DIGITAL DATA TRANSMISSION USING

SINGLE-SIDEBAND MODULATION

DIGITAL DATA TRANSMISSION USING

SINGLE-SIDEBAND MODULATION

by

Nicholas Ouzas

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Engineering

McMaster University

October 1979

Master of Engineering (1979)   McMaster University
                                 Hamilton, Ontario

TITLE: Digital Data Transmission Using
                    Single-Sideband Modulation

AUTHOR: Nicholas Ouzas, B. Eng.
                    McMaster University

SUPERVISOR: Dr. S.S. Haykin

NUMBER OF PAGES: xv , 256

ABSTRACT

The feasibility of using a partial response-encoded single-sideband (SSB) modulated signal for transmission of digital data in a radio system is considered. The principal methods of SSB signal generation and demodulation are examined, and the effects of carrier synchronization are determined. The effect of steady-state carrier phase errors on the error rate of an SSB partial response receiver is analyzed theoretically and by means of computer simulation. The analysis of a decision-directed SSB carrier phase tracking loop is presented and its performance is evaluated using computer simulation. The performance of the SSB partial response system after amplification by means of travelling-wave-tube (TWT) amplifiers is analyzed by using computer simulation. Finally, a comparison of an SSB partial response system and a quadrature partial response system (QPRS) is made.

ACKNOWLEDGEMENTS

The job what takes the longest to
finish is the one that never gets started.

- The Gaffer

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1 The Case for Digital Radio

Since the early 1950's, analog microwave radio systems
have been extensively used to provide high capacity links
between major urban centres. This widespread use is due to
the fact that under many conditions, analog microwave radio
is the most economical form of long-haul transmission.

In these analog systems, the input signals consist
mainly of frequency division multiplexed telephone channels
and television signals. The amplitude, frequency or phase
of the carrier is modulated according to the input signal.
Since the input signals to be modulated are in analog form,
the use of analog modulation techniques follows quite
naturally.

In recent years, however, digital microwave radio
systems have become increasingly attractive. There are
numerous economic and technical reasons for this trend [1].

Increased use of digital switching techniques has necessitated the provision of digital trunks between tandem and toll digital switches. This, combined with new digital data services, has significantly increased the volume of digital traffic between centres. Digital transmission performs better than analog modulation because the use of regeneration at each repeater site permits transmission over very long distances with negligible degradation of the original signal. The low cost and high performance of digital devices allow a more efficient use of available channels and equipment through time division multiplexing. This results in substantial savings in cost, space, and wiring. Finally, a single digital network can be used to provide a variety of services, and new growth can easily be accommodated. The motivation for digital radio systems is not difficult to understand given the aforementioned benefits.

1.2 The Need for Efficient Modulation

While the advantages of digital modulation are numerous, they are obtained at the expense of a greater bandwidth requirement for transmission of the same information digitally than in analog form. The

proliferation of digital transmission combined with a general increase in the number of users, each with a greater amount of information to be transmitted, has resulted in a severe crowding of the available electromagnetic spectrum.

A variety of complementary solutions exist for this problem [2]. As microwave and signal processing technology advance, allocations in new, higher (> 10GHz) frequency bands can be made while frequency reuse and better spectrum management can relieve congestion in existing bands. A proper choice of modulation format, however, is the most important consideration for any given system, and we shall concern ourselves with this solution in this thesis.

The primary objective of an efficient modulation scheme is to maximize the transmitted data rate through a given bandwidth. A measure of this bandwidth efficiency is the ratio of data rate to bandwidth used which has the dimensions of bits/S/Hz. In many situations, the modulation scheme to be used must meet certain legal minimum efficiency requirements. For example, the FCC requires that a digital radio operating in the 11 GHz band have a minimum efficiency of 2 bits/S/Hz, while operation in the more crowded 4 GHz band requires a minimum efficiency of 4 bits/S/Hz [1].

A graphical comparison of several modulation schemes is shown in Fig. 1.1. The ordinate is the bandwidth efficiency in bits/S/Hz while the abscissa is the average signal to noise power ratio (SNR) required to obtain an average error rate of 10^{-4}. This value is considered the highest error rate permissible for digitally encoded voice transmission [3]. In the figure we have included the Shannon capacity of the channel, which defines the upper bound for the data rate through a channel of a given bandwidth, and hence the bandwidth efficiency, at a given SNR for an arbitrarily low error rate. Also shown are some of the better known modulation schemes. Specifically, we have included: (i) M-ary coherent phase shift keying (CPSK), in which the carrier phase takes one of M values depending on the modulating data sequence; (ii) M-ary frequency shift keying (FSK) where the carrier frequency takes one of M values [4]; and (iii) the quadrature modulation schemes such as quadrature amplitude modulation (QAM) or amplitude-phase keying (APK) where the inphase and quadrature components of the carrier are modulated by separate, independent data streams [5]. Single-sideband (SSB) modulation can be considered as a particular case of QAM. The numbers in the diagram refer to the number of discrete amplitude levels. A baseband or SSB system with L amplitude levels would transmit data at the same efficiency as a QAM system with L
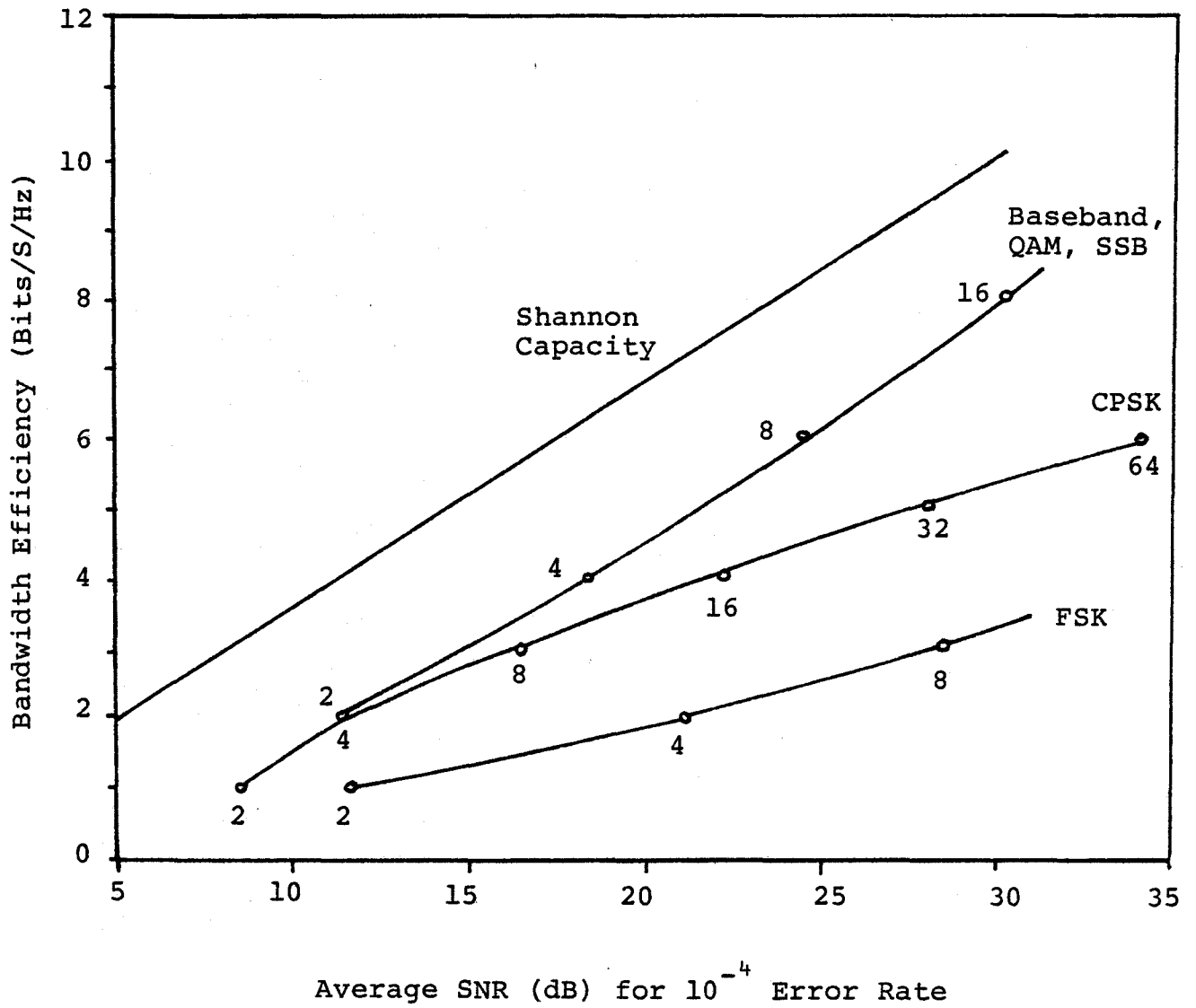
Figure 1.1 Comparison of Several Digital Modulation Schemes

levels in each component of the carrier, or L^2 total signal levels.

At low efficiencies (< 2 bits/S/Hz), the QAM schemes and PSK provide comparable efficiencies at roughly the same SNR. However, PSK has been chosen over QAM for most systems for its ease of implementation and its excellent performance over non-linear channels. In particular, 8-PSK has been used extensively in existing radio systems [6] - [8] and bandwidth efficiencies up to 3 bits/S/Hz have been reported.

At high efficiencies, PSK is no longer competitive with the quadrature schemes. For example, a 4 bits/S/Hz QAM system requires about 18 dB SNR for a 10^{-4} error rate whereas a 16-PSK system at the same efficiency requires 4 dB more SNR for the same error rate. This difference becomes even greater when higher efficiencies are required. It becomes attractive, therefore, to examine the feasibility of these quadrature modulation formats, and SSB in particular, as a possible solution to the problem of high efficiency digital communications.

1.3 The Elements of a Digital Radio System

The composition of a typical digital radio transmitter

and receiver is shown in Fig. 1.2. Several low speed data streams are interleaved to provide a single high speed stream typically of the order of 90 Mb/s. This data stream is scrambled to break up any periodic sequences of data which can result in spikes in the output spectrum of the signal, and parity bits are inserted to provide bit error rate monitoring capabilities at the receiver. The actual digital modulation occurs at intermediate frequencies (IF) with a carrier of about 70 MHz. The digitally modulated signal is then shifted into the microwave frequency band by the upconverter where it is amplified and filtered. It is this RF filter which determines the bandwidth efficiency of the system. The amplified and filtered signal is then passed by means of waveguides to the antenna, where the signal is radiated into the atmosphere.

Due to reflection and refraction in the atmosphere, the original transmitted signal, may in fact follow several different paths before arriving at the receiver. These paths are in general of different lengths, so that each version of the signal will arrive with a different phase shift resulting in partial cancellation of the signal. This effect is known as multipath fading [9] and can result in a significant increase in the error rate of the receiver. To combat this degradation, a second receiver antenna,
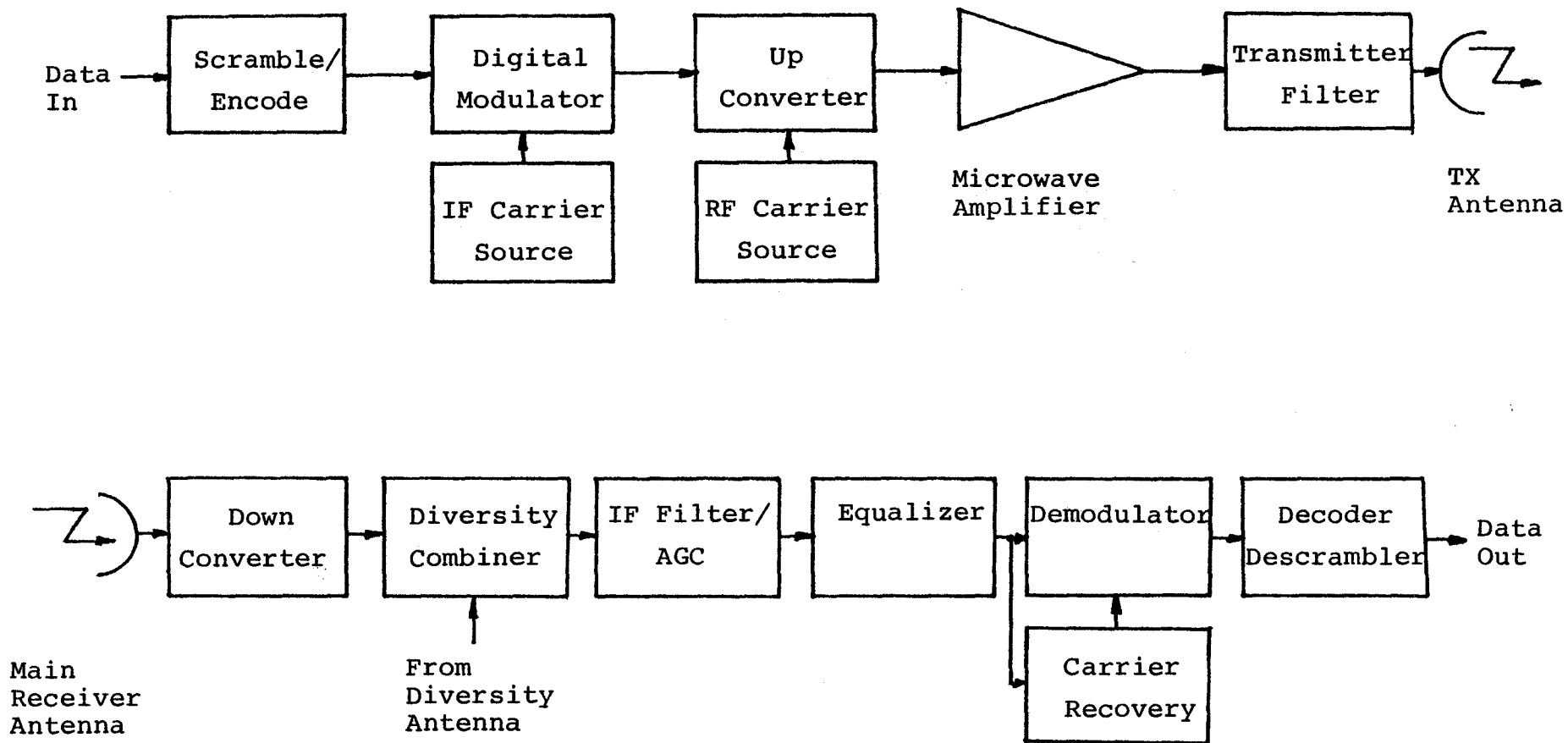
Figure 1.2   Typical Digital Radio Transmitter and Receiver

physically separated from the main antenna, is used. As the fading on the main antenna is generally uncorrelated with the fading on the diversity antenna, there is almost always a usable signal present at one of the two antennas. A combination of these two signals before demodulation therefore greatly improves the reliability of the system. In practical radio systems, these reliability requirements can be quite stringent: for the Northern Telecom RD-3 radio, for example, only two hours of outages per year are allowed [3]. This outage time is further broken down into one hour due to equipment failure, and one hour due to fading. This translates into an availability requirement of 99.98%.

Both the diversity and main receivers have an input bandpass filter which has a bandwidth large enough to pass the received signal with a minimum of distortion but small enough to limit the noise entering the receiver. After down conversion, the main and diversity signals are combined in some optimum fashion and then filtered at IF to further reduce the effects of noise and interference from adjacent radio channels. An automatic gain control (AGC) minimizes the variation in signal level, so that an essentially constant signal is available to the rest of the system. After equalization to compensate for the various degradatory effects accumulated up to that point, the signal is

demodulated to baseband. This requires a local source of the carrier frequency and phase. In addition, the demodulator comprises a symbol timing recovery system, an additional baseband filter, a baseband equalizer, and a decision device which regenerates the original data stream. After regeneration, the data stream is descrambled and the parity bits inserted at the transmitter are used to test for outage conditions. The recovered high speed data stream can be demultiplexed into the original low speed streams so that channels can be dropped or inserted at different stations along the route, or the high speed stream can be directly fed into another digital modulator. The station then acts simply as a repeater.

1.4 Single-Sideband Modulation

For analog signals, SSB is the most efficient form of modulation in that the modulated signal bandwidth is the same as the baseband signal bandwidth, and the transmitted power is the minimum possible. Consequently, SSB has been extensively used in frequency-division multiplex (FDM) voice transmission systems such as found in the telephone network [10], [11]. For the same reasons, SSB has found widespread application in high frequency radio [12].

Microwave transmission of FDM signals, however, does not in general make use of SSB. Rather, frequency modulation (FM) is used because of its greater tolerance of the non-linearities encountered in RF amplifiers. Technological advances and spectral crowding have nevertheless resulted in the development of several SSB-FDM microwave radio systems [13], [14] operating in the 2 and 6 GHz bands.

For digital data transmission, SSB has been used mainly over telephone voice channels [15] where the limited bandwidth necessitates the use of high efficiency multilevel modulation. Similarly, SSB has also been used for data transmission over FDM groups at a rate of 48 Kb/s [16], and a 576 Kb/s system using a 300 MHz carrier has been developed [17]. In general, the use of SSB for data transmission over channels other than voice channels has been rather limited.

1.5 Scope of the Thesis

In this thesis, we shall attempt to analyze the feasibility of using SSB modulation in a high capacity digital radio system. We shall initially address ourselves to the problems involved in the generation and demodulation

of the SSB signal. We shall illustrate the need for shaping the modulating signal spectrum, and hence shall make use of partial response encoding. Two major areas of concern will be examined: the effects of carrier synchronization errors and the effects of the non-linear RF amplifiers used in most digital radio systems. The degradations due to these effects will be quantified and solutions for overcoming these degradations will be proposed and analyzed. Both theoretical and computer simulation methods will be used in the analysis, and in the majority of cases, the bit error rate (BER) will be used as the criterion of comparison.

CHAPTER 2. GENERAL SSB CONSIDERATIONS

2.1 Introduction

In this chapter we shall introduce some of the more important properties of SSB modulation-demodulation systems. The discussion will focus on the operation of the various modulators and demodulators that may be used as well as the complexity and implementability of each configuration. In addition, the performance of the various demodulators in the presence of both time-variant and steady state receiver phase errors will be analyzed. Finally, the properties of the noise encountered in SSB systems will be discussed and the differences with other forms of modulation outlined. Unless otherwise noted, the discussion will be valid for both analog and digital systems.

2.2 Physical and Mathematical Formulation

If a lowpass signal $a(t)$ is multiplied by a carrier $\cos \omega_c t$, then the spectrum of $a(t)$ is shifted to $\pm \omega_c$ as shown in Fig. 2.1(b). It can be seen that the resulting signal requires twice the bandwidth of the original baseband sideband This is referred to as a double-sideband (DSB) modulated signal. Due to the symmetry of the sidebands about ω_c, one of them can be eliminated with no loss of information as shown in Fig. 2.1(c). This is the underlying principle of SSB modulation.

The mathematical description of the resulting SSB waveform is

$$s(t) = a(t) \cos \omega_c t + \hat{a}(t) \sin \omega_c t \qquad (2.1)$$

$$= \text{Re}[(a(t) - j\,\hat{a}(t))\, e^{j\omega_c t}] \qquad (2.2)$$

where $\hat{a}(t)$ is the Hilbert Transform of $a(t)$ and is given by
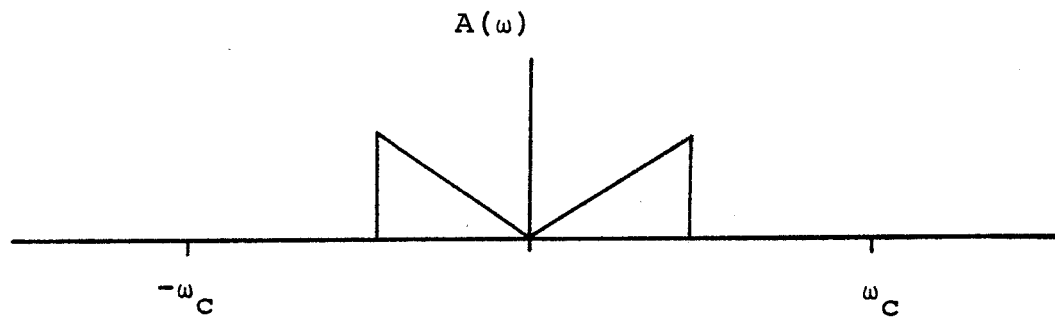
(i) Baseband Signal
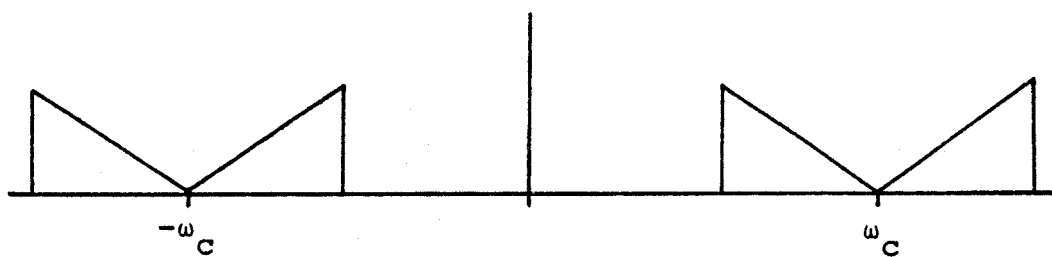
(ii) DSB Signal
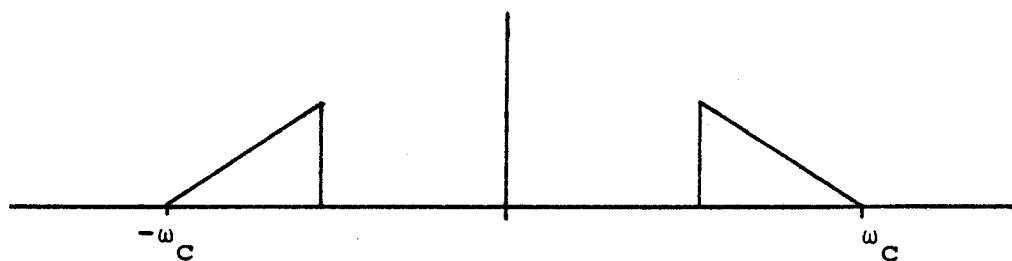
(iii) SSB Signal With Lower Sideband Retained

Figure 2.1 Baseband, DSB and SSB Signals
in the Frequency Domain

$$\hat{a}(t) \triangleq \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{a(\tau)}{t-\tau} d\tau \tag{2.3}$$

where the integral is interpreted in the Cauchy principal value sense.

In this case the lower sideband was retained but the upper sideband could equally well have been retained instead. The resulting SSB waveform in this case is

$$s_U(t) = a(t)\cos\omega_c t - \hat{a}(t)\sin\omega_c t \tag{2.4}$$

$$= \text{Re}[(a(t) + j\hat{a}(t))e^{j\omega_c t}] \tag{2.5}$$

with a(t) defined as before. Rewriting (2.2) and (2.5) in polar form gives

$$s_L(t) = \text{Re}[r(t)e^{j(\omega_c t - \theta(t))}] \tag{2.6}$$

$$s_U(t) = \text{Re}[r(t)e^{j(\omega_c t + \theta(t))}] \tag{2.7}$$

where

$$r(t) \triangleq (a^2(t) + \hat{a}^2(t))^{1/2} \tag{2.8}$$

$$\theta(t) \overset{\Delta}{=} \tan^{-1}[\hat{a}(t)/a(t)]$$ (2.9)

and the subscripts L and U denote the lower and upper sideband signals respectively. We see that the carrier is simultaneously amplitude modulated by $r(t)$ and phase modulated by $\theta(t)$. We also note that for a lower sideband signal the carrier phase is delayed by the modulating signal while the carrier phase is advanced for the upper sideband signal. In both cases, the bandwidth of the modulated signal is the same as that of the original baseband signal.

2.3 SSB Modulation Methods

In order to generate an SSB signal, one of the two sidebands must be removed. This sideband elimination can be accomplished by filtering, by quadrature processing, or by a combination of both. In quadrature processing, the sideband to be removed undergoes two successive 90° phase shifts resulting in a polarity opposite of the sideband to be retained. Simple addition with an ordinary non-phase shifted version of the signal results in cancellation of the unwanted sideband and reinforcement of the other. The Hartley modulator [18] utilizes phase shifting alone while

the Weaver modulator [19] relies on a combination of filtering and phase shifting.

2.3.1 Filter Method

The filter or frequency discrimination method [20] is the simplest way of producing an SSB signal. The modulator is shown in Fig. 2.2(a) while the frequency domain representation of the modulation process is shown in Fig. 2.2(b). Here a DSB signal $a(t)\cos\omega_c t$ is passed through a bandpass filter which is designed to attenuate the unwanted sideband while passing the desired sideband undistorted. This requires a very small filter transition region so that all of the unwanted sideband is removed. Presence of significant energy near zero frequency (i.e. DC) in the spectrum of the baseband signal $a(t)$ makes the filtering very difficult to realize physically. The filter cutoff at the other band edge need not be so steep. This method is commonly used in FDM systems where the stringent filtering requirements are met with high Q crystal filters [10]. Lack of energy below 200 Hz in voice signals greatly simplifies the filtering task.
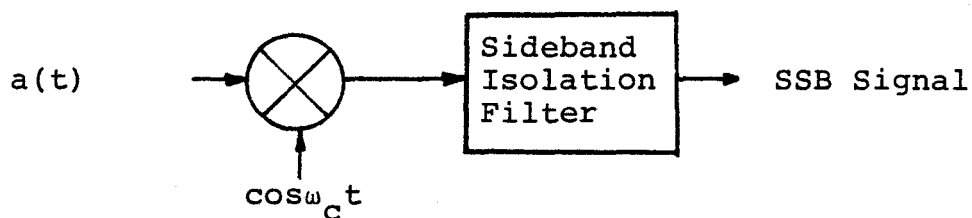
Figure 2.2a   Filter Method of SSB Signal Generation

(i)   Spectrum of a(t)

(ii)   Filter Input Spectrum

Filter Attenuation

(iii)   Filter Output Spectrum
(Lower Sideband Retained)

Filter Attenuation

(iv) Filter Output Spectrum
(Upper Sideband Retained)
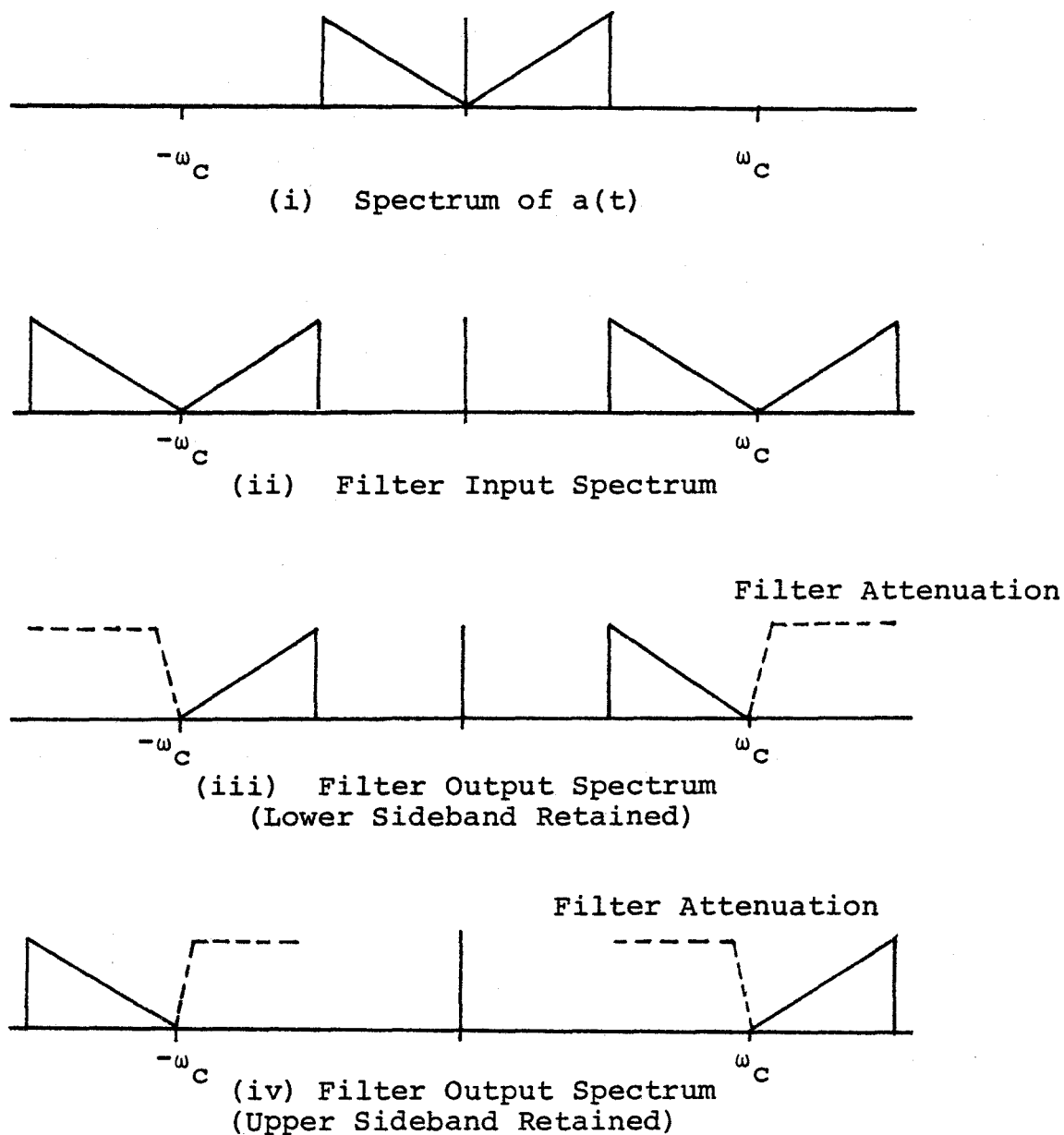
Figure 2.2b   Frequency Domain Representation
of Filter Method

2.3.2 Hartley Method

The Hartley or phase shift method is a direct implementation of (2.1). The modulator is shown in Fig. 2.3(a). The baseband signal a(t) is passed through a filter which effects the Hilbert Transform of the signal. This filter has a frequency response

$$\hat{H}(\omega) = + j \qquad \omega < 0$$

$$= 0 \qquad \omega = 0 \qquad (2.10)$$

$$= - j \qquad \omega > 0$$

Thus the positive frequencies undergo a -90° phase shift while the negative frequencies are shifted by $+90^\circ$. Further multiplication by $\sin \omega_c t$ results in an additional -90° phase shift of the upper sideband thereby reversing the polarity of the sideband. Addition with the output $a(t) \cos \omega_c t$ of the inphase arm of the modulator cancels the upper sideband while subtraction cancels the lower sideband.

The time response corresponding to the filter
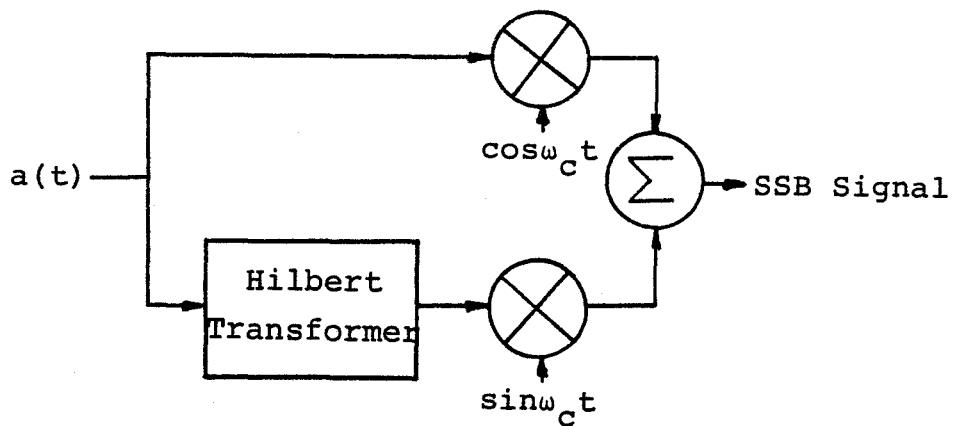
Figure 2.3a   Hartley Modulator
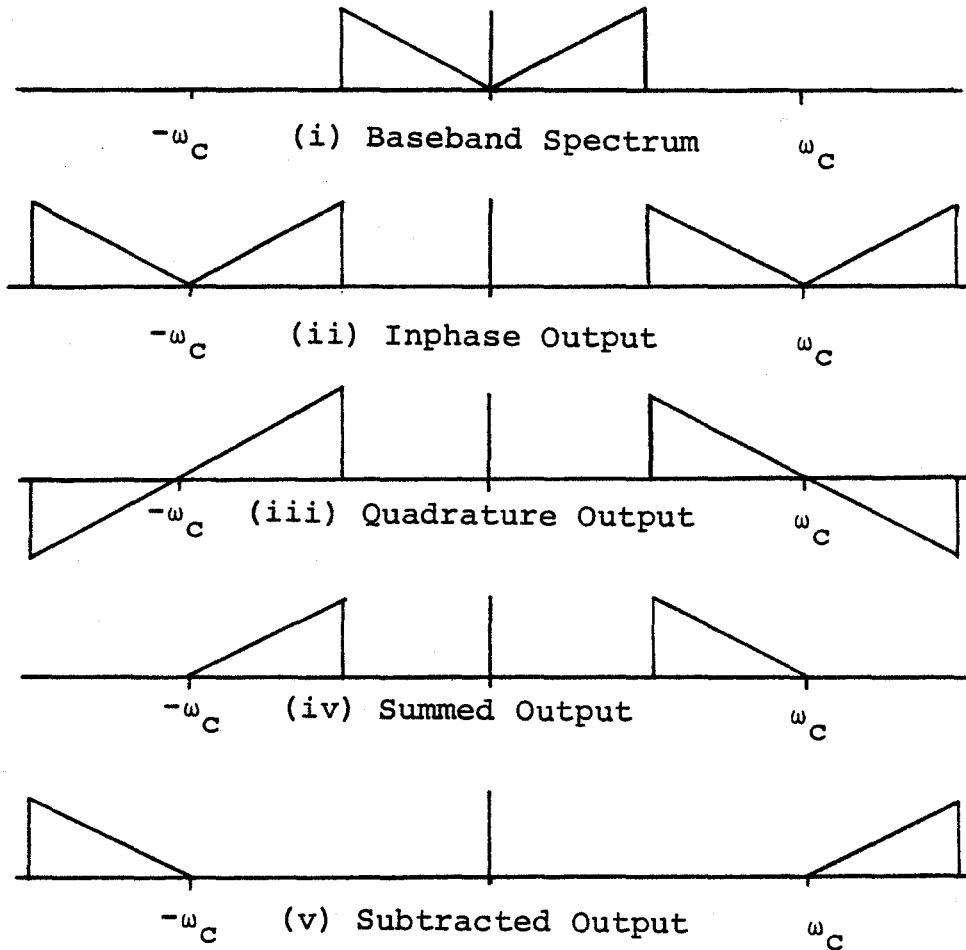
Figure 2.3b   Frequency Domain Representation of
Hartley Modulation Process

function of (2.10) is

$$\hat{h}(t) = \frac{1}{\pi t} \qquad\qquad (2.11)$$

In this form, this filter is clearly not realizable. However, in digital systems where the spectrum of the signal is shaped, use of certain techniques permit an approximation that is sufficiently close for most applications. This is discussed further in Chapter 3. Once again, presence of low frequency energy creates problems in sideband separation as the practical realization of the Hilbert transformer cannot phase shift these components properly.

2.3.3 Weaver Method

The final and least known SSB modulation method is the Weaver modulator illustrated in Fig. 2.4. The unique feature of this method is the premodulation of the baseband signal by $\cos\omega_m t$ and $\sin\omega_m t$ where ω_m is usually taken at the midpoint of the passband of a(t). That is, if a(t) is bandlimited to [-W, W] HZ, then

$$\omega_m = \pi W \qquad\qquad (2.12)$$
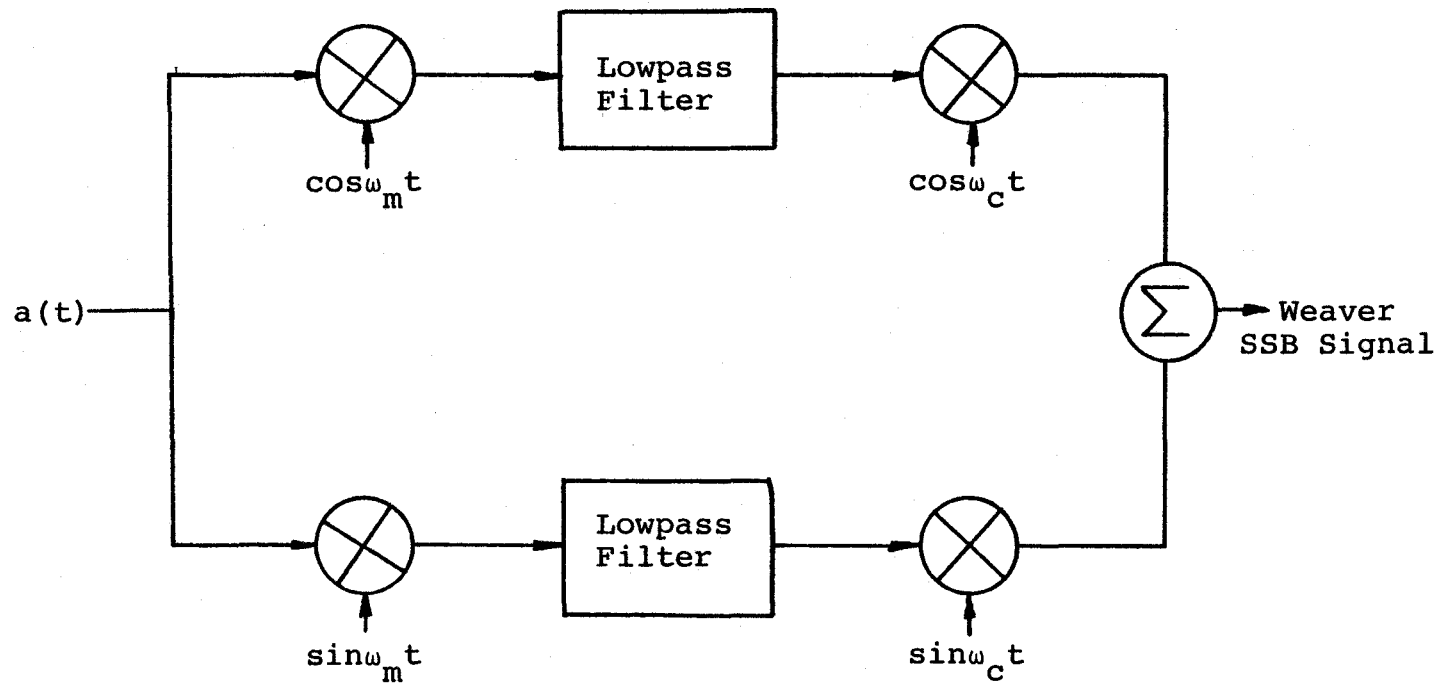
Figure 2.4   Weaver Modulator

The premodulator outputs are then lowpass filtered. The cutoff frequency of the filter is selected once again as ω_m, creating a sideband overlap in the band $[-\omega_m, \omega_m]$ rad/s as shown in Fig. 2.5.

The frequency domain output of the inphase lowpass filter is

$$X_1(\omega) = A_+(\omega+\omega_m) + A_-(\omega-\omega_m) \tag{2.13}$$

while the quadrature lowpass output is

$$Y_1(\omega) = jA_+(\omega+\omega_m) - j A_-(\omega-\omega_m) \tag{2.14}$$

where $A_+(\omega)$ and $A_-(\omega)$ are the positive and negative frequency portions of $a(t)$.

Further multiplication by $\cos\omega_c t$ and $\sin \omega_c t$ gives

$$X(\omega) = A_+(\omega+\omega_m-\omega_c) + A_-(\omega-\omega_m-\omega_c)$$
$$+ A_+(\omega+\omega_m+\omega_c) + A_-(\omega-\omega_m+\omega_c) \tag{2.15}$$

$$A(\omega)$$

(i) Baseband Spectrum

(ii) Inphase Premodulator Output

(iii) Inphase Modulator Output

(iii) Quadrature Modulator
Output

(iv) Summed Output

(v) Subtracted Output
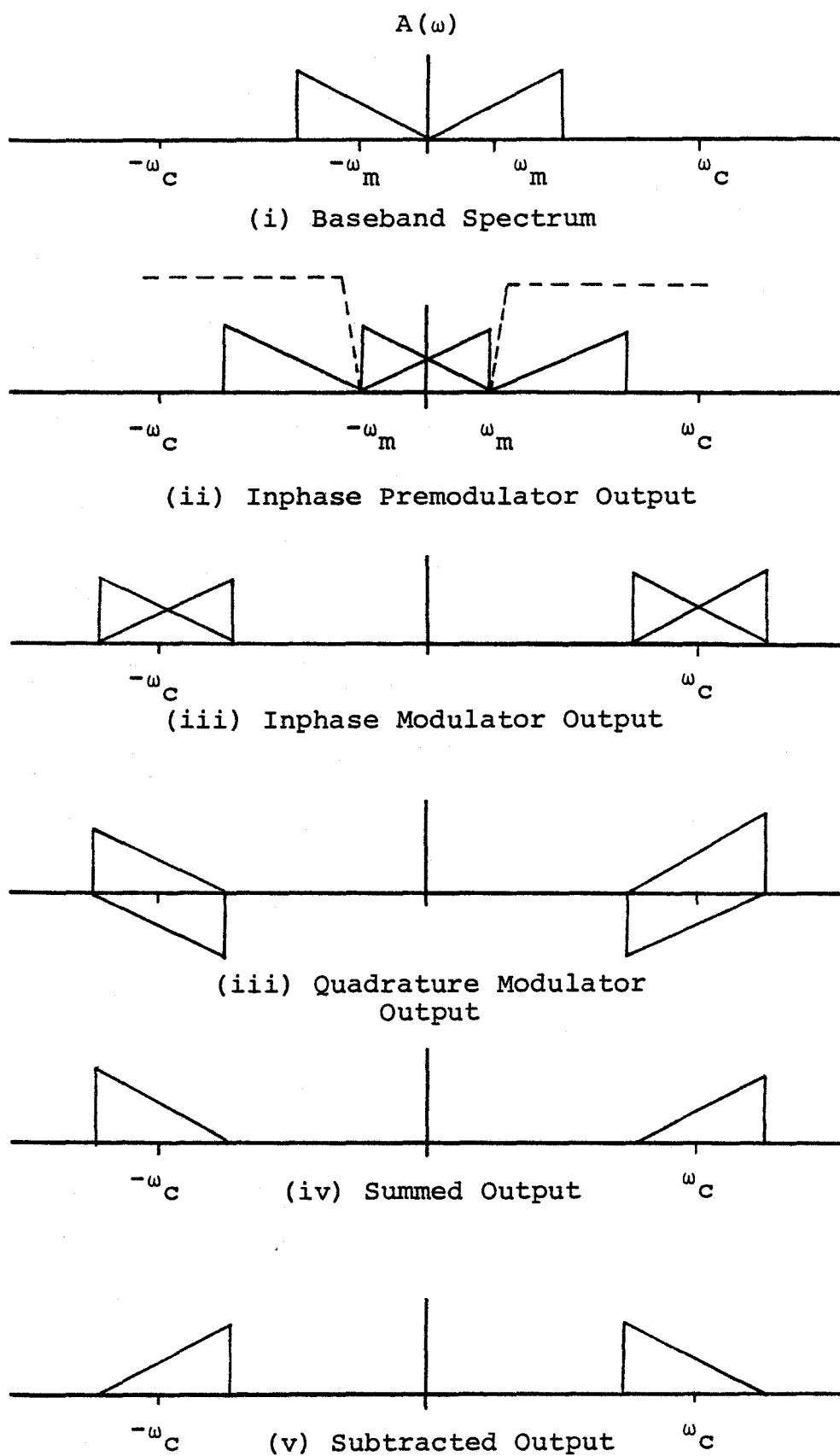
Figure 2.5   Frequency Domain Representation of
Weaver Modulation Process

and

$$Y(\omega) = A_+(\omega + \omega_m - \omega_c) - A_-(\omega - \omega_m - \omega_c)$$

$$- A_+(\omega + \omega_m + \omega_c) + A_-(\omega - \omega_m + \omega_c) \qquad (2.16)$$

as the inphase and quadrature outputs respectively. It can be seen, therefore, that the Weaver modulator processes the quadrature signal in a fashion similar to the Hartley modulator in that one of the sidebands undergoes two 90° phase shifts.

Addition of the branch outputs yields

$$S_1(\omega) = A_+(\omega + \omega_m - \omega_c) + A_-(\omega - \omega_m + \omega_c) \qquad (2.17)$$

which in the time domain is

$$s_1(t) = a(t)\cos(\omega_c - \omega_m)t - \hat{a}(t)\sin(\omega_c - \omega_m)t \qquad (2.18)$$

This is a conventional SSB signal with a carrier $\omega_c - \omega_m$.

Similarly, subtraction of the two branch outputs gives

$$S_2(\omega) = A_-(\omega - \omega_m - \omega_c) + A_+(\omega + \omega_m + \omega_c) \qquad (2.19)$$

or

$$s_2(t) = a(t)\cos(\omega_c + \omega_m)t + \hat{a}(t)\sin(\omega_c + \omega_m)t \qquad (2.20)$$

which is a conventional SSB signal with a carrier $\omega_c + \omega_m$. Note that in both cases, the sideband occupies the same bandwidth [$\omega_c - \omega_m$, $\omega_c + \omega_m$] rad/S while the carrier changes value, whereas in the Hartley modulator the carrier remains the same and the sideband position is changed.

The advantage of the Weaver modulator is that sideband cancellation is much more readily accomplished than with the other two modulators. Also, the sideband cancellation occurs in the passband of the other sideband rather than outside. However, the out of band signal components are determined by how well the lowpass filters attenuate the premodulated signals. This means that filters with steep cutoffs are still required; moreover, these filters must be closely matched to ensure maximum sideband cancellation. Use of the premodulator requires that an additional stable reference source be supplied. Use of digital signal processing can overcome the filtering and matching requirements [21], but for most applications, the Weaver modulator presents no advantage over the Hartley or filter methods.

2.4 SSB Demodulation Methods

Due to the quadrature component inherent to the SSB
signal, simple envelope detection cannot be used for
demodulation. Rather, the signal must be demodulated
coherently by multiplying the received signal with a local
estimate of the carrier. An SSB signal can also be
demodulated using a Weaver demodulator. In the following
sections, each of these demodulation schemes will be
presented. Also, since the exact carrier frequency and
phase are usually not known at the receiver, the effects of
such synchronization errors will be examined.

2.4.1 Coherent Demodulation

Consider a receiver lower sideband signal:

$$s(t) = a(t)\cos(\omega_c t + \phi_c) + \hat{a}(t)\sin(\omega_c t + \phi_c) \qquad (2.21)$$

where ϕ_c is the phase of the received carrier. The incoming
signal is multiplied with a local estimate of the carrier

$$z_I(t) = 2\cos(\omega_c t + \tilde{\phi}_c) \qquad (2.22)$$

where $\tilde{\phi}_c$ is the local estimate of the carrier phase. The

resulting signal is then lowpass filtered as in Fig. 2.6 to give:

$$r_I(t) = a(t)\cos \Delta\phi + \hat{a}(t)\sin \Delta\phi \qquad (2.23)$$

where

$$\Delta\phi \triangleq \phi_c - \tilde{\phi}_c \qquad (2.24)$$

is the phase error between the received signal carrier and the local carrier estimate.

We see that a quadrature channel dependent distortion term is added to the desired signal. Also, the desired signal component decreases with increasing phase error so that the signal is degraded further. This underlines the importance of a good carrier tracking system.

Suppose that the phase error is of the form

$$\Delta\phi = \Delta\omega t + \phi_0 \qquad (2.25)$$

so that the phase error is dynamic. We find that in the case of the lower sideband signal, the frequency error $\Delta\omega$
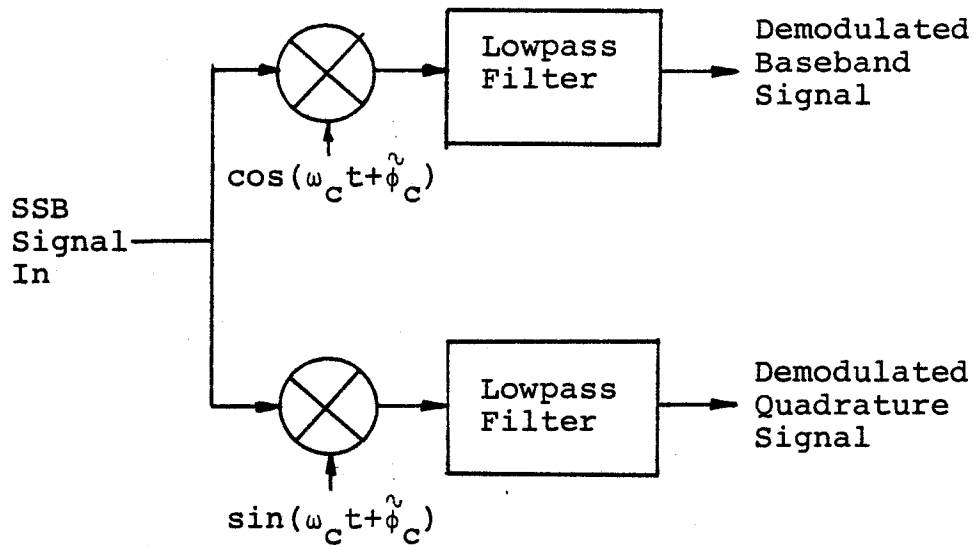
Figure 2.6 SSB Coherent Demodulator
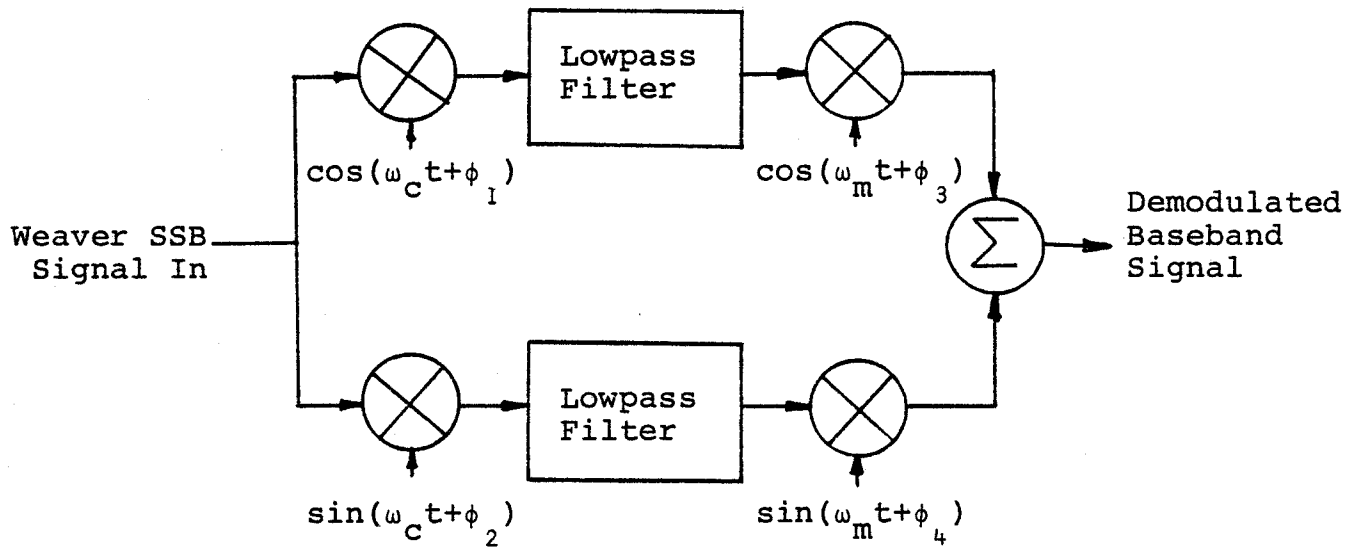
Figure 2.7 Weaver Demodulator

causes the recovered baseband signal sidebands to shift towards each other when $\Delta\omega$ is positive causing the sidebands to overlap. Conversely, a negative $\Delta\omega$ spreads the sidebands further apart.

If the incoming signal is multiplied by

$$Z_Q(t) = 2 \sin(\omega_c t + \tilde{\phi}_c) \qquad (2.26)$$

and then lowpass filtered, the output is

$$r_Q(t) = \hat{a}(t)\cos\Delta\phi - a(t)\sin\Delta\phi \qquad (2.27)$$

with $\Delta\phi$ defined as before. An effect similar to that of the inphase channel occurs. In the case of no frequency or phase errors ($\Delta\phi = 0$), the quadrature output is simply $\hat{a}(t)$, the Hilbert transform of the baseband signal.

2.4.2 Weaver Demodulation

The Weaver demodulation process is similar to the Weaver modulator in reverse order. The demodulator is shown in Fig. 2.7.

Consider a Weaver lower sideband signal:

$$s(t) = a(t) \cos[\,(\omega_c + \omega_m)t + \phi_c\,] + \hat{a}(t) \sin[\,(\omega_c + \omega_m)t + \phi_c\,] \qquad (2.28)$$

The inphase lowpass filter output after multiplication with $2\cos(\omega_c t + \phi_1)$ is

$$x_1(t) = a(t)\cos(\omega_m t + \phi_c - \phi_1) + \hat{a}(t)\sin(\omega_m t + \phi_c - \phi_1) \qquad (2.29)$$

Similarly, the quadrature filter output is

$$y_1(t) = -a(t)\sin(\omega_m t + \phi_c - \phi_2) + \hat{a}(t)\cos(\omega_m t + \phi_c - \phi_2) \qquad (2.30)$$

Note that the receiver lowpass filters act only to remove the double frequency terms resulting from the first multiplication, and hence do not require the steep cutoffs of the modulator filters. After multiplication with $2\cos(\omega_m t + \phi_3)$ and $2\sin(\omega_m t + \phi_4)$, the inphase and quadrature branch outputs are, respectively,

$$x_2(t) = a(t)[\cos(2\omega_m t + \phi_c + \phi_3 - \phi_1) + \cos(\phi_c - \phi_1 - \phi_3)]$$

$$+ \hat{a}(t)[\sin(2\omega_m t + \phi_c + \phi_3 - \phi_1) + \sin(\phi_c - \phi_1 - \phi_3)] \qquad (2.31)$$

and

$$y_2(t) = a(t)[\cos(2\omega_m t + \phi_c - \phi_2 + \phi_4) - \cos(\phi_c - \phi_2 - \phi_4)]$$

$$+ \hat{a}(t)[\sin(2\omega_m t + \phi_c - \phi_2 + \phi_4) - \sin(\phi_c - \phi_2 - \phi_4)] \qquad (2.32)$$

Subtracting the quadrature output from the inphase output gives

$$r_I(t) = a(t)[\cos(2\omega_m t + \phi_c - \phi_1 + \phi_3) - \cos(2\omega_m t + \phi_c - \phi_2 + \phi_4)$$

$$+ \cos(\phi_c - \phi_1 - \phi_3) + \cos(\phi_c - \phi_2 - \phi_4)]$$

$$+ \hat{a}(t)[\sin(2\omega_m t + \phi_c - \phi_1 + \phi_3) - \sin(2\omega_m t + \phi_c - \phi_2 + \phi_4)$$

$$+ \sin(\phi_c - \phi_1 - \phi_3) + \sin(\phi_c - \phi_2 - \phi_4)] \qquad (2.33)$$

In order that the time varying sinusoidal terms cancel, we must have $\phi_1 = \phi_2$ and $\phi_3 = \phi_4$. This gives

$$r_I(t) = a(t)\cos(\phi_c - \phi_1 - \phi_3) + \hat{a}(t)\sin(\phi_c - \phi_1 - \phi_3) \qquad (2.34)$$

Further setting $\phi_1 = \phi_3$ so that all the oscillators are synchronized gives

$$r_I(t) = a(t)\cos(\phi_c - 2\phi_1) + \hat{a}(t)\sin(\phi_c - 2\phi_1) \qquad (2.35)$$

where the factor of 2 has been neglected for convenience.

Thus, a behaviour similar to that of the coherent demodulator is obtained with the exception that the distortion in the Weaver demodulator depends on twice the local carrier phase estimate ϕ_1. This means that the Weaver demodulator with a lower Weaver sideband input degrades more rapidly with carrier misalignment and is therefore inferior to the coherent demodulator.

We shall now develop the equivalent expression for the Weaver upper sideband input,

$$s(t) = a(t)\cos[(\omega_c - \omega_m)t + \phi_c] - \hat{a}(t)\sin[(\omega_c - \omega_m)t + \phi_c] \qquad (2.36)$$

Proceeding as before, the inphase and quadrature lowpass filter outputs are

$$x_1(t) = a(t)\cos(\omega_m t - \phi_c + \phi_1) + \hat{a}(t)\sin(\omega_m t - \phi_c + \phi_1) \qquad (2.37)$$

$$y_1(t) = a(t)\sin(\omega_m t - \phi_c + \phi_2) - \hat{a}(t)\cos(\omega_m t - \phi_c + \phi_2) \qquad (2.38)$$

The output of the second set of multipliers becomes

$$x_2(t) = a(t)[\cos(2\omega_m t - \phi_c + \phi_1 + \phi_3) + \cos(\phi_c - \phi_1 + \phi_3)]$$

$$+ \hat{a}(t)[\sin(2\omega_m t - \phi_c + \phi_1 + \phi_3) - \sin(\phi_c - \phi_1 + \phi_3)] \quad (2.39)$$

$$y_2(t) = a(t)[-\cos(2\omega_m t - \phi_c + \phi_2 + \phi_4) + \cos(\phi_c - \phi_2 + \phi_4)]$$

$$-\hat{a}(t)[\sin(2\omega_m t - \phi_c + \phi_2 + \phi_4) + \sin(\phi_c - \phi_2 + \phi_4)] \quad (2.40)$$

Addition of these two signals gives the demodulated signal

$$r_I(t) = a(t)[\cos(2\omega_m t - \phi_c + \phi_1 + \phi_3) - \cos(2\omega_m t - \phi_c + \phi_2 + \phi_4)$$

$$+ \cos(\phi_c - \phi_1 + \phi_3) + \cos(\phi_c - \phi_2 + \phi_4)]$$

$$+\hat{a}(t)[\sin(2\omega_m t - \phi_c + \phi_1 + \phi_3) - \sin(2\omega_m t - \phi_c + \phi_2 + \phi_4)$$

$$- \sin(\phi_c - \phi_1 + \phi_3) - \sin(\phi_c - \phi_2 + \phi_4)] \quad (2.41)$$

As in the lower Weaver sideband case, setting $\phi_1 = \phi_2$ and $\phi_3 = \phi_4$ is required in order to eliminate the time varying sinusoids. Hence

$$r_I(t) = a(t)\cos(\phi_c - \phi_1 + \phi_3) - \hat{a}(t)\sin(\phi_c - \phi_1 + \phi_3) \quad (2.42)$$

If, in addition, $\phi_1 = \phi_3$, we obtain

$$r_I(t) = a(t)\cos\phi_c - \hat{a}(t)\sin\phi_c \qquad (2.43)$$

In this case, the errors introduced by each oscillator cancel. These errors are not necessarily static in that a frequency error may have to be included. This implies that the effect of a frequency misalignment of the two receiver oscillators would be reduced if they were both misaligned in the same way. However, if the first oscillator frequency ω_c was greater than the corresponding transmitter oscillator while the second oscillator frequency ω_m was lower than the corresponding transmitter oscillator, or vice versa, then the frequency error would be increased, and not decreased.

The primary disadvantage of the Weaver demodulator is the complexity of the implementation. Two receiver frequency sources are needed, as well as four multipliers and a summer. Phase matching of the two channels is once again a problem. While the upper Weaver sideband demodulator shows an error cancellation behaviour, the problem of frequency misalignment becomes crucial. Use of a central frequency reference could alleviate this. Any drift in the central reference would result in a drift in the same direction of both oscillators. The cancellation effect of the demodulator would then reduce the overall frequency

error.

As a final comment, it should be noted that a Weaver modulator signal could also be demodulated coherently in one step if the carrier frequency is properly chosen. For a lower Weaver sideband, this means using $\omega_c + \omega_m$ while for an upper Weaver sideband, $\omega_c - \omega_m$ must be used. Conversely, a conventional SSB signal could be demodulated using the Weaver method if the receiver frequencies are correctly chosen.

2.5 Noise in SSB Systems

Consider now an SSB system with noise present. We assume that the noise is additive white Gaussian noise (AWGN) of zero mean and two sided spectral density $N_0/2$. A lower sideband SSB signal of bandwidth W Hz extends from $\omega_c - 2\pi W$ to ω_c rad/S. Since the signal bandwidth, and hence the noise bandwidth is small compared to the carrier frequency, the narrowband representation [22] of the input noise process n(t) may be used. That is, the input noise is of the form

$$n(t) = n_1(t)\cos(\omega_c - \pi W)t + n_2(t)\sin(\omega_c - \pi W)t \qquad (2.44)$$

where $n_1(t)$ and $n_2(t)$ are independent AWGN processes of zero mean and two-sided spectral density $N_0/2$, and $\omega_c - \pi W$ is the centre frequency of the input signal and of the input bandpass filter used to limit the input noise.

After coherent demodulation with $\cos \omega_c t$ and $\sin \omega_c t$ the noise processes in the inphase and quadrature channels are, respectively,

$$n_I(t) = n_1(t) \cos \pi W t - n_2(t) \sin \pi W t \qquad (2.45)$$

$$n_Q(t) = n_1(t) \sin \pi W t + n_2(t) \cos \pi W t \qquad (2.46)$$

Both $n_I(t)$ and $n_Q(t)$ are AWGN processes of zero mean and variance $\sigma^2 = 2(N_0/2)W = N_0 W$ since the noise is bandlimited to W Hz. The autocorrelation function of the inphase noise is defined as [23].

$$R_{n_I}(\tau) = E[n_I(t) n_I(t+\tau)] \qquad (2.47)$$

where $E[.]$ denotes the expectation over time. Substituting for $n_I(t)$ and $n_I(t+\tau)$ gives

$$R_{n_I}(\tau) = E[n_1(t)n_1(t+\tau)\cos\pi Wt \cos\pi W(t+\tau)$$

$$+ n_2(t)n_2(t+\tau)\sin\pi Wt \sin\pi W(t+\tau)$$

$$- n_1(t)n_2(t+\tau)\cos\pi Wt \sin\pi W(t+\tau) \qquad (2.48)$$

$$- n_2(t)n_1(t+\tau)\sin\pi Wt \cos\pi W(t+\tau)]$$

$$= E[n_1(t)n_1(t+\tau)]\, E[\cos\pi Wt \cos\pi W(t+\tau)]$$

$$+ E[n_2(t)n_2(t+\tau)]E[\sin\pi Wt \sin\pi W(t+\tau)]$$

$$- E[n_1(t)n_2(t+\tau)]\, E[\cos\pi Wt \cos\pi W(t+\tau)] \qquad (2.49)$$

$$- E[n_2(t)n_1(t+\tau)]E[\sin\pi Wt \cos\pi W(t+\tau)]$$

Since $n_1(t)$ and $n_2(t)$ are independent Gaussian processes, the expectations involving products of $n_1(t)$ and $n_2(t)$ are always zero. Also the expectations in $n_1(t)$ only and $n_2(t)$ only are recognized as the autocorrelations of $n_1(t)$ and $n_2(t)$. Hence we may simplify (2.49) to obtain

$$R_{n_I}(\tau) = R_{n_1}(\tau) \; E[\cos \pi W t \; \cos \pi W (t+\tau)]$$

$$+ R_{n_2}(\tau) \; E[\sin \pi W t \; \sin \pi W (t+\tau)] \qquad (2.50)$$

where $R_{n_1}(\tau)$ and $R_{n_2}(\tau)$ are the autocorrelations of $n_1(t)$ and $n_2(t)$ respectively. These two autocorrelations are in fact equal since $n_1(t)$ and $n_2(t)$ are processes of the same type. Thus we have

$$R_{n_I}(\tau) = R_{n_1}(\tau) [E[\cos \pi W t \; \cos \pi W (t+\tau)$$

$$+ \sin \pi W t \; \sin \pi W (t+\tau)]] \qquad (2.51)$$

Expanding the angles and performing the indicated expectations, we get

$$R_{n_I}(\tau) = R_{n_1}(\tau) \cos \pi W \tau \qquad (2.52)$$

Now, $n_1(t)$ is AWGN bandlimited to W Hz, its autocorrelation is therefore given by [19]

$$R_{n_1}(\tau) = \frac{N_0}{2\pi\tau} \; \sin \pi W \tau \qquad (2.53)$$

$$= \frac{\sigma^2}{\pi W \tau} \; \sin \pi W \tau \qquad (2.54)$$

Substituting this into (2.52) gives

$$R_{n_I}(\tau) = \frac{\sigma^2}{\pi W \tau} \sin \pi W \tau \cos \pi W \tau \qquad (2.55)$$

$$= \frac{\sigma^2}{2\pi W \tau} \sin 2\pi W \tau \qquad (2.56)$$

Computation of the autocorrelation of the quadrature noise $n_Q(t)$ yields the same result.

We shall now calculate the cross-correlation of the inphase and quadrature noise,

$$R_{n_I n_Q}(\tau) = E[n_I(t) n_Q(t+\tau)] \qquad (2.57)$$

Substituting once again for $n_I(t)$ and $n_Q(t+\tau)$, we have

$$R_{n_I n_Q}(\tau) = E[n_1(t) n_1(t+\tau) \cos \pi W t \sin \pi W(t+\tau)]$$

$$- E[n_2(t) n_2(t+\tau) \sin \pi W t \cos \pi W(t+\tau)]$$

$$+ E[n_1(t) n_2(t+\tau) \cos \pi W t \cos \pi W(t+\tau)]$$

$$- E[n_2(t) n_1(t+\tau) \sin \pi W t \sin \pi W(t+\tau)]$$

$$(2.58)$$

$$= E[n_1(t) n_1(t+\tau)] E[\cos \pi W t \sin \pi W (t+\tau)]$$

$$- E[n_2(t) n_2(t+\tau)] E[\sin \pi W t \cos \pi W (t+\tau)]$$

$$+ E[n_1(t) n_2(t+\tau)] E[\cos \pi W t \cos \pi W (t+\tau)]$$

$$- E[n_2(t) n_1(t+\tau)] E[\sin \pi W t \sin \pi W (t+\tau)] \quad (2.59)$$

Once again the first two noise expectations are the autocorrelations of $n_1(t)$ and $n_2(t)$ and the last two are zero because $n_1(t)$ and $n_2(t)$ are independent. Also, $R_{n_1}(\tau) = R_{n_2}(\tau)$, and so (2.59) simplifies to

$$R_{n_I n_Q}(\tau) = R_{n_1}(\tau) \{ E[\cos \pi W t \sin \pi W (t+\tau)]$$

$$- E[\sin \pi W t \cos \pi W (t+\tau)] \} \quad (2.60)$$

Expanding the angles and performing the expectations gives

$$R_{n_I n_Q}(\tau) = R_{n_1}(\tau) \sin \pi W \tau \quad (2.61)$$

Substituting for $R_{n_1}(\tau)$ from (2.54) gives

$$R_{n_I n_Q}(\tau) = \frac{\sigma^2}{\pi W \tau} \sin 2 \pi W \tau \quad (2.62)$$

$$= \frac{\sigma^2}{2\pi W\tau} \ (1-\cos \ 2\pi W\tau) \qquad (2.63)$$

This cross-correlation is plotted in Fig. 2.8 along with the autocorrelation (2.56). We see that successive samples of the inphase (or quadrature) noise are independent if the noise is sampled at the Nyquist frequency 2W. As for the autocorrelation, we see that samples taken from the inphase and quadrature channels at the Nyquist rate are correlated if the offset between the two sampling instants is an odd multiple of the sampling period. When the channels are sampled at the same time ($\tau=0$), the two noises are uncorrelated.

A double sideband signal transmitting the same signal requires a bandwidth of 2W Hz. For the DSB signal, the noise centre frequency and the carrier are the same. The narrowband noise at the receiver input would then be

$$n_{DSB}(t) = n_1' \ (t)\cos\omega_c t + n_2' \ (t)\sin \omega_c t \qquad (2.64)$$

where n_1' (t) and n_2' (t) are independent AWGN processes of zero mean and spectral density $N_0/2$ W/Hz. Coherent demodulation gives

$$n_I'' \ (t) = n_1' \ (t) \qquad (2.65)$$
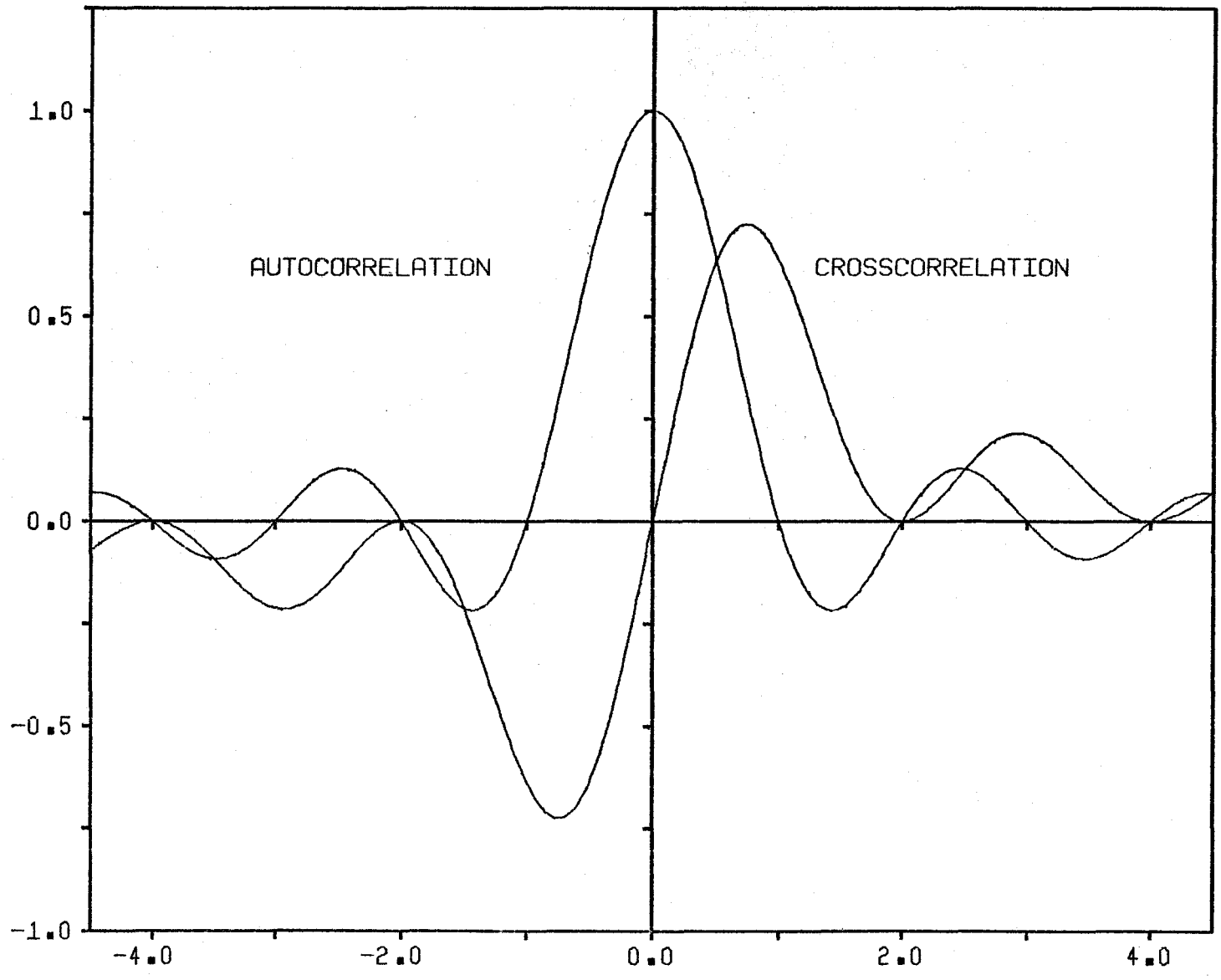
Offset-Nyquist Frequency Product 2Wτ

Figure 2.8   SSB Noise Autocorrelation and Cross-Correlation

$$n_Q{}'(t) = n_2{}'(t) \qquad (2.66)$$

as the inphase and quadrature noise processes. The corresponding autccorrelations are

$$R_{n_I}{}'(\tau) = R_{n_Q}{}'(\tau)$$

$$= \frac{\sigma^2}{2\pi W \tau} \sin 2\pi W \tau \qquad (2.67)$$

which are the same as for the SSB noise. By contrast, the cross-correlaticn of the DSB inphase and quadrature noise is zero for all values of τ since $n_1{}'(t)$ and $n_2{}'(t)$ are independent. The unusual cross-correlation of the SSB noise is a consequence of the noise centre frequency not having the same value as the signal carrier frequency.

2.6 Conclusions

Several methods of modulating and demodulating an SSB signal were considered. It was shown that no matter what modulation method was used, generation of the SSB signal is much easier when no low frequency energy is present in the baseband signal. Thus the baseband signal should be processed to remove the low frequency components before modulaticn.

It was also shown that from the point of view of complexity, the filter or Hartley modulators should be used while coherent detection should be used in the receiver to demodulate the signal.

The autocorrelation and cross-correlation of the inphase and quadrature noise in the SSB receiver were derived. Since the cross-correlation is significantly different from the cross-correlation of noise in a DSB system, care should be exercised when both the inphase and quadrature channels of an SSB receiver are sampled.

CHAPTER 3.   DIGITAL SSB TRANSMISSION

3.1 Introduction

In this chapter we shall discuss the implications of SSB for digital communication. To make the task of sideband isolation easier, we shall need to shape the spectrum of the data signal through the use of partial response encoding. The consequences of this shaping will be presented and the performance of the resulting system in a Gaussian noise environment will be analyzed. The possibility of extracting the data sequence from the quadrature channel will be discussed and a novel method of demodulating an SSB data signal will be presented.

3.2 Partial Response Signalling

Partial response signalling is a technique whereby it is possible to signal at a rate of 1/T symbols/S through a bandwidth 1/2T Hz using realizable and perturbation tolerant filters. The partial response concept was originally introduced by Lender [24] and then extended by Kretzmer [25]. It may be viewed either as a coding

operation or as a filtering operation. A comprehensive performance comparison of various partial response codes is contained in [26]. The technique has found application in a variety of transmission systems ranging from baseband to microwave [27] - [29].

In pulse-amplitude modulation (PAM) systems [30], it is undesirable to have intersymbol interference (ISI) present as this leads inevitably to performance degradations and PAM systems are designed to minimize the amount of ISI present. In partial response (PR) systems, this ISI is introduced intentionally with the rationale that if the ISI is known, it can be much more easily controlled. Additional benefits such as signal spectrum shaping and a decreased sensitivity to symbol timing errors also result. These benefits are obtained at the cost of a slight SNR penalty in the receiver.

A partial response symbol a_k is obtained from a superposition of the N previous input symbols c_k. That is

$$a_k = \sum_{n=0}^{N-1} f_n c_{k-n} \qquad (3.1)$$

where the f_n are the weighting coefficients of the nth delay. The code may also be characterized by the system

polynomial F(D) where

$$F(D) = \sum_{n=0}^{N-1} f_n D^n \tag{3.2}$$

where D is the unit delay operator. The frequency response of the encoder is then

$$F(\omega) = \sum_{n=0}^{N-1} f_n e^{-j\omega nT} \qquad \forall \omega \tag{3.3}$$

where T is the symbol period. If the encoder is followed by a filter $G(\omega)$ restricting the signal to the Nyquist bandwidth $1/2T$ with

$$G(\omega) = T \qquad |\omega| < \pi/T \tag{3.4}$$

$$= 0 \qquad \text{elsewhere}$$

then the overall encoder/filter combination of Fig. 3.1 has a frequency response of

$$H(\omega) = T \sum_{n=0}^{N-1} f_n e^{-j\omega nT} \qquad |\omega| < \pi/T \tag{3.5}$$

$$= 0 \qquad \text{elsewhere}$$

and an impulse response of

$$h(t) = \sum_{n=0}^{N-1} f_n \frac{\sin\pi(t-nT)/T}{\pi(t-nT)/T} \tag{3.6}$$
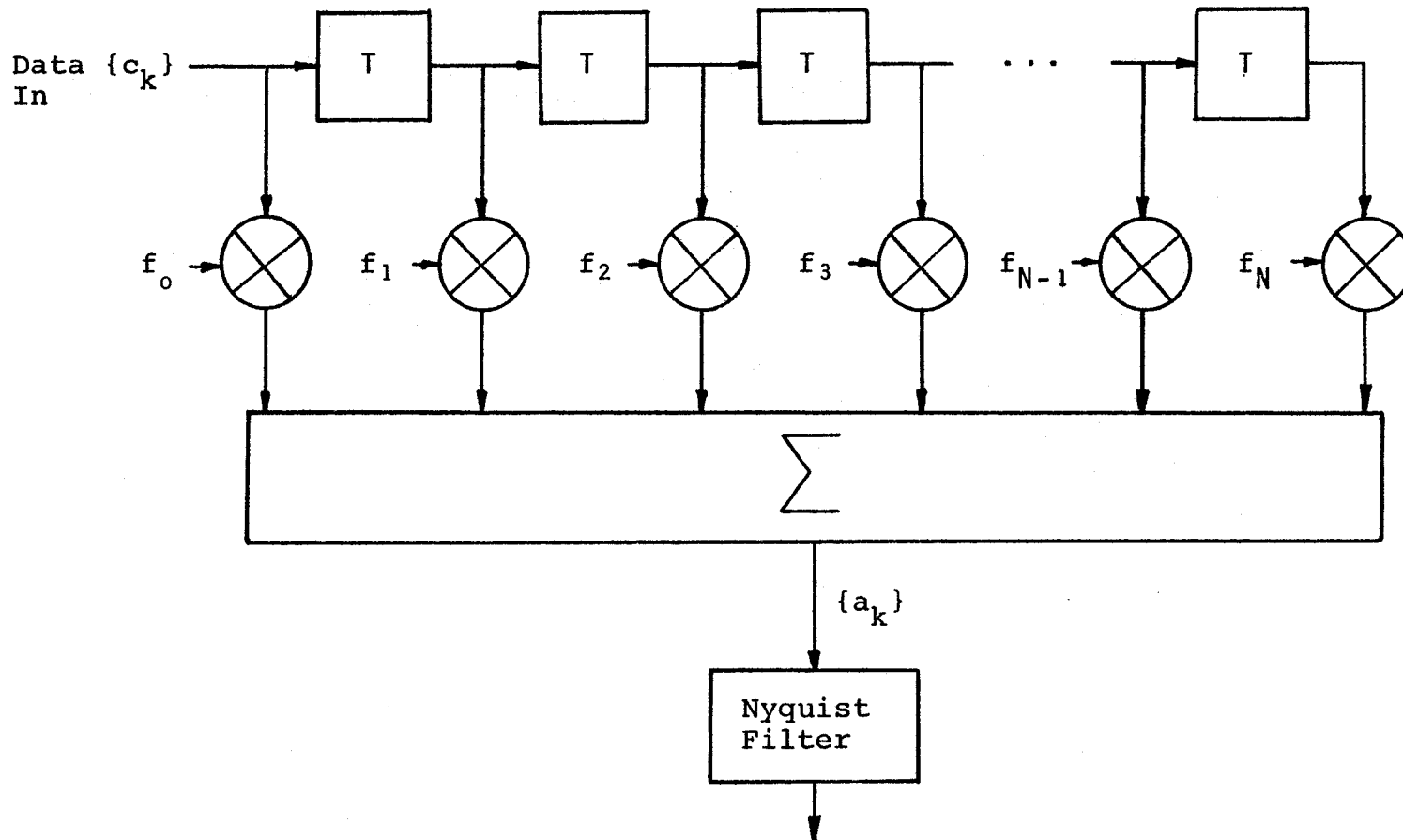
Minimum Bandwidth Encoded Output

Figure 3.1 Generalized Partial Response Encoder

By varying the weighting coefficients $\{f_n\}$ different pulse shapes and frequency characteristics can be obtained. The number of output levels is no longer the same as the number of input levels but is also a function of the weighting coefficients. Table 3.1 illustrates the frequency response of several partial response codes. The corresponding impulse responses as well as the Hilbert Transforms of the impulse responses are shown in Table 3.2 (see Appendix A for details of the pertinent derivations).

It can be seen from Table 3.1 that the 1+D or Class 1 pulse is unsuitable for SSB applications because it concentrates the signal energy in the neighbourhood of DC. The 1-D pulse has the DC null required for easy sideband isolation but cannot be used in the minimum bandwidth (i.e. it cannot be bandlimited to the Nyquist bandwidth) because of the discontinuity at $\omega = \pi/T$. The remaining pulses are all suitable for use in an SSB system due to their spectral shapes. However, the $1-D-D^2+D^3$, $1+D-D^2-D^3$ and $1-2D^2+D^4$ polynomials all produce 4L-3 output levels for an L level input while the $1-D^2$ code produces only 2L-1 levels. Since the susceptibility of the receiver to make errors is greater when the number of levels it must distinguish between increases, the $1-D^2$ code will perform better than the other three. Thus we select the $1-D^2$ or Class 4 code as the code

SYSTEM POLYNOMIAL	NUMBER OF OUTPUT LEVELS	FREQUENCY RESPONSE	$\lvert H(\omega) \rvert$
$1 + D$	$2L - 1$	$2T\cos\omega T/2$	
$1 - D$	$2L - 1$	$j2T\sin\omega T/2$	
$1 - D^2$	$2L - 1$	$j2T\sin\omega T$	
$1 - D - D^2 + D^3$	$4L - 3$	$-4T\sin\dfrac{\omega T}{2}\sin\omega T$	
$1 + D - D^2 - D^3$	$4L - 3$	$j4T\cos\dfrac{\omega T}{2}\sin\omega T$	
$1 - 2D^2 + D^4$	$4L - 3$	$-4T\sin^2\omega T$	

TABLE 3.1   CHARACTERISTICS OF SEVERAL MINIMUM
BANDWIDTH PARTIAL
RESPONSE SYSTEMS

SYSTEM POLYNOMIAL	NUMBER OF OUTPUT LEVELS	IMPULSE RESPONSE	HILBERT TRANSFORM OF IMPULSE RESPONSE
$1 + D$	$2L - 1$	$\dfrac{4T^2}{\pi}\ \dfrac{\cos\pi t/T}{T^2-4t^2}$	$\dfrac{4T}{\pi}\ \dfrac{2t-T\sin\pi t/T}{4t^2-T}$
$1 - D$	$2L - 1$	$\dfrac{8T}{\pi}\ \dfrac{t\cos\pi t/T}{4t^2-T^2}$	$\dfrac{4T}{\pi}\ \dfrac{2t\sin\pi t/T-T}{4t^2-T^2}$
$1 - D^2$	$2L - 1$	$\dfrac{2T^2}{\pi}\ \dfrac{\sin\pi t/T}{t^2-T^2}$	$\dfrac{-2T^2}{\pi}\ \dfrac{1+\cos\pi t/T}{t^2-T^2}$
$1 - D - D^2 + D^2$	$4L - 3$	$\dfrac{16T^2}{\pi}\ \dfrac{(4t^2-3T^2)\cos\pi t/T}{(4t^2-T^2)(4t^2-9T^2)}$	$\dfrac{16T^2}{\pi}\ \dfrac{4Tt+(4t^2-3T^2)\sin\pi t/T}{(4t^2-T^2)(4t^2-9T^2)}$
$1 + D - D^2 - D^3$	$4L - 3$	$\dfrac{-64T^3}{\pi}\ \dfrac{t\cos\pi t/T}{(4t^2-T^2)(4t^2-9T^2)}$	$\dfrac{16T^2}{\pi}\ \dfrac{3T^2-4t^2-4Tt\sin\pi t/T}{(4t^2-T^2)(4t^2-9T^2)}$
$1 - 2D^2 + D^4$	$4L - 3$	$\dfrac{8T^3}{\pi}\ \dfrac{\sin\pi t/T}{t(t^2-4T^2)}$	$\dfrac{8T^3}{\pi}\ \dfrac{1-\cos\pi t/T}{t(t^2-4T^2)}$

TABLE 3.2   IMPULSE RESPONSES OF SEVERAL MINIMUM
BANDWIDTH PARTIAL RESPONSE SYSTEMS AND
THEIR HILBERT TRANSFORMS

which provides the required spectral shape with the minimum number of output levels.

3.2.1 The Class 4 Partial Response Code

We shall now consider the Class 4 (PR4) pulse in more detail. The impulse response of the system is

$$h(t) = \frac{2T^2}{\pi} \frac{\sin\pi t/T}{t^2 - T^2} \tag{3.7}$$

and the Hilbert Transform of this pulse is

$$\hat{h}(t) = - \frac{2T^2}{\pi} \frac{1+\cos\pi t/T}{t^2 - T^2} \tag{3.8}$$

These two pulses are plotted in Fig. 3.2.

The output of a minimum bandwidth PR4 encoder is then

$$a(t) = \sum_{n=-\infty}^{\infty} c_n h(t-nT) \tag{3.9}$$

while its Hilbert Transform is

$$\hat{a}(t) = \sum_{n=-\infty}^{\infty} c_n \hat{h}(t-nT) \tag{3.10}$$
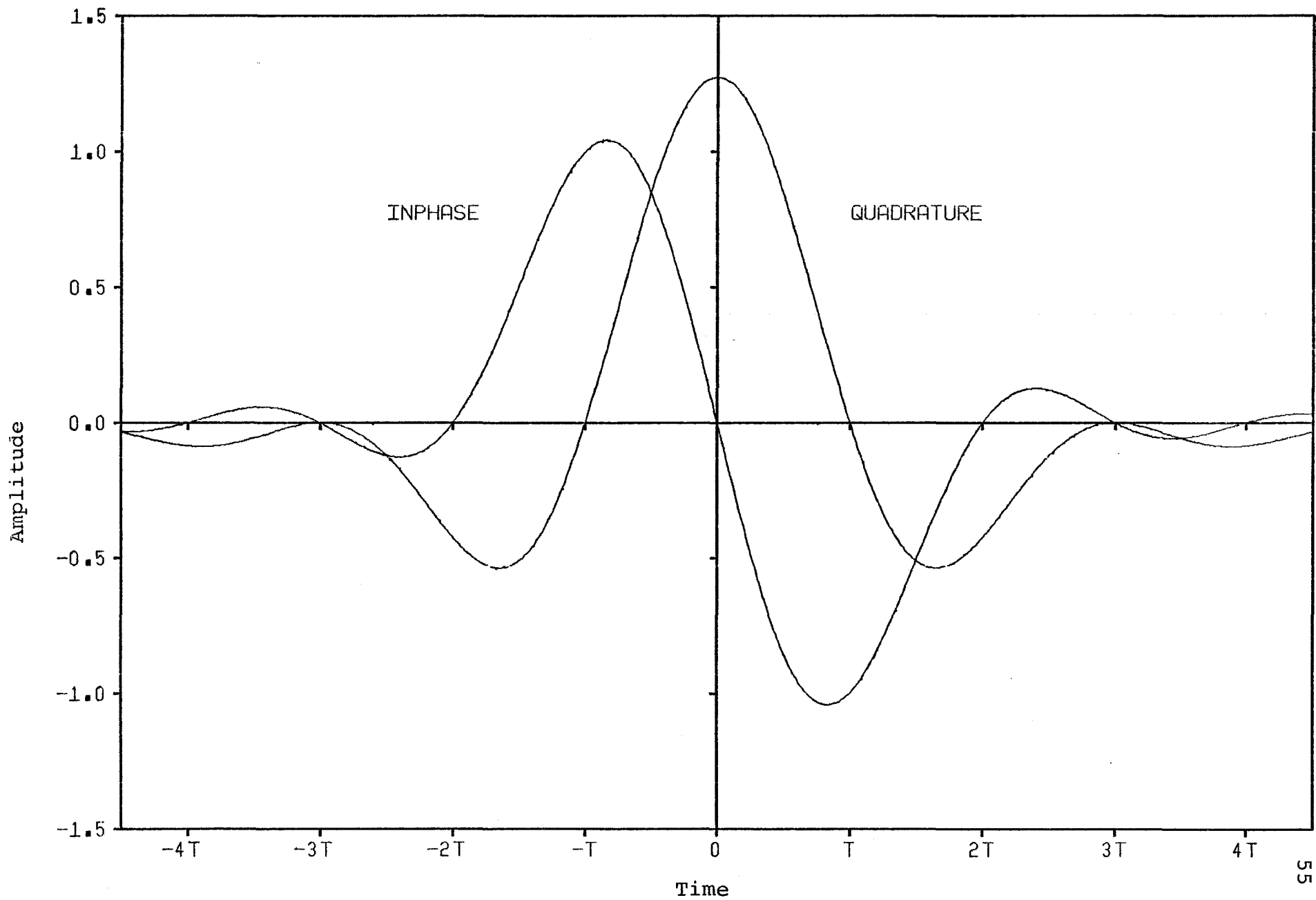
Figure 3.2 Partial Reponse Class 4 Pulse and its Hilbert Transform
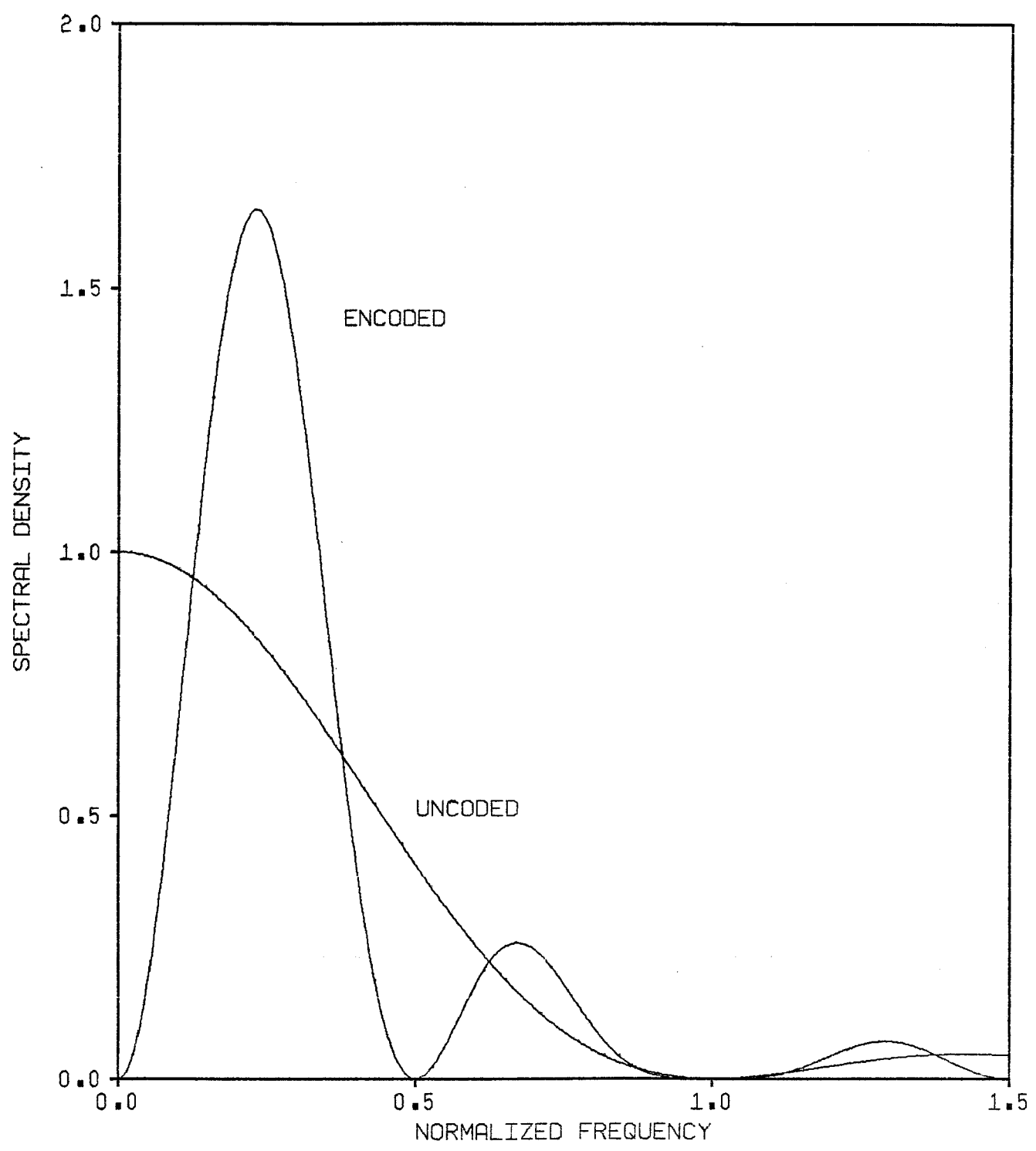
Figure 3.3 Effect of Partial Response Encoding
in the Frequency Domain

where $h(t)$ and $\hat{h}(t)$ are defined by (3.7) and (3.8) and the $\{c_n\}$ are L level equally likely input symbols $\pm 1, \pm 3, \ldots, \pm (L-1)$.

The operation of the non-minimum bandwidth Class 4 encoder in the frequency domain is shown in Fig. 3.3. The encoder input consists of a sequence of rectangular binary pulses normalized to unit power and unit symbol period (T=1). According to Nyquist, this signal could be transmitted in a bandwidth equal to half of the data rate (f = 0.5). This would require a filter with an infinitely steep cutoff. Also, the spectral density attains its maximum value at DC which would require another steep cutoff filter for sideband isolation in order to utilize SSB modulation for transmission of the data.

The encoded spectrum, however, has nulls at DC and at the Nyquist bandwidth. Thus the signal can be limited to the minimum bandwidth through use of a filter of only modest rolloff. The same is true of the sideband isolation filter. The net effect of the encoding operation is to introduce a correlation between output symbols which concentrates the signal energy in a region around $f = 0.25$ rather than at DC.

Sampling the minimum bandwidth encoder output (3.9)

at $t = kT$, we have

$$a(kT) = c_{k+1} - c_{k-1} \tag{3.11}$$

For uniformly distributed input symbols $\{c_k\}$, the output level distribution is no longer uniform but triangular with a probability distribution

$$P[a_k = 2m] = (L - |m|)/L^2 \tag{3.12}$$

for $m = \pm 1$, ± 2, ..., $\pm(L-1)$ where $P[.]$ denotes the probability of the quantity contained within the brackets. Such an L level baseband system operating in a bandwidth W Hz sends data at a rate [30].

$$R = 2W\log_2 L \tag{3.13}$$

The bandwidth efficiency η is the ratio of data rate to bandwidth or

$$\eta = \frac{R}{W}$$

$$= 2 \log_2 L \tag{3.14}$$

A binary input system (L=2) would have an efficiency of 2 bits/S/Hz while a 4 level system would have an efficiency of 4 bits/S/Hz. These correspond to 3 and 7 levels at the output of the encoder. Since an SSB signal occupies the same bandwidth as the baseband signal, the RF efficiency is also given by (3.14). In our study, we shall be concerned primarily with the L=2 and L=4 cases.

If we sample the Hilbert Transform pulse of (3.8) at t=kT, we have

$$\hat{h}(kT) = \frac{-4}{\pi(k^2-1)} \qquad k = 0, \; k \; odd \qquad (3.15)$$

$$= 0 \qquad otherwise$$

This suggests that the Hilbert Transform of the partial response encoded data signal can be implemented in a fashion somewhat similar to the generalized partial response encoder of Fig. 3.1, using a delay line structure followed by a filter limiting the spectrum to the Nyquist bandwidth as shown in Fig. 3.4. The individual top weights are the sampled quadrature impulse response values $\hat{h}_k = \hat{h}(kT)$. In a strict sense, the quadrature pulse is not realizable as an infinite number of delays must be used to synthesize the
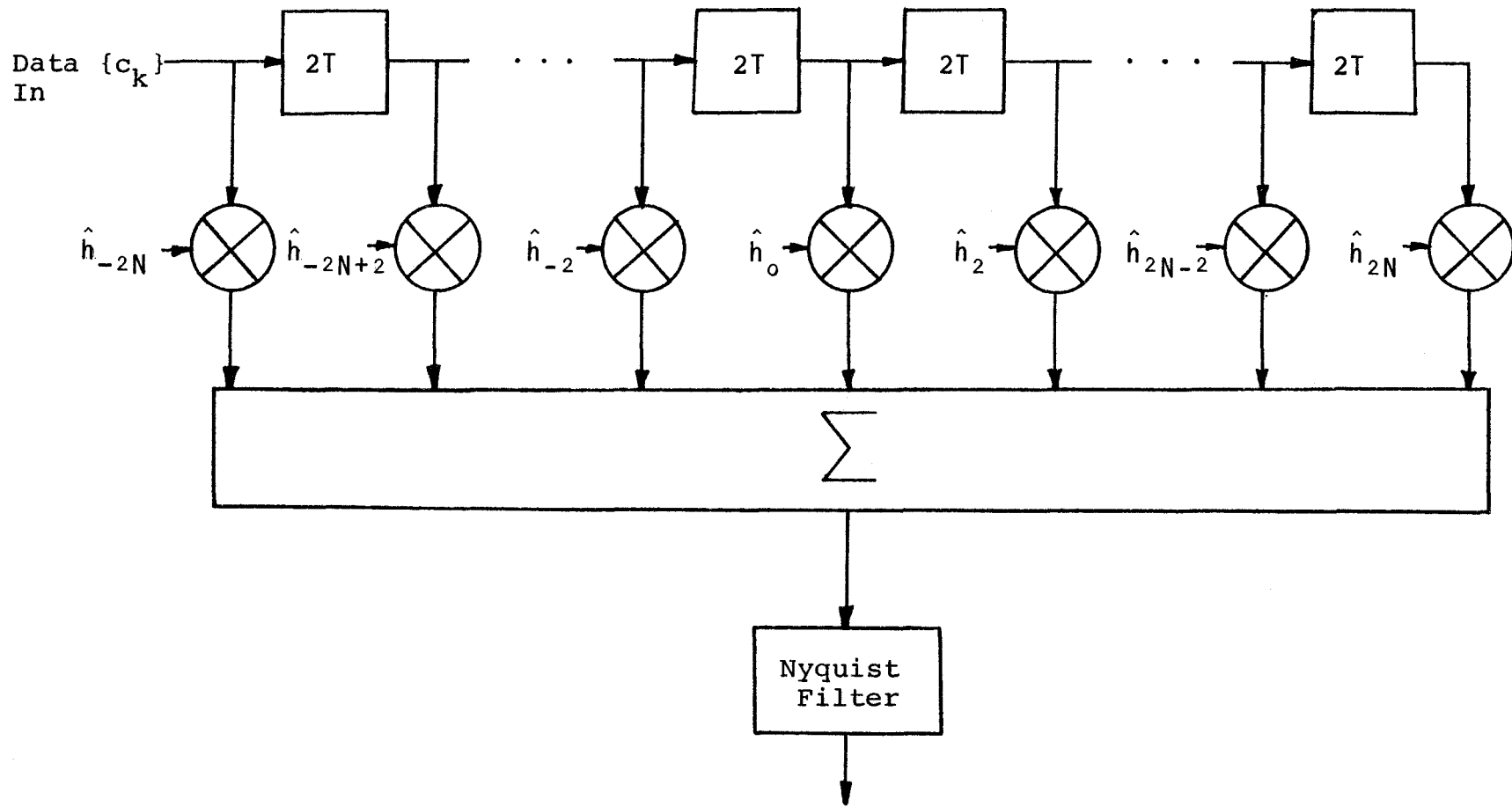
PR4 Hilbert Transorm Signal

Figure 3.4   SSB-PR4 Quadrature Signal Generator

waveform exactly. However, since the pulse decays as $1/t^2$, a reasonable approximation can be obtained with a finite number of delays. Had the $1-2D^2 + D^4$ code been used instead, even fewer delays would be necessary as this pulse decays as $1/t^3$. An ordinary bandlimited PAM pulse decays as $1/t$ so that a very large number of delays would be required for a reasonable sideband rejection. The quadrature shaping would be used in conjunction with a Hartley modulator as shown in Fig. 3.5.

Up to this point, generation of the partial response signal in baseband only has been considered. Since the encoding is a filtering operation, it may be done at RF as well. Although design of the filter would be much more difficult, there are instances where such an approach has been used to bypass non-linear amplifiers [29]. A possible implementation using this approach is shown in Fig. 3.6.

3.2.2 SSB-PR4 Eye Diagram and Signal Space

A useful concept in the study of digital transmission systems is the eye diagram [28]. The eye diagram is obtained by triggering an oscilloscope at the symbol rate $1/T$. What appears on the screen is a superposition of a large number of pulses. From this
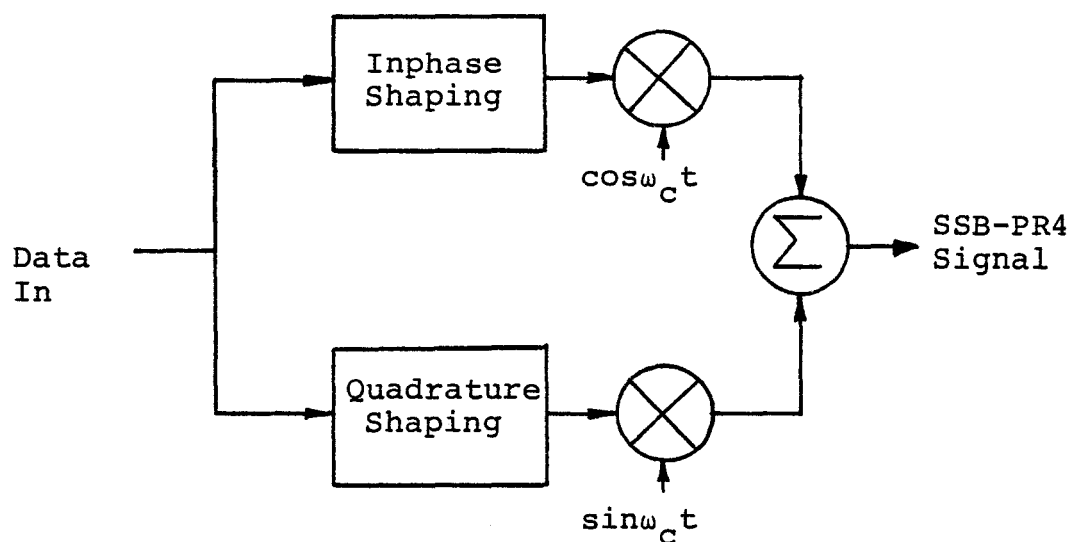
Figure 3.5   Hartley Modulator Using Partial Response
Encoding at Baseband
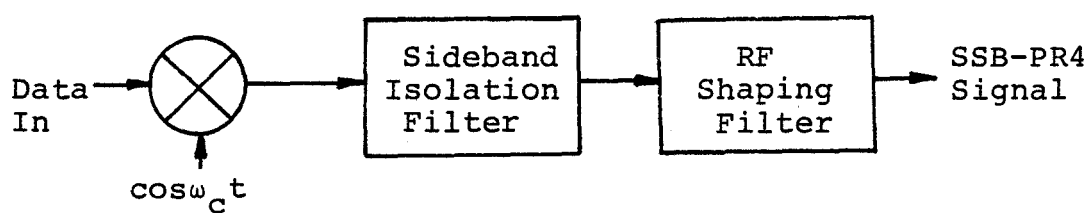
Figure 3.6   SSB Modulator Using Partial Response
Encoding at RF

diagram, certain important parameters can be obtained. The system's sensitivity to noise can be determined from the vertical height or eye opening; the optimum sampling point occurs when the eye opening is at its maximum. The optimum receiver decision threshold may be determined and the sensitivity of the system to timing error can be seen from the slope of the eye boundary at the sampling instant.

Figs. 3.7 and 3.8 are the eye diagrams corresponding to the output of a minimum bandwidth PR4 encoder for binary and four level inputs. We see that the eye opening is maximized if the signal is sampled at the nominal sampling instant t=kT. The multilevel output of the encoder is very well illustrated as is the non-uniform distribution of the output levels. A close examination of the transitions in the eye diagram reveals that certain output sequences are not present. Several transitions are not allowed in the PR4 output sequences so that if these sequences are detected at the receiver, it is known that an error has occured. This permits receiver error rate monitoring [32] and, to a limited extent, receiver error connection [27].

Another useful concept in digital communications systems is that of the signal space [33]. This is simply a plot of the locus of the inphase and quadrature components
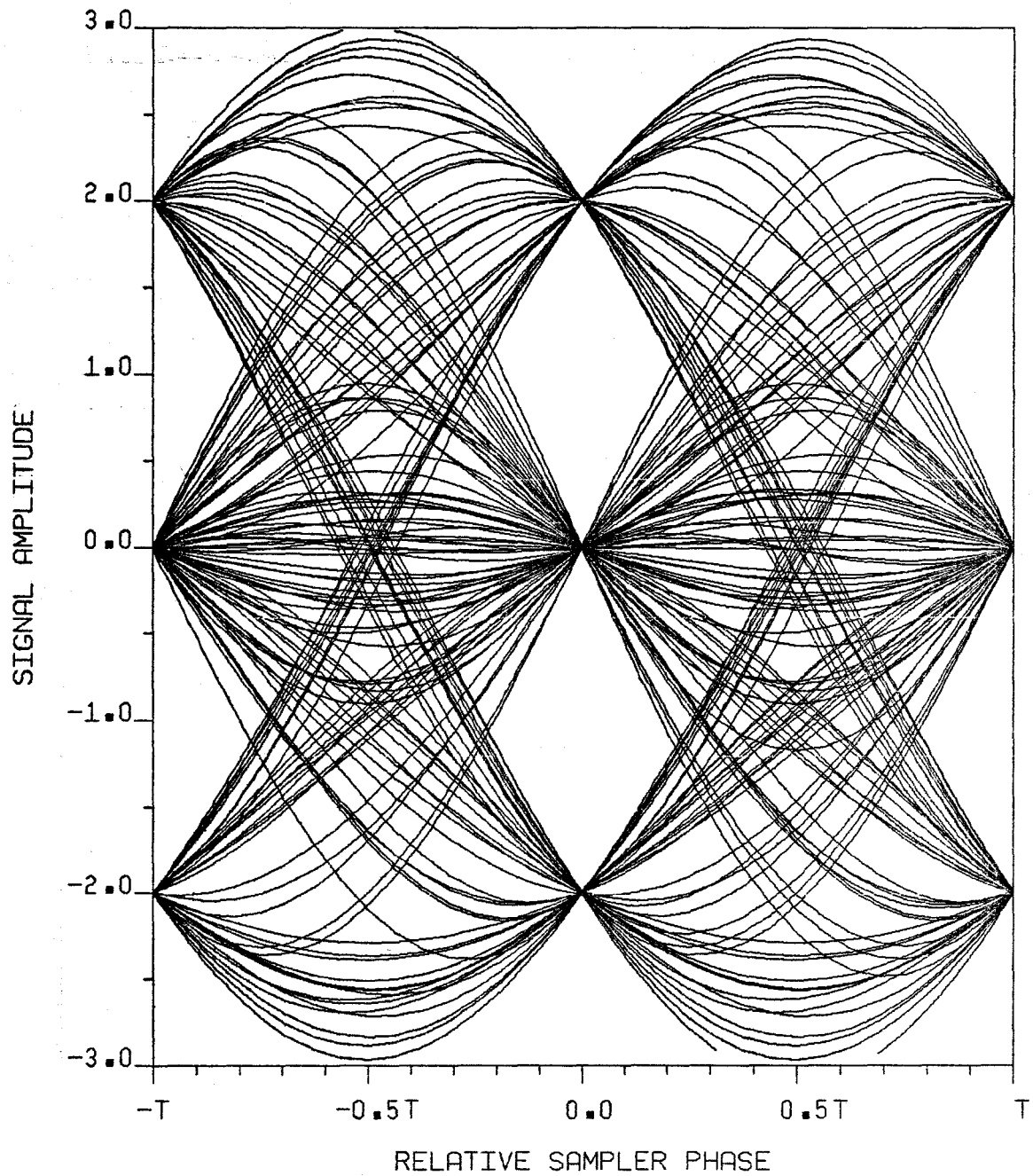
Figure 3.7   Minimum Bandwidth PR4 Encoder Output
Eye Diagram With Binary Input
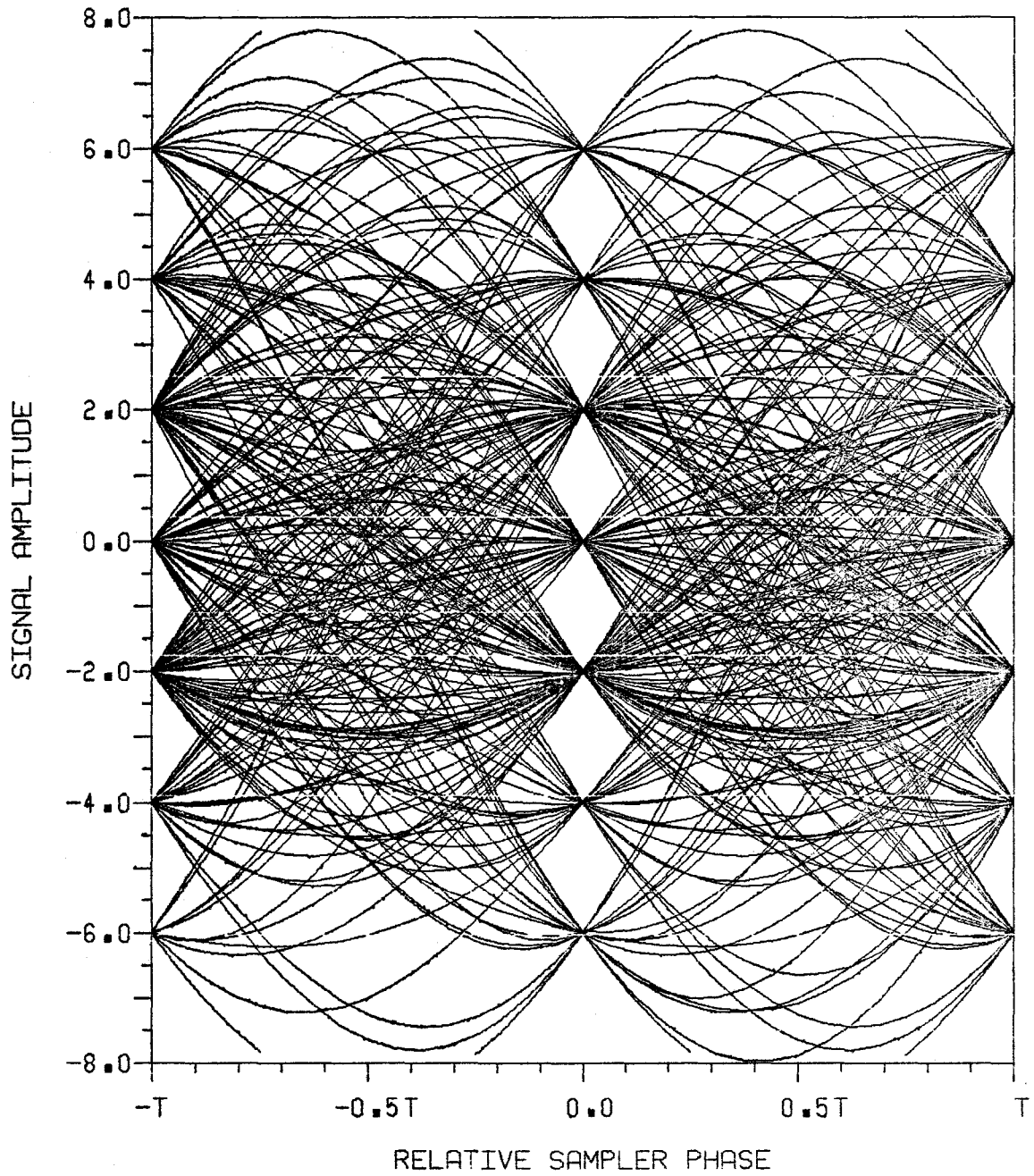
Figure 3.8   Minimum Bandwidth PR4 Encoder Output
Eye Diagram With 4 Level Input

of the transmitted signal for all possible data sequences. For a binary input PR4 signal, the inphase component has three possible values: 0, ±2. The quadrature component is not as easily derived. However, sampling (3.10) at t = kT and making use of (3.15), we have

$$\hat{a}(kT) = \frac{4}{\pi} c_k - \frac{4}{\pi} \sum_{\substack{m \text{ even}}}^{\infty} \frac{c_{k-m}}{m^2-1}$$

$$= \frac{4}{\pi} c_k - \frac{4}{\pi} \sum_{\ell=1}^{\infty} \frac{c_{k+2\ell} + c_{k-2\ell}}{4\ell^2-1} \qquad (3.16)$$

Since this expression depends on an infinite number of symbols, evaluation of all values of a(kT) is not possible. However, it is possible to develop a bound for this quadrature component. Setting $c_k = +1$ and $c_{k+2\ell} = c_{k-2\ell} = -1$ for $\ell \neq 0$ we have

$$\hat{a}(kT) \leqslant \frac{4}{\pi} \left(1 + 2 \sum_{\ell=1}^{\infty} \frac{1}{4\ell^2-1} \right) \qquad (3.17)$$

$$= \frac{8}{\pi}$$

where we have used the fact that the series converges uniformly to 1/2 [34]. If we next set $c_k = -1$ and $c_{k+2\ell} = c_{k-2\ell} = +1$ we obtain
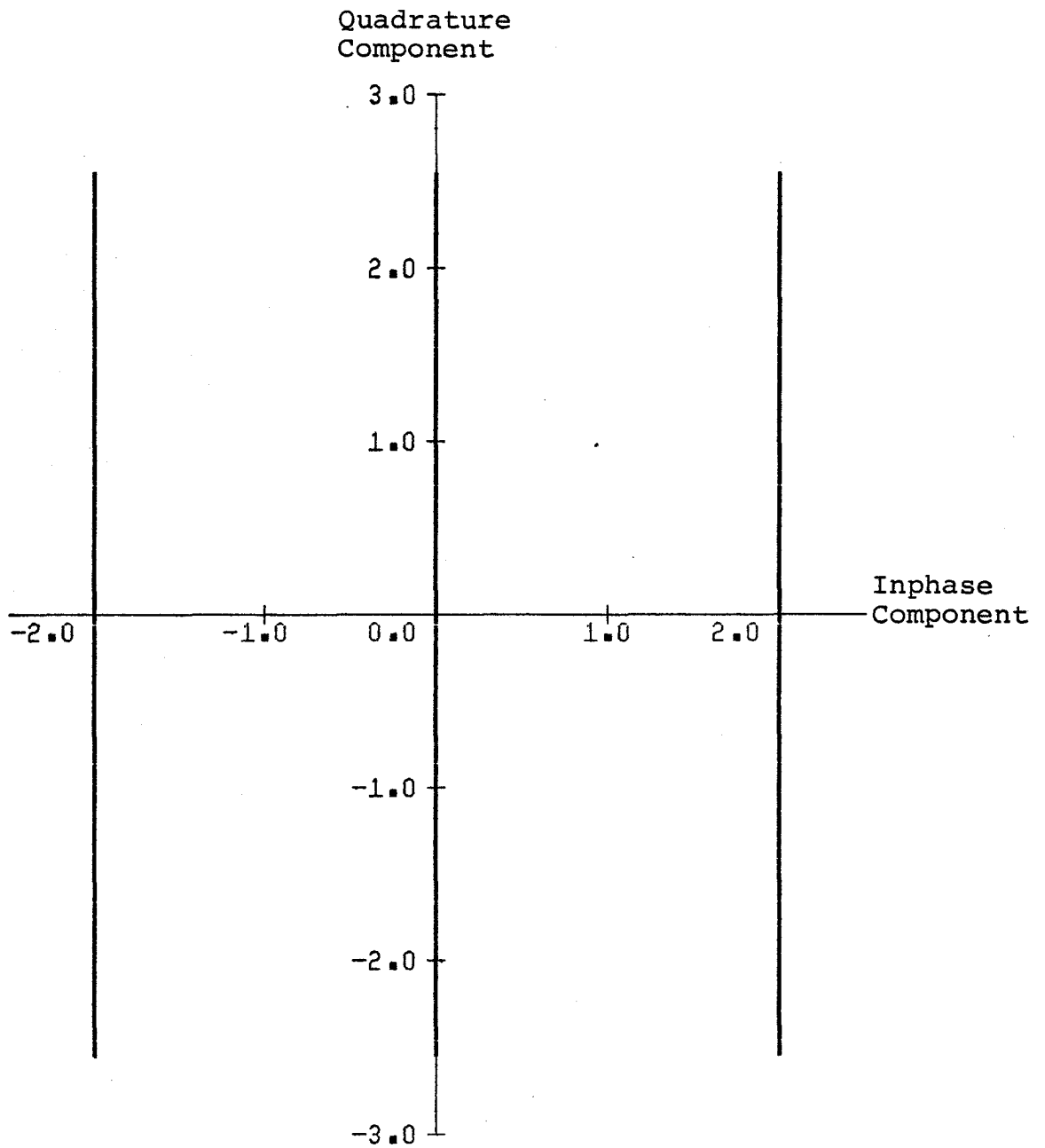
Figure 3.9 SSB-PR4 Signal Space With Binary Input

$$\hat{a}(kT) \geqslant \frac{4}{\pi} \left(1 + 2 \sum_{\ell=1}^{\infty} \frac{1}{4\ell^2 - 1} \right) \tag{3.18}$$

$$= \frac{8}{\pi}$$

Thus we have $|\hat{a}(kT)| \leqslant 8/\pi$.

The corresponding signal space is plotted in Fig. 3.9. The quadrature component is shown as having a continuous value ranging from $-8/\pi$ to $8/\pi$ but it should be remembered that the quadrature component actually consists of an infinite number of discrete points between these values.

3.3 Partial Response Decoding

Equation (3.11) suggests a direct method for recovery of the original L level data sequence from the received partial response encoded symbols. We see that if a value for c_{k-1} has been decided, then it may be added to the current received symbol a_k to yield the original L-ary symbol c_{k+1}. This is known as "decision feedback decoding" and the implementation of such a decoder is shown in Fig. 3.10. An L level decision is made, delayed by 2 symbol periods and then added to the current received symbol at which point another decision is made. The disadvantage of
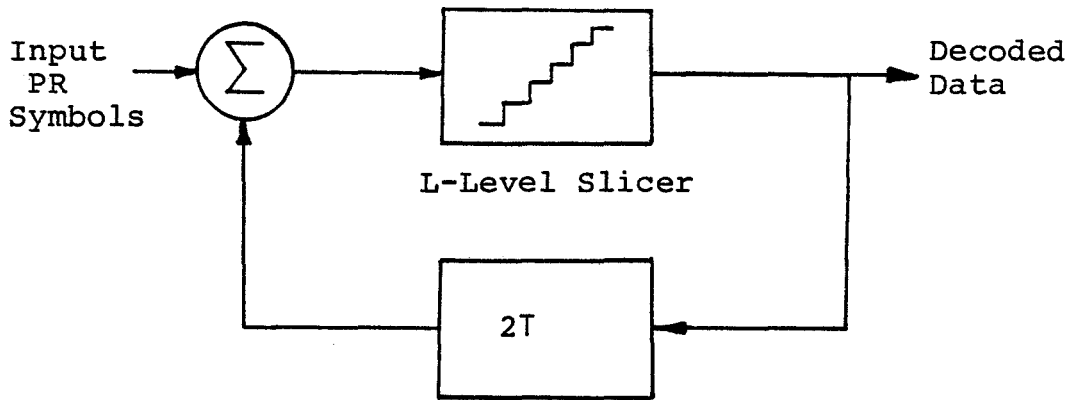
Figure 3.10   Decision Feedback Partial Response Decoder
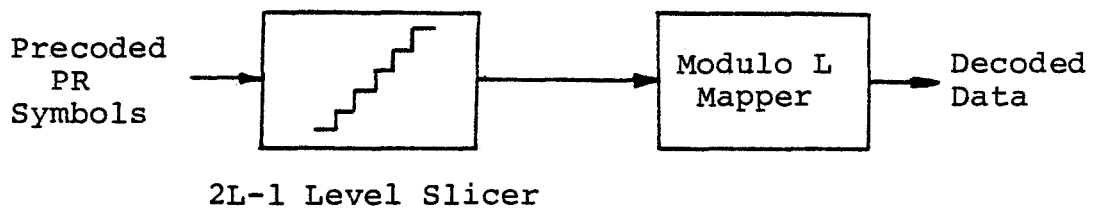
Figure 3.11   Precoded Symbol Partial Response Decoder

this decoder is that once an error is made, subsequent decisions are likely to be in error as well.

Lender [24] devised a method for eliminating this error propagation effect by precoding the input sequence $\{c_k\}$ according to the rule

$$b_k = c_k + b_{k-2} \quad , \text{ mod } L \qquad (3.19)$$

The transmitted PR symbol is then

$$a_k = b_{k+1} - b_{k-1} \qquad (3.20)$$

For binary inputs, $L = 2$ and the precoding operation is a modulo 2 addition. If $c_k = -1$ then $b_k = b_{k-2}$ so that a_{k-1} will always be zero. Alternatively, if $c_k = 1$, then b_k is the reverse of b_{k-2} and a_{k-1} will be ± 2. Therefore, decoding can be accomplished according to the rule

$$c_k = -1 \text{ if } a_{k-1} = 0$$

$$c_k = +1 \text{ if } a_{k-1} = \pm 2 \qquad (3.21)$$

A binary version of this decoder can easily be

implemented by means of a full-wave rectifier followed by a threshold detector. For an L level input signal, the received symbols are decoded modulo L to recover the original sequence $\{c_k\}$. The received symbols are sliced to one of 2L-1 levels and then mapped into the original sequence as shown in Fig. 3.11.

Maximum-likelihood sequence (MLS) decoding can also be used to decode PR codes [35]. These decoders utilize the Viterbi Algorithm [36] for the decoding process and can perform much better than the two simple decoders which have just been outlined as they examine a received sequence of symbols in order to make a decision rather than making decisions on a symbol by symbol basis. This improved performance is obtained at the expense of a significant increase in circuit complexity.

3.3.1 Performance in Gaussian Noise

We shall now analyze the performance of the precoded and decision feedback decoders in Gaussian noise.

The inphase output of a perfectly synchronized SSB demodulator is

$$r_I(t) = a(t) + n_I(t) \tag{3.22}$$

where

$$n_I(t) = n_1(t)\cos \pi t/2T + n_2(t)\sin \pi t/2T \tag{3.23}$$

is AWGN of zero mean and variance σ^2 and $a(t)$ is defined by (3.9). Sampling this output at $t = kT$ gives

$$r_I(kT) = R_I + n_K \tag{3.24}$$

where

$$R_I \stackrel{\Delta}{=} a(kT) \tag{3.25}$$

$$n_K \stackrel{\Delta}{=} n_I(kT) \tag{3.26}$$

From (3.11), we have

$$R_I = c_{k+1} - c_{k-1} \tag{3.27}$$

with $c_k = \pm d, \pm 3d, \ldots, \pm(L-1)d$

Since the c_k are in odd increments of d, the received PR symbol $R_I = c_{k+1} - c_{k-1}$ will always be an even

multiple of d,

$$R_I = 2md, \quad m = 0, \pm1, \pm2, \ldots, \pm(L-1) \qquad (3.28)$$

We shall first consider the precoded receiver. Letting $\lambda_k = r_I(kT)$ denote the decision device input, the decision boundaries for a correct decision for the received symbol R_I are

$$(2m-1)d < \lambda_k < (2m+1)d \qquad (3.29)$$

for all received symbols except the outer levels ($|m| \neq L-1$). Therefore, the probability of error for any of the inner levels is

$$P[E|R_I = 2md, |m| \neq L-1] = 1 - P[(2m-1)d < \lambda_K < (2m+1)d]$$

$$= 1 - P[-d < n_k < d] \qquad (3.30)$$

where $P[.]$ denotes the probability of .. Since the noise is AWGN of zero mean and variance σ^2, we have

$$P[-d < n_k < d] = \int_{-d}^{d} \frac{1}{\sqrt{2\pi}} \exp \frac{-Z^2}{2\sigma^2} \, dZ \qquad (3.31)$$

$$= 1 - 2Q\left(\frac{d}{\sigma}\right) \qquad (3.32)$$

where $Q(x)$ is a normalized form of the complementary error

function defined by

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_X^\infty \exp - \frac{Z^2}{2} \, dZ \qquad (3.33)$$

Thus, we have after substitution,

$$P[E|R_I = 2md, \ |m| \neq L-1] = 2Q\left(\frac{d}{\sigma}\right) \qquad (3.34)$$

For the outer levels, the decision regions are different. An incorrect decision is made for $R_I = 2(L-1)d$ if

$$\lambda_k < 2(L-1)d-d = (2L-3)d \qquad (3.35)$$

so that

$$P[E|R_I = 2(L-1)d] = P[\lambda_k < (2L-3)d]$$

$$= P[n_k < -d]$$

$$= Q\left(\frac{d}{\sigma}\right) \qquad (3.36)$$

Similarly for $R_I = -2(L-1)d$, an error occurs when

$$\lambda_k \quad > -2(L-1)d+d = -(2L-3)d \tag{3.37}$$

Hence

$$P[E \mid R_I = -2(L-1)d] = P[\lambda_k > -(2L-3)d]$$

$$= P[n_k > d]$$

$$= Q\left(\frac{d}{\sigma}\right) \tag{3.38}$$

The average probability of error is then

$$P[E] = \sum_{m=-(L-1)}^{L-1} P[E \mid R_I=2md]P[R_I = 2md] \tag{3.39}$$

The second probability in the summation is the probability density of the PR4 encoder output levels and is given by (3.12). Substituting for this quantity and performing the summation, we obtain

$$P[E] = 2\left(1 - \frac{1}{L^2}\right)Q\left(\frac{d}{\sigma}\right) \tag{3.40}$$

as the expression for the average error rate at the output of the slicer. However, this is not the error rate for the entire decoder as the modulo L mapper has yet to be

considered. Consider a binary input system where $R_I = -2d$ is transmitted. Suppose that the noise is such that the symbol is sliced as +2d. This symbol is incorrect at the slicer output but would be correctly decoded by the mod 2 mapper. The probability of this event occuring is

$$P[R_I + n_k > d \mid R_I = -2d] = P[n_k > 3d] \qquad (3.41)$$

$$= Q\left(\frac{3d}{\sigma}\right)$$

Similarly, $R_I = 2d$ may be sliced as $-2d$ and then correctly mapped into the original binary symbol. The error rate at the output of the mod 2 mapper is then

$$P[E] = \frac{3Q\left(\frac{d}{\sigma}\right)}{2} - P[R_I + n_k > d \mid R_I = -2d]P[R_I = -2d] \qquad (3.42)$$

$$- P[R_I + n_k < -d \mid R_I = 2d]P[R_I = 2d]$$

$$= \frac{3Q\left(\frac{d}{\sigma}\right)}{2} - 1/2 Q\left(3\frac{d}{\sigma}\right)$$

The correction term due to this level coalescion effect is very small when compared to the error rate at the slicer output. We conclude that this effect may be ignored when Gaussian noise is the only form of interference present. A similar analysis for an L level signal would

yield a comparable result.

In Appendix B, the parameter d/σ is derived for the two spectral shaping models. We find that the ratio d/σ depends on how the partial response shaping is split between the transmitter and receive filters. When all of the signal shaping is done at the transmitter, we have

$$\frac{d}{\sigma} = \left(\frac{3}{2(L^2-1)} \quad \frac{P_C}{P_N} \right)^{1/2} \tag{3.43}$$

where P_C/P_N is the received (channel) SNR. When the spectrum shaping is split equally between the transmitter and the receiver, we have

$$\frac{d}{\sigma} = \frac{\pi}{4}\left(\frac{3}{L^2-1} \quad \frac{P_C}{P_N} \right)^{1/2} \tag{3.44}$$

The two models for the spectrum shaping split are shown in Fig. 3.12.

For L-level PAM operating in the minimum bandwidth with no spectrum shaping, the average error rate is [30].

$$P[E] = 2\left(1-\frac{1}{L}\right) Q\left(\frac{d}{\sigma}\right) \tag{3.45}$$

with

n(t)   t=kT

Data ──►┌──────────────┐   ┌───┐   ┌──────────────┐   ┌──────────────┐ Decoded
        │ Transmitter ├──►(Σ)──►│   Receiver   ├──── ✗ ───►│   Partial   ├──► Data
        │   Filter   │   └───┘   │   Filter   │   │   Response   │
        └──────────────┘   └──────────────┘   │   Decoder   │
                                                                   └──────────────┘

$H_T(\omega) = |H(\omega)|^{\frac{1}{2}}$ $H_R(\omega) = |H(\omega)|^{\frac{1}{2}}$
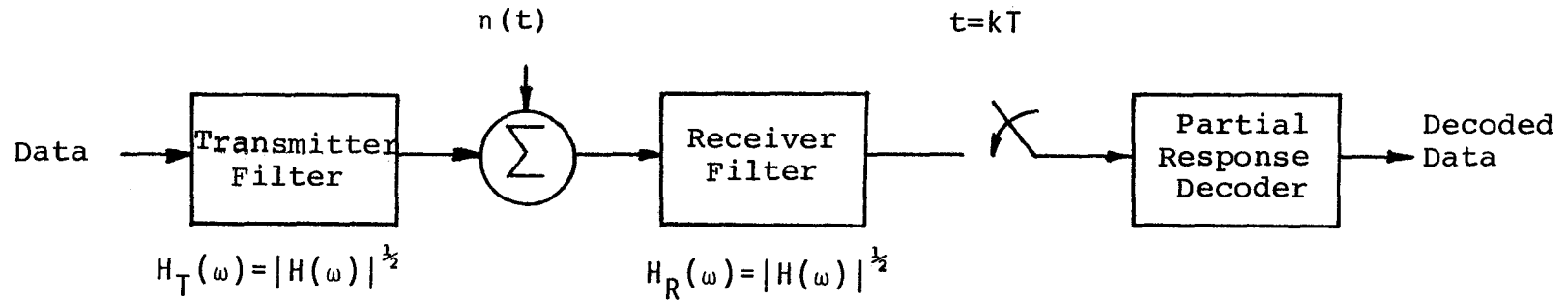
Model 1 : Spectrum Shaping Split Between Receiver and Transmitter

n(t)   t=kT

Data ──►┌──────────────┐   ┌───┐   ┌──────────────┐   ┌──────────────┐ Decoded
        │ Transmitter ├──►(Σ)──►│   Receiver   ├──── ✗ ───►│   Partial   ├──► Data
        │   Filter   │   └───┘   │   Filter   │   │   Response   │
        └──────────────┘   └──────────────┘   │   Decoder   │
                                                                   └──────────────┘

$H_T(\omega) = H(\omega)$ $H_R(\omega) = 1, |\omega| \leq \pi/T$
                                  $\qquad\quad = 0 \text{ elsewhere}$
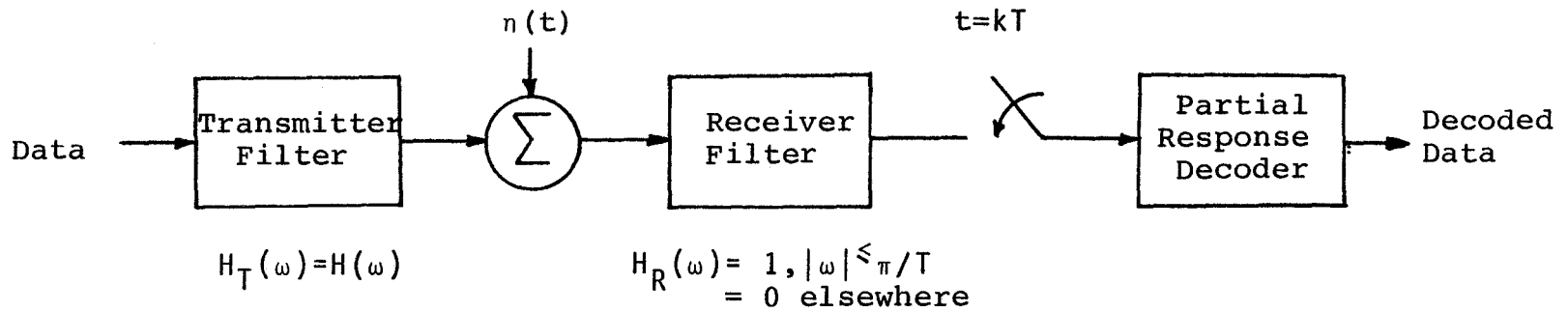
Model 2 :   Complete Spectrum Shaping at Transmitter

Figure 3.12   System Models for Spectral Shaping

$$\frac{d}{\sigma} = \left(\frac{3}{L^2 - 1} \quad \frac{P_C}{P_N} \right)^{1/2} \tag{3.46}$$

In Fig. 3.13 we have presented a graphical comparison of the error rate performance of a binary PAM system, a binary input PR4 system with split shaping, and the same PR4 system but with full transmitter shaping. The split shaping PR4 system requires 2.1 dB more SNR than the uncoded binary system while the full transmitter shaped PR4 signal needs 3 dB more than the binary PAM system. This extra SNR is required because the receiver must distinguish between 3 levels rather than 2.

From (3.40) and (3.43), doubling the bandwidth efficiency of the PR4 system by using a 4 level input signal instead of a binary input signal requires an additional 7 dB SNR for the same error rate. Combined with a 3 dB loss due to full transmitter shaping, this means that a 4 level input PR4 encoded system with full transmitter shaping requires 10 dB more SNR than a binary PAM system for the same error rate.

We shall now turn our attention to the error rate performance of the decision feedback decoder. The sampled received signal at $t = kT$ is
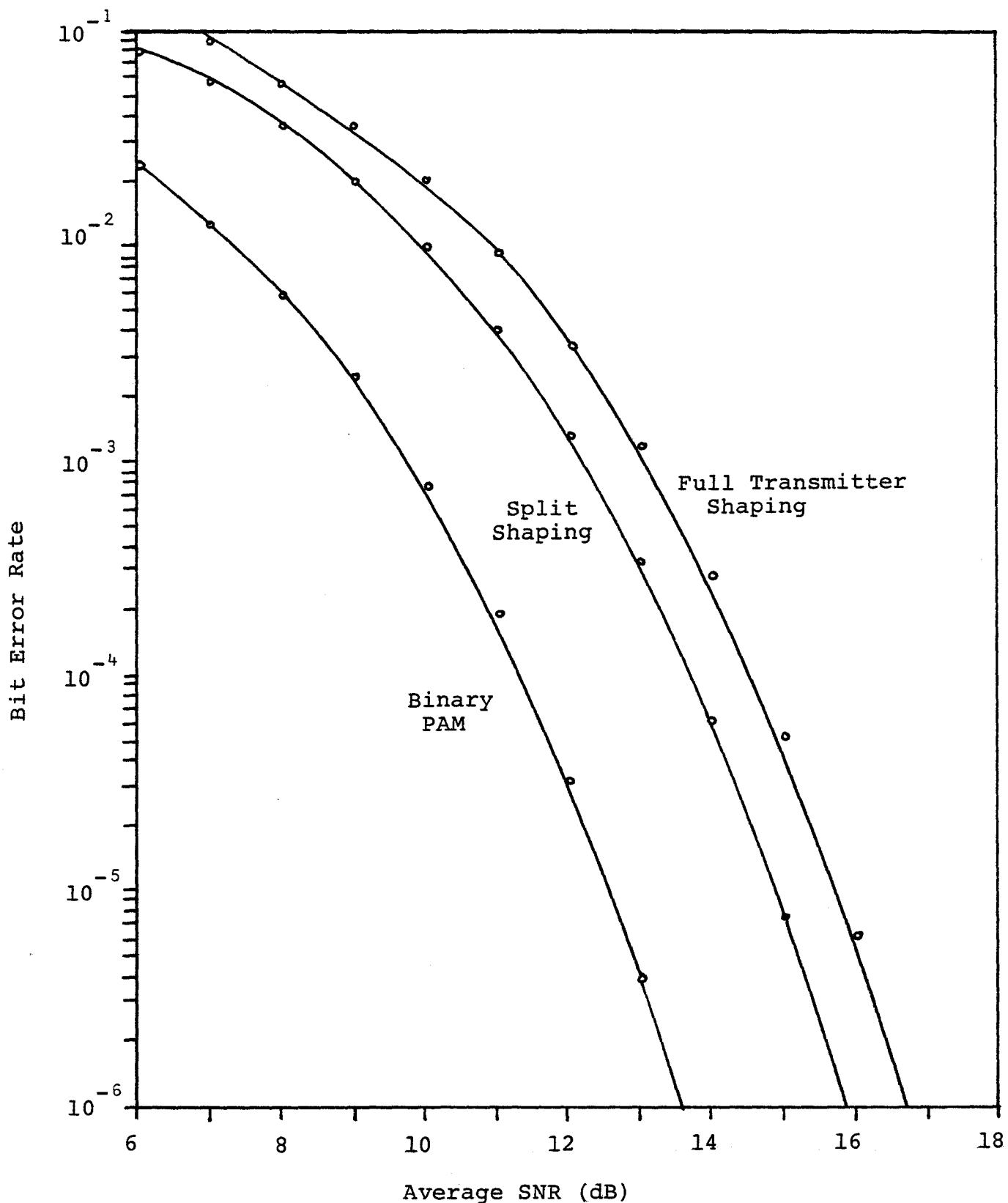
Figure 3.13   Bit Error Rate Performance of Binary PAM
and PR4 Systems

$$r_I(kT) = c_{k+1} - c_{k-1} + n_k \tag{3.47}$$

The slicer input signal is

$$\lambda_k = r_I(kT) + \tilde{c}_{k-1} \tag{3.48}$$

where $\tilde{c}_{k-1}$ is the estimate of the symbol c_{k-1}. We may rewrite (3.48) as

$$\lambda_k = c_{k+1} + \Delta_{k-1} + n_k \tag{3.49}$$

where $\Delta_{k-1} \triangleq \tilde{c}_{k-1} - c_{k-1}$ represents an error propagation term. When the previous decision is correct, this term is zero. The error propagation effect makes an exact analysis of the decoder performance very difficult. However, a lower bound is readily obtained by assuming that all previous decisions are correct i.e. $\Delta_{k-1} = 0$. Thus we have

$$\lambda_k = c_{k+1} + n_k \tag{3.50}$$

We now have an L level slicer with decision boundaries set at 0, $\pm 2d$, $\pm 4d$, $\ldots$, $\pm(L-2)d$. As for the precoded receiver, the probability of error for an inner level $(c_k \neq \pm(L-1)d)$ is $2Q(\frac{d}{\sigma})$ and for an outer level $(c_k = \pm(L-1)d)$, $Q(\frac{d}{\sigma})$. The average probability of error

with no error propagation is [26]

$$P_{e_L} = 2 \quad (1 - \frac{1}{L}) \, Q \, (\frac{d}{\sigma}) \tag{3.51}$$

This is the probability of error for a PAM system with d/σ changed to that of a PR4 system. Once again, this parameter depends on the shaping split.

An upper bound on the performance of a decision feedback equalizer with N feedback inputs is developed in [37]. This bound, extended to include L-ary input symbols [26] is

$$P_{e_U} = \frac{L^N \, P_{e_L}}{\frac{L}{L-1} P_{e_L} (L^N - 1) + 1} \tag{3.52}$$

where P_{e_L} is the error rate with no error propagation and is given by (3.51). For a PR4 decoder, N = 1, and so (3.52) reduces to

$$P_{e_U} = \frac{L \, P_{e_L}}{L \, P_{e_L} + 1} \tag{3.53}$$

Thus the error rate is increased by at most a factor of L due to error propagation. The two decoders presented

	MODEL 1 SPLIT SHAPING		MODEL 2 FULL TRANSMITTER SHAPING	
	PRECODED	DECISION FEEDBACK	PRECODED	DECISION FEEDBACK
SNR DEGRADATION (dB)	2.3	2.1 (LOWER BOUND)	3.2	3.0 - 3.3

TABLE 3.3   SNR DEGRADATION (dB) OVER PAM AT $P_E = 10^{-5}$ FOR PR4 ENCODED BINARY

SYMBOLS

here are compared in Table 3.3 A more comprehensive comparison of the other PR codes is found in [26].

3.4 The Quadrature Channel

Single sideband modulation differs from other forms of modulation in that the same signal is used to produce both the inphase (I) and quadrature (Q) components of the modulated signal. Consequently, it is possible to derive the original data sequence from either the I or Q demodulator output. The Q output of a coherent demodulator with no receiver phase error is

$$r_Q(t) = \hat{a}(t) + n_Q(t) \tag{3.54}$$

with $\hat{a}(t)$ given by (3.10) and $n_Q(t)$ given by (2.46). Sampling this signal at $t = kT$ gives

$$r_Q(kT) = \frac{4}{\pi} c_k + z_k + n_k \tag{3.55}$$

where

$$z_k \triangleq -\frac{4}{\pi} \sum_{\ell=1}^{\infty} \frac{c_{k+2\ell} + c_{k-2\ell}}{4\ell^2 - 1} \tag{3.56}$$

and

$$n_k \overset{\Delta}{=} n_Q (kT) \qquad\qquad (3.57)$$

The z_k term is an intersymbol interference term resulting from the overlapping of tails and heads from other Hilbert transform pulses. For binary inputs ($c_k = \pm d$), the receiver sets $\tilde{c}_k$, the receiver estimate of c_k, to $+d$ if $r_Q(kT) > 0$ and sets $\tilde{c}_k = -d$ if $r_Q(kT) < 0$. The two possible error events are defined by

$$\frac{4d}{\pi} + z_k + n_k < 0 \qquad \text{for } c_k = +d$$

$$-\frac{4d}{\pi} + z_k + n_k > 0 \qquad \text{for } c_k = -d$$

The probability of these two error events are, respectively,

$$P[E| z_k, c_k=d] = Q[(\frac{4d}{\pi}+z_k)/\sigma] \qquad\qquad (3.58)$$

$$P[E| z_k, c_k=-d] = Q[(\frac{4d}{\pi}-z_k)/\sigma] \qquad\qquad (3.59)$$

so that the average probability of error conditioned on the ISI z_k is

$$P[E | z_k] = P[E | z_k, c_k=d]P[c_k=d]$$

$$+ \ P[\ E | \ Z_k \ , c_k = -d \ ] P[\ c_k = -d \ ]$$

$$= 1/2Q[\ (\tfrac{4}{\pi}d + Z_k)/\sigma \ ] \ + \ 1/2 \ Q[\ (\tfrac{4}{\pi}d - Z_k)/\sigma \ ] \qquad (3.60)$$

where we have assumed that both error events are equally likely.

A lower bound is obtained by setting $Z_k = 0$, so that

$$P[\ E \ ] \geqslant Q (\tfrac{4}{\pi} \ \tfrac{d}{\sigma}) \qquad (3.61)$$

which is 2.1 dB better than the inphase channel with precoding. An upper bound is obtained by setting $c_{k+2\ell} = c_{k-2\ell} = -d$ in (3.56) so that the ISI has the maximum value

$$Z_{k_{max}} = - \frac{4}{\pi} \sum_{\ell=1}^{\infty} \frac{-d-d}{4\ell^2 - 1}$$

$$= \frac{4d}{\pi} \qquad (3.62)$$

Substituting this maximum value into (3.60) gives an upper bound of $Q(0) = 1/2$ for the error rate. Actual performance is between these two bounds, and computer simulation was used to measure the real error rate performance of the channel. The simulation results are

shown in Fig. 3.14. Clearly the large amount of SNR degradation over this inphase channel makes direct use of the Q channel undesirable.

To reduce the ISI and decrease the error rate, an equalizer, shown in Fig. 3.15 was developed. It is a feedforward-feedback [38] structure in which previous decisions are fed back and future samples are fed forward in order to cancel the tails and heads of the interfering pulses. The decision device input is

$$\lambda_k = r_Q(kT) + \sum_{\ell=1}^{NF} g_{2\ell} r_Q((k+2\ell)T) \qquad (3.63)$$

$$+ \sum_{m=1}^{NR} g_{-2m} \tilde{c}_{k-2m}$$

where the $\{g_{2k}\}$ are the equalizer tap weights and NF and NR are the number of equalizer feedforward and feedback taps respectively. Simulation results show that the best performance is obtained for 2 forward and 2 reverse taps with the following values:

$$g_4 = 1/15$$

$$g_2 = 1/3$$
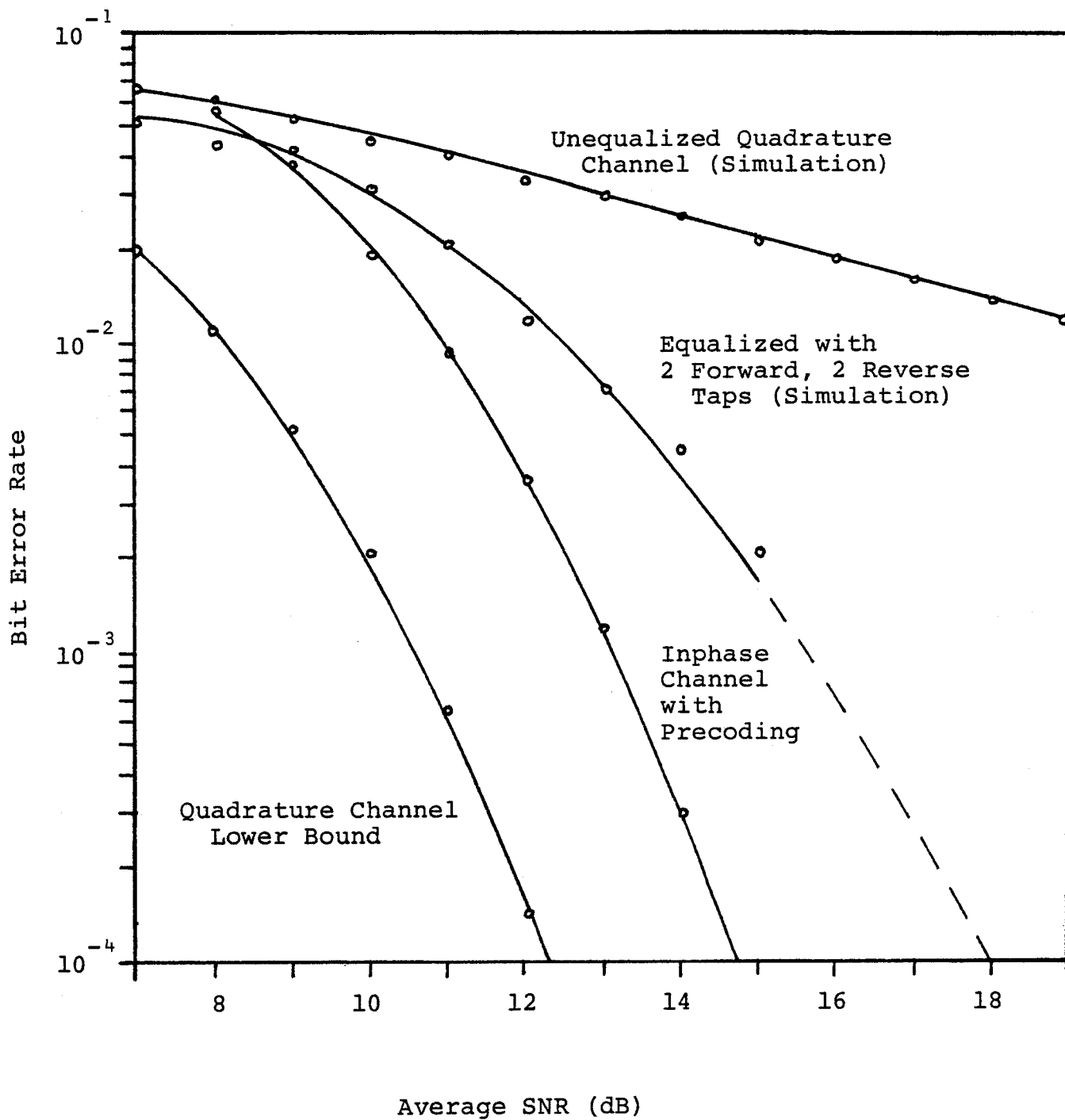
Figure 3.14 Bit Error Rate Performance of
Quadrature Channel with Binary Input
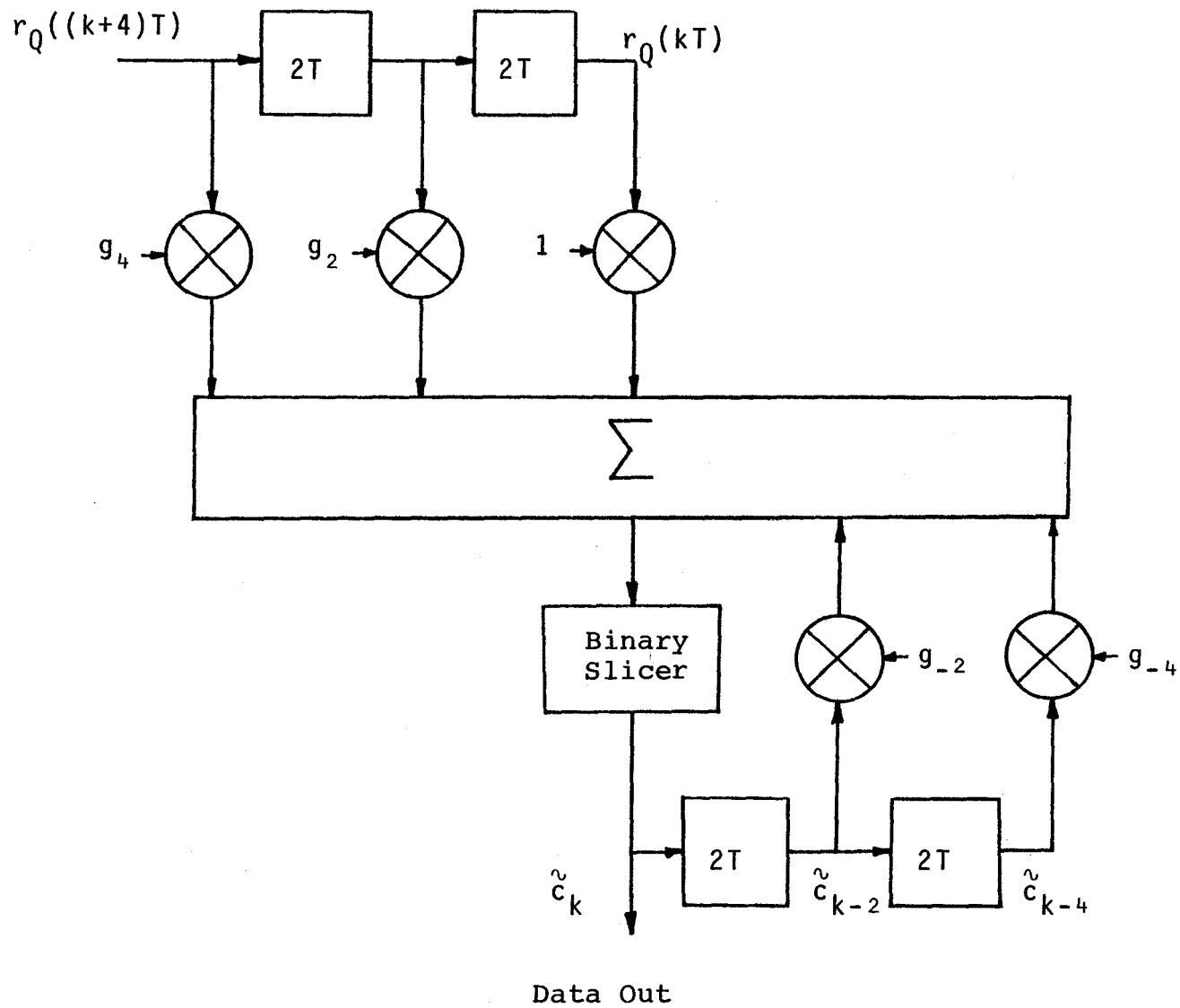
Figure 3.15 Equalizer for Quadrature Channel

$$g_{-2} = 4/3 \ \pi$$

$$q_{-4} = 4/15 \ \pi$$

These values are obtained from the quadrature pulse samples. The performance of this receiver is shown in Fig. 3.14. Although a significant improvement was obtained over the unequalized Q channel, the performance was still several dB worse than the inphase channel.

Attention was then turned to the possibility of combining the I and Q channels since the same data is present in both channels. Sampling the inphase output at t = (k-1) T gives

$$r_I ((k-1)T) = c_k - c_{k-2} + n_I ((k-1)T) \qquad (3.64)$$

while sampling the quadrature output at t = kT gives

$$r_Q (kT) = \frac{4}{\pi} c_k + z_k + n_Q (kT) \qquad (3.65)$$

A receiver combining both channels is shown in Fig. 3.16. The Q signal is equalized as just outlined and then added to the output of a decision feedback equalizer for the inphase channel. A decision is made based on the sum of the
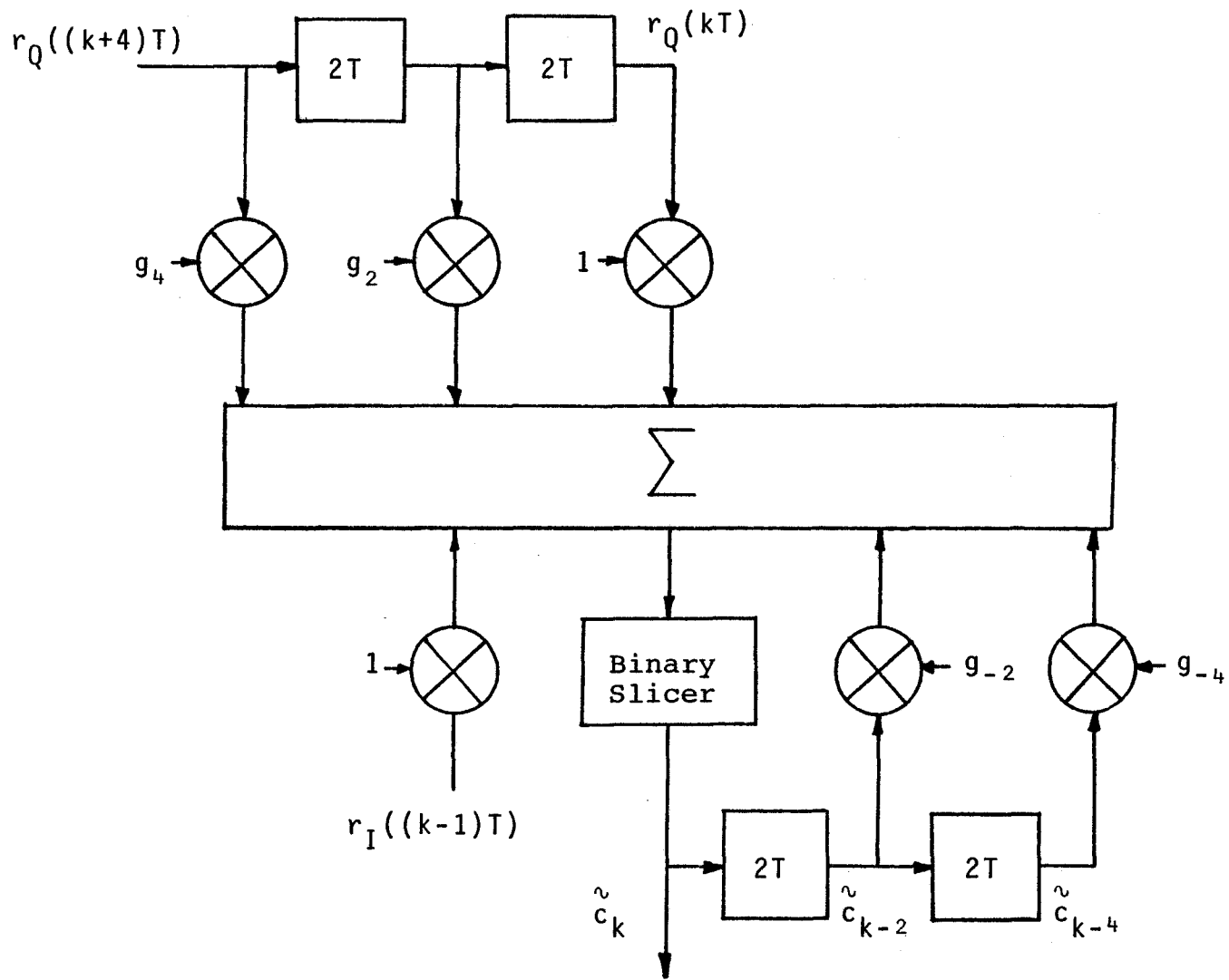
Figure 3.16   Receiver Using Inphase and Quadrature Information

unquantized equalizer outputs. Since the desired signal component c_k is now multiplied by the factor $(1+\underline{4})$ a better π performance should be obtained. However, this is not the case. As the quadrature noise is the Hilbert transform of the inphase noise, the noise samples $n_I(kT)$ and $n_I((k-1)T)$ ar no longer independent. Setting $W = 1/2T$ in (2.63) gives the noise cross-correlation as

$$R_{n_I n_Q}(\tau) = \frac{T \sigma^2}{\pi \tau} (1 - \cos\frac{\pi \tau}{T}) \qquad (3.66)$$

For our case $\tau = T$, so that

$$R_{n_I n_Q}(T) = \frac{2\sigma^2}{\pi} \qquad (3.67)$$

Thus the two noise samples are correlated. Furthermore, the inphase noise sample and the feedforward tap signal samples $r_Q((k+2)T)$ and $r_Q((k+4)T)$ would also be correlated. The implication of this result is that while the desired signal component is increased through the processing of the receiver, the noise component is also increased.

The performance of the receiver of Fig. 3.16 is shown in Fig. 3.17. Also shown is the simulated error rate performance of an ordinary decision feedback decoder operating on the I component only. There is no significant
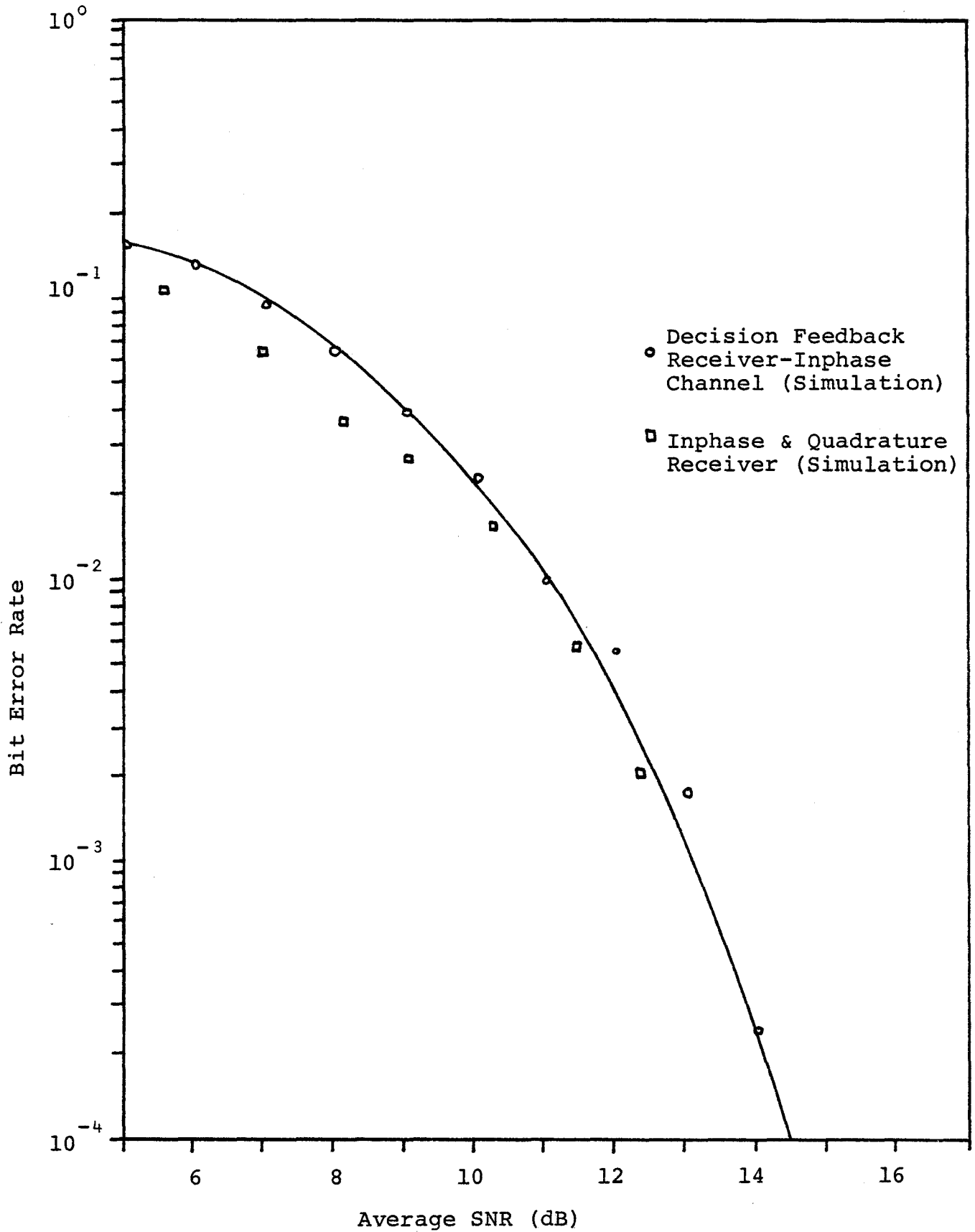
Figure 3.17   Bit Error Rate Performance of
I-Q Receiver

difference between the two receivers due to the noise cross-correlation just outlined. Thus we conclude that while the modulating data sequence can be obtained from either the I or the Q channel, any attempt to combine the two channels will not result in a significant improvement over the inphase channel alone.

3.5 Weaver Demodulation

The I and Q outputs of the first stage of a Weaver demodulator are, respectively

$$x(t) = a(t)\cos\omega_m t + a(t)\sin\omega_m t \qquad (3.68)$$

$$y(t) = a(t)\sin\omega_m t - a(t)\cos\omega_m t \qquad (3.69)$$

where an upper Weaver sideband input is assumed. If $a(t)$ is a PR4 encoded data signal then $\omega_m = \pi/2T$.

In the normal Weaver demodulation process, these signals would be multiplied by $\cos\pi t/2T$ and $\sin\pi t/2T$ and then added as described in an earlier chapter. Sampling and quantization would then follow. However, under certain conditions, the signals $x(t)$ and $y(t)$ may be sampled and quantized directly after the first stage of demodulation.

If $x(t)$ is sampled at $t = 2kT$ while $y(t)$ is sampled at $t = (2k+1)T$, the sampler outputs are

$$x(2kT) = a(2kT)\cos k\pi + \hat{a}(2kT)\sin k\pi$$

$$= (-1)^k a_{2k} \tag{3.70}$$

$$y((2k+1)T) = a((2k+1)T)\sin(2k+1)\pi/2 - \hat{a}((2k+1)T)\cos(2k+1)\pi/2$$

$$= (-1)^k a_{2k+1} \tag{3.71}$$

If the a were binary symbols, then the data could not be recovered directly due to the multiplicative factor $(-1)^k$. Suppose, however, that the a_k are precoded PR4 symbols. Since the recovery of the original binary data from a precoded PR4 sequence is done according to the rule $c_k = $ "0" if $a_k = 0$ and $c_k = $ "1" if $a_k = \pm 2d$, we see that the factor $(-1)^k$ has no effect on the decoding process and the data is recovered as usual. The binary outputs of each arm would be interleaved to reconstruct the original signal. The complete demodulator is shown in Fig. 3.18 while the eye diagram of the output of one branch of the demodulator is shown in Fig. 3.19.

This demodulation method is somewhat similar to an
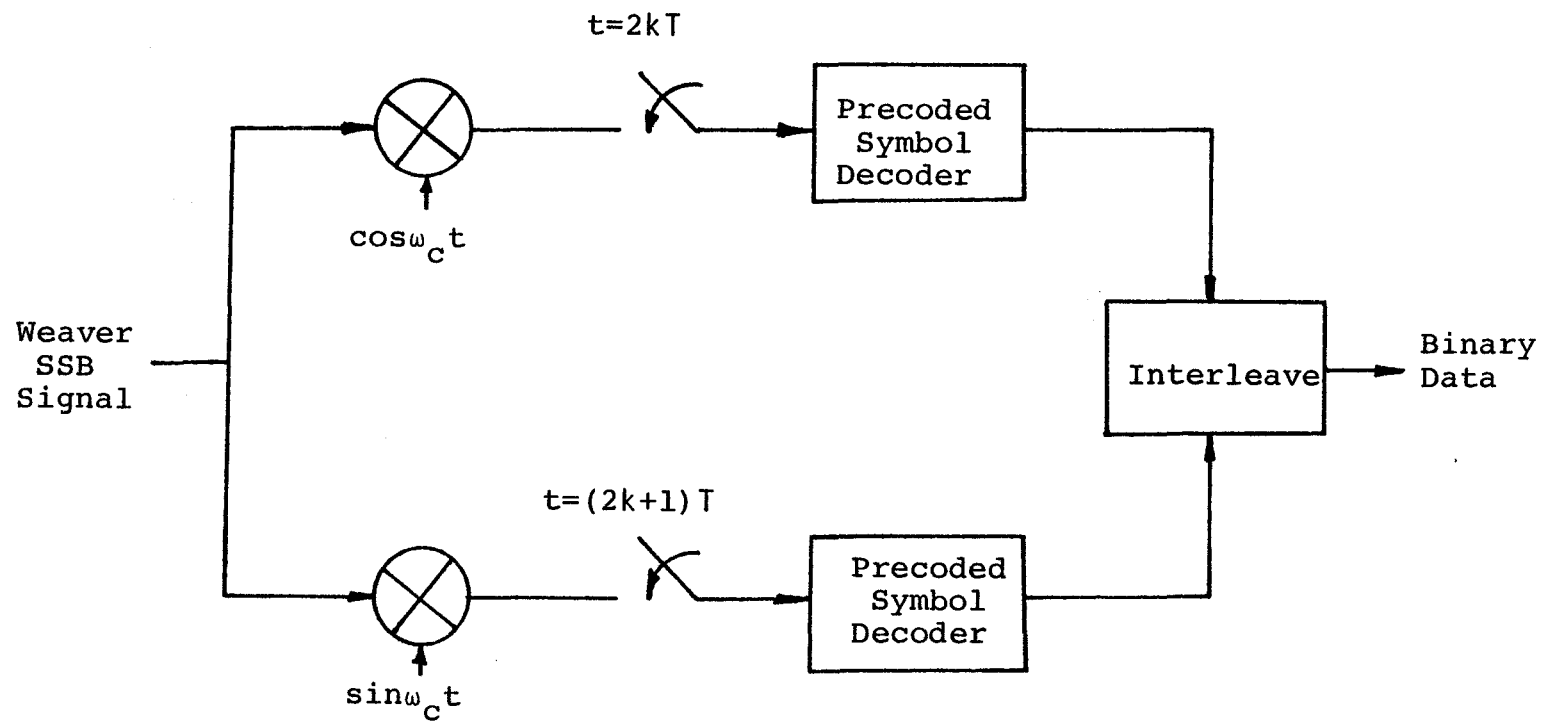
Figure 3.18   Single Stage Weaver Demodulator
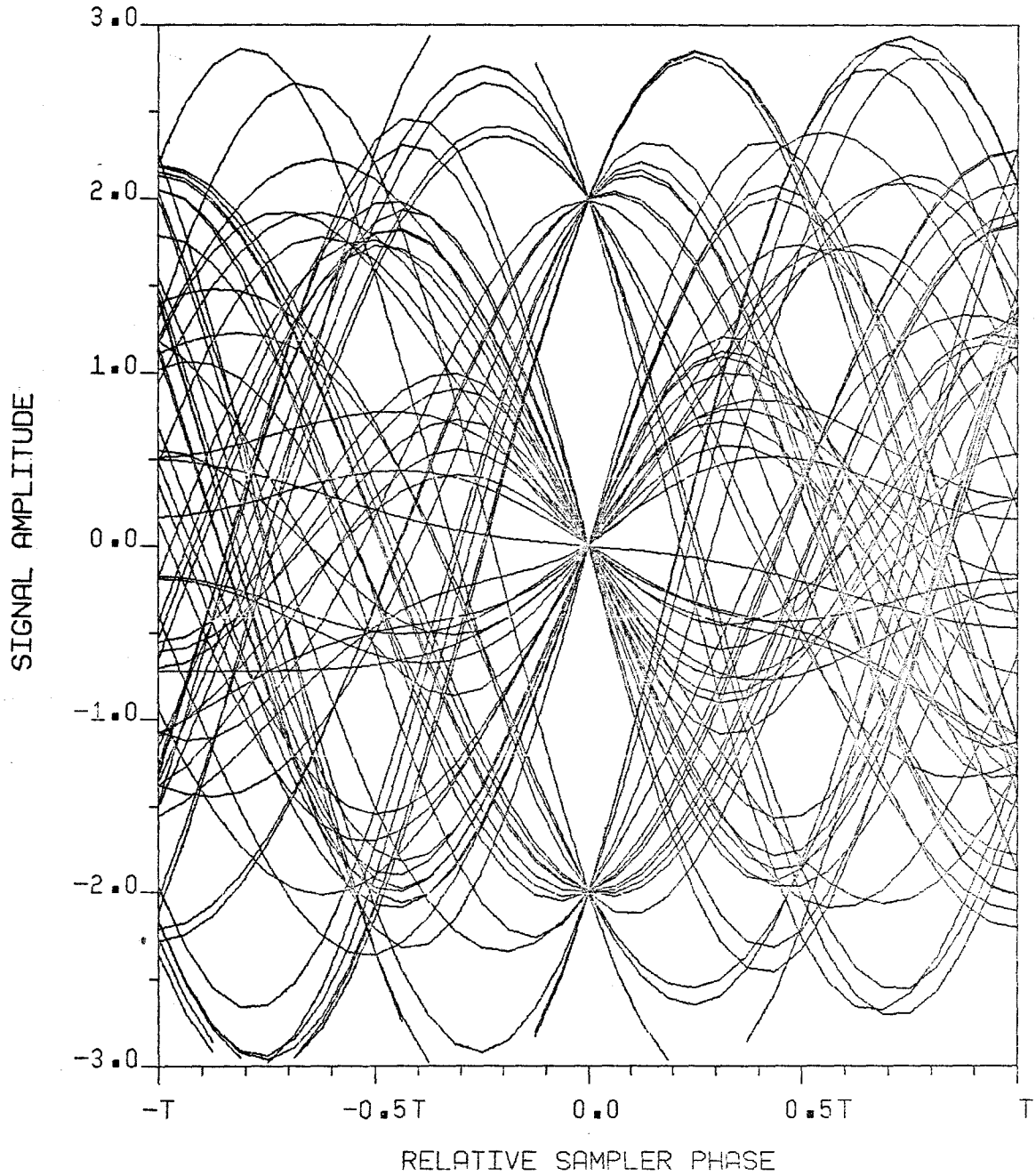
Figure 3.19 Single Stage Weaver Demodulator Output
Eye Diagram With Binary Input

offset quadrature partial response system (QPRS) receiver except for the $(-1)^k$ factors. High performance carrier phase tracking loops for such receivers are well known [39] On the other hand, it may be difficult to obtain symbol timing using this method, and the effects of symbol timing error may become important as the steep slope of the eye diagram boundary indicates. Another disadvantage is that this scheme works only for binary input precoded PR4 symbols. Use of L level inputs may not be possible. For these reasons, this method of demodulation will not be considered further in this report.

3.6 Conclusions

The partial response signalling concept was outlined and the Class 4 partial response pulse was selected for our system because it satisfies the desired spectral requirements with the least number of output levels. The performance of this code in AWGN was analyzed, and the possibility of using quadrature channel information for decreasing the error rate was discussed. It was shown that the performance of a receiver using inphase channel information alone could not be improved by using quadrature information as well as due to the cross-correlation of the noise in the two channels. A novel method of demodulating a

Weaver SSB signal was presented but due to its inherent lack of flexibility with respect to the type of signal that may be used in the system, it was not considered a viable alternative to the standard SSB coherent demodulator. Further investigation of this method may reveal ways of circumventing the flexibility problem.

CHAPTER 4.   CARRIER SYNCHRONIZATION

4.1 Introduction

The problem of carrier frequency and phase recovery in a coherent receiver is crucial for satisfactory operation of the receiver. The local reference must be strong and noise-free, and must be able to track any variations in the received carrier phase so that the signal is correctly demodulated and the probability of error minimized. Furthermore, the receiver reference signal coherence must be independent of the data modulating the carrier.

Carrier recovery in SSB systems is particularly difficult due to the suppressed carrier nature of the signal. In addition, since SSB may be viewed as a combined form of amplitude-phase modulation, the carrier phase is modulated by the data. Unlike APK or PSK, this phase modulation varies over a range of values which must be removed in order that the recovered carrier be useful for demodulation. It is for this reason that linear tracking loops such as the Costas loop [40] are inadequate for SSB applications.

A simple solution to the carrier recovery problem is to transmit a pilot tone with the required carrier frequency along with the SSB signal. A local oscillator would be phaselocked to the output of a narrowband filter centered at the carrier frequency. The lack of low frequency components in the SSB-PR4 signal makes it ideal for the insertion of such a pilot tone because there would be a minimum of interference between the pilot and the data signal. Such a tracking system is described in [41]. From the point of view of transmitter power utilization, a pilot tone system is wasteful in that power that could be used for transmission of the data signal is instead used for pilot transmission. In addition, use of a pilot would put a spike in the transmitter output spectrum which may not be permitted by regulatory bodies such as the FCC.

An alternative structure which has been used for a variety of modulation schemes is the decision-directed carrier tracking loop [42]-[44]. The carrier frequency and phase are derived from the data-bearing signal itself and no pilot need be used, although the two methods may be combined if desired [42]. Kobayashi has developed a combined carrier phase and symbol timing recovery circuit for use in an SSB-PR4 receiver [45]. The design of such loops is also considered in [46].

In the first part of this chapter, we shall examine the effects of a steady state phase error on the error rate performance of the receiver. Upper and lower bounds on the error rate will be developed and compared to simulation results. This will be done for both binary and 4 level input signals. In the second part, a decision directed SSB-PR4 carrier tracking system similar to Kobayashi's is proposed and analyzed for binary input symbols. The theoretical results obtained are compared with simulation to verify the analysis.

4.2 Effects of Steady State Phase Errors

4.2.1 Error Rate Effects

The inphase output of a coherent demodulator with a steady state phase error ϕ is

$$r_I(t) = a(t)\cos\phi + \hat{a}(t)\sin\phi$$

$$+ n_I(t)\cos\phi + n_Q(t)\sin\phi \tag{4.1}$$

with $a(t)$, $\hat{a}(t)$, $n_I(t)$ and $n_Q(t)$ defined by (3.9), (3.10), (2.45) and (2.46) respectively. Letting

$$N(t, \phi) = n_I(t)\cos\phi + n_Q(t)\sin\phi \qquad (4.2)$$

then

$$r_I(t) = a(t)\cos\phi + \hat{a}(t)\sin\phi + N(t,\phi) \qquad (4.3)$$

Since $n_I(t)$ and $n_Q(t)$ are AWGN processes of zero mean and variance $\sigma^2 = N_0/2T$, $N(t,\phi)$ is also a zero-mean AWGN process of variance

$$\sigma_N^2 = E[N(t,\phi)N(t,\phi)] \qquad (4.4)$$

$$= E[(n_I(t)\cos\phi + n_Q(t)\sin\phi)^2] \qquad (4.5)$$

$$= \cos^2\phi\, E[n_I^2(t)] + \sin^2\phi\, E[n_Q^2(t)] $$

$$+ 2\sin\phi\cos\phi\, E[n_I(t)n_Q(t)] \qquad (4.6)$$

The last expectation is zero since $n_I(t)$ and $n_Q(t)$ are uncorrelated for zero time shift. Thus we have

$$\sigma_N^2 = \sigma^2 \qquad (4.7)$$

$$= \frac{N_0}{2T}$$

Sampling the inphase output at $t = kT$ gives us the decision device input

$$\lambda_k = R_I \cos\phi + R_Q \sin\phi + n_k \qquad (4.8)$$

where

$$\lambda_k \overset{\triangle}{=} r_I(kT) \qquad (4.9)$$

$$R_I \overset{\triangle}{=} a(kT) \qquad (4.10)$$

$$R_Q \overset{\triangle}{=} \hat{a}(kT) \qquad (4.11)$$

$$n_k \overset{\triangle}{=} N(kT,\phi) \qquad (4.12)$$

with

$$a(kT) = c_{k+1} - c_{k-1} \qquad (4.13)$$

$$\hat{a}(kT) = \frac{4}{\pi}\left(c_k - \sum_{\ell=1}^{\infty} \frac{c_{k+2\ell} + c_{k-2\ell}}{4\ell^2 - 1} \right) \qquad (4.14)$$

Denoting the quadrature distortion term $R_Q \sin\phi$ by Z_k, we have

$$\lambda_k = (c_{k+1} - c_{k-1})\cos\phi + z_k + n_k \qquad (4.15)$$

The $c_k = \pm d, \pm 3d, \ldots, \pm (L-1)d$ so that $R_I = 2md$ for $|m| = 0, 1, 2, \ldots, L-1$. Since we are considering the precoded receiver only, the decision thresholds will be set at $\pm d, \pm 3d, \pm 5d, \ldots, \pm(2L-3)d$. For a transmitted symbol $R_I = 2md$, a correct decision occurs if

$$(2m-1)d \leqslant \lambda_k \leqslant (2m+1)d \qquad (4.16)$$

for the inner levels ($|m| \neq L-1$). Substituting (4.15) into (4.16) shows that a correct decision is made if

$$d(2m(1-\cos\phi)-1)-z_k \leqslant n_k \leqslant d(2m(1-\cos\phi)+1)-z_k \qquad (4.17)$$

Hence the probability of error for an inner level is

$$P[E \mid z_k, R_I = 2md, |m| \neq L-1]$$

$$= 1-P[d(2m(1-\cos\phi)-1)-z_k \leqslant n_k$$

$$\leqslant d(2m(1-\cos\phi)+1)-z_k] \qquad (4.18)$$

$$= 1 - \int_{d(2m(1-\cos\phi)-1)-Z_k}^{d(2m(1-\cos\phi)+1)-Z_k} \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{x^2}{2\sigma^2} dx \qquad (4.19)$$

$$= Q[(d(1-2m(1-\cos\phi)) + Z_k)/\sigma]$$

$$+ Q[(d(1+2m(1-\cos\phi)) - Z_k)/\sigma] \qquad (4.20)$$

The decision regions are different at the edges. For $R_I = 2(L-1)d$, an error occurs if

$$\lambda_k < (2L-3)d \qquad (4.21)$$

or

$$n_k < d((2L-3)-2(L-1)\cos\phi)-Z_k \qquad (4.22)$$

The probability of this event is

$$P[E|Z_k, R_I = 2(L-1)d]$$

$$= P[n_k < d((2L-3) - 2(L-1)\cos\phi) - Z_k] \qquad (4.23)$$

$$= Q[(d(2(L-1)\cos\phi - (2L-3)) + Z_k)/\sigma] \qquad (4.24)$$

Similarly, at the other edge, $R_I = -2(L-1)d$ and an error occurs if

$$\lambda_k > -(2L-3)d \qquad (4.25)$$

or

$$n_k > d(2(L-1)\cos\phi - (2L-3)) - z_k \qquad (4.26)$$

The probability of this event is

$$P[E|Z_k, R_I = -2(L-1)d]$$

$$= P[n_k > d(2(L-1)\cos\phi - (2L-3)) - z_k] \qquad (4.27)$$

$$= Q[(d(2(L-1)\cos\phi - (2L-3)) - z_k)/\sigma] \qquad (4.28)$$

The total average error probability conditioned on the phase error and the quadrature distortion Z_k is

$$P[E|\phi, Z_k] = \sum_{m=-(L-1)}^{L-1} P[E|\phi, Z, R_I=2md]P[R_I=2md]$$

$$= \sum_{m=-(L-1)}^{L-1} P[E|\phi, Z_k, R_I=2md]\left(\frac{L-|m|}{L^2}\right) \qquad (4.29)$$

Substituting for the conditional probabilities from (4.20), (4.24) and (4.28) gives

$$P\{E \mid \phi, Z_k\} = \sum_{m=-(L-2)}^{L-2} (L- \mid m \mid) \{Q[(d(1-2m(1-\cos \phi)) + Z_k)/\sigma]$$

$$+Q[(d(1+2m(1-\cos\phi))-Z_k)/\sigma]\} \qquad (4.30)$$

$$+ \frac{1}{L^2} \{ Q[(d(1-2(L-1)(1-\cos\phi))+Z_k)/\sigma]$$

$$+Q[(d(1-2(L-1)(1-\cos\phi))-Z_k)/\sigma]\}$$

Setting $\phi = 0$ (and hence $Z_k = 0$) gives

$$P[E \mid \phi = 0] = \frac{2}{L^2} \sum_{m=-L+1}^{L-1} (L- \mid m \mid) Q \left(\frac{d}{\sigma}\right) \qquad (4.31)$$

$$= 2(1-\frac{1}{L^2})Q \left(\frac{d}{\sigma}\right) \qquad (4.32)$$

which is the result obtained earlier.

A lower bound for a given phase error may be obtained by setting $Z_k = 0$ which gives

$$P[E \mid \phi] \geq \frac{1}{L^2} \sum_{m=-L+2}^{L-2} (L- \mid m \mid) \{Q[(d(1-2m(1-\cos\phi))/\sigma]$$

$$+ Q[d(1+2m(1-\cos \phi))/\sigma]\} \qquad (4.33)$$

$$+ \frac{2}{L^2} Q[d(1-2(L-1)(1-\cos\phi))/\sigma]$$

An upper bound is obtained by setting Z_k equal to

its worst-case value. This occurs when

$$c_{k+2\ell} = c_{k-2\ell}$$

$$= -(L-1)d \qquad (4.34)$$

and

$$c_k = (L-1)d \qquad (4.35)$$

in (3.16). This gives the worst-case value of z_k as

$$z_k = \frac{4}{\pi} [(L-1)d + \sum_{\ell=1}^{\infty} \frac{(L-1)d + (L-1)d}{4\ell^2 - 1}]\sin\phi$$

$$= \frac{8}{\pi} (L-1)d\sin\phi \qquad (4.36)$$

Thus the average error rate conditioned on a steady state phase error ϕ is

$$P[E|\phi] < \frac{1}{L^2} \sum_{m=-L-2}^{L-2} (L-|m|) \{Q [\frac{d}{\sigma}(1-2m(1-\cos\phi) + \frac{8}{\pi}(L-1)\sin\phi)]$$

$$+ Q[\frac{d}{\sigma}(1-2m(1-\cos\phi) - \frac{8}{\pi}(L-1)\sin\phi)]\} \qquad (4.37)$$

$$+ \frac{1}{L^2}\{Q[\frac{d}{\sigma}(1-2(L-1)(1-\cos\phi) + \frac{8}{\pi}(L-1)\sin\phi)]$$

$$+ Q[\frac{d}{\sigma}(1-2(L-1)(1-\cos\phi) - \frac{8}{\pi}(L-1)\sin\phi)]\}$$

This upper bound is plotted for several values of

phase error in Fig. 4.1 for binary inputs while Fig. 4.2 illustrates the bound for a 4 level input. Simulation results for these two cases are shown in Figs. 4.3 and 4.4. Finally, the SNR degradation at an error rate of 10^{-5} as a function of phase error obtained from both simulation results and the upper bound of (4.37) is shown in Fig. 4.5. We see that the bound is quite tight for a small phase error or a small number of input levels but when large amounts of intersymbol interference (ISI) due to a large number of input levels and/or a large phase error are present, the bound's usefulness becomes limited. The lower bound (4.33) is not shown as it does not deviate significantly from the probability of error with no phase error present (i.e. 0 dB degradation) until the phase error becomes larger than 10°. Full transmitter shaping was used, so d/σ was calculated according to (3.43).

As expected, the binary input system tolerates phase errors much better than the 4 level input system. A 10° error in the binary case results in only a 3.3 dB degradation while such a phase error in the 4 level system would be intolerable. To keep the SNR degradation less than 3 dB, the phase error must be kept less than 9.3° for the binary system and less than 3.8° for the 4 level system. The degradation of the SNR with phase error illustrates the
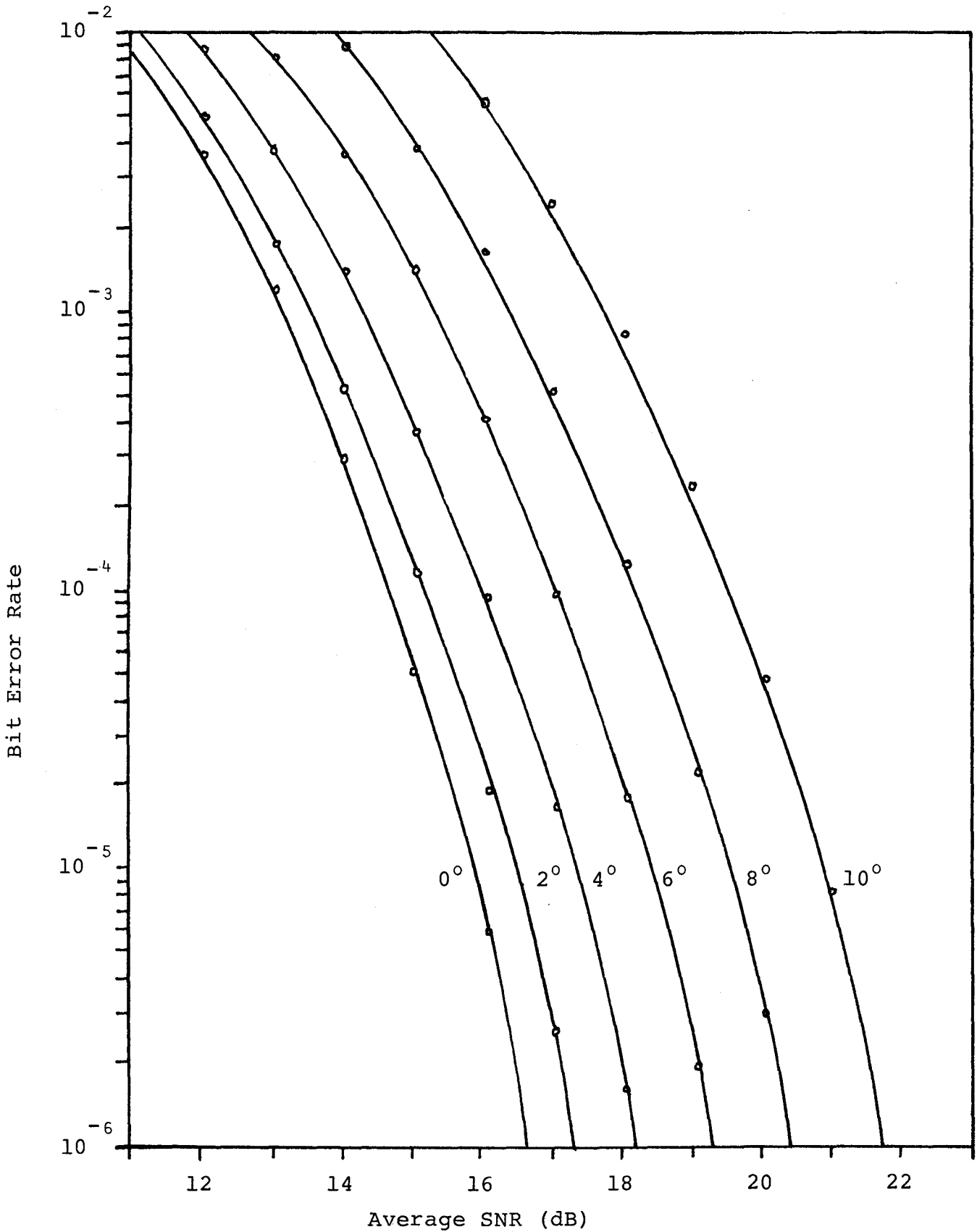
Figure 4.1   Upper Bound on Bit Error Rate with Steady State
Phase Error Present and Binary Input
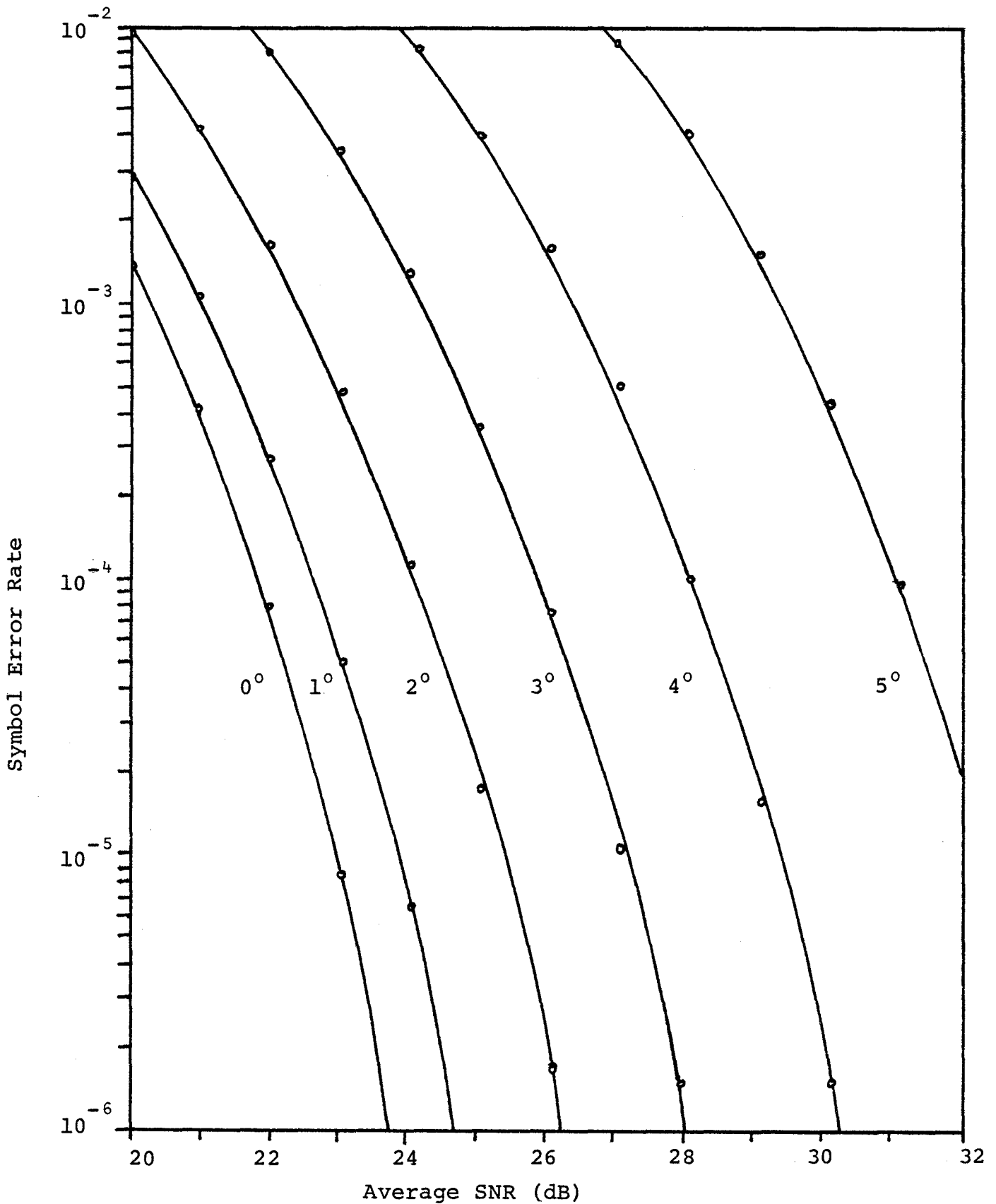
Figure 4.2   Upper Bound on Symbol Error Rate with
Steady State Phase Error Present and 4 Level Input
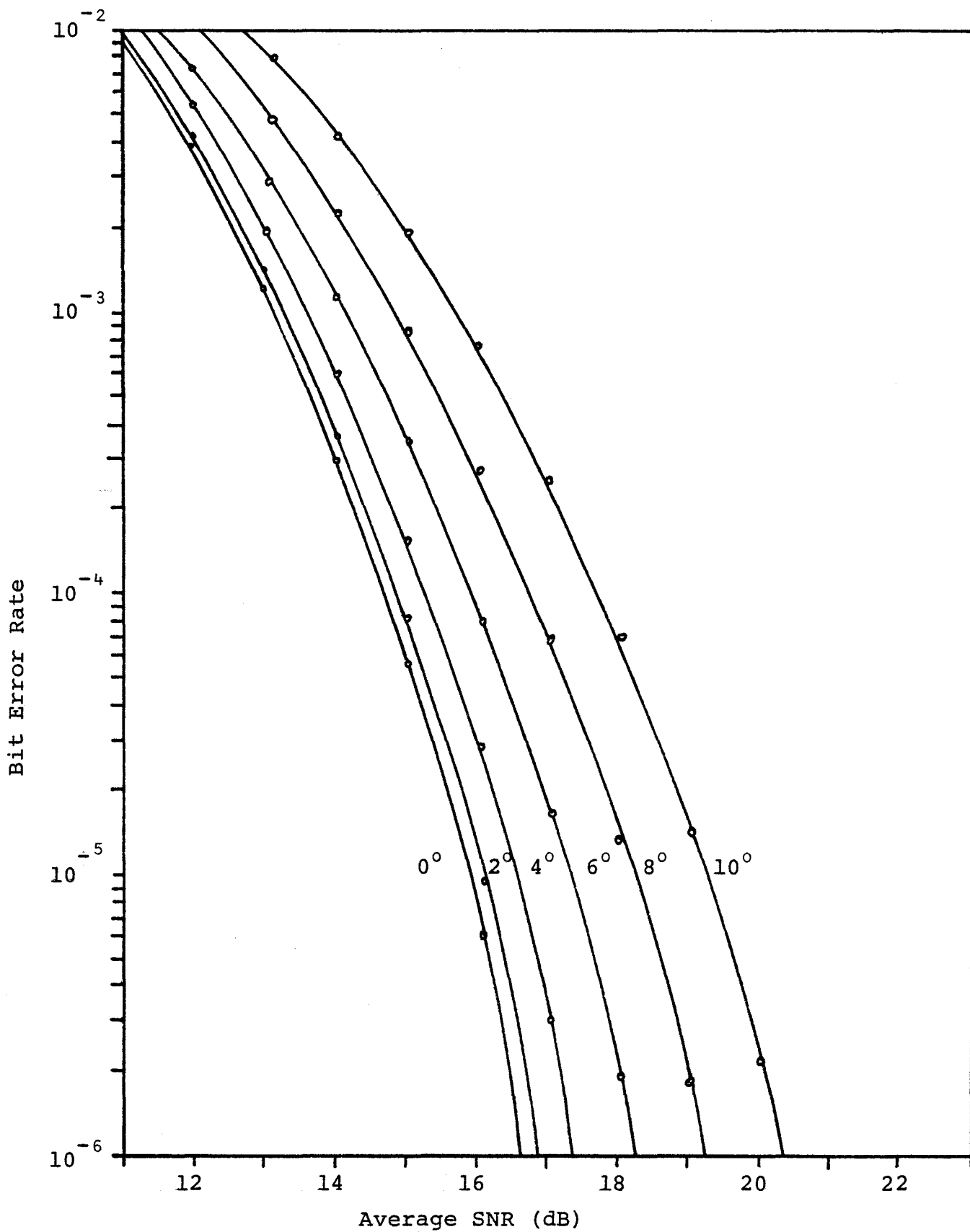
Figure 4.3   Simulated Bit Error Rate Performance with
Steady State Phase Error Present and Binary Input
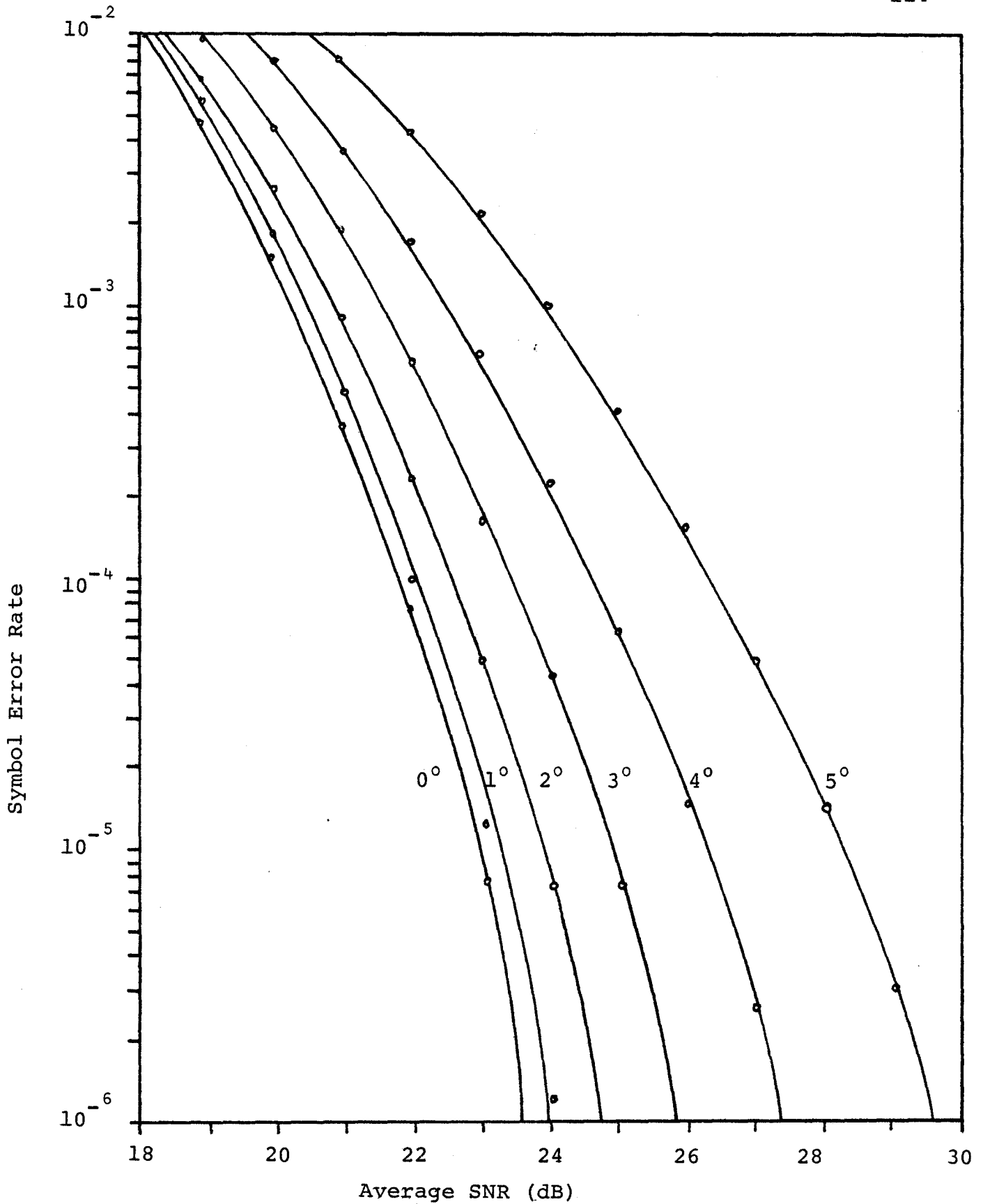
Figure 4.4   Simulated Symbol Error Rate Performance with
Steady State Phase Error Present and 4 Level Input
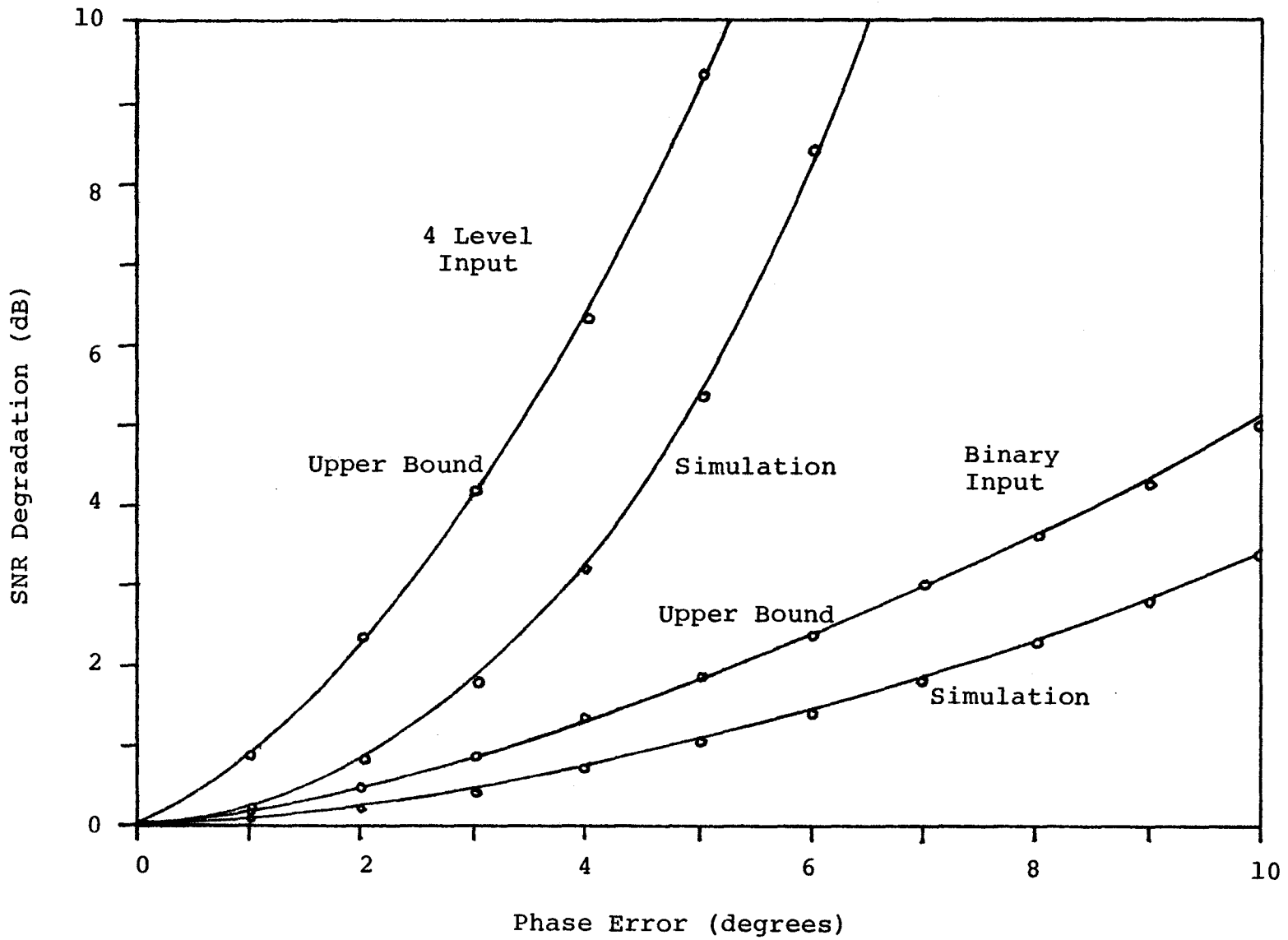
Figure 4.5   SNR Degradation at $P_E = 10^{-5}$ vs. Steady State Phase Error

need for a good carrier tracking system.

Once again, it should be noted that the level coalescion phenomenon was ignored in the derivation of the upper and lower bounds. A general treatment of the probability of coalescion is difficult, but we shall derive an expression for the binary input case.

The two possible coalescion events are: (i) $R_I = -2d$ transmitted, and sliced as 2d and (ii) $R_I = 2d$ transmitted, and sliced as $-2d$. The probability of the first event occuring is

$$P[\lambda_k > d \mid Z_k, \phi, R_I = -2d]$$

$$= P[n_k > d(2\cos\phi + 1) - Z_k]$$

$$= Q[(d(2\cos\phi + 1) - Z_k)/\sigma] \qquad (4.38)$$

while the probability of the second event is

$$P[\lambda_k < -d \mid Z_k, \phi, R_I = 2d]$$

$$= P[n_k < -d(2\cos\phi + 1) - Z_k]$$

$$= Q[(d(2\cos\phi + 1) + Z_k)/\sigma] \tag{4.39}$$

Thus the probability of coalescion is

$$P = P[\lambda_k > d | Z_k, \phi, R_I = -2d] P[R_I = -2d]$$

$$+ P[\lambda_k < -d | Z_k, \phi, R_I = 2d] P[R_I = 2d] \tag{4.40}$$

Substituting (4.38) and (4.39) into (4.40) and noting that $P[R_I = -2d] = P[R_I = 2d] = 1/4$, we get

$$P_c = 1/4 \{Q[(d(2\cos\phi+1)-Z_k)/\sigma]$$

$$+ Q[(d(2\cos\phi+1)+Z_k)/\sigma]\} \tag{4.41}$$

If we now set Z_k to its maximum value $8d/\pi$, we find that the upper bound on the probability of coalescion is

$$P_c < 1/4 \{Q[\frac{d}{\pi}(2\cos\phi+1+\frac{8\sin\phi}{\pi})]$$

$$+ Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{8}{\pi}\sin\phi)]\} \tag{4.42}$$

The upper bound on the error rate including the effects of coalescion is simply the error rate neglecting coalescion (4.37) for $L = 2$ minus the probability of
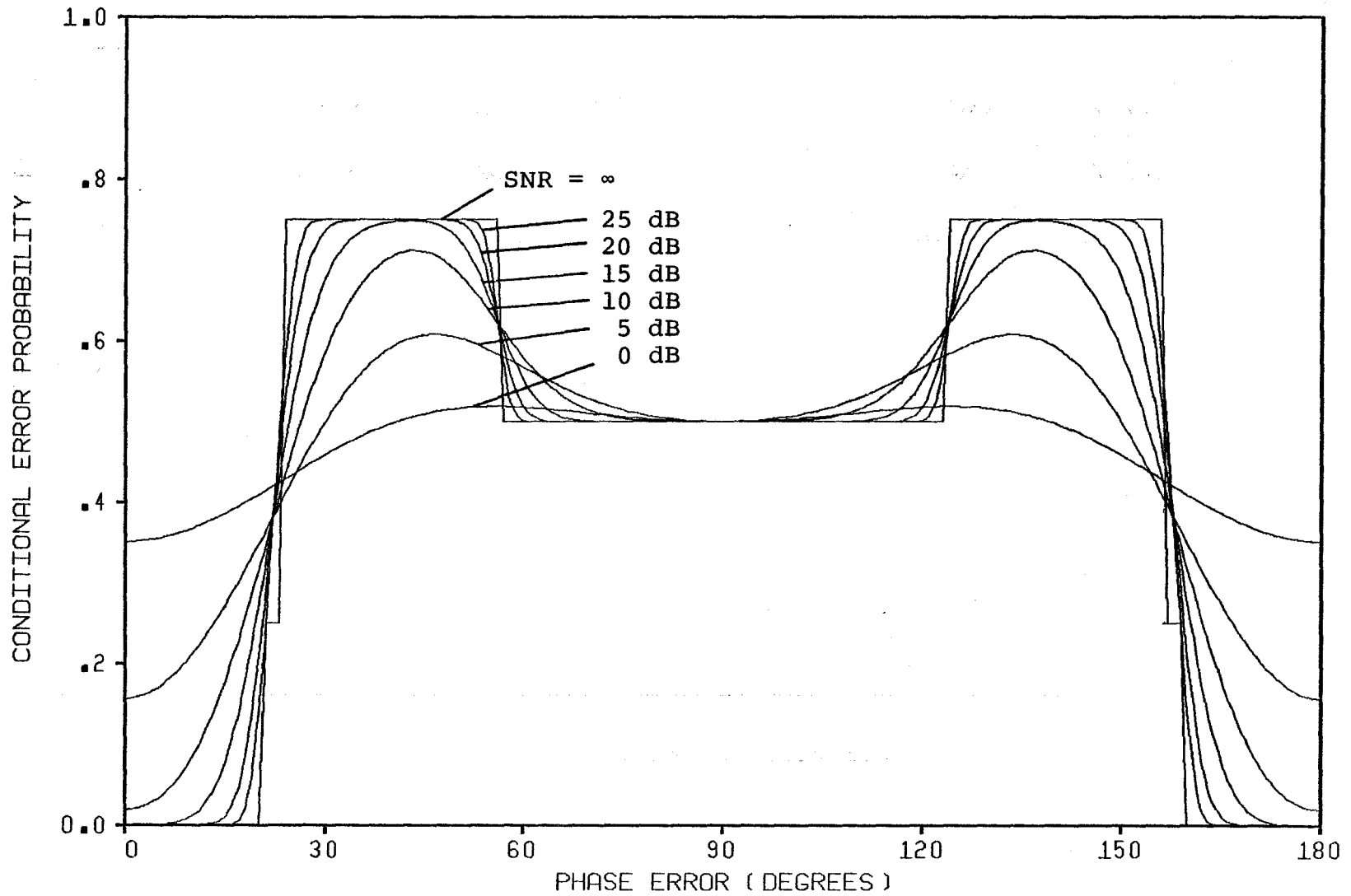
Figure 4.6   Conditional Error Probability vs. Steady State Phase Error

coalescion (4.42). This gives

$$P*[E|\phi] \leq 1/2\{ Q[\frac{d}{\sigma}(1+\frac{8}{\pi}\sin\phi)] + Q[\frac{d}{\sigma}(1-\frac{8}{\pi}\sin\phi)]\}$$

$$+ 1/4\{Q[\frac{d}{\sigma}(2\cos\phi-1+\frac{8}{\pi}\sin\phi)] + Q[\frac{d}{\sigma}(2\cos\phi-1-\frac{8}{\pi}\sin\phi)]\}$$

$$- 1/4\{Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{8}{\pi}\sin\phi)] + Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{8}{\pi}\sin\phi)]\}$$

$$(4.43)$$

The additional terms due to coalescion do not become significant until the phase error becomes quite large. They may therefore be neglected for the value of phase errors considered earlier ($\phi < 10°$).

The function (4.43) is plotted as a function of for various values of SNR in Fig. 4.6. This once again illustrates the need to keep the phase error small.

4.2.2 Eye Diagram Effects

Several eye diagrams were generated to examine the effect of a constant phase error on the partial response eye. Figs. 4.7 and 4.8 show the resulting eye diagrams for a binary input system with a 5° and 10° phase error respectively. Figs. 4.9 and 4.10 show the eye diagrams for
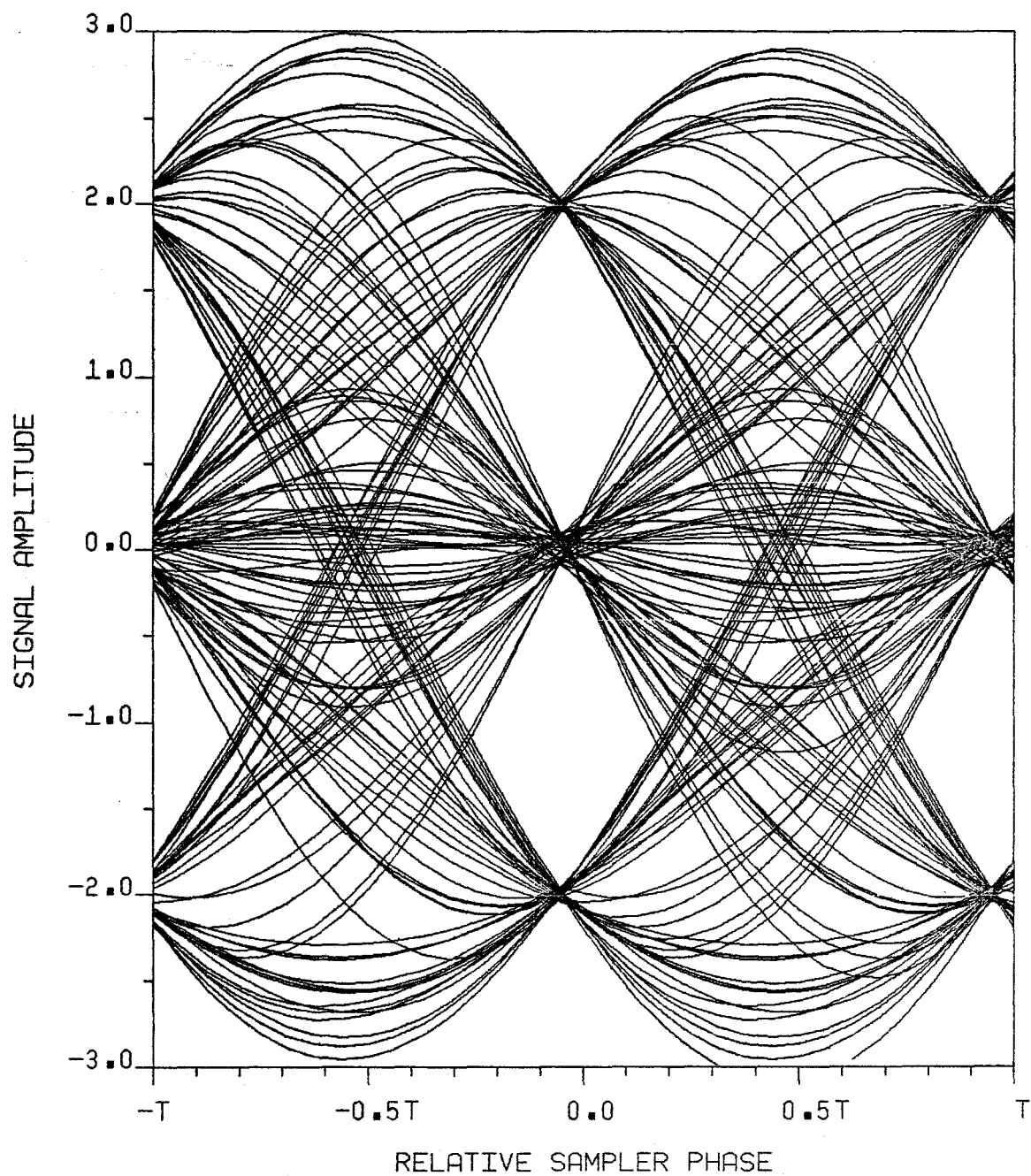
Figure 4.7   Demodulator Output Eye Diagram With
5° Phase Error and Binary Input
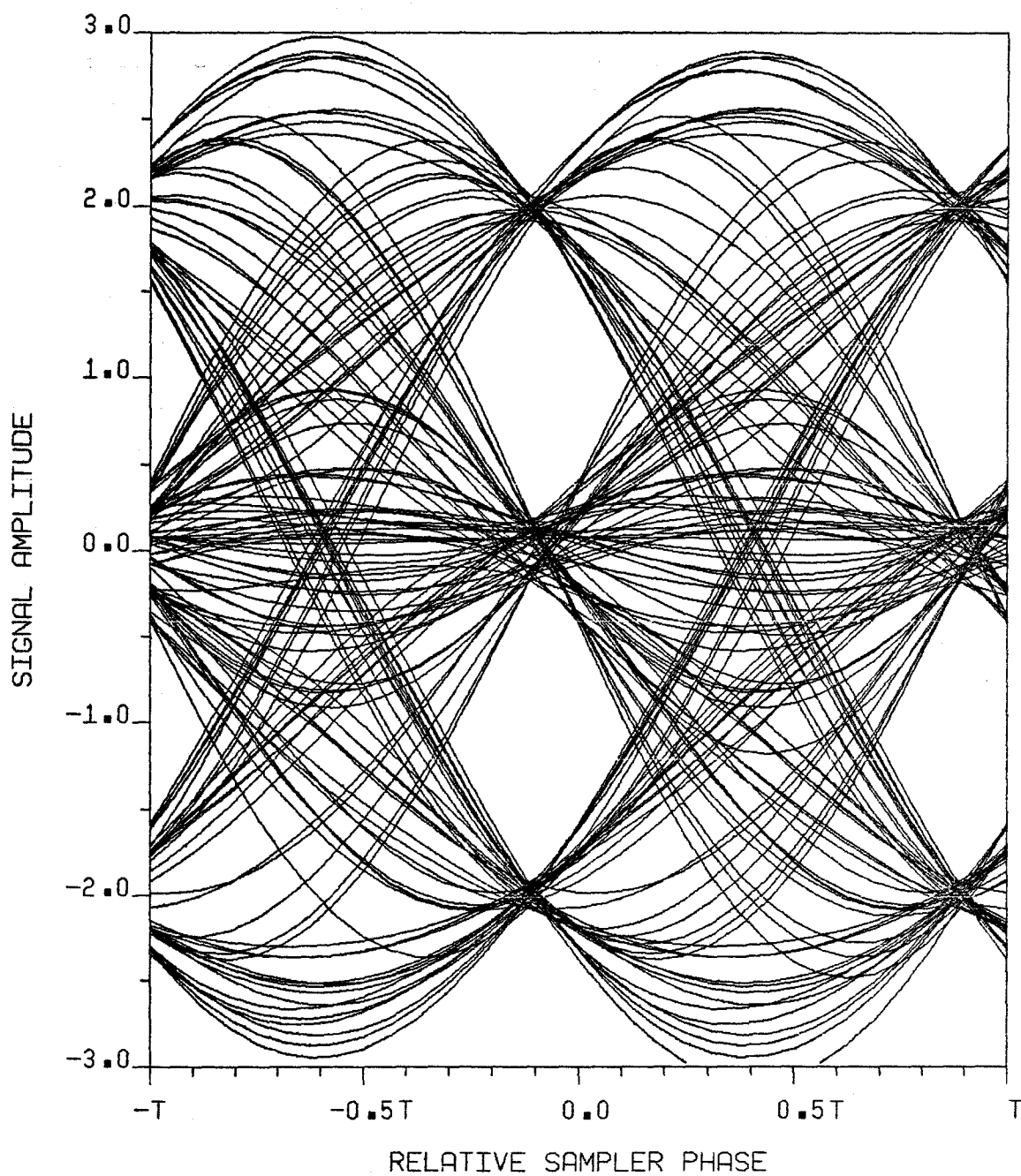
Figure 4.8 Demodulator Output Eye Diagram With
10° Phase Error and Binary Input
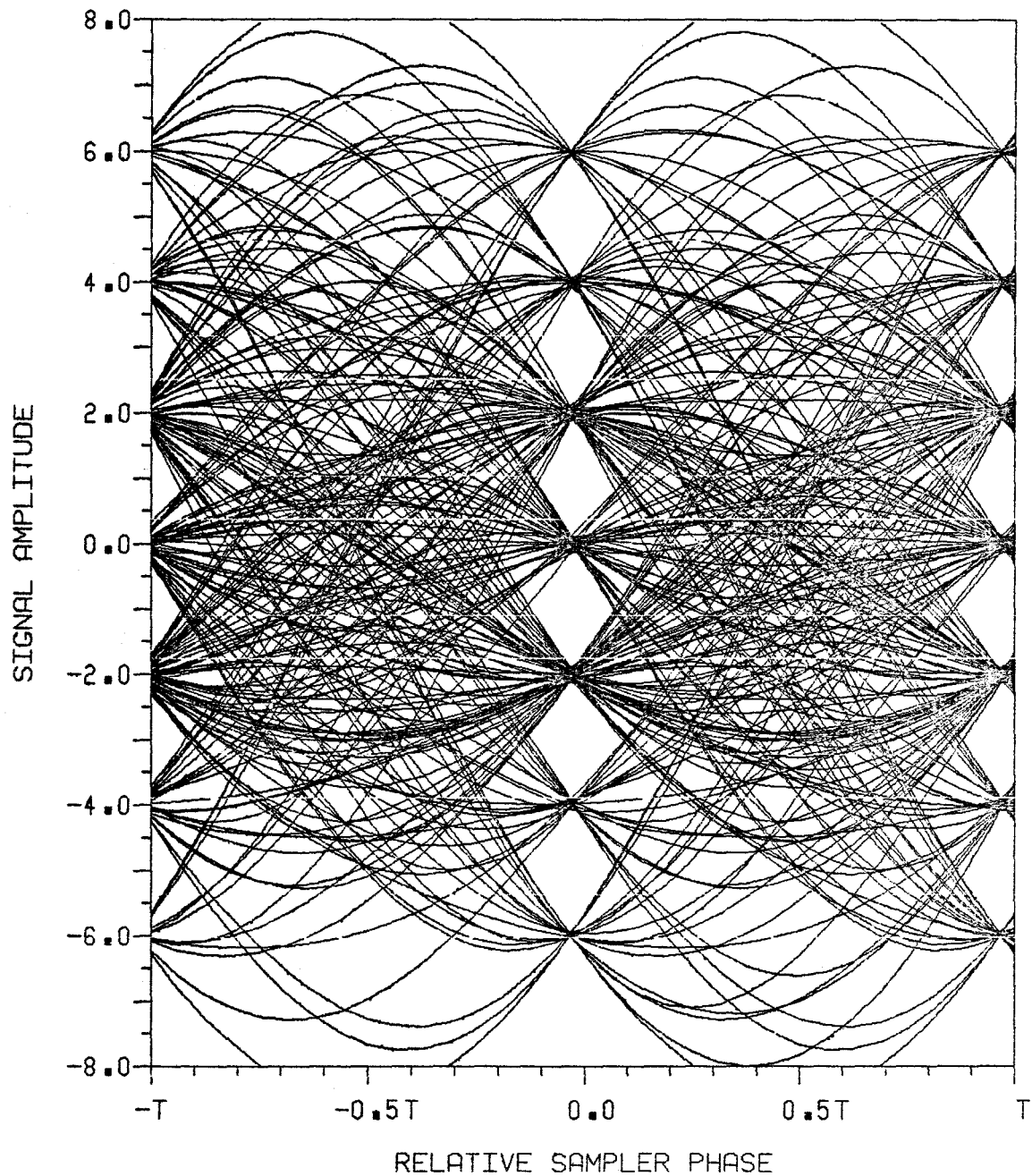
Figure 4.9 Demodulator Output Eye Diagram With
3° Phase Error and 4 Level Input
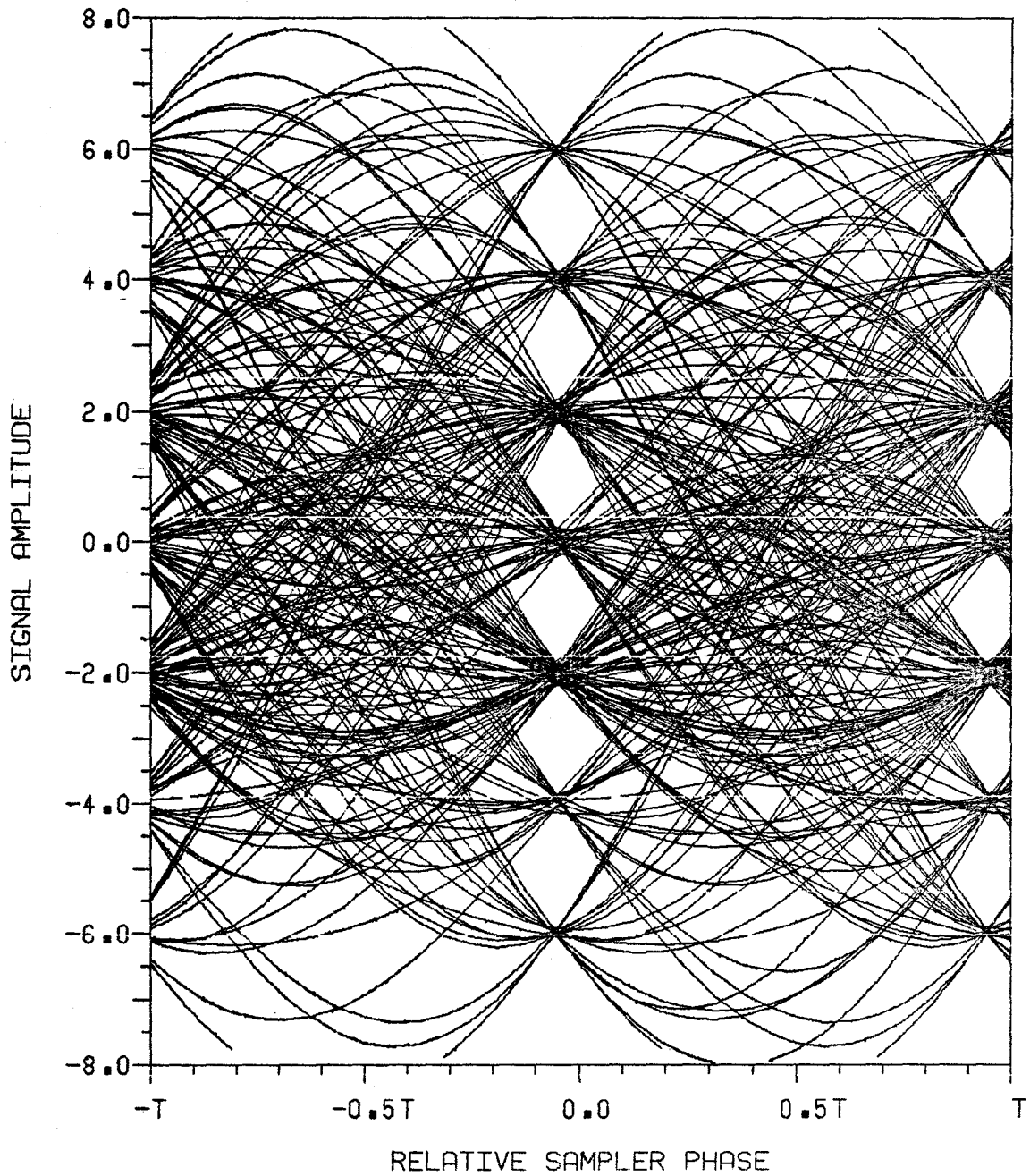
Figure 4.10 Demodulator Output Eye Diagram With
5° Phase Error and 4 Level Input

a 4-level system with phase errors of 3° and 5°. We notice
that the eye closes somewhat at the nominal sampling instant
t = kT for both the binary and 4 level systems. A 5° phase
error in the 4 level case causes a significant eye closure,
which confirms the degradations obtained in the previous
section. Even for a 10° phase error, the binary input eye
is still more than 50% open.

An additional effect which manifests itself in the
eye diagrams is that the optimum sampling instant where the
eye has its maximum opening no longer occurs at t = kT, the
optimum sampling instant for no phase error. This effect
was predicted by Smith in [47], and it shows that by
changing the sampler phase slightly, the effects of a phase
error may be compensated resulting in only a small
performance degradation rather than the large values shown
in Fig. 4.5. If the phase error is kept small through
proper design of a phase tracking loop, however, there is no
need for shifting the sampler instant and hence no need for
the additional complex timing circruitry.

4.3 An SSB Carrier Tracking Loop

The SSB carrier phase tracking loop to be considered
is shown in Fig. 4.11. Both the inphase and quadrature
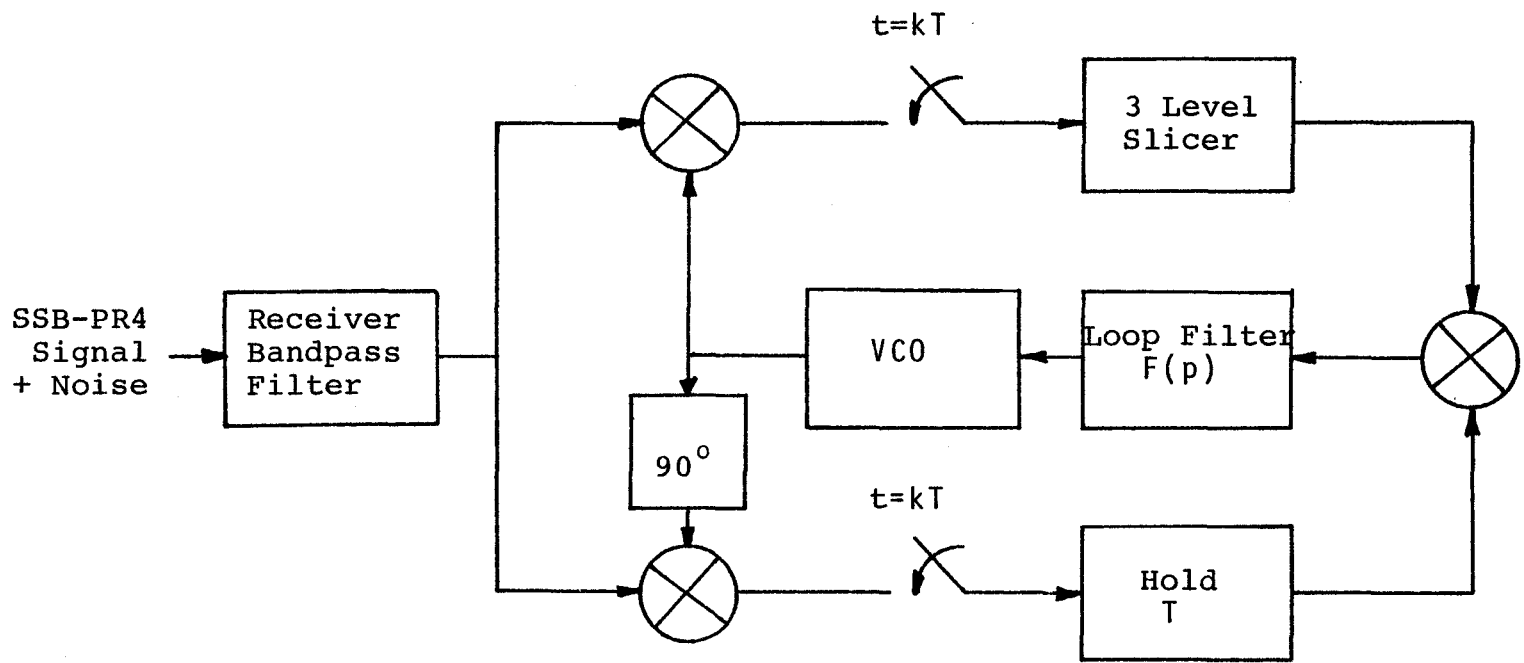
Figure 4.11   Decision-Directed SSB Carrier Tracking Loop

channels are sampled at the nominal sampling time $t = kT$. A 3-level decision is made in the inphase channel and then cross-multiplied by the quadrature sample to provide an error signal to the loop filter and voltage controlled oscillator (VCO). The multiplication operation is effectively the cross-correlation between the local estimate $\tilde{R}_I$ of the transmitted PR symbol and the quadrature sample. The loop acts to drive the cross-correlation to zero, which only occurs when there is no phase error present.

In the following analysis we shall assume perfect symbol timing and amplitude control (AGC) but the means for providing such are outside the scope of this thesis and will not be considered.

4.3.1 Equation of Loop Operation

The received SSB-PR4 signal at the input of the bandpass filter in Fig. 4.11 can be written as

$$s(t) = a(t)\cos[\omega_c t + \theta(t)] - \hat{a}(t)\sin[\omega_c t + \theta(t)]$$

$$+ n_i(t) \qquad (4.44)$$

where

$$n_i(t) \triangleq n_1(t)\cos\left(\omega_c + \frac{\pi}{4T}\right)t + n_2(t)\sin\left(\omega_c + \frac{\pi}{4T}\right)t \qquad (4.45)$$

is a narrow-band noise of zero mean, spectral density $N_0/2$ and centre frequency $\omega_c + \pi/4T$ for a PR4 upper sideband signal input, and where

$$\theta(t) \triangleq \Omega_0 t + \theta_0 \qquad (4.46)$$

with Ω_0 the Doppler shift of the carrier frequency from its nominal value of ω_c and θ_0 a uniformly distributed phase angle, and where $a(t)$ and $\hat{a}(t)$ are given by (3.9) and (3.10).

The input signal is multiplied by the locally generated signals

$$Z_I(t) = \sqrt{2}K_1 \cos[\omega_c t + \tilde{\theta}(t)] \qquad (4.47)$$

$$Z_Q(t) = \sqrt{2}K_1 \sin[\omega_c t + \tilde{\theta}(t)] \qquad (4.48)$$

where K_1 is the VCO RMS amplitude and $\tilde{\theta}(t)$ is the local estimate of $\theta(t)$. The inphase and quadrature outputs after multiplication are

$$r_I(t) = K_1 K_m \{a(t)\cos\phi(t) - \hat{a}(t)\sin\phi(t) + N_I[t,\phi(t)]\} \qquad (4.49)$$

$$r_Q(t) = K_1 K_m \{a(t)\sin\phi(t) + \hat{a}(t)\cos\phi(t) + N_Q[t,\phi(t)]\} \quad (4.50)$$

where K_m is the gain of the multiplier, $\phi(t) \triangleq \theta(t) - \tilde{\theta}(t)$ is the loop phase error and where the two noise processes are given by

$$N_I[t,\phi(t)] \triangleq n_I(t)\cos\phi(t) - n_Q(t)\sin\phi(t) \quad (4.51)$$

$$N_Q[t,\phi(t)] \triangleq n_I(t)\sin\phi(t) + n_Q(t)\cos\phi(t) \quad (4.52)$$

with $n_I(t)$ and $n_Q(t)$ defined as in (2.45) and (2.46).

Denoting $R_I \triangleq a(kT)$ and $R_Q \triangleq a(kT)$, the error signal $e(t)$ at the loop filter input after the sampling and decision operation is

$$e(t) = K_1 K_m \tilde{R}_I \{R_I \sin\phi(t) + R_Q\cos\phi(t) + N_Q[t,\phi(t)]\}$$

$$= K_1 K_m \{\tilde{R}_I R_I \sin\phi(t) + \tilde{R}_I R_Q\cos\phi(t) + \tilde{R}_I N_Q[t,\phi(t)]\}$$

$$\text{for } kT < t < (k+1)T \quad (4.53)$$

Now the instantaneous frequency of the VCO output is related to the loop filter output according to the relation

$$\frac{d\hat{\theta}(t)}{dt} = K_V F(p)e(t) \quad (4.54)$$

where K_V is the VCO gain in rad/V/S and $F(p)$ is the transfer function of the loop filter written in differential notation with $p = d/dt$.

Since $\phi(t) = \theta(t) - \tilde{\theta}(t)$, we may rewrite the instantaneous frequency difference as

$$\dot{\phi}(t) = \dot{\theta}(t) - K_V F(p)e(t)$$
$$= \Omega_0 - K_0 F(p)\{\tilde{R}_I R_I \sin\phi(t) + \tilde{R}_I R_Q \cos\phi(t)$$
$$+ \tilde{R}_I N_Q[t, \phi(t)]\} \tag{4.55}$$

where $K_0 \triangleq K_I K_m K_V$ is the open loop gain.

In order to proceed further, we must make several assumptions concerning the nature of the loop. We assume that the loop filter bandwidth is small compared to the data bandwidth and that the noise process correlation time is small compared to the signalling period T. We may then consider the two baseband noise processes N_I and N_Q as essentially white noise Gaussian processes independent of the phase error $\phi(t)$. Also, we assume that the phase error varies much more slowly than the data or noise. Consequently, we may take the statistical average of (4.55) as in [43] to obtain the stochastic differential equation of loop operation.

$$\dot{\phi}(t) = \Omega_0 - K_0 F(p)\{G(\phi) + H^{1/2}(\phi) N_e(t)\} \tag{4.56}$$

where $N_e(t)$ is approximately white Gaussian noise of two-sided spectral density $N_0/2$ W/Hz and

$$G(\phi) = E\{\tilde{R}_I R_I | \phi(t)\}\sin\phi(t) + E\{\tilde{R}_I R_Q | \phi(t)\}\cos\phi(t) \qquad (4.57)$$

$$H(\phi) = E\{\tilde{R}_I^2 | \phi(t)\} \qquad (4.58)$$

where $E\{.\}$ denotes the expectation operator.

The function $G(\phi)$ is the phase detector characteristic or S-curve of the loop and determines the locking properties of the loop while $H(\phi)$ is the so-called noise function of the loop. We normalize $G(\phi)$ to unit slope at $\phi = 0$ and $H(\phi)$ to unit amplitude at $\phi = 0$ and define the functions

$$g(\phi) \triangleq G(\phi)/G'(o) \qquad (4.59)$$

$$h(\phi) \triangleq H(\phi)/H(o) \qquad (4.60)$$

The equation of loop operation (4.55) can then be written in canonical form as

$$\dot{\phi}(t) = \Omega_0 - K_\ell F(p)\{Ag(\phi) + h^{1/2}(\phi) N_e(t)\} \qquad (4.61)$$

where

$$K_\ell \triangleq K_0 H^{1/2}(0) \qquad (4.62)$$

$$A \triangleq G'(0)/H^{1/2}(0) \qquad (4.63)$$

In the following sections, we evaluate the functions $g(\phi)$ and $h(\phi)$.

4.3.2 Loop Phase Tracking Characteristic Evaluation

The phase characteristic $G(\phi)$ may be written as

$$G(\phi) = G_1(\phi)\sin\phi + G_2(\phi)\cos\phi \qquad (4.64)$$

where

$$G_1(\phi) \triangleq E\{\tilde{R}_I R_I | \phi\} \qquad (4.65)$$

$$G_2(\phi) \triangleq E\{\tilde{R}_I R_Q | \phi\} \qquad (4.66)$$

For convenience, we have dropped the time dependence of ϕ.

The first expectation will now be evaluated. Noting

that $\tilde{R}_I = 0, \pm 2d$, we have

$$E\{\tilde{R}_I R_I | \phi\} = E\{2d\ R_I P[\tilde{R}_I = 2d] + (0)R_I P[\tilde{R}_I = 0]$$

$$-2dR_I P[\tilde{R}_I = 2d]\}$$

$$= 2d\ E\{R_I[P[\tilde{R}_I = 2d] - P[\tilde{R}_I = -2d]]\} \qquad (4.67)$$

where the expectation is evaluated over all values of R_I and R_Q. From (4.49) the decision device input at $t = kT$ is

$$\lambda_k = R_I \cos\phi - R_Q \sin\phi + N_I(k) \qquad (4.68)$$

with

$$N_I(k) \triangleq N_I[kT, \phi(kT)] \qquad (4.69)$$

For $R_I = 2d$, we must have $\lambda_k > d$ or

$$N_I(k) > d - R_I \cos\phi + R_Q \sin\phi \qquad (4.70)$$

so that the probability of the received symbol being $+2d$ conditioned on the transmitted symbol R_I, the quadrature signal sample R_Q, and the loop phase error ϕ is

$$P[\tilde{R}_I = 2d | R_I, R_Q, \phi] = Q[(d - R_I \cos \phi + R_Q \sin \phi) / \sigma] \qquad (4.71)$$

Similarly

$$P[\tilde{R}_I = -2d | R_I, R_Q, \phi] = Q[(d + R_I \cos \phi - R_Q \sin \phi) / \sigma] \qquad (4.72)$$

The expectation of (4.67) can be expanded to the following double summation:

$$E\{\tilde{R}_I R_I | \phi\} = 2d \sum_{\{R_Q\}} \sum_{\{R_I\}} R_I \{P[\tilde{R}_I = 2d | R_I, R_Q, \phi]$$

$$- P[\tilde{R}_I = -2d | R_I, R_Q, \phi]\} P[R_I] P[R_Q] \qquad (4.73)$$

with the summations evaluated over all possible values of R_I and R_Q. The quadrature distortion term R_Q is given by (3.16) and is dependent upon an infinite number of input symbols c_k. Furthermore, the probability distribution $P[R_Q]$ of this quantity is unknown for the same reason. Therefore, to permit further analysis, we approximate (3.16) by retaining the first two terms in the summation:

$$R_Q \simeq \frac{4}{\pi}(c_k - 1/3 (c_{k+2} + c_{k-2})) \qquad (4.74)$$

This effectively assumes that the ISI in the quadrature channel is dominated by the neighbouring pulse values.

Since the c_k are identically distributed binary symbols, we have eight possible cases for R_Q which are summarized in the following table:

R_Q	$P[R_Q]$
$20/3\pi$	$1/8$
$4/\pi$	$1/4$
$4/3\pi$	$1/8$
$-4/3\pi$	$1/8$
$-4/\pi$	$1/4$
$-20/3\pi$	$1/8$

We are now in a position to evalute (4.73). Substituting (4.71) and (4.72) into (4.73) and then summing over R_I and R_Q, we get

$$E\{\tilde{R}_I R_I | \phi\} = \frac{d^2}{4}\{8 - \{2Q[\frac{d}{\sigma}(2\cos\phi - 1 + \frac{4}{\pi}\sin\phi)] + 2Q[\frac{d}{\sigma}(2\cos\phi - 1 - \frac{4}{\pi}\sin\phi)]$$

$$+ Q[\frac{d}{\sigma}(2\cos\phi - 1 + \frac{4}{3\pi}\sin\phi)] + Q[\frac{d}{\sigma}(2\cos\phi - 1 - \frac{4}{3\pi}\sin\phi)]$$

$$+Q[\frac{d}{\sigma}(2\cos\phi-1+\frac{20\sin\phi}{3\pi})]+Q[\frac{d}{\sigma}(2\cos\phi-1-\frac{20\sin\phi}{3\pi})]$$

$$+2Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{4\sin\phi}{\pi})]+2Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{4\sin\phi}{\pi})]$$

$$+Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{4\sin\phi}{3\pi})]+Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{4\sin\phi}{3\pi})]$$

$$+Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{20\sin\phi}{3\pi})]+Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{20\sin\phi}{3\pi})]\}\} \quad (4.75)$$

We now turn our attention to the evaluation of $G_2(\phi)$. Again, since $R_I = 0, \pm 2d$, we have

$$E\{\widetilde{R}_I R_Q|\phi\} = 2d\ E\{R_I[P[\widetilde{R}_I = 2d]-P[\widetilde{R}_I=-2d]]\}$$

$$= 2d\ \sum_{\{R_Q\}} R_Q\ \sum_{\{R_Q\}} \{P[\widetilde{R}_I=2d|R_I,R_Q,\phi]-P[\widetilde{R}_I=-2d|R_I,R_Q,\phi]\}$$

$$\cdot P[R_I]P[R_Q] \quad (4.76)$$

The conditional probabilities in the summation are given by (4.71) and (4.72) and R_Q is approximated as in (4.74). Substitution for R_I and R_Q and performing the required summations give

$$E\{\widetilde{R}_I R_Q|\phi\}=\frac{d^2}{6\pi}\{12Q[\frac{d}{\sigma}(1+\frac{4\sin\phi}{\pi})]-12Q[\frac{d}{\sigma}(1-\frac{4\sin\phi}{\pi})]$$

$$+2Q[\frac{d}{\sigma}(1+\frac{4\sin\phi}{3\pi})]-2Q[\frac{d}{\sigma}(1-\frac{4\sin\phi}{3\pi})]$$

$$+10Q[\frac{d}{\sigma}(1+\frac{20\sin\phi}{3\pi})]-10Q[\frac{d}{\sigma}(1-\frac{20\sin\phi}{3\pi})]$$

$$+6Q[\frac{d}{\sigma}(2\cos\phi-1+\frac{4\sin\phi}{\pi})]-6Q[\frac{d}{\sigma}(2\cos\phi-1-\frac{4\sin\phi}{\pi})]$$

$$+Q[\frac{d}{\sigma}(2\cos\phi-1+\frac{4\sin\phi}{3\pi})]-Q[\frac{d}{\sigma}(2\cos\phi-1-\frac{4\sin\phi}{3\pi})]$$

$$+5Q[\frac{d}{\sigma}(2\cos\phi-1+\frac{20\sin\phi}{3\pi})]-5Q[\frac{d}{\sigma}(2\cos\phi-1-\frac{20\sin\phi}{3\pi})]$$

$$+6Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{4\sin\phi}{\pi})]-6Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{4\sin\phi}{\pi})]$$

$$+Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{4\sin\phi}{3\pi})]-Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{4\sin\phi}{3\pi})]$$

$$+5Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{20\sin\phi}{3\pi})]-5Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{20\sin\phi}{3\pi})]\}$$

$$(4.77)$$

The slope of the S curve at $\phi = 0$ is

$$G'(0) = G_1(0) + G_2'(0) \qquad (4.78)$$

From (4.75), we have

$$G_1(0) = 2d^2 (1-Q(\frac{d}{\sigma})-Q(\frac{3d}{\sigma})) \qquad (4.79)$$

while differentiating (4.77) and then setting $\phi = 0$, we obtain

$$G_2'(0) = - \frac{176}{9\pi\sqrt{2\pi}} \, d^2\Delta \, (3 \exp - \frac{\Delta^2}{2} + \exp - 9 \frac{\Delta^2}{2})$$

(4.80)

with $\Delta \triangleq d/\sigma$. Making use of (4.59), (4.75), (4.77), (4.78) (4.79), and (4.80) we may calculate the normalized S-curve of the loop. This function is plotted in Fig. 4.12 for various values of SNR, with $\Delta = d/\sigma$ related to the SNR by means of (3.43) since full transmitter shaping was assumed.

The normalized S-curve is an odd periodic function of ϕ with period π radians. Since the SSB signal space is unchanged after a rotation of π radians, we would expect a two-fold phase ambiguity, as is indeed the case. This ambiguity does not affect the performance of a precoded partial response system as the precoding is a form of differential encoding which makes the data signal independent of the absolute phase of the carrier. Since a received symbol 0 is mapped into binary "-1" and the received symbols ± 2 are mapped into a binary "1", a change in signal polarity (and hence carrier phase) does not affect the decoder performance.
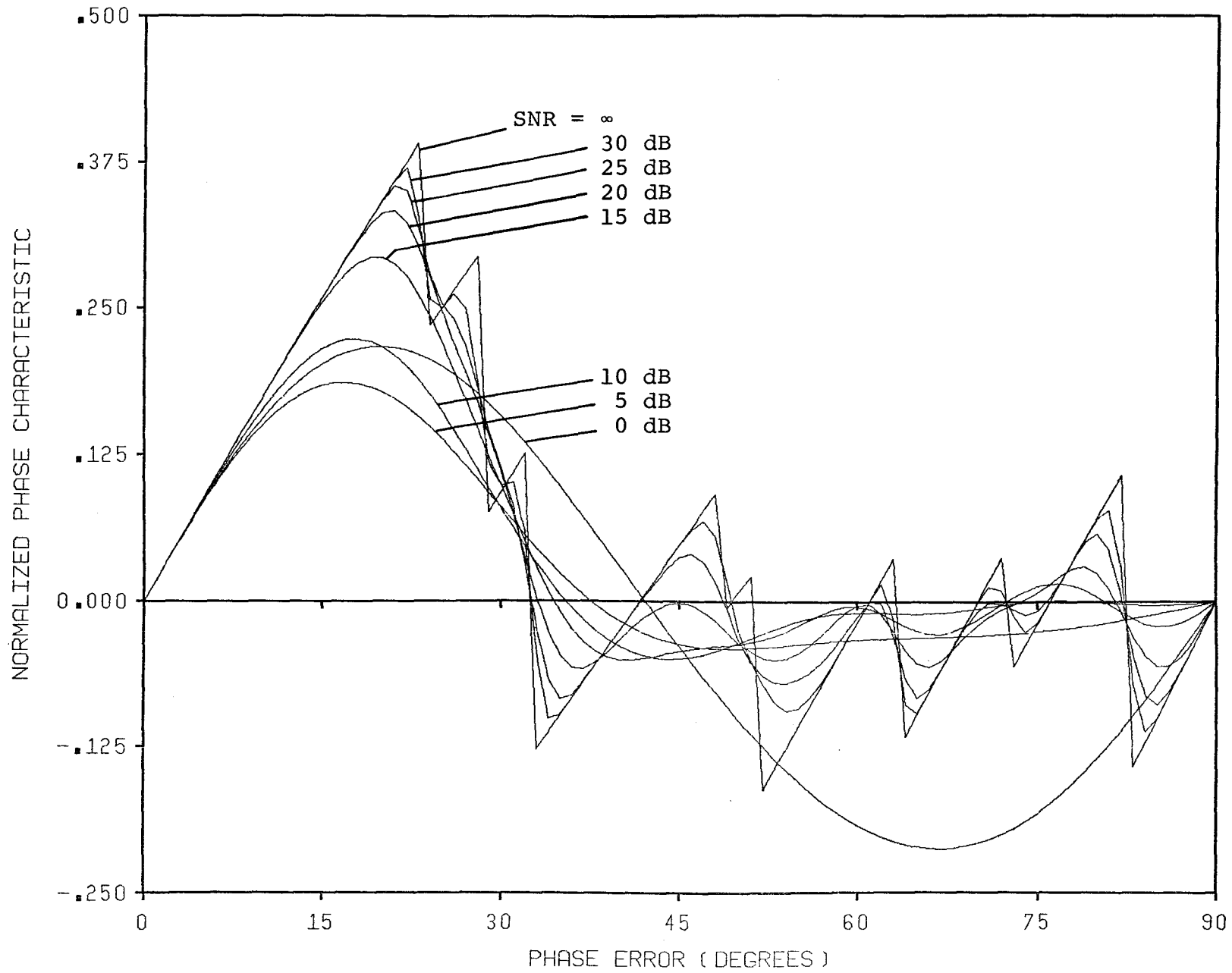
Figure 4.12   Tracking Loop Normalized Phase Detector Characteristic

The S-curve indicates the presence of several false lock points between 0° and 90°. The precoding operation cannot eliminate these lock points so that transmission of a prefix sequence is necessary to ensure that the receiver is locked at the point $\phi = 0°$. This would be necessary at startup and wherever phase lock is lost.

4.3.3 Noise Function Evaluation

The noise function is defined as

$$H(\phi) \triangleq E\{\tilde{R}_I^2 \mid \phi\} \tag{4.81}$$

where the expectation is once again performed over all values of R_I and R_Q. Expanding (4.81) as for the S-curve evaluation, we have

$$E\{\tilde{R}_I^2 \mid \phi\} = 4d^2 \{P[\tilde{R}_I = 2d \mid \phi] + P[\tilde{R}_I = -2d \mid \phi]\}$$

$$= 4d^2 \sum_{\{R_Q\}} \sum_{\{R_I\}} \{P[\tilde{R}_I = 2d \mid R_I, R_Q, \phi] + P[\tilde{R}_I = -2d \mid R_I, R_Q, \phi]\}$$

$$\cdot P[R_I] P[R_Q] \tag{4.82}$$

Proceeding as before, we approximate R_Q as in (4.74) and perform the required substitutions to obtain

$$H(\phi) = 2d^2 + \frac{d^2}{4}\{4Q[\frac{d}{\sigma}(1+\frac{4\sin\phi}{\pi})]+4Q[\frac{d}{\sigma}(1-\frac{4\sin\phi}{\pi})]$$

$$+2Q[\frac{d}{\sigma}(1+\frac{4\sin\phi}{3\pi})]+2Q[\frac{d}{\sigma}(1-\frac{4\sin\phi}{3\pi})]$$

$$+2Q[\frac{d}{\sigma}(1+\frac{20\sin\phi}{3\pi})]+2Q[\frac{d}{\sigma}(1-\frac{20\sin\phi}{3\pi})]$$

$$-2Q[\frac{d}{\sigma}(2\cos\phi-1+\frac{4\sin\phi}{\pi})]-2Q[\frac{d}{\sigma}(2\cos\phi-1-\frac{4\sin\phi}{\pi})]$$

$$-Q[\frac{d}{\sigma}(2\cos\phi-1+\frac{4\sin\phi}{3\pi})]-Q[\frac{d}{\sigma}(2\cos\phi-1-\frac{4\sin\phi}{3\pi})]$$

$$-Q[\frac{d}{\sigma}(2\cos\phi-1+\frac{20\sin\phi}{3\pi})]-Q[\frac{d}{\sigma}(2\cos\phi-1-\frac{20\sin\phi}{3\pi})]$$

$$+2Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{4\sin\phi}{\pi})]+2Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{4\sin\phi}{\pi})]$$

$$+Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{4\sin\phi}{3\pi})]+Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{4\sin\phi}{3\pi})]$$

$$+Q[\frac{d}{\sigma}(2\cos\phi+1+\frac{20\sin\phi}{3\pi})]+Q[\frac{d}{\sigma}(2\cos\phi+1-\frac{20\sin\phi}{3\pi})]\} \quad (4.83)$$

(4.74) and perform the required substitutions to obtain

Setting $\phi = 0$ in this expression gives

$$H(0) = 2d^2[1+Q\frac{(d)}{\sigma}+Q\frac{(\beta d)}{\sigma}] \quad (4.84)$$

Substituting (4.83) and (4.84) into (4.60) gives the
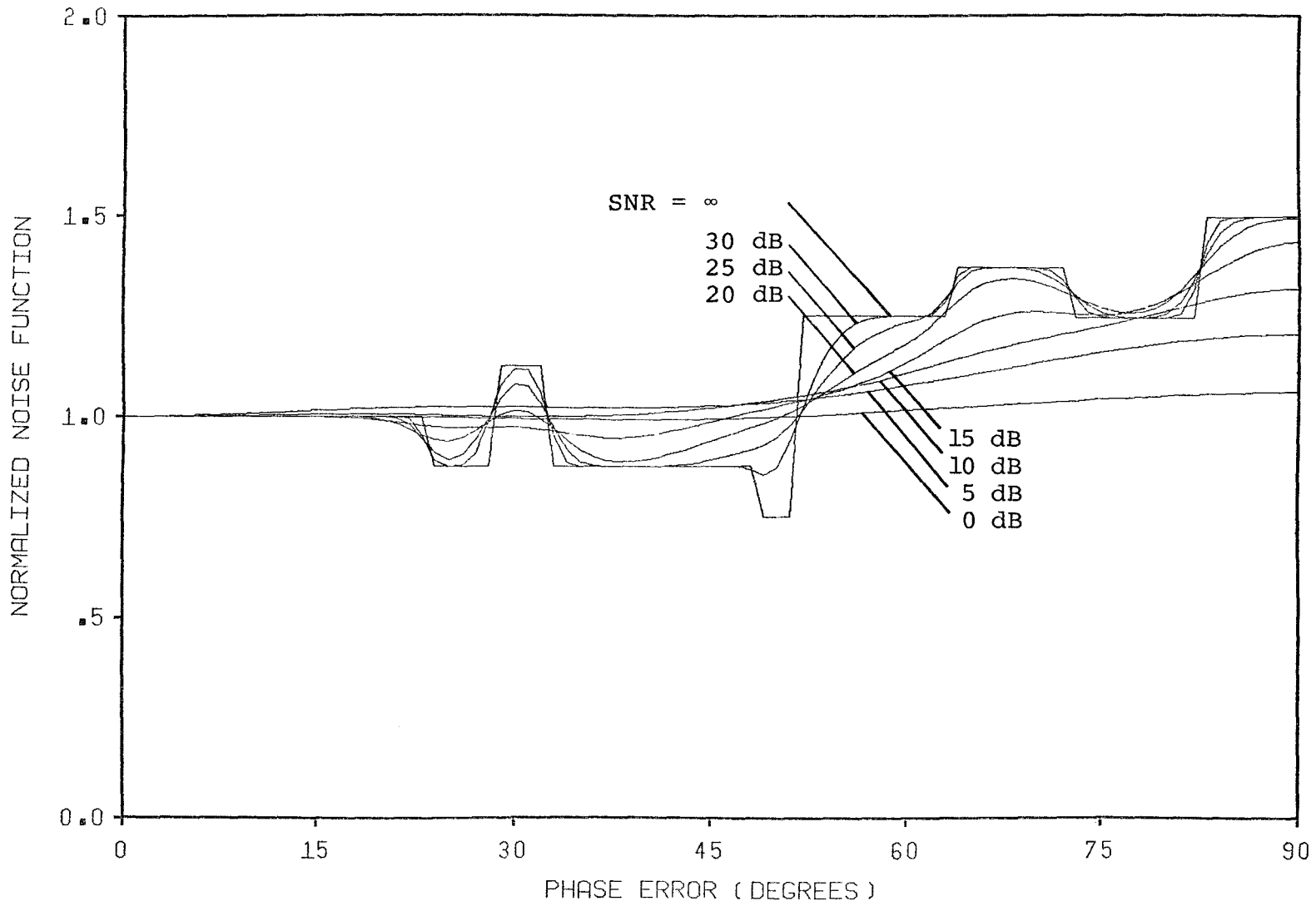
Figure 4.13 Tracking Loop Normalized Noise Function

normalized noise function h(ϕ) which is plotted for various values of SNR in Fig. 4.13.

4.3.4 Probability Density Function of the Phase Error Process

The assumptions made in the derivation of the equation of loop operation permit the use of Fokker-Planck methods [48] to obtain the steady state probability density function (pdf) of the phase error process. To allow a tractable analysis, we assume a first-order loop filter (F(p) = 1) and zero frequency detuning (Ω_0 = 0). Under these conditions, the equation of loop operation becomes

$$\dot{\phi}(t) = -K_\ell \{ Ag(\phi) + h^{1/2}(\phi) N_e(t) \} \qquad (4.85)$$

Since the noise $N_e(t)$ is approximately Gaussian, the steady state pdf of the phase error p(ϕ) satisfies the differential equation

$$\frac{d}{d\phi}[A_0(\phi) p(\phi)] = 1/2 \frac{d^2}{d\phi^2}[B_0(\phi) p(\phi)] \qquad (4.86)$$

where

$$A_0(\phi) \triangleq -K_\ell Ag(\phi) \qquad (4.87)$$

$$B_0(\phi) = \frac{N_0}{2} K_\ell^2 \, h(\phi) \tag{4.88}$$

The solution of this differential equation is [43]

$$p(\phi) = c_0 \exp \left[\int_0^\phi \frac{2A_0(x) - B_0'(x)}{B_0(x)} \, dx \right] \tag{4.89}$$

where c_0 is a normalization constant for unit area under the pdf and the prime denotes differentiation. Substituting for $A_0(x)$ and $B_0(x)$ gives

$$p(\phi) = c_0 \exp \left\{ -\left[\int_0^\phi \frac{\alpha g(x) + h'(x)}{h(x)} \, dx \right] \right\} \tag{4.90}$$

where

$$\alpha \triangleq \frac{4A}{N_0 K_\ell} \tag{4.91}$$

is the loop SNR as defined by Simon and Smith [43].

Now the noise variance at the decision device input is $\sigma^2 = N_0/2T$ so we may rewrite (4.91) as

$$\alpha \triangleq \frac{2A}{\sigma^2 T K_\ell} \tag{4.92}$$

Defining the two side loop bandwith W_L as

$$W_L \overset{\Delta}{=} \frac{AK_\ell}{2} \tag{4.93}$$

the loop SNR becomes

$$\alpha = \frac{A^2}{\sigma^2 T W_L} \tag{4.94}$$

The two-sided data bandwidth is $1/T$, so that the data-to-loop bandwidth ratio may be defined as

$$\delta_S \overset{\Delta}{=} \frac{1}{W_L T} \tag{4.95}$$

which gives

$$\alpha = \frac{A^2 \delta_S}{\sigma^2} \tag{4.96}$$

Substituting for A^2 from (4.63), (4.78) and (4.81) the loop SNR becomes

$$\alpha = 2 \Delta^2 \delta_S \beta(\Delta) \tag{4.97}$$

where

$$\beta(\Delta) = \frac{[1 - Q(\Delta) - Q(3\Delta) - \frac{88}{9\pi\sqrt{2\pi}} \Delta(3 \exp - \frac{\Delta^2}{2} + \exp - \frac{9\Delta^2}{2})]^2}{1 + Q(\Delta) + Q(3\Delta)} \tag{4.98}$$

where $\Delta = d/\sigma$ is computed according to (3.43) to relate it to the SNR.

Equation (4.90) can be readily evaluated numerically and the pdf related to the loop parameters via (4.97) and (4.98).

4.3.5 RMS Phase Error

Once the pdf has been obtained, the variance of the phase error process σ_ϕ^2 can be derived using

$$\sigma_\phi^2 = 2 \int_{-\pi/2}^{\pi/2} p(\phi) \, \phi^2 \, d\phi \tag{4.99}$$

where resolution of the phase ambiguity has been assumed.

Fig. 4.14 illustrates the RMS phase error as a function of SNR calculated using (4.90), (4.97) and (4.98) for a data-to-loop bandwidth ratio (δ_s) of 100. Also shown are the RMS phase error values obtained by computer simulation of the loop for the same set of parameters. We observe that there is close agreement between the two curves. The higher values of RMS phase error obtained in the simulation are due to the effects of data noise which we neglected in the loop analysis. As the loop error signal
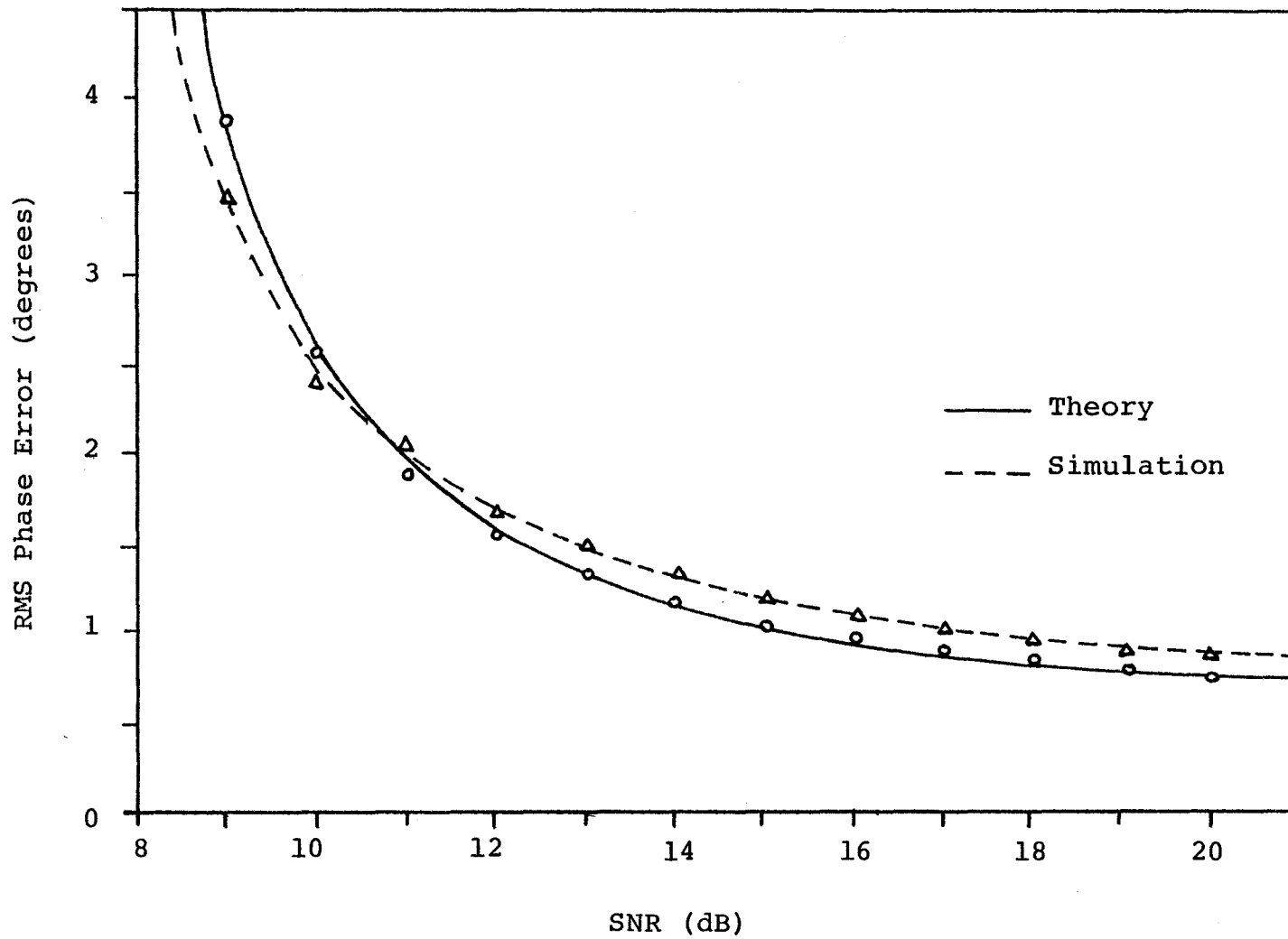
Figure 4.14   Tracking Loop RMS Phase Error for δ_s = 100

depends on the received data sequence the local phase estimate jitters about its mean value when the received data changes. For certain sequences, the loop phase error can become quite large due to this effect. Data noise will be discussed in greater detail in Section 4.3.7.

4.3.6 Average Bit Error Rate

Since the pdf of the phase error process can be easily evaluated, we may determine the average bit error rate of a receiver using this particular tracking loop. Once the phase ambiguity has been resolved, the average error rate is

$$P_E = 2 \int_{-\pi/2}^{\pi/2} P[E \mid \phi]\, p(\phi)\, d\phi$$

(4.100)

where $P[E \mid \phi]$ is the probability of error conditioned on the phase error. This conditional error probability was bounded by (4.43). Thus, the upper bound for the average error rate is defined by

$$P_E \leqslant 2 \int_{-\pi/2}^{\pi/2} P^*[E \mid \phi]\, p(\phi)\, d\phi \qquad (4.101)$$
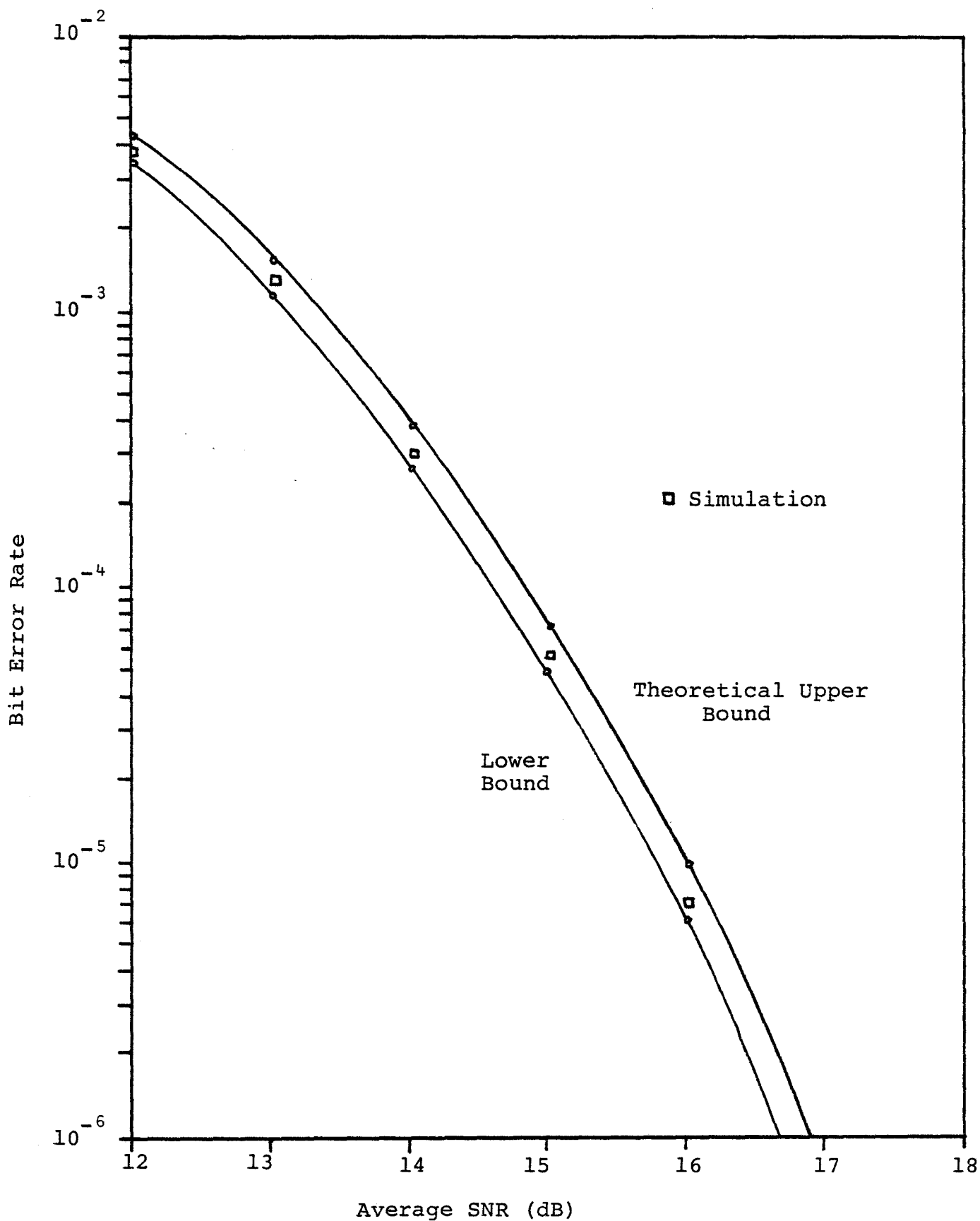
Figure 4.15   Tracking Loop Error Rate Performance for $\delta_s = 100$

with P*[E|φ] given by (4.43). In this case, the effects of
level coalescion have been accounted for as large amounts of
phase error can make this phenomenon significant.

Once   again,   recourse   is   made   to   numerical
integration. The bound on the error rate (4.101) is plotted
in Fig. 4.15 along with the measured error rate obtained by
computer simulation for δ_S = 100. There is close agreement
between theory and simulation.

4.3.7 Data Noise Effects

While data noise has a minor effect on the loop
performance for the loop parameters used in the previous
sections, this   is not always   the case.   Fig. 4.16
illustrates the   RMS phase   error as   a function   of
data-to-loop bandwidth ratio δ_S obtained by simulation of the
tracking loop at an SNR of 100 dB. Under such conditions
the effects of Gaussian noise are negligible. For values of
δ_S < 40, the RMS phase error is quite large and the error
is so large for δ_S< 10 that the loop could be considered
unstable. Thus a large value of δ_S is needed to alleviate
the data noise problem. In an actual system, the value of
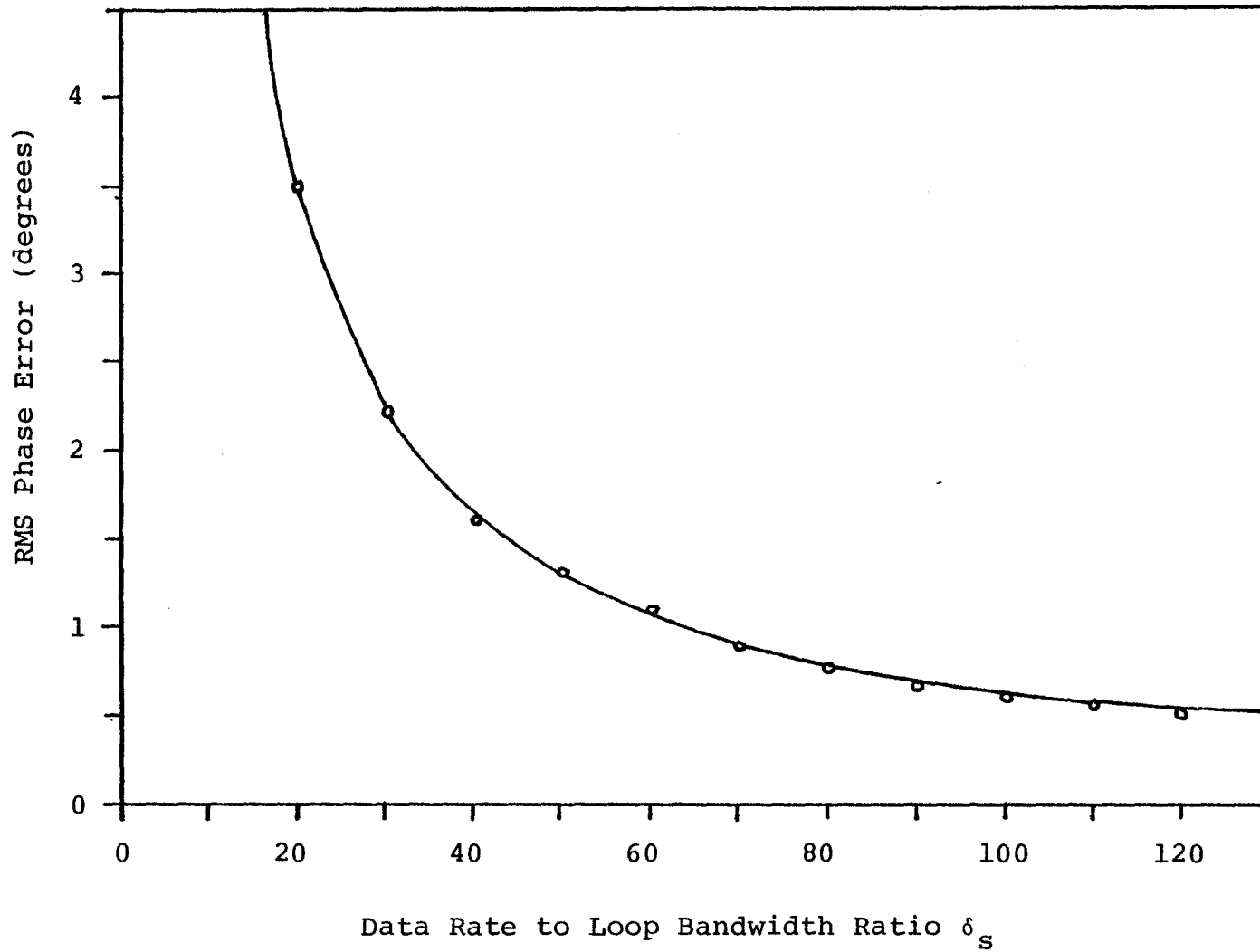
Figure 4.16   Tracking Loop RMS Phase Error Due to Data Noise at SNR = 100dB

RMS phase error due to data noise would be traded off against the transient behaviour of the loop as a large δ_S implies a poor transient response since the open loop gain K_0 would become small (see (4.62), (4.93) and (4.95)).

The large amount of data noise in this particular loop is due to its asymmetrical structure. In an equivalent QAM tracking loop, there would be another decision device in the quadrature channel, the output of which would be multiplied by a sample of the unquantized inphase output and then subtracted from the inphase decision-quadrature sample product. At large SNR, the two cross-multiplier outputs would be approximately the same so that the loop error signal would be small since it is formed from the difference of the multiplier outputs. Since there is no 3 level data signal in the SSB quadrature channel, such subtraction is not possible. However, if a loop structure using a binary decision device such as a decision feedback decoder in the inphase channel were to be devised, then a quadrature decision-inphase sample multiplier stage could be added since it was shown earlier that the data sequence could be extracted from the quadrature channel. Since the signal samplers would be offset by one symbol period, the noise cross-correlation would have to be considered in the loop analysis.

4.4 Conclusions

The performance of an SSB-PR4 system in the presence of a steady-state carrier phase error was analyzed and simulated. To limit the SNR degradation at $P_E = 10^{-5}$ to less than 1 dB requires that the steady state phase error be no greater than 5° for binary inputs and 2° for 4 level inputs. For an SNR degradation of 3 dB, the phase error must be limited to 9.3° for the binary case and 3.8° for the 4 level system. The performance of a decision-directed tracking loop was analyzed and simulated. Data noise was found to be a problem with the particular loop in question and other tracking loop configurations which reduce the data noise should be investigated. Nevertheless, it was shown that carrier tracking loops for SSB which require no pilot tone transmission can in fact be designed. As an extension of this work, the corresponding 4 level input symbol tracking loop should be analyzed.

CHAPTER 5. PERFORMANCE OVER NON-LINEAR CHANNELS

5.1 Introduction

In order to transmit the data signal over large distances power amplification at the transmitter is necessary so that a usable signal level is present at the receiver input. For most microwave and satellite applications, these RF amplification requirements are met through the use of travelling-wave-tube (TWT) amplifiers [49]. Transistor amplifiers may also be used but the TWT is superior in terms of efficiency, bandwidth and range of operating frequencies. To maximize the amplifier output power, these devices must be operated in the saturation or non-linear region of their characteristics. Since SSB is a form of linear modulation, such non-linear distortion will inevitably degrade the performance of the system. In this chapter, we shall determine the extent of this degradation and shall suggest means of overcoming it.

Several digital radio systems using partial response encoding combined with QAM (QPRS) do exist [29], [31], but the encoding is done by filtering the signal after RF

amplification, thus bypassing the non-linearity. In [50], the PR filtering is done before amplification and the performance measured. The effects of SSB and PR encoding together through a radio channel are largely unknown, however.

The presence of the non-linearity creates a significant mathematical obstacle in that an exact theoretical solution is difficult to obtain. By contrast, use of computer simulation techniques provides a very efficient means of obtaining useful results. Also, changes in the system can easily be accomodated. It is for these reasons that computer simulation will be used as the primary tool for measuring the performance of the SSB-PR4 system over a TWT channel.

5.2 The Travelling-Wave-Tube Amplifier

The TWT is basically a device designed for the amplification of weak RF signals. A simplified representation of a typical TWT is shown in Fig. 5.1. The helix and collector are given a positive bias which attracts the electrons emitted by the cathode. An axial magnetic field supplied by an external magnetic coil focuses this electron beam down the centre of the tube.
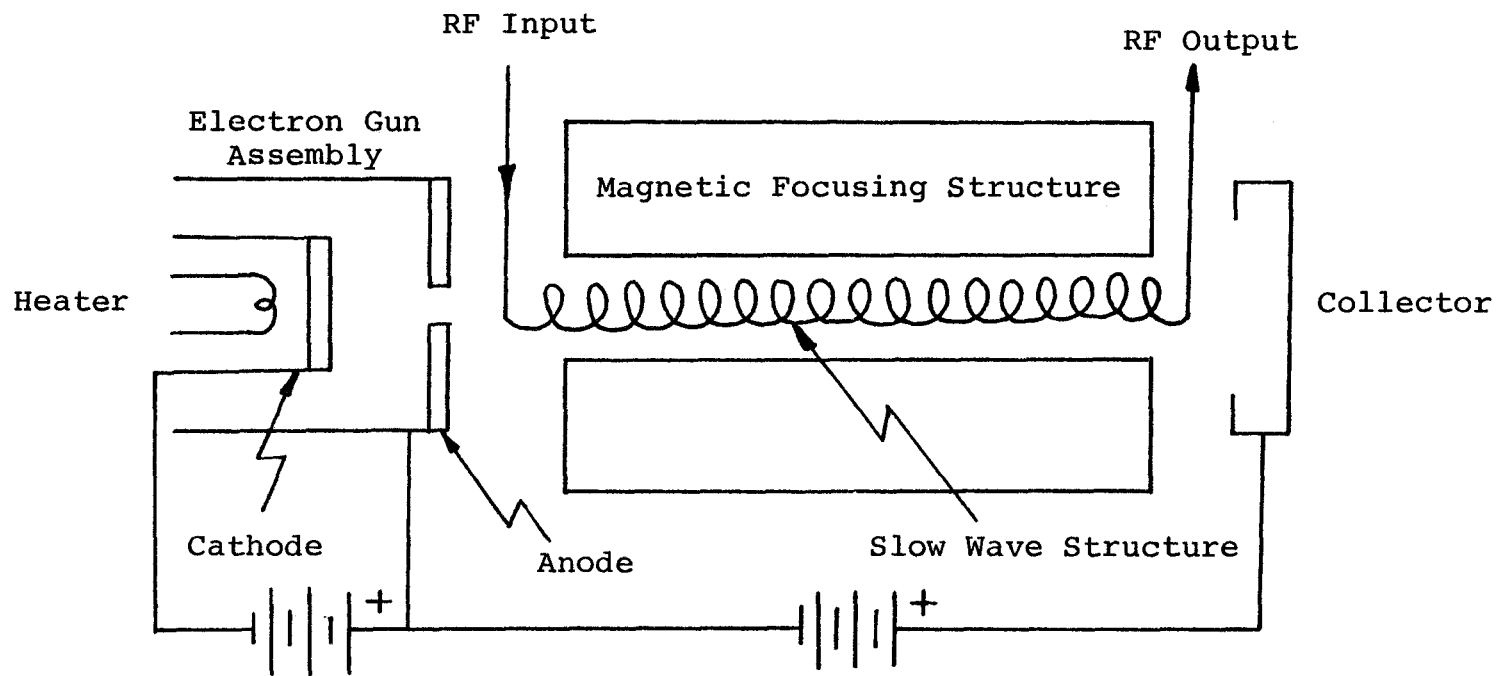
Figure 5.1   Travelling-Wave-Tube Amplifier Structure

The signal to be amplified is fed into the so-called slow-wave structure which often consists simply of a tightly wound wire helix. The RF signal travels through the slow-wave structure at the speed of light but due to the helical path it follows, its velocity in the axial direction is much lower. If the helix is designed such that the phase velocity of the RF wave is the same as the velocity of the electron beam, interaction between the beam and the RF signal occurs and the electrons form bunches as they travel down the length of the tube. These electron bunches induce electric fields in the helix which further interact with the RF wave. The net effect is that the electron beam energy is transferred to the RF wave resulting in amplification of the RF signal.

The power transfer characteristic of a Hughes 261H TWT is shown in Fig. 5.2. We note that at low input power levels, the output power is approximately a linear function of input power. As the input drive level is increased, the output power no longer increases linearly but begins to level off until a certain point where a change in input power gives no change in output power. The device is then said to be in saturation. For maximum output power, it is desirable to operate in this region. The non-linear power transfer behaviour is known as amplitude
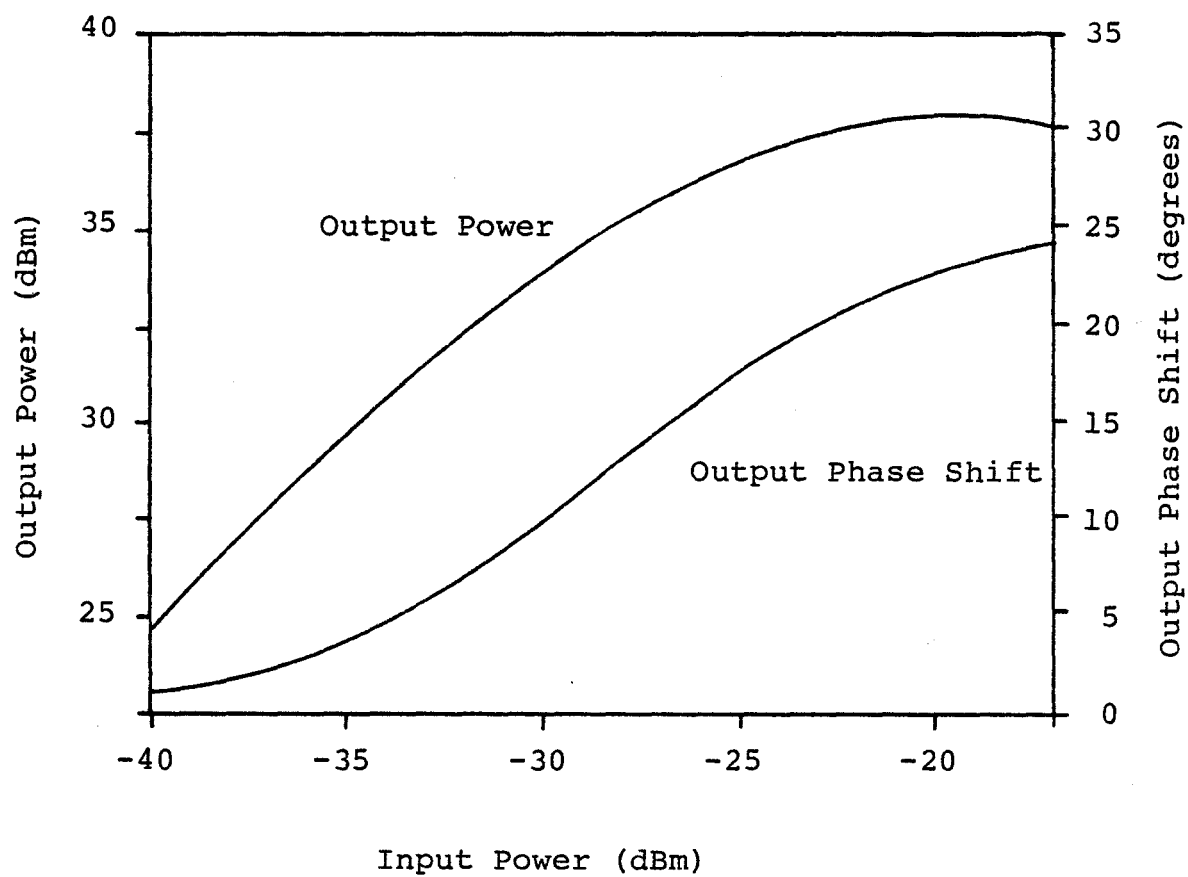
Figure 5.2 Power Transfer Characteristic of
Hughes 261H TWT With Single Carrier Input

modulation-to-amplitude modulation conversion (AM/AM).

An additional effect which manifests itself is that the output phase shift of the device depends on the input power level. This effect is known as amplitude modulation-to-phase modulation conversion (AM/PM) and occurs in a TWT primarily because the average electron beam velocity decreases as the input signal is amplified due to the energy exchange between the beam and the RF wave. This decrease in beam velocity implies that a large input signal will take longer to travel through the tube than a small input signal.

Since SSB is a form of combined amplitude-phase modulation, both the AM/AM and AM/PM effects will have a degradatory effect on the signal.

It should be noted that the TWT characteristic of Fig. 5.2 is a single carrier characteristic in that only a single sinusoidal carrier is used as the input. The non-linear nature of the TWT creates output harmonics at multiples of the input carrier frequency. When multicarrier input signals are present, additional frequency components depending on the frequency differences between the various harmonics and the fundamental result. These intermodulation

(IM) products can distort the individual signals much more than indicated in the figure particularly when the device is operating in the non-linear region. In this study, we shall only consider a single carrier input. We shall also assume that zonal bandpass filtering takes place after amplification so that signal components at harmonics of the carrier frequency may be ignored.

5.2.1 Modelling of the TWT

For the purpose of analysis, we have to develop a mathematical model for the TWT. This model must accurately describe both the AM/AM and AM/PM characteristics of the device. One possible approach, developed in [51], is to consider the TWT as a cascade of an amplitude non-linearity and a phase non-linearity. We shall not make use of this approach; rather, the TWT will be modelled with an inphase nonlinearity Z_p and a quadrature nonlinearity Z_q operating on the input envelope $R(t)$ as in Fig. 5.3 [52].

For a general input signal

$$s(t) = R(t)\cos[\omega_c t + \theta(t)]$$ (5.1)

the TWT output may be written as
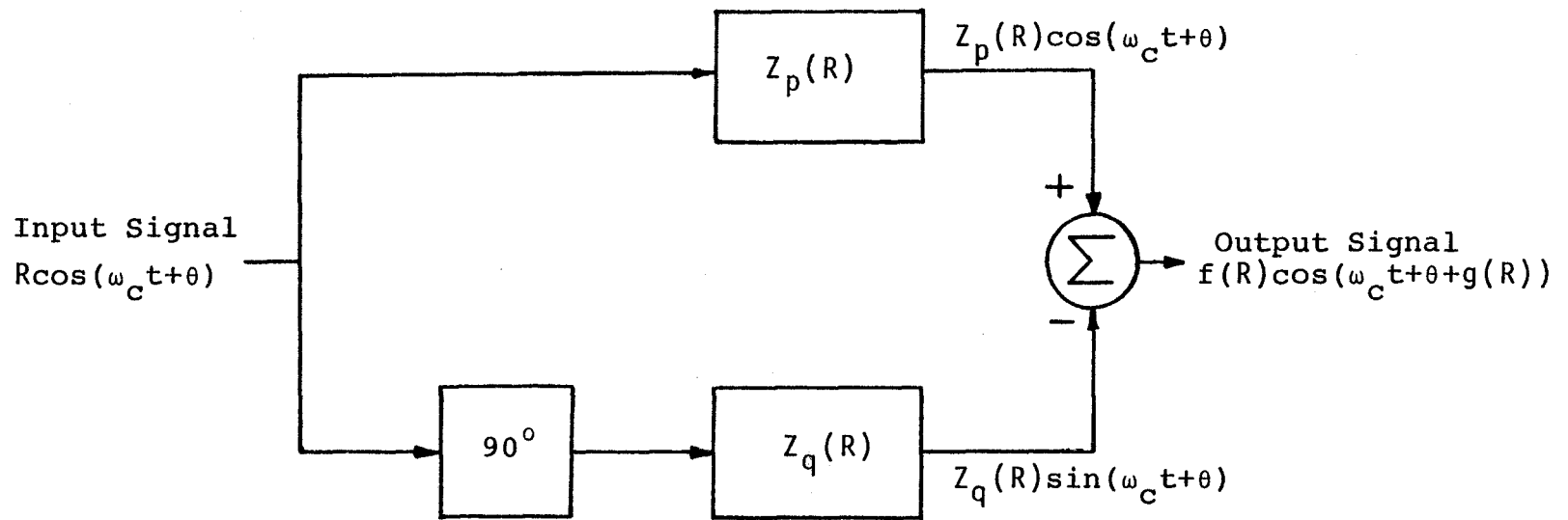
Figure 5.3 Quadrature Model of TWT

$$V(t) = Z_p(R)\cos(\omega_c t+\theta) - Z_q(R)\sin(\omega_c t+\theta) \qquad (5.2)$$

where the time dependence of R and ϕ has been suppressed for clarity. Use of trignometric identities permits us to rewrite the output as

$$V(t) = f(R)\cos[\omega_c t+\theta+g(R)] \qquad (5.3)$$

where

$$f(R) \overset{\Delta}{=} [Z_p{}^2(R)+Z_q{}^2(R)]^{1/2} \qquad (5.4)$$

is the AM/AM distortion term, and

$$g(R) \overset{\Delta}{=} \tan^{-1} \frac{Z_q(R)}{Z_p(R)}$$

$$(5.5)$$

corresponds to the AM/PM conversion. For a linear amplifier, $f(R) = KR$ where K is the amplifier gain and $g(R)$ is a constant phase shift.

The individual non-linearities $Z_p(R)$ and $Z_q(R)$ can be modelled in various ways. We shall use the model developed by Hetrakul and Taylor [52] where the non-linearities take the form

$$Z_p(R) = C_1 R e^{-C_2 R^2} I_0[C_2 R^2] \tag{5.6}$$

$$Z_q(R) = S_1 R e^{-S_2 R^2} I_1[S_2 R^2] \tag{5.7}$$

where $I_n(.)$ is the modified Bessel function of order n. The coefficients C_1, C_2, S_1 and S_2 are obtained from a least-squares fit to the actual tube characteristics. For the Hughes 261H tube, these coefficients are

$$C_1 = 1.61245 \qquad C_2 = .053557$$

$$S_1 = 1.71850 \qquad S_2 = .242218$$

The actual tube characteristics are compared to the model in Fig. 5.4. The advantage of this model over, say, a polynomial approximation of $Z_p(R)$ and $Z_q(R)$ is that only four coefficients are required to completely specify the tube characteristics.

5.2.2 TWT Linearization Techniques

Clearly, a typical TWT characteristic would severely distort an SSB signal. It would be desirable then to modify the characteristic so that the TWT behaves in a fashion
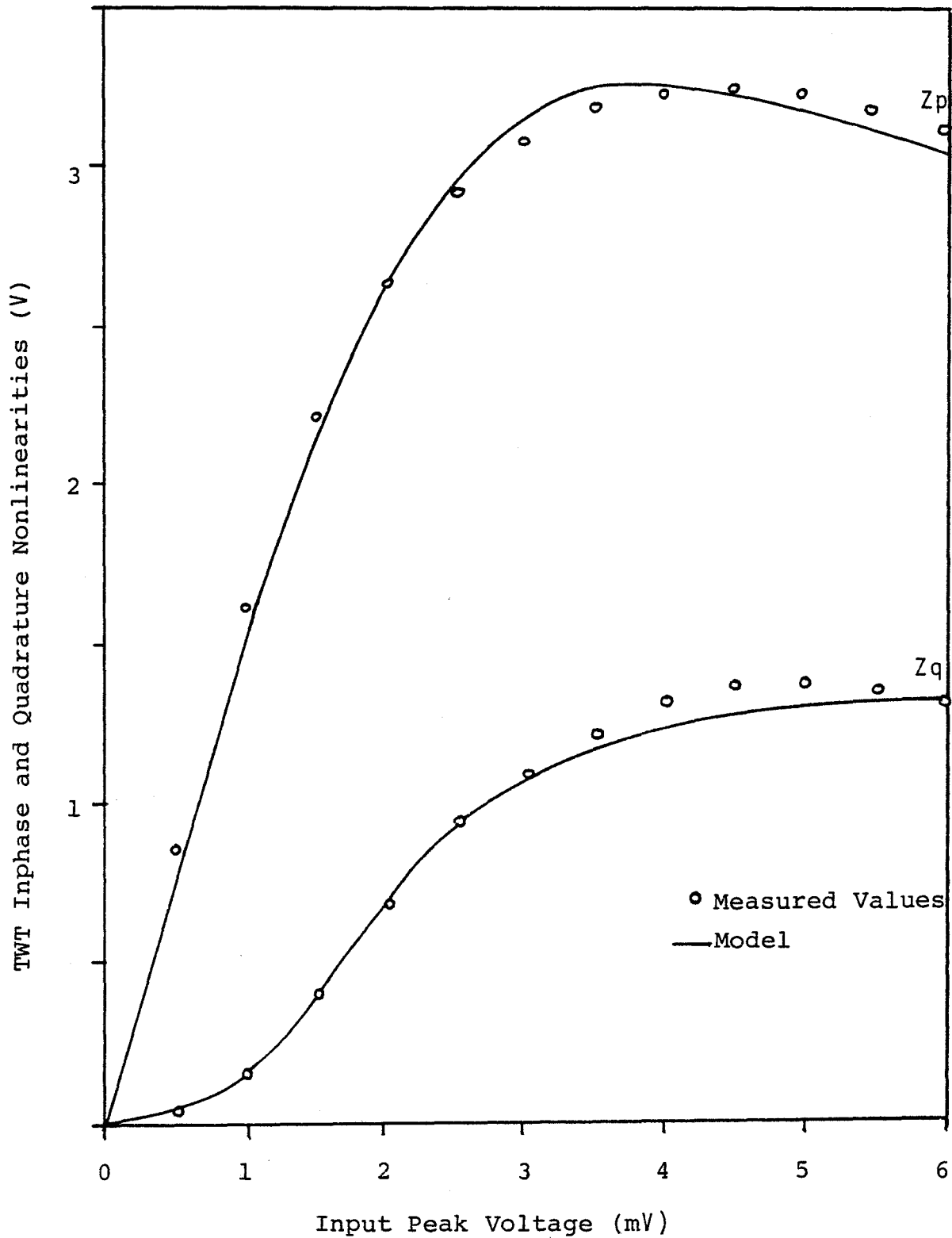
Figure 5.4   Inphase and Quadrature Characteristics
of Hughes 261H TWT

approaching that of a linear amplifier. There exists a variety of linearization techniques [13] of which we shall discuss only two: feed-forward compensation and signal predistortion.

The application of feedforward to microwave amplifiers was extensively studied by Seidel in [53] and has also been considered for satellites in [54]. The system requires two TWT's as shown in Fig. 5.5. The input signal is split between two paths, one of which consists of a TWT while the other consists of a delay line which matches the TWT transit time. The amplified (distorted) signal and the undistorted delayed signal are then fed into an error determining coupler. This coupler is essentially an interferometer with two outputs: the amplified output of TWT1, and the difference between this output after attenuation and the undistorted input. This difference signal forms an error signal which is amplified by TWT2 and then subtracted from the amplified input signal after a delay corresponding to the transit time through the second TWT.

This method is attractive for several reasons. First of all, since no feedback is used, the system is absolutely stable. Secondly, the distortion can be reduced
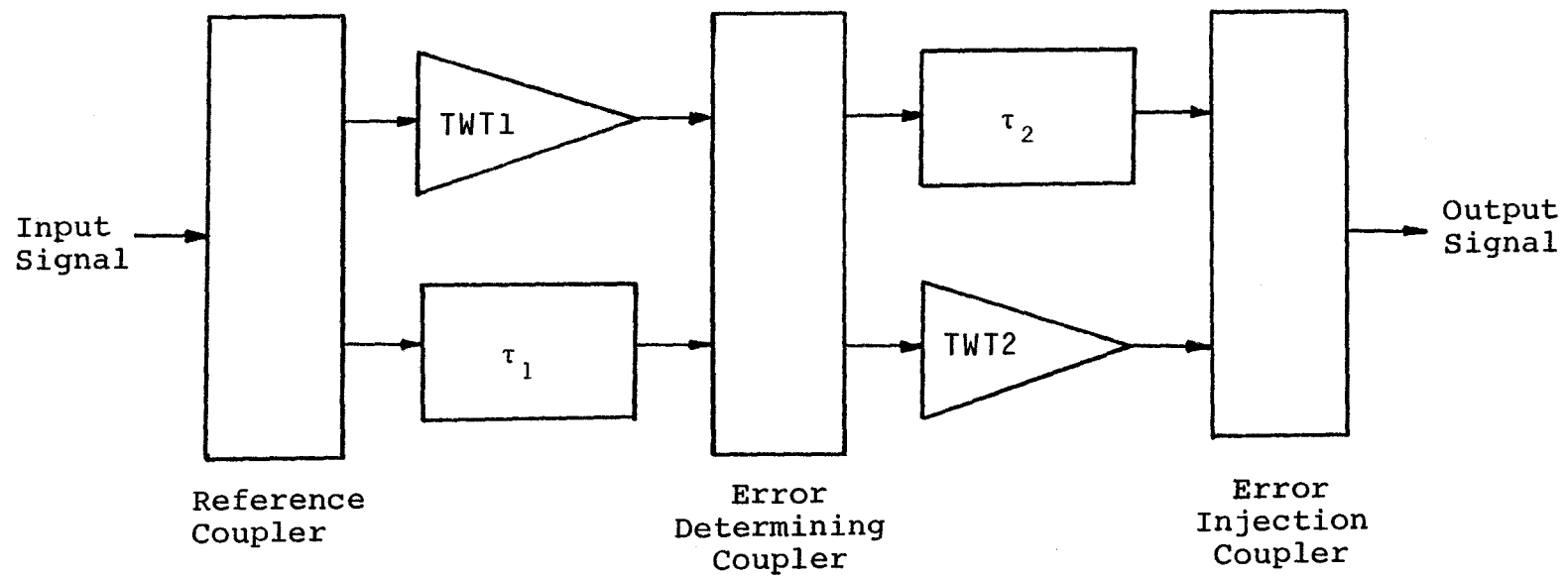
Figure 5.5   Feedforward TWT Linearization

to arbitrarily small values by using a multistage configuration in which each amplifier is embedded in another two-loop correction stage. As the original input signal is used as a reference for each additional stage, there is no cumulative error.

For proper operation, a precise knowledge of the TWT gains and transit times is necessary. Also, the two TWT's must be closely matched and the TWT helix voltages closely regulated. To compensate for any variation in the device characteristics with time, complex circuitry must be introduced. The degree of complexity required is in fact the greatest disadvantage of feed-forward compensation.

In the Bell System AR6A SSB-FDM radio system [14], predistortion was selected as the more economical linearization technique. In this method, an additional amplitude and phase non-linearity is cascaded with the TWT. The non-linearity is carefully chosen so that the combination of distortion due to the TWT and that of the predistorter approximates a piece-wise linear amplifier with no amplitude-dependent phase shift. Once again, a precise knowledge of the TWT characteristics is required. Since the predistorter operates on the signal envelope, it may be implemented at baseband, IF or RF. An additional advantage

is that only a single TWT is required.

The predistorter to be used here was developed in [52]. Two separate envelope non-linearities are used in the configuration of Fig. 5.6. Denoting these non-linearities as $G_p(R)$ and $G_q(R)$ we may write the compensator output after zonal filtering as

$$V_c(t) = G_p(R)\cos(\omega_c t + \theta) - G_q(R)\sin(\omega_c t + \theta)$$

$$= G'(R)\cos[\omega_c t + \theta + \phi(R)] \tag{5.8}$$

where

$$G'(R) \overset{\Delta}{=} [G^2_p(R) + G^2_q(R)]^{1/2} \tag{5.9}$$

$$\phi(R) \overset{\Delta}{=} \tan^{-1} \frac{G_q(R)}{G_p(R)} \tag{5.10}$$

The TWT acts on the new envelope $G'(R)$ to give the output

$$V_o(t) = Z_p\{G'(R)\}\cos[\omega_c t + \theta + \phi(R)]$$

$$- Z_q\{G'(R)\}\sin[\omega_c t + \theta + \phi(R)]$$

$$= f'(R)\cos[\omega_c t + \theta + \alpha(R)] \tag{5.11}$$

$$V_c(t) = G'(R)\cos(\omega_c t + \theta + \phi(R))$$
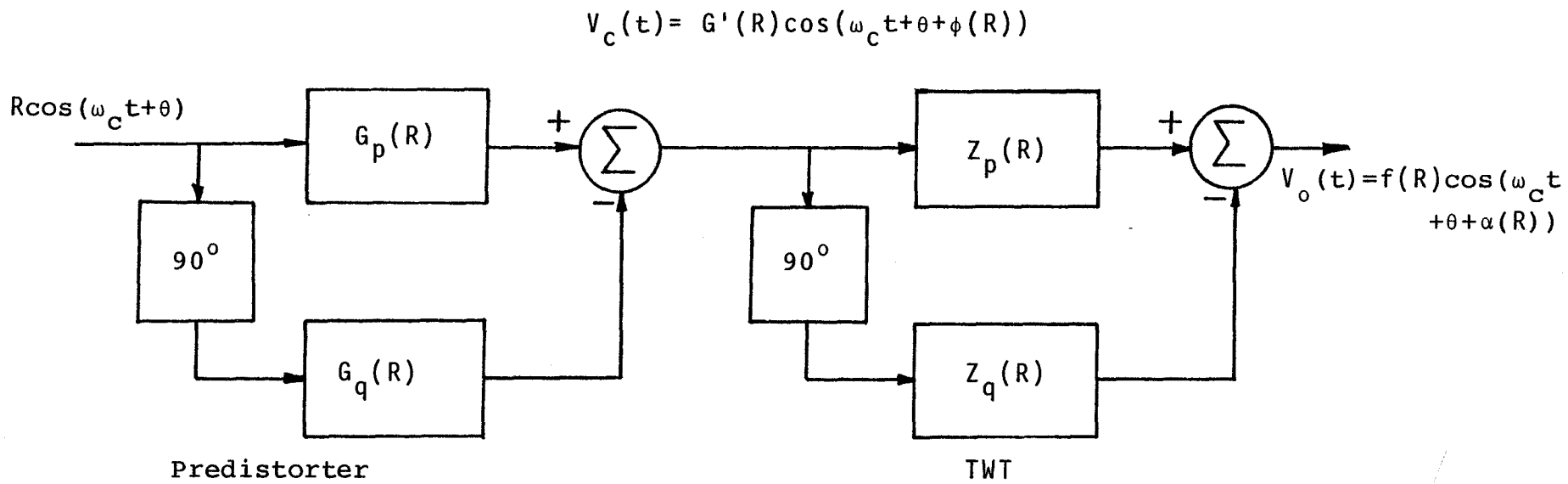
Figure 5.6   TWT Linearization Using Predistortion Compensation

where

$$f'(R) \overset{\Delta}{=} [Z_p^2\{G'(R)\} + Z_q^2\{G'(R)\}]^{1/2} \qquad (5.12)$$

and

$$\alpha(R) \overset{\Delta}{=} \phi(R) + \tan^{-1} \frac{Z_q\{G'(R)\}}{Z_p\{G'(R)\}} \qquad (5.13)$$

We must find $G_p(R)$ and $G_q(R)$ such that the following conditions are satisfied:

$$f'(R) = R_o \frac{R}{R_i} \quad , 0 < R < R_i$$

$$= R_o \quad , R > R_i \qquad (5.14)$$

and

$$\alpha(R) = 0 \quad , \forall R \qquad (5.15)$$

where R_o is the output voltage at saturation, and R_i is the input voltage which saturates the device.

Clearly an exact solution would be difficult to derive, but approximating $G_p(R)$ and $G_q(R)$ as finite order polynomials can lead to a result which is sufficiently close

for our purposes. That is, we let

$$G_p(R) = \sum_{k=1}^{N} a_k R^{2k-1} \qquad (5.16)$$

$$G_q(R) = \sum_{k=1}^{N} b_k R^{2k-1} \qquad (5.17)$$

where the polynominal coefficients a_k and b_k are chosen such that (5.14) and (5.15) are satisfied according to some criterion of goodness of fit.

The compensator coefficients with $N = 6$ for a Hughes 261H tube, derived in [52], are shown in Table 5.1. The corresponding compensator non-linearities are shown in Fig. 5.7 and the overall characteristic of both the compensated and uncompensated tube are shown in Fig. 5.8. The linearized TWT closely approximates a piece-wise linear amplifier and the AM/PM effect has been reduced considerably as well. Use of a higher order predistorter polynomial would result in an even better approximation. A detailed discussion of the implementation of the compensator is found in [52].

K	a_K MILLIVOLT/(MILLIVOLT)$^{2K-1}$	b_K
1	1.00184	- 1.06837E-2
2	1.53746E-1	- 1.76412E-1
3	- 1.64945E-2	1.88326E-2
4	4.53263E-4	- 7.70929E-4
5	1.84654E-8	1.31815E-5
6	- 9.50101E-8	- 7.35811E-8

TABLE 5.1   PREDISTORTION COMPENSATOR COEFFICIENTS
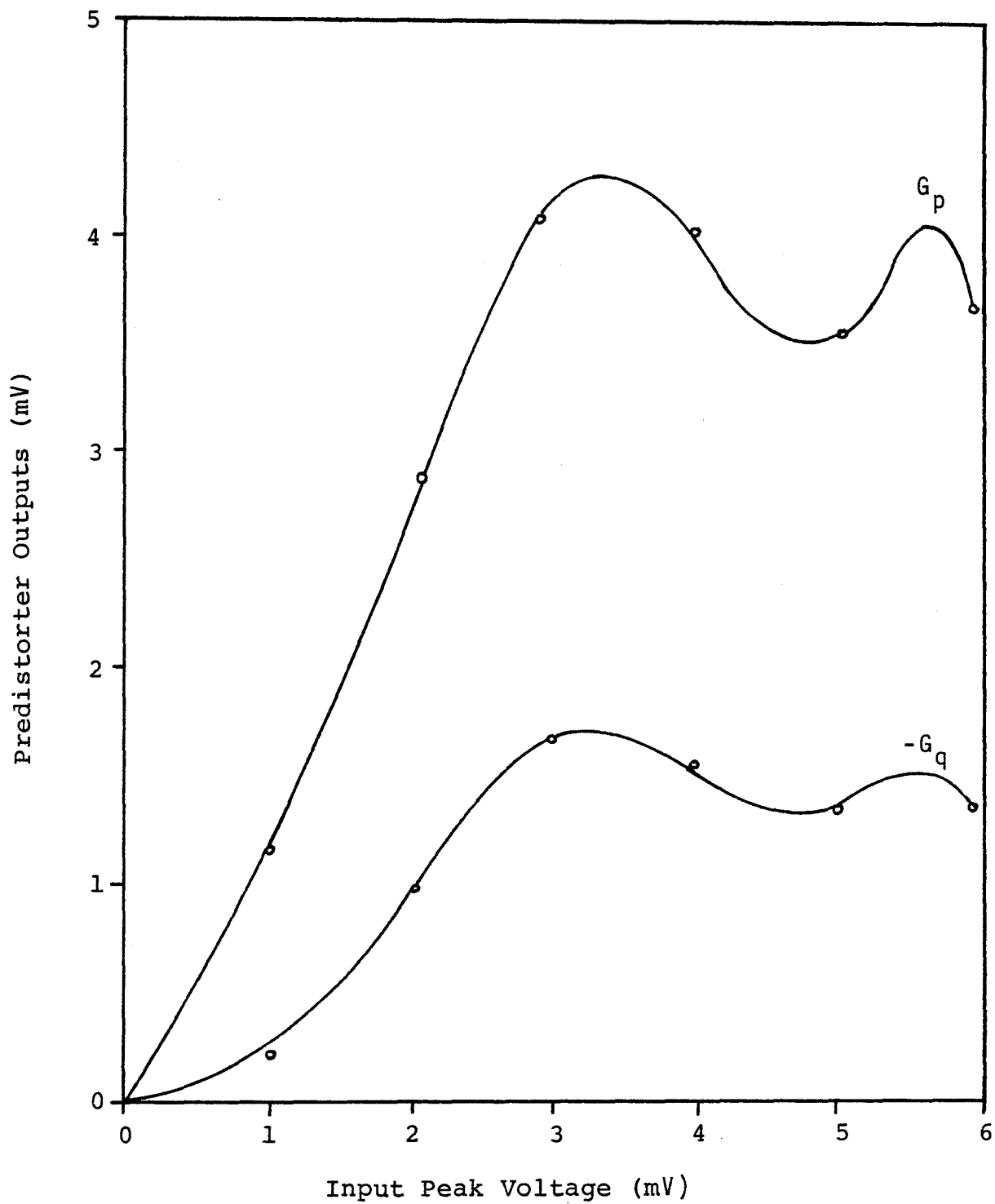
Figure 5.7   Inphase and Quadrature Characteristics
of Predistorter
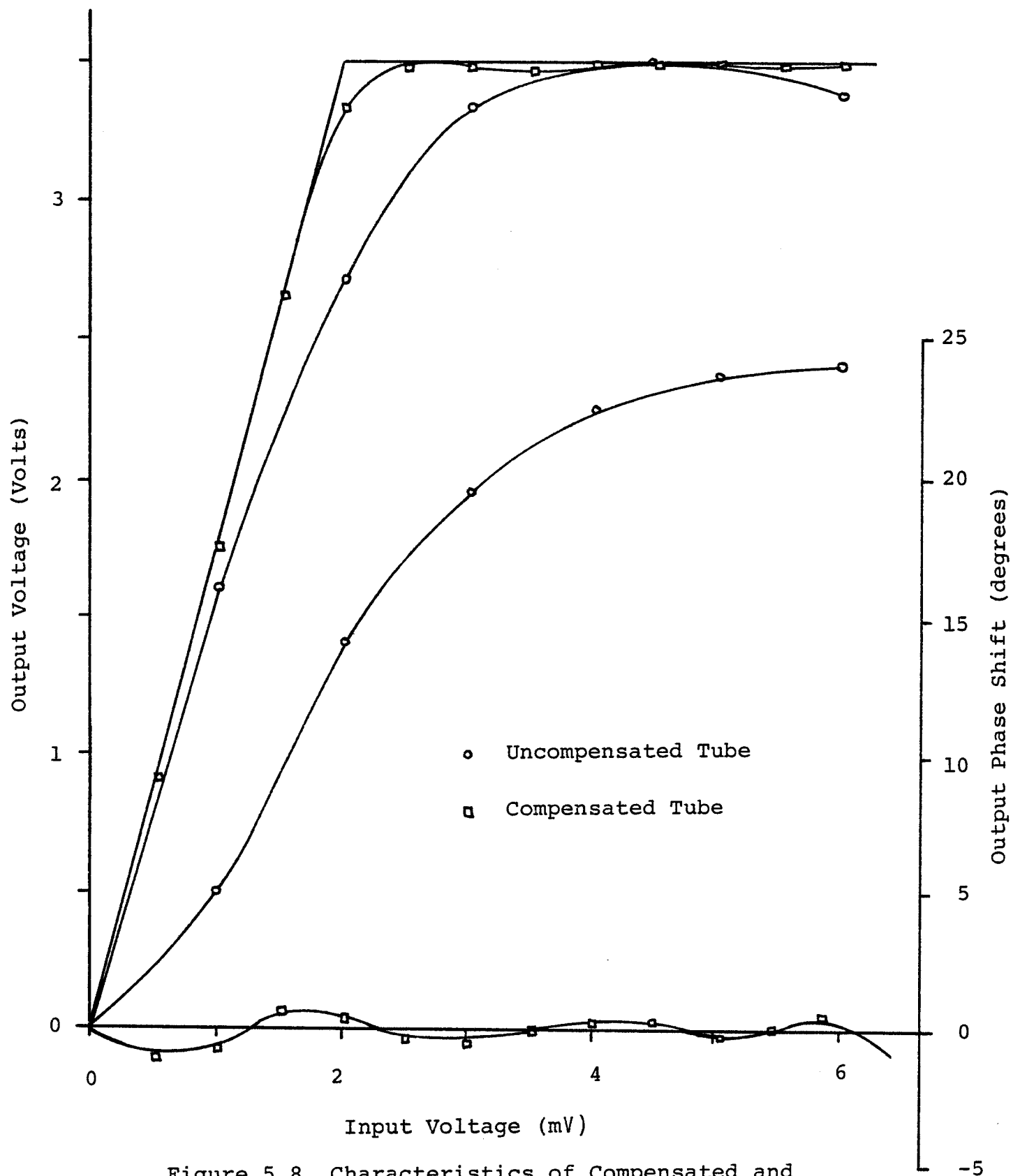
Figure 5.8 Characteristics of Compensated and
Uncompensated Hughes 261H TWT

5.3 System Model

The system model used in the simulation is shown in Fig. 5.9. The binary input data is PR encoded at baseband with full transmitter shaping. After SSB modulation, the signal is passed through the TWT amplifier. Three possible amplifiers were considered: (i) a Hughes 261H TWT operating in the normal mode (ii) a Hughes 261H TWT with no AM/PM and (iii) a Hughes 261H TWT linearized using the predistorter of the previous section. The bandpass transmit and receive filters were assumed ideal rectangular filters which restrict the amplified signal to the same bandwidth as the amplifier input signal. The system considered is an 80 Mbit/s system operating in a 40 MHz bandwidth.

The TWT operating points were selected so that the TWT was saturated for an input envelope of 2. From Fig. 5.8 we see that this corresponds to an input voltage of 4.5mV and an output voltage of 3.5V for (i) and (ii) while for (iii), saturation occurs at a 2.5mV input, with an output voltage of 3.5V. These points were defined as 0 dB input backoff. A 1 dB input backoff would then mean that an input envelope of 2 corresponds to $4.5 \times 10^{-1/20} = 4.24$ mV for (i) and (ii), and $2.5 \times 10^{-1/20} = 2.35$ mV for (iii). A precoded PR4 receiver was used for simplicity and ease of simulation.
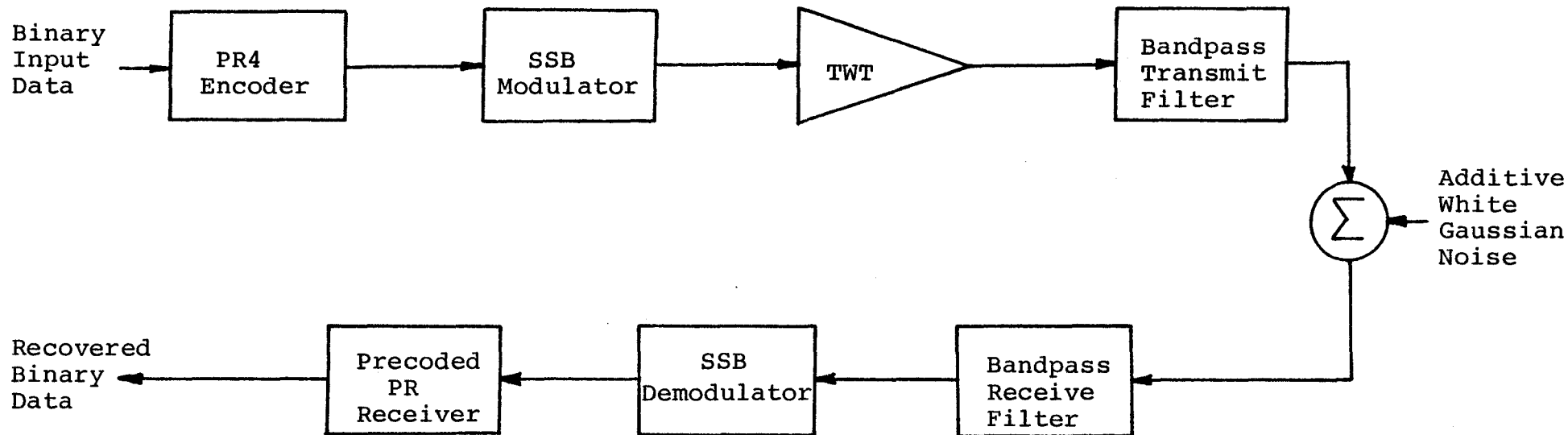
Figure 5.9   System Model for Non-Linear Channel Simulation

5.4 Effects of the Non-Linear Channel

5.4.1 Signal Space Effects

Consider an upper sideband input to the TWT.

$$s(t) = a(t)\cos\omega_c t - \hat{a}(t)\sin\omega_c t$$

$$= R(t)\cos[\omega_c t + \theta(t)] \tag{5.18}$$

where

$$R(t) \triangleq [a^2(t) + \hat{a}^2(t)]^{1/2} \tag{5.19}$$

$$\theta(t) \triangleq \tan^{-1}\frac{\hat{a}(t)}{a(t)} \tag{5.20}$$

with $a(t)$ and $\hat{a}(t)$ defined by (3.9) and (3.10) respectively. The TWT output is then given by (5.2). Neglecting noise and the effects of the transmit and receive filters, demodulation with $\cos\omega_c t$ and $\sin\omega_c t$ gives

$$r_I(t) = f(R)\cos[\theta + g(R)] \tag{5.21}$$

$$r_Q(t) = f(R)\sin[\theta + g(R)] \tag{5.22}$$

as the I and Q outputs of the demodulator.   Expanding the

angles gives

$$r_I(t) = f(R) \cos g(R) \cos\theta - f(R) \sin g(R) \sin\theta \qquad (5.23)$$

$$r_Q(t) = f(R) \cos g(R) \sin\theta + f(R) \sin g(R) \cos\theta \qquad (5.24)$$

From (5.18), (5.19) and (5.20), the following identities may be deduced:

$$\cos\theta \triangleq a(t)/R(t) \qquad (5.25)$$

$$\sin\theta \triangleq \hat{a}(t)/R(t) \qquad (5.26)$$

$$\cos g(R) \triangleq Z_p(R)/f(R) \qquad (5.27)$$

$$\sin g(R) \triangleq Z_q(R)/f(R) \qquad (5.28)$$

Substituting these results into (5.23) and (5.24) gives as the demodulator outputs

$$r_I(t) = Z_p(R) \frac{a(t)}{R(t)} - Z_q(R) \frac{\hat{a}(t)}{R(t)} \qquad (5.29)$$

$$r_Q(t) = Z_q(R) \frac{a(t)}{R(t)} + Z_p(R) \frac{\hat{a}(t)}{R(t)} \qquad (5.30)$$

Sampling these outputs at t = kT and plotting the locus of $(r_I(kT), r_Q(kT))$, we obtain the signal space after amplification. The signal space corresponding to the TWT in the normal mode of operation at 0 dB backoff is shown in Fig. 5.10. The TWT output voltage was normalized so that there was no average power gain through the device. When compared with the undistorted signal space of Fig. 3.9, we see that the AM/AM warps the lines while the AM/PM rotates the space in the counter-clockwise direction with respect to the original axes.

The demodulator outputs for a TWT with no AM/PM can be obtained by setting $g(R) = 0$ in (5.21) and (5.22) to obtain

$$r_I(t) = f(R)\, a(t)/R(t) \qquad (5.31)$$

$$r_Q(t) = f(R)\, \hat{a}(t)/R(t) \qquad (5.32)$$

The signal space for no AM/PM at 0 dB input backoff is shown in Fig. 5.11. Since there is no AM/PM, there is no rotation, only warping of the outer lines.

Similarly, for the linearized TWT we have

Figure 5.10   Output Signal Space for TWT With
AM/AM and AM/PM at 0 dB Backoff
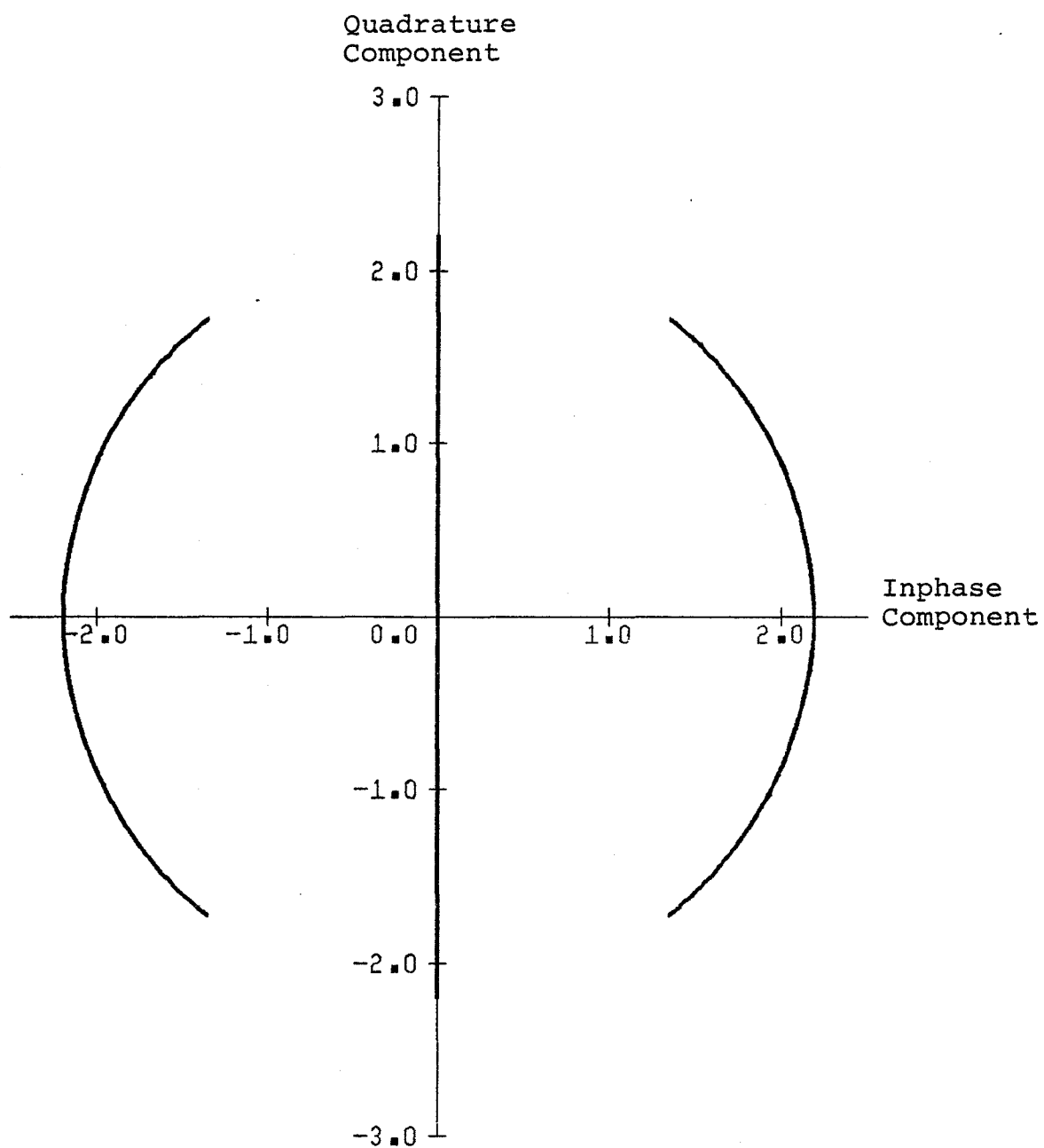
Figure 5.11 Output Signal Space for TWT With
AM/AM Only at 0 dB Backoff
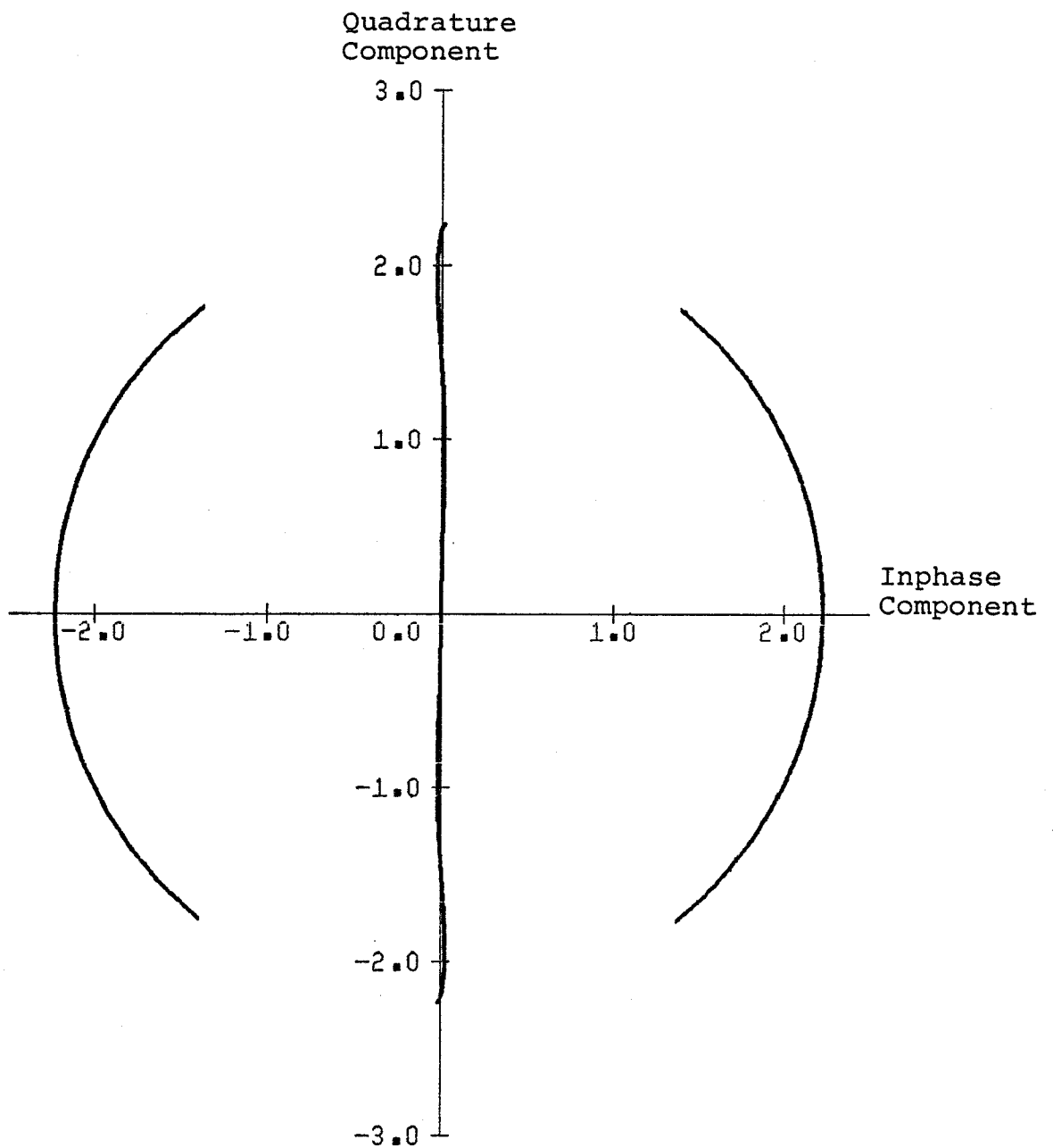
Figure 5.12   Output Signal Space for Linearized TWT
at 0 dB Backoff

$$r_I(t) = Z_p\{G'(R)\}\frac{a(t)}{R(t)} - Z_q\{G'(R)\}\frac{\hat{a}(t)}{R(t)} \qquad (5.33)$$

$$r_Q(t) = Z_q\{G'(R)\}\frac{a(t)}{R(t)} + Z_p\{G'(R)\}\frac{\hat{a}(t)}{R(t)} \qquad (5.34)$$

The corresponding signal space at 0 dB input backoff is shown in Fig. 5.12. Since the AM/PM effect is greatly reduced, there is only a slight rotation of the space, and a slight bending of the central portion corresponding to $a_k = 0$. The warping of the outer portions occurs despite the linearization of the amplifier because it is operating in the saturation region ($R > R_i$ in (5.14)).

In simulating the system, certain assumptions concerning the carrier tracking mechanism in the receiver must be made. Clearly if the original reference axes in Fig. 5.10 were to be used as the basis for decisions in the I channel, an unacceptable amount of degradation would occur due to the rotation of the space. This is not a realistic model of what the receiver would do given the signal of Fig. 5.10. On the other hand, we could assume that the receiver could instantaneously track any variations in phase due to AM/PM conversion. The receiver would then behave exactly as the system with no AM/PM present. However, this assumption is not realistic either; the real answer lies between these
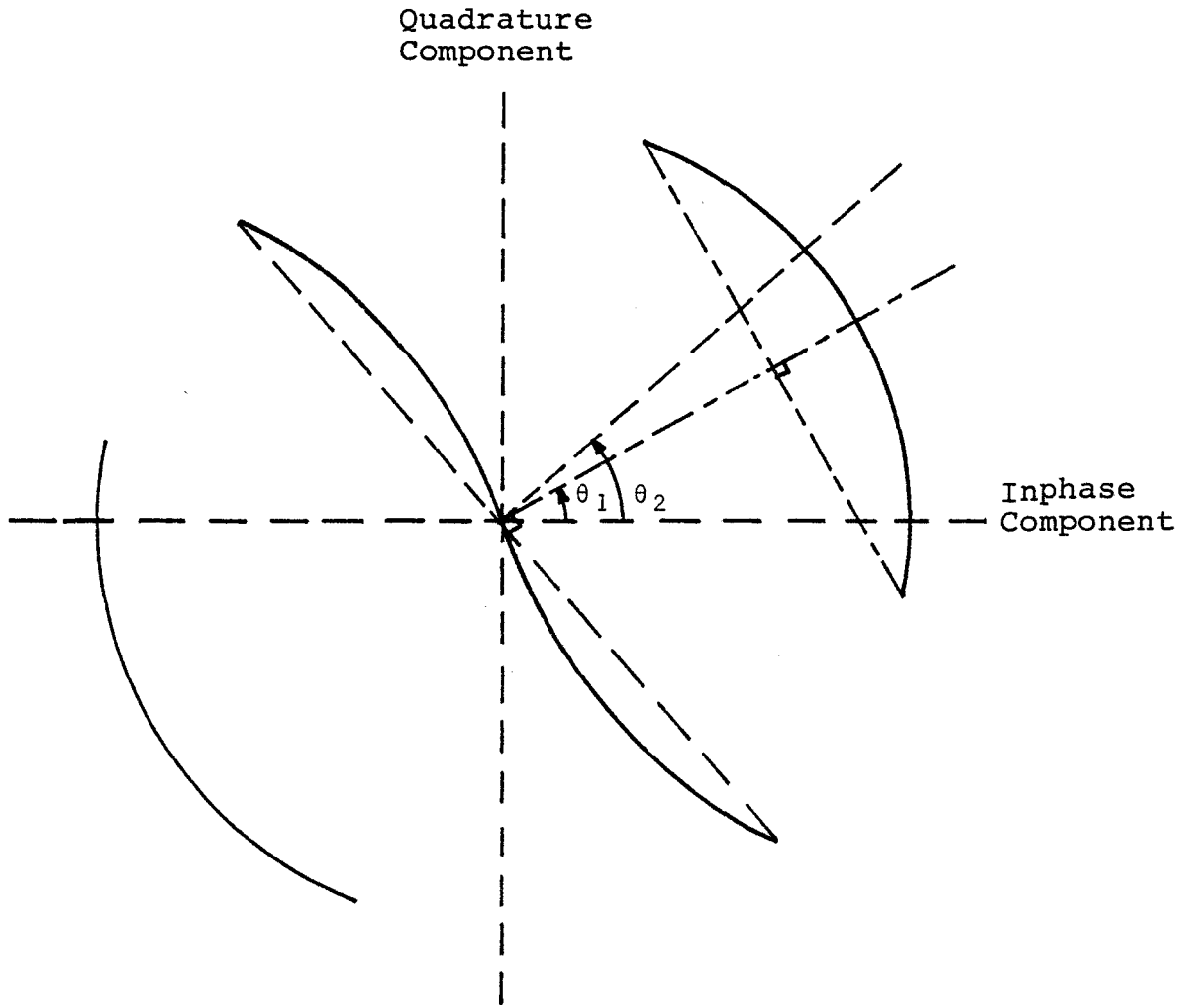
Figure 5.13 Phase Compensation of Signal Space

two extremes. As a compromise it was assumed that the carrier tracking loop would track around a certain "mean" value of carrier phase $\bar{\theta}$, and then jitter about this value due to the AM/PM. This corresponds to defining a new reference axis at an angle $\bar{\theta}$ to the old axis as shown in Fig. 5.13. The carrier loop would attempt to select the value of $\bar{\theta}$ in such a way that the signal space would become symmetric about this axis. In the figure, rotating the space by θ_1, would make the portion of the signal space corresponding to ±2 symmetric about the new axis while rotating by θ_2 would make the portion corresponding to $a_k = 0$ symmetric. In general, $\theta_1 \neq \theta_2$ so their mean value $\bar{\theta} = (\theta_1 + \theta_2)/2$ was selected to define the new reference axis.

Since there is no signal space rotation when there is no AM/PM present, such an approach was not necessary for the TWT with no AM/PM. The new reference axis was used with the linearized TWT as well although its effect is minimal due to the small amount of rotation.

5.4.2 Eye Diagram Effects

The simulation program was designed to generate eye diagrams if desired, several of which are shown in Figs. 5.14-5.19. The bandpass filters in Fig. 5.9 were excluded
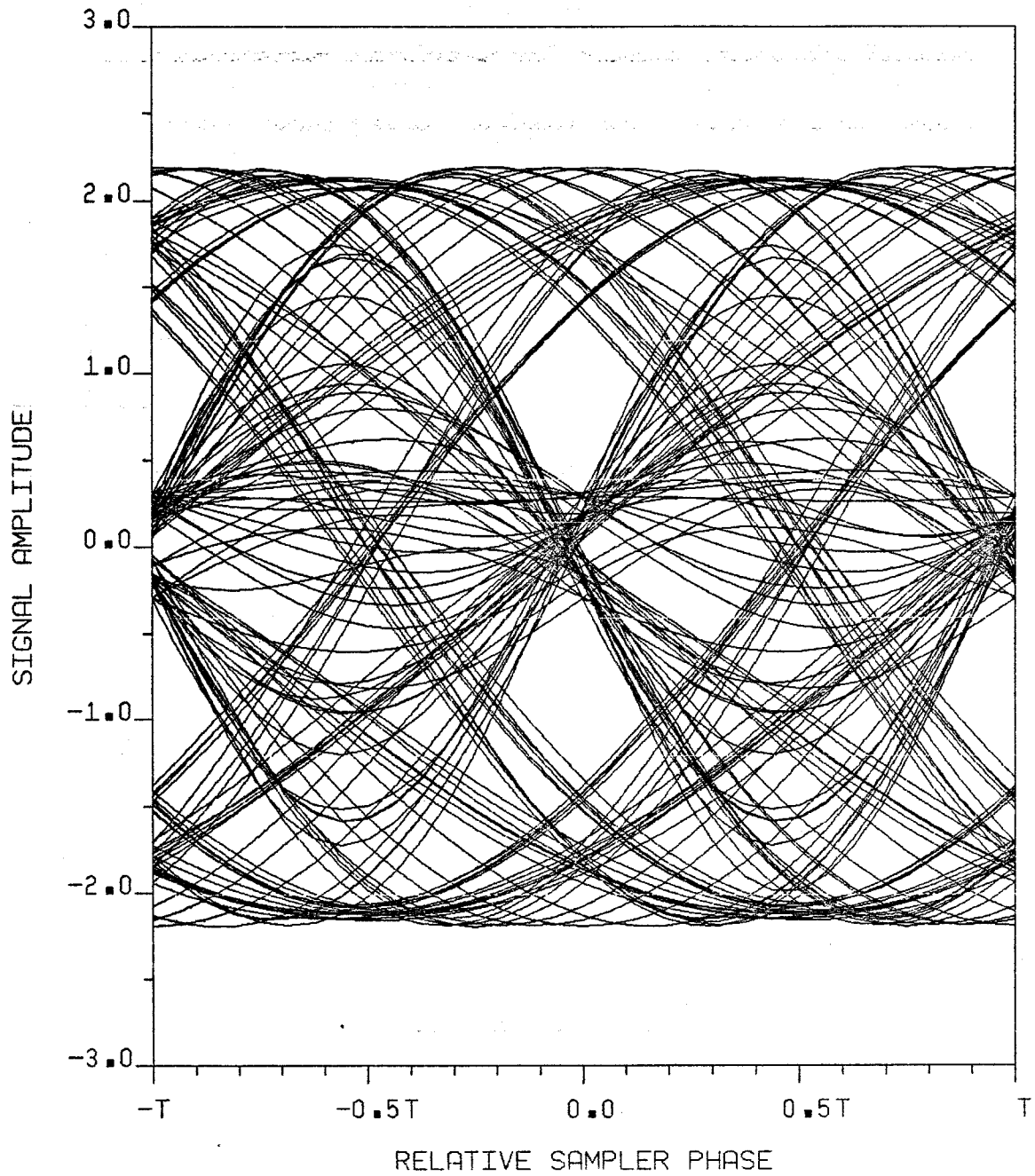
Figure 5.14   Output Eye Diagram for TWT With
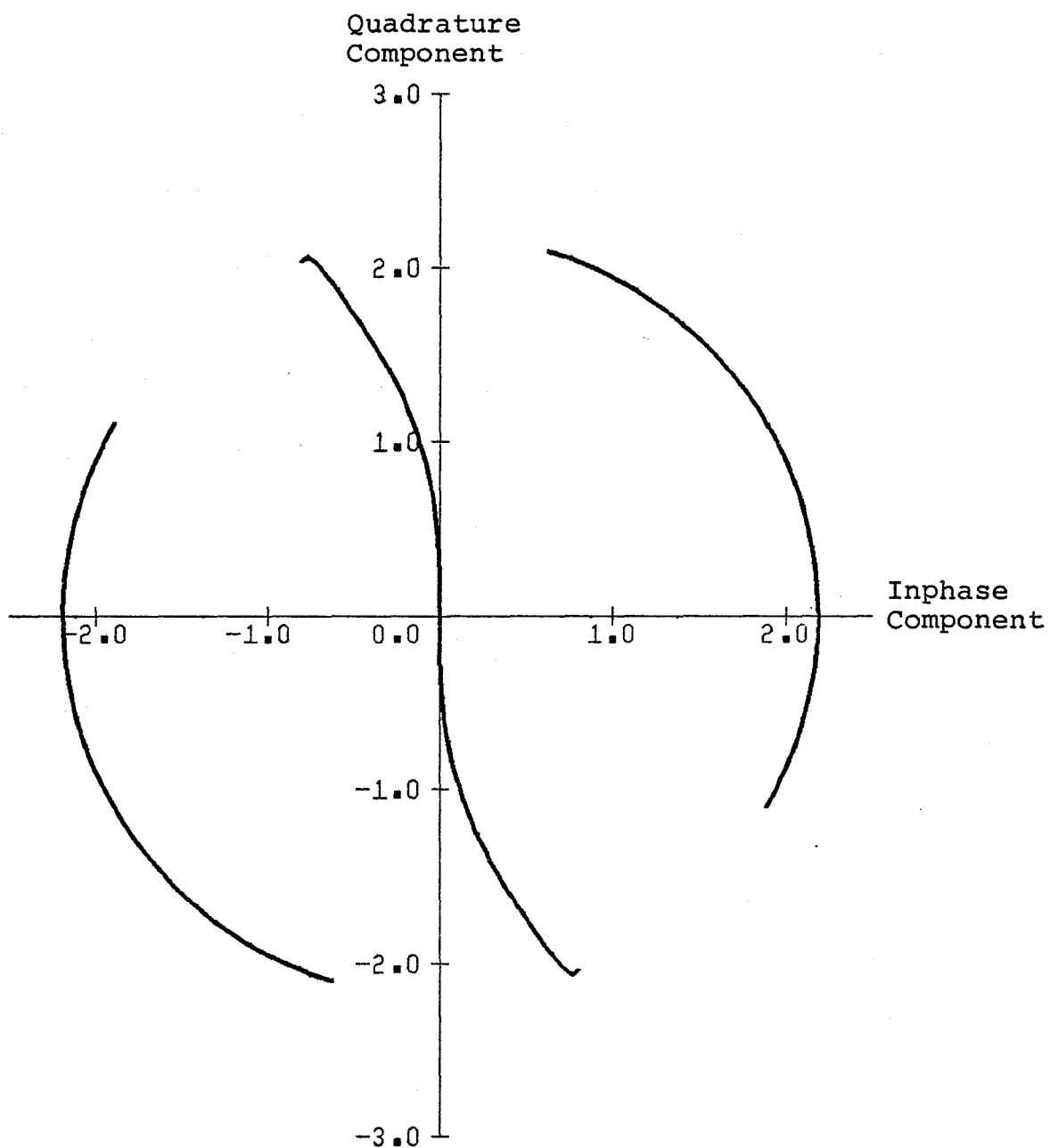AM/AM and AM/PM at 0 dB Backoff
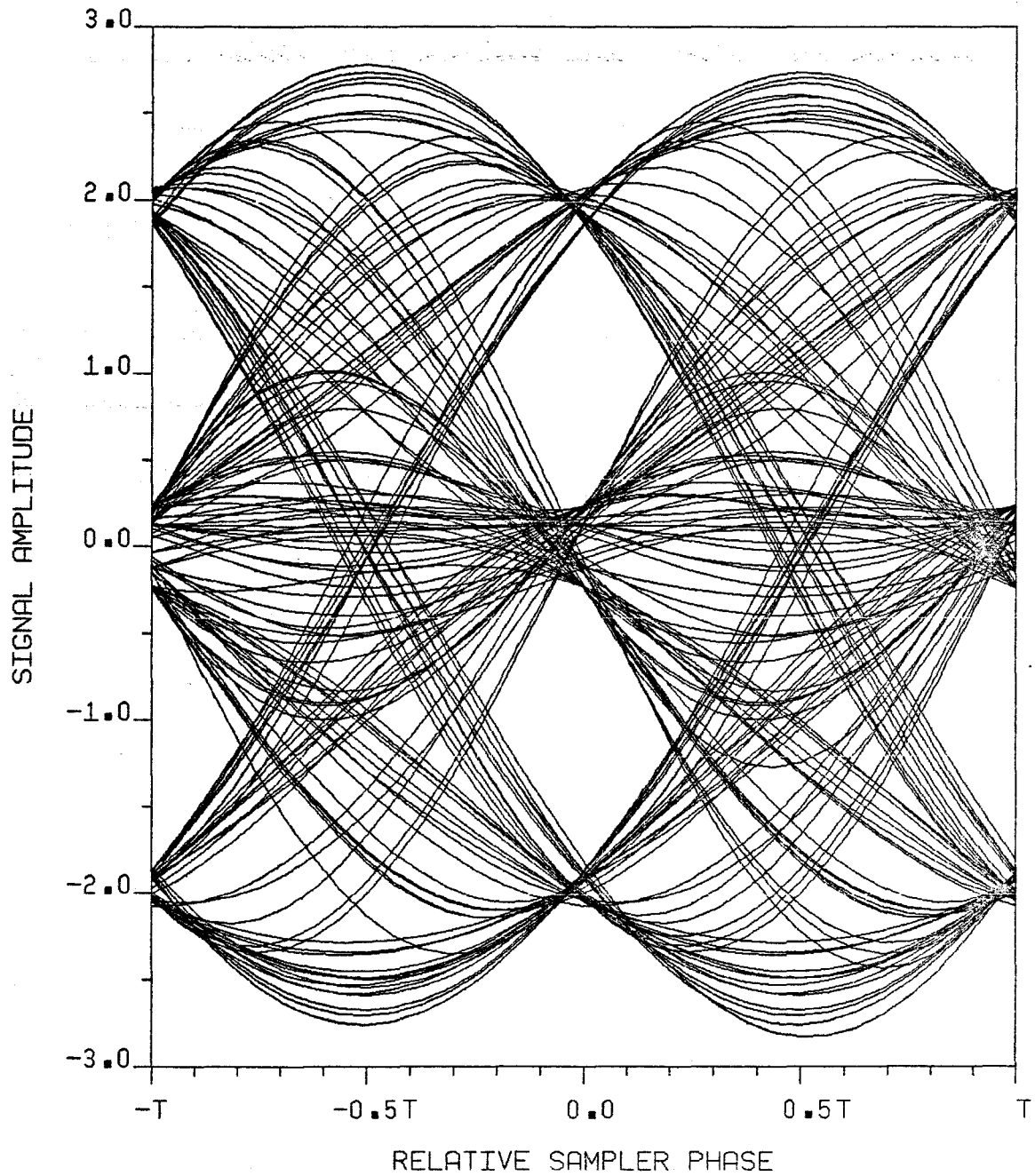
Figure 5.15   Output Eye Diagram for TWT With
AM/AM and AM/PM at 10 dB Backoff
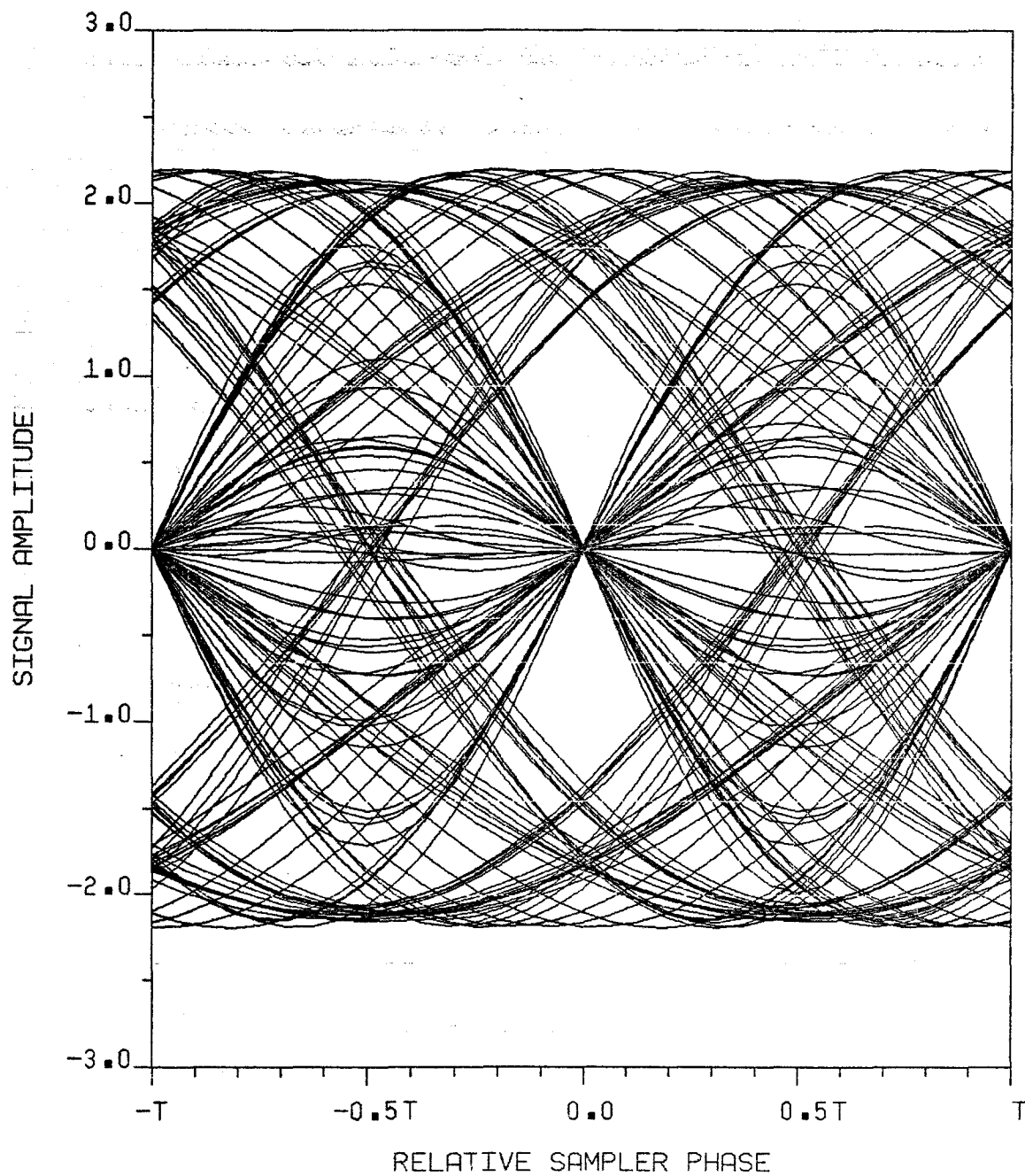
Figure 5.16   Output Eye Diagram for TWT With
AM/AM Only at 0 dB Backoff
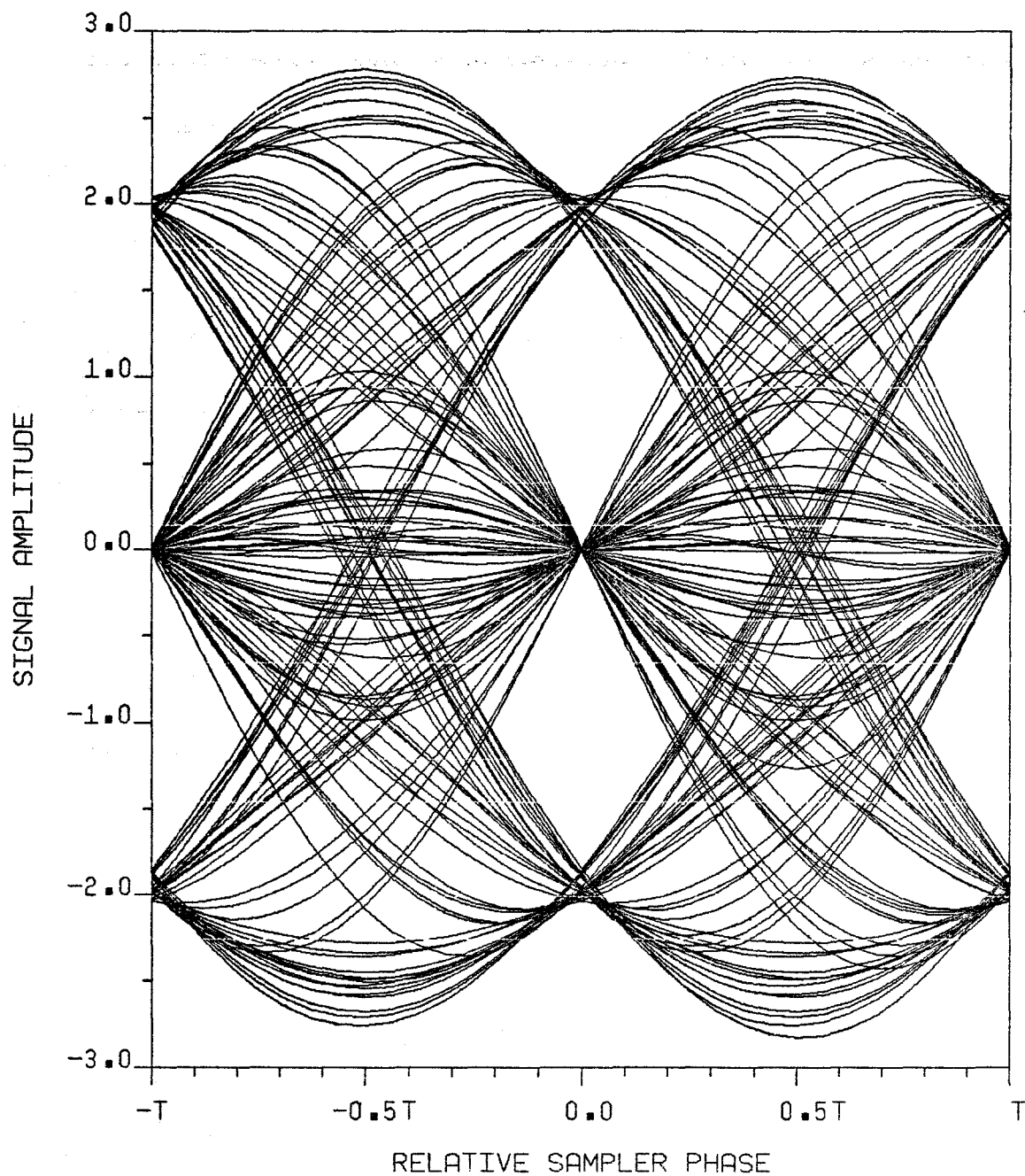
Figure 5.17 Output Eye Diagram for TWT With
AM/AM Only at 10 dB Backoff
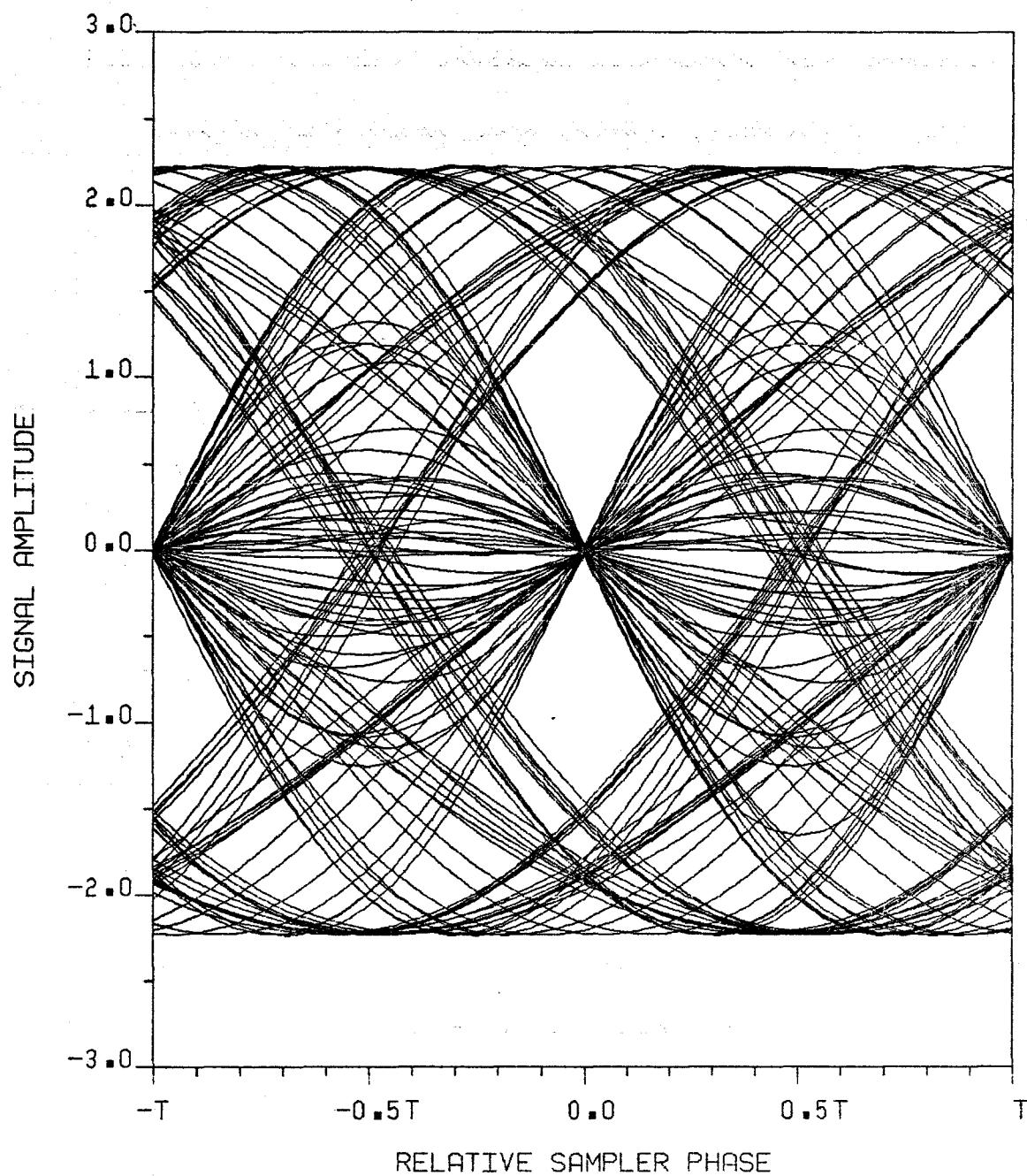
Figure 5.18 Output Eye Diagram for Linearized TWT
at 0 dB Backoff
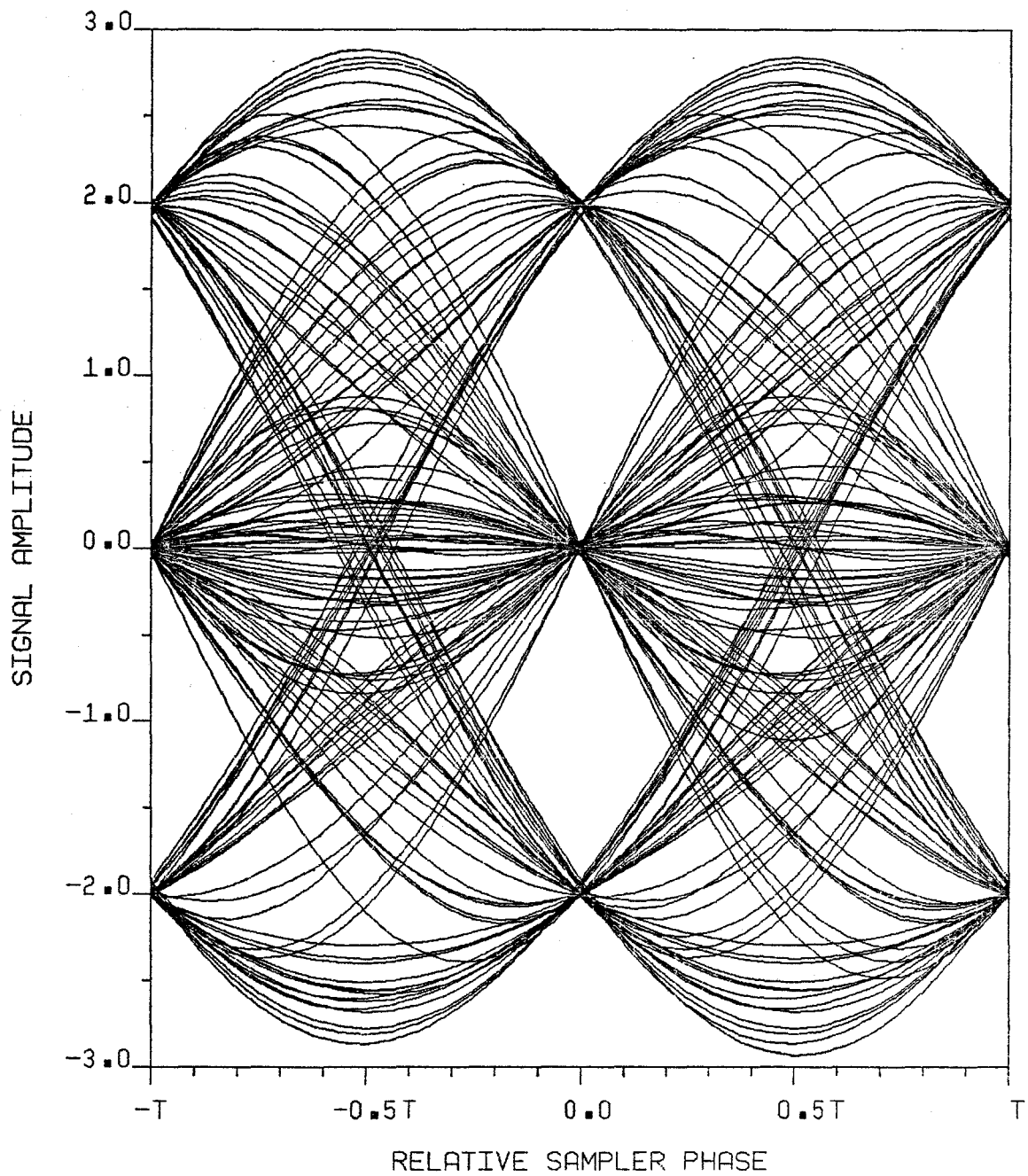
Figure 5.19   Output Eye Diagram for Linearized TWT
at 6 dB Backoff

in the eye diagram generation. The first two correspond to a TWT with both AM/AM and AM/PM at 0 and 10 dB backoff, while Figs. 5.16 and 5.17 are the eye diagrams with no AM/PM present at the same input backoffs. The signal space was rotated as described in the previous section. The saturation effect of the TWT is very clear while the AM/PM has two major effects on the eye: firstly, the zero crossings no longer occur at a single point, and secondly, the eye is warped towards the direction of increasing time. The latter phenomenon is consistent with the AM/PM effect as large input signals take longer to pass through the amplifier. With no AM/PM present, the zero crossings are well defined and the eye is symmetric about the nominal sampling point $t = kT$. In both cases, it can be seen that the eye opening is more or less maximized at the nominal sampling instant. Also, at 10 dB input backoff, the two signals resemble the undistorted eye of Fig. 3.7.

Figs. 5.18 and 5.19 are the eye diagrams of the linearized TWT at 0 and 6 dB backoff. The latter eye is practically indistinguishable from the eye of Fig. 3.7. The predistorter improves the symmetry of the eye as well as the tightness of the zero-crossings.

We shall now derive expressions for the eye openings

for the three cases with no filtering after amplification.

With the signal space rotated as in the previous section, the inphase demodulator output is

$$r_I(t) = f(R)\cos[\theta + g(R) - \bar{\theta}] \qquad (5.35)$$

where $\bar{\theta}$ is the angle of rotation. Expanding the cosine term in (5.35) and using the definitions in (5.25) and (5.28), we obtain

$$r_I(t) = \frac{a(t)}{R(t)}[Z_p(R)\cos\bar{\theta} + Z_q(R)\sin\bar{\theta}]$$

$$- \frac{\hat{a}(t)}{R(t)}[Z_p(R)\sin\bar{\theta} + Z_q(R)\cos\bar{\theta}] \qquad (5.36)$$

To determine the vertical eye opening (VEO), we must determine the minimum value of $r_I(kT)$ when $a(kT) = 2$ is transmitted, and the maximum value of $r_I(kT)$ when $a(kT) = 0$ is transmitted. Denoting these values as λ_1 and λ_2, we have

$$VEO = \lambda_1 - \lambda_2$$

where

$$\lambda_1 \overset{\Delta}{=} \min_{\hat{a}_k}\left\{ \frac{2}{\sqrt{4+\hat{a}_k^2}}\left(Z_p(\sqrt{4+\hat{a}_k^2})\cos\bar{\theta} + Z_q(\sqrt{4+\hat{a}_k^2})\sin\bar{\theta}\right) \right.$$

$$\left. - \frac{\hat{a}_k}{\sqrt{4+\hat{a}_k^2}}\left(Z_p(\sqrt{4+\hat{a}_k^2})\sin\bar{\theta} + Z_q(\sqrt{4+\hat{a}_k^2})\cos\bar{\theta}\right) \right\}$$

$$0 \leq \hat{a}_k \leq 8/\pi \qquad (5.37)$$

and

$$\lambda_2 \triangleq \max_{\hat{a}_k} \{Z_p(\hat{a}_k)\sin\bar{\theta} + Z_q(\hat{a}_k)\cos\bar{\theta}\}$$

$$0 \leqslant \hat{a}_k \leqslant 8/\pi \qquad (5.38)$$

The case for no AM/PM is somewhat simpler. Setting $g(R) = 0$ in (5.21) gives

$$r_I(t) = f(R)\cos\theta$$
$$= (z_p^2(R) + z_q^2(R))^{1/2} \frac{a(t)}{(a^2(t)+\hat{a}^2(t))^{1/2}} \qquad (5.39)$$

For $a(kT) = 0$, $r_I(kT) = 0$, so that the vertical eye opening is

$$VEO = \min_{\hat{a}_k} \{(z_p^2(\sqrt{4+\hat{a}_k^2}) + z_q^2(\sqrt{4+\hat{a}_k^2}))^{1/2}$$

$$\cdot \frac{2}{\sqrt{4+\hat{a}_k^2}}\} \qquad 0 \leqslant \hat{a}_k \leqslant 8/\pi \qquad (5.40)$$

For the linearized TWT, we have

$$VEO = \mu_1 - \mu_2$$

where

$$\mu_1 \triangleq \min_{\hat{a}_k} \{ \frac{2}{\sqrt{4+\hat{a}_k^2}} (z_p\{ G'(\sqrt{4+\hat{a}_k^2}) \}\cos\bar{\theta}$$

$$+ z_q\{ G'(\sqrt{4+\hat{a}_k^2}) \}\sin\bar{\theta}) \qquad (5.41)$$

$$- \frac{\hat{a}_k}{\sqrt{4+\hat{a}_k^2}} (z_p\{ G'(\sqrt{4+\hat{a}_k^2}) \}\sin\bar{\theta}$$

$$+ z_q\{ G'(\sqrt{4+a_k^2}) \}\cos\bar{\theta})\}, 0 \leq a_k \leq 8/\pi$$

and

$$\mu_2 \triangleq \max_{\hat{a}_k} \{ z_p\{ G'(\hat{a}_k) \}\sin\bar{\theta} + z_q\{ G'(\hat{a}_k) \}\cos\bar{\theta} \}$$

$$0 \leq \hat{a}_k \leq 8/\pi \qquad (5.42)$$

Here $\bar{\theta}$ is the rotation angle of the predistorter - TWT combination and is usually small.

The solutions to these equations for various backoffs were obtained using a searching type program. The VEO's obtained were scaled so that there was no net average power gain through the amplifier. The output eye opening as a function of backoff is shown in Fig. 5.20. As expected, the TWT with both AM/AM and AM/PM distorts the signal the most. Removal of the AM/AM effect improves the eye opening somewhat, while the linearized TWT distorts the signal the least. Even at large backoffs (> 10dB), the AM/PM effect in the normal mode TWT is still present as the eye is still partially closed while for the other two cases, the eye is effectively open completely.
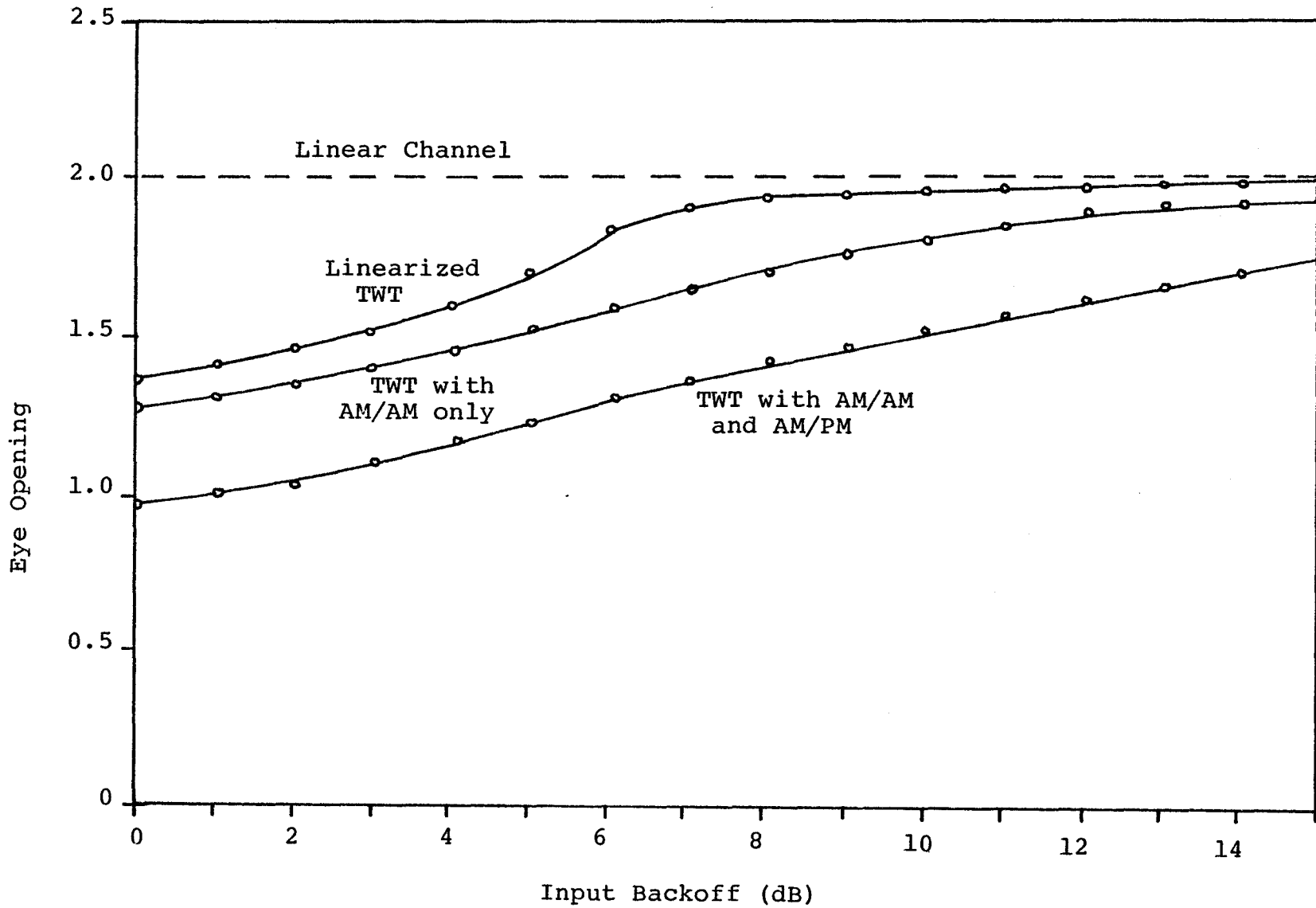
Figure 5.20   TWT Output Eye Opening vs. Input Backoff

5.4.3 Bit Error Rate Effects

The probability of error curves obtained for the three cases under consideration for various input backoffs are shown in Figs. 5.21-5.23. These curves have been corrected to remove a 0.8 dB simulation error. Fig. 5.24 shows the SNR degradation over a linear channel at an error rate of 10^{-5} as a function of backoff. At 0 dB backoff, a very large amount of degradation is present, particularly for the TWT operating in the normal node. Removal of the AM/PM results in about a 2 dB improvement in SNR while linearization improves performance by about 6 dB at 0 dB backoff. For a 1 dB SNR degradation the TWT with both AM/AM and AM/PM requires a backoff greater than 11 dB, while the TWT with AM/AM only requires a 7 dB input backoff. The linearized TWT requires an input backoff of only 4 dB for a 1 dB SNR degradation. At large input backoffs (> 8 dB), the AM/PM effect still has a significant effect on the SNR degradation, which is consistent with the eye opening results obtained earlier. The reason for this effect is that at high backoffs, the output phase shift changes more rapidly with input amplitude in Fig. 5.8. Thus there is a non-negligible amount of jitter about the angle $\bar{\theta}$ due to the AM/PM phenomenon, even though the amplitude characteristic is more or less linear.
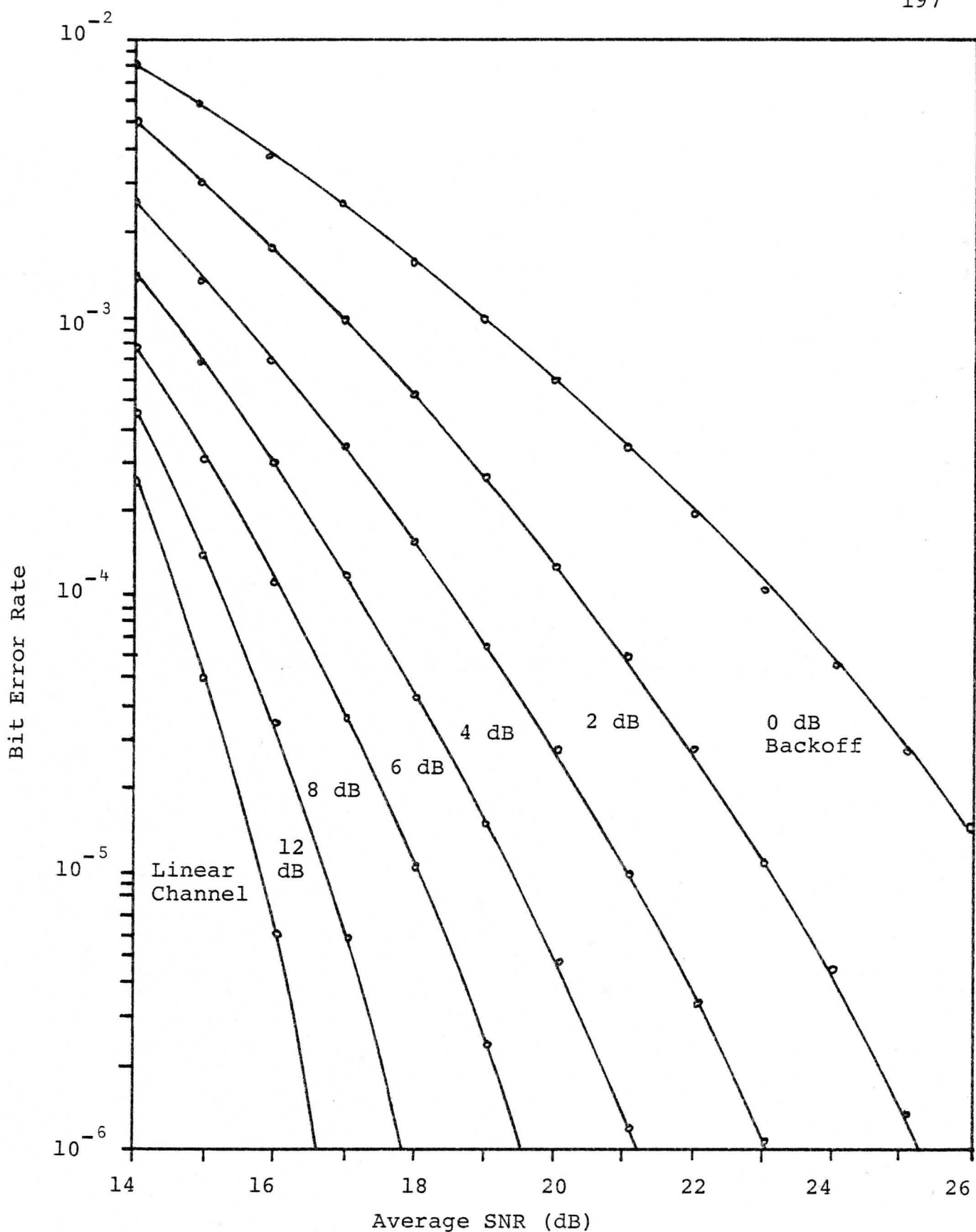
Figure 5.21   Simulated Bit Error Rate Performance of
TWT with AM/AM and AM/PM
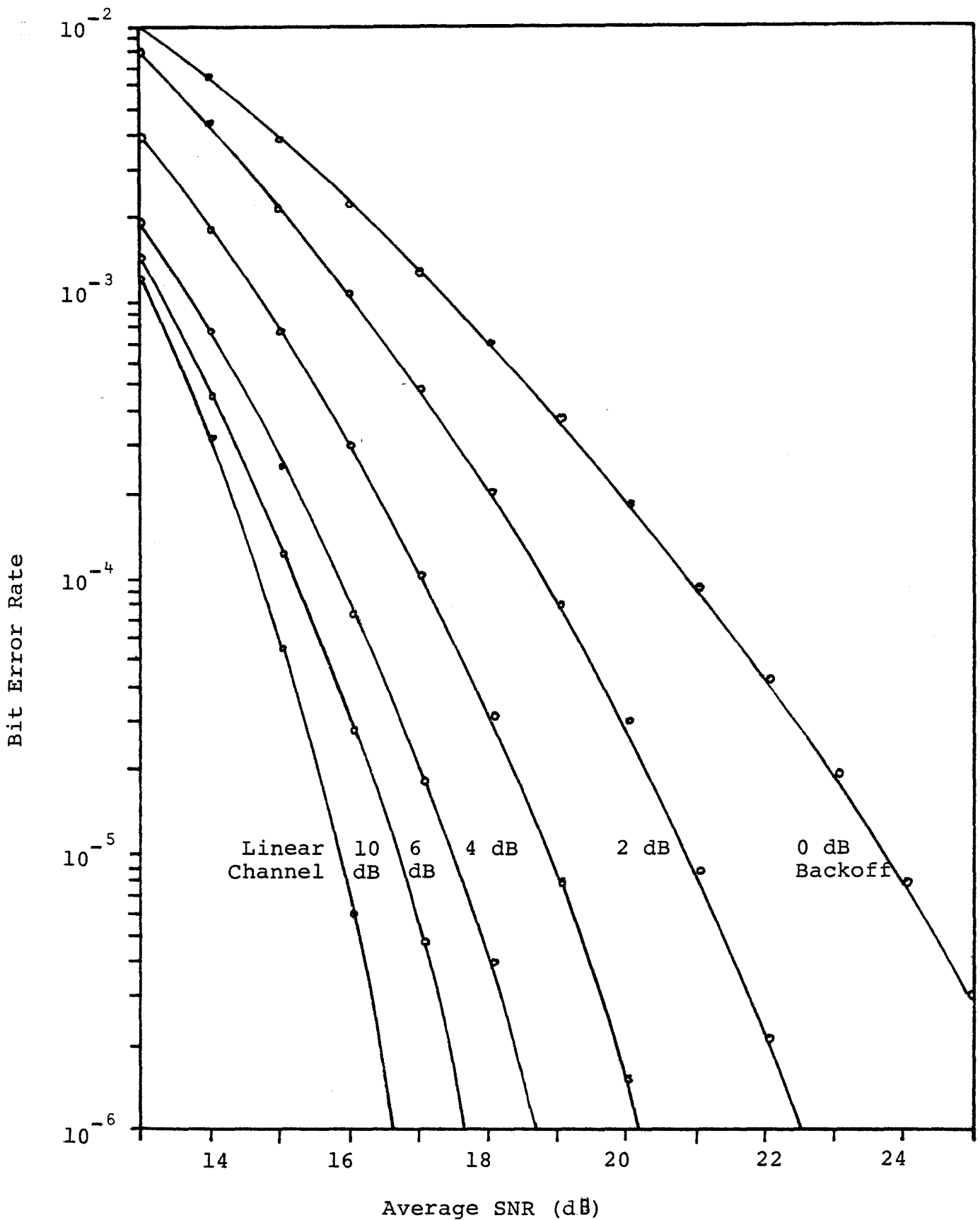
Figure 5.22   Simulated Bit Error Rate Performance of
TWT with AM/AM only
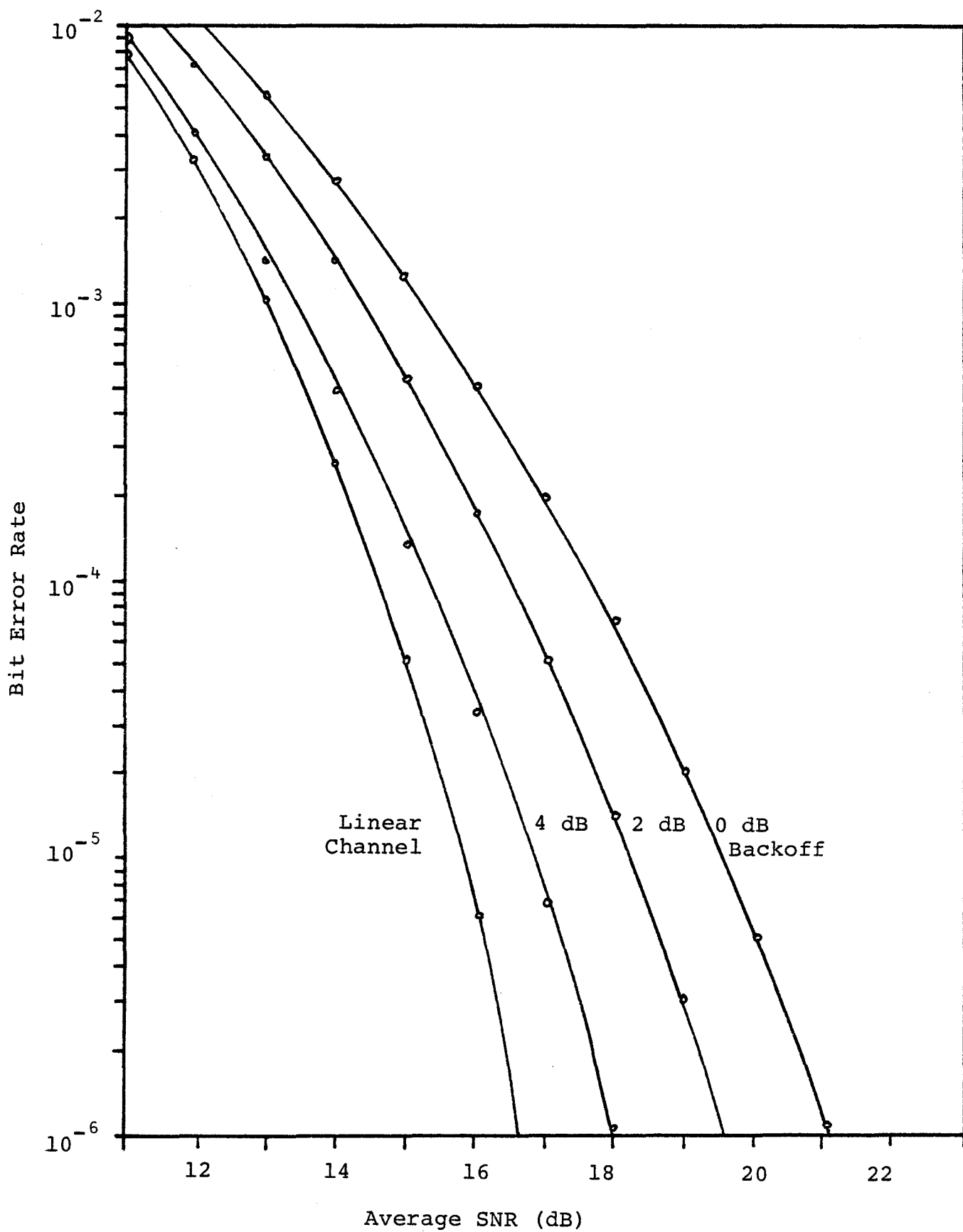
Figure 5.23   Simulated Bit Error Rate Performance of
Linearized TWT
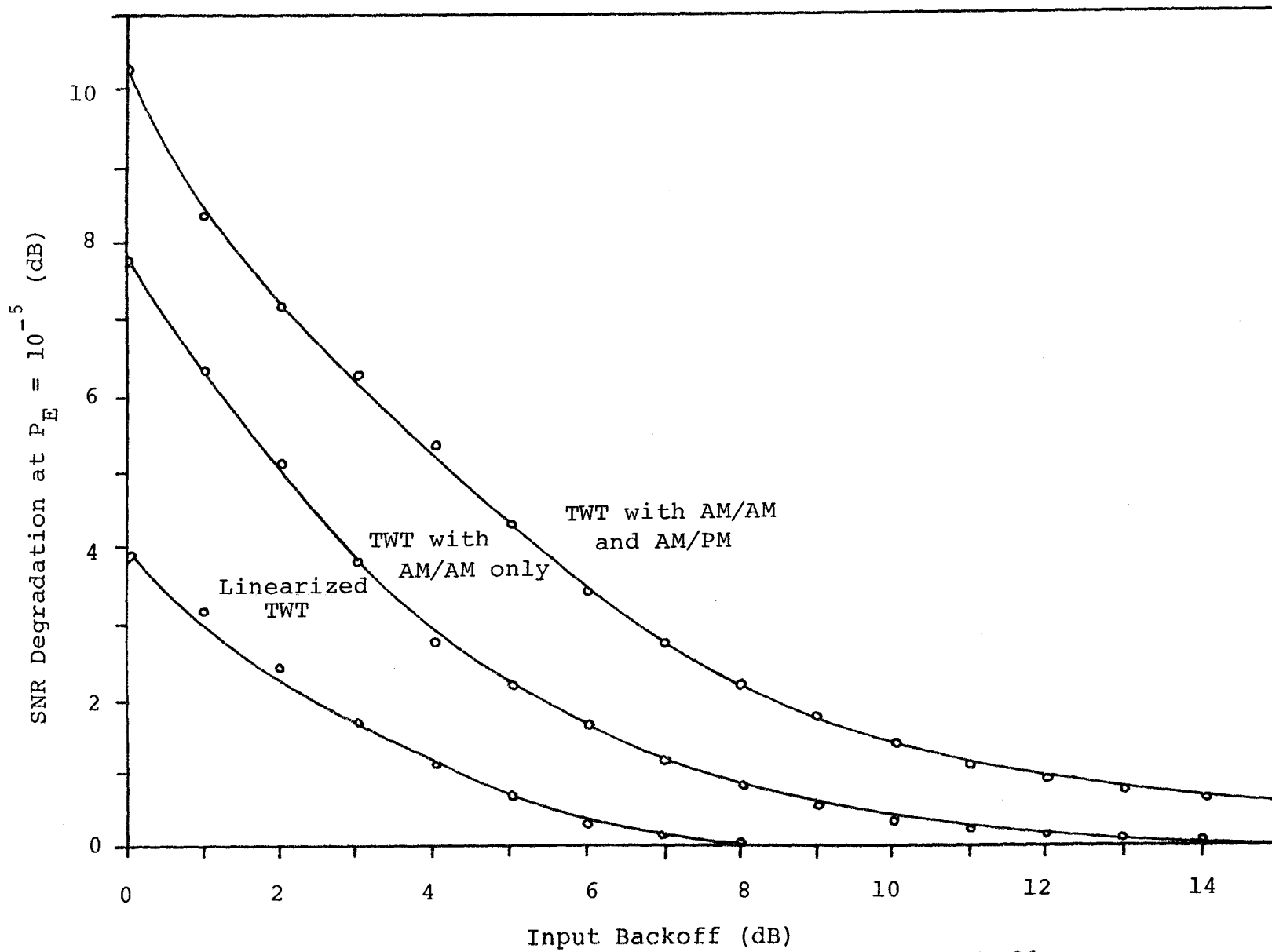
Figure 5.24   SNR Degradation vs. TWT Input Backoff

5.5 System Gain Considerations

An important parameter in digital radio system design is the system gain which is defined as

$$G_{SYS} = P_{OUT} - P_T \qquad (5.43)$$

where P_{OUT} is the TWT output power in dBm and P_T is the required received RF signal achieve a certain bit error rate. We shall nowproceed to calculate the system gain for the 80 Mb/s, 40 MHz system, making use of the results obtained in the previous section.

The thermal noise power is a bandwidth of B Hz is given by [12]

$$P_N = kTB \qquad (5.44)$$

where k = Boltzmann's constant = 1.3803 x 10^{-23} J/°K and T is the absolute temperature of the noise source. The noise power in a 1 Hz band is then kT watts which has the value -174.0 dBm at 290°K (17°C). Our receiver bandwidth is 40 MHz = 76.02 dBHz. Thus the receiver input noise power is

$$P_N = -174.0 \text{ dBm} + 76.02$$

$$= -97.98 \text{ dBm} \qquad (5.45)$$

Assuming a front-end amplifier noise figure of 8 dB, the noise power at the decision device input is

$$P'_N = -97.98 \text{ dBm} + 8 \text{ dB}$$

$$= -89.98 \text{ dBm} \qquad (5.46)$$

From Fig. 3.13 we see that an SNR of 15.8 dB is required for a BER of 10^{-5} over a linear channel, so that the required received signal level for a BER of 10^{-5} is

$$P_T = -89.98 \text{ dBm} + 15.8 \text{ dB}$$

$$= -74.18 \text{ dBm} \qquad (5.47)$$

Making use of the SNR degradation curves Fig. 5.24, we may compute the received signal level for a BER of 10^{-5} as a function of input backoff. Letting P_D denote the SNR degradation at a particular value of input backoff, we have

$$P_T = -74.18 + P_D \qquad (5.48)$$

Thus the system gain can be computed according to the relation

$$G_{SYS} = P_{OUT} - P_D + 74.18 \qquad (5.49)$$

where P_{OUT} is the TWT output power in dBm at a particular value of input backoff. This parameter can be obtained from Fig. 5.25 which shows the TWT output power for various values of input backoff as obtained by simulation. Since changing the phase angle does not change the output power, both the TWT with AM/AM and AM/PM and the TWT with AM/AM only are described by the same curve in Fig. 5.25. From this curve, the TWT output backoff can be related to the TWT input backoff.

The system gain plotted as a function of input backoff for the three cases under consideration is shown in Fig. 5.26. Also shown is the extrapolated behaviour of a linear channel to provide a reference for comparison. The TWT operating in its normal mode has the lowest overall gain, with a maximum value of 106.6 dB at 7 dB input backoff. The maximum system gain for the TWT with no AM/PM has a maximum at 6 dB input backoff of 108.3 dB. The linearized TWT has a maximum gain of 108.5 dB at 4 dB input backoff. The presence of these maxima is a consequence of
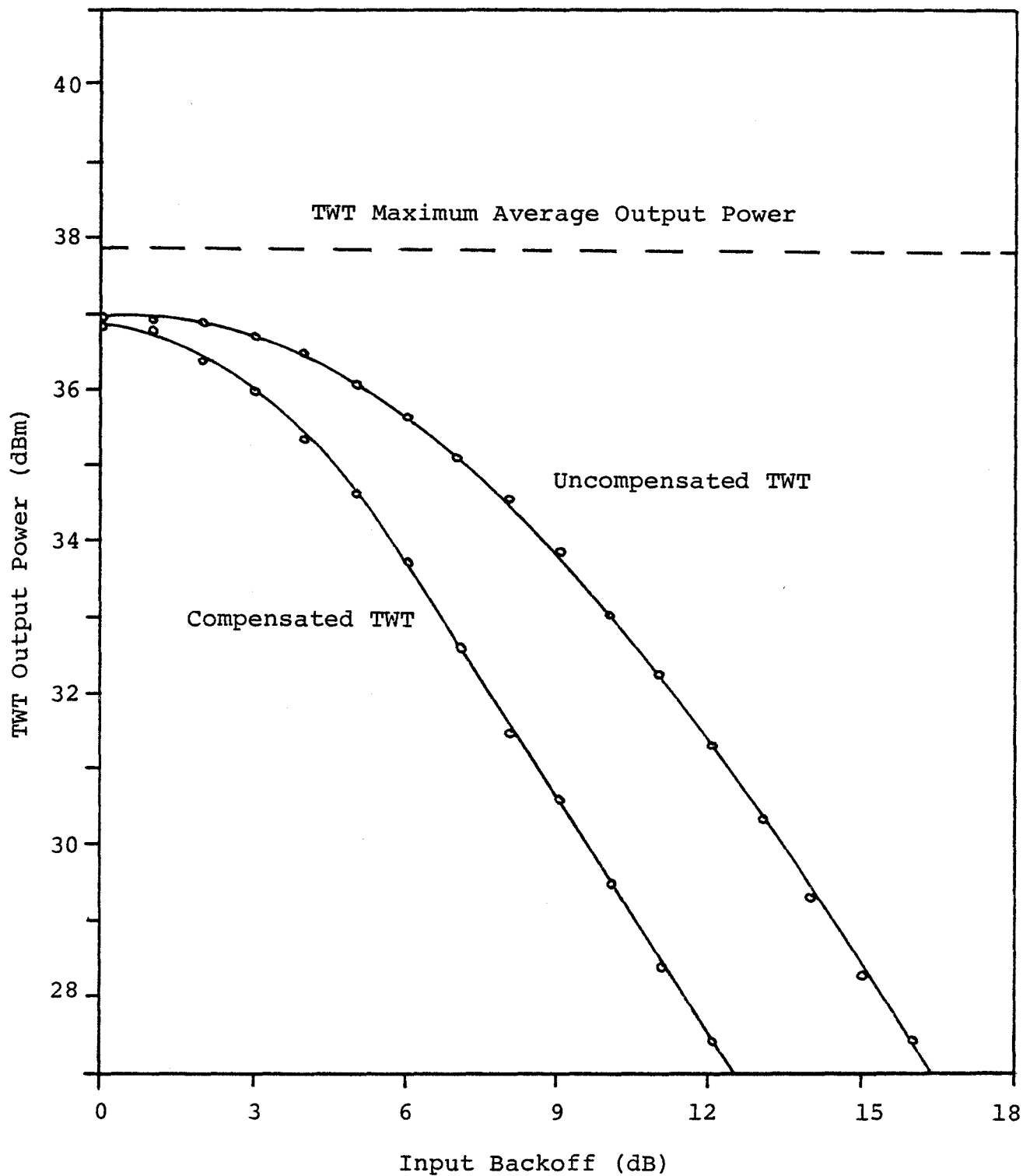
Figure 5.25 TWT Output Power vs. Input Backoff
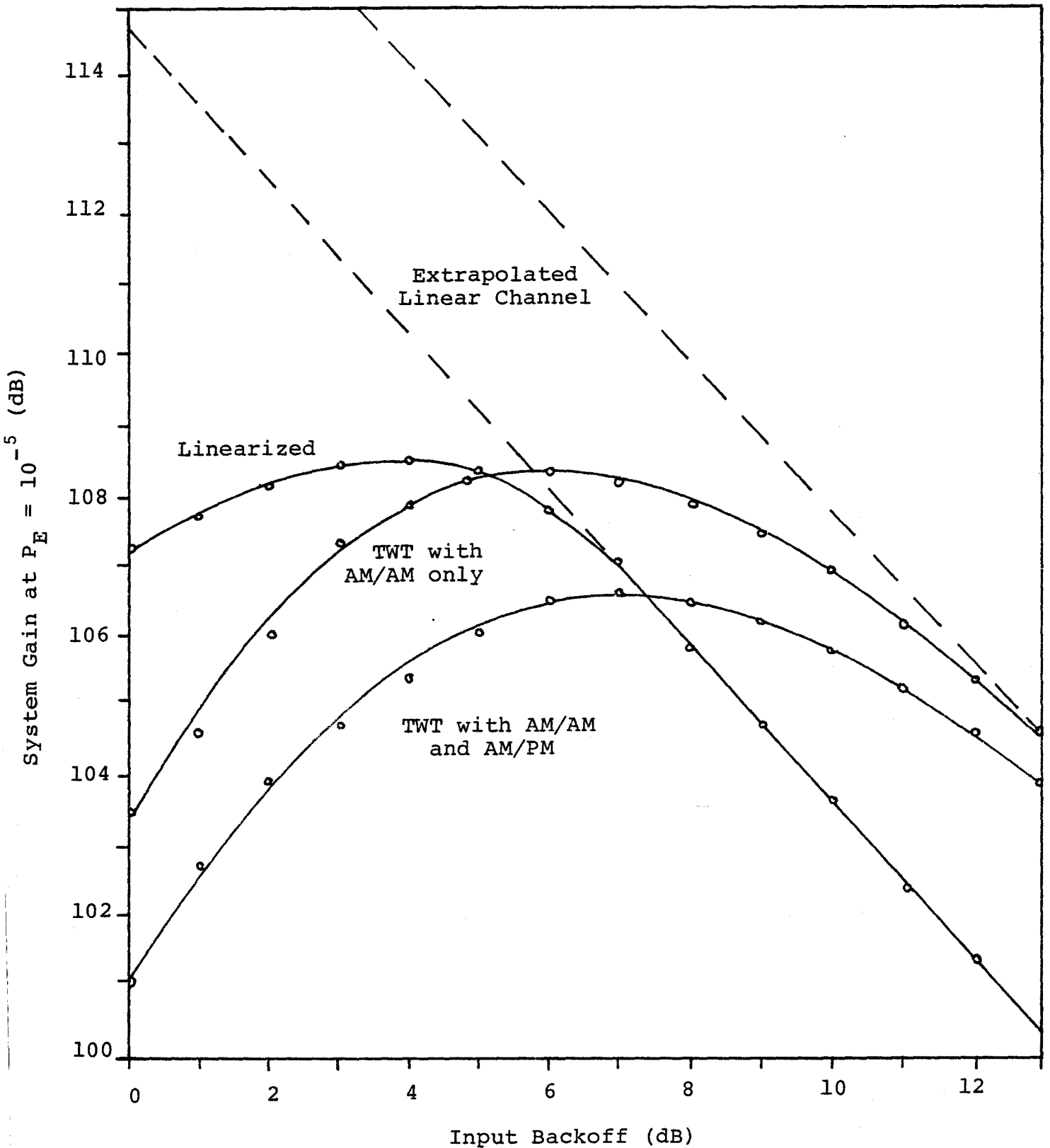
Figure 5.26   System Gain at $P_E = 10^{-5}$ vs. Input Backoff

two behaviours in the system. On the one hand, changing the TWT operating point changes the SNR degradation in the receiver; on the other hand, changing the operating point changes the TWT output power. When backing off from 0 dB, the SNR degradation decreases much more rapidly than the TWT output power, so that it is to be expected that there would be some optimum operating point where the difference between these two parameters is maximized.

Another striking feature of Fig. 5.26 is that the maximum linearized TWT system gain is not significantly higher than the maximum system gain for the TWT with no AM/PM. This would suggest that an equivalent system gain performance could be obtained if a TWT with only amplitude distortion were to be used rather than a linearized TWT. The reason for this apparent anomaly is that while the TWT distorts the signal more than the linearized TWT, the TWT has a higher output power which overcomes the degradation, at least under the conditions described here.

Fig. 5.27 illustrates the system gain degradation vs. output backoff for the 3 cases of our study. The system gain degradation was obtained from the difference between the actual system gain and the linear channel system gain in Fig. 5.26 while the output backoff was related to the input
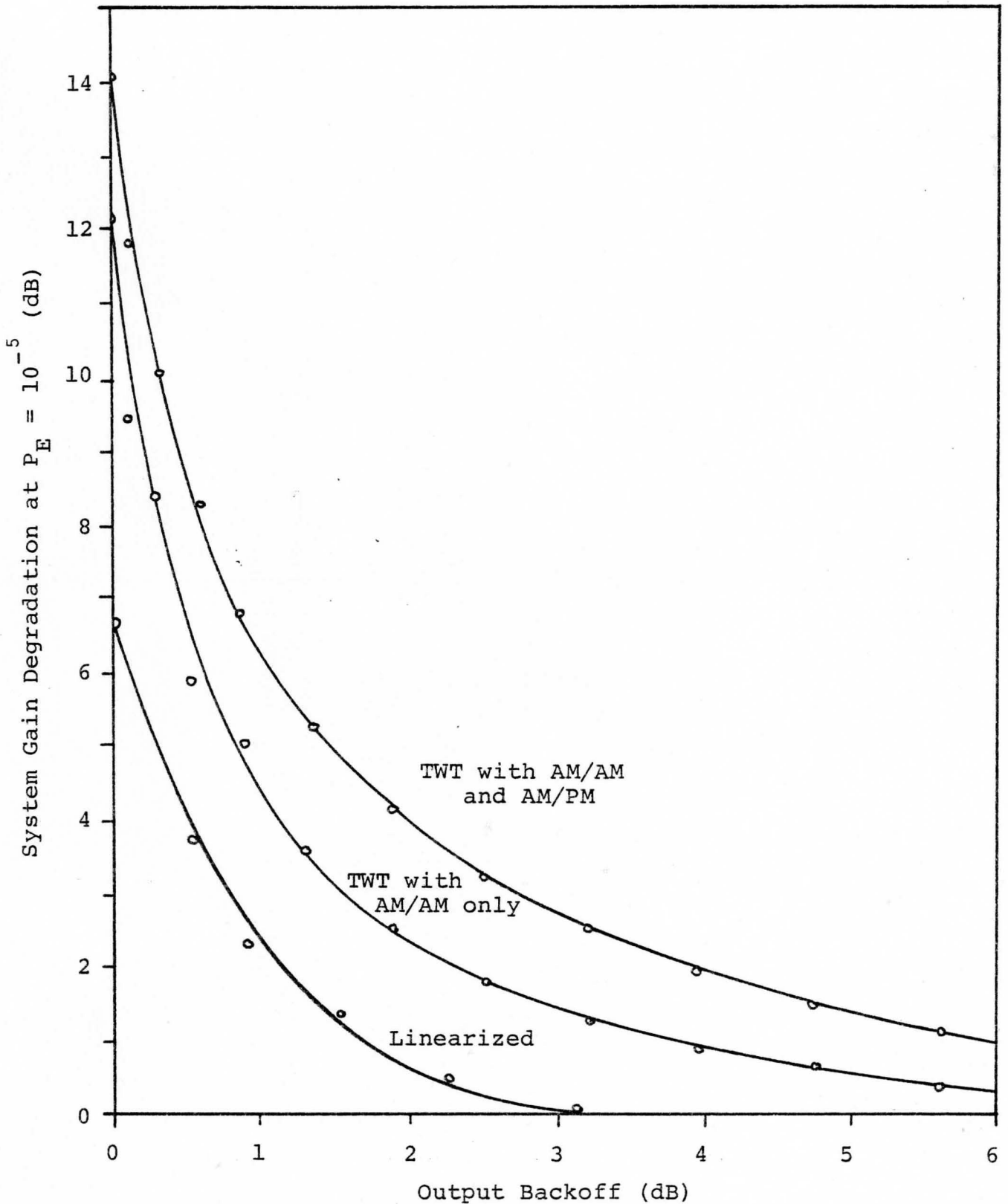
Figure 5.27   System Gain Degradation vs. TWT
Output Backoff

backoff via Fig. 5.25. We see that by backing off the TWT output by 1 dB results in a much smaller degradation with respect to a linear channel. In an actual system design, however, the most important parameter is the actual value of the system gain rather than the degradation with respect to a linear channel.

Typically, most digital radio system require a fade margin of about 40 dB [3]. Subtracting 40 dB from the ordinary TWT maximum system gain leaves 66.6 dB from which additional degrading effects due to carrier phase error, timing error and so on, must be subtracted. If the transmitting and receiving antenna gains are known, then the maximum spacing between repeaters can be determined from the maximum allowable free space loss.

5.6 Multilevel Signalling

If a multilevel signal were used as the TWT input, the output would be severely distorted to the point of being unusable. An alternative method of producing a multilevel signal must be found. Such an alternative for a 4-level signal, shown in Fig. 5.28, is a variation of a multilevel QAM transmitter developed in [55].
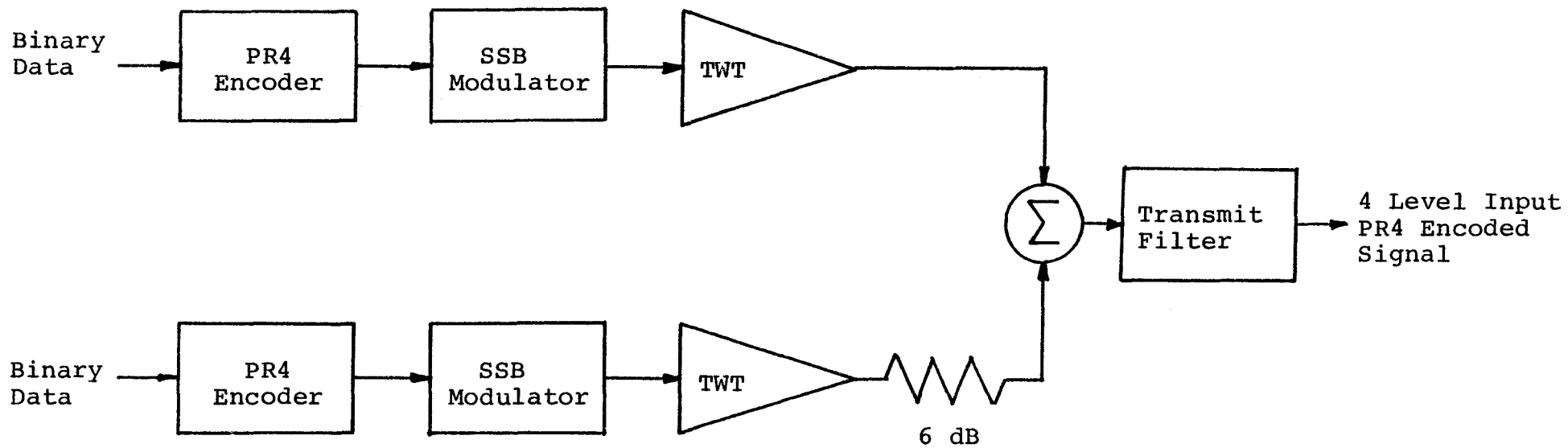
Figure 5.28   Generation of a 4 Level SSB-PR4 Signal at RF

A 4-level signal y(t) may be generated at baseband from a linear combination of two binary signals $x_1(t)$ and $x_2(t)$ according to the relation

$$y(t) = x_1(t) + 2x_2(t) \tag{5.50}$$

If $x_1(t)$ and $x_2(t)$ are independent binary data streams taking on the values ±1, then y(t) will be a 4-level data signal with values ±1, ±3. In Fig. 5.28, the combination of the two streams occurs at RF rather than at baseband, so that the amplifier distortion is reduced. The receiver would treat the signal as a 7-level PR4 encoded signal. While this system is not analyzed here, it is proposed as a possible solution to the problem of generating a 4 bits/S/Hz digital radio signal.

5.7 Conclusions

The performance of an SSB-PR4 signal when applied to a typical digital radio channel was considered. It was shown that linearizing the TWT used in the radio transmitter was an effective way of reducing the signal distortion of the amplifier. From the point of view of system gain, however, it was shown that a linearized TWT performed only marginally better than a TWT with amplitude distortion only.

Finally, a solution to the problem of generating a 4-level input SSB-PR4 signal was proposed, which deserves to be given further consideration and analysis.

CHAPTER 6. CONCLUDING REMARKS AND SUGGESTIONS FOR FURTHER RESEARCH

In this thesis, we have examined the feasibility of using single-sideband modulation in conjunction with partial response encoding for a high capacity digital radio system. Three different aspects were examined: implementability, the effects of carrier synchronization errors, and the effects of non-linear amplifiers. It was shown that while the latter two effects can seriously degrade the performance of the system, they can be overcome if additional complexity is not a design issue.

The problem of generating a 4-level input SSB-PR4 signal should be investigated further, as should carrier tracking loops which reduce data noise. Also, in a real system design, a multiplicity of issues must be considered in addition to what was discussed here, namely the effects of fading, bandlimiting, adjacent channel interference, and cochannel interference.

The original reason for using SSB in digital modulation was that higher capacities could be obtained in

the same bandwidth than with PSK systems. We have indeed shown that this is true and that the implementation problems associated with SSB can be overcome, given enough complexity. During the initial discussion of types of modulation, it was also mentioned that the bandwidth efficiency of SSB was the same as that of a QAM system. We must now determine the relative advantages and disadvantages of using SSB and QAM.

The QAM equivalent of SSB-PR4 is QPRS, where two duobinary encoded data streams independently modulate the inphase and quadrature components of the carrier. Each of the two streams would have a rate of one half the data rate of the SSB-PR4 system so that the combination of the two streams yields the same data rate in the same bandwidth, and hence the same bandwidth efficiency. Implementation of two such systems are described in [29] and [31].

The first point to be noted is that QPRS has none of the generation problems that SSB has. No additional processing is necessary to generate the QPRS signal other than the PR encoding while one of the sidebands must be eliminated in SSB.

Secondly, the two systems are equivalent from the

point of view of SNR performance. As the spectral densities of the two signals are almost identical, we would expect the two systems to perform similarly under fading conditions. For the same reason, bandlimiting the two signals would have the same effect on both systems. The carrier tracking loop for the QPRS signal would have the symmetric structure discussed in Chapter 4 so that the data noise problem would be much less severe than for SSB.

In [31] the QPRS signal is generated by duobinary filtering a QPSK signal after amplification. Since the QPSK signal has a constant envelope, there is no signal distortion due to the RF amplifier. Such a technique could be used to produce an SSB-PR4 signal but two stages of filtering would be needed. To eliminate one of the sidebands, a 3 dB loss in transmitter output power would occur, and additional losses would occur when PR4 filtering the signal. The QPRS signal would lose several dB of signal power due to the duobinary filtering only so that the QPRS signal would be 3 dB higher at the transmitter output. Summing two of these signals in a fashion analagous to Fig. 5.28 would produce a 4-level input QPRS signal with a 4 bits/S/Hz efficiency without the need for linearizing the TWT. The SNR degradation would be independent of the TWT operating point and the system gain would vary only with the

TWT output power for a fixed bandwidth system.

We conclude from the points just outlined that, at best, the SSB-PR4 system performs as well as the QPRS. In the areas where QPRS is better, additional complexity in the SSB system can overcome the performance deficiency but will not make SSB perform better than QPRS. While SSB can produce higher efficiencies than PSK, QPRS can produce the same efficiencies as SSB, but with much lower complexity. It would seem, then, that QPRS would be the superior overall for digital radio applications. SSB-PR4 could still have applications in areas where SSB is already used such as in FDM groups. However, for the design of new systems, SSB-PR4 is not competitive with QPRS.

REFERENCES

1. R.G. DeWitt, "Digital microwave radio -- another building block for the integrated digital network", Proc. Int. Conf. Comm, , June 1975, pp. 21.5-21.9.

2. J.G. Smith, "Spectrally efficient modulation", Proc. Int. Conf. Comm, June 1977, pp 3.1-37 - 3.1-41.

3. I. Godier, "DRS-8 digital radio for long-haul transmission", Prcc. Int. Conf. Comm, June 1977, pp 5.4-102 - 5.4-105.

4. W.R. Bennett, J.R. Davey, "Data transmission", McGraw-Hill 1965.

5. C.N. Compopiano, B.G. Glazer, "A coherent digital amplitude and phase modulation scheme" IRE Trans. Comm, Syst., Vol. CS-10, pp 90-95, March 1962.

6. P.R. Hartman, J.A. Crossett, "A 90 Mbs digital transmission system at 11 GHz using 8 PSK modulation", Proc. Int. Conf. Comm, June 1976, pp 18.8 -18.13.

7. R. Mitchell et al, "An eight level PSK microwave radio for long haul data communications", Proc. Int. Conf. Comm, June 1975, pp. 21.15 - 21.19.

8. W.A.H. Wood, "Modulation and filtering techniques for 3 bits/Hz operation in the 6 GHz frequency band" Proc. Int. Conf. Comm, June 1977, pp. 5.3-97 - 5.3-101.

9. R.S. Kennedy, "Fading dispersive communication channels", J. Wiley 1969.

10. C.F. Kurth, "Generation of single-sideband signals in multiplex communication systems", IEEE Trans. Circuits and Systems, vol. CAS-23, No. 1, pp 1-17 Jan. 1976.

11. "Transmission systems for communications", Bell Telephone Laboratories, Dec. 1971.

12. R.L. Freeman, "Telecommunication Transmission Handbook", J. Wiley 1975.

13. M. Harp, "New developments in radio systems", Proc. Int. Conf. Comm, June 1973, pp 28.7 - 28.11.

14. R.E. Markle, "The AR6A single-sideband long haul radio system", Proc. Int. Conf. Comm., June 1977, pp 40.1-78 - 40.1-82.

15. A.E. Chertok, "A multiple speed partial response modem featuring a unique SSB modulation-demodulation system", Proc. Int. Conf. Comm., June 1969, pp 29.21 - 29.29.

16. R.J. Sluyter, P.J. Van Gerwen, "Dual single-sideband modulation for wideband data transmission", Electronics Letters, Vol. 7 No. 21, pp 640-642, 21st Oct. 1971.

17. R.T.A. Standford, "Single-sideband digital communications", Proc. IEE, Vol. 114 No. 12, Dec. 1967, pp. 1825-1830.

18. D.E. Norgaard, "The phase-shift method of single sideband signal generation", Proc. IRE, Vol. 44, pp. 1718-1735, Dec. 1956.

19. D.K. Weaver, "A third method of generation and detection of single-sideband signals", Proc. IRE, vol. 44, pp. 1703-1705, Dec. 1956.

20. S.S. Haykin, "Communication systems", J. Wiley 1978.

21. S. Singh et al, "Digital single-sideband modulation", IEEE Trans. Comm., Tech., Vol CCM-21, pp. 255-262, March 1973.

22. W.C. Lindsay, M.K. Simon, "Telecommunication systems engineering", Prentice-Hall 1973.

23. A. Papoulis, "Probability, random variables and stochastic processes", McGraw-Hill 1965.

24. A. Lender, "The duobinary technique for high-speed data transmission" IEEE Trans. Comm Electron., Vol. 82, pp. 214-218, May 1963.

25. E.R. Kretzmer, "Generalization of a technique for binary data communication" IEEE Trans. Comm, Technol., Vol COM-14, pp. 67-68, Feb. 1966.

26. P. Kabal, S. Pasupathy, "Partial-response signalling", IEEE Trans. Comm, , Vol COM-23, pp. 921-934, Sept. 1975.

27. J. Baart, D. Baker, "DG-1 equipment for parallel data under voice systems", Telesis, Fall 1976 pp. 46-52.

28. A. Lender, "Seven level correlative digital transmission over radio", Proc. Int. Conf. Comm, June 1976, pp 18.22 - 18.26.

29. H. Kurematsu et al, "The QAM2G-10R digital radio equipment using a partial response system", Fujitsu Sci. and Technol. J., Vol. 13, No. 2, pp 27-48, June 1977.

30. R. Lucky, J. Salz, E.J. Weldon Jr. "Principles of data communication", McGraw-Hill 1968.

31. C.W. Anderson, S.G. Barber, "Modulation considerations for a 91 Mbit/s digital radio", IEEE Trans. Comm, , Vol COM-26, pp. 523-527, May 1978.

32. J.F. Gunn, J.A. Lombardi, "Error detection for partial response systems" IEEE Trans. Comm, Tech., Vol. COM-16, pp. 734-737, Dec. 1969.

33. J.M. Wozencraft, I.M. Jacobs, "Principles of communications engineering", J. Wiley, 1965.

34. I.S. Gradshteyn, I.M. Ryzhik, "Tables of integrals, series and products", Academic Press 1965.

35. H. Kobayashi, "Correlative encoding and maximum likelihood decoding", IEEE Trans. Inf. Theory, vol. IT-18, pp. 586-594, Sept. 1971.

36. G.D. Forney Jr., "The Viterbi Algorithm", Proc. IEEE, Vol. 61, pp. 268-278, March 1973.

37. D.L. Duttweiler et al, "An upper bound on the error probability in decision feedback equalization", IEEE Trans. Inf. Theory, vol IT-20, pp 490-497, July 1974.

38. M.E. Austin, "Decision feedback equalization for digital communication over dispersive channels", MIT/RLE Tech. Rep. 461, Aug. 1967.

39. D. Cheung, "Signal processing analysis and carrier

recovery circuits for a digital radio system", CRL Report No. 37, McMaster University May 1976.

40. J.J. Stiffler, "Theory of synchronous communications", Prentice-Hall 1971.

41. U. Mazzei, "Carrier phase tracking loop computer simulation and performance evaluation in high speed SSB data transmission", CSELT Rapporti tecnici, No. 3, pp. 49-54, April 1974.

42. W.C. Lindsay, M.K. Simon, "Data-aided carrier tracking loops", IEEE Trans. Comm, Tech., vol. COM-19, pp. 157-168, April 1971.

43. M.K. Simon, J.G. Smith, "Carrier Synchronization and detection of QASK signal sets", IEEE Trans. Comm, vol. COM-22, pp. 98-106, Feb. 1974.

44. M.K. Simon, J.G. Smith, "Offset quadrature communications with decision feedback carrier synchronization", IEEE Trans. Comm, vol COM-22, pp. 1576-1584, Oct. 1974.

45. H. Kobayashi, "Simultaneous adaptive estimation and decision algorithm for carrier modulated data transmission systems", IEEE Tran. Comm, Tech., vol COM-19, pp. 268-280, June 1971.

46. G. Schollmeier and A. Schatz, "The design of nonlinear phase-tracking loops by simulation", IEEE Trans. Comm, , Vol. COM-23, pp. 296-299, Feb. 1975.

47. B.M. Smith, "Some results for the eye pattern of class 4 partial response data signals", IEEE Trans. Comm, vol. COM-22, pp. 696-698, May, 1974.

48. A.J. Viterbi, "Principles of coherent communication", McGraw-Hill, 1966.

49. J.R. Pierce, "Travelling Wave Tubes", D. Van Nostrand, 1950.

50. Y. Tan et al, "The 8 level PSK modem with cosine roll-off spectrum for digital microwave communications", Proc. Int. Conf. Comm, June 1969, pp. 29.13-29.18.

51. C.M. Thomas et al,"Digital amplitude-phase keying with M-ary alphabets", IEEE Trans. Comm, Tech., Vol. COM-22, pp. 168-180, Feb. 1974.

52. P. Hetrakul, "Performance analysis of CPSK transmission through non-linear channels", Ph.D. Thesis, Dept. Elec. Eng., McMaster University, March 1976.

53. H. Seidel, "A microwave feed-forward experiment", Bell System Technical Journal, Vol. 50, No. 9, pp. 2879-2926, Nov. 1971.

54. D. Chakraborty, "Consideration of 8 phase CPSK-TDMA signal transmission through bandlimited satellite channels", IEEE Trans. Comm, , Vol. COM-25, pp. 1233-1237, Oct. 1977.

55. K. Miyauchi, S. Seki and H. Ishio, "New technique for generating and detecting multilevel signalling formats", IEEE Trans. Comm, , vol. COM-24, pp. 263-277, Feb. 1976.

56. D.P. Taylor et al, "Comparative evaluation of digital modulation techniques", CRL Report No. 18, Vol. 3, McMaster University, April 1974.

APPENDIX A. HILBERT TRANSFORMS OF PARTIAL RESPONSE
PULSES

From (3.6), the impulse response of a minimum bandwidth partial response encoder is

$$h(t) = \sum_{n=0}^{N-1} f_n \frac{\sin \pi (t-nT)/T}{\pi (t-nT) T} \qquad (A.1)$$

The corresponding Hilbert Transform pulse is

$$\hat{h}(t) = H \left[\sum_{n=0}^{N-1} f_n \frac{\sin \pi (t-nT)/T}{\pi (t-nT)/T} \right]$$

$$= \sum_{n=0}^{N-1} f_n H \left[\frac{\sin \pi (t-nT)/T}{\pi (t-nT)/T} \right] \qquad (A.2)$$

where H [.] denotes the Hilbert Transform of the quantity within the brackets. The Fourier Transform of $g(t) = \sin \pi t/T \,/\pi t/T$ is

$$G(f) = 1 \quad , |f| < 1/2T \tag{A.3}$$

$$= 0 \quad , \text{ elsewhere}$$

The Hilbert Transform of the Fourier Transform is

$$\hat{G}(f) = -j \, \text{sgn}(f) G(f) \tag{A.4}$$

Hence, the Hilbert Transform of the Fourier Transform of the pulse $g(t)$ is

$$\hat{G}(f) = -j \quad , 0 < f < 1/2T$$

$$= 0 \quad , \quad f = 0$$

$$= +j \quad , -1/2T < f < 0 \tag{A.5}$$

Taking the inverse Fourier Transform of (A.5) we have

$$\hat{g}(t) = \int_{-\infty}^{\infty} \hat{G}(f) e^{j2\pi ft} \, df$$

$$= \frac{1 - \cos \pi t/T}{\pi t/T} \tag{A.6}$$

The Hilbert Transform of a generalized PR pulse is therefore

$$\hat{h}(t) = \sum_{n=0}^{N-1} f_n \frac{1 - \cos \pi (t-nT)/T}{\pi (t-nT)/T}$$

$$= \frac{T}{\pi} \sum_{n=0}^{N-1} f_n \frac{1 - (-1)^n \cos \pi t/T}{t-nT} \tag{A.7}$$

In order to obtain the responses of Table 3.2, we simply shift the time origin in (A.7) by $(N-1)T/2$ so as to make the pulse symmetric about $t = 0$.

APPENDIX E. DERIVATION OF PARAMETER d/σ FOR TWO SPECTRAL SHAPING MODELS

MODEL 1: SPLIT SHAPING

Denote the average symbol power at the shaping filter input as P_S. For L input amplitudes $\pm d$, $\pm 3d$, ..., $\pm (L-1)d$ with a symbol period T, we have

$$P_S = \frac{1}{T} \frac{1}{L} \sum_{i=-L/2}^{L/2} [(2i-1)d]^2$$

$$= \frac{d^2}{3T} (L^2 - 1) \tag{B.1}$$

The power input to the channel, P_c, is given by

$$P_c = \frac{P_S}{2\pi} \int_{\infty}^{\infty} |H_T(\omega)|^2 \, d\omega$$

$$= \frac{P_S}{2\pi} \int_{-\pi/T}^{\pi/T} 2T |\sin\omega T| \, d\omega$$

$$= \frac{4P_S}{\pi}$$

$$= \frac{4d^2}{3T} (L^2 - 1) \tag{B.2}$$

Rearranging (B.2) gives the decision distance as a function of channel input power and the number of levels L, as shown by

$$d = \left(\frac{3TP_c}{4(L^2-1)} \right)^{1/2} \tag{B.3}$$

The noise power at the input of the receiver is

$$P_N = 2 \ (\frac{N_o}{2}) \ (\frac{1}{2T})$$

$$= \frac{N_o}{2T} \tag{B.4}$$

because the noise is AWGN over a bandwidth 1/2T.

The noise power at the receiver filter output is

$$P_{N_o} = \frac{N_o}{2} \int_{-\infty}^{\infty} |H_R(\omega)|^2 \ \frac{d\omega}{2\pi}$$

$$= \frac{N_o}{2} \int_{-\pi/T}^{\pi/T} 2T |\sin\omega T| \ \frac{d\omega}{2\pi}$$

$$= \frac{2}{\pi} N_o \tag{B.5}$$

This is the variance σ^2 of the noise at the decision device input. Rewriting (B.5) we may express the output noise variance in terms of the received noise power at the filter input as

$$\sigma^2 = \frac{4T}{\pi} P_N \tag{B.6}$$

Combining (B.3) and (B.6), we obtain

$$\frac{d}{\sigma} = \frac{\pi}{4} \left(\frac{3}{L^2-1} \ \frac{P_c}{P_N} \right)^{1/2} \tag{B.7}$$

where the ratio P_c/P_N is the received (channel) average signal to noise ratio.

MODEL 2 : FULL TRANSMITTER SHAPING

The average symbol power P_S is the same as for the first model. However, the channel input power for model 2 is

$$P_c = \frac{P_S}{2\pi} \int_{\pi/T}^{\pi/T} 4T^2 \sin^2 \omega T \, d\omega$$

$$= 2TP_S$$

$$= \frac{2}{3} d^2 (L^2 - 1) \qquad (B.8)$$

Therefore, the decision distance as a function of channel input power is

$$d = \left(\frac{3P_c}{2(L^2 - 1)} \right)^{1/2} \qquad (B.9)$$

For this model, the noise power at the output of the receiver filter is the same as the noise power at the receiver filter input. Hence

$$P_N = P_{N_o}$$

$$= \frac{N_o}{2T} \qquad (B.10)$$

which gives, for this model,

$$\frac{d}{\sigma} = \left(\frac{3}{2(L^2-1)} \quad \frac{P_C}{P_N} \right)^{1/2}$$

(B.11)

APPENDIX C. DESCRIPTION OF COMPUTER SIMULATIONS

C.1 Common Features

Several different simulation programs were used depending on the type of simulation done. Certain features are common to all the programs and will be discussed now.

All simulations were done in FORTRAN on a CDC 6400 computer. Baseband simulation techniques were used since RF simulations are highly impractical. All operations were performed on the complex pre-envelope [20] of the signal rather than on the signal itself. These pre-envelopes were readily represented using the COMPLEX variable type in FORTRAN.

Pseudo-random binary data sequences were generated by producing an array of pseudo-random variables uniformly distributed between 0 and 1 using the IMSL subroutine GGUB. During the initial entry to the subroutine, an integer seed number must be supplied but the subroutine generates its own seeds for

subsequent calls. Each element of the array supplied by
the subroutine is assigned the value 1 or -1 depending
upon whether the random variable is greater or less than
0.5. Four level sequences were produced through a
linear combination of independent binary sequences.

Filtering operations were performed using the
Fast Fourier Transform (FFT). For linear channel
simulations, a 256 point FFT was used with 1
sample/symbol so that the program processes an array of
256 symbols at any one time. When non-linearities are
present, the frequency spectrum of the signal spreads
and space must be allocated in the frequency domain to
accomodate this spread. This is accomplished by
oversampling in the time domain. For the non-linear
channel simulations used here, 16 samples/symbol were
used with a 256 point FFT so that the signal was
synthesized over 16 symbol periods. However, in the
frequency domain, 16 times the bandwidth of the 1
sample/symbol signal is available which is ample room
for the frequency spread.

Sectioning of the data is necessary to simulate
a continuous system on a discrete machine. The binary
data is processed in 256 element blocks, fast Fourier

transformed, multiplied by the appropriate filter function, and inverse transformed. To prevent wraparound of the filtered signal in the time domain due to the periodic nature of the FFT, zeros are inserted in the input data blocks so that the transient response of the filter can be accomodated. In these simulations, 128 zeros were inserted into the 256 element filter input blocks. The filter output is split into two parts: the portion corresponding to the head of the filter-data convolution, and the portion corresponding to the tail. The tail of the previous filter output calculation is added to the head of the current filter output block and the current tail saved for the next calculation. The filtering and tail addition are performed by two subroutines FILTER and ADDTAIL. In this fashion, we create a continuous stream of filtered data which can be used in subsequent processing.

Two possible approaches exist for the simulation itself: the Monte Carlo method and the calculated noise method [56]. In the Monte Carlo method an actual noise sample (i.e. a pseudo-random Gaussian number) is added to the signal and a decision is made based on the received signal plus noise. The decision result is compared to the original symbol and errors counted. The

advantage of this method is its flexibility in that any theory of noise may be accomodated. The primary disadvantage is that at high SNR's, error events occur infrequently and large numbers of symbols must be processed for a small simulation error. The standard deviation of an error rate P obtained from a Monte Carlo simulation is given by [56]

$$\sigma_p = [P(1-P)/N]^{1/2} \qquad (C.1)$$

where N is the number of symbols processed. Suppose that we require that $P = 10\,\sigma_p$ i.e. that there only be a 10% simulation error. We would then need to process

$$N = 100(1-P)/P \qquad (C.2)$$

symbols for the required accuracy. In order to measure an error rate of 10^{-4} with an error of 10%, we must then process 10^6 symbols, which clearly requires an inordinate amount of computing time.

To reduce the time required for a given simulation, particularly at high SNR's, a computed noise approach is required. Here the noise variance σ^2 at

the decision device input is calculated according to theory, that is according to the noise spectral density and receiver input bandwidth. For example, for a binary PAM system in a Gaussian noise environment, if the received signal level is V_i and the receiver threshold is T, then the probability of error for that particular signal level is

$$P[E|V_i] = Q[(V_i-T)/\sigma]$$ (C.3)

if V_i is on the correct side of the threshold. Under certain conditions, the signal may be so distorted that errors are made even when noise is not present. Under these conditions, V_i is on the incorrect side of the threshold so that the probability of error for that signal level is

$$P[E|V_i] = 1 - Q[(V_i-T)/\sigma]$$ (C.4)

By averaging over all the received signal levels, the average probability of error can be computed using

$$P[E] = \frac{1}{N} \sum_{i=1}^{N} P[E|V_i]$$ (C.5)

where N is the number of symbols processed. To obtain an accurate measure of the error rate using this method, enough symbols must be processed so that all possible signal levels are considered and are present in their correct distribution. Once this requirement is met, error rates can be determined just as easily at high SNR's as at low SNR's as the signal level distribution is independent of SNR. In most cases, though, due to the large number of possible signal levels, the use of calculated noise at high SNR's is limited.

Calculated noise is used whenever possible in conjunction with a precoded PR receiver. Use of a receiver with any sort of a decision feedback mechanism dictates the use of Monte Carlo methods, however, as error propagation effects cannot be simulated using a calculated noise approach.

In the design of the simulation programs themselves, top-down techniques were used, which involves decomposing the program into a hierarchy of subroutines. This decomposition is accomplished with relatively little difficulty as the various simulation functions such as data generation, filtering, tail addition, etc. are already logically distinct.

Correctness of the software was verified using bottom-up testing, where each of the individual subroutines is tested from the lowest level up to ensure that the particular function is question works correctly. The correctness of the ensemble of subroutines is verified by comparison of simulation results with well-defined theoretical results. For example, the correctness of the phase error simulation was verified by running the system with no phase error and then comparing with the theoretical result (3.40).

C.2 Equalizer and I-Q Receiver Simulation

Since a decision feedback equalizer was used, a Monte Carlo simulation was required. Input parameters to the simulation consist of (i) the number of SNR points to be simulated and their values (ii) the number of equalizer taps and their values (iii) a weighting coefficient for the inphase channel (iv) the number of points to be considered and (v) seed numbers for the data sequence and noise generation routines. For lower SNR's only several thousand points (symbols) were necessary but higher SNR values (> 13 dB) required the use of up to ten thousand points. Computing time

restrictions limited the number of points to this value so that error rate measurements below 10^{-3} were not possible.

The equalizer coefficients are stored in an array COEFF with the corresponding signal values in an array EQIN. Multiplication of each element of COEFF with each element of EQIN and then summing the result provides the new decision variable upon which the next decision is based. The higher order locations of EQIN contain the previous decisions corresponding to the feedback portion of the equalizer, while the lower order locations contain the signal plus noise samples of the feedforward portion. The exact proportion of each depends on the number of feedforward/feedback taps specified at run time. Once a new decision is made, the previous decisions are shifted by one element with the new decision occupying the first feedback position. The signal plus noise samples are also shifted and a new value entered into the lowest order element.

At the start of the program, there are no decisions to be fed back and the initial decisions made by the equalizer are likely to to be in error. Some provision must be made for this equalizer transient

response so the equalizer was allowed to run for 500 symbols before error counting began.

When the inphase-quadrature receiver of Fig. 3.16 was simulated, it was necessary to generate noise described by (2.45) and (2.46). Since $n_1(t)$ and $n_2(t)$ are narrowband processes of one half the bandwidth of the baseband data signal, they are generated by filtering two arrays of independent Gaussian pseudo-random variables generated by subroutine GGNOR. After transformation into the frequency domain by means of the FFT, the two arrays are limited to one half their original bandwidth and then inverse transformed. Multiplication by $\sin \pi t/2T$ and $\cos \pi t/2T$ in their appropriate sampled form provides the required noise processes. The accuracy of this noise generator was determined by measuring the autocorrelation of the noises, as well as the cross-correlation of the inphase and quadrature noise for various offsets. Close agreement between the measured values and the values of Fig. 2.8 was obtained.

The simulation flowchart is shown in Fig. C.1.
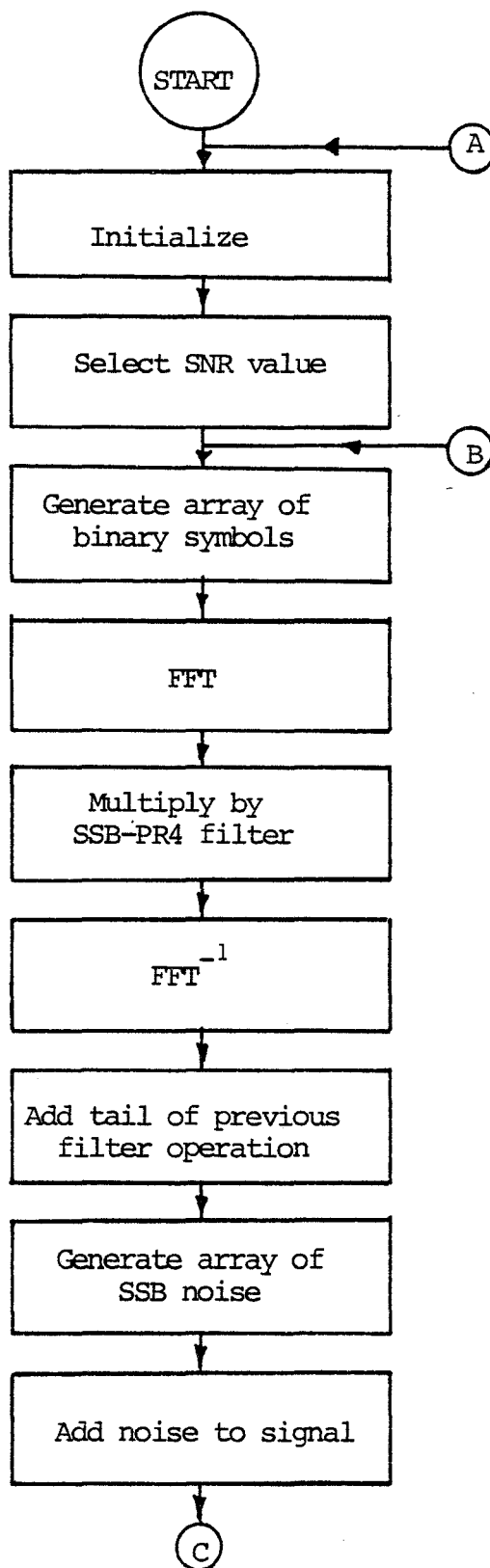
Figure C.1 Flowchart of Quadrature Channel
Simulation

238
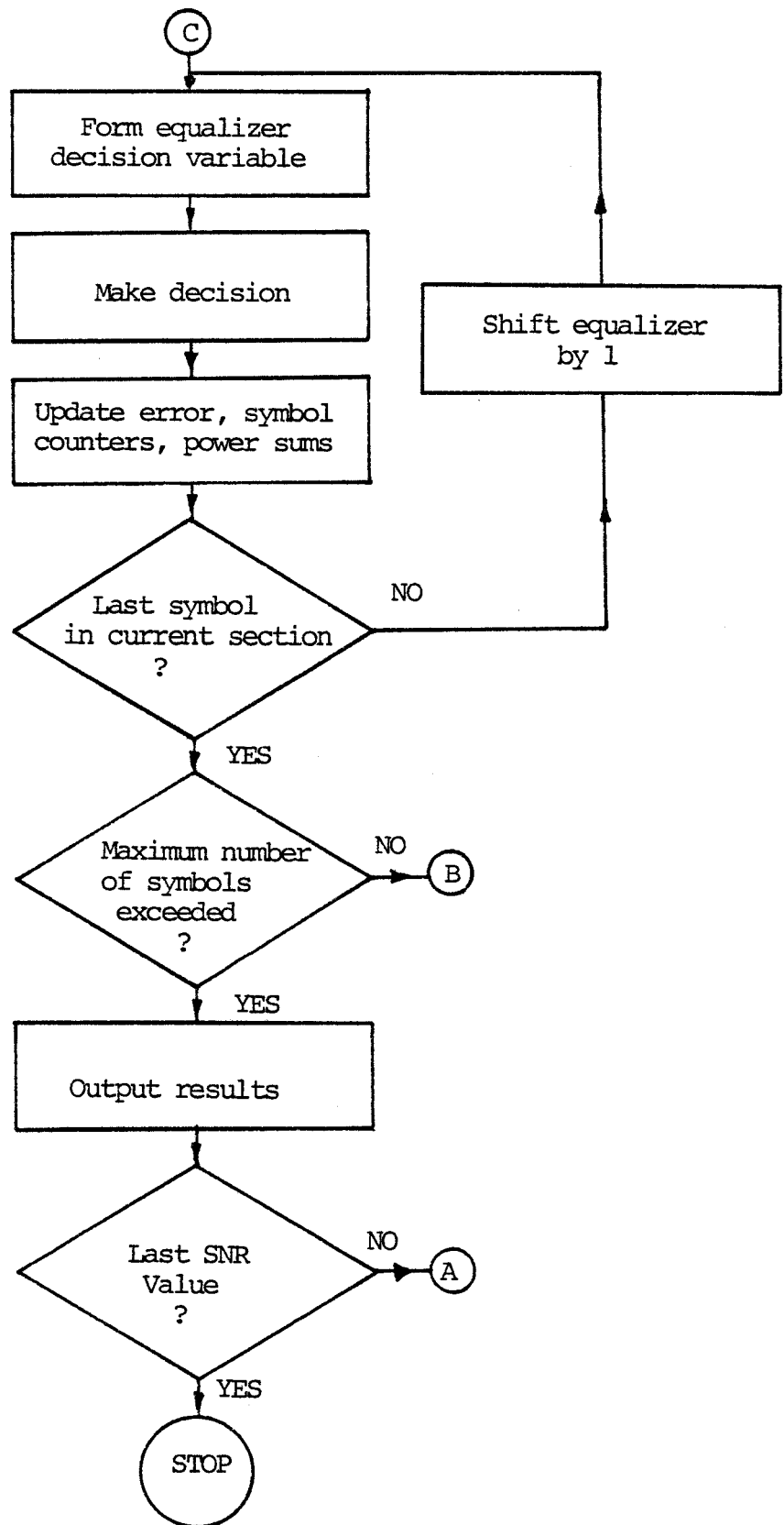
Figure C.1 (cont.) Flowchart of Quadrature Channel
Simulation

C.3 Phase Error and Tracking Loop Simulation

C.3.1 Phase Error Simulation

A precoded receiver was used in conjunction with a calculated noise simulation. Simulation input parameters include: (i) number of SNR points and their values, (ii) number of symbols to be processed, (iii) steady-state phase error to be simulated, and (iv) seed number for the data generator.

The SSB-PR4 filter output is sectioned as outlined earlier to provide a stream of filtered data. The demodulator inphase output component is formed from the SSB I and Q signal components and the phase error via (4.8). For the binary input case, the demodulator output is rectified as would be in a precoded receiver. The distance to the binary decision threshold is calculated and the probability of error calculated for all supplied values of SNR using (C.3) or (C.4) as appropriate with the value of computed from the SNR via (3.43). The simulation flowchart for binary input signals is shown in Fig. C.2. For the 4 level input case, the probability of coalescion was assumed negligible so that the receiver simulated was simply a 7 level decision device. Processing of one thousand symbols was found adequate for convergence of the
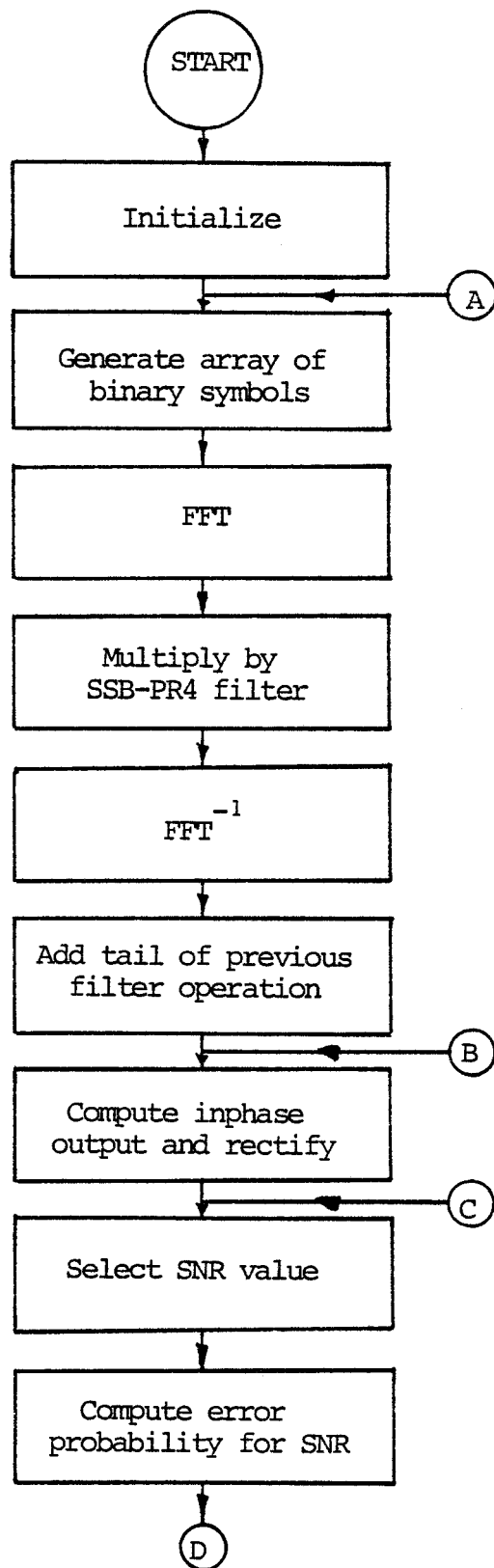
Figure C.2 Flowchart of Steady-State Phase Error
Simulation
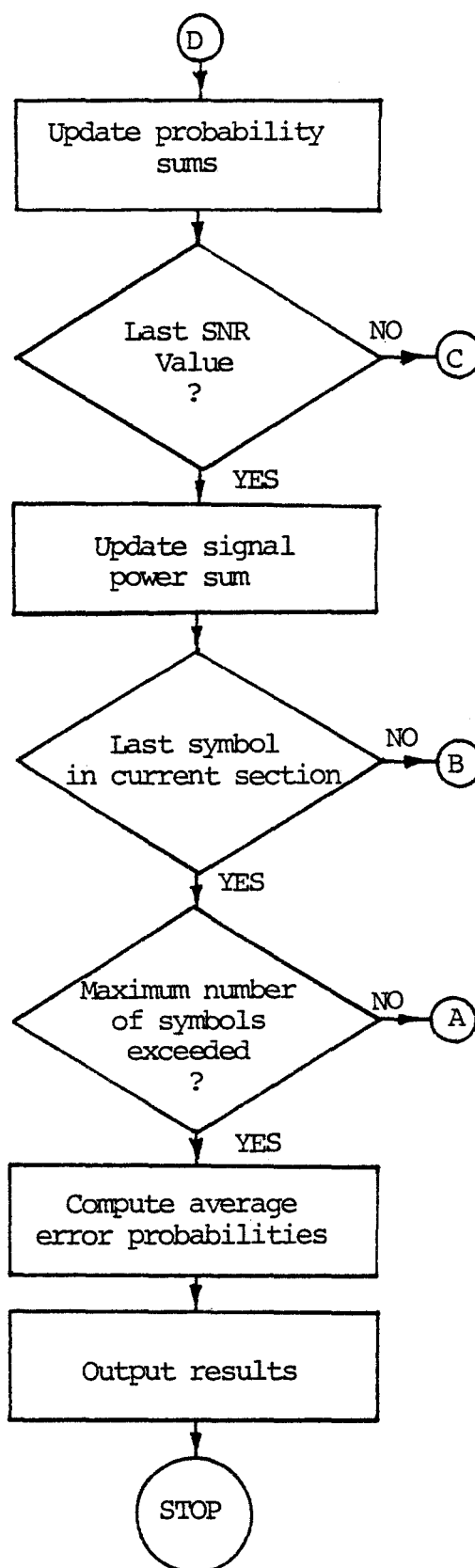
Figure C.2 (cont.) Flowchart of Steady-State Phase
Error Simulation

measured error probabilities to stable values.

C.3.2 Tracking Loop Simulation

A Monte Carlo simulation was necessary due to the decision feedback nature of the loop. After generation of the SSB-PR4 signal, independent Gaussian noise samples are added to the I and Q signal components. Since no sampler offset between the I and Q channels is involved, and since the I and Q noises are uncorrelated for zero offset, adding such independent samples is justified. The received carrier phase was assumed constant (i.e. 0 °) and the jitter of the receiver oscillator phase about this value was measured.

The receiver decision device input was calculated from the current received signal and the previous carrier phase estimate. A three level decision is made and then cross-multiplied with the quadrature receiver output. The new phase estimate is calculated according to the relation

$$\phi_{k+1} = \phi_k - \tilde{R}_I r_Q(kT)/\delta_s \qquad (C.6)$$

with $\tilde{R}_I$, $r_Q(kT)$ and δ_s defined as in Sections 4.3.1 and

4.3.4. The mean and variance of the phase error were measured to provide the RMS phase error of the loop. Five thousand symbols were used.

The receiver error rate can easily be measured by mod 2 mapping the 3 level symbols into binary symbols and then comparing with the original transmitted symbols. At large SNR's few errors occur, however, and computation time becomes excessive. To eliminate this problem, calculated noise was used for the purposes of the error rate measurement. The I and Q components of the SSB-PR4 signal and the previous phase estimate are used to generate the decision device input with no noise samples added. After rectification, the distance to the binary threshold is calculated and the probability of error for that received signal value computed as for the steady-state phase error simulation.

The tracking loop simulation flowchart is shown in Fig. C.3.

C.4 Non-Linear Channel Simulation

The system model of Fig. 5.9 was used in the non-linear channel simulation. Simulation input
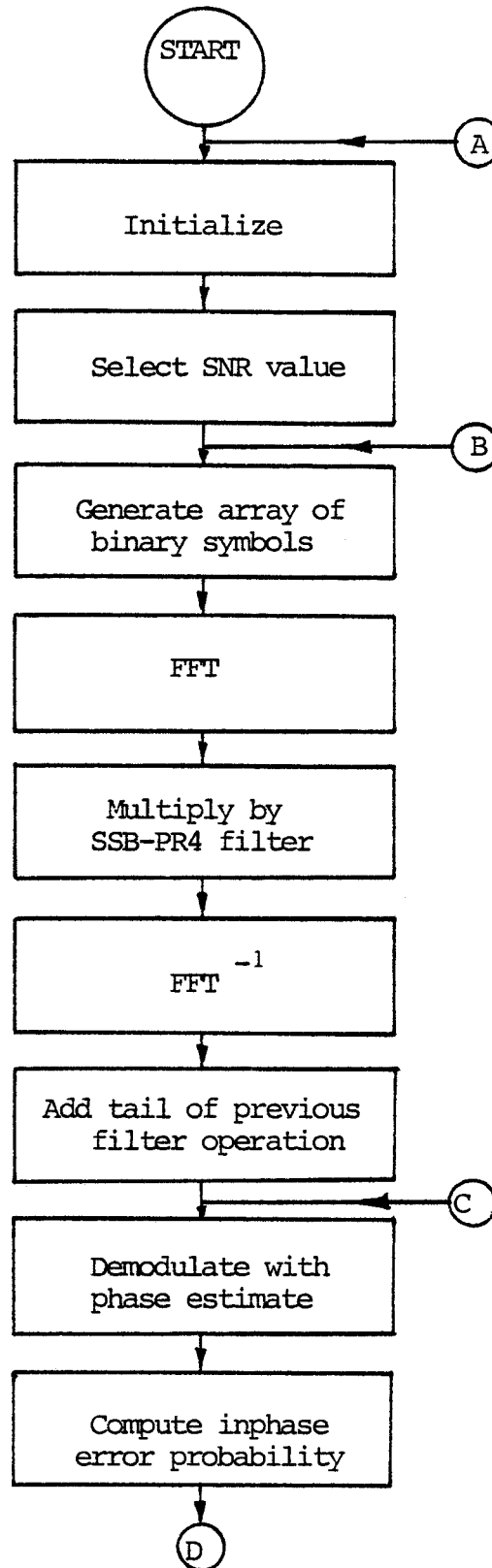
Figure C.3 Flowchart of Tracking Loop Simulation
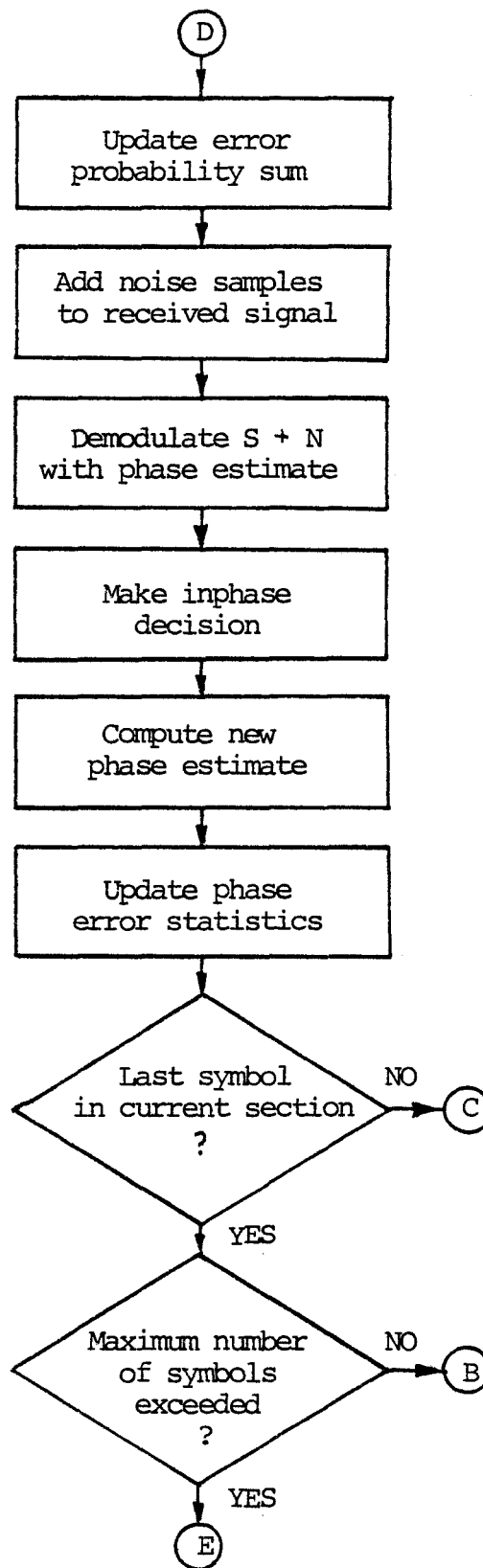
Figure C.3 (cont.) Flowchart of Tracking Loop Simulation
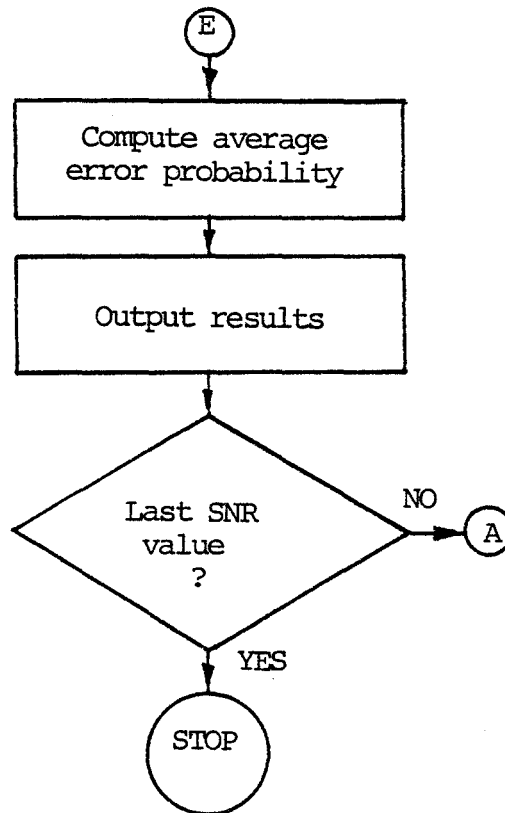
Figure C.3 (cont.) Flowchart of Tracking Loop Simulation

parameters are (i) number of SNR points and their values
(ii) mode of TWT operation ie. normal, AM/AM only, or
linearized (iii) TWT input backoff (iv) number of
symbols to be processed and (v) seed number for the data
generation routine. The simulation flowchart is shown
in Fig. C.4.

The SSB-PR4 signal is generated as before and
then passed to the TWT subroutine. During the first
entry to the subroutine, the TWT input voltage scaling
factor is computed from the input backoff. The mean
angle of rotation is then computed to realign the signal
space as outlined in Section 5.4.1, if AM/PM is to be
included. These parameters remain constant throughout
the rest of the simulation. Subsequent values of input
envelope are converted to input voltages using the
voltage scaling factor. These input voltages are used
in subroutine DISTORT to compute the TWT output envelope
and phase shift using the Bessel function model of the
TWT, (5.6) and (5.7). The signal is resolved input
inphase and quadrature components before passing the
amplified signal to the postfiltering subroutine
POSTFIL.

Use of the linearized TWT necessitates a

Figure C.4 Flowchart of Non-Linear Channel Simulation

249

```
              (C)
               │
               ▼
    ┌──────────────────────┐
    │    Compute TWT        │
    │     output            │                      (B)
    └──────────────────────┘                        │
               │                                     ▼
    ┌──────────────────────┐        ┌──────────────────────┐
    │    Rotate signal      │        │  Compute amplitude    │
    │      space            │        │  distorted output     │
    └──────────────────────┘        └──────────────────────┘
               │◄──────────────────────────────┘
    ┌──────────────────────┐
    │         FFT           │
    └──────────────────────┘
               │
               ▼
    ┌──────────────────────┐
    │    Multiply by        │
    │  postfilter response  │
    └──────────────────────┘
               │
               ▼
    ┌──────────────────────┐
    │        FFT$^{-1}$      │
    └──────────────────────┘
               │
               ▼
    ┌──────────────────────┐
    │  Add tail of previous │
    │   postfilter output   │
    └──────────────────────┘
               │◄────────────(D)
    ┌──────────────────────┐
    │   Compute inphase     │
    │  output and rectify   │
    └──────────────────────┘
               │◄────────────(E)
    ┌──────────────────────┐
    │   Compute error       │
    │  probability for SNR  │
    └──────────────────────┘
               │
               ▼
    ┌──────────────────────┐
    │   Update error        │
    │  probability sum      │
    └──────────────────────┘
               │
               ▼
              (F)
```
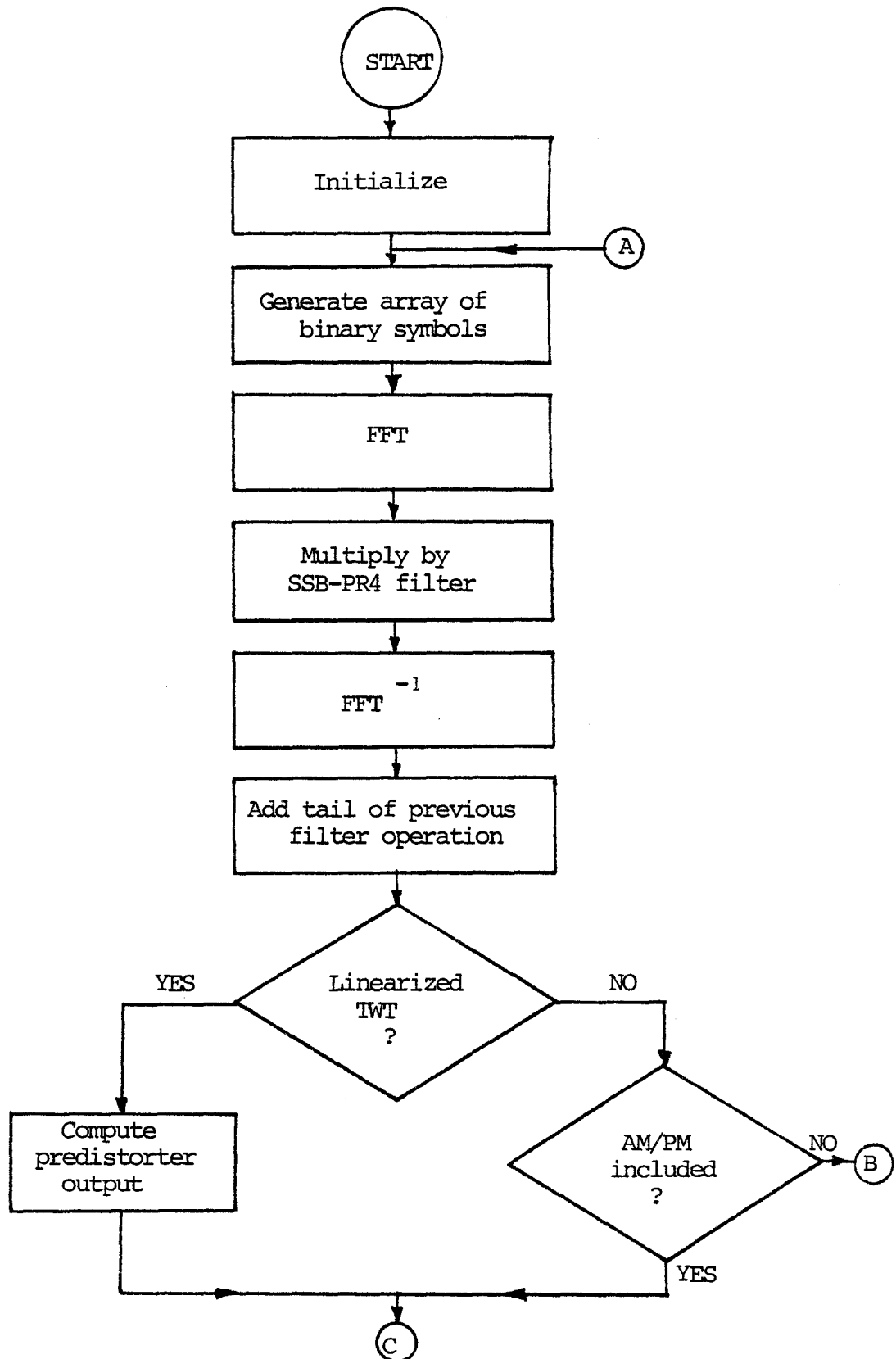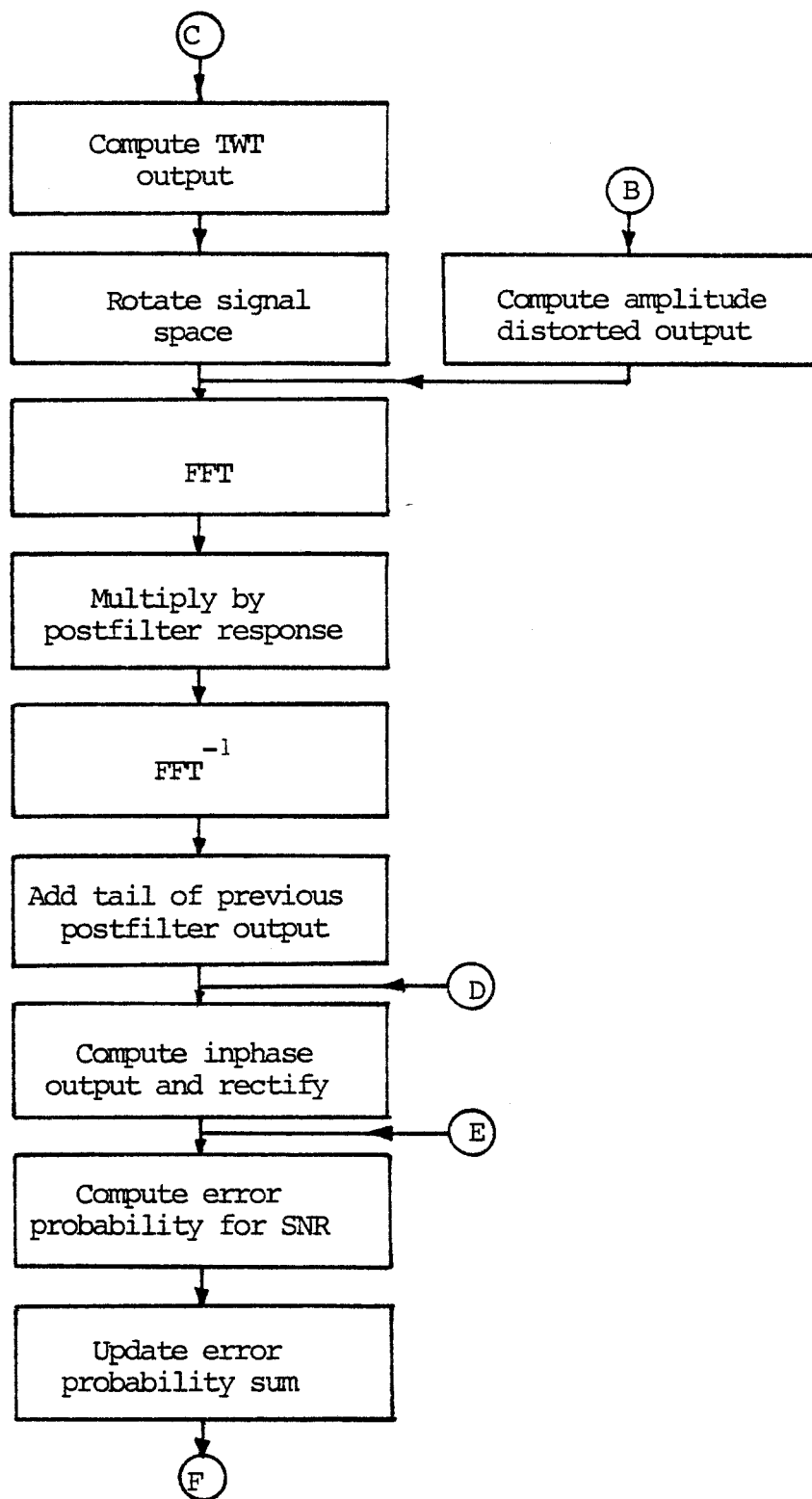
Figure C.4 (cont.) Flowchart of Non-Linear Channel
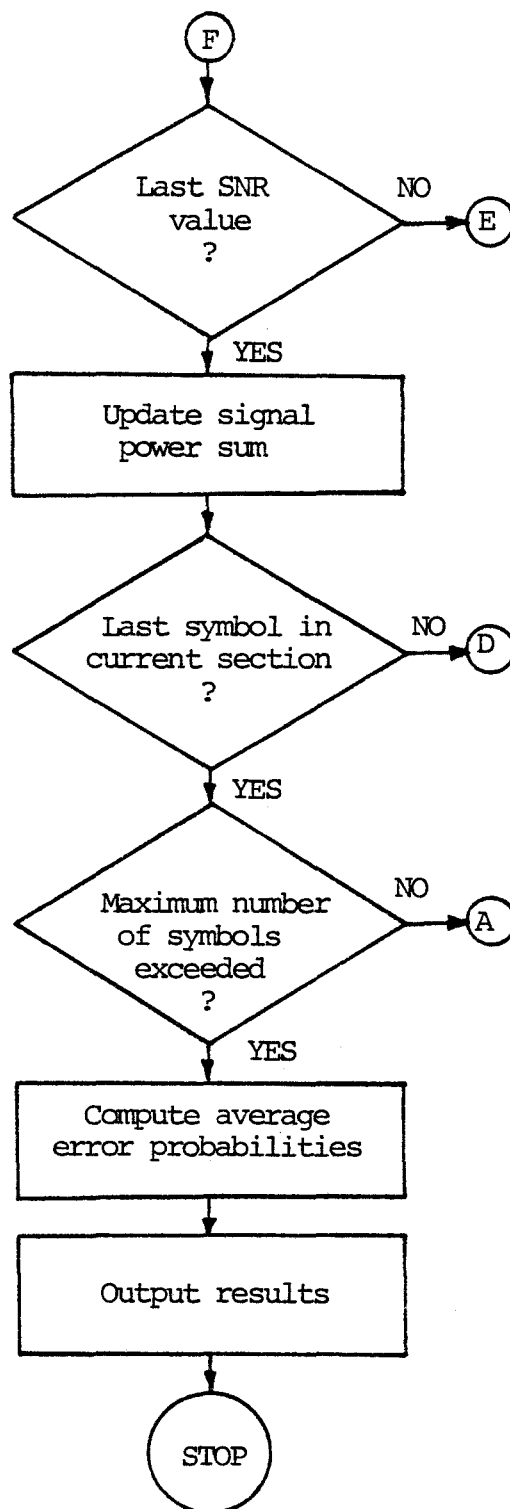Simulation

Figure C.4 (cont.) Flowchart of Non-Linear Channel
Simulation

slightly different procedure. The input envelope voltage is calculated as before, but using a different value of peak input voltage. The signal is predistorted in subroutine PREDIST using the polynomial model of (5.16) and (5.17). The predistorter output voltage is then used as the TWT input voltage in DISTORT, and the corresponding output voltage and phase shift computed. The overall phase shift is the sum of the predistorter phase shift and the TWT phase shift. Once again, the signal is resolved into its I and Q components.

Once passed through the TWT, the distorted signal is postfiltered in subroutine POSTFIL. An ideal rectangular filter was used to restrict the output spectrum to the same bandwidth as the input spectrum but other filter responses could easily be substituted. After use of the inverse FFT, tail addition is again necessary to provide a continuous section of filtered data. This is done in subroutine ATAIL.

The signal is now ready to be demodulated and the error rate determined. Since the received signal level depends on the TWT backoff, the optimum receiver thresholds also become backoff-dependent. Due to the action of the postfilter, it becomes extremely difficult

to theoretically determine the optimum receiver threshold. However, it is possible to determine this quantity during the simulation itself. This was done by measuring the upper and lower eye boundaries after rectification at the nominal sampling instant. The difference between these two values is the vertical eye opening while their mean gives the optimum threshold. The first 175 symbols generated during the simulation are used towards this end, they are not used for error probability calculations. Once the threshold is determined, it remains constant for the remainder of the simulation. An additional 1000 symbols were used for error rate measurements. Pertinent information printed during the simulation includes the SNR's and corresponding error probabilities as well as the measured TWT output power in dBm and watts, TWT voltage scale factors and angle of signal space rotation for the supplied value of input backoff.

A simulation error of 0.8 dB was determined from the simulation results obtained at 20 dB backoff. Under such a condition, the channel is essentially linear and thus should exhibit the error rate performance of a linear channel. In reality, the measured curve was 0.8 dB worse than the theoretical linear channel behaviour.

This value was used to correct the curves shown in Figs. 5.21 - 5.23. The simulation error is due primarily to the use of a rectangular postfilter response and also to the fact that the PR filter response was synthesized over 16 symbol periods only.

C.5 Eye Diagram Generation

Eye diagrams were generated on a VERSATEC plotter using modified versions of the aforementioned simulation programs. Data generation and filtering were identical, with the exception that all filtering was done using a 16 sample/symbol FFT to provide increased time domain detail. Also, postfiltering of the TWT output was not done. Rather than passing the processed symbols to an error probability calculation subroutine, the symbols are given to a plotting subroutine EYE which partitions the data into 32 element arrays (i.e. two symbols) and then calls the VERSATEC software for the actual line drawing. A judicious selection of vertical and horizontal scales produced curves sufficiently smooth without need for a spline or interpolating polynomial.

To produce the TWT output eye diagrams of Figs.

5.14-5.19, the TWT output voltages were scaled using a normalization factor obtained from the non-linear channel error rate simulation. This factor normalizes the TWT output power to the same value as the input power. Use of this scale factor is also made in the eye opening calculations of Section 5.4.2.

APPENDIX D. LIST OF ACRONYMS

AGC	Automatic gain control
AM/AM	Amplitude modulation-to-amplitude modulation conversion
AM/PM	Amplitude modulation-to-phase modulation conversion
APK	Amplitude-phase keying
AWGN	Additive white Gaussian noise
BER	Bit error rate
CPSK	Coherent phase shift keying
DSB	Double-sideband modulation
FCC	Federal Communications Commission
FDM	Frequency division multiplex
FFT	Fast Fourier Transform
FM	Frequency modulation
FSK	Frequency shift keying
I	Inphase
IF	Intermediate frequency
IM	Intermodulation
ISI	Intersymbol interference
PAM	Pulse-amplitude-modulation
PDF	Probability density function
PR	Partial response
PR4	Partial response Class 4 code

PSK   Phase-shift-keying

Q   Quadrature

QAM   Quadrature amplitude modulation

QPRS   Quadrature partial response system

RF   Radio frequency

RMS   Root-mean-square

SNR   Signal power to noise power ratio

SSB   Single-sideband modulation

SSB-PR4 Single-sideband modulation with partial
response Class 4 coding

TWT   Travelling-wave-tube

VCO   Voltage-controlled oscillator

VEO   Vertical eye opening