ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

TECHNOLOGY FACULTY

ELECTRICAL AND COMPUTER ENGINEERING DEPARTMENT

SIMULATION OF A MOBILE RADIO COMMUNICATION

By

ROBEL HADDIS WELDEGEBREAL

A thesis submitted to

the School of Graduate Studies of Addis Ababa University

in partial fulfillment of the requirements for the Degree of Master of Science in

Electrical Engineering

Advisor

PROFESSOR WOLDE-GHIORGIS WOLDEMARIAM

Feb. 2003

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

TECHNOLOGY FACULTY
ELECTRICAL AND COMPUTER ENGINEERING DEPARTMENT

SIMULATION OF A MOBILE RADIO COMMUNICATION

By

ROBEL HADDIS WELDEGEBREAL

Signature of the Board of Examiners for Approval

_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

ACKNOWLEDGEMENT

First and foremost, to God, who makes everything possible. I would also like to extend my sincere gratitude to the following; Ethiopian Airlines Enterprise that gave time to my studies at this school. my advisor Professor Wolde-ghiorgis Woldemariam for his constructive reviews of my work and Dr Mohammed Abdo for his support during the research, my family who have been supporting me in so many ways, all Electrical Engineering Department staff who have facilitated access to resources at the school, and helped me in any way, fellow classmates for their moral support and also constructive ideas during the research, and all those who have contributed in one way or other to the success of this paper.

Table of Contents

ABSTRACT

The implementation of a simulation environment for mobile communication is very significant for Engineers and researchers of a mobile communication since it enables them to work in a environment that is almost similar with the practical one. Because multipurpose programmable digital signal processing have made it possible to implement digital communication like modulators and demodulators completely in software.

The aim of this project work is to create and verify a channel model for a mobile radio channel. This model will later be used in the project to simulate digital transmission of data through a source(speech) coder, Channel coder and the modulator.

When radio systems are simulated it is usually not practical to simulate the bandpass carrier signals. The high sampling frequency leads to unnecessarily long simulation time. Instead we use an equivalently complex low pass-model of the whole system.

This study was designed in such a way that first the detail background theory of the subject matter is discussed in with alternative means of approaching the subject matter and why one technique is preferred amongst existing techniques is discussed after mathematical and logically presenting of the efficiency and effectiveness of the technique. The chosen technique will be discussed in detail and a software algorithm is formulated.

Chapter 1 presents radio propagation path loss, link-budgets and log normal shadowing and describes a way to model and predict the fading, time delay spread and Dopler spread and describes how to measure and model the impact that signal bandwidth and motion have on the instantaneous received signal through the multipath channel of radio propagation in many operating environments. The various characteristics that produce and affect the channel will be simulated using the Spectrum method.

Chapter 2 provides an introduction to speech coding. The goal of speech coding systems is to transmit speech with the highest possible quality using the least possible channel capacity. Linear predictive coding techniques are simulated in this thesis.
Channel coding, error control coding and interleaving are presented in chapter 3 for error correcting capability that resist mobile channel impairments. Amongst these interleaving and reed-Solomon coding and decoding are discussed for software simulation.

Chapter 4 covers the most common and bandwidth efficient digital modulation techniques used in wireless communications and demonstrates the tradeoff that must be made in selecting a modulation technique. I tried to simulate the bandwidth efficient and error prone $\pi/4$ Differential Quadratic phase shift keying modulation scheme(QPSK) and compared for performance with the binary phase shift keying and the famous and efficient GSM modulation technique the Gausian minimum phase shift keying(GMSK).

There are a number of factors that enter into the choice of a modulation scheme for use in

a wireless application. Performance of a cellular system is dependent on the efficiency of the modulation scheme employed. Constant envelope modulation techniques, such as QPSK and GMSK, will be used to examine the features of each scheme and illustrate their uses in the cellular environment. The study of QPSK and GMSK is crucial to second and third generation cellular systems where achieving high capacity is the supreme factor. The goal of a modulation technique is not only to transport a message signal through a radio channel, but to achieve this with the best quality, power efficiency, and the least amount of bandwidth possible. In order to study the techniques for occupying less bandwidth and reducing power consumption per channel, a closer study of transmission techniques are explored in order to determine a favorable modulation technique for a particular wireless application.

In portable radio systems where people communicate while walking or driving, these methods are used individually or in tandem to improve the quality (i.e speech coding reduce the bit error rate) of digital mobile radio communications in the presence of fading and noise.

CHAPTER ONE

MOBILE RADIO CHANNEL

1.1 Introduction

A mobile radio system is usually cellular, i.e. several transmitters and receivers are placed in a particular pattern which is given by the appearance of the environment, the transmitted power and the desired capacity in the area. The reason why cellular systems are dominating today is that the capacity for the whole system depends on the distance between the base stations and not just on the availability of spectrum, and that the transmitted power can be reduced, which leads to a longer standby time and smaller and lighter cellular phones.

Since it is a very complex operation to simulate a large system with many users, simulations are usually done in two steps. The first step is to make a simulation of the air Interface(channel), with one isolated cell, which, of course, limits the number of users and the complexity of the simulation. After that, a simulation of the system is conducted, for which the results from the first simulation (typically required signal-to-noise-ratio for acceptable performance) are used to simulate multiple users and cells, In this project first the air interface will be simulated.

To simulate the air interface, a complete description of the system from transmitter to receiver is needed, which most often follows Shannon's classical model (see Figure 1.1). This model consists of a source which transmits signals over the channel via the transmitter and of a receiver which decodes the signal to the sink. The transmitter is usually divided into source encoder, channel encoder and modulator, while the receiver consists of the inverse blocks demodulator, channel decoder and source decoder.

The *source* can be a speech signal from a microphone or a binary data sequence from a computer. Since speech usually consists of a lot of redundant (superfluous) information, the speech is usually coded(compressed) in a speech coder- special form of *source coder*. The speech coder produces a series of bits, which for simplicity can be assumed to be

independent, and with equal probability of ones and zeros. This assumption allows us to simulate only one source model, regardless of whether we are transmitting speech or data.

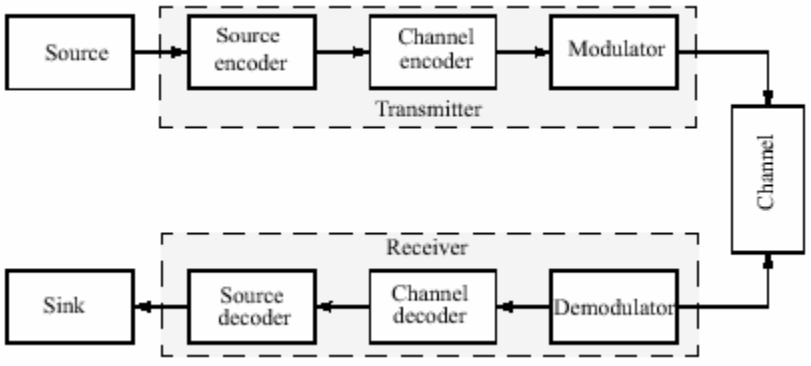


Figure 1.1 Shannon's communication model.

However, speech is not as sensitive for errors on the channel, and this usually means that we can allow a maximum bit error probability of 10^{-3} , while data usually requires 10^{-6} or lower.

The purpose of the *channel coder* is to add redundancy in a controlled way, so that the errors that may appear on the channel are counteracted (are detected and/or corrected). Many different types of codes with different levels of error protection, for different channels and with complexity (in coders and decoders) exist.

The *modulator* transforms a bit sequence into a format that is suited for the channel. For digital radio communication this means that, among other things, the signal is modulated on a carrier before it is transmitted. Appropriate frequencies are found in the interval 800 kHz up to maybe 30 GHz (usually called the radio spectrum). Since there is a shortage of frequencies (spectrum) the trend for new systems is to use higher and higher frequencies. A higher frequency leads to increased path loss, and therefore reduced interference from the surrounding cells. This means that the cell size can be reduced, the capacity of the whole system increases and the power consumption's reduced. However, this is at the cost of more expensive infrastructure (more base stations are required to cover a certain area) and more expensive electronic devices.

The *channel* is a model of the actual physical channel, i.e. the relation between a transmitted and received signal. For many realistic applications, i.e. mobile radio, the physical channel is so complicated that I have to use a highly simplified model. The result of this is a model which certainly can not predict performance in a given situation, but still gives a good picture of how the system as a whole will work.

When radio systems are simulated it is usually not practical to simulate the bandpass carrier signals. For simulation of a GSM-system this would mean that all the signals are upconverted) to a carrier at approximately 900.MHz, and the high sampling frequency then leads to unnecessarily long simulation time. Instead we use an equivalently complex low pass-model of the whole system. Subsequently, we do not simulate the carrier signal, but instead we are transforming the carrier signal down to frequencies at the base band. However, this means that the signals have to be complex valued, but this is a price which is usually worth paying [5].

In the project, most of the components of a mobile radio system are simulated. However, to reduce the simulation time, in most problems only one user will be simulated and a simple (but quite realistic) channel model will be used. The aim of this project work is to create and verify a channel model for a mobile radio channel. This model will later be used in the project to simulate digital transmission of data through a source(speech) coder, Channel coder and the modulator.

1.2 Radio wave Propagation

In a typical macro cell (approximately. 1-20 km radius) the base stations are placed high, which means that there exist no or few local objects which can reflect or spread the received or transmitted electromagnetic waves. Around the mobile station, however, which is usually placed on ground level, there are often many local objects which block the line of sight (LOS) signal between the base station and the mobile phone. Reflections and diffraction from these objects create many different electromagnetic waves which are received in the mobile antenna. These waves usually come from many different direction's and the delay varies, depending on the distance the waves have been travelling. In the

receiver the waves are added, either constructively or destructively and create a received signal which may vary rapidly in both amplitude and phase, depending on the loca1 objects or how the mobile terminal moves. We call this phenomenon the multipath propagation. Since the amplitude of the signal varies, the name fading is also used, and usually also small scale fading to distinguish this phenomenon (multipath propagation) from path loss (shadow fading or large scale fading) [3] .

Unlike wired channels that are stationary and predictable, radio channels are extremely random and do not offer easy analysis. Even the speed of motion impacts how rapidly the signal level fades as a mobile terminal moves in space. Modeling the radio channel is one of the most difficult parts of mobile radio system design, and is typically done in a statistical fashion, based on measurements made specifically for an intended communication system or spectrum allocation.

Propagation models have traditionally focused on predicting the average received signal strength at a given distance from the transmitter, as well as the variability of the signal strength in close spatial proximity to a particular location. Propagation models that predict the mean signal strength for an arbitrary transmitter-receiver (T -R) separation distance are useful in estimating the radio coverage area of a transmitter and are called large-scale propagation models, since they characterize signal strength over large, T -R separation distances (several hundreds or thousands of meters). On the other hand, propagation models that characterize the rapid fluctuations of the received signal strength over very short travel distances (a few wavelengths) or short time durations (on the order of seconds) are called small-scale or fading models.

As a mobile receiver moves over very small distances, the instantaneous received signal strength may fluctuate rapidly giving rise to small-scale fading. The reason for this is that the received signal is a sum of many contributions coming from different directions. Since the phases are random, the sum of the contributions varies widely; for example, obeys a Rayleigh fading distribution. In small-scale fading, the received signal power may vary by as much as three or four orders of magnitude (30 or 40 dB) when the receiver is moved by

only a fraction of a wavelength. As the mobile moves away from the transmitter over much larger distances; the local average received signal will gradually decrease, and it is this local average signal level that is predicted by large-scale propagation models. Typically, the local average received power is computed by averaging signal measurements over a measurement track of 5 λ to 40 λ. For cellular and PCS frequencies in the 1 GHz to 2 GHz band, this corresponds to measuring the local average received power over movements of 1 m to 10 m.

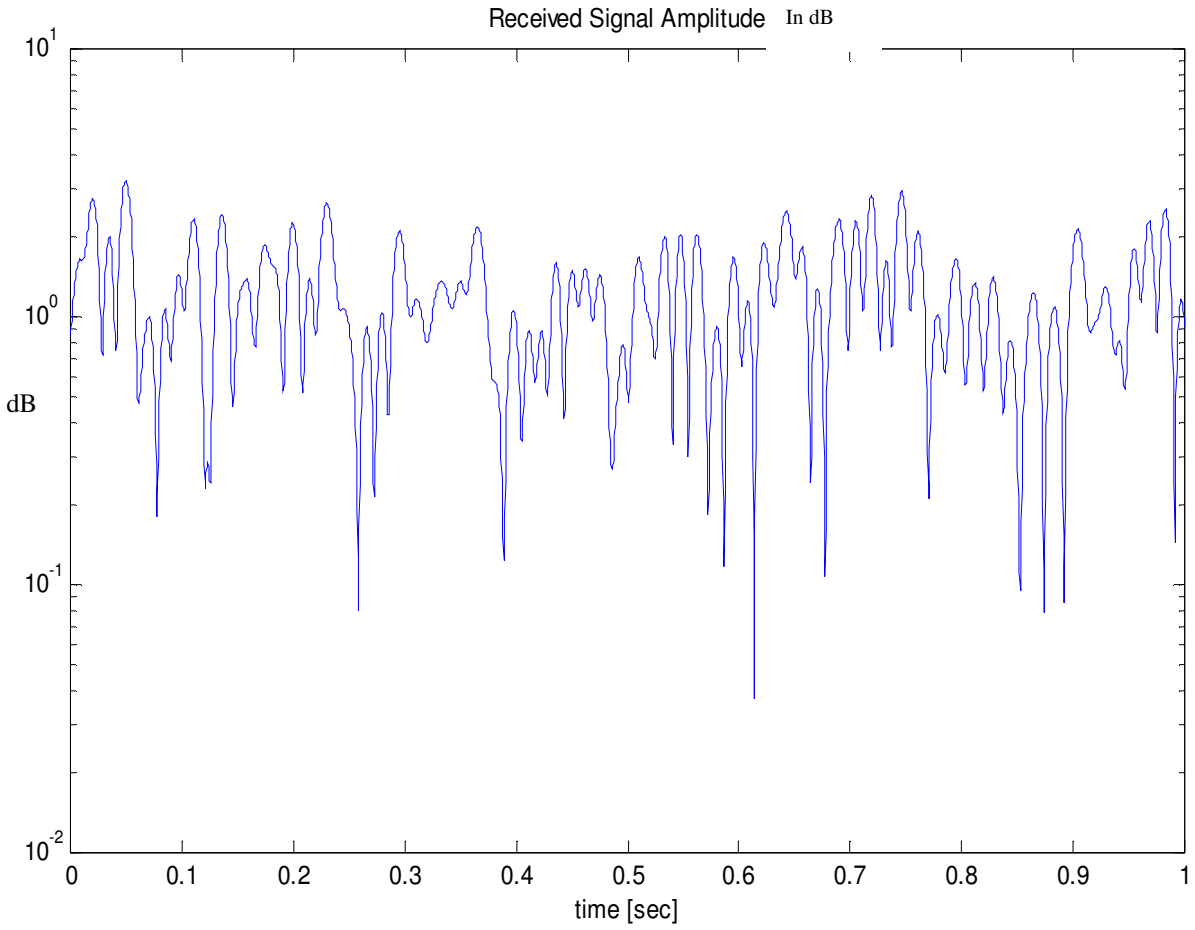


Figure 1.2 small-scale fading and the more gradual large-scale variations for an indoor radio communication system.

Notice in the figure that the signal fades rapidly (small-scale fading) as the receiver moves, but the local average signal changes much more gradually with distance.

1.3 Free Space Propagation Model

The free space propagation model is used to predict received signal strength when the transmitter and receiver have a clear, unobstructed line-of-sight path between them. Satellite communication systems and microwave line-of-sight radio links typically undergo free space propagation. As with most large-scale radio wave propagation mode's, the free space model predicts that received power decays as a function of the T -R separation distance raised to some power (i.e. a power law function). The free space power received by a receiver antenna which is separated from a radiating transmitter antenna by a distance d, is given by the Friis free space equation,

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d\ L} \qquad 1.1$$

where P_t is the transmitted power, $P_r(d)$ is the received power which is a function of the T-R separation, G_t is the transmitter antenna gain, G_r is the receiver antenna gain, d is the T -R separation distance in meters, L is the system loss factor not related to propagation ($L \geq 1$), and λ is the wavelength in meters. The gain of an antenna is related to its effective aperture, A_e by

$$G = \frac{4\pi A_e}{\lambda^2} \qquad 1.2$$

The effective aperture A_e is related to the physical size of the antenna, and λ is related to the carrier frequency by

$$\lambda = \frac{c}{f} = \frac{2\pi c}{\omega_c} \qquad 1.3$$

where f is the carrier frequency in Hertz, ω_c is the carrier frequency in radians per second, and c is the speed of light given in meters/s. The values for P_t and P_r must be expressed in the same units, and G_t and G_r are dimensionless quantities. The miscellaneous losses L ($L \geq 1$) are usually due to transmission line attenuation, filter losses, and antenna losses in the communication system. A value of L = 1 indicates no loss in the system hardware. The Friis free space equation of (1.1)shows that the received power falls off as the square of

the T-R separation distance. This implies that the received power decays with distance at a rate of 20 dB/decade.

The *path loss* which represents signal attenuation as a positive quantity measured in dB, is defined as the difference (in dB) between the effective transmitted power and the received power and may or may not include the effect of the antenna gains. The path loss for the free space model when antenna gains are included is given by

$$PL(dB) = 10\log\frac{P_t}{P_r} = -10\log\left[\frac{G_t G_r \lambda^2}{(4\pi)^2 d^2}\right] \quad 1.4$$

When antenna gains are excluded, the antennas are assumed to have unity gain. and path loss is given by

$$PL(dB) = 10\log\frac{P_t}{P_r} = -10\log\left[\frac{\lambda^2}{(4\pi)^2 d^2}\right] \quad 1.5$$

The Friis free space model is only a valid predictor for P_r, for values of d which are in the far-field of the transmitting antenna. The far-field, or Fraunhofer region, of a transmitting antenna is defined as the region beyond the far-field distance d_f which is related to the largest linear dimension of the transmitter antenna aperture and the carrier wavelength. The Fraunhofer distance is given by

$$d_f = 2D^2/\lambda \quad 1.6$$

where D is the largest physical linear dimension of the antenna. Additionally, to be in the far field region. d_f must satisfy

$$d_f >> D \quad 1.7$$

$$d_f >> \lambda \quad 1.8$$

Thus, using Equation (1.1). the received power in free space at a distance greater than do is given by

$$P_r(d) = P_r(d)\left(\frac{d_0}{d}\right)^2 \quad where \qquad d \geq d_0 \geq d_f \qquad 1.9$$

1.4 Relating Power to Electric Field

The free space path loss model of Section 1.2 is readily derived from first principles. It can be proven that any radiating structure produces electric and magnetic fields **[8]**, **[9]**. Consider a small linear radiator of length L , that is placed coincident with the z-axis and has its center at the origin, as shown in Figure 1.3.

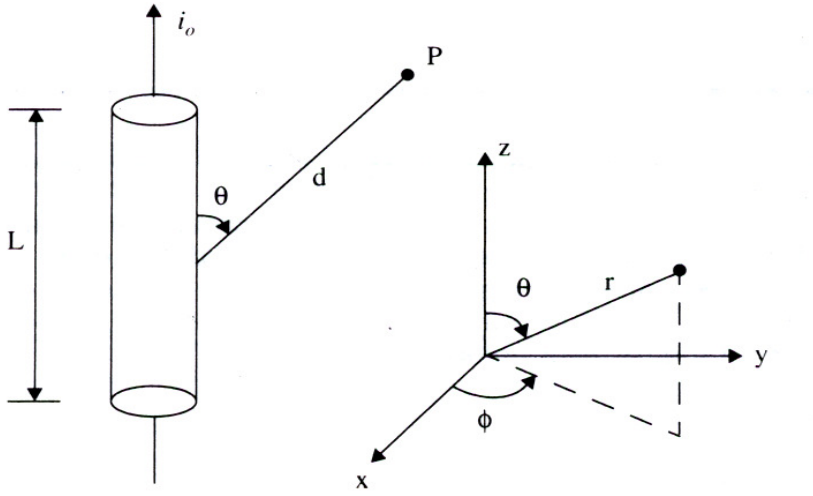


Figure 1.3 A linear radiator of length L(L « A), carrying a current of amplitude i_o and making an angle θ with a point, at distance d.

If a current flows through such an antenna, it launches electric and magnetic fields that can be expressed as

$$E_r = \frac{i_0 L\cos\theta}{2\pi\varepsilon_0 c}\left(\frac{1}{d^2} + \frac{c}{j\omega_c d^3}\right)e^{j\omega_c(t-d/c)} \qquad 1.10$$

$$E_\theta = \frac{i_0 L\sin\theta}{2\pi\varepsilon_0 c^2}\left(\frac{j\omega}{d} + \frac{c}{d^2} + \frac{c^2}{j\omega_c d^3}\right)e^{-j\omega_c(t-d/c)} \qquad 1.11$$

$$H_\phi = \frac{i_0 L\sin\theta}{2\pi c}\left(\frac{j\omega_c}{d} + \frac{c}{d^2}\right)e^{j\omega_c(t-d/c)} \qquad 1.12$$

with $E_\phi = H_r = H_\theta = 0$. In the above equations, all 1/ d terms represent the radiation field component, all 1 / d^2 terms represent the induction field component, and all 1/ d^3 terms represent the electrostatic field component. As seen from Equations (1.10) to (1.12), the electrostatic and inductive fields decay much faster with distance than the radiation field. At regions far away from the transmitter (far-field region), the electrostatic and inductive fields become negligible and only the radiated field components of E_θ and H_ϕ need be considered.

In free space, the power flux density P_d (expressed in W/m^2) is given by

$$P_d = \frac{PG}{4\pi d^2} = \frac{E^2}{R_{fs}} = \frac{E^2}{\eta} W/m^2 \qquad 1.13$$

where R_{fs} is the intrinsic impedance of free space given by $\eta = 120\pi\Omega$ (377 Ω). Thus, the power flux density is

$$P_d = \frac{|E|^2}{377\Omega} W/m^2 \qquad 1.14$$

where |E| represents the magnitude of the radiating portion of the electric field in the far field. Figure 1.4a illustrates how the power flux density disperses in free space from an isotropic point

source. P_d may be thought of as the EIRP divided by the surface area of a sphere with radius d.

The power received at distance d, $P_r(d)$, is given by the power flux density times the effective aperture of the receiver antenna, and can be related to the electric field using Equations (1.1),(1.2), (1.13), and (1.14).

$$P_r(d) = P_d A_e = \frac{|E|^2}{120\pi} A_e = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} W \qquad 1.15$$

$$P(d) = \frac{V^2}{R_{ant}} = \frac{[V_{ant}/2]^2}{R_{ant}} = \frac{{V_{ant}}^2}{4R_{ant}} \qquad 1.16$$

Equation (1.15) relates electric field (with units of V/m) to received power (with units of watts), and is identical to Equation (1.1) with L = 1 .

Often it is useful to relate the received power level to 'a receiver input voltage, as well as to an induced E-field at the receiver antenna. If the receiver antenna is modeled as a matched resistive load to the receiver, then the receiver antenna will induce a voltage into the receiver which is half of the open circuit voltage at the antenna. Thus, if V is the rms voltage at the input of a receiver (measured by a high impedance voltmeter), and R_{ant} is the resistance of the matched receiver, the received power is given by

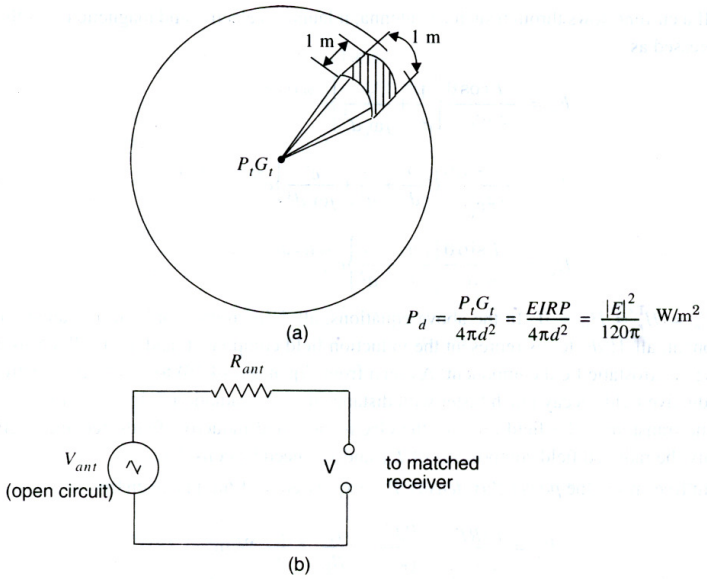


Figure 1.4 (a) Power flux density at a distance d from a point source; (b) model for voltage applied to the input of a receiver.

Through Equations (1.14) to (1.16), it is possible to relate the received power to the received E-field or the open circuit rms voltage at the receiver antenna terminals. Figure 1.4b illustrates an equivalent circuit model. Note Vant = V when there is no load.

1.5 The Three Basic Propagation Mechanisms

Reflection, diffraction, and scattering are the three basic propagation mechanisms which impact propagation in a mobile communication system. Received power (or its reciprocal, path loss) is generally the most important parameter predicted by large-scale propagation models based on the physics of reflection, scattering, and diffraction. Small-scale fading

and multipath propagation may also be described by the physics of these three basic propagation mechanisms.

Reflection occurs when a propagating electromagnetic wave impinges upon an object which has very large dimensions when compared to the wavelength of the propagating wave.
Reflections occur from the surface of the earth and from buildings and walls. When a radio wave propagating in one medium impinges upon another medium having different electrical properties, the wave is partially reflected and partially transmitted. If the plane wave is incident on a perfect dielectric, part of the energy is transmitted into the second medium and part of the energy is reflected back into the first medium, and there is no loss of energy in absorption. If the second medium is a perfect conductor, then all incident energy is reflected back into the first medium without loss of energy. The electric field intensity of the reflected and transmitted waves may be related to the incident wave in the medium of origin through the Fresnel reflection coefficient (Γ). The reflection coefficient is a function of the material properties, and generally depends on the wave polarization, angle of incidence, and the frequency of the propagating wave.
In general, electromagnetic waves are polarized, meaning they have instantaneous electric field components in orthogonal directions in space. A polarized wave may be mathematically represented as the sum of two spatially orthogonal components, such as vertical and horizontal, or left-hand or right-hand circularly polarized components. For an arbitrary polarization, superposition may be used to compute the reflected fields from a reflecting surface.

Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a surface that has sharp irregularities (edges). The secondary waves resulting from the obstructing surface are present throughout the space and even behind the obstacle, giving rise to a bending of waves around the obstacle, even when a line-of-sight path does not exist between transmitter and receiver. At high frequencies, diffraction, like reflection, depends on the geometry of the object, as well as the amplitude, phase, and polarization of the incident wave at the point of diffraction.

Scattering occurs when the medium through which the wave travels consists of objects with dimensions that are small compared to the wavelength, and where the number of obstacles per unit volume is large. Scattered waves are produced by rough surfaces, small objects, or by other irregularities in the channel. In practice, foliage, street signs, and lamp posts induce scattering in a mobile communications system.

1.6 Small-Scale Multipath Propagation

Small-scale fading or simply fading, is used to describe the rapid fluctuations of the amplitudes, phases, or multipath delays of a radio signal over a short period of time or travel distance, so that large-scale path 1oss effects may be ignored. Fading is caused by interference between two or more versions of the transmitted signal which arrive at the receiver at slightly different times. These waves, called multipath waves, combine at the receiver antenna to give a resultant signal which can vary widely in amplitude and phase, depending on the distribution of the intensity and relative propagation time of the waves and the bandwidth of the transmitted signal.

Multipath in the 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 below the height of surrounding structures, so there is no single line-of-sight path to the base station. Even when a line-of-sight exists, multipath still occurs due to reflections from the ground and surrounding structures. The incoming radio waves are 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, phases, and angles 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 objects 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 a long period of time. Antenna space diversity can prevent deep fading nulls.

Figure 1.2 shows typical rapid variations in the received signal level due to small-scale fading as a receiver is moved over a distance of a few meters.

Due to the relative motion between the mobile and the 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 multipath wave.

1.7 Factors Influencing Small-Scale Fading

Many physical factors in the radio propagation channel influence small-scale fading.

These include the following:

Multipath propagation -The presence of reflecting objects and scatterers 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 time and spatial orientation. The random phase and amplitudes of the different multipath components cause fluctuations 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 whether the mobile receiver is moving toward or away from the 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 a greater rate than the mobile, then this effect dominates the small-scale fading. Otherwise, motion of surrounding objects may be 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 the multipath channel, the received signal will be distorted, but the received signal strength will not fade much over a local area (i.e., 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 a measure of the maximum frequency difference for which signals 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 the transmitted signal.

1.8 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 1.5 shows a Rayleigh distributed signal envelope as a

function of time. The Rayleigh distribution has a probability density function (pdf) given by

$$p(r) = \left\{ \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2} \right), (.0 \le r \le \infty) \right\} \qquad 1.17$$

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

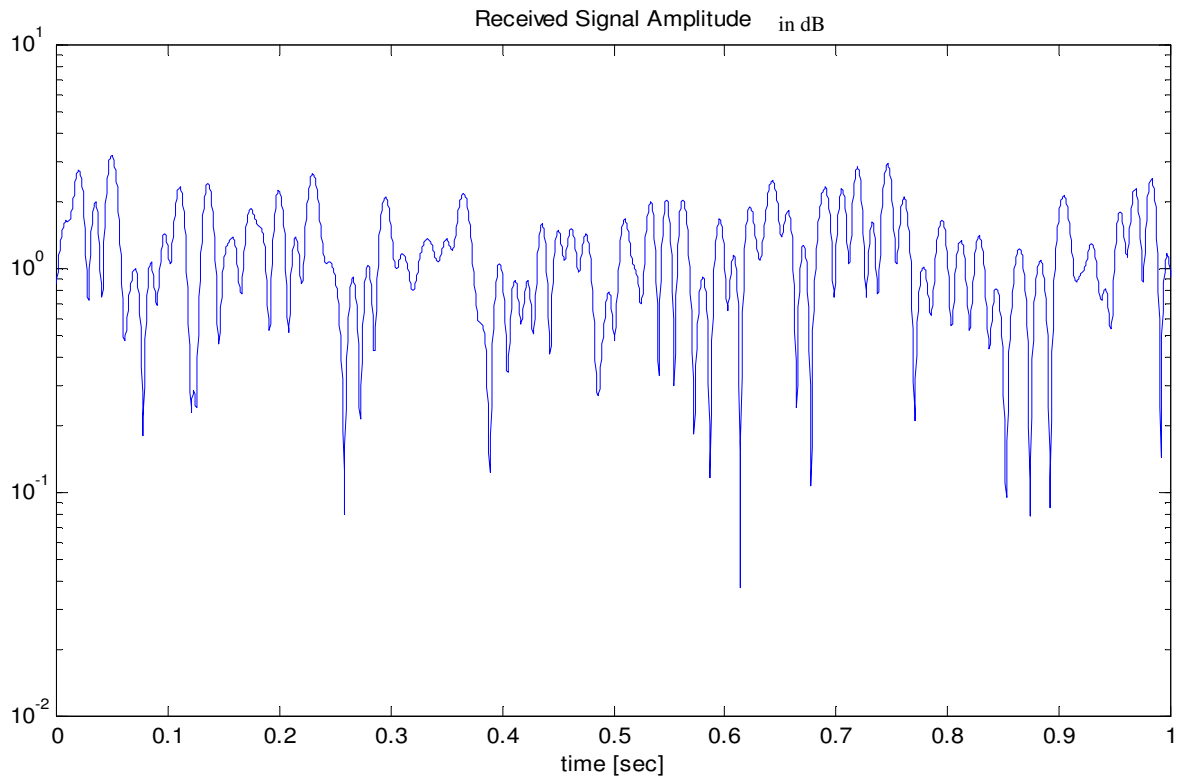


Figure 1.5 A typical Rayleigh fading envelope at 900 MHz

The probability that the envelope of the received signal does not exceed a specified value R is given by the corresponding cumulative distribution function (CDF)

$$P(R) = P(r \le R) = \int_0^R p(r)dr = 1 - \exp\left(-\frac{R^2}{2\sigma^2}\right) \qquad 1.18$$

The mean value r_{mean} of the Ray1eigh distribution is given by

$$rmean = E[r] = \int_0^\infty rp(r)dr = \sigma\sqrt{\frac{\pi}{2}} = 1.2533\sigma \qquad 1.19$$

and the variance of the Rayleigh distribution is given by σ , which represents the ac power in the signal envelope

$$\sigma_r^2 = E[r^2] - E^2[r] = \int_0^\infty r^2 p(r)dr - \frac{\sigma^2\pi}{2} \qquad 1.20$$

$$= \sigma^2\left(2 - \frac{\pi}{2}\right) = 0.4292\sigma^2$$

The rms value of the envelope is the square root of the mean square, or $\sqrt{2}\sigma$, where σ is the standard deviation of the original complex Gaussian signal prior to envelope detection. The median value of r is found by solving ,

$$\frac{1}{2} = \int_0^{r_{median}} p(r)dr \qquad 1.21$$

$$r_{median} = 1.177\sigma$$

Thus the mean and the median differ by only 0.55 dB in a Rayleigh fading signal. Note that the median is often used in practice. since fading data are usually measured in the field and a particular distribution cannot be assumed. By using median values instead of mean values, it is easy to compare different fading distributions which may have widely varying means.

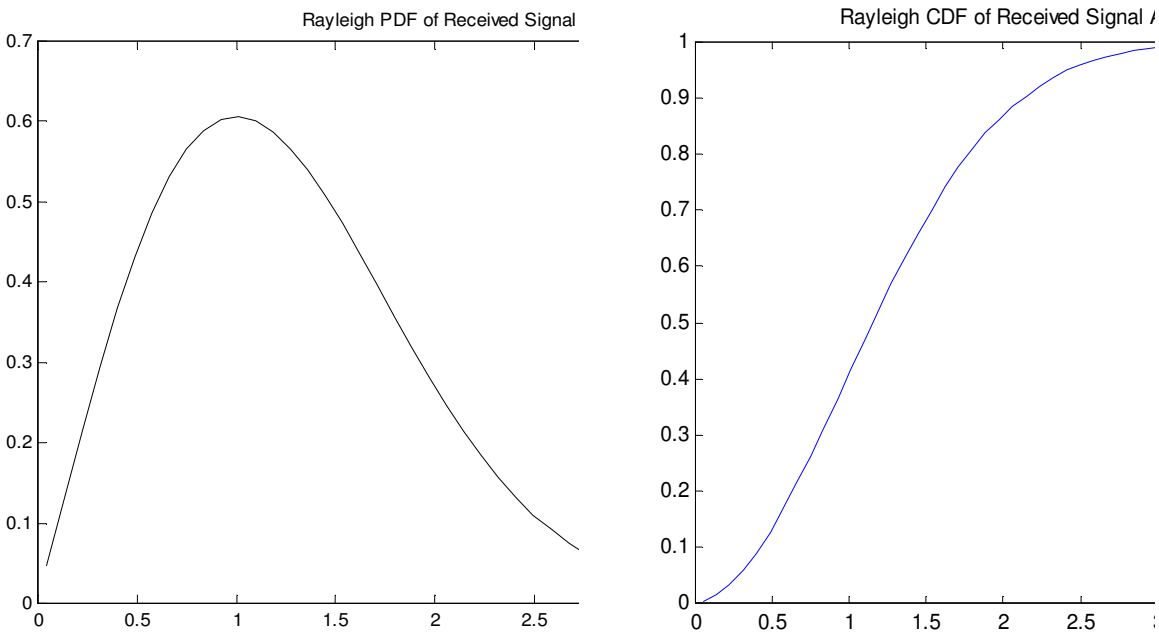


Fig 1.6 Rayleigh probability density function(pdf) and Cumulative density function(cdf)

1.9 Level Crossing and Fading Statistics

Rice computed joint statistics for a mathematical problem and thereby provided simple expressions for computing the average number of level crossing and the duration of fades. The level crossing rate (LCR) and average fade duration of a Rayleigh fading signal are two important statistics which are useful for designing error control codes and diversity schemes to be used in mobile communication systems, since it becomes possible to relate the time rate of change of the received signal to the signal level and velocity of the mobile. The level crossing rate (LCR) is defined as the expected rate at which the Rayleigh fading envelope, normalized to the local rms signal level, crosses a specified level in a positive-going direction. The number of level crossings per second is given by

$$N_R = \int_0^\infty \dot{r} p(R, \dot{r}) d\dot{r} = \sqrt{2\pi} f_m \rho e^{-\rho^2} \qquad 1.22$$

where r is the time derivative of r(t) (i.e., the slope), p(R,r) is the joint density function of r and r at r = R, f_m the maximum Doppler frequency, and $\rho = R/ R_{rms}$ is the value of the

specified level R, normalized to the local rms amplitude of the fading envelope [7]. Equation (1.22) gives the value of N_R the average number of level crossings per second at specified R. The level crossing rate is a function of the mobile speed as is apparent from the presence of f_m in Equation (1.22). There are few crossings at both high and low levels, with the maximum rate occurring at $\rho = 1/\sqrt{2}$, (i.e., at a level 3 dB below the rms level). The signal envelope experiences very deep fades only occasionally, but shallow fades are frequent.

The average fade duration is defined as the average period of time for which the received signal is below a specified level R. For a Rayleigh fading signal. this is given by

$$\bar{\tau} = \frac{1}{N_R} \Pr[r \le R] \qquad 1.23$$

where $\Pr[r \le R]$ is the probability that the received signal r,. is less than R and is given by

$$\Pr[r \le R] = \frac{1}{T} \sum \tau_i \qquad 1.24$$

where τ_i is the duration of the fade and T is the observation interval of the fading signal. The probability that the received signal r is less than the threshold R is found from the Rayleigh distribution as

$$P_r[r \le R] = \int_0^R p(r)dr = 1 - \exp(-\rho^2) \qquad 1.25$$

where p(r) is the pdf of a Rayleigh distribution. Thus. using Equations (1.22). (1.23), and (1.24). the average fade duration as a function of ρ and f_m can be expressed as

$$\bar{\tau} = \frac{e^{\rho^2} - 1}{\rho f_m \sqrt{2\pi}} \qquad 1.26$$

The average duration of a signal fade helps determine the most likely number of signaling bits that may be lost during a fade. Average fade duration primarily depends upon the speed of the mobile, and decreases as the maximum Doppler frequency f_m becomes large. If there is a particular fade margin built into the mobile communication system, it is appropriate to evaluate the receiver performance by determining the rate at which the input signal falls below a given level R, and how long it remains below the level, on

average. This is useful for relating SNR during a fade to the instantaneous BER which results.

1.10. A general channel model

The transmitted wave from a base station is usually vertically polarized, i.e. the waves radiate horizontally. Since the mobile terminal is usually moving, the received signal's frequency content is affected due to Doppler shift. A horizontal wave with an angle of incidence at the mobile terminal of θ relative to the mobile's direction (in the horizontal plane) is received with a positive frequency shift (Doppler shift)

$$f_d = f_m \cos\theta, \tag{1.27}$$

where $f_m = v/\lambda$, λ is the carrier wavelength of the signal and v is the speed of the mobile. Signals at the same angle of arrival but with opposite moving direction will generate a negative frequency shift of the same size. If we assume that there are many local objects surrounding the mobile, the received signal will consist of a sum of many signals which arrives with different angles and therefore different Doppler shifts. Furthermore, it is reasonable to assume that these different signals have been attenuated unequally on their way towards the receiver. Starting from a transmitted bandpass signal.

$$x(t) = c_I(t)\cos 2\pi f_c t - c_Q(t)\sin 2\pi f_c t = \mathrm{Re}\{c(t)e^{j2\pi f c t}\} \tag{1.28}$$

where $c(t) = c_I(t) + jc_Q(t)$ is the complex baseband signal and fc is the frequency of the carrier, a received signal $y(t) = \mathrm{Re}\{r(t)e^{j2\pi ft}\}$ consisting of N different arriving waves can be written as

$$y(t) = \mathrm{Re}\left\{\sum_{i=0}^{N-1} a_i(t) e^{j2\pi(f_c + f_{d,i}(t))(t-\tau_i(t))} c(t-\tau_i(t))\right\} \tag{1.29}$$

Here, $a_i(t)$ is amplitude, $\tau_i(t)$ is time delay and $f_{d,i}$ (t) is Doppler shift of the ith wave. Equation (1.29) can be simplified if we introduce

$$\theta_i(t) = 2\pi\{(fc + f_{d,i}(t,\tau))\,\tau_i(t) - f_{d,i}(t,\tau)\,t\} \tag{1.30}$$

Then we get

1.31

$$y(t) = \mathrm{Re}\left\{\sum_{i=0}^{N-1} a_i(t) e^{-j\theta_i(t)} c(t-\tau_i(t)) e^{j2\pi f_c t}\right\} \tag{1.31}$$

$$r(t) = \left\{\sum_{i=0}^{N-1} a_i(t) e^{-j\theta_i(t)} c(t-\tau_i(t))\right\} \tag{1.32}$$

We now have a general model for a mobile radio channel: The channel can be modeled as a time variant linear filter with the impulse response (in complex baseband form)

$$h_b(t,\tau) = \sum_{i=0}^{N-1} a_i(t) e^{-j\theta_i(t)} \delta(\tau - \tau_i(t)) \tag{1.33}$$

As seen in 1.33 a small change in the delay $\tau_i(t)$ means that the phase is changed a lot, because f_c is usually very large. However, the amplitude $a_i(t)$ does not change as fast, since it mainly depends on the area of the reflecting object. This mean, that what we normally refer to as small scale fading mainly depends on the fast shifts in the received waves and not their signal strength.

1.11. Flat fading

If the time delays in the channel $\tau_i(t)$ are approximately equal and much smaller than a symbol interval, the channel is called *flat fading*. We can then assume that the whole symbol which is transmitted over the channel is affected in the same way the spectrum of the channel is constant (flat) over the bandwidth of the transmitted signal.

A simplification of (1.31) can then be done, since the bandwidth of the transmitted signal is very small as compared to the carrier frequency: We assume that only one unmodulated carrier wave is transmitted (c(t) = 1). The received signal can then be written in complex baseband I/Q form (In-phase and Quadrature-phase)

$$x(t) = r_I(t)\cos 2\pi f_c t - r_Q(t)\sin 2\pi f_c t \qquad 1.34$$

where $r(t) = r_I(t) + jr_Q(t)$ and

$$r_I(t) = \sum_{i=0}^{N-1} a_i(t)\cos\theta_i(t) \qquad 1.35a$$

$$r_Q(t) = \sum_{i=0}^{N-1} a_i(t)\sin\theta_i(t)\,. \qquad 1.35b$$

For large N we can refer to the central limit theorem, which then says that $r_I(t)$ and $r_Q(t)$ can be regarded as independent stochastic process with normally distributed samples. If we also assume that these processes are wide sense stationary ($a_i(t) = a.$ and $\theta_i(t) = \theta.$), we arrive at normally distributed samples with the mean value zero and variances $\sigma_{rI}^2 = \sigma_{rQ}^2 = \sigma^2$.It is then possible to show that the envelope $z(t) = |r(t)| = \sqrt{r_I(t)^2 + r_Q(t)^2}$ is Rayleigh distributed with probability density

$$p_z(r) = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right) \quad r \geq 0 \qquad 1.36$$

and the phase <r(t) equally distributed between 0 and 2π .We now have a model for the envelope z(t) = |r(t)| and the phase <r(t) , but we also have to find the correlation in time, i.e. Rrr(τ}= E[r(t)r*(t + τ)] .However, this involves a lot of calculations [1]. Instead, we can-calculate the spectrum of r(t) which by definition is the Fourier transform of the autocorrelation Rrr(τ) (actually, the spectrum of y(t) is calculated, but there is a simple relationship between these spectra).

The result is

$$S(f) = \frac{K}{\pi f_m \sqrt{1 - \left(\frac{f - f_c}{f_m}\right)^2}} \qquad | f | \leq f_m \tag{1.37}$$

where K is a constant.

But the Analysis follows. Let p(α)dα denote the fraction of the total incoming power within dα of the angle α, and let A denote the average received power with the respect to an isotropic antenna. As N→∞, p(α)dα approaches a continuous, rather than a discrete, distribution. If G(α) is the azimuthal gain pattern of the mobile antenna as a function of the angle of arrival, the total received power can be expected as

$$P_r = \int A\mathrm{G}(\alpha)\mathrm{p}(\alpha)d\alpha \tag{1.38}$$

where AG(α)p(α)dα is the differential variation of received power with angle. If the scattered signal is a CW signal of frequency f_c, then the instantaneous frequency of the received signal component arriving at an angle is obtained using equation (1.39)

$$f(\alpha) = f = \frac{v}{\lambda}\cos(\alpha) + f_c \tag{1.39}$$

where f_m is the maximum Doppler shift. It should be noted that f(α) is an even function of α.

If S(f) is the power spectrum of the received signal, the differential variation of received power with frequency is given by

$$S(f)\,|\,df\,| \tag{1.40}$$

Equating the differential variation of received power with frequency to the differential variation in received power with angle, we have

$$S(f)\,|\,df\,| = A[p(\alpha)G(\alpha) + p(-\alpha)G(-\alpha)]\,|\,d\alpha\,| \qquad 1.41$$

Differentiating Equation (1.39) and rearranging the terms, we have

$$|\,df\,| = |\,d\alpha\,|\,|-\sin\alpha\,|\,f_m \qquad 1.42$$

Using equation (1.39), α can be expressed as a function of f as

$$\alpha = \cos^{-1}\left[\frac{f - f_c}{f_m}\right] \qquad 1.43$$

This implies that

$$\sin\alpha = \sqrt{1-\left(\frac{f - f_c}{f_m}\right)^2} \qquad 1.44$$

Substituting Equation 1.43 and 1.45 into both sides of 1.42 the power spectral density S(f) can be expressed as

$$S(f) = \frac{A[p(\alpha)G(\alpha) + p(-\alpha)G(-\alpha)]}{f_m\sqrt{1-\left(\frac{f - f_c}{f_m}\right)^2}} \qquad 1.45$$

where

$$S(f)=0, \quad |f-f_c| > f_m \qquad 1.46$$

The spectrum is centered on the carrier frequency and is zero outside the limits of $f_c \pm f_m$. Each of the arriving waves has its own carrier frequency (due to its direction of arrival) which is slightly offset from the center frequency. For the case of a vertical $\lambda/4$ antenna ($G(\alpha)=1.5$), and a uniform distribution $p(\alpha) = 1/2\pi$ over 0 to 2π, the output spectrum is given by

$$S(f) = \frac{1.5}{\pi f_m\sqrt{1-\left(\frac{f - f_c}{f_m}\right)^2}} \qquad 1.47$$

1.12. Frequency selective fading

If the bandwidth of the transmitted signal is of the same size or larger than the coherence bandwidth of the channel $1/(2\pi\sigma_\tau)$, where σ_τ is the delay spread (standard deviation of the time delays $\tau(t_i)$), different parts of the signal spectrum will be affected differently by the channel. This happens when the symbol period of the signal is small relative to σ_τ.This is called frequency selective fading. An appropriated model for this is obtained if we group the time delays which are approximately of the same size (their time differences should be

small relative to a symbol interval). Each and every of these groups of rays can then still be modeled as flat fading with a Rayleigh fading amplitude and a spectrum according to (12). The total received signal can then be modeled by an L-tap channel

$$r(t) = \sum_{l=0}^{L-1} r_l(t)c(t-\tau_l(t)) \qquad 1.48$$

where $\{r_l(t)\}$ are independent Rayleigh fading processes with power spectrum as given by (12).

Usually the time delays $\tau_n(t)$ are not an integer multiple of the symbol time, and therefore we have to over sample the transmitted signal. To avoid this increase of complexity, it is common to introduce an FIR channel model instead, where the distance between the channel taps is constant and equal to the symbol period. The different taps will then be correlated, but are usually modeled as independent anyway.

1.13. Simulation of a mobile radio channel

We have now found models for a mobile radio channel for both flat and frequency selective fading. Since frequency selective fading turned out to be a generalization of a flat fading channel, we will concentrate here on simulating a flat fading channel. Over the years several methods have been suggested, some more appropriate for hardware implementation and other more appropriate for software implementation. Here it will be presented two different methods appropriate for software implementation; the filter method and the spectrum method.

The filter method

The envelope of the received signal is Rayleigh distributed at every point in time. This Rayleigh distribution results from the signal's I and Q components, respectively, being independent and normally distributed with mean value zero and the variance σ^2.

Further, we have shown that the complex baseband signal has a correlation over time which is given by its spectrum. A natural simulation method (see Figure 1.8 for block diagram would then be to first generate a complex normally distributed white process and then filter it to get a desired spectrum. We call this method the filter method.

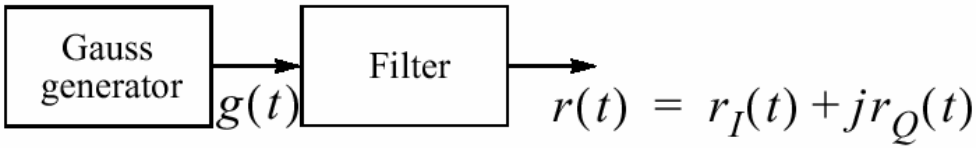


Figure 1.8 Block diagram for the filter method.

The impulse response H(f) of the filter, is obtained by using the Super formula

$$S_{rr}(f) = S_{gg}(f) \mid H(f) \mid^2 \qquad 1.49$$

where Sgg(f) is the power density spectrum of the normally distributed process. We also need a filter with frequency response equal to the square root of Srr(f) (since Sgg(f) is constant). The impulse response of this filter is given by the inverse transform

$$\sqrt{S_{rr}(f)} \subset K \frac{J_{1/4}(2\pi f_m |t|)}{\sqrt[4]{|t|}} \qquad 1.50$$

where k is a constant and J1/4(x) the bessel function of order ¼. When t→0 we have to use the following limiting value

$$\lim_{t \to 0} \frac{J_{1/4}(2\pi f_m |t|)}{\sqrt[4]{|t|}} = \frac{\sqrt[4]{f_m \pi}}{\Gamma(5/4)} \qquad 1.51$$

where $\Gamma(x)$ is the Gamma function

When one implement the filter in discrete time, the impulse response must be sampled, truncated, windowed and normalized. Since the bandwidth of $r_I(t)$ and $r_Q(t)$, respectively, is f_m , a sampling frequency $f_s \geq 2f_m$ should be used. Observe also that the complex process has the same sampling frequency f_s but that its bandwidth is $2f_m$. A proper window could be a Hamming window. The normalization is done so that the input signal and the output signal of the filter have the same energy.

The spectrum method

The dual operation to convolution in the time domain is multiplication in the frequency domain. The spectrum method is based on a complex normally distributed vector with independent samples which is Fourier transformed to another vector with the same characteristics (the Fourier transform is linear and all linear operations on uncorrelated and normally distributed data generate uncorrelated and normally distributed results). Instead of generating the time signal directly through convolution, we generate the signal's Fourier transform and use the inverse Fourier transform to get back to the time domain. Therefore, the spectrum method is in its nature block based. To generate a block or a vector consisting of N samples with the desired correlations and spectrum characteristics, we sample [12] with N samples in the frequency domain and multiply it with a vector of complex normally distributed independent samples. The spectrum method can be summarized in the three paragraphs below and is implemented in the project .

1. Specify the number of samples N needed to represent $S_{rr}(f)$ and the desired sampling frequency

2. Calculate the frequency resolution $\Delta f = f_s / (N - 1)$ and sample $\sqrt{S_{rr}(f)}$ at N points between $-f_s/2$ and $f_s/2$ at frequency spacing Δf. Normalize the vector such that its total energy is 1. Denote this vector $\sqrt{\mathbf{S}_{rr}}$.

3. Multiply component wise the vector $\sqrt{\mathbf{S}_{rr}}$ with a vector G containing N complex normally distributed independent random variables with mean 0 and variance 1, and inverse transform to get

$$r = DFT^{-1}\{\sqrt{\mathbf{S}_{rr}} \bullet \mathbf{G}\}$$

where • denoted component wise multiplication.

If one happen to sample [12] exactly at the frequency corresponding to the maximum Doppler frequency f_m , you have to truncate it in some smart way since $S_{rr}\{f_m\} = \infty$ Another hint is to choose N as a power of 2. The spectrum method is, in many cases, faster than the filter.

The big disadvantage with the filter model is that the impulse response in most cases of practical interest decreases very slowly in time, which leads to very long filters (if FIR-filters are assumed) and long simulation times. This is especially the case if the Doppler frequency is small as compared to the sampling frequency. The spectrum method avoids this problem.

CHAPTER TWO

SPEECH CODERS

2.1 Introduction

speech coders have assumed considerable importance in communication systems as their performance, to a large extent, determines the quality of the recovered speech and the capacity of the system. In wireless communication systems, bandwidth is a precious commodity, and the challenge of accommodating more users within a limited allocated bandwidth must be achieved. Low bit-rate speech coding offers a way to meet this challenge. The lower the bit rate at which the coder can deliver toll-quality speech, the more speech channels can be compressed within a given bandwidth.

In mobile communication systems, without low data rate speech coding, digital modulation schemes offer little in the way of spectral efficiency for voice traffic. To make speech coding practical, implementations must consume little power and provide tolerable, if not excellent, speech quality. The goal of speech coding systems is to transmit speech with the highest possible quality using the least possible channel capacity. This has to be accomplished while maintaining certain required levels of complexity of implementation and communication delay. In general, there is a positive correlation between coder bit-rate efficiency and the algorithmic complexity required to achieve it. The more complex an algorithm is, the more its processing delay and cost of implementation. A balance needs to be struck between these conflicting factors, and it is the aim of this speech processing developments to shift the point at which this balance is made toward lower bit rates.

Speech coders differ widely in their approaches to achieving signal compression. Based on the means by which they achieve compression, speech coders are broadly classified into two categories: waveform coders and vocoders. Waveform coders essentially strive to reproduce the time waveform of the speech signal as closely as possible. They are, in principle, designed to be source independent and can hence code equally well a variety of signals. They have the advantage of being robust for a wide range of speech characteristics and for noisy environments. All these advantages are preserved with minimal complexity, and in general this class of coders achieves only moderate economy in transmission bit rate. Example of waveform coders include pulse code modulation (PCM), adaptive differential pulse code modulation (ADPCM), and adaptive predictive coding(APC).

Pulse code modulation systems do not attempt to remove the redundancies in the speech signal. Adaptive differential pulse code modulation is a more efficient coding scheme which exploits the redundancies present in the speech signal. practically, adjacent samples of speech waveform are often highly correlated. This means that the variance of the difference between adjacent speech amplitudes is much more smaller than the variance of the speech signal itself. ADPCM allows speech to be encoded at a bit rate of 32kbps, which is half the standard 64 kbps PCM rate, while retaining the same voice quality.

Vocoders on the other hand achieve very high economy in transmission bit rate and are in general more complex. They, are based on using a priori knowledge about the signal to be coded, and for this reason, they are, in general, signal specific. It is this coding scheme that is implemented in this work. Before dealing with this type of coding in detail let ‘s see the general charestristics of speech.

2.2 Characteristics of Speech Signals

Speech waveforms have a number of useful properties that can be exploited when designing efficient coders. Some of the properties that are most often utilized in coder design include the nonuniform probability distribution of speech amplitude, the nonzero autocorrelation between successive speech samples, the nonflat nature of

the speech spectra, the existence of voiced and unvoiced segments in speech, and the quasiperiodicity of voiced speech signals. The most basic property of speech waveforms that is exploited by all speech coders is that they are bandlimited. A finite bandwidth means that it can be time-discretized (sampled) at a finite rate and reconstructed completely from its samples, provided that the sampling frequency is greater than twice the highest frequency component in the low pass signal. While the band limited property of speech signals makes sampling possible, the aforementioned properties allow quantization, the other most important process in speech coding, to be performed with greater efficiency.

Probability density function (pdf) –The nonuniform probability density function of speech amplitudes is perhaps the next most exploited property of speech. The pdf of a speech signal is in general characterized by a very high probability of near-zero amplitudes, a significant probability of very high amplitudes, and a monotonically decreasing function of amplitudes between these extremes. The exact distribution, however, depends on the input bandwidth and recording conditions. The two-sided exponential (Laplacian) function given in Equation (2.1) provides a good approximation to the long-term pdf of telephone quality speech signals

$$p(x) = \frac{1}{\sqrt{2\sigma_x}} \exp(-\sqrt{2}\,|x|/\sigma_x) \qquad 2.1$$

Note that this pdf shows a distinct peak at zero, which is due to the existence of frequent pauses and low level speech segments. Short-time pdfs of speech segments are also single peaked functions and are usually approximated as a Gaussian distribution. Nonuniform quantizers, including the vector quantizers, attempt to match the distribution of quantization levels to that of the pdf of the input speech signal by

allocating more quantization levels in regions of high probability and fewer levels in regions where the probability is low.

Autocorrelation Function (ACF)- Another very useful property of speech signa1s is that there exists much correlation between adjacent samples of a segment of speech. This implies that in every sample of speech, there is a large component that is easily predicted from the value of the previous samples with a small random error. All differential and predictive coding schemes are based on exploiting this

$$C(k) = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x(n)x(n+|k|) \qquad 2.2$$

property. The autocorrelation function ,(ACF) gives a quantitative measure of the closeness or similarity between samples of a speech signal as a function of their time separation. This function is mathematically defined as
where x(k) represents the kth speech sample.
The autocorrelation function is often normalized to the variance of the speech signal and hence is constrained to have values in the range { -1, 1} with C(0) = 1. Typical signals have an adjacent sample correlation, C(l) , as high as 0.85 to 0.9.

Power Spectral Density function (PSD) -The nonflat characteristic of the power spectral density of speech makes it possible to obtain significant compression by coding speech in the frequency domain. The nonflat nature of the PSD is basically a frequency domain manifestation of the nonzero autocorrelation property. Typical long-term averaged PSDs of speech show that high frequency components contribute very little to the total speech energy. This indicates that coding speech separately in different frequency bands can lead to significant coding gain. However, it should be noted that the high frequency

components, though insignificant in energy are very important carriers of speech information and hence need to be adequately represented in the coding system.

2.3 Vocoders

Vocoders are a class of speech coding systems that analyze the voice signal at the transmitter, transmit parameters derived from the analysis, and then synthesize the voice at the receiver using those parameters. Vocoder systems attempt to model the speech generation process as a dynamic system and try to quantify certain physical constraints of the system. These physical constraints are used to provide a parsimonious description of the speech signal. Vocoders achieve very high economy in transmission bit rate. The most popular vocoder is the linear predictive coder (LPC), which is the subject of this work.

Figure 2.2 shows the traditional speech generation model that is the basis of all vocoding systems. The sound generating mechanism forms the source and is linearly separated from the intelligence modulating vocal tract filter which forms the system. The speech signal is assumed to be of two types: voiced and unvoiced. Voiced sound ("m", "n", "v" pronunciations) are a result of quasiperiodic vibrations of the vocal chord and unvoiced sounds (“l”, "s", " sh " pronunciations) are fricatives produced by turbulent air flow through a constriction. The parameters associated with this model are the voice, the pitch frequencies of the modulating filter and the corresponding amplitude parameters, The pitch frequency for most speakers is below 300 Hz. The pole frequencies correspond to the resonant frequencies of the vocal tract and are often called the formants of the speech signal. For adult speakers, the formants are centered around 500Hz, 1500 Hz, 2500 Hz, and 3500 Hz. By meticulously adjusting the parameters of the speech generating model, good quality speech can be synthesized.

2.3 **Linear Predictive Coders**

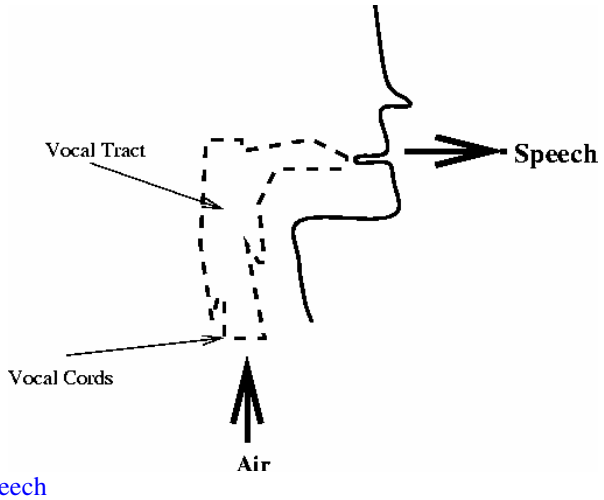


Fig 2.1 Physical Model of a human speech

When one speaks:

- Air is pushed from lung through vocal tract and out of your mouth comes speech.
- For certain *voiced* sound, your vocal cords vibrate (open and close). The rate at which the vocal cords vibrate determines the *pitch* of your voice. Women and young children tend to have high pitch (fast vibration) while adult males tend to have low pitch (slow vibration).
- For certain *fricatives and plosive (or unvoiced)* sound, your vocal cords do not vibrate but remain constantly opened.
- The shape of your vocal tract determines the sound that you make.
- As you speak, your vocal tract changes its shape producing different sound.
- The shape of the vocal tract changes relatively slowly (on the scale of 10 msec to 100 msec).

The amount of air coming from your lung determines the loudness of your voice.

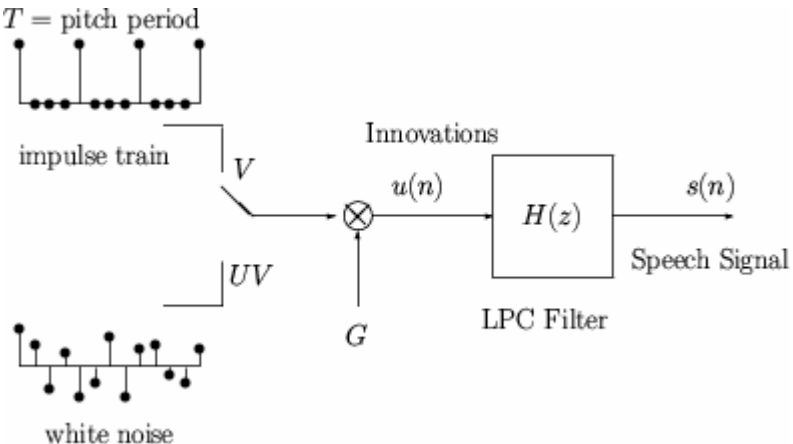


Fig 2.2 The LPC Model.

The LPC model says that the digital speech signal is the output of a digital filter (called the LPC filter) whose input is either a train of impulses or a white noise sequence.
The *relationship* between the physical and the mathematical models:

Vocal Tract	=	**H(z)** (LPC Filter)
Air	=	**u(n)**(Innovations)
Vocal Cord Vibration	=	**V** (voiced)
Vocal Cord Vibration Period	=	**T**(pitch period)
Fricatives and Plosives	=	**UV** (unvoiced)
Air Volume	=	**G** (gain)

Linear predictive coders (LPCs) attempts to extract the significant features of speech from the time waveform. With LPC, it is possible to transmit good quality voice at 4.8 kbps and poorer quality voice at even lower rates. The linear predictive coding system models the vocal tract as an all pole linear filter with a transfer function described by

$$H(z) = \frac{G}{1 + \sum b_k z^{-k}} \qquad 2.4$$

where G is a gain of the filter and z^{-1} represents a unit delay operation. The excitation to this filter is either a pulse at the pitch frequency or random white noise depending on whether the speech segment is voiced or unvoiced. The coefficients of the all pole filter are obtained in the time domain using linear prediction techniques. The code works, instead of transmitting quantized values of the error signal representing the difference between the predicted and actual waveform, the LPC system transmits only selected characteristics of the error signal. The parameters include the gain factor, pitch information, and the voiced/unvoiced decision information, which allow approximation of the

correct error signal. At the receiver, the received information about the error signal is used to determine the appropriate excitation for the synthesis filter. That is, the error signal is the excitation to the decoder. The synthesis filter is designed at the receiver using the received predictor coefficients. In practice, many LPC coders transmit the filter coefficients which already represent the error signal and can be directly synthesized by the receiver. Fig 2.2 shows a block diagram of an LPC system.

2.4 Determination of Predictor Coefficients

The linear predictive coder uses a weighted sum of p past samples to estimate the present sample, where p is typically in the range of 10–15. Using this technique, the current

$$s_n = \sum_{k=1}^{p} a_k s_{n-k} + e_n \qquad 2.5$$

sample S_n can be written as a linear sum of the immediately preceding samples S_{n-k}

where e_n is the prediction error (residual). The predictor coefficients are calculated to minimize the average energy E in the error signal that represents the difference between the predicted and actual speech amplitude

where $a_o = -1$; Typically, the error is computed for a time

$$s_n = \sum_{n=1}^{N} e_n{}^2 = \sum_{n=1}^{N} \left(\sum a_k s_{n-k} \right)^2 \qquad 2.6$$

window of 10 ms, which corresponds to a value of N = 80 .To

$$\frac{\partial E}{\partial a_m} = \sum_{n=1}^{N} 2 s_{n-m} \sum_{k=0}^{p} a_k s_{n-k} = 0 \qquad 2.7a$$

minimize E with respect to a_m, it is required to set the partial derivatives equal to zero

The inner summation can be recognized as the correlation

$$= \sum_{n=1}^{N} \sum_{k=0}^{p} a_k s_{n-m} s_{n-k} = 0 \qquad 2.7b$$

coefficient C_{rm} and hence the above equation can be rewritten as

$$\sum_{k=0}^{p} C_{mk} a_k = 0 \qquad 2.8$$

After determining the correlation coefficients C_{rm} Equation (2.8) can be used to determine the predictor coefficients. equation (2.8) is often expressed in matrix notation and the predictor coefficients calculated using matrix inversion. Normally, the predictor coefficients are not coded directly, as they would require 8 bits to 10 bits per coefficient for accurate representation. The accuracy requirements are lessened by transmitting the reflection coefficients (a closely related parameter), which have a smaller dynamic range. These reflection coefficients can be adequatly represented by 6 bits per coeficient. Thus for a 10^{th} order predictor , the total number of bits assigned to the model parameters per frame is 72, which includes 5 bits for a gain parameter and 6 bits for the pitch period. If the parameters

$$LAR_n(k) = \tanh^{-1}(R_n(k)) \log_{10}\left[\frac{1+R_n(k)}{1-R_n(k)}\right] \qquad 2.9$$

are estimated every 15 ms to 30 ms, the resulting bit rate is in the range of 2400 bps to 4800 bps. The coding of the reflection coefficients can be further improved by

performing a non linear transformation of the coefficients prior to coding. This non linear transformation reduces the sensitivity of the reflection coefficients to quantization errors. This is normally done through a log-area ratio (LAR) transform which performs an inverse hyperbolic tangent mapping of the reflection coefficients, $R_n(k)$

various LPC schemes differ in the way they recreate the error signal (excitation) at the receiver. Here used is the most popular means . It uses two sources at the receiver, one of white noise and the other with a series of pulses at the current pitch rate. The selection of either of these excitation methods is based on the voiced / unvoiced decision made at the transmitter and communicated to the receiver along with the other information. This technique requires that the transmiter extract pitch frequency.

2.5 Generalization of LPC Modeling for Software algorithm

The LPC filter is given by

$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{10} z^{-10}} \quad 2.10$$

which is equivalent to saying that the input-output relationship of the filter is given by the linear difference equation:

$$s(n) + \sum_{i=1}^{10} a_i s(n-i) = u(n) \quad 2.11$$

The LPC model can be represented in vector form as:

$$\mathbf{A} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, G, V/UV, T) \quad 2.12$$

A changes every 20 msec or so. At a sampling rate of 8000 samples/sec, 20 msec is equivalent to 160 samples.

The digital speech signal is divided into *frames* of size 20 msec. There are 50 frames/second.

The model says that

$$\mathbf{A} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, G, V/UV, T) \quad 2.13$$

is equivalent to

$$\mathbf{S} = (s(0), s(1), \ldots, s(159))$$

Thus the 160 values of **S** is compactly represented by the 13 values of **A**

There's almost no perceptual difference in if:

- **For Voiced Sounds (V):** the impulse train is shifted (insensitive to phase change).
- **For Unvoiced Sounds (UV):**} a different white noise sequence is used.

LPC Analysis: Given **S**, find the best **A**(this is described below).

LPC Synthesis: Given **A**, generate **S** (this is done using standard filtering techniques).

LPC Analysis

$$\mathbf{S} = (s(0), s(1), \ldots, s(159))$$

Consider one frame of speech signal. The signal **s(n)** is related to the innovation **u(n)** through the linear difference equation

$$s(n) + \sum_{i=1}^{10} a_i s(n-i) = u(n) \qquad 2.14$$

The ten LPC parameters are chosen to minimize the energy of the innovation

$$f = \sum_{n=0}^{159} u^2(n) \qquad 2.15$$

Using standard calculus, we take the derivative of with respect to a_i and set it to zero:

$$\begin{aligned} df/da_1 &= 0 \\ df/da_2 &= 0 \\ &\cdots \\ df/da_{10} &= 0 \end{aligned} \qquad 2.16$$

We now have 10 linear equations with 10 unknowns:

$$\begin{bmatrix} R(0) & R(1) & R(2) & R(3) & R(4) & R(5) & R(6) & R(7) & R(8) & R(9) \\ R(1) & R(0) & R(1) & R(2) & R(3) & R(4) & R(5) & R(6) & R(7) & R(8) \\ R(2) & R(1) & R(0) & R(1) & R(2) & R(3) & R(4) & R(5) & R(6) & R(7) \\ R(3) & R(2) & R(1) & R(0) & R(1) & R(2) & R(3) & R(4) & R(5) & R(6) \\ R(4) & R(3) & R(2) & R(1) & R(0) & R(1) & R(2) & R(3) & R(4) & R(5) \\ R(5) & R(4) & R(3) & R(2) & R(1) & R(0) & R(1) & R(2) & R(3) & R(4) \\ R(6) & R(5) & R(4) & R(3) & R(2) & R(1) & R(0) & R(1) & R(2) & R(3) \\ R(7) & R(6) & R(5) & R(4) & R(3) & R(2) & R(1) & R(0) & R(1) & R(2) \\ R(8) & R(7) & R(6) & R(5) & R(4) & R(3) & R(2) & R(1) & R(0) & R(1) \\ R(9) & R(8) & R(7) & R(6) & R(5) & R(4) & R(3) & R(2) & R(1) & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \\ a_{10} \end{bmatrix} = \begin{bmatrix} -R(1) \\ -R(2) \\ -R(3) \\ -R(4) \\ -R(5) \\ -R(6) \\ -R(7) \\ -R(8) \\ -R(9) \\ -R(10) \end{bmatrix}$$

where

$$\begin{aligned} R(k) &= \sum_{n=0}^{159-k} s(n)s(n+k) \\ &= \text{autocorrelation of } s(n) \end{aligned} \qquad 2.17$$

The above matrix equation could be solved using:

- The Gaussian elimination method.
- Any matrix inversion method or best with.

- The Levinson-Durbin recursion (described below).

Levinson-Durbin Recursion:

Solve the above for i=1,2,..10, and then set

$$\begin{aligned} E^{(0)} &= R(0) \\ k_i &= [\rfloor a_i = -\alpha_i^{(10)} \alpha_j^{(i-1)} R(i-j)]/E^{(i-1)} \qquad i = 1, 2, \ldots, 10 \\ \alpha_i^{(i)} &= k_i \\ \alpha_j^{(i)} &= \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \qquad j = 1, 2, \ldots, i-1 \\ E^{(i)} &= (1 - k_i^2) E^{(i-1)} \end{aligned} \tag{2.18}$$

To get the other three parameters:**V/UV,T,G**, we solve for the innovation

$$u(n) = s(n) + \sum_{i=1}^{10} a_i s(n-i) \tag{2.19}$$

Then calculate the autocorrelation of **u(n)**

$$R_u(k) = \sum_{n=0}^{159-k} u(n)u(n+k) \tag{2.20}$$

Then make a decision based on the autocorrelation

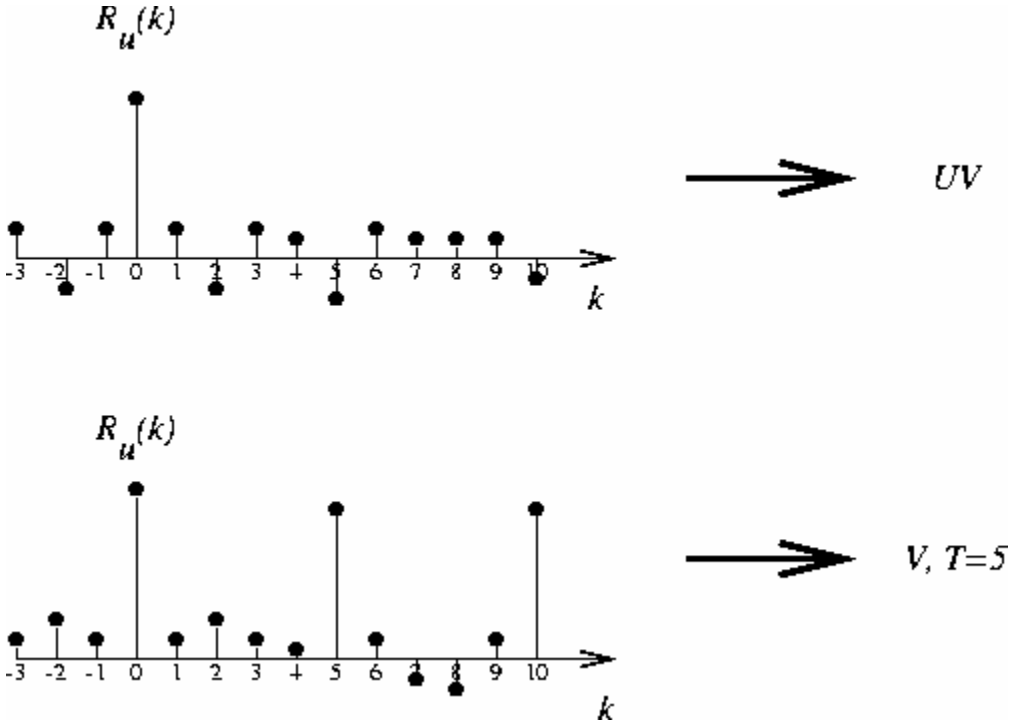


Fig 2.3 voiced and unvoiced Autocorelations

2.4kbps LPC Vocoder

- The following is a block diagram of a 2.4 kbps LPC Vocoder:

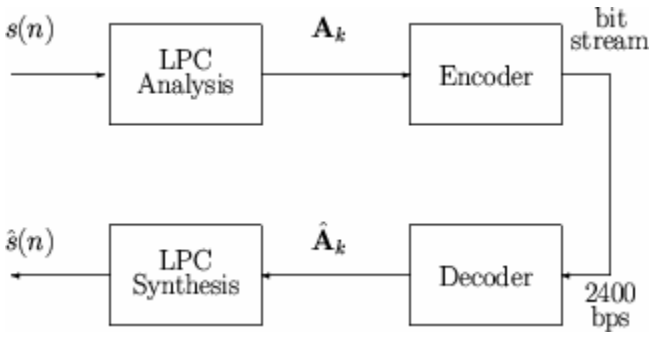


Fig 2.5 block diagram of a 2.4 kbps LPC Vocoder

The LPC coefficients are represented as *line spectrum pair* (LSP) parameters. LSP are mathematically equivalent (one-to-one) to LPC. LSP are more amenable to quantization. LSP are calculated as follows:

$$\begin{aligned} P(z) &= 1 + (a_1 - a_{10})z^{-1} + (a_2 - a_9)z^{-2} + \ldots + (a_{10} - a_1)z^{-10} - z^{-11} \\ Q(z) &= 1 + (a_1 + a_{10})z^{-1} + (a_2 + a_9)z^{-2} + \ldots + (a_{10} + a_1)z^{-10} + z^{-11} \end{aligned} \quad 2.21$$

Factoring the above equations, we get:

$$\begin{aligned} P(z) &= (1 - z^{-1}) \prod_{k=2,4,\ldots,10} (1 - 2\cos\omega_k z^{-1} + z^{-2}) \\ Q(z) &= (1 + z^{-1}) \prod_{k=1,3,\ldots,9} (1 - 2\cos\omega_k z^{-1} + z^{-2}) \end{aligned} \quad 2.22$$

$\{\omega_k\}_{k=1}^{10}$

$0 < \omega_1 < \omega_2 < \ldots < \omega_{10} < \pi$

are called the LSP parameters. LSP are *ordered* and *bounded*:

LSP are more correlated from one frame to the next than LPC. The frame size is 20 msec. There are 50 frames/sec. 2400 bps is equivalent to 48 bits/frame. These bits are allocated

Parameter Name	Parameter Notation	Rate (bits/frame)
LPC (LSP)	$\{a_k\}_{k=1}^{10}$ ($\{\omega_k\}_{k=1}^{10}$)	34
Gain	G	7
Voiced/Unvoiced & Period	$V/UV, T$	7
Total		48

as follows:

Table 2.1 Parameter of he 34 bits for the LPC are allocated as follows

Table 2.2 Parameter of the 34 bits for the LSP are allocated as follows

The gain G , is encoded using a 7-bit non-uniform scalar quantizer (a 1-dimensional vector quantizer). For voiced speech, values of ranges from 20 to 146. are jointly encoded as

LSP	No. of Bits
ω_1	3
ω_2	4
ω_3	4
ω_4	4
ω_5	4
ω_6	3
ω_7	3
ω_8	3
ω_9	3
ω_{10}	3
Total	34

follows:

V/UV	T	Encoded Value
UV	—	0
V	20	1
V	21	2
V	22	3
V	23	4
⋮	⋮	⋮
⋮	⋮	⋮
V	146	127

Table 2.3 Voiced frequency notations

CHAPTER THREE

INTERLEAVING AND CHANNEL CODING

3.1. Introduction

Interleaving is used to obtain time diversity in a digital communications system without adding any overhead. Interleaving has become an extremely useful technique in all second and third generation wireless systems, due to the rapid proliferation of digital speech coders which transform analog voices into efficient digital messages that are transmitted over wireless links .

Because speech coders attempt to represent a wide range of voices in a uniform and efficient digital format, the encoded data bits (called source bits) carry a great deal of information and some source bits are more important than others and must be protected from errors. It is typical for many speech coders to produce several "important" bits in succession, and it is the function of the interleaver to spread these bits out in time so that if there is a deep fade or noise burst, the important bits from a block of source data are not corrupted at the same time. By spreading the source bits over time, it becomes possible to make use of error control coding (called channel coding) which protects the source data from corruption by the channel. Since error control codes are designed to protect against channel errors that may occur randomly or in a bursty manner, interleavers scramble the time order of source bits before they are channel coded. An interleaver can be one of two forms ,a block structure or a convolutional structure. A block interleaver formats the encoded data into a rectangular array of m rows and n columns, and interleaves nm bits at a time. Usually, each row contains a word of source data having n bits. An interleaver of degree m (or

depth m) consists of m rows. Source bits are placed into the interleaver by sequentially increasing the row number for each successive bit, and filling the columns. The interleaved source data is then read out row-wise and transmitted over the channel. This has the effect of separating the original source bits by m bit periods.

At the receiver, the de-interleaver stores the received data by sequentially increasing the row number of each successive bit, and then clocks out the data row-wise, one word (row) at a time: Convolutional interleavers can be used in place of block interleavers in much the same fashion. Convolutional interleavers are ideally suited for use with convolutional codes. There is an inherent delay associated with an interleaver since the received message block cannot be fully decoded until all of the nm bits arrive at the receiver and are de-interleaved. In practice, human speech is tolerable to listen to until delays of greater than 40 ms Occur. It is for this reason that all of the wireless data interleavers have delays which do not exceed 40ms. The interleaver word size and depth are closely related to the type of speech coder used, the source coding rate, and the maximum tolerable delay.

3.2 Fundamentals of Channel Coding

Channel coding protects digital data from errors by selectively introducing redundancies in the transmitted data. Channel codes that are used to detect errors are called error detection codes, while codes that can detect and correct errors are called error correction codes. By properly encoding of the information, errors induced by a noisy channel can be reduced to any desired level without sacrificing the rate of information transfer. The Shannon's channel capacity formula is applicable to the AWGN channel and is given by

$$C = B\log_2(1 + P/N_oB) = B\log_2(1 + S/N)$$

3.1

where C is the channel capacity (bits per second), B is the transmission bandwidth (Hz), P is the received signal power (W), and N_o is the single-sided noise power density (W/Hz). The received power at a receiver is given as

$$P=E_bR_b$$

3.2

where E_b is the average bit energy and R_b is the transmission bit rate.

The basic purpose of error detection and error correction techniques is to introduce redundancies in the data to improve wireless link performance. The introduction of redundant bits increases the raw data rate used in the link, and, hence, it increases the bandwidth requirement for a fixed source data rate. This reduces the bandwidth efficiency of the link in high SNR conditions, but provides excellent BER performance at low SNR values. It is well known that the use of orthogonal signaling allows the probability of error to become arbitrarily small by expanding the signal set, i.e., by making the number of waveforms $M \rightarrow \infty$, provided that the SNR per bit exceeds the Shannon limit of $SNR_b \geq -1.6$ dB. In the limit, Shannon's result indicates that extremely wideband signals could be used to achieve error free communications, as long as sufficient SNR exists, and this is partly why wideband CDMA is being adopted for 3G. Error control coding waveforms, on the other hand, have bandwidth expansion factors that grow only linearly with the code block length. Error correction

coding thus offers advantages in bandwidth limited applications, and also provides link protection in power limited applications. A channel coder operates on digital message (source) data by encoding the source information into a code sequence for transmission through the channel. There are three basic types of error correction and detection codes block codes, convolution codes, and turbo codes, but block codes discussed

3.3 Block Codes and Finite Fields

Block codes are forward error correction (FEC) codes that enable a limited number of errors to be detected and corrected without retransmission. Block codes can be used to improve the performance of a communications system when other means of improvement (such as increasing transmitter power or using a more sophisticated demodulator) are impractical. In block codes, parity bits are added to blocks of message bits to make codewords or code blocks. In a block encoder k information bits are encoded into n code bits. A total of n –k redundant bits are added to the k information bits for the purpose of detecting and correcting errors. The block code is referred to as an (n, k) code, and the rate of the code is defined as R= k/n and is equal to the rate of information divided by the raw channel rate. The ability of a block code to correct errors is a function of the code distance. Many families of codes exist that provide varying degrees of error protection. Besides the code rate, other important parameters are the code-distance and the weight of particular codewords. These are defined below.

Distance of a Code –The distance between two codewords is the number of elements in which two codewords Ci and Cj differ.

The length of each codeword is N elements or characters. If the code used is binary the distance is known as the Hamming distance. The minimum distance dmin is the smallest distance for the given codeword set and is given as

$$d_{min} = Min\{d(C_i, C_i)\}$$

3.3

Weight of a Code –The weight of a codeword of length N is given by the number of nonzero elements in the codeword. For a binary code, the weight is basically the number of 1 s in the codeword

Properties of Block Codes

Linearity –Suppose C_i and C_j are two codewords in an (n, k) block code. Let α_1 and α_2 be any two elements selected from the alphabet. Then the code is said to be linear if and only if $\alpha_1 C_1 + \alpha_2 C_2$ is also a code word. A linear code must contain the all–zero code word Consequently, a constant–weight code is nonlinear.

Systematic –A systematic code is one in which the parity bits are appended to the end of the information bits. For an (n,k) code, the first k bits are identical to the information bits, and the remaining n –k bits of each code word are linear combinations of the k information bits.

Cyclic –Cyclic codes are a subset of the class of linear codes which satisfy the following cyclic shift property: If $C = [C_{n-1}, C_{n-2}, \ldots C_0]$ is a codeword of a cyclic code, then $[C_{n-2}, C_{n-3}, \ldots C_0, C_{n-1}]$, obtained by a cyclic shift of the elements of C, is also a code word. That is, all cyclic shifts of C are code words. As a consequence of the cyclic propeerty, the codes possess a considerable amount of structure which can be exploited to greatly simplify the encoding and decoding operations.

Encoding and decoding techniques make use of the mathematical constructs know as finite fields. Finite fields are algebraic systems which contain a finite set of elements. Addition, subtraction,multiplication, and division of finite field elements is accomplished without leaving the set (i.e., the sum or product of two field elements is a field element). Addition and multiplication must satisfy the commutative, associative, and distributive laws. A formal definition of a finite field is:

Let F be a finite set of elements on which two binary operations–addition and multiplication are defined. The set F together with the two binary operations is a field if the following conditions are satisfied:

- F is a commutative group under addition. The identity element with respect to addition is called the zero element and is denoted by 0.
- The set of nonzero elements in F is a commutative group under multiplication. The identity element with respect to multiplication is called the unit element and is denoted by 1.
- Multiplication is distributive over addition; that is, for any three elements a, b, and c in F:

  $a.(b+c) = a.b+a.c$

The additive inverse of a field element a, denoted by –a, is the field element which produces the sum 0 when added to a [so that a + (–a) = 0]. The multiplicative inverse of a , denoted by a^{-1} , is the field element which produces the product 1 when multiplied by a [so that a. a^{-1} = 1].

Four basic properties of fields can be derived from the definition of a field. They are as follows:

Property I: a. 0 = 0 = 0 .a

Property II: For nonzero field elements a and b, a. b ≠ 0

Property III: a. b = 0 and a ≠ 0 imply b = 0

Property IV: $-(a.\ b) = (-a)\ .b = a.\ (-b)$

For any prime number p, there exists a field which contains p elements. This prime field is denoted as GF(p) because finite fields are also called Galois fields. It is also possible to extend the prime GF(p) to a field of p^m elements Which is called an extension field of GF(p) and lS denoted by $GF(p^m)$, where m is a positive integer. Codes with symbols from the binary field GF(2) or its extension field $GF(2^m)$ are most commonly used in digital data transmission and storage systems, since information in these systems is always encoded in binary form in practice.

In binary arithmetic, modulo-2 addition and multiplication are used. This arithmetic is actually equivalent to ordinary arithmetic except that two is considered equal to 0 ie (1 + 1 = 2 = 0) .Note that since 1 + 1 = 0, it follows that 1= -1, and hence for the arithmetic used to generate error control codes, addition is equivalent to subtraction.

Reed-Solomon codes make use of nonbinary fields $GF(2^m)$. These fields have more than two elements and are extensions of the binary field GF(2) = {0, 1 }. The additional elements in the extension field $GF(2^m)$ cannot be 0 or 1, since all of the elements must be unique, so a new symbol α is used to represent the other elements in the field. Each nonzero element can be represented by a power of α.

The multiplication operation " ." for the extension field must be defined so that the remaining elements of the field can be represented as sequence of powers of α. The multiplication operation can be used to produce the infinite set of elements F shown below

$$F = \{0,\ 1,\ \alpha,\ \alpha^2,\ \ldots\ ,\ \alpha^j,\ \ldots\ \} = \{0,\ \alpha^0,\ \alpha,\ \alpha^2,\ \ldots\ ,\ \alpha^j,\ \ldots\ \} \qquad 3.4$$

To obtain the finite set of elements of GF(2^m) from F, a condition must be imposed on F so that it may contain only 2^m elements and is a closed set under multiplication (i.e.,

$$\alpha^{(2^m-1)} + 1 = 0 \qquad 3.5$$

$$\alpha^{(2^m-1)} = \alpha^0 = 1$$

multiplication of two field elements is performed without leaving the set). The condition which closes the set of field elements under multiplication is known as the irreducible polynomial, and it typically takes the following

$$\alpha^{(2^m+n)} = \alpha^{(2^m-1)} \cdot \alpha^{n+1} = \alpha^{n+1}$$

form or equivalently Using the irreducible polynomial of Equation (3.5), any element which has a power greater than 2^m – 2 can be reduced to an element with a power less than 2^m – 2 as folows:

The sequence of elements F thus becomes the following sequence F* , whose nonzero terms are closed under multiplication: 3.6

$$F = \{0, \alpha^0, \alpha, \alpha^2, \ldots , \alpha^{2m-2}, \alpha^0, \alpha, \alpha^2, \ldots \} \qquad 3.7$$

Take the first 2^m terms of F* and you have the elements of the finite field GF(2^m) in their power representation

$$GF(2^m) = \{0, \alpha^0, \alpha, \alpha^2, \ldots , \alpha^{2m-2}\} = \{0, 1, \alpha, \alpha^2, \ldots , \alpha^{2m-2}\} \qquad 3.8$$

It can be shown that each of the 2^m elements of the finite field can be represented as distinct polynomial of degree m-1 or less (the element 0 is represented by the zero polynomial, a polynomial with no nonzero terms). Each of the

nonzero elements of GF(2^m) can be denoted as a polynomial $a_i(x)$ where at least one of the m coefficients is nonzero

$$\alpha^i = a_i(x) = a_{i,0} + a_{i,1}x + a_{i,2}x^2 + a_{i,m-1}x^{m-1} \quad (3.9)$$

Addition of two elements of the finite field is then defined as the modulo-2 addition of each of the polynomial coefficients of like powers. Thus, GF(2^m) may be constructed, and using Equations (3.4)and (3.7) the polynomial representation for the 2^m elements of the field may be obtained.

3.4 Examples of Block Codes

Hamming Codes

Hamming codes were among the first of the nontrivial error correction codes. These codes and their variations have been used for error control in digital communication systems. There are both binary and nonbinary Hamming codes. A binary Hamming code has the property that,

$$(n,k) = (2^m-1, 2^m-1-m) \quad (3.10)$$

where k is the number of information bits used to form a n bit codeword, and m is any positive integer. The number of parity symbols are n-k.

Hadamard Codes

Hadamard codes are obtained by selecting as codewords the rows of a Hadamard matrix. A Hadamard matrix A is a N x N matrix of I s and Os, such that each row differs from any other row in exactly N /2 locations. One row contains all zeros with the remainder containing N/2 , zeros and N/2 ones. The minimum distance for these codes is N/2 . In addition to the special case considered above when N = 2^m (m being a positive integer), Hadamard codes of other block lengths are possible, but the codes are not linear.

Golay Codes

Golay codes are linear binary (23, 12) codes with a minimum distance of seven and a error correction capability of three bits . This is a special, one of a kind code in that this is the only nontrivial example of a perfect code. (Hamming codes and some repetition codes are also perfect.) Every codeword lies within distance three of any codeword, thus making maximum likelihood decoding possible.

Cyclic Codes ,

Cyclic codes are a subset of the class of linear codes which satisfy the cyclic property as discussed before. As a result of this property, these codes possess a considerable amount of structure which can be exploited. A cyclic code can be generated by using a generator polynomial g(p) of degree (n -k). The generator polynomial of an (n, k) cyclic code is a factor of p^n+1

Encoding for a cyclic code is usually performed by a linear feedback shift register based on either the generator or parity polynomial.

BCH Codes

BCH cyclic codes are among the most important block codes, since they exist for a wide range of rates, achieve significant coding gains, and can be implemented even at high speeds . The block length of the codes is $n = 2^m - 1$ for $m \geq 3$, and the number of errors that they can correct is bounded by $t < (2^m - 1)/2$. The binary BCH codes can be generalized to create classes of nonbinary codes which use m bits per code symbol. The most important and common class of nonbinary BCH codes is the family of codes known as Reed-Solomon codes in which this project mainly implements. The (63,47) Reed-Solomon code in US Cellular Digital Packet Data (CDPD) uses m = 6 bits per code symbol.

Reed-Solomon Codes

Reed-Solomon (RS) are nonbinary codes which are capable of correcting errors which appear in bursts and are commonly used in concatenated coding systems . The block length of these codes is $n = 2^m - 1$. These can be extended to 2^m or $2^m + 1$. The number of parity symbols that must be used to correct e errors is $n - k = 2e$ .The minimum distance $d_{min} = 2e + 1$. RS codes achieve the largest possible d_{min} of any linear code.

Implementation: Reed-Solomon Codes

In this project , the field choice will be GF(64) m = 6, so each of the 64 field elements is represented by a 6 bit symbol.
A finite field entity p(X) is introduced in order to map the 64 distinct 6-bit symbols to the elements of the field. An irreducible polynomial p(X) of degree m is said to be primitive if the , smallest integer n for which p(X) divides $x^n + 1$ is $n = 2^m - 1$ .This primitive polynomial p(X) is vital for all coding operations as it defines the field and is typically of the form

$$p(X) = 1 + X + X^m \qquad 3.11$$

Primitive polynomials have been tabulated for a wide range of field sizes, so p(X) is a known quantity and is part of the definition of any code. For m=6 the primitive polynomial is

$$P(X) = 1 + X + X^6 \qquad 3.12$$

In order to map symbols to field elements, set the primitive polynomial $p(\alpha) = 0$. This yields the following result, which closes the set of field elements:

$$1 + \alpha + \alpha^6 = 0$$

3.13

This shows the proper mapping of 6-bit symbols to field elements and this mapping is used to govern all computations. The field elements are generated by first starting with the element 1 (α^o) and then by multiplying by α to obtain the next field element. Any element which contains an α^5 term will yield an α^6 term in the next element, but α^6 is not in GF(64) . The primitive polynomial rule is used to convert α^6 to $\alpha + 1$. Also note that α^{62}. $\alpha = \alpha^{63} = \alpha^o = 1$ This simple result is critical when implementing finite field multiplication in software. Multiplication can be accomplished quickly and efficiently by using modulo $2^m - 1$ addition of the powers of the elemental operands. For the (63, 47) Reed-Solomon code multiplication of two field elements corresponds to adding the powers of the two operands modulo-63 to arrive at the power of the product.

Addition in GF(2^m) corresponds to modulo-2 adding the coefficients of the polynomial representation of the elements. Since the cgefficients are either 1s or Os (because the field is an extension of the binary field GF(2)), this can simply be implemented with the bit-wise exlusive OR of the 6-bit symbol representation of the elemental opeands. Some examples of finite field addition in GF(64) are shown below:

$\alpha^{27} + \alpha^5 = (001110)_2 \text{XOR} (100000)_2 = (101110)_2 = \alpha^{55}$

$\alpha^{19} + \alpha^{62} = (011110)_2 \text{XOR} (100001)_2 = (111111)_2 = \alpha^{58}$

3.5 Reed-Solomon Encoding

In the discussion of a Reed-Solomon encoder, the following polynomials are used frequently ,

d(x) : raw information polynomial

p(x) : parity polynomial

c(x) : codeword polynomial

g(x) : generator polynomial

q(x) : quotient polynomial

r(x) : remainder polynomial

Let $d(x) = c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \ldots + c_{2t+1}x^{2t+1} + c_{2t}x^{2t}$ 3.14

Be the information polynomial which is received before decoding and represents the user data, let

$$P(x)=c_o+C_1x+ \ +c_{2t-1}x^{2t-1} \quad 3.15$$

Be the parity polynomial(ci are all elements of GF(64)). The encoded RS polynomial can thus be expressed as

$$c(X) = d(x)+p(x) \quad 3.16$$

A vector of n field elements (c_o, c_1, c_2,... c_{n-1}) is a codeword if and only if it is a multiple of the generator polynomial g(x). The generator polynomial for a t-error, correcting, Reed-Solomon code has the form ,

$$g(x) = (x+\alpha)(x+\alpha^2)\ldots (x+\alpha^{2t}) \quad 3.17$$

A common method for encoding a cyclic code like an RS code is to derive p(x) by dividing d(x) by g(x). This yields an irrelevant quotient polynomial q(x) and an important polynomial r(x) as follows:

$$d(x) = g(x)q(x) + r(x) \quad (3.17)$$

Thus, using equation 3.17 the codeword polynomial can be expressed as

$$c(x) = p(x) + g(x)q(x) + r(x) \quad (3.18)$$

If the parity polynomial is defined as being equal to the negatives of the coefficients of r (x), then it follows that ,

$$c(x) = g(x)q(x) \quad (3.19)$$

Thus, by ensuring that the codeword polynomial is a multiple of the generator polynomial, a Reed-Solomon encoder can be constructed by using Equation (3.17) to obtain p(X). The key to encoding and decoding is to find r(x),the remainder polynomial, which maps to the transmitted data, A straightforward method of obtaining the remainder from the division process by using the monic polynomial g(x) is to connect a shift register according to g(x) as shown in Figure 3.1. Each "+" represents an exclusive-OR of two m-bit numbers, eacb X represents a multiplication of two m-bit numbers under GF(2^m) where each m-bit register contains an m-bit number denoted by b_i,

Initially, all registers are set to 0, and the switch is set to the data position. Code symbols C_{n-1} through C_{n-k} are sequentially shifted into the circuit and simultaneously transmitted to the output line. As soon as code symbol C_{n-k} enters the circuit, the switch is flipped to the parity position and the gate to the feedback network is opened so that no further feedback is provided. At that same instant. the registers b_o through b_{2t-1} contain the parity symbols P_o through P_{2t-1} which correspond directiy to the coefficients

of the parity polynomial. They can be sequentially shifted to the output to complete the Reed-Solomon encoding process.

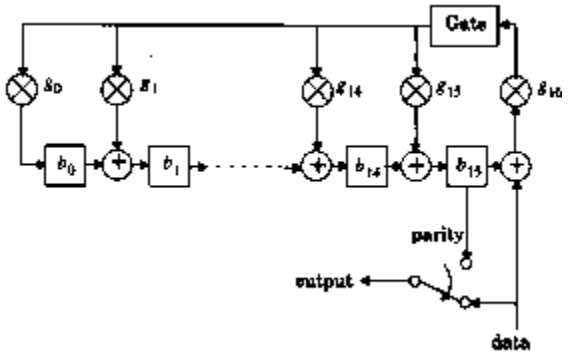


Figure 3.1 Reed Solomon Encoding Circuit

3.6 Reed-Solomon Decoding

Suppose that a codeword

$$c(x) = v_o + v_1 x + \ldots + v_{n-1} x^{n-1} \quad (3.20)$$

is transmitted and that channel errors result in the received codeword

$$r(x) = r_o + r_1 x + \ldots + r_{n-1} x^{n-1} \quad (3.21)$$

The error pattern e(x) is the difference between c(x) and r(x) using Equations (3.20) and (3.21),

$$e(x) = r(x) - c(x) = e_o + e_1 x + \ldots + e_{n-1} x^{n-1} \quad (3.22)$$

Let the 2t partial syndromes S_i, $1 < i \leq 2t$, be defined as $S_i = r(\alpha^i)$. Since $\alpha^1, \alpha^2, \ldots, \alpha^{2t}$ are roots of each transmitted codeword c(x) (because each codeword is a multiple of the generator polynomial g(x)), it follows that $c(\alpha^i) = 0$. and $S_i = c(\alpha^i) + e(\alpha^i) = e(\alpha^i)$. Thus, it is clear that the 2t

partial syndromes S_i depend only on the error pattern e(x) and not on the specific received codeword r(x).
Suppose the error pattern contains k errors ($k \leq t$) at locations x^{j1}, x^{j2}, ..., x^{jk}, where $0 \leq j_1 < j_2 < \ldots < j_k < n-1$ .Let the error magnitude at each location x^{ji} be denoted as e_j. Then e(x) has the form

$$e(x) = e_{j1}x^{j1} + e_{j2}x^{j2} + \ldots + e_{jk}x^{jk} \qquad 3.23$$

Define the set of error locator numbers $\beta_i = \alpha^{ji}$, i = 1, 2, , k. Then the set of 2t partial syndromes yields the following system of equations:

$S_1 = e_{j1}\beta_1 + e_{j2}\beta_2 + \ldots + e_{jk}\beta_k$
$S_2 = e_{j1}\beta_1^2 + e_{j2}\beta_2^2 + \ldots + e_{jk}\beta_k^2$
……..
$S_{2t} = e_{j1}\beta_1^{2t} + e_{j2}\beta_2^{2t} + \ldots + e_{jk}\beta_k^{2t}$

Any algorithm which solves this system of equations is a Reed-Solomon decoding algorithm. The error magnitudes e_{ji} are found directly, and the error locations x^{ji} can be determined from β_i.

Reed-Solomon decoders can be implemented in hardware, software or a mix of hardware and software. Hardware implementations are typically very fast, but they cannot be used for a wide variety of Reed-Solomon code sizes. For example, there are several single-chip, Reed-Solomon decoders available which decode codes commonly used in satellite communications, digital video applications, and compact disk technology. Hardware decoders can operate at speeds in excess of 50 Mbits per second, but specific hardware solutions operate with specific symbol definitions, such as 8-bit or 9-bit symbols from GF(255) , whereas the(63,47)Reed-Solomon code operates with a 6-bit symbol.

Since most wireless comm operates at a slow data rate of 19.2 kbps, a real-time software implementation of a (63,47) Reed-Solomon decoder can be achieved. A software approach may be more attractive to a modern developer, because it will have a shorter development time, lower development cost, and greater flexibility. That is why I chose software implementation for this project work.

For this typical Reed-Solomon decoder I used five distinct algorithms. But the code can be expanded to implement any (n,k) BCH codes.

- The first algorithm calculates the 2t partial syndromes S_i.
- The second step in the RS decoding process is the Berlekamp- Massey algorithm, which calculates the error locator polynomial $\sigma(x)$. This polynomial is a function of the error locations in the received codeword r(x), but it does not directly indicate which symbols of the received codeword are in error.
- A Chien search algorithm is then used to calculate these specific error locations from the error locator polynomial.
- The fourth step in the decoding process is the calculation of the magnitude of the error at each location.
- Finally, knowing both the error locations in the received codeword and the magnitude of the error at each location, an error correction algorithm may be implemented to correct up to t errors successfully .

Syndrome Calculation The syndrome of a cyclic code is typically defined as the remainder obtained by dividing the received codeword r(x) by the generator polynomial g(x).

However, for Reed-Solomon codes, 2t partial syndromes are computed. Each partial syndrome Si is defined as the remainder obtained when dividing the received codeword r(x) by x + α^i

$$S_i = \text{rem}[r(x)/(x + \alpha^i)], \quad i = 1, 2, \ldots, 2t \tag{3.24}$$

The division of two polynomials results in a quotient polynomial q(x) and a remainder polynomial rem(x). The degree of the remainder rem(x) must be less than the degree of the dividing polynomial p(x). Thus, if p(x) has degree 1 (i.e., p(x) = x + α^i), rem(x) must have degree 0. In other words, rem(x) is simply a field element and can be denoted as rem. Thus, the determination of the 2t partial syndromes begins with the calculation

$$r(x) = q(x).(x+\alpha^i) + \text{rem} \tag{3.25}$$

Letting x = α^i

$$r(x) = q(\alpha^i).(\alpha^i + \alpha^i) + \text{rem}$$
$$= \text{rem} = S_i$$

Thus, the calculation of the 2t partial syndromes S_i can be simplified from a full-blown polynomial division (which is computationally intense) to merely evaluating the received polynomial r(x) at x = α^i

$$S_i = r(\alpha^i), \quad i = 1, 2, \ldots, 2t \tag{3.26}$$

where

$$r(x) = r_o + r_1 x + \ldots + r_{n-1} x^{n-1} \quad (3.27)$$

Thus, $r(\alpha^i)$ has the form

$$r(\alpha^i) = r_o + r_1 \alpha^i + \ldots + r_{n-1} \alpha^{i(n-1)} \quad (3.28)$$

The evaluation of $r(\alpha^i)$ can be implemented very efficiently in the software by arranging the function so that it has the following form

$$r(\alpha^i) = (\ldots ((r_{n-1}\alpha^i + r_{n-2})\alpha^i + r_{n-3})\alpha^i + \ldots)\alpha^i + r_o \quad (3.29)$$

Error Locator Polynomial Calculation

The Reed-Solomon decoding process is simply any implementation which solves Equations 3.11 through 3.29. These 2t equations are symmetric function in $\beta_1, \beta_2, \ldots, \beta_k$ known as power-sum symmetric functions. We now define the polynomial

$$\sigma(x) = (1 + \beta_1 x)\ (1 + \beta_2 x) \ldots (1 + \beta_k x) \quad (3.30)$$

The roots of $\sigma(x)$ are $\beta_1^{-1}, \beta_2^{-1}, \ldots, \beta_k^{-1}$, which are the inverses of the error location numbers β_i. Thus, $\sigma(x)$ is called the error locator polynomial because it indirectly contains the exact locations of each of the errors in r(x). Note that $\sigma(x)$ is an unknown polynomial whose coefficients must also be determined during the Reed-Solomon decoding process. The coefficients of $\sigma(x)$ and the error-location numbers β_i are related by the following equations

$\alpha_0 = 1$

$\alpha_1=\beta_1 + \beta_2 + ... + \beta_k$

$\alpha_2=\beta_1\beta_2 + \beta_2\beta_3 + ... + \beta_{k-1}\beta_k$

.....

$\alpha_k=\beta_1\beta_2\beta_3 ... \beta_k$

The unknown quantities α_i and β_I can be related to the known partial syndromes S_j by the following set of equations known as Newton's Identities.

$S_1 + \alpha_1 = 0$

$S_2 + \alpha_1 S_1 + 2\alpha_2 = 0$

$S_3 + \alpha_2 S_2 + \alpha_1 S_1 + 2\alpha_3 = 0$

$S_k + \alpha_1 S_{k-1} + ... + \alpha_{k-1} S_1 + k\alpha_k = 0$

The most common method for determining $\sigma(x)$ is the Berlekamp-Massey algorithm .

CHAPTER FOUR

MODULATION TECHNIQUES

4.1 Introduction

Modulation is the process of encoding information from a message source in a manner suitable for transmission. It generally involves translating a baseband message signal (called the source) to a bandpass signal at frequencies that are very high when compared to the baseband frequency. The bandpass signal is called the modulated signal and the baseband message signal is called the modulating signal. Modulation may be done by varying the amplitude, phase, or frequency of a high frequency carrier in accordance with the amplitude of the message signal. Demodulation is the process of extracting the baseband message from the carrier so that it may be processed and interpreted by the intended receiver (also called the sink). There are various modulation techniques that are used in mobile communication systems. Since digital modulation offers numerous benefits and is already being used to replace conventional analog systems, the primary emphasis of this project is on digital modulation schemes. However, analog systems are in widespread use, and will continue to exist. Here, the coverage focuses on modulation and demodulation as it applies to mobile radio systems. A large variety of modulation techniques have been used in mobile radio communications systems. Given the hostile fading and multipath conditions in the mobile radio channel, designing a modulation scheme that is resistant to mobile channel impairments is the task of this thesis. Since the ultimate goal of a modulation technique is to transport the message signal through a radio channel with the best possible quality while occupying the least amount of radio spectrum, new advances in digital signal processing continue to bring about new forms of modulation and demodulation. I tried to simulate the vastly used Quadratic phase shift keying practical modulation scheme, receiver architecture, design tradeoffs, and the performance under various types of channel impairments.

4.2 Digital Modulation

Modern mobile communication systems use digital modulation techniques. Advancements in very large scale integration and digital signal processing technology have made digital modulation more cost effective than analog transmission systems. Digital modulation offers many advantages over analog modulation. Some advantages include greater noise immunity and robustness to channel impairments, easier multiplexing of various forms of information (eg voice, data, and video), and greater security. Furthermore, digital transmissions accommodate digital error control codes which detect and/or correct transmission errors, and support complex signal conditioning and processing techniques such as source coding, encryption, and equalization to improve the performance of the overall communication link. Multipurpose programmable digital signal processing have made it possible to implement digital modulators and demodulators completely in software. Instead of having a particular modem design permanently frozen as hardware, embedded software implementations now allow alterations and improvements without having to redesign or replace the modem.

Example is phase shift keying (PSK) which may be either coherently or differentially detected ; and may have two, four, eight or more possible levels (states) (eg n=1,2,3, or more bits) per symbol, depending on the manner in which information is transmitted within a single symbol.

Several factors influence the choice of digital modulation scheme. A desirable modulation scheme provides low bit error rates at low received signal to noise ratio, performs well in multipath and fading conditions, occupies a minimum of bandwidth, and easy and cost effective to implement. The performance of a modulation scheme is often measured in terms of its power efficiency and Bandwidth efficiency.

In the design of a digital communication system, very often there is a tradeoff between bandwidth efficiency and power efficiency. For example adding error control coding to a message increases the bandwidth occupancy (and this, in turn, reduces the bandwidth efficiency), but at the same time reduces the required received power for a particular bit error rate, and hence trades bandwidth efficiency for power efficiency. On the other hand,

higher level modulation schemes decrease bandwidth occupancy but increase the required received power, and hence trade power efficiency for bandwidth efficiency.

While power and bandwidth efficiency considerations are very important, other factors also affect the choice of a digital modulation scheme. For example, for all personal communication systems which serve a large user community, the cost and complexity of the subscriber receiver must be minimized, and a modulation which is simple to detect is most attractive. The performance of the modulation scheme under various types of channel impairments such as Rayleigh and Ricean fading and multipath time dispersion, given a particular demodulator implementation, is another key factor in selecting a modulation. In cellular systems where interference is a major issue, the performance of a modulation scheme in an interference environment is extremely important. Sensitivity to detection of timing jitter, caused by time-varying channels, is also an important consideration in choosing a particular modulation scheme. In general, the modulation, interference, and implementation of the time varying effects of the channel as well as the performance of the specific demodulator are analyzed as a complete system using simulation to determine relative performance and ultimate selection.

For a given S/N = 20dB = 100 and RF Bandwidth B=30000 Hz Using Shannon's channel capacity formula the maximum possible data rate

$C = B \log_2(1+S/N) = 199.75$ kbps

The United States Digital Communication data rate is 48.6 kbps, which is only about one fourth the theoretical limit under 20 dB SNR conditions.

For a given GSM S/N = 10dB = 10 and RF Bandwidth B=200 kHz Using Shannon's channel capacity formula the maximum possible data rate

$C = B \log_2(1+S/N) = 691.886$ kbps

The GSM data rate is 270.833 kbps, which is only about 40% the theoretical limit for 10 dB SNR conditions.

4.3 Geometric Representation of Modulation Signals

Digital modulation involves choosing a particular signal waveform $s_i(t)$, from a finite set of possible signal waveforms (or symbols) based on the information bits applied to the

modulator. For binary modulation schemes, a binary information bit is mapped directly to a signal, and S will contain only two signals. For higher-level modulation schemes (M-ary keying) the signal set will contain more than two signals, and each signal (or symbol) will represent more than a single bit of information per symbol. With a signal set size M, it is possible to transmit a maximum of $\log_2$ M bits of information per symbol.

It is instructive to view the elements of S as points in a vector space. The vector space representation of modulation signals provides valuable insight into the performance of particular modulation schemes. The vector space concepts are extremely general and can be applied to any type of modulation.

Basic to the geometric viewpoint is the fact that any finite set of physically realizable waveforms in a vector space can be expressed as a linear combination of N orthonormal waveforms which form the basis of that vector space. To represent the modulation signals on a vector space, one must find a set of signals that form a basis for that vector space. Once a basis is determined any point in that vector space can be represented as a linear combination of the basis signals $\{\phi_j(t)/j = 1,2,\ldots, N\}$ The basis signals are orthogonal to one another in time Each of the basis signals is normalized to have unit energy. The basis signals can be thought of as forming a coordinate system for the vector space. The Gram-Schmidt procedure provides a systematic way to obtaining the basis signals for a given set of signals[1].

For Example, consider the set of BPSK signals $s_t(t)$ and $s_t(t)$ given by

$$s_{BPSK}(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \theta_c) \qquad 0 \le t \le T_b (binary 1) \tag{4.2}$$

or

$$s_{BPSK}(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \pi + \theta_c)$$

$$= -\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \theta_c) \qquad 0 \le t \le T_b (binary.0) \tag{4.3}$$

Where Eb is the energy per bit, Tb is the bit period, and a rectangular pulse shape is assumed.$\phi i(t)$ for this signal set simply consists of a single waveform $\phi_1(t)$.

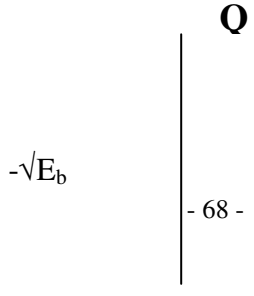
Using the basis signal, the BPSK signal set can be represented as 4.4

$$S_{BPSK}=\{\sqrt{E_b}\phi_1(t), -\sqrt{E_b}\phi_1(t)\}$$

This signal set can be shown geometrically in Figure 4.1. such a representation in called a constellation diagram which provides a graphical representation of the complex envelope of each possible symbol state. The x-axis of a constellation diagram represents the in-phase component of the complex envelope, and the y-axis represents the quadrature component of the complex envelope. The distance between signals on a constellation diagram relates to how different the modulation waveforms are, and how well a receiver can differentiate between all possible symbols when random noise is present.

It should be noted that the number of basis signals will always be less than or equal to the number of signals in the set. The number of basis signals required to represent the complete modulation signal set is called the dimension of the vector space. If there are as many basis signals as there are signals in the modulation signal set, then all the signals in the set are necessarily orthogonal to one another. For example, the bandwidth occupied by the modulation signals decreases as the number of signal points/dimension increases. Therefore, if a modulation scheme has constellation that is densely packed, it is more bandwidth efficient than modulation scheme with a sparsely packed constellation. However, it should be noted that the bandwidth occupied by a modulated signal increases with the dimension N of the constellation.

The probability of bit error is proportional to the distance between the closest points in the constellation. This implies that a modulation scheme with a constellation that is densely packed is less energy efficient that a modulation scheme that has a sparse constellation.



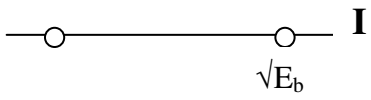


Fig 4.1 constellation diagram

A simple upper bound for the probability of symbol in an additive white Gaussian noise channel (AWGN) with a noise spectral density N_o for an arbitrary constellation can be obtained using the union bound [1]. The union bound provides a representative estimate for the average probability of error for a particular modulation signal, $P_s(\varepsilon \backslash s_i)$

$$P_s(\varepsilon \mid s_i) \leq \sum Q\left(\frac{d_{ij}}{\sqrt{2N_0}}\right) \qquad 4.5$$

Where d_{ij} is the Euclidean distance between the ith and jth signal point in the constellation. And Q(x) is the Q-function defined as

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx \qquad 4.6$$

If all of the M modulation waveforms are equally likely to be transmitted, then the average probability of error for a modulation can be estimated by

$$P_s(\varepsilon) = P_s(\varepsilon \mid s_i)P(s_i) = \frac{1}{M}\sum_{i=1}^{M} P_s(\varepsilon \mid s_i) \qquad 4.7$$

For symmetric constellations, the distance between all constellation points are equivalent, and the conditional error probability $Ps(E\backslash s_i)$ is the same for all i. Hence the above Equation 4.7 gives the average probability of symbol error for a particular constellation set.

4.3 Linear Modulation Techniques

Digital modulation techniques may be broadly classified as linear and non linear. In linear modulatation techniques, the amplitude of the transmitted signal s(t) varies linearly with the modulating signal m(t). Linear modulation techniques are bandwidth efficient and hence are attractive for use in wireless communication systems where there is an

increasing demand to accommodate more and more users within a limited spectrum.

In a linear modulation scheme, the transmitted signal s(t) can be expressed as

$$s(t) = Re[Am(t)\exp(j2\pi f_c t)] = A[m_R(t)\cos(2f_c t) - m_I(t)\sin(2\pi f_c t)] \quad (4.8)$$

where A is the amplitude, f_c is the carrier frequency, and $m(t) = m_R(t) + jm_I(t)$ is a complex envelope representation of the modulated signal which is in general complex form. From Equation (4.8), it is clear that the amplitude of the carrier varies linearly with the modulating signal. Linear modulation schemes, in general, do not have a constant envelope. As shown subsequently, some nonlinear modulations may have either linear or constant carrier envelopes, depending on whether or not the baseband waveform is pulse shaped. While linear modulation schemes have very good spectral efficiency, they must be transmitted using linear RF amplifiers, which have poor power efficiency. Using power efficient nonlinear amplifiers leads to the regeneration of filtered sidelobes, which can cause severe adjacent channel interference, and results in the loss of all the spectral efficiency gained by linear modulation. However, clever ways have been developed to get around these difficulty. The most popular linear modulation techniques include pulse-shaped QPSK, OQPSK and $\pi/4$ QPSK. Even if this work emphasizes $\pi/4$ QPSK for simulation , all are discussed subsequently.

4.4 Binary Phase Shift Keying(BPSK)

In binary phase shift keying (BPSK) the phase of a constant amplitude carrier signal is switched between two values

$$s_{BPSK}(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \theta_c) \qquad 0 \le t \le T_b (binary 1)$$

or

$$s_{BPSK}(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \pi + \theta_c)$$

$$= -\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \theta_c) \qquad 0 \le t \le T_b (binary .0)$$

according to the two possible signals m_1 and m_2 corresponding to binary 1 and 0, respectively. Normally, the two phases are separated by 180° .If the sinusoidal carrier has amplitude A_c and energy per bit $E_b = 1/2A_c^2T_b$, then the transmitted BPSK signal is either

It is often convenient to generalize m_1 and m_2 as a binary data signal m(t) , which takes on one of two possible pulse shapes. Then the transmitted signal may be represented as

$$s_{BPSK}(t) = m(t)\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \theta_c) \qquad 4.11$$

The BPSK signal is equivalent to a double sideband suppressed carrier amplitude modulated waveform, where $\text{Cos}(2\pi f_c t)$ is applied as the carrier, and the data signal m(t) is applied as the modulating waveform. Hence a BPSK signal can be generated using a balanced modulator.

Spectrum and Bandwidth of BPSK

The BPSK signal using a polar baseband data waveform m(t) can be expressed in complex envelope form as

$$S_{BPSK}(t) = \text{Re}\{g_{BPSK}(t)\exp(j2\pi f_c t)\} \qquad 4.12$$

where $g_{BPSK}(t)$ is the complex envelope of the signal given

$$g_{BPSK}(t) = \sqrt{\frac{2E_b}{T_b}}m(t)e^{j\theta_c} \qquad 4.13$$

by,

The power spectral density (PSD) of the complex envelope can be shown to be

$$P_{g_{BPSK}}(f) = 2E_b\left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2 \qquad 4.14$$

The PSD for the BPSK signal at RF can be evaluated by translating the baseband spectrum to the carrier frequency using the relation given in Equation (4.15).

Hence, the PSD of a BPSK signal at RF, is given

$$P_{BPSK}(f) = \frac{E_b}{2}\left[\left(\frac{\sin\pi(f-f_c)T_b}{\pi(f-f_c)T_b}\right)^2 + \left(\frac{\sin\pi(-f-f_c)T_b}{\pi(-f-f_c)T_b}\right)^2\right] \quad 4.15$$

The PSD of the BPSK signal for both rectangular and raised cosine rolloff pulse shapes is plotted in Figure 4.2. The null-to-null bandwidth is found to be equal to twice the bit rate (BW = $2R_b$ = $2/T_b$). From the plot, it can also be shown that 90% of the BPSK signal energy is contained within a bandwidth approximately equal to $1.6R_b$ for rectangular pulses, and all of the energy is within $1.5R_b$ for pulses with α= 0.5 raised cosine filtering.

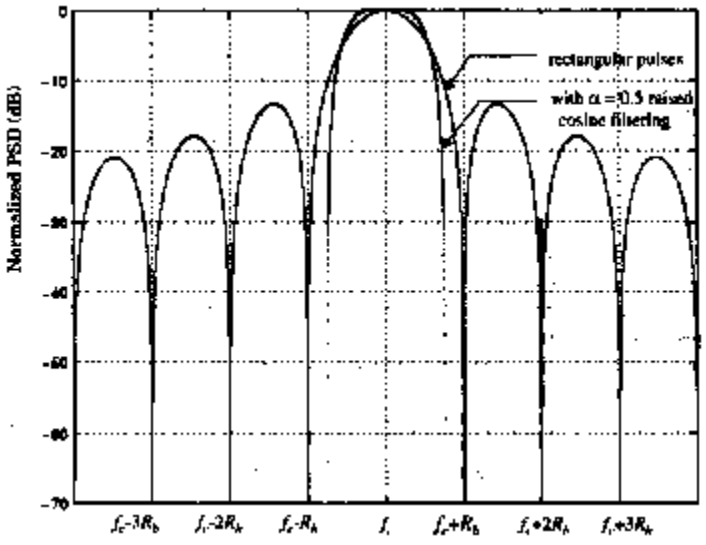


Figure 4.2 Power spectral density (PSD) of a BPSK signal

BPSK Receiver

If no multipath impairments are induced by the channel, the received BPSK signal can be expressed as

$$s_{BPSK}(t) = m(t)\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \theta_c + \theta_{ch}) \quad 4.16$$

$$s_{BPSK}(t) = m(t)\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \theta) \quad 4.17$$

where θ_{ch} is the phase shift corresponding to the time delay in the channel. BPSK uses coherent or synchronous demodulation, which requires that information about the phase and frequency of the carrier be available at the receiver. If a low level pilot carrier signal is transmitted along with he BPSK signal, then the carrier phase and frequency may be recovered at the receiver using a phase locked loop (PLL). If no pilot carrier is transmitted, a Costas loop or squaring loop may be used to synthesize the carrier phase and frequency from the received BPSK signal.
The received signal $\cos(2\pi f_c t + \theta)$ is squared to generate a

$$m(t)\sqrt{\frac{2E_b}{T_b}}\cos^2(2\pi f_c t+\theta) = m(t)\sqrt{\frac{2E_b}{T_b}}\left[\frac{1}{2}+\frac{1}{2}\cos 2(2\pi f_c t+\theta)\right] \qquad 4.18$$

dc signal and an amplitude varying sinusoid at twice the carrier frequency. The dc signal is filtered out using a bandpass filter with center frequency tuned to $2f_c$. A frequency divider is then used to recreate the waveform $\cos(2\pi f_c t + \theta)$. The output of the multiplier after the frequency divider is given by 4.18

This signal is applied to an integrate and dump circuit which forms the low pass filter segment of a BPSK detector. If the transmitter and receiver pulse shapes are matched, then the detection will be optimum. A bit synchronizer is used to facilitate sampling of the integrator output precisely at the end of each bit period. At the end of each bit period, the switch at the output of the integrator closes to dump the output signal to the decision circuit. Depending on whether the integrator output is above or below, a certain threshold, the decision circuit decides

that the received signal corresponds to a binary 1 or 0. The threshold is set at an optimum level such that the probability of error is minimized. If it is equally likely that a binary 1 or 0 is transmitted, then the voltage level corresponding to the midpoint between the detector output voltage levels of binary 1 and 0 is used as the optimum threshold.

4.5 Differential Phase Shift Keying (DPSK)

Differential PSK is a noncoherent form of phase shift keying which avoids the need for a coher ent reference signal at the receiver. Noncoherent receivers are easy and cheap to build, and hence are widely used in wireless communications. In DPSK systems, the input binary sequence is first differentially encoded and then modulated using a BPSK modulator. The differentially encoded sequence $\{d_k\}$ is generated from the input binary sequence $\{m_k\}$ by complementing the modulo-2 sum of m_k and d_{k-1}. The effect is to leave the symbol d_k unchanged from the previous symbol if the incoming binary symbol m_k is 1, and to toggle d_k if m_k is 0. The generation of a DPSK signal for a sample sequence m_k which follows the relationship $d_k = m_k \oplus d_{k-1}$ It consists of a one bit delay element and a logic circuit interconnected so as to generate the differentially encoded sequence from the input binary sequence. The output 1s passed through a product modulator to obtain the DPSK signal. At the receiver, the original sequence is recovered from the demodulated differentially encoded signal through a complementary process.

while DPSK signaling has the advantage of reduced receiver complexity, its energy efficiency is inferior to that of coherent PSK by about 3 dB. The average probability of error for DPSK in additive white Gaussian noise is given by

$$P_{e,DPSK} = \tfrac{1}{2}\exp(-E_b/N_o) \qquad 4.19$$

4.6 Quadrature Phase Shift Keying (QPSK)

Quadrature phase shift keying (QPSK) has twice the bandwidth efficiency of BPSK, since two bits are transmitted in a single modulation symbol. The phase of the carrier takes on one of four equally spaced values, such as 0, $\pi/2$, π, and

$$s_{QPSK}(t) = \sqrt{\frac{2E_S}{T_S}}\cos\left[2\pi f_c t + (i-1)\frac{\pi}{2}\right] \qquad 4.20$$

$3\pi/2$, where each value of phase corresponds to a unique pair of message bits. The QPSK signal for this set of symbol states may be defined

Where $0 \le t \le T_s$ $\quad i = 1,2,3,4$

where T_s is the symbol duration and is equal to twice the bit period. Using trigonometric identities, the above equations

$$s_{QPSK}(t) = \sqrt{\frac{2E_S}{T_S}}\cos\left[(i-1)\frac{\pi}{2}\right]\cos(2\pi f_c t) - \sqrt{\frac{2E_S}{T_S}}\sin\left[(i-1)\frac{\pi}{2}\right]\sin(2\pi f_c t) \qquad 4.21$$

can be rewritten for the interval $0 \le t \le T_s$ as

If basis functions $\Phi_1(t) = \sqrt{2/T_s}\cos(2\pi f_c t)$, $\Phi_2(t) = \sqrt{2/T_s}\sin(2\pi f_c t)$ are defined over the interval $0 \le t \le T_s$ for the QPSK signal set, then the four signals in the set can be expressed in terms of the basis signals as

4.22

$$s_{QPSK}(t) = \sqrt{E_s}\cos\left[(i-1)\frac{\pi}{2}\right]\phi_1(t) - \sqrt{E_s}\sin\left[(i-1)\frac{\pi}{2}\right]\phi_2(t)$$

Where i = 1, 2, 3, 4

Based on this representation, a QPSK signal can be depicted using a two-dimensional constellation diagram with four points as shown in Figure 4.3(b).

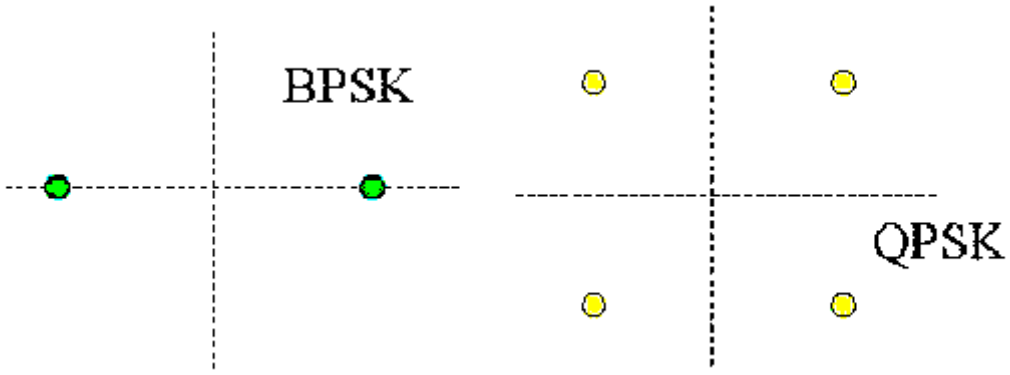


Figure 4.3 constellation BPSK and QPSK signal set

It should be noted that different QPSK signal sets can be derived by simply rotating the constellation. As an example, Figure 4.3(b) shows another QPSK signal set where the phase values are $\pi/4$, $3\pi/4$, $3\pi/4$, and $7\pi/4$. From the constellation diagram of a QPSK signal, it can be seen that the distance between – adjacent points in the constellation is $\sqrt{2E_s}$ .Since each symbol corresponds to two bits. then $E_s = 2\ E_b$, thus the distance between two neighboring points in the QPSK constellation is equal to $2\sqrt{E_b}$. Substituting this in the union bound estimation the average probability of bit err 4.23 in the additive white Gaussian noise (AWGN) channel is obtained as

$$P_{e.QPSK} = Q\left(\sqrt{\frac{2E_b}{N_o}}\right)$$

A striking result is that the bit error probability of QPSK is identical to BPSK, but twice as much data can be sent in the same bandwidth. Thus when compared to BPSK, QPSK provides twice the spectral efficiency with exactly the same energy efficiency. Similar to BPSK, QPSK can also be differentially encoded to allow noncoherent detection

Spectrum and Bandwidth of QPSK Signals

The power spectral density of a QPSK signal can be obtained in a manner similar to that used for BPSK. with the bit periods T_b replaced by symbol periods T_s. Hence, the PSD of a

$$P_{QPSK}(f) = \frac{E_s}{2}\left[\left(\frac{\sin\pi(f-f_c)T_s}{\pi(f-f_c)T_s}\right)^2 + \left(\frac{\sin\pi(-f-f_c)T_s}{\pi(-f-f_c)T_s}\right)^2\right] \qquad 4.24$$

QPSK signal using rectangular pulses can be expressed as

The PSD of a QPSK signal for rectangular and raised cosine filtered pulses is plotted in Figure 4. 4. The null-to-null RF bandwidth is equal to the bit rate R_b, which is half that of a BPSK signal.

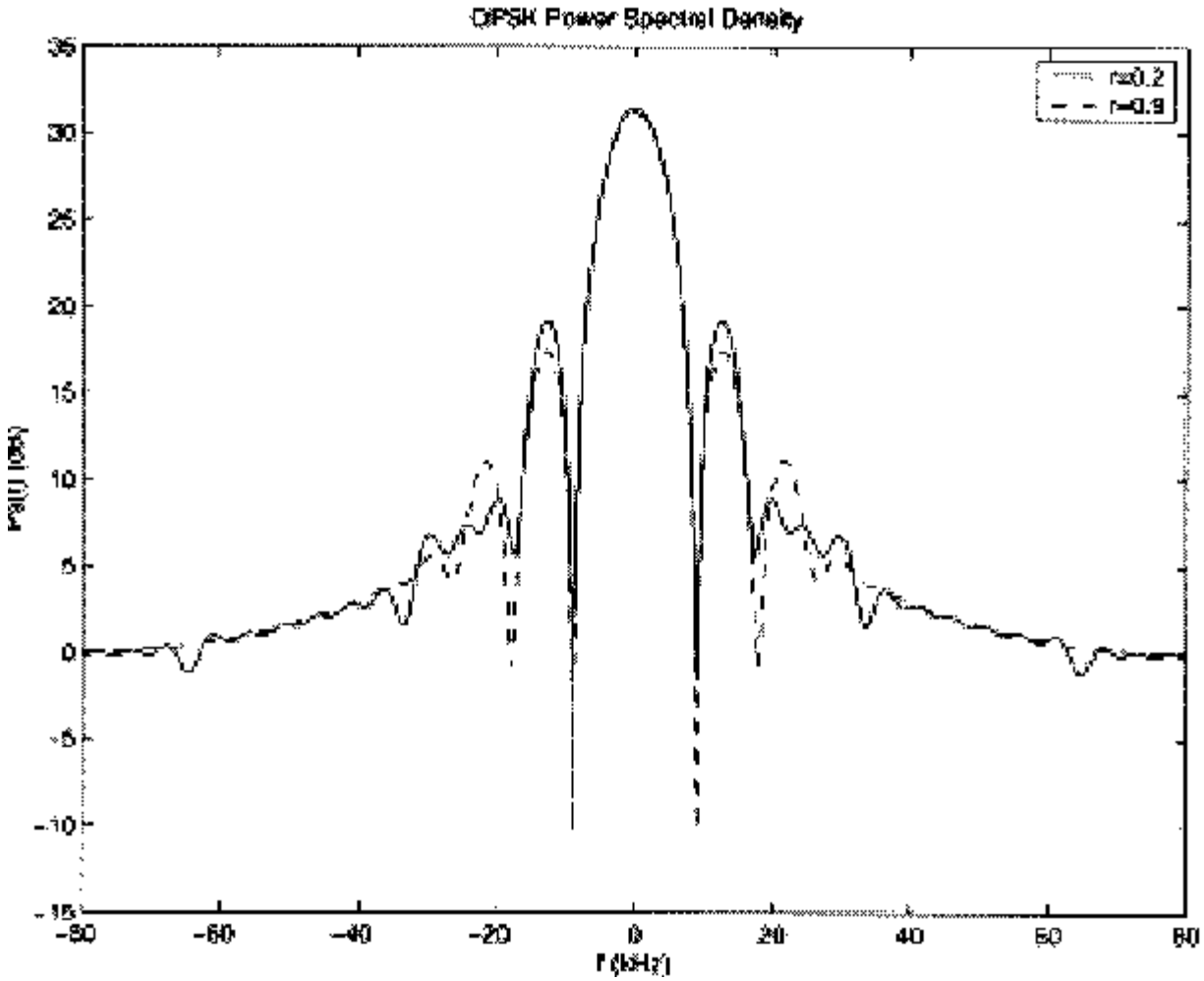


Figure 4.4 PSD of a QPSK signal for rectangular and raised cosine filtered pulses

4.6 QPSK Transmission and Detection Techniques

The unipolar binary message stream has bit rate R_b and is first converted into a bipolar non-return-to-zero (NRZ) sequence using a unipolar to bipolar converter. The bit stream m(t) is then split into two bit streams $m_I(t)$ and $m_Q(t)$ (in-phase and quadrature streams), each having a bit rate of $R_s = R_b/2$. The bit stream $m_I(t)$ is called the "even" stream and $m_Q(t)$ is called the "odd" stream. The two binary sequences are separately modulated by two carriers and , which are in quadrature. The two modulated signals, each of

which can be considered to be a BPSK signal are summed to produce a QPSK signal. The filter at the output of the modulator confines the power spectrum of the QPSK signal within the allocated band. This prevents spill-over of signal energy into adjacent channels and also removes out-of-band spurious signals generated during the modulation process. In most implementations, pulse shaping is done at baseband to provide proper RF filtering at the transmitter output .

The frontend bandpass filter removes the out-of-band noise and adjacent channel interference. The filtered output is split into two parts, and each part is coherently demodulated using the in-phase and quadrature carriers. The coherent carriers used for demodulation are recovered from the received signal using carrier recovery circuits of the type described in Figure 4.5. The outputs of the demodulators are passed through decision circuits which generate the in-phase and quadrature binary streams. The two components are then multiplexed to reproduce the original binary sequence.

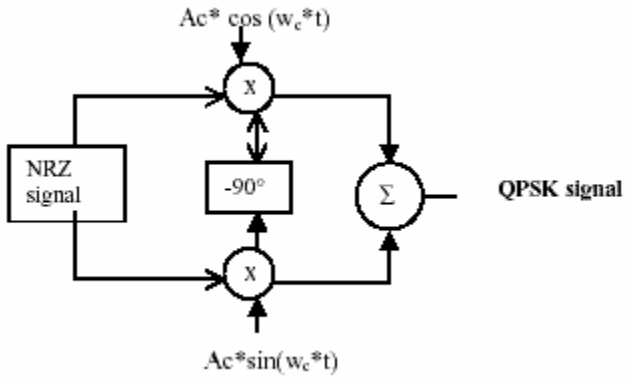


Fig 4.5 Generating a QPSK signal

4.8 π/4 Quadrature Phase Shit Keying (QPSK)

The π/4 shifted QPSK modulation is a quadrature phase shift keying technique which offers a compromise between QPSK in terms of the allowed maximum phase transitions. It may be demodulated in a coherent or noncoherent fashion. In π/4 QPSK, the maximum phase change is limited to ±135°, as compared to 180° for QPSK. Hence, the bandlimited π/4 QPSK signal preserves the constant envelope property better than bandlimited QPSK. An extremely attractive feature of π/4 QPSK is that it can be noncoherently detected, which greatly simplifies receiver design. Further, it has been found that in the presence of multipath spread and fading, π/4 QPSK performs better than OQPSK. Very often, π/4 QPSK signals are differentially encoded to facilitate easier implementation of differential detection or coherent demodulation with phase ambiguity in the recovered carrier. When differentially encoded, π/4 QPSK is called π/4 DQPSK. In a π/4 QPSK modulator, signaling points of the modulated signal are selected from two QPSK constellations which are shifted by π/4 with respect to each other. Figure 4.3 shows the two constellations along with the combined constellation where the links between two signal points indicate the possible phase transitions. Switching between two constellations, every successive bit ensures that there is at least a phase shift which is an integer multiple of π/4 radians between successive symbols. This ensures that there is a phase transition for every symbol, which enables a receiver to perform timing recovery and synchronization.

4.8.1 π/4 QPSK Transmission Techniques

A block diagram of a generic $\pi/4$ QPSK transmitter is shown in Figure 4.6. The input bit stream is partitioned by a serial-to-parallel (SIP) converter into two parallel data streams $m_{I,k}$ and $m_{Q,k}$, each with a symbol rate equal to half that of the incoming bit rate. The k^{th} in-phase and quadrature pulses, I_k and Q_k are produced at the output of the signal mapping circuit over time $kT \leq t \leq (k + 1)T$ and are determined by their previous values, I_{k-1} and Q_{k-1} as well as θ_k which itself is a function of Φ_k which is a function of the current input symbols m_{1k} and m_{Qk}. I_k and Q_k represent rectangular pulses over one symbol duration having amplitudes given by

$$I_k = \cos\theta_k = I_{k-1}\cos\varphi_k - Q_{k-1}\sin\varphi_k$$
$$Q_k = \sin\theta_k = I_{k-1}\sin\varphi_k - Q_{k-1}\cos\varphi_k \qquad 4.25$$

where

$$\theta_k = \theta_{k-1} + \varphi_k$$

and Q_k and Q_{k-1} are phases of the kth and k-1st symbols. The phase shift φ_k is related to the input symbols m_{1k} and m_{Qk} according to Table 4.3.

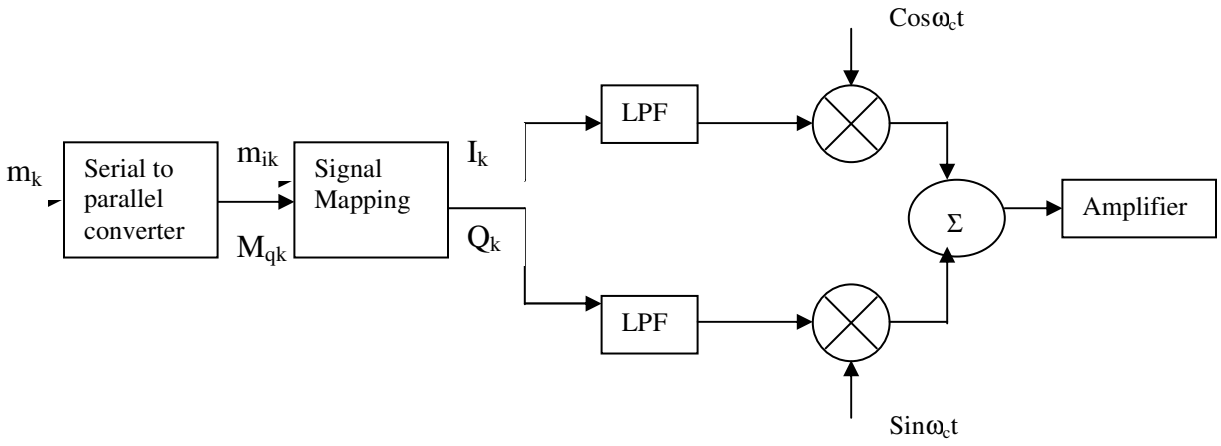


Figure 4.6 $\pi/4$ QPSK Generic transmitter

Table 4.3 Phase shift of sequential bits

Information Bits mik,mQk	Phase Shift
00	$-3\pi/4$
01	$3\pi/4$
10	$-\pi/4$
11	$\pi/4$

Just as in a QPSK modulator, the in-phase and quadrature bit streams I_k and Q_k are then separately modulated by two carriers which are in quadrature with one another, to produce the $\pi/4$ QPSK waveform given by 4.25

$$S_{\pi/4QPSK}(t) = I(t)\cos\omega_c t - Q(t)\sin\omega_c t$$

Both I_k and Q_k are usually passed through raised cosine rolloff pulse shaping filters before modulation, in order to reduce the bandwidth occupancy. The function p(t) corresponds to the pulse shape, and T_s is the symbol period. Pulse shaping also reduces the spectral restoration problem which may be significant in fully saturated, nonlinear amplified systems. It should be noted that the values of I_k and Q_k and the peak amplitude of the waveforms I(t) and Q(t) can take one of the five possible values, 0, +1, −1, $+1/\sqrt{2}$, $-1/\sqrt{2}$.

From the above discussion, it is clear that the information in a $\pi/4$ QPSK signal is completely contained in the phase difference φ_k of the carrier between two adjacent symbols. Since the information is completely contained in the phase difference, it is possible to use noncoherent differential detection even in the absence of differential encoding.

4.8.2 π/4 QPSK Detection Techniques

Due to ease of hardware implementation, differential detection is often employed to demodulate $\pi/4$ QPSK signals. In an AWGN channel, the BER performance of a differentially detected $\pi/4$ QPSK is about 3 dB inferior to QPSK, while

coherently detected $\pi/4$ QPSK has the same error performance as QPSK. In low bit rate, fast Rayleigh fading channels, differential detection offers a lower error floor since it does not rely on phase synchronization. There are various types of detection techniques that are used for the detection of $\pi/4$ QPSK signals. They include baseband differential detection, IF differential detection, and FM discriminator detection. While both the baseband and IF differential detector determine the cosine and sine functions of the phase difference, and then decide on the phase difference accordingly, the FM discriminator detects the phase difference directly in a noncoherent manner. Interestingly, simulations have shown that all three receiver structures offer very similar bit error rate performances, although there are implementation issues which are specific to each technique.

Baseband Differential Detection

Figure 4.6 shows a block diagram of a baseband differential detector. The incoming $\pi/4$ QPSK signal is quadrature demodulated using two local oscillator signals that have the same frequency as the unmodulated carrier at the transmitter, but not necessarily the same phase. If $\phi_k = \tan^{-1}(Q_k/I_k)$ is the phase of the carrier due to the k^{th} data bit, the output w_k and Z_k from the two low pass filters in the in-phase and quadrature arms of the demodulator can be expressed as

$$W_k = \cos(\phi_k - \gamma)$$
$$Z_k = \sin(\phi_k - \gamma) \qquad 4.26$$

where γ is a phase shift due to noise, propagation, and interference. The phase γ is assumed to change much slower

than ϕ_k so it is essentially a constant. The two sequences Wk and Zk are passed through a differential decoder which operates on the following rule.

$$x_k = w_k w_{k-1} + z_k z_{k-1} \qquad 4.27$$

$$y_k = z_k w_{k-1} - w_k z_{k-1}$$

Let me illustrate how the data is transmitted from the binary sequence of data

Assume that $\theta_0 = 0^\circ$. The bit stream 0 0 1 0 1 1 is to be sent using DQPSK. The leftmost bits are first applied to the transmitter. Determine the phase of , and the values of I_k, Q_k during transmission.

Given $\theta_0 = 0^\circ$, the first two bits are 0 0 , which implies that

$\theta_1 = \theta_0 + \phi_1 = -3\pi/4$ from table 4.2

This implies I_1, Q_1 are (−0.707, −0.707) from 4.12 and 13 .

The second two bits are 1 0 , which maps from table 4.2 into $\phi 2 = -\pi/4$.

Thus from equation 4.13 θ_2 becomes $-\pi$, and I_2 , Q_2 are (−1,0) from equation 4.25 .

The bits 1 1 induce $\phi_3 = -\pi/4$ and thus $\theta_3 = -3\pi/4$.

Thus , I3 , Q_3 are (−0.707,−0.707).

Using the $\pi/4$ QPSK signal of the above , demonstrate how the received signal is detected properly using a baseband differential detector

Assume the transmitter and receiver are perfectly phase locked, and the receiver has a front end gain of two. Using eq 4.21 and 22 the phase difference between the three transmitted phases ield (x1,y1) = (−0.707,−0.707); (x2,y2)=(0.707,−0.707) ; (x3,y3) =(0.707,0.707). Applying

the decision rules of Eq 4.23 24 the detected bit stream is $(S_1, S_2, S_3, S_4, S_5, S_6) = (0, 0, 1, 0, 1, 1)$

It has been simulated the π/4 QPSK modulation using a random sequence of binary bits and converted it into a QPSK signal using the mapping table and added a noise of SNR 14 dB in the channel after modulating it using π/4 QPSK modulation technique.
Then the noised signal is demodulated and is shown in the following figures.

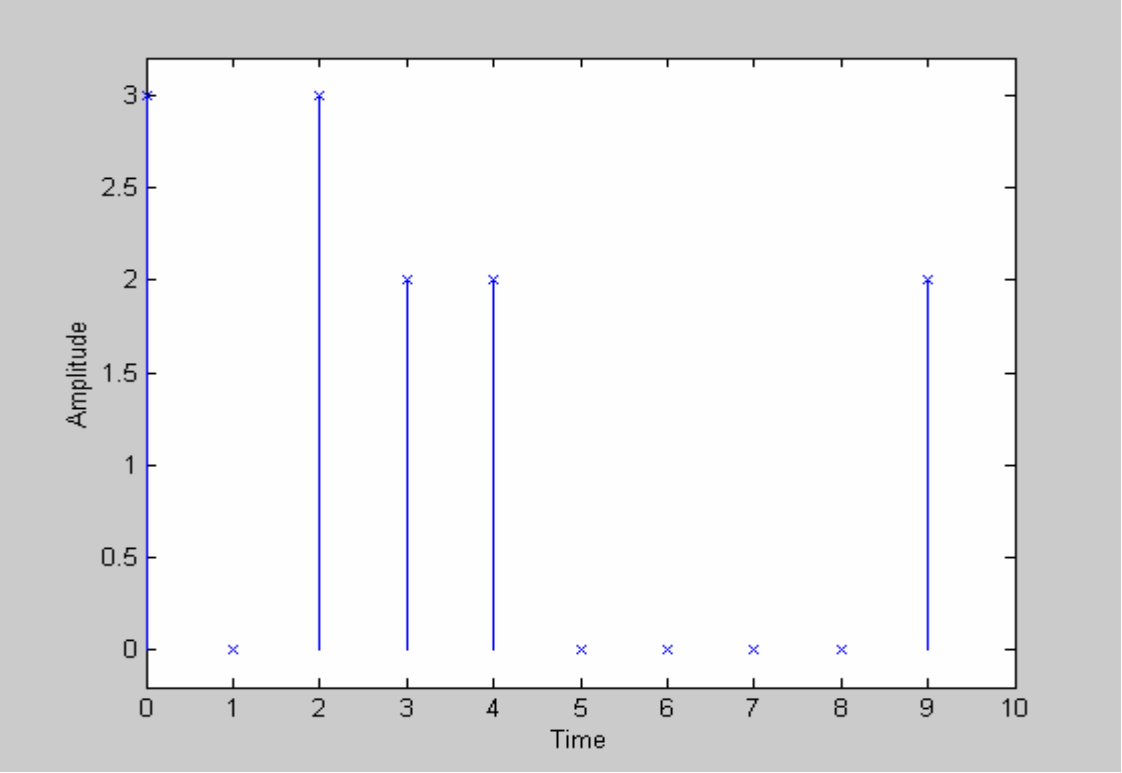


Fig 4.7 Random sequence of QPSK message signals

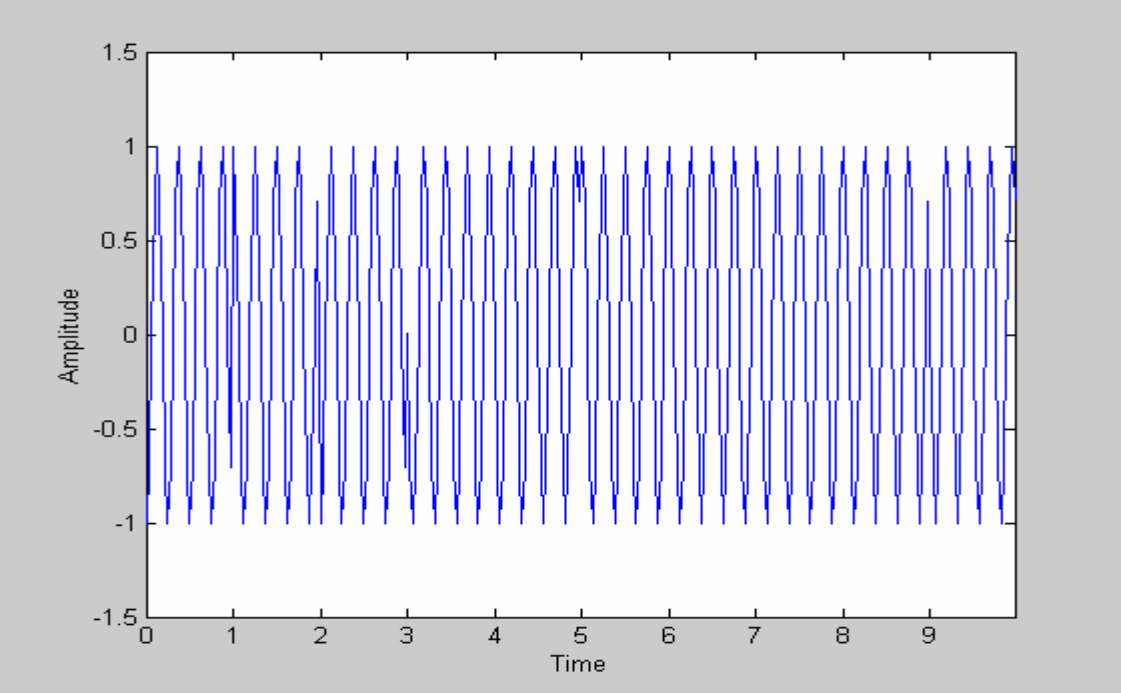


Fig 4.8 Modulated QPSK message signals

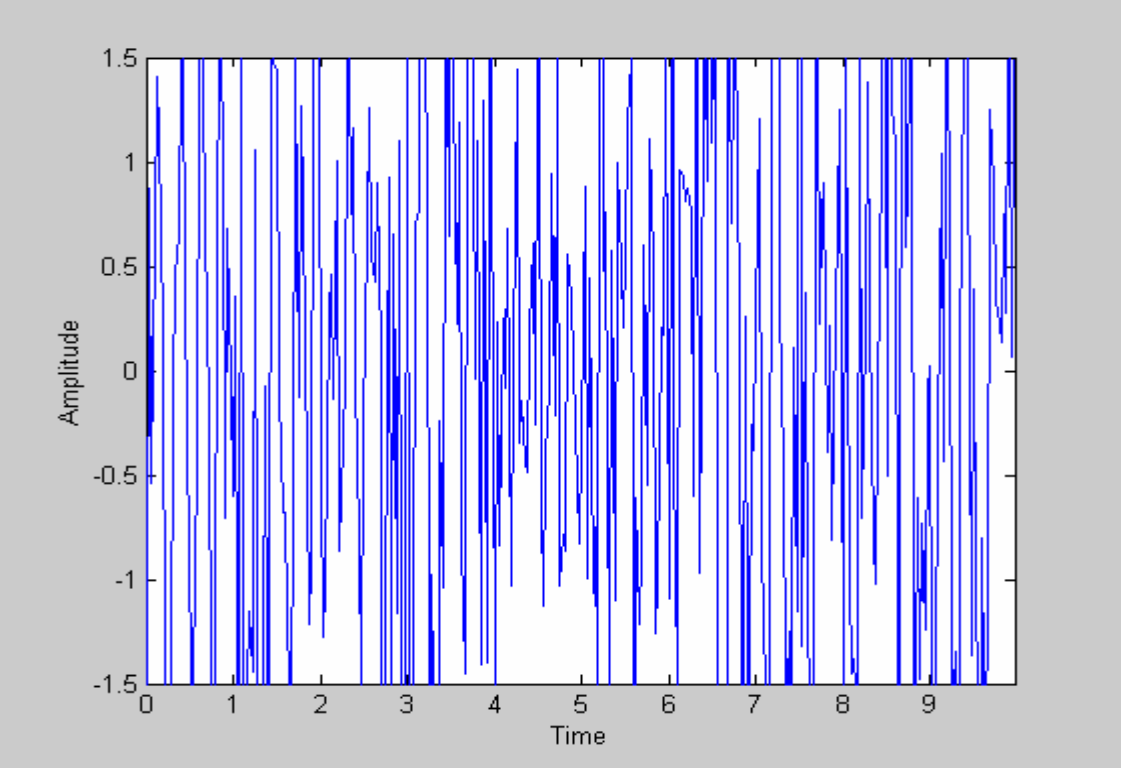


Fig 4.9 Modulated QPSK message signals when passed through a noisy channel

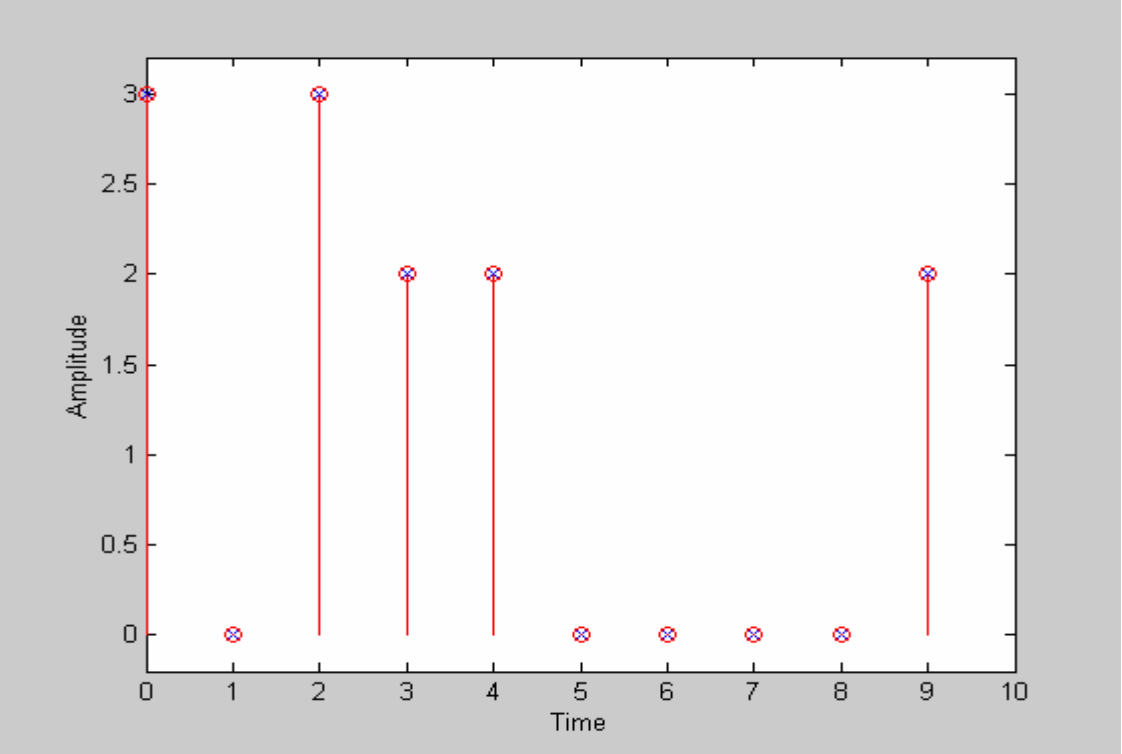


Fig 4.10 Demodulated QPSK message signals

As seen in the simulation the π/4 QPSK modulation technique has the power of error correcting capability even if it passes through a channel noise . This is mainly because Digital modulation and transmission techniques offer a great deal of error resisting capability because the symbol difference is not continuous unlike analog signal and transmission system.

4.9 Minimum Shift keying (MSK)

Minimum shift keying (MSK) is a special type of continuous phase-frequency shift keying wherein the peak frequency deviation is equal to ¼ the bit rate. In other words, MSK is continuous phase Frequency Shift Keying (FSK) with a modulation index of 0.5. The modulation index of an FSK signal is similar to the FM modulation index, and is defined as $k_{FSK} = (2\Delta F)/Rb$, where ΔF is the peak RF frequency deviation and R_b is the bit rate. A modulation index of 0.5 corresponds to the minimum frequency spacing that allows two FSK signals to be coherently orthogonal and the name minimum shift keying implies the

minimum frequency separation (i.e bandwidth) that allows orthogonal detection. Two FSK signals $v_H(t)$ and $v_V(t)$ are said to be orthogonal if

$$\int_0^T v_H(t)v_L(t)dt = 0 \qquad 4.28$$

MSK is sometimes referred to as fast FSK, as the frequency spacing used is only half as much as that used in conventional noncoherent FSK.

MSK is a spectrally efficient modulation scheme and is particularly attractive for use in mobile radio communication systems. It possess properties such as constant envelope, spectral efficiency, good BER performance, and self-synchronizing capability.

An MSK signal can be thought of as a special form of QPSK where the baseband rectangular pulses are replaced with half-sinusoidal pulses. These pulses have shapes like the Church bell arch during a period of $2T_b$.

There are a number of variations of MSK that exist. One version of MSK uses only positive half-sinusoids as the basic pulse shape, another version uses alternating positive and negative half sinusoids as the basic pulse shape. However, all variations of MSK are continuous phase FSK employing different techniques to achieve spectral efficiency.

It can be concluded that the MSK signal is an FSK signal with binary signaling frequencies of f_c + ¼T and f_c - ¼T. It can further be seen from

$$S_{MSK}(t) = \sqrt{\frac{2E_b}{T_b}} \cos\left[2\pi f_c t - m_I(t)m_Q(t)\frac{\pi t}{2T_b} + \phi_k \right] \qquad 4.29$$

that the phase of MSK signal varies linearly during the course of each bit period.

MSK Power Spectrum

The RF power spectrum is obtained by frequency shifting the magnitude squared of the fourier transform of the baseband phase-shaping function . For MSK, the baseband pulse shaping function is given by

4.30

$$p(t) = \begin{cases} \cos\left(\frac{\pi t}{2T}\right) & |t| < T \\ 0 & elsewhere \end{cases}$$

Thus the normalized power spectral density for MSK is given by

$$P_{MSK}(f) = \frac{16}{\pi^2}\left(\frac{\cos 2\pi(f+f_c)T}{1.16f^2T^2}\right)^2 + \frac{16}{\pi^2}\left(\frac{\cos 2\pi(f-f_c)T}{1.16f^2T^2}\right)^2$$

The power spectral density of an MSK signal. The spectral density of QPSK are also drawn for comparison. From the figure 4.7 it can be seen that the MSK spectrum has lower sidelobes than QPSK. 99% of the MSK bandwidth B is equal to 8/T. The faster rolloff of the MSK spectrum is due to the fact that smoother pulse functions are used. Figure 4.7 also shows that the main lobe of MSK is wider than that of QPSK, and hence when compared in terms of first null bandwidth, MSK is less spectrally efficient than tha phase-shift keying techniques.

4.10 Gaussian Minimum Shift keying (GSMK)

GMSK is a simple binary modulation scheme which may be viewed as a derivative of MSK. In GMSK, the sidelobe levels of the spectrum are further reduced by passing the modulating NRZ data waveform through a premodulation Gaussian pulse shaping filter. Baseband Gaussian pulse shaping smoothes the phase trajectory of the MSK signal and hence stabilizes the instantaneous frequency variations over time. This has the effect of considerably reducing the sidelobe levels in the transmitted spectrum.

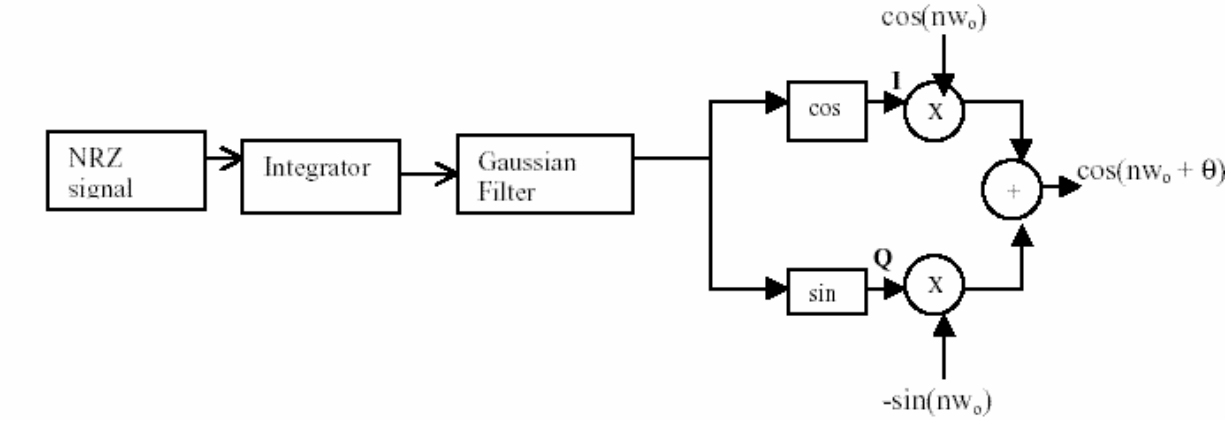


Fig 4.11 Generating a GMSK Signal

Premodulation Gaussian filtering converts the full response message signal(where each symbol occupies a single bit period T) into a partial response scheme where each transmitted symbol spans several bit periods. However, since pulse shaping does not cause the pattern-averaged phase trajectory to deviate from that of simple MSK, GMSK can be coherently detected just as an MSK signal, or noncoherently detected as simple FSK. In practice GMSK is most attractive for its excellent power efficiency (due to the constant envelope) and its excellent spectral efficiency. The premodulation Gaussian filtering introduces ISI in the transmitted signal, but it can be shown that the degradation is not severe if the 3dB bandwidth bit duration product (BT) of the filter is greater than 0.5. GMSK sacrifices the irreducible error rate caused by partial response signaling in exchange for extremely good spectral efficiency and constant envelope properties.

The Gmsk premodulation filter has an impulse response given by

$$h_G = \frac{\sqrt{\pi}}{\alpha}\exp\left(-\frac{\pi^2}{\alpha^2}t^2\right) \qquad 4.30$$

and the transfer function given by

$$HG(f) = \exp(-\alpha^2 f^2)$$

The parameter α is related to B, the 3 dB baseband bandwidth of *HG(f)* , by

$$\alpha = 0.5887/B$$

and the GMSK filter may be completely defined from B and the baseband symbol duration T. It is therefore customary to define GMSK by its BT product.

Figure 4.12 shows the simulated RF power spectrum of the GMSK signal for various values of BT. The power spectrum of MSK, which is equivalent to GMSK with a BT product of infinity , is also shown for comparison purposes. It is clearly seen from the graph that the BT product decreases, the sidelobe levels fall off very rapidly. For example for a BT=0.5, the peak of the second lobe is more than 30dB below the main lobe, whereas for simple MSK, the second lobe is only 20 dB below the main lobe. However reducing BT increases the irreducible error rate produced by the low pass filter due to ISI.

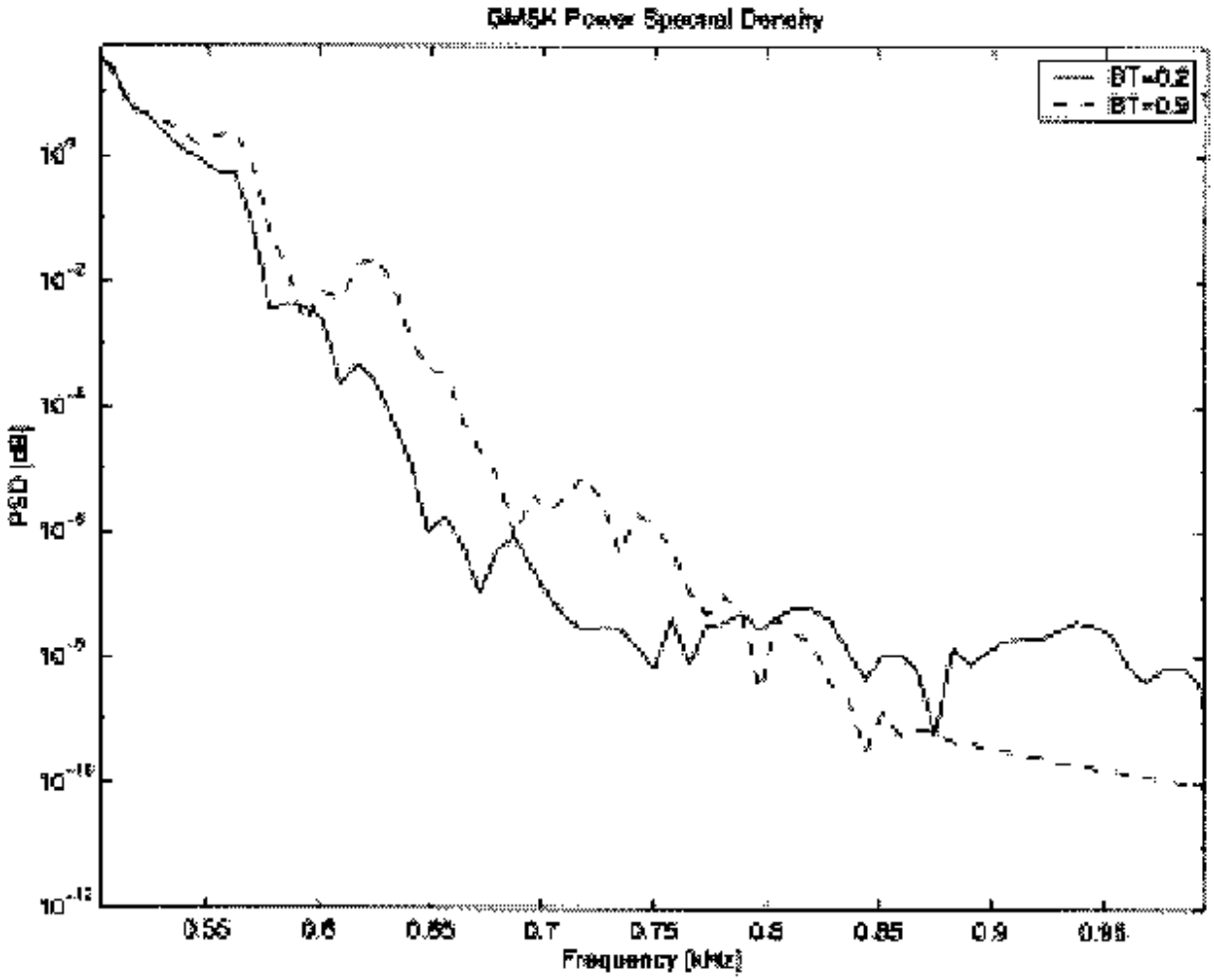


Figure 4.12 The simulated RF power spectrum of the GMSK

As shown in section 1.6, mobile radio channels induce an irreducible error rate due to mobile velocity, so as long as the GMSK irreducible *error rate is less than that produced by the mobile channel, there is no penalty in using GMSK.*

Table 4.4 shows occupied bandwidth containing a given percentage of power in a GMSK signal as a function of the BT product.

Table 4.4 occupied bandwidth containing a given percentage of power in a GMSK

BT	90%	99%	99.9%	99.99%
0.2 GMSK	0.52	0.79	0.99	1.22
0.25 GMSK	0.57	0.86	1.09	1.37
0.5 GMSK	0.69	1.04	1.33	2.08
MSK	0.78	1.20	2.76	6.00

While GMSK spectrum becomes more and more compact with decreasing BT value, the degradation due to ISI increases. It was shown that the BER degradation due to ISI caused by filtering is minimum for a BT value of 0.5887, where the degradation in the required E/N is only 0.14 dB from the case of no ISI.

GMSK Bit Error Rate

The bit error rate for GMSK was first found for AWGN channels, and was shown to offer performance within 1 dB of optimum MSK when BT=0.25. The bit error probability is a function of BT, since the pulse shaping impacts the ISI. The bit error probability for GMSK is given by.

$$P_e = Q\left\{\sqrt{\frac{2\gamma E_b}{N_0}}\right\} \qquad 4.31$$

where γ is a constant related to BT by

γ= 0.68 for GMSK with BT=0.25 and 0.85 for simple MSK(BT=∞)

4.11 Comparison of QPSK and GMSK

Quadrature Phase Shift Keying (QPSK) is currently used in applications such as cable modems and the IS-95 (CDMA) system. QPSK is predominantly noted for its power efficiency and robustness against phase noise. Another type of modulation scheme used in mobile radio systems is Gaussian Minimum Shift Keying (GMSK),which is vastly popular in Europe's GSM cellular standard. In addition, narrow bandwidth and its ability to use coherent detection characterize GMSK, a constant envelope modulation technique. QPSK originates from Quadrature Amplitude Modulation (QAM). QAM combines phase changes with signal amplitude variations that enable more information to be carried over a limited channel bandwidth. Several varieties of QAM modulation exist, such as PSK, BPSK, and QPSK, providing various levels of bandwidth efficiency. In QPSK, two data

channels modulate the carrier. Transitions in the data cause the carrier to shift by either 90° or 180°. This allows two discrete data streams, identified as I channel (in phase) and Q channel (quadrature) data.

GMSK is from the Minimum Shift Keying (MSK) modulation family. GMSK differs from MSK in the aspect of filter use, hence the name Gaussian MSK. GMSK resulted from an attempt to improve the MSK power spectrum. However, one of the advantages of MSK is that it does not produce Intersymbol Interference (ISI). The transmitted pulse is confined within its bit duration resulting in no adjacent channel interference. Nevertheless, GMSK possesses a more compact spectrum, with the application of a lowpass filter, helping to reduce its spectral sidelobes.

Most digital transmitters operate their power amplifiers at or near saturation to achieve maximum power efficiency. At saturation, it poses a threat to the signal, exposing it to phase and angle distortions. These distortions spread the transmitted signal into the adjacent channel, causing interference. To resolve this issue, a filter is used to suppress the sideband lobes.

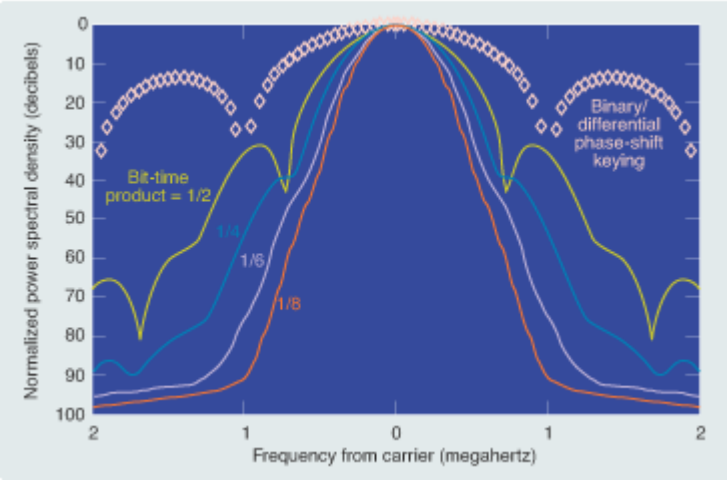


Figure 4.13 Comparison of QPSK and GMSK Power Spectral Density

Nyquist pulse-shaping techniques, such as the Raised Cosine (RC) filter and Gaussian filter, are used to reduce ISI.

Taking a closer look at the respective modulation schemes reveals a need for a filter.
First, the system converts a bit stream into a Non-Return-to-Zero (NRZ) signal which is multiplied by an in-phase (I) and quadrature (Q) signal, keeping in mind that each carrier phase is separated by 90°. The two components are then summed to achieve the desired QPSK signal output. In order to optimize the signal, QPSK uses the RC filter. This prevents the signal from spreading its energy into the adjacent channels. Ideally, the Nyquist filter is free of ISI. However, all practical Low Pass Filters (LPF) exhibit phase and amplitude distortions so special pulse shaping filters are needed to ensure that the total transmitted signal arrives at the receiver. For the case of QPSK, the RC filter is used for the ideal LPF since the roll-off factor, r, controls the bandwidth of the Nyquist filter. This formula signifies the impulse response of the RC filter for changing values of r, the roll-off factor. According to the equation, as r increases, the bandwidth of the filter also increases while maintaining low sidelobes. As the roll-off factor decreases, the spectrum becomes more compact. This requires a more complex receiver at demodulation. In order to create a Gaussian modulator, an NRZ signal comprised of –1's and 1's is made from a binary sequence of 0's and 1's and passed through an integrator. Then the signal is

convoluted via a Gaussian filter. The I and Q components are multiplied with their respective carriers to get the resulting output.

In the case of GMSK, a Gaussian filter is applied prior to modulation so as to maintain the constant envelope property of the modulation scheme. The impulse response equation of the Gaussian filter, h(t), is show below. In order to reduce the sidelobes and produce a compact spectrum, the BT factor controls the effects of the Gaussian filter. In the BT product, B is the 3 Db (or half-power) bandwidth, while T is the symbol period.

$$h(t) := K \cdot \sqrt{2 \cdot \frac{\pi}{\ln(2)}} \cdot (BT) \cdot e^{-2 \cdot \frac{((BT) \cdot t \cdot \pi)^2}{\ln(2)}} \qquad 4.32$$

Gaussian minimum shift keying is a form of continuous phase modulation, a technique that achieves smooth phase transitions between signal states, thereby reducing bandwidth requirements. With Gaussian minimum shift keying, input bits with rectangular (+1, -1) representation are converted to Gaussian (bell-shaped) pulses by a Gaussian filter before further smoothing by a frequency modulator. Also, in most cases, the Gaussian pulse is allowed to last longer than one bit time—the amount of time a binary 1 is in the "on" position. Consequently, the pulses overlap, giving rise to a phenomenon known as intersymbol interference. The extent of this overlap is determined by the product of the bandwidth of the Gaussian filter and the data-bit duration; the smaller the bandwidth–bit-time product, the more the data bits or pulses overlap.

The resulting carrier signal is very smooth in phase-particularly in comparison to waveforms generated through standard binary or quarternary phase-shift keying. This is important because signals with smooth phase transitions require less bandwidth to transmit. On the other hand, this very smooth phase makes the receiver's job much harder. With Gaussian minimum shift keying, there are no well-defined phase transitions to detect for bit synchronization, and the energy from each bit is mixed with the energy from several other bits. The transmitter output looks nothing like the data input, and on the receiver side, a special demodulator of increased complexity is needed to extract the data bits. For the receiver to achieve a given bit-error rate, the transmitter must generate more power to overcome the receiver noise in the presence of the intersymbol interference. In other words, the Gaussian minimum shift keying waveform is usually less power-efficient

than more traditional waveforms such as binary phase-shift keying and requires a more complex receiver, but this potential reduction in power efficiency and increase in receiver complexity could be rewarded with a very significant enhancement of bandwidth efficiency. So, with Gaussian minimum shift keying, there is a trade-off between bandwidth efficiency and power efficiency.

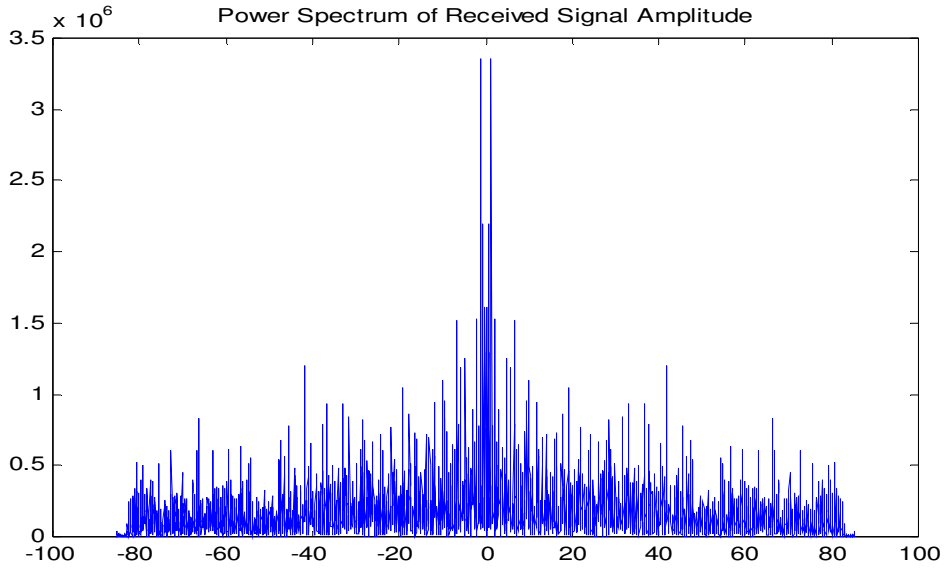
Gaussian minimum shift keying is not new—the technique has been used extensively in Europe for cell-phone applications with a bandwidth–bit-time product of 0.4. But system designs using very small bandwidth–bit-time products such as 1/5 or 1/8 are new—and challenging. Mobile Communication became interested in these smaller bandwidth–bit-time products because of their narrow bandwidth occupancy and the rapid roll-off of their power spectra. These two factors strongly influence the ability to pack many different channels into a limited amount of bandwidth. The Gaussian minimum shift keying waveform exhibits a steep power spectrum and therefore coexists well with adjacent channels in a frequency-division multiple-access system.

CHAPTER FIVE
RESULTS AND DISCUSSION

A striking result that have been found out in this project is that almost all Elecrical Engineering techniques especially in digital signal processing and transmission has been completely been converted to software implementation rather than being specified only in a hardware due to new multipurpose programmable digital signal processing have made it possible to implement digital communication like modulators and demodulators completely in software.

In this project only the channel is simulated the rest speech coding, modulation and error control coding are programmed as what is adopted in a standard company coding. Except for the way I programmed the algorithm , the concept of implementing these digital signal processing algorithms and concepts were programmed like what is implemented in any practical wireless mobile communication system.

This shows that except for the difficulty of miniaturization and mass production of tight RF filters and low noise, front end receiver amplifiers ,thanks to the technology of programmable digital signal processing , Ethiopian Engineers are not far away of mass producing these multipurpose programmable digital signal processing completely in



softw Fig 5.1 Power Spectrum of Received Signal Amplitude verses frequency

A Doppler effect that is noticed during the simulation process is that was sent a baseband signal in the channel, the output of the channel simulation was a signal in which the frequency being randomly modulated as seen in the figure 5.1 and 5.2, where the maximum value of modulation being the Doppler frequency. This indicates except for the interleaving and channel coding which counteract the channel impairments like the deep fading, the frequency of the received signal is shifted at the maximum of the Doppler frequency.

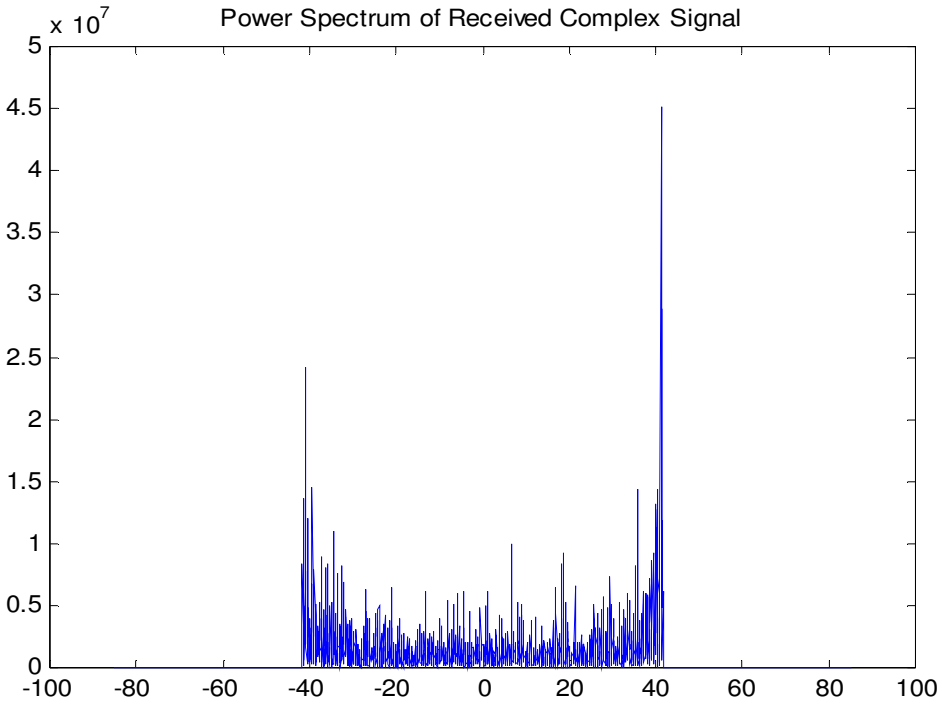


Fig 5.2 Power Spectrum of Received Complex Signal

The speech quality degradation of the present wireless communication system is due to these random modulation by Doppler frequencies. And this phenomenon was tried to be counteracted as a software solution in this study using interleaving and channel coding and finally using channel efficient $\pi/4$ Differential Quadrature phase shift keying modulation.

The astonishing result from the simulator which cannot be left undiscussed is the the phenomena of speech coder especially the simulated Linear Predictive coding . The speech signal occupies 64 Kbps is sampled at 8 KHz in an 8 bit stream . Bit the result of the bit rate reduction rate of the Linear Predictive Speech coder is almost $1/27^{th}$ of the actual signal sampling . This is a compression rate that the simulator program outputed at a rate of 2.4 Kbps from a speech signal of 64 Kbps . One can imagine how the bandwidth which is the expensive commodity in mobile wireless communication system is efficiently utilized. Had not been redundancy being added for error correction this compression meant it is possible to transmit 27 voice channels in place of the obsolete single channel in which a speech is transmitted uncompressed. Usually this margin won't be achievable due to error control coding which are necessary since each parameters extracted from the Linear predictive coding are significant and important that they should be protected from errors with very much redundancy. Hence in practice speech coding technology increase the capacity to six users per channel in the same bandwidth.

The goal of speech coding systems is to transmit speech with the highest possible quality using the least possible channel capacity. Because speech coders attempt to represent a wide range of voices in a uniform and efficient digital format, the encoded data bits (source bits) carry a great deal of information and some source bits are more important than others and must be protected from errors. It is typical for speech coders to produce several "important" bits in succession, and it is the function of the interleaver to spread these bits out in time so that if there is a deep fade or noise burst, the important bits from a block of source data are not corrupted at the same time. By spreading the source bits over time, it becomes possible to make use of error control coding (called channel coding) which protects the source data from corruption by the channel. Since error control codes are designed to protect against channel errors that may occur randomly or in a bursty manner, interleavers scramble the time order of source bits before they are channel coded.

Channel coding protects digital data from errors by selectively introducing redundancies in the transmitted data. By properly encoding of the information, errors

induced by a noisy channel can be reduced to any desired level without sacrificing the rate of information transfer .

The basic purpose of error detection and error correction techniques is to introduce redundancies in the data to improve wireless link performance. The introduction of redundant bits increases the raw data rate used in the link, and, hence, it increases the bandwidth requirement for a fixed source data rate. This reduces the bandwidth efficiency of the link in high SNR conditions, but provides excellent BER performance at low SNR values.

Several factors influence the choice of digital modulation scheme. A desirable modulation scheme provides low bit error rates at low received signal to noise ratio, performs well in multipath and fading conditions , occupies a minimum of bandwidth, and easy and cost effective to implement. one modulation schemes do not simultaneously satisfy all of these requirements. Some modulation schemes are better in terms of the bit error rate performance, while others are better in terms of bandwidth efficiency. Depending on the demands of the particular application, tradeoffs are made when selecting a digital modulation.

As discussed in chapter 1, the mobile radio channel is characterized by various impairments such as fading, multipath and the Doppler spread. In order to evaluate the effectiveness of any modulation scheme, it is required to evaluate the performance of the modulation scheme in a mobile radio environment over such channel conditions. Although bit error rate (BER) evaluation gives a good indication of the performance of a particular modulation scheme, It does not provide information about the type of errors. For example, it does not give incidents of bursty errors. In a fading mobile radio channel, it is likely that

a transmitted signal will suffer deep fades which can lead to outage or a complete loss of the signal.

Evaluating the probability of outage is another means to judge the effectiveness of the signaling scheme in a mobile radio channel. An outage event is specified by a specific number of bit errors occurring in a given transmission. Bit error rates and probability of outage for various modulation schemes under various types of channel impairments can be evaluated either through analytical techniques or through simulation. While simple analytical techniques for computing bit error rates in slow flat fading channels exist, performance evaluation in frequency selective channels and computation outage probabilities are often made through computer simulation. Computer simulations are based on convolving the input bit stream with a suitable channel impulse response and counting the bit errors at the output of the receiver decision circuit.

A striking result is that the bit error probability of QPSK is identical to BPSK as shown in the simulation results of figures 5.3 and 5.4, but twice as much data can be sent in the same bandwidth. Thus when compared to BPSK, QPSK provides twice the spectral efficiency with exactly the same energy efficiency.

.

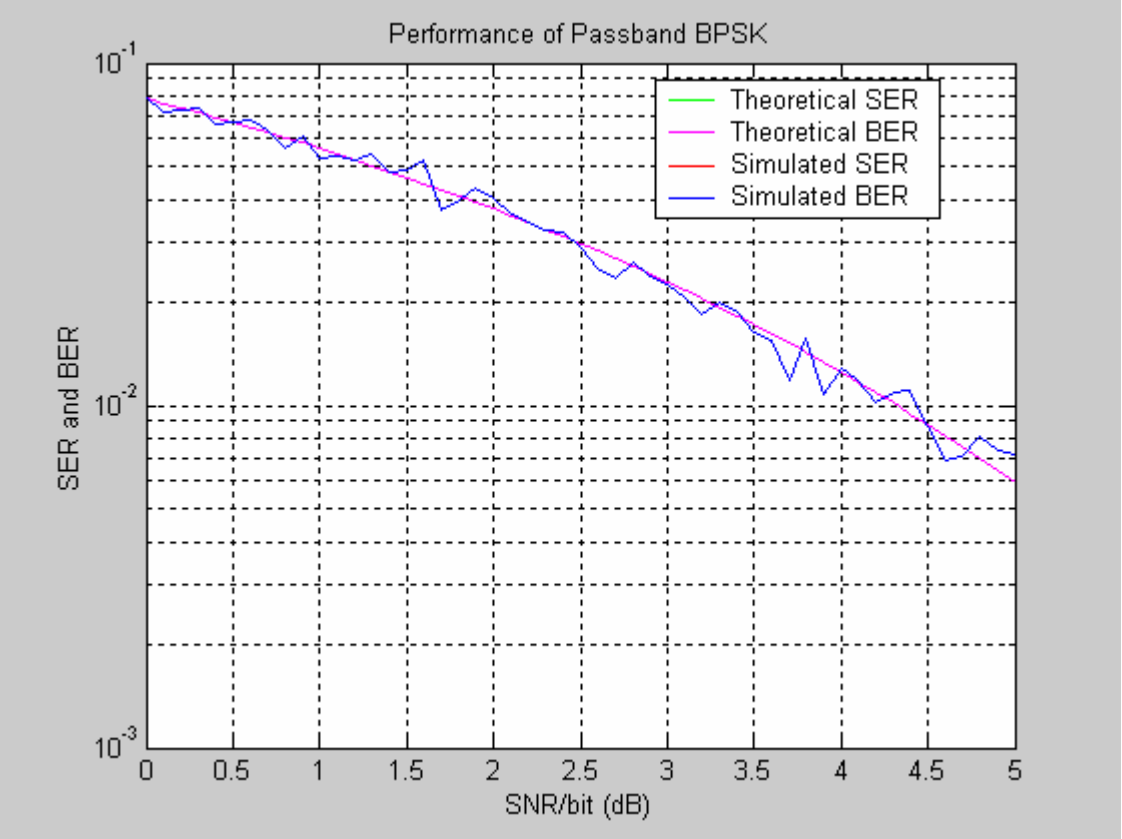


Fig 5.3 BER versus SNRSimulation result of a BPSK modulation

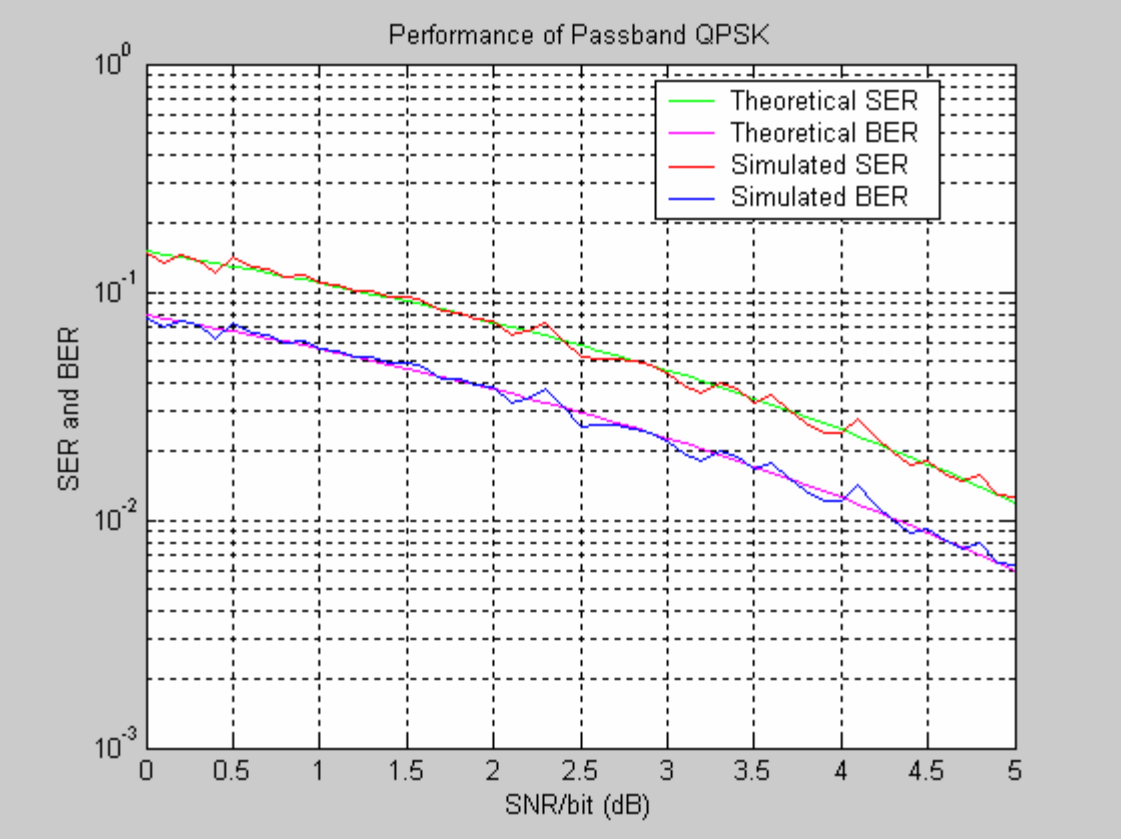


Fig 5.4 BER versus SNRSimulation result of a QPSK modulation

Due to ease of implementation, differential detection is often employed to demodulate π/4 QPSK signals. In an AWGN channel, the BER performance of a differentially detected π/4 QPSK is about 3 dB inferior to QPSK, while coherently detected π/4 QPSK has the same error performance as QPSK. In low bit rate, fast Rayleigh fading channels, differential detection offers a lower error floor since it does not rely on phase synchronization.

The π/4 QPSK modulation using a random sequence of binary bits and converted it into a QPSK signal using the mapping table and added a noise of SNR 14 dB in the channel after modulating it using π/4 QPSK modulation technique.

Then the noised signal is demodulated and is shown in the following figures.

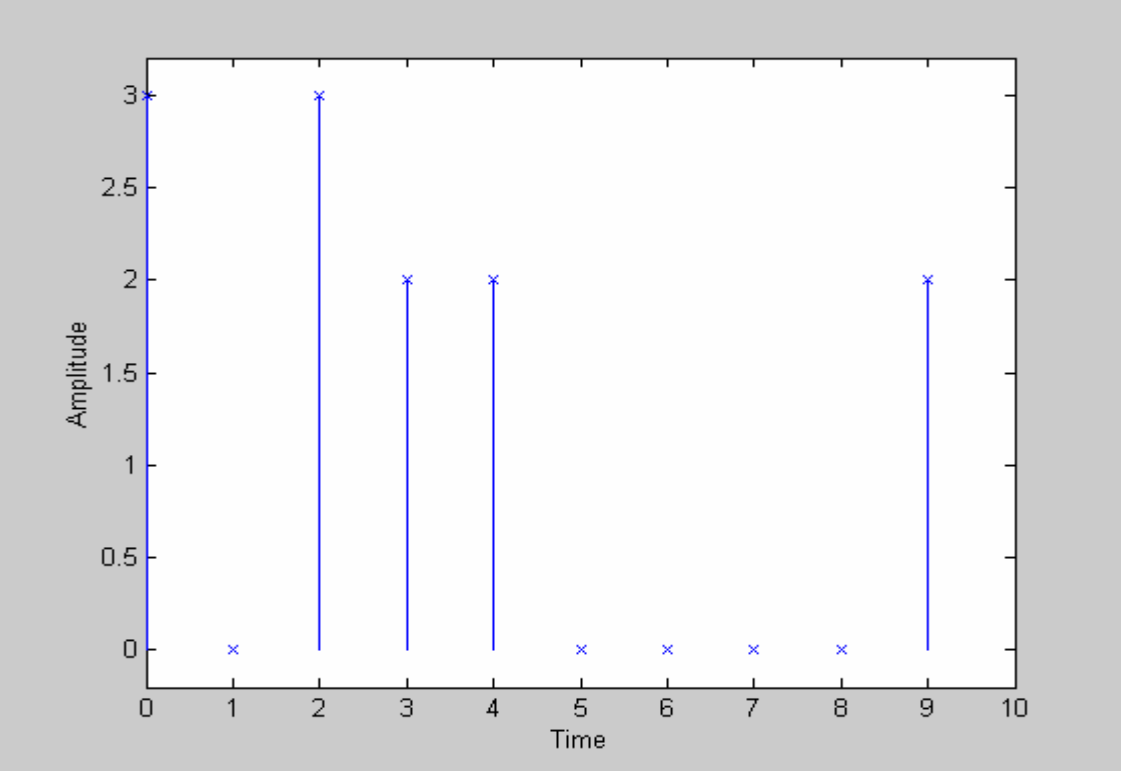


Fig 5.5 Random sequence of QPSK message signals

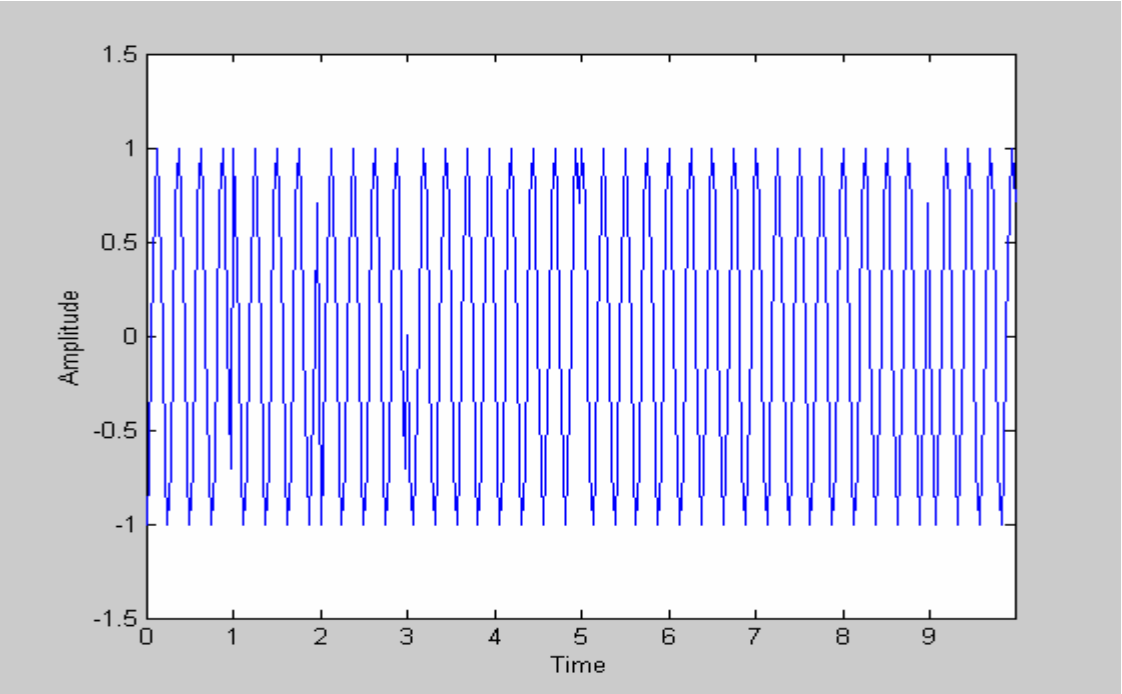


Fig 5.6 Modulated QPSK message signals

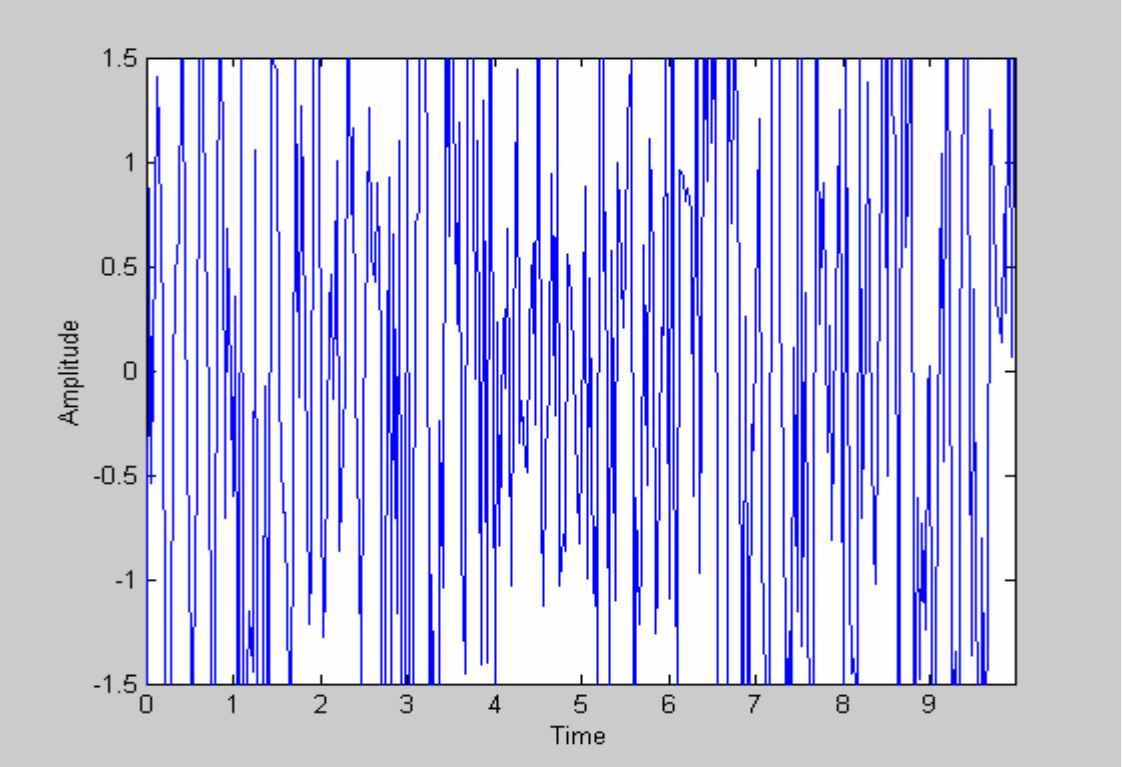


Fig 5.7 Modulated QPSK message signals when passed through a noisy channel

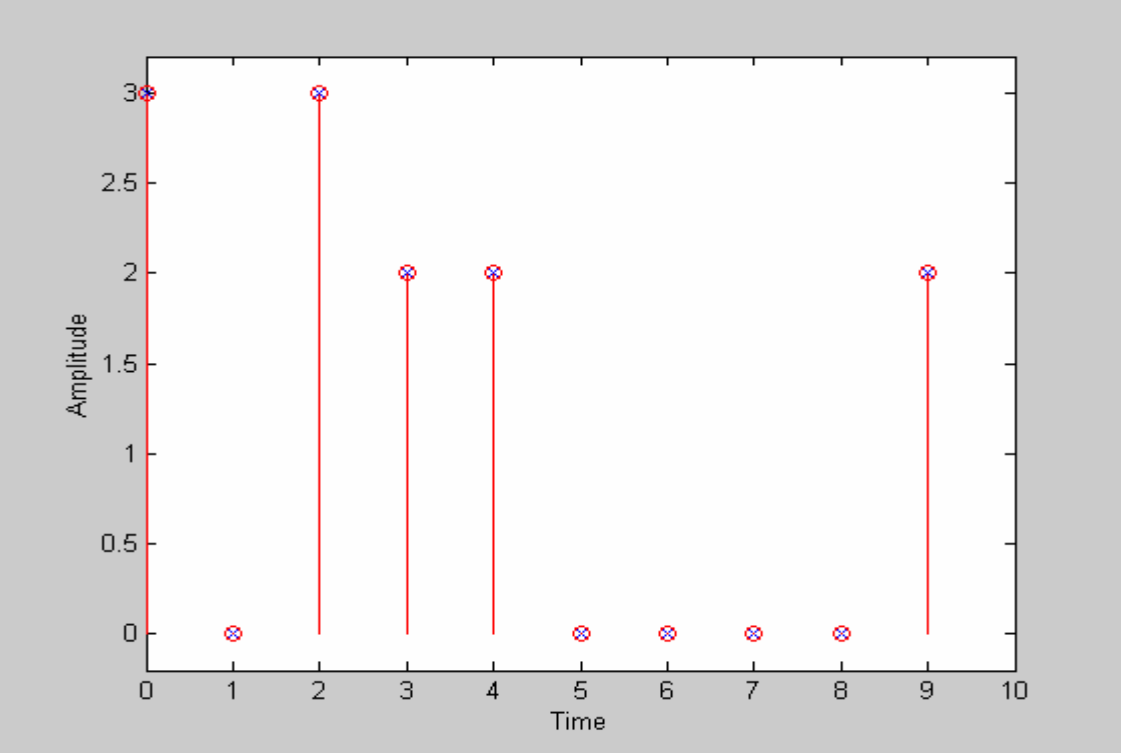


Fig 5.8 Demodulated QPSK message signals

Taking a closer look at the respective modulation schemes reveals a need for a filter. Shown below is the block diagram for generating a QPSK signal.

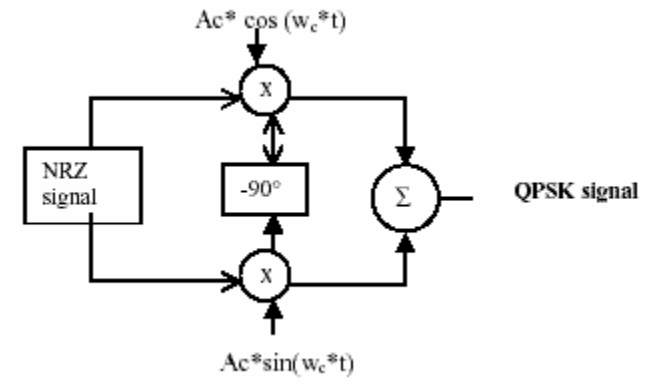


Figure 5.9: Generating a QPSK signal

First, the system converts a bit stream into a Non-Return-to-Zero (NRZ) signal which is multiplied by an in-phase (I) and quadrature (Q) signal, keeping in mind that each carrier phase is separated by 9õ°. The two components are then summed to achieve the desired QPSK signal output. In order to optimize the signal, QPSK uses the RC filter. This prevents the signal from spreading its energy into the adjacent channels. Ideally, the Nyquist filter is free of ISI. However, all practical Low Pass Filters(LPF) exhibit phase and amplitude distortions so special pulse shaping filters are needed to ensure that the total transmitted signal arrives at the receiver. For the case of QPSK, the RC filter is used for the ideal LPF since the roll-off factor, r, controls the bandwidth of the Nyquist filter. The formula used to calculate the RC filter is shown below.

$$Hrc(t) := \left(\frac{\sin\left(\pi \cdot \frac{t}{Ts}\right)}{\pi \cdot t}\right) \cdot \left[\frac{\cos\left(\pi \cdot r \cdot \frac{t}{Ts}\right)}{1 - \left[4 \cdot r \cdot \frac{t}{(2 \cdot Ts)}\right]^2}\right] \qquad 5.1$$

This formula signifies the impulse response of the RC filter for changing values of r, the roll-off factor. According to the equation, as r increases, the bandwidth of the filter also increases while maintaining low sidelobes. As the roll-off factor decreases, the spectrum becomes more compact. This requires a more complex receiver at demodulation.
In order to create a Gaussian modulator, an NRZ signal comprised of –1's and 1's is made from a binary sequence of 0's and 1's and passed through an integrator. Then the signal is convoluted via a Gaussian filter, as shown in Figure 5.10. The I and Q components are multiplied with their respective carriers to get the resulting output.

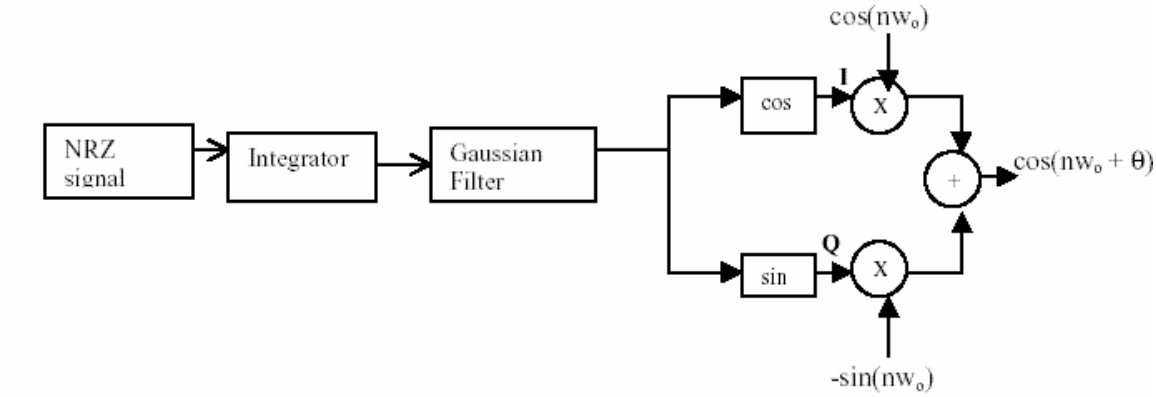


Figure 5.10: Generating a GMSK signal

In the case of GMSK, a Gaussian filter is applied prior to modulation so as to maintain the constant envelope property of the modulation scheme. The impulse response equation of the Gaussian filter, h(t), is show below.

$$h(t) := K \cdot \sqrt{2 \cdot \frac{\pi}{\ln(2)}} \cdot (BT) \cdot e^{-2 \cdot \frac{((BT) \cdot t \cdot \pi)^2}{\ln(2)}} \qquad 5.2$$

In order to reduce the sidelobes and produce a compact spectrum, the BT factor controls the effects of the Gaussian filter. In the BT product, B is the 3 Db (or half-power) bandwidth, while T is the symbol period.

Using the RC and Gaussian filter, the linear and constant envelope modulation schemes are implemented in Matlab to yield its performance and drawback issues. All of the following results show simulations for values of BT and roll-off factor r equal to 0.2 and 0.9. For the case of GMSK, as BT decreases, the sidelobe levels fall off very rapidly causing an increase in bandwidth efficiency, as seen in Figure 5.3. However, as BT increases in the Power Spectral Density (PSD) implementation, as in the case of BT=0.9, the graph yields a wider spectrum, indicating a loss in power efficiency.

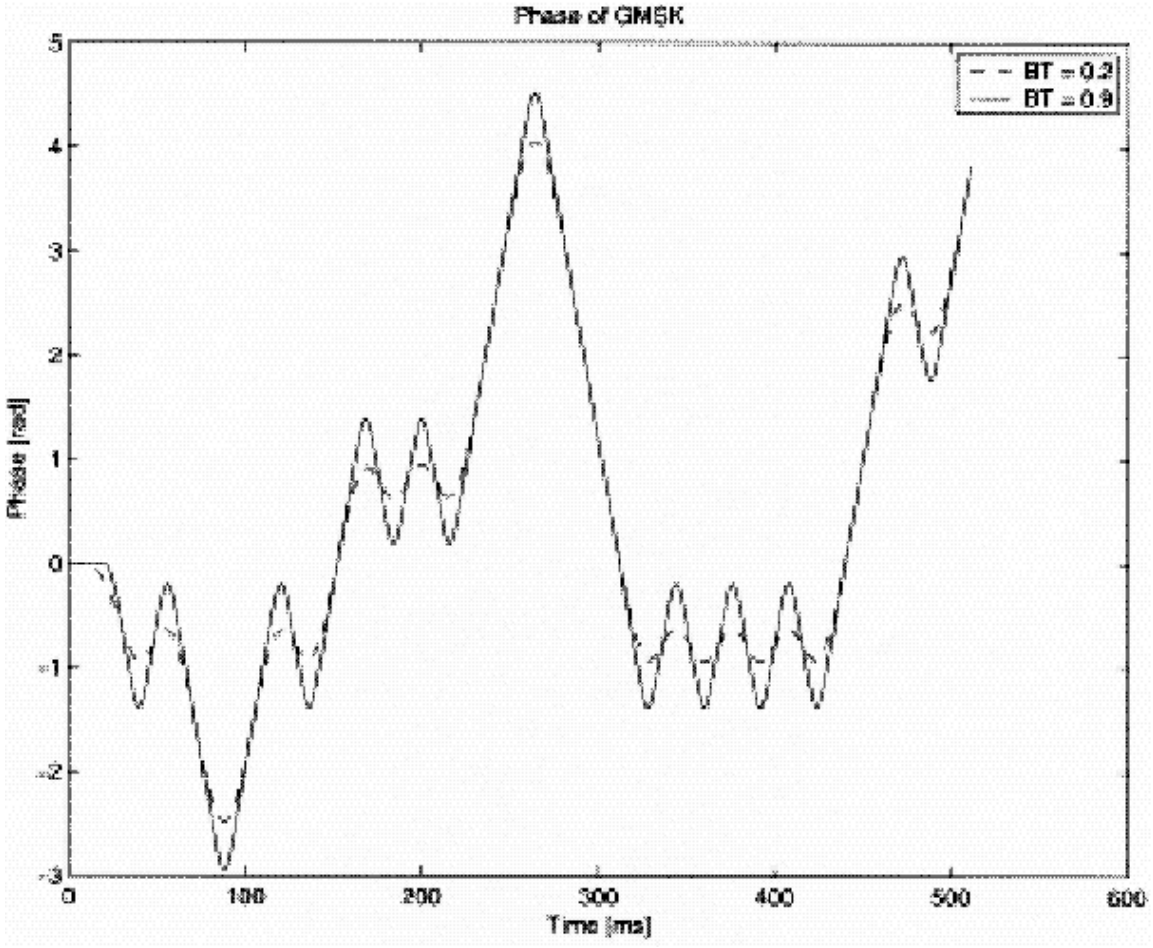


Figure 5.11: GMSK Phase Transitions

With regards to the phase of GMSK, illustrated in Figure 5.11, the spectrum becomes more compressed, the phase transitions tracing a smoother curve, for decreasing values of BT. This is beneficial because it better utilizes the power spectrum. Next, referring to the total GMSK signal output, the graph reveals a wider eye diagram for a increasing BT, as shown in Figures 5.12 and 5.13. This means that it is easier to recover the carrier at the demodulator.

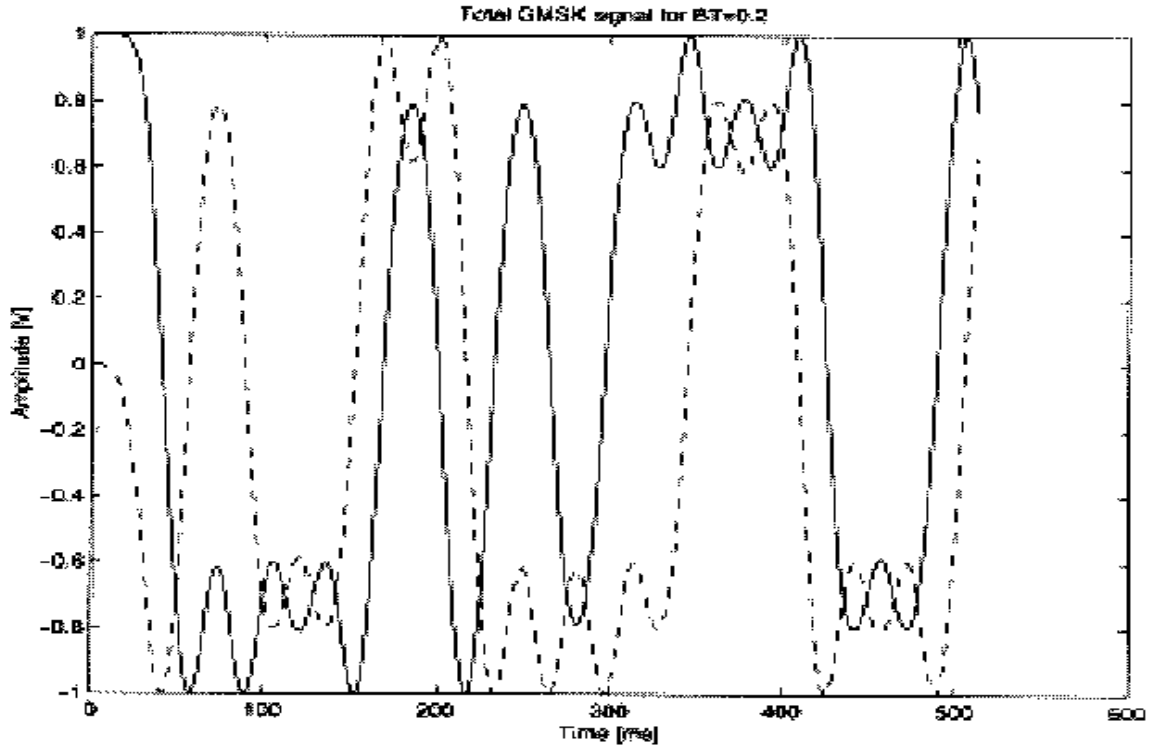


Figure 5.12: Total GMSK Signal, BT=0.2

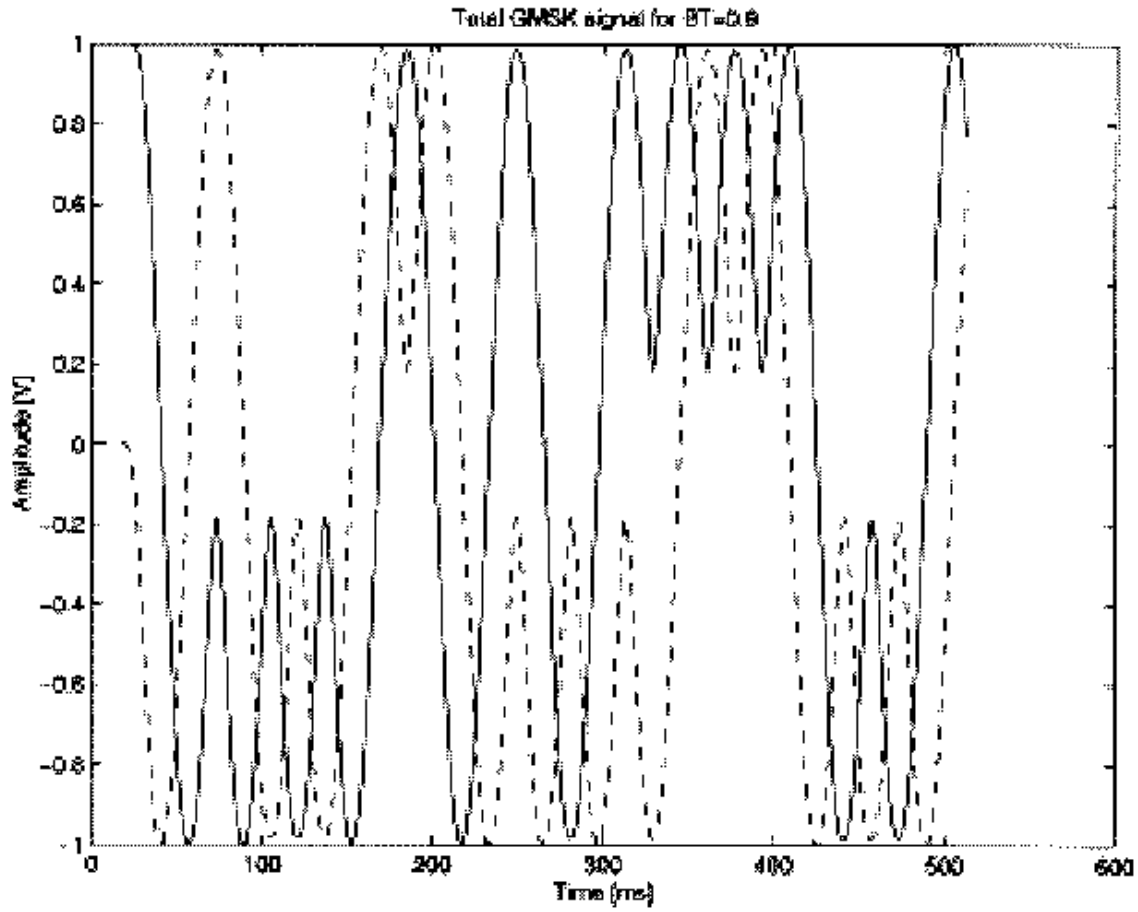


Figure 5.13: Total GMSK Signal, BT=0.9

Implementing QPSK's RC filter shows that increasing the roll-off factor r, makes the spectrum more compact, causing a faster decay of the signal response as seen in Figure 5.10. However, our interest lies in the slow decay and this exists for smaller roll-off values, where ISI is fought more effectively.

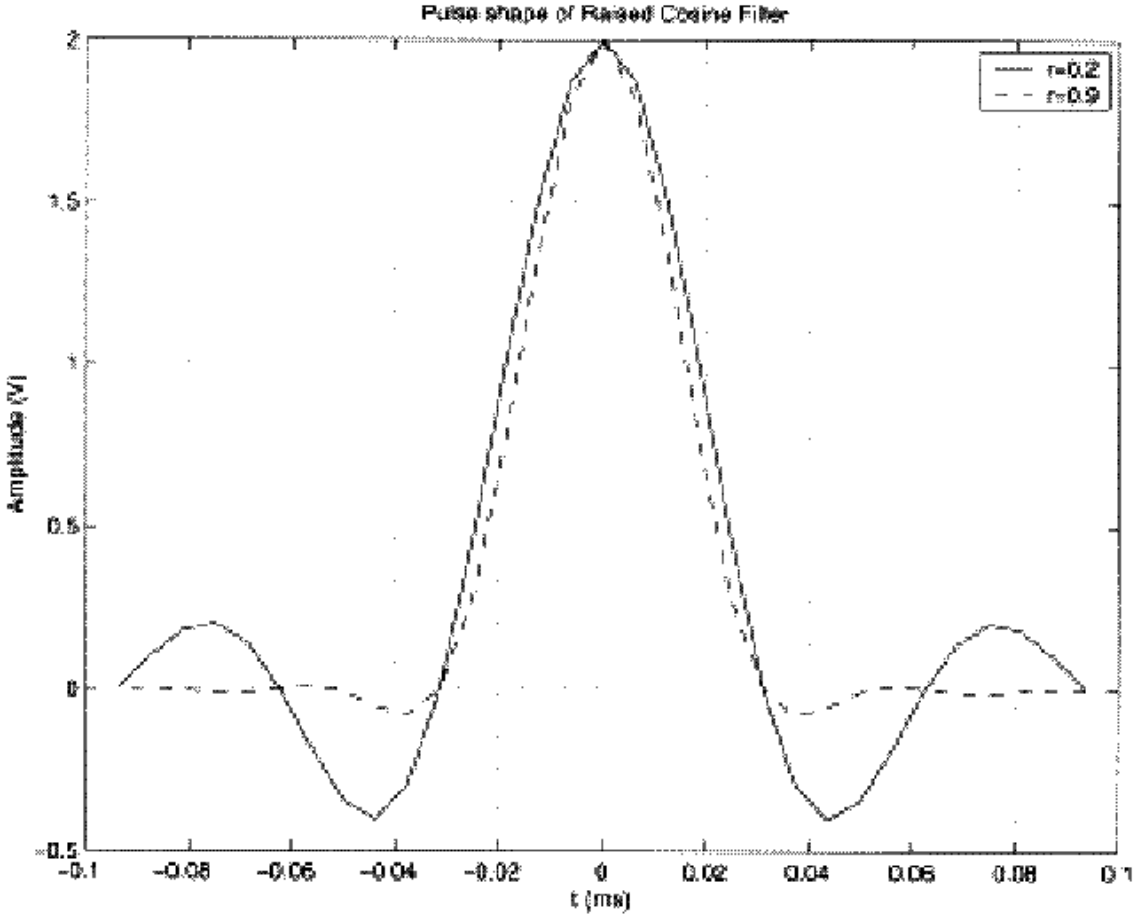


Figure 5.14: QPSK RC Filter Response

Examining the PSD of QPSK in Figure 5.14, decreasing the roll-off factor r expands the spectrum, requiring more bandwidth, and thus, more power. However, since QPSK is predominantly noted for its bandwidth efficient feature, it is preferable to operate at higher values of roll-off in order to accommodate for the increasing demand for more users within a limited channel bandwidth. Nevertheless, another tradeoff is that a complex receiver is needed at the end of the filter to recover the carrier. In order to achieve easier

carrier recovery, a wider eye diagram is necessary as shown in Figure 5.15 with a roll-off factor decreasing from 0.9 to 0.2.

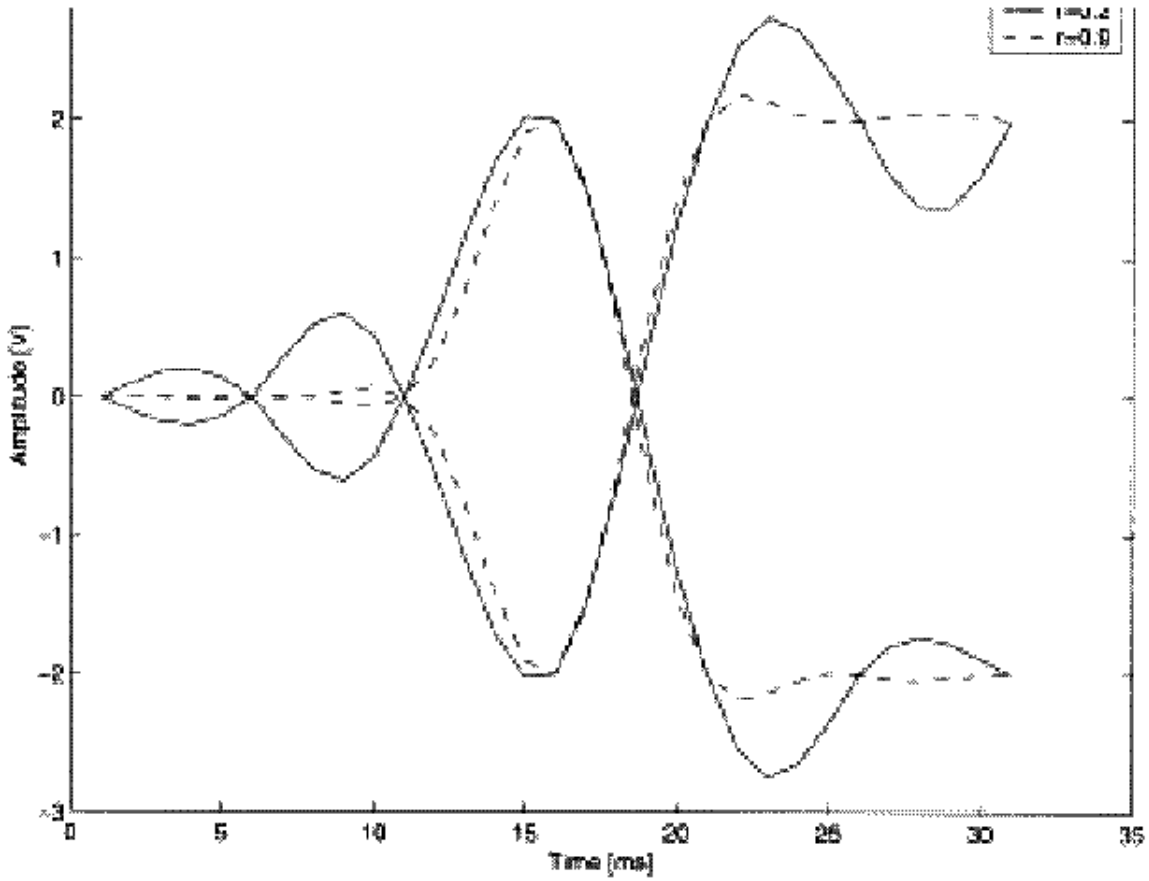


Figure 5.15: Total QPSK Signal

This effect yields less power efficiency; yet since the eye is wide, this shows that minimal ISI is present.

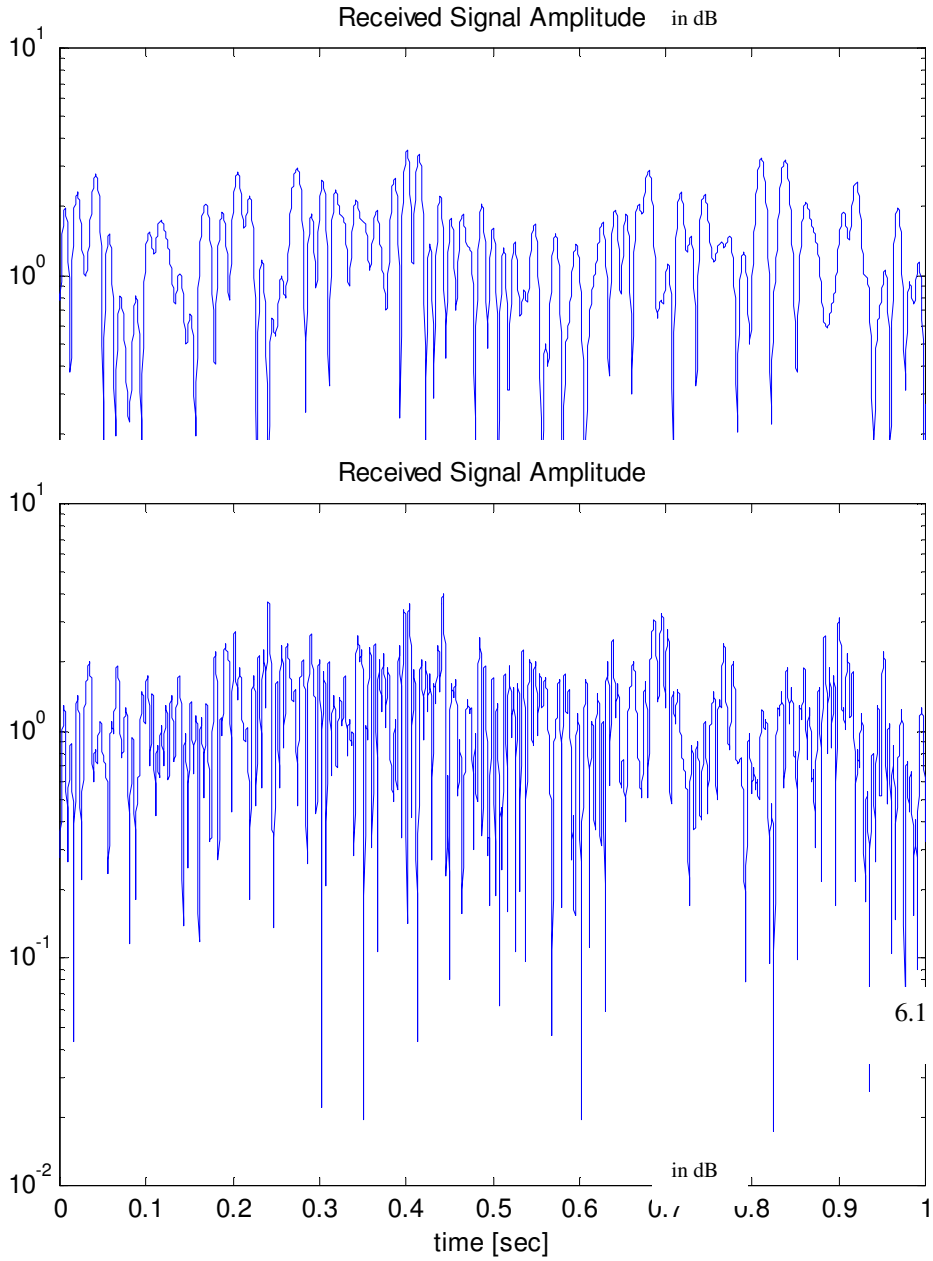
CHAPTER SIX

CONCLUSION AND RECOMENDATION

It is known that the rate of signal fluctuation increases in direct proportion to the speed of motion and that as the mobile speed increases it will receive more and more fades as seen in figures 6.1 and 6.2 for a vehicle speed of 50Km/h and 126 Km/h. The reverse of this enables us to estimate the speed of any moving mobile unit from the received rayleigh faded signal . A joint statistics for a mathematical problem was used to provide a simple expression for computing the average number of level crossing and duration of fades in the simulator. The level crossing rate and average fade duration of a Rayleigh fading signal are two important statistics which are useful for designing error control codes and diversity schemes since it becomes possible to relate the time rate of change of the received signal to the signal level and the velocity of a mobile.

Fig 6.1 Rayleigh fading for a speed of 50 Km/h.

The level crossing rate was defined as the expected rate at which rms s going



6.1

Fig 6.2 Rayleigh fading channel for a mobile speed of 126 Km/hr

The level crossing rate is a function of of the mobile speed as is apparent from the presence of f_m in the equation.

The average fade duration was also defined as the average period of time for which the received signal is below a specified level R.

$$\tau = \frac{e^{\rho^2} - 1}{\rho f_m \sqrt{2\pi}} \qquad 6.2$$

Take as an example The average fade duration for a threshold level of p=0.707 when f=20 Hz.

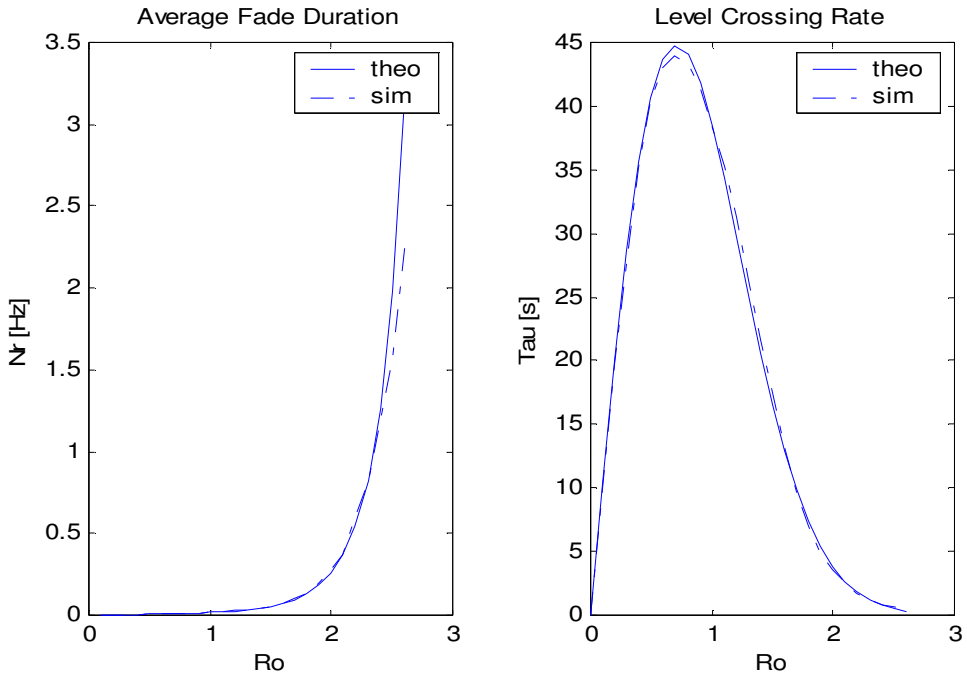
$$\tau = 18.3 \text{ ms} \qquad 6.3$$

For a data rate of 50 bps, the bit period is 20 ms. Since the bit period is greater than the average fade duration, for the given data rate the signal undergoes fast Rayleigh fading.

The average fade duration for p=0.1 is equal to .002 s. This is less than the duration of one bit. Therefore, only one bit on average will be lost during a fade. The number of level crossings for p=0.1 is Nr=4.96 crossings per second. Since a bit error is assumed to occur whenever a portion of a bit encounters a fade, and since average fade duration spans only a fraction of a bit duration, the total number of bits in error is 5 per second, resulting in a BER=5/50=0.1

As seen in the simulation result of figure 6.3 the statistical calculation of Level crossing and the theoretical plot almost align. With the simulation of the statistical estimation of level crossing and the average fade duration values, the normalized error calculated from the above mathematical equation with the simulation output error is 0.01 for level crossing rate and 0.015 for the average fade duration as seen in fig 6.3. This indicates that one can calculate the level crossing and the average fade duration from the received statistical signal of any mobile communication system and hence can easily estimate the speed of a mobile apparatus. hence with this in mind one can calculate the velocity of any moving mobile unit from a base station , to be exact from three base stations .

Fig 6.3 Simulation results of average phase duration and Level crossing rate



With this figure as an initial its possible to design an interleaver and channel coder to counteract this fading. For these reason the channel simulator was important for the design of error correction coder. Not only the channel but also the speech coding affects the design of error correction coder.

Another thing that is found from this study is that almost all Electrical Engineering techniques especially in digital

signal processing and transmission has completely been converted to software implementation rather than being specified only in hardware due to multipurpose programmable digital signal processing which made it possible to implement digital communications like modulators and demodulators completely in software.

In this study only the channel is simulated , the rest speech coding, modulation and error control coding are programmed as what is adopted in standard company coding except for the way the programming algorithm is adopted. The concepts applied are standard for any practical wireless communication system.

This shows that except for the difficulty of miniaturizing and mass production of tight filters and low noise amplifiers, thanks to the technology of programmable digital signal processing, Ethiopian Engineers are not a distant away of mass producing these multipurpose programmable digital signal processing in software.

The astonishing result from the simulator which canot be left undiscussed is the the phenomena of speech coder especially the simulated Linear Predictive coding . The speech signal occupies 64 Kbps is sampled at 8 KHz in an 8 bit stream . Bit the result of the bit rate reduction rate of the Linear Predictive Speech coder is almost $1/27^{th}$ of the actual signal sampling . This is a compression rate that the simulator program outputed at a rate of 2.4 Kbps from a speech signal of 64 Kbps . One can imagine how the bandwidth which is the expensive commodity in mobile wireless communication system is efficiently utilized. Had not been redundency being added for error correction this compression meant it is possible to transmit 27 voice channels in place of the obsolete one channel in which a speech is transmitted uncompressed. Usually this margin won't be achievable due to error control coding which are necessary since each parameters extracted from the Linear predictive coding are significant and important that they should be protected from errors with very much redundancy. Hence in practice speech coding technology increase the capacity to six users per channel in the same bandwidth.

The goal of speech coding systems is to transmit speech with the highest possible quality using the least possible channel capacity. Because speech coders attempt to represent a wide range of voices in a uniform and efficient digital format, the encoded data bits (source bits) carry a great deal of information and some source bits are more important than others and must be protected from errors. It is typical for speech coders to produce several "important" bits in succession, and it is the function of the interleaver to spread these bits out in time so that if there is a deep fade or noise burst, the important bits from a block of source data are not corrupted at the same time. By spreading the source bits over time, it becomes possible to make use of error control coding (called channel coding) which protects the source data from corruption by the channel. Since error control codes are designed to protect against channel errors that may occur randomly or in a bursty manner, interleavers scramble the time order of source bits before they are channel coded.

Channel coding protects digital data from errors by selectively introducing redundancies in the transmitted data. By properly encoding of the information, errors induced by a noisy channel can be reduced to any desired level without sacrificing the rate of information transfer .

The basic purpose of error detection and error correction techniques is to introduce redundancies in the data to improve wireless link performance. The introduction of redundant bits increases the raw data rate used in the link, and, hence, it increases the bandwidth requirement for a fixed source data rate. This reduces the bandwidth efficiency of the link in high SNR conditions, but provides excellent BER performance at low SNR values. It is well known that the use of orthogonal signaling allows the probability of error to become arbitrarily small by expanding the signal set, i.e., by making the number of waveforms M -> ∞, provided that the SNR per bit exceeds the Shannon limit of $SNR_b \geq -1.6$ dB. In the limit, Shannon's result indicates that extremely wideband signals could be used to achieve error free communications, as long as sufficient SNR exists, and this is partly why wideband CDMA

is being adopted and is recommended for future 3G wireless communication link , as its seen in signal orthogonality in the error detection and correction simulation, for its

1. Supports a variable number of users
2. Operate at much larger interference levels because of their inherent interference resistance properties.
3. Low co-channel interference
4. Operate with a much smaller signal-to-noise ratio (SNR)
5. Use the same set of Frequency in every cell, which provide a large improvement in capacity

Error control coding waveforms, on the other hand, have bandwidth expansion factors that grow only linearly with the code block length. Error correction coding thus offers advantages in bandwidth limited applications, and also provides link protection in power limited applications. A channel coder operates on digital message (source) data by encoding the source information into a code sequence for transmission through the channel.

The performance of a modulation scheme is often measured in terms of its power efficiency and Bandwidth efficiency. The results in the study indicate that there is no one prominent modulation scheme between linear and constant envelope modulation methods. Both QPSK and GMSK have strong features that provide a desirable cellular environment. When it comes to any one particular application, it is important to look at the tradeoffs involved. Most mobile products are designed with Class C power amplifiers, which offer the highest power efficiency, yet because they are nonlinear, require the amplified signal to have a constant envelope. This reduces the desirability of implementing QPSK in this situation. However, QPSK effectively utilizes bandwidth; whereas, GMSK requires more bandwidth to effectively recover the carrier. Due to its linear amplification feature, QPSK is able to maintain low spectral sidelobes; thus providing good adjacent channel performance. This is an important contribution to wireless systems because it enables a higher channel reuse factor. Furthermore, QPSK's importance in CDMA is evident with its efficient bandwidth use, enabling more users within a limited channel bandwidth. GMSK makes its contribution to cellular systems in communications from the mobile to the base station. In the case of uplink, power is drained significantly from the mobile, necessitating a power efficient amplifier. GMSK fulfills this need. Furthermore, due to its frequency modulating characteristic, GMSK shows a greater immunity to signal fluctuations. QPSK and GMSK each provide beneficial features, and although neither dominates the other, both contribute to the advancement of wireless telecommunication systems.

As seen in the simulation for lower error rates all five modulation techniques exhibit an inverse algebraic relation between error rate and mean SNR. This is in contrast with the exponential relationship between error rate and SNR in an AWGN channel. According to this result, it is seen that operating at BER's of 10.3 to 10.6 requires roughly a 30dB to 60dB mean SNR. This is significantly larger than that required when operating over nonfading Gaussian noise channel (20 dB to 50 dB more link is required). However, it can be easily be shown that the poor error performance is due to the non zero probability of very deep fades, when the instantaneous BER can become as low as 0.5. Significant improvement in BER can be achieved by using efficient techniques such as diversity or error control coding to totally avoid the probability of deep fades. That is why Gausian minimum shift keying modulation is preferred for Global system of mobiles (GSM) where the detail is left for the next researcher.