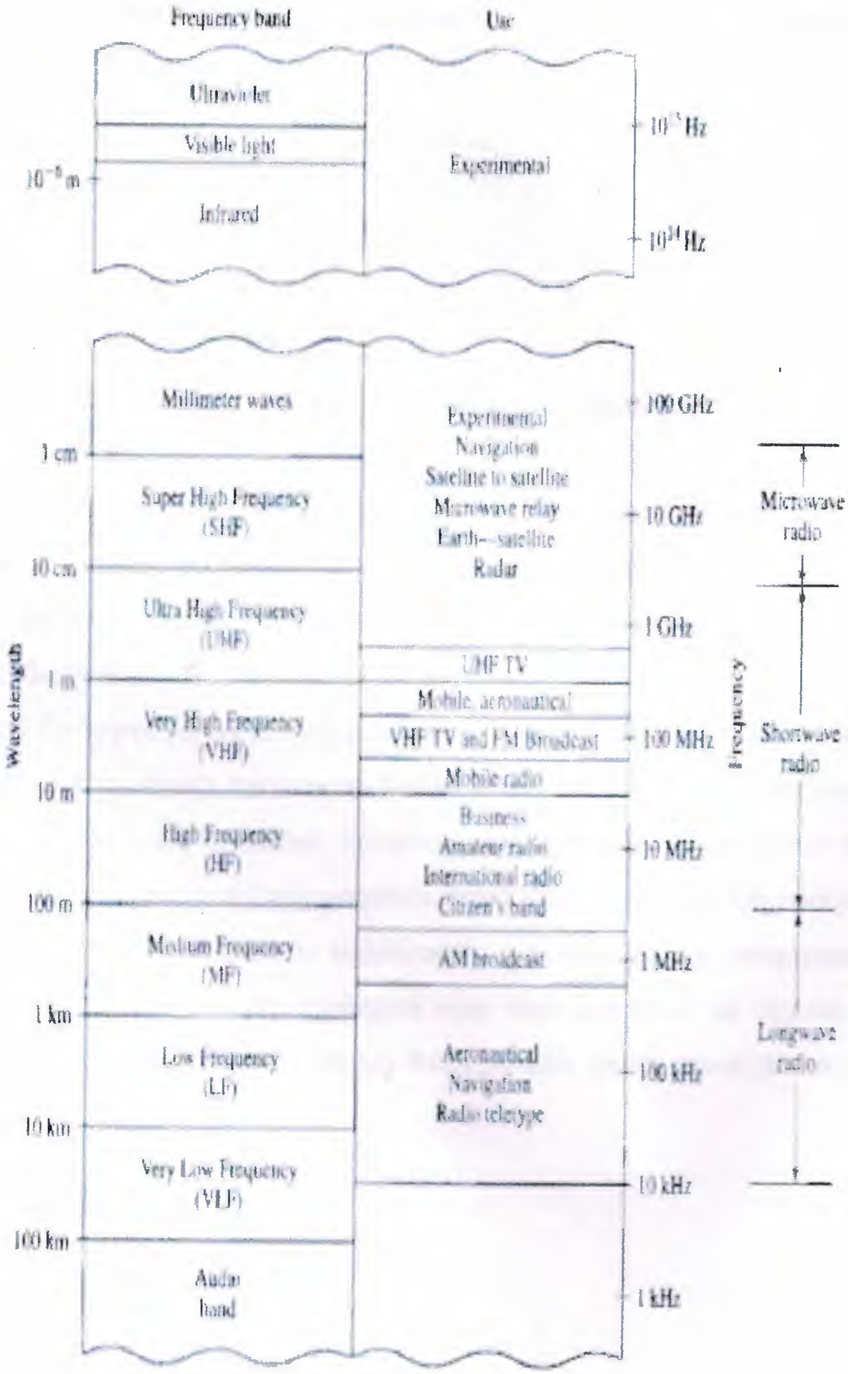

Figure 1.4 Frequency range for wireless electromagnetic channels [3].

Ground-wave propagation, illustrated in Figure 1.5, is the dominant mode of propagation for frequencies in the MF band (0.3-3 MHz). This is the frequency band used for AM broadcasting and maritime radio broadcasting. In AM broadcast, 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 electromagnetic components at the receiver are dominant disturbances for signal transmission of MF.

Figure 1.5 Illustration of ground-wave propagation [1].

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 30-250 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 serve to absorb frequencies below 2 MHz, thus, severely 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 day time 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 90-250 miles above the surface of the earth.

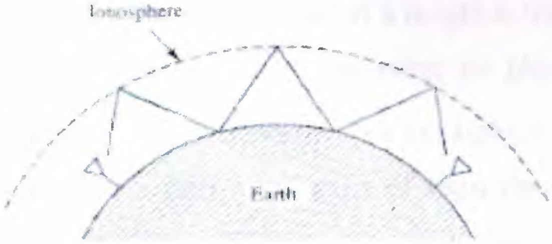

Figure 1.6 Illustration of sky – wave propagation [1].

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

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 30-60 MHz, resulting from signal scattering from the lower ionosphere. It is also possible to communicate over distances of several hundred miles by use of tropospheric scattering at frequencies in the range of 40-300 MHz. Troposcatter 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 VHF 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 in order 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 a mountains, is approximately d = $\sqrt{2h}$ miles. For example, a TV antenna mounted on a tower of 1000 ft in height provides a 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.

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 GHz, atmospheric conditions play a major role in signal propagation. Figure 1.7 illustrates the signal attenuation in dB/mile due to precipitation for frequencies in the range of 10-100 GHz. We observe that heavy rain introduces extremely high propagation losses that can result in service outages (total breakdown in the communication system).

1.3.4 Underwater Acoustic Channels

Over the past few decades, ocean exploration activity has been steadily increasing. Coupled with this increase in ocean exploration is the need to transmit data, collected by sensors placed underwater, 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 underwater, 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 $\delta = 250/\sqrt{h}$, where f is expressed in Hz and δ is in meters. For example, at 10 kHz, the skin depth is 2.5 m. In contrast, acoustic signals propagate over distances of tens and even hundreds of kilometres.

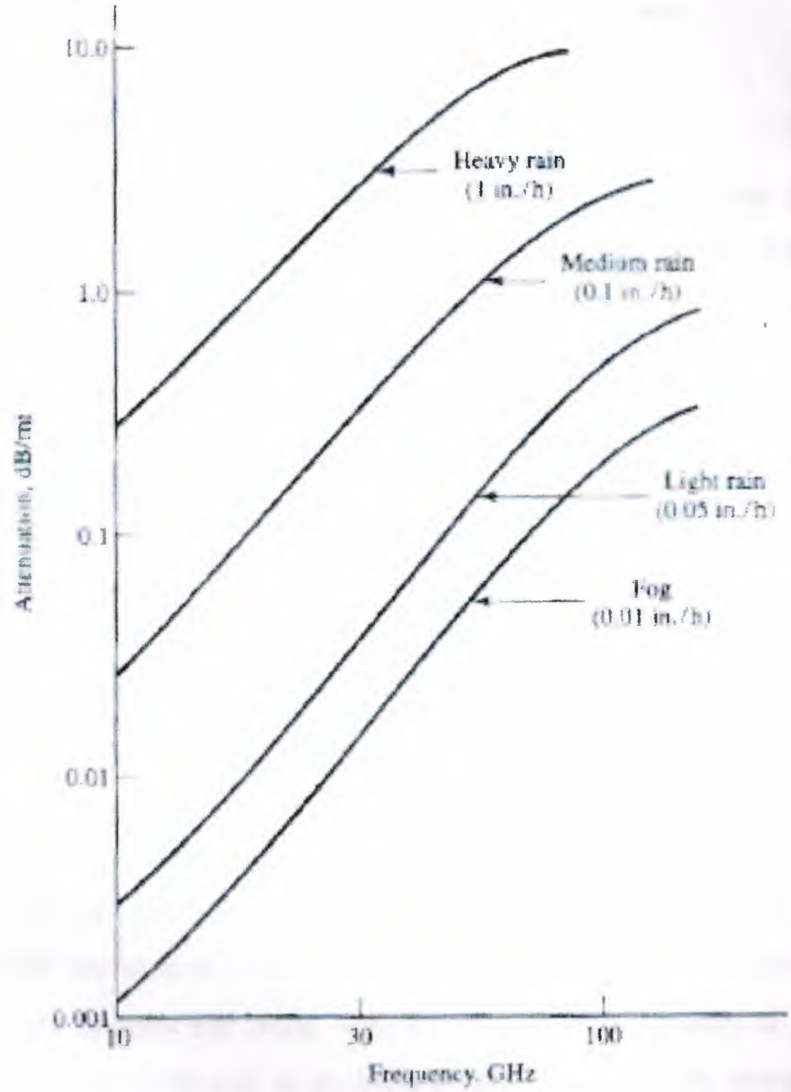

Figure 1.7 Signal attenuation due to precipitation [1].

A shallow water acoustic channel is characterized as a multipath channel due to signal reflections from the surface and the bottom of the sea. Due to 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 our 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, and optical disks used for computer data storage, music (compact disks), and video 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/sq, in 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. Next, 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.

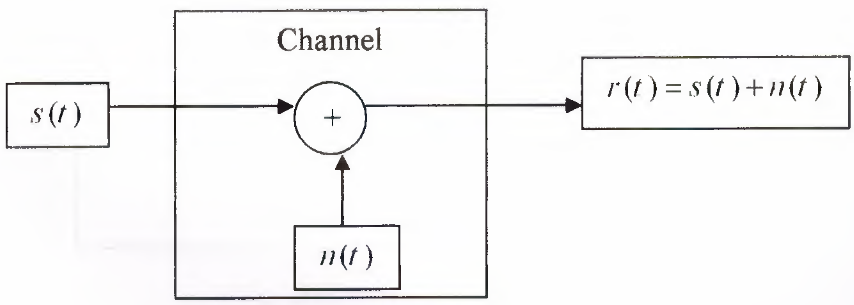

Figure 1.8 The additive noise channel [1].

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 process. Hence, the resulting mathematical model for the channel is usually called the additive Gaussian noise channel. Because this channel 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 communication system analysis and design. Channel attenuation is easily incorporated into the model. When the signal undergoes attenuation in transmission through the channel, the received signal is

$$R(t) = aS(t) + n(t) \tag{1.1}$$

where a 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

$$R(t) = S(t)*h(t)+n(t) = \int_{-\infty}^{+\infty} h(\tau)S(t-\tau)d\tau + n(t) \qquad (1.2)$$

where h(t) is the impulse response of the linear filter and * denotes convolution .

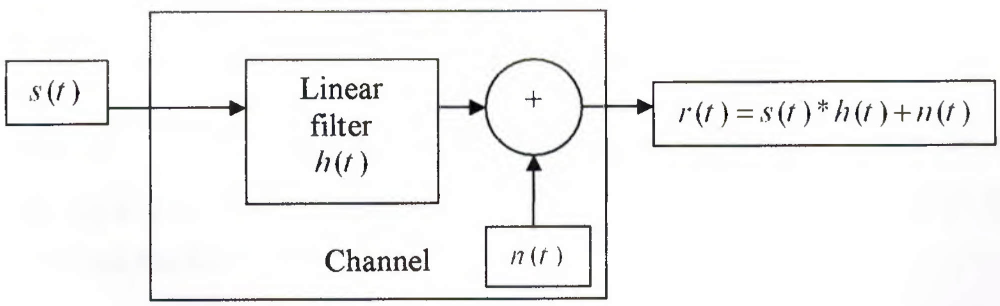

Figure 1.9 The linear filter channel with additive noise [1].

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 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 - τ. 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

$$R(t) = S(t)*h(\tau;t)+n(t) = \int_{-\infty}^{+\infty} h(\tau;t)S(t-\tau)d\tau + n(t) \qquad (1.3)$$

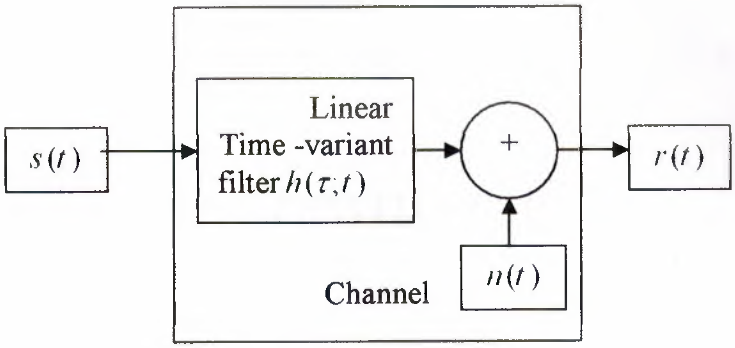

Figure 1.10 Linear time variant filter channel with additive noise [1].

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 Equation (1.3) in which the time-variant impulse response has the form

$$h(\tau;t) = \sum_{K=1}^{L} a_K(t)\delta(\tau - \tau_K) \qquad (1.4)$$

where the $\{a_K(t)\}$ represent the possibly time-variant attenuation factors for the L multipath propagation paths. If Equation (1.4) is substituted into Equation (1.3), the received signal has the form

$$R(t) = \sum_{K=1}^{L} a_K(t)s(t - \tau_K) + n(t) \qquad (1.5)$$

Hence, the received signal consists of L multipath components, where each component is attenuated by $\{a_K\}$ 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 next for the analysis and design of communication system.

CHAPTER TWO

MULTIPATH FADING

2.1 Small-Scale Multipath Propagation

Multipath in radio channel creates small-scale fading effects. The three most important effects are:

- Rapid changes in signal strength over a small travel distance or time interval.
- Random frequency modulation due to varying Doppler shifts on different multipath signals.
- Time dispersion (echoes) caused by multipath propagation delays.

In built-up urban areas, fading occurs because the height of the mobile antennas are well bellow the height of surrounding structures, so there is no single line-of-sight path to base station. Even when a line-of-sight exists, multipath still occurs due to reflection from the ground and surrounding structures. The incoming radio waves arrive from different directions with different propagation delays. The signal received by the mobile at any point in space may consist of a large number of plane waves having randomly distributed amplitudes, phase, and angels of arrival. These multipath components combine vectorially at the receiver antenna, and can cause the signal received by the mobile to distort or fade. Even when a mobile receiver is stationary, the received signal may fade due to movement of surrounding objects in the radio channel.

If object in the radio channel are static, and motion is considered to be only due to that of the mobile, then fading is purely a spatial phenomenon. The spatial variations of the resulting signal are seen as temporal variations by the receiver as it moves through the multipath field. Due to the constructive and destructive effects of multipath waves summing at various points in space, a receiver moving at high speed can pass through several fades in a small period of time. In a more serious case, a receiver may stop at a particular location at which the received signal is in a deep fade. Maintaining good communications can then become very difficult, although passing vehicles or people walking in the vicinity of the mobile can often disturb the field pattern, thereby diminishing the likelihood of the received signal remaining in a deep null for along period of time. Antenna space diversity can prevent deep fading nulls.

Due to the relative motion between the mobile and base station, each multipath wave experiences an apparent shift in frequency. The shift in received signal frequency due to motion is called the Doppler shift, and is directly proportional to the velocity and direction of motion of the mobile with respect to the direction of arrival of the received multipathe wave.

2.1.1 Factors Influencing Small-Scale Fading

Many physical factors in the radio propagation channel influence small-scale fading. This includes the following:

- Multipath propagation

The presence of reflecting objects and scatters in the channel creates a constantly changing environment that dissipates the signal energy in amplitude, phase, and time. These effects result in multiple versions of the transmitted signal that arrive at the receiving antenna, displaced with respect to one another in and special orientation. The random phase and amplitude of different multipath components cause fluctuation in signal strength, thereby inducing small-scale fading, signal distortion, or both. Multipath propagation often lengthens the time required for the baseband portion of the signal to reach the receiver which can cause signal smearing due to intersymbol interference.

- Speed of the mobile

The relative motion between the base station and the mobile results in random frequency modulation due to different Doppler shifts on each of the multipath components. Doppler shift will be positive or negative depending on the whether the mobile receiver is moving toward or a way from base station.

- Speed of surrounding objects

If objects in the radio channel are in motion, they induce a time varying Doppler shift on multipath components. If the surrounding objects move at greater rate than the mobile, then this effect dominates the small-scale fading. Otherwise, motion of surrounding objects maybe ignored, and only the speed of the mobile need be considered. The coherence time defines the "staticness" of the channel, and is directly impacted by the Doppler shift.

- The transmission bandwidth of the signal

If the transmitted radio signal bandwidth is greater than the "bandwidth" of multipath channel, the received signal will be distorted, but the received signal strength will not fade much over a local area (the small scale signal fading will not be significant). As will be shown, the bandwidth of the channel can be quantified by the coherence bandwidth which is related to the specific multipath structure of the channel. The coherence bandwidth is measure of the maximum frequency difference for which signal are still strongly correlated in amplitude. If the transmitted signal has a narrow bandwidth as compared to the channel, the amplitude of the signal will change rapidly, but the signal will not be distorted in time. Thus, the statistics of small-scale signal strength and the likelihood of signal smearing appearing over small-scale distances are very much related to the specific amplitudes and delays of the multipath channel, as well as the bandwidth of transmitted signal.

2.1.2 Doppler Shift

When a wave source and a receiver are moving relative to one another the frequency of the received signal will not be the as the source, when they are moving toward each other the frequency of the received signal is higher then the source, and when they are approaching each other the frequency decreases. This is called the Doppler effect, this effect becomes important when developing mobile radio system.

The amount the frequency changes due to the Doppler shift depends on the relative motion between the source and the receiver, and on the speed of the propagation of the wave, the Doppler shift in frequency can be written

$$\Delta f \cong \pm f_o \frac{v}{c} \tag{2.1}$$

where Δf is the change in the frequency of the source seen at the receiver, "f_o" is the frequency of the source, "v" is the speed difference between the source and the transmitter, and "c" is the speed of light.

Doppler shift can cause significant problems if the transmission technique is sensitive to carrier frequency offsets, or the relative speed is very high as is the case for low earth orbiting satellite.

2.2 Parameters of Mobile Multipath Channels

Many multipath channel parameters are derived from the power delay profile. Power delay profiles are measured using the techniques discussed in Section 2.2 and are generally represented as plots of relative received power as a function of excess delay with respect to a fixed time delay reference. Power delay profiles are found by averaging instantaneous power delay profile measurements over a local area in order to determine an average small-scale power delay profile. Depending on the time resolution of the probing pulse and the type of multipath channels studied, researchers often choose to sample at spatial separations of a quarter of a wavelength and over receiver movements no greater than 6 m in outdoor channels and no greater than 2 m in indoor channels in the 450 MHz-6 GHz range. This small-scale sampling avoids large-scale averaging bias in the resulting small-scale statistics.

2.2.1 Time Dispersion Parameters

In order to compare different multipath channels and to develop some general design guidelines for wireless systems, parameters which grossly quantify the multipath channel are used. The mean excess delay, rms delay spread, and excess delay speard (X dB) are multipath channel parameters that can be determined from a power delay profile. The time dispersive properties of wide band multipath are most commonly quantified by their mean excess delay ($\bar{\tau}$) and rms delay spread (σ_τ). The mean excess is the first moment of the power delay profile an is defined to be

$$\bar{\tau} = \frac{\sum_k a_k^2 \tau_k}{\sum_k a_k^2} = \frac{\sum_k P(\tau_k)\tau_k}{\sum_k P(\tau_k)} \tag{2.2}$$

The rms delay spread is the square root of second central moment of the power delay profile and is defined to be

$$\sigma_\tau = \sqrt{\overline{\tau^2} - (\bar{\tau})^2} \tag{2.3}$$

where

$$\overline{\tau^2} = \frac{\sum_k a_k^2 \tau_k^2}{\sum_k a_k^2} = \frac{\sum_k P(\tau_k)\tau_k^2}{\sum_k P(\tau_k)} \tag{2.4}$$

These delays are measured relative to first detectable signal arriving at the receiver at $\tau_0 = 0$. Equation (2.2)-(2.3) do not rely on the absolute power level of P(τ), but only the relative amplitudes of the multipath components within P(τ). Typical values of rms delay spread are on the order of microsecond in outdoor mobile radio channels and on the order of nanoseconds in indoor radio channels.

It is important to note that the rms delay spread and mean excess delay are defined from a single power delay profile which is temporal or spatial average of consecutive impulse response measurements collected and averaged over a local area. Typically, many measurements are made at many local areas in order to determine a statistical range of multipath channel parameters for a mobile communication system over a large scale-area .

The maximum excess delay (X dB) of the power delay profile is defined to the time delay during which multipath energy falls to X dB below the maximum. In other words, the maximum excess delay is defined as $\tau_x - \tau_0$, where τ_0 is the first arriving signal and τ_x is the maximum delay at which a multipath component is within X dB of the strongest arriving multipath signal (which dose not necessarily arrive at τ_0). Figure 2.1 illustrates the computation of the maximum excess delay for multipath components within 10 dB of the maximum. The maximum excess delay (X dB) defines the temporal extent of the multipath that is above a particular threshold. The value of τ_x is sometimes called the excess delay spread of a power delay profile, but in all cases must be specified with a threshold that relates the multipath noise floor to the maximum received multipah component.

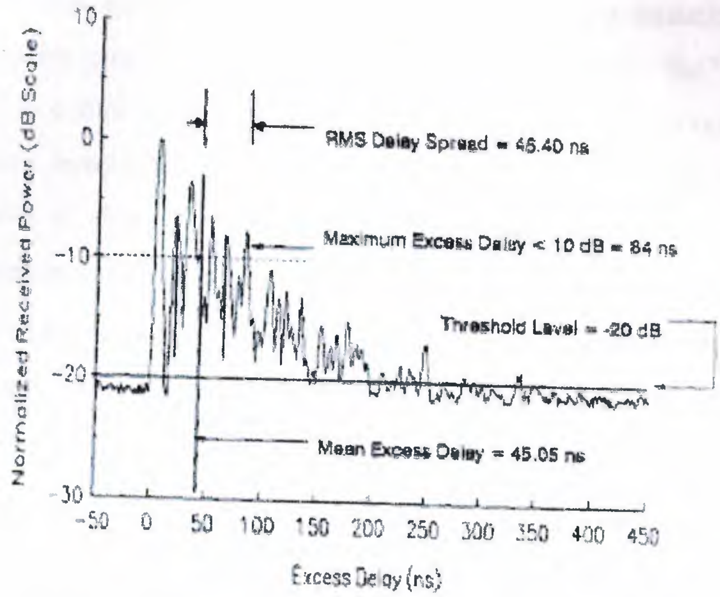

Figure 2.1 Example of an indoor power delay profile [2].

In practice, values for $\bar{\tau}$, $\overline{\tau^2}$ and σ_τ depend on the choice of the noise used to process P(τ). The noise threshold is used to differentiate between received multipath components and thermal noise. If the noise threshold is set too low, then noise will be processed as multipath, thus giving rise to values of $\bar{\tau}$, $\overline{\tau^2}$, and σ_τ that are artificially high.

It should be noted that the power delay profile and the magnitude frequency response (the spectral response) of a mobile radio channel are related through the Fourier transform. It is therefore possible to obtain to an equivalent description of the channel in the frequency domain using its frequency response characteristic. Analogous to the delay spread parameters in the time domain, coherence bandwidth is used to characterize the channel in the frequency domain. The rms delay spread and coherence bandwidth are inversely proportional to one another, although their exact relationship is a function of the exact multipath structure.

2.2.2 Coherence Bandwidth

While the delay spread is a natural phenomenon caused by reflected and scattered propagation paths in the radio channel, the coherence bandwidth, B_c, is a defined relation derived from the rms delay spread. Coherence bandwidth is a statistical measure of the range of frequencies over which the channel can be considered "flat" (a channel which passes all spectral components with approximately equal gain and linear phase). In other words, coherence bandwidth is the range of frequencies over which two frequency components have a strong potential for amplitude correlation. Two sinusoids with frequency separation greater than B_c are affected quite differently by the channel. If the coherence bandwidth is defined as the bandwidth over which the frequency correlation function is above 0.9, then the coherence bandwidth is approximately.

$$B_c \approx \frac{1}{50\sigma_\tau} \tag{2.5}$$

If the definition is relaxed so that the frequency correlation function is above 0.5, then the coherence bandwidth I approximately

$$B_c \approx \frac{1}{5\sigma_\tau} \tag{2.6}$$

It is important to note that an exact relationship between coherence bandwidth and rms delay spread is a function of specific channel impulse responses and applied signals, and Equations (2.5) and (2.6) are "ball park estimates." In general, spectral analysis techniques and simulation are required to determine the exact impact that time varying multipath has on a particular transmitted signal. For this reason, accurate multipath channel models must be used in the design of specific modems for wireless applications.

2.2.3 Doppler Spread and Coherence Time

Delay spread and coherence bandwidth are parameters which describe the time dispersive nature of the channel in a local area. However, they do not offer information about the time varying nature of the channel caused by either relative motion between the mobile and base station, or by movement of objects in the channel. Doppler spread and coherence time are parameters which describe the time varying nature of the channel in a small-scale region.

Doppler spread B_D is a measure of the spectral broadening caused by the time rate of change of the mobile radio channel and is defined as the range of frequencies over which the received Doppler spectrum is essentially non-zero. When a pure sinusoidal tone of frequency f_c is transmitted, the received signal spectrum, called the Doppler spectrum, will have components in the range $f_c - f_d$ to $f_c + f_d$, where f_d is the Doppler shift. The amount of spectral broadening depends on f_d which is function of the relative velocity of the mobile, an the angle θ between the direction of the motion of the mobile and direction of the arrival of the scattered waves. If the baseband signal bandwidth is much greater than B_D, the effects of Doppler spread are negligible at the receiver. This is a slow fading channel.

Coherence time T_C is the time domain dual of Doppler spread and is used to characterize the time varying nature of the frequency dispersiveness of the channel in the time domain. The Doppler spread and coherence time are inversely proportional to one anther. That is,

$$T_C \approx \frac{1}{f_m} \tag{2.7}$$

Coherence time is actually a statistical measure of the duration over which the channel impulse response is essentially invariant, and quantifies the similarity of the channel response at different times. In other words, coherence time is the time duration over which two received signals have strong potential for amplitude correlation. If the reciprocal bandwidth of the baseband signal is the greater than the coherence time of the channel, then the channel will change during the transmission of the baseband message, thus causing distortion at the receiver. If the coherence time is defined as the over which the time correlation function is above 0.5, then the coherence time is approximately

$$T_C \approx \frac{9}{16\pi f_m} \tag{2.8}$$

Where f_m is the maximum Doppler shift given by $f_m = v / \lambda$. In practice, (2.7) suggest a time duration during which a Rayleigh fading signal may fluctuate widely, and (2.8) is often too restrictive. A popular rule of thumb for modern digital communications is to define the coherence time as the geometric mean of Equations (2.7) and (2.8). That is,

$$T_C = \sqrt{\frac{9}{16\pi f_m^2}} = \frac{0.423}{f_m} \tag{2.9}$$

The definition of coherence time implies that two signals arriving with a time separation greater than T_C are affected differently by the channel. For example, for a vehicle traveling 60 mph using a 900 MHz carrier, a conservative value of T_C can be shown to be 2.22 ms from Equation (2.8). If a digital transmission system is used, then as a long as the symbol rate is greater than $1/T_C = 454$ bps, the channel will not cause distortion due to motion (however, distortion could result from multipath time delay spread, depending on the channel impulse response). Using the practical formula of (2.9), $T_C = 6.77$ ms and the symbol rate must exceed 150 bit/s in order to avoid distortion due to frequency dispersion.

2.3 Types of Small-Scale Fading

The previous section demonstrated signal propagating through a mobile radio channel depends on the nature of the transmitted signal with respect to the characteristics of the channel. Depending on the relation between the signal parameters (such as bandwidth, symbol period, etc.) and the channel parameters (such as rms delay spread and Doppler spread), different transmitted signal will undergo different types of fading.

The time dispersion and frequency dispersion mechanisms in a mobile radio channel lead to four possible distinct effects, which are manifested depending on the nature of the *transmitted signal, the channel, and the velocity. While multipath delay spread leads to time* dispersion and frequency selective fading, Doppler spread leads to frequency dispersion and time selective fading. The two propagation mechanisms are independent of one another. Figure 2.2 shows a tree of the four different types of fading.

2.3.1 Fading Effects Due to Multipath Time Delay Spread

Time dispersion due to multipath causes the transmitted signals to undergo either flat or frequency selective fading.

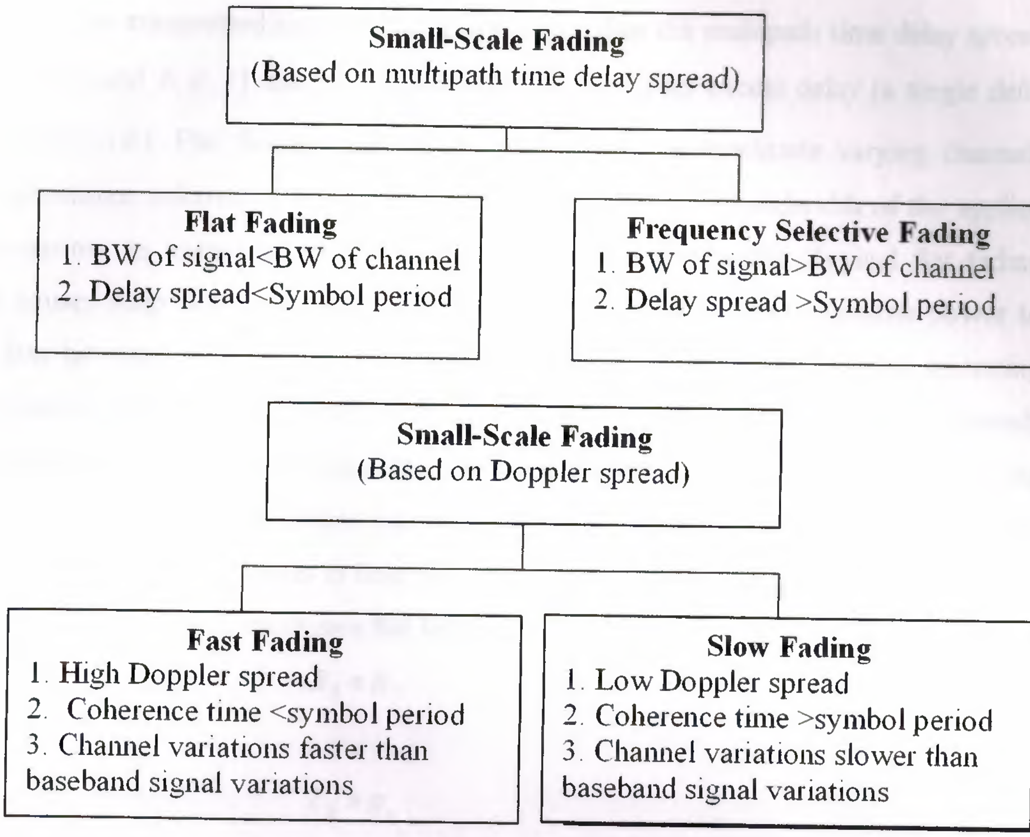

Figure 2.2 Types of small-scale fading [2].

2.3.1.1 Flat Fading

If the mobile radio channel has a constant gain and linear phase response over a bandwidth which is greater than the bandwidth of the transmitted signal, then the received signal will undergo flat fading. This type of fading is historically the most common type of fading described in the technical literature. In flat fading, the multipath structure of the channel is such that the spectral characteristics of the transmitted signals are preserved at the receiver. However the strength of the received signals changes with time, due to fluctuations in the gain of the channel caused by multipath. The characteristic of the flat fading channel are illustrated in figure 2.3.

It can be seen from figure 2.3 that if the channel gain changes over time, a change of amplitude occurs in the received signal. Over time, the received signal R(t) varies in gain, but the spectrum of the transmission is preserved. In a flat fading channel, the reciprocal bandwidth of the transmitted signal is the much larger than the multipath time delay spread of the channel, and $h_b(t,\tau)$ can be approximated as having no excess delay (a single delta function with $\tau = 0$). Flat fading channels are also known as amplitude varying channels and are sometimes referred to as narrowband channels, since the bandwidth of the applied signal is narrow as compared to the channel flat fading bandwidth. Typical flat fading channels causes deep fades, and thus may require 20 or 30 dB more transmitter power to achieve low bit error rates during times of deep fades as compared to systems operating over non-fading channels. The distribution of the instantaneous gain of flat fading channels is important for designing radio links, and the most common amplitude distribution is the Rayleigh distribution. The Rayleigh flat fading channel model assumes that the channel induces an amplitude which varies in time according to the Rayleigh distribution.

To summarize, a signal undergoes flat fading if

$$B_S \ll B_C \tag{2.10}$$

and

$$T_S \gg \sigma_\tau \tag{2.11}$$

where T_S is the reciprocal bandwidth (symbol period) and B_S is the bandwidth, respectively, of transmitted modulation, and σ_τ and B_C are the rms delay spread and coherence bandwidth, respectively, of the channel.

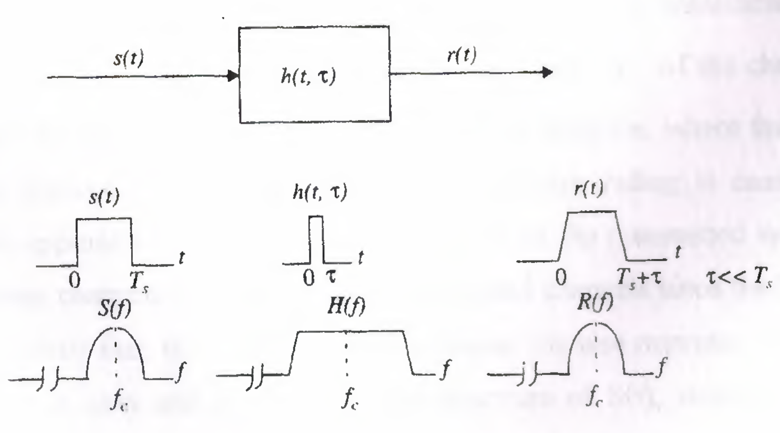

Figure 2.3 Flat fading channel characteristics [2].

2.3.1.2 Frequency Selective Fading

If the the channel possesses a constant-gain and linear phase response over a bandwidth that is smaller then the bandwidth of transmitted signal, then the channel creates frequency selective fading on the received signal. Under such conditions, the channel impulse response has a multipath delay spread which greater than the reciprocal bandwidth of transmitted message waveform. When this occurs, the received signal includes multiple versions of the transmitted waveform which are attenuated (faded) and delayed in time, and hence the received signal is distorted. Frequency selective fading is due to time dispersion of the transmitted symbols with in the channel. Thus the channel induces intersymbol interference (ISI). Viewed in the frequency domain, certain frequency components in the received signal spectrum have greater gains then others.

Frequency selective fading channels are much more difficult to model than flat fading channels since each multipath signal must be modeled and the channel must be considered to be a linear filter. It is for this reason that wideband multipath measurements are made, and models are developed from this measurements. When analyzing mobile communication systems, statistical impulse response models such as the two-ray Rayleigh fading model (which considers the impulse response to be made up of two delta functions which independently fade and have sufficient time delay between them to induce frequency selective fading upon the applied signal), or computer generated or measured impulse responses, are generally used for analyzing frequency selective small-scale fading. Figure 2.4 illustrates the characteristics of a frequency selective fading channel.

For frequency selective fading, the spectrum $S(f)$ of transmitted signal has a bandwidth which is greater than the coherence bandwidth B_C of the channel. Viewed in the frequency domain, the channel becomes frequency selective, where the gain is different for different frequency components. Frequency selective fading is caused by multipath delays which approach or exceed the symbol period of the transmitted symbol. Frequency selective fading channels are also known as wideband channels since the bandwidth of the signal S(t) is wider than the bandwidth of the channel impulse response. As time varies, the channel varies in gain and phase across the spectrum of S(t), resulting in time varying distortion in the received signal r(t). To summarize, a signal undergoes frequency selective fading if

$$B_S > B_C \tag{2.11}$$

and

$$T_S < \sigma_\tau \tag{2.12}$$

A common rule of thumb is that a channel is flat fading if $T_S \geq 10\sigma_\tau$ and a channel is frequency selective if $T_S < 10\sigma_\tau$, although this dependent on the specific type of modulation used.

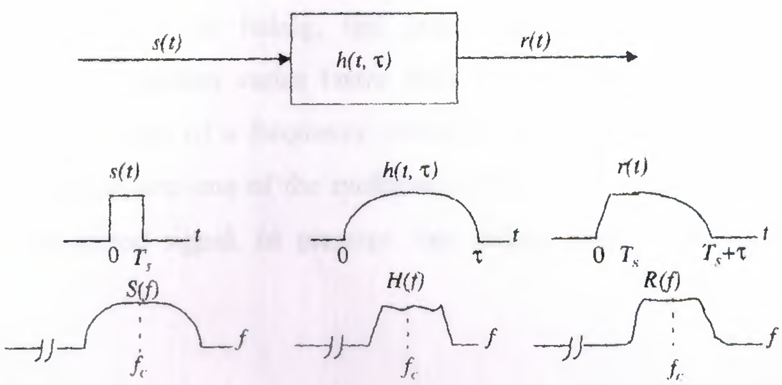

Figure 2.4 Frequency selective fading channel characteristics [2].

2.3.2 Fading Effects Due to Doppler Spread

2.3.2.1 Fast Fading

Depending on how rapidly the transmitted baseband signal changes as compared to the rate of change of channel, a channel may be classified either as a fast fading or slow fading channel. In fast fading channel, the channel impulse response changes rapidly within the symbol duration. That is, the coherence time of the channel is smaller than the symbol period of transmitted signal. This causes frequency dispersion (also called time selective fading) due to Doppler spreading, which leads to signal distortion. Viewed in the frequency domain, signal distortion due to fast fading increases with increasing Doppler spread relative to the bandwidth of the transmitted signal. Therefore, a signal undergoes fast fading if

$$T_S > T_C \tag{2.13}$$

and

$$B_S < B_D \tag{2.14}$$

It should be noted that when a channel is specified as a fast or slow fading channel, it does not specify whether the channel is flat fading or frequency selective in nature. Fast fading only deals with the rate of change of the channel due to motion. In the case of the flat fading channel, we can approximate the impulse response to be simply a delta function (no time delay). Hence, a flat fading, fast fading channel is a channel in which the amplitude of the delta function varies faster then the rate of change of the transmitted baseband signal. In the case of a frequency selective, fast fading channel, the amplitudes, phase and time delays of any one of the multipath components vary faster than the rate of change of the transmitted signal. In practice, fast fading only occurs for very low data rates.

2.3.2.2 Slow Fading

In a slow fading channel, the channel impulse response changes at a rate much slower than the transmitted baseband signal S(t). In this case, the channel may be assumed to be static over one or several reciprocal bandwidth intervals. In the frequency domain, this implies that the Doppler spread of the channel is much less then the bandwidth of the baseband signal. Therefore, a signal undergoes slow fading if

$$T_S \ll T_C \tag{2.15}$$

and

$$B_S \gg B_D \tag{2.16}$$

It should be clear that velocity of the mobile (or velocity of objects in the channel) and the baseband signaling determines whether a signal undergoes fast fading or slow fading.

The relation between the various multipath parameters and the type of fading experienced by the signal are summarized in figure 2.5. It should be emphasized that fast and slow fading deal with relationship between the time rate of the changes in the channel and transmitted signal, and not with propagation path loss models.

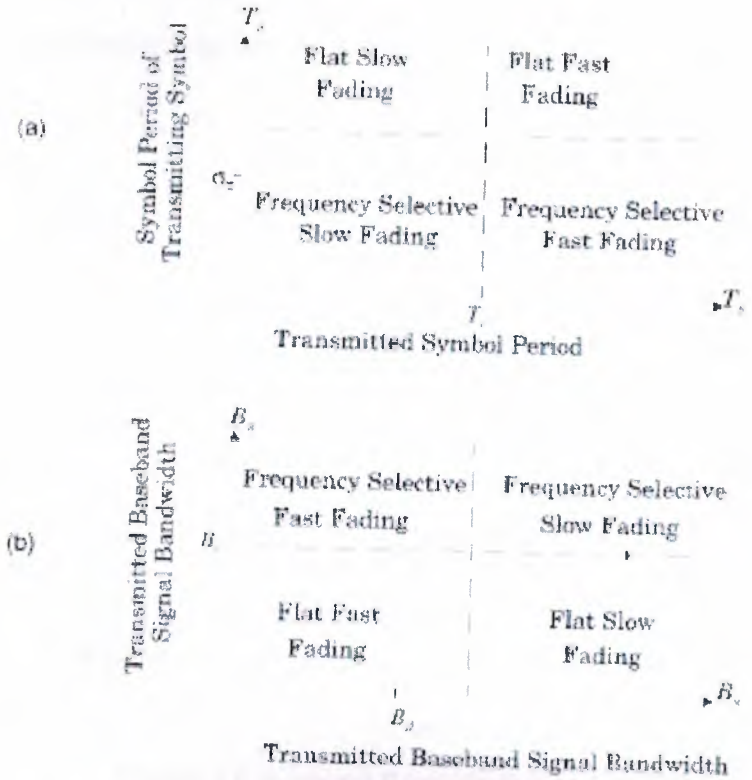

Figure 2.5 fading as function of a) symbol period b) base band signal bandwidth

2.4 Rayleigh and Ricean Distribution

2.4.1 Rayleigh Fading Distribution

In mobile radio channels, the Rayleigh distribution is commonly used to describe the statistical time varying nature of the received envelope of a flat fading signal, or the envelope of an individual multipath component. It is well known that the envelope of the sum of two quadrature Gaussian noise signals obeys a Rayleigh distribution. Figure 2.6 shows a Rayleigh distributed signal envelope as a function of time. The Rayleigh distribution has a probability density function (pdf) given by

$$p(r) = \begin{cases} \dfrac{r}{\sigma_2} \exp\left(-\dfrac{r^2}{2\sigma^2} \right) & (0 \le r \le \infty) \\ 0 & (r < 0) \end{cases} \tag{2.6}$$

where σ is the rms value of the received voltage signal before envelope detection, and σ^2 is the time-average power of the received signal before detection.

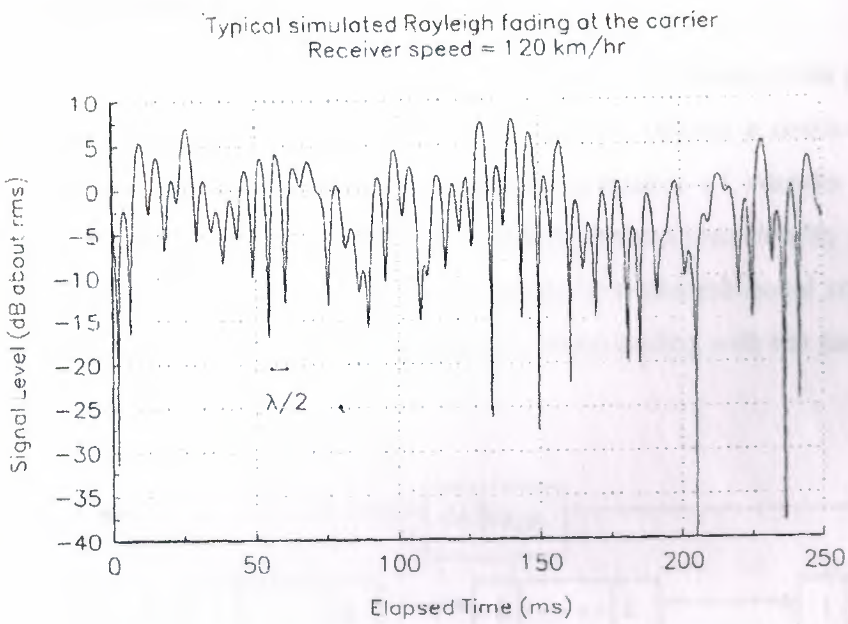

Figure 2.6 A typical Rayleigh fading envelop at 900 MHz

2.4.2 Ricean Fading Distribution

When there is a dominant stationary (nonfading) signal component present, such as line-of-sight propagation path, the small scale fading envelope distribution is Ricean. In such situation, random multipath components arriving at different angels are superimposed on a stationary dominant signal. At the output of an envelope detector, this has the effect of adding a dc component to the random multipath.

Just as for the case of detection of a sine wave in thermal noise, the effect of a dominant signal arriving with many weaker multipath signals gives rise to the Ricean distribution. As the dominant signal becomes weaker, the composite signal resembles a noise signal which has an envelope that is Rayleigh. Thus, the Ricean distribution degenerates to a Rayleigh distribution when the dominant component fades away.

CHAPTER THREE

CONVOLUTIONAL CODING

3.1 Convolutional Codes

Convolutional codes are fundamentally different from block codes in that information sequences are not grouped into distinct and encoded. Instead a continuous sequences of information bits is mapped into a continuous sequence of encoder output bits. This mapping is highly structured, enabling a decoding method considerably different from that of block codes to be employed. It can be argued that convolutional coding can achive a larger coding gain than can be achieved using a block coding with the same complexity [6].

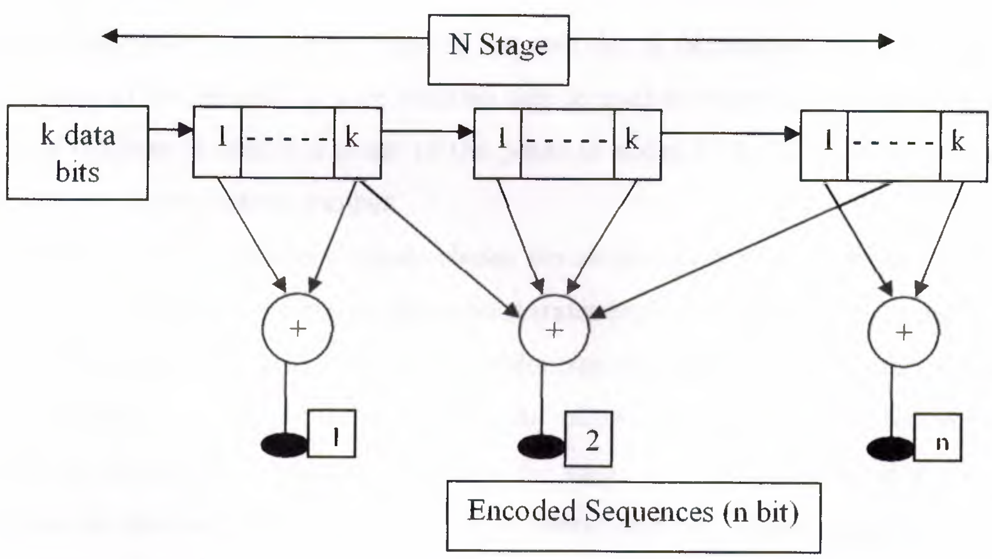

Figure 3.1 General block diagram of convolutional encoder [2].

A convolutional code is generated by passing the information sequence through a finite state shift register. In general, the shift register contains N k-bits and m linear algebraic function generators based on the generator polynomials as shown in figure 4.1. The input data is shifted into and along the shift register, k bits at time. The number of output bits for each k bit user input data sequence is n bits. The code rate $R_c = k / n$. The parameters N is called the constraint length and indicates the number of input data bits that the current output is dependent upon. The constraint length determines how powerful and the complex the code is. Following is an outline of the various ways of representing convolutional codes.

Generator matrix – The generator matrix for aconvolitional code can define the code and is semi-infinite since the input is semi-infinite in length. Hence, this is not convenient way of representing a convolutional code.

Generator polynomials – For convolitional codes, we specify a set of n vectors, one for each of the n module-2 adder. A 1 in the i th position of the vector indicates that the corresponding shift register stage is connected and a 0 indicates no connection.

Logic table – A logic table or lookup table can be built showing the outputs of the convolutional encoder and the state of the encoder for all of the specific input sequences present in the shift register.

State diagram – since the output of the encoder is determined by the input and the current state of the encoder, a state diagram can be used to represent the encoding process. The state diagram is simply a graph of the possible states of the encoder and the possible transitions from one state to another.

Tree diagram – The tree diagram shows the structure of the encoder in the form of a tree with the branches representing the various states and the outputs of the coder.

Trellis diagram – Close observation of the tree reveals that the structure repeat itself once the number of stage is greater than the constraint length. It is observed that all branches emanating from two nodes having the same state are identical in the sense that they generate identical output sequences. This means that two nodes having the same label can be merged. By doing this throughout the tree diagram, we can obtain another diagram called a trellis diagram which is a more compact representation.

3.1.1 Decoding of Convolutional Codes

The function of the decoder is to estimate the encoded input information using a rule or method that results in the minimum possible number of errors. There is a one-to-one correspondence between the information sequence and the code sequence. Further, any information and a code sequence pair is uniquely associated with a path through the trellis. Thus, the job of the convolutional decoder is to estimate the path through the trellis. Thus, the job of the convolutional decoder is to estimate the path through the trellis that was followed by the encoder.

There are a number of techniques for decoding convolutional codes. The most important of these methods is the viterbi algorithm which performs maximum likelihood decoding of convolutional codes. Both soft and hard decision decoding can be implemented for convolutional codes. Soft decision decoding is superior by about 2 – 3 dB [7].

3.1.1.1 The Viterbi Algorithm

The viterbi algorithm can be described as follows:

Let the trellis node corresponding to state S_j at time i be denoted $S_{j,i}$. Each node in the trellis is to be assigned a value V($S_{j,i}$) based on a metric. The node values are computed in the following manner.

1. Set V($S_{o,o}$) = 0 and i =1.
2. At time i, compute the partial path metrics for all paths entering each node.
3. Set V($S_{j,i}$) equal to smallest partial path metric entering the node corresponding to state S_j at time i. Ties can be broken by previous node choosing a path randomly. The nonsurviving branches are deleted from the trellis. In this manner, a group of minimum paths is created from $S_{o,o}$.
4. If i< L + m, where L is the number of input code segments (k bits for each segment) and m is the length of the longest shift register in the encoder, let i = i + 1 and go back to step 2.

Once all node values been computed, start at state S_o, time i = L + m, and follow the surviving branches backward through the trellis. The path thus defined is unique and corresponds to the decode output. When hard decision decoding is performed, the metric used is the Hamming distance, while the Eculidean distance is used for soft decision decoding.

3.1.1.2 Other Decoding Algorithms for Convolutional Codes

3.1.1.2.1 Fano's Sequential Decoding

Fano's algorithm searches for the most probable path through the trellis by examining one path at a time. The increment added to the metric along each branch is proportional to the probability of the received signal for that branch, as in Viterbi decoding, with the exception that an additional negative constant is added to each branch metric. The value of this constant is selected such that the metric for correct path will increase on the average while the metric for any incorrect path will decrease on the average. By comparing the metric of candidate path with an increasing threshold, the algorithm detects and discards incorrect paths. The error rate performance of the Fano algorithm is comparable to that of Vitrebi decoding. However, in comparison to Viterbi, in comparison to Viterbi decoding, sequential decoding has a significantly larger delay. Its advantage over is that it requires less storage, and thus codes with larger constraint lengths can be employed [8].

3.1.1.2.2 The Stack Algorithm

In contrast to the Viterbi algorithm which keeps track of $2^{(N-1)k}$ paths and their corresponding metrics, the stack algorithm deals with fewer paths and their corresponding metrics. In a stack algorithm, the more probable paths are ordered according to their metrics, with the path at the top of the stack is extended by one branch. This yields 2^k successors and their metrics. These 2^k successors along with the other paths are then reordered according to the values of the metrics, and all paths with metrics that fall below some preselected amount from the metric of the top path may be discarded. Then the process of extending the path with the largest metric is repeated. In comparison with Viterbi decoding, the stack algorithm requires fewer metric calculations, but this computational savings is offset to a large extent by the computations involved in reordering the stack after every iteration. In comparison with the Fano algorithm, the stack algorithm is computationally simpler since there is no retracing over the same path, but on the other hand, the stack algorithm requires more storage then the Fano algorithm.

3.1.1.2.3 Feedback Decoding

Here, the decoder makes a hard decision on the information bit at stage j based on the metrics computed from stage j to stage j + m, where m is preselected positive integer. Thus, the decision on whether the information bit was 1 or a 0 depends on whether the minimum Hamming distance path which beings at stage j and ends at stage j + m contains 0 or 1 in the branch emanating from stage j. Once a decision is made, only that part of the tree which stems from the bit selected at stage j is kept and the reminder is discarded. This is the feedback feature of the decoder. The next step is to extend the part of the tree that has survived to stage j + 1 + m and to consider the paths from stage j + 1 to j + 1 + m in deciding on the bit at stage j + 1. This procedure is repeated at every stage. The parameter m is simply the number of stage in the tree that the decoder looks ahead before making a hard decision. Instead of computing the metrics, the feedback decoder can be implemented by computing the syndrome from the received sequence and using a table look up method to correct errors. For some convolutional codes, the feedback decoder can be simplified to majority logic decoder or a threshold decoder.

4.2 Coding Gain

The advantage for error control codes, whether they block codes or convolutional codes, is that they provide a coding gain for the communications link. The coding gain describes how much better the user's decode message performs as compared to the raw bit error performance of the coded transmission within the channel. Coding gain is what allows a channel error rate of 10^{-2} to support decoded user data rates which are 10^{-5} or better.

Each error control code has a particular coding gain, which depends on the particular code, the decoder implementation, and the channel BER probability, p_c. It can be shown that a good approximation for the decode message error probability, P_B is given by

$$P_B \cong \frac{1}{n}\sum_{i=t+1}^{n} i \binom{n}{i} P_c^{\,i}(1-P_c)^{n-i} \qquad (4.1)$$

where t denotes the number of errors that can be corrected in an (n, k) block code. Thus, given known channel BER, it is possible to determine the user's decoded message error rate easily. The coding gain measures the amount of additional SNR that would be required to provide the same BER performance for uncoded message signal in the same channel conditions.

CHAPTER FOUR

RESULTS

4.1 Additive White Gaussian Noise (AWGN) Channel Result

4.1.1 Simulink Model of AWGN Channel using BPSK Modulation

Figure 4.1 shows the block diagram of the model that is used for simulation of single carrier communications over AWGN.

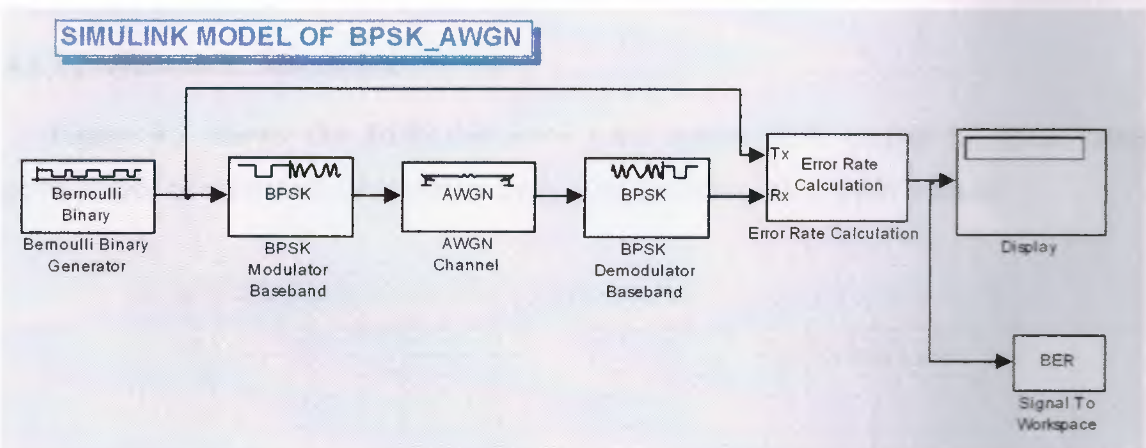

Figure 4.1 Simulink model of AWGN channel.

- **Bernoulli Binary Generator** - The Bernoulli Binary Generator block generates random binary numbers using a Bernoulli distribution.
- **BPSK (Binary Phase Shift Keying) Modulator Baseband** – The BPSK Modulator Baseband block modulates using the binary phase shift keying method. The output is a baseband representation of the modulated signal.
- **AWGN Cannel** - The AWGN Channel block adds white Gaussian noise to a real or complex input signal. When the input signal is real, this block adds real Gaussian noise and produces a real output signal. When the input signal is complex, this block adds

complex Gaussian noise and produces a complex output signal.This block inherits its sample time from the input signal.

- **BPSK Demodulator Baseband** - The BPSK Demodulator Baseband block demodulates a signal that was modulated using the binary phase shift keying method. The input is a baseband representation of the modulated signal.
- **Error Rate Calculations -** The Error Rate Calculation block compares input data from a transmitter with input data from a receiver. It calculates the error rate as a runningstatistic, by dividing the total number of unequal pairs of data elements by the total number of input data elements from one source.
- **Signal to Workspace -** The To Workspace block writes its input to the workspace.
- **Display** - The Display block shows the value of its input on its icon.

4.1.2 Performance Simulation

Figure 4.2 shows the BER (bit error rate) versus SNR (signal to noise ratio) performance of communication system using a single carrier on AWGN channel.

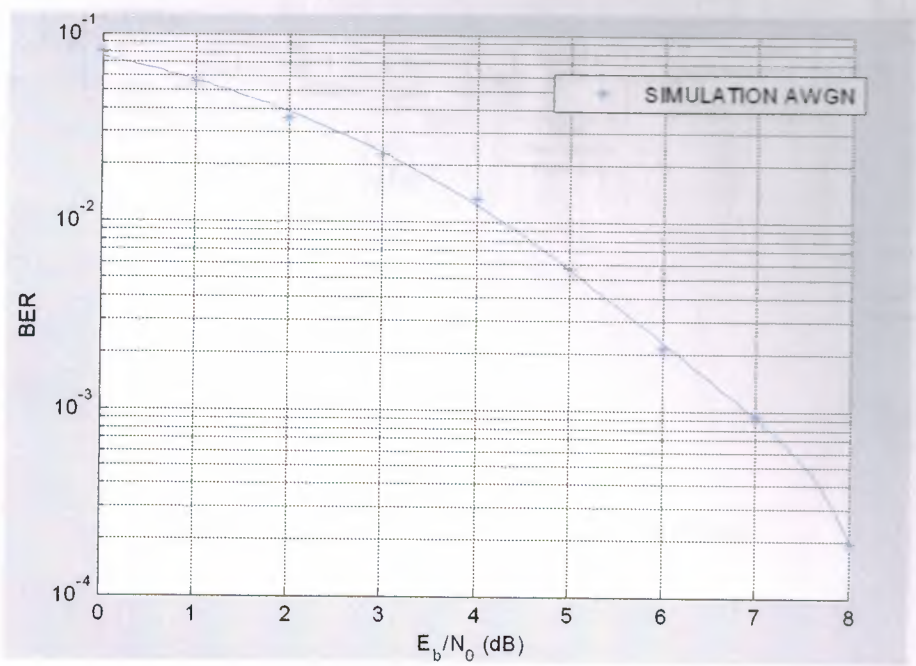

Figure 4.2 Performance of single carrier communication on AWGN channel.

SNR is inversely proportional to the noise variance (equation 4.1). Therefore as SNR increases, the noise corrupts less number of data and BER decreases [1].

$$variance = \frac{1}{2(SNR)} \tag{4.1}$$

4.2 Fading Result

4.2.1 Simulink Model of Rayleigh Fading

Figure 4.3 shows the model of figure 4.1 with an additional Rayleigh noise generator block, for generating multipath fading.

- **Rayleigh Noise Generator -** The Rayleigh Noise Generator block generates Rayleigh distributed noise.

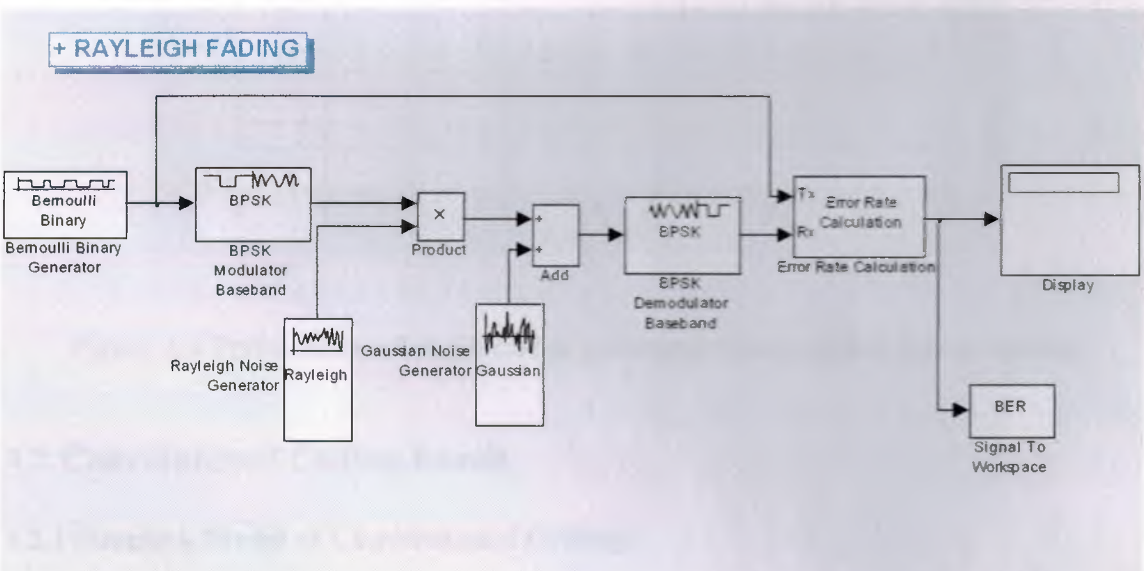

Figure 4.3 Simulink model of Rayleigh fading

4.2.2 Performance Simulation

Figure 4.3 shows the performance of simulation of a communication system over AWGN channel and rayleigh (flat) fading channel. It is observed that the performance is worse than the obtained over AWGN channel.

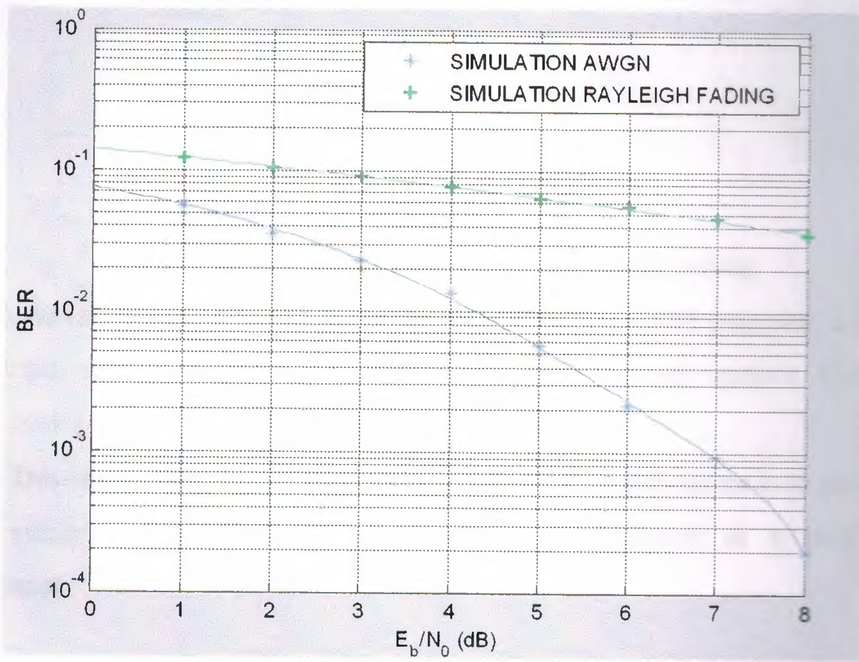

Figure 4.3 Performance of single carrier communication over flat fading channel.

4.3 Convolutional Coding Result

4.3.1 Simulink Model of Convolutional Coding

Figure 4.4 shows the same communication model as before with convolutional encoder and viterbi decoder blocks, added for error correction.

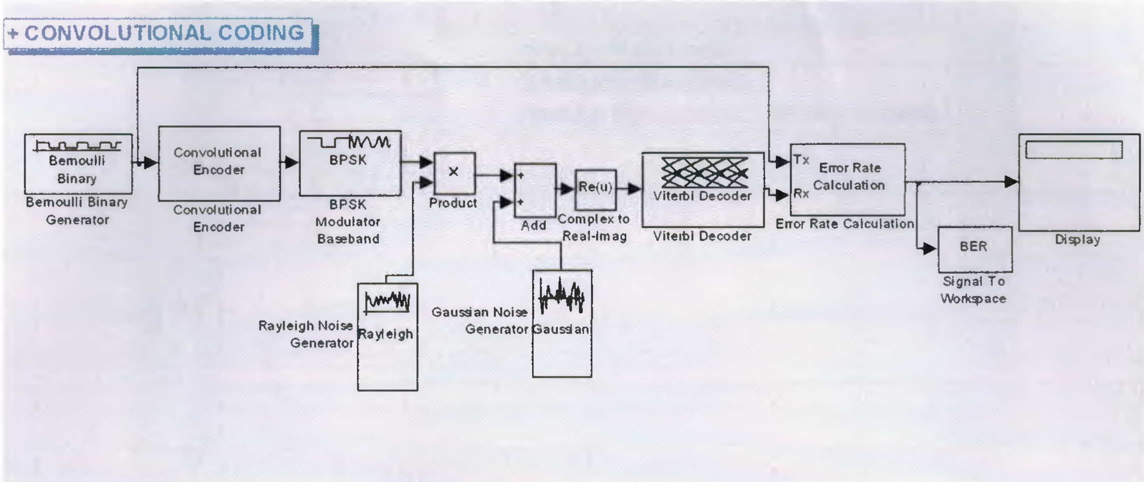

Figure 4.4 simulink model of convolutional coding.

- **Convolutional Encoder -** The Convolutional Encoder block encodes a sequence of binary input vectors to produce a sequence of binary output vectors. This block can process multiple symbols at a time.
- **Vitrebi Decoder -** The Viterbi Decoder block decodes input symbols to produce binary output symbols. This block can process several symbols at a time for faster performance.

4.3.2 Simulation of Convolutional Coding

The idea of coding is to improve received signal quality and link performance over small-scale times and distance. Here, we use convolutional coding to reduce the bit errors that occur in the channel. Figure 4.5 shows performance of flat fading using convolutional coding.

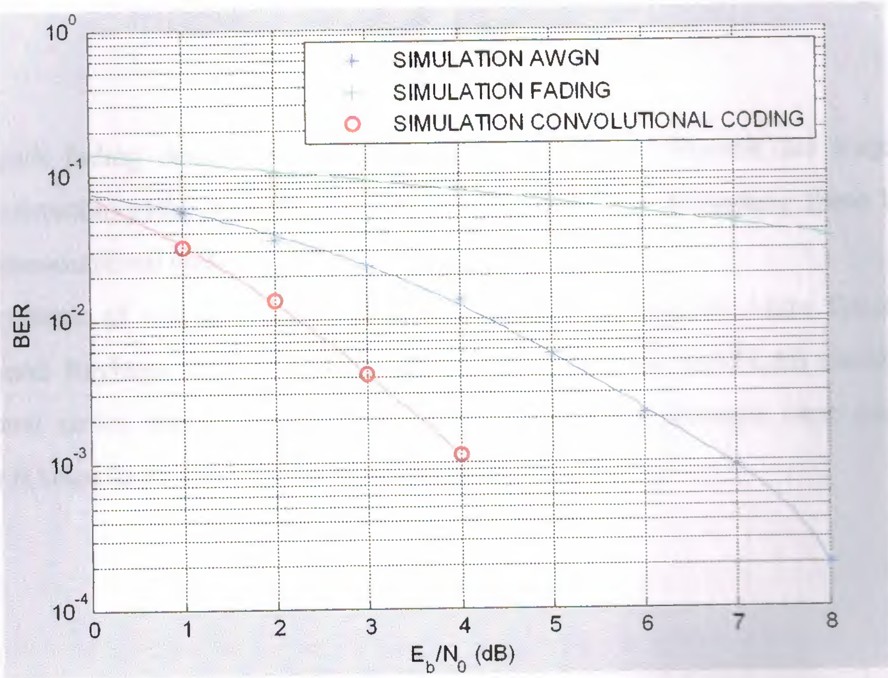

Figure 4.5 Performance of flat fading using convolutional coding

CONCLUSION

Multipath fading occurring in wireles and bandlimited channels are major limiting factors in communication systems. Various methods exist for eliminating these limitations, including convoultional codes.

Performance of convolutional codes are analyzed in additive white Gaussian noise (AWGN) and Reyleigh fading channels. It is shown, through MATLAB simulation, that convolutional codes should be preferred in Reyleigh fading channels over single carrier systems as it helps to increase the performance.

REFERANCES

[1] Proakis, J.G, Salehi, M., Communication system Engineering, Upper Saddele River, New Jersy 07458, 2002.

[2] Repappert, T.S., Wireless Communications, prentice Hall PTR, Upper Saddle River, NJ 07458, 2002.

[3] Carlson, A. B. (1986), Communication Systems, 3rd Ed., McGraw-Hill, New York.

[4] Lee, W.C.Y., Mobile Telecommunication systems, McGraw Hill publications, New York, 1989.

[5] Steel, R, ed., Mobile Radio Communications, IEEE press, 1994.

[6] Viterbi, A.J., and Omura, J.K., Principles of Digital Communication and Coding, McGraw Hill, New York, 1979.

[7] Foney, G.D., "The Viterbi Algorithm," Proceedings of the IEEE, Vol. 61, No. 3, pp. 268-278, March 1973.

[8] Fano, R.M., "A Heuristic Discussion of probabilistic coding," IEEE Transaction on information Theory, Vol. IT-9, pp. 64-74, Appril 1963.