MILITARY COLLEGE OF ELECTRONICS AND MECHANICAL ENGINEERING

Faculty of Electronics
Secunderabad - 500 015

DISSERTATION

ON

SIMULATION OF COMMUNICATION SYSTEMS USED IN ARMY EQUIPMENTS USING DIGITAL SIGNAL PROCESSING

81p + fig.

Guide :

Dr.D.P. Roy, VSM
Faculty of Electronics
MCEME

By:

Maj Narendra Kumar Yadav

Jawaharlal Nehru University, New Delhi
1998

MILITARY COLLEGE OF ELECTRONICS AND MECHANICAL ENGINEERING

Faculty of Electronics
Secunderabad - 500 015

CERTIFICATE

Certified that this is a bonafide report of the dissertation work done by Maj. Narendra Kumar Yadav during the year 1998 in partial fulfilment of the requirment for the award of the Degree of Master of Technology in Electrical Engineering by the Jawaharlal Nehru University, New Delhi.

Guide:

Dr.D.P. Roy, VSM
Faculty of Electronics
MCEME

External Examiner

Dr Rameshwar Rao
Dept of Communication engineering
Osmania University,
Hyderabad

ACKNOWLEDGEMENT

I consider it a pleasant duty to express my heartfelt gratitude, appreciate and indebtedness to Dr. D P Roy, VSM of Facility of Electronics for his valuable guidance, supervision and assistance for successful completion of this dissertation.

I am grateful to Dean, FEL and Col G.Ankaiah, Head, Communication Engg. Deptt. for valuable help and Support.

I must also acknowledge my gratitude to Dr. D.Nag Chaudhry, Head of Electrical Engg.Deptt. IIT, Delhi and Mr.V Subramaniam of Centre of Applied Research in Electronics, IIT Delhi, for extending complete cooperation and facilities for long hours of work.

***I would like to thank Col Mrigendra Kumar, Commanding Officer 609 EME Bn for providing me valuable time from the hectic activities of the Bn for completion** of this dissertation.*

Maj Narendra Kumar Yadav

SYNOPSIS

1. Simulation may be defined as the discipline where objective is to imitate one or more aspects of reality in a way that is as close to that reality as possible. Simulation can play an important role during all phases of the design and engineering of communication systems, from the early stages of conceptional design through the various stages of implementation, testing, and fielding of the system.

2. Simulation of a communication system involves generating sampled values of signals and noise, processing these sampled values through discrete time models of functional blocks in communication systems and estimating performance measures. The validity and accuracy of simulation results will depend on the correctness of modeling and estimation techniques and the length of simulation.

3. One of the central activities of simulation is to develop "models" for the various building blocks of a communication system, as well as for the signal and noise process that are the stimuli for any system. The modeling of communication systems is the theme of this dissertation work. A model is any description of the functioning of an element of the system that forms the basis for the representation of that element in a simulation. Ultimately, a simulation requires an executable model, which is the actual code in some programming language. This thesis work concerns with abstract models, which are the conceptual descriptions that are ultimately transformed into executable models. An abstract model may be a set of equations, an algorithm, a table, or some other procedure. In this thesis work we discuss the modeling of a number of functional blocks that may be found in a communication system, although we do not attempt to give a complete exposition of all the available topics due to constraint of time and space. We choose a representative subset to illustrate modeling and simulation approaches.

CHAPTER - I

INTRODUCTION

1.1 INTRODUCTION

There are large number of communication equipments, (63 types to be precise as per data summary complied by Headquarters Technical Group EME), used by Armed Forces in all types of terrain viz plains, desserts, Jungles, Mountains and snow bound areas. All these communication equipments are designed and tested through various trials in all the terrains of the country for their optimum performance before inducting in Armed Forces. All these equipments involve Analog and Digital systems.

Simulation can play an important role during all phases of communication systems, from the early stages of conceptual design through the various stages of implementation, testing and fielding of the system .

The validated simulation model can be used to predict the End of Life (EOL) performance of the system using postulated characteristics of key components due to aging. The validated simulation model can also be used during operational stages for trouble shooting and for providing answers to "what if" scenarios e.g a simulation program including digital filter to model the transmitter, detector and the transmission channel may attempt to answer the following questions. What frequencies should be used for more reliability? What detector is suitably reliable but reasonably expensive or fast?. To investigate the complicated effects of interference can be added to the transmitted signal and the resulting characteristics of a given detector can be examined. All these procedures can be performed on computer without taking the equipment to farflung places.

1.2 AIM

The objective of this thesis is to develop "models" for the various building blocks of a communication system, which forms the basis for simulation.

1.3 METHODOLOGY

Chapter I gives the introduction, aim and outline of the thesis.

Chapter II encompasses modeling of generic communication system in its section as follows.

Section 2.1 gives introduction to modeling of communication systems with a blocks diagram of a "generic" communication system in the sense that virtually any communication system can be specified as a particular case of this block diagram.

Section 2.2 gives the stimuli or driving functions modeling in communication system for Analog and Digital signals both.

Section 2.3 give the models of source Encoders of Decoders.

Section 2.4 gives the models of baseband Modulation.

Section 2.5 gives models of RF modulation

Section 2.6 gives models of Demodulation including coherent and Noncoherant Demodulation both

Section 2.7 gives models of Filtering

Section 2.8 gives models of Tropospheric, Rainy atmosphere and Ionospheric Phase Channels.

Section 2.9 gives the models of Interference.

Chapter III gives the conclusion and further scope of studies for simulation of communication systems and at the end Bibliography is given.

CHAPTER - II

MODELING OF COMMUNICATION SYSTEMS

2.1 INTRODUCTION

The simulation of a communication system requires a software representable description of the system. The standard description of a system is a block diagram, where each block represents a signal-processing operation. The block diagram, as such, is really only a signal flow diagram in the sense that it merely indicates the generic type of operations that the signal(s) and noise(s) that drive the system are subjected to. While there are many different types of communication systems using a wide range of technologies, information transmission in all communication systems takes place through a series of basic (or generic") signal-processing operations. The following operations are fundamental to all communication systems (although, in a given system, not all operations need appear); source encoding and decoding, modulation and demodulation, multiplexing, error control coding/decoding, filtering (fixed and adaptive), and synchronization. Although many of these terms are usually associated with digital transmission, they can in some cases also be given a meaningful interpretation for analog. Figure 2.1 (adapted from Ref.1) shows a block diagram of a "generic" communication system in the sense that virtually any communication system can be specified as a particular case of this block diagram.

We take the functions to be modeled as those generic-processing operations given in Figure 2.1. indicate, where applicable, the ways in which we can characterize nonideal behavior.

2.2 INFORMATION SOURCES

The stimuli or driving functions in communication systems are the outputs of various sources of information, noise, and interference. Outputs of these sources may be random processes or deterministic functions and they may be analog or digital in nature. The term "signal" is traditionally used to refer to the output of information sources, and the "signal component" of an observed waveform contains information of interest. Thus the terms "signal" and "information" are synonymous. Examples of analog signals include a sinusoidal tone, the output of a microphone or of a TV camera. Digital signals are those which contain embedded digital sequences; that is, the information to be sent is digital (discrete), but the method of transmission may involve analog waveforms.

Noise and interference represent the undesirable components of a waveform. Noise arises due to natural causes and interference is man made. Interference may be unintentional (e.g. hum due to harmonics in a power supply, cross talk due to nonlinearities, etc.) or it may be intentional (e.g, jamming).

The commonly used models of signals and noise are discussed in the following sections. While there is a clear dichotomy between the contents of signals (information) and noise (useless), the models of signals and noise used in simulations are not necessarily dichotomous. For example, a tone might be used to represent a signal or interference, Similarly, we may use a band-limited Gaussian processes to model both signals and noise.

2.2.1 Analog Signals

Analog signals are usually specified in terms of their power spectral density and amplitude distribution. For simulation purposes, an analog signal is normally modeled either by sampled values of a single tone, the sum of tones, or a filtered random process. Sometimes a combination is also used to create a test signal.

Single Tone. Single tones (usually called test tones) are perhaps the most common test signal used in communication systems. For simulation purposes we must of course use sampled values of the test tone. Thus, the discrete-time signal

$$X(t_k) = A\cos(2\pi f_0 t_k + \theta) \qquad (2.2.1)$$

is used to represent a test tone within a simulation. The sequence of sampling instants $\{t_k\}$ is always equidistant in simulation, and separated by the sampling interval T_s, i.e., $t_k = kT_s$. The amplitude A, and frequency f_0 are varied to obtain the so-called swept frequency response and the swept power response of the system. In theoretical models, the angle θ is usually assumed to be uniformly distributed over $[0,2\pi]$. However, the response of a system is not normally sensitive to the value of θ, hence, for simulation purpose the angle θ is usually fixed at an arbitrary value. In bandpass systems, the tone frequency f_0 is offset from the center frequency f_c. That is, the sampled tone has the form

$$X(t_k) = X(k) = A\cos[2\pi(f_0 + f_c)t_k + \theta]$$

In either case, the sampled complex envelope representation used for simulation is

$$\tilde{X}(k) = A\exp(2\pi jkf_0/f_s)\exp(j\theta) \qquad (2.2.2)$$

where $f_s = T_s^{-1}$ is the sampling frequency. Note that f_s is typically set at 8 to 16 times f_0. It is necessary that f_s be an integer multiple of f_0 in order that the sampled sinusoid be periodic.

Multiple Tones : A set of multiple tones is typically used to evaluate intermodulation distortion in nonlinear systems. The complex envelope representation of a set of tones has the form

$$\tilde{X}(k) = \sum_{n=1}^{M} A_n \exp\left[2\pi k \frac{f_n}{f_s} + j\theta_n\right] \tag{2.2.3}$$

where a_n, θ_n, and f_n represent the amplitude, phase, and frequency of the nth tone

$$X_n(t) = A_n \cos[2\pi(f_0+f_n)t + \theta_n] \tag{2.2.4}$$

Unless there is known to be a particular relationship among the θ_n, they would be assumed independent and uniformly distributed over $[0,2\pi]$. Thus, a uniform random number generator would be called to generate the θ_n.

Filtered Random Processes. Analog signals that are random (such as audio and video signals) can be simulated using the random process models. These signals, which are typically low pass prior to modulation, are specified in terms of their power spectral density and their amplitude distribution, which is assumed to be Gaussian in many cases.

With the Gaussian assumption, sampled values of signals can be generated by applying a linear transformation to sequences of independent gaussian variables. The linear transformation preserves the Gaussian amplitude distribution and alters only the power spectral density. By choosing the coefficients of the transformation appropriately, one can shape the spectral density of the transformed sequence in such a way as to match some desired shape as closely as possible. The coefficients of the transformation can be determined by fitting an autoregressive (AR) or autoregressive moving-average (ARMA) model to the given spectral density or by factoring the given spectral density (assuming it is given by a ratio of polynomials) and constructing a filter that contains the poles and zeros of the factored spectral density that lie inthe left half of the complex frequency domain. The transformation can also be implemented with an FIR filter.

It is usually very difficult to generate sampled values of random processes with arbitrary power spectral density and arbitrary amplitude distribution. While it is easy to control either one of these two attributes, no easily implemented general procedures are available to provide control of both attributes except for the Gaussian case.

2.2.2 Digital Signals

Digital signals contain embedded digital sequences. A typical example of a digital signal is

$$X(t) = \sum_{k=-\infty}^{\infty} A_k p(t - kT_b - D) \qquad (2.2.5)$$

where $\{A_k\}$ is a digital sequence, T_b represents the time between successive elements of the sequence, D is a random delay, and p(t) is a suitable "pulse" waveform which "carries" the digital sequence. The elements A_k are drawn from a finite set of real numbers, and are usually referred to as symbols, or as bits when the set has only two possibilities. X(t) is often referred to as a baseband signal. in much of the literature, p(t) is taken as a rectangular pulse T seconds long. This is an idealization which is easy to analyse but cannot be realized in practical equipment. More realistic versions of the idealized rectangular pulse train are not so simple to analyze but can be simulated relatively straightforwardly. For example, one might use as an approximation to the behavior of real equipment the trapezoidal waveform shown in Figure 2.2. It is to be noted in this representation that p(t) is actually not a unique pulse, but depends on the logical value of the previous as well as the current symbol. In principle, the duration of p(t) need not be limited to a symbol duration, nor even be of finite duration. If we impose certain conditions on the zero crossings of p(t), it turns out that p(t) is not time limited although its effective duration is finite.

The sequence $\{A_k\}$ might have been mapped from the output of a continuous source or from another sequence $\{b_k\}$. Indeed, there can be a succession of mappings between the information source and the channel input sequence $\{A_k\}$. For example, $\{A_k\}$ might be the sequence of quantized outputs of an analog source and would then represent an M-ary digital source. As another example, $\{b_k\}$ might represent the quantized M-ary sequence and $\{A_k\}$ might be a different sequence better suited for transmission, e.g., binary PCM, Manchester, etc.; this type of mapping is sometimes referred to as (baseband) formatting or line coding. Another common formatting process is a mapping called differential encoding, which is frequently used in digital communications. Sometimes the situation arises where $\{b_k\}$ is a sequence from a naturally digital M-ary alphabet, such as the English language,which has an inherent redundancy, and $\{A_k\}$ is a corresponding sequence with the redundancy removed; this type of mapping. is called source coding. A different type of mapping is error-correction coding in which, typically, a binary sequence $\{b_k\}$ has redundancy added to it in a structured way to from an encoded sequence $\{A_k\}$. The recovery of $\{b_k\}$ from $\{A_k\}$ for any of the above mappings is done by an appropriate inverse operation of "decoding". The details of the mappings and inverse mappings between $\{b_k\}$ and $\{A_k\}$ will be discussed subsequently.

The embedding of A_k onto X(t), as already implied, is done by associating with each symbol A_k a time waveform p(t). This process is known as baseband or pulse modulation. Recovering the logical sequence from the time sequence is done in the detection process. For carrier modulation systems, the digital signal X(t), 2.2.5, modulates the carrier in some fashion and is recovered inthe receiver through the inverse function of demodulation. Details of the modulation, demodulation,and detection operations are described later.

Within a simulation, the digital signal X(t), given by (2.2.5), itself has to be sampled in order to generate a sampled value X(n). The sampling rate is usually set at 8 to 16 times the rate of the sequence A_k(i.e., f_s = 8 to 16 times $1/T_b$.

2.3 SOURCE ENCODERS/DECODERS

A source encoder maps the output of an information source into a binary (or M-ary) sequence. In perhaps the most common case, this operation consists of sampling the analog output of the information source, quantizing the sampled values and encoding the qantized values into binary digits. The quantizing operation, per se, is central to the transmission of analog sources by digital means since it transforms an analog source to a digital source. This process is frequently referred to as analog-to-digital (A/D) conversion. As an example, if the information source is speech, then the audio signal is typically sampled at a rate of 8,000 samples per second, each sample is quantized in 2^k levels and each quantized sample is represented (encoded) by a group of k binary digits.

Values of k range from 6 to 10 bits/sample, which produces a bit rate of 48,000 to 80,000 bits per second. The decoder (or D/A converter) performs the complementary operation of mapping groups of k bits of the encoder output into 2^k levels. A filter is used to reconstruct an analog signal (which is an approximation to the original) from the sampled values (Figure 2.3).

The quantizing operation typically maps an analog variable X into a discrete variable X_q. (This could also be a mapping of a discrete variable with N values to a new variable X_q with Q values, $Q < N$.) This operation introduces a quantizing error X_q - X, and it is not possible to recover the exact value of X from X_q. The objective of the quantizer designer is to develop a mapping algorithm from X to X_q such that some performance measure sûch as the mean squared error, $E\{(X - X_q)^2\}$ is minimized for a given value of Q.

If the sampled sequence $X(kT_s)$ is uncorrelated, then each sample in the sequence is quantized independently using a uniform or nonuniform quantization algorithm. If the sequence is correlated, then the correlated samples are first processed to create an uncorrelated sequence prior to quantizing. We describe below an example of a differential quantization algorithm for quantizing correlated samples.

Encoding of a sequence of discrete variables $X_q(kT_s)$ into a set of M-ary variables is done such that exact reconstruction (or decoding) is possible, i.e., encoding is a one-to-one transformation, whereas quantizing is not. The choice of encoding algorithm depends on the (joint) probability distribution of $X_q(kT_s)$. The simplest encoding algorithm maps each member of the sequence $X_q(kT_s)$ in a group of n, M-ary symbols. At the other end are fairly complex encoding algorithms which map blocks of $X_q(kT_s)$

into a variable number of M-ary symbols. These variable-length encoding algorithms are derived using the framework of information theory and we present one example of a variable length, block encoding algorithm.

2.3.1 Quantization

Quantization is the heart of the A/D conversion process. In the classical scheme, the input voltage range is divided into equal intervals of size Δ; the ith interval corresponds to $i\Delta \pm (\Delta/2)$. If the input voltage falls in the ith interval, it is assigned the integer i; if i is written in binary form, the output of the process is a bit sequence and the result is termed pulse code modulation or PCM. Because the step sizes are equal, this scheme is called uniform quantization.

Nonuniform quantization divides the input range into unequal intervals. In practice, this is accomplished in two steps (Figure 2.4). First, the signal is processed through a nonlinear device called a compressor, and the result is fed to a uniform quantizer. This dual process is clearly equivalent to nonuniform quantization of the original input. In practice, there are two standard compression schemes, referred to as μ-law and A-law compression. Let ν_c be the compressor output, ν the input, and V the maximum (or 'overload") input voltage, and define $x_c = \nu_c/V$ and $x = \nu/V$. Then, the compressed output for μ-law is given by

$$x_c = (\text{sgn } x) \frac{\log(1 + \mu|x|)}{\log(1 + \mu)} \qquad (2.3.1)$$

The expression(2.3.1) is also called logarithmic compression; a commonly used value for μ is 255 [2].

The A-law characteristic is described by

$$x_c = (\mathrm{sgn}\,x)\,\frac{A|x|}{1+\ln A},\quad 0 \le |x| \le A^{-1} \qquad (2.3.2a)$$

$$= (\mathrm{sng}\,x)\,\frac{1+\ln A|x|}{1+\ln A},\quad A^{-1} \le |x| \le 1 \qquad (2.3.2b)$$

The source decoder is shown in Figure 2.4b. the estimate of x_c is simply the inverse mapping from bits to the corresponding integer, i.e.,

$$\hat{x}_c = (\mathrm{sgn}x)\,(\Delta/2 + i\Delta,),\ \ \Delta = 2^{-m}, i = 0,1,\ldots,2^m - 1$$

where m is the number of bits per sample (excluding the sign bit). An estimate of the original input is then provided by the expander, which performs the inverse of compression. Thus, if we express (2.3.1) or (2.3.2) as (sgn x) g(x), than the expander output is

$$\hat{x} = \mathrm{sgn}(\hat{x}_c)g^{-1}(\hat{x}_c) \qquad (2.3.3)$$

2.3.2 Differential Quantization

It can be shown that the MSE due to quantizing X (kT_s) into $X_q(kT_s)$ using Q levels is

$$E\{[X(kT_s) - X_q(kT_s)]^2\} = \alpha\frac{Var[X_q(kT_s)]}{Q^2}$$

where α is a proportionality constant. If $\{X(kT_s)\}$ is a correlated sequence, then the variance of $X(kT_s)$ - $X((k-1)\,T_s)$ will be small compared to the variance of $X(kT_s)$. This observation suggests quantizing $[X(kT_s) - X((k-1)\,T_s)]$ instead of $X(kT_s)$ to reduce MSE.

A more general version of this algorithm using a predictor is shown in Figure 2.5 assuming $T_s = 1$. The predicted value $\hat{X}(k)$ of $X(k)$ is a linear combination of previous values, that is

$$\hat{X}(k) = \sum_{i=1}^{M} a_i X(k-i) \tag{2.3.4}$$

The predictor coefficients a_i that minimize $E\{[\hat{X}(k) - X(k)]^2\}$ can be obtained by solving the set of simultaneous equations

$$R_{xx}(j) = \sum_{i=1}^{M} a_i R_{xx}(j-i), \quad j = 1,2,\ldots M \tag{2.3.5}$$

assuming that $R_{xx}(j)$, $j = 1,2\ldots, M$, are given. Adaptive, recursive methods for solving this problem when $R_{xx}(j)$ is not known (and/or slowly changing) are equivalent to adaptive equalization. More general techniques are given in Refs. 3 and 4.

Equation 2.3.4 uses X(k-i), whereas the receiver has to operate on $\hat{X}(k-i) + [X(k-i) - \hat{X}(k-i)]_q$. If we assume that the quantizing error is small, then the latter can be substituted for the former. Hence the predicted values are formed in the transmitter and receiver using $\hat{X}(k-i) + [X(k-i) - \hat{X}(k-i)]_q$ as shown in Figure 2.5.

2.3.3 Encoding the Output of Discrete Information Sources

A discrete information source emits sequences of symbols from an alphabet of symbols. The alphabet of symbols may be the letters of the alphabet of a language such as English, or they may represent the output levels of a quantizer. Source encoding is the process by which the output of the information source is converted to a binary (or M-ary) sequence. In order to conserve bandwidth it is desirable that the encoding be done with the output bit rate as low as possible and the constraint that the encoding process be unique, i.e., such that the sequence of symbols can be recovered without errors from the encoder output.

There are many algorithms for encoding the output of a discrete information source. We present below an algorithm first proposed by Shannon [5] which is representative of source encoding algorithms. Shannon's algorithm maps sequences from an M-ary alphabet source into sequences of binary digits.

Let us assume that the input to the encoder consists of blocks or "messages' of N symbols. Suppose there are q(N) possible messages $m_1, m_2, \ldots m_{q(N)}$ containing N symbols. Let $p_1, p_2 \ldots, p_q(N)$ be the probabilities of these messages, $(p_1 + p_2 + \ldots P_{q(N)}(N) = 1$, and let us a assume that $p_1 \geq p_2 \ldots \geq p_{q(N)}$. Shannon has shown that the minimum average number of bits per symbol needed to encode blocks of N symbols is

$$H_n = \frac{1}{N} \sum_{i=1}^{q(N)} p_i \log_2 (1/p_i)$$

He has also shown that $H_{N+1} \leq H_N$, that is, as the block size or the message length N is increased, the average number of bits per symbol needed to encode the source output decreases monotonically to a nonzero limit H, which is called the "source entropy". H_N and H can be computed for stationary sources if the probability distributions of the source output are known. Shannon proposed the following algorithm to encode blocks of N symbols. Suppose the messages $m_1, m_2 \ldots$, are arranged such that $p_1 \geq p_2 \geq \ldots \geq p_q$. Let

$$F_i = \sum_{k=1}^{i-1} p_k \quad \text{with } F_1 = 0 \qquad (2.3.6)$$

and n_i be an integer such that

$$\log_2 (1/p_i) \leq n_i < 1 + \log_2 (1/p_i)$$

Then the binary encoding for the message m_i is the binary expansion of the fraction F_i up to n_i bits.

This algorithm produces unique encoding for each message and it assigns a variable number of bits to each group of N symbol. Messages with higher probability are assigned a shorter code. The average number of bits per symbol used by this algorithm is

$$\hat{H}_N = \frac{1}{N} \sum_{i=1}^{1(N)} n_i p_i \qquad (2.3.7)$$

It can be shown that

$$H_N \leq \hat{H}_N < H_N + \frac{1}{N}$$

i.e., the algorithm is asymptotically optimal. An example of this algorithm with M = 3, N = 2, and q(N) = 7 is shown in Table 2.1.

Table 2.1 Example of Variable Length Source Encoding Algorithm

Message	p_i	n_i	c_i
AA	9/32	2	00
BB	9/32	2	01
AC	3/32	4	1001
CB	3/32	4	1010
BC	3/32	4	1000
CA	3/32	4	1101
CC	2/32	4	1111

$\hat{H}_2 = 1.44$ bits/symbol

$H_2 = 1.279$ bits/symbol

If the symbols emitted by the sequence are independent, that is X(k), X(k-1),... are independent, and the letters of the alphabet occur with equal probability, then each symbol in the sequence can be encoded independently (there is no need to take groups of N symbols for encoding). IF $Q = 2^k$, each symbol is then encoded into a simple k-bit binary representation.

With regard to simulating the source encoding or decoding process we note first that the physically idealized A/D operation (essentially sampling and quantizing), whether it is uniform or nonuniform quantization, is basically a memoryless nonlinear transformation. Such a transformation is straightforward to implement, namely, read the value of the input signal at the sampling instant and compute the corresponding output. Real A/D converters depart from this idealization in a number of ways depending upon the actual hardware design and the processing speed. For most system-oriented applications it will suffice to assume ideal A/D converters. At the other end of the link there will usually be digital-to-analog conversion, which reconstructs an analog waveform. This usually involves two steps, a mapping from some digital sequence to another digital sequence with the same number of symbols as the originally quantized sequence, and an analog filtering operation. These are deterministic steps, which, once specified, are straightforwardly implemented. All other aspects of source encoding and decoding are basically digital-to-digital mappings and, as such, can be reproduced without approximation within the simulation context.

Thus, for further discussion from this point on, we are justified in assuming that the source output has been encoded into an independent binary sequence with the digits '0' and '1' occurring with equal probability, or an indpendnet M-ary sequence with each symbol occurring equally probably. With this assumption, we can replace the source and the source encoder with a random binary (M-ary) or PN sequence for simulation purposes. Of course, the sampling, quantizing, and encoding/decoding operations may have to be simulated in detail if the primary focus of the design is on these operations.

2.4 BASEBAND MODULATION : FORMATTING; LINE CODING

As we saw in Section 2.2 a digital signal is described by a real modulating sequence $\{A_k\}$ and an associated pulse waveform. The sequence $\{A_k\}$ is typically obtained through one or more operations on a source sequence. It is convenient to think of the source sequence as a logical sequence, i.e., one drawn from the integer set $\{0,1,...L-1\}$, for an L-ary alphabet. For present purposes we can think of the source sequence as the output of the source encoder.

The process of creating the digital baseband signal we refer to as baseband (pulse) modulation. Variations and generalizations of baseband pulse modulation (also called line coding or baseband formatting) have been devised for various purposes, such as spectrum shaping or inducing a certain density of zero-crossings to aid synchronization. These desirable properties are put to use not only for baseband transmission but also in carrier modulated as well as optical systems.Inthe following, we present a few models for some of the more commonly encountered baseband modulation techniques. For simulation (as well as conceptual) purposes, it is useful to divide the process into two parts. The first we refer to as a logical-to-logical mapping, which converts one binary or M-ary sequence into another sequence with desired properties. This may be viewed as a form of "coding". The second part we refer to as logical-to-real mapping. its function is to associate chosen pulse waveforms with particular logical sequences.

2.4.1 Logical-to-logical Mapping I : Binary Differential Encoding

To avoid confusion, we shall label logical sequences by lower case letters. This mapping converts an input (0,1) binary sequence $\{a_m\}$ into a new (0,1) sequence $\{b_m\}$ which conveys the information through changes between the current output bit and the next source bits. Symbolically, the relationship between the sequences is given by

$$b_m = a_m \oplus b_{m-1} \qquad (2.4.1)$$

where $\oplus$ represents modulo-2 addition.

2.4.2 Logical-to-Logical Mapping II: Correlative Coding

This mapping converts an M-ary sequence $\{b_m\}$ into a Q-ary sequence $\{c_m\}$ with properties considered to be desirable, particularly in the frequency domain (see e.g., Ref. 6 and 7 for a good discussion). Typically, the M-ary source sequence $\{a_m\}$ is first "precoded," to avoid error propagation, into sequence $\{b_m\}$. Precoding with $M <= 2$ is equivalent to differential encoding. Pulse waveforms associated with the encoded sequence are referred to as partial response signaling, and will be considered in Section 2.4.8. The precoded sequence is obtained through the operation [8].

$$b_m = \frac{1}{g_o}\left\{a_m - \sum_{i=1}^{N-1} g_i b_{m-i}\right\} \bmod M \qquad (2.4.2)$$

and the coded sequence is given by

$$c_m = \sum_{i=0}^{N-1} g_i b_{m-i} \qquad (2.4.3)$$

where Q, the alphabet size of $\{c_m\}$, is given by

$$Q = (M-1)\sum_{i=0}^{N-1} |g_i| + 1 \qquad (2.4.4)$$

The coefficients $\{g\}$ in 2.4.3 are called partial response coefficients, and determine the properties of the resulting waveform. The parameter N is referred to as the memory of the encoding process.

Remark. A variety of partial response signals have been devised. These differ from one another by the set of coefficients $\{g_i\}$. The most common classes of sets are capsuled *in Table 2.2. A concise way of describing these coefficients is through a polynomial in the delay variable D, i.e,* $F(D) = \sum g_i D^i$

2.4.3 Logical-to-Real Mapping 1 : Non-Return-to-Zero (NRZ) Binary Signaling

NRZ signaling is the "traditional" logical-to-real mapping and is commonly used for carrier modulation schemes such as PSK. The logical sequence can be directly the source output a coded sequence. The mapping is given by

$$0 \rightarrow A_p(t) \qquad (2.4.5a)$$

$$1 \rightarrow -A_p(t) \qquad (2.4.5b)$$

Generally, any pulse shape p(t) with desirable properties can be used. The choice of p(t) will generally depend on the nature of the channel following the pulse modulator. The most commonly assumed pulse is the unit rectangular pulse

$$p_T(t) = \begin{cases} 1 & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \qquad (2.4.6)$$

Table 2.2 List of Partial Response Polynomials

Nomenclature	F(D)
Duobinary, or Class 1 P.R.	$1 + D$
Dicode, bipolar,or AMI	$1 - D$
Class II P.R.	$1 + 2D + D^2$
Class III P.R.	$2 + D - D^2$
Modified doubinary, or Class IV P.R.	$1 - D^2$
Class V P.R	$1 - 2D^2 + D^4$

The notatiaon $P_T(t)$ as in (2.4.6) will be used consistently in this subsection.

2.4.4 Logical-to-Real Mapping II : NRZ M-ary Signaling (PAM)

This mapping is an extension of binary NRZ to M-ary alphabets. It produces what is commonly called pulse amplitude modulation (PAM) namely.

$$S_m(t) = A[(M - 1) - 2a_m]p_T(t) \tag{2.4.7}$$

This mapping yields (M/2) pairs of antipodal signals, that is, $A_k \in \} \pm A, \pm 3A, \ldots, \pm(M - 1)A\}$.

2.4.5 Logical-to-Real Mapping III: Return-to-Zero (RZ) Binary Signaling

RZ signaling is another method of associating a pulse waveform with a sequence. The signal always contains a discrete spectral component. The mapping takes the form

$$0 \rightarrow 0 \tag{2.4.8a}$$

$$1 \rightarrow \begin{cases} -Ap_{T/2}(t), & 0 \le t \le T/2 \\ 0, & T/2 \le t \le T \end{cases} \tag{2.4.8b}$$

Thus $A_k \in \{0, A\}$

2.4.6 Logical-to-Real Mapping IV : Biphase Signaling or Manchester Code

This method of generating real waveforms corresponding to a logical sequence has three useful properties. It has precisely zero average value, i.e., no dc wander; it has a spectral null near dc, which is desirable for channels with poor dc response, or allows insertion of a discrete carrier component in carrier transmission; and it is self-synchronizing inthe sense that there is a zero crossing in the middle of every bit. The mapping is given by

$$0 \rightarrow A[p_{T/2} - p_{T/2}(t - T/2)] = A_{pM}(t) \qquad (2.4.9a)$$

$$1 \rightarrow -A_{pM}(t). \qquad (2.4.9b)$$

2.4.7 Logical-to-Real Mapping V : Miller Code or Delay Modulation

Another method of associating waveforms with binary sequences, the Miller code, differs from the preceding ones in that the sequence of waveforms is desirable by a Markov chain [9]. The spectrum is highly concentrated near about half the baud rate. The mapping is given in table 2.3 where (within the factor A).

$$p_1(t) = - p_4(t) = p_T(t) \qquad (2.4.10a)$$

$$p_2(t) = - p_3(t) = p_M(t) \qquad (2.4.10b)$$

2.4.8 Logical-to-Real Mapping VI : Partial Response Signaling

This mapping may be thought of as the method of associating real waveforms with correlative encoding. The typical pulse is given by

$$s_m(t) = A[(Q-1) - 2c_m]p(t) \qquad (2.4.11a)$$

Assuming $c_m \in (0,1,2,...Q - 1)$, the above equation Q/2 pairs of antipodal signals. The range of c_m will vary depending upon the coefficients g_i. If they are naturally balanced around zero, then we have simply.

$$s_m(t) = Ac_m(t) \qquad (2.4.11b)$$

The pulse p(t) is often assumed to be a Nyquist pulse, which has desirable zero-crossing properties. This pulse describes the wanted overall system response, which is a composite of the responses of

Table 2.3 Mapping Rule for Symbols to Waveforms for Miller Code

Input bit	Previous transmitted signal	Current transmitted signal
0	$p_1(t)$	$p_4(t)$
1	$p_1(t)$	$p_2(t)$
0	$p_2(t)$	$p_4(t)$
1	$p_2(t)$	$p_3(t)$
0	$p_3(t)$	$p_1(t)$
1	$p_3(t)$	$p_2(t)$
0	$p_4(t)$	$p_1(t)$
1	$p_4(t)$	$p_3(t)$

the individual elements of the system. When we simulate the system, we do not directly produce the system response because we simulate on a block-by block basis. Thus, an equation like (2.4.11b) describes the transmitter output, where the pulse p(t) is appropriately chosen, but is not the overall response.

The above equation can be represented in alternative fashion. Let

$$q(t) = \sum_{i=0}^{N-1} g_i p(t-iT) \tag{2.4.12}$$

represent a partial response pulse. Then, we have

$$s_m(t) = A[(M-1) - 2b_m]q(t) \tag{2.4.13}$$

Table 2.4 Definition of Digital Line Coding Formats

Noreturn to zero-level (NRZ-L)

1 = high level

0 = low level

Nonreturn to zero-mark (NRZ-M)

1 = transition at beginning of interval

0 = no transition

Nonreturn to zero-space (NRZ-S)

1 = no transition

0 = transition at beginning of interval

Return to zero (RZ)

1 = pulse in first half of bit interval

0 = no pulse

Biphase-level (Manchester)

1 = transition from high to low in middle or interval

0 = transition from low to high in middle of interval

Biphase-mark

Always a transition at beginning of interval

1 = transition in middle of interval

0 = no transition in middle of interval

Differential Manchester

1 = no transition at beginning of interval

0 = transition at beginning of interval

Always a transition in middle of interval

Delay modulation (Miller)

1 = transition in middle of interval

0 = no transition if followed by 1

Transition at end of interval if followed by 0

Bipolar

1 = pulse in first half of bit interval, alternating polarity from pulse to pulse

0 = no pulse

where $\{b_m\}$ is the precoded sequence, assumed to be in the set (0,1,2,...,M-1).

A summary of the definitions of the line-coding formats discussed earlier, as well as some others, is given in Table 2.4. A pictorial representation of the effect of each of these formats on a particular logical sequence is shown in Fig.2.6a. As was mentioned earlier, one of the objectives of line coding is spectral shaping. As an illustration of the way in which line coding can have such an effect, Figure 2.6b shows the power spectral density of several of these formats.

2.5 RF MODULATION

Modulation is one of the most important signal processing operations that takes place in a communication system.It can be effectively used to match a signal with channel characteristics, to minimize the effects of channel noise, and to provide the capability to multiplex many signals. Perhaps the most important function is to transmit a low-pass signal X(t) over a bandpass channel centered at f_c. For example, we can translate the spectrum of X(t) by multiplying X(t) with a carrier (C(t) of the form.

$$C(t) = A \cos (2\pi f_c t + \theta) \qquad (2.5.1)$$

where A is the amplitude of the carrier, f_c is the carrier frequency, and θ is an arbitrary phase constant assumed to be a random variable uniformly distributed in $[-\pi, \pi]$. The product (or modulated) signal Y(t)

$$Y(t) = C(t)\, X(t) = AX(t) \cos (2\pi f_c t + \theta) \qquad (2.5.2)$$

has a power spectral density

$$S_{YY}(f) = \frac{A^2}{4}[S_{xx}(f - f_c) + (S_{xx}(f + f_c)] \quad (2.5.3)$$

Equation (2.5.3) shows that multiplying X(t) by a carrier C(t) translates the spectrum of X(t) by the carrier frequency. Thus the low-pass signal X(t) is transformed to a bandpass signal suitable for transmission over a bandpass channel. All methods of modulating a carrier can be shown to have a similar effect in the sense of centering the modulated spectrum around the carrier frequency, but for nonlinear modulation methods the modulated spectrum is not merely a translation of the baseband spectrum.

2.5.1 Analog Modulation

In analog modulation, the modulating signal is one whose amplitude can take a continuum of values, as in speech or images. The signal

$$Y(t) = AX(t) \cos (2\pi f_c t + \theta) \quad (2.5.4)$$

is said to be amplitude modulated since X(t) modifies the amplitude of C(t).

In a similar fashion, we can modulate th ephase or frequency of C(t) according to

$$Z(t) = A \cos [2\pi f_c t + \theta + k_p X(t)] \quad (2.5.5)$$

or

$$W(t) = A\cos\left[2\pi f_c t + k_f 2\pi \int X(t)dt + \theta\right] \quad (2.5.6)$$

where the instantaneous frequency deviation is given by

$$f_i(t) = k_f X(t) \qquad (2.5.7)$$

and k_p and k_f are constants. [When X(t) is normalized so that $|X(t)| \leq 1$, k_p, and k_f are referred to as the phase deviation and frequency deviation, respectively; and the ratio of the peak frequency deviation to highest frequency in the modulating signal is called the modulation index.] Z(t) is called a phase modulated carrier and W(t) is called a frequency modulated carrier. The simulator should bear in mind that when the modulation index is large, the bandwidth of W(t) can be many times larger than that of X(t). Therefore, the simulation sample rate should be chosen accordingly.

In general, a modulated carrier can be represented in quadrature form as

$$Y(t) = X_1(t)\cos(2\pi f_c t + \theta) - X_2(t)\sin(2\pi f_c t + \theta) \qquad (2.5.8)$$

where $X_1(t)$ and $X_2(t)$ are low-pass processes with bandwidth B, and the carrier frequency f_c is typically > > B. $X_1(t)$ and X_2 (t) could be two independent analog signals or they may be uniquely related to a common signal X(t) as in the case of some modulation schemes such as vestigial side band modulation or single sideband modulation. In the case of (2.5.5), for example, $X_1(t) = a\cos[k_p X(t)]$ and $X_2(t) = a\sin[k_p X(t)]$. Table 2.5 lists $X_1(t)$ and $X_2(t)$ for a number of analog modulation schemes.

For simulation purposes, we represent modulated waveforms in the complex envelop from

$$\tilde{Y}(t) = [X_1(t) + jX_2(t)]e^{j\theta} \qquad (2.5.9)$$

Table 2.5 Quadrature Representation of Analog Modulation Schemes.

Carrier Phase Offset $\theta = 0$

Modulation scheme	$X_1(t)$	$X_2(t)$	Comments
Amplitude modulation(AM)	$a[1+k_a X(t)]$	0	K_a modulation index
Quadrature AM	$X_1(t)$	$X_2(t)$	
Double sideband (DSB)	$X(t)$	0	
Single sideband X(t)	$X(t)$	$\dot{X}(t)$-Hilbert transform of X(t)	
Phase modulation	$\cos[k_p X(t)]$	$\sin[k_p X(t)]$	k_p- modulation sensitivity
Frequency modulation	$\cos[k_f \int X(\alpha)d\alpha]$	$\sin[k_f \int X(\alpha)d\alpha]$	k_f- modulation sensitivity

The random phase offset of the carrier, θ, is often assumed to be zero. The complex envelop $\tilde{Y}(t)$ can also be expressed as

$$\tilde{Y}(t) = R(t)\, e^{j\psi(t)} \qquad (2.5.10a)$$

where that real envelope R(t) is

$$R(t) = [X_1^2(t) + X_2^2(t)]^{1/2} \qquad (2.5.10b)$$

and

$$\psi(t) = \theta + \tan^{-1}\frac{X_2(t)}{X_1(t)} \qquad (2.5.10c)$$

Sampled values of $\tilde{Y}(t)$ are used in simulations. While we do not sample the bandpass modulated signal, effects such as frequency offsets and frequency selective behavior can be modeled using a low-pass equivalent.

2.5.2 Digital Quadrature Modulation

An important class of digital modulation schemes can be naturally represented in the form (2.5.8) or (2.5.9), where $X_1(t)$ an $X_2(t)$ might be two different pulse waveforms of the form

$$X_1(t) = \sum_k A_k p_1(t - kT_1 - D_1) \tag{2.5.11a}$$

and

$$X_2(t) = \sum_k B_k p_2(t - kT_2 - D_2) \tag{2.5.11b}$$

where $p_1(t)$ and $P_2(t)$ are finite energy pulses (rectangular, filtered rectangular or sincϕ, for example). A_k and B_k are sequences of discrete random variables with symbol rates of $1/T_1$ and $1/T_2$, respectively, and D_1 and D_2 are possible delays. If $X_1(t)$ and $X_2(t)$ are unsynchronized, a proper model for D_1 and D_2 is to assume they are random and uniformly distributed on the intervals $[0, T_1]$ and $[0, T_2]$, respectively. For simulation purposes, an equivalent result is obtained if we set D_1 = 0, say, and let D_2 be random. In many important cases, $T_1 = T_2$ and D_1 and D_2 have a fixed, nonrandom relationship. Many digital schemes of practical interest are subsumed under the representation (2.5.11) in particular, QPSK, OQPSK, UQPSK, M-QAM, and MSK, among others (see Table 2.6). Particular examples are sketched in Figure 2.7.

Table 2.6 Some Digital Modulation Scheme Defined by Equation (2.5.11)

$T_1 = T_2 = T$, $D_1 = D_2 = 0$, and $\theta = 0$

Modulation Scheme	(A_k, B_k)	$p_1(t)$, $p_2(t)$
Amplitude shift keying (M-ray ASK)	$A_k = \pm nd$, $n = 1,2\ldots, M/2$ $B_k = 0$	$p_1(t) = 1, 0 \leq t \leq T$ $p_2(t) = 0$
Phase shift keying (M-PSK)	$A_k + B_k = e^{j\phi_k}$ $\phi_k = 2\pi n/M, n = 0,1\ldots,M-1$	$p_1(t) = 1, 0 \leq t \leq T$ $p_2(t) = p_1(t)$
QPSK (M-PSK, M = 4)	$(A_k B_k) = (\pm 1, \pm 1)$ or $\phi_k = 45°, 135°, 225°, 315°$	$p_1(t)$, $p_2(t)$ same as M-PSK
O-QPSK	$(A_k B_k) = (\pm 1, \pm 1)$	$p_1(t) = 1, 0 \leq t \leq T$ $p_2(t) = 1, ½ T \leq t \leq 3/2\, T$
Minimum shift keying	$(A_k, B_k) = (\pm 1, \pm 1)$	$p_1(t) = \sin(\pi t/2T - T/2)$ $0 \leq t \leq T$
(MSK)		$p_2(t) = \sin(\pi t/2T - T/2)$ $½ T \leq t \leq 3/2\, T$
M-ary Quadrature amplitude modulation (M-QAM)	$(A_k, B_k) \epsilon (\pm 1, \pm 3, \ldots, \pm \sqrt{M} - 1)$	$p_1(t)$, $p_2(t)$ same as M-PSK

2.5.3 Continuous Phase Modulation (CPM) : CPFSK; MSK

An important class of digital modulation schemes, not readily interpretable in the form (2.5.8) is continuous phase modulation, or CPM. Both superior performance and efficient spectrum utilization are available in principle by proper choice of the parameters. An important subclass of CPM is continuous-phase frequency-shift-keying, CPFSK, and a popular special case of the latter is minimum-shift-keying (MSK).

2.5.3.1 Continuous Phase Modulation

CPM is a form of digital phase modulation where that phase is constrained to remain continuous; that is, the phase cannot jump discontinuously between symbols, as it can (in principle) in QPSK (see, for example Ref. 9). A general CPM signal is given by

$$Y(t) = A\cos[\omega_c t + \phi(t) + \phi_0] \qquad (2.5.12a)$$

$$\tilde{Y}(t) = A\exp[j\phi(t)+j\phi_0] \qquad (2.5.12b)$$

$$\phi(t) = 2\pi \int_{-\infty}^{t} \sum_{k=-\infty}^{n} d_k h_k \alpha(\tau - kT) d\tau \qquad (2.5.13a)$$

$$= 2\pi \sum_{k=-\infty}^{n} d_k h_k q(t - kT)\,, \quad nT \le t \le (n+1)T \qquad (2.5.13b)$$

where $\alpha(t)$ is a 'frequency pulse", and

$$q(t) = \int_0^t \alpha(\tau)\, dt$$

The constraint imposed in (2.5.13) established the continuity of the phase. The parameter T is symbol duration; $\{d_k\}$ is the data sequence, where $d_k \in \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$; and h_k is called the modulation index. Commonly, we set $h_k = h$, a fixed value. In some instances, h_k varies k in a cyclic manner; this situation is referred to as multi-h CPM. In practice, $\alpha(t)$ is finite in extent

$$\alpha(t) = 0, \quad t<0, \quad t>LT$$

and the normalization

$$\int_{0}^{LT} \alpha(\tau)\, d_T = \frac{1}{2}$$

is used. The following terminology is common when L = 1, full response CPM; and when L ≥ 2, partial response CPM.

For L finite, we have

$$\phi(t) = 2\pi \sum_{k=n-L+1}^{n} d_k h_k q(t-kT)$$

$$+\pi \sum_{k=-\infty}^{n-L} d_k h_k, \quad nT \le t \le (n+1)T \tag{2.5.14}$$

2.5.3.2 Continuous-Phase Frequency-Shift-Keying; CPFSK

When the instantaneous frequency in each signaling interval is fixed and chosen from a set of M values, we have the subclass called continuous phase frequency-shift-keying[10] or CPFSK. To obtain the desired condition, we set

$$\alpha(t) = \frac{1}{2T} p_T(t) \tag{2.5.15}$$

so that

$$\phi(t) = \frac{1}{T}\pi h d_n(t-nT) + \pi h \sum_{k=-\infty}^{n-1} d_k, \quad nT \le t \le (n+1)T \tag{2.5.16}$$

It is customary to set $h_k = h$, a fixed value for all k, although that is not strictly necessary to create an FSK signal. From (2.5.16) we see that the instantaneous frequency is given by

$$f_i(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} = \frac{hd_n}{2T}, nT \leq t \leq (n+1)T$$

Notice that the increment of frequency shifts is $f_d = h/2T$. Since h represents the ratio of the minimum separation $2f_d$ to the symbol rate, it is also referred to as the deviation ratio. It may also be noted that the second term (2.5.16)

$$\psi_n = \pi h \sum_{k=-\infty}^{n-1} d_k$$

can be easily implemented recursively since $\psi_n + 1 = \psi_n + \pi h d_n$.

2.5.3.3 Minimum-Shift-Keying: MSK

A form of CPFSK, known as MSK, which has recently found wide application, is that for which M = 2 and h = 0.5. Under this specialization the phase function becomes.

$$\phi(t) = \frac{\pi}{2T} d_n(t - nT) + \frac{\pi}{2} \sum_{k=-\infty}^{n-1} d_k, \quad nT \leq t \leq (n+1)T \qquad (2.5.17)$$

It turns out for this case that expanding (2.5.12a) in quadrature form does lead to a useful expression, which in turn gives rise to a practical alternative implementation. Specifically, substituting (2.5.17) into (2.5.12b) yields

$$\tilde{Y}(t) = [A \cos u_n \cos(\pi t/2T)$$

$$+ jd_n \cos u_n \sin(\pi t/2T)]\exp(j\phi_o), \quad nT \le t \le (n+1)T \qquad (2.5.18)$$

where

$$u_n = \frac{\pi}{2}\left\{\sum_{k=-\infty}^{n-1} dk - nd_n\right\}$$

Upon closer examination, it can be shown that the "equivalent" I and Q data streams, cos u_n and d_n cos u_n are such that each (I or Q) data symbol has duration 2T, and the data streams are offset from one another by T seconds.

The (2.5.18) can be meaningfully implemented in quadrature form. However, this particular form has some drawbacks, as it stands. First, the equivalent data streams have to be computed, and second the "subcarriers" cos (πt/2T) and sin (πt/2T) reverse the sign of every other of these bits. Further reflection shows that an instantaneously different, but statistically equivalent, form of (2.5.18) results in a more conveniently implemented version. This form is obtained simply by replacing cos u_n and d_n cos u_n by d_{2n} an d_{2n-1}, respectively. In other words, demultiplex the input bit stream into "even" and "odd" bits, stretch each bit to length 2T, and multiply by the subcarriers. The latter are aligned with their respective bit streams so that each bit pulse is a half-sinusoid. The alternate bit inversion caused by the subcarriers makes detection somewhat awarded in simulation, but this bit signs unchanged. A block diagram of a possible implementation would thus appear as in Figure 2.7. In this block diagram, another variation of (2.5.18) is implicit. We have previously conceived of the I and Q data streams as demultiplexed versions of a single high rate bit stream. It is equally valid to suppose that we initially have two independent bit streams, as is done in Figure 2.8. The resulting MSK signal would not be pointwise identical to one in which the subcarrier is not rectified, but in it would be statistically identical. Note further that, in the rectified form, the MSK signal can be represented precisely in the form (2.5.8) with the quadrature components given by (2.5.11), where $p_1(t) = p_2(t) = \sin(\pi t/2T)$, $T_1 = T_2 = 2T$, $D_1 = 0$, and $D_2 = T$.

2.6 DEMODULATION

Demodulation is the complementary or inverse operation of modulation and is performed in the receiver. However, the demodulation method is not unique in the sense that, given a specific modulated signal, it may be possible to demodulate it using more than one distinct implementation. Nevertheless, nearly all demodulators may be put into one of two classes; coherent demodulate it using more than one distinct implementation. Nevertheless, nearly all demodulators may be put into one of two calsses; coherent demodulators, which use a locally generated carrier signal for demodulation, and noncohernt demodulators, which do not require a local carrier.

2.6.1 Coherent Demodulation

Suppose that we can generate a "local carrier" (or local oscillator) signal in the receiver of the form

$$\hat{C}(t) = 2\cos(2\pi\hat{f}_c t + \hat{\theta})$$

where $\hat{f}_c$ an $\hat{\theta}$ are estimates, respectively, of the carrier frequency and phase generated by the "carrier recovery" mechanism in the receiver. Then, for the general quadrature representation (2.5.8), we can form estimates of $X_1(t)$ and $X_2(t)$ using the "generic" quadrature demodulator arrangement shown in Figure 2.9. The predetection filter in the figure limits the amount of noise but ideally does not distort the signal. The post detection (LPF) filters shown condition the single (such a matched filtering) for further processing; this will be discussed shortly. It is easy to show that

$$\hat{X}_1(t) = X_1(t)\cos 2\pi[(f_c - \hat{f}_c)\,t + (\theta - \hat{\theta})] \qquad (2.6.1a)$$

$$-X_2(t)\sin 2\pi[(f_c - \hat{f}_c)\,t + (\theta - \hat{\theta})]$$

$$\hat{X}_2(t) = X_2(t)\cos[2\pi(f_c - \hat{f}_c)t + (\theta - \hat{\theta})] \qquad (2.6.1b)$$

$$+X_1(t)\sin[2\pi(f_c - \hat{f}_c)t + (\theta - \hat{\theta})]$$

If $\hat{f}_c = f_c$ and $\hat{\theta} = \theta$ i.e., if the carrier recovery mechanism provides error free estimates of f_c and θ, then $\hat{X}_1(t) = (X_1(t)$ and $\hat{X}_2(t) = X_2(t)$ (assuming no noise). As was the case with modulators, we can incorporate possible imperfections in the hardware. For example, we can visualize that $\hat{X}_1(t)$ (the I-channel), is recovered by using the local oscillator (LO) 2 cos $(2\pi\hat{f}_c t + \hat{\theta})$ and ~~$\hat{X}_2(t)$ (the Q-channel), is obtained by using the LO 2a cos $(2\pi\hat{f}_c t - \hat{\theta})$ and~~ $\hat{X}_2(t)$ (the Q-channel), is obtained by using the LO

$2a\cos(2\pi\hat{f}_c t - \pi/2 - \alpha)$ where a and α represent departures from ideal. The associated modifications to (2.6.1) are straightforward.

In practical systems, the input to the receiver is a distorted and noisy version of Y(t). An additive model of the form

$$\hat{Z}(t) = \tilde{W}(t) + \tilde{N}(t) \tag{2.6.2}$$

where $\tilde{W}(t)$ is the complex envelop of the distorted signal component and $\tilde{N}(t)$ is the complex envelope of the noise, is used to represent the input to the demodulator.

In terms of the compiled envelope representation, the demodulator operation can be represented as

$$\hat{X}_1(t) + j\hat{X}_2(t) = \{W_R(t) + n_s(t) + j[W_1(t) + n_s(t)]\} \tag{2.6.3}$$

$$\times \exp[j2\pi(f_c - \hat{f}_c)t + j(\theta - \hat{\theta})]$$

where $\tilde{W}(t) = W_R(t) + jW_I(t)$ and $\tilde{N}(t) = n_c(t) + jn_s(t)$.. Sampled values of the preceding equation is what is implemented in simulating a coherent demodulator. The demodulated signals $\hat{X}_1(t)$ and $\hat{X}_2(t)$ are often referred to a baseband signals.

If the demodulated signals $\hat{X}_1(t)$ and $\hat{X}_2(t)$ are digital signals, they must now be processed in some fashion so as to recover the original sequences $\{A_k\}$ and $\{B_k\}$ with as few errors as possible. We assume the original digital modulating signals $X_1(t)$ and $X_2(t)$ are given by (2.5.11) but for simplicity in the present context we assume $T_1 = T_2 = T$ and $D_1 = D_2 = 0$. Typically, two operations are performed on the demodulated signal, a filtering operation usually referred to as "matched filtering", followed by sampling (observing) the signal. (We are using the term "matched filter" loosely here; strictly speaking, a filter that is "matched" to a pulse has a transfer function that is the complex conjugate of the pulse spectrum.However, this usage is also widespread for a filter that attempts to do the matching.) These operations are sometimes collectively referred to as detection. The sampled signal values from the basis for making decisions as to what symbol was sent. In order to illustrate the points we wish to make, let us assume first that $\hat{f}_c = f_c$ and $\hat{\theta} = \theta$. Let us assume also that the distortion arises from a channel with impulse response $h_c(t)$, and also denote by $h_m(t)$ the matched filter impulse response. Then the waveforms at the input to the sampler can be written as

$$S_1(t) = \sum_{k=-\infty}^{\infty} A_k p(t - kT - D) * [h_c(t) * h_m(t)] + n_1(t) \qquad (2.6.4a)$$

$$S_2(t) = \sum_{k=-\infty}^{\infty} B_k p(t-kT-D) * [h_c(t) * h_m(t)] + n_2(t) \qquad (2.6.4.b)$$

where D is a fixed but unknown transmission delay, * represents convolution, and $n_1(t) = n_c(t)*h_m(t), = n_2(t) = n_s(t)*h_m(t)$. In order to recover $\{A_k\}$ and $\{B_k\}$ the waveform (2.6.4) must be sampled at the appropriate instants. Thus, the sampling operation takes a place at the instants.

$$t_k = K\,T_b + \hat{D} \qquad (2.6.5)$$

where $\hat{D}$ is an estimate of the "system" delay. The estimate $\hat{D}$ is provided by the "timing recovery" system in the receiver (Figure 2.10). In the simulation context there is normally no propagation delay, and the effective delay due to filtering can be estimated fairly accurately beforehand. Thus, an adequate "guess" at $\hat{D}$ can often be made without explicitly simulating timing recovery circuitry.

In Section 2.5 we pointed out that the generic quadrature modulator might be topologically altered in particular instances, to reflect properties specific to one or another modulation scheme belonging to the general class. This idea applies even more so in the case of demodulators, since there can be many variations of demodulator architectures, even for the same modulation scheme. As an example of this point. Figure 2.11 shows a modification of the topology of the generic quadrature receiver that applies specifically to the synchronization functions for an MSK signal. This architecture takes advantage of the properties peculiar to an MSK signal, in particular the fact that such a signal can be viewed either as a QAM signal, of the form (2.5.8), or as a binary FSK signal with $h = 0.5$. Viewed in the latter form, it can be shown that squaring the MSK signal generates two spectral lines, at $2f_c \pm (1/2T)$. Where 2T is the bit duration in the I or Q channel. Because of this particular relationship, the manipulations shown in Figure 2.10 result directly is separate I and Q channel carrier or "subcarrier" references, as well as clock (timing) recovery. The main implication for simulation is that different structures are potentially subject to different kinds of impairments. Hence, it is important to study the intended implementation, if known, or perhaps study several to develop a sense of their sensitivity to imperfections. In early stages of system design, however, a generic receiver may be quite adequate to study the trades between different types of modulation schemes.

2.6.2 Noncoherent Demodulation

The term noncoherent demodulation implies that the demodulator does not require a coherent local carrier to perform its function. Coherence in this context refers to knowledge of the carrier phase. Although phase is not needed, approximate knowledge of carrier frequency is needed in order to properly place the receiver passband around the signal spectrum. Note that some modulation schemes, such as FSK, may be demodulated by coherent or noncoherent structures.

2.6.2.1 Amplitude Demodulation

Consider the case when Y(t) is an amplitude modulated signal of the form.

$$Y(t) = [1+kX(t)]\cos(2\pi f_c t+\theta) \qquad (2.6.6)$$

If $|kX(t)| \leq 1$ then by observing the real envelop of Y(t) which is [1 = kX (t)] we can extract the modulating signal X(t), That is, since

$|1+kX(t)| = 1+kX(t)$ when $|kX(t)| < 1$ except for the dc offset of 1 and scale factor, k, the envelope contains X(t) intact. A demodulation operation that retrieves the envelope is called envelope detection. Because of noise or distortion. $\tilde{Y}(t)$ is generally complex, hence, the envelope detection operation simply involves taking the absolute value of the complex envelope.

Another noncoherent amplitude demodulation method involves taking the square of the magnitude of the complex envelope. This method is called square law demodulation. If the input signal is given by (2.6.6) square law demodulation takes the form. $[1+kX(t)]^2 = 1+k^2X^2(t)+2kX(t) \approx 1+2kX(t)$ for $|X(t)| \ll 1$

Thus, this method requires a small modulation index. In terms of the complex envelope, both of the above noncoherent methods can be expressed as

$$\hat{X}(t)=\begin{cases} |\tilde{Z}(t)| & \text{envelope demodulation} \\ |\tilde{Z}(t)|^2 & \text{square law demodulation} \end{cases} \qquad (2.6.7a\&b)$$

where $\tilde{Z}(t)$ is the complex envelope at the input to the demodulator and $\hat{X}(t)$ is the real output of the demodulator.

2.6.2.2 Discriminator Detection of PM/FM Signals

Frequency modulated and phase modulated signals can be demodulated using a variety of schemes. The classical scheme for FM signals is termed discriminator detection. (Since phase is the integral of frequency, a PM demodulator is a discriminator followed by an integrating filter). In principle, the operation is intended to be precisely the inverse of FM modulation. Looking at the FM signal (2.5.6) we see that the modulating 2.5.6 signal is recovered by differentiating the phase. Thus, as in (2.5.10a) if.

$$\tilde{W}(t) = R(t)\, e^{j\psi(t)}$$

is the complex envelope of the modulating signal, we see that

$$X(t) = \frac{d}{dt}\,\psi(t) \tag{2.6.8}$$

The actual received signal is a distorted and noisey version, say

$$\tilde{W}_0(t) = R_0(t)\, e^{j\psi_0(t)} \tag{2.6.9}$$

so that, in the simulation context, ideal discriminator detection can be implemented by differentiating the argument of the complex envelope, viz.,

$$\hat{X}(t) = \frac{d}{dt}\,\psi_0(t) \tag{2.6.10}$$

Although 2.6.10 is simple in appearance, differentiation is not simple to simulate, especially if one wishes to study operation below threshold, where FM clicks are significant contributors to noise [12].

2.6.2.3 PLL Demodulation of PM/FM Signals

Probably the most widely used demodulator for PM and FM signals is the phase-locked loop or PLL, which when used for this purpose, is referred to as a PLL demodulator; see Figure 2.12. Strictly speaking, PLL demodulation is a coherent process, or at least partially coherent. but in this case the demodulation method is such that no advantage can be taken of coherence, in terms of the SNR above threshold, although the threshold is lowered. The PLL demodulator relies on a device called a voltage controlled oscillator (VCO) whose output is

$$V(t) = \sqrt{2}\,K_1 \cos\left[2\pi f_c t + k_2 \int Z(\alpha)\, d\alpha\right] \qquad (2.6.11)$$

K_1 and K_2 are gain constants, f_c is the quiescent operating frequency (assumed equal to the unmodulated input carrier frequency), and Z(t) is the "control" voltage. It can be easily verified that the feedback loop generates an error signal

$$\epsilon(t) = AK_1K_m \sin[\phi(t) - \hat{\phi}(t)]$$

If the loop filter and the constants are chosen carefully when the lop can be made to "lock" onto, or "track", the input phase, i.e., the loop established the conditions.

$$\hat{\delta}(t) \approx \phi(t),\ \epsilon(t) \approx 0$$

Since the error signal $\epsilon(t)$ is nonlinear in $\phi(t) - \hat{\phi}(t)$, the loop is nonlinear and the analysis of the analysis of the nonlinear loop is difficult; simulation, although more straightforward, can still be quite tricky. A linear assumption of the form

$$\sin[\phi(t) - \hat{\phi}(t)] \approx \phi(t) - \hat{\phi}(t) \qquad (2.6.12)$$

is customarily made to simplify the analysis. With this assumption the PLL is an ordinary feedback loop whose response can be analysed using the closed loop transfer function. The nonlinear operation of the PLL can be simulated using the low-pass equivalent model shown in Figure 2.13

The PLL is a device of fundamental importance in communication systems. We had already mentioned that the PLL is the basic structure for carrier recovery as well as for timing recovery in digital systems, and we see here its application in angle demodulation. In order to properly simulate a PLL using the sampled (discrete-time) version of the complex low-pass equivalent model shown in Figure 2.13. It is necessary to understand the relationship between this model and the actual (continuous structure), which is described by a nonlinear differential equation. For now it suffices to point out that when a PLL is simulated using the complex low-pass equivalent model shown in Figure 2.13 special attention must be paid to the following:

1. The sampling rate must be much larger than the loop bandwidth in order to accurately simulate the nonlinear behavior.
2. FFT-type block processing operations cannot be used inthe loop since such processing introduces a delay which may make the loop unstable.

2.7 FILTERING

The primary applications of filtering in communication systems are to select desired signals, minimize the effects of noise and interference, modify the spectra of signals, and shape the time domain properties (zero crossings and pulse shapes) of digital waveforms.

An example of the first application, namely, the selection of a desired signal, occurs in radio and TV receivers when we "tune" to pick up one of many stations that are broadcasting simultaneously. Receivers also use filters to reject "out-of-band" noise. Transmitters, on the other hand, have to meet regulatory constraints on the shape of the transmitted spectra or on "out-of-band"power, and filtering is used for controlling these.

Time domain properties such as Zero-crossings and pulse shapes are important in digital communication systems. These properties can also be controlled via appropriate pulse-shaping filters.

In the following sections, we discuss the form of filter transfer functions for various types of applications in communication systems.

Filters used in communications system are sometimes required to be adaptive. These filters are required to change their response with the properties of the input signals change. For example, a filter designed to remove signal distortion introduced by a channel should change its response as the channel characteristics change. The most commonly used adaptive filter structure (also commonly referred to as an equalizer) is the adaptive tapped delay line (TDL). Details of the adaptive TDL are also presented in this section.

All the filters described in this section are linear and time invariant except for the adaptive TDL filter, which is linear and time varying.

2.7.1 Filters for Spectral Shaping

It can be that the power spectral density $S_{YY}(f)$ of the output Y(t) of a linear time -invariant system is

$$S_{YY}(f) = S_{xx}(f)\,|H(f)|^2 \qquad (2.7.1)$$

where $S_{xx}(f)$ is the PSD of the input signal X(t) an H(f) is the transfer function of the system. By carefully selecting H(f) we can emphasize or emphasize selected spectral components of the input signals. This frequency selective filtering operation is used extensively incommunications systems. If $S_{xx}(f)$ and H(f) are given, calculation of $S_{YY}(f)$ is straightforward. On the hand, suppose that $S_{YY}(f)$ is a desired output PSD given that $S_{xx}(f)$ is an input PSD. Equation (2.7.1) can be rewritten as

$$|H(f)|^2 = \frac{S_{YY}(f)}{S_{xx}(f)} \tag{2.7.2}$$

and the problem here is to synthesize a realizable filter H(f) that satisfied (2.7.2). This can be done by factoring the right-hand side of Equation (2.7.2) and including only poles and zeros in the left-half of the s-plane for specifying H(s). (The transformation $s = j2\pi f$ is applied before factoring). This is called spectral factorization. Of course, it is implied that the right-side of (2.72) is a ratio of polynomials in s. If this is not so, which is frequently the case, then one must first develop an approximation that is a ratio of polynomials.[13]

2.7.2 Filters for Pulse Shaping

The individual pulses in the waeforms used in digital transmission systems are typically required to satisfy two important conditions; band-width and zero-crossings. if the embedded digital sequence has a symbol rate of R_b, then the bandwidth of the digital waveform should be of the order of R_b and it might b required to have zero crossings in the time domainonce every T_b seconds where $T_b = 1/R_b$ Nyquist has shown that a time domain pulse p(t) will have zero crossings once every T_b seconds if its transform P(f) meets the following constraint:

$$\sum_{k=-\infty}^{\infty} P(f+kR_b) = T_b \quad \text{for } |f| < R_b/2 \tag{2.7.3}$$

If p(f) satisfied equation (2.7.3) then p(t) has the following properties:

$$p(0) = 1$$

$$p(kT_b) = 0, \quad k = \pm 1, \pm 2 \tag{2.7.4}$$

Zero crossing requirement are imposed to provide zero intersymbol interference (ISI) and assist in the timing recovery process.

A family of P(f) which meets the Nyquist criterion given in Equation (4.8.3) is the raised cosine (or cosine roll-off) family with

$$P(f) = \begin{cases} T_b, & |f| \le R_b/2-\beta \\ T_b\cos^2\frac{\pi}{4\beta}\left(|f|-\frac{R_b}{2}+\beta\right), & \frac{R_b}{2}-\beta < |f| \le \frac{R_b}{2}+\beta \\ 0, & |f| > \frac{R_b}{2}+\beta \end{cases} \tag{2.7.5}$$

where β is the "excess bandwidth" parameter. Note that the raised cosine P(f) is band-limited to $\beta + (R_b/2)$ and hence it can meet selected bandwidth constraints. (See Figure 2.14a).

The raised cosine family has an impulse response of the form

$$p(t) = \frac{\cos 2\pi\beta t}{1-(4\beta t)^2}\left(\frac{\sin R_b t}{\pi R_b t}\right) \tag{2.7.6}$$

(see Figure 2.14 b) and it has zero crossings once every T_b seconds.

The raised cosine family has an impulse then, in order to produce an output in the raised cosine family, the filter transfer function should be $H(f) = P(f)/G(f)$, where $G(f)$ is the Fourier transform of $g(t)$.

In most applications the filtering operation to produce $P(f)$ is split between two filters, one at the transmitter, with transfer function denoted $H_T(f)$, and one at the receivr, with transfer function denoted $H_R(f)$.

With an impulse at the input to $H_T(f)$ and an ideal channel, we still must have $H_T(f)H_R(f) = P(f)$. In the relatively simple case where we have an additive noise source at the receiver input, it can be shown that the optimum partition, in the sense of optimizing the signal-to-noise ratio at the receiver output, is to split $P(f)$ equally between the transmitter and the receiver, i.e.,

$$H_T(f) = H_R(f) = H_R(f) = [P(f)]^{1/2} \qquad (2.7.7)$$

2.7.3 Filters for Minimizing Noise and Distortion

If we, can-use two filters in the system, a transmit (preemphasis) filter $H_T(f)$ and a receive (deemphasis) filter $H_R(f)$ as shown in Figure 2.15, then we can choose their transfer functions such that $X_0(t) = kX(t - t_d)$ (where K is a constnat and t_d is an arbitrary delay) and $E[N_0^2(t)]$ is minimized. That is, we can minimize the noise power at the output of the receive filter subject to the constraint of no signal distortion. With two filters we can meet the dual criteria of distortionless transmission and minimum noise power (or maximum signal-to-noise ratio).

The transfers function of $H_T(f)$ and $H_R(f)$ can be shown to be equal to [14].

$$|H_R(f)|^2 = \left[\frac{S_{xx}(f)}{S_{NN}(f)}\right]^{1/2} \frac{1}{|H_c(f)|} \tag{2.7.8a}$$

$$|H_T(f)|^2 = \left[\frac{S_{NN}(f)}{S_{XX}(f)}\right]^{1/2} \frac{1}{|H_c(f)|} \tag{2.7.8b}$$

with $k = 1$ The phase response of each of the filters is chosen to be linear. Note that spectral factoriation is involved in obtaining $H_T(f)$ and $H_R(f)$ from $|H_T(f)|^2$ and $|H_R(f)|^2$.

The transmit filter given in (2.7.8a) amplified or emphasizes weaker spectral components of the signal that might be asked by strong spectral components of the noise. The receive filter deemphasizes the same components inversely so that the net result preserves the spectral properties of the input signal X(t). This arrangement of preemphasis/deemphasis filters is used in FM broadcast and audio recording.

In digital transmission systems, the transmit and receive filters used to shape an input pulse waveform

$$X_T(t) = \sum_{k=-\infty}^{\infty} a_k p_T(t - kT_b)$$

into an output waveform

$$X_0(t) = \sum_{k=-\infty}^{\infty} A_k p_R(t - kT_b), \quad A_k = K_c a_k$$

while maximizing $E(A_k^2/E[N_0^2(t)]$, have the following transfer function [14]

$$|H_R(f)|^2 = \frac{K|P_R(f)|}{|H_c(f)|[S_{NN}(f)]^{1/2}} \tag{2.7.9a}$$

$$|H_T(f)|^2 = \frac{K_c^2|P_R(f)}{K|P_T(f)|^2}\frac{|[S_{NN}I(f)]^{1/2}}{|H_c(f)|} \tag{2.7.9b}$$

where K is an aribitrary positive constant. Figure 2.14 also applies topologically to this situation, but the interpretation and form of the transfer functions is evidnetly differnt for the case that minimizes the MSE.

If $p_T(t) = \delta(t)$ and $S_{NN}(f)$ is constant over the bandwidth of $P_R(f)$, then it can be shown that

$$|H_R(f)|^2 = K_1\frac{|P_R(f)|}{|H_c(f)|} \tag{2.7.10a}$$

$$|H_T(f)|^2 = K_2\frac{|P_R(f)|}{|H_c(f)|} \tag{2.7.10b}$$

If $H_c(f)$ is constant over the bandwidth of $P_R(f)$ then

$$|H_T(f)| = |H_R(f)| = K_3\,[P_R(f)]^{1/2}$$

which is precisely the "square-root filtering" discussed in section 2.7.2 . Of course, in order to produce the desirbale zero-crossing properties mentioned there, $P_R(f)$ must be a member of the raised cosine family (or satisfy the Nyquist criterion).

2.7.4 Matched Filters

Consider the problem of detecting the presence or absence of a pulse with a known shape p(t) by observing p(t) + N(t) when N(t) is zero mean, statonary additive noise.

Let us assume that p(t) is of duration T in the interval [0,T] and

$$Y(t) = \begin{cases} p(t) + N(t) \\ \text{or} \\ N(t) \end{cases}$$

We observe Y(t) for T seconds in the interval [0,T] and determine whether or not p(t) is present by processing Y(t). This is a fundamental problem in digital communication systems, radar, and sonar.

Suppose we base our decision about the presence or absence of p(t) based on the output of a filter at time t = T, that is we process Y(t) and obtain

$$Z = \int_0^T Y(\tau)h(T-\tau)dT$$

and base our decision on Z.

It can be shown that the filter that minimizes the average probability of incorrect decisions has the transfer function [14].

$$H(f) = K\frac{P*(f)\exp(j2\pi T)}{S_{NN}(f)} \qquad (2.7.11)$$

where K is a real, positive constant. It N(t) is white noise, then

$$H(f) = KP^*(f)\exp(j2\pi fT) \qquad (2.7.12)$$

By taking the inverse transform of H(f), it can be shown that the impulse response of the filter is

$$h(t) = P(T - t) \qquad (2.7.13)$$

so that

$$Z = \int_0^T Y(\tau)h(T - \tau)\, d\tau = \int_0^T Y(\tau)p(\tau)d\tau \qquad (2.7.14)$$

Equation (2.7.13) and (2.7.14) show that the impulse response of the filter is "matched" to p(t) and for this reason this filter is called the matched filter. Note that when p(t) is the unit rectangular pulse, p(t) = 1, for $0 \leq t \leq T$ and zero elsewhere, Z is just the integral of the input. If a decision has to be made in successive intervals (2.7.14) implies that this integration starts anew at the beginning of each such interval. Thus, the value of the integral at the end of the preceding interval must be discarded (or "dumped"). A mechanism of this process is the well-known integrate and -dump (I&D) filter.

The solution (2.7.14) has been cast in the context of detection but it obviously applies equally to decisioning for data transmission; the pulse detection problem is completely equivalent to on-off keying modulation and a little thought shows it is also equivalent to antipodal signaling.

All digital communication systems incldue a low-pass filter inthe receiver which is intended to perform the task of matched filtering. However, the matched filter must be matched to the received pulse, not the transmitted pulse. The former is inevitably a distorted version of the latter. Thus, even if a rectangulr pulse is sent, it will be received diffrently, and the I&D filter is no longer the matched filter. Furthermore, as we have seen (e.g., Figure 2.2), a digital signal is not always representable as a sequence of pulses with a single shape. Even if that were the case, ISI and nonlinearities would combine to produce the equivalent of multiple pulse shapes at the receiver. Thus, in practice, the idea of a matched filter to a single pulse shape is not realistic in many cases. if it were possible to actually determine a single received pulse, one could still, in principle, determine the matchd filter (2.7.13), but such a filter may not be easy to realize. Most distorted pulses can, in fact, be matched relatively closely by a low-order classical filter like a 2-pole or 3-pole Buttterworth or Chebyshev filter set an appropriate bandwidth. For simultion purposes, therefore, we recommended having these filters, as well as the classical I&D, available as choices for "matched" filtering when doing system studies.

Matched filtering can, in principle, compensate for linear distrotion in pulses when that distortion is knwon, but it cannot compensate for intersymbol interference. As we have seen, one could control ISI (ideally eliminate it) if it were possible to design all the filters so that the received pulse belongs to the raised cosine family. An examination of (2.7.9) shows that such filters may not be simple to synthesize especially since they depend on the "channel" transfer function $H_c(f)$. The latter can represent complex system, as well as the intervening medium; moreover, the system may be nonlnear, for which we cannot strictly speak of a transfer function.

Furthermore, the system and channel may not be well defined or may be time varying. This suggests looking for a self-adaptive solution to the problem discussed.

2.8 COMMUNICATION CHANNELS AND MODELS

In its most general sense the word "channel" can be used to mean everything between the source and the sink of a signal. Referring back to Figure 2.1, this general definition would include all of the equipment shown as well as the physical medium between the transmitter and the receiver through which the signal is radiated or conducted. In order to simulate such a channel when it is explicity represented by a sequence of blocks we need to have a model for each block (as well as for the medium), and in fact we do discuss models for all of these blocks in various parts. In this section, however, we discuss channel models, which needs to be distinguished from the general definition of a channel given above. Basically, a channel model may be thought of as a representation, in mathematical or algorithmic form, for the transfer characteristics of any continguous subset of the general block diagram. This representation is generally not based on the underlying physical phenomena, but rather on fitting external (empirical) observations. Inpractice, there are actually two quite distinct entities that are referred to as channel models. one of these is the physical medium only, which may be free space (as an idealiaton), the atmosphere, wires, waveguide, or optical fibers. Often, one implicitly includes in this definition of channel certain other physical conditions or geometrical constraints that have a strong bearing on the effective transfer function that the medium creates between transmitter and receiver. Some of these conditions or constraints include the carrier frequency, the bandwidth, and the physical environment. Depending upon the particular combination of controlling factors, the "atmospheric channel" maybe described in significantly different ways. Under some conditions, communication to or from satellites may be described as passing through an "almost free-space" channel. On the other hand, radio-relay or mobile systems are known to operate in an environment that produces multiple paths, hence is appropriately described as a "multipath" channel.

2.8.1 The Almost Free-Space Channel

Virtually any channel will be benign if the signal has sufficiently small bandwidth, and conversely almost any benign channel will exhibit nontrivial distrotion when the signal had sufficiently large bandwidth. Propagation in or through the atmosphere is an extremely complex phenomenon, [15], which can take on a wide range of behavior depending on circumstances. At one end of this range, the atmosphere has frequently been regarded as well approximated by free space, i.e., an ideal channel. This approximation is fairly good for satellite systems using large ground antennas operating in the 4-6 GHz range and with elevation angles that are not too small. However, as we increase the carrier frequency and increase the bandwidth correspondingly, a nonnegligible filtering effect begins to manifest itself. We will refer to such a case as the almost free-space channel. In this channel we consider only this effective filtering to exist: that is, we assume there is no multipath nor scintillation. Although this filtering characteristic is in reality time-variant, it is reasonable to treat it on a quasistatic basis because at the very wide bandwidths for which the filtering effect becomes significant the channel does vary very slowly with respect to the signal. The time variations can be separately modeled as phase noise whose spectrum is extremely narrow. We consider breifly three contributions to this filtering effect: (a) the clear-air atmosphere, in which significant filtering can occur around specific absorption "lines": (b) the rainy atmosphere, in which absorption is a function of frequency; and (c) phase distortion due to the ionsphere. In the rainy atmosphere we also briefly consider the depolarizing effect of rain.

2.8.1.1 Clear-Air Atmospheric (Tropospheric) Channel

In a clear-air atmosphere an electromagnetic wave interacts with the oxygen and water vapor that are present, in a way that depnds on the frequency of the wave. At certain frequencies there are resonances, resulting in peaks of absorption, (16) An example of atmospheric absorption curves is shown in Figure 2.16. Clearly, over a large enough bandwidth, the atmosphere acts as a filter. The possible effects of such a filter have been reported in Ref.[17]. In order to simulate this effect, one needs to know the transfer function H(f). If one has a set of measurements, it is strightforward to include them as a filter in a simulation, as discussed in Chapter 2. An analytical form is preferable, however, and for this purpose, one can use Liebe's model

$$H(f)=H_o\exp[j0.02096(10^6+N)L] \qquad (2.8.1)$$

where N is the complex refractivity in parts per million (a function of f), $N=N_o+D(f)+jN''(f)$;H_o is a constant determined from table look-up; N_o is the frequency-independent refractivity; D(f) is the refractive absorption; N"(f) is the absorption; and L is the distance in km.

The above parameters are dependent on frequency and on atmospheric conditions, namely, temperature, barometric pressure, and relative humidity. Values of these parameters are tabulated in Ref. (16) Liquid water in clouds or fog will also affect the absorption. Of course, the major liquid water effect occurs during rain, which we discuss next,

2.8.1.2 The Rainy-Atmospheric Channel

At microwave freqencies, say above 10 GHz, rain can become a dominant effect an atmospheric propagation. Certainly, at high enough frequencies and rain rate, rain attenuation is much more significant than atmospheric absorption, except possibly at the resonance lines. The effect of rain is well known [18] and is usually accounted for as attenuation which is independent of frequency across the bandwidth of the signal. However, as mentioned, for sufficiently wide bandwidth this attenuation will be frequency dependent, and hence effectively act as a filter. Figure 2.17 illustrates this point. As an example, we have computed rain attenuation as a function of frequency over the band from 17.7 to 21.2 GHz allocated to the space-to-earth path for satellite communications. At these frequencies rain attenuation is not only a major attenuating factor, but as can be seen, it exhibits a significant effetive gain slope across the band. The curves in Figure 2.17 have been computed using the Crane model [18] for three different elevation angles and rainfall corresponding to an outage robability of 0.2%. The receiving station is assumed at Washington, D.C. which has a relatively temperate climate. The effects would be more severe in rainier climates. one could interpret Figure 2.17 as a filter for use in simulation and simply input points from the curves.

Another effect of rain which can be significant in dual-polarized systems is the depolarization of radio waves. That is, some of the energy in each polarization is transformed into energy of the opposite (orthogonal) Polarization Let

$$S_1(t) = \rho_1(t) \cos[3\omega_1 t + \phi_1(t)] \tag{2.8.2a}$$

$$S_2(t) = \rho_2 (t) \cos[\omega_2 t + \phi_2(t)] \tag{2.8.2b}$$

represent signals on polariztion "1" and "2" respectively; these polarizations could be linear, say horizontal and vertical, or left hand and right hand circular; the carrier frequencies f_1 and f_2 may not be the same, but the spectra overlap. The simplest model for depolarization takes the form

$$R_1(t) = \alpha_{11} S_1(t) + \alpha_{21} S_2(t) \tag{2.8.3a}$$

$$R_2(t) = \alpha_{22}(t) S_2(t) + \alpha_{12}(t) * S_1(t) \tag{2.8.3b}$$

where $R_1(t)$ and $R_2(t)$ are the signals received on polarizations "1" and "2", respectively, and the coefficients represent the relative magnitude of each term. Basically, the cross-polarized leakage introduces interfernce into each signal. This intergerence is referred to as XPI. In signal "1", for example the XPI is 20 log

$(\alpha_{11}/\alpha_{21})$. Situations of this type are easily simulated, and there is also a good deal of literature on the effect of interference on the BER of digital signals, as well as on methods to compensate for XPI. However, simulation becomes increasingly simpler than analysis as the system departs further from the ideal. Equations (2.8.3) can be generalized by replacing each of the coefficients α_{ij} by a linear frequency dependent transfer characteristic. The result would be

$$R_1(t) = h_{11}(t) * S_1(t) + h_{21}(t) * S_2(t) \tag{2.8.4a}$$

$$R_2(t) = h_{22}(t) * S_2(t) + h_{12}(t) * S_1(t) \tag{2.8.4b}$$

which, again, is straight forward to simulate if we know the impulse responses $h_{ij}(t)$

2.8.1.3 The Ionospheric Phase Channel

In the lower-frequency bands, say of the order of a few hundred megahertz and below, the effect of the ionosphere on propagation is extremely complex and best characterized as a time-varying multipath channel. At frequencies above a few hundred megahertz, for example in the bands allocated to satellite communications, and assuming the absence of anomalous conditions (solar flares, nuclear events, extrmely low evaluation angles) the ionosphere can be approximately modeled by an all-pass filter with a nonideal phase characteristic.

It can be shown [19] that the phase shift expeienced by a wave of frequency f due to free electrons in the ionosphere, over and above the free-space propagationlag, is given by

$$\phi(f)=\frac{2\pi 40\times 10^6}{cf}\int_{s_1}^{s_2} N_e(s)ds \text{ (rad)} \qquad \text{(rad)} \tag{2.8.5}$$

where c is the speed of light (cm/s), N_e is the areal electron concentration (electrons/cm^2) at any point along the path, s, and the integral represents the integrated ("columnar") electron density along the signal path. The differential phase shift between any two frequencies f_0 and $f_0 + \Delta f$ is therefore given by

$$\psi(\Delta f)=\phi(f_o+\Delta f)-\phi(f_o) \tag{2.8.6a}$$

$$\frac{2\pi 40\times 10^6 \Delta f}{cf_o(f_o+\Delta f)}\int_{s_1}^{s_2} N_e(s)ds \tag{2.8.6b}$$

$$=k(f_0,N_e,s)\frac{\Delta f}{f_o+\Delta f} \tag{2.8.6c}$$

in which we have lumped into the constant K all the factors not depending on Δf.In fact, since we can consider Δf to be the departure from any given centre frequency f_o, it can be seen that (with $\Delta f \rightarrow f$)

$$\psi(f)=K\frac{f}{f_0+f} \tag{2.8.7}$$

is the complex low-pass equivalent filtering characteristic. We can simplify (2.8.7) still further by defining the normalized frequency $v = f/f_o$, so that

$$\psi(\nu) = K\frac{\nu}{1+\nu} \tag{2.8.8}$$

Obviously, the actual phase characteristic depends in scale on f^{-1}_o and on the integrated electron density, which can vary by two or three orders of magnitude depending upon path length (elevation angle), time of day,or solar activity. The frequency-dependent characteristic $\nu/(1+\nu)$ is very simple and can be entered directly as (in sampled form) for use in simulation; this function is plotted in Figure 2.18. However, it is even simpler to implement if we realize, as the figure suggests, that this function is dominated by parabolic phase. That is, if we expand $\Psi(\nu)$ in a power series, say, or a least-squares polynomial decomposition, then the ν^2 term is the major one (other than ν, which does not lead to distortion) for typical fractional bandwidth values, say $|v| \leq (B/f_0) < 0.1$. Because of the simple form of $\Psi(\nu)$ it is easy to demonstrate the point using a Taylor series about $\nu = 0$. Thus, one obtains

$$\Psi(\nu)/K = \nu/(1 + \nu) = \nu - \nu^2 + R_3 \qquad 2.8.9$$

where the remainder

$$R_3 \leq Mv^3/3!$$

and $M \geq \psi'''(u)/k$ where $|u| \leq B$. Since $\psi'''(u)/K = 6(1 + u)^4$, its maximum value is reached when $u = -B$, which we can take as -0.1. Thus, $R_3 = \nu^3/(0.9)^4 \approx 1.5v^3$, which means that at $\nu = -0.1$, the magnitude of the remainder, relative to the squared term, is $1.5v^3/v^2 = 1.5v = 0.15$. This means, basically, that the parabolic phase term v^2 dominates over any reasonable fractional bandwidth.Hence,a simple expedient to model the effect of the ionosphere is to replace it by a parabolic phase functional filter. Notice in this particular case that the parabola is concave down.

2.9 NOISE AND INTERFERENCE

Ideal transmission of information is impeded by two classes of phenomena; imperfect equipment (distortion), and the presence of "generalized noise,' i.e., any waveform, random or deterministic, other than the desired signal itself. Of course, in nature, both phenomena are unavoidable, but in principle and under some circumstances, this need not impair the perfect flow of information. For example, in Section 2.7 we saw that, in principle, linear distortion could be totally compensated for by the design of certain pulse shapes or by equalizers. Similarly, it is known that, even in the presence of noise, digital transmission can be error-free if the rate of transmission does not exceed the channel capacity. However, all known transmission schemes are adversely affected by the presence of unwanted waveforms. The latter comprise the usual thermal, or''front-end'' noise of radio amplifiers; the corresponding shot noise or impulsive noise at the ''front-end'' of optical receivers; a large class of external noise sources usually described as ''impulsive'', which may arise from natural or man-made sources; and another class that may be referred to as interference, which includes other signals, modulated or not, that are present (inadvertently or intentionally) at the input of a wanted signal's receiver. In this section we describe interference sources, primarily from the point of view of modeling them for the purpose of simulation.

2.9.1 Interference

Interference can be initially categorized in two broad categories: intentional (e.g., jamming) and unintentional. Implied in these terms is also a certain power level relative to the desired signal. Thus, intentional connotes that the interference is relatively high powered, on the order of the desired signal power or orders of magnitude larger. Also implied in this category is that the interferer may have a strategy to defeat measures in the receiver that are intended to mitigate its presence. Such measures generally fall under the heading of spread spectrum techniques, although such techniques are also useful in nonjamming environments, as in CDMA.

Unintentional interference, on the other hand, generally connotes relatively low-level interference, and is generally present as a result of the sharing of a communications resource, where signals cannot be perfectly isolated from one another. Frequency-division multiple access can produce adjacent-channel interference (ACI); sharing of the same frequency allocation by geostationary satellites and terrestrial radio-relay system results in cochannel interference (CCI) as well as ACI. In this subsection we discuss this type of interference.

The general interference scenario can thus be modeled simply by

$$S(t) = S_0(t) + \sum_{i=1}^{N} \alpha_i S_i(t) \tag{2.9.1}$$

where $S(t)$ is the total received waveform at the desired receiver input, $S_o(t)$ is the desired signal, and $\{S_i(t)\}$ are the interfering signals, possibly attenuated by the factors α_i representing isolation mechanisms such as antenna discrimination.

From the simulation viewpoint, this situation is easy to reproduce. The S_i (t) can be more or less arbitrary modulated signals, with carrier frequencies generally different from that of S_0 (t). We from the complex envelope of S(t) as usual and use sampled values. For example, if the S_i (t) are given by

$$S_i(t) = \alpha_i A_i(t) \cos [\omega_{ci} (t) + \phi_i + \theta_i] \tag{2.9.2}$$

and $S_o(t)$ by

$$S_0(t) = A_0(t) \cos[\omega_c t + \phi_0(t) + \theta_0] \tag{2.9.3}$$

then

$$\tilde{S}(t) = A_0(t)\, e^{j[\phi_0(t)+\theta_0]} + \sum_{i=1}^{N} \alpha_i A_i(t)\, e^{j[\phi_i(t) + \theta_i]} e^{j(\omega_{ci} - \omega_c)t} \tag{2.9.4}$$

is the complex equivalent low-pass signal of the receiver input. To simulate 2.9.4 we use sampled values. In principle,we could use different sampling rates for the different signals involved,depending on their individual bandwidth. However, it is more convenient to use a single sampling rate f_s if the bandwidth in question are not too different. It should be recognized, however, that, the condition $(\omega_{ci}-\omega_c) \neq 0$ has the effect of increasing the "effective" bandwidth associated with the received signal,hence will require a higher sampling rate than if the interference were cochannel. This increase in bandwidth stems from the fact that the equivalent low-pass spectrum is still modulated on a "carrier" frequency f_{ci} -f_c. In general, it is only necessary to account for the co-channel interference and the immediately adjacent channels. The effective bandwidth is, of course, at most the sum of the individual bandwidth. Generally, the most straightforward approach is simply to use form 2.9.4, and then take values $\tilde{S}(kT_s)$ at an appropriate valve of T_s, recognizing that $T_s = f^{-1}_s$ will usually be smaller than in the interference-free case.

Once we have values of $\tilde{S}(kT_s)$ the simulation proceeds in the same way whether interference is present or not. That is, simulation of the given system itself is,of course, independent of the input signal. Naturally, we expect to have some degradation, due to the presence of interference.

We note that the presence of unintentional interference generally induces a performance degradation that is relatively small, and this degradation can normally be compensated for by fairly straightforward and not too costly means, for example, by increasing the signal power by a moderate amount, or perhaps by improving the receiver front-end filtering characteristics. On the other hand, as the interference level becomes large, say as large as implied by the "intentional" label, the measures for mitigating the interference cannot stay simple.

CHAPTER III

CONCLUSION AND FURTHER SCOPE OF STUDY

3.1 CONCLUSION

In this thesis work, we have given an introduction to the notion of the "modeling" of communication system. A system can be considered to be composed of subsystems where conventional representation is the block diagram, each block of which stands for a subsystem. If a subsystem were to be the role object of study, it too could be considered a system which, in general, would be representable in terms of its own systems. Thus, modeling in general can be thought of in this hierarchical fashion.

In this thesis work, definition of system is that represented by Figure 2.1. Therefore modeling" in the context of this dissertation means to express the functioning of each subsystem in a way that can be actualized in the computer. The actual expression can be one or several equation, an algorithm, a set of tables, etc. Almost always, the form of expression is functional, meaning that what is represented is some net effect of interest, and not the basic physical building block.

3.2 FURTHER SCOPE OF STUDY

While we have given a fairly substantial sample of communication subsystem models, it is clear that a reasonably complete treatment would require a large volume in itself. Following are the topics which needs to be addressed for bringing a clear simulation images of communication systems.

(a) Basic properties of signals and systems with emphasis on their representation and generation in the simulation context. This includes linear time invariant (LTI), linear time variant (LTV) and non-linear (NL) systems.

(b) Information bearing signals, noise and interference in communication system are random in nature. Also, the characteristics of some components of a communication system might change randomly over time e.g. change in the response of a filter due to aging of components. These phenomena as well as signals,noise and interference can be modeled using random variable an random processes and their application to Monte Carlo simulation of communication systems.

(c) Complete modeling of communication system would further require models of Adaptive filtering,multipath channels, Discrete Channel, multiplexing/multiple access,Noise, Error Control Coding, Synchronization, spread spectrum Techniques and coded Modulation.

(d) Another discipline required for study is the statistical aspect of simulation for estimation of Performance Measures.

(e) A good portion of the simulation discipline deals with a number of issues that are as much "art" as science. These include software structure, reduction of block diagrams to simpler equivalent, the replacement of processes passed through chains of equipments by an equivalent process, sources and magnitude of errors, and the important related subject of validation.

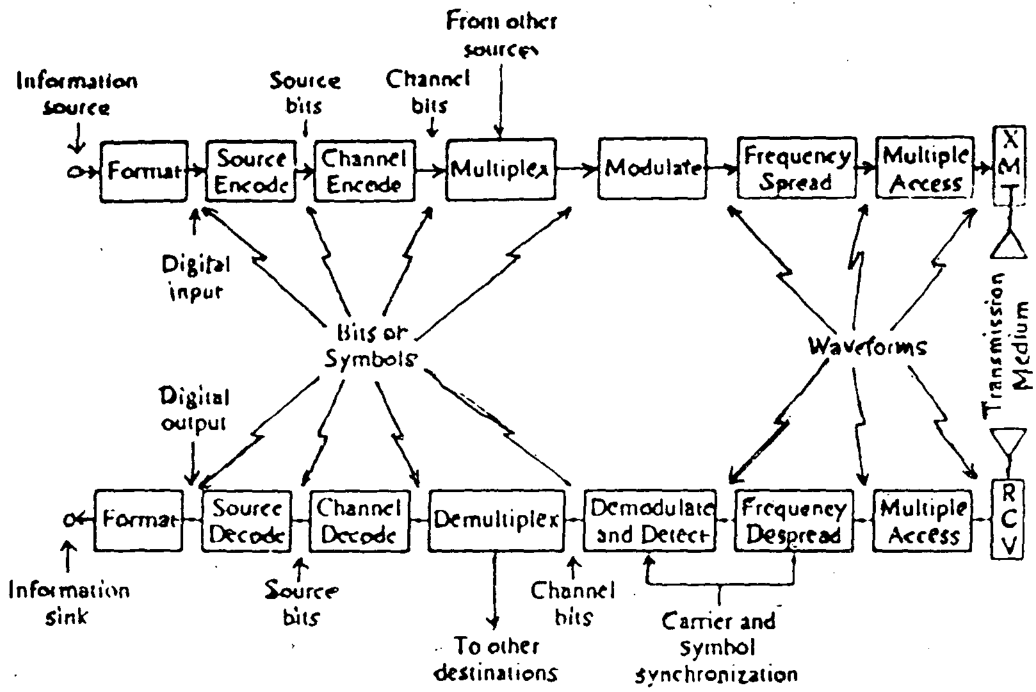


Fig.2.1 Block diagram of generic communication system (from Ref.1).

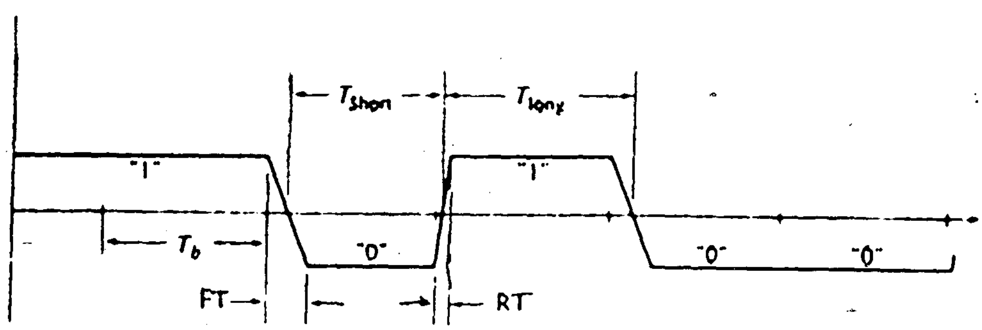


Fig.2.2 One possible model for simulating digital source waveform. RT, Risetime, FT, falltime, nominal bit duration $T_b = (T_{short} + T_{long})/2$)

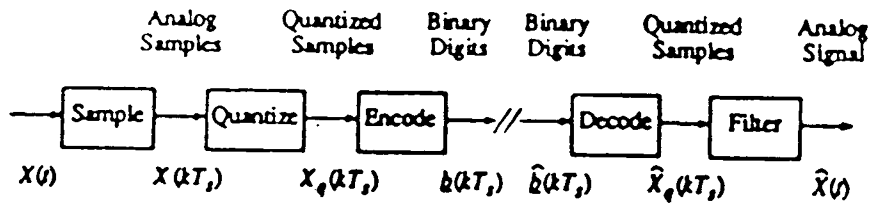


Fig.2.3 Block diagram of source encoding/decoding

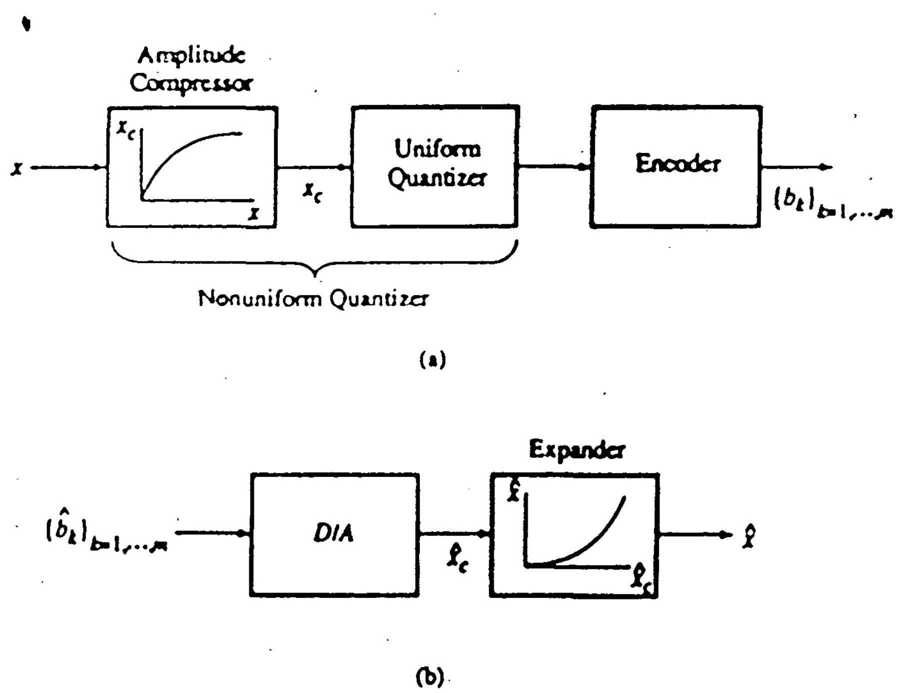


Fig.2.4 Nonuniform quantization process and its inverse. (a) Nonuniform analog-to-digital conversion; (b) nonuniform digital-to-analog conversion.

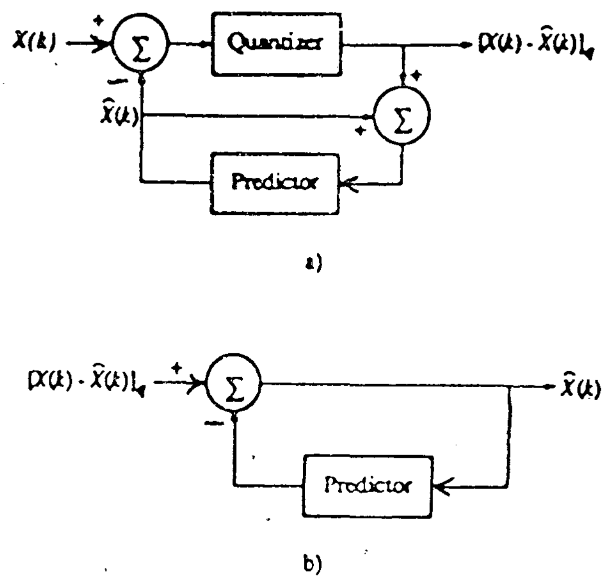


Fig.2.5 Differential encoding. (a) Transmitter; (b) receiver

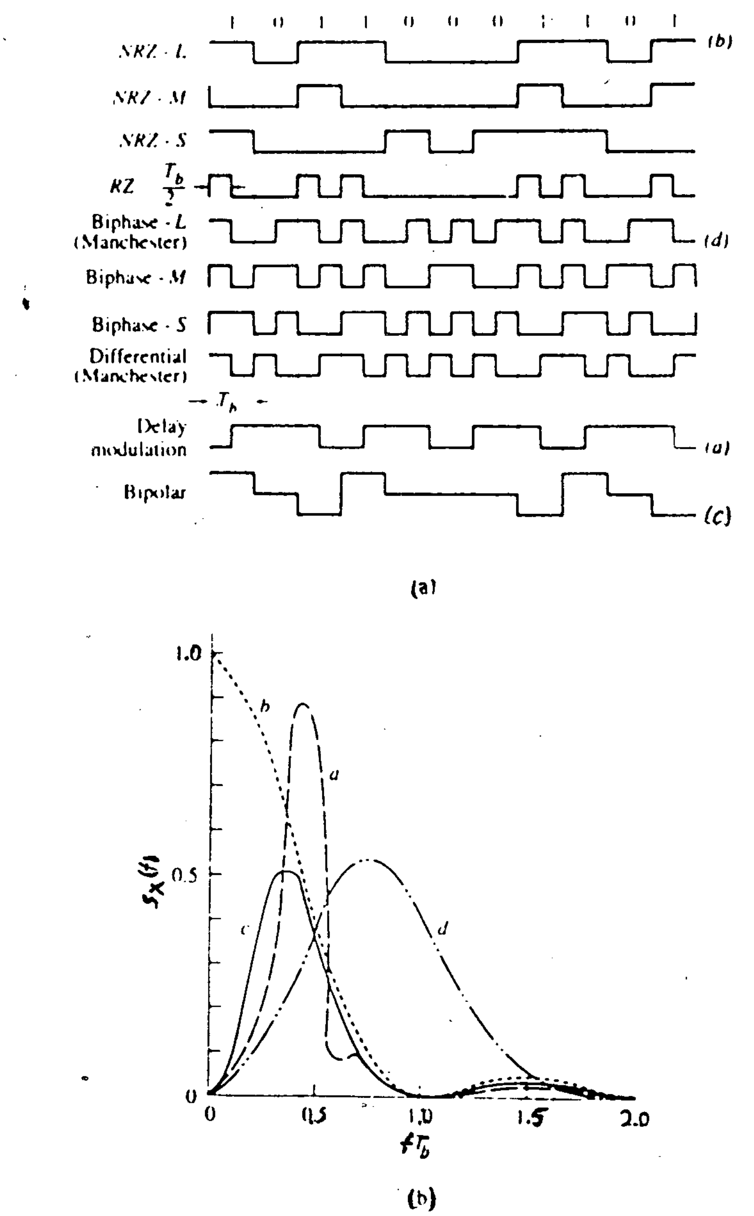


Fig.2.6 (a) Some digital signaling line coding formats. (b) Power spectra of several binary data line coding formats, a, Delay; b, NRZ-L; c, NRZ bipolar; d, Manchester.

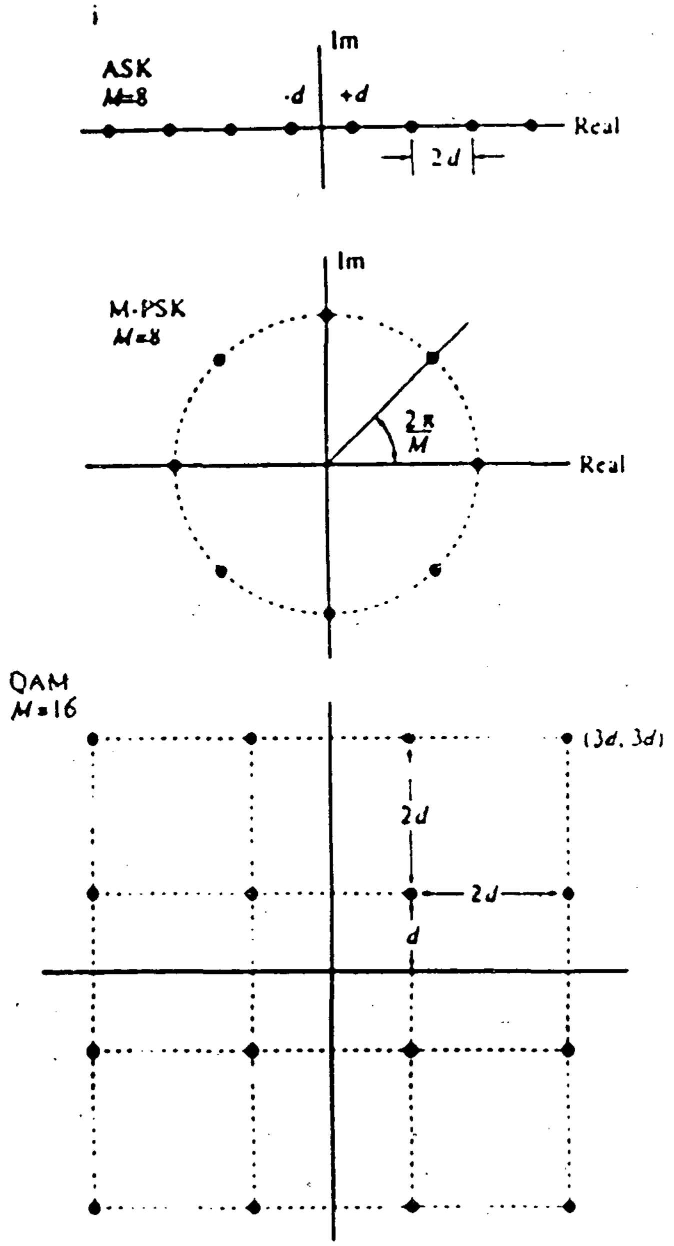


Fig.2.7 Some examples of signals "constellations" for digital modulations

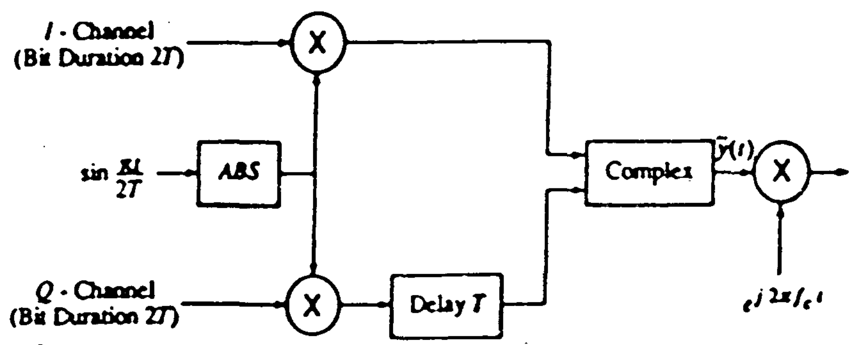


Fig.2.8 One implementation of MSK convenient for simulation purposes. The I and Q channels are initially synchronized with one another and with the zero-crossings of sin (πt/2T)

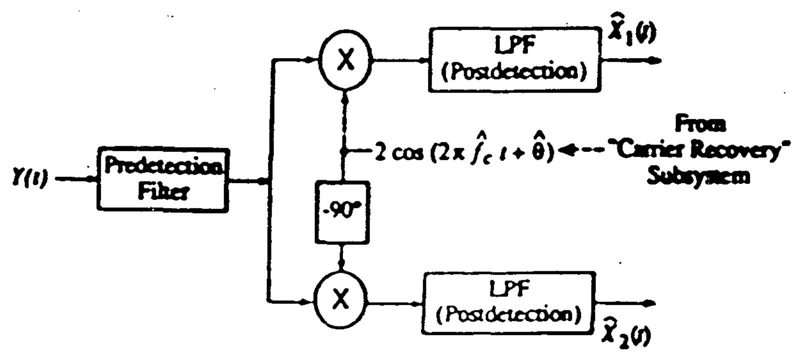


Fig. 2.9 "Generic" quadrature demodulator

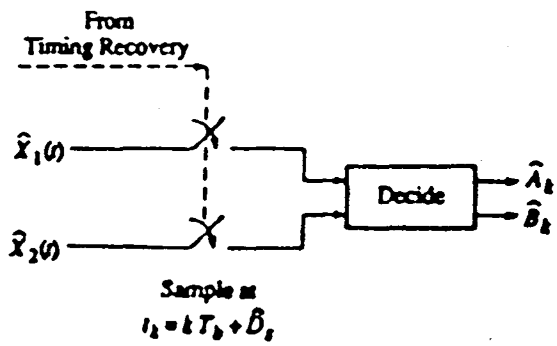


Fig.2.10 Sample and decide arrangement

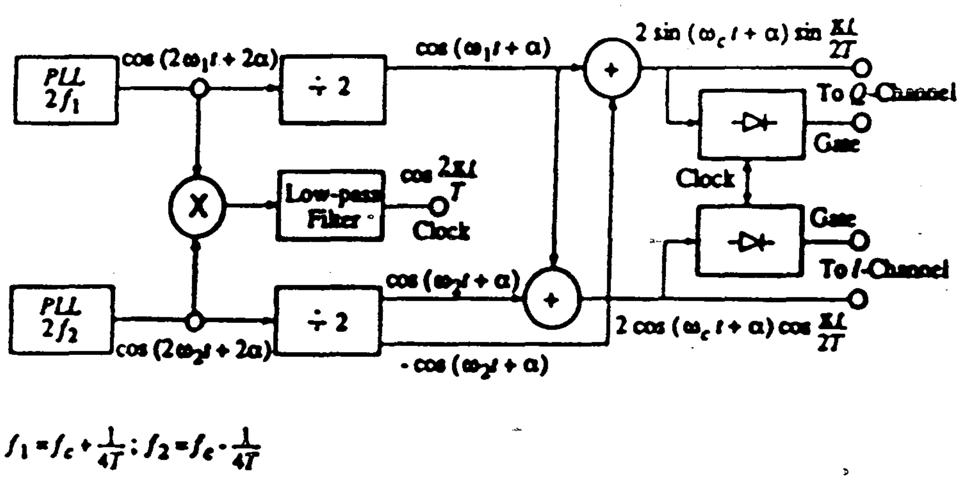


Fig.2.11 Integrated structure for the generation of carrier and timing references for MSK

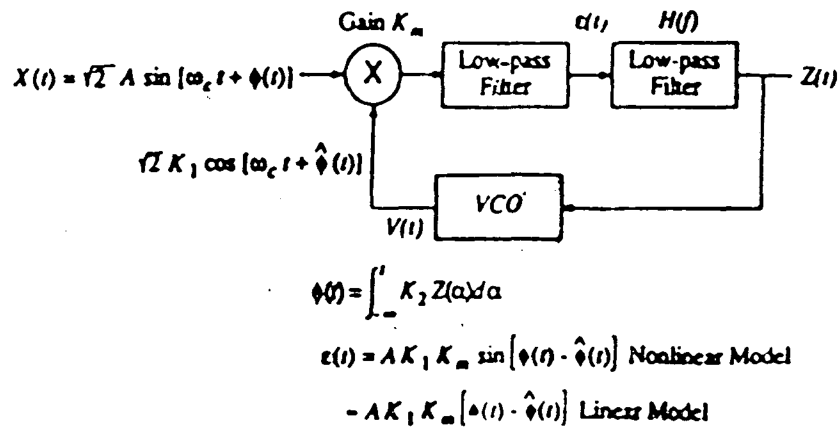


Fig.2.12 Phase locked loop

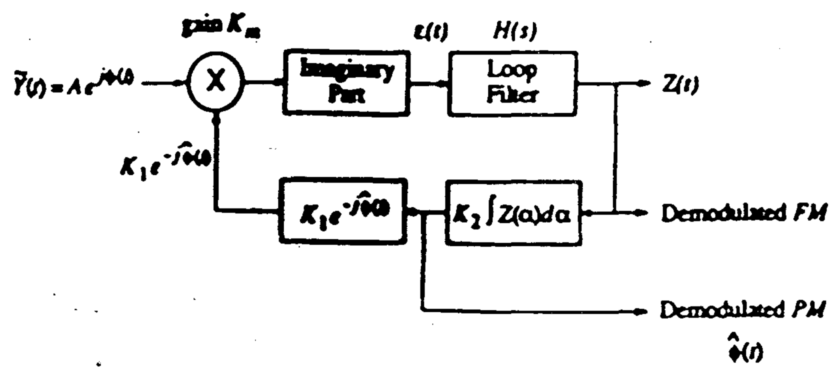
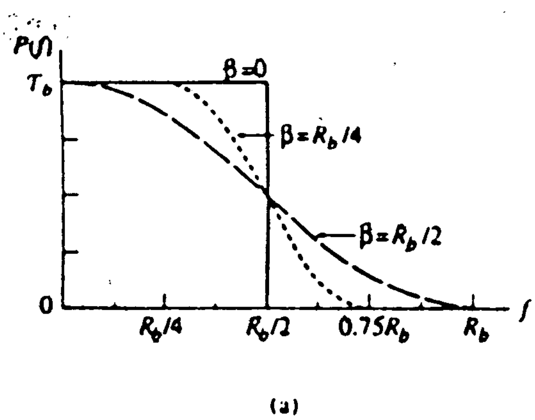


Fig.2.13 Complex low-pass equivalent of PLL



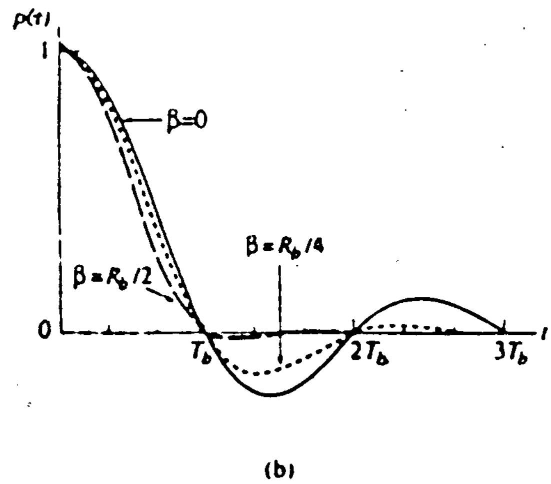


Fig.2.14 Pulses with raised cosine frequency characteristics. (a) P(f) for three values of β (Note that p(f) = P(-f)]; (b) p(t) for three values of β

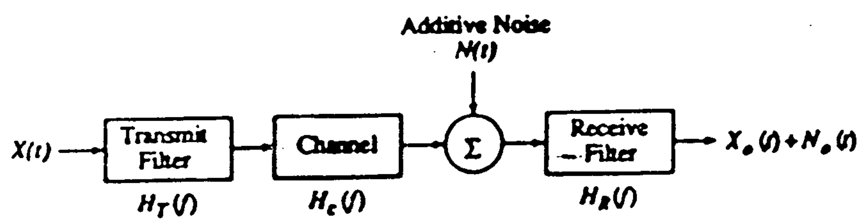


Fig.2.15 Channel model showing preemphasis (or transmit) filter-and deemphasis (or receive) filter.

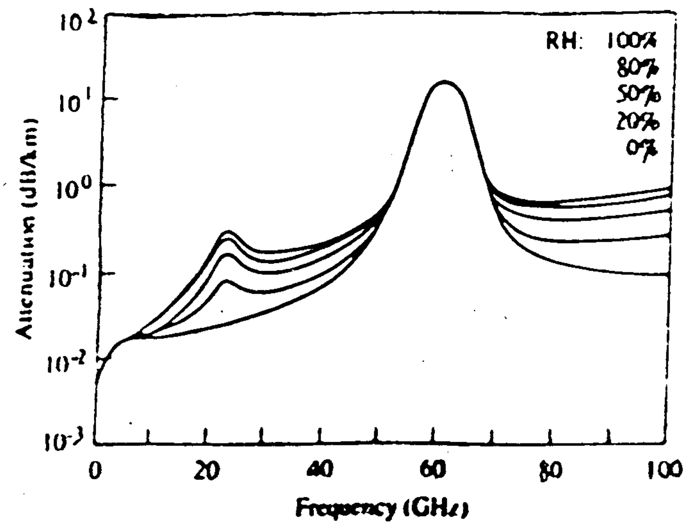


Fig.2.16 Atmospheric absorption at sea level as a function of frequency and relative humidity (from Ref.17)

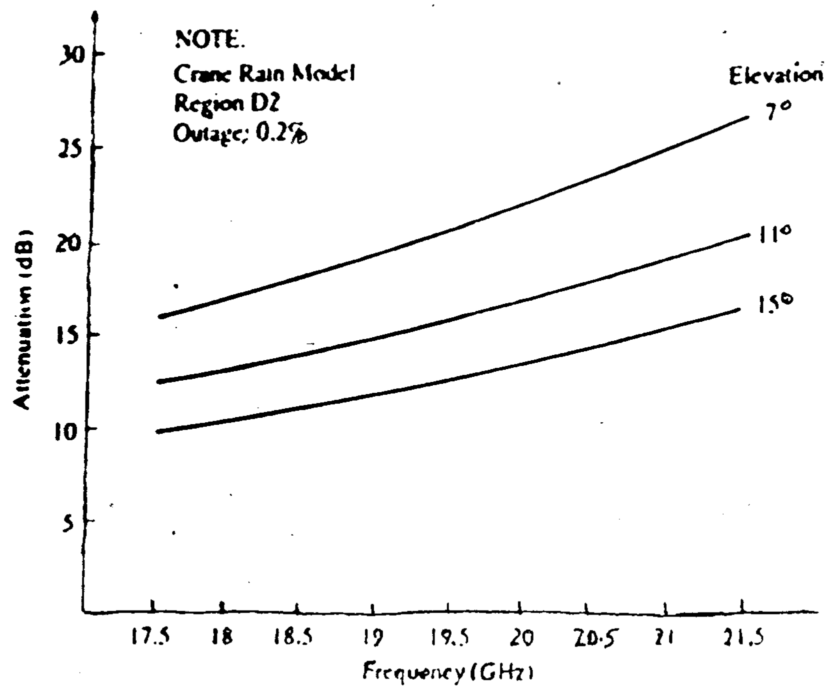


Fig.2.17 Rain attenuation as a function of frequency

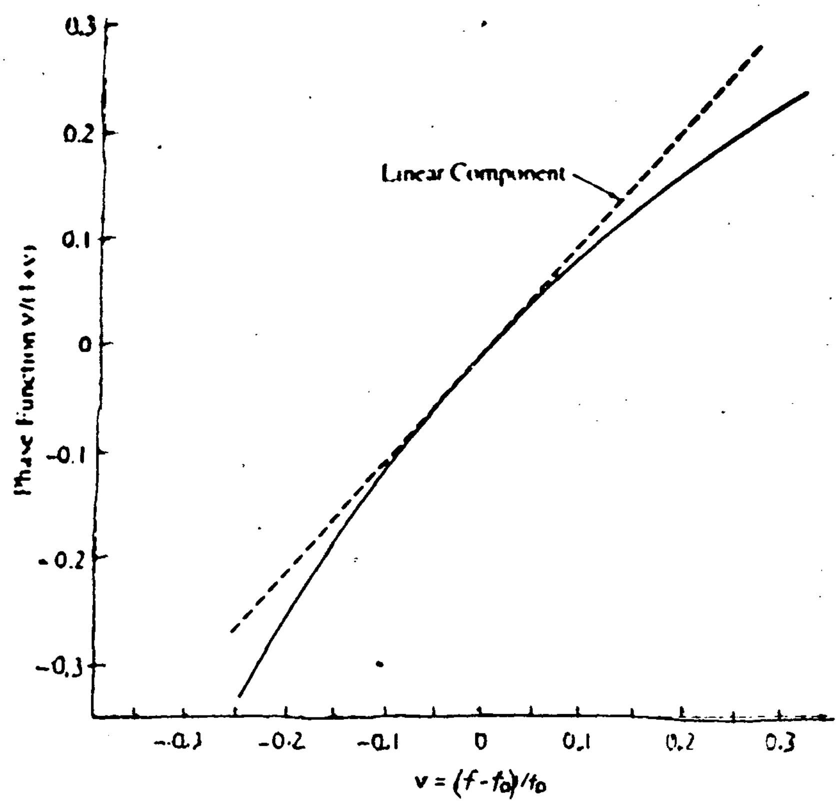


Fig.2.18 Ionospheric low-pass equivalent phase characteristic

BIBLIOGRAPHY

I K.S.Shanmugan Digital and Analog Communication Systems McGrawHill 1979.

II S. Haykin Communication Systems Wiley 1983

III J.J. Spilker Digital Communication by Satellite, Prentice Hall 1977

IV Digital Simulatation of Physical Systems, Joseph. S. Rosko

V Mathematical Modelling and Digital Simulation for Engineers and Scientists Jon M Smith

1. B. Sklar A structured overview of digital communications - A tutorial review IEEE common Mag 1983
2. B.Smith, Instantaneous companding of quantized signals
3. M.L. Hosing & D.G.Messerchmitt, Adaptive filters Klnwer Academic Press, Boston.
4. B.Widrow and S.D. Stearns, Adaptive Signal Processing, Prentice Hall
5. C.E. Shanman and W.waver, Mathematical Theory of communications, University of Illinois Press
6. S.Pasupathy, Correlative Coding IEEE Commun Mag (1977).
7. P.Kabal and S.Pasupathy, Partal response signalling IEEE Trans Commun 1975

8. S.Bandetto, Digital Transmission Theory.

9. J.G. Proakis, Digital Communications McGrawHill 1988.

10. W.P. Osborne, Coherent and noncoherant detection of CPFSK, IEEE Trans Commun 1974.

11. R.DeBuda, Coherent demodulation of FSK IEEE Trans Commun 1972

12. G.Lindgran, Shape and duration of clicks in modulated FM Transmission, IEEE Trans Inf. Theory (1984)

13. MT Jong and K.S. Shanmugan, Determination of transfer function from amplitude response data Int.J. Control 1977.

14. K.S. Shanmugan, Digital and Analog Communication Systems McGraw Hill, 1979.

15. G.H. Millman, atmospheric and Extraterrestrial Effects on Radio Wave Propogation. GEC Technical Info Series 1961.

16. H.J.Liebe. Modeling attenuation and phase of radio waves in air at frequencies below 1000 GHz Radio Sci 1981.

17. K.S. Shanmugan, V.S.Frost, Wideband digital transmission through the atmosphere at EHF Frequencies, Proc globecom Conference, Atlanta 1984.

18. R.K. Crave Prediction of attenuation by rain, IEEE Trans Commun 1980.

19. G.H. Millman, Tropospheric & Ionospheric phase Perturbations. GEC Technical Info Series 1984.