INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University
Microfilms
International

300 N. Zeeb Road
Ann Arbor, MI 48106

Pauw, Christoff Karl

DIFFERENTIAL 16-ARY PHASE-AMPLITUDE MODULATION SCHEMES FOR DATA TRANSMISSION ON FADING CHANNELS

City University of New York

PH.D. 1983

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Other _____

DIFFERENTIAL 16-ARY PHASE-AMPLITUDE MODULATION SCHEMES FOR DATA TRANSMISSION ON FADING CHANNELS.

by

CHRISTOFF K. PAUW

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

1983

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

December 15, 1993
date

Professor D. L. Schilling
Chairman of Examining Committee

Dec 15, 1993
date

Dean P. Karmel
Executive Officer

Professor R. L. Lunayach

Professor T. Saadawi

Professor N. Scheinberg

Professor H. Taub

Supervisory Committee

The City University of New York

Abstract

DIFFERENTIAL 16-ARY PHASE-AMPLITUDE MODULATION SCHEMES FOR DATA TRANSMISSION ON FADING CHANNELS.

by

Christoff K. Pauw

Adviser: Professor Donald L. Schilling

Three digital modulation formats for transmitting 4 bits per symbol on a fading channel are compared to one another and to binary DPSK. The results are applicable to Rayleigh fading channels in general, but parameters typical of the HF (ionosperic) radio channel were used when evaluating equations. The three schemes are normal 16-ary DPSK, differential phase-amplitude shift keying (DPASK) with 8 phases and 2 amplitudes, and DPASK with 8 phases, 2 amplitudes and an additional 22.5° phase shift between the two amplitude levels. Exact expressions for the probability of bit error were derived for the first two schemes, on both the AWGN and the slowly Rayleigh fading channels. The probability of bit error was also found by simulation for conditions of slow fading, fast fading, multipath time spread and additive Gaussian noise. Differential phase-amplitude shift keying was found to outperform 16-DPSK under most circumstances.

It is shown (by simulation only) that an optimum signaling

rate (baud rate) exists for a given combination of fading rate and multipath time spread, and that the HF channel should be treated as a fast-fading channel for the three schemes involved here.

A simplified analysis leading to a special case of a previously published equation for the probability of bit error for binary DPSK on the fast-fading channel is presented and is used to validate the simulation of the fast-fading channel.

Aan Irma, vir haar liefde, geduld en ondersteuning en omdat sy saam met my na hierdie vreemde land gekom het.

To Irma, for her love, patience and support, and for coming with me to this foreign country.

ACKNOWLEDGEMENTS

Many people and institutions helped me on my way to writing this dissertation. I would like to thank them all and specifically the following:

Professor Donald L. Schilling, for accepting me as his student and helping me in many ways.

Dean Paul R. Karmel, for solving all those little administrative problems in a warm, friendly manner.

The members of my committee and other members of the faculty at City College, from whom I have learnt a great deal.

Johann Ribbens and Hannes van Schalkwyk, my mentors and friends at the University of Pretoria in South Africa, who introduced me to communications and signal processing.

The South African Council for Scientific and Industrial Research, for a grant that paid my tuition and part of my living expenses during three of my five semesters at the City College of the City University of New York.

CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1 BACKGROUND

The study of digital communications on fading channels is a major subsection of the study of communications in general. It has drawn a great deal of attention in the literature and is discussed in a number of textbooks. In theory and in practice the emphasis has always been on binary and M-ary FSK, DPSK and DQPSK. A paper by Harper [10] did address the possibility of having simultaneous differential modulation of amplitude and phase. The purpose of the proposed research project is to produce a comparison for various 16-ary differential phase-amplitude modulation schemes, to see if any improvement on the results in [10] was possible.

One would be tempted to try to find the optimum signal constellation as was done for the non-fading additive Gaussian noise case [7]. The optimum can be found for a given signaling scheme, but one would be hard pressed to prove that another scheme with a better performance does not exist. That was not attempted.

The research grew out of a need to improve the spectral efficiency of a digital HF radio communication system. In the specific application, it was necessary to know the probability of bit error up to very high error rates. Therefore the results in [10], which applies mostly to low

error rates, had to be extended. Although the parameters of the HF channel was used for calculating specific numerical results, the theoretical results are applicable to other fading channels as well.

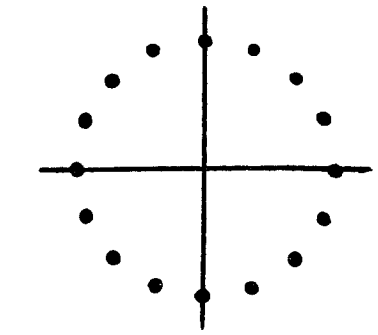
Three specific effects of the channel on the probability of error were investigated. They are (a) slow Rayleigh fading in the presence of additive white Gaussian noise, (b) fast fading in the absence of additive noise and (c) multipath time spread with slow Rayleigh fading and additive white Gaussian noise.

To limit the scope of the research, some specific assumptions were made. Only 16-ary modulation formats were considered. Bit-synchronization was assumed to be perfect at all times. The use of carrier tracking loops and equalizers was not considered.

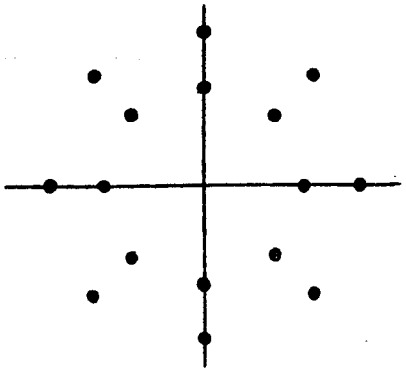
1.2 DIFFERENTIAL 16-ARY SIGNALING SCHEMES

In this section the reader is introduced to the three schemes for differential 16-ary data transmission that are investigated in this dissertation.

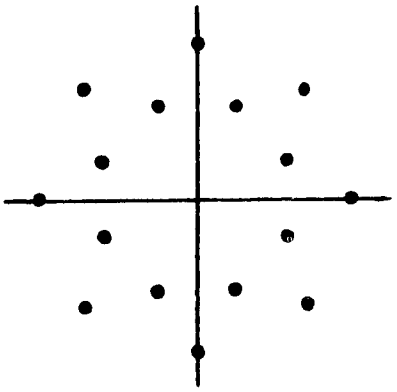
Referring to Fig. 1.1 we define Scheme 1 as 16-ary differential phase shift keying. If the transmitter wants to transmit the m-th symbol (m = 1,2,.....,16), it adds m times



Scheme 1
16-DPSK



Scheme 2
8/2-DPASK



Scheme 3
8/2-DPASK with 22.5°
phase shift.

Fig. 1.1: Three differential 16-ary signaling schemes.

22.5° to the previously transmitted phase. The receiver detects the phase shift between consecutive symbols and makes a decision based that.

Scheme 2 utilizes two amplitude levels and eight phases. We refer to it as 8/2 differential phase-amplitude shift keying (8/2-DPASK). The three phase bits are encoded normally as in Scheme 1. The fourth bit is differentially encoded on the amplitude. This is not a new idea [10] , but it is unusual and deserves some more explanation. The transmitter uses the following algorithm:

To transmit a 1, do not change the amplitude. To transmit a 0, change the amplitude (or vice versa). Since we are treating phase and amplitude as two independent variables, the receiver has to measure the phase shift between consecutive symbols in order to detect the three bits encoded on the phase. The problem of detecting the amplitude bit is a little different since amplitude does not have the "modulo 2" property of phase. The receiver really has to know which one of the two amplitude levels was transmitted, in order to know if the threshold should be set higher or lower than the previously received amplitude, which is the reference for detection of the presently received amplitude. But the receiver does not know which amplitude level was transmitted because of the fading channel. That is exactly

why the data has to be encoded differentially. The best way out of the problem is for the receiver to set thresholds, one above and one below the previously received amplitude. If the presently received amplitude is between those thresholds, a 1 is detected. Otherwise it is assumed that the amplitude changed and a 0 is detected. Fig. 1.2 illustrates the point.

Scheme 3 is a modification of Scheme 2 where an additional 22.5° areadded to the phase whenever a changed in amplitude occurs. This increases the distance between adjacent points on the signal constellation somewhat without affecting the transmitter power. The price for the gain is a more complicated receiver. To take advantage of the additional phase shift, one can no longer make separate decisions on phase and on amplitude. The receiver has to generate a reference signal constellation from the previously received symbol, containing eight signal points on the same amplitude level as that symbol, eight signal points shifted by 22.5° on a lower amplitude level, and likewise on a higher amplitude level. Then the distance between the presently received symbol and each of the 24 reference points has to be calculated. The decision is made based on which reference point is closest to the received symbol, or some other suitable criterion.

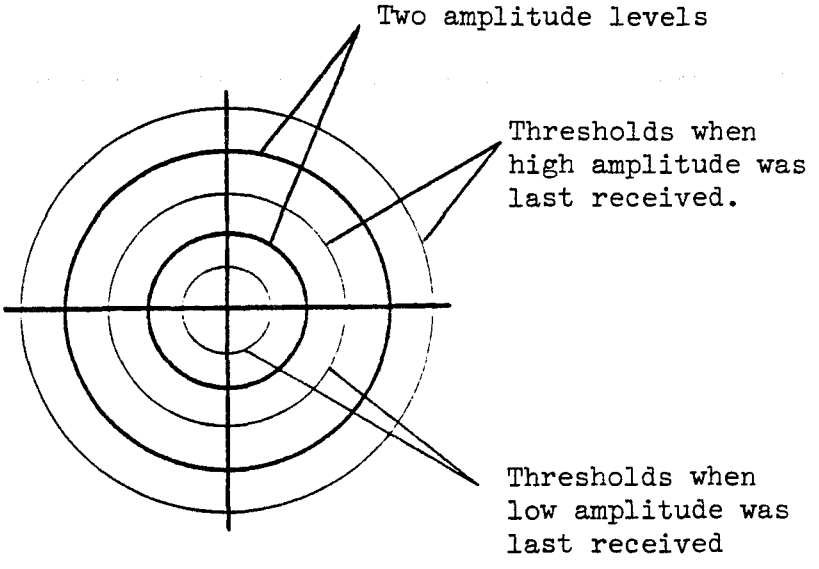


Fig. 1.2: Detection of the amplitude-modulated bit for "Scheme 2".

1.3 PREVIEW.

We now give a short description of the chapters to follow.

Chapter 2 is a description of the properties of the HF channel. It serves as an introduction for the reader who may not be familiar with the HF channel. Since it is only a summary of what is already available in the literature, it does not form part of the main thrust of the dissertation.

Chapter 3 addresses a problem in the simulation of the HF channel. The channel is modeled by a tapped delay line filter with tap-gains that are stationary random processes. The chapter is mostly about how to generate these random processes with a prescribed autocorrelation function. The simulation is a useful tool for checking some of the work in later chapters. In some cases, where the analyses became too cumbersome, only simulation results were presented.

Chapters 4, 5 and 6 are about the additive white Gaussian noise (AWGN) channel and the slowly Rayleigh fading channel. In chapter 4 we derive expressions for the probability of bit error for M-ary DPSK with Gray coding, using an equation from [16]. The results thus obtained are not new, but the expressions are needed in chapter 5, where differential amplitude modulation is imposed on the DPSK (M=8) to form 8/2-DPASK ("Scheme 2"). The results of chapter 4 are

extended here to include the amplitude modulation.

Chapter 6 adds an additional 22.5° phase shift between the amplitude levels of "Scheme 2" so that it becomes "Scheme 3". In this case we could not find an exact expression for the probability of error. An approximate expression, based on the union bound, was derived instead.

Chapter 7 is about the fast-fading channel. An expression for the unminimizable probability of error for binary DPSK is derived. This expression is a special case of results published in [1]. Our analysis is made easier to follow by some specific assumptions. Simulation results for the three 16-ary schemes (section 1.2) are also presented.

Chapter 8 investigates by simulation only the combined effect of additive noise, multipath time spread and slow as well as fast Rayleigh fading.

Chapter 9 is a short discussion of some additional schemes that were considered and found to be unsuitable for the HF channel.

Conclusions are drawn in chapter 10, followed by appendices containing material that was thought to be distracting in the development of the arguments in the main body of the text. That is followed by a list of references.

2. PROPERTIES OF THE HF CHANNEL

2.1 INTRODUCTION

High frequency radio refers to radio transmission in the frequency range 3 MHz to 30 MHz. In this frequency range, electromagnetic (EM) waves are reflected from the ionosphere, making possible communications over the horizon. At higher frequencies that is not possible because the EM waves propagate in straight lines and only a very small fraction of the transmitted energy reaches the ground beyond the horizon by diffraction. In contrast to that, in the HF range communications between points on opposite sides of the earth is made possible by multiple reflections between the surface of the earth and the ionosphere.

Because of the involvement of the ionospere, the term "HF radio channel" is often used interchangeably with the term "ionosperic radio channel". That is not quite correct because the ionosphere usually comes into play only with beyond-line-of-sight distances of 300 km or more. Line-of-sight transmission is also possible in the HF range, of course, but we will not consider that. We will use the term HF channel as if it was synonymous with the term "ionosperic channel".

The HF channel has many severe limitations. The most obvious

is the narrow bandwidth of only 27 MHz. With world-wide communications occupying many billions of MHz today, that is but a drop in the ocean. The HF channel was used for the first trans-atlantic communications systems, but was soon replaced by undersea cables, which were later supplemented by satellite systems.

Other problems include the fact that between any two points usually only a portion of the spectrum between 3 MHz and 30 MHz can sustain communications, depending on the time of day and a host of other factors. Frequency management becomes a major problem.

Furthermore, the HF channel is a prime example of a fading channel with multipath time spread. In some circumstances, it should also be characterized as a fast-fading rather than a slowly fading channel. We show in chapter 8 how the combined effect of fast fading and multipath time spread degrades the probability of bit error.

With all these limitations, the HF channel can sustain a signaling rate of only 100 baud, unless equalization is employed to mitigate the effect of multipath. Equalization, however, is also difficult because of the fast-fading nature of the channel. When higher data rates are needed, the solution is usually to add many subcarriers which are

modulated individually at 100 baud.

The question is now if anyone would want to use such a channel. The answer is that it is used only when nothing else is available. It is used professionally only by people in remote areas who do not have access to satellite or other channels for long-distance communications, and by the military as a back-up for other systems. It is also used heavily by amateur radio enthusiasts.

Many students who "grew up" with systems operating in the Giga-Hz frequency range are perplexed by the term "high frequency" as it is used here. The explanation is simply that it is a relic of the early days of radio when frequencies above 1 MHz were indeed the highest that could be handled by the equipment of the time.

Finally, we would like to comment that an introduction such as this does no justice to the vast amount of information on the HF channel that was published in the last 40 years or so. New information is still being added almost monthly in journals in the fields of geophysics and electrical engineering. The information in this chapter was gathered from an excellent textbook [5] and a number of other references [8,12,13,17,18,23,25,35].

2.2 THE IONOSPHERE

The ionosphere is the upper part of the atmosphere where a significant fraction of the rarified gases are ionized, mostly by radiation from the sun.

For the purposes of communication systems, there are three regions that can be identified as having significant influence on the propagation of radio waves. They are called "layers" and do indeed appear as "layers" in their effect on HF radio signals.

The D region is at a height of approximately 75 to 80 km, and is mostly responsible for the attenuation (over and above that of the range equation) that radio waves suffer. The mechanism for the loss of energy is that the atmosphere is still dense enough so that electrons accelerated by the EM waves have a significant chance of colliding with other particles, thus expending or at least scattering some of the energy from the wave, rather than aiding in the reflection process. The absorbtion increases with decreasing frequency because the distance that the electrons travel under the influence of the wave, is inversely proportional to the square of the frequency of the wave. The longer the distance traveled, the greater the probability of a collision and the more energy is extracted from the EM wave.

The process of reflection is actually better described as refraction, since the change in the direction of propagation is gradual rather than suddenly as in the case of reflection from a conducting plane.

The E and F regions occur between 90 to 100 km and 140 to 400 km or higher, respectively. Here the probability of collision between ions accelerated by the EM waves is very low, and the waves are refracted without much loss of energy. If the frequency is low enough, the refraction will be enough to return the waves to the earth.

An important parameter of the ionospheric channel is the so-called maximum usable frequency (MUF). That is the highest frequency at which reflection of a wave will occur at a given angle of incidence. It is proportional to the square root of the maximum electron density of the given layer of the ionosphere and inversely proportional to the cosine of the angle of incidence. (The angle of incidence is the smallest angle between the initial direction of the EM wave and the normal to the reflecting layer at the "point" of reflection.)

Multipath time spread is introduced when favorable conditions for the refraction of the EM waves exist at different heights in the ionospere, so that the wave

literally travel along different paths of different lengths from the transmitter to the receiver. As an example, there could be one component travelling via the E-layer, and another component, delayed by a ms of two, travelling via the F layer. Also, in many cases it can be shown that there are two angles between the incident wave and any layer of the ionosphere that will support communications between two given points at a given frequency. The two resulting paths are called the low ray and the high ray, and can be thought of in rather inaccurate terms as one reflection from the bottom of a layer, and another from the top of the layer. The separation between the paths increases with decreasing frequency. The two paths coincide at the MUF for a particular layer.

Each of the possible multipath components described above, can sometimes be divided into two more components, called the two magnetoionic components, or the ordinary wave and the extraordinary wave. That happens because we are dealing with the propagation of an EM wave in an ionized medium, in the presence of a stationary magnetic field, that of the earth. The motion of charged particles is affected by the stationary magnetic field, and therefore the propagation of an EM wave depends on its direction of travel and on its polarization with respect to the magnetic field of the earth. The ordinary wave is polarized parallel to the earth's B field, the extraordinary wave is polarized

perpendicular to the B field.

Since the ionosphere is a gaseous medium much like the lower atmosphere, it is not uniform. Furthermore, winds of up to 1000 km/h are present. That is the main cause of the fading that is characteristic of the HF channel. The properties of the ion gas at the "point" of reflection vary with time due to the winds and the irregularities of ion concentrations. The mechanism of fading is constructive or destructive interference between the many multipath components of the received signal. The movement of the non-uniform medium causes random changes in the phase of each component, which is observed at the receiver as a random amplitude and phase because of the interference between the components.

Typical parameters of the HF channel are time spreads of the order of 0.1 to say 4 ms, and fading rates of 0.05 to 2 Hz. Higher values of time spread and frequency spread are sometimes experienced, especially in transauroral transmissions.

2.3 DIURNAL AND OTHER VARIATIONS

Since the ionosphere exists mostly because of radiation from the sun, it should come as no surprise that the ionosphere changes drastically with the time of day.

The D layer is virtually absent during the night. It's electron density reaches a peak at noon. The peak density is larger during the summer than during the winter. Ionization in the D layer is caused mostly by X-ray radiation from the sun. This type of radiation increases during solar flares, and that is part of the reason for HF blackouts which occur after some solar flares.

Ionization in the E layer also reaches a peak at noon, but it decays only gradually during the night to a low just before sunrise. The MUF reaches a maximum of about 20 MHz for communications channels of about 2000 km in length during the day. During the night it can drop as low as 2 or 3 MHz.

The F region is considered to consist of two layers during the day, the F1 and the F2 layers. During the night they merge to form a single layer, the F layer. Apart from that, its behavior is complex and not as easily linked to the sun as the D and E layers. During periods of very high ionization, frequencies of as high as 50 MHz can be reflected from the F2 layer. The F2 layer is usually involved in transmission distances of more than 3000 km. It is much more difficult to establish reliable communications over such long distances, because of the complexity of the behavior of the F region.

Ionospheric storms coupled with magnetic disturbances tend to occur between 20 to 40 hours after a solar flare and may last for several days. They are caused by interaction between the magnetic field of the earth, electrical currents of up to one million amperes in the ionosphere, and particle radiation from the sun. During these storms the variation as a function of position of electron densities in the ionosphere is greatly increased. The result is that the layers of the ionosphere no longer act like rather smooth reflectors, but tend to scatter EM waves much as a rough reflector would do. This is another mechanism whereby HF communications can be completely disrupted.

The electron densities in all three layers of the ionosphere also vary with a 27-day and an 11-year cycle. The first is related to the rotation of the sun, and the second to an eleven-year cycle in the number of sunspots that can be observed from the earth. The relative positions of the earth, the sun and the moon also have some effect because of interaction of their magnetic fields.

This was a very rough and intuitive explanation of the nature of the ionosphere. We believe that it will give the reader enough background to be comfortable with the "black box" model of the HF channel in chapter 3.

3. SIMULATING THE HF CHANNEL.

3.1 INTRODUCTION

Any communications channel can be thought of as a "black box" with an input and an output. When the effect of the channel on signals passed through it is specified, the communications engineer can go about his business without further concern about the physical processes involved, however complex they may be.

In communications such black boxes are usually called filters when they affect the spectral contents of the signal. The HF channel, and in general most fading channels, do affect the spectral contents of the signal, so they can be thought of as filters. The only unusual aspect is that these "filters" are not time-invariant. Keeping that in mind, the problem of modeling the channel becomes an ordinary filtering problem.

All that needs to be done is to specify the parameters of the filter. That, however is easier said than done because of the fact that it is time-varying, because it can not be put in an electrically quiet laboratory for proper measurements, and because of the multitude of factors influencing it's properties, as was discussed in chapter 2. An important study of this problem was published in [29]. A

specification for HF channel simulators resulted from that study and is available in [30]. It is widely used today. We now proceed to discuss the specifics of the filter-model and it's parameters.

Fig. 3.1 shows a diagram of a transversal filter. It is also the model of the simulator of [30]. If the tap gains $g_n(t)$ are kept constant, any transfer function can be realized with an appropriate choice of the tap gains and the number of taps (with the usual restrictions applying to sampled-data systems, of course). Any time-varying impulse response can also be realized, provided that a system for appropriate variation of the tap gains can be found. In this sense the channel model is completely general. Any channel, time-varying or not, can be modeled by appropriate specification of the parameters of this transversal filter.

In the simulation of the HF channel, the taps from the delay line have a physical interpretation, apart from just being part of a transversal filter. We have seen in chapter 2 that the ionosphere contains several layers that can individually reflect radio transmissions. The outputs from the tap-gain multipliers in the transversal filter can be thought to be representing these reflections. This statement tacitly assumes that an impulse reflected from each layer will be received as a delayed impulse, which is not quite true. Some

"smearing" of the impulse will occur, but in most practical cases that effect is not resolvable because the signals transmitted through the ionosperic channel are bandlimited. Therefore we think, in the rest of this chapter, of the output from each tap-gain multiplier as a reflection from one of the layers of the ionosphere. That includes multiple reflections between layers or between the ionosphere and the earth. A word of caution is in order, though. In wide-band systems such as radar systems or spread-spectrum communication systems this "smearing" may be resolvable and should be taken into consideration.

Before proceeding to the specification of the parameters, a note on the use of complex numbers in Fig. 3.1 is in order here. The signal transmitted through the channel is modulated on a carrier, so that the physical process that is being modeled is a bandpass process. When doing a digital simulation, however, one needs to keep the sampling rate as low as possible. That is why we have opted for a low-pass instead of a bandpass simulation. All information pertaining to the bandpass process, except the actual carrier frequency, is retained by representing all variables by complex numbers. A further discussion of how the complex numbers relate to physical systems, can be found in section 4.2.

The complete specifications of the parameters of the model

are given in [30]. A discussion on how these were implemented in the COMSIM simulation that was used for all simulations reported on in this dissertation, can be found in [4]. We repeat here only those specifications that are necessary to follow the rest of the material presented in this chapter, which is mostly concerned with the time-varying nature of the tap gains.

Now for the specifications of the parameters. At least three taps should be used, in order to be able to simulate at least three different propagation paths. The differential delay between the first and the last tap should be adjustable up to at least 10 ms, and the differential delay between any pair of paths should not differ from the nominal value by more than 1 percent.

The tap-gain functions $g_n(t)$ should have complex Gaussian probability density functions with equal-powered, uncorrelated, real and imaginary parts. The expected (or mean) values of the $g_n(t)$ will usually be zero to produce a Rayleigh distribution for the amplitude of the tap-gain[15]. The exception is when the first tap is used to model a ground-wave propagation path, which will have a constant gain. In that case the amplitude of a sine wave passed through the filter (or the channel) will have a Ricean amplitude distribution.[33]

The spectrum of the real and imaginary part of each tap-gain function should also have a Gaussian shape. The half-power bandwidth of the tap-gains should be adjustable from about 0.005 to 250 Hz. Note that this has nothing to do with the bandwidth of the channel or of the signal being passed through it. It merely specifies the rate of change of the tap-gains, i.e. the rate of change of one of the components of the time-varying impulse response of the channel. The presence of two totally different Gaussian functions is also confusing. We would like to stress again that the first one refers to a probability density function, and the second one to a power spectral density, and the two have nothing to do with each other.

Finally, each of the tap-gain functions should also have a frequency offset in order to be able to simulate the Doppler offsets which occur, especially at sunrise when the hight of the layers of the ionospere may change continuously for up to half an hour. It can also be used to model relative motion of the transmitter and receiver. The offset will, however, in most cases be set to zero.

It should be mentioned here that [30] actually requires the tap-gains to be the sum of <u>two</u> functions, each with a Gaussian spectrum, but with different bandwidths and Doppler offsets. This is to be able to separately model the two

magnetoionic components of the signal. It was decided not to include this feature since it happens only rarely that these two components are both resolvable and of comparable magnitude.

The rest of this chapter is devoted to a discussion of a method of generating the proper spectral shape for the tap-gain functions. A block diagram of the system used for generating the tap gains is shown in Fig. 3.2. Our specific concern is the proper choice of the impulse response h(t).

The problem of obtaining the required spectral shape (or, equivalently, autocorrelation function) has been discussed in the literature [22,31,61].

3.2 GENERATING TAP GAINS WITH GAUSSIAN AUTOCORRELATION

Generating a random variable with a given autocorrelation function can be done by passing white Gaussian noise through a filter with an appropriate transfer function. Noise with Gaussian autocorrelation also has a Gaussian power spectral density. Therefore the transfer function of the filter required to produce a Gaussian autocorrelation is not realizable. It has an infinite number of poles as can be seen by using the Taylor series expansion of the Gaussian function as a function of frequency. Since the required

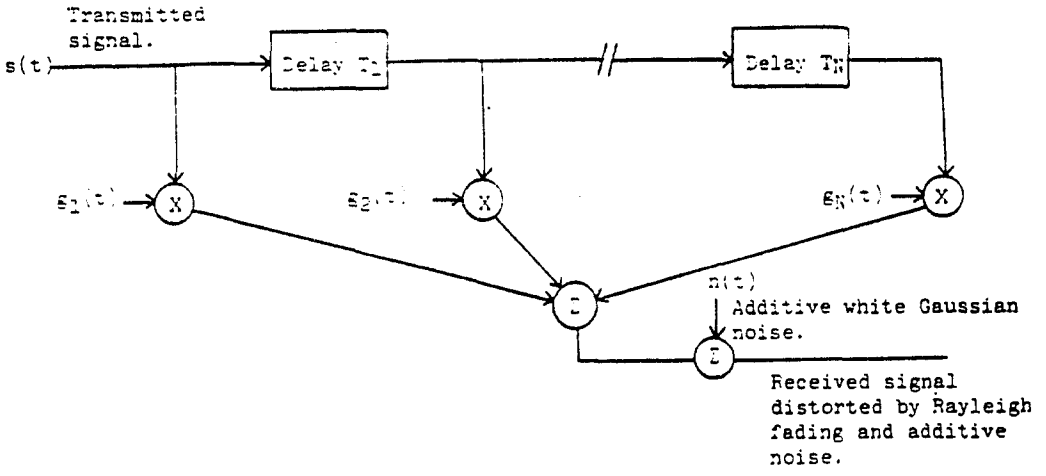


Fig. 3.1: Low-pass simulation of HF channel. All variables are complex numbers.

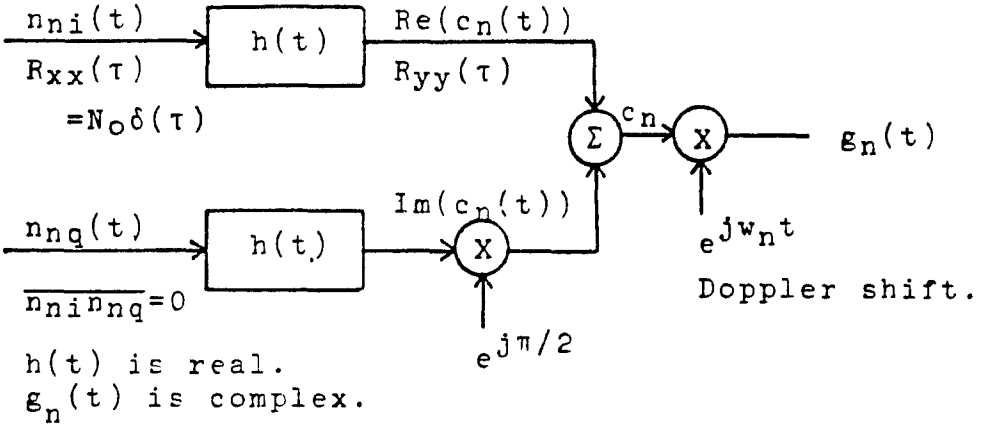


Fig. 3.2 Generating the tap gains.

filter is unrealizable, one has to make do with some approximation.

The question is now what criterion to use to find an appropriate approximation. Our application is the simulation of digital communication systems where the signal bandwidth is much greater than the channel fading rate. The signal degradation due to fading for digital communication systems is dependent on the autocorrelation of the tap gains in the interval $0 < \tau < T$ or $0 < \tau < 2T$, depending on the system, where $1/T$ is the baud rate [1]. It is therefore natural to perform the analysis in the time domain rather than in the frequency domain and the primary concern is the behavior at small values of time. When any form of tracking loop is used, the behavior at other values of time should be considered, but only as a matter of secondary importance.

By the Central Limit Theorem the repeated convolution of a large number of variables will under certain conditions approach the Gaussian function. Thus any all-pole low-pass filter of sufficiently high order will produce the desired autocorrelation function. In this paper we show that all low-pass filters of order $N > 1$ produce the desired autocorrelation in the limit as τ approaches 0. In cases other than this limit, some filters have more desirable characteristics than others.

3.3 CALCULATING AUTOCORRELATION FUNCTIONS

If the input to a filter has an autocorrelation $R_{xx}(\tau)$ then the output of the filter will have an autocorrelation

$$R_{yy}(\tau) = R_{xx}(\tau) * h(\tau) * h^*(-\tau) \tag{3.3.1}$$

where h(t) is the impulse response of the filter, * denotes convolution, and h^* denotes the conjugate of h [15].

If the input is white noise, then

$$R_{xx}(\tau) = N_o \delta(\tau) \tag{3.3.2}$$

and

$$R_{yy}(\tau) = N_o h(\tau) * h(-\tau) \tag{3.3.3}$$

The output power spectral density is

$$S_{yy}(\omega) = N_o H(j\omega) H^*(j\omega) = N_o |H(j\omega)|^2 \tag{3.3.4}$$

An N-th order lowpass filter has a transfer function

$$H_N(s) = G / \prod_{n=1}^{N} (s+a_n) \tag{3.3.5}$$

Then

$$S_{yy}(\omega) = N_o G^2 / \prod_{n=1}^{N} (a_n + j\omega)(a_n^* - j\omega) \tag{3.3.6}$$

The impulse response for the case of N unequal poles is found after a partial fraction expansion of H(s):

$$h_N(t) = (-1)^{N+1} \sum_{n=1}^{N} \left[u(t) \exp(-a_n t) / \prod_{\substack{j=1 \\ j \neq n}}^{N} (a_n - a_j) \right] \tag{3.3.7}$$

$R_{yy}(\tau)$ is best calculated by evaluating the convolution in Eq. 3.3.3. For convenience let

$$A_n = 1/\prod_{\substack{j=1 \\ j\neq n}}^{N} (a_n - a_j) \tag{3.3.8}$$

Then

$$R_{yy}(\tau) = N_o G^2 \sum_{n=1}^{N} \left[\sum_{k=1}^{N} A_n A_k \exp(-a_n|\tau|)/(a_n+a_k) \right] \tag{3.3.9}$$

Now that we have an expression for the power spectral density and the autocorrelation of the signal at the output of the filter, we can evaluate it to see how closely it resembles the desired Gaussian autocorrelation.

The desired autocorrelation is given by

$$R_g(\tau) = S \exp(-B^2\pi^2\tau^2/\ln 2) \tag{3.3.10}$$

where B is the half-power bandwidth in Hz [2]. We will refer to B as the fading rate of the channel.

The Gaussian function is characterized at small values of τ by the fact that it has a smooth maximum at $\tau = 0$:

$$R_g(0) = S \tag{3.3.11}$$

$$R'(0) = 0 \tag{3.3.12}$$

$$R''_g(0) = -2SB^2\pi^2/\ln 2 \tag{3.3.13}$$

$R_{yy}(\tau)$ has similar properties:

$$R_{yy}(0) = N_o G^2 \sum_{n=1}^{N} \left[\sum_{k=1}^{N} A_n A_k/(a_n+a_k) \right] \tag{3.3.14}$$

$$R'_{yy}(0) = 0 \tag{3.3.15}$$

$$R''_{yy}(0) = N_o G^2 \sum_{n=1}^{N} \left[\sum_{k=1}^{N} a_n^2 A_n A_k/(a_n+a_k) \right]; \quad N = 1 \tag{3.3.16}$$

Eq. 3.3.15 is easy to prove in the frequency domain and Eq. 3.3.16 follows by inspection from Eq. 3.3.9 by calculating the limit $R''_{yy}(\tau)$ as τ approaches zero. This can be done because $R''_{yy}(\tau)$ is not discontinuous at $\tau=0$, since the limit as τ approaches zero from the left and from the right is the same, and because $R''_{yy}(\tau)$ does not contain an impulse at the origin (except for N=1). This last statement is true because it can be seen in the frequency domain that $R''_{yy}(0)$ has a finite value (except for N=1).

It can be seen from Eq. 3.3.11 through 3.3.16 that any filter of order 2 or higher will be a good approximation of the Gaussian for small values of time if we let $R_{yy}(0) = R_g(0)$ and $R''_{yy}(0) = R''_g(0)$. This is done by manipulating the gain factor G and the half-power bandwidth of the filter. Note that the half-power bandwidth of $S_g(\omega)$ and $S_{yy}(\omega)$ will not necessarily be the same.

Comparison for other values of τ is best done graphically. Behavior in the limit as $\tau\to\infty$ is not very interesting and a little thought will show that the only way to improve the match between $R_{yy}(\tau)$ and $R_g(\tau)$ is to increase the order N.

$R_{yy}(\tau)$ was also calculated for a simple third-order filter

with transfer function

$$H(s) = G/(s+a)^3 \tag{3.3.17}$$

Then

$$R_{yy}(\tau) = (3N_oG^2/16a^5)\exp(-a|\tau|)(a^2\tau^2/3 + a|\tau| + 1) \tag{3.3.18}$$

$R_{yy}(\tau)$ was evaluated for different types of filters in order compare them to each other and to the ideal Gaussian function. Filter design tables from [32] were used. The results are now discussed.

Fig. 3.3 shows $R_{yy}(\tau)$ for different types of filters all with half-power bandwidth of 1 rad/s. Fig. 3.4 shows the same for sixth-order filters. It can be seen that the Transitional Gaussian filter design gives the best result in both cases, while a Bessel filter is not much worse.

Fig. 3.5 shows the behavior of $R_{yy}(\tau)$ for small τ on a logarithmic scale. It can be seen that all the filters have the same slope. If the filters are designed as described previously the curves will all fall on top of one another. This serves to stress the point that the filter design should not be done to match the desired half-power bandwidth, if one is interested in the asymptotic behavior of the autocorrelation function as τ goes to zero.

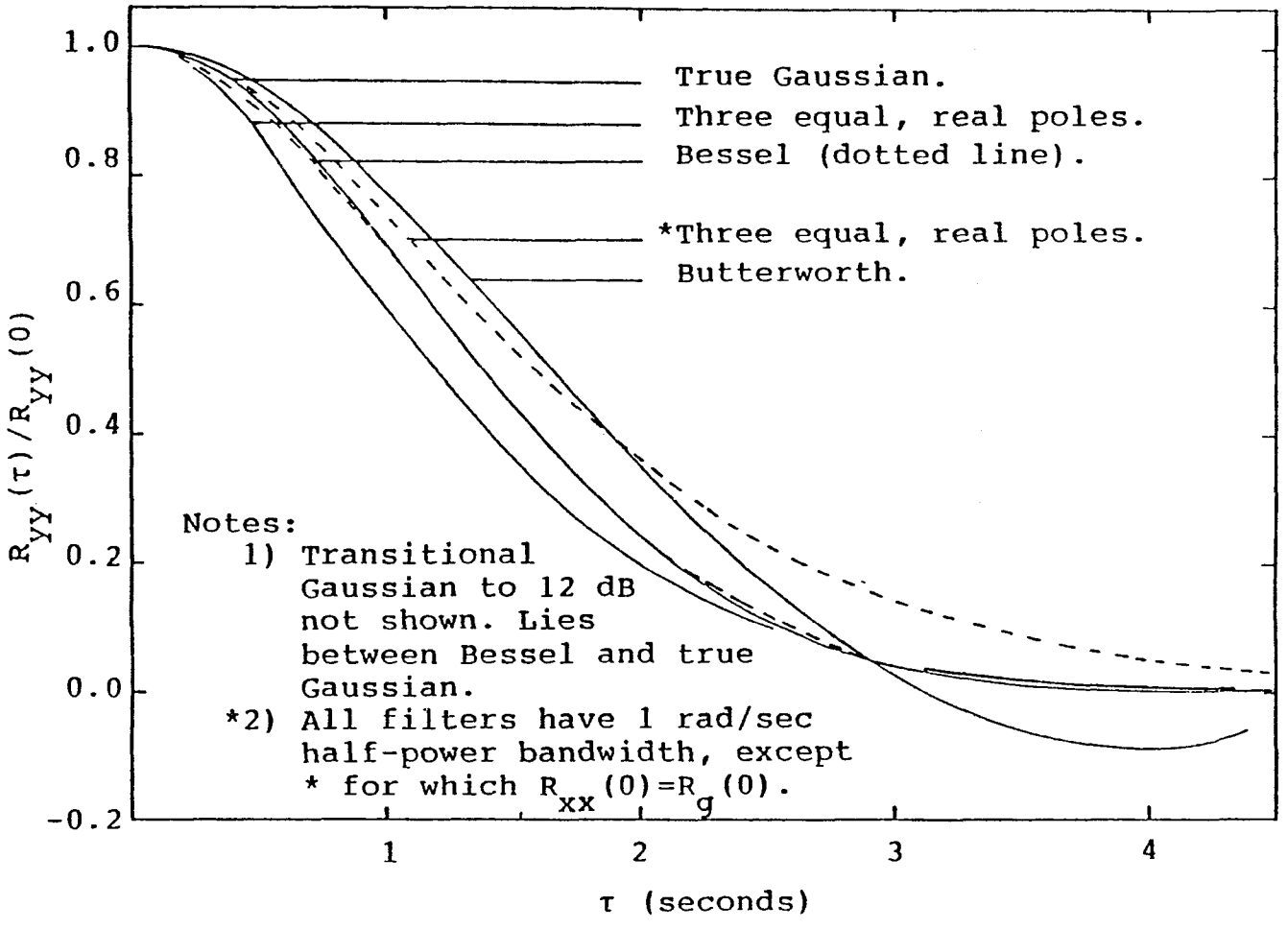


Fig. 3.3: Autocorrelation at output of third-order filters compared to true Gaussian.

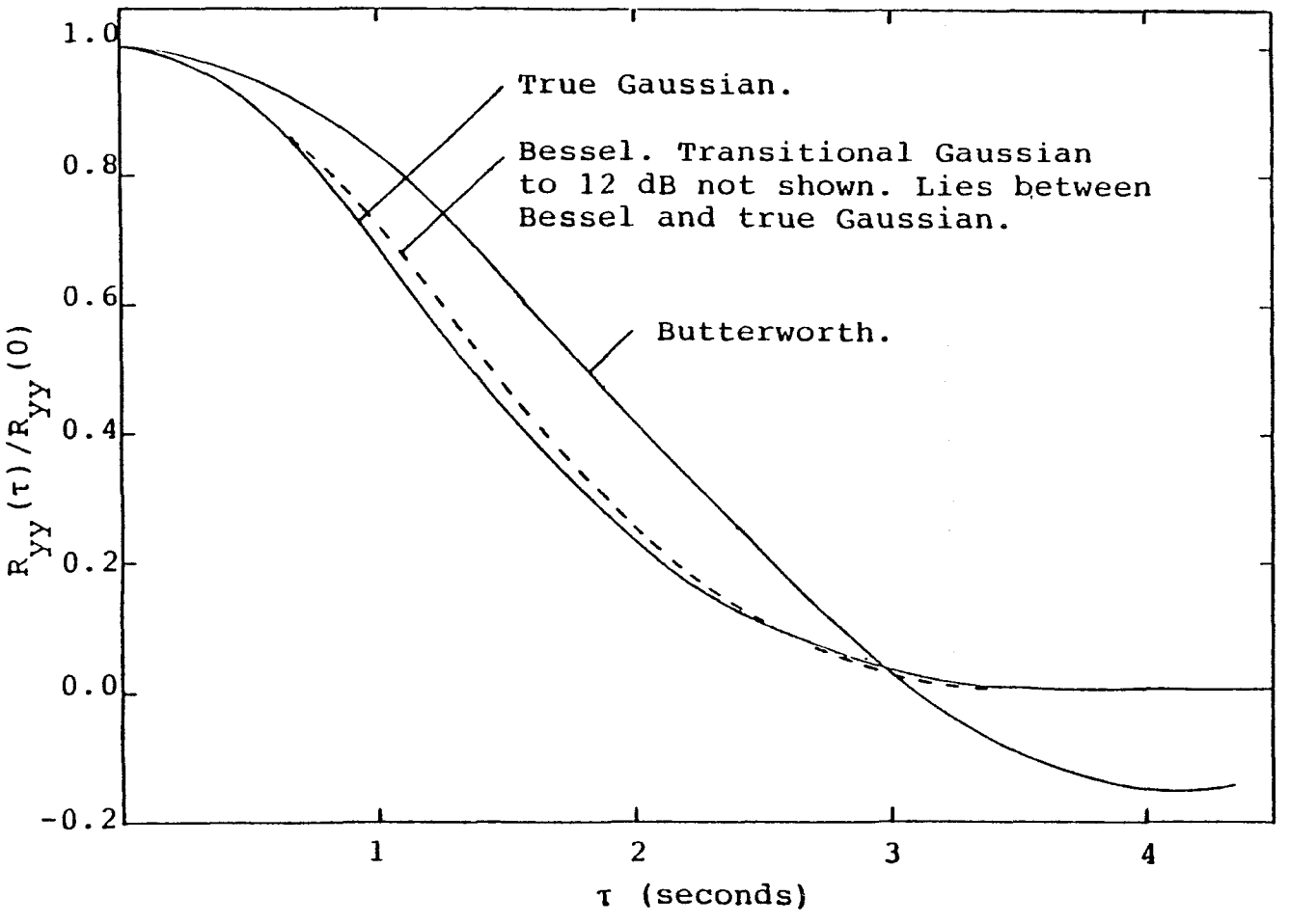


Fig. 3.4 : Autocorrelation at output of sixth-order filters compared to true Gaussian.

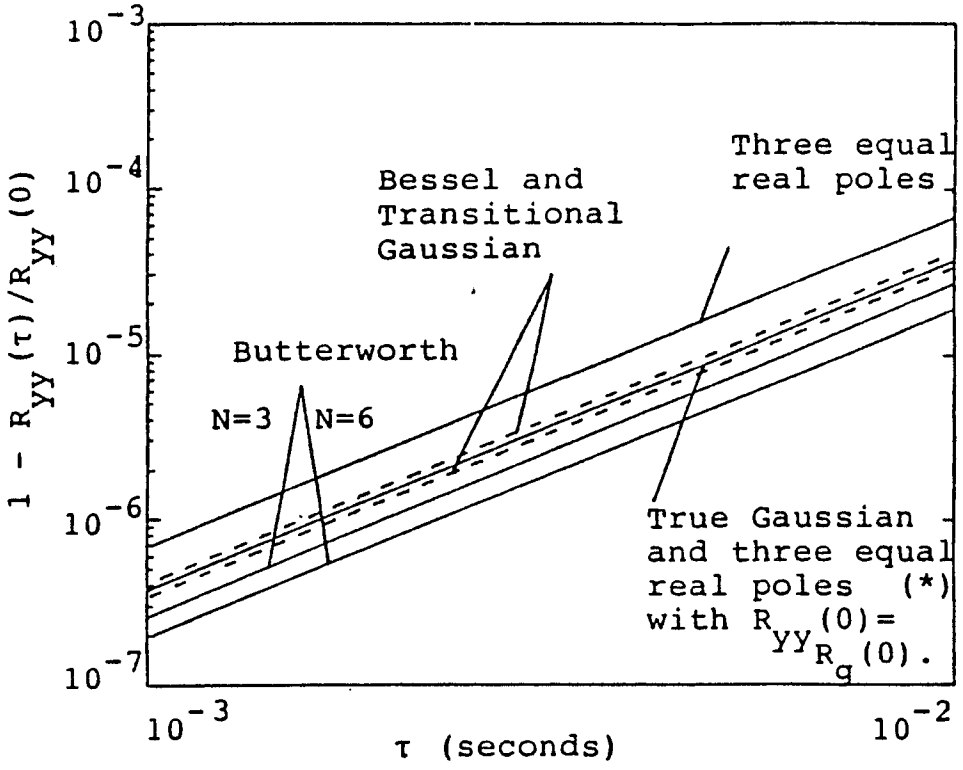


Fig. 3.5: Asymptotic behavior of autocorrelation functions at small values of time. All filters except (*) have half-power bandwidth of 1 rad/sec.

3.4 ON THE CHOICE OF FILTER DESIGN

In a digital simulation one has to realize a discrete-time rather than a continuous-time filter. We accomplished this by using the bilinear transformation [14].

In our simulation we finally decided to use the simple third-order filter with three equal real poles. This was done because it enabled us to use three first-order sections instead of a second-order followed by a first-order section. The second-order section caused problems due to the finite precision of floating-point arithmetic when the simulation called for very low fading rates. Realization of a digital lowpass filter with very narrow bandwidth involves finding small differences between large numbers, and the difference is raised to approximately the power of the order of the filter section, therefore our preference for first-order sections. Another possible remedy is to use double precision arithmetic, but we felt that was not worth the extra cost.

These practical considerations are not the last word on the choice of filter design, however. This chapter has so far addressed the problem of generating the tap-gains only in the time domain, but the specification in [30] is in the frequency domain. It reads: ".... the tap-gain spectrum shall not differ from the ideal Gaussian shape by more than 1 percent of the maximum value".

Compared to this standard, the simple 3-rd order filter is not a good choice, as can be seen in Fig. 3.6. The maximum deviation is about 20%, instead of 1%. A third-order transitional Gaussian filter has a maximum deviation of about 2%, and a fourth-order transitional Gaussian filter makes the grade with some margin to spare.

This discussion, together with section 3.3, indicates that a transitional Gaussian filter of order N=4 should be used. Our interest, however, was in the range of very low fading rates where the finite precision of the computer is a limitation. For that reason our choice remains the simple 3rd-order filter with three equal real poles.

3.5 DESIGNING THE DIGITAL FILTER

This section is about the design of a digital filter that is the equivalent of the simple third-order analogue filter with the transfer function of Eq. 3.3.17.

The first requirement is to find the position of the three poles, i.e. the parameter a. From Eq. 3.3.18:

$$R_{yy}(0) = 3N_o G^2/16a^5 \qquad (3.5.1)$$

We require:

$$R_{yy}(0) = R_g(0) \qquad (3.5.2)$$

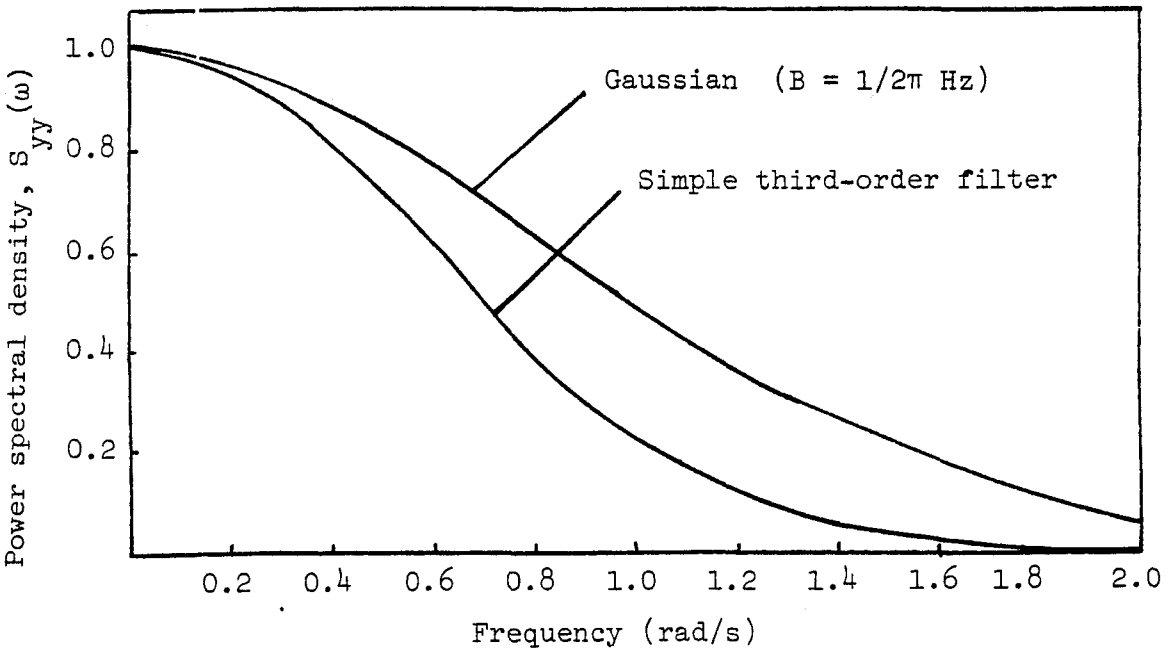


Fig. 3.6: Normalized power spectral densities: Ideal Gaussian shape and simple third-order filter. For design see section 3.4.

Therefore, from Eq. 3.3.11:

$$3N_oG^2/16a^5 = S \qquad (3.5.3)$$

It was shown in the discussion following Eq. 3.3.16 that $R''_{yy}(0)$ can be found by taking the limit of $R''_{yy}(\tau)$ as τ approaches zero. Doing that, it is found from Eq. 3.3.18:

$$R''_{yy}(0) = (3N_oG^2/16a^5)(-a^2/3) \qquad (3.5.4)$$

We require:

$$R''_{yy}(0) = R''_g(0) \qquad (3.5.5)$$

Substituting Eq. 3.5.4, 3.5.3 and 3.3.13 into Eq. 3.5.5, we can solve for a:

$$a = 2\pi B\sqrt{6/4\ln 2}$$
$$= 1.471 \times 2\pi B \qquad (3.5.6)$$

It should be pointed out that with this value of a, the half-power bandwidth occurs at 0.7 B Hz, not B Hz, but it is of no concern to us since we pointed out earlier we care about the autocorrelation function of the noise more than about the spectrum.

The transfer function of the corresponding digital filter can be found by using the bilinear transformation ([14] p. 207). We also make use of pre-warping to minimize distortion of the transfer function. For pre-warping let

$$b = (2/T)\tan^{-1}(aT/2) \qquad (3.5.7)$$

where T is the sampling period of the digital filter. Let

$$H'(s) = K/(s+b)^3 \qquad (3.5.8)$$

Now using the bilinear transformation

$$H(z) = H'(s)\Big|_{s = (2/T)[(1-z^{-1})/(1+z^{-1})]} \qquad (3.5.9)$$

To simplify the expression that will result from this substitution, let

$$G = (kT^3)/(2+bT^3) \qquad (3.5.10)$$

and

$$D = (2-bT)/(2+bT) \qquad (3.5.11)$$

Then

$$H(z) = G[(1+z^{-1})/(1-Dz^{-1})]^3 \qquad (3.5.12)$$

The next step is to find a suitable structure for implementing the filter. The most effective way would be to expand the third power in H(z) and then to use one of the standard canonical forms, but we have already pointed out that this will cause a problem. We want to be able to simulate channels with very low fading rates. That makes D a number very close to one. It could realistically differ from 1 in the fifth decimal position. Expanding H(z) results in a feedback term proportional to D^3, which now differs from 1 in the fifteenth decimal position. The computer rounds that

off to 1 and the filter becomes unstable. To avoid the problem, the filter is implemented by simply cascading three first-order filters as shown in Fig. 3.7.

The gain of the filter G has yet to be determined. In the computer simulation it is convenient to have the power in the output sequence equal to the power in the input sequence. Using that relation the desired gain can be found by writing the discrete-time autocorrelation function of the output sequence in terms of that of the input sequence:

$$R_{yy}(m) = R_{xx}(m) * h(m) * h^*(-m)$$
$$= \sum_{\ell=-\infty}^{\infty} R_{xx}(m) \sum_{k=-\infty}^{\infty} h(k)h(\ell+k) \qquad (3.5.13)$$

But the input sequence is white noise with

$$R_{xx}(m) = \sigma_x^2 \delta(m) \qquad (3.5.14)$$

Therefore

$$R_{yy}(m) = \sigma_x^2 \sum_{k=-\infty}^{\infty} h(k)h(m+k) \qquad (3.3.15)$$

and the output power is given by

$$\sigma_y^2 = R_{yy}(0) = \sigma_x^2 \sum_{k=0}^{\infty} h^2(k) \qquad (3.5.16)$$

Now the requirement for G is that

$$\sum_{k=0}^{\infty} h^2(k) = 1 \qquad (3.5.17)$$

The closed-form expression for the infinite summation is a formidable equation and extremely tedious to calculate. We

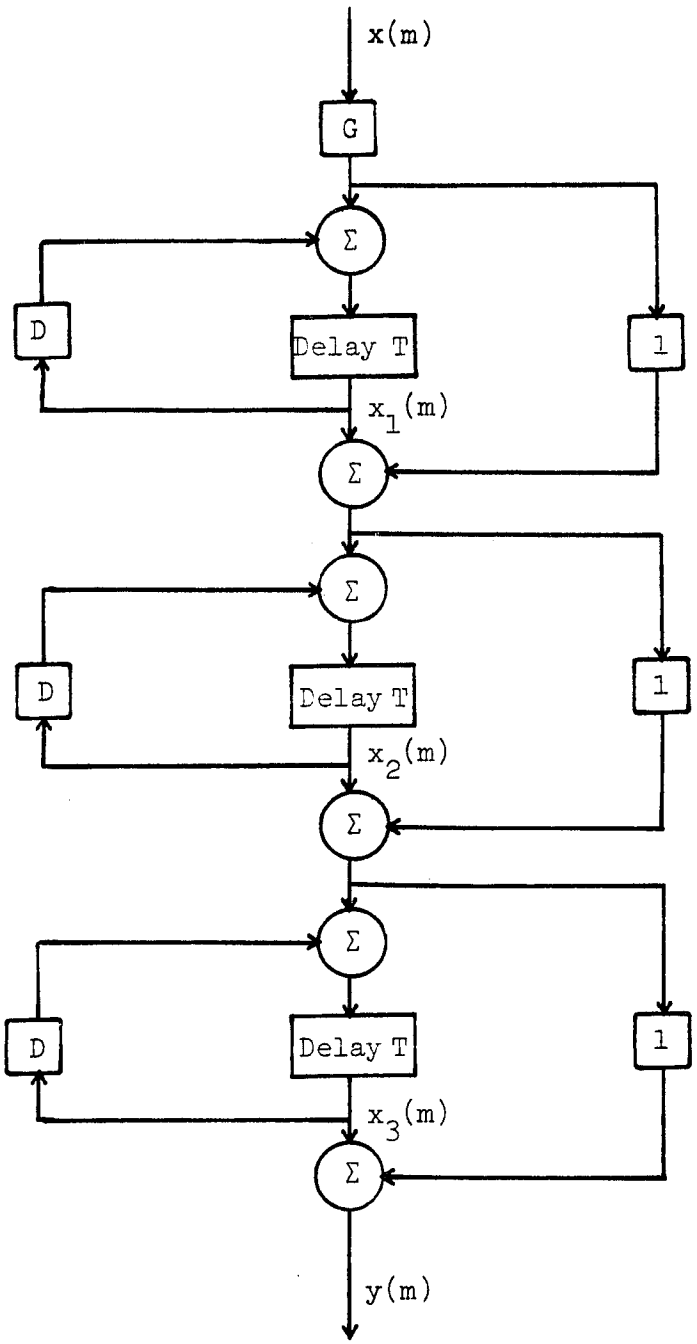


Fig. 3.7: Structure of the digital filter for generating noise with Gaussian autocorrelation function.

therefore use the following procedure in the computer simulation:

1) Specify B.
2) Calculate D.
3) Let G=1, drive the filter with an impulse so that the output will be h(k).
4) Accumulate the $h^2(k)$ for say ten times the time-constant of the filter to approximate the infinite summation.
5) Now

$$G = \sqrt{1/\sum_{k=0}^{\infty} h^2(k)} \qquad (3.5.18)$$

Figure 3.8 shows simulation results obtained using this filter design in the simulation of the fading channel, for the unminimizable probability of error due to fading for DPSK compared with the theoretical probability of error [1]. The theoretical error rate is derived in chapter 7. (The unminimizable error rate is the error rate at infinitely large signal to noise ratio.) The curves were plotted as a function of the ratio of the bit rate to the fading rate. It can clearly be seen that the limiting behavior for small fading rates, corresponding to small values of τ in $R_{yy}(\tau)$, is correct, while there is some deviation at larger fading rates. To illustrate the point, let us assume that the bit rate is 100 bits/s. If the fading rate is also 100 Hz, then the simulated probability of error

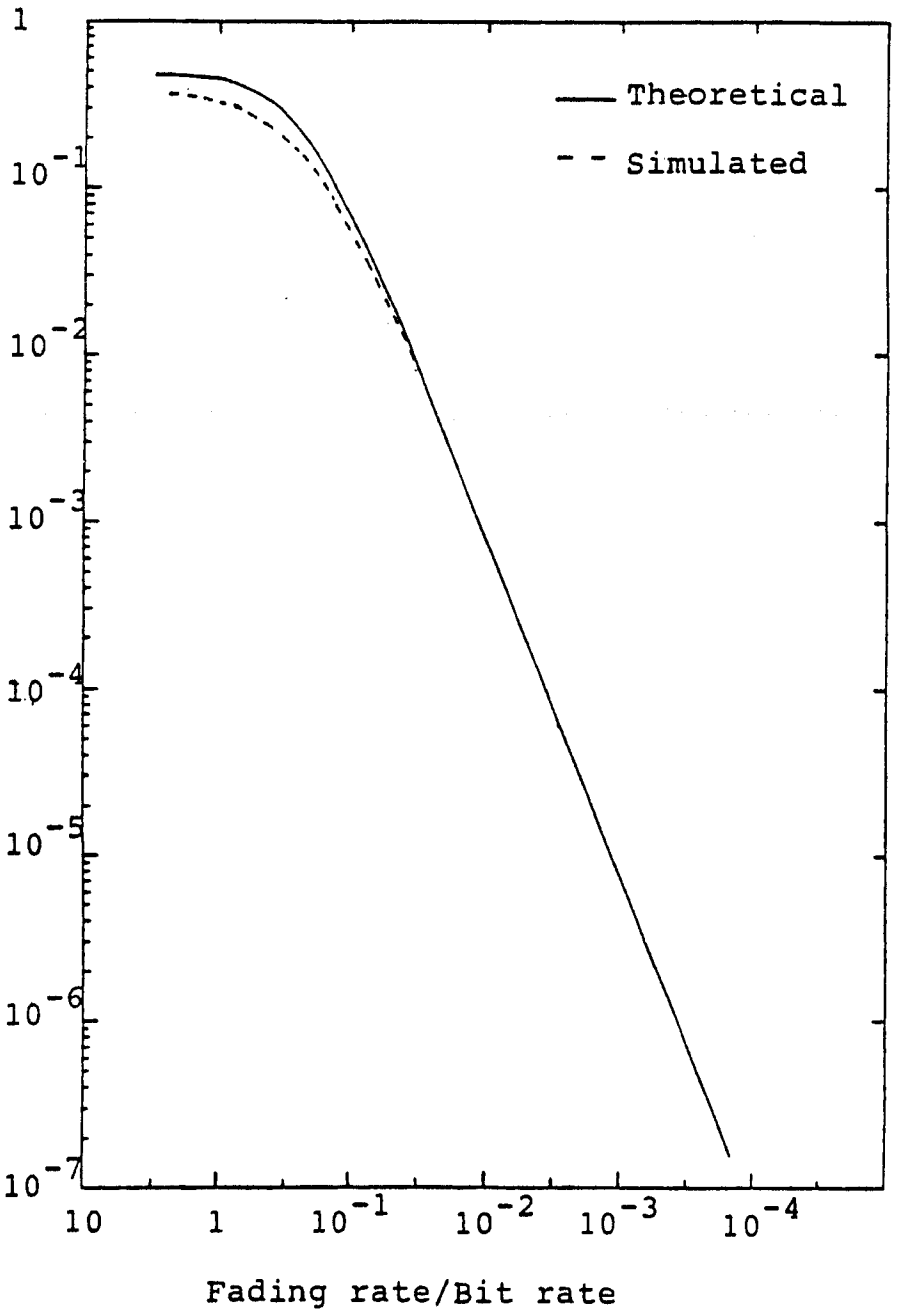


Fig. 3.8: Probability of error at infinite signal-to-noise ratio for DPSK over Rayleigh fading channel as a function of the fading rate.

is 0.33 instead of 0.46. When the theoretical probability of error is 0.1, then the simulation yields 0.078. At a more practical fading rate of 1 Hz, the probability of error drops to 7.1×10^{-4} and the deviation between the two curves is well within the statistical error of the simulated experiment.

3.6 AVOIDING TRANSIENTS IN THE DIGITAL FILTER.

The tap gains in the simulation of the fading channel should be stationary processes. Therefore transients should be avoided. This is a problem when the simulation is first started and when frequency hopping is in effect. When the system hops to a new frequency, the assumption is that there is no correlation between the tap gains just after the hop and those just before the hop.

One solution is simply to wait for transients to die away by letting the filter run for several time-constants before starting the simulation and between the last sample on the old frequency and the first sample on the new frequency. This solution is impractical when the fading rate is very low.

A more elegant solution is to find the variance of the

contents of each of the three registers in the filter (see Fig. 3.7) and the correlation between them. The problem of setting or resetting the filter then becomes one of simply generating three random variables with a given covariance matrix. We first describe the procedure for generating three such variables and then the procedure for finding the covariance matrix.

The calculations that now follow are with reference to [15] pp. 225 and 253 to 257.

Let the contents of the three taps be x_1, x_2 and x_3 as indicated in Fig. 3.7. They have a joint probability density function

$$f(x_1,x_2,x_3) = 1/\sqrt{(2\pi)^3|\mu|}\ \exp[(-1/2)\underline{x}A\underline{x}^T] \qquad (3.6.1)$$

where

$$\underline{x} = [x_1,x_2,x_3] \qquad (3.6.2a)$$

$$\mu = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \\ \mu_{31} & \mu_{32} & \mu_{33} \end{bmatrix} \qquad (b)$$

$$\mu_{ij} = \overline{x_i x_j} = R_{ij} \text{ (covariance)} \qquad (c)$$

$$A = \mu^{-1}$$

The generating procedure is the following:

1) Choose a value for x_1. Since x_2 and x_3 are not

known x_1 is a random variable with probability density function

$$f(x_1) = 1/\sqrt{2\pi\sigma_1^2}\ \exp[-(x_1^2)/(2\sigma_1^2)] \qquad (3.6.3a)$$

where

$$\sigma_1^2 = R_{11} \qquad (b)$$

Therefore

$$x_1 = n_1 \qquad (3.6.4)$$

where n_1 is a Gaussian random variable with variance R_{11} and zero mean. It can be generated using any of the standard methods for generating Gaussian random variables.

2) Choose a value for x_2. Since x_1 is known and x_3 is not known, x_1 is a random variable conditional on x_1 and

$$f(x_2|x_1) = 1/\sqrt{2\pi e_2}\ \exp[-(x_2-a_1x_1)^2/(2e_2)] \qquad (3.6.5a)$$

where

$$a_1 = r\sigma_1\sigma_2/\ \sigma_1^2 = R_{12}/R_{11} \qquad (b)$$

and

$$e_2 = \sigma_2^2 - r^2\sigma_1^2\sigma_2^2\ /\ \sigma_1^2 \qquad (c)$$

r is the correlation coefficient. Now

$$x_2 = a_1x_1 + n_2 \qquad (3.6.6)$$

where n_2 is a Gaussian random variable with variance e_2 and zero mean.

3) Choose x_3. Since x_1 and x_2 are known, x_3 is a random variable conditional on x_1 and x_2 and

$$f(x_3|x_1,x_2) = (1/2\pi e_3)\exp[-(x_3-b_1x_1-b_2x_2)^2/2e_3] \qquad (3.6.7a)$$

with

$$b_1 = (R_{31}R_{22} - R_{12}R_{32})/(R_{11}R_{22} - R_{12}R_{21}) \qquad (b)$$

$$b_2 = (R_{11}R_{32} - R_{31}R_{21})/(R_{11}R_{22} - R_{12}R_{21}) \qquad (c)$$

$$e = R_{33} - (b_1R_{31} + b_2R_{32}) \qquad (d)$$

Therefore

$$x_3 = b_1x_1 + b_2x_2 + n_3 \qquad (3.6.8)$$

where n_3 is a Gaussian random variable with variance e_3 and zero mean.

Proof of the validity of the generating procedure is offered in Appendix 2. It was also demonstrated in the simulation that no transients occur after resetting the filter.

The covariance matrix is calculated in the same way as in Eq. 3.5.16. The reader is referred to the discussion following that equation and to Fig. 3.7 for more detailed explanations.

Let

$$x_1(m) = x(m)*h_1(m) \tag{3.6.9a}$$

$$x_2(m) = x(m)*h_2(m) \tag{b}$$

$$x_3(m) = x(m)*h_3(m) \tag{c}$$

Then

$$R_{ij}(m) = x_i(k)x_j(k+m) = R_{xx}(m)*h_i(m)*h_j^*(-m) \tag{3.6.10}$$

and

$$\mu_{ij} = R_{ij}(0) = \sum_{k=0}^{\infty} h_i(k)h_j(k) \tag{3.6.11}$$

If the filter is driven with a unit impulse, then

$$x_i(k) = h_i(k) \tag{3.6.12}$$

and

$$\mu_{ij} = \sum_{k=0}^{\infty} x_i(k)x_j(k) \tag{3.6.13}$$

The infinite summation is approximated as in section 3.5.

This concludes all calculations necessary to explain the procedure for resetting, without generating transients, the digital filters generating the tap gains for the fading channel.

3.7 CONCLUSION

It was shown that the random process with Gaussian autocorrelation function that is required for the tap gains

in the simulation of the HF channel can be generated with a high degree of accuracy by using a third-order transitional Gaussian or Bessel low-pass filter. To meet the specification of [30], fourth-order filters have to be used. An adequate degree of accuracy for the purpose of this dissertation could be obtained by using a simple third-order filter with three equal real poles. Practical considerations dictated that this simple third-order filter be used. The probability of error for DPSK as a function of the fading rate of the channel, was found by simulation and compared with the theoretically expected values. At large fading rates there was some deviation. This corresponds to fading rates greater than 10 Hz when the bit rate is 100 bits/s. At more practical values of fading rate below 1 Hz, with the same bit rate, the simulation results followed the theoretical curve with no perceptible deviation.

4. PROBABILITY OF ERROR FOR M-ARY DPSK WITH M=8 AND M=16

4.1 INTRODUCTION.

This chapter deals with the evaluation of the probability of bit error for DPSK with 8 and 16 phases, with Gray coding, in the presence of white Gaussian noise (AWGN channel) and on the slowly Rayleigh fading channel.

The evaluation is based on formulas for the cumulative distribution function of the phase difference between two vectors perturbed by Gaussian noise, from a paper by Pawula, Rice and Roberts [16].

It should be pointed out that this analysis does not lead to new results. The probabilities of error were derived by Proakis [19], but in a form that is not suitable for expansion to the case where amplitude modulation is superimposed on the phase modulation, as will be required in chapter 5.

4.2 SYSTEM DESCRIPTION FOR M-ARY DPSK WITH M=8

A block diagram of the system is given in Fig. 4.1. Complex notation is used for mathematical convenience. Some people find the transition from complex algebra to signals in the real world problematical. A complex number in a block

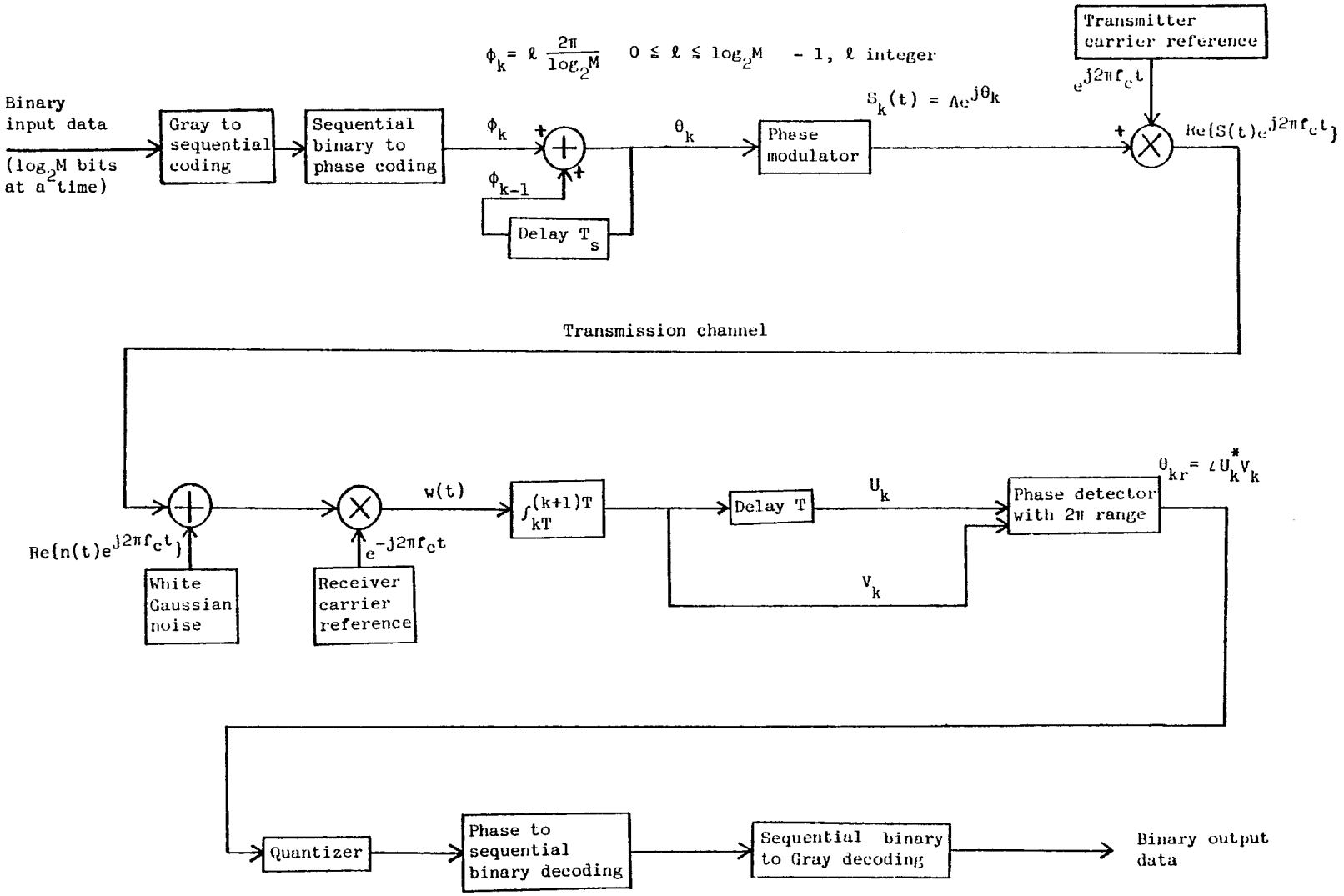


Fig. 4.1: Block diagram of system for M-ary DPSK.

diagram of a system presupposes the existence of two lines (in an analogue system) or two registers (in a digital system), one containing the real part of the number, the other containing the imaginary part of the number. The imaginary number j does not exist as such in the real system, but it is a mathematical operator that determines how the signals on the two lines shall interact with one another, and with other signals on similar pairs of lines. Fig. 4.2 shows an example of how complex multiplication will be done in a real system.

When real transmission channels are involved, only the real part of the signal modulated on the complex carrier is transmitted, as can be seen from Fig. 4.1. Many authors simply assume the reader knows that and do not bother to indicate that only the real part is transmitted.

Normally there will be a phase offset and a small frequency offset between the transmitter and receiver carrier references. The phase offset does not matter because of the differential phase modulation, and it is assumed that the frequency offset is small enough to cause negligible phase drift during the duration of a symbol. Therefore, these offsets are not shown in Fig. 4.1. It is also assumed that perfect symbol and bit synchronization is available.

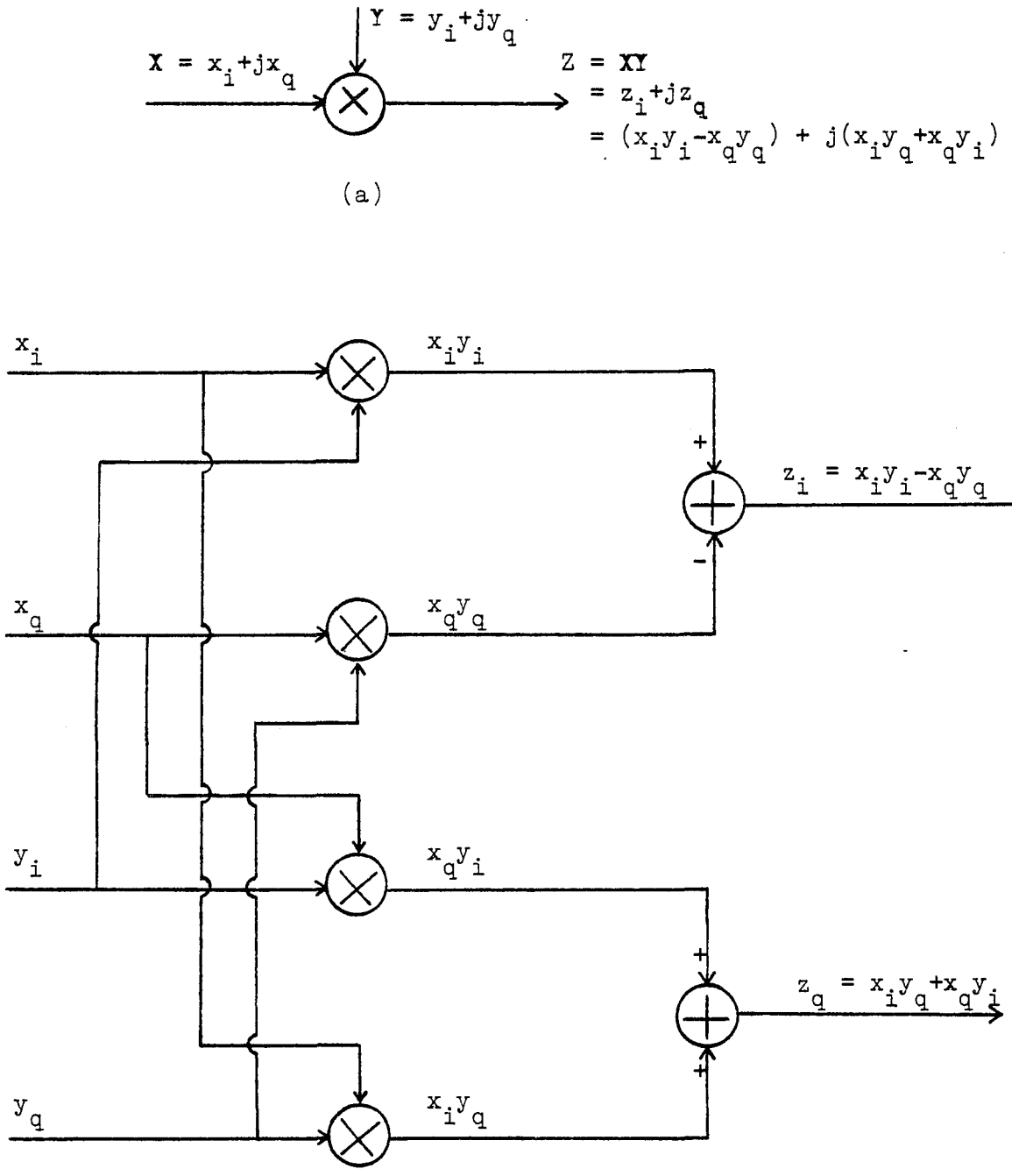


Fig. 4.2: (a) Block diagram of multiplier using complex notation. (b) Same system in real world, using real numbers only.

4.3 PROBABILITY OF BIT ERROR WITH GRAY CODING GIVEN THE TYPE OF SYMBOL ERROR (M=8)

It can be seen from Fig. 4.3 that if a certain symbol (phase) is received, and an error is made to the nearest neighbor, then only one of the three bits will be in error. That is a property of the Gray code.

Errors to symbols other than the nearest neighbor can cause one, two or three bit errors, but on average it always causes 2 bit errors when M=8, as can be seen from Fig. 4.4.

4.4 PROBABILITY OF SYMBOL ERROR (M=8)

Let us assume that the signal is transmitted over a channel where the only detrimental effect is additive white Gaussian noise (the AWGN channel). If $\phi_k = 0$, then the probability of a symbol error is given by (see Fig. 4.1 and 4.3):

$$P(se) = \int_{-\pi}^{\pi/8} f_{\theta_{kr}}(\psi)\,d\psi + \int_{\pi/8}^{\pi} f_{\theta_{kr}}(\psi)\,d\psi \qquad (4.4.1)$$

where $f_{\theta_{kr}}(\psi)$ is the probability density function of the received phase θ_{kr} given that $\phi_k = 0$ was transmitted. It is shown in [16] that $f_{\theta_{kr}}(\psi)$ is an even function, therefore

$$P(se) = 2\int_{\pi/8}^{\pi} f_{\theta_{kr}}(\psi)\,d\psi$$

$$= 2[F(\pi) - F(\pi/8)] \qquad (4.4.2)$$

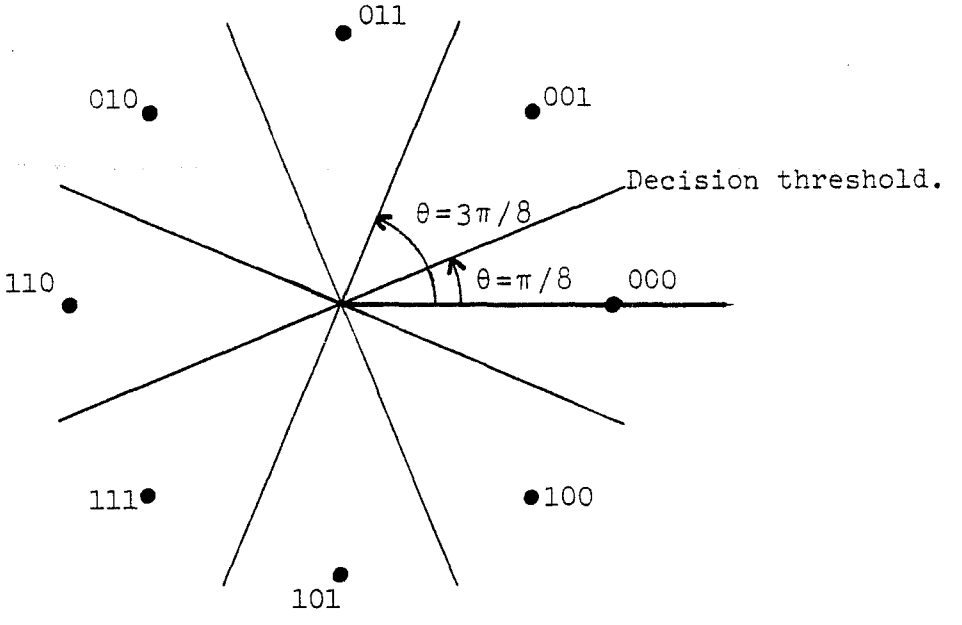


Fig. 4.3: Gray code for M-ary DPSK (M = 8).

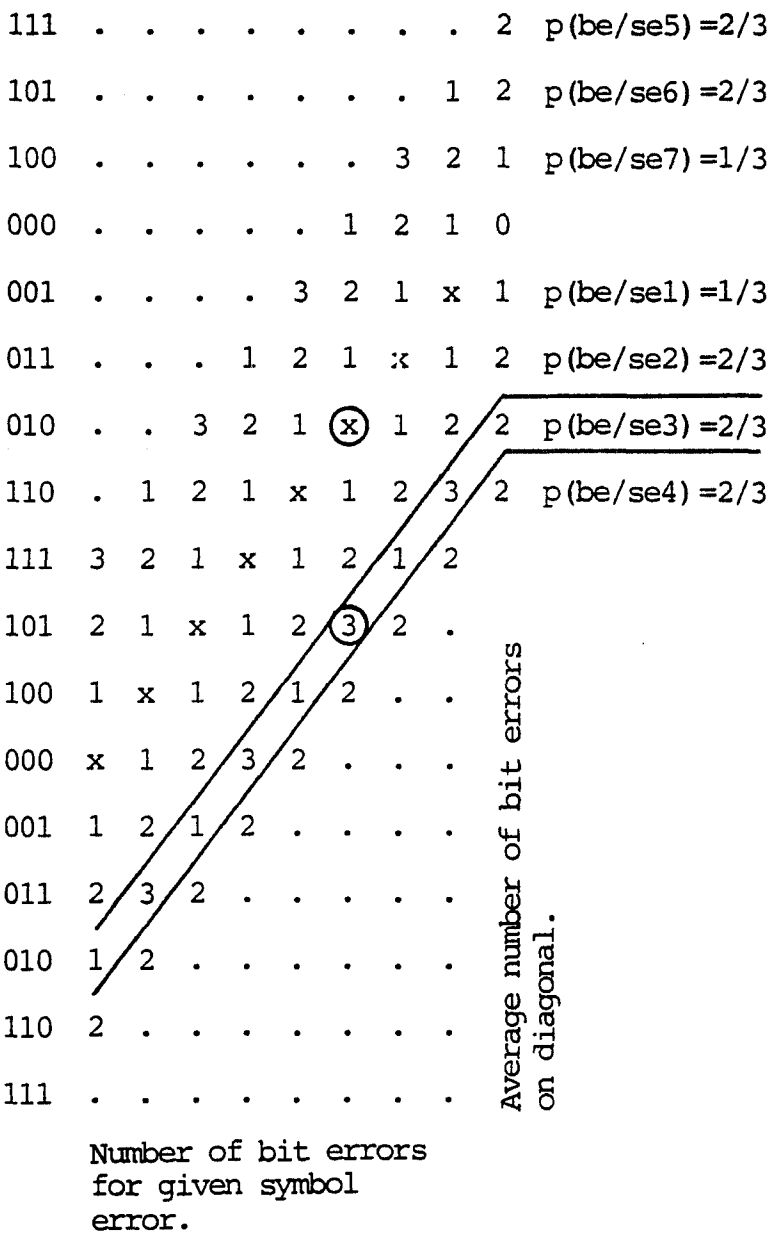


Fig. 4.4: Table for calculating the average number of bit errors for a given symbol error.

where $F(\theta)$ is the cumulative distribution function:

$$F(\theta) = \int_{-\pi}^{\theta} f(\Psi)d\Psi \qquad (4.4.3)$$

$F(\theta)$ is given by [16] (their Eq. 11, modified to suit our purpose):

$$G(\theta) = [-\gamma\sin(\theta)/4\pi] \int_{-\pi/2}^{\pi/2} dt \cdot \frac{\exp[-\alpha+\beta\sin(t)-\gamma\cos(\theta)\cos(t)]}{\alpha-\beta\sin(t)-\gamma\cos(\theta)\cos(t)} \qquad (4.4.4)$$

and

$$F(\theta) = \begin{cases} G(\theta) & -\pi<\theta\leq 0 \\ 1+G(\theta) & 0<\theta\leq\pi \end{cases} \qquad (4.4.5)$$

Stated in this form $F(\theta)$ is a true cumulative distribution function in the range $-\pi<\theta\leq\pi$. The parameters α, β, and γ are given by:

$$\alpha = (\rho_v+\rho_u)/2 \qquad (4.4.6a)$$

$$\beta = (\rho_v-\rho_u)/2 \qquad (b)$$

$$\gamma = \sqrt{\rho_v\rho_u} = \sqrt{\alpha^2-\beta^2} \qquad (c)$$

$$\rho_v = |V_k|^2/2\sigma^2 = E_{sv}/\eta \qquad (d)$$

(ρ_v is the SNR when V_k is transmitted. See Fig. 4.1)

$$\rho_u = |U_k|^2/2\sigma^2 = E_{su}/\eta \qquad (e)$$

η is the single-sided noise power spectral density. The noise power in a bandwidth B Hz is given by the product ηB. The fact that $|V_k|^2/2\sigma^2 = E_{sv}/\eta$ follows from the theory of matched filter receivers [27].

Since we are dealing with pure phase shift keying, $E_{sv}=E_{su}=3E_b$, where E_b is the energy per bit. Therefore $\beta=0$.

The integral in the expression for $G(\theta)$ has to be evaluated using numerical methods. In this case the program package DCADRE from [11] was used.

As a practical matter, the probability of bit error is more important than the probability of symbol error. The next section is devoted to finding the probability of bit error.

4.5 PROBABILITY OF BIT ERROR (M=8)

The probability of bit error is given by

$$P(be) = \sum_{j=1}^{7} P(be/Sej)P(Sej) \qquad (4.5.1)$$

where P(Sej) is the probability of a symbol error to the j-th symbol, when the intended symbol is numbered 0 as in Fig. 4.3. P(be/Sej) was found in section 4.3. P(Sej) can be found from the probability density function of the received phase. As an example:

$$P(Sel) = \int_{\pi/8}^{3\pi/8} f_{\theta_{kr}}(\psi)\,d\psi \qquad (4.5.2)$$

After simplifying,

$$P(be) = 2[(1/3)\int_{\pi/8}^{3\pi/8} f_{\theta_{kr}}(\psi)\,d\psi + (2/3)\int_{3\pi/8}^{\pi} f_{\theta_{kr}}(\psi)\,d\psi]$$

$$= (2/3)[2F(\pi) - F(3\pi/8) - F(\pi/8)] \qquad (4.5.3)$$

$F(\theta)$ and all the parameters necessary to evaluate it is given in section 4.4. A computer program was written to evaluate P(be). It is plotted in Fig. 4.7.

4.6 PROBABILITY OF BIT ERROR ON A SLOWLY RAYLEIGH FADING CHANNEL (M=8)

The probability of bit error on the AWGN channel as derived in section 4.5 is a function of the SNR. We now change our notation to stress that fact. Let

$$P_{be}(\rho)=P(be) \tag{4.6.1}$$

where P(be) is given by Eq. 4.5.3.

To find the probability of error in the slowly Rayleigh fading channel one has to multiply $P_{be}(\rho)$ with the probability that that particular SNR will occur, and then average that over all possible values of SNR.

The amplitude of the received signal has the Rayleigh distribution ([33], p. 529)

$$f(x) = (2x/b)\exp(-x^2/b)u(x) \tag{4.6.2}$$

and the average received power is

$$\overline{x^2} = b \tag{4.6.3}$$

Normalizing so that $2\sigma^2= 1$ or $\eta = 1$ (Eq. 4.4.6e), we can let $b = \overline{E}_s/\eta$, the average SNR (remember x is the square root of the instantaneous SNR.)

The probability of error on the Rayleigh fading channel is now given by:

$$P_R(be) = \int_0^\infty (2x/b)\exp(-x^2/b)P_{be}(x^2)\,dx \tag{4.6.4}$$

The result of a computer evaluation of this equation is given in Fig. 4.8.

Note on evaluation of $P_R(be)$: Since $P_{be}(\rho)$ contains an integral (Eq. 4.5.3 and 4.4.4), $P_R(be)$ contains a double integral. The subroutine DBLINT [11] can be used for that, but in later chapters P_{be}'s containing double integrals will be encountered, requiring evaluation of a triple integral in P_R, for which no ready-made program was available. It was therefore decided to minimize programming effort by writing a program that would calculate P_R, given a table of values for any $P_{be}(\rho)$. The table consisted of 17 values of $\log(P_{be}(\rho))$ spaced at 2dB intervals from ρ=-10dB to ρ = 22db. Next, for every interval of SNR, a parabolic curve was fitted to the nearest point at the next higher SNR in the table.

In calculating P_R, the appropriate parabolic was used to find by interpolation first $\log(P_{be}(\rho))$, then $P_{be}(\rho)$ itself, for any value of ρ required by the DCADRE integration routine. For values outside of the table, extrapolation from the first and the last parabolic is used, taking care to observe the limit $P_{be}(\rho) \leq \frac{1}{2}$.

This procedure was tested against the closed-form expression for P_R for binary DPSK, and was found to give results to an error of smaller than 1%, which is adequate for the type of

graph presented in this dissertation. Obviously, the error can be reduced by adding more values to the interpolation table.

4.7 PROBABILITY OF BIT ERROR GIVEN THE TYPE OF SYMBOL ERROR (M=16).

The procedure is exactly as in section 4.3. It is sufficient to present here in Fig. 4.5 and Fig. 4.6 only the Gray code for M=16. and the resulting average probability of bit error for a given type of symbol error.

4.8 PROBABILITY OF BIT ERROR (M=16)

Using the same procedure as in section 4.5,

$$P(be) = \sum_{j=1}^{15} P(be/Sej)P(Sej) \qquad (4.8.1)$$

P(be/Sej) is obtained from Fig. 4.6. P(Sej) is calculated using Eq. 4.4.4 and 4.4.5. for example:

$$P(Sel) = F(3\pi/16) - F(\pi/16) \qquad (4.8.2)$$

After simplifying:

$$\begin{aligned} P(be) = (1/4)[&-2F(\pi/16) - 2F(3\pi/16) - F(5\pi/16) \\ &+ F(7\pi/16) + F(9\pi/16) - F(11\pi/16) \\ &- 2F(13\pi/16) - 2F(15\pi/16) + 8F(\pi)] \end{aligned} \qquad (4.8.3)$$

Eq. 4.8.3 was evaluated and plotted in Fig. 4.6. A simulation experiment using a modified version of the COMSIM program [4] was also done. The results are also shown

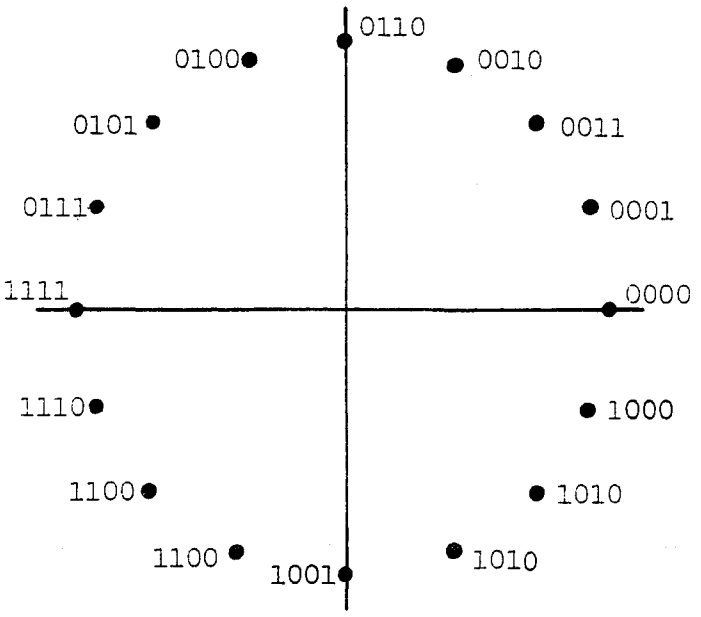


Fig. 4.5: Gray code for M-ary DPSK (M = 16).

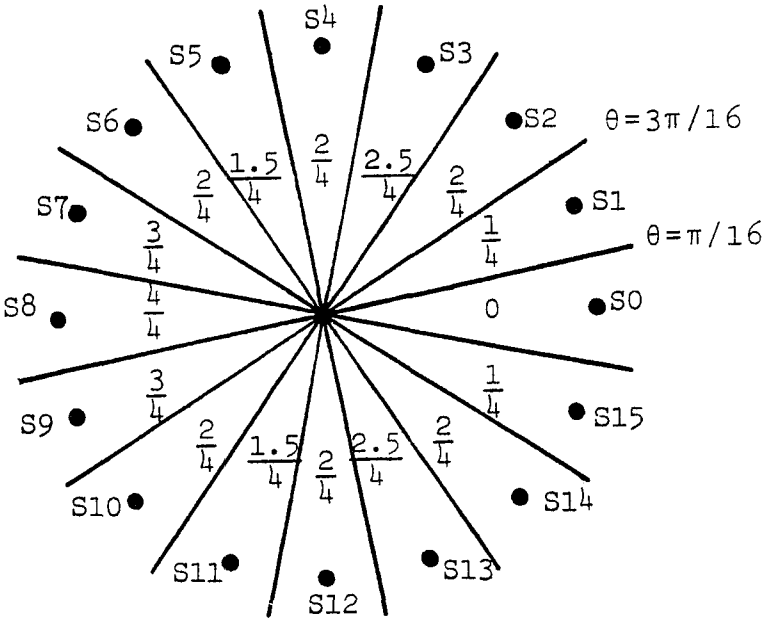


Fig. 4.6: Average probability of bit error for a given symbol error.

in Fig. 4.7 and are in good agreement with the theoretical curve.

4.9 PROBABILITY OF BIT ERROR ON A SLOWLY RAYLEIGH FADING CHANNEL (M=16).

The same procedure as in section 4.6 was used. Eq. 4.6.4 was evaluated, with $P_{be}(x^2)=P(be)$ as given by Eq. 4.8.3. The resulting curve is found in Fig. 4.8. A simulation experiment was also done. The simulation results vary somewhat around the theoretical curve. The reason for that is that not enough uncorrelated sample points were obtained due to the cost of running the simulation long enough at low fading rates. The pertinent input and output parameters of the simulation program are given in Fig. 4.9. The effect of the fading rate is discussed in chapter 8.

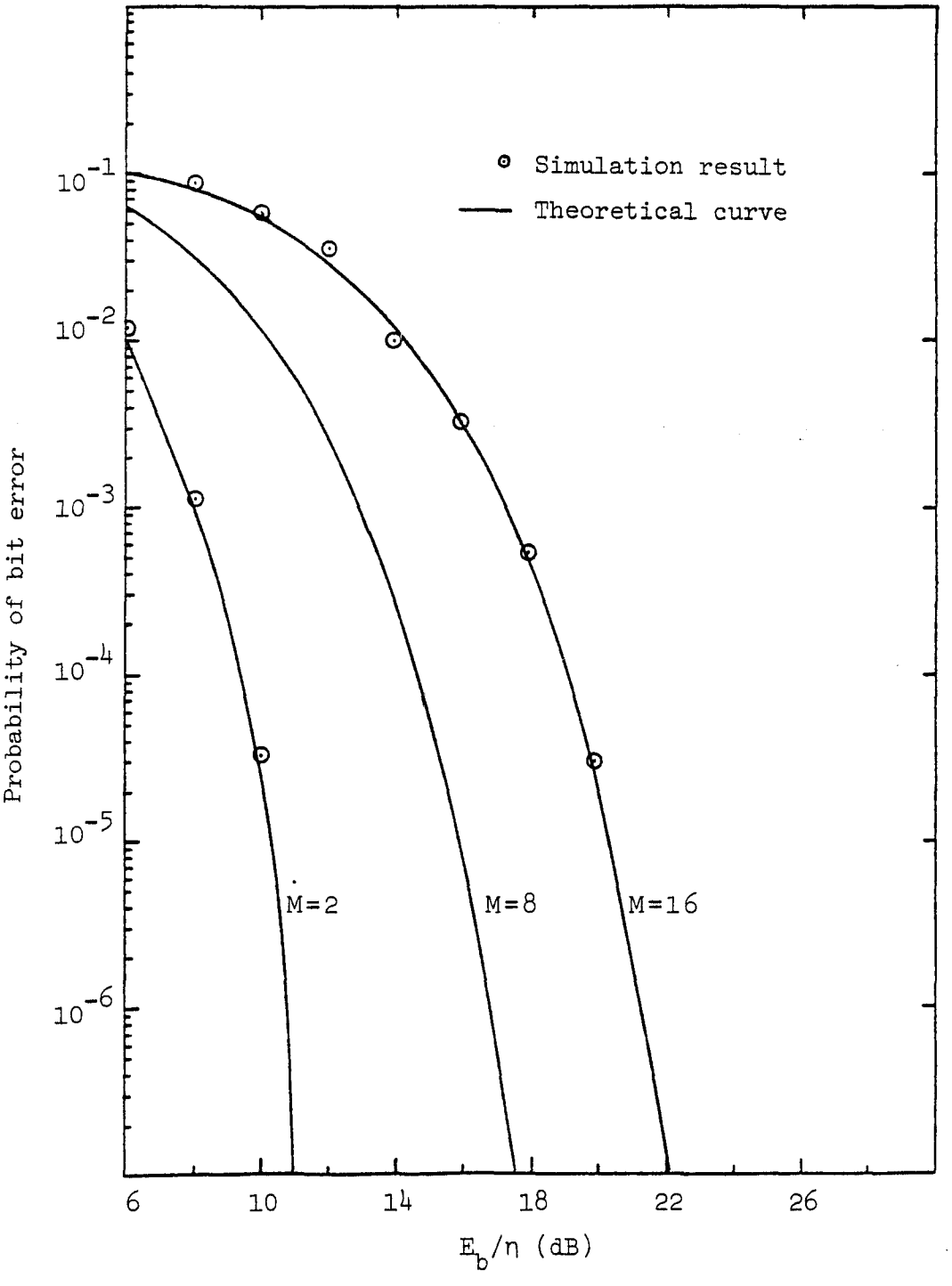


Fig. 4.7: Probability of bit error for AWGN channel. M-ary DPSK with M = 2,8,16.

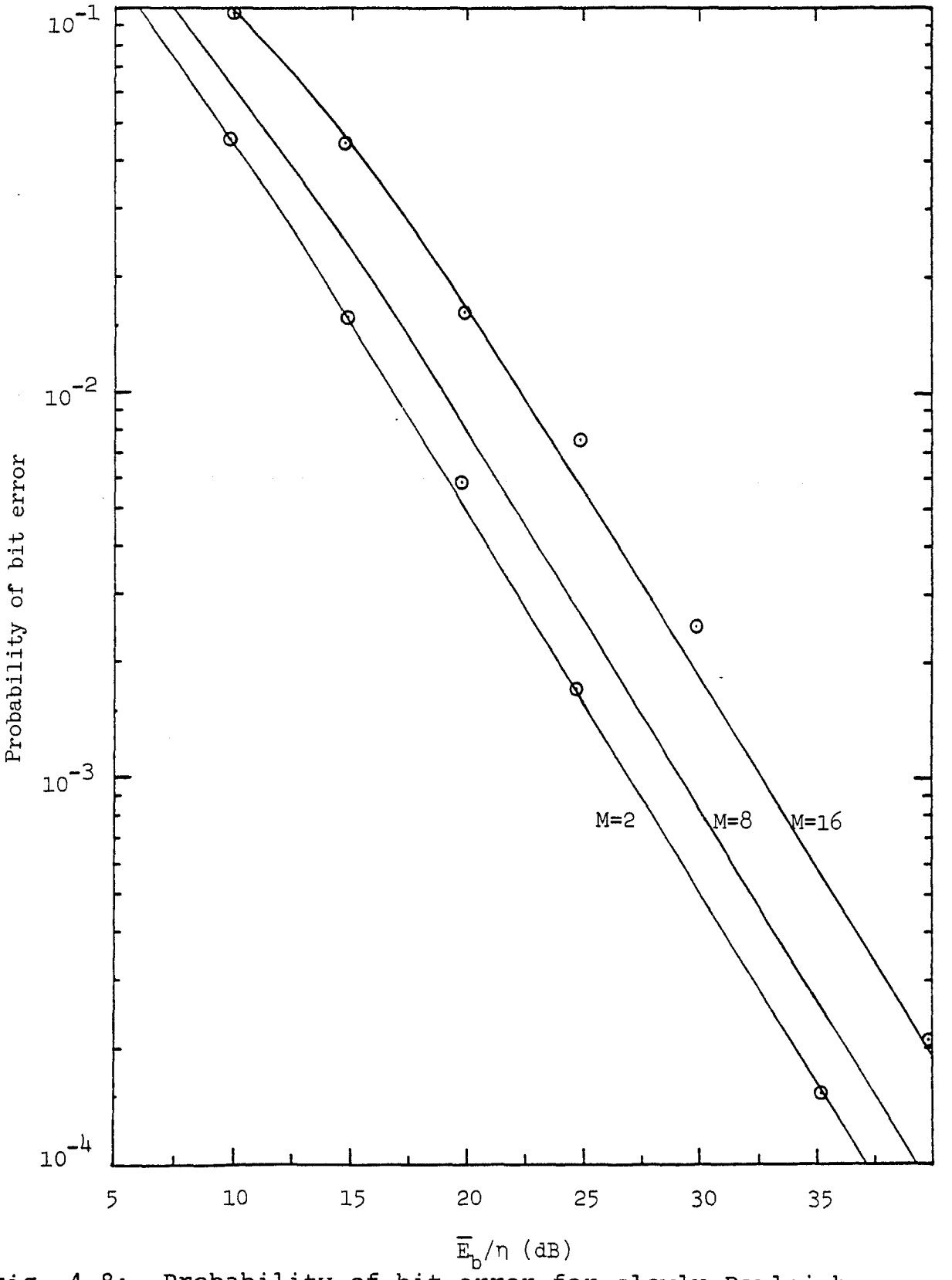


Fig. 4.8: Probability of bit error for slowly Rayleigh fading channel. M-ary DPSK with M= 2,8,16.

Bit rate: 300 bps.

Multipath spread: 10 μs.

SNR (dB)	Fading Rate (Hz.)	Duration in simualated real time, seconds	P(be)
0	0.1	300	.290
10	0.03	1000	.103
15	0.03	1000	4.55×10^{-2}
20	0.03	1000	1.64×10^{-2}
25	0.01	3000	7.55×10^{-3}
30	0.01	3000	2.42×10^{-3}
40	0.003	3000	1.98×10^{-4}

Fig. 4.9: Table of input and output parameters for simulation of M-ary DPSK (M=16) on slowly Rayleigh fading channel.

5. PROBABILITY OF ERROR FOR 8/2-DPASK ("SCHEME 2").

5.1 INTRODUCTION AND SYSTEM DESCRIPTION.

The reader is referred to section 1.2, where the notion of differential amplitude shift keying is explained, and to section 4.2, where the details of the differential 8-phase modulator will be found.

Fig. 5.1 shows how the differential amplitude and differential phase modulation is integrated into one system.

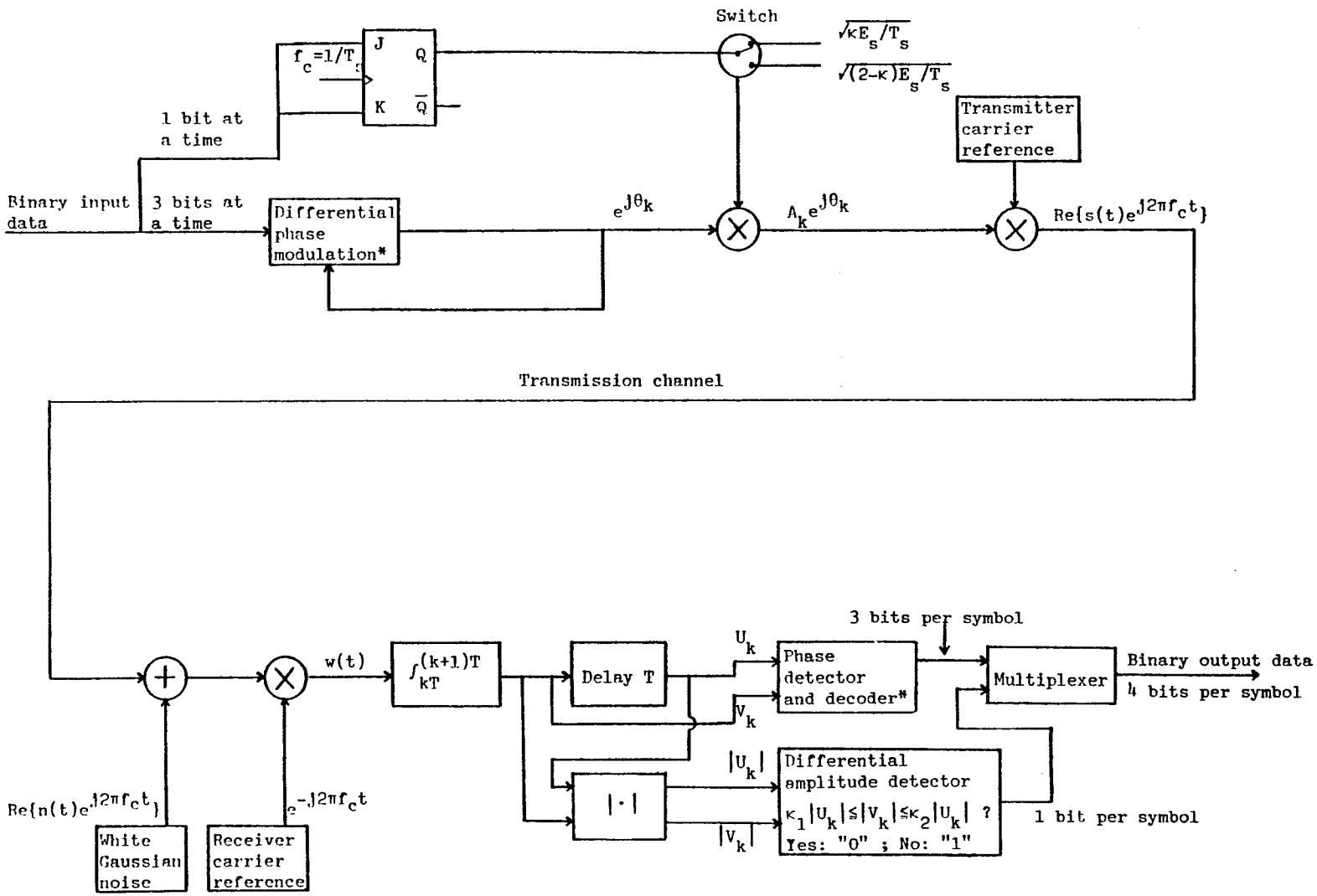
The probability of bit error can be found by separate analyses for the bits carried on the phase and the bit carried on the amplitude. The probability of error is given by

$$P(be) = (3/4)P(pbe) + (1/4)P(abe) \qquad (5.1.1)$$

where P(pbe) is the probability of phase bit error and P(abe) is the probability of amplitude error. P(pbe) and P(abe) are derived in sections 5.2 and 5.3 respectively.

5.2 PROBABILITY OF PHASE BIT ERROR.

The base for this analysis was laid out in sections 4.3, 4.4 and 4.5. The difference now is that the amplitude can change from symbol to symbol.



*As in Fig. 4.1 with M=8.

Fig. 5.1 : Block diagram of system for 8/2-DPASK ("Scheme 2").

It can be seen from Fig. 5.1 that the transmitted energy per symbol can be either κE_s or $(2-\kappa)E_s$. Since the two energy levels are equally likely, the average energy per symbol remains unchanged at E_s.

Likewise, the average SNR is given by $\rho = E_s/\eta = 4E_b/\eta$, while the SNR during any one symbol interval can be either

$$\rho_1 = \kappa E_s/\eta = \kappa\rho \qquad (5.2.1a)$$

or

$$\rho_2 = (2-\kappa)E_s/\eta = (2-\kappa)\rho \qquad (b)$$

Now, let P_{ij} be the probability of phase bit error when

$$\rho_V = |V_k|^2/2\sigma^2 = \rho_i \qquad (5.2.2a)$$

and

$$\rho_U = |U_k|^2/2\sigma^2 = \rho_j \qquad (b)$$

Since ρ_1 and ρ_2 can appear with equal probability, the average probability of phase bit error, is given by

$$P(pbe) = (1/4)\sum_{i=1}^{2}\sum_{j=1}^{2} P_{ij}(pbe) \qquad (5.2.3)$$

P_{ij} is derived by following the procedure leading to Eq. 4.5.3, and is given by

$$P_{ij}(pbe) = (2/3)[2F_{ij}(\pi) - F_{ij}(3\pi/8) - F_{ij}(\pi/8)] \qquad (5.2.4)$$

$F_{ij}(\theta)$ is the same as $F(\theta)$ (Eq. 4.4.4 and 4.4.5) except for the replacement of the parameters α, β and γ by α_{ij}, β_{ij}

and γ_{ij}. These parameters are as follows:

$$\alpha_{11} = (1/2)(\rho_1 + \rho_1) = \kappa\rho \qquad 0 < \kappa < 1 \qquad (5.2.5)$$
$$\beta_{11} = (1/2)(\rho_1 - \rho_1) = 0$$
$$\gamma_{11} = \sqrt{\rho_1\rho_1} = \kappa\rho$$

$$\alpha_{12} = (1/2)(\rho_1 + \rho_2) = \rho$$
$$\beta_{12} = (1/2)(\rho_1 - \rho_2) = (\kappa - 1)\rho$$
$$\gamma_{12} = \sqrt{\rho_1\rho_2} = \kappa\rho$$

$$\alpha_{21} = (1/2)(\rho_2 + \rho_1) = \rho$$
$$\beta_{21} = (1/2)(\rho_2 - \rho_1) = (1 - \kappa)\rho$$
$$\gamma_{21} = \sqrt{\rho_2\rho_1} = \rho\sqrt{\kappa(2-\kappa)}$$

$$\alpha_{22} = (1/2)(\rho_2 + \rho_2) = (2 - \kappa)\rho$$
$$\beta_{22} = (1/2)(\rho_2 - \rho_2) = 0$$
$$\gamma_{22} = \sqrt{\rho_2\rho_2} = (2 - \kappa)\rho$$

The same computer program mentioned in section 4.5, with some modifications, was used to evaluate the P_{ij} (pbe). Fig. 5.2 shows results of analysis and simulation with $\kappa = 0.5$. The choice of the value of κ is discussed in section 5.4.

5.3 PROBABILITY OF AMPLITUDE BIT ERROR.

The data which is differentially modulated on the amplitude, is detected by deciding if the amplitude of the present symbol ($|V_k|$) falls within a range of values derived from

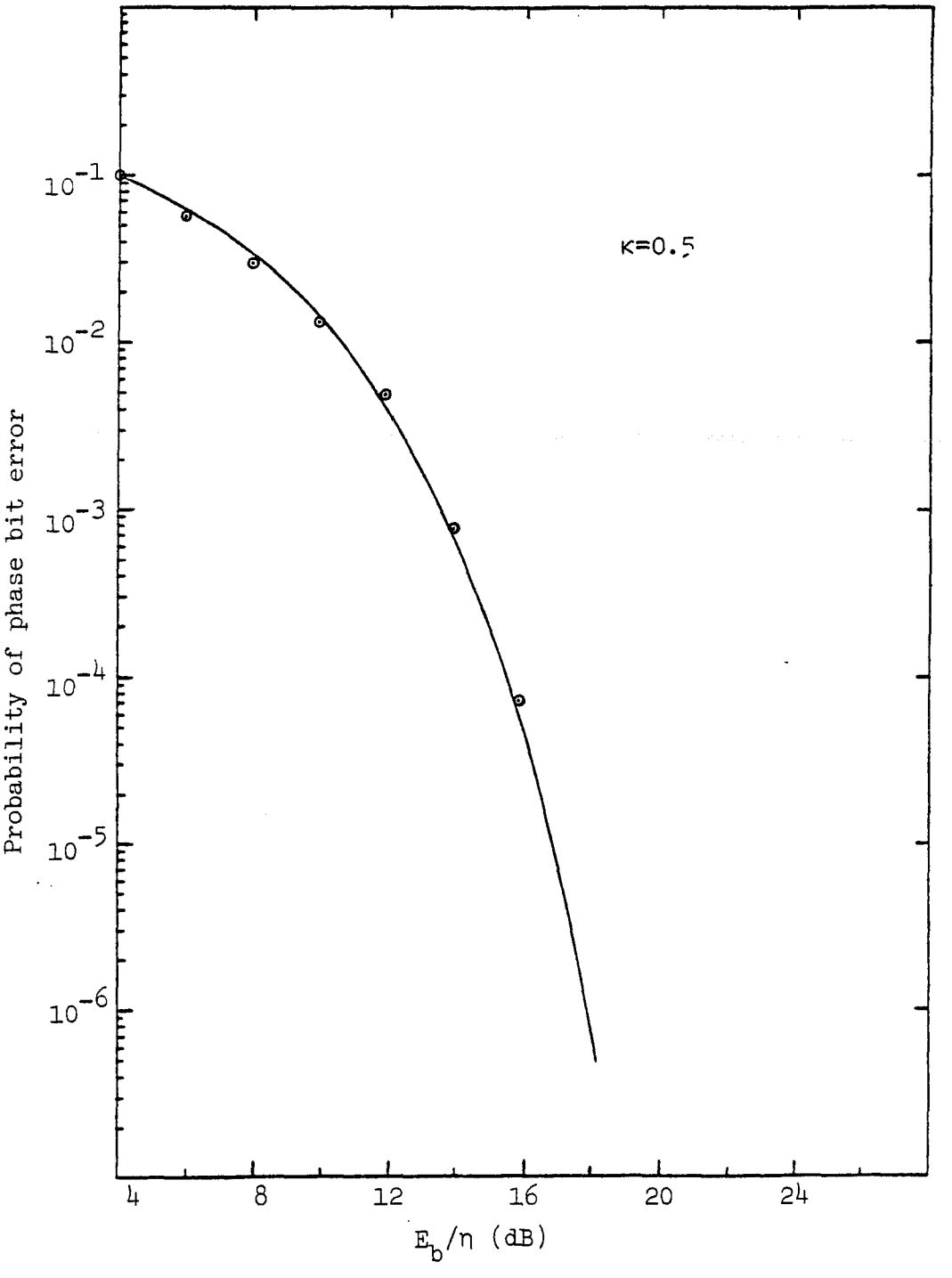


Fig. 5.2: Probability of phase bit error for 8/2-DPASK ("Scheme 2").

the amplitude of the previous symbol ($|U_k|$). In order to derive an expression for the probability of error, we need an expression for the joint probability density function of $|V_k|$ and $|U_k|$.

First of all, $|V_k|$ and $|U_k|$ are the outcomes of independent experiments. Therefore the joint probability density function will be the product of the individual probability density functions ([15], section 6.3)

The amplitude of a vector, phasor or sinusoid perturbed by Gaussian noise, has a Rician probability density function. From [19], Eq. 1.1.134 :

$$f_z(z) = (z/\sigma^2)\exp[-(z^2+s^2)/2\sigma^2]I_o(zs/\sigma^2) \qquad z \geq 0 \tag{5.3.1}$$

where z is the random variable representing the amplitude of the vector, s is the amplitude of the vector before the noise is added, $2\sigma^2$ is the noise power and $I_o(x)$ is the modified Bessel function of the first kind of order 0:

$$I_o(x) = (1/2)\int_0^{2\pi} \exp[x\cos(\theta)]d\theta \tag{5.3.2}$$

The SNR is given by

$$\rho = s^2/2\sigma^2 \tag{5.3.3}$$

If the parameters are normalized such that $2\sigma^2 = 1$, i.e. $s^2 = \rho_i$ then Eq. 5.3.1 becomes:

$$f_z(z) = 2z\exp[-(z^2+\rho_i)]\ I_o(2z\sqrt{\rho_i}) \tag{5.3.4}$$

Now, the value of ρ_i depends on which of the two amplitude

levels was transmitted, and is given in Eq. 5.2.1. If we let $z_1 = |U_k|$ and $z_2 = |V_k|$, then, using the statistical independence discussed before,

$$f_{ij}(z_1,z_2) = 2z \exp[-(z_1^2+\rho_i)]\, I_0(2z_1\sqrt{\rho_i}\,) \cdot 2z \exp[-(z_2+\rho_j)]\, I_0(2z_2\sqrt{\rho_j}\,) \qquad (5.3.5)$$

The probability of amplitude bit error is given by

$$P(abe) = (1/4) \sum_{i=1}^{2} \sum_{j=1}^{2} P_{ij}(abe) \qquad (5.3.6)$$

where

$$P_{ij}(abe) = \iint_{A_{ij}} f_{ij}(z_1,z_2)\,dz_1\,dz_2 \qquad (5.3.7)$$

A_{ij} is that area in the (z_1,z_2) plane where the signal will be in error. As an example (See Fig. 5.3):

$$P_{11}(abe) = \int_0^\infty \int_0^{\kappa_1 z_1} f_{11}(z_1,z_2)\,dz_2\,dz_1 + \int_0^\infty \int_{\kappa_2 z_1}^\infty f_{11}(z_1,z_2)\,dz_2\,dz_1 \qquad (5.3.8)$$

The P_{ij}'s were evaluated using a computer program for double integration [11] that could only handle fixed boundaries. Therefore 5.3.7 was converted from rectangular to polar coordinates :

$$dz_2\,dz_1 = r\,dr\,d\theta \qquad (5.3.9)$$

$$f_{ij}(r,\theta) = 4r^2 \sin(\theta)\cos(\theta)\exp[-r^2-\rho_1-\rho_2] \cdot I_0[2r\sqrt{\rho_1}\cos(\theta)]\, I_0[2r\sqrt{\rho_2}\sin(\theta)] \qquad (5.3.10)$$

Finally, $P_{ij}(abe)$ is given by:

$$P_{\substack{ij \\ i=j}}(abe) = \int_0^\infty \int_0^{\tan^{-1}(\kappa_1)} f_{ij}(r,\theta)\,r\,d\theta\,dr + \int_0^\infty \int_{\tan^{-1}(\kappa_2)}^\infty f_{ij}(r,\theta)\,r\,d\theta\,dr \qquad (5.3.11a)$$

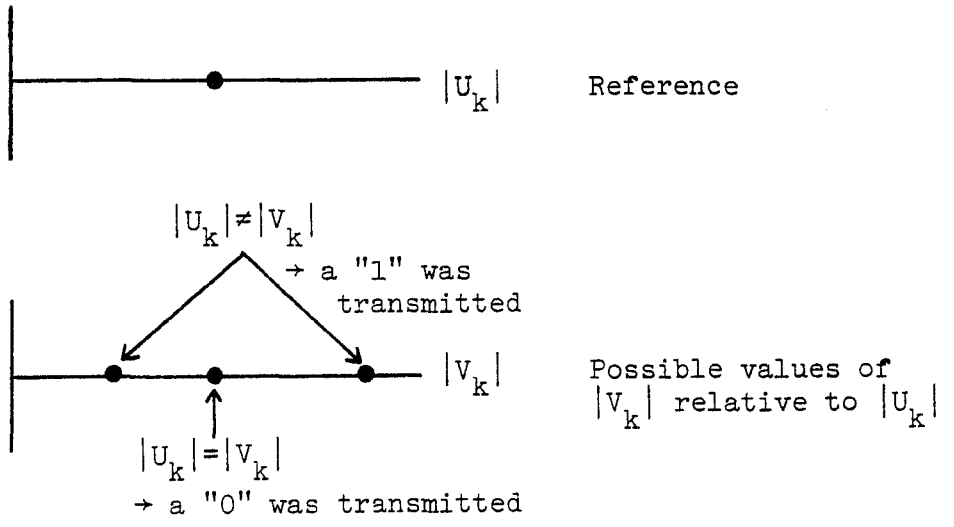


Fig. 5.3: Decision regions for amplitude modulated bit.

$$P_{\substack{ij \\ i\neq j}}(abe) = \int_0^{\infty} \int_{\tan^{-1}(\kappa_1)}^{\tan^{-1}(\kappa_2)} f_{ij}(r,\theta)\, r\, d\theta\, dr \qquad (5.3.11b)$$

The $P_{ij}(abe)$ were evaluated using digital computation methods, after which P(abe) was easy to calculate from Eq. 5.3.6. The result is plotted along with simulation results in Fig. 5.4 for $\kappa = 0.5$, $\kappa_1 = 0.789$ and $\kappa_2 = 1.365$.

NOTE ON EVALUATION OF EQ. 5.3.11: The equations were evaluated using the DBLINT subroutine from [11]. A number of problems were encountered. The first, concerning variable boundaries, and its solution, was already mentioned. The second problem concerns the upper limit of integration on r, infinity. It is obviously impossible to integrate to infinity, but since the integrand approaches zero exponentially for large values of r, it is possible to set the upper limit to some value that will give a result for the integral arbitrarily close to the true value of the integral. A practical value that gave good results for most situations was $10(\rho_1 + \rho_2)$.

The next problem has to do with overflow in the calculation of the Bessel function. The integrand contains the two Bessel functions, that are exponentially increasing with r and the exponential itself decreasing with r^2, so that the individual terms could be outside the range of values that the computer can handle, while the integrand as a whole

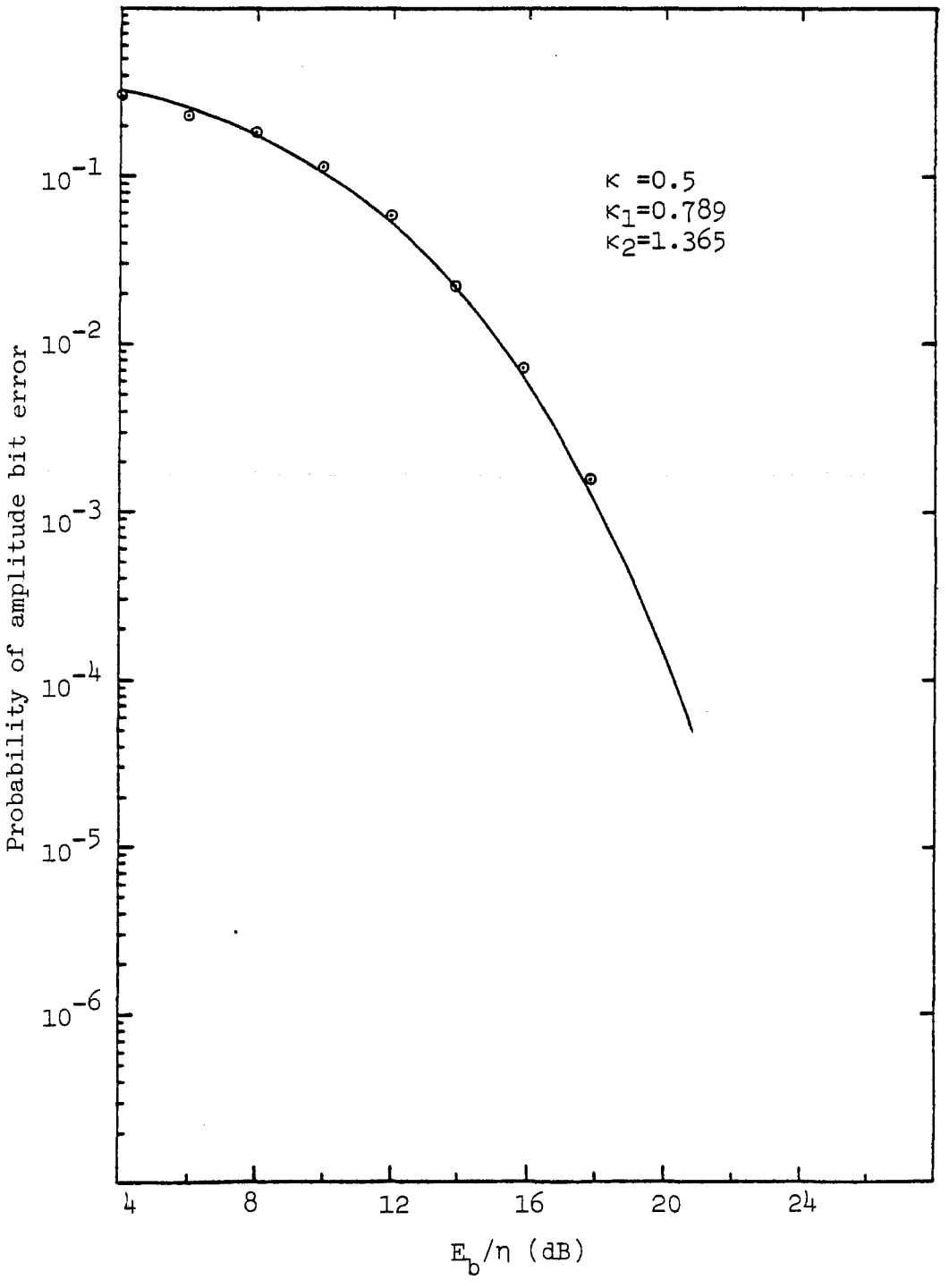


Fig. 5.4: Probability of amplitude bit error for 8/2-DPASK ("Scheme 2").

would still have a significant value.

This problem was overcome by using the approximation [26]

$$I_n(x) \cong \exp(x)/\sqrt{2\pi x} \qquad (5.3.12)$$

for large values of x (x > 178). The approximation differs from the true value by less than 0.1% at x = 178, and the difference decreases with increasing x. Using the approximation does not by itself solve the overflow problem, but it does allow the problem to be solved by first calculating the logarithm of the integrand.

The final problem concerns the fact that $f_{ij}(r,\theta)$ approaches a two-dimensional impulse for large values of SNR. The integration routine calculates values of the function at a certain set of locations (not controlled by the user) in the two-dimensional region of integration. If all those values are insignificantly small relative to some user-supplied number, then the integration routine declares the value of the integral to be zero. The problem now is that the location of the pseudo-impulse usually does not coincide with a member of this set of locations, so that the program usually fails to calculate the proper value.

The problem was solved by writing a subroutine to find the location of the maximum value of the integrand, and then to set up the limits of integration around that point in such a

way that the integration routine would always find enough significant values of the function to allow it to calculate the proper value of the integral.

5.4 OPTIMIZING THE PARAMETERS κ, κ_1 and κ_2.

The probability of bit error at a given SNR depends on κ, κ_1 and κ_2. κ specifies the separation of the two amplitude levels (Fig. 5.1.) while κ_1 and κ_2 specify the location of the decision thresholds for detection of the amplitude modulated bit (Fig. 5.3). κ depends on κ_1 and κ_2. κ_1 and κ_2 are independent of one another but not of κ.

A program was written to find by digital computation methods the partial derivative of P(be) with respect to κ, κ_1 and κ_2 The method of linear interpolation was then used to find the location of the zero of the partial derivative, given the other two parameters. P(be) was found to stabilize in the first two digits after about five such iterations. No further improvement could be obtained due to noise in the calculated value of the derivative, which in turn was due to noise from the double integral for P(be).

In such a search for the minimum of a function, it is always possible that the search can produce a local minimum instead of the global minimum of the function. To check for such a possibility, the search was done repeatedly at SNR = 8dB

with different starting values of the three parameters. Since the result was always the same, it was assumed that there are no local minima in the vicinity of the global minimum.

A more elegant mathematical solution for the optimum values of κ_1 and κ_2 for a given κ is certainly possible, but since the programs for calculating P(be) had to be written in any case, it was decided to use the method of digital computation. Furthermore, the mathematical solution for the optimum κ will be extremely tedious to find.

The optimum values of κ, κ_1 and κ_2 were calculated and plotted in Fig. 5.5. The resulting probability of error is shown in Fig. 5.6. Simulation results obtained by fixing the values of κ, κ_1 and κ_2 at the optimum value for SNR = 8dB, for all values of SNR, are also shown.

5.5 PROBABILITY OF ERROR IF AMPLITUDE MODULATION IS NOT DIFFERENTIAL.

If the receiver knew whether the previously received amplitude was high or low, it would be possible to drop the requirement of differential modulation, while still using the previously received amplitude as the reference for the present one. Some improvement in the probability of error could then be obtained.

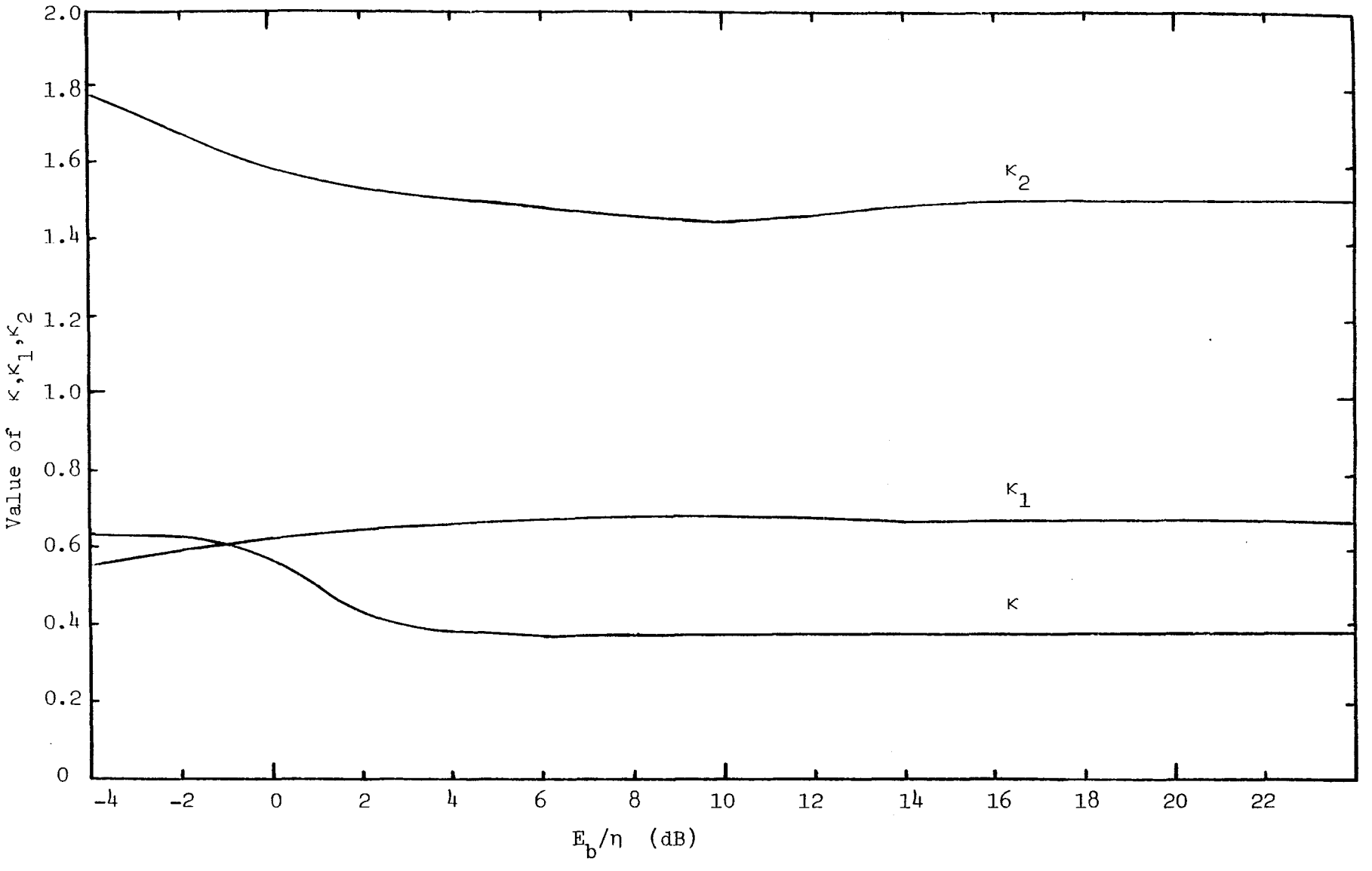


Fig. 5.5: Optimum values of κ, κ_1 and κ_2 on AWGN channel for 8/2-DPASK.

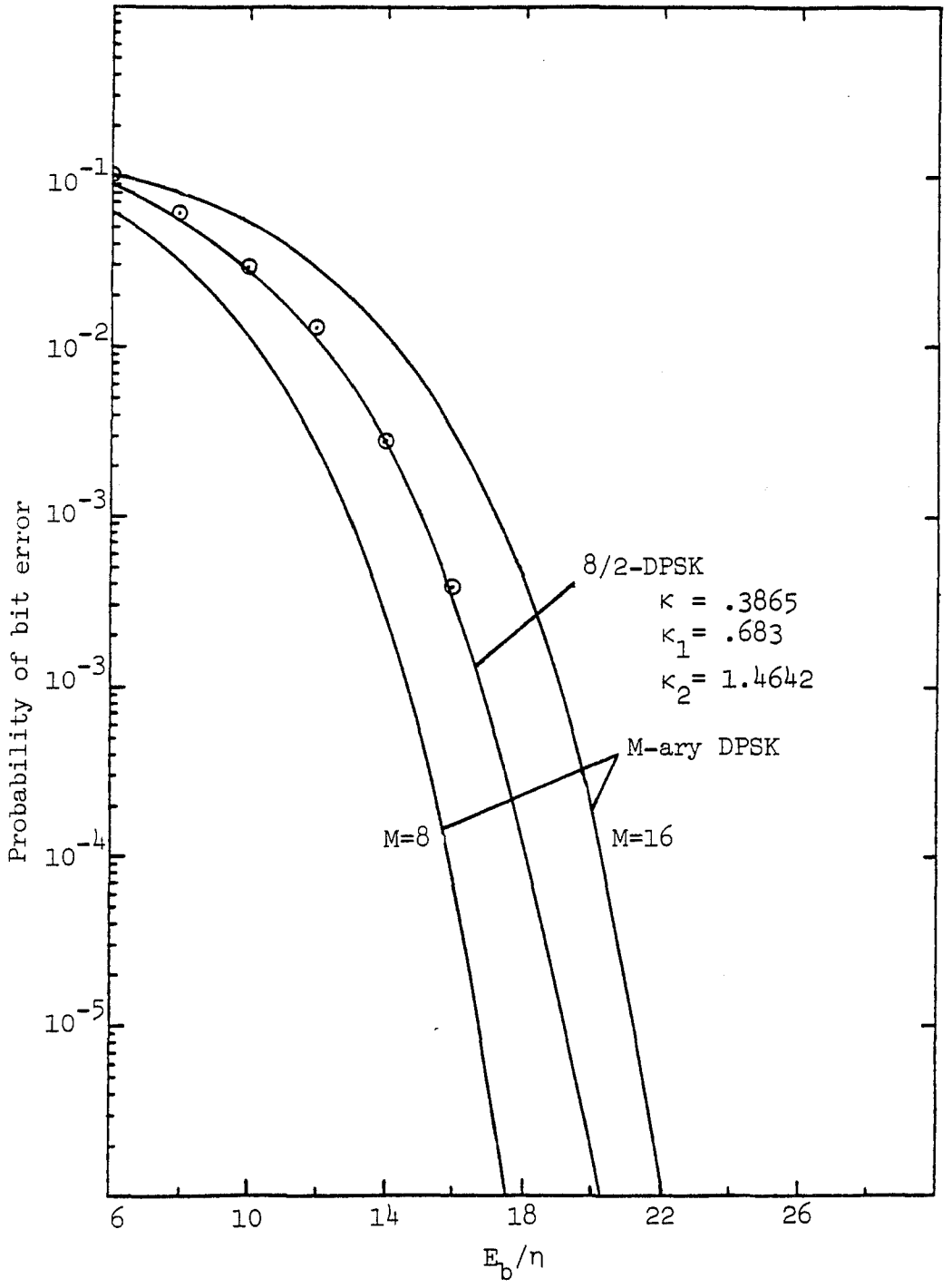


Fig. 5.6: Probability of bit error on AWGN channel for 8/2-DPASK ("Scheme 2").

The only way that the receiver can know what level was received, is to keep track of it on a symbol-by-symbol basis. As soon as an error is made with respect to the amplitude level, a burst of bit errors will follow. The length of the burst will be determined by the particular data sequence following the first error. For example, if a low level is mistaken for a high level, the received bits will remain locked in a state of either "1" or "0" until the next high to low transition is received, or until noise causes a second amplitude error to be made. The problem could be overcome by employing an error detection code, but the use of the code will have a cost by requiring some fraction of the available energy per bit.

A lower bound on the probability of error can be obtained by ignoring the effect of multiple errors and the energy required for error detection. After making the appropriate changes in the limits of integration in Eq. 5.3.11, the procedure of section 5.4 was followed to obtain the optimum values of κ, κ_1 and κ_2 and the resulting lower bound on the probability of error was calculated at a SNR of 16dB. The results are tabulated in Fig. 5.6.

The improvement in probability of bit error amounts to a gain in SNR of 0.2dB, but that is in the lower bound which

SNR = 16 dB.

	K	K_1	K_2	P(be)
*Differential:	0.385	0.663	1.508	3.32×10^{-4}
Not differential:	0.409	0.753	1.471	2.33×10^{-4}

*From Fig. 5.5 and 5.6.

Fig. 5.7: Comparative results for the case when the amplitude modulation is not differential.

is not very tight. One would therefore expect the actual performance to be worse than before. Harper [10] came to a similar conclusion. His argument is based on results for fading channels, while our analysis so far was only for the AWGN channel. This approach is thus abandoned.

5.6 PROBABILITY OF BIT ERROR ON SLOWLY RAYLEIGH FADING CHANNEL.

The procedure of section 4.6 was used, with P(be) given by Eq. 5.1.1., and the parameters κ, κ_1 and κ_2 as optimized in section 5.4 for each individual SNR. This approach therefore assumes that both the transmitter and the receiver have accurate knowledge of the instantaneous SNR on the fading channel and are adjusting the three parameters accordingly. It is clearly impractical for most HF applications, but it is a lower bound on the probability of error.

The probability of error was calculated a second time, now using a single set of κ's, as optimized in section 5.4 for SNR = 8dB. The result is less than 0.04dB worse than the lower bound calculated above (at high values of average SNR). The degradation will be more at very low values of average SNR (below 10dB, i.e. P(be) > 10^{-1}), but the conclusion remains the same: It is not necessary to adapt the parameters of the amplitude modulation to the

instantaneous SNR on the fading channel.

The numerically calculated probability of error and simulation results are shown in Fig. 5.8. The simulation was done with fixed κ, κ_1 and κ_2 as optimized at SNR = 8 dB for the AWGN channel (Fig. 5.5).

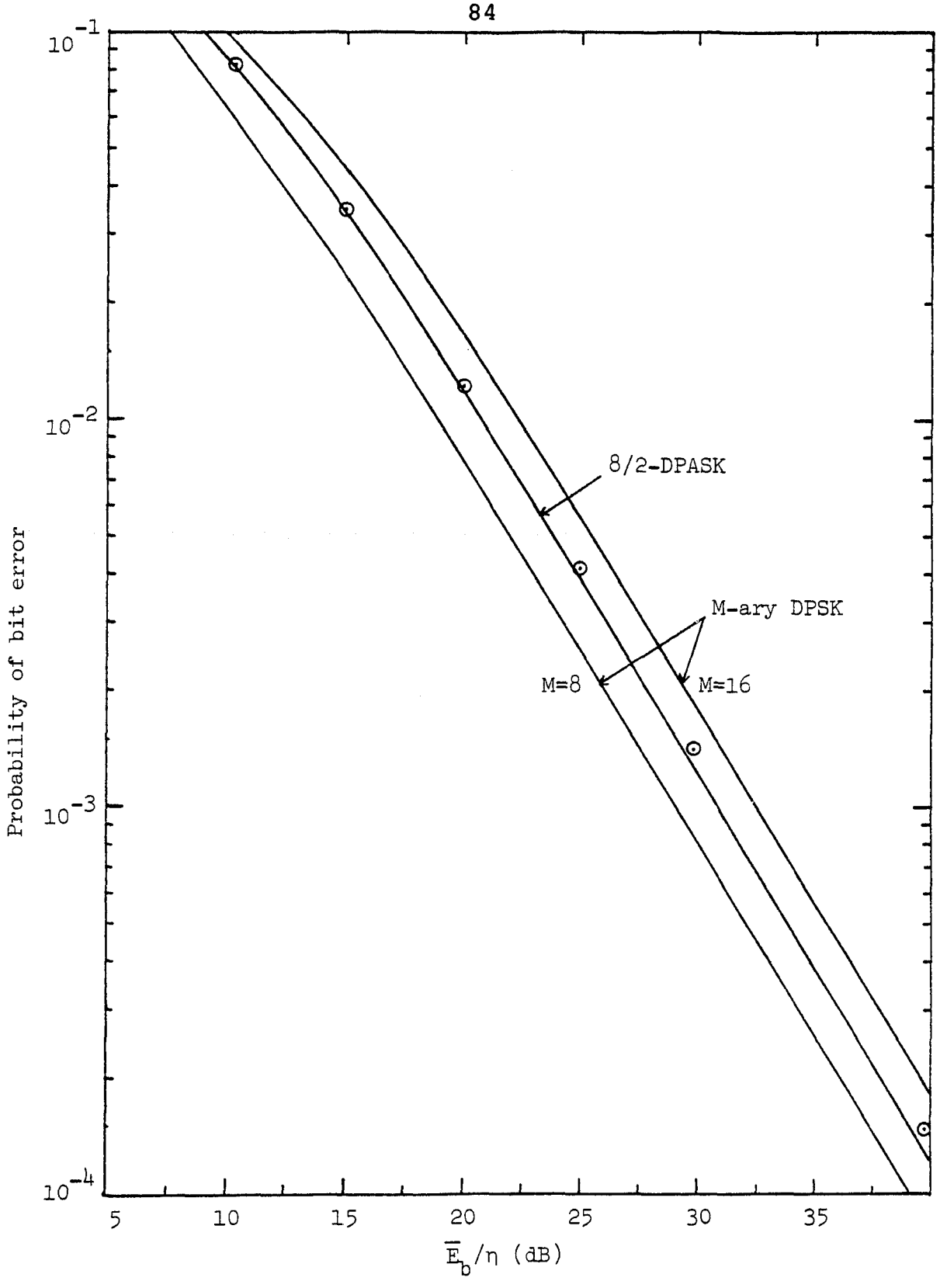


Fig. 5.8: Probability of bit error on slowly Rayleigh fading channel for 8/2-DPASK ("Scheme 2").

6. PROBABILITY OF ERROR FOR 8/2-DPASK WITH 22.5° PHASE SHIFT BETWEEN AMPLITUDE LEVELS.("SCHEME 3")

6.1 INTRODUCTION AND SYSTEM DESCRIPTION.

The reader is once more referred to section 1.2, where a description of the modulation format for "Scheme 3" is given.

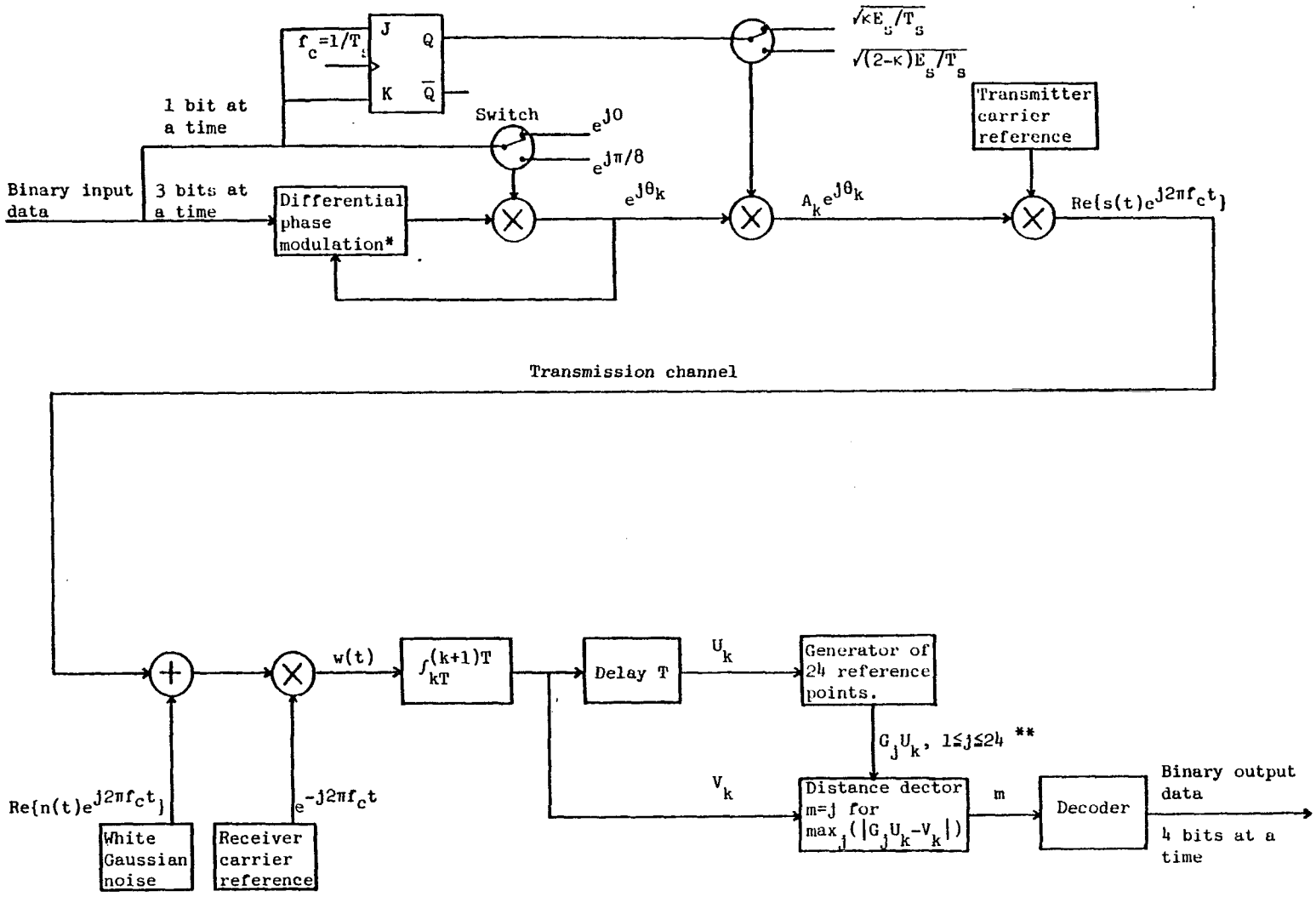
The system description is given in Fig. 6.1. The transmitter does not need further explanation, but the generation of the 24 reference points at the receiver needs some discussion. The resulting "template" is shown in Fig. 6.2. The generating constants are given by:

$$G_m = \sqrt{\kappa/(2-\kappa)}\ \exp(jm\pi/4) \qquad m = 1,8$$
$$G_m = \exp[j(m\pi/4 + \pi/8)] \qquad m = 9,16$$
$$G_m = \sqrt{(2-\kappa)/\kappa}\exp(jm\pi/4) \qquad m = 17,24 \qquad (6.1.1)$$

The decision as to which symbol was received is based on the weighted length $K_m|G_mU_k - V_k|$, where the weighting factor is given by:

$$K_m = \kappa_1 \qquad m = 1,8$$
$$K_m = 1 \qquad m = 9,16$$
$$K_m = \kappa_2 \qquad m = 17,24 \qquad (6.1.2)$$

The three values given for the G_m's and the K_m's are applicable respectively to the three cases when the amplitude changes from high to low, when no amplitude change occurs and when the amplitude changes from low to high.



*As in Fig. 4.1 with M=8.
**See text.

Fig. 6.1: Block diagram of system for 8/2-DPASK ("Scheme 3").

The weighting factor K_m is introduced to allow the decision thresholds between reference points of differing amplitude to be moved around. That is necessary because the variance of the noise associated with the T_m's varies with the amplitude level (because of the multiplication with G_m). The decision thresholds are sections of an ellipse when they are between T_m's of different amplitudes. Fig. 6.2 shows the thresholds for the special case when $\kappa_1 = \kappa_2 = 1$.

6.2 ON FINDING AN ESTIMATE OF THE PROBABILITY OF ERROR.

To find and evaluate an exact expression for the probability of bit error is a major problem. Some of the difficulties will be discussed in this section.

First some symbols have to be defined. The meaning of U_k and V_k is clear from Fig. 6.1. Then let

$$U_k = X_1 + N_1 \qquad (6.2.1)$$

where

$$X_1 = x_1 + jy_1$$

$$N_1 = n_{x1} + jn_{y1}$$

and let

$$V_k = X_2 + N_2 \qquad (6.2.2)$$

where

$$X_2 = x_2 + jy_2$$

$$N_2 = n_{x2} + jn_{y2}$$

The four noise components are uncorrelated and each has a mean value of zero and a variance of σ^2.

The reference points T_m (Fig. 6.1 and 6.2) are given by:

$$T_m = G_m U_k \qquad (6.2.3)$$

and the G_m's are given by Eq. 6.1.1.

Now we are ready for the discussion. Say the transmitter signal is such that the receiver ought to decide that T_9 was received (i.e. $V_k = T_9$ in the absence of noise). The probability of bit error is the probability that V_k will be closest to each of the other T_m's, times the average probability of bit error for each of those T_m's. The joint probability density of T_m and V_k is a function in four variables. The probability that V_k will be closest to some T_m is therefore given by an integral in four dimensions, and the limits of integration in two of those dimensions are given by the rather comolicated shapes shown in Fig. 6.2. It was decided that doing such an analysis was not worth the trouble. Instead, the union bound will be modified and then used to find an estimate of the probability of bit error.

The union bound is a well known upper bound for the probability of error for any M-ary signal constellation. [33]. It consists of calculating the probability of bit error for M-1 pairs of reference symbols, where one member of the pair is the correct symbol, and the other is each of

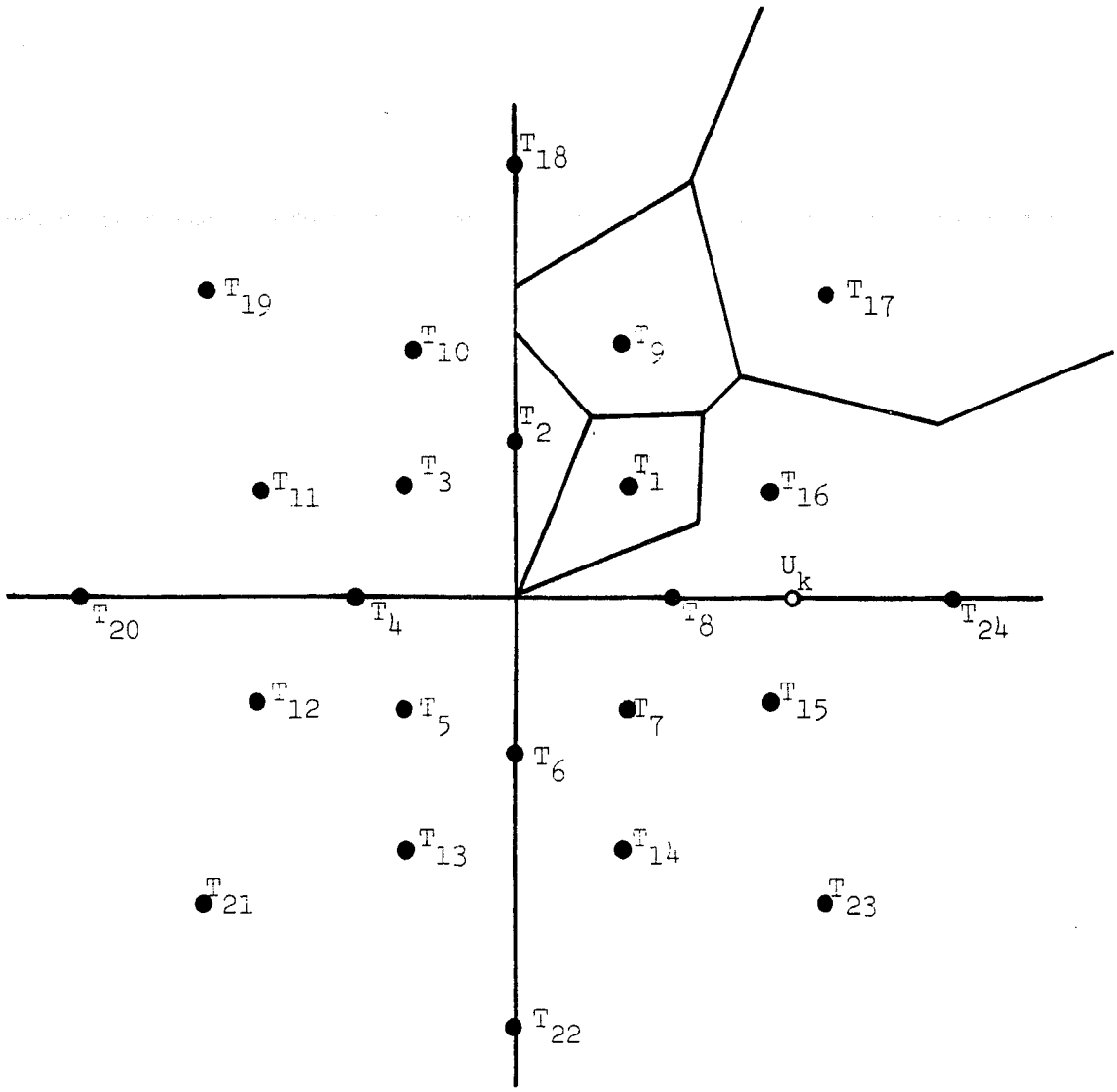


Fig. 6.2: "Template" with 24 reference points generated from previously received signals.

the other M-1 symbols taken one at a time. This adds up to all possible ways in which an error can be made, and many types of error are counted more than once.

We modify the union bound to include only the probability of error to nearest neighbors. For example, the probability of an error from T_{16} to T_9 (Fig. 6.2) includes the probability of an error from T_{16} to T_{18}. Therefore the latter is not included in the calculation. This slightly tightens the union bound.

We further modify the union bound to take the probability of bit error given a symbol error into account. With these two modifications, some symbol errors are counted up to three times, but then multiplied by a conditional probability of bit error which could be as low as 1/4 instead of 4/4. This makes the end result no longer an upper bound but merely an estimate of the probability of bit error.

6.3 CALCULATING AN ESTIMATE OF THE PROBABILITY OF BIT ERROR ON THE AWGN CHANNEL.

In this section a need will arise for several auxiliary variables. The following convention will be used:

1) The symbols $X_\ell = x_\ell + jy_\ell$ will be used for deterministic signals.

2) The symbols $N_\ell = n_{x\ell} + jn_{y\ell}$ will be used for complex Gaussian random variables (CGRV's). N_ℓ is usually associated with X_ℓ. $n_{x\ell}$ and $n_{y\ell}$ are also uncorrelated and have zero mean unless it is specified otherwise. ℓ will be a numeral 1,2,3 etc.

Before proceeding, the reader should be familiar with the definition of symbols in sections 6.1 and 6.2.

Now consider the reference points T_i and T_m and let $V_k = T_i$ in the absence of noise, i.e. let

$$X_2 = G_i X_1 \tag{6.3.1}$$

To calculate the union bound, it is necessary to find the probability

$$P(e/i,m) = P[K_i |T_i - V_k| > K_m |T_m - V_k|] \tag{6.3.2}$$

For ease of notation, let

$$D_i = d_{xi} + jd_{yi} = T_i - V_k$$
$$D_m = d_{xm} + jd_{ym} = T_m - V_k \tag{6.3.3}$$

and

$$z_i = |T_i - V_k|$$
$$z_m = |T_m - V_k| \tag{6.3.4}$$

Then, from Eq. 6.3.2:

$$\begin{aligned} P(e/i,m) &= P(K_i z_i > K_m z_m) \\ &= P(z_m < K_i z_i / K_m) \\ &= \int_0^\infty \int_0^{K_i z_i / K_m} f(z_i, z_m)\, dz_m\, dz_i \end{aligned} \tag{6.3.5}$$

$f(z_i,z_m)$ is the joint probability density function of the two lengths z_i and z_m, and can be found using Baye's theorem [15]:

$$f(z_i,z_m) = f(z_i)\ f(z_m \mid z_i) \tag{6.3.6}$$

$f(z_i)$ can easily be seen to be a Rayleigh density. z_i is the magnitude of D_i, which is a CGRV.

Using 6.3.3 and 6.2.1 through 6.2.3:

$$D_i = G_i N_1 - N_2 \tag{6.3.7}$$

d_{xi} and d_{yi} are uncorrelated and

$$\sigma_i^2 = d_{xi}^2 = d_{yi}^2$$

$$= \sigma\,(|G_i|^2 + 1) \tag{6.3.8}$$

Since $z_i = |D_i|$, then from [15] Eq. 7-17:

$$f(z_i) = (z_i / \sigma_i^2)\ \exp(-z_i^2/2\sigma_i^2) \qquad z_i > 0 \tag{6.3.9}$$

$f(z_m/z_i)$ is much more difficult to calculate. It is first necessary to calculate $f(z_m|D_i)$, and to do that we need D_m as a function of D. Using Eq. 6.3.3 and 6.2.1 through 6.2.3:

$$D_m = G_m X_1 - X_2 + G_m N_1 - N_2 \tag{6.3.10}$$

But from Eq. 6.3.7

$$N_2 = G_i N_1 - D_i \tag{6.3.11}$$

Substituting into Eq. 6.3.10:

$$D_m = G_m X_1 - X_2 + D_i + (G_m - G_i) N_1 \tag{6.3.12}$$

Eq. 6.3.12 contains two CGRV'S, D_i and N_1. They are

correlated, since D_i contains a term proportional to N_1(Eq. 6.3.7). But we are interested in the density function of $z_m = |D_m|$ when D_i is given, so we consider D_i to be deterministic. $N_1|D_i$ becomes a CGRV with nonzero mean and some variance that can be calculated. The following procedure to find that mean and variance is the same as in section 3.6, so the detailed explanations are not repeated here.

We need the three covariances:

$$R_{DD} = \overline{D_i D_i^*} = 2(1+|G_i|^2)\sigma^2$$

$$R_{NN} = \overline{N_1 N_1^*} = 2\sigma^2$$

$$R_{ND} = \overline{N_1 D_i^*} = 2\sigma^2 G_i^* \qquad (6.3.13)$$

Then

$$\overline{N_1|D_i} = R_{DN} D_i / R_{DD} = G_i^* D_i/(1+|G_i|^2) \qquad (6.3.14)$$

($\overline{N_1|D_i}$ is the mean value of N_1 when D_i is given)

The variance of $N_1|D_i$ is given by

$$|(N_1|D_i - \overline{N_1|D_i})|^2 = R_{NN} - |R_{DN}|^2/R_{DD}$$
$$= 2\sigma^2[1-|G_i|^2/(1+|G_i|^2)]$$
$$= 2\sigma^2/(1+|G_i|^2) \qquad (6.3.15)$$

Now, when D_i is given, let

$$N_1 = X_4 + N_4 \qquad (6.3.16)$$

where from Eq. 6.3.14

$$X_4 = G_i^* D_i/(1+|G_i|^2) \tag{6.3.17}$$

and N_4 is a CGRV with zero mean and variance of each quadrature component given by (from 6.3.15)

$$\sigma_4^2 = \sigma^2/(1+|G_i|^2) \tag{6.3.18}$$

Finally, substituting Eq. 6.3.16 and 6.3.17 into 6.3.12

$$D_m|D_i = G_m X_1 - X_2 + D_i[1+(G_m-G_i)G_i^*/(1+|G_i|^2)] + (G_m-G_i)N_4 \tag{6.3.19}$$

This is a deterministic complex number of length

$$\eta = |G_m X_1 - X_2 + D_i[1+(G_m-G_i)G_i^*/(1+|G_i|^2)] \tag{6.3.20}$$

perturbed by a CGRV with variance of each quadrature component given by

$$\begin{aligned}\sigma_m^2 &= |G_m-G_i|^2\sigma_4^2 \\ &= \sigma^2|G_m-G_i|^2/(1+|G_i|^2)\end{aligned} \tag{6.3.21}$$

Therefore the length of $D_m|D_i$ (i.e. $z_m|D_i$) has a Ricean probability density function:

$$f(z_m|D_i) = (z_2/\sigma_m^2)\exp[-(z_2^2+\eta^2)/2\sigma_m^2]I_o(z_2\eta/\sigma_m^2) \tag{6.3.22}$$

However, the distribution is needed not for D_i, but for $Z_i = |D_i|$. To that end, in Eq. 6.3.19, let

$$D_i = r\cos(\theta) + jr\sin(\theta) \tag{6.3.23}$$

Then average Eq. 6.3.22 over θ :

$$f(z_m|z_i) = (1/2\pi)\int_0^{2\pi} f(z_m|D_i)\, d\theta \tag{6.3.24}$$

P(e/i,m) (Eq. 6.3.5) can now be evaluated using Eq. 6.3.6, 6.3.9, and 6.3.25. The integral in Eq. 6.3.25 is evaluated using the DCADRE routine in [11], and Eq. 6.3.6 is evaluated in the same manner as the double integral discussed in section 5.3.

To complete the calculation, it is necessary to calculate the probability of error for the four possible amplitude sequences: low-low, low-high, high-low and high-high. The practice of chapter 5 of using a "1" subscript to indicate a low and a "2" to indicate a high, is continued.

It is necessary to specify i and m in Eq. 6.3.2, and X_1 and X_2 in Eq. 6.3.19 (or 6.2.1 and 6.2.2).

The calculations are normalized as follows: Let the SNR be given by

$$\rho = \overline{E}_s / 2\sigma^2 \tag{6.3.25}$$

Let

$$2\sigma^2 = 1 \tag{6.3.26}$$

Then

$$\overline{E}_s = \rho \tag{6.3.27}$$

and the two possible signal amplitude are given by

$$A_1 = \sqrt{E_1} = \sqrt{\kappa\rho}$$
$$A_2 = \sqrt{E_2} = \sqrt{(2-\kappa)\rho} \tag{6.3.28}$$

For P_{11}:

Let i = 16. Then, from Fig. 6.2, the nearest neighbors are given by (with the probability of bit error given that symbol error in brackets):

m = 8 (2/4), 1 (1/4), 9 (1/4), 17 (1/4), 24 (2/4), 15 (1/4).

But the probability of symbol error is the same when m = 8 and when m = 1, also for the pairs 9,15 and 24,17.

Therefore

$$P_{11} \cong (3/4)P_{11}(16,8) + (1/2)P_{11}(16,9) + (1/2)P_{11}(16,24)$$

and

$$X_1 = \sqrt{\kappa\rho} + j0$$

$$X_2 = G_1\sqrt{\kappa\rho} \qquad (6.3.29)$$

$P_{11}(16,8)$ is given by Eq. 6.3.6 with i =16, m = 8 and the X's as above.

For P_{22}:

Follow the same argument as for P_{11}, then

$$P_{22} \cong (3/4)P_{22}(16,8) + (1/2)P_{22}(16,9) + (1/2)P_{22}(16,24)$$

$$X = \sqrt{(2-\kappa)\rho} + j0$$

$$X = G_{16}\sqrt{(2-\kappa)\rho} \qquad (6.3.30)$$

For P_{12}:

Let i=17. Then the nearest neighbors are given by m = 18 (1/4), 9 (1/4), 16 (2/4) and 24 (1/4). The probability of symbol error is the same for m=18 and m=24, also for the pair 9,16. Therefore

$$P_{12} \cong (1/2)P_{12}(17,18) + (3/4)P_{12}(17,9)$$
$$X_1 = \sqrt{\kappa\rho} + j0$$
$$X_2 = G_{17}\sqrt{\kappa\rho} \qquad (6.3.31)$$

For P_{21}:

Let i =1. Then the nearest neighbors are given by m =2 (1/4), 9 (1/4), 16 (2/4) and 8 (1/4). The probability of symbol error is the same for the pairs 2,8 and 9,16. Therefore

$$P_{21} \cong (1/2)P_{21}(1,2) + (3/4)P_{21}(1,9)$$
$$X_1 = \sqrt{(2-\kappa)\rho} + j0$$
$$X_2 = G_1\sqrt{(2-\kappa)\rho} \qquad (6.3.32)$$

Finally the upper bound on the probability of bit error is given by:

$$P(be) \cong (1/4)(P_{11} + P_{12} + P_{21} + P_{22}) \qquad (6.3.33)$$

Eq. 6.3.33 was evaluated and is shown in Fig. 6.3 with some simulation results. It can be seen that the estimate is higher than the simulation results at low SNR's, while the

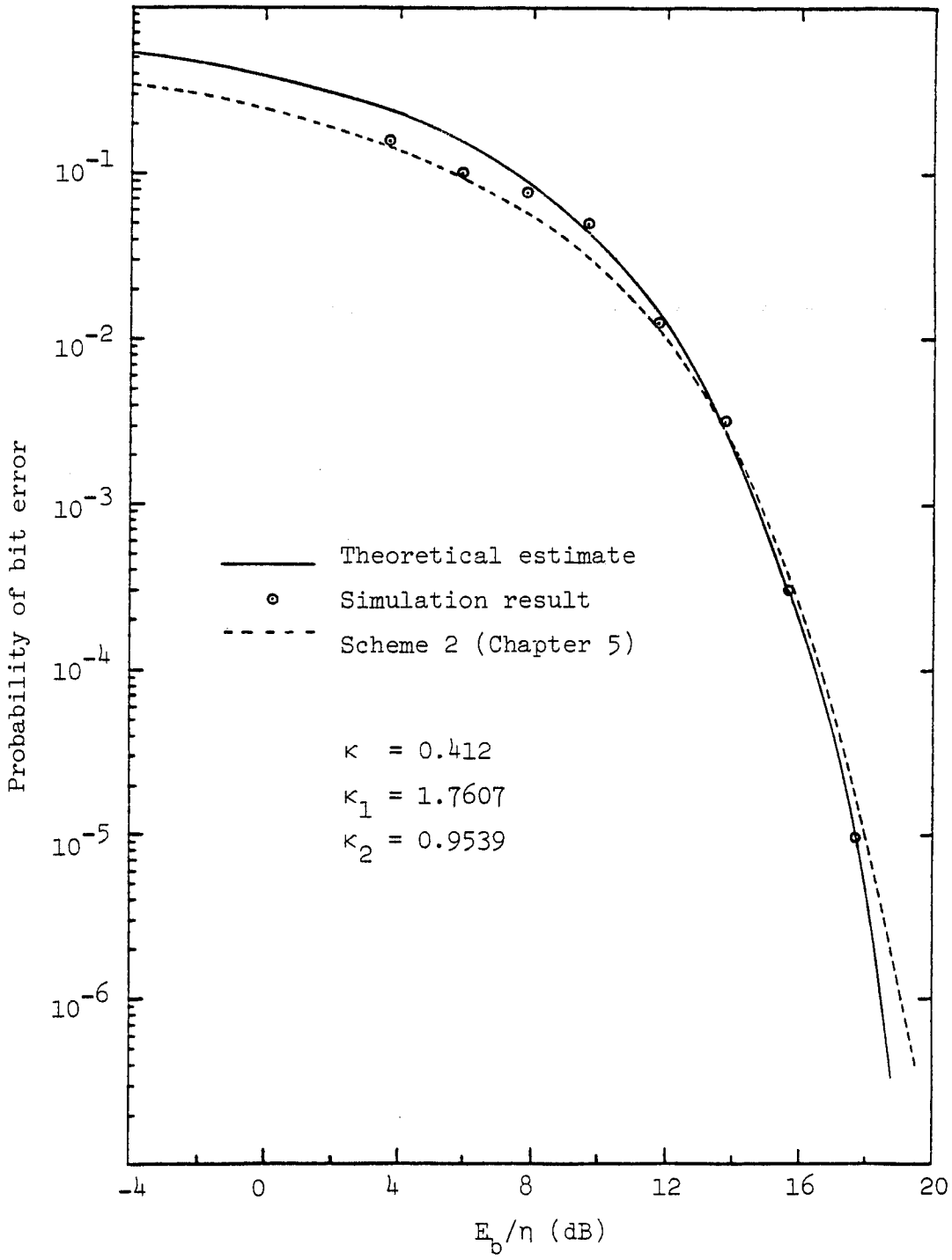


Fig. 6.3: Probability of bit error on AWGN channel for 8/2-DPASK with 22.5° phase shift between amplitude levels ("Scheme 3").

agreement between estimated and simulated results is quite good at high SNR's.

6.4 OPTIMIZING THE PARAMATERS κ, κ_1 AND κ_2.

The method of section 5.4 can, in theory, be used in the present case as well, but there is a practical problem. The presence of the single integral of Eq. 6.3.24 within the double integral of Eq. 6.3.5, makes the double integral rather expensive to evaluate (5 minutes of CPU time on an IBM 3033). As a result, the method of section 5.4 will require more than 10 hours of CPU time to find the optimum set of parameters at any one SNR. For our purpose, that is clearly out of the question.

It should also be kept in mind that Eq. 6.3.33 is merely an estimate of the actual probability of bit error. Especially at low SNR's, parameters derived from the estimate should be used with care.

The values of κ, κ_1 and κ_2 that were used for Fig. 4.3, was found simply by adjusting them by hand and making repeated runs of the simulation program, counting only 50 errors to determine the probability of error. This was done at a SNR of 16 dB. The adjustments were stopped when the first two digits of the probability of bit error stabilized.

Comparing "Scheme 3" (simulated) to "Scheme 2" (theoretical) shows that their performance is virtually the same, except for the two points at 8 and 10 dB. Since we do not know how close the adjustment of the three κ parameters came to the true optimum, it is possible that "Scheme 3" could be made to perform slightly better than "Scheme 2". It is the opinion of the author that the difference will be insignificant at best. In practical situations "Scheme 3" should be disqualified because of it's complexity.

6.5 PROBABILITY OF BIT ERROR ON THE SLOWLY RAYLEIGH FADING CHANNEL.

An estimate of the probability of bit error on the slowly Rayleigh fading channel was found using the method of section 4.6. $P_{be}(\rho)$ in Eq. 4.6.4 was replaced by the estimate of Eq. 6.3.33.

The results are shown in Fig. 6.4. The estimate of the probability of bit error is always higher than the results obtained by simulation. On the basis of the simulation there is again little difference between "Scheme 2" and "Scheme 3" and the comments of section 6.4 still apply.

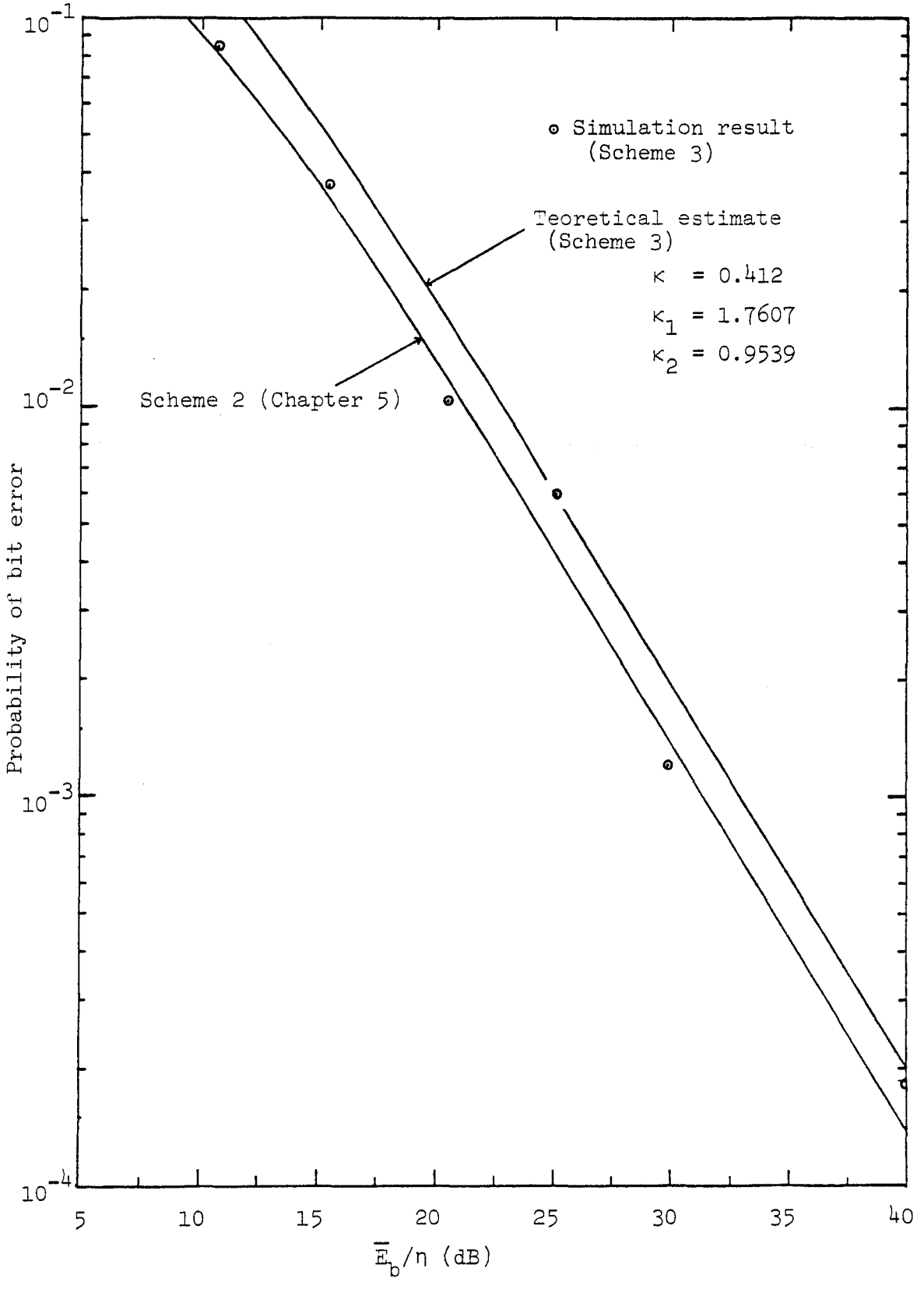


Fig. 6.4: Probability of bit error on slowly Rayleighy fading channel for 8/2-DPASK with 22.5° phase shift between amplitude levels ("Scheme 3")

7. THE UNMINIMIZABLE PROBABILITY OF ERROR DUE TO FADING.

7.1 INTRODUCTION.

The unminimizable probability of error due to fading is the probability of error at infinite average signal to noise ratio on a fading channel when the duration of the multipath is negligible with respect to the duration of a bit or symbol.

This effect was first mentioned in 1962 in a paper by Bello and Nelin [1]. It is not very well known, however. The author knows of only one communications textbook that even mentions it [34]. It is usually ignored by assuming that the channel is fading slowly - slow enough so that the change from bit to bit is negligible.

In the case of the three differential 16-ary modulation formats discussed in this dissertation, it was found by simulation that the unminimizable probability of error is an important limitation on the HF channel.

We did not attempt to extend the results in [1] to the 16-ary modulation formats. An approximate analysis (asymptotically accurate at low probability of error) can be done using equations in [10]. We present only simulation results. Chapter 3 was devoted to proving the validity of

the relevant aspects of the fading channel model used in the simulation.

The analysis in [1] is repeated in section 7.2, with some modifications and many specific assumptions. The author believes that this simplified analysis will give the reader a better intuitive understanding of the mechanism that produces the unminimizable probability of error.

7.2 DERIVATION OF THE UNMINIMIZABLE PROBABILITY OF ERROR DUE TO FADING FOR BINARARY DPSK.

It was already pointed out that the unminimizable probability of error was first derived in [1]. That analysis was general enough to include both FSK and DPSK, as well as diversity receivers and the presence of additive noise.

In this section some specific assumptions are made. Only DPSK is considered, with no diversity and no additive noise. The problem is also set up in a slightly different way, leading to somewhat different mathematical procedures. The form of the final result differs from the result in [1]. Computer evaluations of the two equations give the same numerical value, however.

Fig. 7.1 shows a diagram of the system. The data signal s(t)

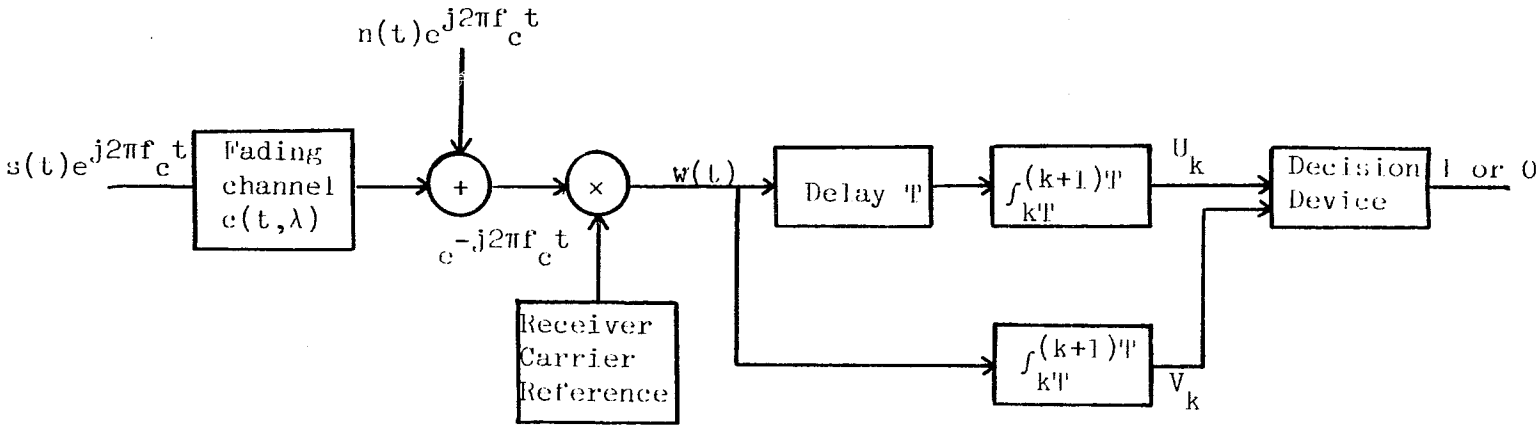


Fig. 7.1: Block diagram of system.

is modulated on a carrier of frequency f_c and

$$s(t) = \sum_k d_k(t-kT) \tag{7.2.1}$$

d_k is the complex envelope of the data signal and is given by

$$d_k(t) = \begin{cases} A \exp(j\theta_k) & 0 \leq t < T \\ 0; \text{ all other } t \end{cases} \tag{7.2.2a}$$

$$\theta_k = \begin{cases} 0 \\ \pi \end{cases} \tag{b}$$

The fading channel is assumed to have no time spread and no time delay. The channel impulse response is then given by

$$c(t,\lambda) = c(\lambda)\delta(t-\lambda) \tag{7.2.3}$$

where

$$c(\lambda) = c_i(\lambda) + jc_q(\lambda) \tag{7.2.4}$$

$c(\lambda)$ is a complex Gaussian random process and $c_i(\lambda)$ is uncorrelated with $c_q(\lambda)$. $|c(\lambda)|$ has a Rayleigh distribution.

The autocorrelation of $c_i(t)$ and $c_q(t)$ is given by [2,29]

$$R_{cc}(\tau) = (1/2)\exp(-\pi^2 B^2 \tau^2 / \ln 2) \tag{7.2.5}$$

B is the half-power bandwidth and the channel gain is assumed to be unity. $R_{yy}(\tau)$ is equivalent to $R_{cc}(\tau)$ in section 3.3. With the above assumptions, only $g_1(t)$ will have a value in Fig. 3.1. All the other tap-gains will be zero.

The transmitter and receiver carrier references are assumed to be derived from stable local oscillators. Therefore one should expect a small frequency offset Δf. We assume however, that the offset is very small, i.e. $T\Delta f << 1$. It is then only necessary to consider a slowly varying phase difference . But this is inconsequential since information is differentially encoded on the phase of the carrier. So θ is also not included in the system model.

With all the above assumptions, the received signal is given by (see Fig. 7.1):

$$w(t) = c(t)s(t) \tag{7.2.6}$$

Because we are dealing with DPSK, U_k is the reference for detection of V_k .

An error will occur if the channel introduces a relative phase shift of more than $\pi/2$ between U_k and V_k. Assuming for the moment that the same symbol is transmitted twice (i.e. $d_{k-1} = d_k$), an error will occur if, in Fig. 7.2, V_k is in the shaded region. The probability of error does not depend on the data sequence, so in the rest of this section it is assumed that

$$\theta_{k-1} = \theta_k = 0. \tag{7.2.7}$$

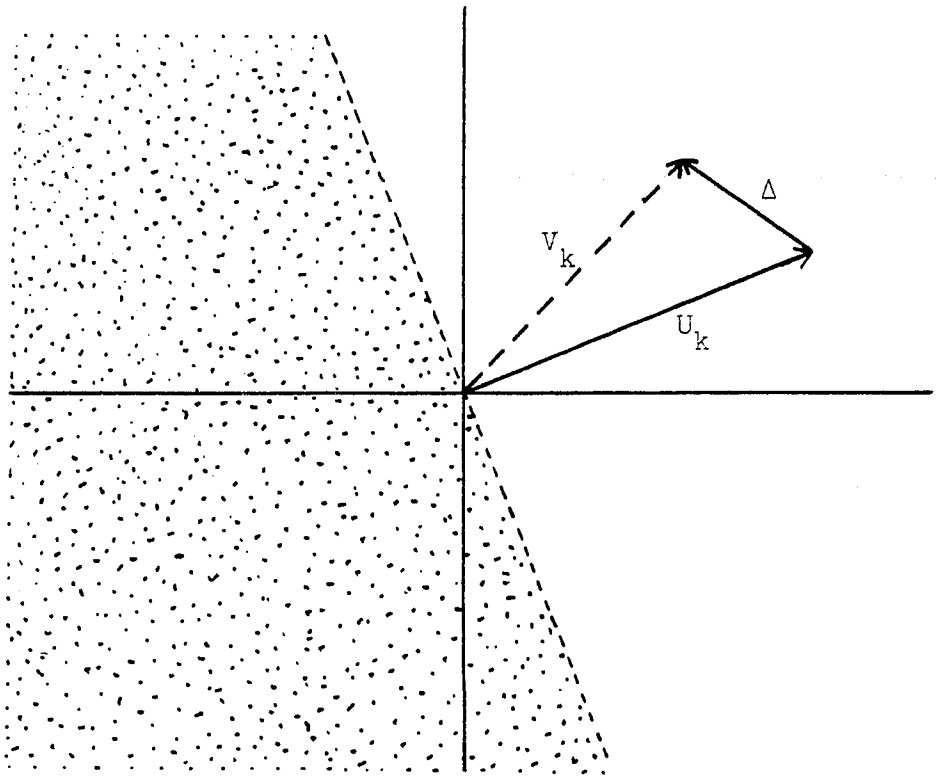


Fig. 7.2: U_k and V_k. Error if V_k in shaded region.

The next step is to find a joint probability density function for U_k and V_k in order to find the probability of error.

We introduce first two auxiliary variables:

$$\Delta c(t) = c(t) - c(t-T) \tag{7.2.8}$$

$$\Delta = V_k - U_k$$

$$= \int_{(k-1)T}^{kT} \Delta c(t)\, dt$$

$$= \Delta_i + j\Delta_q \tag{7.2.9}$$

also let

$$U_k = U_{ki} + jU_{kq}$$

$$V_k = V_{ki} + jV_{kq}$$

V_k, U_k and Δ are generated by linear operations on $c(t)$, therefore they are complex Gaussian random variables (CGRV's).

Inspecting Fig. 7.2, it is clear that the portion of Δ which is perpendicular to U_k, does not contribute to the probability of error, so in the analysis one has to consider only the part of Δ which is in phase with U_k. It is now possible to simplify the analysis by introducing a change of variables that will rotate U_k, V_k and Δ such that U_k is in the positive real direction.

Let

$$U'_k = U_k \exp(-j\angle U_k) \qquad (7.2.10a)$$

$$V'_k = V_k \exp(-j\angle U_k) \qquad (b)$$

$$\Delta' = \Delta \exp(-j\angle U_k) \qquad (c)$$

clearly

$$\mathrm{Re}(U'_k) = U'_{ki} = |U_k| \qquad (7.2.11a)$$

$$\mathrm{Im}(U'_k) = U'_{kq} = 0 \qquad (b)$$

$|U_k|$ has a Rayleigh probability density function:

$$f(|U_k|) = (|U_k|/\sigma_u^2)\exp(-|U_k|^2/2\sigma_u^2), \qquad |U_k| \geq 0 \qquad (7.2.12a)$$

where

$$\sigma_u^2 = \overline{U_{ki}^2} = \overline{U_{kq}^2} \qquad (b)$$

The value of σ_u^2 is derived in appendix A1.1

An error will occur if $\mathrm{Re}(V'_k) < 0$. We therefore need the probability density function of $\mathrm{Re}(V'_k)$ conditional on U'_{ki}, or $|U_k|$. This can be done by writing V'_k as:

$$V'_k = U'_k + \Delta' \qquad (7.2.13)$$

We need $f(V'_{ki} | U'_{ki})$, therefore U'_k is a deterministic number in Eq. 7.2.13. Δ' is a CGRV that is correlated with U'_k and we need to consider only the real parts of these numbers. The conditional density function is therefore Gaussian, with a non-zero expected value (from [15], Eq. 7-105):

$$f(V'_{ki}|U'_{ki}) = f(V'_{ki} \mid |U_k|)$$

$$=[1/\sqrt{2\pi\sigma_\Delta^2(1-r^2)}]\exp\{-[V'_{ki}-(1+r\sigma_\Delta/\sigma_u)|U_k|]^2/2\pi\sigma_\Delta^2(1-r^2)\} \quad (7.2.14a)$$

where

$$\sigma_\Delta^2 = \overline{\Delta_i^2} = \overline{\Delta_q^2} \quad (b)$$

and

$$r = \overline{\Delta_i U_{ki}}/\sigma_\Delta\sigma_u$$

$$= \overline{\Delta_q U_{kq}}/\sigma_\Delta\sigma_u \quad (c)$$

It should be pointed out that σ_Δ, σ_u and r are the same before and after the rotation of Eq. 7.2.10. That is because U_k, V_k and Δ have uncorrelated, equal-powered real and imaginary parts. σ_Δ^2 and r are derived in appendix A1.2.

The probability of error conditional on $|U_k|$ is given by (an error occurs if V'_{ki} is negative):

$$p(e||U_k|) = \int_{-\infty}^{0} f(V'_{ki}||U_k|)dV'_{ki} \quad (7.2.15)$$

The total probability of error is found by averaging Eq. 7.2.15 over $|U_k|$:

$$p(e) = \int_0^\infty f(|U_k|)\int_{-\infty}^{0} f(V'_{ki}||U_k|)dV'_{ki}d|U_k| \quad (7.2.16)$$

After substituting Eq. 7.2.12 and 7.2.14 into Eq. 7.2.16, the integrals are evaluated in appendix A3. The final result is:

$$p(e) = (1/2)[1 - (1 + r\sigma_\Delta/\sigma_u)/(\sigma_\Delta^2 + 2r\sigma_\Delta\sigma_u + \sigma_u^2)] \tag{7.2.17}$$

If

$$C = \pi^2 B^2 T^2/\ln 2 \tag{7.2.18}$$

then the parameters are given by (from appendix A1):

$$\sigma_u^2 = \overline{U_{ki}^2}$$
$$= A\{[(2T/B)\sqrt{\ln 2/\pi}\,\mathrm{erf}(\sqrt{2C})] - (\ln 2/\pi^2 B^2)[1-\exp(-C)]\} \tag{7.2.19}$$

$$\sigma_\Delta^2 = \overline{\Delta_i^2}$$
$$= A\{(4T/B)\sqrt{\ln 2/\pi}\,[2\mathrm{erf}(\sqrt{2C}) - \mathrm{erf}(2\sqrt{2C})] + (\ln 2/\pi^2 B^2)[-3 + 4\exp(-C) - \exp(-4C)]\} \tag{7.2.20}$$

$$r = (A^2/\sigma_\Delta\sigma_u)\{(2T/B)\sqrt{\ln 2/\pi}[\mathrm{erf}(2\sqrt{2C}) - \mathrm{erf}(\sqrt{2C})] + (\ln 2/2\pi^2 B^2)[1 - 2\exp(-C) - \exp(-4C)] - \sigma_u^2/A^2\} \tag{7.2.21}$$

where

$$\mathrm{erf}(z) = (1/\sqrt{2\pi})\int_0^z \exp(-x^2/2)\,dx \tag{7.2.22}$$

Eq. 7.2.17 was evaluated on a digital computer, as was the equivalent result from [1], which can be found in appendix A4. The two equations gave identical numerical values and were plotted in Fig. 3.8.

7.3 COMPARISON BY SIMULATION OF DIFFERENT SCHEMES WITH RESPECT TO THE UNMINIMIZABLE PROBABILITY OF ERROR.

Fig.7.3 shows simulation results for the unminimizable probability of bit error for DPSK, 16-DPSK, 8/2 DPASK (Scheme 2) and 8/2-DPASK (Scheme 3). (The different schemes are described in section 1.2).

First, a note about the simulation results for DPSK. The observant reader will notice the results in Fig. 7.3 are closer to the theoretical curve than in Fig. 3.8. That has to do with the sampling rate that was used in the simulation. The experiment of Fig. 3.8 used only 2 samples per bit, while 10 samples per bit were used in Fig. 7.3. The discrepancy between simulation and theory in Fig. 3.8, thus seems to be due to the distortion of the bilinear transformation at low sampling rates, rather than due to the fact that the autocorrelations of the tap-gains are not truly Gaussian in the continuous time domain.

Of the three 16-ary scheme, 8/2-DPASK (Scheme 1) can be seen to perform the best, while 16-DPSK is the worst. There is a severe penalty in the probability of bit error for using the 16-ary scheme. At low fading rates 8/2-DPASK is worse than DPSK by a factor of 120. This comparison is not quite fair, however, because theoretically the probability of error depends on the duration of a symbol, which is 4 times longer

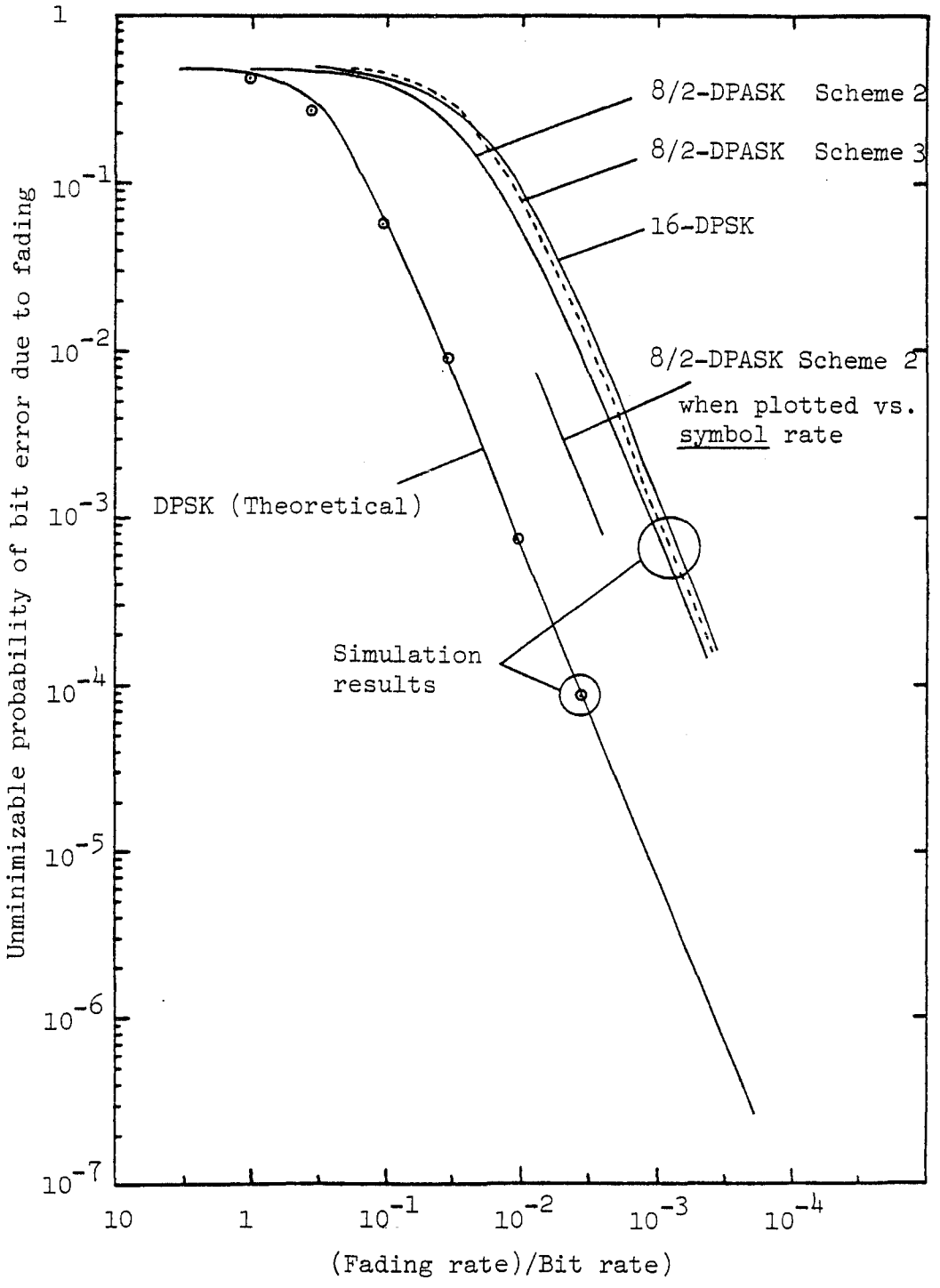


Fig. 7.3: Simulation results for the unminimizable probability of bit error due to fading as a function of the ratio between channel fading rate and the bit rate.

for the 16-ary scheme than for DPSK. Correction for this factor of four on the abscissa of Fig. 7.3 will move the 16-ary curves to the left. Then the penalty is reduced to a factor of 24 in the bit error rate.

From Fig. 7.3 it seems that the bit rate should be made as high as possible. But in chapter 8 we will show that the bit rate should be as low as possible to avoid the detrimental effects of multipath. There will therefore be an optimum bit rate when both effects are present. Simulation results to illustrate this point will be presented in chapter 8.

8. THE EFFECT OF MULTIPATH.

8.1 PROBABILITY OF ERROR AS A FUNCTION OF CHANNEL TIME SPREAD.

This chapter presents only simulation results.

Fig. 8.1 shows the probability of bit error as a function of the channel time spread, when the channel is fading slowly and when the SNR is 27dB.

On the left-hand side, the curves approach a constant value of $P(e) = 1/2$ when the time spread becomes larger than the symbol duration (not the bit duration). On the right-hand side, the curves approach a constant probability of error when the intersymbol interference due to the multipath becomes negligible. The constant values can be found from the appropriate figures in chapters 4, 5 and 6.

8/2-DPASK (Scheme 2) has the lowest probability of error of the three 16-ary schemes. From the curves it would appear that the three 16-ary schemes erform better than DPSK at some values of time spread. If the comparison is made on the basis of symbol duration rather than bit duration, the 16-ary schemes are always worse than DPSK. The location of the curve for 8/2-DPASK (Scheme 2) when plotted vs. symbol duration, is indicated on Fig. 8.1.

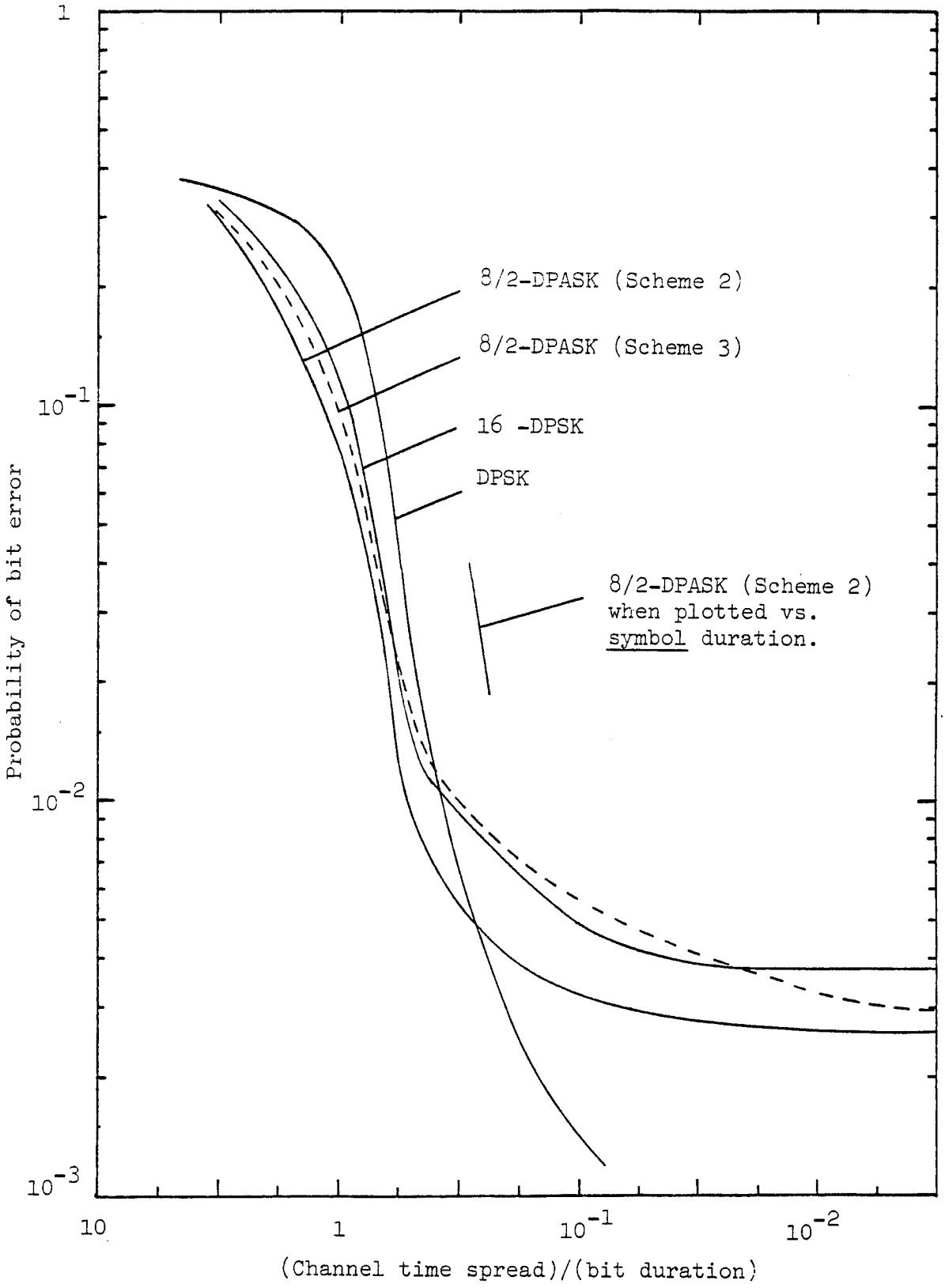


Fig. 8.1: Probability of bit error as a function of the ratio between the multipath time spread and the bit duration, at a SNR of 27 dB. Simulation results only.

The K parameters of the two DPASK schemes were as indicated in Fig. 5.6 and 6.3.

8.2 THE OPTIMUM BIT RATE WHEN BOTH FAST FADING AND MULTIPATH IS PRESENT.

Fig. 8.1 shows that the bit duration should be as long as possible to avoid the effect of multipath. But Fig. 7.3 shows the opposite, that the bit rate should be as high as possible to avoid the effect of fast fading.

Since the HF channel has fast fading and multipath time spread it is important to choose a bit rate that will minimize the effect of both. Fig. 8.2 shows simulation results for DPSK and 8/2-DPASK (Scheme 2) when the SNR is 27dB, the time spread is 1 ms and the fading rate is 1 Hz. (By fading rate we mean the value of B in Hz in Eq. 7.2.5) These values are fairly typical of the HF channel.

It can be seen that a well-defined minimum exists at 100 bps for DPSK, and at 400 bps or 100 baud (symbols per second) for 8/2-DPASK. The best probability of error for DPSK is 2.4×10^{-3}. 8/2-DPASK performs much worse with a best probability of error of 1.2×10^{-2}. By comparison, the probability of error in the absence of time spread and fast fading at a SNR of 27 dB are 1.0×10^{-3} and 2.5×10^{-3}, respectively. (From Fig. 5.8)

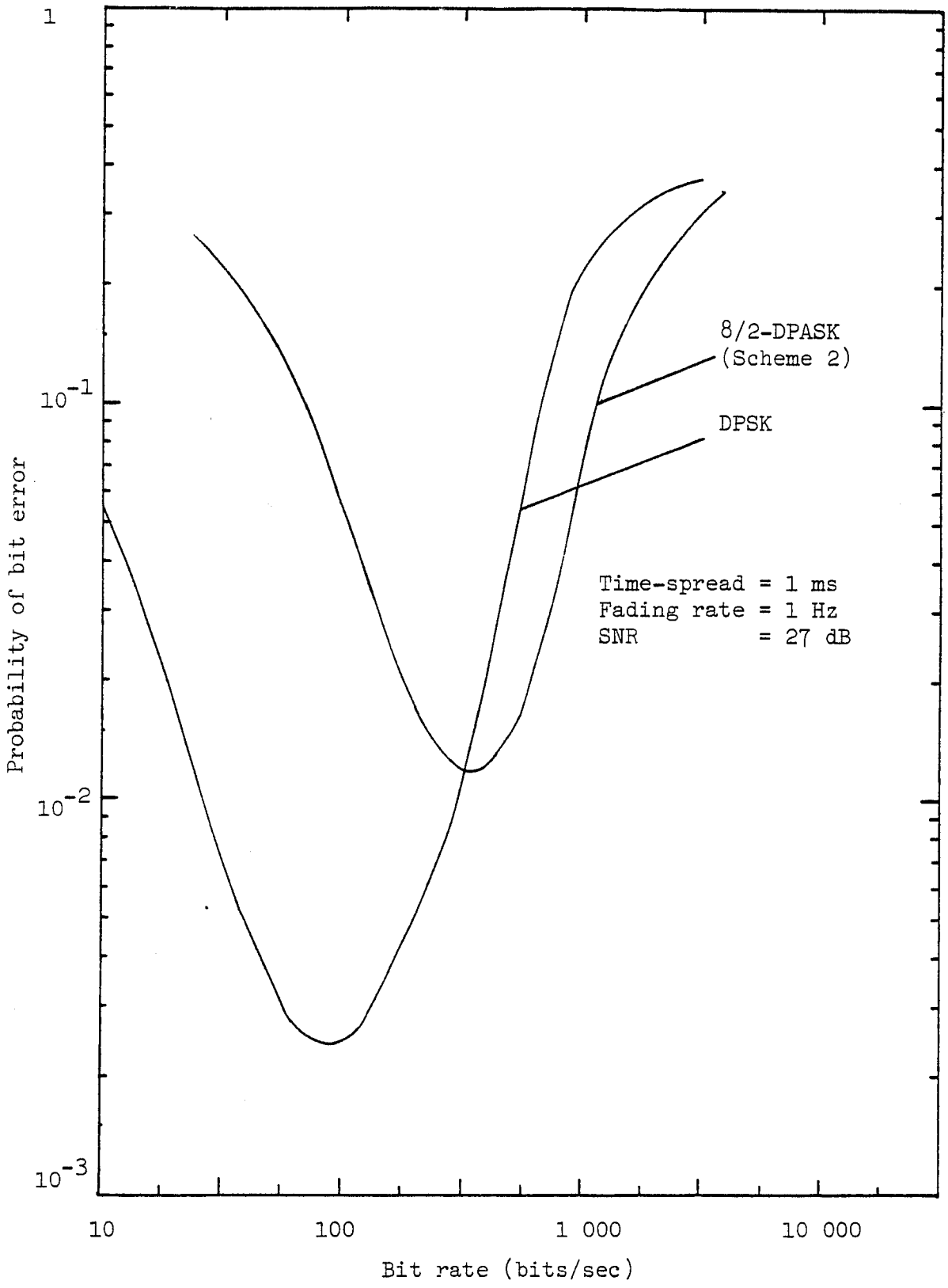


Fig. 8.2: Simulation results illustrating the combined effect of fast fading, multipath time spread and additive noise on the probability of bit error for DPSK and 8/2-DPASK.

To summarize we note that in this experiment the joint effect of time spread, fast fading and additive noise degrades the probability of bit error of 8/2-DPASK by a factor of 5 with respect to that of DPSK. When only additive noise is present, the degradation is only a factor of 2.5.

One can draw the following conclusion from this chapter: Under a typical set of conditions on the HF channel, when 16-ary 8/2-DPASK is used instead of binary DPSK, the price for a four-fold increase in the bit rate, is an increase by a factor of four of the transmitter power (to keep the energy per bit constant) and an increase by a factor of five of the probability of bit error.

9. OTHER MODULATION FORMATS

9.1 INTRODUCTION

This chapter presents a brief discussion of a few additional modulation formats that will occupy approximately the same bandwidth (at equal bit rates) as the three 16-ary schemes discussed so far. These additional schemes were found to be either impractical or completely unworkable in the HF environment.

9.2 ARGUMENT AGAINST USING MORE THAN TWO AMPLITUDE LEVELS

The following observation is obvious, but very important when contemplating differential amplitude modulation:

> The phase of a phasor is naturally measured modulo 2π, but amplitude does not have a modulus.

The practical consequence of this is that differential phase modulation works because the phase can be increased or decreased indefinitely without any problems, but when differential amplitude modulation calls for an indefinite increase of the amplitude, amplifiers at the transmitter will soon saturate. The signal-to-noise ratio will also change. That is not the way that communication systems are usually operated. Normally one requires either that the transmitter operates at the maximum power available, or at

some specific level for the conservation of energy. The transmitter power is not allowed to wander around of it's own accord. But that is what will happen if more than two amplitude levels are used in a differential modulation format.

We illustrate this point by considering a scheme with three amplitude levels, as in Fig. 9.1. It is not suitable for the normal binary representation of data, but it simplifies the argument and can easily be extended to four or eight amplitude levels.

With these three amplitudes, an alphabet containing three symbols can be sent, say S_1, S_2 and S_3. The previous amplitude is the reference, say it is called A . The differential encoding rule is the following:

To send S_1: Reduce the amplitude.

$$A = k_1 A_r \qquad (k_1 < 1)$$

To send S_2: Send the same amplitude.

$$A = A_r$$

To send S_3: Increase the amplitude.

$$A = k_2 A_r \qquad (k_2 > 1)$$

The problem arises when the information to be transmitted calls for for repeatedly sending only S_1 or S_3 . The

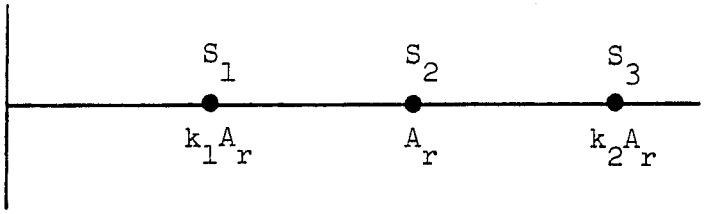


Fig. 9.1: Alphabet containing three symbols, represented by three amplitude levels.

amplitude grows when S_3 is sent. If it is sent repeatedly, the power amplifier at the transmitter will soon saturate. Likewise, if S_1 is sent repeatedly the signal will soon become vanishingly small in comparison with the noise at the receiver.

Based on this discussion, no modulation format requiring more than two amplitude levels was considered.

9.3 ARGUMENT AGAINST USING OFFSET KEYING

Offset keying refers to quadrature modulation formats where the data on the in-phase and quadrature-phase carriers are changed respectively on odd and even multiples of half of the symbol period, instead of simultaneously at the beginning of each symbol period as is normally done. In order to decode the data at the receiver, it is therefore necessary to identify the in-phase and the quadrature carriers. But it is exactly to avoid that requirement that we set out investigating differential signaling schemes in the first place.

On the AWGN channel it is possible to acquire carrier synchronization and to resolve at the beginning of a transmission the inevitable phase ambiguities of tracking loops for M ary phase-modulation formats. If the SNR is high enough the phase can then be tracked accurately and with a

low probability of cycle slipping, so that no new ambiguities will be introduced.

On a slowly fading channel, it is still possible to track the phase, but after every fade-out there will be a phase ambiguity to resolve, and since the transmitter does not know when that happens, the receiver is left to it's own resources. It can resolve the ambiguity by observing the transitions of the data, but in the mean time the probability of error will be 1/2.

The HF channel, moreover, can often not be treated as a slowly fading channel. In that case the probability of error will be increased by imperfect tracking of the carrier phase.

Because of the problems of phase ambiguity in the tracking loops and imperfect tracking of the phase on the fast-fading channel, modulation formats involving offset keying are not considered.

9.4 QASK, QPR AND MSK

Quadrature amplitude shift keying, quadrature partial response keying and minimum shift keying are three well known modulation formats that are not suitable for use on

fast-fading channels.

Inspection of Fig. 9.2 shows that QASK requires three amplitude levels. Therefore QASK is disqualified on the basis of the discussion in section 9.2.

Fig. 9.3 shows the signal constellation for QPR. The correlative coding [19] at the transmitter is done separately on the in-phase and the quadrature-phase carriers, and likewise for the decoding at the receiver. Therefore the receiver has the problem of accurately identifying the two components of the carrier, and the discussion on ambiguity in section 9.3 applies. The data is differentially encoded on each of the two quadrature carriers, so there is protection against a 180° phase slip, but not against a 90° slip. A further problem has to do with the presence of the point at the origin of the signal constellation. If no power is transmitted, then there is no amplitude reference for the following symbol. The receiver then has to use the symbol before the zero as the reference for the symbol after the zero. That is a severe limitation on a fast-fading channel, especially when several zero's occur in a row. Therefore QPR is disqualified both by the problem of phase ambiguity and by the problem of the zero amplitude level.

MSK is a special form of shaped offset QPSK, so it is

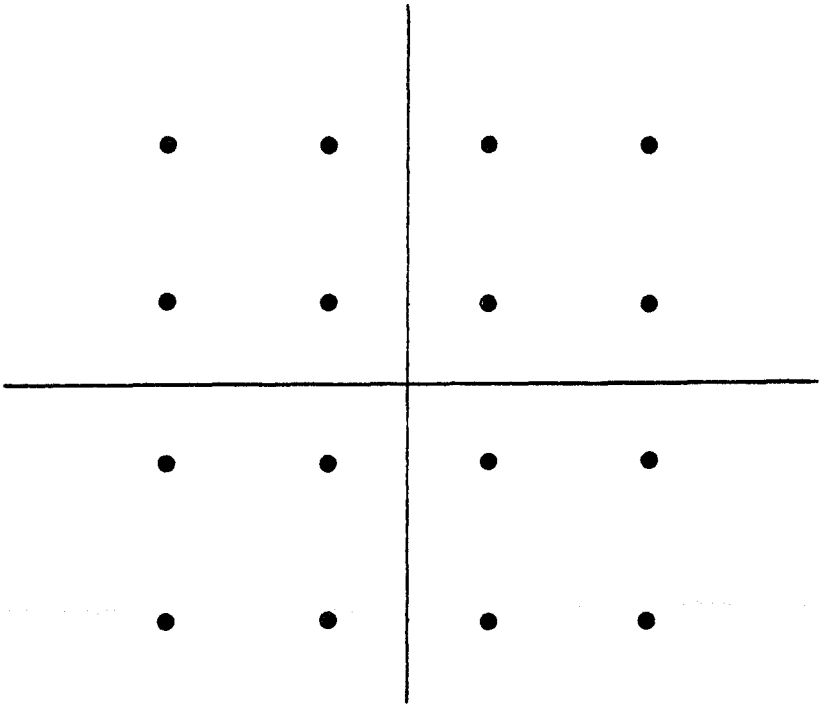


Fig. 9.2: Signal constellation for 16-ary QASK.

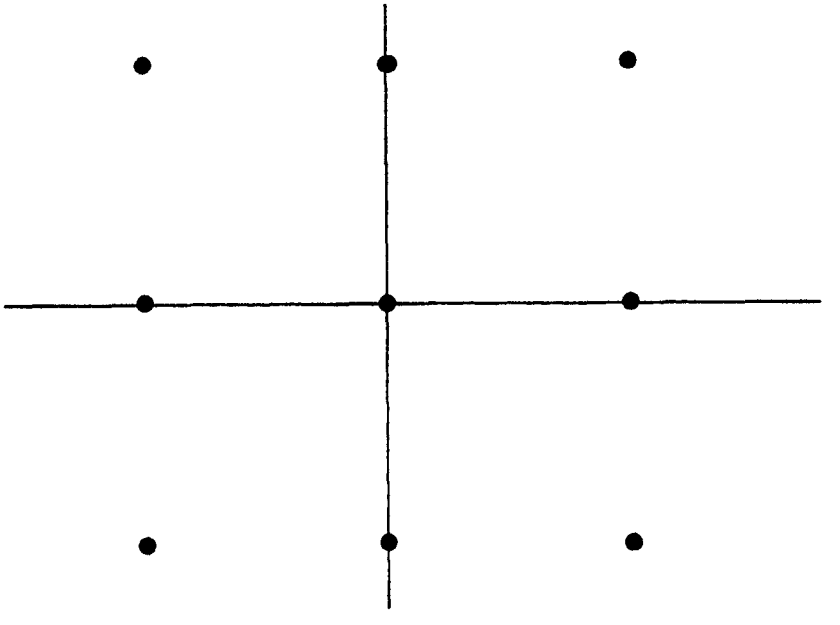


Fig. 9.3: Signal constellation for QPR.

disqualified on the basis of the discussion in section 9.3. But the idea of using pulse shaping should be kept in mind. The three 16-ary schemes that we have been discussing all along have the same sin(x)/x spectrum shape of PSK, which contains too much out-of-band energy for many applications. That can be improved by using pulse shaping.

10. CONCLUSIONS.

Three differential 16-ary digital signaling formats were compared to each other and to binary DPSK. Some comparisons were done theoretically. The theoretical work was then checked against a digital computer simulation of the communication systems involved [4,21]. A considerable amount of time was spent to modify the simulation to be able to do these simulated experiments

The theoretical expressions were usually not in a closed form, and extensive computer programming was necessary to evaluate them. In some cases only simulation results were presented.

On the slowly Rayleigh fading channel, differential phase-amplitude modulation with 8 phases and 2 amplitude levels (8/2-DPASK) was found to be 1.6 dB better than regular 16-ary DPSK. It was 1.7dB worse than 8-ary DPSK and 4.0 dB worse than binary DPSK. When an additional phase shift of 22.5^{o} was introduced between the two amplitude levels of 8/2-DPASK, the results were slightly worse.

On the fast Rayleigh fading channel, comparisons were made on the basis of the unminimizable error rate due to fading. 8/2-DPASK was found to be able to tolerate a fading rate of 1.3 times faster than 16-DPSK, but DPSK, by comparison could

tolerate a fading rate of 4.3 times faster than 8/2-DPASK. The addition of the 22.5^{o} phase shift again made matters slightly worse. These comparisons are on the basis of equal symbol rate and equal probability of bit error.

When multipath time spread occurs on the slowly Rayleigh fading channel, 8/2-DPASK can again tolerate more time spread than the other two 16-ary schemes. On the basis of equal bit rate it can tolerate more time spread than binary DPSK, but on the basis of equal symbol rate, it is worse.

The final comparison was for binary DPSK and 8/2-DPASK on a channel with typical HF parameters, i.e. a fading rate of 1 Hz, a time spread of 1 ms, and a signal-to- noise ratio of 27 dB. It was found that the best bit rate to use was 100 bps for binary DPSK and 400 bps (i.e. 100 symbols per second) for 8/2-DPASK. The resulting probabilities of error was 2.4×10^{-3} for binary DPSK and 1.2×10^{-2} for 8/2-DPASK, so that the price for a four-fold increase in bit rate, is a five-fold increase in the probability of bit error.

Finally, chapter 3 presented a method of generating the required autocorrelation function for the random processes that are used for the tap gains in the transversal filter that simulates the HF channel. This was necessary in order to perform the simulated experiments in the other chapters.

APPENDIX A1

A1.1 DERIVATION OF σ_u^2

This is the derivation of σ_u^2 as it appears in Eq. 7.2.12.

$$\sigma_u^2 = \overline{U_{ki}^2} = \overline{U_{kq}^2} \tag{A1.1}$$

We make use of a property of linear systems ([15] p. 346).

Fig. A1.1 shows the integrate-and-dump matched filter that is used to generate U_{ki} (or U_{kq}).

The input to the filter is $c_i(t)$ which has an autocorrelation function $R_{cc}(\tau)$ (Eq. 7.2.5). U_{ki} is the output of the filter sampled at time $t = (k+1)T$. Since $c_i(t)$ is a stationary random process it does not really matter when the integration starts and stops, as long as the integral is performed over a time T. Therefore we can write the impulse responds of the integrate-and-dump filter without reference to any particular starting time:

$$h(t) = u(t) - u(t-T) \tag{A1.2}$$

where u(t) is the familiar step function. Writing the impulse response in this form makes the output due to $c_i(t)$, which we call $u_i(t)$, a stationary process. This is not true for a regular integrate-and-dump filter. Usually only the sampled process U_k is stationary.

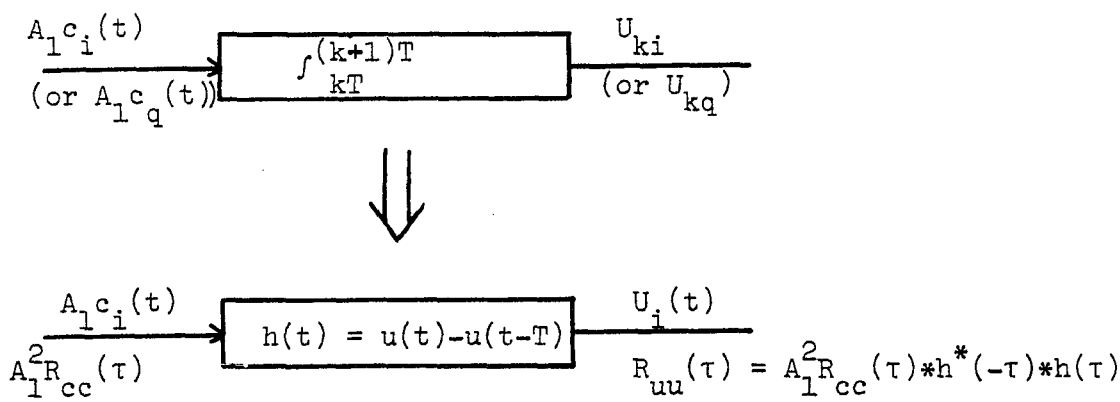


Fig. A1.1: The integrate-and -dump filter and relation between autocorrelation functions.

The autocorrelation of u (t) is given by

$$R_{uu}(\tau) = A_1^2 R_{cc}(\tau) * h^*(-\tau) * h(\tau) \tag{A1.3}$$

and

$$\sigma_u^2 = R_{uu}(0) \tag{A1.4}$$

With h(t) from Eq. A1.2 and referring to Fig. A1.2 let

$$g(\tau) = h^*(-\tau) * h(\tau)$$

Then

$$g(\tau) = \int_{-\infty}^{\infty} [u(\lambda) - u(\lambda-T)]\ [u(\lambda-T) - u(\lambda-T-\tau)]\, d\lambda$$

$$= \begin{cases} \int_0^{T+\tau} 1\, d\lambda = T+\tau & -T < \tau < 0 \\ \int_\tau^T 1\, d\lambda = T-\tau & 0 < \tau < T \\ 0 & \text{elsewhere} \end{cases}$$

$$= [T-|\tau|][u(\tau+T) - u(\tau-T)] \tag{A1.5}$$

Now

$$R_{uu}(\tau) = A_1^2 R_{cc}(\tau)\ g(\tau)$$

and

$$R_{uu}(0) = A_1^2 \int_0^{\infty} R_{cc}(\lambda) g(-\lambda) d\lambda$$

$$= 2A_1^2 \int_0^{\infty} R_{cc}(\lambda) g(\lambda) d\lambda$$

$$= 2A_1^2 \int_0^{T} R_{cc}(\lambda) (T-\lambda) d\lambda$$

$$= A_1^2 \{ (2T/B) \sqrt{\ln 2/\pi}\ \mathrm{erf}(\sqrt{2/\ln 2}\,\pi BT)$$

$$-(\ln 2/\pi^2 B^2)[1-\exp(-\pi^2 B^2 T^2/\ln 2)]\} \tag{A1.6}$$

The second step is possible because the functions are even.

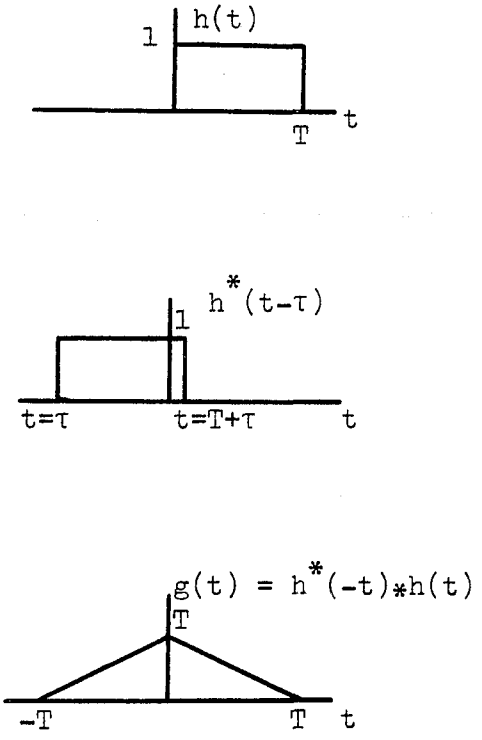


Fig. A1.2: Convolution for Eq. A1.3.

Also

$$\mathrm{erf}(z) = (1/\sqrt{2\pi}) \int_0^z \exp(-x^2/2)\,dx$$

This concludes the derivation of σ_u^2.

A1.2 DERIVATION OF σ_Δ^2 AND OF r

This is the derivation of σ_Δ^2 and r as they appear in Eq. 3.2.17.

Proceeding as in section A1.1 and referring to Fig. A1.3, the impulse response of the filter that would be used to produce $\Delta(t)$ is given by

$$h_3(t) = u(t) - 2u(t-T) + u(t-2T) \qquad \text{(A1.7)}$$

and

$$\sigma_\Delta^2 = R_{\Delta\Delta}(0) \qquad \text{(A1.8)}$$

where

$$R_{\Delta\Delta}(\tau) = R_{cc}(\tau) * h_3^*(-\tau) * h_3(\tau) \qquad \text{(A1.9)}$$

Referring to Fig. A1.4 let

$$g_3(\tau) = h_3^*(-\tau) * h_3(\tau) \qquad \text{(A1.10)}$$

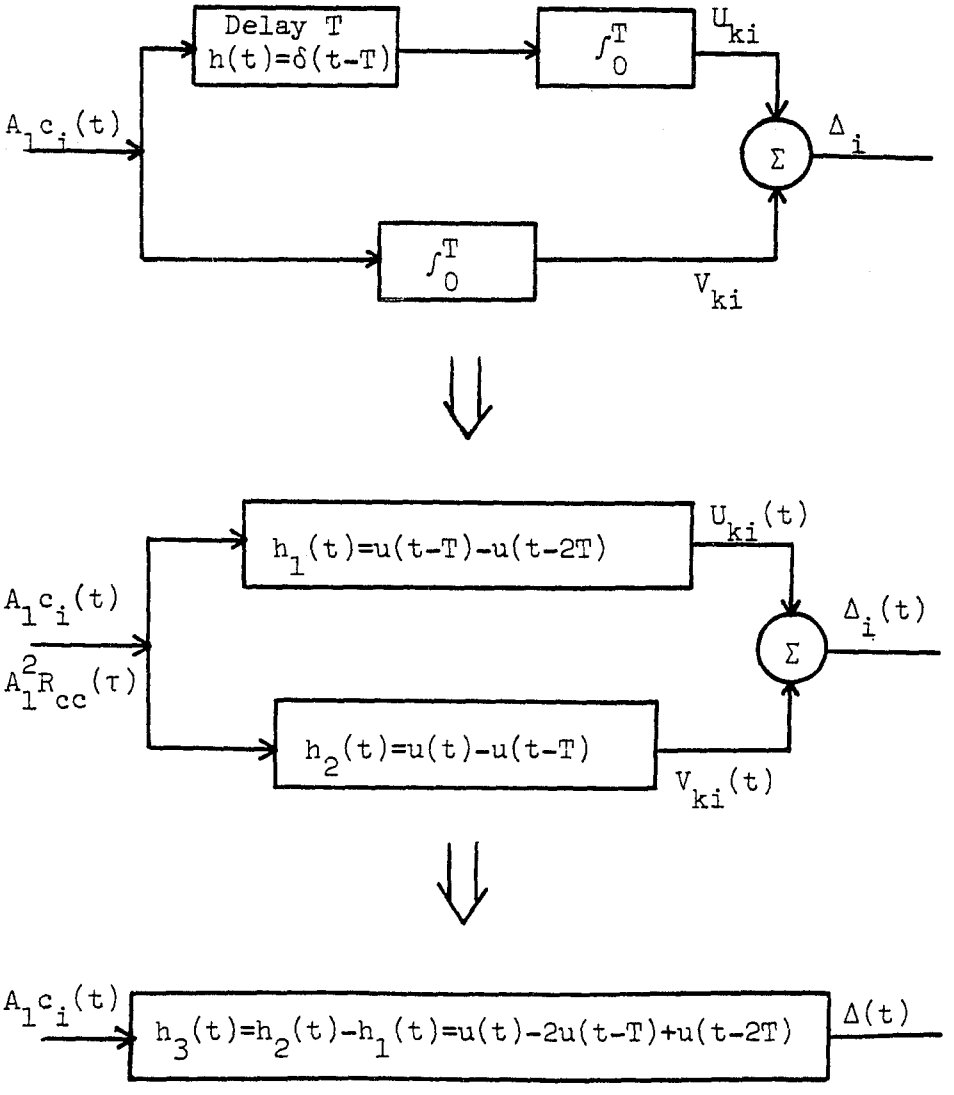


Fig. A1.3: Filter producing $\Delta(t)$.

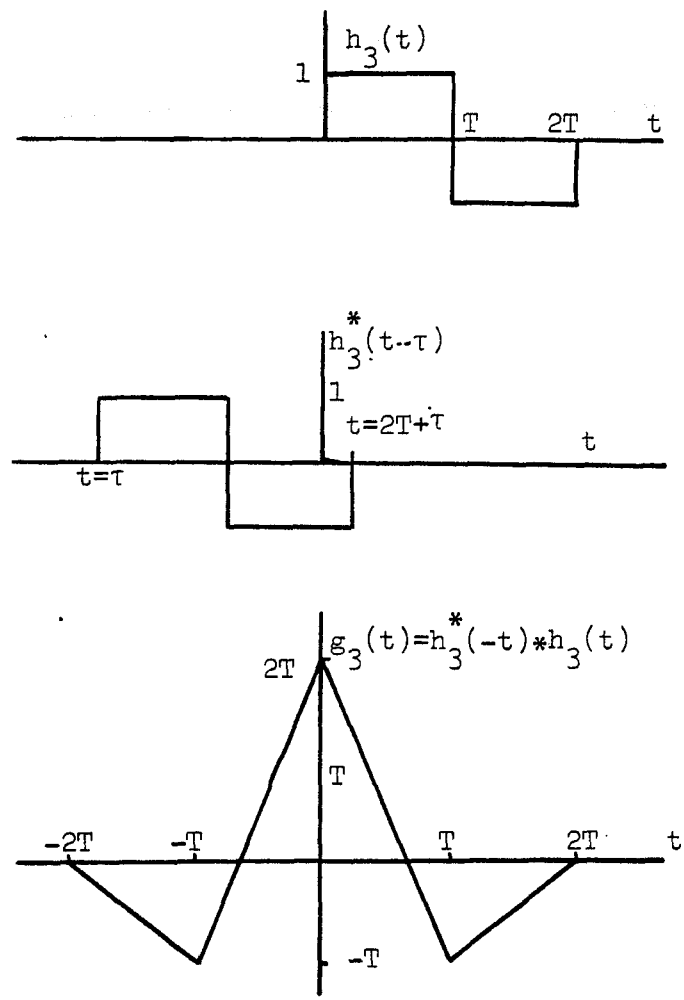


Fig. A1.4: Convolution for Eq. A1.10

Now

$$R_{\Delta\Delta}(0) = A_1^2 \int_{\infty}^{\infty} R_{cc}(\lambda) g_3(-\lambda) d\lambda$$

$$= 2A_1^2 \int_0^{\infty} R_{cc}(\lambda) g_3(\lambda) d\lambda$$

$$= 2A_1^2 \left[\int_0^T (2T-3\lambda) R_{cc}(\lambda) d\lambda + \int_T^{2T} (\lambda-2T) R_{cc}(\lambda) d\lambda\right]$$

$$= A_1^2 \{(4T/B)\sqrt{\ln 2/\pi}\ \mathrm{erf}(\sqrt{2/\ln 2}\pi BT)$$
$$- (3\ln 2/\pi^2 B^2)[1 - \exp(-\pi^2 B^2 T^2/\ln 2)]$$
$$- 4(T/B)\sqrt{\ln 2/\pi}[\mathrm{erf}(2\sqrt{2/\ln 2}\pi BT) - \mathrm{erf}(\sqrt{2/\ln 2}\pi BT)]$$
$$+ (\ln 2/\pi^2 B^2)[\exp(-\pi^2 B^2 T^2/\ln 2) - \exp(-4\pi^2 B^2 T^2/\ln 2)]\}$$

$$= A_1^2 \{(4T/B)\sqrt{\ln 2/\pi}[2\mathrm{erf}(\sqrt{2/\ln 2}\pi BT) - \mathrm{erf}(2\sqrt{2/\ln 2}\pi BT)]$$
$$+ (\ln 2/\pi^2 B^2)[-3 + 4\exp(-\pi^2 B^2 T^2/\ln 2)$$
$$- \exp(-4\pi^2 B^2 T^2/\ln 2)]\} \qquad \text{(A1.11)}$$

This concludes the derivation of σ_Δ^2 .

We now proceed to derive r. From Eq. 7.2.14c:

$$r = \overline{\Delta_i U_{ki}}/\sigma_\Delta \sigma_u \qquad \text{(A1.12)}$$

Now

$$\overline{\Delta_i U_{ki}} = \overline{(V_{ki} - U_{ki})U_{ki}} = R_{uv}(0) - R_{uu}(0)$$
$$= R_{uv}(0) - \sigma_u^2 \qquad \text{(A1.13)}$$

Referring to Fig. A1.3 (see also Fig. A1.1):

$$R_{uv}(\tau) = A_1^2 R_{cc}(\tau) * h_1(\tau) * h_2^*(-\tau) \qquad \text{(A1.14)}$$

(From [15] p. 353)

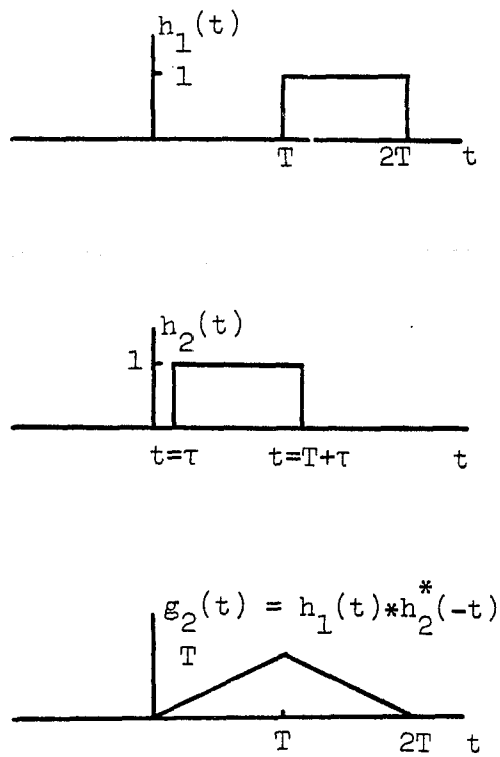


Fig. A1.5: Convolution for Eq. A1.15.

Now

$$R_{uv}(0) = A_1^2 \int_{-\infty}^{\infty} R_{cc}(\lambda) g_2(-\lambda)\, d\lambda$$

$$= A_1^2 \int_0^{\infty} R_{cc}(\lambda) g_2(\lambda)\, d\lambda$$

$$= A_1^2 \int_0^{T} \lambda R_{cc}(\lambda)\, d\lambda + A_1^2 \int_T^{2T} (2T-\lambda) R_{cc}(\lambda)\, d\lambda$$

$$= A_1^2 \{ (\ln 2/2\pi^2 B^2)[\, 1 - \exp(-\pi^2 B^2 T^2/\ln 2)]$$
$$+ (2T/B)\sqrt{\ln 2/\pi}[\operatorname{erf}(2\sqrt{2/\ln 2}\,\pi BT) - \operatorname{erf}(\sqrt{2/\ln 2}\,\pi BT)]$$
$$- (\ln 2/2\pi^2 B^2)[\exp(-\pi^2 B^2 T^2/\ln 2) - \exp(-4\pi^2 B^2 T^2/\ln 2)]\}$$

$$= A_1^2 \{ (\ln 2/2\pi^2 B^2)[\, 1 - 2\exp(-\pi^2 B^2 T^2/\ln 2)$$
$$+ \exp(-4\pi^2 B^2 T^2/\ln 2)]$$
$$+ (2T/B)\sqrt{\ln 2/\pi}[\operatorname{erf}(2\sqrt{2/\ln 2}\,\pi BT) - \operatorname{erf}(\sqrt{2/\ln 2}\,\pi BT)]\} \qquad \text{(A1.15)}$$

Finally

$$r = (R_{uv}(0) - \sigma_u^2)\,/\,\sigma_\Delta \sigma_u \qquad \text{(A1.16)}$$

This concludes the derivation of r.

APPENDIX A2

ON GENERATING A SET OF RANDOM VARIABLES WITH A GIVEN COVARIANCE MATRIX.

This appendix refers to the procedure for generating three random variables with a given covariance matrix as described in section 3.6.

That the generating procedure is valid, can be proved by writing x_2 and x_3 in terms of x_1 and the independent variables n_2 and n_3 and then to calculate each element of the resulting covariance matrix, which should then be equal to the desired covariance matrix. We proceed now, by way of example, to prove just that for two elements of the matrix, R_{22} and R_{31}.

$$\overline{x_2 x_2} = [(R_{12}/R_{11})x_1 + n_2]^2$$

$$= (R_{12}/R_{11})^2 \overline{x_1^2} + 2(R_{12}/R_{11})\overline{x_1 n_2} + \overline{n_2^2}$$

$$= (R_{12}/R_{11})^2 R_{11} + 0 + (R_{22} - R_{12}^2/R_{11})$$

$$= R_{22}$$

$$x_3 = b_1 x_1 + b_2(R_{12}/R_{11}x_1 + n_2) + n_3$$

$$\overline{x_3 x_1} = b_1 \overline{x_1^2} + b_2(R_{12}/R_{11})\overline{x_1^2} + b_2\overline{x_1 n_2} + \overline{x_1 n_3}$$

$$= R_{11}(R_{31}R_{22} - R_{12}R_{32})/(R_{11}R_{22} - R_{12}R_{21})$$

$$+ R_{11}(R_{12}/R_{11})(R_{11}R_{32} - R_{31}R_{21})/(R_{11}R_{22} - R_{12}R_{21})$$

$$+ 0 + 0$$

$$= R_{31}$$

APPENDIX A3

DERIVATION OF EQ. 7.2.17

This is the evaluation of the double integral in Eq. 7.2.16. Let

$$x = |U_k| \tag{A3.1a}$$

$$y = V'_{ki} \tag{b}$$

$$a = 1/2\sigma_u^2 \tag{c}$$

$$b = (1 + r\sigma_\Delta/\sigma_u)/\sqrt{2\sigma_\Delta^2 (1-r^2)} \tag{d}$$

$$c = 1/\sqrt{2\sigma_\Delta^2 (1-r^2)} \tag{e}$$

Substitute Eq. A3.1, 7.2.12(a) and 7.2.14 into Eq. 7.2.16. Then

$$p(e) = \int_0^\infty 2ax \exp(-ax^2) \int_{-\infty}^{\infty} (c/\sqrt{\pi}) \exp[-(cy-bx)^2] dy dx$$

$$= (2ac/\sqrt{\pi}) \int_0^\infty x \exp[-(a+b^2)x^2] \int_{-\infty}^{0} \exp(-c^2y^2 + 2cbxy) dy dx \tag{A3.2}$$

The inner integral can be evaluated using a table of integrals ([9], Eq. 3.322(2)):

$$p(e) = (2ac/\sqrt{\pi}) \int_0^\infty x \exp[-(a+b^2)x^2] \{\sqrt{\pi/4c^2} \exp(b^2x^2) [1-\Phi(bx)]\} dx$$

$$= a \int_0^\infty x \exp(-ax^2) [1-\Phi(bx)] dx \tag{A3.3}$$

where

$$\Phi(x) = (2/\sqrt{\pi}) \int_0^x \exp(-t^2) dt \tag{A3.4}$$

Eq. A3.3 can be evaluated in two parts:

$$I_1 = \int_0^\infty x \exp(-ax^2)\,dx$$

$$= (-1/2a)\exp(-ax^2)\,\Big|_0^\infty$$

$$= 1/2a \qquad \text{(A3.5)}$$

$$I_2 = \int_0^\infty x \exp(-ax^2)\,\Phi(bx)\,dx \qquad \text{(A3.6)}$$

Integrating by parts, and making use of the relation

$$\Phi'(bx) = (2b/\sqrt{\pi})\exp(-b^2x^2) \qquad \text{(A3.7)}$$

we find

$$I_2 = (-1/2a)\exp(-ax^2)\,\Phi(bx)\,\Big|_0^\infty$$

$$+ b/a\sqrt{\pi}\int_0^\infty \exp[-(a+b^2)x^2]\,dx$$

$$= 0 + (b/a\sqrt{\pi})(1/2)\sqrt{\pi/(a+b^2)}$$

$$= (b/2a)\sqrt{1/(a+b^2)} \qquad \text{(A3.8)}$$

Finally, from Eq. A3.3:

$$p(e) = a(I_1 - I_2)$$

$$= 1/2(1 - b/\sqrt{a+b^2}) \qquad \text{(A3.9)}$$

Eq. 7.2.17 follows from Eq. A3.9 after substitution of a and b from Eq. A3.1.

APPENDIX A4.

EQUATIONS FOR THE UNMINIMIZABLE PROBABILITY OF ERROR FROM BELLO AND NELIN [1]:

Eq. 7.2.17 was first derived by Bello and Nelin in [1], but in a different form. After the simplifications following from the assumptions of our chapter 7, their equations reduce to the following:

$$p(e) = 1/(2+\gamma) \tag{A4.1}$$

where

$$\gamma = (2m_{10})/(m_{11}-m_{10}) \tag{A4.2}$$

$$m_{11} = k\sqrt{\pi/C}[\Phi(\sqrt{C}) + \exp(-C) - 1] \tag{A4.3}$$

$$m = k\sqrt{\pi/C}[\Phi(2\sqrt{C}) - \Phi(\sqrt{C})] + (1/2C)[1 - 2e^{-C} + e^{-4C}] \tag{A4.4}$$

where

k is some constant that divides out in Eq. A4.2

$C = \pi^2 B^2 T^2 / \ln 2$

and

$$\Phi(x) = (2/\sqrt{\pi}) \int_0 \exp(-t^2)\,dt$$

For the sake of the reader trying to derive these expressions from [1], we point out that there is a typographical error in [1], Eq. (57). The first line should read:

$$m_{11}^{11} = m_{00}^{00} = 32E^2\sigma^2[(1/a)\{\sqrt{a\pi}\Phi(\sqrt{a}) + e^{-a} - 1\} + 1/\rho]$$

BIBLIOGRAPHY

[1] P.A. Bello and B.D. Nelin, "The influence of fading spectrum on the binary error probabilities of incoherent and differentially coherent matched filter receivers", IRE Trans. on Commun. Syst., vol. CS-10, pp. 160-169, June 1962.

[2] P.A.Bello and B.D. Nelin, "Predetection diversity combining with selectively fading channels", IRE Trans. Comm. Syst., vol. CS-10, March 1962.

[3] J.J. Bussgang and M. Leiter, "Error rate approximation for differential phase shift keying", IEEE Trans. on Comm. Syst., vol. CS-12, March 1964.

[4] B.H. Coetzer, "A comparison of spread spectrum techniques in a fading channel - a simulation", Ph. D. dissertation, City University of New York, 1982.

[5] K. Davies, "Ionospheric Radio Waves", Blaisdell Publishing Co., Waltham, 1969.

[6] L. Ehrman, et. al., "Real-time software simulation of the HF radio channel", IEEE Trans. Commun., vol COM-30, pp. 1809-1817, Aug. 1982.

[7] G.J. Foschini, R.D. Gitlin and S.B. Weinstein, "Optimization of two-dimensional signal constelletions in the presence of Gaussian noise", IEEE Trans. on Comm., vol. COM-22, no.1, Jan.1974.

[8] B. Goldberg, "300 kHz - 30 MHz MF/HF", IEEE Trans. Comm. Tech., vol. COM-14, Dec. 1966.

[9] I.S. Gradshteyn and I.M. Ryzhik, "Table of Integrals, Series, and Products", Academic Press, New York, 1081.

[10] R.C. Harper, "Adaptive phase and amplitude modulation on a frequency dispersive fading channel", IEEE Trans. on Comm., vol. COM-22, no. 6, June 1974.

[11] Internasional Mathematical Subroutine Library (IMSL)7500 Bellaire Blvd. Houston, Texas 77036.

[12] E.C. Jordan and K.G. Balmain, "Electromagnetic Waves and Radiating Systems", Second Edition, Prentice-Hall, Engelwood Cliffs, 1968.

[13] J.M. Kelso,"Radio Ray Propagation in the Ionosphere", McGraw-Hill, New York 1964.

[14] A.V. Oppenheim and R.W. Schafer, "Digital signal processing", Prentice-Hall, Engelwood Cliffs, 1975.

[15] A. Papoulis, "Probability, Random Variables and Stochastic Processes", McGraw-Hill, New York, 1965.

[16] R.F. Pawula, S.O. Rice and J.H. Roberts, "Distribution of the phase angle between two vectors perturbed by Gaussian noise," IEEE Trans. Commun., vol. COM-30, Aug. 1982.

[17] L.W. Pickering, "The calculation of ionospheric doppler spread on HF communication channels", IEEE Trans. on Comm., vol. COM-23, no. 5, May 1975.

[18] R. Price, "The autocorrelogram of a complete carrier wave received over the ionosphere at oblique incidence", IRE Proc.., vol. 45, June 1957.

[19] J.G. Proakis, "Digital Communications", McGraw-Hill, New York, 1983.

[20] J.G. Proakis, "Probabilities of error for adaptive reception of M-phase Signals," IEEE Trans. Comm. Tech., Vol. COM-16, Feb. 1968.

[21] S.S. Rappaport, D.L. Schilling, et. al., "Digital simulation of a spread spectrum communication system, ICC 1981.

[22] N. Ream, "Simulation of Rayleigh fading by digital generation of amplitude and phase", The Radio and Electronic Engineer, vol. 48, no. 11, pp. 567-572, Nov. 1978.

[23] P.N. Saveskie, "Radio Propagation Handbook", Tab Books Inc., Blue Ridge Summit, PA. 17214, 1980.

[24] M. Schwartz, Information Transmission, Modulation and Noise, Third Edition, McGraw-Hill, New York, 1980.

[25] R.A. Shepherd and J.B. Lomax, "Frequency spread in ionospheric radio propagation", IEEE Trans. on Comm. Tech., vol. COM-15, no. 2, April 1967.

[26] M.R. Spiegel, "Mathematical Handbook", Schaum's Outline Series, McGraw-Hill, New York, 1968.

[27] H. Taub and D.L. Schilling, "Principles of Communication Systems", McGraw-Hill, New York, 1971.

[28] H.B. Voelcker, "Phase-shift keying in fading channels",

Proceeding of the IEE, Jan 1960, Paper no. 3125E.

[29] C.C. Watterson, J.R. Juroshek, and W.D. Bensema, "Experimental confirmation of an HF channel model", IEEE Tans. Commun. Tech., vol. COM-18, pp. 792-803, Dec. 1970.

[30] C.C. Watterson and R.M. Coon, "Recommended specifications for ionospheric channel and atmospheric noise simulators", U.S. Dept. of Commerce/Environmental Science Services Administration, Publication no. ERL 127-ITS 89, September 1969, Bouder, CO. Obtainable from Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402.

[31] K. Wesokowski, "Computer generation of a slowly varying pseudorandom process", ICC 1983.

[32] A.B. Williams, "Electronic Filter Design Handbook", McGraw-Hill, New York,1981.

[33] J.M. Wozencraft and I.M. Jacobs, "Principles of Communication Systems", John Wiley & Sons, New York, 1965.

[34] M. Schwartz, W.R. Bennett and S. Stein, "Communication Systems and Techniques", McGraw-Hill, New York, 1966.

[35] A.A. Barrios, "High-frequency skywave system performance in a nuclear environment," IEEE Trans. on Military Electronics, Vol. MIL-9, April 1965.