JPL PUBLICATION 81-73

Modulation/Demodulation Techniques for Satellite Communications

Part I: Background

Jim K. Omura
Marvin K. Simon

November 15, 1981

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

JPL PUBLICATION 81-73

Modulation/Demodulation Techniques for Satellite Communications

Part I: Background

Jim K. Omura
Marvin K. Simon

November 15, 1981

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

FOREWORD

This report represents Part I of a series of reports to be published under the same title with the following subtitles:

Part II:	Advanced Techniques - The Linear Channel
Part III:	Advanced Techniques - The Nonlinear Channel
Part IV:	Appendices

ABSTRACT

Part I of this report provides the background necessary for a complete understanding of Parts II, III, and IV. Here we summarize the error probability performance and spectral characteristics of various modulation/demodulation techniques commonly used or proposed for use in radio and satellite communication links. Forward error correction with block or convolutional codes is also discussed along with the important coding parameter, channel cutoff rate. Our purpose here is to review those known and published results that serve as a prelude to the new research results to be reported in Parts II and III.

CONTENTS

CONTENTS (Con't)

Figures

Figures (Con't)

Figures (Con't)

Figures (Con't)

Tables

1.0 Introduction

The primary goal of Part I is to summarize in a partly tutorial manner the various modulation/demodulation techniques that are suitable for use on radio and satellite communication links. Throughout, the emphasis will be on

(1) Analytical characterization of the modulation technique and its spectral occupancy,

(2) Structure and implementation of the optimum demodulator,

(3) Error probability performance of the modulation/demodulation for the ideal linear additive white Gaussian noise channel model.

Forward error correction with block and convolutional codes will also be discussed along with concatenation coding techniques. The use of coding with the various modulations discussed will be examined by the evaluation of computational cutoff rate parameters [1].

In summary, Part I will provide the necessary background material for later parts of this report that will emphasize new modulation/demodulation techniques for both linear and nonlinear satellite channel models.

2.0 System Overview

Before examining the details of various modulation and coding techniques it is important to discuss some basic characteristics of digital data transmission systems. This includes the characterization of the physical communication links, the notion of bandwidth including the FCC regulations, and performance measures such as bit rates, bit error probabilities, throughputs, and delays.

2.1 Transmitter/Receiver System

Figure 1 shows the basic components of a digital data transmission system. Only the transmitter is shown here since the receiver merely reverses the processing of the transmitter. The most important part of digital data transmission systems are the modulation and coding subsystems. These two work together to determine the basic performance of satellite and terrestrial radio links. Other parts shown here can be designed independent of the modulation and coding.

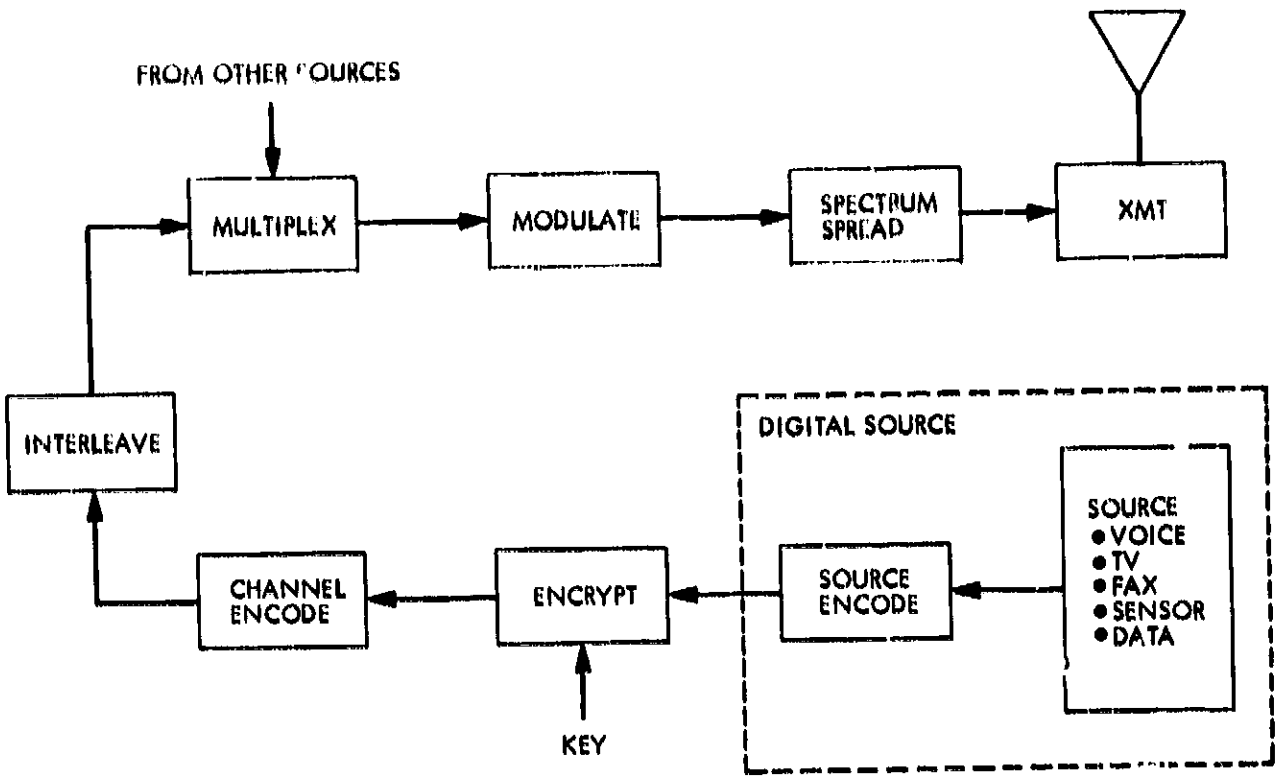


Figure 1. Digital Data Transmission System

2.1.1 Source Coding

The original sources such as voice, TV, or computer data must first be converted into digital data symbols suitable for transmission. Typically, the transmission symbols are binary. This source encoding process allows many diverse types of sources to share a common communication link. Analog sources such as voice require data compression techniques which involve some signal distortions. Sources already in digital form, such as computer outputs, may require redundancy reducing algorithms. Generally when these various source encoding techniques are done efficiently the resulting data bits for transmission appear as equal probable independent binary symbols.

2.1.2 Encryption

To protect the data from unauthorized users the data can be encrypted. Typically, this is a one-to-one mapping of binary sequences to binary sequences which require knowledge of a "key" to map back from the encrypted data bits to

the original data bits. The Data Encryption Standard (DES) [2] and the Public Key Cryptosystem (PKC) [3] are currently two highly secure systems that are commonly considered. The DES system requires distribution of keys whereas the PKC system requires a common public library of encryption functions. Currently the high speed operation of DES makes it desirable for most applications while the PKC has been proposed as a key distribution system for the DES system.

2.1.3 Channel Coding and Interleaving

To achieve high reliability, channel coding may be required. This, of course, depends on the bit error probability requirements. Typically for 64K bps PCM digitized voice only 10^{-3} bit error probabilities are required whereas computer messages may require 10^{-8} or lower bit error probabilities. For some channels with memory an interleaver may be required to aid the coding technique. Most common codes are effective for independent channel disturbances and the interleaving/deinterleaving process helps convert dependent channel disturbances to independent ones.

2.1.4 Multiplexing

Other users may share a common communication link. The multiplexing operation merges several data streams into one data sequence for transmission. Time division multiplexing (TDM) is the most common form.

2.1.5 Modulation and Spectrum Spreading

The modulator converts digital data sequences into waveforms that are suitable for transmission over the communication channel. The modulation waveform bandwidths, bit error rates, and data rates are typical performance measures of modulation techniques.

For military applications wide bandwidth spread spectrum waveforms are often used. Although inefficient in bandwidth utilization these techniques protect the communication system from intentional jamming [4]. Also, it provides protection against multipath in radio links. For these reasons the spread spectrum waveforms have been proposed for packet radios and commercial mobile radios. They are not practical for most other commercial digital communication systems.

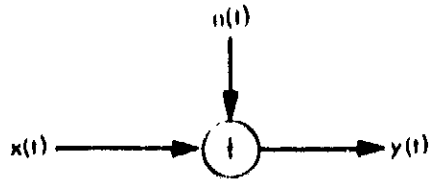
Modulation is necessary for all data transmission systems. Coding techniques work together with modulation to improve overall bit error rates. These two basic parts of the system shown in Figure 1 will be the focus of part 1.

2.2 Communication Channels

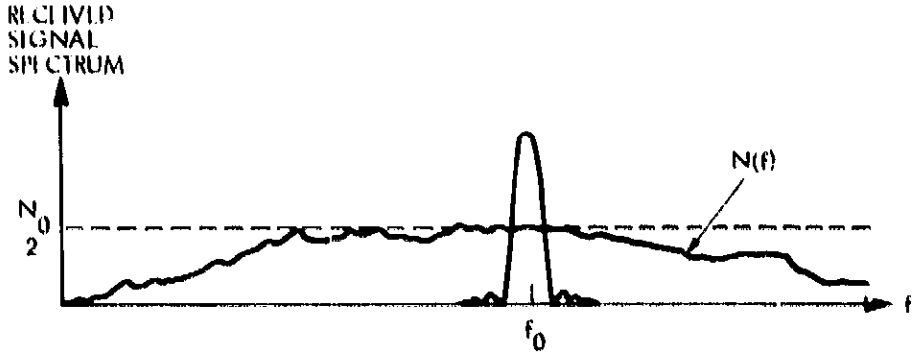
The design and analysis of almost all satellite and terrestrial communication systems are based on the additive white Gaussian noise (AWGN) channel model. This model accurately characterizes the front end receiver noise that always exists due to random electron motions and as such is temperature dependent. Figure 2a shows this ideal channel model which assumes the noise spectrum is approximately flat over the signal waveform bandwidth. Throughout part 1, we assume this ideal AWGN channel model.

In many real communication channels there are various forms of interferences as shown in Figure 2b. Intersymbol interference, multipath, and co-channel interference are most commonly encountered in commercial communication channels. In almost all cases, however, the communication system design is based on the ideal additive white Gaussian noise channel. This is partly due to the fact that designing systems adapted to each specific channel is not practical. A good design should be robust or not too sensitive to specific channel conditions. Coding techniques typically make the overall communication system more robust and able to perform well in a wide range of non-ideal channels.

The satellite channel shown in Figure 3 has an uplink, satellite transponder, and downlink. Typically the uplink is an interference channel as shown in Figure 2b while the downlink has less interference except possibly co channel interference. The satellite transponder is basically an amplifier with appropriate filters. A satellite system usually consists of several such transponders. The SBS 12/14 GHz satellite system, for example, consists of 10 transponders of 10 channels as shown in Figure 4. To avoid interference with each other uplink and downlink channel frequencies are separated. Uplink frequencies are larger to allow more narrow uplink antenna beams at the ground stations so that interference to other geostationary satellites using the same frequencies is minimized.



$$E\{n(t+\tau)\,n(t)\} = \frac{N_0}{2}\,\delta(\tau)$$



$\frac{N_0}{2}$ IS THE DOUBLE-SIDED RECEIVER FRONT END NOISE SPECTRAL DENSITY AND IS TEMPERATURE DEPENDENT

Figure 2a. Mathematical Models for Satellite Channels; The Additive White Gaussian Noise (AWGN) Channel

- SOURCES OF INTERFERENCE
  - Intersymbol, Multipath } Related to x(t)
  - Co-channel
  - RFI (Radar)
  - Jamming
    - Noise (Broadband, Partial Band)
    - CW
    - Multitone
    - Pulse
    - Repeat Back

Figure 2b. Mathematical Models for Satellite Channels; The Interference Channel

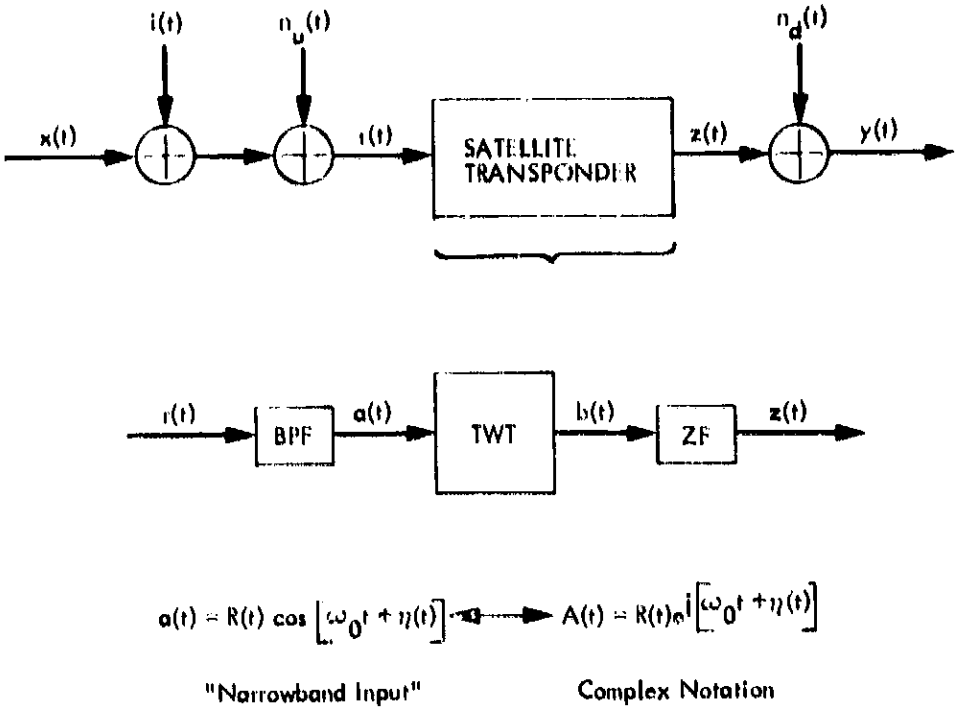


Figure 3. The Transponder Satellite Channel

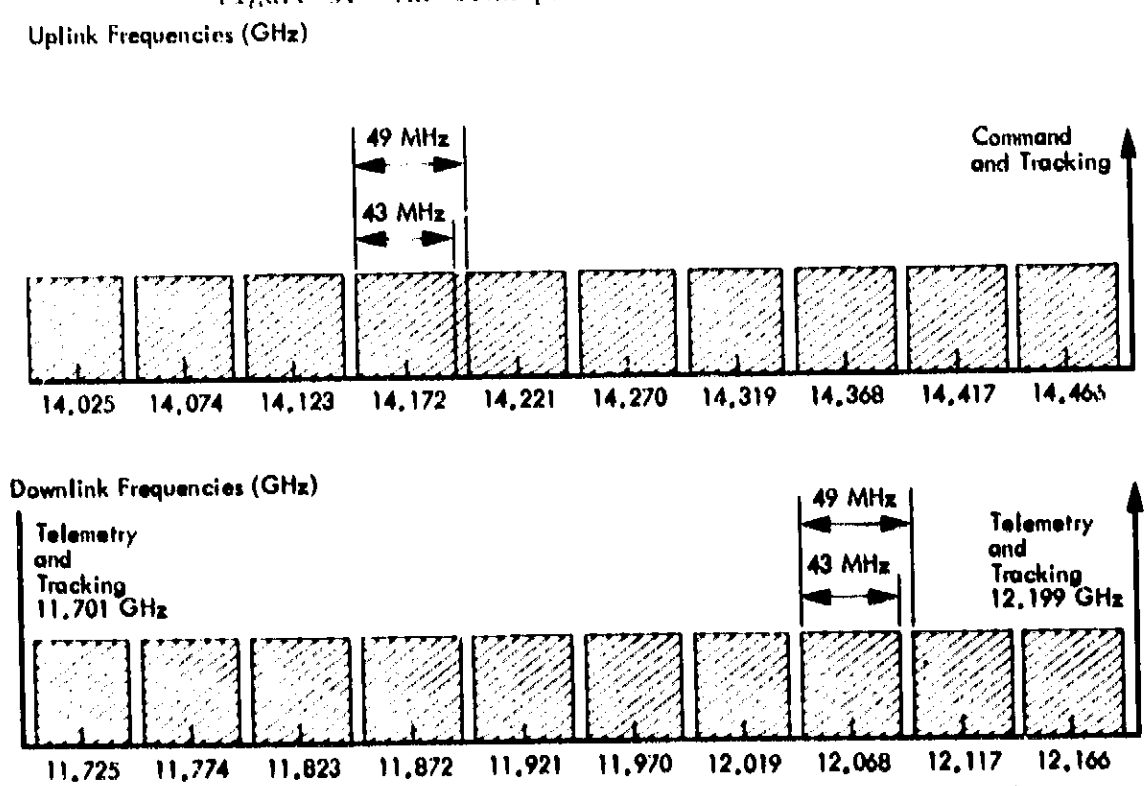


Figure 4. Transponder Frequency Allocations for the SBS 12/14 GHz Satellite System

Each satellite transponder is a nonlinear amplifier that can cause signal distortions. A transponder input of the form

$$x(t) = A(t) \cos [\omega_0 t + \eta(t)] \tag{1.2.1}$$

is amplified at the output as

$$z(t) = f(A(t)) \cos [\omega_0 t + g(A(t)) + \eta(t)] \tag{1.2.2}$$

where

$f(A)$ = AM/AM distortion

$g(A)$ = AM/PM distortion.

Figure 5 shows typical AM/AM and AM/PM characteristics of a transponder. As an illustration, if $x(t)$ consists of two separate frequency tones, then the output $z(t)$ can consist of many intermodulation terms as sketched in Figure 6.

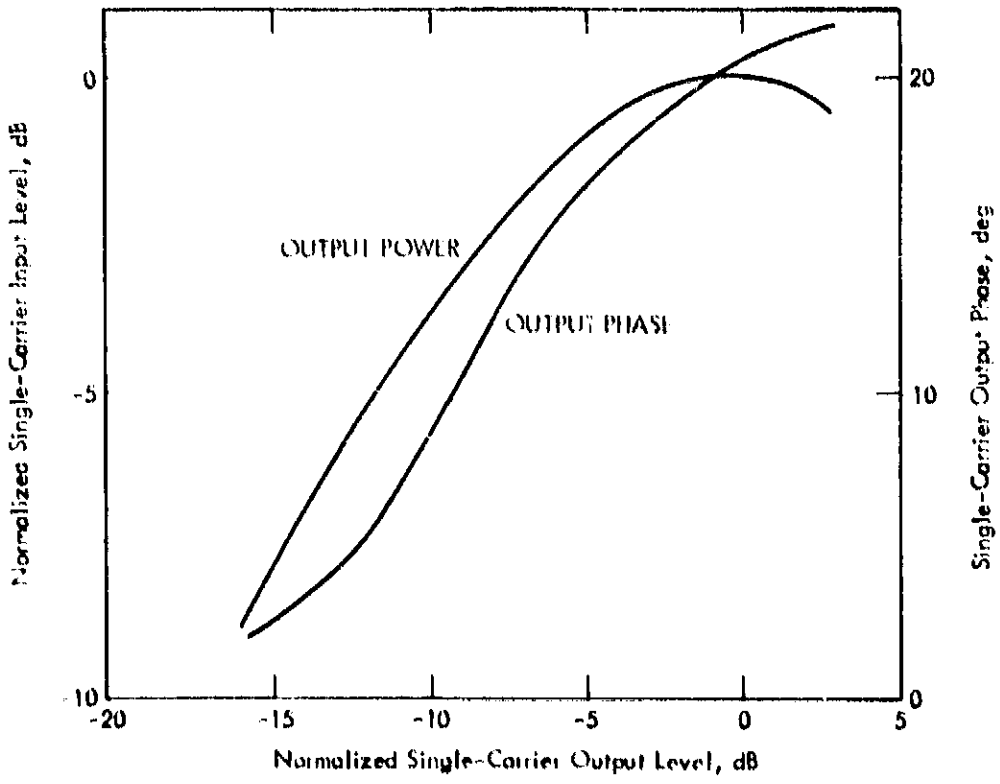


Figure 5. FWT Power and Phase Characteristics

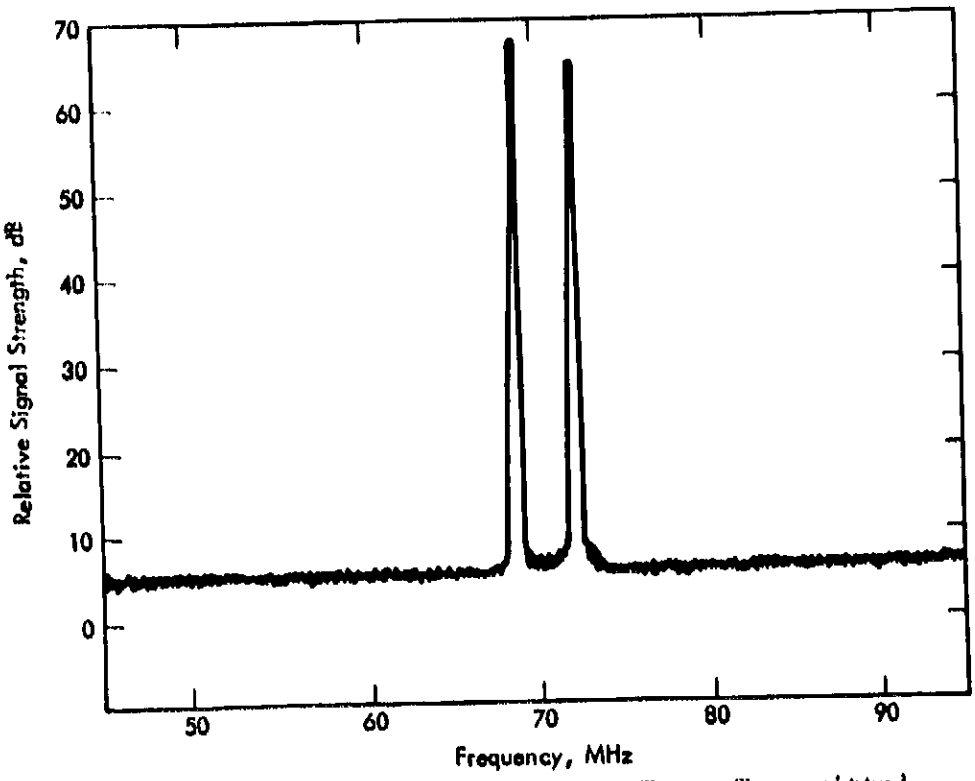


Figure 6a. Power Spectrum of Two Tones Transmitted To The Satellite

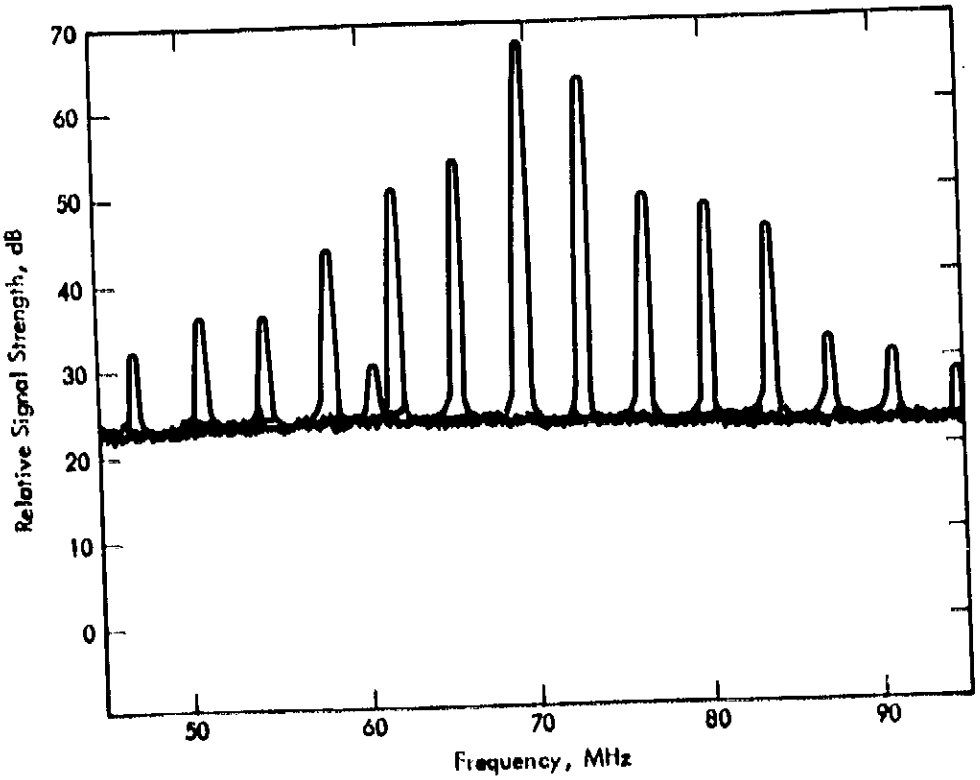


Figure 6b. Power Spectrum of Received Signal at Ground Stations Showing How Intermodulation Has Occurred Between The Two Tones

If A(t) = A is constant then there is no distortion. Hence, for satellite channels most modulation waveforms have constant envelope. If constant envelope signals are not used or if multiple input signals are used in each transponder then the transponder must operate in the less efficient "back-off" mode where the amplifier is more linear and there is less distortion.

For terrestrial links where more power is available, linear amplifiers are more commonly used. If, however, constant envelope signals are transmitted, less expensive nonlinear amplifiers could be used resulting in overall savings. Thus, there is a strong efficiency and cost motivation to use constant envelope signals in both satellite and terrestrial links.

2.3 Bandwidth and FCC Regulations

To avoid interference with other users, digital radio signals are required to stay within an assigned bandwidth. Suppose the transmitted signal is

$$s(t) = A(t)\cos[\omega_0 t + \theta(t)] \tag{1.2.3}$$

where $f_0 = \omega_0/2\pi$ is the carrier frequency in Hz. The signal power spectrum is

$$S_s(f) = \lim_{T\to\infty}\frac{2}{T}E\left\{\left|\int_0^T s(t)e^{-j2\pi ft}dt\right|^2\right\} \tag{1.2.4}$$

where $E\{\cdot\}$ is the expectation over all possible sequences. Here the mean power in the frequency interval $[f_1,f_2]$ is given by

$$\int_{f_1}^{f_2} S_s(f)df \tag{1.2.5}$$

and the total mean power is

$$P = \int_0^\infty S_s(f)df \tag{1.2.6}$$

It is convenient to define a normalized signal power spectral density.

$$S_0(f) = \frac{1}{PT_s} S_x\left(\frac{f}{T_s} + f_0\right) \qquad (1.2.7)$$

where T_s is the signal pulse symbol duration. The Federal Communication Commission (FCC) places a requirement on radio signals when it authorizes bandwidth for commercial communication within the United States. These requirements are described here in terms of this normalized signal power spectral density.

If the authorized bandwidth is $B = 2W$ then in general

$$\int_{-WT_s}^{WT_s} S_0(f)df = 2\int_{0}^{WT_s} S_0(f)df \geq 0.99 \qquad (1.2.8)$$

is required. In addition, assuming the signal power spectral density is essentially constant for any 4KHz band, we have the following requirements:

2.3.1 $f_0 \leq 15$ GHz

For carrier frequencies below 15 GHz there is a maximum of 80 dB attenuation required outside of the authorized bandwidth. A general weaker requirement is that $10 \log_{10} S_0(f)$ does not have to be below

$$-116 - 10 \log_{10} T_s$$

for f outside the authorized bandwidth. The primary requirement is the rate of out-of-band drop in power given by

$$10 \log_{10} S_0(f) \leq -11-40\left[\frac{f}{WT_s} - 1\right] - 10 \log_{10} 2WT_s \qquad (1.2.9)$$

for all $f > WT_s$.

2.3.2 $f_0 > 15$ GHz

Above 15 GHz there is never more than 56 dB attenuation required out of band and in fact $10 \log_{10} S_0(f)$ again does not have to be below

$$-116 - 10 \log_{10} T_s.$$

Here the required rate of out-of-band drop in power is given by

$$10 \log_{10} S_0(f) \leq -11 - 20 \left[\frac{f}{WT_s} - 1 \right] - 10 \log_{10} 2WT_s \qquad (1.2.10)$$

for all $f \geq WT_s$.

Most modulations produce signals that do not meet the FCC regulations and must therefore be filtered before transmission, typically at radio frequencies (RF). RF filtering is generally complicated, costly, and adds distortions and intersymbol interference to the transmitted signal. Even with a constant envelope modulation, RF filtering will result in signals with amplitude variations. If such a signal is followed by a nonlinear amplifier, distortions such as those described earlier can occur and result in amplifier output signals which do not meet the FCC regulations. This can be avoided by using highly linear amplifiers which again is more costly. Thus, meeting bandwidth requirements can add considerable cost to a system due to RF filtering and linear amplifier designs. Figure 7 shows a typical example of the FCC mask superimposed on a given $S_0(f)$.

2.4 Performance Measures

Assuming that a communication channel with authorized bandwidth is available, and the bandwidth requirements are met, then the key technical performance parameters of a satellite or terrestrial communication system are

R = data rate in bits/second

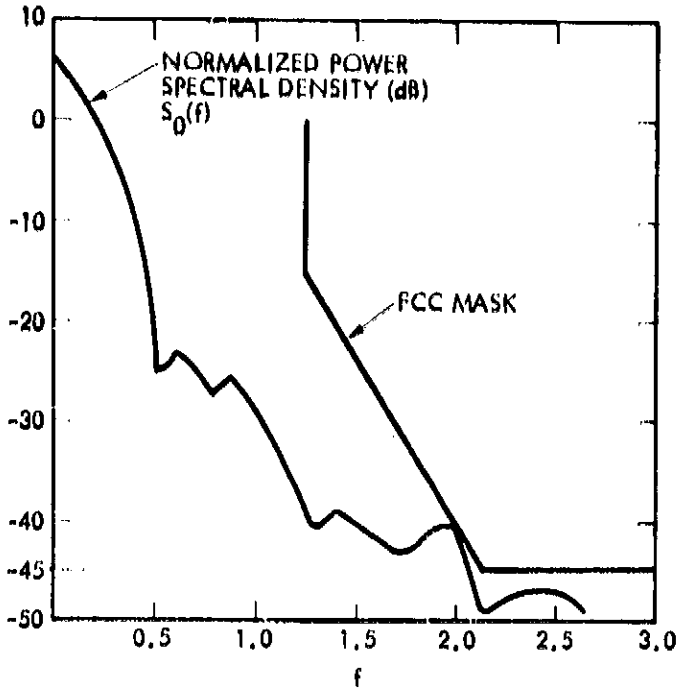


Figure 7. The FCC Mask Superimposed on a Given Normalized Power Spectral Density

and

P_b = bit error probability.

The relationship between R and P_b is very complex and depends on the choice of modulation and coding. Generally, higher data rates imply lower bit error probabilities.

Among those modulation and coding techniques that meet technical requirements, cost of implementation will determine the final choice. Cost is a time-varying parameter which is a function of available technology. Digital processing, for example, has become faster, less expensive, and more powerful and will have an increasingly important impact on modulation and coding implementations.

2.3.2 $f_0 > 15$ GHz

Above 15 GHz there is never more than 56 dB attenuation required out of band and in fact $10 \log_{10} S_0(f)$ again does not have to be below

$$-116 - 10 \log_{10} T_s.$$

Here the required rate of out-of-band drop in power is given by

$$10 \log_{10} S_0(f) \leq -11 - 20 \left[\frac{f}{WT_s} - 1 \right] - 10 \log_{10} 2WT_s \qquad (1.2.10)$$

for all $f \geq WT_s$.

Most modulations produce signals that do not meet the FCC regulations and must therefore be filtered before transmission, typically at radio frequencies (RF). RF filtering is generally complicated, costly, and adds distortions and intersymbol interference to the transmitted signal. Even with a constant envelope modulation, RF filtering will result in signals with amplitude variations. If such a signal is followed by a nonlinear amplifier, distortions such as those described earlier can occur and result in amplifier output signals which do not meet the FCC regulations. This can be avoided by using highly linear amplifiers which again is more costly. Thus, meeting bandwidth requirements can add considerable cost to a system due to RF filtering and linear amplifier designs. Figure 7 shows a typical example of the FCC mask superimposed on a given $S_0(f)$.

2.4 Performance Measures

Assuming that a communication channel with authorized bandwidth is available, and the bandwidth requirements are met, then the key technical performance parameters of a satellite or terrestrial communication system are

R = data rate in bits/second

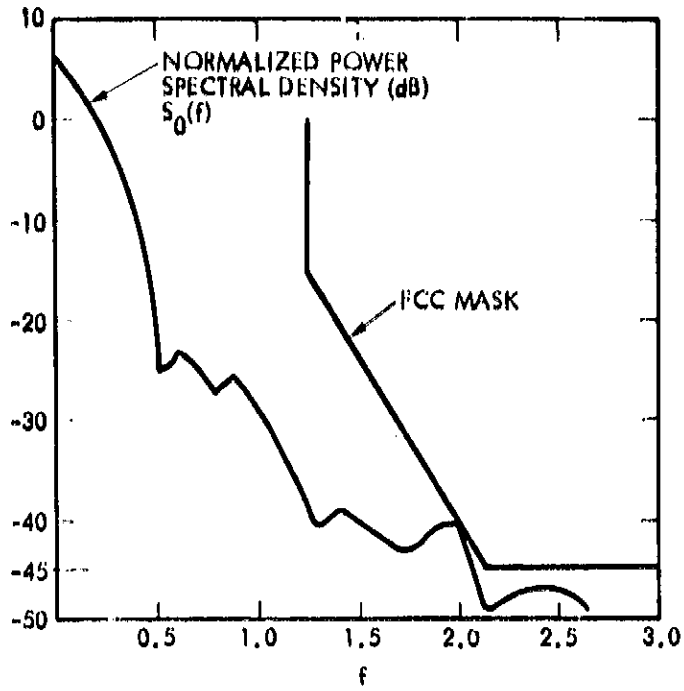


Figure 7. The FCC Mask Superimposed on a Given Normalized Power Spectral Density

and

P_b = bit error probability.

The relationship between R and P_b is very complex and depends on the choice of modulation and coding. Generally, higher data rates imply lower bit error probabilities.

Among those modulation and coding techniques that meet technical requirements, cost of implementation will determine the final choice. Cost is a time-varying parameter which is a function of available technology. Digital processing, for example, has become faster, less expensive, and more powerful and will have an increasingly important impact on modulation and coding implementations.

2.5 Network Considerations

The discussion to this point has concentrated on point-to-point satellite and terrestrial radio communication links. A communication network consists of many such links and a large number of various types of users. Often it is impractical to provide dedicated links to users and many users must share communication links. Most new digital communication networks will do this in a TDMA format with packetized data [5].

The use of packets introduces new requirements on the choice of modulation and coding. A packet of data consists of a burst of symbols of fixed finite length. During the reception of the initial part of the packet, the receiver must estimate

τ = delay

$\Delta\omega$ = oscillator drift

θ = phase

of the packet's modulated signal. The choice of modulation and demodulation design can have an important impact on the number of initial overhead symbols necessary for this purpose of acquiring τ, $\Delta\omega$, and θ.

Since packets are fixed length blocks of symbols, it is natural to consider forward error correction (FEC) block coding techniques rather than the more commonly used convolutional codes with Viterbi decoding which are used for continuous data sequences. Also with packets the use of error detection (ED) and automatic retransmission request (ARQ) is natural to consider. Some combination of FEC and ED/ARQ should be used in most packetized data systems.

In a large digital network where channel links are shared by many users, the choice of modulation and coding techniques must include consideration of how data is to be packaged as a string of modulated/coded symbols.

3.0 Modulation

To avoid distortions and bandwidth expansion caused by amplitude variations in signals through nonlinear amplifiers, constant envelope signals are commonly used for satellite links. For terrestrial radio links where linear amplifiers are easier to develop, multi-amplitude modulation techniques are also used. In telephone cables the most common digital modulation format is the 16-ary quadrature amplitude modulation (QAM) technique.

At the receiver, estimation of the transmitted signal phase adds to implementation costs. Coherent modulations refer to those techniques where the receiver continuously estimates the phase of the received signal using a phase locked loop (PLL) or a Costas loop [6]. With a slight loss in performance one can use differentially coherent modulation techniques where during any given symbol time of T_s seconds the transmitted phase uses the previously transmitted phase as a reference. Hence, the receiver uses the previous T_s second signal phase as a reference in demodulation and does not require any loop for phase estimation. This is easier to implement in a receiver. Noncoherent schemes assume no phase reference and result in the simplest and most robust design but at a cost in performance. Modulations used in military applications are typically noncoherent or differentially coherent since a robust design is more desirable even at a cost of up to 3 dB in performance.

In this section we summarize the various modulation techniques commonly used as well as some new bandwidth efficient modulation techniques that will likely be used in future digital communication systems.

3.1 Coherent Modulation

For commercial applications coherent modulation techniques are most commonly used particularly QPSK modulation for satellite and terrestrial radio channels and 16-ary QAM for telephone channels. QAM is also used in some terrestrial radio applications.

3.1.1 M-ary Phase Shift Keying (MPSK)

M-ary phase shift keying (MPSK) consists of taking K data bits and converting these bits to one of $M = 2^K$ phase angles $0, 2\pi/M, 2(2\pi/M), \ldots, (M-1)(2\pi/M)$.

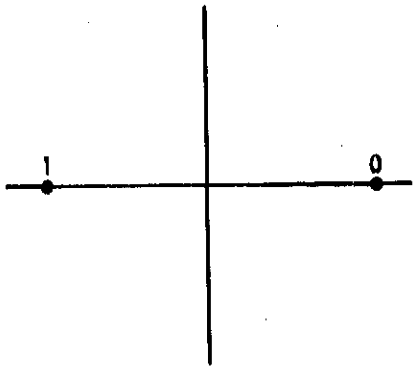
This conversion from bits to phase angles is done using Gray coder [7] in accordance with the following:

	Data Bits	Phase Angles
M = 2:	0	0
	1	π
M = 4:	00	0
	01	$\frac{1}{2}\pi$
	11	π
	10	$\frac{3}{2}\pi$
M = 8:	000	0
	001	$\frac{1}{4}\pi$
	011	$\frac{1}{2}\pi$
	010	$\frac{3}{4}\pi$
	110	π
	111	$\frac{5}{4}\pi$
	101	$\frac{3}{2}\pi$
	100	$\frac{7}{4}\pi$

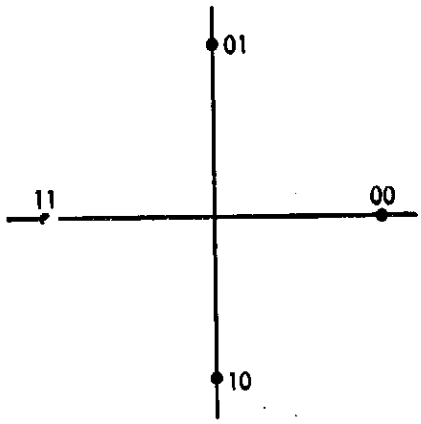
The basic idea is to have only one binary number change in the assignment of adjacent phase angles. This is shown in Figure 8.

The phase Θ is then modulated on a carrier resulting in the MPSK signal of duration T_s seconds,

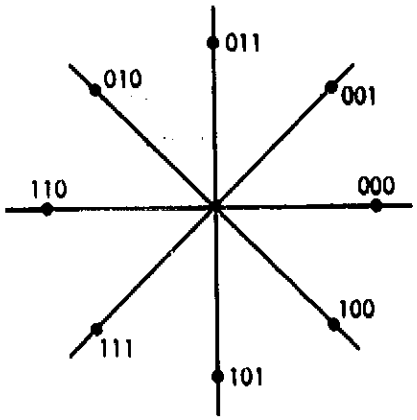
$$x(t) = \sqrt{2S}\cos[\omega_0 t + \theta]; \quad 0 \leq t \leq T_s \tag{1.3.1}$$



a) BPSK, K=1



b) QPSK, K=2



c) 8-PSK, K=3

Figure 8. Data Bit Assignment

This is repeated every T_s seconds so that during the n^{th} interval $(n-1)T_s < t < nT_s$ the n^{th} set of K bits determine the n^{th} modulation phase. For this modulation the data rate is

$$R = \frac{K}{T_s} \text{ bits/second}$$
$$= \frac{\log_2 M}{T_s} \text{ bits/second.} \qquad (1.3.2)$$

For a fixed symbol time T_s and transmitter power P we have the same signal bandwidth for all values of M. Larger values of M means higher data rates but at a cost of higher bit error probabilities. The energy per symbol is

$$E_s = PT_s \qquad (1.3.3)$$

while the energy per bit for $M = 2^K$ is

$$E_b = \frac{E_s}{K} = \frac{PT_s}{K} \qquad (1.3.4)$$

The bit error probability in a white Gaussian noise channel with single-sided noise spectral density N_0 is as follows:*

$$M = 2 \; : \; P_b = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$
$$M = 4 \; : \; P_b = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \qquad (1.3.5)$$

---

*Here we define $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ which is tabulated or can be computed using simple algorithms.

and for larger values of $M = 2^K$ we have the tight upper bounds

$$P_b \leq \frac{1}{K} Q\left(\sqrt{\frac{2KE_b}{N_0}} \sin \left[\pi 2^{-K}\right]\right) \tag{1.3.6}$$

3.1.2 16-Quadrature Amplitude Modulation (QAM)

16-QAM is common in digital telephone cable transmission and appears in some terrestrial digital radios. Recall that MPSK has the form

$$\begin{aligned} x(t) &= \sqrt{2S} \cos \left[\omega_0 t + \theta\right] \\ &= \sqrt{2S} \cos \theta \cos \omega_0 t - \sqrt{2S} \sin \theta \sin \omega_0 t \end{aligned} \tag{1.3.7}$$

This can be viewed as binary amplitude modulation of the quadrature carrier components $\cos \omega_0 t$ and $\sin \omega_0 t$. 16-QAM is a 4-level amplitude modulation of these quadrature components where the signal has the form

$$x(t) = a(t) \cos \omega_0 t + b(t) \sin \omega_0 t \tag{1.3.8}$$

Here $a(t)$ and $b(t)$ each take on one of four levels, i.e.,

$$a(t), b(t) = \{-3A, -A, A, 3A\}$$

during each T_s second interval. Here two data bits determine the amplitude $a(t)$ and another two data bits determine the amplitude $b(t)$ during each symbol time T_s. Hence, the data rate is

$$R = \frac{4}{T_s} \quad \text{bits/second} \tag{1.3.9}$$

Figure 9 shows the 16 possible pairs of quadrature amplitudes in the quadrature space. Unlike MPSK this results in a signal where there are three distinct envelope levels as measured by the radius from the center in Figure 9. Defining E_b as the average energy per data bit, we have the symbol error probability

$$P_s = 3Q_1 - \frac{9}{4} Q_1^2 \tag{1.3.10}$$

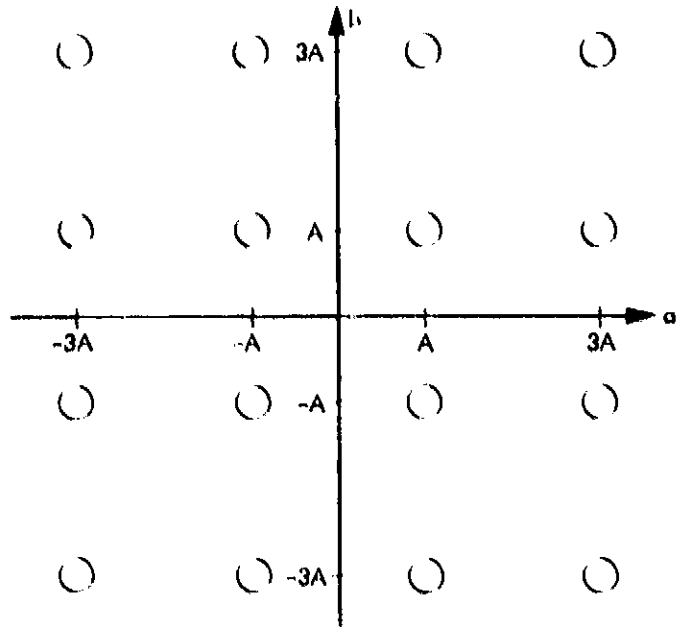


Figure 9. Signal Point Constellation for 16-QAM

where

$$q_m = Q\left(m\sqrt{\frac{4E_b}{5N_0}}\right); \quad m = 1,3,5 \tag{1.3.11}$$

The corresponding bit error probability P_b is given by

$$P_b = \frac{3}{4}q_1 + \frac{1}{2}q_3 - \frac{1}{4}q_5 \tag{1.3.12}$$

The variable envelope of this modulation means linear amplifiers are required to avoid distortions.

1.1.3 Offset Quadrature Modulation

A class of modulation techniques that are less sensitive to channel distortions are the offset quadrature modulations. These modulations are particularly well suited for channels with nonlinearities and are now being proposed for future satellite systems.

Recall that conventional MPSK and QAM can be viewed as amplitude modulations on quadrature components $\cos \omega_0 t$ and $\sin \omega_0 t$. Both quadrature amplitudes modulate during the same T_s second intervals. In offset quadrature modulation, the $\cos \omega_0 t$ amplitudes modulate during the time intervals

$$(n-1)\ T_s \leq t \leq nT_s$$

$$n = \ldots, -1, 0, 1, \ldots.$$

while the $\sin \omega_0 t$ amplitudes modulate during the offset time intervals

$$\left(n-\frac{1}{2}\right) T_s \leq t \leq \left(n+\frac{1}{2}\right) T_s$$

$$n = \ldots, -1, 0, 1, \ldots .$$

This $T_s/2$ second offset or relative delay between the two quadrature modulations results in a signal that experiences less bandwidth expansion or distortion after passing through nonlinearities [8].

The most common form of this modulation is Offset QPSK (OQPSK) where the amplitudes in both quadratures are binary valued ($\pm A$) much like QPSK. In fact, the bandwidths and bit error probabilities for QPSK and OQPSK are the same for the ideal additive white Gaussian noise channel. The primary difference is the impact of nonlinearities on these two modulations [8].

The bandwidth characteristics of the offset and normal quadrature modulations can be modified by using different pulse shapes for the quadrature modulations. QPSK and OQPSK, for example, correspond to using rectangular shaped pulses on each quadrature component. Different pulse shapes, however, can lead to a signal with varying envelope whereas the rectangular pulses result in constant envelope signals. Generally the more envelope variations a signal has the more nonlinear channel distortions it encounters.

Figure 10 illustrates a raised cosine pulse shaping with binary modulation on each offset quadrature pulse. Here the pulse duration is $2T_s$ seconds so some pulse overlap occurs resulting in "intersymbol interference." This is called Staggered Quadrature Overlapped Raised Cosine (SQORC) modulation. Without the offset or staggered quadratures this is simply called Quadrature Overlapped Raised Cosine (QORC) [8,9]. Figure 11 shows the power spectral densities of QPSK and QORC. The power spectral densities are primarily determined by the pulse shaping function and not the offset or staggered property. Hence, QPSK and OQPSK have the same power spectral density as do QORC and SQORC. Figure 12 illustrates the impact of a hard limiter on a filtered QORC signal [10]. This shows that even though QORC signals have a rapidly dropping out of band power spectrum before any nonlinearity, after passing through the nonlinearity this desirable spectral characteristic is destroyed. On the other hand, by using offset or staggered quadrature modulation this desirable spectral characteristic is maintained [10]. This is shown in Figure 13 for SQORC.

Despite the fact that the raised cosine pulses cause envelope variations when the quadrature pulses are offset relative to each other by $T_s/2$ seconds, the resulting signal is less sensitive to channel nonlinearities. For many radio channels, this is a desirable characteristic.

3.1.4 Duobinary Modulation

MPSK and QAM signals do not generally meet required bandwidth constraints such as the FCC mask and thus they are usually filtered. The filtering causes symbol pulses to interfere with each other, causing "intersymbol interference." This causes some distortions of the signal and corresponding degradation in performance. Channels like telephone cables also act like filters and cause considerable signal distortions.

To minimize the impact of channel distortions, the duobinary modulation introduces controlled amounts of intersymbol interference which tend to shape the transmitted signal in a way that minimizes distortions. This modulation technique is sketched in Figure 14. Here the data bits are denoted

$$a_n \in \{-1, 1\} .$$

To avoid error propagation which is a characteristic of this modulation, differential encoding of the data is required. This is described by the equation

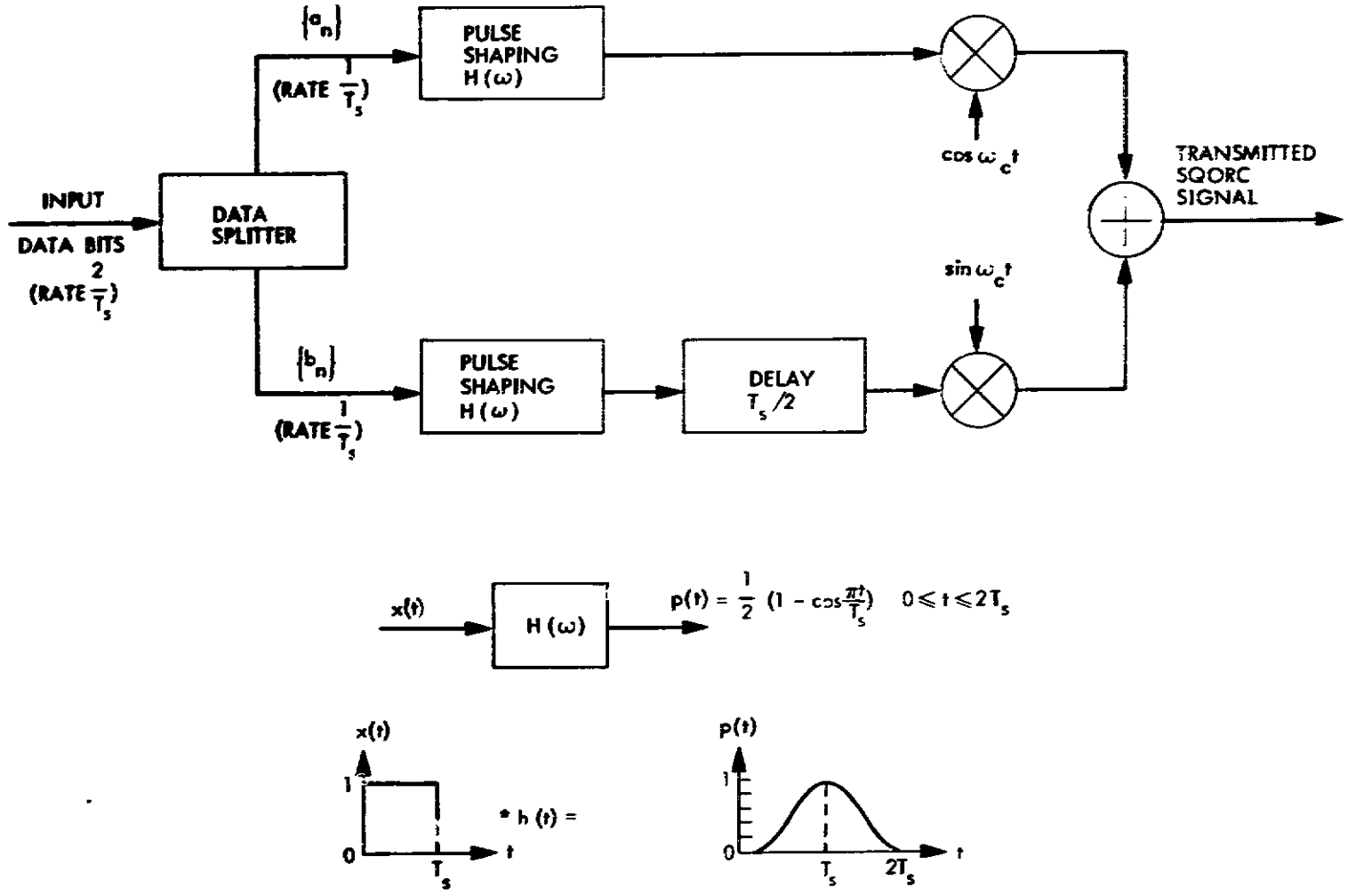


Figure 10. A Staggered Quadrature Overlapped Raised Cosine (SQORC) Modulator

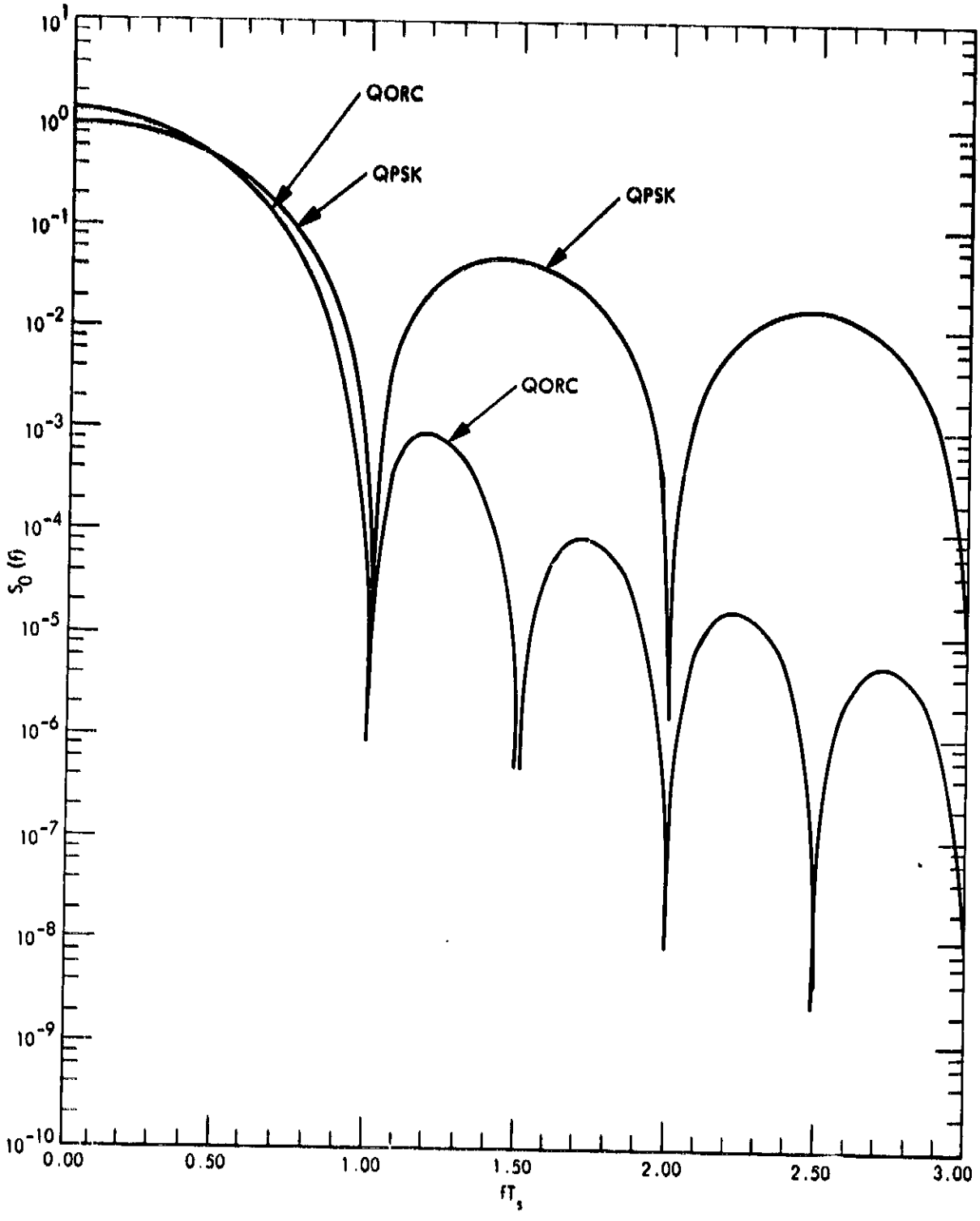


Figure 11. Power Spectral Densities of QPSK and QORC

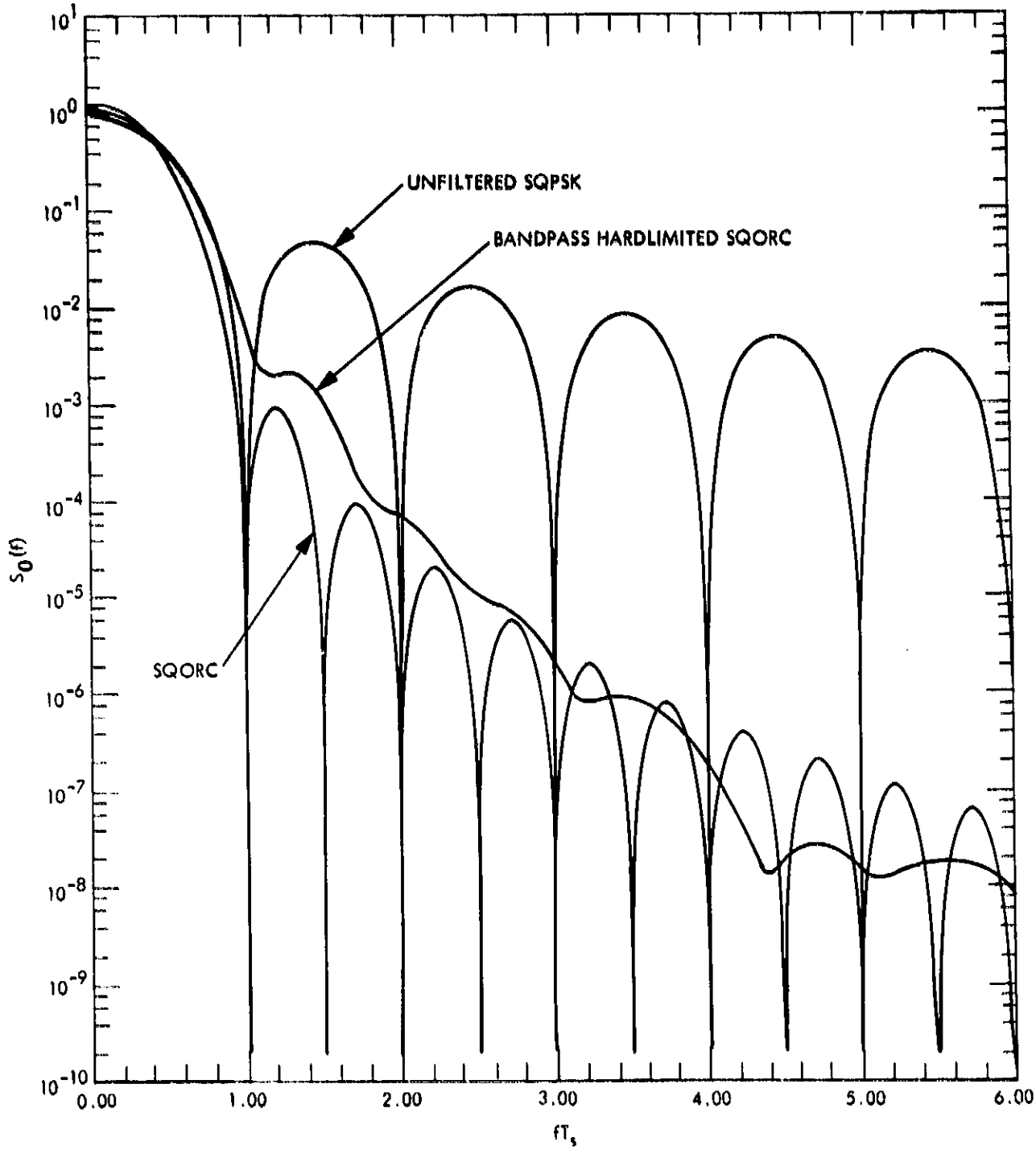


Figure 12. A Comparison of the Power Spectral Density of Several Quadrature Modulation Techniques

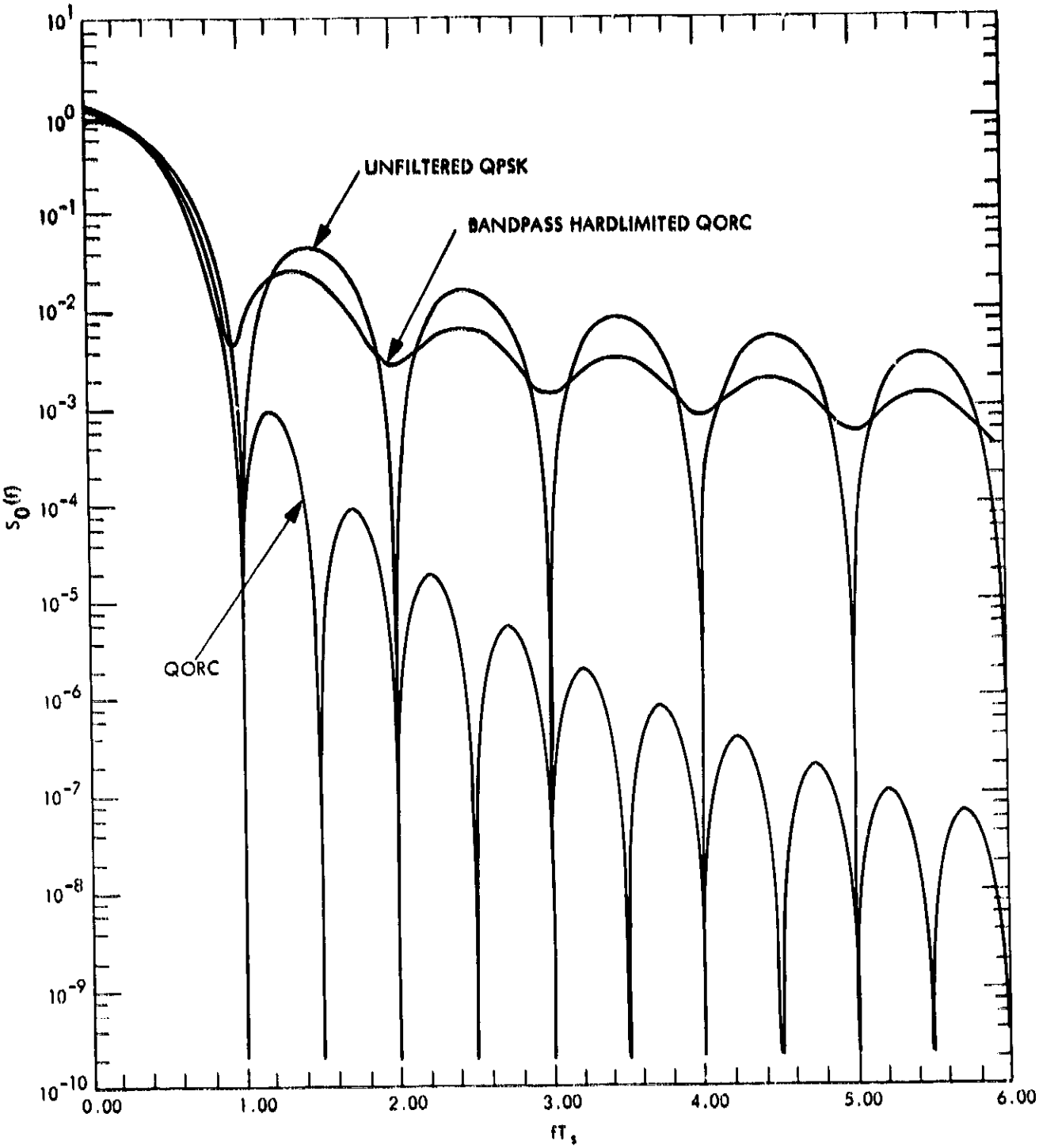


Figure 13. A Comparison of the Power Spectral Density of Several Staggered Quadrature Modulation Techniques

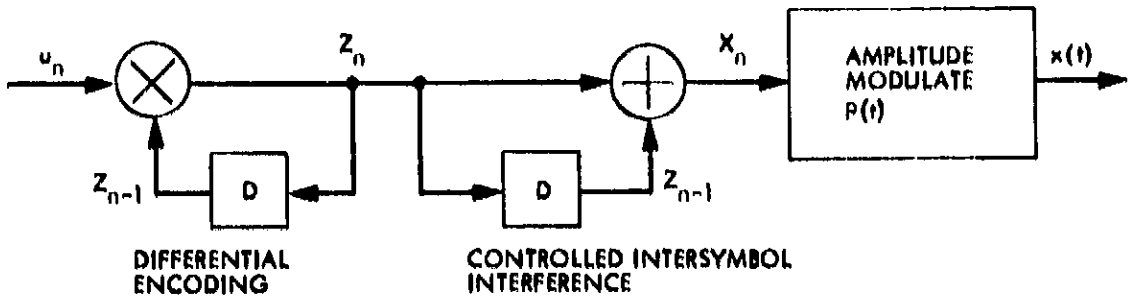


Figure 14. Duobinary Modulation

$$Z_n = u_n Z_{n-1} \in \{-1, 1\} . \tag{1.3.13}$$

Next the controlled intersymbol interference is introduced by generating the amplitudes

$$x_n = Z_n + Z_{n-1} \in \{-2, 0, 2\} \tag{1.3.14}$$

which modulates a carrier pulse p(t) [7].

Since there is only one data bit entering the differential encoder each time an amplitude modulated symbol is generated, the data rate is

$$R = \frac{1}{T_s} \quad \text{bits/second} \tag{1.3.15}$$

The bit error probability (assuming no further channel distortions) is given by

$$P_b = \frac{3}{2} Q\left(\frac{\pi}{4}\sqrt{\frac{2E_b}{N_0}}\right) - \frac{1}{2} Q\left(\frac{3\pi}{4}\sqrt{\frac{2E_b}{N_0}}\right) \tag{1.3.16}$$

where E_b is the average signal energy per pulse.

3.1.5 Continuous Phase Modulation (CPM)

Recently there has been a considerable amount of interest in a class of bandwidth efficient modulation techniques that maintain phase continuity from symbol to symbol [11-19]. This class which we refer to as Continuous Phase Modulation (CPM) includes as special cases Minimum Shift Keying (MSK), Fast Frequency Shift Keying (FFSK), Smoothed Frequency Shift Keying (SFSK), Tamed Frequency Shift Keying (TFM), and many others.

Figure 15 illustrates this class of modulations. CPM is essentially a phase or frequency modulation technique where the data is represented as a binary or higher level square wave which is then passed through a filter with impulse response h(t). The filtered signal is the input to an FM modulator resulting in a constant envelope signal of the form

$$x(t) = \sqrt{2S}\cos\left[\omega_0 t + \theta(t)\right] . \qquad (1.3.17)$$

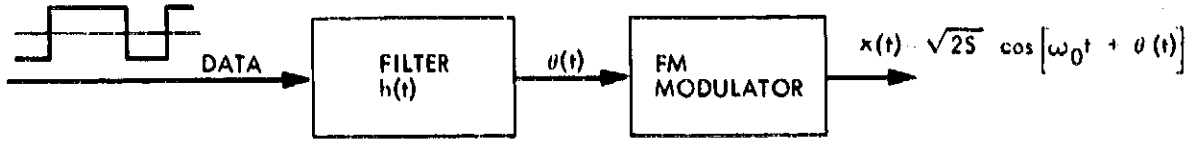


Figure 15. A Block Diagram Illustration of the Class of Continuous Phase Modulations

where

$$\dot{\theta}(t) = \sum_{n=-\infty}^{\infty} a_n \, p\left(t - nT_s\right) \tag{I.3.18}$$

with

$$p(t) = \int_0^{T_s} h(t - \alpha)d\alpha \triangleq h\psi(t) \tag{I.3.19}$$

representing the convolution of the filter impulse response with a unit square, $\{a_n\}$ the data sequence, and h the modulation index. Generally the gain of p(t), which is proportional to the modulation index, determines the main lobe of the resulting signal power spectral density while the "smoothness" of p(t) determines how rapidly it drops outside the signal band.

To illustrate CPM with an example, consider unfiltered binary data, in which case, p(t) is the rectangular pulse

$$p(t) = \begin{cases} \frac{\pi}{T_s} h \ , & 0 \le t \le T_s \\ 0 \ , & \text{elsewhere} \end{cases} \tag{I.3.20}$$

and $T_s = T_b$. The set of all possible phase sequences $\theta(t)$ corresponding to all possible data sequences is illustrated in Figure 16. MSK is the special case of this example corresponding to the modulation index h = .5. MSK is known to have the same* bit error probability, P_b, as BPSK, QPSK, or OQPSK [see (I.3.5)]. For various other choices of modulation indices and M-ary data symbols, we have the signal power spectral densities shown in Figure 17. Larger values of M tend to smooth out the signal spectral variations.

An obvious choice for the pulse p(t) is the raised cosine pulse which is sketched in Figure 18a for 6 different pulse durations. These pulses result in a smoother phase function $\theta(t)$ as illustrated in Figure 18b for the binary data sequence (M = 2) and the raised cosine pulse that spans two symbol times, $2T_s$. The raised cosine pulses result in a more rapid rolloff of out-of-band signal

*More precisely, the input data a_n must first be precoded by a differential decoding operation in order to produce the performance given in (I.3.21).

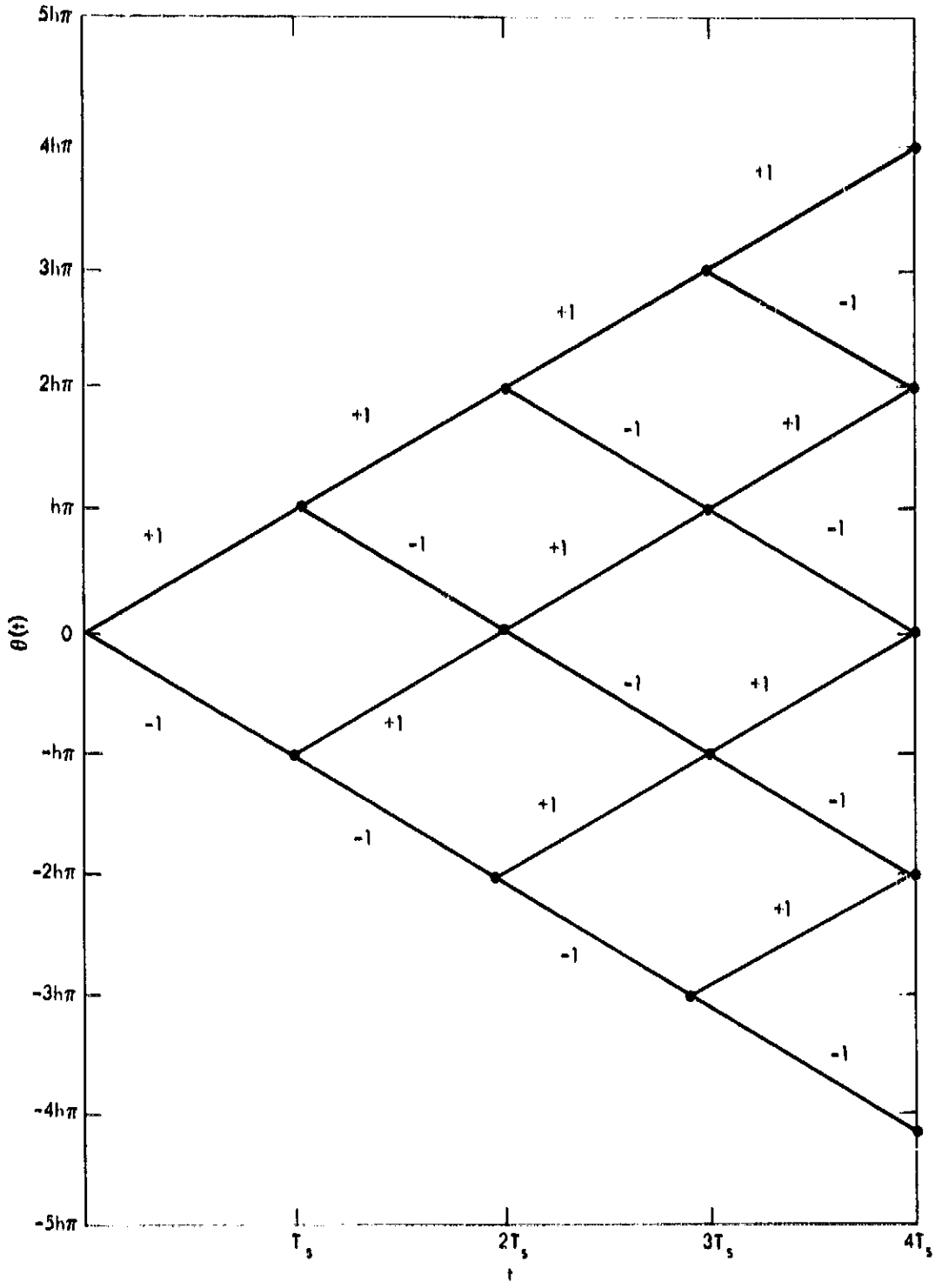


Figure 16. Phase trajectories for a binary full response CPM system. Four bit time intervals are shown. (Reprinted from [12])

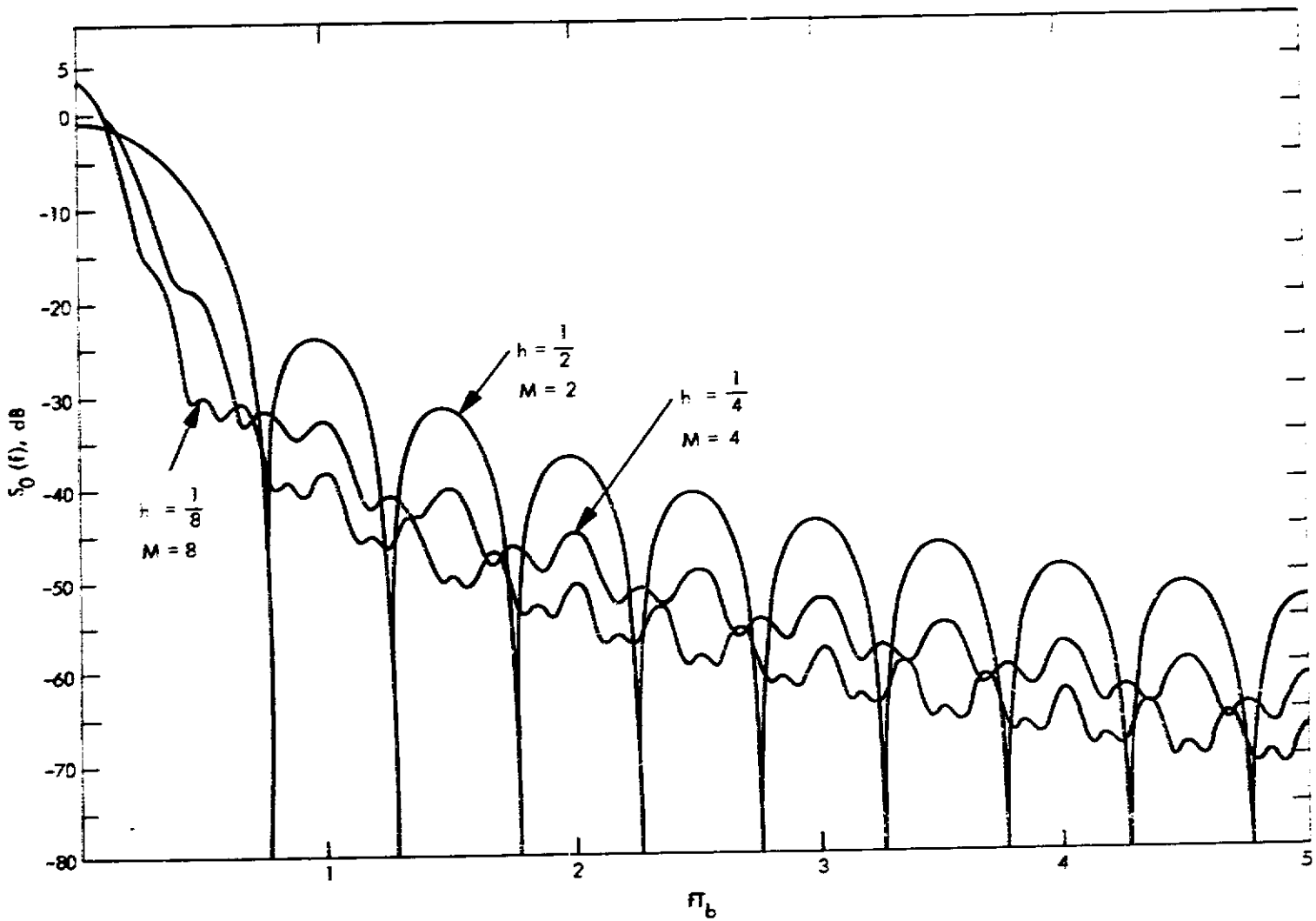


Figure 17. Normalized power spectral density in dB versus normalized frequency for M-ary (M = 2,4 and 8) CPM with modulation indices h=1/2, 1/4 and 1/8 respectively. (Reprinted from [14])

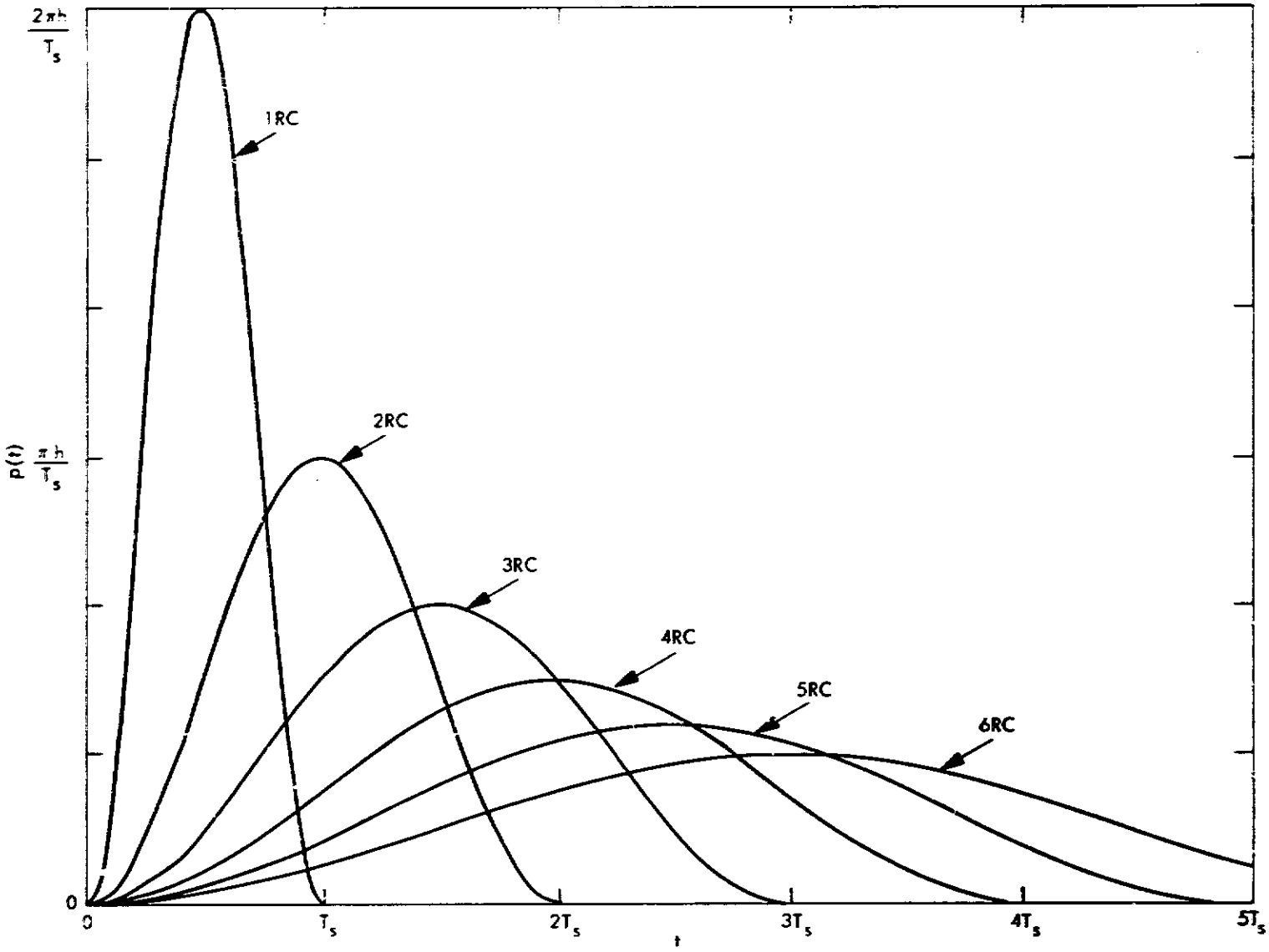


Figure 18a. Frequency pulses p(t) for the Raised Cosine class of CPM systems. (Reprinted from [14])

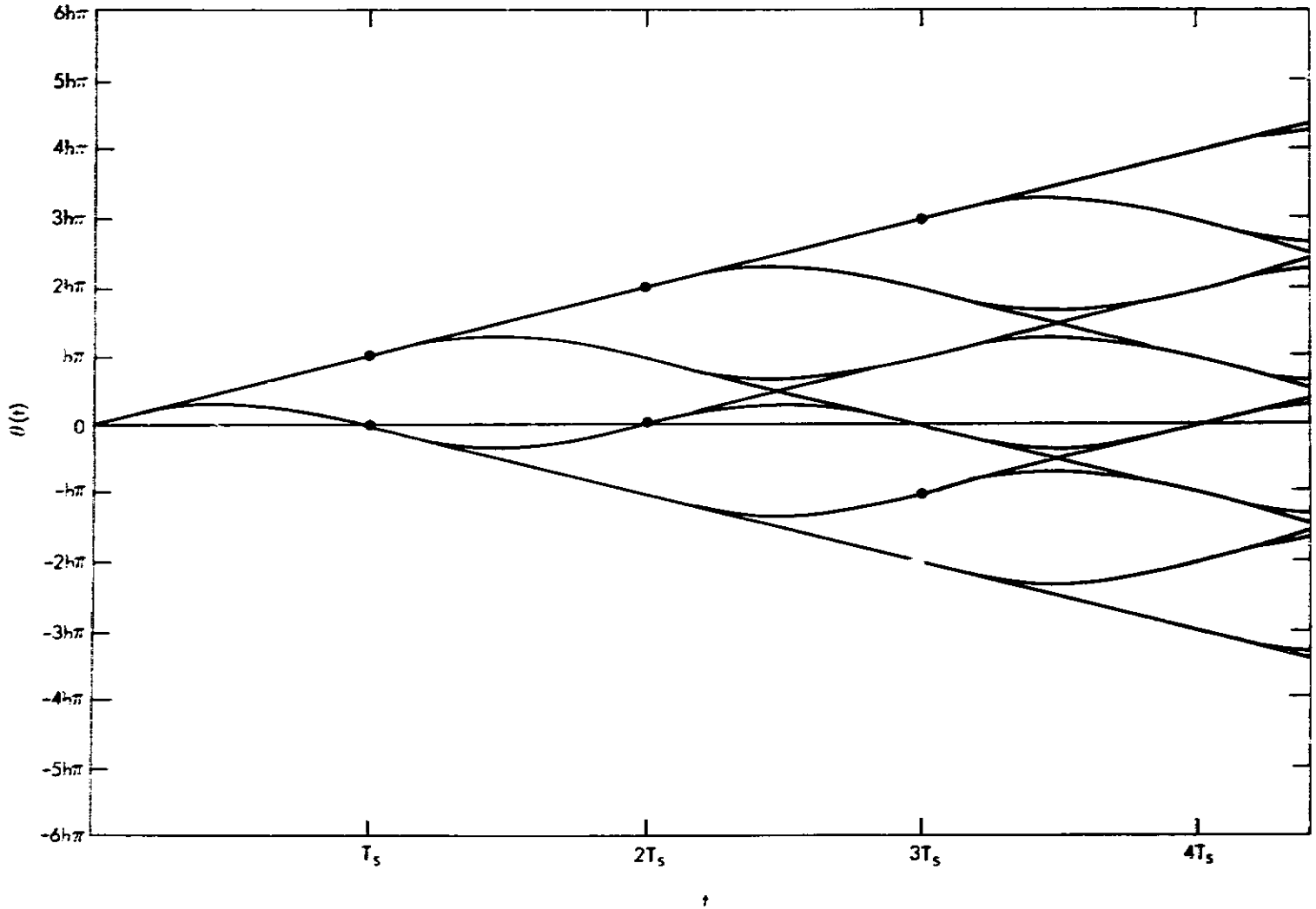


Figure 18b. Binary phase tree for 2RC pulses. (Reprinted from [14])

power spectral density. Figures 19 and 20 present some examples of this. In Figure 19 we see that for fixed modulation index h = .5, longer memory pulses result in smoother and more rapidly dropping signal power spectral density out of the signal band. Figure 20 shows the effects of different modulation indices for a raised cosine pulse that spans $4T_s$. Here we see the increase in width of the main lobe with an increase in the modulation index. Figure 21 has the same type of results for M = 4 and a $3T_s$ raised cosine pulse duration.

These new bandwidth efficient modulation techniques called CPM have the advantages:

(a) They are constant envelope signals and can be used with inexpensive nonlinear amplifiers.

(b) Their power spectral densities can be shaped to meet bandwidth constraints such as the FCC regulations without further RF filtering which is generally expensive.

(c) Together with coding, CPM can achieve [17,18] excellent performance measured in R bits/second and bit error probability P_b.

The key drawbacks of CPM are:

(a) Phase estimation at the receiver can be difficult.

(b) The demodulator must use complex digital signal processing in the form of a Viterbi algorithm.

Omura and Jackson [17] have shown that phase estimation can be done efficiently using a generalized Viterbi algorithm that simultaneously estimates carrier phase and the data. This is a generalization of the use of the Viterbi algorithm for phase estimation [20-27]. As for the complexity of the digital signal processing, it is obvious that these costs are rapidly declining. The whole communication industry has felt the tremendous impact that less expensive digital signal processing capabilities have had on reducing system complexity.

3.1.6 Orthogonal Signals

A less common form of modulation is to use orthogonal signals. The usual way to construct orthogonal signals is to use different frequency tones. For example, suppose that for each K data bits we send one of $M = 2^K$ tones of duration

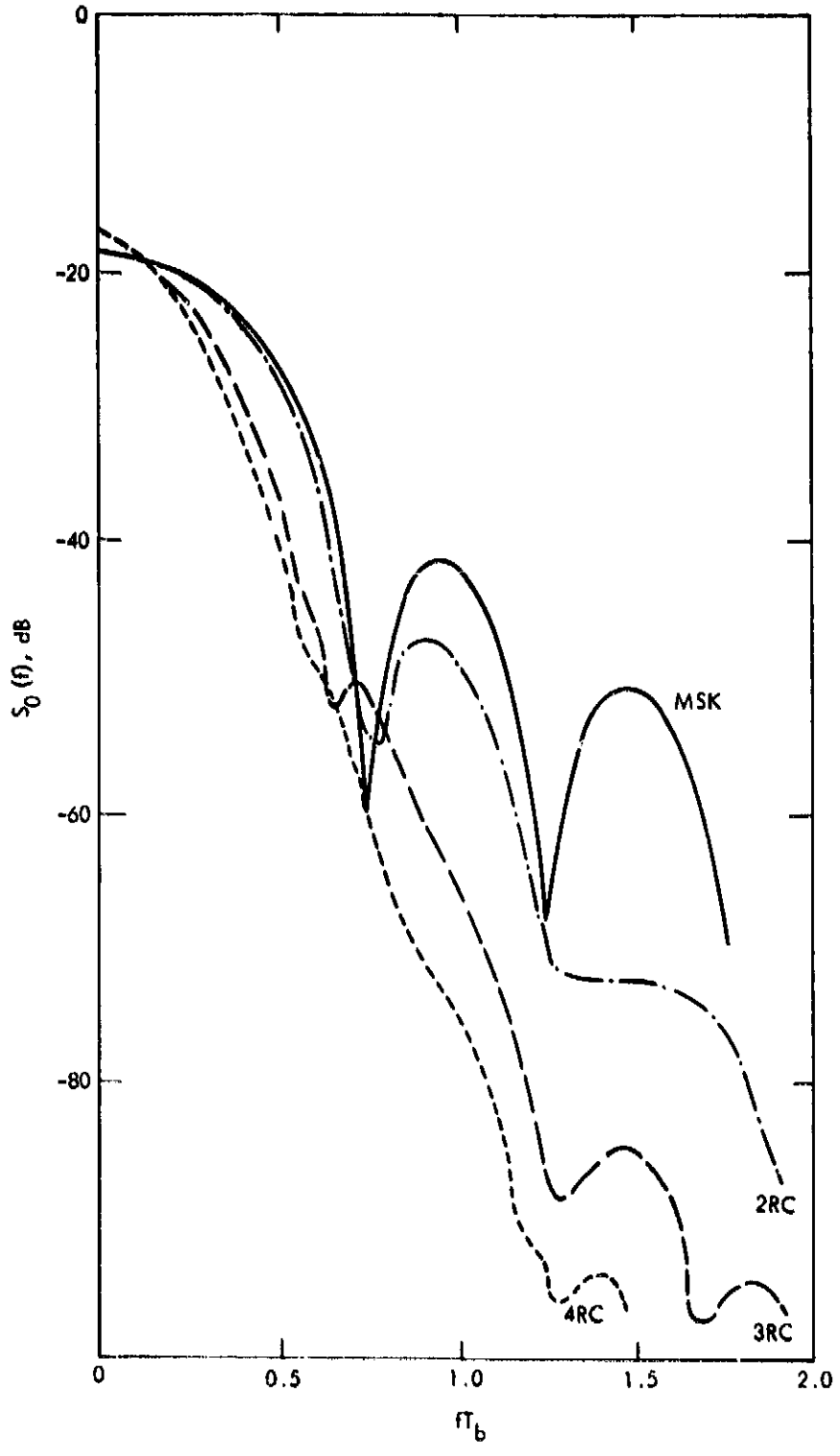


Figure 19. Normalized power spectra for binary CPM schemes with various baseband pulses; h=0.5. (Reprinted from [14])

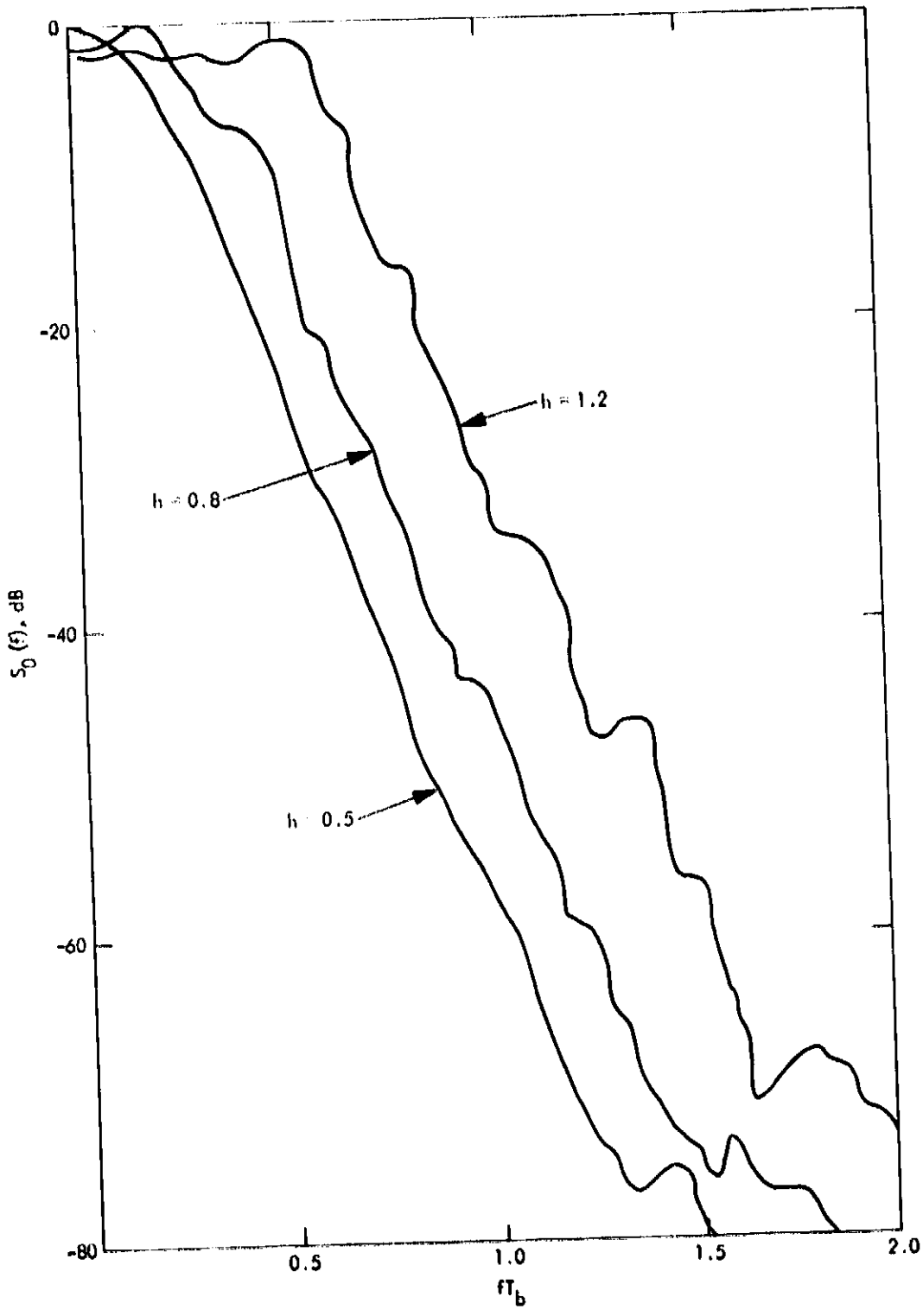


Figure 20. Normalized power spectra for M=2, 4RC pulses; h=0.5, 0.8 and 1.2. (Reprinted from [13])

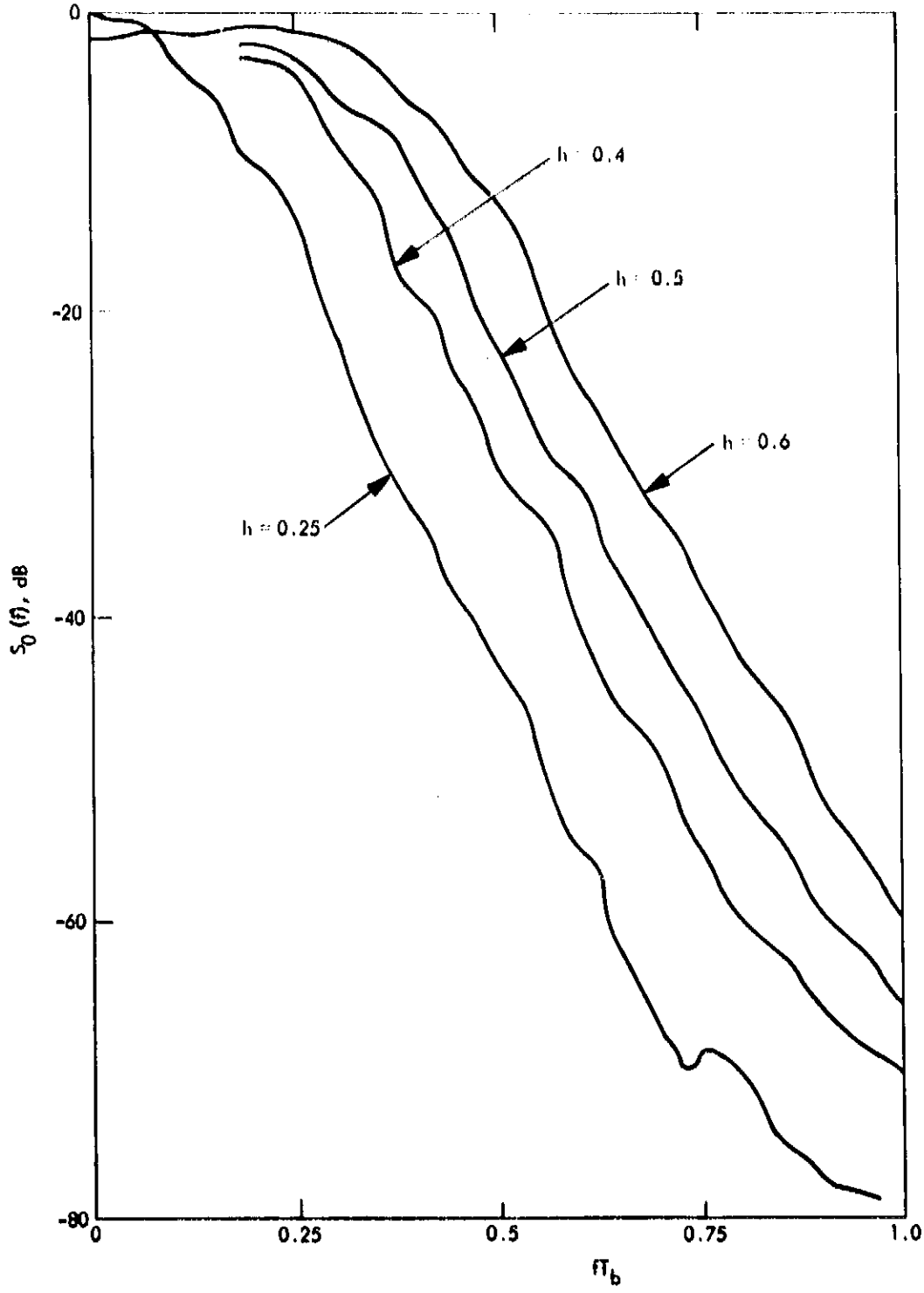


Figure 21. Normalized power spectra for M=4, 3RC pulses. (Reprinted from [13])

T_s seconds. The transmitted signal is then

$$x(t) = \sqrt{2S} \cos \omega t \tag{1.3.21}$$

where

$$\omega \in \left\{ \omega_0 , \omega_0 + \frac{\pi}{T_s} , \omega_0 + 2 \frac{\pi}{T_s} , \ldots , \omega_0 + (M-1) \frac{\pi}{T_s} \right\} \tag{1.3.22}$$

Here $\Delta\omega = \frac{\pi}{T_s}$ is the smallest frequency separation that guarantees the tones are all orthogonal to each other where

$$\int_0^{T_s} \cos \omega_k t \cos \omega_\ell t \, dt = 0 \qquad \ell \neq k \tag{1.3.23}$$

where

$$\omega_k = \omega_0 + k \frac{\pi}{T_s} \qquad k = 0,1,2,\ldots, M-1. \tag{1.3.24}$$

The bit error probability is tightly upper bounded by

$$P_b \leq 2^{K-1} \left(Q \sqrt{\frac{KE_b}{N_0}} \right) \tag{1.3.25}$$

Note that as long as*

$$\frac{E_b}{N_0} > 2 \ell n \, 2 \tag{1.3.26}$$

---

*This can be seen from the bound

$$Q(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}} .$$

this bound converges to zero as K increases. Hence, one would expect this to be a good modulation technique. Unfortunately, the signal bandwidth also increases with K. Note that

$$\Delta f = \frac{\Delta\omega}{2\pi}$$

$$= \frac{1}{2T_s} \qquad (I.3.27)$$

is approximately the bandwidth occupied by each tone. Since there are $M = 2^K$ tones the total bandwidth W occupied by this orthogonal signal set is

$$W = M\,\Delta f$$

$$= \frac{2^K}{2T_s}$$

$$= \frac{2^{K-1}}{T_s} = \frac{2^{K-1}}{KT_b} \qquad (I.3.28)$$

This total bandwidth grows rapidly with K.

One might wonder if there are other orthogonal signals whose total bandwidth does not grow as rapidly with K. In general given bandwidth W and time T_s the number of orthogonal signals of bandwidth W and duration T_s seconds is

$$M = 2WT_s. \qquad (I.3.29)$$

This is also satisfied by our set of orthogonal frequency tones. Hence, we cannot do any better in terms of bandwidth.

3.1.7 Summary of Coherent Modulation Techniques

To compare various modulation techniques we should consider the performance parameters data rate and bit error probability; the latter being a function of the energy per data bit-to-noise spectral density ratio. These would be the key performance parameters if there were virtually no bandwidth limitations. This

situation occurs in deep space communication and most spread spectrum military communications. Generally in commercial applications there are bandwidth constraints which must be considered. When strict bandwidth limitations such as the FCC regulations are imposed, the CPM techniques appear to be the most attractive. Currently, however, RF filtered MPSK, QAM, and some variations of duobinary modulations are the most common forms of coherent modulation techniques.

Keeping in mind that bandwidth limitations and data rate are important considerations, we next show the bit error probabilities for various coherent modulations. Figure 22 shows the bit error probability as a function of E_b/N_0 in dB for orthogonal signaling where $M = 2^K$ is the number of orthogonal tones. In Figure 23, we have the symbol error probabilities for coherent MPSK indicated by the solid lines. The dotted lines are for M-ary differentially coherent MPSK (DMPSK) which we shall consider in the next section. The duobinary modulation has a bit error probability that is nearly the same as the symbol error probability for MPSK with M = 4.

To examine the bandwidth efficient CPM techniques, one must take into consideration the bandwidth of the signals. If we define the 99% bandwidth W as in (1.2.8) with the equality sign, then we can compare all the CPM techniques with each other. In general, we shall use the particular CPM, namely MSK, as a baseline for comparison. As already mentioned previously, MSK is known to have essentially the same bit error probability as BPSK, which is shown in Figure 23 as the M = 2 line. Coherent BPSK is the most commonly used baseline.

Since it is analytically difficult to exactly calculate the error probability performance for coherent receivers of CPM, it is common to make the performance comparison on the basis of minimum squared Euclidean distance over say N_β symbol intervals normalized by the symbol energy [12]. In mathematical terms, we evaluate the minimum of

$$\frac{D^2}{2E_s} = \sum_{i=n}^{n+N_\beta+1} \int_{iT_s}^{(i+1)T_s} \left[x\left(t;a^{(1)}\right) - x\left(t;a^{(2)}\right) \right]^2 dt \qquad (1.3.30)$$

with respect to the two arbitrary but different sequences $a^{(1)}$ and $a^{(2)}$ where $x\left(t;a^{(1)}\right)$ and $x\left(t;a^{(2)}\right)$ respectively represent x(t) of (1.3.17) for

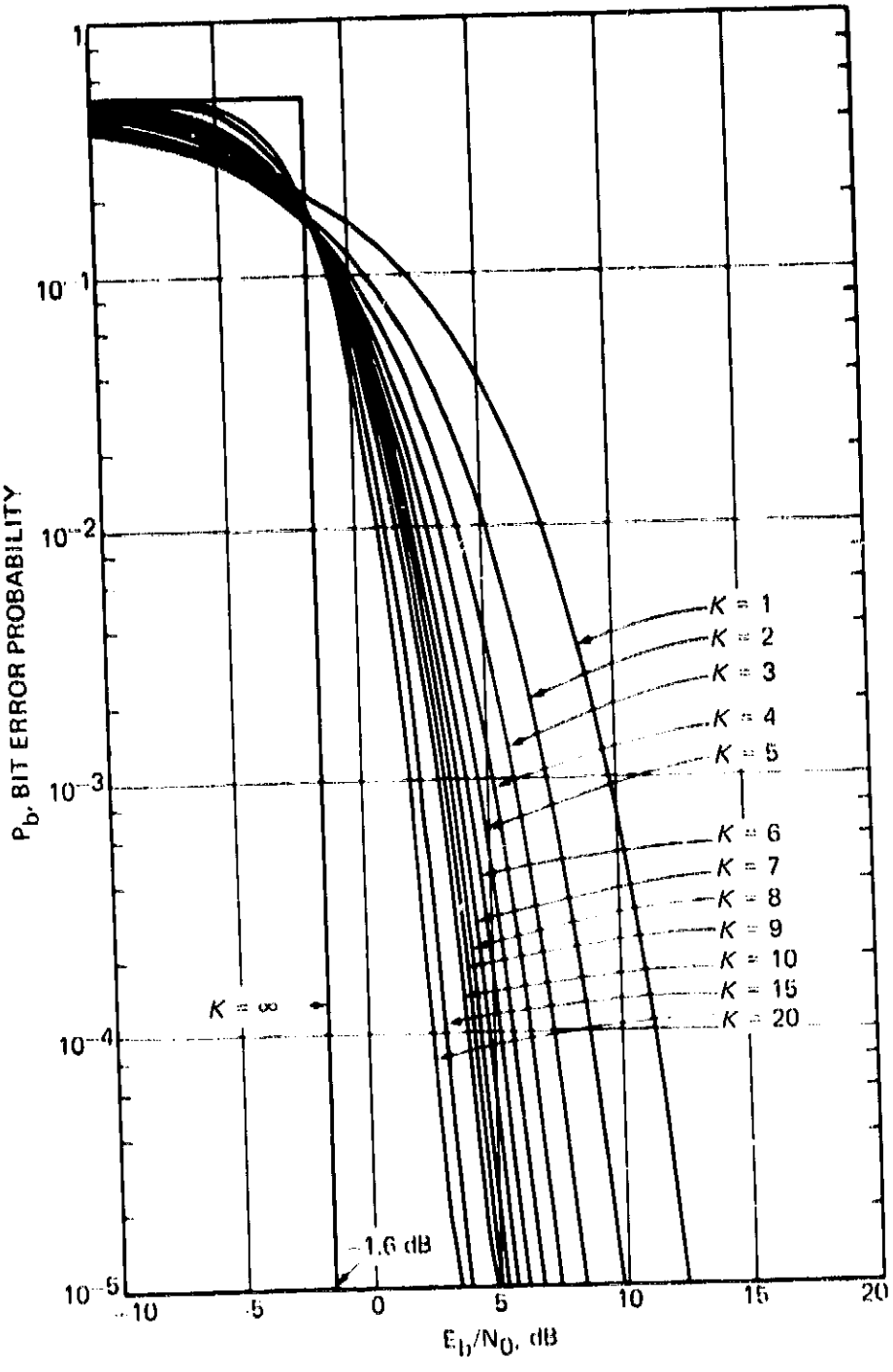


Figure 22. Bit Error Probability Performance of a System Transmitting an Orthogonal Signal Set (Reprinted from [28])

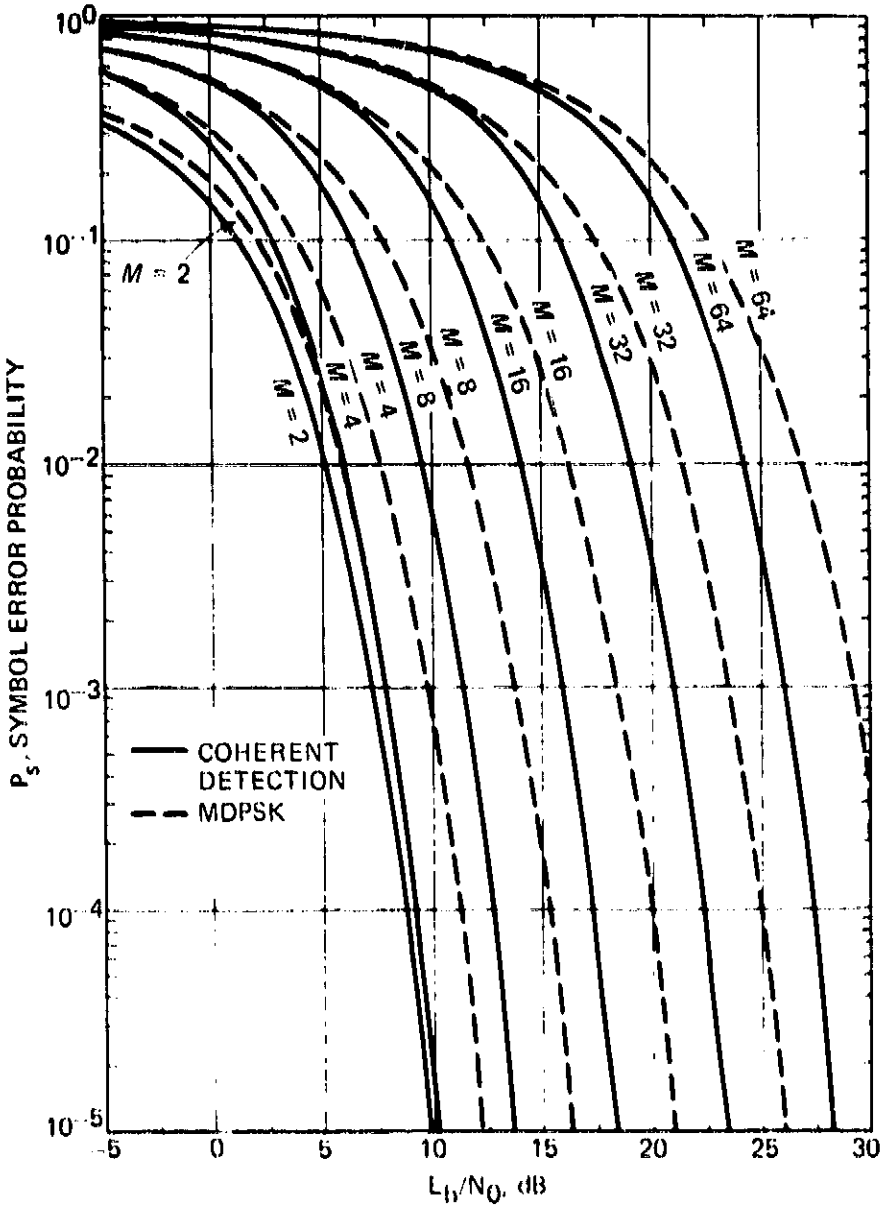


Figure 23. Symbol Error Probability Performance of MPSK and DMPSK (Reprinted from [28])

these two sequences. Typically then one plots the ratio (in dB) of $D^2_{min}/2E_b = (D^2_{min}/2E_s)\log_2 M$ for a given CPM technique to that of MSK versus the normalized bandwidth $2WT_b$.

In Figure 24 we consider the case where p(t) is defined in (1.3.20) with various values of the modulation index h and M = 2, 4, and 8. This particular form of CPM is called Continuous Phase Frequency Shift Keying (CPFSK), with MSK again corresponding to the special case of M = 2, h = .5. For the Raised Cosine (RC) pulse that spans L symbol times (see Figure 18a), Figure 25 shows these modulations for various modulation indices h, M = 2, 4, 8, and pulse memory lengths L = 3, 4, 5 symbols. Also shown in this figure are curves for the less realistic shaping pulse

$$p(t) = \frac{2\pi}{LT_s}h\left(\frac{\sin\left(\frac{2\pi t}{LT_s}\right)}{\left(\frac{2\pi t}{LT_s}\right)}\right)\left(\frac{\cos\left(\frac{2\pi t}{LT_s}\right)}{1-4\left(\frac{2t}{LT_s}\right)^2}\right) \qquad (1.3.31)$$

which has infinite duration. These curves are labeled LSRC (L=1,2,3,4,5,6). Both Figures 24 and 25 show that considerable improvement over MSK is possible at the cost of more receiver digital processing. It should be noted that we assume no RF filtering on these signals while conventional MPSK would require some RF filtering to satisfy a bandwidth requirement and thus suffer some additional signal distortion. Also, note that the FCC mask places additional limitations on the signal power density spectrum than the 99% bandwidth limitation assumed in these results.

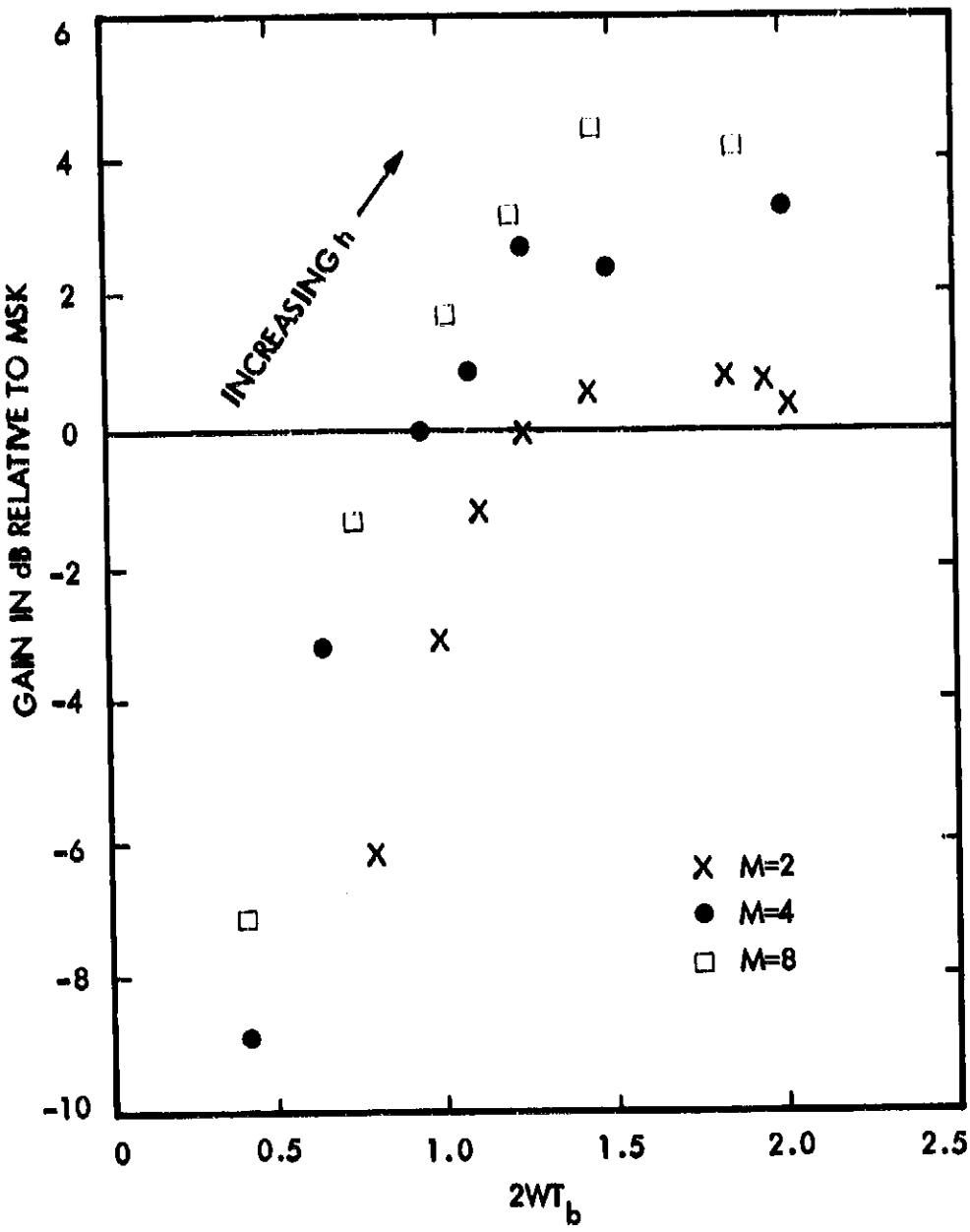


Figure 24. Bandwidth/Performance Comparison Relative to MSK for Various CPFSK Systems (99 Percent Fractional Out of Band Power Bandwidth, See Table I-1)
(Reprinted from [14])

Table 1-1. Bandwidth/Distance Trade-Off for Some M-ary CPFSK Systems

CPFSK Scheme	99% Bandwidth $2WT_b$	$D^2_{min}/2E_b$	Gain over MSK, dB	N_B Symbols
M=2 h=.5	1.20	2.0	0	2
M=4 h=.25	0.80	1.45	-1.38	2
M=8 h=.125	0.54	.60	-5.23	2
M=4 h=.40	1.08	3.04	1.82	4
M=4 h=.45	1.18	3.60	2.56	5
M=8 h=.30	1.00	3.0	1.76	2
M=8 h=.45	1.40	5.40	4.31	5

3.2 Differentially Coherent Modulation

In differentially coherent modulation the transmitted signal carrier phase during any symbol time, T_s, uses the previously transmitted T_s second signal carrier phase as a reference. This is generally done with MPSK modulation. The primary reason for considering this modulation is to allow the receiver to demodulate a symbol without the need for a continuous phase estimator such as a phase-locked-loop (PLL) or a Costas loop. [28].

The transmitted signal for differentially coherent MPSK (DMPSK) is of the form in (1.3.17) where now

$$\theta(t) = \theta_I(t) + \theta_R(t) \tag{1.3.32}$$

with

$$\theta_I(t) \in \{0, \tfrac{2\pi}{M}, 2(\tfrac{2\pi}{M}), \ldots, (M-1)\tfrac{2\pi}{M}\};$$

$$(n-1)T_s < t < nT_s; \; n = 1, 2, \ldots \tag{1.3.33}$$

and

$$\theta_R(t) = \theta_I(t - T_s) \tag{1.3.34}$$

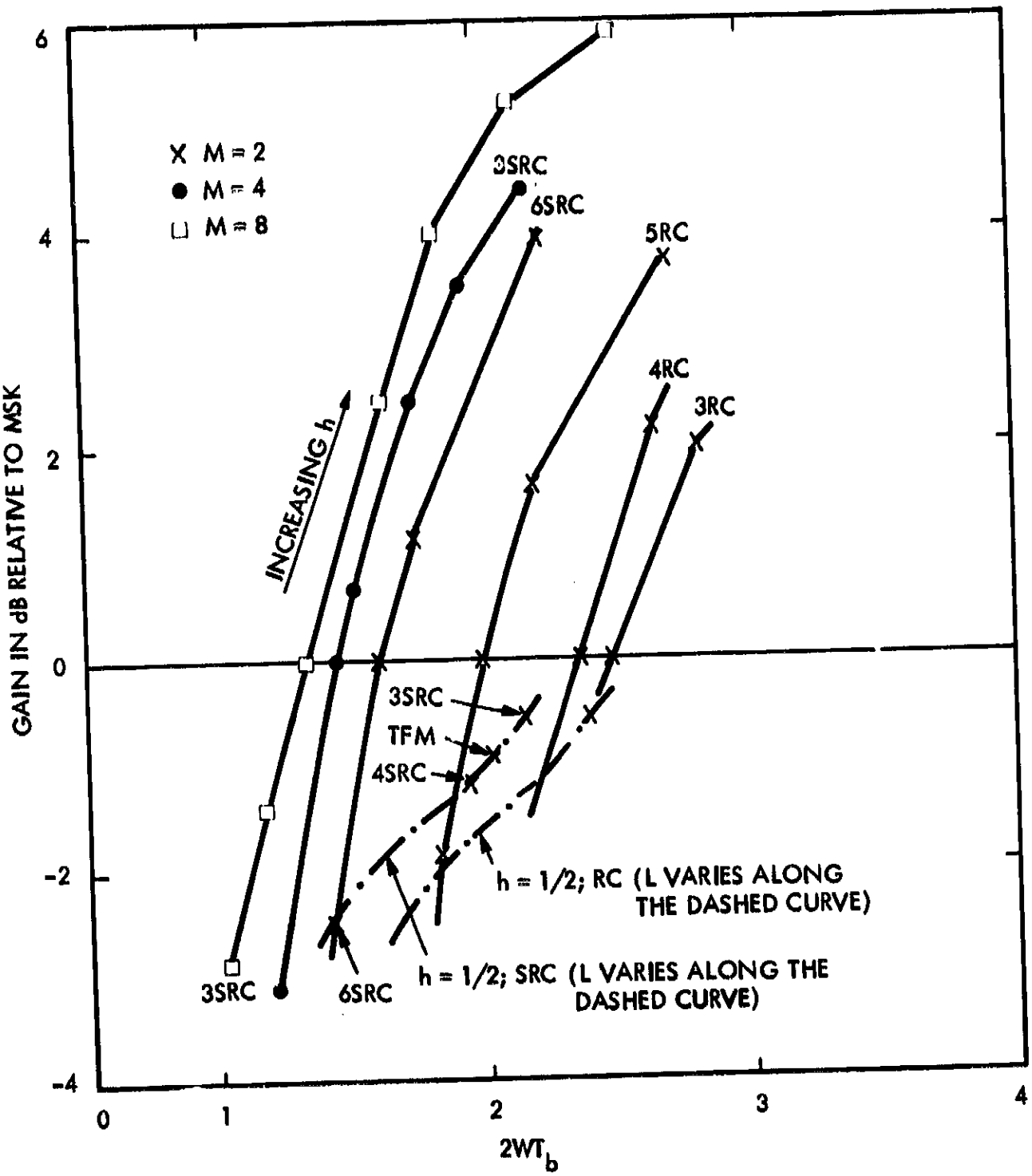


Figure 25. Bandwidth/Power Comparison Between Various Partial Response CPM Systems (Reprinted from [14])

Figure 26 illustrates a DMPSK modulation for the simple case of M = 2 which is commonly referred to as DBPSK or even simpler DPSK. In this figure, $u_k \in \{-1, 1\}$ is the binary data sequence, while

$$x_n = x_{n-1} u_n = \begin{cases} x_{n-1} ; & u_n = 1 \\ -x_{n-1} ; & u_n = -1 \end{cases} \tag{1.3.35}$$

is the differentially encoded binary sequence that is modulated by a conventional BPSK modulator. The receiver for this is shown in Figure 27 while the general

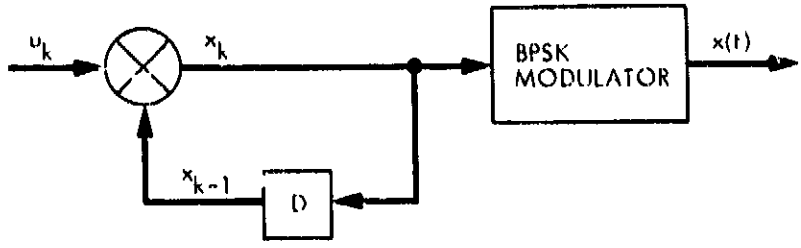


Figure 26. A DBPSK Modulator

DMPSK receiver is illustrated in Figure 28. These receiver structures replace a more complex phase tracking loop with some simple digital processing (differential decoders). This generally results in a more robust system since phase acquisition and loss of lock problems associated with closed loop systems are eliminated.

Clearly, the above advantages of DMPSK over coherent MPSK must come at the expense of some loss in error probability performance. For DMPSK transmitted over the ideal additive white Gaussian noise channel, the symbol error probability is given by [29]

$$P_s = \frac{\sin \frac{\pi}{M}}{\pi} \int_0^{\pi/2} \frac{\exp\left\{-\frac{E_s}{N_0}\left[1 - \cos\frac{\pi}{M}\cos\phi\right]\right\}}{1 - \cos\frac{\pi}{M}\cos\phi} d\phi \tag{1.3.36}$$

with E_s/N_0 related to E_b/N_0 by (1.3.4). An excellent, easy to evaluate asymptotic formula for the above symbol error rate has recently been derived as [29]

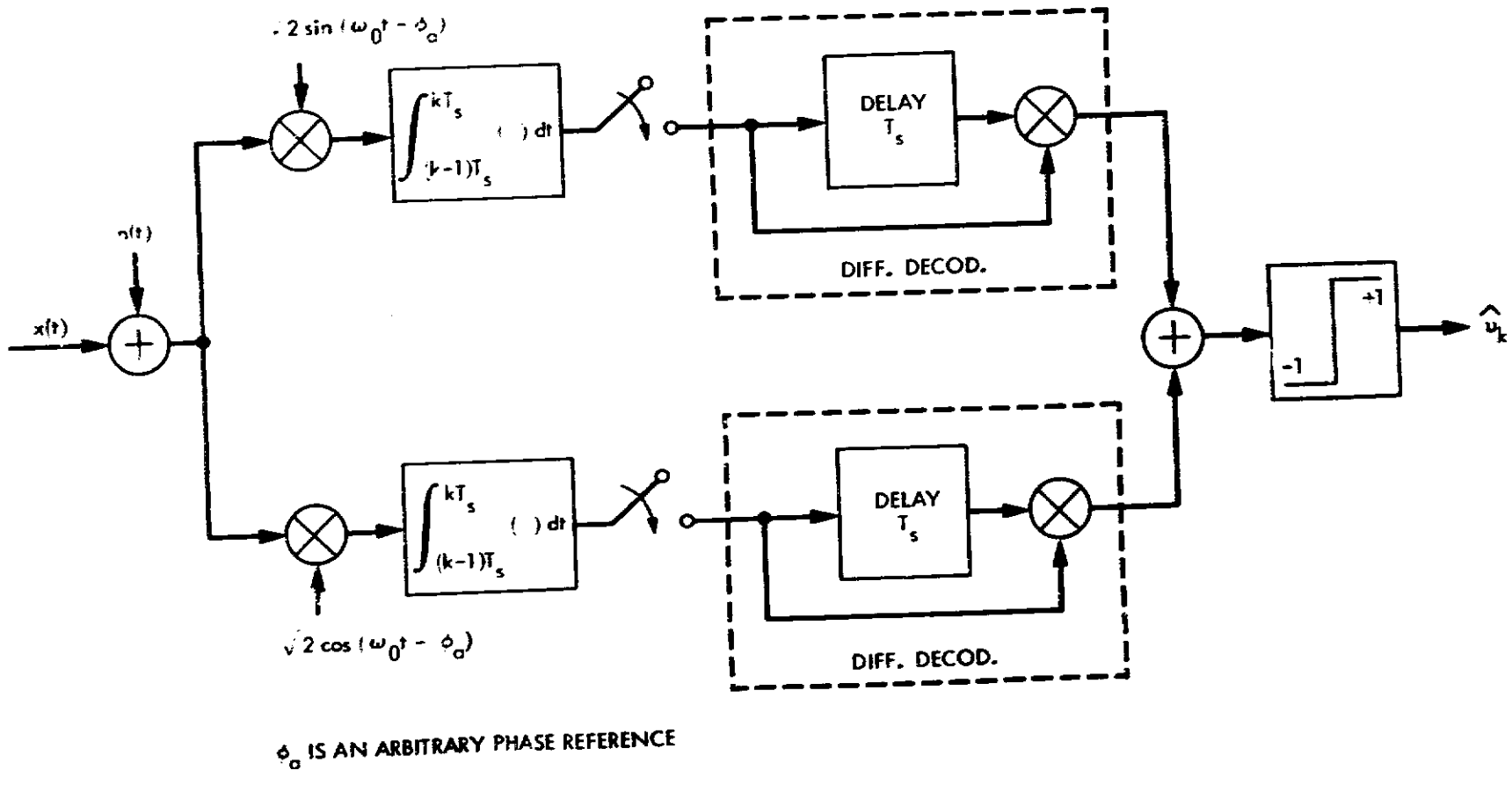


Figure 27. A DBPSK Receiver

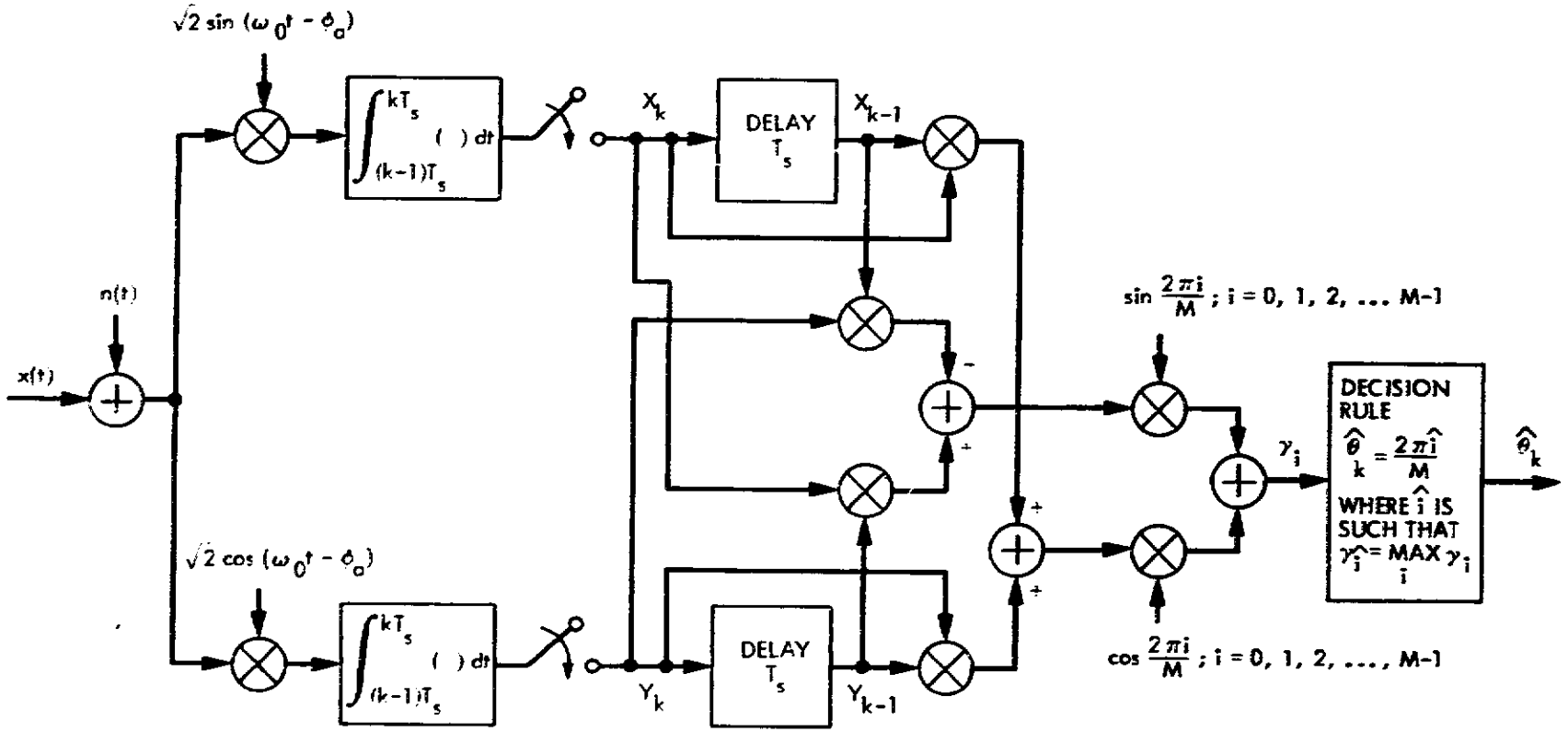


Figure 28. A DMPSK Receiver

$$P_s = \sqrt{\frac{1 + \cos\frac{\pi}{M}}{2\cos\frac{\pi}{M}}}\, Q\sqrt{2\frac{E_s}{N_0}\left(1 - \cos\frac{\pi}{M}\right)} \qquad (1.3.37)$$

Figure 23 illustrates the exact computation of P_s as in (1.3.36) (dashed lines) and compares it with the ideal coherent MPSK modulation which requires a phase tracking loop in the receiver.

3.3 Noncoherent Modulation

An even simpler and more robust class of modulation techniques are those that use no phase estimation at all in the demodulation process. Generally, these modulations use orthogonal signals. The most common example are the M-ary Frequency Shift Keying (MFSK) signals of the form given in (1.3.21) where now*

$$\omega_c\{\omega_0 - (M-1)\Delta\omega,\ \omega_0 - (M-3)\Delta\omega,\ldots,\ \omega_0 - \Delta\omega,\ldots,\ \omega_0 + (M-1)\Delta\omega\} \qquad (1.3.38)$$

Here $2\Delta\omega = 2\pi/T_s$ is the smallest frequency separation that guarantees the different frequency tones are all orthogonal to each other for all relative phase shifts, i.e.,

$$\int_0^{T_s} \cos[\omega_k t + \theta_k]\cos[\omega_\ell t + \theta_\ell]dt = 0 \qquad \ell \neq k \qquad (1.3.39)$$

for all θ_k, θ_ℓ and

$$\omega_k = [2k - 1 - M]\Delta\omega + \omega_0 \quad ; \quad k = 1,2,\ldots, M \qquad (1.3.40)$$

*Note that the frequency spacing in (1.3.38) is twice that given in (1.3.22) for coherent MFSK. This is required in order to satisfy the orthogonality condition of (1.3.39) as opposed to the less stringent orthogonality condition of (1.3.23).

Thus, by comparison with (I.3.28) the total bandwidth is now

$$W = M \frac{2A\omega}{2\pi}$$

$$= M \frac{1}{T_s}$$

$$= \frac{2^K}{KT_b} \qquad \text{(I.3.41)}$$

or

$$M = WT_s \qquad \text{(I.3.42)}$$

is roughly the number of possible noncoherently orthogonal signals of bandwidth W Hz and duration T_s seconds.

The optimum receiver for noncoherent binary FSK (BFSK) is shown in Figure 29. This receiver consists of two energy detectors centered at the two frequencies ω_1 and ω_2. The decision rule is to choose the signal with the larger energy detector output. This receiver assumes no knowledge of the signal carrier phase and in fact its performance does not depend on the actual phase. With M frequencies we would require M energy detectors.

In Figure 29 we show the energy detection process as first obtaining the quadrature components of each possible carrier frequency and then summing the square of these components to get the energy detected at each carrier frequency. This can be accomplished more simply with matched filters followed by envelope detectors whose outputs are sampled. This is, in fact, the form used in general especially when the orthogonal signals are not necessarily MFSK.

As an example of this point, another common way to generate orthogonal signals is to use binary sequences of length $M = 2^K$ to generate M orthogonal binary sequences wich are then BPSK modulated. This yields signals that consist of sequences of M BPSK pulses. The binary sequences are generated recursively as follows:

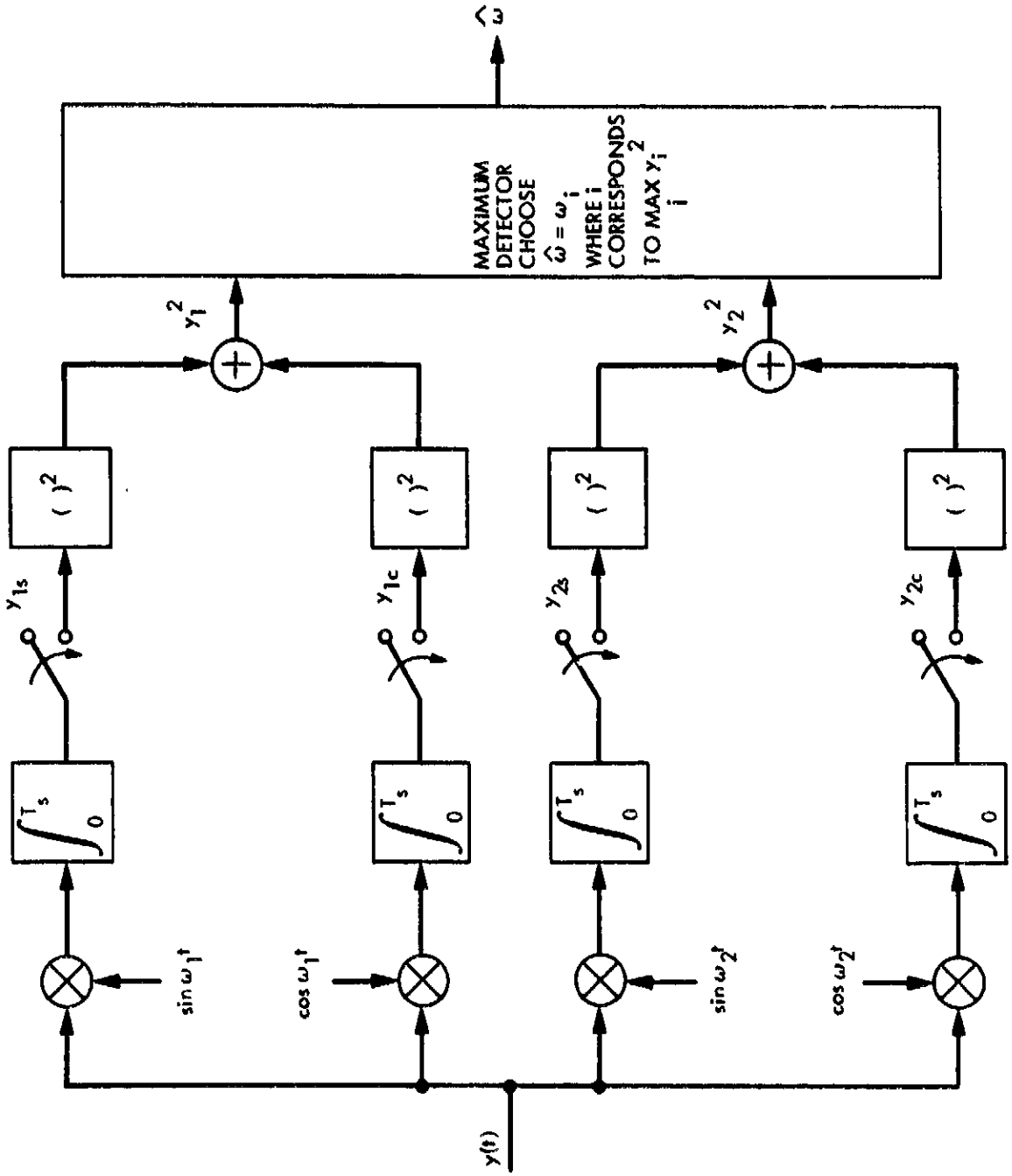


Figure 29. Optimum-Receiver for Noncoherent BFSK

$$H_{k+1} = \begin{bmatrix} H_k & H_k \\ H_k & -H_k \end{bmatrix}; \quad k=1,2,\ldots \tag{1.3.43}$$

where

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{1.3.44}$$

The $M = 2^K$ orthogonal binary sequences are then the rows of H_K. Each binary (+1) symbol of the binary sequence of length M is used to multiply a pulse of duration T_s/M. Typically, this is a BPSK modulation with pulse or "chip" duration $T_c = T_s/M$. The receiver consists of M matched filters (matched to the binary sequences with arbitrary carrier phase) followed by envelope detectors whose outputs are sampled. Again, this results in the detection of energy in each of M orthogonal signal coordinates and then deciding the signal transmitted is the one corresponding to the largest detected energy.

For all orthogonal signaling schemes where there are $M = 2^K$ signals for every K data bits, the bit error probability of the noncoherent receiver is given by [28]

$$P_b = \frac{M}{2(M-1)} P_s = \frac{M}{2(M-1)} \left[1 - \int_0^\infty \exp\left[-\left(x + \frac{E_s}{N_0}\right)\right] I_0\left(\sqrt{4\left(\frac{E_s}{N_0}\right)x}\right)\right.$$

$$\left. x\,[1 - \exp(-x)]^{M-1}\, dx\right]$$

$$= \sum_{k=1}^{M-1} \frac{(-1)^{k+1}}{k+1} \left(\frac{(M-1)!}{k!(M-1-k)!}\right) \exp\left(-\frac{k\,\frac{E_s}{N_0}}{k+1}\right) \tag{1.3.45}$$

and tightly bounded by

$$P_b \leq 2^{K-2} \exp\left(-\frac{E_s}{N_0}\right) \tag{1.3.46}$$

Figure 30 shows these bit error probabilities as computed from (1,3,45) as a function of $E_b/N_0 = E_s/KN_0$ with K the number of data bits used to determine the number of orthogonal signals as a parameter. Figure 31 compares coherent and noncoherent receivers for orthogonal signals for the special case of K = 5. The uncoded cases correspond to using BFSK 5 times with the symbol error probability being the probability that one or more of the 5 BFSK "chips" are incorrect. The coded cases are where we have M = 32 orthogonal signals (32-ary FSK for example). Similar results are shown in Figure 32 for K = 10 (M = 1024).

With larger values of M the difference between coherent and noncoherent orthogonal signals becomes small. The noncoherent receiver, however, is simpler since it does not need to include carrier phase estimation. Such noncoherent modulation/demodulation techniques are robust and common in spread spectrum military communication systems where bandwidth efficiency is not a consideration. For commercial applications, however, bandwidth efficiency is very important and these orthogonal signalling techniques are impractical.

One noncoherent modulation technique that has been proposed for commercial applications [30] is the M = 4 level CPFSK modulation discussed earlier together with a noncoherent FM discriminator receiver. This M = 4 FSK modulation differs from the MFSK modulations considered above in that it has continuous phase transitions from symbol to symbol and the 4 possible frequency tones are not orthogonal to each other. The noncoherent FM discriminator receiver has been proposed because it is one of the least expensive receivers available. The performance of such a noncoherent receiver of CPFSK modulation is not precisely known. Its symbol error probability, however, is certainly lower bounded by the M = 4 coherent CPFSK symbol error probability (see Figure 24).

For binary signals (M=2), we compare the various modulation techniques in Figure 33. Here we consider coherent modulation/demodulation of binary PSK, differentially encoded PSK, and binary FSK. This is compared with differential PSK using the DBPSK receiver of Figure 27 and the noncoherent BFSK receiver of Figure 29. Coherent detection of differentially encoded PSK generated as in Figure 26 asymptotically (large E_b/N_0) results in a small loss (a factor of two in error probability) when compared with coherent detection of conventional binary

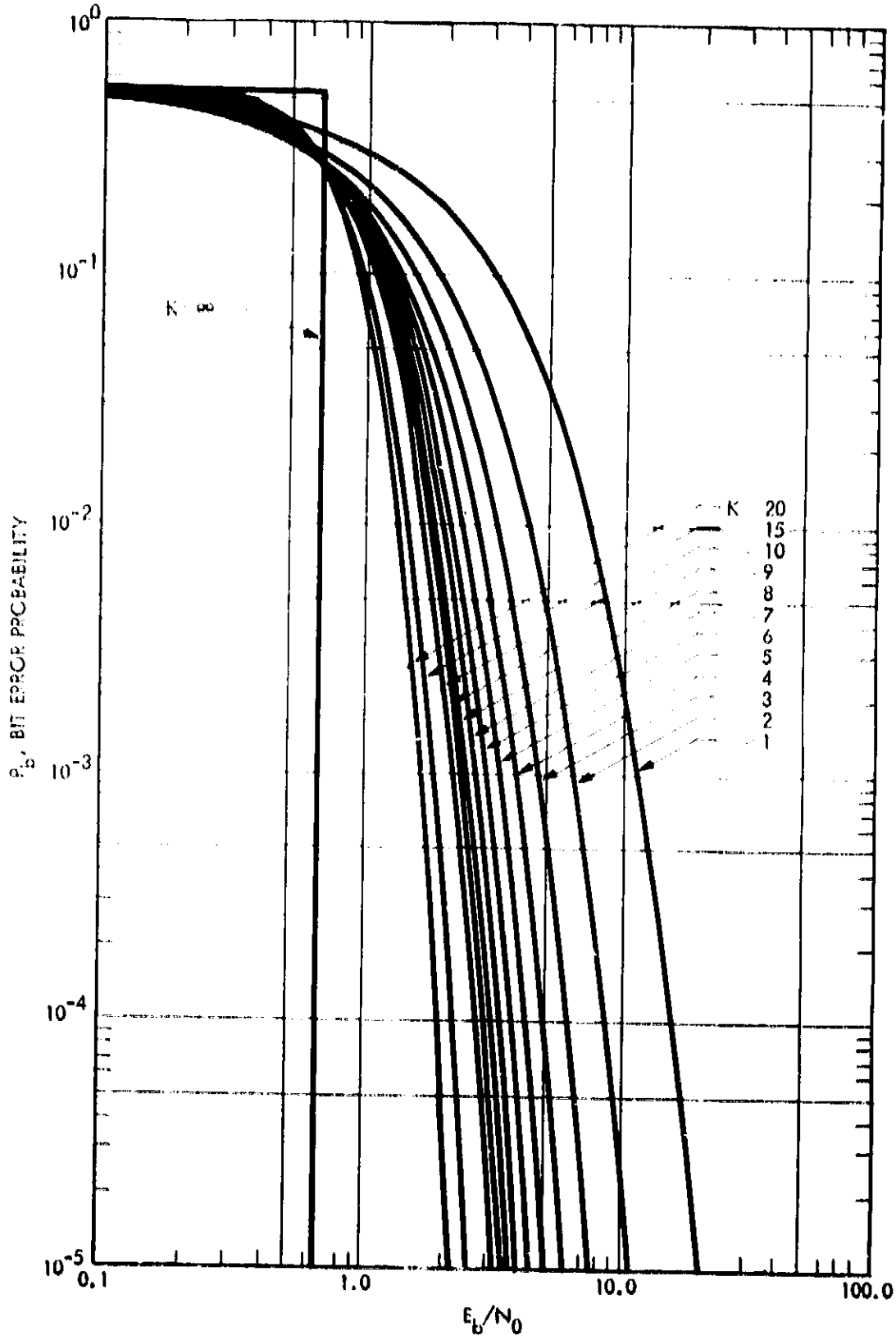


Figure 30. Bit error probability performance of an ideal noncoherent receiver (orthogonal signaling) (Reprinted from [28])

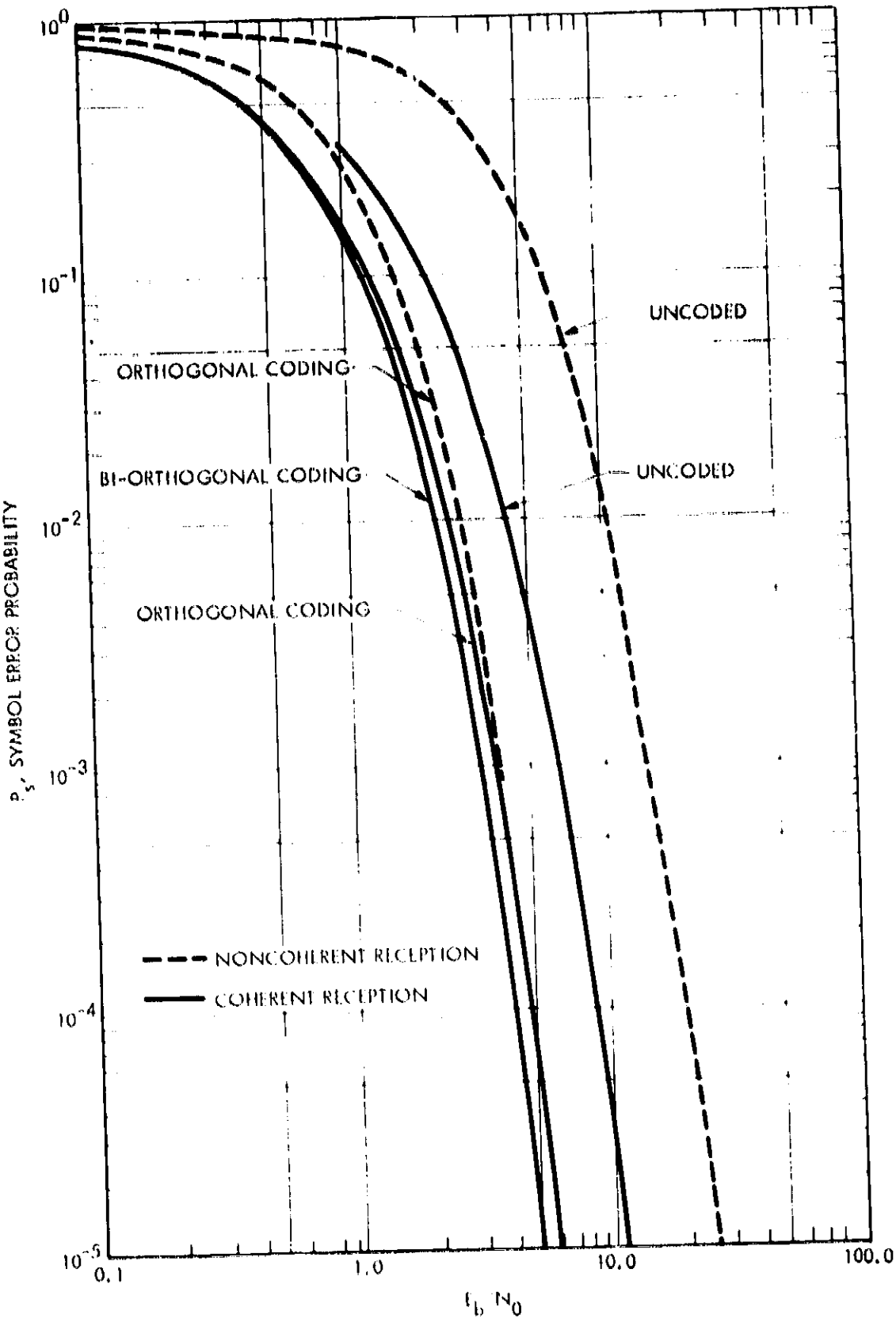


Figure 31. A comparison of coded and uncoded, coherent and noncoherent error probabilities (K = 5) (Reprinted from [28])

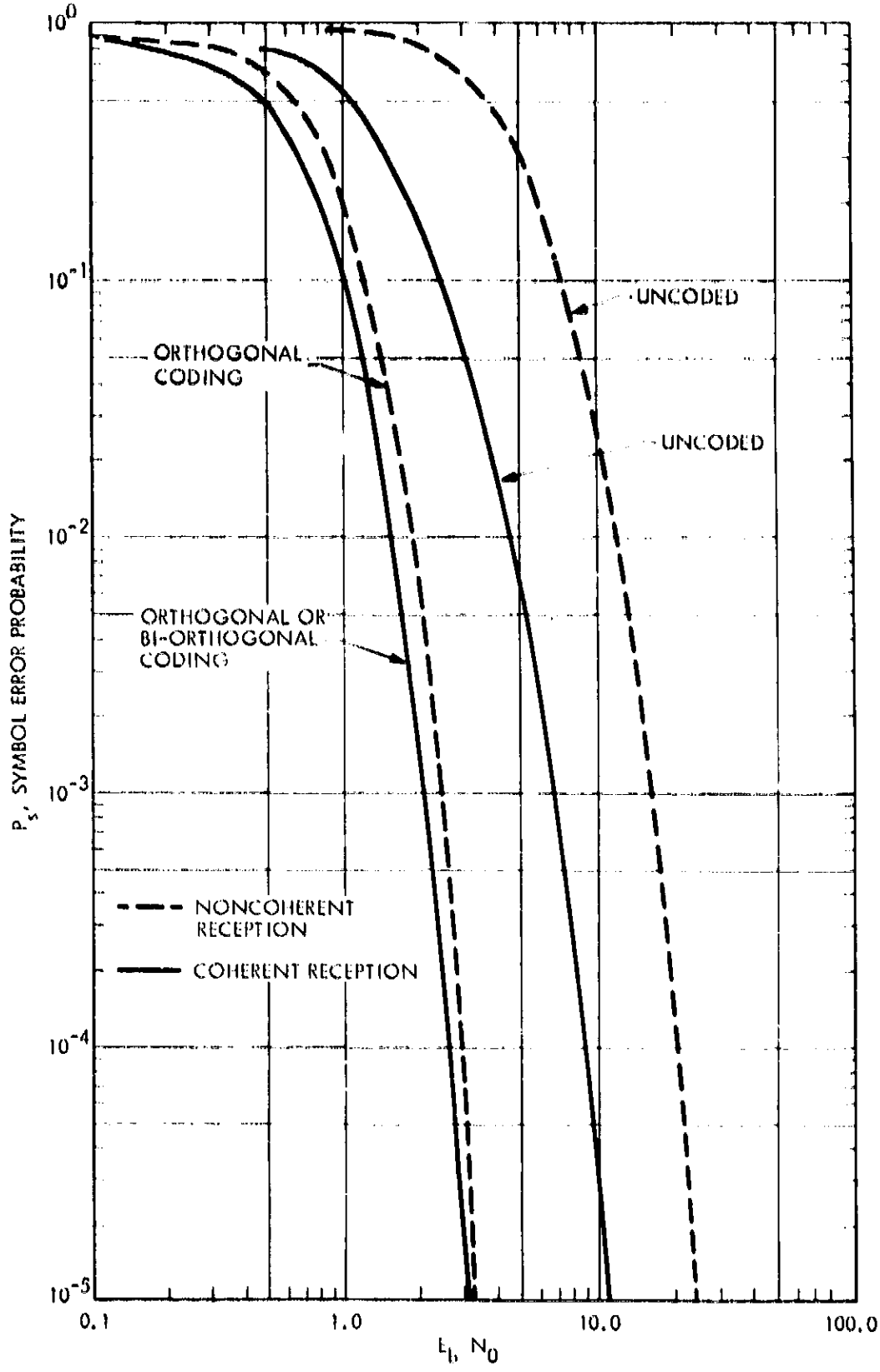


Figure 32. A comparison of coded and uncoded, coherent and noncoherent error probabilities (K = 10) (Reprinted from [28])

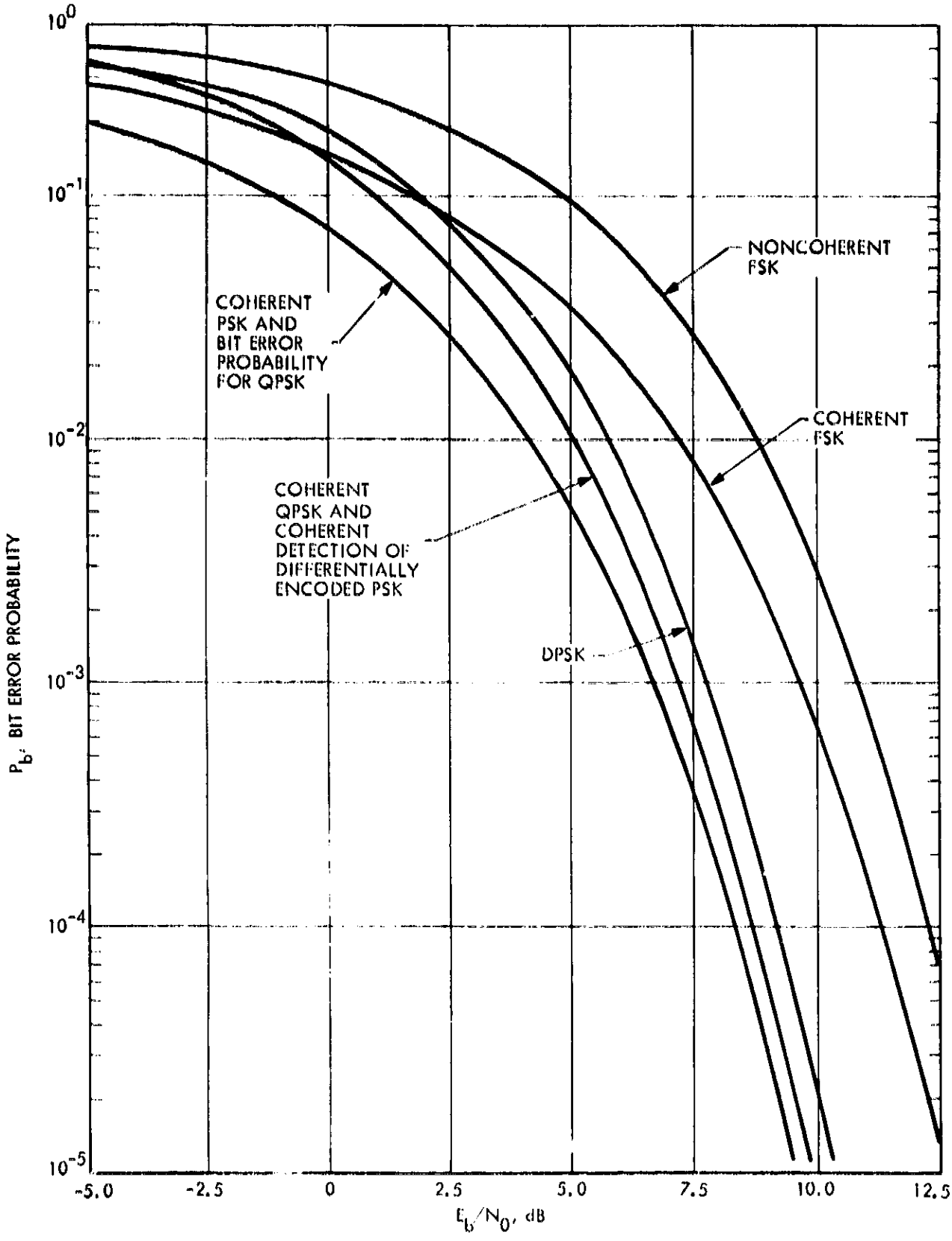


Figure 33. A comparison of error probabilities for coherent, differentially coherent, and noncoherent reception of several signaling formats (Reprinted from [28])

PSK. The differential (encoding) and decoding operations in these systems are used for resolving the 180° phase ambiguity inherent in carrier tracking loops for suppressed carrier binary modulations, e.g., PSK. A further loss occurs when the differentially coherent receiver of Figure 27 is used with the DBPSK modulator of Figure 26. Finally, the FSK curves correspond to BFSK where the two tones are orthogonal.

4.0 Forward Error Correction

In 1949, Shannon [31] first showed that it is possible to send data over a noisy communication channel with arbitrarily small error probabilities as long as that data rate is less than the "channel capacity." This can be achieved by first encoding the data sequence into codewords and at the receiver decoding the received channel output symbols back into a data sequence. Only in the last ten years has Shannon's promise begun to be realized. This was partly due to the development of the Viterbi algorithm [32] which together with moderate size convolutional codes proved to be quite effective in correcting channel errors. Perhaps more important was the evolution of the solid-state electronic technology at this same time that allowed increasingly fast and complex digital processing that was necessary to implement the elaborate techniques that coding theory developed in the years since Shannon.

Today coding techniques are used extensively to improve the performance of communication systems. For the most part, coding has been used with coherent BPSK and QPSK modulations and usually as an addition to an existing system to lower the bit error probability. However, when designing a communication system, the choice of coding and modulation techniques should be made together to achieve overall better performance [33]. Surprisingly, the modulations yielding the smallest uncoded error probabilities are not always the best modulations when used together with coding.

In this section we first examine the most commonly used codes with coherent BPSK and QPSK modulations. Then we discuss a systematic procedure for evaluating modulation techniques that will be used together with coding.

4.1 Coding Channel Models

Recall that for BPSK modulation the bit error probability is given by

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{1.4.1}$$

where E_b is the energy per data bit entering the BPSK modulator. This bit error probability (M=2 curve in Figure 23) may not be small enough for some users and so we might consider the options:

(a) increase power

(b) increase symbol duration.

Both of these have the effect of increasing E_b which would lower the bit error probability. Another possibility is to introduce coding to correct errors caused by the noisy channel. Here we can decrease the bit error probability by keeping E_b unchanged but at a cost of more digital processing at the transmitter in the form of an encoder and at the receiver in the form of a decoder.

Encoding is done prior to modulation while decoding follows demodulation as shown in Figure 34. From a coding point of view, the portion of the system between the input to the modulator and the output of the demodulator is regarded as the "coding channel." For BPSK and QPSK modulations we have the coding channel models shown in Figure 35. For BPSK the probability ε of making a channel error (referred to as the crossover probability of the binary symmetric channel (BSC)) is given by

$$\varepsilon = Q\left(\sqrt{\frac{2E_c}{N_0}}\right) \tag{1.4.2}$$

where E_c is the energy per coded binary symbol. Note that this is identical to the BPSK bit error probability but now each binary symbol entering the BPSK modulator is a coded bit which is an output of the channel encoder.

The channel encoder takes data bits and outputs coded channel symbols. For the BPSK channel these would be binary symbols while for the QPSK modulation

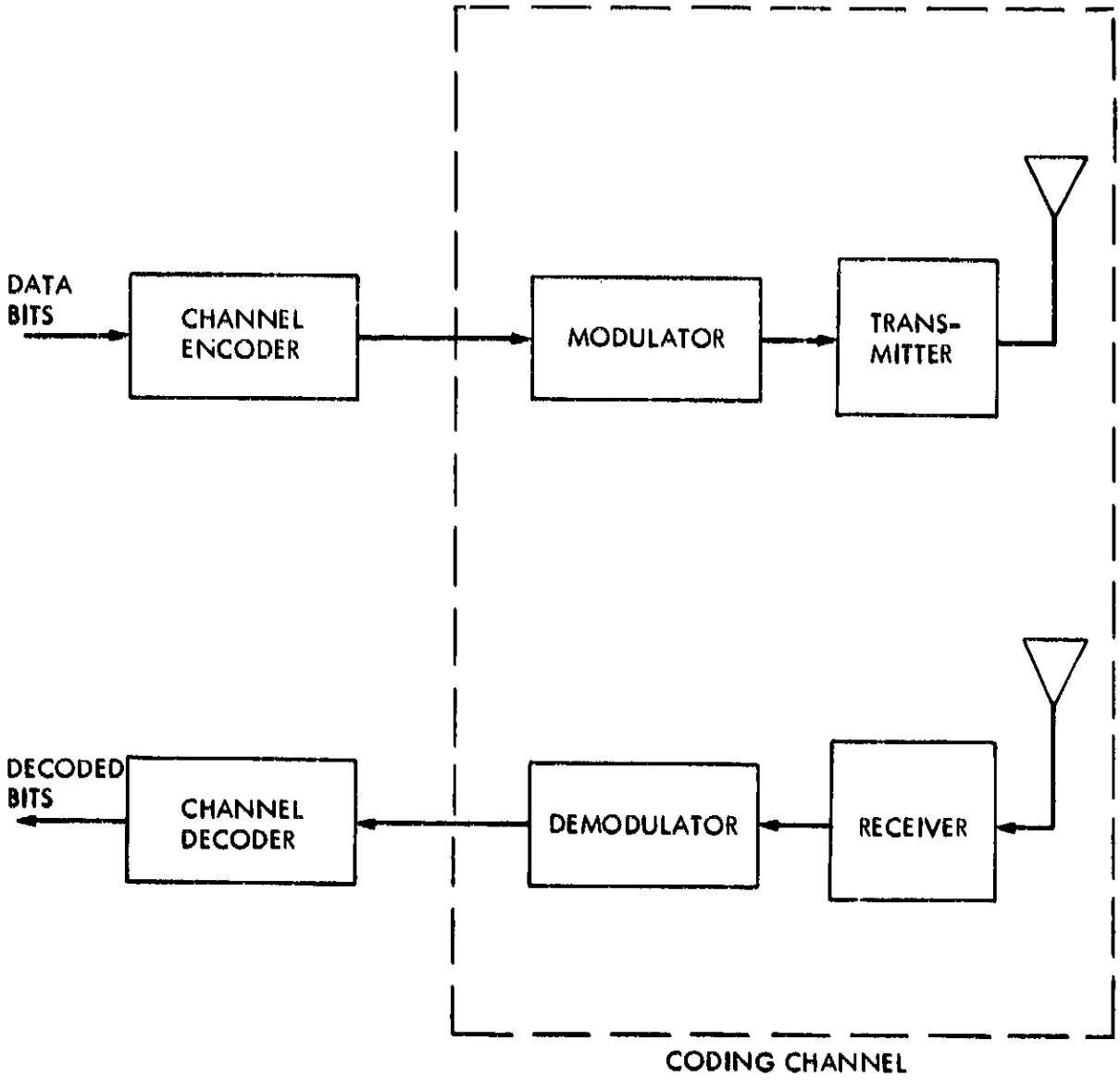
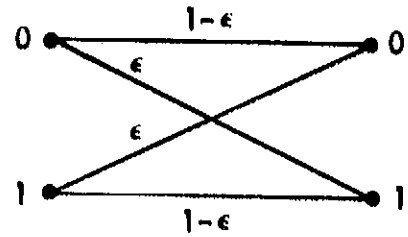
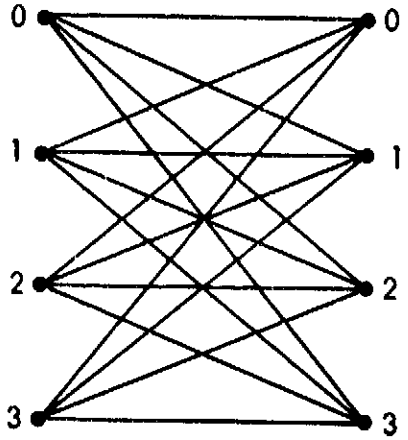


Figure 34. The Coding Channel



(a) BPSK



(b) QPSK

Figure 35. BPSK and QPSK Coding Channels

these are 4-ary symbols or pairs of binary symbols.* To achieve error correction capabilities, some redundancy is introduced by the channel encoder. For the BPSK coding channel we define the code rate as r bits/coded bit while in general r is defined in bits per coded channel symbol. For BPSK modulation the energy per data bit and energy per coded bit are related by

$$E_c = rE_b \tag{1.4.3}$$

4.2 Block Codes

Block Codes refer to the encoding technique where K data bits are collected by the encoder and then a "block" of N channel symbols are sent over the coding channel to represent the K data bits. For the binary channel in Figure 35a, we might have an $r = \frac{1}{2}$ block code with K = 3 and N = 6 (see Table T-2).

Table 1-2.
A K=3, N=6 Block Code

data bits	channel bits (codewords)
000	000000
001	001111
010	010110
011	011001
100	100011
101	101100
110	110101
111	111010

Keeping $r = \frac{1}{2}$ we might have a code with K = 6 and N = 12 or possibly K = 12 and N = 24. In general for a fixed ratio r = K/N we can achieve smaller bit error probabilities as we increase K and N in this proportion. This is true as long as the code rate r is less than channel capacity. For the binary channel of Figure 35a the channel capacity is [1]

$$C = 1 + \epsilon \log_2 \epsilon + (1-\epsilon)\log_2(1-\epsilon) \quad \text{bits/coded bit} \tag{1.4.4}$$

*The QPSK coding channel is equivalent to two BPSK coding channels.

A practically more important parameter than channel capacity is the channel cutoff rate R_0 (also referred to as computational cutoff rate from its usage in sequential decoding) which for this channel is given by [1]

$$R_0 = 1 - \log_2\left[1 + \sqrt{4\varepsilon(1-\varepsilon)}\right] \quad \text{bits/coded bit} \tag{1.4.5}$$

This is sketched in Figure 36. In general we can always find codes of rate R and block length N (K=RN) such that the probability of a data bit error at the receiver is bounded by

$$P_b \leq 2^{-N(R_0 - R)} \tag{1.4.6}$$

This bit error probability bound is shown in Figure 37 as a function of R_0/R for various values of K = RN. Clearly as long as $R < R_0$ we can achieve as small a bit error probability as desired by choosing large enough values of N. Unfortunately the processing complexity C at the encoder is roughly

$$C = 2^K = 2^{RN} \tag{1.4.7}$$

Today, digital processing capabilities and the speed of computations is rapidly increasing so that we are able to develop coding systems with ever increasing block lengths N.

4.3 Convolutional Codes

The $R = \frac{1}{2}$, K = 3, and N = 6 block code shown in Table 1-2 can have an encoder illustrated in Figure 38a. Here the three data bits are shifted into the top shift register or discrete delay line and then the code switches close to form the N = 6 bits that form the codeword. The bottom discrete delay line is shifted out at twice the data rate. Mathematically, the 6 codeword bits $x = (x_1, x_2, x_3, x_4, x_5, x_6)$ are formed from the 3 data bits $u = (u_1, u_2, u_3)$ by

$$x = u\,G \tag{1.4.8}$$

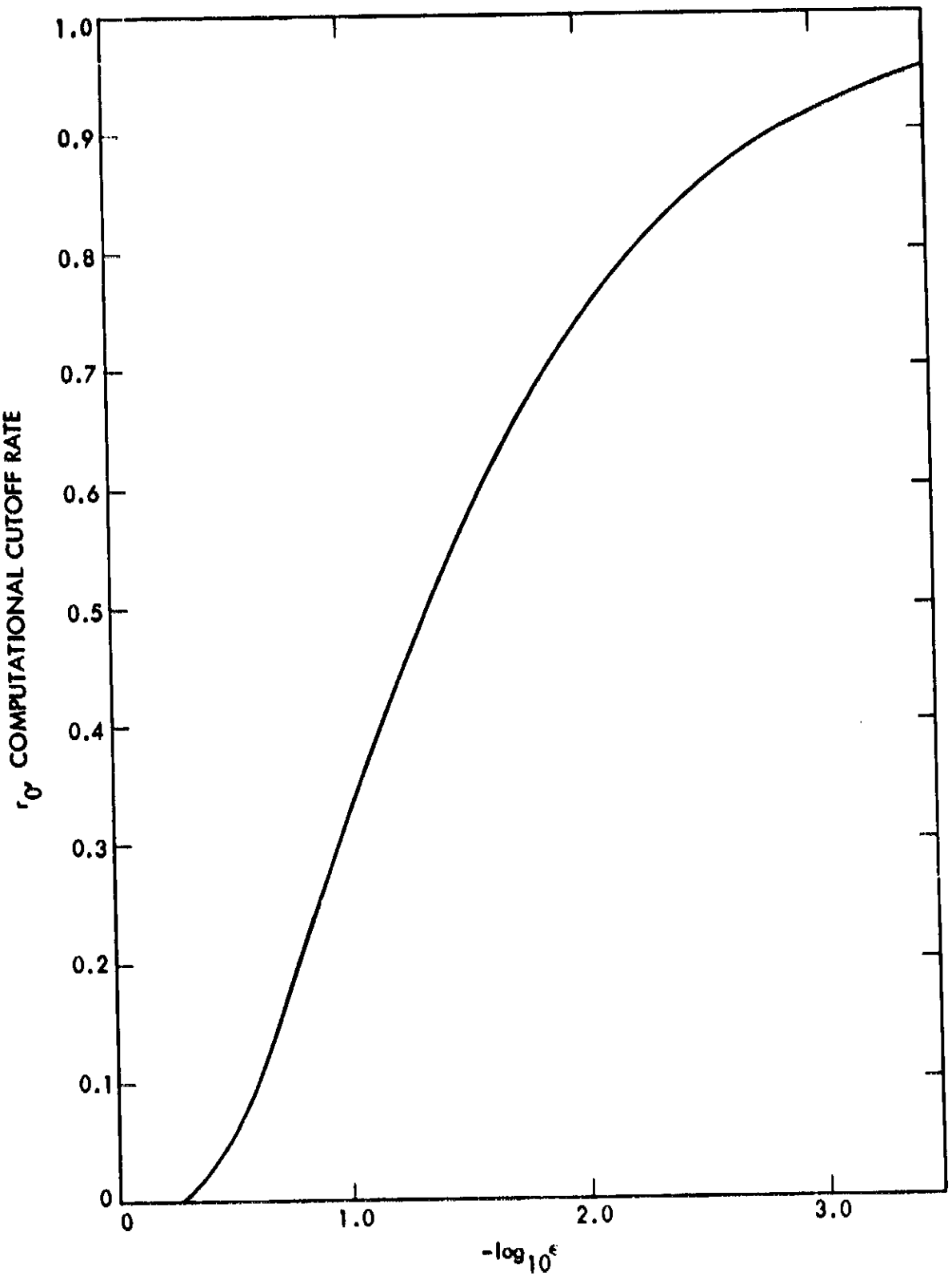


Figure 36. Computational Cutoff Rate versus BPSK Error Rate Exponent

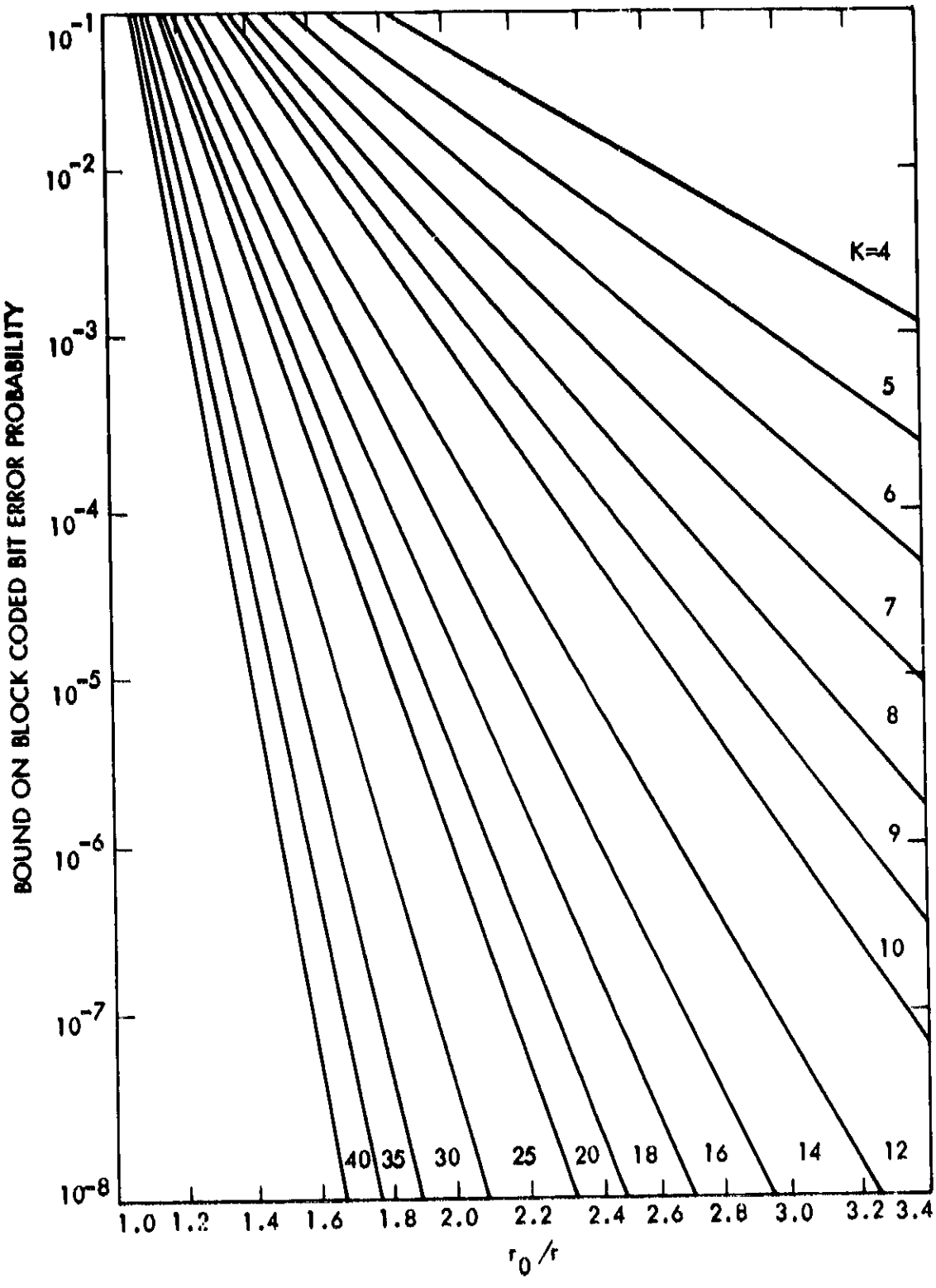


Figure 37. Block Code Bit Error Probability Bound

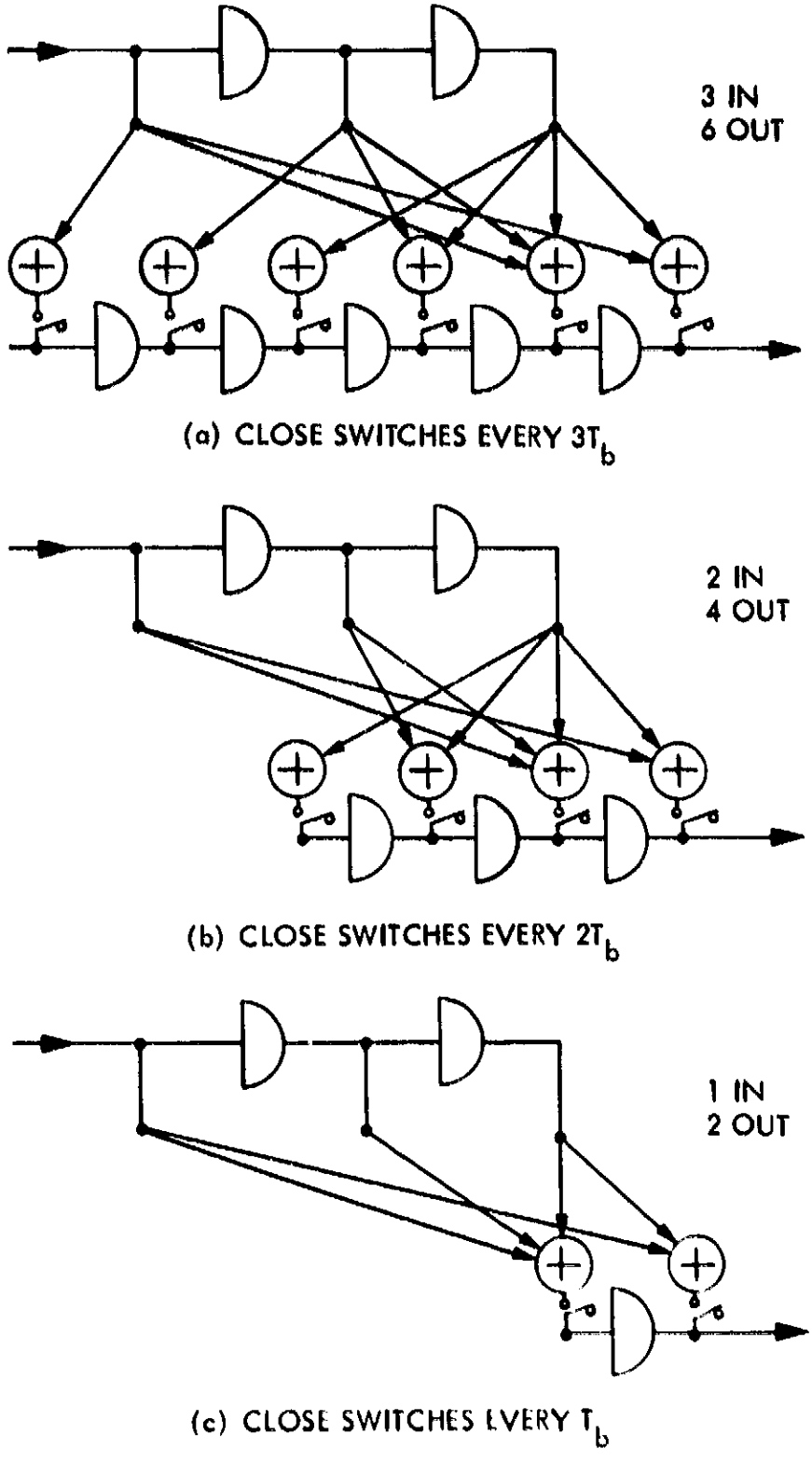


Figure 38. The Effect of Varying the Sample Times

where

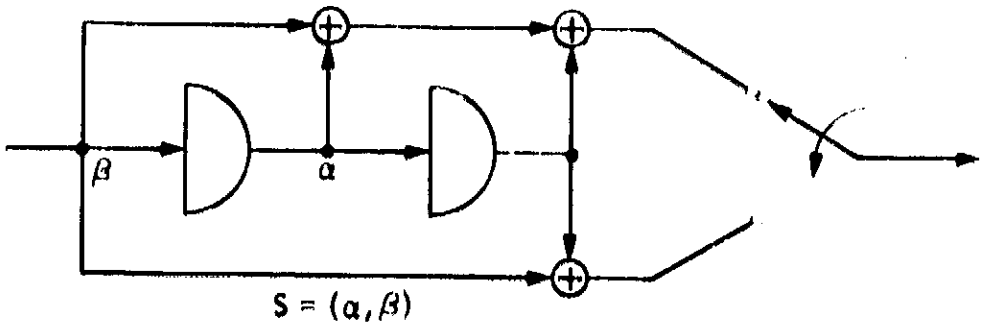
$$G = \begin{bmatrix} 10 & 00 & 11 \\ 01 & 01 & 10 \\ 00 & 11 & 11 \end{bmatrix} \qquad (1.4.9)$$

Codes like this where all codewords are merely linear combinations of K row vectors of a "generator matrix," are called linear codes. Besides having a very simple encoder structure, these codes have the property that the modulo-2 sum of any two codewords is another codeword.

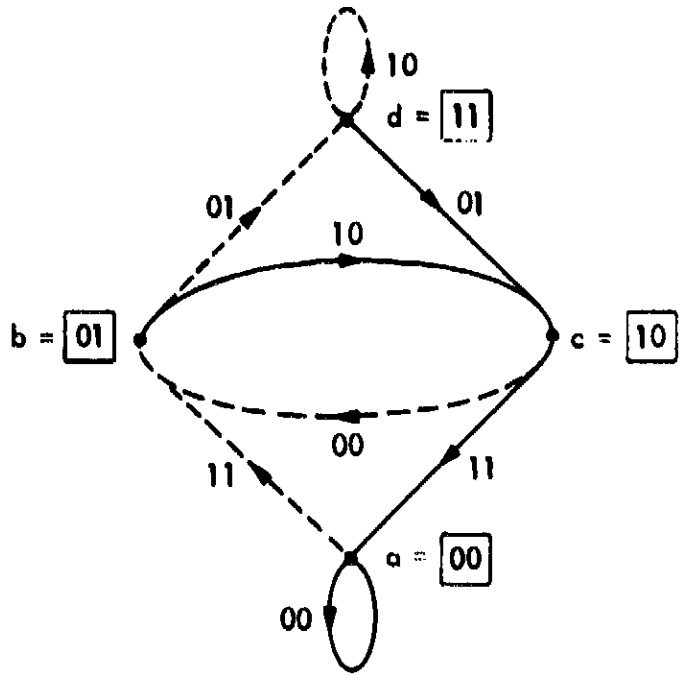
Suppose next, as shown in Figure 38b, we shift only two data bits into the top discrete delay line before closing the switches to form a codeword of 6 bits. If we close the switches every time two data bits are shifted in, the left-most two coded bits never get shifted out and thus we can eliminate the last two coded bits. Next, suppose the switches are closed every time a data bit is shifted in. In this case only two coded bits are shifted out before the switches close again. Hence, we can forget all but the right-most two adders.

Notice that in Figure 38 all encoders have outputs that depend on at most K = 3 bits. The encoders of Figures 38b and 38c, however, no longer have outputs that naturally separate into independent blocks or codewords of N = 6 bits as in the encoder of Figure 38a. These encoders are called convolutional encoders [1] and generate convolutional codes.

Convolutional codes cannot be simply described by listing for each K data bits the corresponding N coded bits as in block coding. For the convolutional code of Figure 38c, for example, it is more convenient to show the encoder as a finite state machine (FSM) described by Figure 39a where the state of the encoder is the previous two data bits that are shifted into the encoder and is denoted by $S = (\alpha,\beta)$. The encoder state transitions are described by the state diagram of Figure 39b where dotted lines indicate a "1" data bit entering the encoder and the solid lines correspond to a "0" data bit. The branches are labelled with the two corresponding coded bits that enter the coding channel. Another description of the inputs and outputs of the convolutional encoder is the tree diagram of Figure 40. Here the branches moving up on the tree corresponds to "0" data bits and along the branches we show the coded output bits of the encoder. Note that parts of the tree are identical to other parts of the tree. This observation allows us



(a) ENCODER



(b) STATE DIAGRAM FOR ENCODER

Figure 39. Representation of a Convolutional Encoder As a Finite State Machine

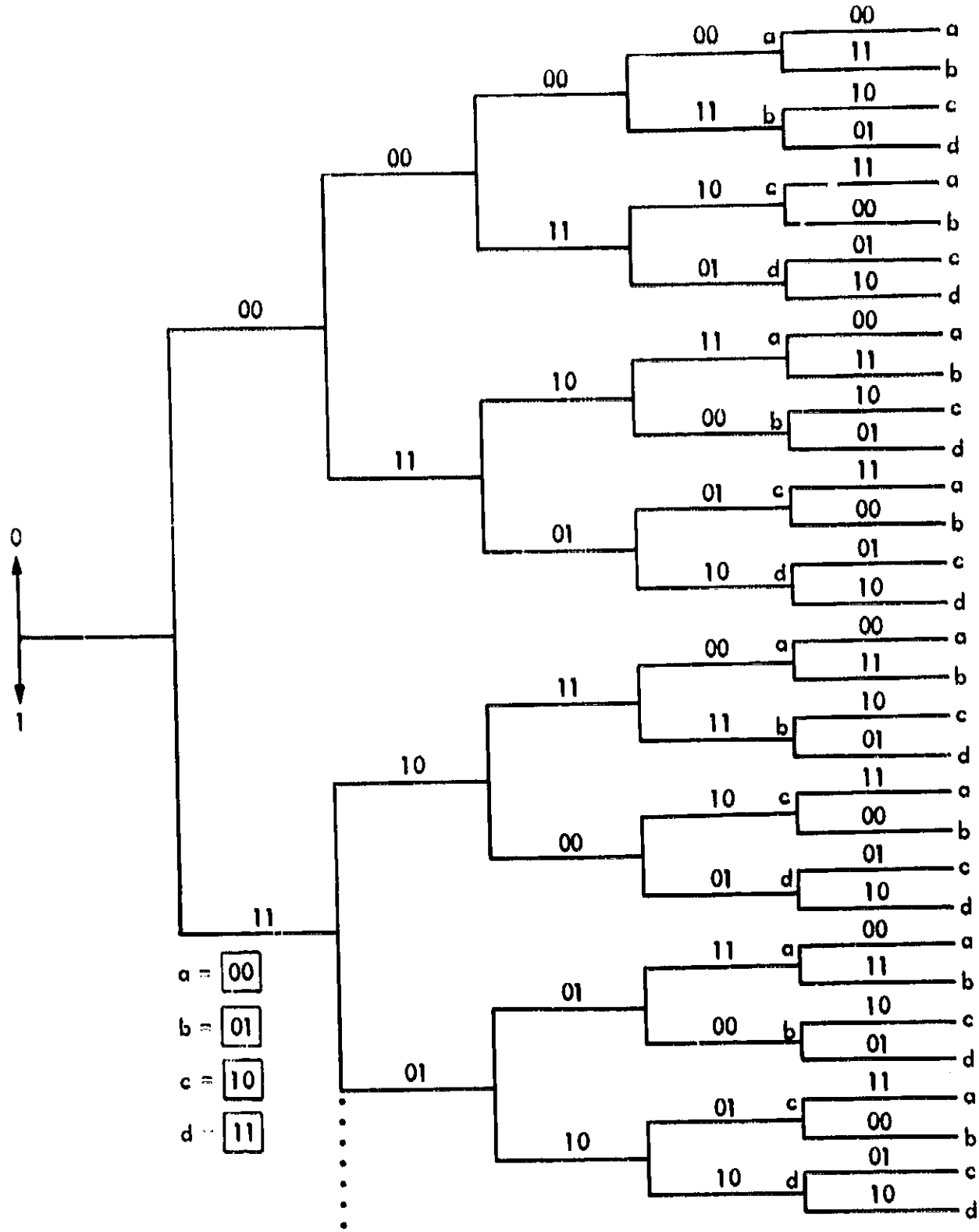


Figure 40. Tree-Code Representation for Convolutional Encoder

to use the more compact form of the tree diagram referred to as a trellis diagram shown in Figure 41. Note that the trellis diagram is also a time sequence representation of the state diagram of Figure 39b.

Because of the "convolved" dependence of the convolutional encoder outputs on the data bits, the performance of convolutional codes tends to be better than that of corresponding block codes with the same rate r and number of data bits K = rN. This may seem surprising since the encoder structure for the convolutional codes of Figure 38 looks simpler than that of the block code.

For the BPSK-generated coding channel of Figure 35a, using convolutional codes of rate r and data bit memory K = rN, there exist codes whose bit error probability satisfies the bound

$$P_b \leq \frac{2^{-Nr_0}}{\left(1-2^{-\left[\left(r_0/r\right)-1\right]}\right)^2} \qquad (I.4.10)$$

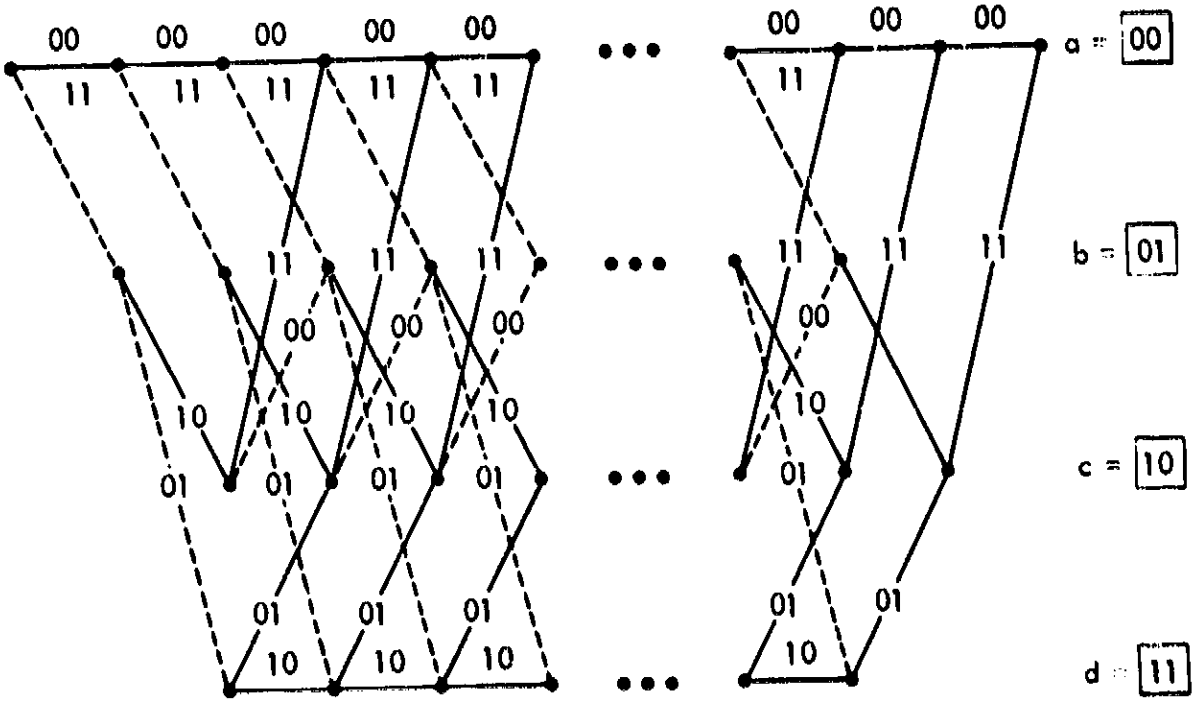


Figure 41. Trellis Diagram for Convolutional Encoder

Figure 42 illustrates the bit error bound of (1.4.10) as a function of r_0/r for various values K. Comparing Figures 37 and 42 we observe verification of the previous statement that, for the same rate r and data bit memory K the convolutional codes generally have smaller bit error probabilities than block codes.

The optimum decoding of convolutional can be done using the Viterbi algorithm which also has complexity on the order of (1.4.7). This algorithm searches all paths in the trellis diagram of Figure 41 which requires roughly 2^K computations per depth into the trellis. Hence, in comparing block codes with convolutional codes for the same complexity we should use the same rate r and data bit memory K.

As mentioned earlier, one of the reasons coding first became practical was the fact that convolutional codes with Viterbi decoding were quite effective for values of K that were manageable from a complexity and speed viewpoint. Using computer search techniques, good convolutional codes were found by Odenwalder [34] for values of K up to 10 and $r = \frac{1}{2}, \frac{1}{3}$, and $\frac{1}{4}$. For the best convolutional codes of $r = \frac{1}{2}$, Figure 43 shows the bit error bounds.

Convolutional codes can also be decoded using a tree search algorithm called sequential decoding [1]. Indeed, sequential decoding preceded the discovery of the optimum Viterbi algorithm by about 10 years [35]. The computational complexity of sequential decoding is insensitive to the encoder data bit memory K, but has a random amount of computation per data bit that can cause buffer overflows at the receiver if the algorithm is spending too much time searching. Since large K is possible, the decoded bit error probability using sequential decoding is quite small. However, with not so small a probability, any receiver data buffer storage will overflow and data will be lost. With ever increasing computational speeds and larger buffers available, sequential decoding is becoming more commonly used in practice.

4.4 Soft Decision Decoding

Up to this point we have assumed that the coding channel as shown in Figure 34 is defined from the transmitter's modulator input to the receiver's demodulator output. For BPSK modulation this results in the binary symmetric channel of Figure 35a. In general for any M-ary modulation this would result in

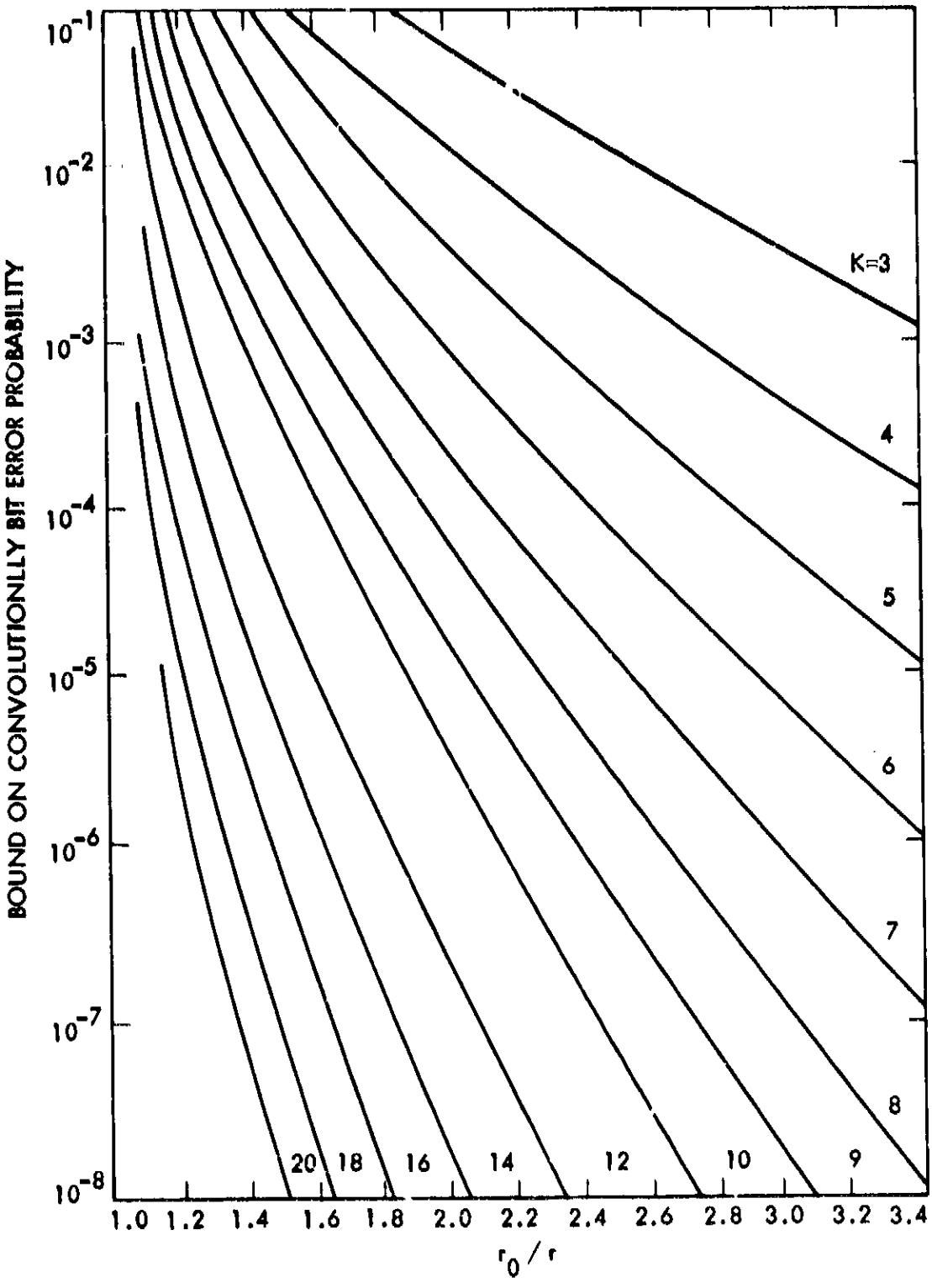


Figure 42. Convolutional Code Bit Error Probability Bound

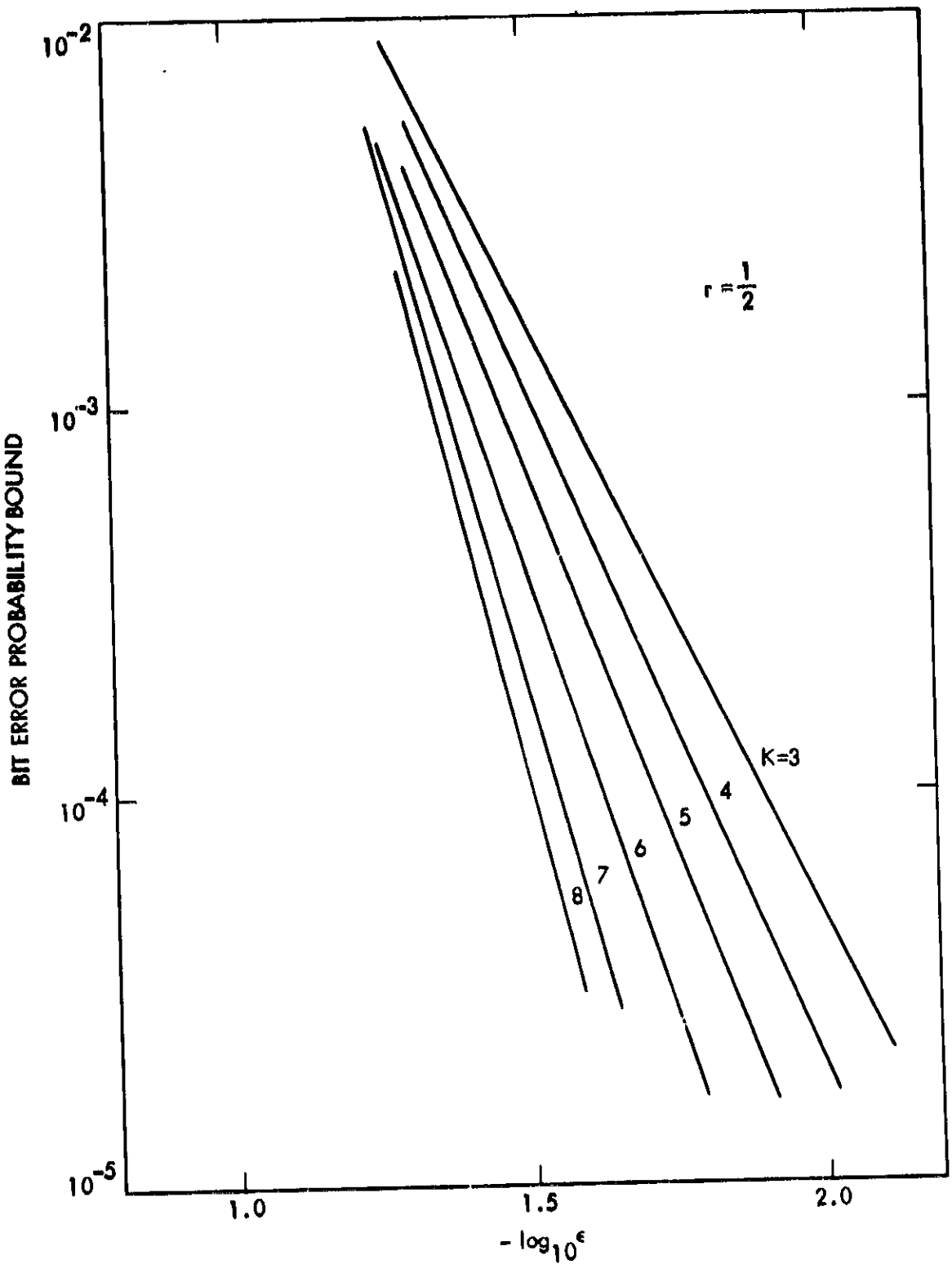


Figure 43. Bounds on the Bit Error Probability Performance of Optimum Binary Convolutional Codes

a coding channel that has M inputs and M outputs such as the M = 4 QPSK modulation coding channel of Figure 35b. These coding channels are called "hard decision" channels.

When the set of modulator input M-ary sequences are restricted to be coded sequences, improved performance can be achieved by using the radio channel output statistics before the demodulator decision circuit. A coding channel that directly uses these channel output variables prior to any demodulator decision circuit is called a "soft decision" coding channel. Consider BPSK modulation in an additive white Gaussian noise channel. As shown in Figure 44a, this channel can be modelled as one with an input of $\pm \sqrt{E}$ and an output y consisting of the input plus a Gaussian random variable n. With a BPSK demodulator a decision is made based on the sign of y resulting in the equivalent binary symmetric channel shown in Figure 44b.

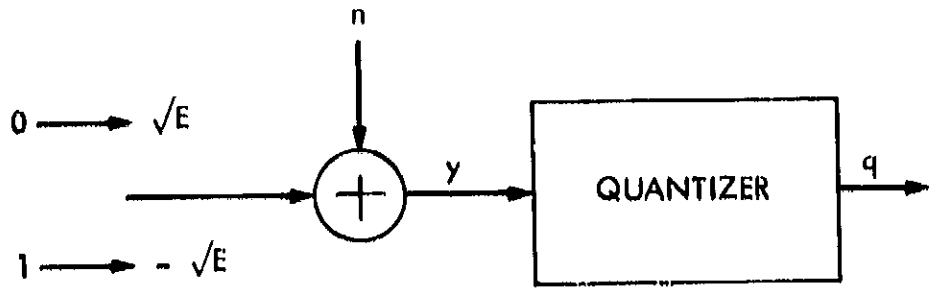
Instead of "hard quantizing" the channel output y to obtain the hard decision coding channel of Figure 44b, suppose we now assume y is the coding channel output. Under this assumption, we then have a "soft decision" channel with a cutoff rate given by

$$r_0 = 1 - \log_2 \left[1 + e^{-rE_b/N_0} \right] \quad \text{bits/coded bit.} \qquad (I.4.11)$$

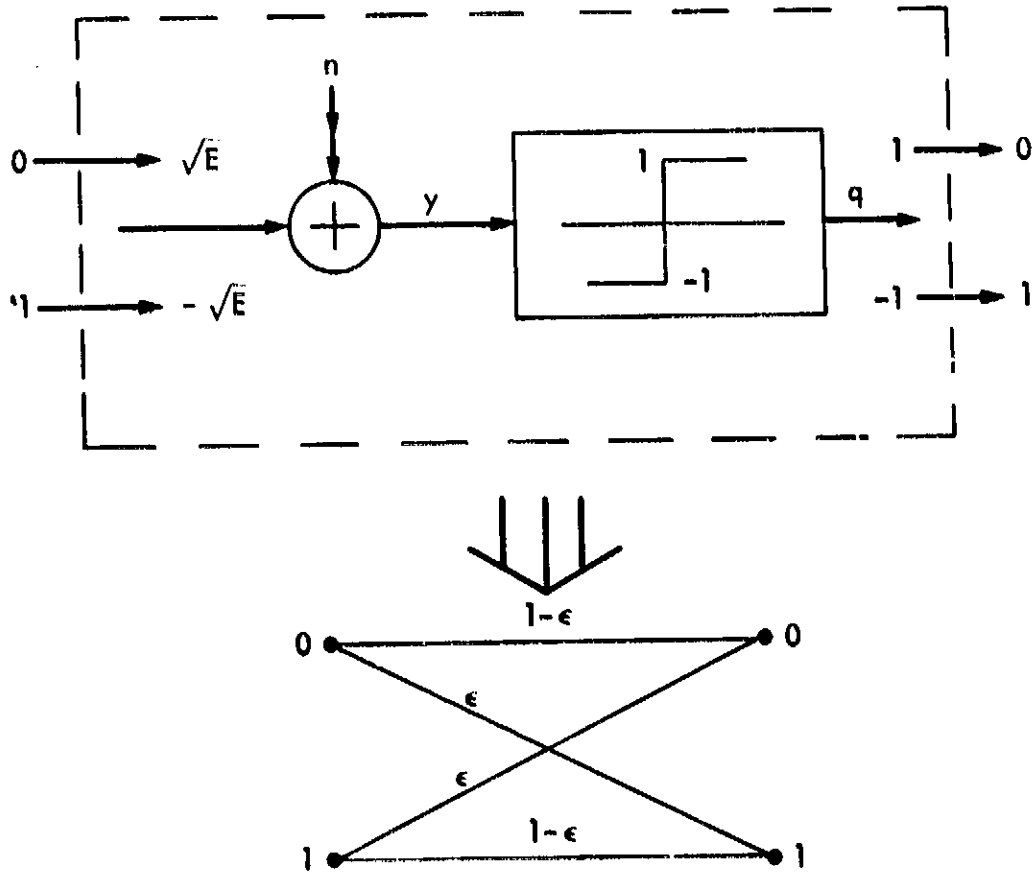
This soft decision cutoff rate yields the same cutoff rate as the hard decision channel (see (I.4.5) together with (I.4.2)) with approximately 2 dB less E_b/N_0. Thus, in terms of overall performance for BPSK modulations, the soft decision channel results in approximately 2 dB better performance than the corresponding hard decision channel.

In practice it is difficult to work with real numbers such as the channel output y if lots of digital processing is required. Typically, y is quantized to 8 levels or 3 bits as shown in Figure 45, which is drawn for the specific case of a hard decision channel crossover probability $\varepsilon = .06$. In general, 3 bit quantization results in only approximately .1 dB loss compared to no quantization at all.

Examples of the bit error probability versus E_b/N_0 for various rate 1/2 codes are shown in Figure 46. These are all for the BPSK coding channel with either hard or soft decisions. The Golay and Quadratic Residue codes are block



(a) GENERAL BPSK CODING CHANNEL



(b) HARD DECISION BPSK CODING CHANNEL

Figure 44. BPSK coding channels

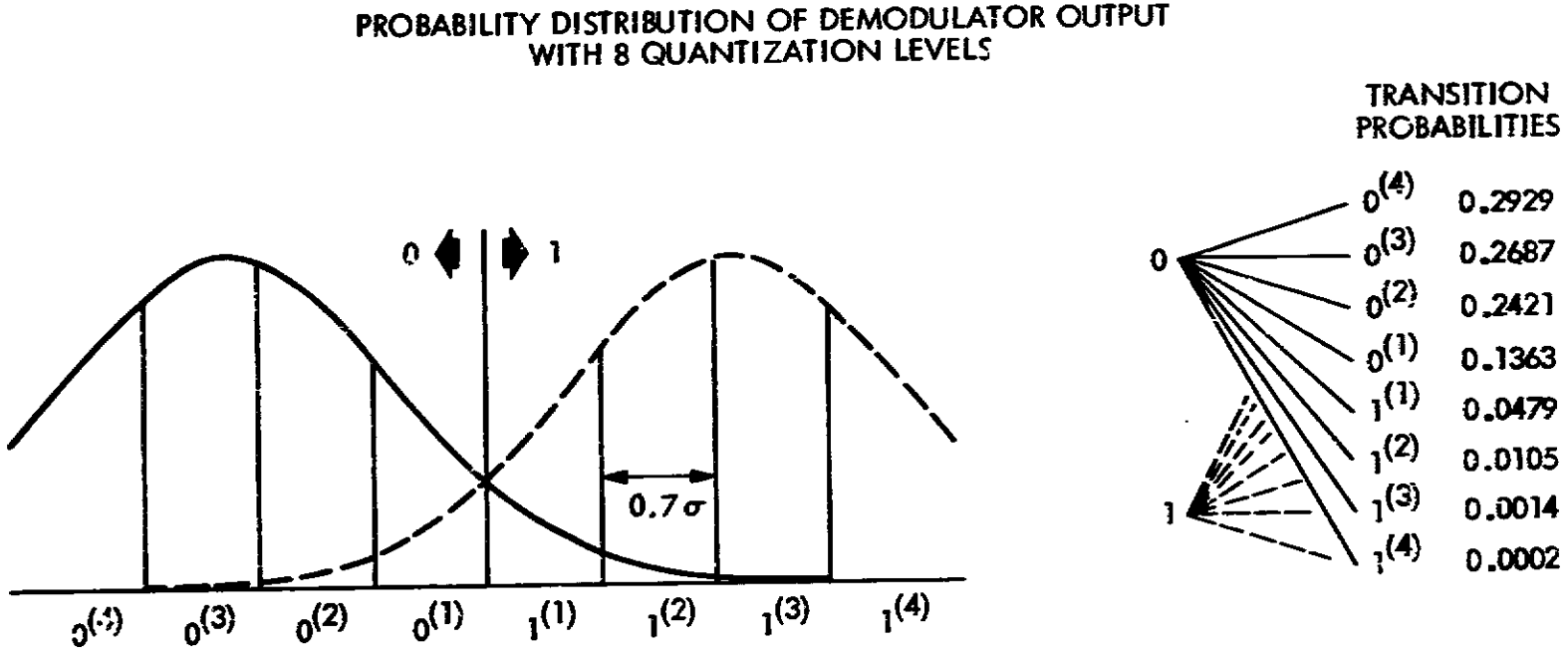


Figure 45. 3 bit quantized BPSK coding channel; $\epsilon = 0.06$

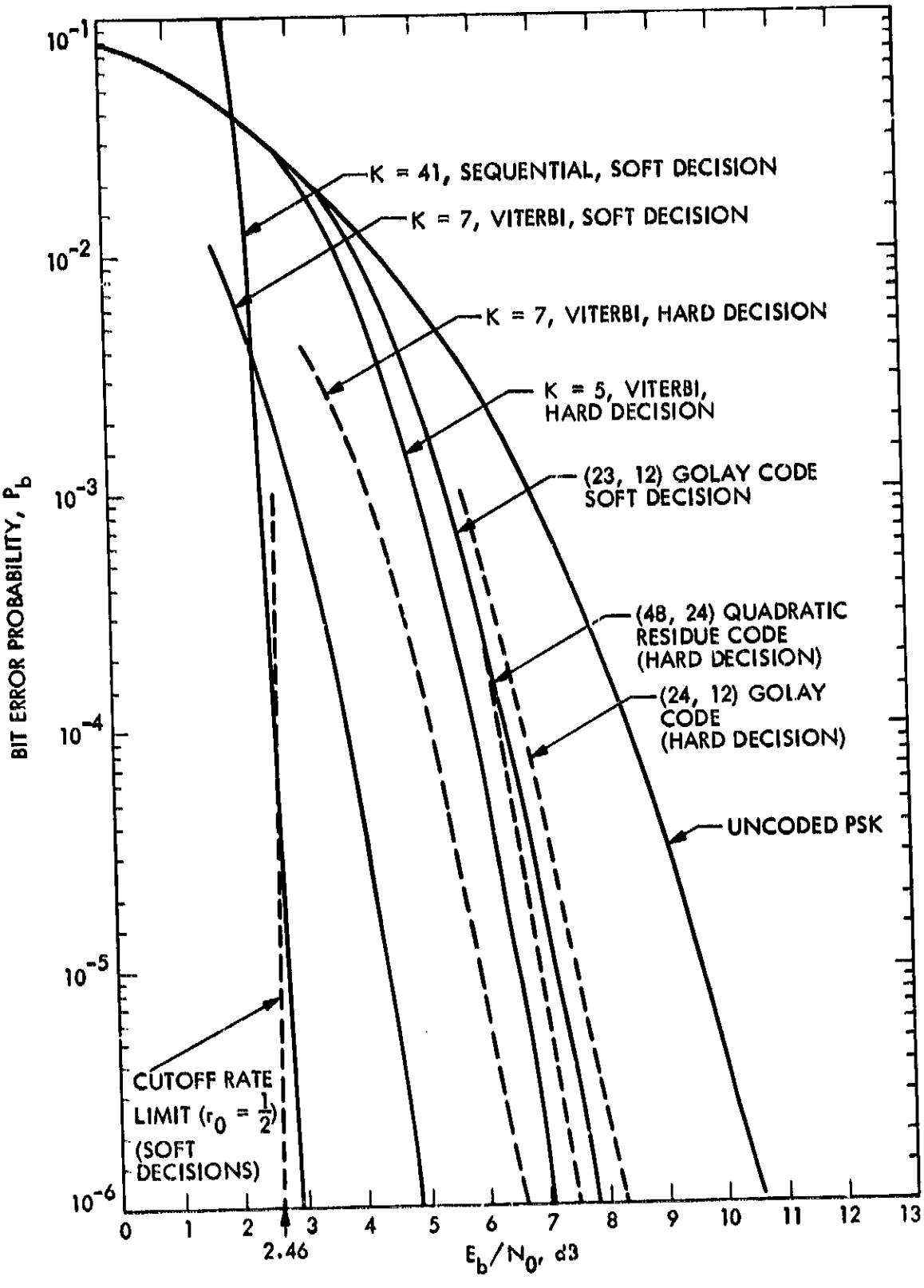


Figure 46. Bit error probability performance of various rate 1/2 codes

codes parameterized by (N,K) where there are $M = 2^K$ codewords of block length N. The remaining curves show various convolutional codes with Viterbi decoding and sequential decoding. The uncoded BPSK curve is also shown. In Figure 47 we compare rates $r = \frac{1}{2}$ and $r = \frac{1}{3}$ for various convolutional codes. Note that at bit error rates of 10^{-6} there is 4 to 7 dB in E_b/N_0 to be gained by applying currently used forward error correction coding techniques to BPSK modulation. Complexity and speed requirements determine the choice of the particular coding technique.

The performance of commercially available sequential and Viterbi decoders is tabulated in Table 1-3. Listed here are the E_b/N_0 values in dB that are required to achieve the given bit error probabilities. These are compared with the values for uncoded BPSK or QPSK modulations.

4.5 Modulation Choice and Cutoff Rates

For the most part, coding techniques have been used with BPSK and QPSK modulations. In particular, the coding results presented above assumed only these two modulations. If, instead, we use any modulation technique but only consider the end-to-end modulation bit error probability P_b (as discussed in Section 2), then we have an equivalent binary symmetric channel as in Figure 35a where $\varepsilon = P_b$. With this substitution all the hard decision coding curves can be applied with slight modifications to reflect the new relationship between ε and E_b/N_0 through $\varepsilon = P_b$. Another approach is to compare the E_b/N_0 differences for a fixed $\varepsilon = P_b$ for the various modulations of interest. The hard decision coding curves of Figures 46 and 47 would then be shifted by the same amount.

A more fundamental approach to examining the impact of coding techniques for various modulations including all types of hard, soft, and quantized channel outputs is based on evaluating the cutoff rates for these various channels. Consider any modulator input alphabet $\mathcal{X}$ and coding channel output $\mathcal{Y}$. Here $\mathcal{X}$ is typically some M-ary alphabet and $\mathcal{Y}$ is determined by the channel outputs which are typically correlator outputs followed by some quantization. The coding channel is then characterized by the conditional probabilities $p(y|x), y \in \mathcal{Y}, x \in \mathcal{X}$. These can be obtained for each type of modulation and channel correlator output quantization. The cutoff rate is defined as

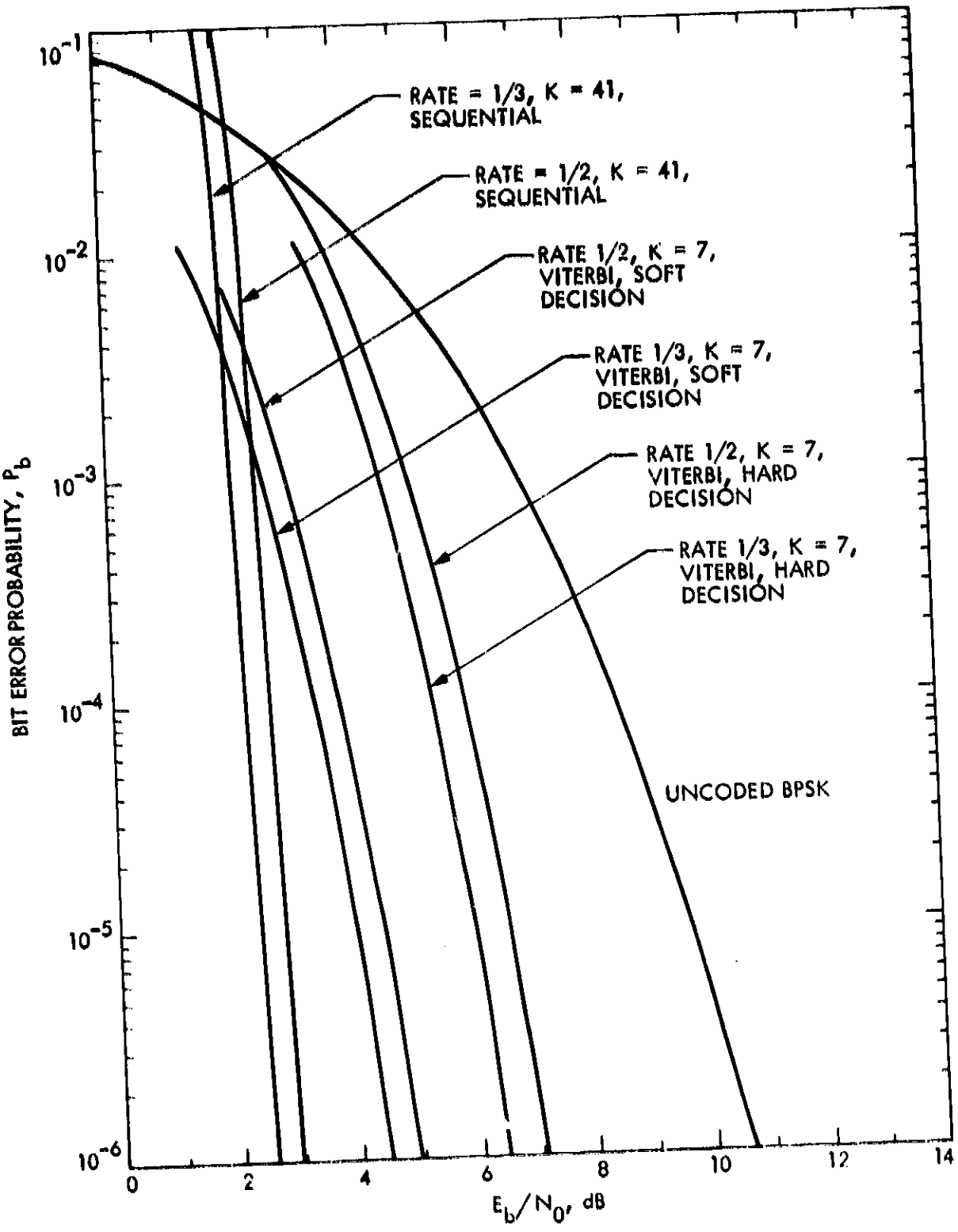


Figure 47. A comparison of the bit error probability performances of rate 1/2 and 1/3, hard and soft decision viterbi and sequential decoders

Table I-3. Performance of Commercially Available Sequential and Viterbi Decoders (Information Courtesy of Linkabit Corporation)

Bit Error Probability	E_b/N_0, dB					
	Sequential Decoder			Viterbi Decoder		Uncoded BPSK or QPSK
	Rate 1/2	Rate 3/4	Rate 7/8	Rate 1/2	Rate 3/4	
10^{-2}				2.0		4.2
10^{-3}		4.6	5.4	3.0	3.9	6.8
10^{-4}	4.1	5.0	5.8	3.8	4.7	8.3
10^{-5}	4.5	5.4	6.3	4.5	5.4	9.6
10^{-6}	4.9	5.8	6.7	5.2	6.1	10.5
10^{-7}	5.3	6.2	7.2	5.8	6.7	11.3
10^{-8}	5.7	6.6	7.6	6.4	7.4	12.0
10^{-9}	6.1	7.0	8.0	6.9	8.0	12.6
10^{-12}	7.2	8.1	9.0	8.1	9.2	14.0

<u>SEQUENTIAL DECODER</u>

50 Mbps

2 Bit Soft Quantization

Decoder Speed =

75 Mega Computations/sec

Main Memory Size: 6K Bits

<u>Viterbi Decoder</u>

10 Mbps

3 Bit Soft Quantization

K = 7 Rate 1/2

K = 9 Rate 3/4

$$r_0 = -\min_{q(x)} \left\{ \log_2 \sum_y \left[\sum_x q(x) \sqrt{p(y|x)} \right]^2 \right\} \quad \text{bits/coded bit.} \qquad (I.4.12)$$

We have already seen that for block codes, there exist codes of rate r bits per channel symbol belonging to $\mathscr{A}$ and block length N such that the bit error probability is bounded as in (I.4.6). Similarly, as a generalization of (I.4.10), there exist convolutional codes of rate r = b/n whose bit error probability has the bound

$$P_b \leq A(b,n,r)2^{-Nr_0} \qquad (I.4.13)$$

where

$$N = nK$$

$$A(b,n,r) = \frac{2^b - 1}{\left(1-2^{-b}\left|\left(r_0/r\right)-1\right|\right)^2}$$

$$K = \text{constraint length} \qquad (I.4.14)$$

Hence, for any modulation there exist codes that have bit error bounds identical to those for the BPSK modulation where the only difference is the value or r_0. Thus, using the r_0 versus ϵ curve for BPSK (Figure 36) as a baseline, we can measure the effectiveness of coding with other modulations by comparing their cutoff rates, as computed from (I.4.12).

For coherent MPSK with soft decisions, we have the cutoff rates illustrated in Figure 48. As E_b/N_0 increases, it is clear that higher order alphabets yield higher values of cutoff rate and thus higher coded data rates. For noncoherent MFSK with hard decisions, similar results are illustrated in Figure 49.

For the CPM techniques, it is important to normalize with respect to some common bandwidth. In Figure 50 we use the 99% bandwidth criterion defined by (I.2.8) with the equality sign. This shows the use of a rectangular filter for several modulation indices, h, alphabets, $\mathscr{A}$, and pulse memory length, L,

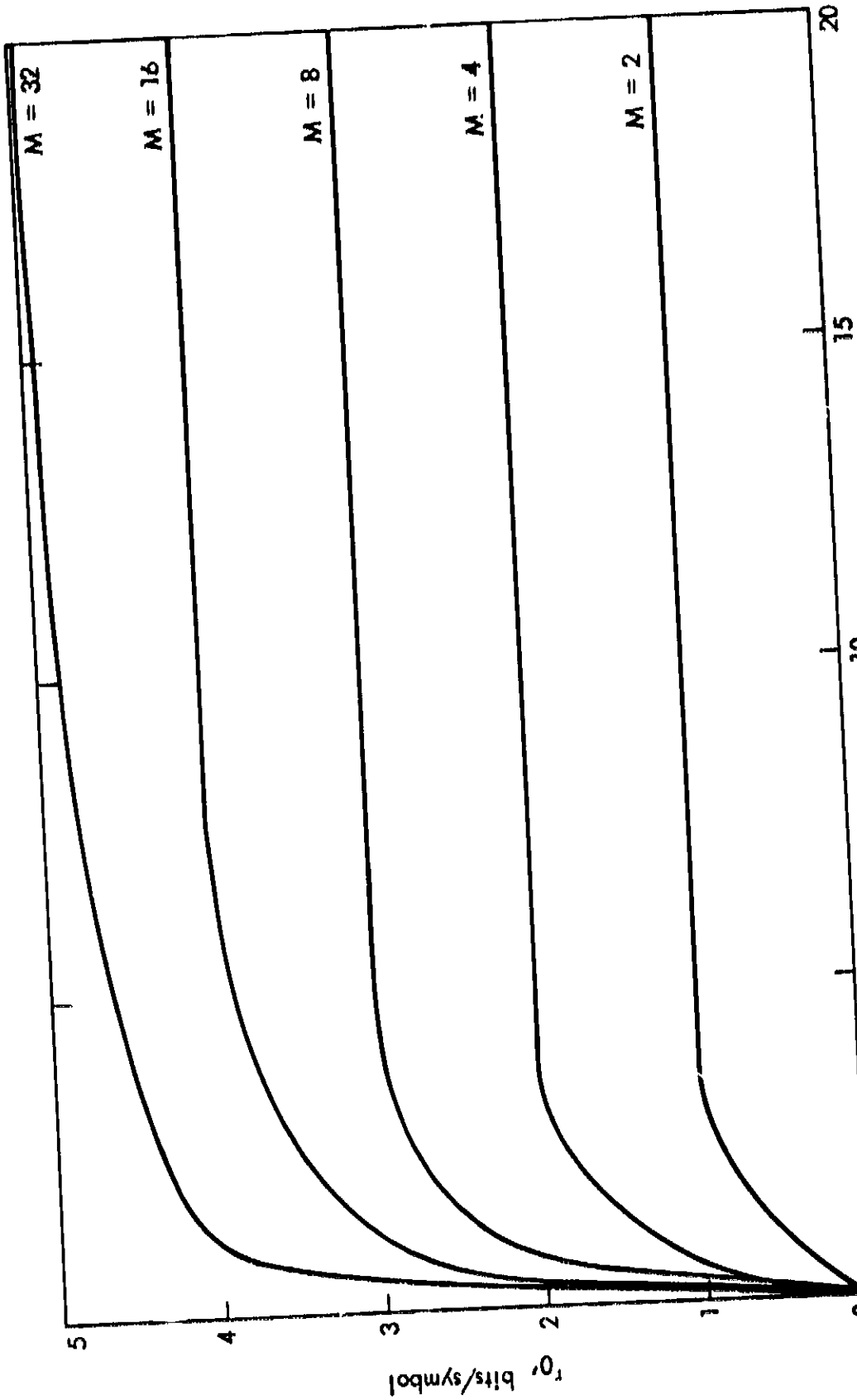


Figure 48. Soft decision cutoff rate for MPSK

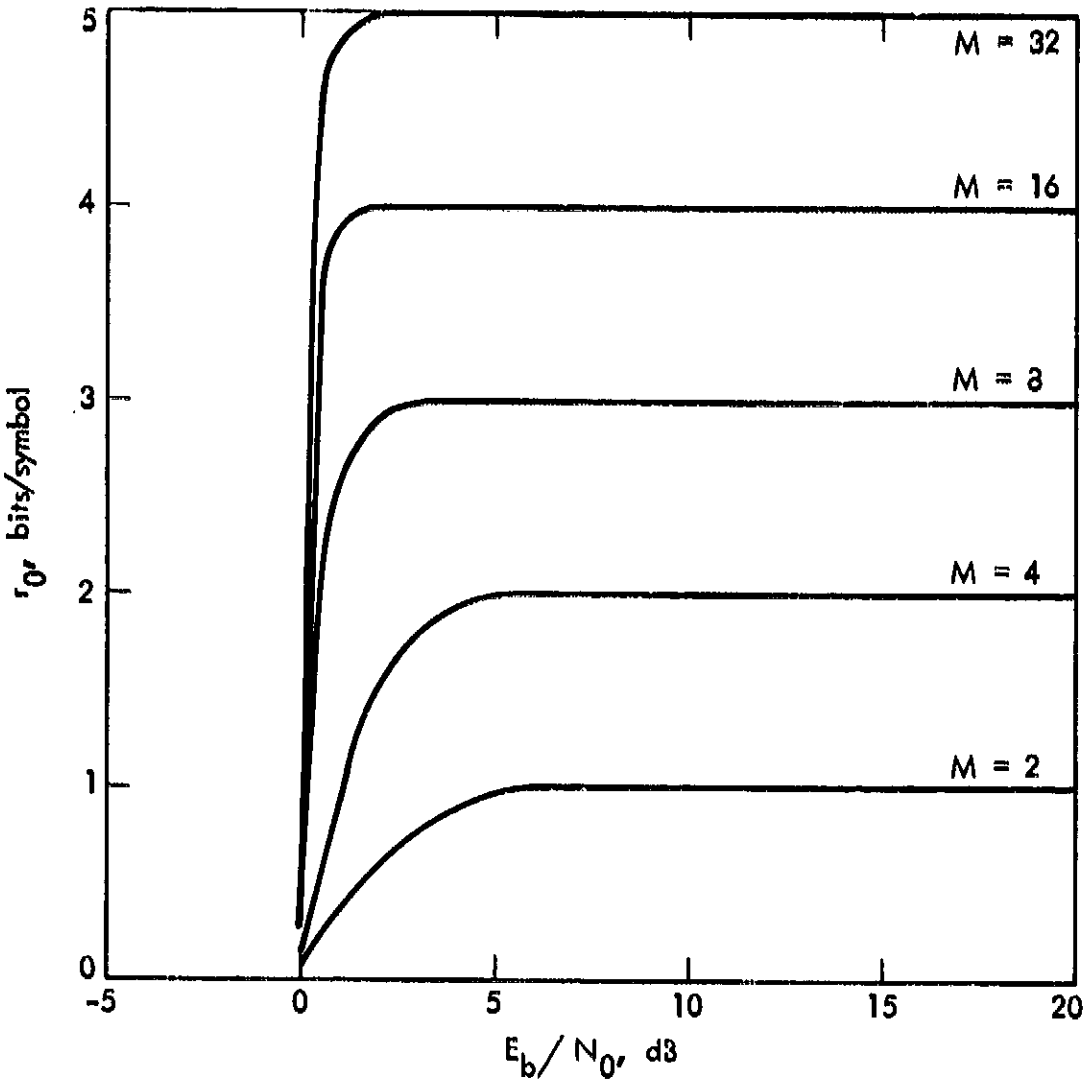


Figure 49. Hard decision cutoff rate for noncoherent MFSK

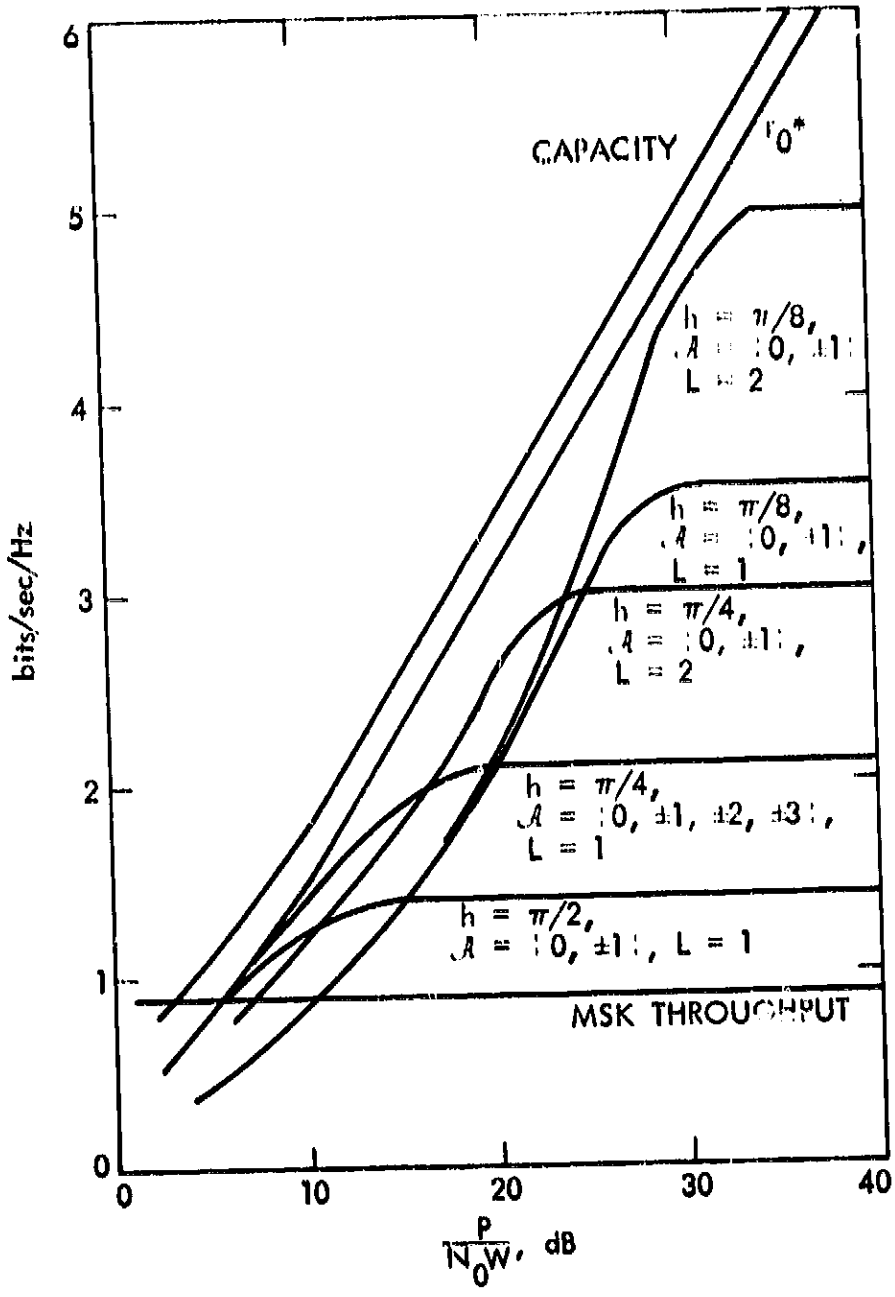


Figure 50. r_0 versus signal-to-noise ratio for coherent, maximum likelihood demodulation, 99 percent bandwidth, rectangular filter (reprinted from [16])

in symbols. The cutoff rates here have been normalized by the bandwidth and as such are measured in bits per second per Hz. It is important to note in this figure how close these CPM techniques are to the theoretically maximum value denoted r_0^* even though these signals are restricted to have constant envelope.

If we were to impose the more restrictive FCC bandwidth constraints discussed in Section 2.3, then the cutoff rate parameters for rectangular, triangular, and raised cosine filters are shown in Figures 51 through 53. In these figures the curves are labeled XZZY where

$$
\begin{aligned}
Y &= \text{L, pulse memory in symbols} \\
ZZ &= \begin{cases} \text{RE, rectangular pulse} \\ \text{TR, triangular pulse} \\ \text{RC, raised cosine pulse} \end{cases} \\
X &= \begin{cases} 1,\ h=\pi/2,\ \mathcal{A}=\{0,\pm 1\} \\ 2,\ h=\pi/4,\ \mathcal{A}=\{0,\pm 1,\pm 2,\pm 3\} \\ 3,\ h=\pi/8,\ \mathcal{A}=\{0,\pm 1,\pm 2,\pm 3,\pm 4,\pm 5\} \end{cases}
\end{aligned}
\qquad \text{(I.4.15)}
$$

4.6 Concatenation of Codes

Although in theory it is possible to achieve arbitrarily small coded error probabilities as long as the data rate is less than the channel capacity or cutoff rate, in practice we are limited by the processing complexity and speed required in channel decoding. For this reason convolutional codes are generally limited to constraint lengths $K \leq 10$. To achieve even smaller error rates with forward error correction codes, Forney [36] proposed using a concatenation of codes.

Figure 54 shows the most popular form of concatenation of codes. Basically, this approach is to now create a new coding channel as in Figure 34, only here the new coding channel begins at the input to a convolutional encoder and ends at the Viterbi decoder output. The convolutional encoder with the Viterbi decoder that forms part of the new coding channel is referred to as the "inner code." This inner code does not have to be restricted to a convolutional code with a Viterbi decoder. Likewise, the outer code does not have to be a Reed-Solomon code. However, these are the most common types of inner and outer codes.

C-2

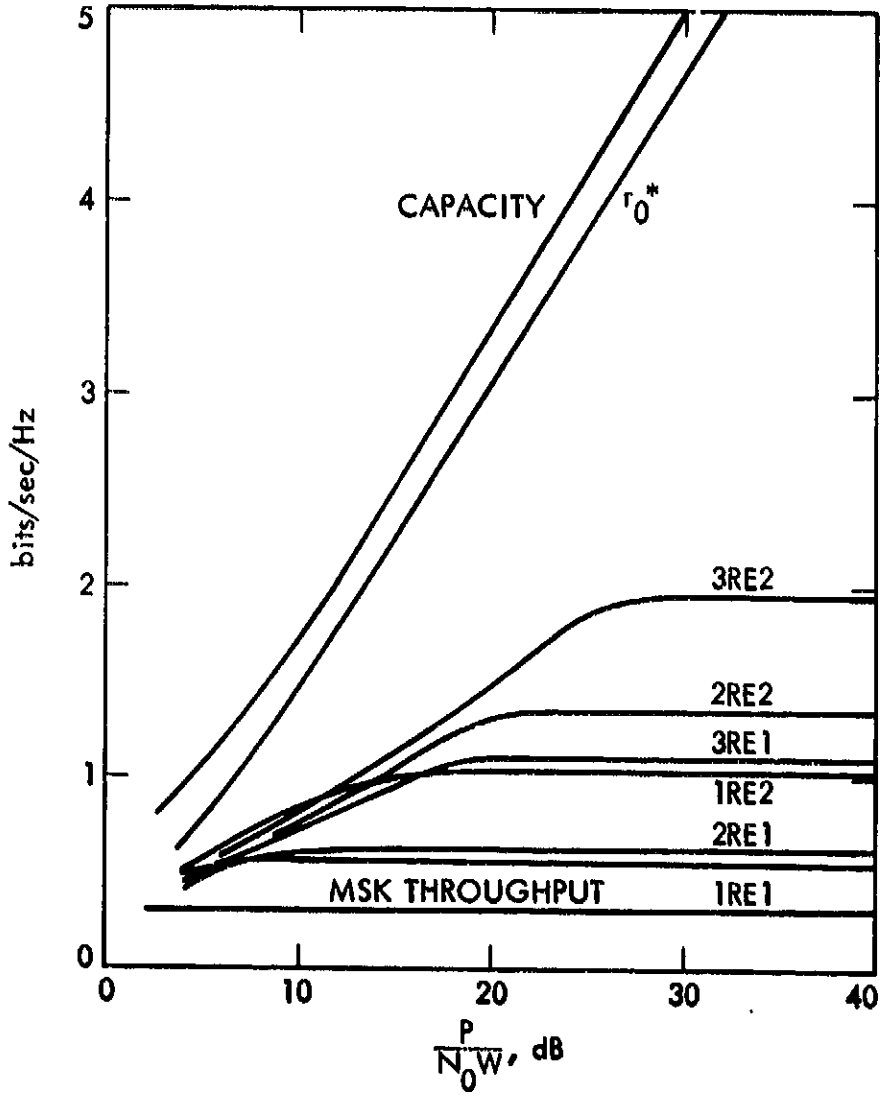


Figure 51. r_0 versus signal-to-noise ratio for coherent, maximum likelihood demodulation, FCC bandwidth, rectangular filter (reprinted from [16])

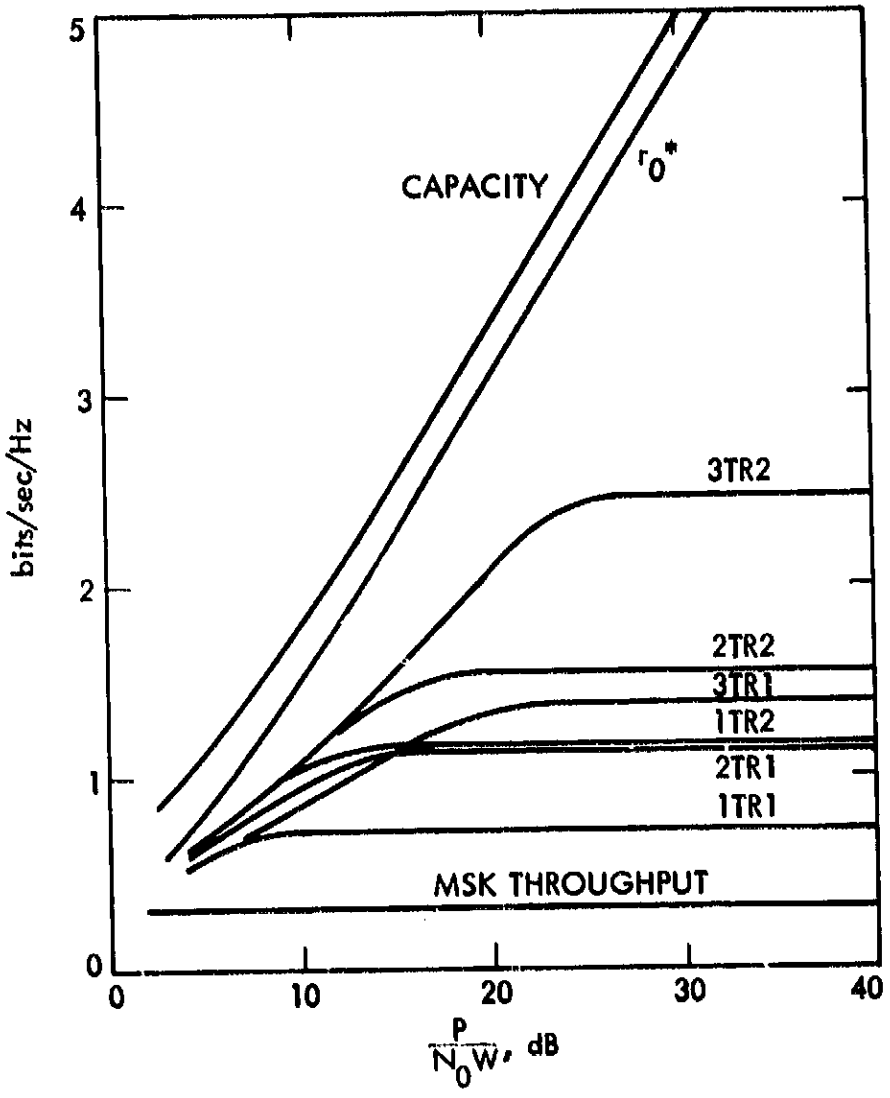


Figure 52. r_0 versus signal-to-noise ratio for coherent, maximum likelihood demodulation, FCC bandwidth, triangular filter (reprinted from [16])

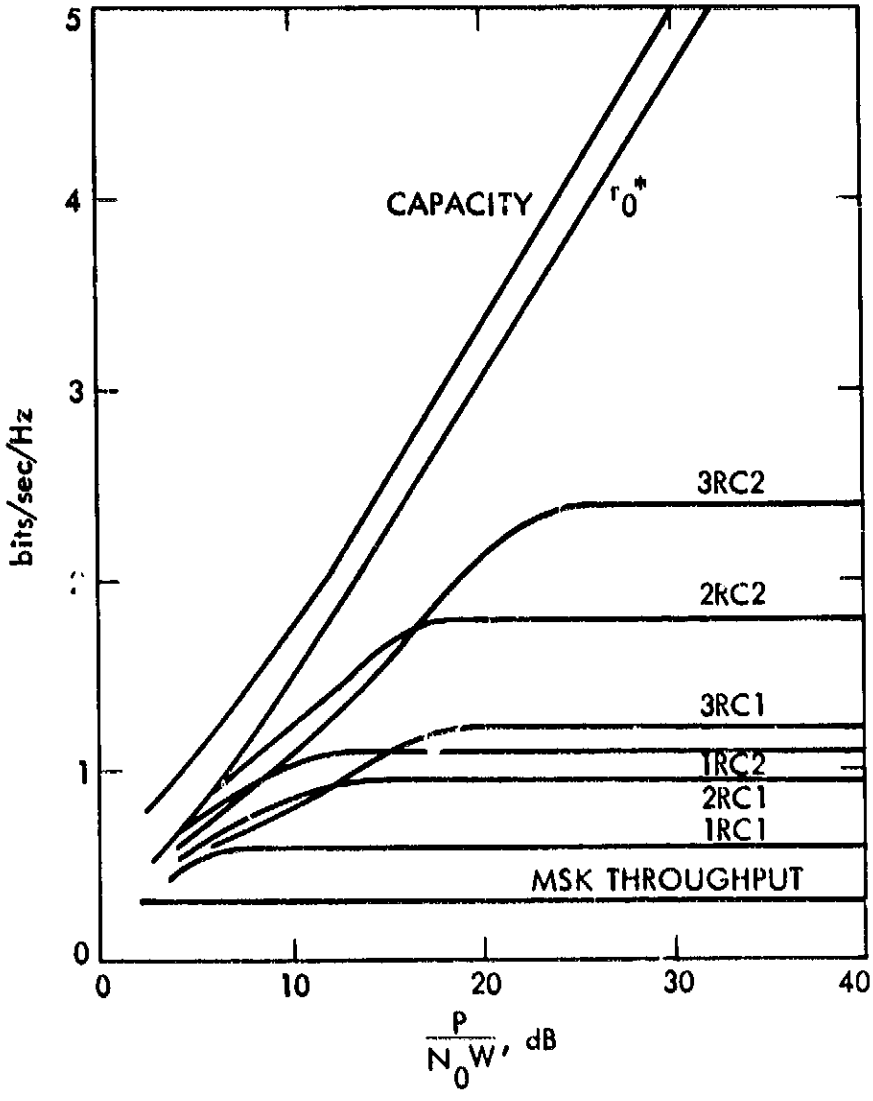


Figure 53. r_0 versus signal-to-noise ratio for coherent, maximum likelihood demodulation, FCC bandwidth, raised cosine filter (reprinted from [16])

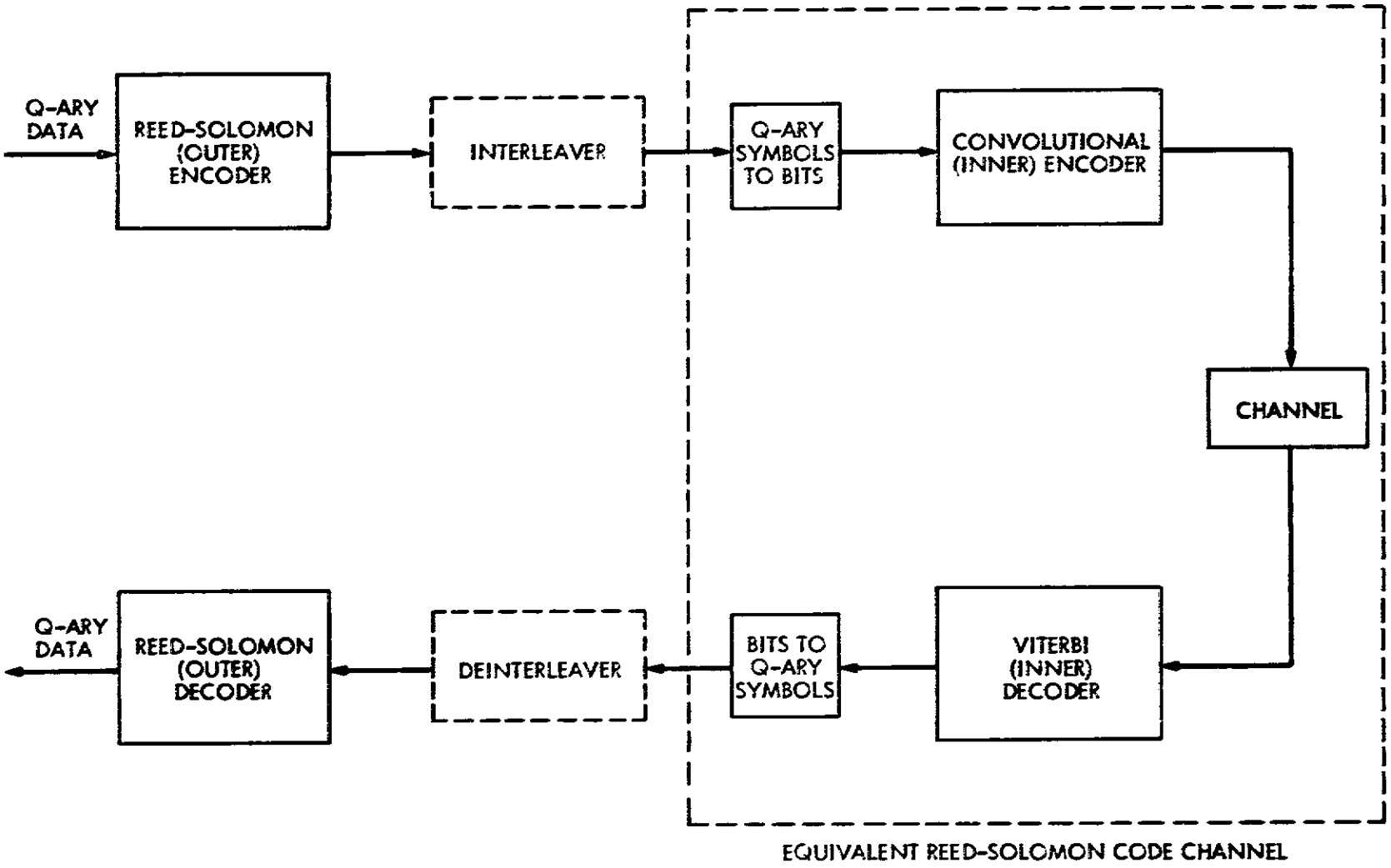


Figure 54. Concatenated coding channel

The inner code reduces the error probabilities of the first coding channel (see Figure 34) which consists of the modulator, radio channel, and demodulator. This then forms a new coding channel for another "outer code" which can further reduce the error rate. Thus, we can use the most powerful available coding systems and reduce error rates by concatenation of these codes.

Figures 55 and 56 show how decoded bit errors out of a Viterbi decoder tend to occur in bursts of various lengths. Here we assume the BPSK or QPSK modulation with 3 bit quantization. Generally, the decoded bits are error free for a while and then when a decoding bit error occurs the errors occur in a burst or string of length L_b with a probability distribution that is geometric [37], i.e.,

$$\Pr\{L_b=m\} = p(1-p)^{m-1} \quad ; \; m = 1,2,... \tag{1.4.16}$$

where

$$p = \frac{1}{\bar{L}_b} \tag{1.4.17}$$

and $\bar{L}_b$ is the average burst length in data bits. The waiting time, W, between bursts has the empirical distribution

$$\Pr\{W = n\} = q'(1-q')^{n-K+2} , \qquad n = K+1,K+2,... \tag{1.4.18}$$

where

$$q' = \frac{1}{\bar{W}-K+2} . \tag{1.4.19}$$

In (1.4.19), $\bar{W}$ is the average waiting time and K is again the convolutional code constraint length.

By choosing a Reed-Solomon (RS) outer code we can take advantage of the bursty nature of the bit errors out of the inner code. RS codes use higher order q-ary symbols where typically $q = 2^m$ for some integer m. In most applications m = 8 (q = 256). By taking m = 8 bits to form a single q-ary symbol, a burst of errors in the 8 bits results in only one q-ary symbol error. In this way we tend to reduce the impact of bit error bursts. In Figure 57 we show the conversion from q-ary symbols to

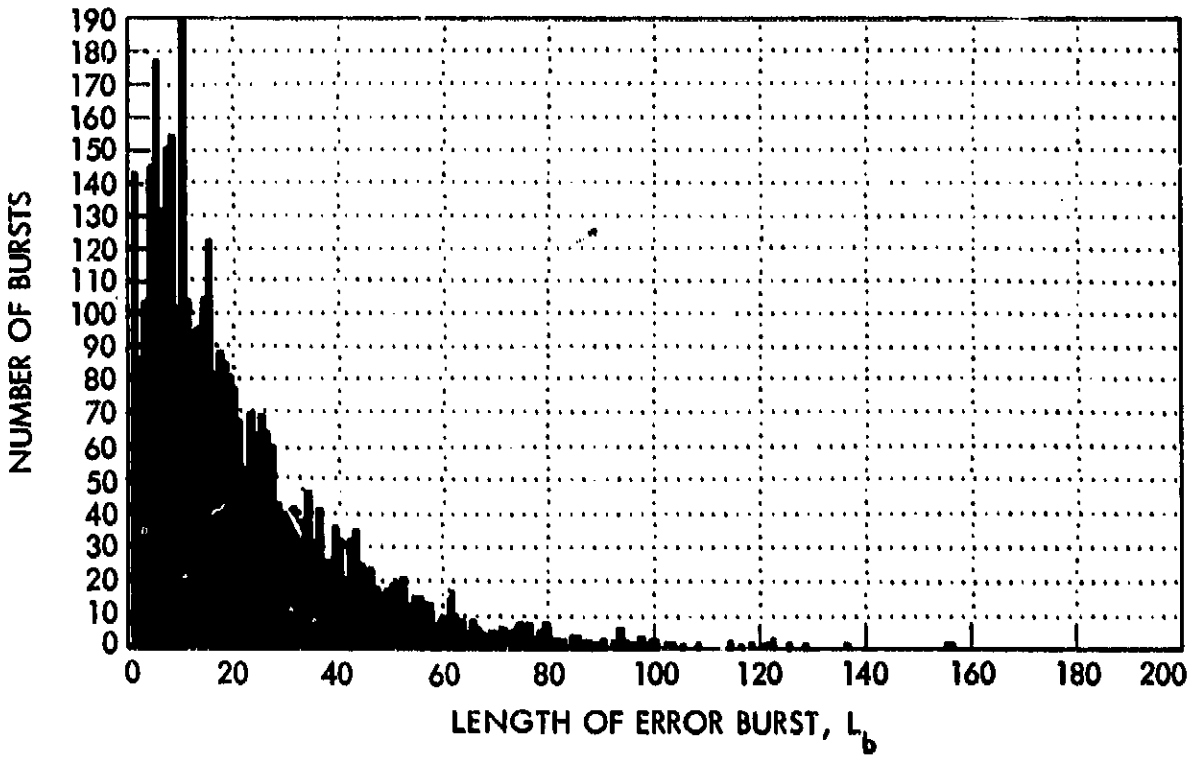


Figure 55. Histogram of burst lengths; Viterbi decoded constraint length 7, rate 1/2 convolutional code; $E_b/N_0 = 1.0$ (courtesy of R.B. Miller, JPL)

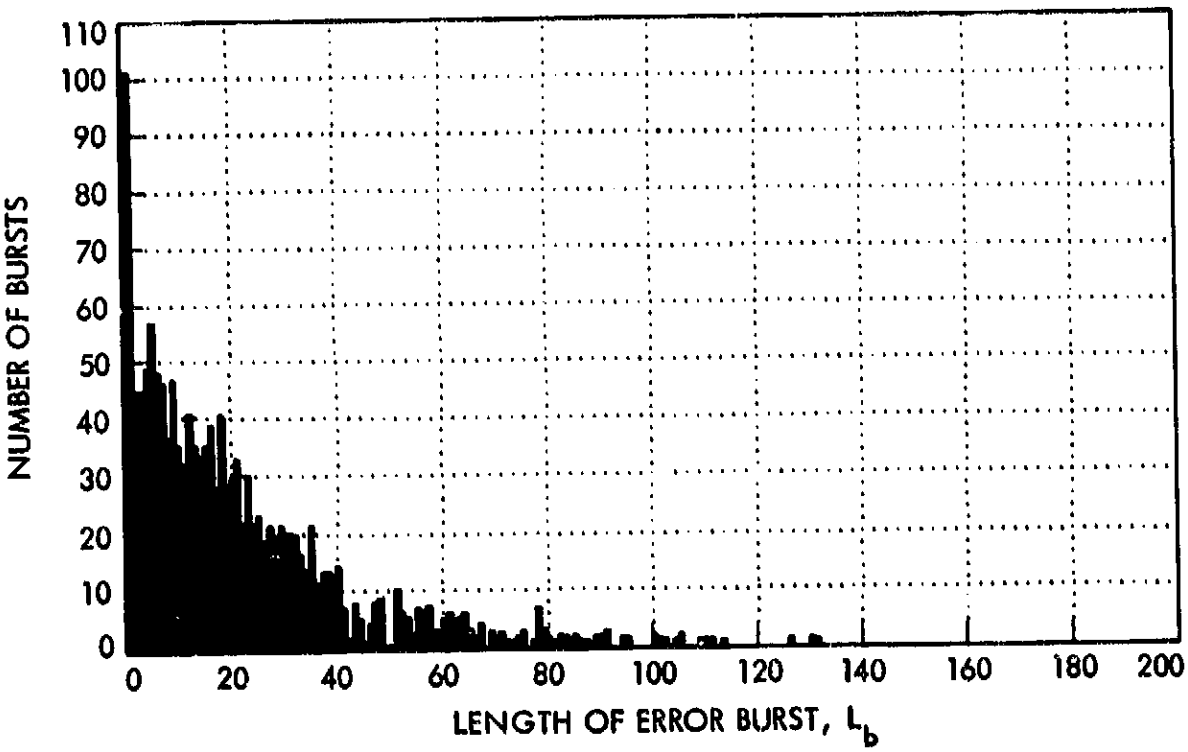


Figure 56. Histogram of burst lengths; Viterbi decoded constraint length 10, rate 1/3 convolutional code; $E_b/N_0 = 0.75$ (courtesy of R.B. Miller, JPL)

bits and back again for the coding channel of the RS outer code. In addition, to avoid bursts of q-ary symbols we may introduce interleavers and deinterleavers to provide a memoryless (non-bursty) q-ary coding channel for the RS code.

As mentioned above, the most commonly employed RS code has m = 8 (q = 256) and a block length of N = q - 1 = 255 q-ary symbols. To be able to correct up to t = 16 q-ary symbols we choose the number of data q-ary symbols to be K = q - 1 = 2t = 223. Thus we have (255,223) block code using 256-ary symbols that can correct up to t = 16 symbols. This is equivalent to a (2040,1784) block code that can correct up to t = 128 bits in error as long as these bits are confined to at most 16 q-ary symbols where q = 256.

In Figures 57 and 58, we show the performance of the concatenation system with no interleaving for two convolutional codes with Viterbi decoding. Ideal interleaving is assumed in Figure 59. These curves show the currently most powerful (non-sequential) coding technique available. Sequential decoding of convolutional codes with large constraint lengths can also achieve similar performance but with the possibility of losing data due to buffer overflows.

5.0 Discussion

All the results presented here have assumed the ideal additive white Gaussian noise channel of Figure 2a. In practice, for terrestrial radio links and telephone cables, the real channel can be more accurately modelled as the interference channel of Figure 2b, whereas for the satellite links, we have the nonlinear channel model of Figure 3. Partial results for these more complex channels have previously been obtained primarily for the simple BPSK and QPSK modulations. These results along with those for more sophisticated modulation and coding techniques employed on these more general channel models are the subject of the next parts of the report.

In the meantime, the results presented here for the ideal additive white Gaussian noise channel have provided us with insights into choosing modulation and coding techniques for digital terrestrial and satellite links and allow us to draw the following conclusions:

(a) Modulation and coding ought to be jointly optimized rather than separately selected as is too often done today.

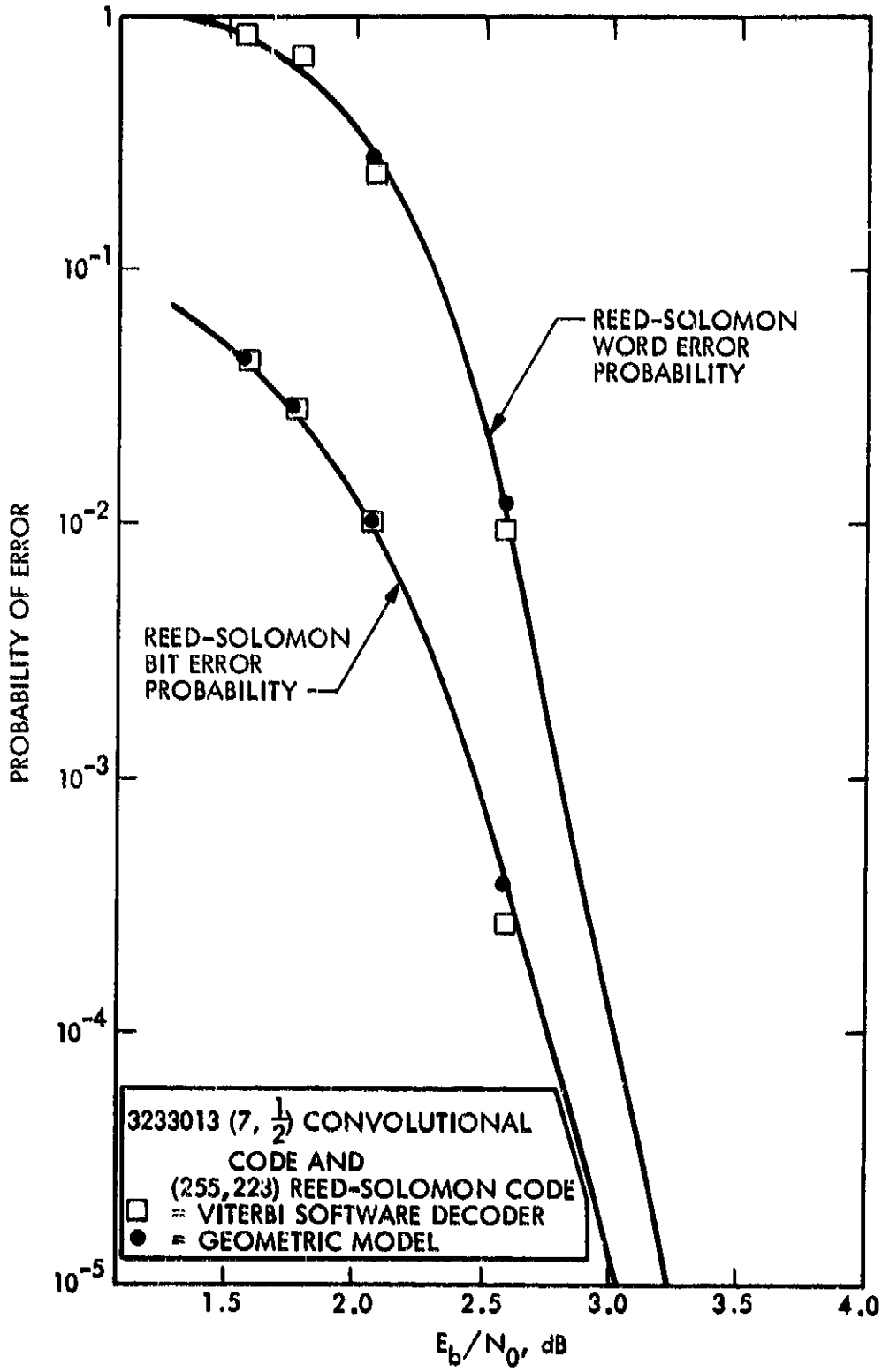


Figure 57. Non-interleaved performance statistics for concatenated coding scheme assuming no system losses; (7, 1/2) convolutional code (reprinted from [37])

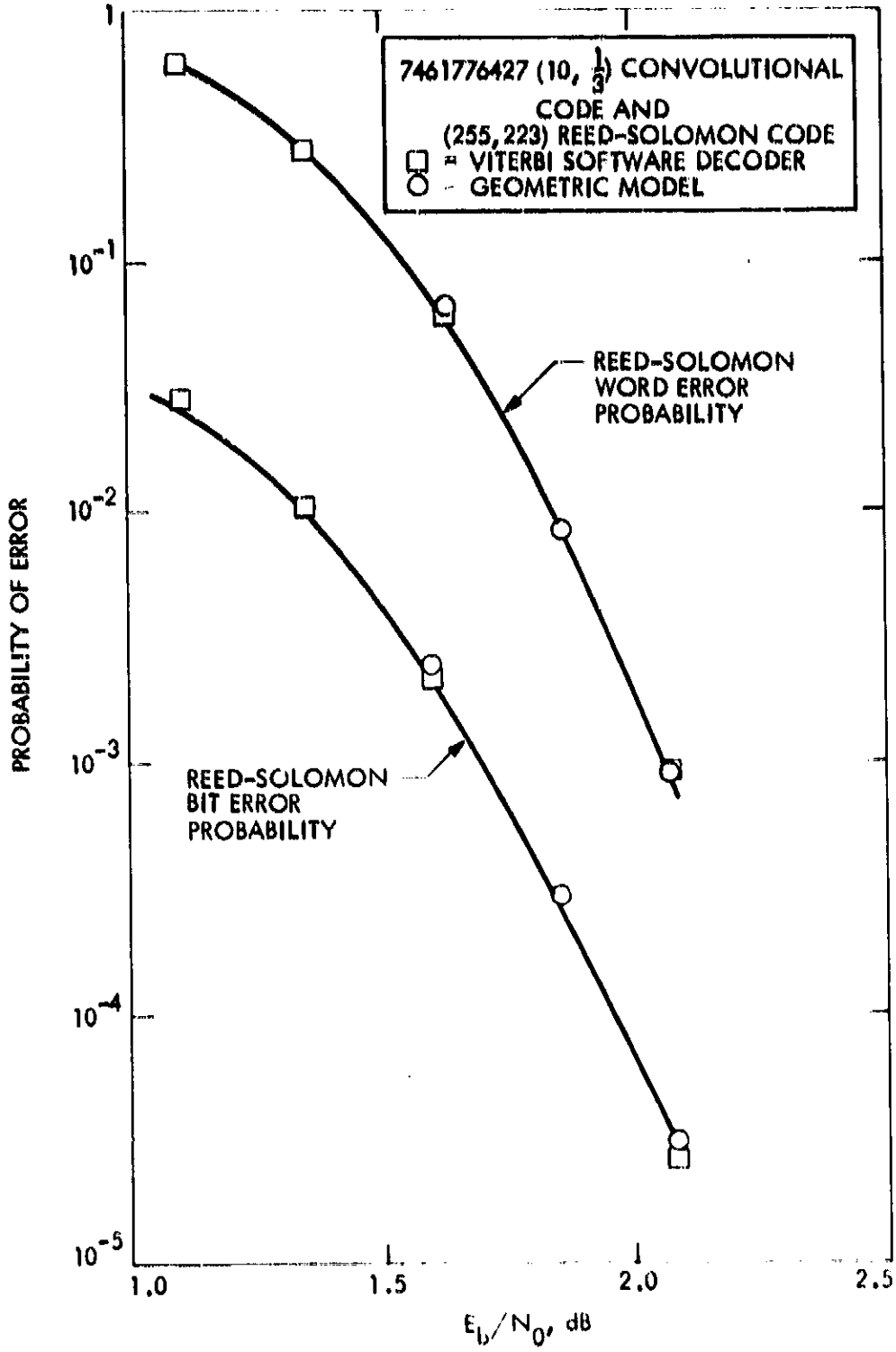


Figure 58. Non-interleaved performance statistics for concatenated coding scheme assuming no system losses; (10, 1/3) convolutional code (reprinted from [37])

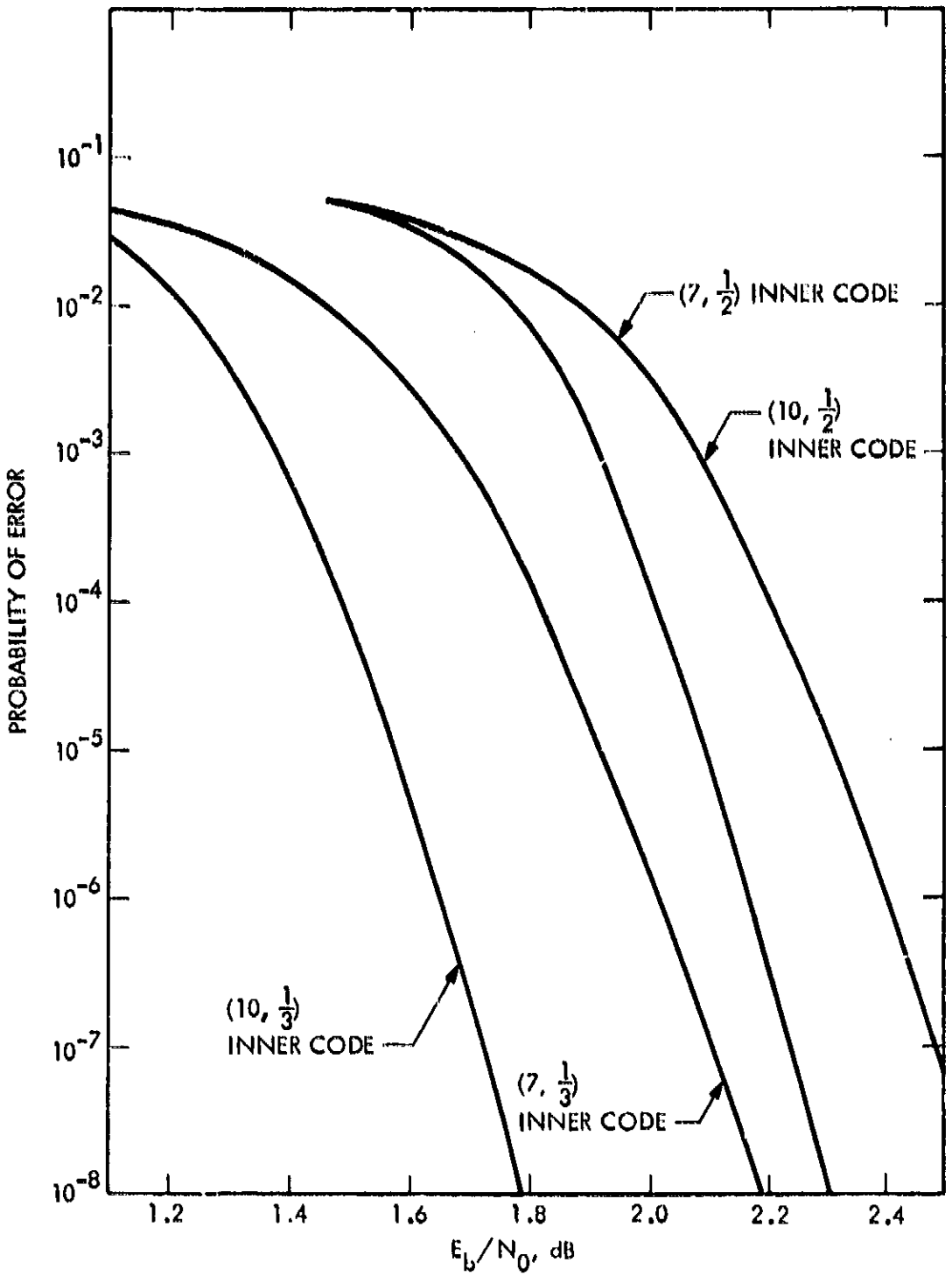


Figure 59. Comparison of concatenated channel decoder bit error rates for several convolutional inner codes and a Reed-Solomon (255,223) outer code with ideal interleaving assuming no system losses (reprinted from [37])

(b) Constant envelope modulations, particularly the staggered types, have less filter/nonlinearity distortions and should be used with nonlinear channels such as satellite channels.

(c) Continuous phase modulations (CPM) can achieve bandwidth requirements (99% or FCC) without RF filtering. These constant envelope signals can achieve high throughputs with coding but require complex digital processing at the receiver.

(d) Convolutional coding with Viterbi decoding can usually provide adequate reduction in bit error rates with minimal complexity. New convolutional codes, however, must be found for use with CPM.

(e) Very low error rates can be achieved in practice with sequential decoding of long constraint length convolutional codes and with concatenation of codes. Speed, buffer overflow, and complexity limit these applications.

(f) The rapid evolution in solid-state electronic technology will continue to have a big impact on the design of communication systems. This will likely mean that CPM and more powerful coding techniques will become more commonly used in communication systems to achieve higher data rates at lower error rates.

REFERENCES

1. A. J. Viterbi and J. K. Omura, Principles of Digital Communication and Coding, McGraw-Hill, N.Y., 1979.

2. R. Morris, et al., "Assessment of the NBS Proposed Federal Data Encryption Standard," Cryptologia, Vol. 1, No. 3, July 1977.

3. W. Diffie and M. E. Hellman, "New Directions in Cryptography," IEEE Trans. on Information Theory, Vol. IT-22, No. 6, November 1976, pp. 644-654.

4. R. C. Dixon, Spread Spectrum Systems, John Wiley, N.Y., 1976.

5. L. G. Roberts, "Computer Network Development to Achieve Resource Sharing," AFIPS Conference Proceedings Spring Joint Computer Conference, Vol. 36, 1970, p. 543.

6. A. J. Viterbi, Principles of Coherent Communication, McGraw-Hill, N.Y., 1966.

7. R. W. Lucky, J. Salz, and E. J. Weldon, Principles of Data Communication, McGraw-Hill, N.Y., 1968.

8. M. C. Austin and M. U. Chang, "Quadrature Overlapped Raised-Cosine Modulation," IEEE Transactions on Communications, Vol. COM-29, No. 3, pp. 237-249.

9. D. Divsalar and M. K. Simon, "Performance of Quadrature Overlapped Raised-Cosine Modulation over Nonlinear Satellite Channels," ICC'81 Conference Record, Denver, Colorado, June, 1981, pp. 2.3.1-2.3.7.

10. D. Divsalar and M. K. Simon, "The Power Spectral Density of Digital Modulations Transmitted Over Nonlinear Channels," to be published in the IEEE Transactions on Communications.

11. J. B. Anderson and D. O. Taylor, "A bandwidth-efficient class of signal-space codes," IEEE Trans. on Information Theory, Vol. IT-24, Nov. 1978, pp. 703-712.

12. C-E. Sundberg, T. Aulin, and N. Rydbeck, "Recent Results on Spectrally Efficient Constant Envelope Digital Modulation Methods"; "M-ary CPFSK Type of Signalling With Input Data Symbol Pulse Shaping - Minimum Distance and Spectrum"; "Bandwidth Efficient Digital FM With Coherent Phase Tree Demodulation," ICC'79 Conference Record, Boston, MA, June 1979, 42.1.1-42.1.6, 42.3.1-42.3.6, 42.4.1-42.4.6.

13. J. B. Anderson, "Error Bounds for Phase Modulation Codes With a Severe Bandwidth Constraint," NTC'79 Conference Record, Washington, DC, Nov. 1979, pp. 55.2.1-55.2.5.

14. T. Aulin, "Three Papers on Continuous Phase Modulation," Ph.D. Dissertation, Telecommunication Theory Dept., Univ. of Lund, Lund, Sweden, 1979.

15. J. B. Anderson, "Error Bounds for Smooth Phase Modulation Codes," ICC'80 Conference Record, Seattle, WA, June 1980, pp. 26.3.1-26.3.5.

16. D. E. Jackson, "Bandwidth Efficient Modulation and Coding," Ph.D. Dissertation, UCLA, 1980.

17. J. K. Omura and D. E. Jackson, "Cutoff Rates for Channels Using Bandwidth Efficient Modulations," NTC'80 Conference Record, Houston, Texas, Nov. 1980, pp. 14.1.1-14.1.11.

18. J. B. Anderson, C-E. Sundberg, T. Aulin, and N. Rydberg, "Smoothed Phase Modulation Codes: Power Vs. Bandwidth," NTC'80 Conference Record, Houston, Texas, Nov. 1980, pp. 14.6.1-14.6.6.

19. T. Aulin, N. Rydbeck, C-E. Sundberg, "Performance of Constant Envelope M-ary Digital FM Systems and Their Implementation," NTC'79 Conference Record, Washington, DC, Nov. 1979, pp. 55.1.1-55.1.6.

20. C. R. Cahn, "Phase Tracking and Demodulation With Delay," IEEE Trans. on Information Theory, Vol. IT-20, January 1974, pp. 50-58.

21. G. Ungerboeck, "New Applications for the Viterbi Algorithm: Carrier Phase Tracking in Synchronous Data-Transmission Sytems," NTC'74 Conference Record, San Diego, Calif., Dec. 1974, pp. 734-738.

22. L. L. Scharf, "A Viterbi Algorithm for Modulo-2π Phase Tracking in Coherent Data Communication Systems," ONR Tech. Rep. #25, December 1977.

23. L. L. Scharf, D. D. Cox and C. J. Masreliez, "Modulo-2π Phase Sequence Estimation," ONR Tech. Rep. #27, February 1978.

24. L. L. Scharf, D. D. Cox and C. J. Masreliez, "Modulo-2π Phase Sequence Estimation," IEEE Trans. on Information Theory, IT-26, pp. 615-620, September 1980.

25. T. Aulin, "Viterbi Detection of Continuous Phase Modulated Signals," NTC'80 Conference Record, Houston, Texas, Nov. 1980, pp. 14.2.1-14.2.7.

26. D. E. Meer, "Phase Sequence Estimation for Laser Line-Scan Imagery in the Presence of Rayleigh Fading," M.S. Thesis, CE-79D-24, Department of Electrical Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH 45433, December 1979, ADA 080368.

27. S. R. Robinson and D. E. Meer, "Phase Sequence Estimation in the Presence of Rayleigh Fading," NTC'80 Conference Record, Houston, Texas, Nov. 1980, pp. 58.2.1-58.2.6.

28. W. C. Lindsey and M. K. Simon, Telecommunication Systems Engineering, Prentice-Hall, Englewood Cliffs, N.J., 1973.

29. R. F. Pawula, S. O. Rice, and J. H. Roberts, "Distribution of the Phase Angle Between Two Vectors Perturbed by Gaussian Noise," submitted for publication.

30. FCC Filing, TYMNET Corp., May 12, 1981.

31. C. E. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal, Vol. 27, 1948, pp. 379-423.

32. A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," IEEE Transactions on Information Theory, Vol. IT-13, 1967, pp. 260-269.

33. J. L. Massey, "Coding and Modulation in Digital Communications," 1974 International Zurich Seminar on Digital Communications, Zurich, Switzerland, pp. E2(1) - E2(4).

34. J. P. Odenwalder, "Optimal Decoding of Convolutional Codes," Ph.D. Dissertation, University of California, Los Angeles, 1970.

35. J. M. Wozencraft, "Sequential Decoding for Reliable Communication," IRE National Convention Record, Vol. 5, 1957, pp. 11-25.

36. G. D. Forney, Jr., Concatenated Codes, The M.I.T. Press, Cambridge, Mass., 1966.

37. L. J. Deutsch, and R. L. Miller, "Burst Statistics of Viterbi Decoding," TDA Progress Report 42-64, May-June 1981, pp. 187-193.