Order Number 9315443

Constrained FM, a narrow band FM signaling technique

Aghazadeh Alavi, Reza, Ph.D.

City University of New York, 1993

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

CONSTRAINED FM, A NARROW BAND FM SIGNALING TECHNIQUE

by

REZA AGHAZADEH ALAVI

A Dissertation submitted to the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1993

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

1/16/93
Date

Chair of Examining Committee

1/20/93
Date

Executive Officer

Dr. Joseph Barba

Dr. Radomir Bozovic

Dr. Svetislav Maric

Dr. Tarek Saadawi
Supervisory Committee

The City University of New York

Abstract

CONSTRAINED FM

by

Reza Aghazadeh Alavi

Advisor: Dr. Donald L. Schilling

There has been a great deal of research done on modulation schemes that are spectral efficient, such as M-ary PSK, MSK and Duobinary FM, to name a few.

In this research, another spectrally efficient method utilizing CPFSK as the modulation scheme, is presented. Here the data bits, d(t), are encoded such that the ratio of the maximum to minimum phase that the modulated FM signal can attain is bounded.

The power spectra as well as the error probability of the system utilizing a non-coherent detector has been investigated. It has been shown that the system has a power spectral efficiency of 1 b/s/Hz. Due to encoding of d(t), the modulation waveform exhibits a trellis structure. A coding gain of about 2 dB has been realized in decoding the received signal by utilizing this structure in decoding process. In addition, it has been shown that, other trellis coded modulation systems adhering to CNFM principles result in an improved probability of error at a cost of increasing the bandwidth necessary to transmit the coded signal.

ACKNOWLEDGEMENTS

There have been several people whom have had contributed either directly or indirectly toward the completion of this thesis, for which I am grateful and wish to thank.

I would like to thank Dr. Donald L. Schilling not only for his input, criticism, and constructive suggestions, without which this work would not have been possible, but also for his positive influence on me as a teacher, friend and role model that has inspired me.

I am also grateful to Mr. Kamal Shahrabi for his help and suggestions as well as Dr. Gary Lomb of the SCS Telecom who helped me understand trellis coded modulation.

Finally, I wish to thank my wife, Yasaman, and my son, Amir, whom have spent many lonely hours, however, provided me with endless support and encouragement. I am also thankful to my brother , Ali, and my sisters Shifteh, Shiva, and Sharareh, for their understanding and support.

TABLE OF CONTENTS

Chapter 1 ***An Overview***

1.0 Introduction 1

Chapter 2 ***Frequency Modulation***

2.0 Introduction 4
2.1 Bandwidth of frequency modulated signal 5

Chapter 3 ***Digital FM***

3.0 Introduction 16
3.1 Continues phase binary frequency shift keying (CPFSK) 17
3.2 Duobinary FM 24
3.3 Non-coherent detection of FSK 28
3.4 Performance of FSK signaling 29

Chapter 4 ***Coding***

4.0 Introduction 34
4.1 Linear block codes 35
4.2 Matrix description of linear codes 35
4.3 Decoding of received codeword 37
4.4 Cyclic codes 38
4.5 Reed-Solomon codes 40
4.6 Convolutional codes 41
4.7 Trellis-coded modulation 44

Chapter 5 ***Constrained FM***

5.0 Introduction 48
5.1 Generation of constrained FM 48
5.2 On the performance of CNFM 55
5.3 Trellis coded CNFM 62

Chapter 6 ***Summary and conclusion***

6.0 Summary and conclusion 81

Bibliography 83

LIST OF FIGURES

Figures

2.1 Block diagram for generating a FM signal 6

2.2 The FM modulating and modulated waveforms 7

2.3 Estimation of an FM signal bandwidth 12

3.1 Phase trajectory for FSK signaling 22

3.2 Power spectral density of FSK signaling 23

3.3 Power spectral density of rectangular pulse 23

3.4 Duobinary encoder with a precoder 26

3.5 Power spectral density of a duobinary signal 26

3.6 Power spectrum of a duobinary FM signal 27

3.7 Block diagram for a noncoherent detector 28

3.8 Probability of error versus SNR in dB for FSK 32

3.9 Comparison of FSK and DBFM 33

4.1 Rate 1/3 convolutional encoder 43

4.2 Trellis diagram for the rate 1/3 encoder 43

4.3 Set partitioning of an 8-PSK 45

4.4 Four-state trellis coded 8-PSK modulation 46

4.5 Four state trellis 47

5.1 Power spectral density of CPFSK for h=1/2 and 1/4 51

5.2 Power spectral density of Modified CNFM for h=1/2 and h=1/4 51

5.3 Block diagram for a CNFM transmitter 53

5.4 The phase trellis for a CNFM for h=0.5 54

5.5 The modulating waveform for a CNFM system 55

5.6 Power spectral density of CNFM for h=1 58

5.7 Power spectral density of CNFM for h=0.5 59

5.8 Power spectral density of CNFM for h=0.25 60

5.9 Power spectral density of CNFM for h=0.125 61

5.10 A comparison in PSD for MSK, DBFM, and CNFM for h=1/2 63

5.11 Probability of error vs. SNR for MSK, DBFM, and CNFM for h=1/2 . . . 64

5.12 Probability of error for CNFM for h=1/2, 1/4, and 1/8 65

5.13 Trellis structure for CNFM 66

5.14 A comparison in Pe between a CNFM signal being decoded conventionally or by utilizing its trellis structure in the decoding process 68

5.15 A four state trellis for a CNFM system 69

5.16 An eight state trellis for a CNFM system 70

5.17 Another eight state trellis for a CNFM system 71

5.18 A sixteen state trellis for a CNFM system 72

5.19 Power spectral density of the four state trellis CNFM 73

5.20 Power spectral density of the eight state trellis coded CNFM 74
5.21 PSD of the second 8 state trellis coded CNFM 75

5.22 Power spectral density of the sixteen state trellis coded CNFM 76

5.23 Block diagram for a transmitter/ receiver system used to generate the trellis coded CNFM 77

5.24 A multi-amplitude signal 78

5.25 Probability of error vs. SNR for the trellis coded CNFM system 79

CHAPTER 1

AN OVERVIEW

1.0 INTRODUCTION

In frequency modulation, **FM**, the carrier frequency changes in proportion to the modulating signal m(t). Therefore the carrier frequency changes continuously with time. The maximum change in frequency from the carrier frequency is known as the frequency deviation, (Δf). Depending upon the amount of frequency deviation with respect to bandwidth of the modulating signal, FM, can be categorized into either wideband FM (**WBFM**) or narrowband (**NBFM**).

When the frequency deviation is much larger than the bandwidth of the modulating waveform, WBFM results. The bandwidth necessary to accommodate such scheme of transmitting a signal is approximately twice the frequency deviation (Δf).

However, if Δf is much less than the bandwidth of the modulating waveform, NBFM occurs. In this case the bandwidth of the modulated waveform (B_{f_m}) is about twice the bandwidth of the modulating waveform.

Since the instantaneous frequency is the derivative of its phase, $\phi(t)$, a change in the instantaneous frequency causes a change in the phase of the modulated waveform. In fact, in binary shift keying (**BFSK**) or M-ary frequency shift keying (**MFSK**) the phase of the modulated waveform changes linearly with time within each period. During each period maximum change in phase is $\pm 2h\pi$, where h is the modulation index. If the instantaneous frequency changes continuously, the resulting frequency modulated signal is phase continuous and hence it is known as continuous phase FSK (**CPFSK**).

An FSK signal can be generated either continuously, varying the instantaneous frequency in proportion with the modulating waveform or by selecting one frequency from a bank of oscillators tuned to the desired frequencies. The latter suffers from the fact that it has larger side lobes outside its main-lobe, due to phase discontinuity as result of switching from one oscillator to another. Therefore CPFSK is preferred, since the band of frequencies needed to transmit a signal is smaller. In addition, in CPFSK, the constraint for the signal to have a continuous phase causes the system to have memory. This can be used to detect an error that occurs during the transmission of the signal.

There exists several CPFSK systems that are considered bandwidth efficient. They are minimum shift keying (**MSK**), duobinary signaling, and tamed FM. Although there exits several bandwidth efficient amplitude modulated schemes, such as quadrature phase shift keying (**QPSK**), M-ary phase shift keying (**MPSK**), or quadrature amplitude modulation (**QAM**), CPFSK is preferred in systems where non-linear elements exists.

In addition detection of FSK signals can be done either coherently or non-coherently. In non-coherent demodulation the information about the phase of the modulated signal is not used in the decoding process. Although this degrades the performances of the system, the advantage is a simpler receiver design.

The research in this thesis is in the characterization, design and development of a continuous phase narrowband digital FM system called **CONSTRAINED FM**. It is intended to operate at about 1 b/s/Hz. In this system the ratio of the maximum to minimum phase that the modulated FM signal can attain is bounded.

The objective of this thesis is to investigate:

1) The bandwidth efficiency of *constrained* FM.

2) The performance (P_e) of the system using non-coherent detection when the signal is corrupted by white gaussian noise.

3) To realize the net gain in performance once the structure of the code is utilized in decoding process.

4) To investigate the trade off between bandwidth and P_e utilizing other trellis encoded modulation systems while adhering to the concept of constrained FM.

CHAPTER 2

FREQUENCY MODULATION

2.0 INTRODUCTION

Frequency modulation (**FM**) is a form of angle modulation in which the instantaneous frequency changes with respect to the modulating signal. This is a nonlinear scheme of modulation. That is, the spectrum of the modulated waveform is not a frequency translated version of the baseband signal as in the case of amplitude modulation (**AM**). In fact, the spectrum of a FM signal depends on amplitude of the baseband signal as well as the frequency of the spectral components. A frequency modulated signal is given by

$$v(t) = A \cos\left(\omega_c t + 2\pi k_f \int_{-\infty}^{t} m(\lambda)\, d\lambda\right) \tag{2.1}$$

where ω_c is the angular velocity of the carrier, k_f a constant representing frequency sensitivity of the modulator in hertz per volt, and m(t) is the modulating signal. For a sinusoidal function the instantaneous angular velocity is equal to the time rate of change of the argument. That is

$$\omega_i(t) = \frac{d}{dt}\left[\omega_c t + 2\pi k_f \int_{-\infty}^{t} m(\lambda)\, d\lambda\right] \qquad (2.2)$$
$$= \omega_c + 2\pi k_f m(t)$$

or

$$f_i(t) = f_c + k_f m(t) \qquad (2.3)$$

From equation (2.3), it can be observed that the instantaneous frequency is directly proportional to the modulating waveform. As result the frequency the frequency modulated signal changes frequency whenever the modulation changes level. A block diagram for generating a frequency modulated signal is shown in figure 2.1 . A frequency modulated signal is shown in figure 2.2 .

2.1 BANDWIDTH OF A FREQUENCY MODULATED SIGNAL

To consider the bandwidth of a frequency modulated signal, define

$$\hat{v}(t) = A \exp\left[j(\omega_c t + k_f \theta(t))\right] \qquad (2.4)$$

where

$$\theta(t) = 2\pi \int_{-\infty}^{t} m(\lambda)\, d\lambda \qquad (2.5)$$

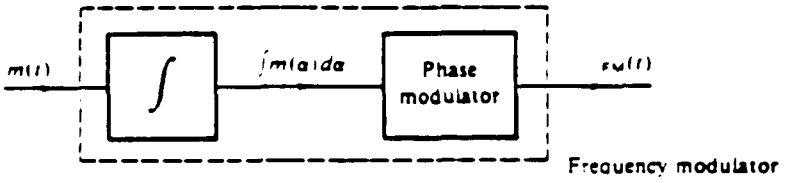


Figure 2.1 Block diagram for generating a FM signal.

Then

$$v(t) = Re\ [\ \hat{v}\ (t)\ ] \tag{2.6}$$

Equation (2.4) can be expanded using the Taylor series expansion formula, thus

$$\hat{v}(t) = A\ [\ 1 + jk_f\theta(t) - \frac{k_f^2}{2!}\theta^2(t) + \ldots + j^n\frac{k_f^n}{n!}\theta^n(t) + \ldots\ ] \tag{2.7}$$

so

$$v(t) = A\ [\ \cos\omega_c t - k_f\,\theta(t)\sin\omega_c t - \frac{k_f^2}{2!}\theta^2(t)\cos\omega_c t + \ldots\ ] \tag{2.8}$$

From equation (2.8), it can be seen that the spectrum of the modulated signal contains the carrier frequency in addition to sum of the spectrums of $\theta(t)$, $\theta^2(t)$ and etc. If m(t) is a bandlimited signal, then $\theta(t)$ is also bandlimited. This is due to the fact that integration is a linear process. So it multiplies the spectrum of m(t), M(f), by $1/j\omega$. Although the

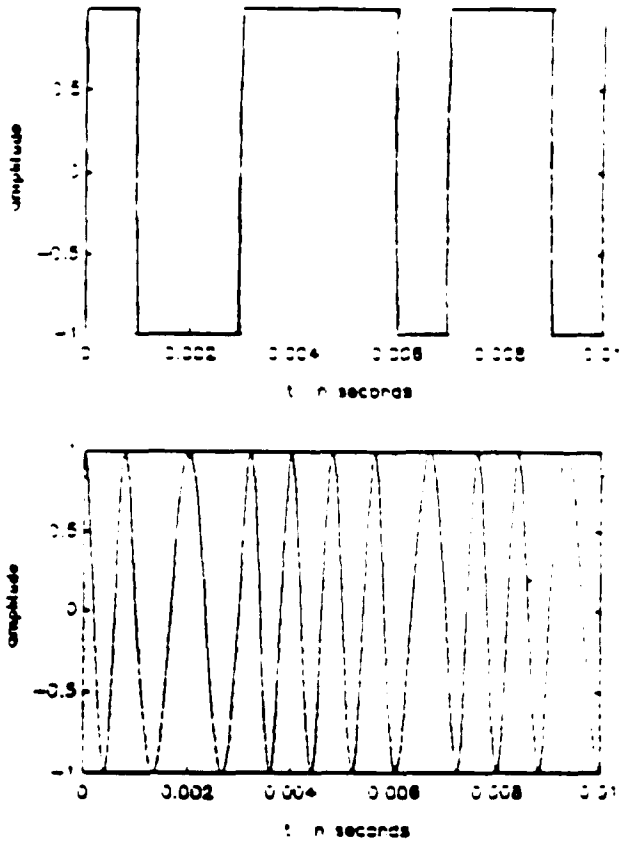


Figure 2.2 a)Modulating waveform b)Modulated signal

spectrum of $\theta(t)$ is bandlimited, the spectrum of v(t) is not. This can shown by using the fact that multiplication in time domain is convolution in frequency domain. Thus, the spectrum of $\theta^2(t)$ is $\theta(t) * \theta(t)$ and is 2B where B is the bandwidth of $\theta(t)$ and the symbol * represents convolution. Using the above argument then $\theta^n(t)$ has a bandwidth of nB.

Since the bandwidth of v(t) is the sum of the spectrums of $\theta'(t)$, it is not bandlimited. Although the theoretical bandwidth of FM is infinite, it turns out that most of the power of a FM signal is contained in a finite bandwidth. In fact, it has experimentally been determined that as long as 98 percent of the power of an FM signal is passed through a bandliniting filter, tolerable distortion occurs.

As an example, if the modulating signal is a sinusoidal signal, that is

$$m(t) = v_m \cos \omega_m t \tag{2.9}$$

then

$$v(t) = A \cos \left(\omega_c t + \frac{k_f v_m}{f_m} \sin \omega_m t \right)$$

or

$$v(t) = A \cos \left(w_c t + \frac{\Delta f}{f_m} \sin w_m t \right) \tag{2.10}$$

where

$$\Delta f = k_f v_m \tag{2.11}$$

The quantity Δf is called the frequency deviation, and it represents the maximum change in the instantaneous frequency from the carrier frequency. The ratio of the frequency deviation to the modulating frequency, f_m, is known as modulation index (β).

$$\beta = \frac{\Delta f}{f_m} \tag{2.12}$$

Thus

$$v(t) = A \cos (\omega_c t + \beta \sin \omega_m t) \tag{2.13}$$

By using equation 2.4, equation 2.13 can be written as

$$\begin{aligned} \hat{v}(t) &= A \exp [j(\omega_c t + \beta \sin \omega_m t)] \\ &= A \exp(j\omega_c t) [\exp (j\beta \sin \omega_m t)] \end{aligned} \tag{2.14}$$

If the term in the bracket is expanded using Fourier series, then

$$\exp (j\beta \sin \omega_m t) = \sum_{n=-\infty}^{\infty} c_n \exp (jn\omega_m t) \tag{2.15}$$

where

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp [j (\beta \sin\lambda - n\lambda)] d\lambda \tag{2.16}$$

The integral in equation 2.16, is known as the Bessel function of the first kind, order n. It is denoted by $J_n(\beta)$ and is extensively been tabulated. Thus

$$\hat{v}(t) = \sum_{n=-\infty}^{\infty} J_n(\beta) \exp [j (\omega_c + n\omega_m) t] \tag{2.17}$$

and

$$v(t) = A \sum_{n=-\infty}^{\infty} J_n(\beta) \cos (\omega_c + n\omega_m)t \tag{2.18}$$

The representation of v(t) in equation 2.18, is known as the series expansion of the single tone modulation. The spectrum of v(t) is obtained by taking the fourier transform of 2.18.

Then

$$V(f) = \frac{A}{2}\sum_{n=-\infty}^{\infty} J_n(\beta)\,\delta(\omega - n\omega_m - \omega_c) \tag{2.19}$$

from equation 2.19 it can be observed that the spectrum of v(t) consists of a carrier with an amplitude of $J_0(\beta)$ and sidebands spaced f_m apart with amplitudes of $J_n(\beta)$. The power associated with v(t) can be determined by using 2.18. That is

$$P_v = \frac{A}{2}\sum_{n=-\infty}^{\infty} J_n^2(\beta) \tag{2.20}$$

The number of sidebands, n, necessary to have 98 percent of energy always occurs whenever $n=\beta+1$. Thus, the bandwidth of an FM modulated signal (B_{fm}) can be approximated by

$$\begin{aligned} B_{fm} &= 2\,(\,\beta+1\,)\,f_m \\ &= 2\,(\,\Delta f+f_m\,) \end{aligned} \tag{2.21}$$

In the case of an arbitrary modulating signal, the bandwidth necessary to transmit an FM signal can be estimated using worse case tone modulation. That is, replacing f_m by B and Δf by $k_f m_p$, where B is the bandwidth of the modulating signal and m_p is the peak voltage of m(t). Thus

$$\begin{aligned} B_{fm} &= 2\,(\Delta f + B\,) \\ &= 2B\,(\,D + 1\,) \end{aligned} \tag{2.22}$$

where

$$D = \frac{\Delta f}{B} \tag{2.23}$$

The quantity D is known as the deviation ratio, and it plays the same role as the

modulation index (β) in the case of a tone modulated signal. Equation 2.22 is known as the Carson 's rule of bandwidth.

The Carson 's rule of bandwidth underestimates bandwidth for the case where $\Delta f >> B$. Here the bandwidth can be estimated by noting that the carrier frequency is usually much larger than the largest spectral component present in m(t). Thus, it can be thought that the carrier frequency remains a constant over the interval 1/2B. As result m(t) can be represented by the staircase signal shown in figure 2.3a. The sinusoidal waveform gives rise only to a single spectral component if it exist over the time interval of $(-\infty, \infty)$. Once the sinusoidal waveform exists only over a finite interval, as shown in figure 2.3b, the spectrum spreads. This can be shown using the fact that a sinusoidal waveform over a limited time can be generated as result of multiplication of sinewave by the gating function $\Pi(t)$, where the gating function is defined as

$$\Pi(t) = \begin{cases} 1 & for\ -T/2 < t < T/2 \\ 0 & Otw. \end{cases} \tag{2.24}$$

Thus, its spectrum is the convolution of the spectrum of the gating function and that of the sinewave. This is shown in figure 2.3c. This is what happens in FM. At each sampling instant of m(t), $m(t_k)$, the carrier frequency f_k, remains constant over the time interval t_k to t_{k+1}. However over the interval larger than the sampling interval, the carrier frequency appears to change to a new value depending on the value of the sample $m(t_k)$. Hence, each interval over which f_k can be assumed as constant gives rise to spectral spreading. From figure 2.3c, it can be seen that the spectrum lies in the band f_k-2B to

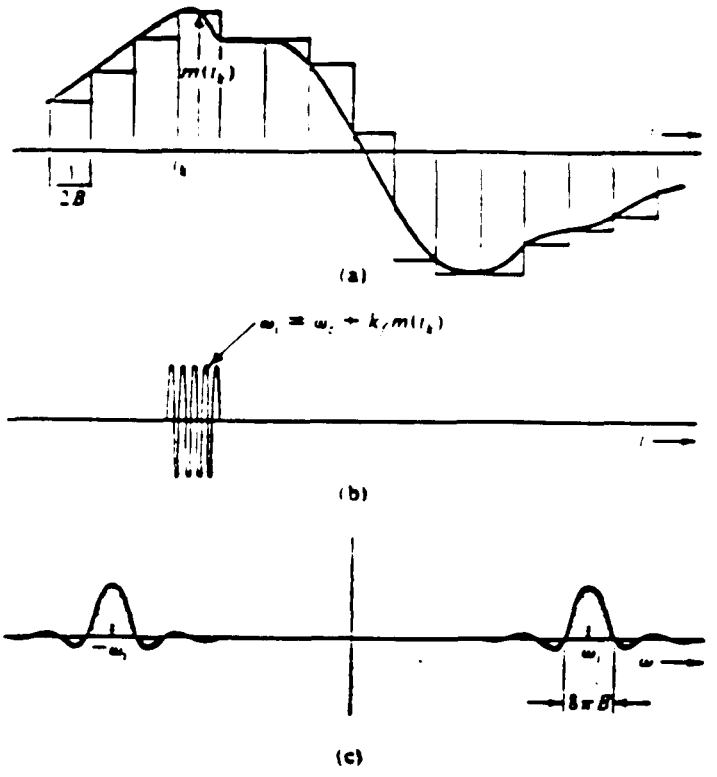


Figure 2.3 Bandwidth estimation of a FM waveform.

f_k+2B. Thus the spectrum of the FM signal is in the range of $f_k - k_f m_p - 2B$ to $f_k + k_f m_p + 2B$ and the bandwidth of the FM signal is given by

$$B_{fm} = 2 (\Delta f + 2B) = 2B (D + 2) \quad (2.25)$$

Thus by comparing equations 2.22 and 2.25 bandwidth of an FM signal can be written in general as

$$B_{fm} = 2B (D + k) \quad (2.26)$$

where k is a number between 1 and 2, and it depends on the value of D, the deviation ratio. It is in terms of D, or β for the case of tone modulation, that FM can be

categorized as either narrowband FM (**NBFM**) or wideband FM (**WBFM**). If D < < 1, NBFM results. In this case $\Delta f << B$. As result, the bandwidth of the FM signal is mainly due to the spreading phenomena, and is in the order of B on both sides of the spectrum. Therefore bandwidth of a NBFM signal is determined by using Carson 's rule for bandwidth and is given by

$$B_{fm} = 2\ (\Delta f + B\ ) \approx 2B \tag{2.27}$$

On the other hand, in the case WBFM, where $\Delta f >> B$, the bandwidth of a FM signal is given by

$$B_{fm} = 2\ (\ \Delta f + 2B\ ) \approx 2\Delta f \tag{2.28}$$

Now it will be shown the value of β for which an FM signal can be considered as a NBFM signal, is about .5 . In NBFM, B_{fm} is approximately about 2B. The same amount of bandwidth necessary to transmit an AM signal that is also bandlimited to B Hz. In addition, narrowband FM is approximately linear and very similar to AM. Therefore the value of β for which an FM signal has the same performance as an AM signal, can be used as the border line between WBFM and NBFM. The performance of an AM signal, denoted by $(S_o/N_o)_{AM}$, having a modulation index of one, is given by

$$\left(\frac{S_o}{N_o} \right)_{AM} = \frac{\overline{m^2}}{m_p^2 + \overline{m^2}} \frac{S_i}{\eta B} \tag{2.29}$$

where $\overline{m^2}$ is the mean square value of the modulating signal and m_p is the peak amplitude of the modulating waveform. The quantity $S_i/\eta B$ is the input signal to noise ratio of the modulated waveform, corrupted by white gaussian noise having a spectral density of

($\eta/2$). It is also referred to as the figure of merit. The performance of an FM signal, also corrupted by white gaussian noise, using discriminator receiver is given by

$$(S_o/N_o)_{FM} = 3\beta^2 \left(\frac{\overline{m^2}}{m_p^2} \right) \frac{S_i}{\eta B} \tag{2.30}$$

Now equating equations 2.29 & 2.30, the value of β is given by

$$\beta^2 = \frac{1}{3} \left[\frac{1}{1 + (\overline{m^2} / m_p^2)} \right] \tag{2.31}$$

since the value of $\overline{m^2} / m_p^2 < 1$, then

$$\beta^2 = \frac{1}{3}$$

or

$$\beta = 0.6 \tag{2.33}$$

Thus for $\beta < .6$, NBFM occurs.

The value of modulation index plays an important role in digital FM where frequency modulated signals are used to transmit digitalized signals. By the virtue of digitalizing signals, the spectrum necessary for transmitting a signal increases. As result bandwidth efficient signaling must be used. Keeping the value of the modulation index, which is represented by h in digital signals, small causes the bandwidth of the transmitted signal to become the same as the bandwidth of the baseband signal. Thus by proper wave shaping the baseband signal, the bandwidth necessary to transmit a signal can be

minimized. In the next chapter several of the methods used to accomplish this task are discussed.

CHAPTER 3

DIGITAL FM

3.0 INTRODUCTION

In digital FM, the frequency of a carrier signal is changed with respect to digital data. The principle feature of digital communication is that during a finite interval of time, it sends a waveform from a finite set of waveforms. Therefore, at the receiver the objective is not to reproduce a transmitted waveform with precision; it is instead to determine from noise contaminated signal which waveform from the finite set of waveforms had been sent by the transmitter. As result digital communication has several advantages over analog communication.

1. Digital communication is more noise immune to channel noise and distortion.

2. Digital signals are easier to regenerate.

3. Digital hardware implementation is more flexible and less subject to distortion and interference.

4. Digital signals can be coded to have excellent performance as well as privacy.

5. In digital transmission, one can trade off performance for bandwidth.

The primary disadvantage of digital communication is that it requires more bandwidth. Thus, narrowband FM becomes an attractive means of transmitting digital data. As stated previously, by limiting the modulation index, h, to be less 0.6, the bandwidth of the system will be the same as that of the baseband signal. Several of these techniques are discussed in this chapter. However the discussion is limited to continuous phase modulation methods, since these are spectrally more efficient.

3.1 CONTINUES PHASE BINARY FREQUENCY SHIFT KEYING (CPBFSK)

In frequency shift keying the data bits modulate the carrier waveform by changing the carrier frequency in an amount in proportion to the modulating signal or the data bit stream, d(t). The data bit stream may be represented by

$$d(t) = \sum_{n} A_n \, u(t - nT) \tag{3.1}$$

where A_n is the amplitude of the modulating waveforms, taking on the value of ± 1, and

u(t) is a rectangular waveform having an amplitude of 1/T, with a duration of T seconds. Now by using equation 2.8, the frequency modulated waveform may be represented by

$$v(t) = A \cos\left(2\pi f_c t + 2\pi T f_d \int_{-\infty}^{t} d(\lambda)\, d\lambda + \phi_0\right) \tag{3.2}$$

where k_f has been replaced by Tf_d. The terms ϕ_o and f_d represent the initial phase and the peak frequency deviation respectively. By substituting equation 3.1 into above equation and assuming that the initial phase is zero, without any loss of generality, then equation 3.2 may be expressed as

$$v(t) = A \cos\left(2\pi f_c t + 2\pi T f_d \int_{-\infty}^{t} \sum_{n} u(\lambda - nT)\, d\lambda\right) \tag{3.3}$$

for $nT < t < (n+1)T$

$$\begin{aligned} v(t) &= A \cos\left(2\pi f_c t + 2\pi T f_d \sum_{k=-\infty}^{n} A_k + 2\pi T f_d A_n q(t-nT)\right) \\ &= A \cos(2\pi f_c t + \theta_n + \pi h A_n q(t-nT)) \end{aligned} \tag{3.4}$$

where

$$h = 2Tf_d$$

$$\theta_n = \pi h \sum_{k=-\infty}^{n} A_k$$

and

$$q(t) = \begin{cases} 0 & \text{for } t < 0 \\ t/T & 0 \le t \le T \\ 1 & t > T \end{cases}$$

The terms h,and θ_n represent the modulation index and the accumulation of all symbols up to (n-1)T, which may be thought of as the memory of the system. In

addition, although the data bit stream is discontinuous, its integral is not. Thus a continuous phase signal results.

Orthogonal signaling plays an important role in the performance of FSK signaling. For the signals to be orthogonal, they must satisfy

$$\int_0^T v_1(\lambda)v_2(\lambda)\,d\lambda = 0 \tag{3.5}$$

For BPFSK signals to be orthogonal, the modulation index, h, takes on the values of either 1 or 1/2, depending if the signal is being detected non-coherently or coherently, respectively. Since the value of h = 1/2, corresponds to the minimum value of h such that orthogonality can be preserved, this is called minimum shift keying (**MSK**) or fast frequency shift keying (**FFSK**). To show this, let

$$f_c + \pi\frac{h}{T} = f_1 \quad \& \quad f_c - \pi\frac{h}{T} = f_2 \tag{3.6}$$

Then, assuming d(t) = 1 equation 3.4 may be rewritten as

$$v(t) = A\cos(2\pi f_1 t + \theta_n) \tag{3.7}$$

so for the transmitted signal and locally generated signal to be orthogonal, they should satisfy

$$\int_0^T \cos(2\pi f_1 t + \theta_n)\cos(2\pi f_2 t)\,dt = 0 \tag{3.8}$$

This yields

$$\cos\theta_n \sin 2\pi(f_1 - f_2)T + \sin\theta_n [\cos 2\pi(f_1 - f_2)T - 1] = 0 \tag{3.9}$$

If the value of θ_n is not known, as in the case of *non-coherent* detection, then to satisfy

equation 3.9 both sin $2\pi(f_1 - f_2)T = 0$, and simultaneously cos $2\pi(f_1 - f_2) = 1$. Thus

$$2\pi (f_1 - f_2)T = 2k\pi$$
$$or \quad f_1 - f_2 = \frac{k}{T} \tag{3.10}$$

where k is an integer. For minimum spacing of the signals k = 1. On the other hand, if the value of θ_n is known, as in the case of *coherent* detection, the locally generated signal includes the offset in phase. so sin $\theta_n = 0$ and only the first term of equation 3.9 must be set equal to zero. Thus

$$\sin 2\pi (f_1 - f_2)T = k\pi$$
$$or \quad f_1 - f_2 = \frac{k}{2T} \tag{3.11}$$

Again for minimum tone spacing of the signals the value of k = 1. However by comparing equations 3.10 with that of 3.11, the spacing required for MSK signaling is half of the non-coherent orthogonal BFSK. Thus, in orthogonal FSK, coherent FSK is the most bandwidth efficient technique. In order to determine the distribution of the signals energy in the frequency domain as means of comparing bandwidth efficiency of different signaling technique, power spectral density (**PSD**) of the signals must be evaluated.

In order to determine the power spectral density of FSK signaling, for $0 < t < T$, equation 3.4 may be rewritten as

$$v(t) = A \cos\left(\theta_n \pm \frac{\pi h t}{T}\right) \cos 2\pi f_c t - A \sin\left(\theta_n \pm \frac{\pi h t}{T}\right) \sin 2\pi f_c t \tag{3.12}$$

For h = 1, θ_n can only take on an integer multiple values of π, or simply 0 or π since all phase shifts are modula 2π, as shown in figure 3.1, so equation 3.12 can be

simplified to

$$v(t) = \pm A \cos\left(\frac{\pi t}{T}\right)\cos 2\pi f_c t \pm A \sin\left(\frac{\pi t}{T}\right)\sin 2\pi f_c t \qquad (3.13)$$

Since both the in-phase and the quadrature components of the carrier frequency are multiplied by signal shaping functions, then the low pass power spectral density of v(t), denoted by $S_{v(lp)}(f)$, is the PSD of these symbol shaping functions, which is

$$S_{v(lp)}(f) = \frac{4\pi A^2 \cos^2(\pi T f)}{\pi(4T^2f^2 - 1)^2} \qquad (3.14)$$

For MSK signaling where h=1/2, θ_n has in interesting feature. That is θ_n can only take on the values of $\pm\pi/2$ at odd multiples of T and only the values of 0 or π at even multiples of T. Thus equation 3.4 may be written as

$$v(t) = \pm A \cos\left(\frac{\pi t}{2T}\right)\cos(2\pi f_c t) \pm A \sin\left(\frac{\pi t}{2T}\right)\sin(2\pi f_c t) \qquad (3.15)$$

for $0<t<T$. By observing equation 3.15, it can be seen that the wave shaping signals are defined over an interval of 2T. This is due to the fact that θ_n can take on the values of 0 and π, or $\pm\pi/2$, at every other bit. The PSD of these symbol shaping function is given by

$$S_{v(lp)}(f) = \frac{16A^2T^2}{\pi^2}\left[\frac{\cos(2\pi Tf)}{16T^2f^2 - 1}\right]^2 \qquad (3.16)$$

The baseband power spectra of equations 3.14 and 3.16 for the FSK signaling where h=1 and h=1/2 is plotted in figure 3.2a. Figure 3.2b is the computer simulated plot of the same power spectra. From Figure 3.2 it can be observed that MSK signaling utilizes only half as much bandwidth as FSK with h=1. This is as result of symbol

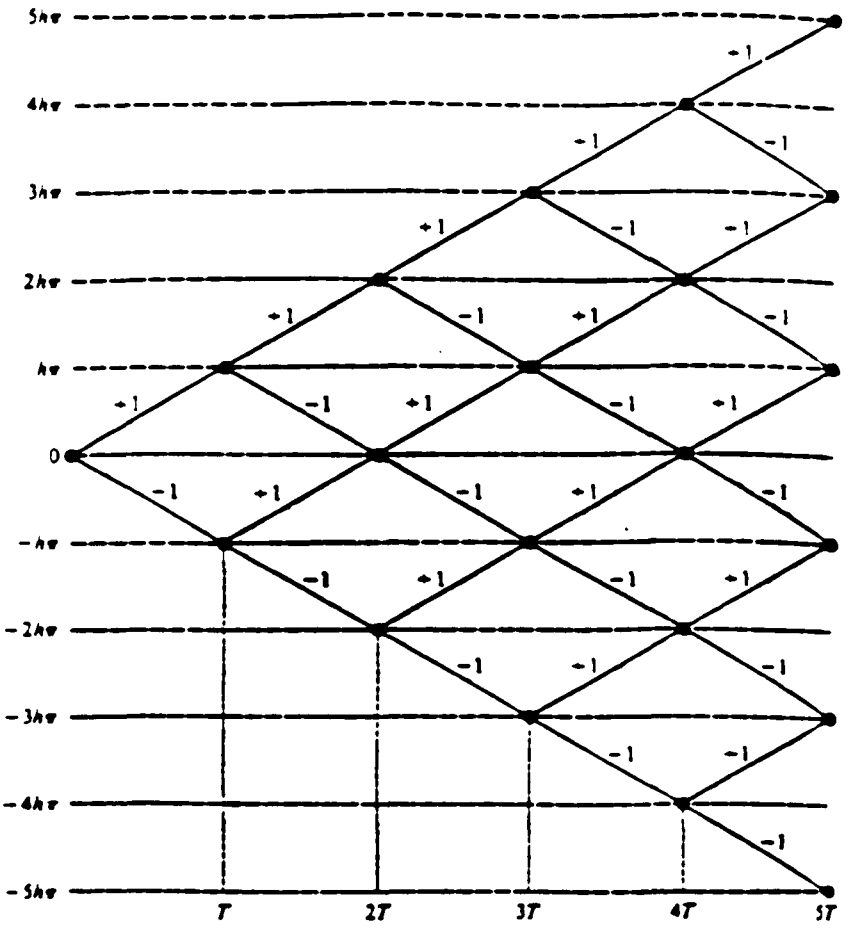


Figure 3.1 Phase trajectory for FSK signaling.

shaping functions being defined over 2T. It is also worthwhile noting that side lobes of CPFSK are reduced by a factor of f^4, as compared to non-continuous FSK signaling where the side lobes reduce by a factor of f^2.

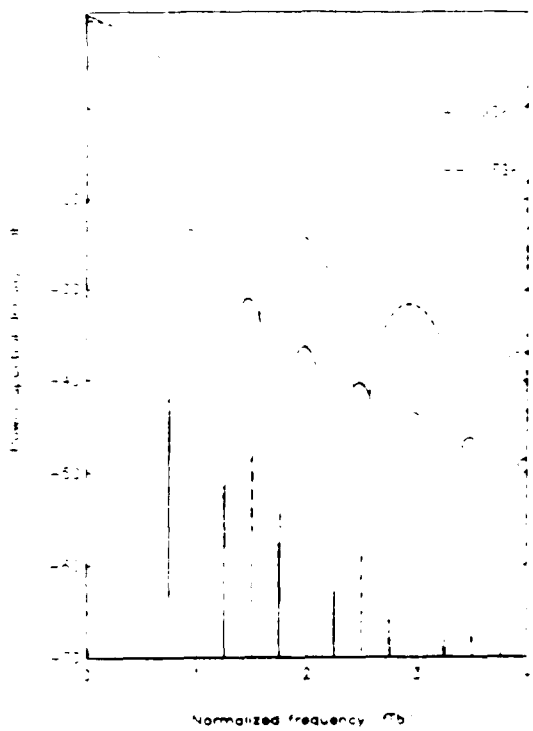


Figure 3.2 Power spectral density of FSK signaling. a)h=1 "-" b)h=1/2 "--"

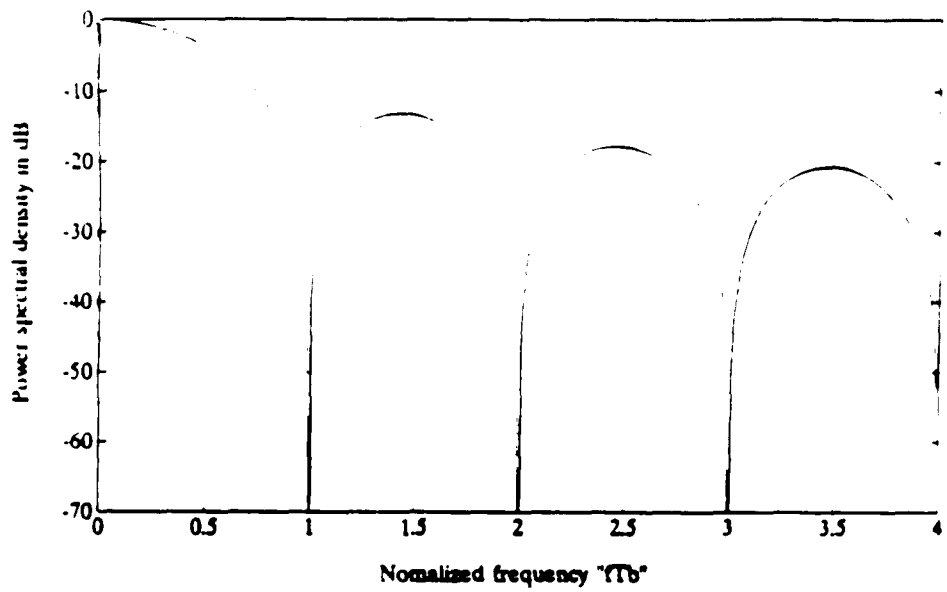


Figure 3.3 Power spectral density of a rectangular pulse having an amplitude of 1/T and duration of T seconds.

3.2 DUOBINARY FM

Thus far the baseband signals used to frequency modulate the carrier signal have been rectangular in nature. A rectangular pulse having an amplitude of 1/T and a duration of T seconds, as defined by u(t) previously, has a power spectral density given by

$$S(f) = \frac{1}{T}\left(\frac{\sin 2\pi Tf}{\pi Tf}\right)^2 \qquad (3.17)$$

From the above equation and Figure 3.3, it can be observed that most of the energy of the pulse is concentrated in the range from 0 to f_b. Thus, the minimum bandwidth necessary to transmit this pulse is f_b Hz. If the baseband signal is frequency modulated, the approximate bandwidth necessary for transmission of the signal is given by Carson's rule. That is

$$BW = 2\,(\Delta f + f_m) \qquad (3.18)$$

However in NBFM, since Δf is either less than or comparable to f_m, the bandwidth necessary for transmission of signal decreases if f_m can be reduced. Duobinary signaling encodes the baseband signal according to the rule shown in figure 3.4. According to this rule, the present pulse is added to the previous pulse to generate the duobinary signal $m_{db}(t)$. Since m(t) can take on only values of $\pm$ 1/T, the resulting duobinary signal is a tri-ary pulse taking values of $\pm$ 1/T and zero. To observe spectral efficiency obtained from this, the duobinary encoder can be thought of as a pre-filter. Since an ideal delay element, producing a delay of T seconds, has a transfer function of

exp(-j2πfT), the transfer function of the encoder, H(f), is written as

$$H(f) = \frac{1}{2}\left[1 + e^{-j2\pi fT}\right] = \cos \pi fT \tag{3.19}$$

Thus, the power spectral density of the duobinary signal is given by

$$S_{db} = H^2(f) \cdot S(f) = \frac{1}{T}\cos^2 \pi fT \cdot \left(\frac{\sin \pi fT}{\pi fT}\right)^2 \tag{3.20}$$

Figure 3.5 shows a plot of the above power spectral density. If a doubinary signal is used to frequency modulate a carrier, the resulting modulated wave form has power spectral density shown in figure 3.6.

As means of comparison between MSK and duobinary FM as far spectral efficiency is concerned, the bandwidth for transmission of data at the rate of 1/T must be determined. From Figures 3.3 and 3.5, it is reasonable to assume that f_m may be taken 1/4T and 1/2T for duobinary And MSK signaling, respectively. With $\Delta f = 1/4T$, the transmitted bandwidth is

$$B|_{dbfm} = 2\left(\frac{1}{4T} + \frac{1}{4T}\right) = \frac{1}{T} = f_b$$
$$B|_{MSK} = 2\left(\frac{1}{2T} + \frac{1}{4T}\right) = \frac{1.5}{T} = 1.5f_b \tag{3.21}$$

Thus MSK signaling requires 50% more bandwidth than duobinary FM. Hence it is spectrally more efficient.

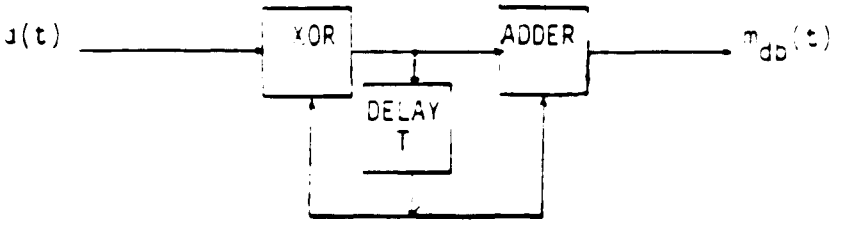


Figure 3.4 Duobinary encoder with a precoder.

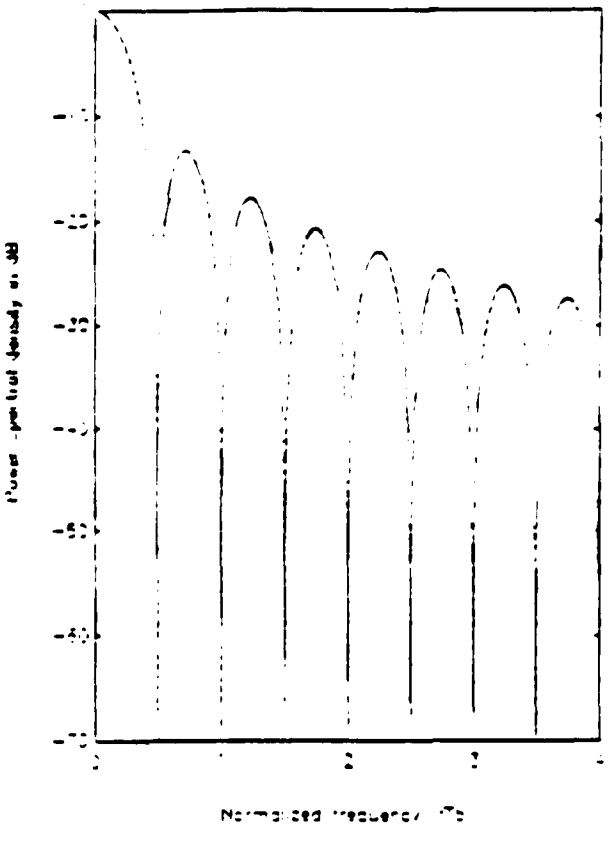


Figure 3.5 Power spectral density of a duobinary signal.

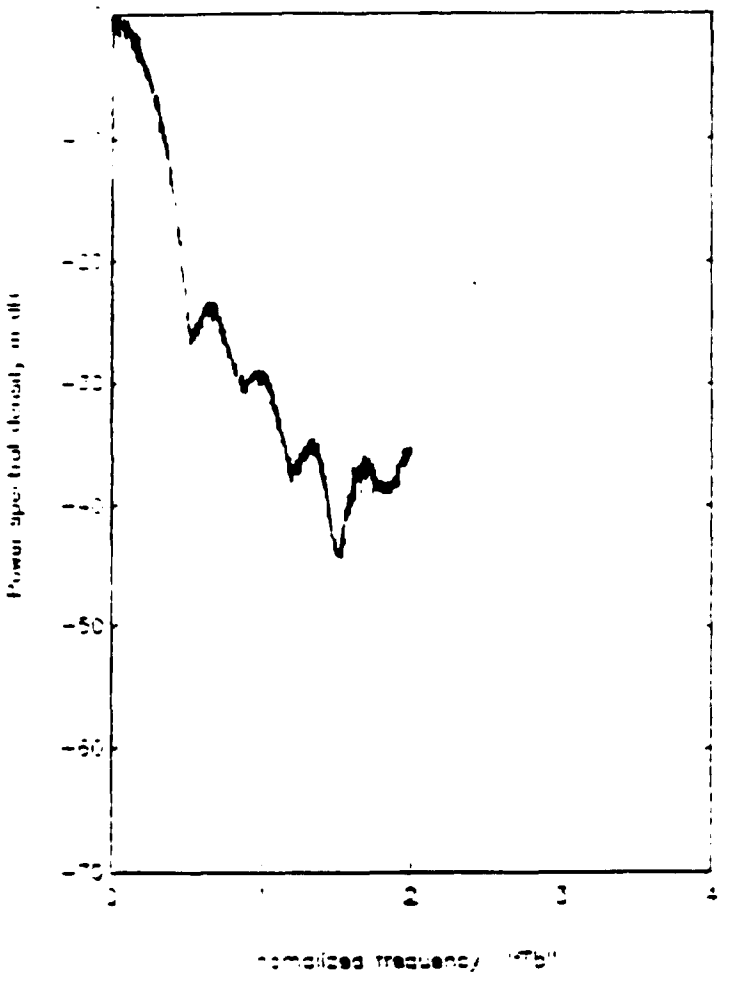


Figure 3.6 Power spectral density of duobinary FM signal.

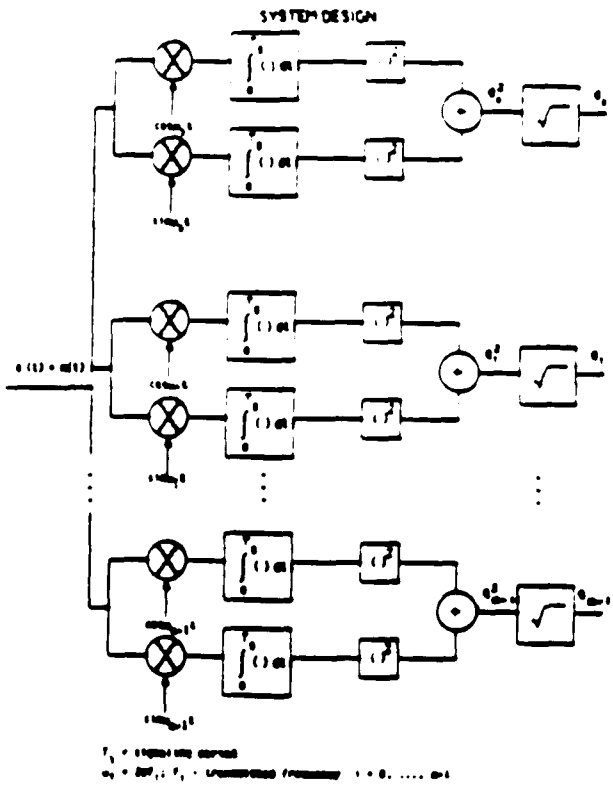


Figure 3.7 Block diagram for a non-coherent detector.

3.3 NON-COHERENT DETECTION OF FSK

In non-coherent detection of signals, the receiver does not exploit the phase of the incoming signal in the decision making process. As shown in figure 3.7, detection of

signals non-coherently, requires two correlators followed by an energy detector for each signal that is transmitted. This can be explained intuitively as follows. At the receiver, the incoming signal has the form of where n(t) is white gaussian noise having a mean of zero and power spectral density of N_o /2. From equation 3.24, it can be observed that the incoming signal partially correlates both $\cos\omega_i t$ and $\sin \omega_i t$. As result, the receiver not knowing anything about the phase of the incoming signal, must resolve the signal into in-phase and quadrature components. Then by integrating, squaring and summing, as shown in figure 3.7, the amount of energy at the output of each correlator due to transmitted signal and noise is determined. Finally, the energy at the output of each bank of correlators are compared, and the branch with the largest energy output will correspond to the transmitted signal. It is said that an error has occurred if the energy at the output of one of the detectors corresponding to noise is larger than the output of the correlator matching the transmitted signal.

3.4 PERFORMANCE OF FSK SIGNALING

Performance of FSK signaling depends on orthogonality of waveforms, as well as the type of detection being done. In the case of non-coherent detections of signals, it was shown that the minimum spacing between two tones that are orthogonal has to be 1/T. However in the case of MSK signaling, since the frequency separation between the two signals is 1/2T, non-coherent detection does not preserve the orthogonality, as oppose to coherent detection, resulting in a degradation in the performance of the system.

In order to investigate the performance of an FSK system using non-coherent detection, the decision variables r_1 and r_2 at the output of detector have the form of

$$r_1 = | E_b + N_1 | \qquad (3.23)$$
$$r_2 = | \rho E b + N_2 |$$

if the signal matched to r_1 was transmitted. Above, the noise components N1 and N2 are gaussian random variables with zero mean and variance $N_o/2$, and ρ represents the correlation coefficient between the two signals. It is given by

$$| \rho | = | \frac{\sin \pi T \Delta f}{\pi T \Delta f} | \qquad (3.24)$$

Since both r_1 and r_2 are the envelopes of gaussian random variables with means of E_b and ρE_b respectively, it can be shown that their probability density function are Rician distributed random variables having the following density function

$$\rho (r_i) = \begin{cases} \frac{r_i}{\eta} e^{-\left(\frac{r_i^2 + \beta_m^2}{\eta}\right)} I_o\left[\frac{2r\beta_m}{\eta}\right] & r_i \geq 0 \\ 0 \quad for & r_i < 0 \end{cases} \qquad 3.27$$

where $\beta_1 = E_b$ and $\beta_2 = \rho E_b$. An error occurs at the receiver whenever the envelope sample due to noise ,r_2, exceeds the envelope sample due to signal plus noise, r_1. Consequently, the probability of error is obtained by integrating $\rho(r_1,r_2)$ with respect to r_2 from r_1 to infinity, and then averaging over all possible values of r_1. That is

$$\int_0^\infty dr_1 \int_{r_1}^\infty \rho(r_1,r_2)\, dr_2 \qquad (3.26)$$

The solution of the above equation is given in terms of the generalized Q function and

is given by

$$P_e = Q(a,b) - \frac{1}{2} e^{(a^2+b^2)/2} I_o(ab) \qquad (3.27)$$

where

$$a = \sqrt{\frac{E_b}{2\eta}(1 - \sqrt{1 - |\rho^2|})}$$

and

$$b = \sqrt{\frac{E_b}{2\eta}(1 + \sqrt{1 - |\rho^2|})}$$

for the case of orthogonal signals, equation 3.29 reduces to

$$P_e = \frac{1}{2} e^{-\left(\frac{E_b}{2\eta}\right)} \qquad 3.30$$

The probability of error versus the signal to noise ratio , Eb/η, is shown in figure 3.8 for the cases $\Delta f = 1/T$ and $\Delta f = 1/2T$. Both the simulated and theoretical results are presented to show the validity of simulated results. In addition, figure 3.9 shows the probability of error for duobinary FM system, and its comparison to MSK signaling, with non-coherent detection.

Both duobinary FM and MSK signaling, are used as the basis of comparison with constrained FM signaling. Specially, the comparison with duobinary FM is of importance since the baseband signal in both systems are tri-ary pulses having similar characteristics.

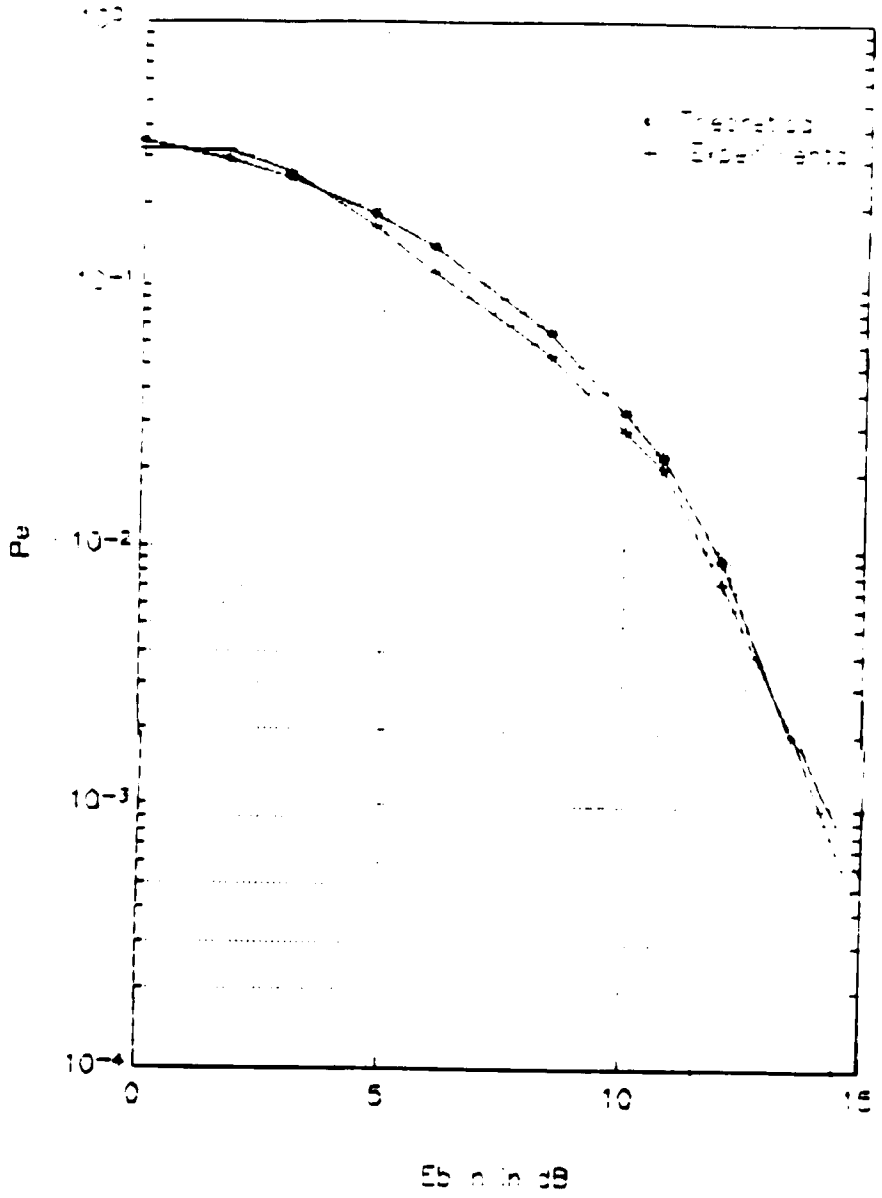


Figure 3.8 Pe vs. SNR in dB for FSK.

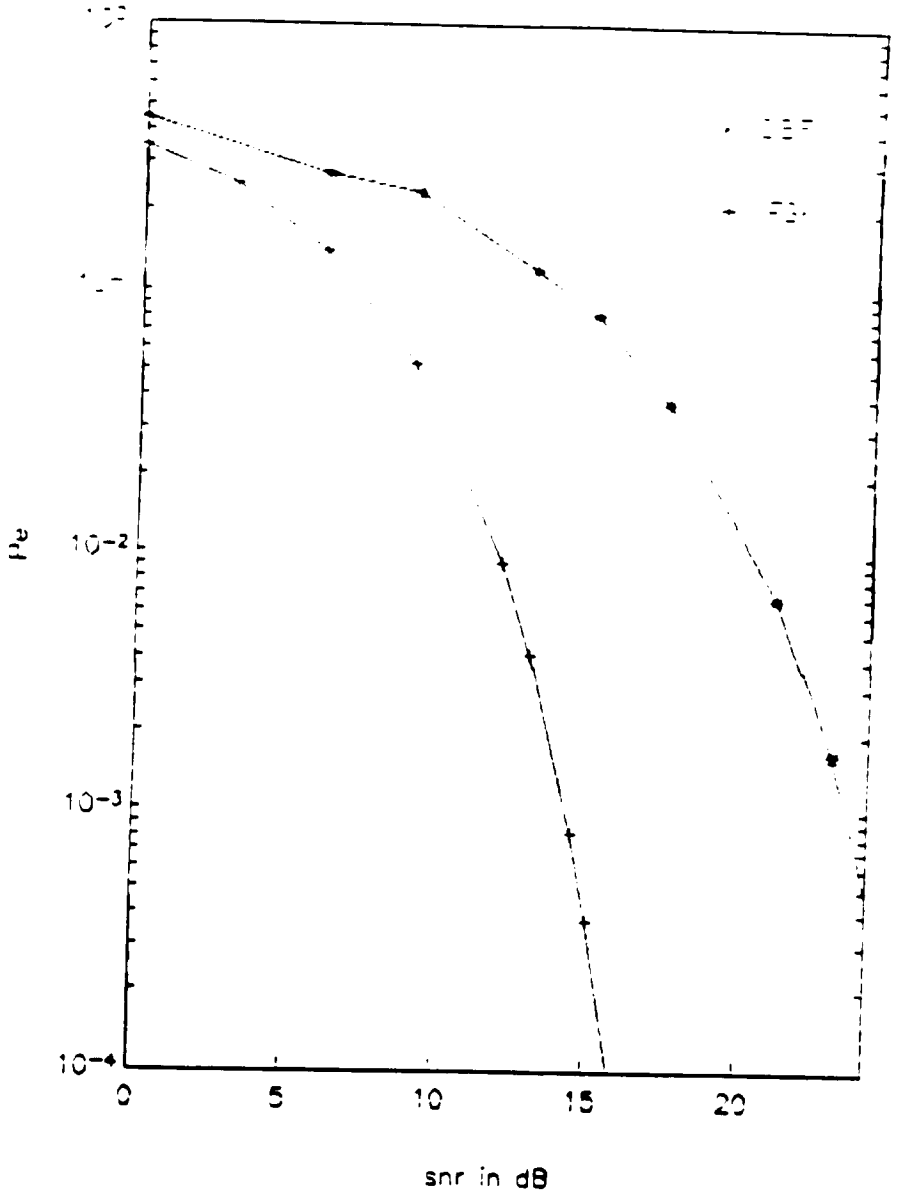


Figure 3.9 Comparison of FSK and DBFM.

CHAPTER 4

CODING

4.0 INTRODUCTION

Digital signals may be encoded to improve the performance of a system, usually at the expense of the bandwidth. The function of the encoder is to take k information digits and map it into n digits. This produces a code of the rate k/n. At the receiver, these n bits are decoded and the original information bits are retrieved. If the mapping at the encoder is done in such a manner that it satisfies certain structure and constraints, these restrictions can be used to correct a number of errors that may have occurred during the transmission of the signal. In general error correcting codes are subdivided into two categories; Block codes and Convolutional codes. Here convolutional and linear block codes, a subset of block codes, are discussed, since linearity reduces the complexity of the encoding process.

4.1 LINEAR BLOCK CODES

Linear block codes are a class of parity check codes that is characterized by an (n,k) notation. At the encoder r parity digits are added to a block of information digits. This produces a codeword having length n, where n=k+r. These parity digits are redundancies appended to the information digits in a controlled manner in order to increase the Hamming distances between codewords.The Hamming distance between two codewords denoted by d(v,w) is the number places in which the two codewords v and w differ. The Hamming distance between two codewords can be described in terms of the Hamming weight, w, of a codeword.Hamming weight is number of non-zero components in a codeword. Thus

$$d(v,w) = d(0,w-v) = w(w-v) = w(c) \tag{4.1}$$

The above equation uses the fact the difference between two codewords is another codeword. The minimum weight of a code w^* is the smallest weight of any non-zero codeword and it determines the number of errors that a code corrects. For a code to correct t errors, one must find a linear code with minimum weight satisfying

$$w^* \geq 2t + 1.$$

4.2 MATRIX DESCRIPTION OF LINEAR CODES

A block of k information digits can be represented by a vector of k tuples, $(u_1,u_2,\ldots,u_k)$, where u_i's represent information digits. Since the information digits are a

member of a finite set, their values can be represented by elements of a finite Galois field called GF(q), where q represents the number of field elements. In the field of GF(q) there exists a k x n matrix called the generator matrix, **G**, such that it maps the information vector **u** under the following operation.

$$C = UG \tag{4.2}$$

where **c** is the n tuple code vector. Since the code vector is a linear combination of the rows of **G**, the rows of **G** are code vectors themselves and form the basis of generator matrix. A desirable property for linear block codes to posses is a systematic structure. That is, the codeword is divided into two parts, the information digits and the parity check digits. Thus, the generator matrix of a systematic linear block code takes the form of

$$G = [\, I_k \mid P_{k\,x\,(n-k)} \,] \tag{4.3}$$

where $\mathbf{I_k}$ and $\mathbf{P_{kx(n-k)}}$ represent a k x k identity matrix and the parity check matrix respectively.

It is well known from the theory of vector spaces and matrices that for any k x n matrix, with k independent rows, there exist a (n - k) x n matrix, **H**, with n - k independent rows, such that for any row $\mathbf{g_i}$ in **G** and $\mathbf{h_j}$ in **H**, satisfies $\mathbf{g_i\, h_j = 0}$. The matrix **H** is called the parity check matrix and is given by

$$H = [\, I_{n-k} \mid -P^T \,] \tag{4.4}$$

From the above it can be observed that the rows of **G** and **H** form the orthogonal

complement of each other. In addition, since every codeword is a linear combination of the rows of **G** then a codeword **c** is also orthogonal to the rows of the parity check matrix. Thus, this can be used as the basis for determining and correcting errors that may have occurred during the transmission of the signal.

4.3 DECODING OF RECEIVED CODEWORD

Let **r** denote a received vector representing the output of a detector, at the receiver, corresponding to detection of the encoded message that has been corrupted by white gaussian noise. As result the received vector may or may not represent the original encoded message. In order to determine if an error has occurred and to correct it, the first step is to define the syndrome **s** of a received vector. That is

$$s = c \cdot H^T \tag{4.5}$$

The syndrome is a zero vector if the transmitted and received vectors are the same. However, if they aren't, this implies that an error has occurred and the received vector can be decomposed as the sum of vectors **c** and **e**, where **e** represents the error vector and has non-zero elements in the positions where **r** differs from **c**. Thus the syndrome may be rewritten as

$$\begin{aligned} s &= r \cdot H^T \\ &= (c + e) \cdot H^T \\ &= c \cdot H^T + e \cdot H^T \\ &= e \cdot H^T \end{aligned} \tag{4.6}$$

From equation 4.6 it can be seen that the syndrome of a received vector corresponds

directly to the error vector. By knowing the syndrome, the decoder can determine the location of the error and its magnitude, provided t or less errors have occurred. One method of accomplishing this is to use a table and determine the error that corresponds to the evaluated syndrome. Then mod-2 add the error to the received signal and the correct code word results. However for large n-k, the implementation of this scheme becomes impractical since the number of stored syndrome is in the order of q^{n-k}. Therefore other decoding methods must be used resulting in additional properties imposed on the code word.

4.4 CYCLIC CODES

Cyclic codes are a subset of linear block codes having the property that a cyclic shift of the code word, produces a code word of a cyclic code. That is if $\mathbf{c}=(c_{n-1}, c_{n-2},\ldots ,c_0)$, then $(c_{n-2}, c_{n-3}, \ldots , c_0, c_{n-1})$ is obtained by a cyclic shift of the elements of $\mathbf{c}$. As a result of this property, the code possesses desirable properties that is exploited in encoding and decoding of cyclic codes.

Cyclic codes can best be dealt with in polynomial form. That is a code word $\mathbf{c}$ of the form $(c_{n-1}, c_{n-2}, \ldots , c_0)$ in polynomial form is given by

$$c(x) = c_{n-1} x^{n-1} + c_{n-2} x^{n-2} + \ldots + c_0 \qquad (4.7)$$

Thus, coefficients of the polynomial correspond to the code word vector $\mathbf{c}$. Now

multiplying c(x) by x^i yields

$$x^i c(x) = c_{n-1} x^{n+i-1} + c_{n-2} x^{n+i-2} + \ldots + c_0 x^i \tag{4.8}$$

which can be written as

$$\begin{aligned} x^i c(x) &= c_{n-i} + c_{n-i+1} x + \ldots + c_{n-1} x^{i-1} + c_0 x^0 i + \ldots + c_{n-i-1} x^{n-1} \\ &\quad + c_{n-i}(x^n - 1) + \ldots + c_{n-1}(x^n - 1) \\ &= c^i(x) + q(x)(x^n - 1) \end{aligned} \tag{4.9}$$

From equation 4.9 it can be observed that the code polynomial c'(x), corresponds to a cyclic shift of order i. As a result of this, an (n,k) cyclic code can easily be generated by means of a generator polynomial **g**(x), if **g**(x) satisfies the following two properties:

1) Degree of **g**(x) is n-k.

2) **g**(x) divides x^n-1.

then **c**(x) can be written as product of **i**(x), the information polynomial, and **g**(x).

$$c(x) = i(x)g(x) \tag{4.10}$$

Since **i**(x) has a degree of k-1, **c**(x) is a polynomial of order n-1. Since **g**(x) divides x^n-1, it can be written as

$$x^n - 1 = g(x)h(x) \tag{4.11}$$

where **h**(x) is called the parity check polynomial, used to detect and correct the receiver errors. Treating the received vector **r** as a polynomial of degree n-1 or less,

$$r(x) = r_{n-1} x^{n-1} + r_{n-2} x^{n-2} + \ldots + r_0 \tag{4.12}$$

then by multiplying **r**(x) by **h**(x) mod (x^n-1), the syndrome polynomial is obtained. Thus,

$$s(x) = r(x)h(x) \bmod (x^n - 1) \tag{4.13}$$

if no error has occurred during transmission, then

$$r(x) = c(x) = i(x)g(x) \tag{4.14}$$

and s(x) mod (x^n-1)=0.

However if an error occurs, s(x) is no longer zero which can be used to correct errors if t or fewer errors occur.

4.5 REED-SOLOMON CODES

Reed-Solomon codes are a subclass of cyclic codes. They are non-binary codes with symbols from a Galois Field GF(q). Usually $q=2^k$, so k information bits are mapped to form one of the q symbols. A t error correcting Reed-Solomon code with symbols from GF(q) has the following properties.

Block length:	$n=q-1=2^k-1$
Number of parity digits:	$n-k=2t$
Minimum distance:	$d_{min}=2t+1$

From above it can be observed the length of the code is one less than the size of the symbols and the minimum distance is one greater than the number of parity -check digits. Thus one of the reason for their importance is their good distance properties. A second reason for their importance is the existence of efficient hard -decision algorithms, such as the Berlekamp[9,10], algorithm which allows the implementation of relatively long

codes in practical applications. Reed-Solomon codes are specially effective in channels where burst of errors occur. Reed-Solomon codes correct up to t error symbols and each symbol is k binary bits, if the errors are consecutive or occur in a burst, then up to kt bits of error can be corrected. On the other hand if errors occur at random, such as each symbol contains only one error, then Reed-Solomon codes can correct up to t bit errors.

4.6 CONVOLUTIONAL CODES

Convolutional codes, introduced as an alternative to block codes, are discussed here, since a convolutional encoder is utilized as part of an encoder for trellis coded modulation codes which are used for bandwidth constrained channels.

Convolutional codes differ from block codes in that the encoder contains memory, and therefore, the encoders output not only depends on the present input but also depends on the previous inputs as well. In general a convolutional code is generated by passing the information sequence to be transmitted through a K stage register and then combining their outputs by using n exclusive-or gates,or modulo two adder. The input to the encoder is assumed to be binary, and since for k input bits, there are n output bits, the rate of the code, R_c, is k/n.

There are three different methods to describe convolutional codes. These are the tree diagram, trellis diagram, and the state diagram. Here the trellis diagram approach, in the context of an example, is used to describe the convolutional code. A rate 1/3 convolutional code and its trellis diagram are shown in figures 4.1 and 4.2 respectively. From figure 4.1, it can be observed that the present output is determined by the present

input and the previous inputs contained in the first two stages of the shift registers, since the contents of the second shift register is lost when the new input bit enters. Thus, the code has four states of a=00, b=01, c=10, and d=11 as shown in figure 4.2. In drawing figure 4.2, the convention is that a solid line corresponds to the output as result of input being zero and the dotted line corresponds to output as result of input being one. It can also be observed that after three inputs the branches may merge. This is because the structure repeats itself after the third stage, which is consistent with the number of shift registers, or memory, being used.

Since the structure of the code repeats itself after three stages, the free distance of the code, d_{free}, to distinguish from the Hamming distance, is said to be at least three, and the code is capable of correcting one error. The free distance plays an important role in convolutional codes, since by increasing d_{free}, the code can correct more errors. From Larson[11], it can seen that, increasing d_{free} corresponds to increasing the number of delays of shift registers, thus increasing the complexity of the code. In addition, use of convolutional codes increases the bandwidth of the transmitted signal by the factor of $1/R_c$.

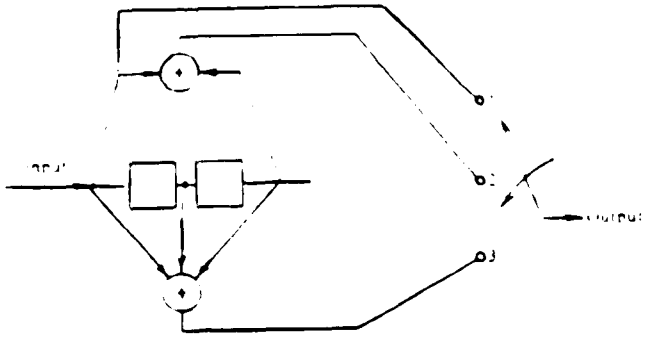


Figure 4.1 Rate 1/3 convolutional encoder.

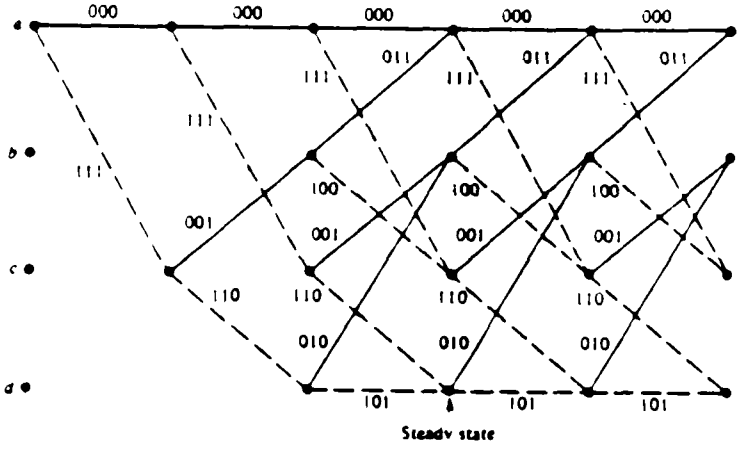


Figure 4.2 Trellis diagram for the rate 1/3.

4.7 TRELLIS-CODED MODULATION

The performance of different modulation schemes, depends upon the Euclidean distance between the signals. In linear block codes or convolutional codes, the modulation is treated separately from the coding process. However in trellis coded modulation , the modulation is treated as part of the encoding process. In this scheme first proposed by Ungerboeck[12,13], the number of signal points are increased, thus reducing the Euclidean distance between the signals; but, in the coded signal, the minimum Euclidean distance increases, and the loss from expansion of the signal set is easily overcome by a notable coding gain.

The encoding of the signal is done on the basis of mapping by set partitioning. For instance, if an uncoded four-phase PSK is used as the basis for which the performance of the system is measured against, it can be shown that the coded eight-phase PSK can provide a gain of 3.6 dB. For the above example, the partitioning of eight-phase PSK is shown in Figure 4.3. It should be noted that the degree to which the signals are subdivided, depends on the attributes of the code.

The encoding process is done by taking a group of k information bits to be subdivided to two groups of lengths k1 and k2. By the means of a convolutional encoder, the k1 bits are mapped into n bits, while the k2 bits are uncoded.Then each of the signal set is assigned to one of the possible 2^n combinations. Then the k2 bits are used as the possible signal points in each subset. Obviously when k2=0, all of the information bits are encoded. The rules of bit to symbol mapping which is based on the method of set partitioning is given by:

(1) All parallel transitions are separated by the maximum Euclidean distance.

(2) All transition emanating from or converging into a state should have the next maximum Euclidean distance separation.

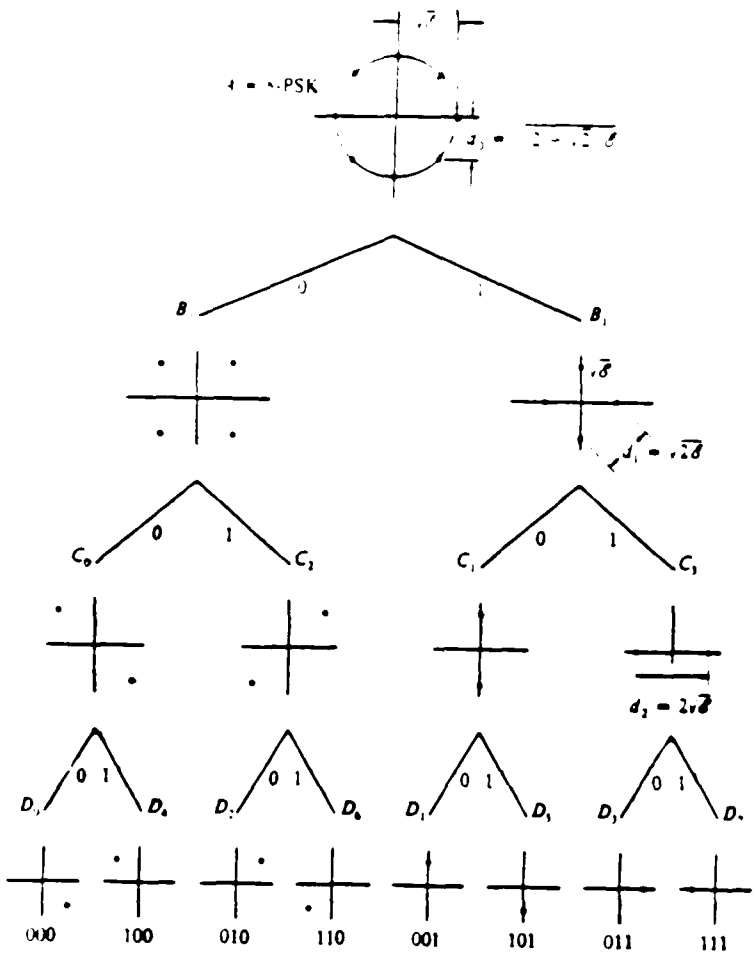
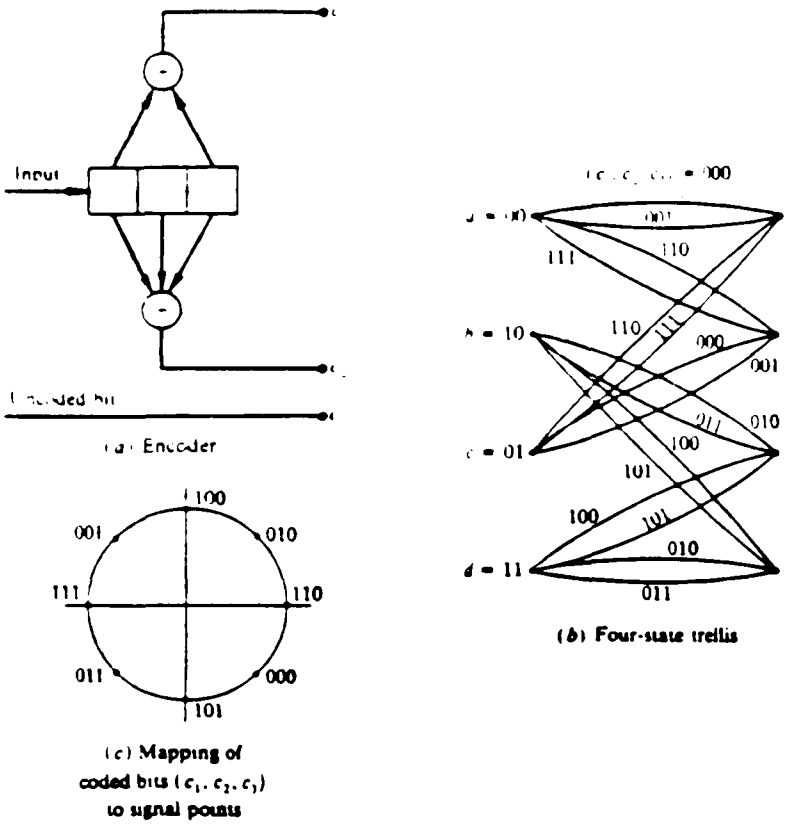


Figure 4.3 Set partitioning of an 8-PSK.



(a) Encoder

(b) Four-state trellis

(c) Mapping of coded bits (c_1, c_2, c_3) to signal points

Figure 4.4 Four-state trellis coded 8-PSK modulation.

An example of a trellis encoder, in which it changes a four-phase PSK to eight-phase PSK is shown in figure 4.4. This 4 state trellis is optimum in the sense that, it provides the largest Euclidean distance for a four state trellis.

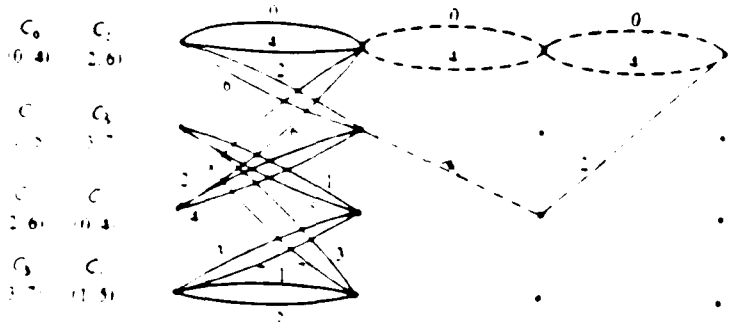


Figure 4.5 Four state trellis.

The decoding of trellis coded modulation as well as convolutional codes are done via Viterbi decoding algorithms. The algorithm is basically a maximum likelihood receiver; however it reduces the computational burden by taking advantage of the structure in the trellis. The algorithm involves computing the distance between the received signal at time t_k, and all the trellis paths entering each state at time t_k. By keeping track of these metrics, the decoder chooses the path with maximum metric or minimum distance, when two or more paths merge at a node, as shown in figure 4.5. There is a surviving path for each state. At the end of transmission, the system is usually forced to go into the zero state and the path with maximum likelihood is selected.

CHAPTER 5

CONSTRAINED FM

INTRODUCTION

Both MSK and Duobinary FM signaling offer spectral efficiency. In MSK, the spectral efficiency is mainly caused by keeping the frequency deviation, Δf, small, which causes most of the modulated signal's energy to be within 75 percent of the bit rate of modulating signal as shown in figure 5.1. However, bandwidth of an FM signal is a function of the frequency deviation and bandwidth of the modulating signal. Thus by controlling both the frequency deviation and the bandwidth of modulating waveform simultaneously, the frequency of modulated signal can acquire significant spectral efficiency. This concept is utilized in the generation of ***constrained FM***.

GENERATION OF CONSTRAINED FM

In this technique, in order to concurrently control both the bandwidth of the modulating waveform and its frequency deviation, the ratio of the maximum to minimum

phase that the modulated FM signal can attain has to be bounded. This restriction implies that the maximum and the minimum phase modulo 2π, that the modulated waveform can attain is bounded, hence acquiring the name ***constrained FM (CNFM)***.

The limiting of the phase can result in two different techniques of generating the CNFM. For the first method, suppose that the maximum and the minimum phase that the modulated signal can have is $\pm\pi$. The result of bounding the signal is to encode the baseband signal. In general, the data bit stream can be represented by

$$d(t) = \sum_n A_n u(t - nT) \tag{5.1}$$

where A_n is the amplitude of the modulating signal, taking on the values of ± 1, and u(t) is a rectangular waveform having an amplitude of 1/T, with a duration of T seconds. The equation of the frequency modulated signal by d(t) was given in equation 3.4. That is

$$v(t) = A \cos(2\pi f_c t + \theta_n + \pi h A_n q(t - nT)) \tag{5.2}$$

where

$$h = 2Tf_d$$

$$\theta_n = \pi h \sum_{k=-\infty}^{n} A_k$$

and

$$q(t) = \begin{cases} 0 & for\ t<0 \\ \frac{t}{T} & 0 \le t \le T \\ 1 & t>T \end{cases}$$

The terms h, and θ_n represent the modulation index and the accumulation of the phases of all symbols up to (n-1)T. Obviously, if θ_n is bounded, then the data stream has to be

encoded according to the following rule such that θ_n does not exceed the bound. That is

$$b(t) = \begin{cases} d(t) & |\theta_n| < \pi \\ 0 & \theta_n = \pm\pi \end{cases} \tag{5.3}$$

Then the modulated signal becomes

$$v(t) = A \cos\left(2\pi f_c t + \theta_n + \pi h B_n q(t - nT)\right) \tag{5.4}$$

where

$$\theta_n = \pi h \sum_{k=-\infty}^{n} B_n \tag{5.5}$$

The term B_n represents the amplitude of the encoded baseband signal b(t), and it takes on the values of A_n when $|\theta_n| < \pi$ and 0 for $\theta_n = \pm\pi$. From equations 5.3 and 5.4, it can be seen that the instantaneous frequency of the modulated signal is f_c when $B_n = 0$. Thus by choosing f_c properly, such that f_cT is an integer, the phase of the modulated waveform does not change during the bit interval that B_n was zero. Once f_c is the detected frequency at the receiver, its mapping onto d(t) can easily be accomplished by assigning f_c to the value of +1 if θ_n is positive and to value of -1 if θ_n is negative. However, when the modulation index,h, is .5, the spectrum of the modulated signal resembles that of an MSK, or for h=.25, the spectrum takes after that of a CPFSK, with $\Delta fT = .25$ as shown in figures 5.1 and 5.2. This is due to the fact that for h being small, the probability that data stream contains several consecutive ± 1 such that, the phase of the modulated waveform reaches either of the boundaries is small. As result, the modulated waveform is similar to that of a FSK, and therefore having comparable spectrums. Due to this, there is on gain in spectral efficiency. Hence, the analysis of this technique was not

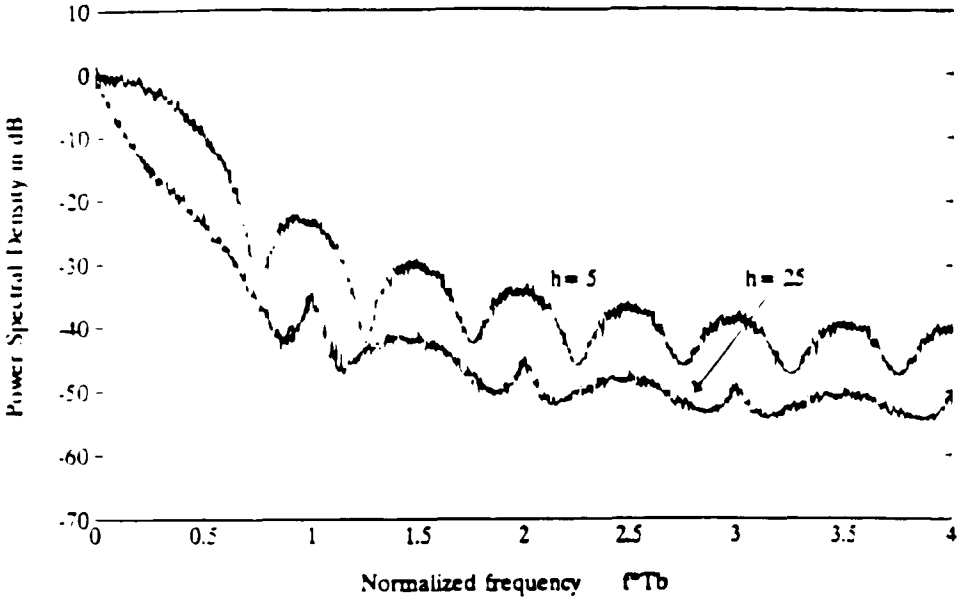


Figure 5.1 Power spectral density of CPFSK for h=1/2 and h=1/4.

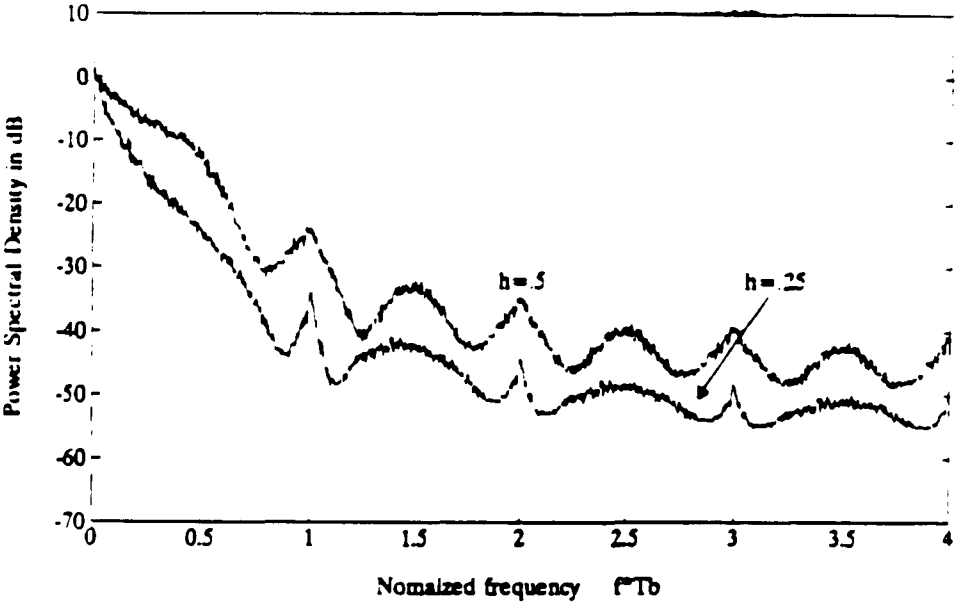


Figure 5.2 Power spectral density of Modified CNFM for h=1/2 and 1/4.

pursuit any further in favor of the second technique, which is spectrally more efficient.

In the second method, the phase of the modulated waveform is polarized. That is, any change in the phase of the modulated FM signal can only be unidirectional. This does not imply that the phase of the FM signal has to be either positive or negative at all times, it merely states that during a certain time interval, the phase of the modulated signal can only increase, then for another period of time it can only decrease, and so forth. The determining factor in the switching time where the direction of the phase change alters are the maximum and minimum bounds that the phase can reach. Again, in order to accomplish this, the baseband signal has to be encoded as shown in figure 5.3, which is the representation of a CNFM transmitter.

The baseband encoder, encodes the data bit stream, d(t), according to the following rule:

$$b(t) = (-1)^k \cos\left(\frac{d(t)\pi + \pi}{4}\right) \tag{5.6}$$

where k is a constant taking on the value of one or negative one, depending on the past history of the phase of the modulated waveform. From equation 5.6 it can be observed that, b(t) is a tri-ary pulse, taking on the values of ± 1, and zero. For as long as the value of k=0, since d(t) takes on the value of ± 1, d(t) is mapped onto zero and one. However, once the value of k=-1, then d(t) maps onto zero and negative one. Referring to these values as B_n and noting that B_n can only take on two values at a time, zero and one or of zero and negative one, it can be seen from equation 5.4 that the phase of the modulated waveform can only increase or decrease at a time. Then by keeping track of the phase of the modulated waveform by the phase monitor, as shown in figure 5.3, the

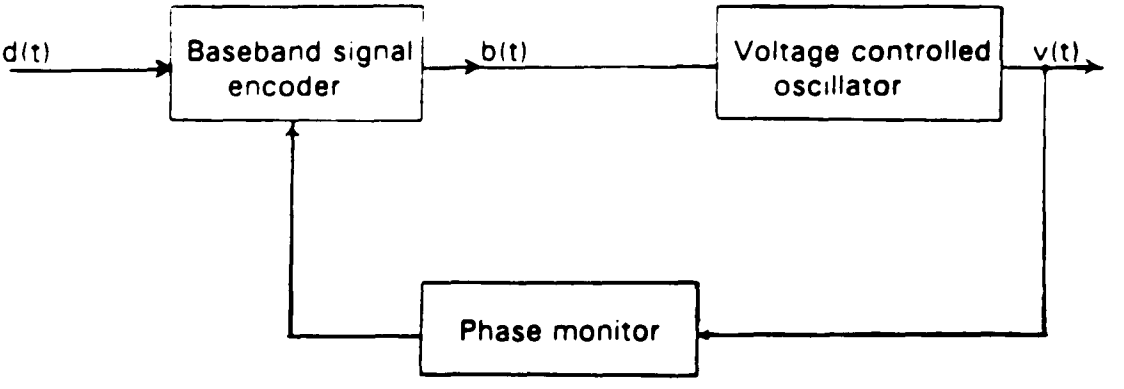


Figure 5.3 Block diagram for a CNFM transmitter.

encoder changes the value of k whenever the phase of the modulated waveform reaches either of the boundaries. Figure 5.4 represents the phase trellis for a CNFM signal.The maximum and minimum phase that the signal can acquire is bounded to $\pm\pi$. The Solid lines conform to d(t)=-1, which maps onto b(t)=$\pm$1, altering the phase by $\pm\pi/2$, corresponding to h=0.25. The broken lines correspond to d(t)=1, or b(t)=0, where the phase remains unchanged.

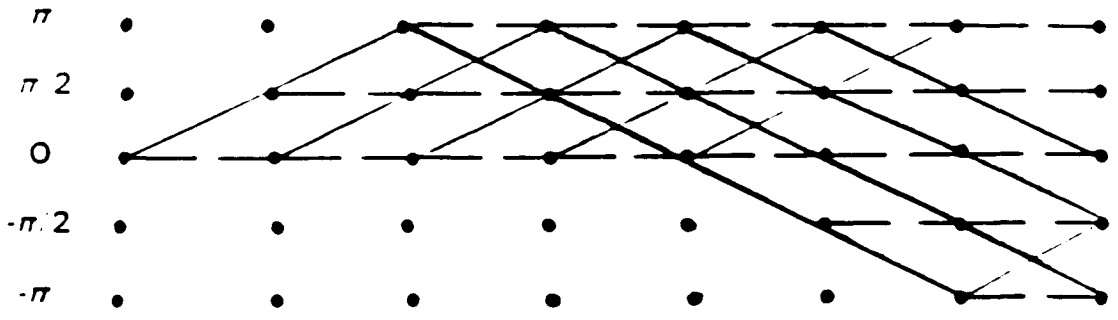


Figure 5.4 The phase trellis for a CNFM for h=0.5.

5.2 ON THE PERFORMANCE OF CNFM

As stated previously, the bandwidth of a FM signal depended on the bandwidth of the modulating signal as well as the frequency deviation of the FM waveform. Since the d(t) has a period of 2T, as shown in figure 5.5, then it is reasonable to assume that the maximum frequency, f_m, it contains is 1/2T. Thus by being able to increase symbol

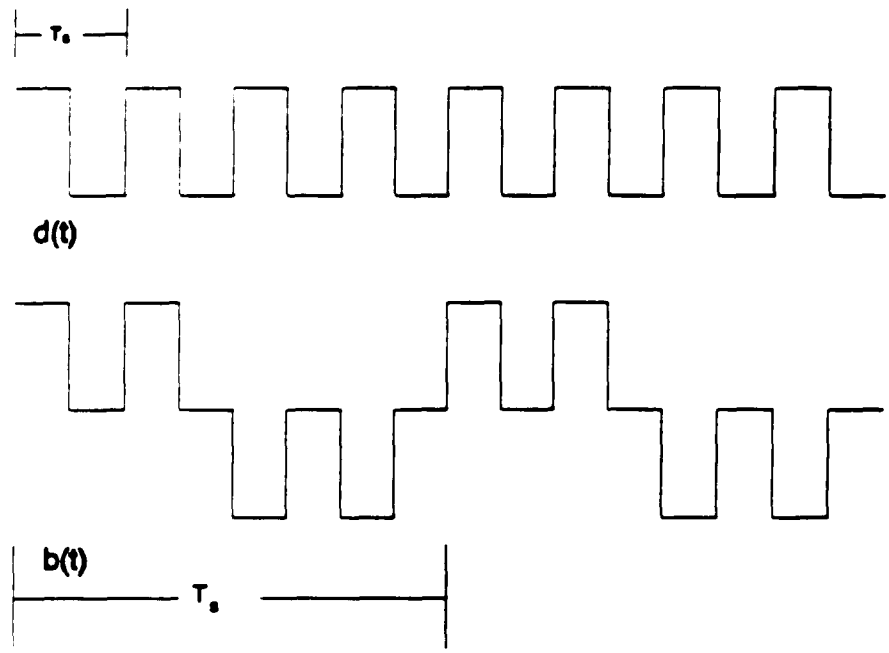


Figure 5.5 The modulating waveform for a CNFM System.

period, T_s, f_m can be decreased. For the modulating waveform of CNFM, b(t), the symbol rate depends on the following two factors:

1) The maximum and minimum value that modulated phase can attain.

2) The modulation index.

By having the minimum and maximum value that the modulated phase can attain set at a certain value, then by decreasing h, the symbol rate can be decreased by a factor of n, thus reducing f_m. The value of n is given by

$$n = \frac{| \theta_{max} | + | \theta_{min} |}{h} \tag{5.7}$$

where θ_{max} and θ_{min} represent the maximum and minimum value that the phase can attain. In figure 5.5, where θ_{max} and θ_{min} are limited to $\pm\pi$, and h=0.5, the symbol rate is reduced by a factor of 4. Next, it has to be determined for what value of θ_{max} and θ_{min}, the spectrum is spectrally more efficient. Through an exhaustive search via computer simulation, it has been determined that the choosing of θ_{max} and θ_{min} have to conform to the following rules.

1) The ratio of the θ_{max} to θ_{min} has to be an integer.

2) The values of θ_{max} and θ_{min} have to be multiples of π.

By choosing the θ_{max} and θ_{min} other $\pm\pi$, such as $\pm 2\pi$, etc. the symbol rate can be reduced . However, by doing so, the energy contained within the sidelobes of baseband signal is such that it results in no additional spectral efficiency. Hence, for this reason the values of θ_{max} and θ_{min} were chosen to be at $\pm\pi$. For these values of θ_{max} and θ_{min}, the power spectral density of CNFM is plotted in figures 5.6 to 5.9.

From those figures it can be seen that the spectrums contain horn like peaks. The

spectrum of CPFSK signals have been studied thoroughly. It has been determined by Rice and Bennet that these horn like peaks occur in CPFSK signals as long as there is a relationship between the harmonics of the frequency shift and the signaling rate, as in our case. However there are certain exceptions such as the case of MSK signaling. As h or f_d decreases, the result is that the spectrum becomes significantly narrow. Obviously, at the expense of the bit error rate. For the case where h=.5, which is of practical interest, the CNFM signal contains over 99 percent of its energy within the band of f_b Hz. where as MSK signaling has only about 90% of its energy within the f_b Hz. respectively. Thus making CNFM spectrally more efficient. In comparing CNFM with DBFM it should be noted that they both exhibit similar characteristics, since they both have the same error performance and both contain 99% of their energy within f_b Hz. Figures 5.10 and 5.11 show a comparison in spectral efficiency and probability of error versus signal to noise ratio for the above systems when h=1/2. It is also worthwhile to note that the 99% bandwidth occurs at $0.9f_b$, $0.32f_b$ and $0.175f_b$ for h taking on the values of 1, 1/4, 1/8 respectively. From the probability of error view point, it can be observed from figure 5.12 that the error rate increase as h decreases, since by decreasing h, the Euclidian distance between the signal sets decreases.

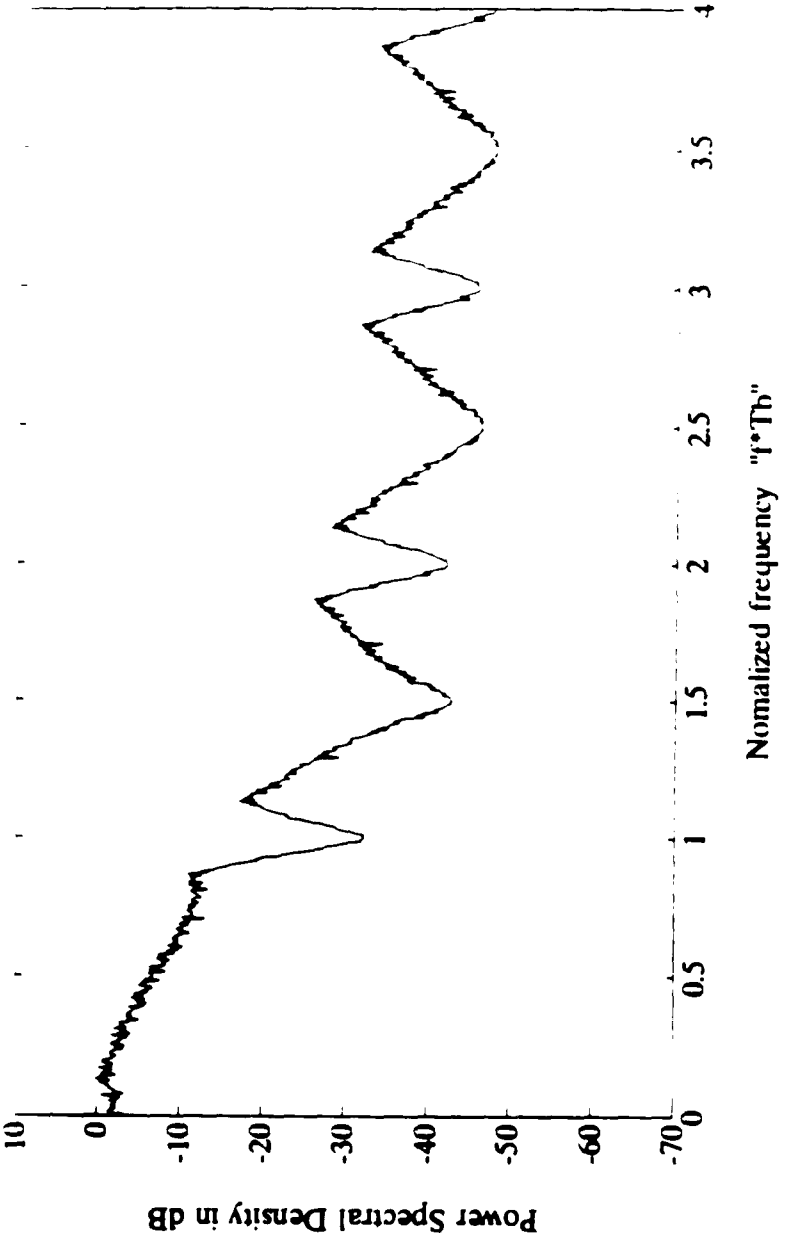


Figure 5.6 Power spectral density of CNFM for h=1.0.

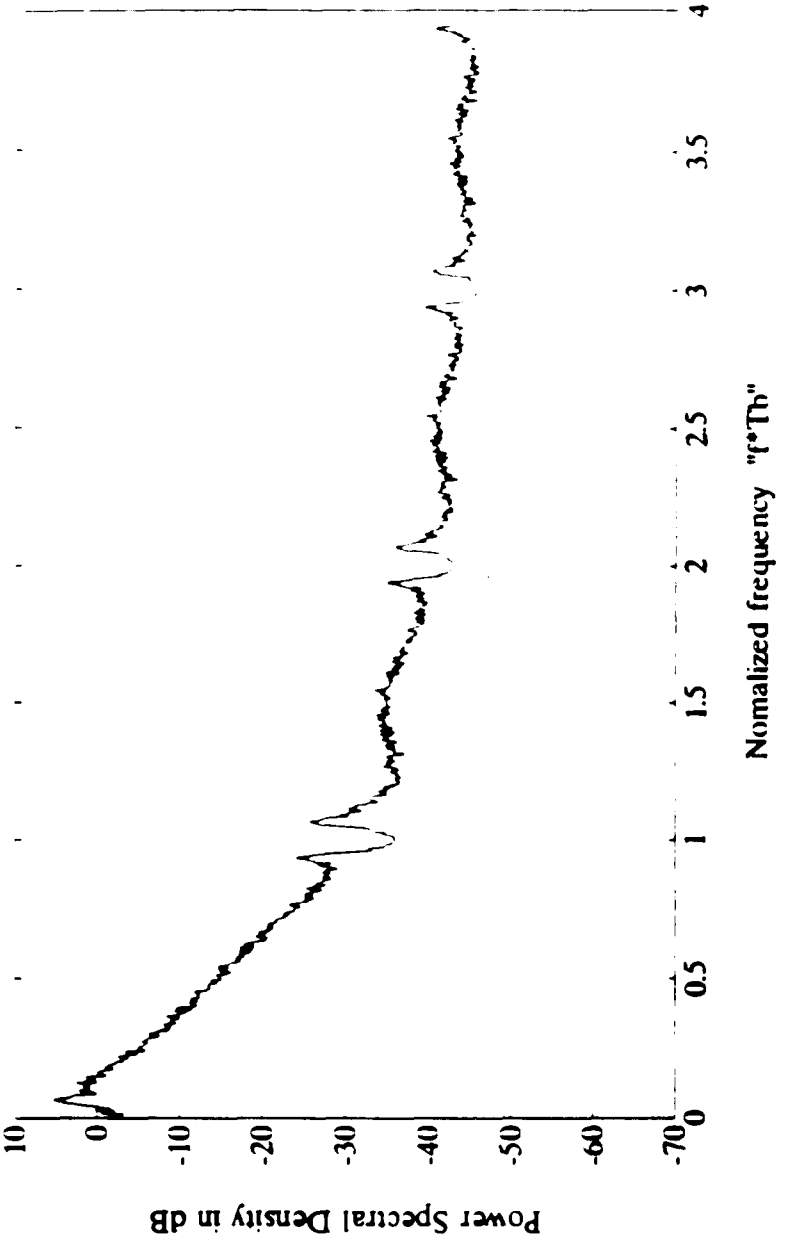


Figure 5.7 Power spectral density of CNFM for h=0.5.

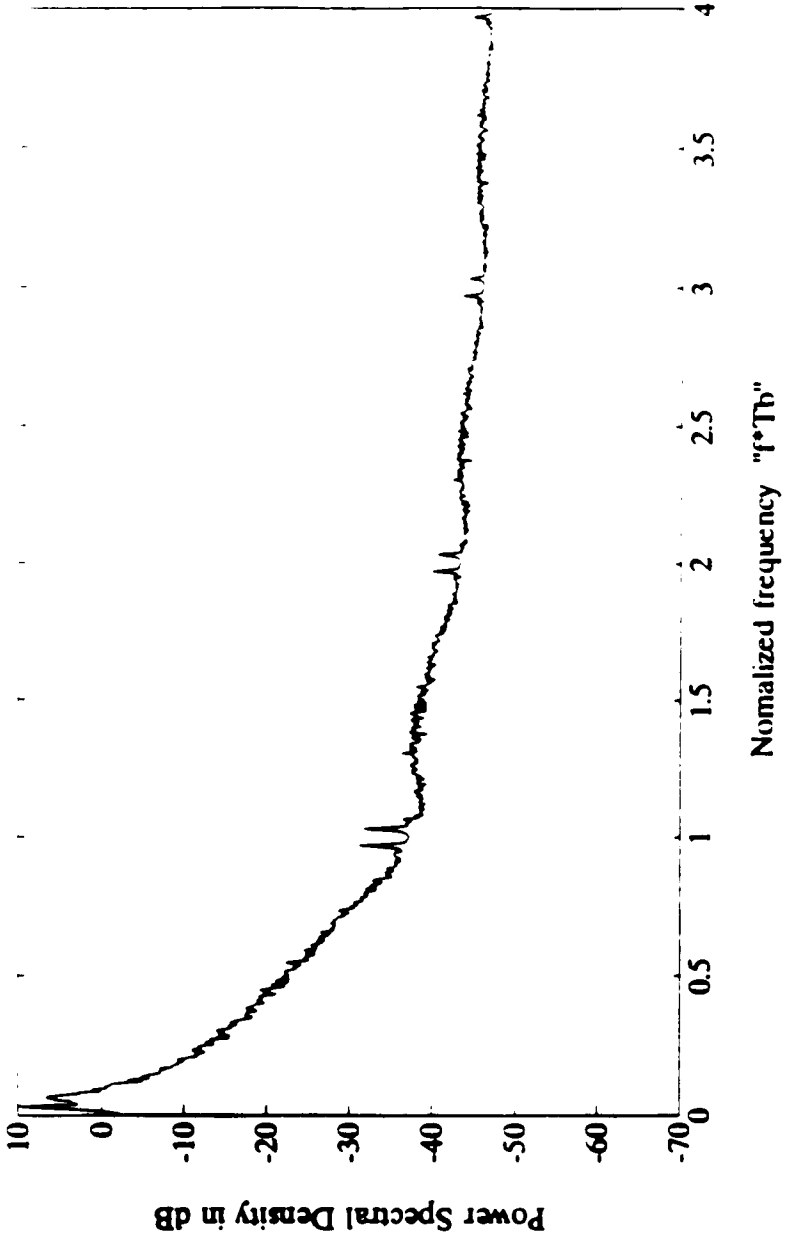


Figure 5.8 Power spectral density of CNFM for h=0.25.

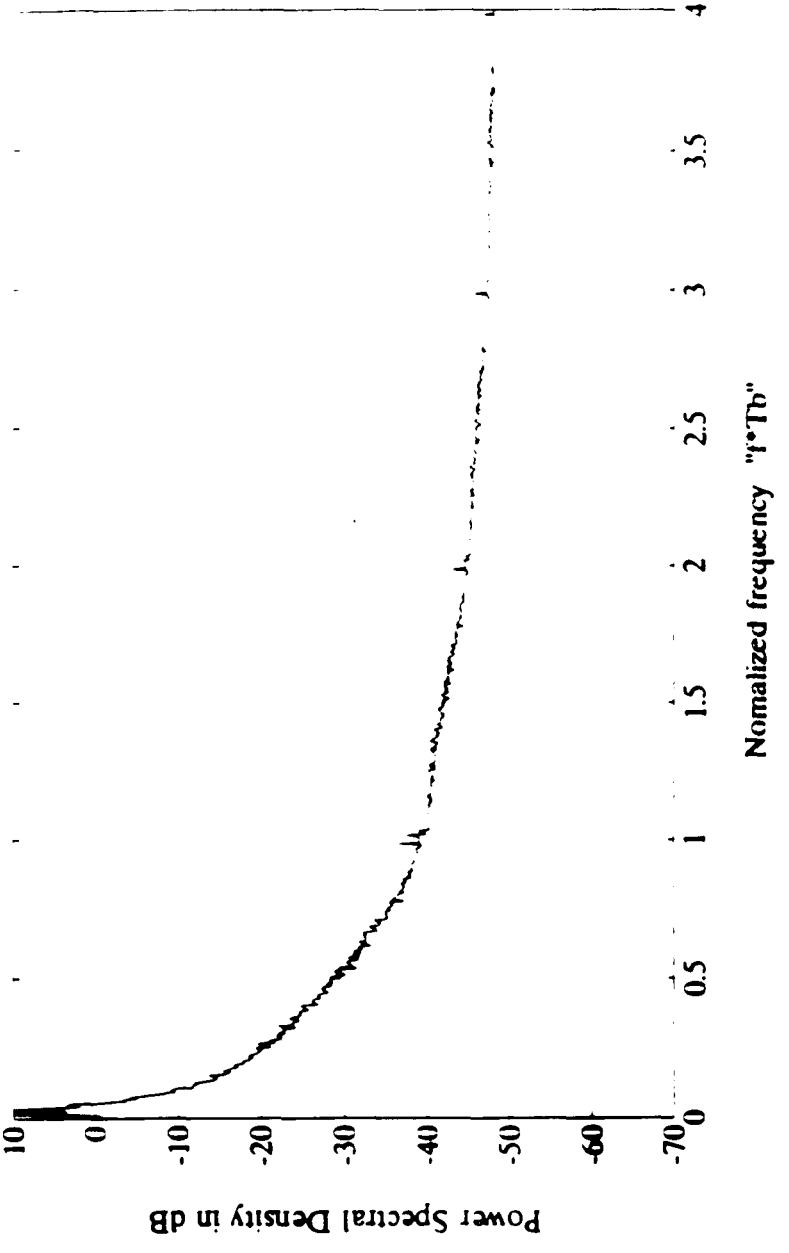


Figure 5.9 Power spectral density of CNFM for h=1/8.

5.3 TRELLIS CODED CNFM

As it was mentioned, the effect of keeping the ratio of the maximum to minimum phase bounded, was to polarize the way in which the phase of the modulated waveform can change its value. For a given h, the phase of the modulated waveform can change by πhB modulo 2π. Since B takes on the value of ± 1 or zero, the maximum phase change in T seconds is $\pm \pi h$. Since the maximum and minimum phase that the modulated waveform can attain is bounded by $\pm \pi$, there are 2/h phase states that the phase of the modulated waveform can have for the phase to change from its minimum value to its maximum value and vice versa. Since the phase change is polarized, each value that the phase can attain at nT, where n is an integer , corresponds to two states. Hence there exists 4/h states in the structure of CNFM. This trellis structure is shown in figure 5.13. The positive sign written as a superscript is used to denote the value of phase at nT is increasing whereas the negative superscript implies that the phase is decreasing in value.

From figure 5.13, it can be observed that the minimum *free distance* in this code structure is two. This suggests that not only the code is capable of detecting one error but also if the trellis structure is used in the decoding process, there is a possibility of some coding gain.

The error detecting capability of the code can be explained as follows. For an error to occur B_n must change value from 0 to ± 1 or vice versa. As result the phase of received signal either exceeds the bounds or switches direction prematurely. Thus the receiver simply by monitoring the phase of the incoming signal, can determine if an

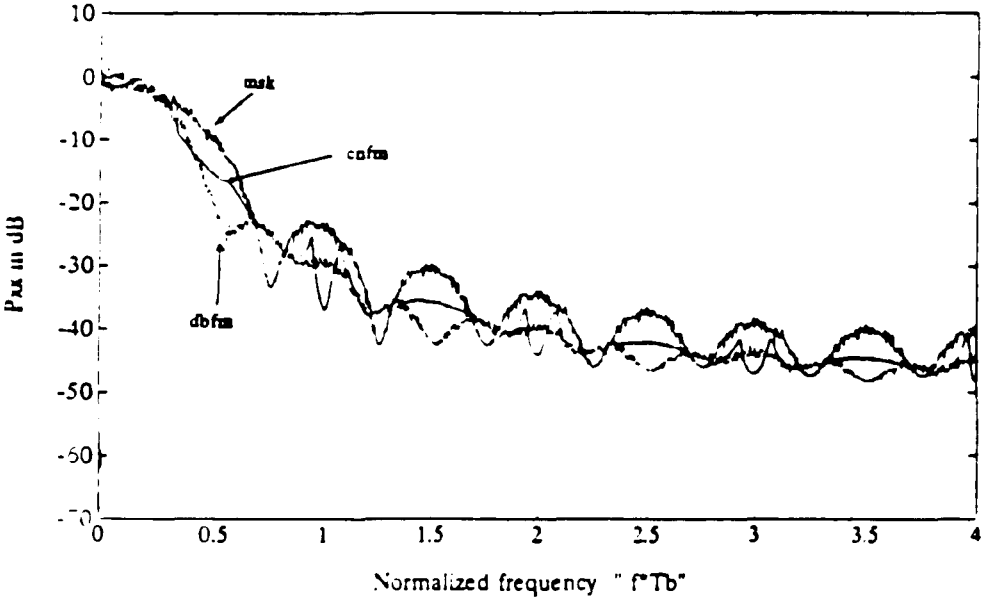


Figure 5.10 A comparison in power spectral density for MSK, DBFM and CNFM for h=1/2.

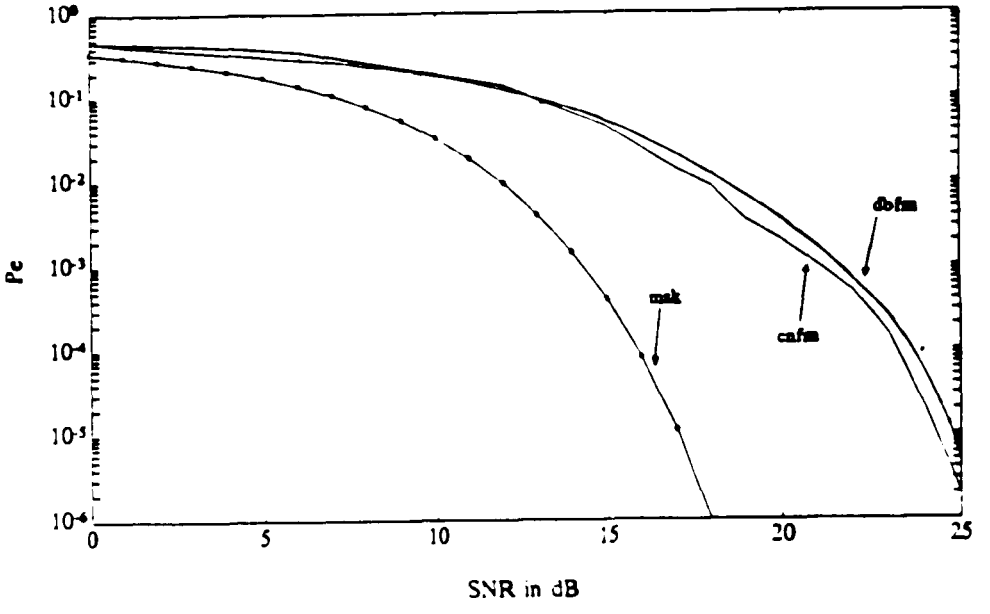


Figure 5.11 Probability of Error vs. signal to noise ratio for MSK, DBFM and CNFM for h=1/2.

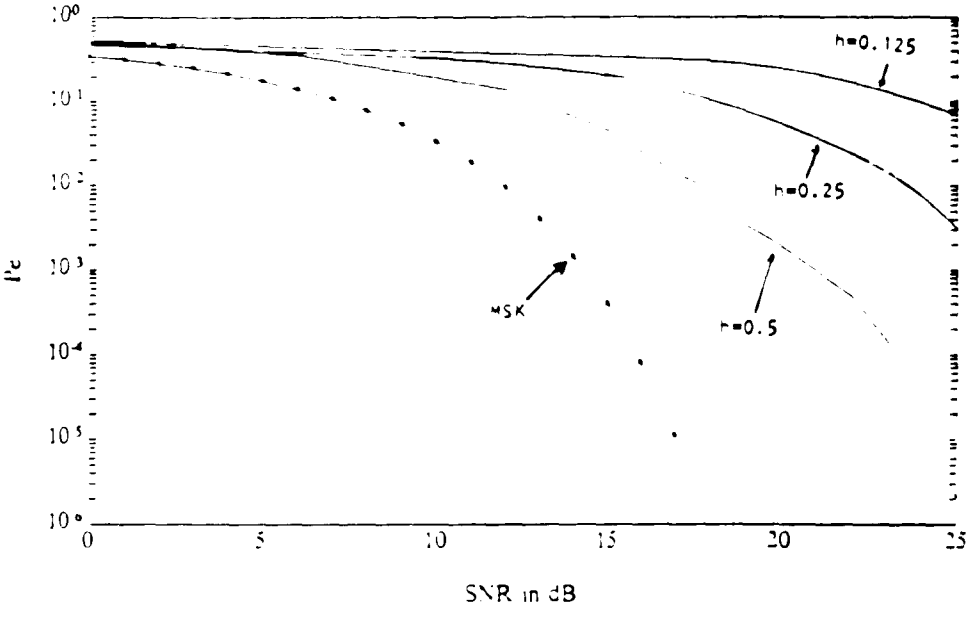


Figure 5.12 Probability of error for CNFM for h=0.5, 0.25 and 0.125

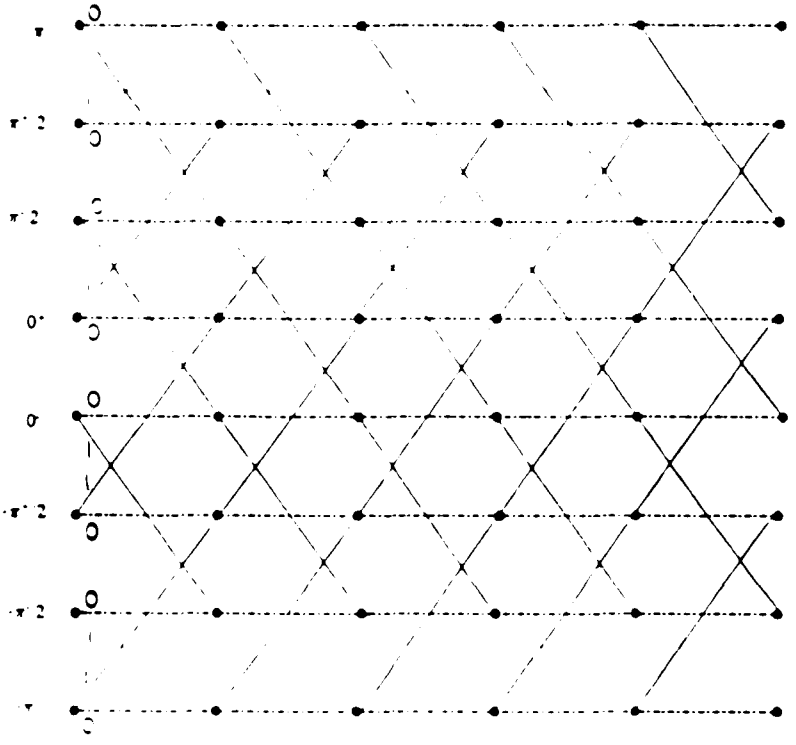


Figure 5.13 Trellis structure for the CNFM signal.

error has occurred. However, since the system does not know where the error occurs, it can not correct it.

The coding gain in utilizing the structure of the trellis in decoding process is shown in figure 5.14. However, this gain may not be large enough to justify complex decoding algorithm. As result other trellis codes that adhere to the concept of CNFM have been developed. These trellis codes are shown in figure 5.15 through 5.18. From the power spectral density of the above codes, as shown in figure 5.19-5.22, it can be observed that the spectrum widens. This spectrum widening is due to increase in maximum phase deviation for increasing the free distance in a code. For the four and eight state trellis CNFM, this corresponds to maximum phase deviation of 2π and π radian respectively. The sixteen and the uncoded CNFM signals, both have a maximum phase deviation of $\pi/2$ radian. Result of increasing the maximum phase deviation is a multi-amplitude baseband signal. This can be explained by noting that since the phase deviation is given by $2\pi f_d T B_n$, for a fixed frequency deviation, f_d, B_n must take on several distinct levels resulting in a multi-amplitude baseband signal. such a signal is shown in figure 5.24.

As stated previously, the increase in free distance results in an increase in bandwidth necessary for transmitting the signals. For these systems the 99% bandwidth is $1.62f_b$, $0.9f_b$, $0.85f_b$, and $0.5f_b$ Hz. for the four, eight, modified eight, and the sixteen states trellis coded CNFM signals as compared to $0.5f_b$ Hz. for the CNFM signaling.

Another measure of comparison is the error rate for the above systems. Figure 5.25 shows the probability of error versus signal to noise ratio all of the aforementioned systems. From this figure it can be observed that although the sixteen state trellis CNFM

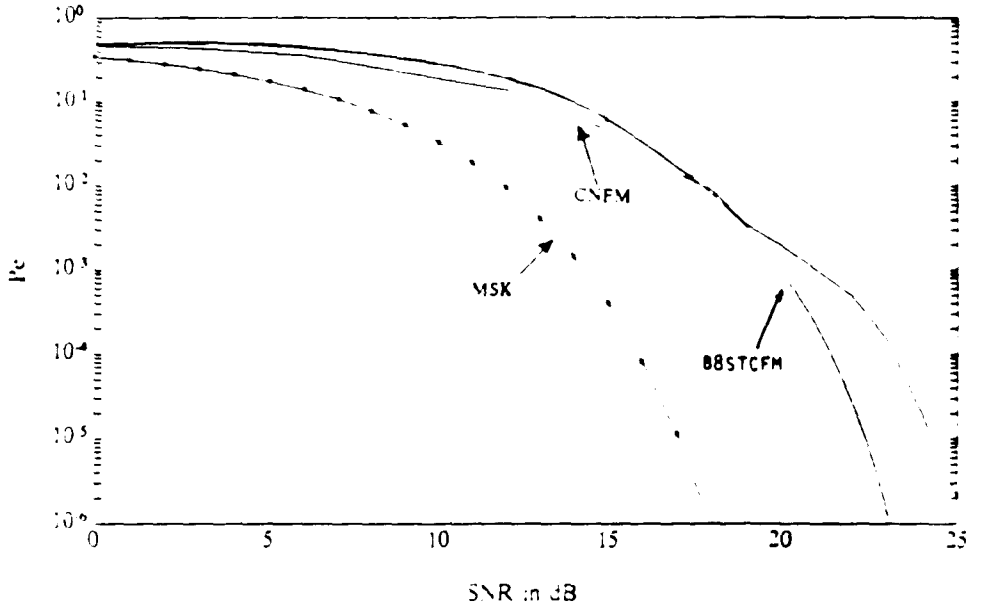


Figure 5.14 A comparison in probability of error between a CNFM signal being decoded conventionally or by utilizing the trellis structure in decoding process.

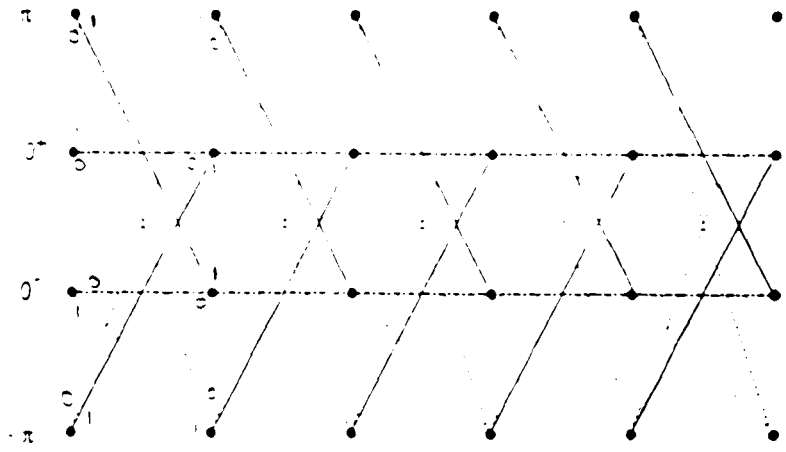


Figure 5.15 A four state trellis for a CNFM system.

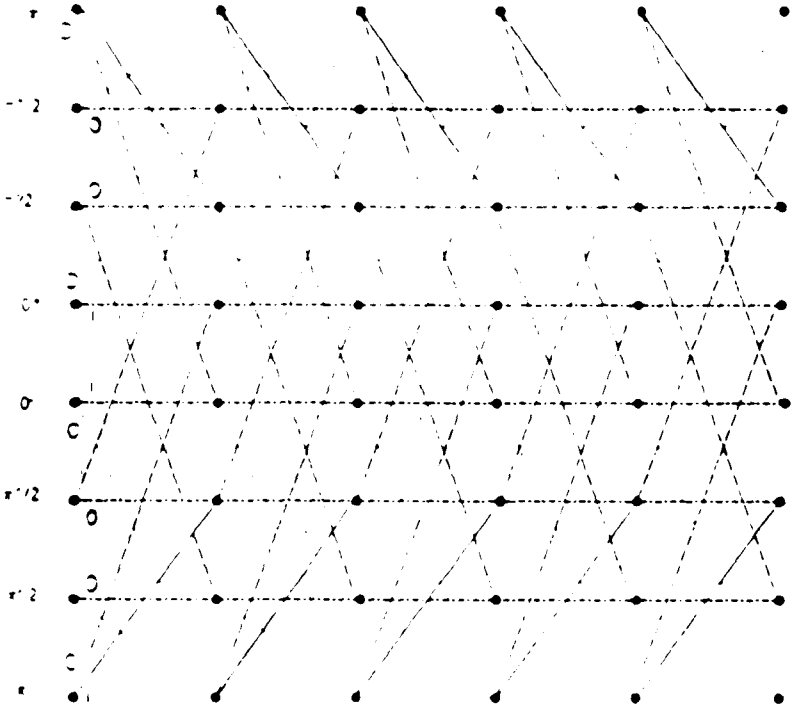


Figure 5.16 An eight state trellis for a CNFM system.

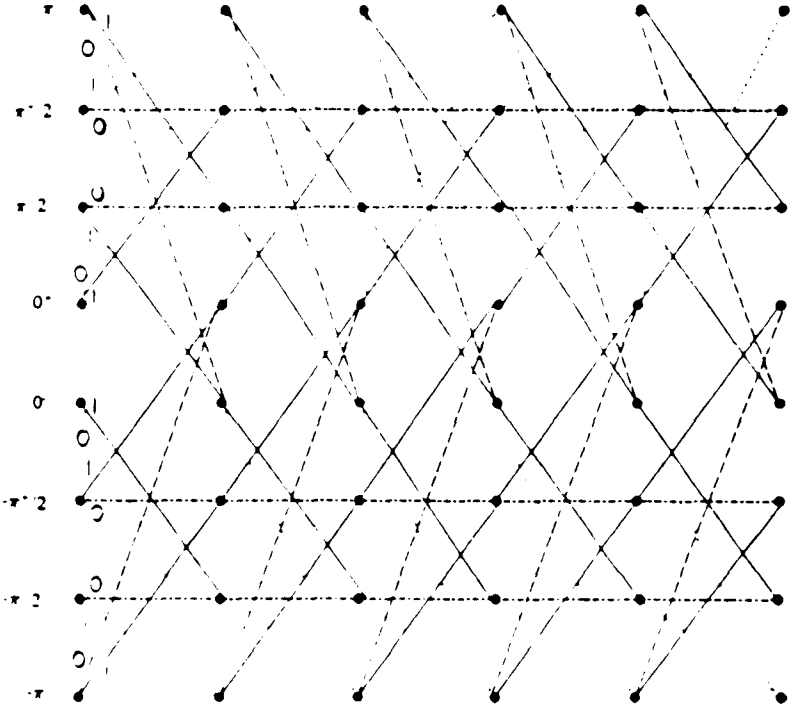


Figure 5.17 Another eight state trellis for a CNFM system.

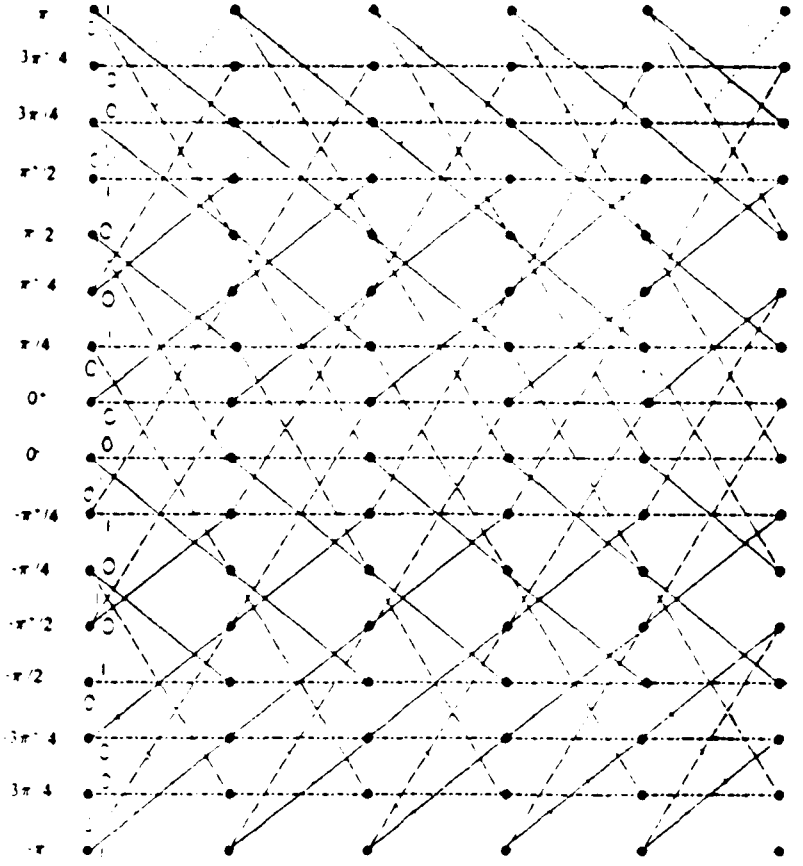


Figure 5.18 A sixteen state trellis for a CNFM system.

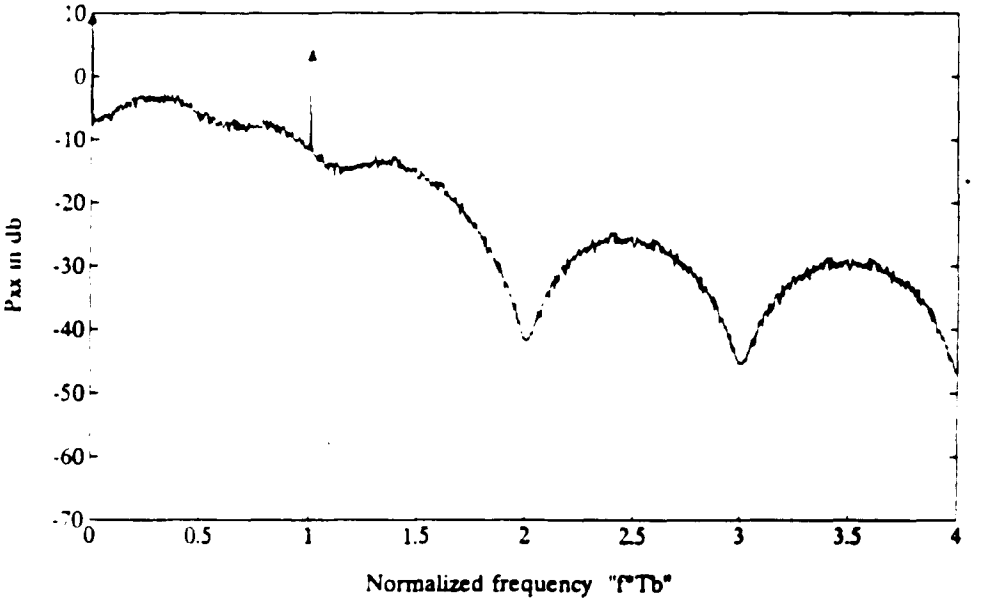


Figure 5.19 Power spectral density of the four state trellis coded CNFM.

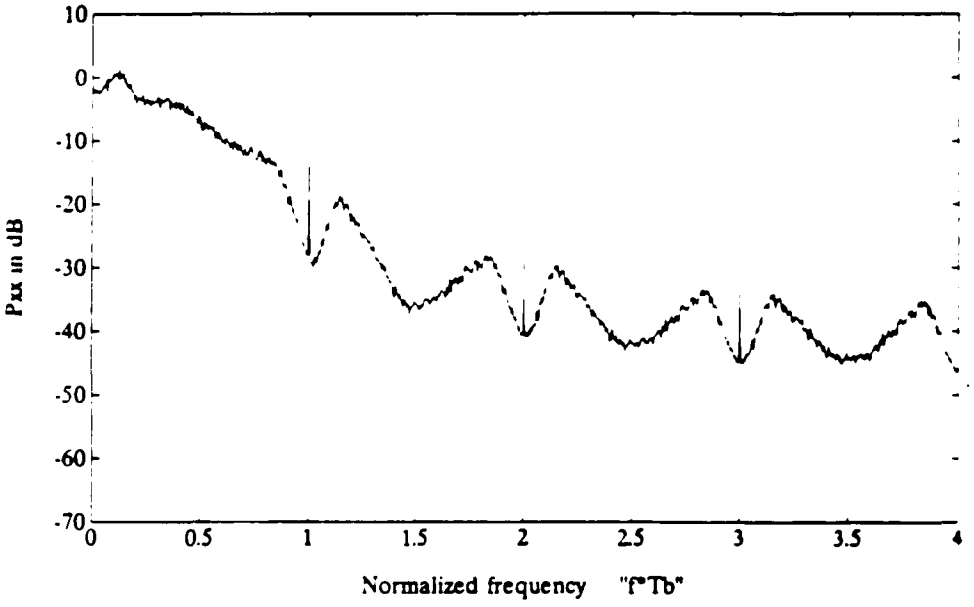


Figure 5.20 Power spectral density of the eight state trellis coded CNFM.

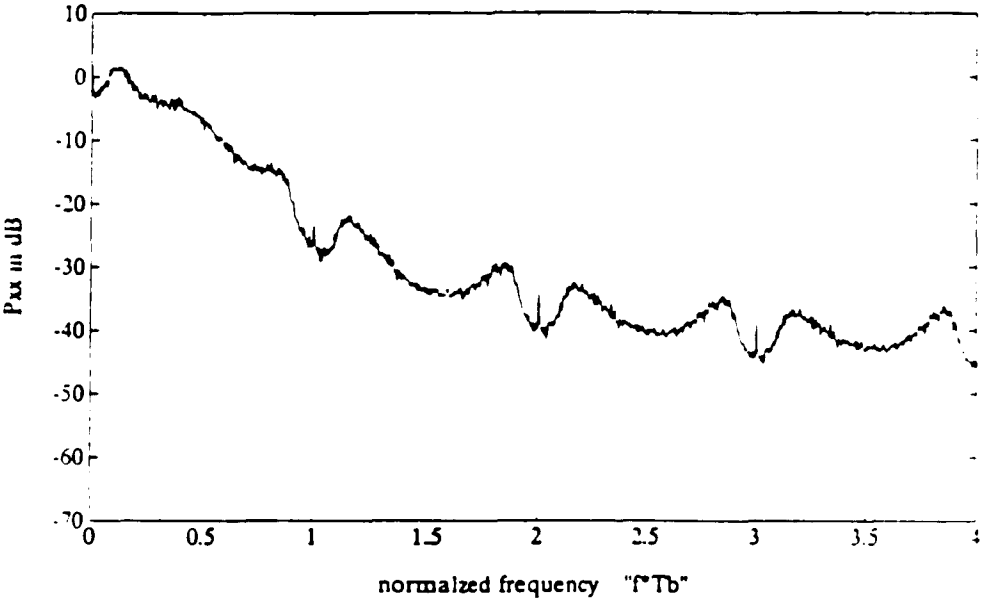


Figure 5.21 Power spectral density of the second 8 state trellis coded CNFM.

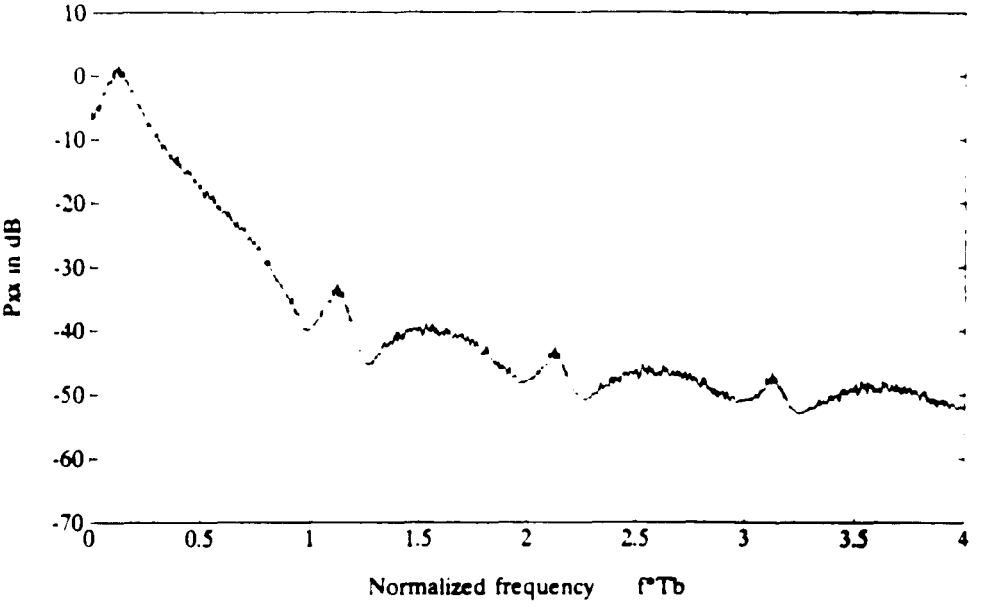


Figure 5.22 Power spectral density of the sixteen state trellis coded CNFM.

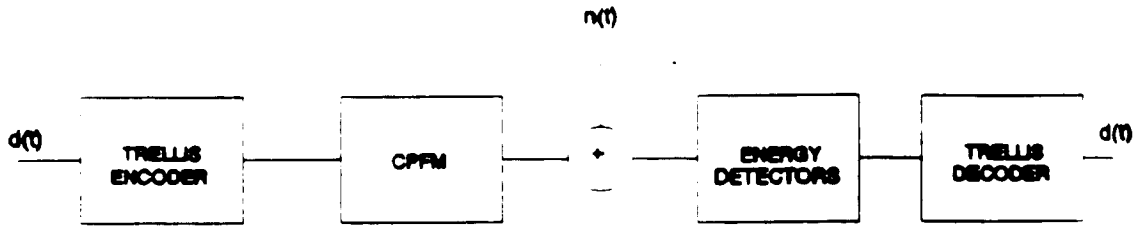


Figure 5.23 Block diagram for a transmitter/ receiver system used to generate the trellis coded CNFM.

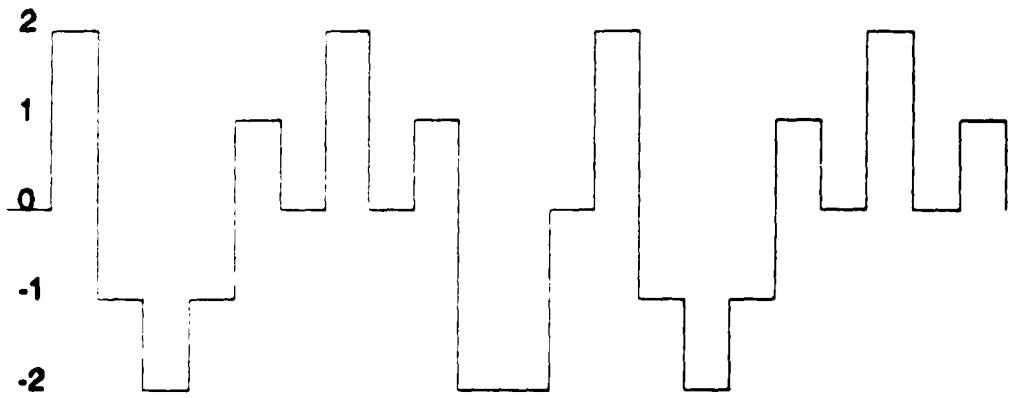


Figure 5.24 A multi-amplitude signal.

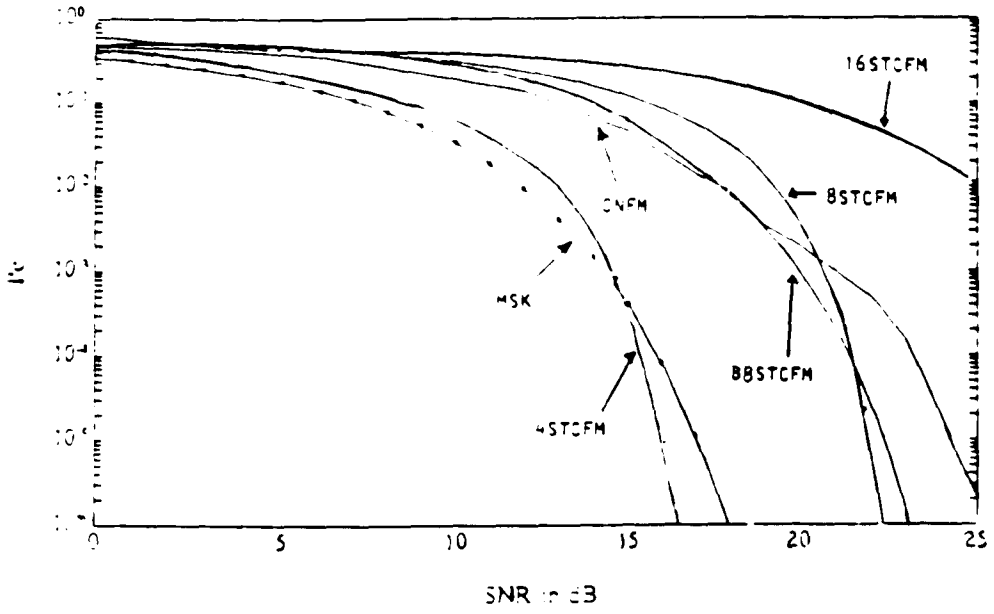


Figure 5.25 Probability of error versus signal to noise ratio for the trellis coded CNFM systems.

has a good spectral efficiency, it has a high error rate. This is mainly is due to the fact the distance between the signal sets is very small. On the other hand, although the four state trellis coded CNFM enjoys a good performance in its error rate, it is not spectrally efficient.

As far as the eight state trellis codes are concerned, decoding the received data by utilizing its trellis structure provides a gain of about 2 dB. Another gain of a dB can be obtained by utilizing the codes which have the minimum free distance of three. However this additional gain of a dB comes at the expense of widening of the spectrum. It is worthwhile to note that, the curve representing probability of error for the trellis coded CNFM has a steeper slope than the basic trellis decoded CNFM. This results in additional gain in signal to noise ratio at probability of errors less than 10^{-6}.

CHAPTER 6

SUMMARY AND CONCLUSION

6.0 SUMMERY AND CONCLUSION

In this thesis, the performance of a narrow band FM system was investigated. In this system, the maximum attainable phase of the frequency modulated waveform was limited. It was found that optimal results were obtained when the maximum phase was set to be $\pm\pi$. By setting such constraint, for $\Delta f T_b = 1/2$, the system has the capability of about 1/b/s/Hz. which is 50% more bandwidth efficient than MSK. for modulation index of 1/4 and 1/8, the system can operate at about 1.56 b/s/Hz. and 2.85 b/s/Hz. respectively. However, by reducing the modulation index, the system's error performance degrades. For example, from figure 5.12, it can be seen that the system

has the probability of error of about 10^{-2} for a SNR of about 24 dB when the modulation index is 1/4. As result of the constraint, the system exhibited a trellis structure that could be used in the decoding process. If this structure is used in the decoding process a gain of about 2 dB is obtained in SNR for the same probability of error. For instance, the system can operate at SNR of about 23 dB, in order to attain a probability of error of about 10^{-6}. This is under the condition that the system utilizes a non-coherent detector in its receiver.

In order to improve, the error performance of the system further, other trellis coded CNFM were also developed. However, the practical use of trellis codes are limited to 8 states. This is a result of widening the spectrum for the four state trellis and poor error performance for the sixteen state trellis coded CNFM. In the four state trellis, the spectrum widens since to meet the requirement of the phase constraint, the modulation index had to be increased. This results in a significant distribution of power over a wide range of frequencies. On the other hand, for the sixteen state trellis,the degradation in error performance is due to the decrease in the Euclidean distance between the transmitted signals. As far as the eight state trellis codes are concerned, both systems have comparable error performance, and provide a 3 dB improvement over the conventionally decoded CNFM system when the modulation index is 1/2. However, the 99% bandwidth for these systems occur at $0.95f_b$ and $0.85f_b$, respectively. Thus, the latter system should be utilized.

BIBLIOGRAPHY

1. Taub, H. Schilling, D. (1986), *Principal of Communication Systems*, McGraw Hill, N.Y.

2. Hykin, S. (1983), *Communication Systems*, Wiley, N.Y.

3. Proakis, J. (1989), *Digital Communication,* McGraw Hill, N.Y.

4. Sklar, B. (1988), *Digital Communications*, Prentice Hall, N.J.

5. Lathi, B.P. (1983), *Modern Digital and Analog Communication Systems*, HRW, N.Y.

6. Blahut, R. (1984), *Theory and Practice of Error Control Codes*, Addison Wesley, M.A.

7. Lin, S. Costello, D.J. (1983), *Error Control Coding*, Prentice Hall, N.J.

8. Cain, C. (1982), *Error-Control Coding for Digital Communication,* Plenum, N.Y.

9. Berlekemp, E.R., "On Decoding Binary Bose-Chaudhuri-Hocquenghem Codes," IEEE Transaction on Information Theory, IT-11, P. 577-580, Oct. 1965

10. Berlekamp, E.R. (1968), *Algebraic Coding Theory,* McGraw Hill, N.Y.

11. Larsen, K.J., " Short Convolutional Codes with Maximal Free Distance for Rates 1/2, 1/3, and 1/4," IEEE Transaction on Information Theory, Vol IT-19, pp. 371-372, May 1973.

12. Ungerboeck, G., " Channel Coding with Multilevel/Phase Signals, IEEE Transaction on Information Theory," IT-28, pp. 55-56, Jan. 1982

13. Ungerboeck, G., " Trellis-Coded Modulation with redundant Signal Sets," Parts I & II, IEEE Communication Mag. Vol. 25, pp. 5-21 Feb. 1987

14. Bennet, W.R. Davey, J.R. (1965), *Data Transmission*, McGraw Hill, NY.