OPTIMIZATION OF COMMUNICATION DATA RATES UNDER FADING CONDITIONS

by Pius Alois Okelo

A THESIS SUBMITTED TO THE UNIVERSITY OF LONDON FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN ENGINEERING

Pius Alois Okelo
Department of Electrical Engineering
University College Nairobi
P.O. Box 30197
Nairobi
Kenya

July 1970

ABSTRACT

This thesis is concerned with the study of techniques of alleviating the detrimental effects of signal fading, that is, fluctuation of the received signal level that occurs over certain radio circuits. To begin with, a brief review of fading communication channels, a description of the channel model used in the thesis and a summary of the techniques which have been used in the past to counteract fading are given.

Techniques in which a transmission parameter is varied in the attempt to reduce the fluctuations of the received signal level have been analysed in the last decade and have been found to lead to significant advantages. Analyses of some of the existing variable-rate techniques have been included for later use in comparing various systems. In previous analyses, mainly continuous parameter variation, which is difficult to implement, has been considered.

Two new techniques of varying the rate of transmission of a system in a discrete manner in order to make easier the problem of implementation are analysed. In the first method, the time-duration of the transmitted signals is

varied in a discrete manner, depending on the conditions in the forward channel. In the second method, the rate of transmission is varied by altering the size of the signalling alphabet, that is, by selecting a signal to be transmitted from m possible signals, where m is a integer variable. The variable-duration technique is analysed with reference to an incoherent FSK system and the variable-size (-of- alphabet) is analysed with reference to a coherent ASK system.

The existing and some new techniques of implementing variable-rate schemes are described. Lastly, extensional ideas of combining the new with the existing variable-rate techniques, over-all discussion of results and suggestions for further work are given.

CONTENTS

1. INTRODUCTION

1.1 Fading Communication Channels

In many communication systems the signal levels at the receiver are relatively constant with time. Mainly, these are line-of-sight type systems. Examples of these are short range UHF, VHF or HF groundwave transmissions. In systems of this type the noise is additive and stationary, and is usually described in terms of the Gaussian (normal) probability law. Though communication using these systems is an important technology, it will not form part of the present study.

Some of the most important communication systems do not have a steady signal level at the receiver. The noise in these systems is not wholly described by the additive and stationary Gaussian model. Two examples of other disturbances which come readily to mind are the signal fading encountered in long-distance beyond-the-horizon radio links and the impulsive noise in wire or cable links.

This thesis will be concerned with the study of transmission of digital waveforms over radio channels and, in particular, with matching techniques for improving the performance of systems using these channels. Some of the channels of fundamental interest in the thesis are now described in a little more detail.

In conventional long-distance HF skywave propagation, fluctuation of the level of the received signal occurs because of multipath propagation conditions. Multipath means simply that there are many paths along which the electromagnetic energy can travel from the transmitter to the receiver such that the total propagation time differs between the signals arriving via the various routes. Physically, multipath propagation arises because the reflectors which bend back the radio waves are constantly moving and changing their reflection characteristics as a result of the dynamic nature of the medium. In tropospheric and ionospheric scatter similar irregularities in the scatterers also lead to multipath propagation conditions.

The use of satellites for communication is increasing. As long as the satellites (passive or active) are used as relay stations at low altitude levels, the signal levels at the receiver end remain fairly steady. However, when active satellites are used in "stationary" (very high altitude) orbit, more severe fading occurs than in either HF or tropospheric scatter. The results derived in this thesis can thus be applied to certain satellite communication systems.

The fluctuation of the received signal level, or fading, that occurs in the above channels is described in some detail in the next paragraph. Before this, it is important to state that the theory explaining the physical phenomena observed in a turbulent medium is not so important for the design of a communication system. Indeed, in tropospheric scatter, no existing theory alone explains all the phenomena observed in this mode of propagation. In this thesis, a fading signal will simply be considered to have certain statistical properties and the specific physical phenomena (some of which are mentioned above) giving rise to these properties will not be discussed.

Fading can be described as either short-term or long-term. A discussion of the relative periods over which the short- and the long-term fading can occur, for the channel model used in the thesis, is given in the following section. Experimental investigations (Ref. 5, pp. 129-130) indicate that short-term fading, also called fast fading, is such that the amplitude of the received signal follows a Rayleigh probability density law. Most anti-multipath or antifading measures are orientated towards improving the short-term quality of transmissions, that is, counteracting fast fading. The long-term performance of a system can be specified by imposing a condition that a minimum quality of transmission should be achieved over some minimum percentage of the total long-term period, e.g. 99% of the total hours of the years or 95% of the hours of the worst month. This thesis is concerned only with the methods of counteracting short-term fading.

The object of this section has been to discuss very briefly some of the more important fading communication media in which the matching techniques that are to be analysed in the thesis can be used. A more thorough discussion of fading communication media can be found in Chapter 9 of Ref. 74. The next section discusses briefly mathematical channel models and gives the model to be used in the thesis.

1.2 Channel Models

1.2.1 A General Model

All communication channels have an input and an output. In a noisy environment, when no signal is applied at the input of a channel, an output termed '~~additive~~ noise' is observed. Also, an applied input signal may arrive at the output distorted, attenuated and delayed. Many of these channels are linear. Fig. 1.1 below is a mathematical model of a time-varying linear channel with additive noise.

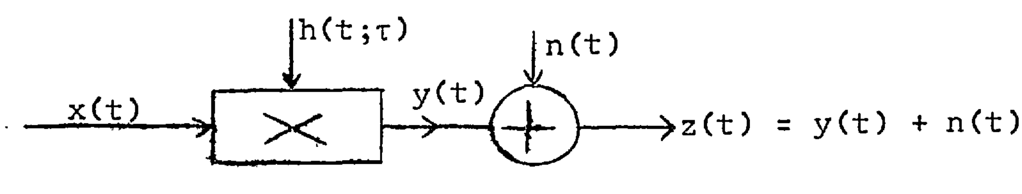


Fig. 1.1 A General Channel Model

In the Figure, $h(t;\tau)$ represents the impulse response of the linear time-varying channel. $h(t;\tau)$ may represent the fading, the distortion or the intersymbol interference in a system. $n(t)$ represents the additive noise.

With reference to Fig. 1.1, let $x_m(t)$ represent the general modulated signal, i.e.

$$x_m(t) = R_e\left\{ x(t)\, e^{j2\pi f_o t} \right\} \qquad \ldots.1.1$$

where f_o is the carrier frequency. $x(t)$, the complex envelope, denotes the amplitude and phase variations resulting from modulation.

If the additive noise $n(t)$ is ignored for the moment, then the received signal is $y(t)$, and for multipath propagation, is of the form

$$y(t) = \sum_k \alpha(t - t_o\,;\, k)\ x_m\left[(t_o - t - \tau_k\ (t)\right] \ldots..1.2$$

where $\alpha(t\,;\,k)$ is the amplitude of the signal received over the kth path at time t, with delay $\tau_k(t)$ relative to some average propagation time t_o. For continuous multipath, equation 1.2 becomes

$$y(t) = \int_{-\infty}^{\infty} \alpha[t - t_o ; \tau(t)] \; x_m [t - t_o - \tau(t)] \, d\tau \quad \ldots 1.3$$

Note: The functional dependence of α on changes with time, and this dependence is expressed by the function $\alpha\{t - t_o ; \tau(t)]$ in equation 1.3.

Substituting Equation 1.1 into 1.3 results in

$$y(t) = \text{Re}\left\{\int_{-\infty}^{\infty} \alpha (t-t_o ; \tau) \, x (t-t_o - \tau) e^{j 2\pi f (t-t_o-\tau)} \, d\tau\right\}$$

$$= \text{Re}\left\{ v(t-t_o) \, e^{j 2\pi f_o (t-t_o)} \right\} \quad \ldots\ldots 1.4$$

where

$$v(t) = \int_{-\infty}^{\infty} \alpha(t ; \tau) \, x (t -\tau) \, e^{-j 2\pi f_o \tau} \, d\tau \quad \ldots\ldots 1.5$$

With the substitution

$$h(t ; \tau) = \alpha (t ; \tau) \, e^{-j 2 \pi f_o \tau} , \quad \ldots\ldots 1.6$$

Equation 1.5 becomes

$$v(t) = \int_{-\infty}^{\infty} h(t ; \tau) \, x (t -\tau) \, d\tau \quad \ldots\ldots 1.7$$

From Equation 1.7, it is clear that the complex output of the received signal is equivalent to that which would be obtained by passing the transmitted signal through a linear time-varying filter. The time-varying functionality is such that $h(t ; \tau)$ can be regarded as a random variable in the t-dependence.

Under certain conditions, specified in the following Section, it is valid to use a single-tone input to estimate the conditions in the forward channel. For such an input, an approximation to the general channel model is made in the following Section.

1.2.2 The Piece-Wise Constant Approximation of the Channel Model for Time- and Frequency-Flat Rayleigh Fading

Experimental results obtained for a number of physical channels indicate that when the input signal x(t) is a sinusoidal waveform, then the output y(t) (in the absence of additive noise) is of the form (Ref. 86, p. 349),

$$y(t) = a(t) \sin\{\omega_0 t + \phi(t)\} \qquad \ldots\ldots 1.8$$

The envelope a(t) and the phase $\phi(t)$ of the received signal vary continuously with time. It has been observed experimentally (Ref. 5, p. 129-130) that a(t) and $\phi(t)$ appearing in Equation 1.8 have a Rayleigh and a uniform distribution respectively.

There are many transmission channels over which the fading is time-flat. This is also frequently referred to as 'slow' fading. It means that changes with time in the channel are slow enough not to be significant over the duration of one pulse (or in general over a transmission epoch). Time-flat fading is still of the short-term type described in Section 1.1 and must not be confused with the long-term (which is also sometimes called 'slow')fading. The time-flat fading occurs in seconds or fractions of seconds with a typical value of the order of six fades per minute. The long-term gross changes in the medium occur over periods of many minutes and, often, hours.

With the above assumption, a(t) in Equation 1.8 can be replaced by a piece-wise constant, a. This means that 'a' does not change over one transmitted pulse but is a random variable from pulse to pulse. The variation follows the Rayleigh distribution, i.e. has a probability density function of the form

$$p(a) = \frac{a}{\sigma_a^2} e^{-\frac{a^2}{\sigma_a^2}} \qquad .....1.9$$

Fig. 1.1 can now be replaced by the following simplified model;

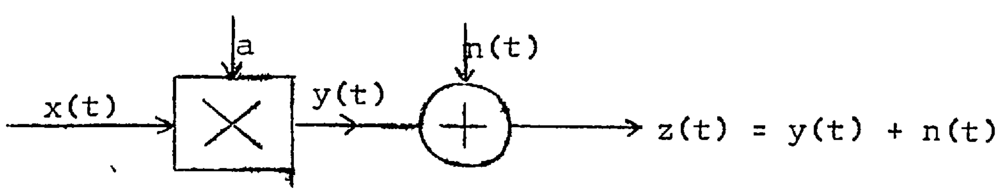


Fig. 1.2 Fading Channel Model with Piece - Wise Constant Approximation

Further, there are time-flat channels over which the fading is frequency-flat, that is, non-selective. This means that the whole spectrum of the signal fades up and down uniformly. For frequency-flat channels it would be valid to initiate a single tone from the transmitter to be used at the receiver for prediction of conditions in the forward channel.

The time- and frequency-flat fading will often be referred to in this thesis simply as flat-flat fading.

The second form of disturbance in the channel models is the additive noise. This includes contributions from the noise generated in the receiver, cosmic noise coupled to the receiving antenna, man-made interference, and intentional and unintentional interference from other systems.

In the analyses in this thesis, the additive noise n(t) is assumed to be a stationary Gaussian process with zero mean. Thus the probability density function of the instantaneous value of n(t) is given by

$$p(n) = \frac{1}{\sqrt{2\pi}\,\sigma_n} e^{-\frac{n^2}{2\sigma_n^2}} \qquad \text{.....1.10}$$

where σ_n^2 is the noise power.

The channel model used in the thesis will be assumed to be of the form shown in Fig. 1.2.

1.3 Background Discussion of Methods of Counteracting Fading

When the received signal level fluctuates, the periods that it exceeds a predetermined value are called up-fades (or surges) and the periods that it is below the reference level are called down-fades, or simply fades. When the term fading is used in this thesis it will be taken to mean down-fades.

The fading of signals causes severe deterioration in the performance of fixed-parameter digital communication systems. This is because there is a tendency for more transmission errors to occur during the weak signal periods with the net effect that the average probability of error is higher than it would have been with a non-fading signal of the same average power. For example, in the absence of fading, the bit probability of error for a binary communications system using matched filter decreases exponentially with increasing detection signal-to-noise ratio (SNR). This result is derived in Chapter 3. With flat-flat Rayleigh fading, the bit probability of error decreases only linearly with increasing SNR. See Table 3.2 in Chapter 3. The methods which have been used in the past to counteract fading will now be discussed briefly.

1.3.1 Early Methods

Until the late 1950's, only two methods were used to overcome the detrimental effects of signal fading. The first was to design the system to operate with a minimum acceptable performance under the worst possible conditions. With systems of this kind, a fixed-parameter technique was used with the transmitter power being set at a sufficiently high level to guarantee that the received signal level was such as to give minimum performance requirements for 99% of the total transmission time. This design for worst-case operation necessitates continuous radiation of large quantities of power. Yeh[93] has shown, experimentally, that the high level of power is only necessary for 0.1% of the transmission time. The excess radiation of power is not only expensive but it also creates severe interference, particularly in congested bands. In addition, it results in a very high rate of deterioration of the equipment employed and thus leads to a high cost of maintenance. This latter problem can be extremely serious in situations where the communication stations have to be placed in sites that are not readily accessible. Also, in satellite systems it is, more often that not, impossible to design for the worst-case operation.

The second method which was developed and is still commonly used to reduce signal fluctuations was diversity. The object of this technique is to arrange for transmission of signals over several paths and arrange for the signals received over the various paths to be reasonably uncorrelated. If this is done then the probability of the signals in any two branches going into a fade simultaneously will be small. Clearly, the greater the number of diversity branches the less will be the fluctuations of the combined or selected signal. Depending on the mode of propagation, the diversity techniques which can be used are: spaced antenna (space), frequency (different carriers), polarisation, angle (-of-arrival) and time (signal repetition) diversities. Some of the methods which have been used to combine the signals in the various branches are: the selection technique, the maximal-ratio, the equal-gain and the square-law combining techniques.

Several problems can arise in practice with diversity techniques. The obvious one is the added cost of implementation, particularly when it is necessary to use a large number of branches in order to reduce the signal fluctuations to an acceptable level.

This situation can easily arise in cases where the correlation between the signals in various branches is not small (Ref. 5, pp. 164-167). Although the requirement that the correlation between the signals in the branches be less than some specified value (Ref. 74, pp. 429-434) is met at present in many practical applications, this cannot be guaranteed in future if the use of diversity increases. For example, in time diversity, the storage capacity limits the time separation between signals as they are repeated and in frequency diversity, spectrum allocation limits the frequency separation between signals in the various branches. Lastly, the use of certain diversity techniques, for example spaced antenna, would lead to severe interference in congested communication bands.

The early methods of counteracting fading will not be discussed further in this thesis.

1.3.2 More Recent Methods

The difficulties with diversity techniques discussed above, and the relative inefficiency of the worst-case design technique, has led in the last decade or so, to a search for new methods of counteracting fading. Some of the most promising methods that have been reported are those which employ in one form or other 'matching' techniques. With these techniques, a signal parameter is varied to compensate for the signal fluctuations.

Montgomery (Ref. 47, pp. 1678-1684) has analysed a continuously-variable-rate technique for a frequency-shift-keyed (FSK) system and pointed out the potential gain that could be derived from adjusting the rate of signal transmission to compensate for the fluctuations in the a received signal level. Yeh[93] and Axelby and Osborne[4] have investigated the use of automatic control of transmitter power on fading circuits and have reported important advantages. Palmer, et al[64] have studied a system in which the rate of transmission is varied continuously when the received SNR lies between a lower and an upper threshold. Above the upper threshold SNR and below the lower threshold SNR, the rate of transmission is fixed at a high and at a low value respectively. From the analysis of this technique, Palmer et al have reported considerable saving in transmitter power. Signal 'repetition-on-request' methods[73,85] have also been analysed and have been shown to have advantages over fixed-parameter transmissions. Lastly, intermittent system operation whereby transmission is halted during fades has been analysed by Montgomery (Ref. 47, pp. 1678-1684), Cowan et al[23] and Martin[60]. They have all reported large savings in the transmitter power that can be obtained by using intermittent system operation.

Without exception, the investigators cited above have pointed out the considerable improvement in system performance or other advantages that can be obtained by channel 'matching' techniques.

Systems have also been built using these principles and tested in the field. The first to be built was a meteor-burst intermittent communication system, commonly known as the JANET[47]. The 'RAKE' principle has also been used in two systems which have both been field tested. The first was known simply as the 'RAKE' receiver[68]. This system works well under poor propagation conditions but is rather wasteful of bandwidth when the channel conditions are good. This is because the system does not operate at a flexible data rate. To overcome this problem, the AN/GSC-10 (KATHRYN)[51,94] system was designed. This system allows the data rate to be varied by using signal redundancy. A report on the field test of this system has recently been published[51]. Lastly, field tests have been reported[6] on a system, called COMET, which employs signal repetition and diversity reception.

These more recent channel matching techniques will be reviewed in greater detail in Chapter 2. In the following Section, some techniques in which the data rate is varied directly as a means of counteracting fading are described.

1.3.3 Varying the Data Rate

The one way of alleviating the effects of signal fading that will receive detailed consideration in this thesis is that of varying the operating data rate. It is possible to do this when the average data rate (measured in bauds) is much higher than the fading rate (measured in Hertz). Under certain conditions*, Pierce[65] has shown that for a variable-rate scheme, the average probability of error is lower bounded by a function $P_e(U_o)$, where U_o is the average SNR. In other words, any choice of data rate variation other than the optimum choice leads to an average probability of error greater than $P_e(U_o)$. From considerations of Pierce's work, it can be seen that the optimum way of varying the data rate, so as to attain the lower bound, is to vary the duration, or some other parameter, of the transmitted signal in order to maintain a SNR of exactly U_o at the receiver.

Pierce has also analysed the intermittent system as a simple nonoptimum method of varying the data rate. For flat-flat Rayleigh fading, he showed that asymptotically (for low error rates), the

*Specifically that the probability of digit error, say $f_e(U)$, as a function of instantaneous SNR, is continuous and concave upwards.

intermittent system requires only 4.34 dB more transmitter power than the optimum method of varying the data rate. The intermittent system is thus theoretically almost as good as the optimum system. However, some problems do arise with the implementation of intermittent systems. These problems will be discussed in Chapter 2.

It is evident that an optimum variable-rate system is an ideal that cannot be achieved in practice since it assumes that the forward channel gain (which is a random variable) is known exactly at the transmitter. Even under ideal conditions, namely, perfect prediction, and delayless and noiseless feedback, when the above requirement could be achieved, there is, further, the implied requirement that the parameter varied should be able to take any value however large or small. This additional requirement is clearly physically unattainable.

A new method of varying the data rate is suggested in this thesis to enable easier practical implementation than that involved in the continuous rate variation scheme. This method is to vary the data rate in discrete steps in one of two ways. One way is to make the duration of the transmitted signals take certain discrete values. The second method is to vary the size of the signalling alphabet, that is, to select the signal to be transmitted from m possible signals, where the number m is variable. The first method will be called here variable-duration technique and the second will be called variable-level technique.

A n-data-rate system is thus defined as that system which is capable of transmitting information at n discrete data rates. When one of the n possible data rates is zero, the system will be called an intermittent system. Otherwise, the system is a non-intermittent one. Evidently, these definitions are such that the class of systems handled include existing intermittent systems and fixed-rate systems which transmit continuously.

The first object of this thesis is to derive the formulae for the performance of the new class of nDR system defined above. The derivations are carried out with application to two types of signalling technique. The variable-duration method of varying the rate is applied to incoherent FSK signalling and the variable-level technique is applied to coherent ASK signalling.

The second object of the thesis will then be to apply the general principles to easily realisable, mainly two-data-rate (2DR) type, systems. The first system to be analysed will be a two-data-rate system in which one of the rates is zero. This is the same intermittent system as that analysed by Montgomery, (Ref. 47 pp. 1678-1684), Pierce[65], Cowan et al[23] and Martin[60].

In all previous analyses (i.e. those referred to above), the operating data rate is allowed to take as high a value as the system equations dictate. It will be shown in Chapter 7 that in order to maintain a constant average data rate, the higher operating data rate increases without bound as SNR increases, for an optimised threshold SNR. Practical limitations, for example, bandwidth restrictions and realisability constraints, require that an upper bound, say R_{max}, be placed on the operating data rate. Analysis of two-data-rate systems with an upper limitation placed on the higher date rate is given in Chapter 7. This analysis is new to the author's knowledge.

A non-intermittent system using two non-zero data rates is somewhat more general than the intermittent system since the former does not pre-assign a value to one of the data rates. One may ask, for example, what the optimum choice of the two rates of a 2DR system is. The answer to this question will be given in Chapter 8, after the analysis of the 2DR systems. The concept of switching between two nonzero rates under fading conditions has not, to the author's knowledge, been analysed before. The advantages of such an operation will be discussed in Chapter 8 when comparing the various types of two-data-rate systems.

Combining more than one technique to counteract fading (hybrid techniques) can be very useful where good system performance is necessary. Several hybrid systems making use of 2DR systems will be suggested in Chapter 10.

A further objective of the thesis will be to suggest some methods by which the nDR class of system could be implemented. The last major objective will be to compare and contrast the digital communication systems that are analysed or quoted in the thesis. The comparison between the new techniques given in this thesis and the established techniques will be stressed. These comparisons are made on the basis to be described in the next section.

1.4 A Note on Methods of Comparing Digital Communication Systems

In designing a digital communication system, the final choice from the available alternatives depends mainly on economic factors. Nevertheless, it is helpful to have a definite way of comparing and ordering systems. The designer is usually concerned with how reliably a system can transmit messages (i.e. the probability of error for a given SNR), how many messages it can transmit per unit time, and what bandwidth it requires. Some rational way by which to compare different systems is thus required. The present section discusses this problem as it will be necessary towards the end of the thesis to compare various digital communication systems that are analysed herein.

The systems to be compared may differ in bandwidth, in the number of possible signals used for transmission (i.e. the level of the system), in the type of coding used, or in the average transmission rate. If systems which differ in average data rates were to be compared, for example, it would be found generally that the lower rates are the more reliable. Therefore, when one finds that a system transmits messages with a lower probability of error than some other system under similar circumstances, one must be aware of the possibility that the 'superior' system has a lower average rate of transmission. Thus from one point of view, in order to make a set of systems strictly comparable, the constraint of equal average data rates between the individual systems should be imposed, if possible. Thus, neglecting bandwidth considerations, the system with the lowest probability of error, for the same transmitter power and average rate (more strictly, for the same energy per signal dimension), is the 'best' of the set of systems being compared.

It must be pointed out that the 'best' system might still have some disadvantages as compared with other systems. For example, the assumption that there is no restriction in the available bandwidth is generally a fictitious one. Thus consider a fixed rate system transmitting continuously and an intermittent system using one fixed rate for transmission and halting from time to time when the conditions in the channel are not favourable. If both systems operate under flat-flat Rayleigh fading conditions it is shown in Chapter 8 that the intermittent system has a better error rate for the same transmitter power and the same average data rate. It is evident that the intermittent system has a larger bandwidth since it employs a higher operating data rate whenever it transmits. This could clearly be a disadvantage when bandwidth is restricted.

The equal-rate comparison is therefore appropriate when there is no strict bandwidth limitations and one tries to attain the most reliable communication for a given average rate of transmission. A second method of comparing digital systems is to adjust the operating data rate for each system so that the systems have equal bandwidths[43]. This is most appropriate when the system to be designed is to transmit within a limited band of frequencies, which is to be used as efficiently as possible.

Some references dealing in detail with comparison of digital communication systems are C. W. Helstrom[43] and I. Jacobs[46].

Circumstances play a large part in choosing which parameters to hold constant when comparing communication systems.

In this thesis the equal-rate comparison will be used. This means that when comparing two systems, the transmitter power saving of the 'better' system for the system will be quoted for the same average data rate. Alternatively, for the same transmitter power and average data rate, the improvement will be given in terms of a reduced probability of error.

Some cases exist where it is necessary to accept a drop in the average data rate. An example of this is the ARQ (automatic repetition on request) system. Another example is a two-data-rate system in which the operating data rate has to be less than a given value. This will be considered in Chapter 7.

In the ARQ and other retransmission techniques, the decrease in the average data rate can be extremely severe for low SNR's. In order not to render the comparison of such a system to other systems meaningless, it is necessary to control the decline in the average, or throughput, rate. A compromise between the performance and the average rate could be made in this case. A method of doing so is analysed in Chapter 4.

1.5 Organisation of the Thesis

The present Chapter is a general introduction to the thesis. It opens with a brief discussion of the physical channels in which fading occurs and then gives a general and a simplified characterisation of these channels. A brief historical survey into the methods of counteracting fading is made in Section 1.3. This background discussion leads to a brief introductory treatment of the new techniques suggested and analysed in this thesis. The basis of comparing digital communications systems, discussed in the last Section, will be required later in the thesis.

In Chapter 2 a literature survey of the recent techniques of counteracting fading is provided. Chapter 3 deals with the classical problem of detecting fading and non-fading signals in additive Gaussian noise. The main object of the Chapter is to demonstrate the deterioration in the performance of a system operating under fading conditions, when compared to its performance in the absence of fading.

In Chapter 4 analysis of a retransmission technique (the ARQ) is carried out. The chief aim being to derive the formulae and the performance curves of this technique in forms which will make them directly comparable to those of the new techniques discussed in the later part of the thesis. A modification is also suggested for the ARQ system.

Chapters 5 through 7 analyse in detail the new methods of varying the data rate which were suggested in Section 1.3.3. Chapters 5 and 6 deal with the general theory of n-data-rate (nDR) systems, which have already been defined. Chapter 7 applies the general theories to two-data-rate systems which are the simplest class of n-data-rate systems. When analysing two-data-rate systems in the first part of Chapter 7, no constraint is placed on the value that the higher rate can take. However, it is important that the higher rate used for an analysis of a two-data-rate system be achievable in practice. Thus, in the second part of Chapter 7, the constraint that the higher rate of the two rates of a two-data-rate system be less than or equal to a pre-fixed value, R_{max}, is introduced.

The main results of the thesis are discussed in Chapter 8. For ease of comparison between the various two-data-rate systems, all the performance curves of these systems are given in that Chapter. In addition, for ease of comparing two-data-rate systems with ARQ and with an interleaved Golay code, the curves of Chapters 3 and 4 are reproduced in Chapter 8.

In Chapter 9, the methods of practical implementation of n-data-rate systems will be discussed. Also some attention will be given to the difficulties that may arise from some of the assumptions that are made in the theoretical derivations of the performance functions of the n-data-rate systems. Chapter 10 gives some extensions to the main study by considering combining the new techniques with some existing techniques, in particular ARQ, in order to obtain further improvement in the system performance. Chapter 11 gives the overall discussion of the main results of the thesis and suggests possible angles of approach in future studies.

On a first reading, Chapters 3 and 4 can be omitted without loss of continuity as long as the results (most of which are standard) that are derived in these chapters are accepted. However, two points need to be noted. The first is that in Chapter 3, a concise method has been evolved for deriving the probability of error for three basic signalling techniques (namely, ASK, FSK, and PSK) operating under conditions of flat-flat Rayleigh fading. The second point is the modification which has been made to the ARQ technique in Chapter 4. The number of repetitions of a given cycle are restricted so as to allow an acceptable through-put rate for all SNRs. This is really a compromise between the performance and the information rate at low SNRs.

2. A REVIEW OF EXISTING VARIABLE-PARAMETER SYSTEMS

This Chapter provides a literature survey of the recent techniques which have been used to counteract fading. These techniques have a factor in common in that they adjust one transmission parameter or another to counteract the detrimental effects of fading. Most of these techniques also require the use of a feedback link to inform the transmitter about the state of the forward channel. The use of feedback is, therefore, very important to some communications systems. For this reason a review of feedback and two-way communications is provided in the next section. In Section 2.2, a method of varying the transmitter power under fading conditions is reviewed.

Varying the rate of transmission can be effected by a direct or an indirect method. Techniques of direct rate variation are reviewed in Section 2.3. Indirect techniques using mainly the retransmission-on-request (also called repeat-request) strategy are considered in Section 2.4. Lastly, forward error control (coding) techniques are reivewed in Section 2.5.

2.1 Feedback and Two-Way Communications

In certain communication situations, the opportunity, and sometimes the necessity exists for the use of a feedback link.

In such situations, some information from the receiver can be made available to the transmitter using the return path. It is well known (Ref. 5, pp. 345-368) that there exists in these cases a possibility of increasing the reliability of the forward transmission or decreasing the equipment complexity, and hence the cost, by exploiting the fed-back information. Clearly, the use of feedback is essential if some parameter of a signal, e.g. the duration, is to be adjusted at the transmitter on evidence obtained by the receiver of fading in forward channel.

For two decades the theoretical aspects of feedback and two-way communications have received detailed analyses. See, for example, Refs. 19, 41, 5 pp. 363-366, 32 pp. 481-482. Alongside this development, but quite independently, practical feedback schemes have been designed for use in communication systems[47,6].

In the following sub-section, definitions of different types of feedback are given. In subsequent sub-sections examples of the application of each type of feedback are reviewed.

2.1.1 Types of Feedback

Feedback techniques can be divided into two broad classes. There are those that supply the transmitter with information about the actual channel and others that supply the transmitter with data relating to the individual receiver decisions or strings of decisions. The former class is known as predecision feedback and the latter is known as postdecision feedback. In post-decision feedback the decision as to the action that should be taken when the performance of a system deteriorates below the acceptable level is generally taken at the receiver. The feed-back signal then requests the transmitter simply to carry out the receiver's wish which might be, for example, that a group of symbols be repeated. In predecision feedback the receiver sends to the transmitter some evidence of the state of the forward channel. On this evidence, the transmitter, and possibly the receiver, modify one or more transmission parameters to match the channel conditions. Green (Ref. 5, pp. 345-368) has given a thorough review of postdecision and predecision feedback systems up to about 1960. Sometimes postdecision and predecision feedback are called ~~'information'~~ 'decision' and ~~'decision'~~ 'information' feedback respectively.

Turin[84] has described a type of feedback technique which he has called 'uncertainty' feedback. This technique is a com-bination of predecision and postdecision feedback. It employs predecision feedback in that the feedback signals are sent before any decision about the received message is taken. Postdecision feedback is also utilised indirectly since the return path is also used to synchronise the transmitter and the receiver.

2.1.2 An Example of the Use of Postdecision Feedback

A form of postdecision feedback is the use of a null-zone detection technique with retransmission[14,73,66]. If the decision level of a received signal falls in a null-zone detection region, which is predetermined by the design of the system, no decision about the received signal is taken. However, a request for the retransmission of the message is made via the feedback link.

It is intuitively obvious that the probability of error decreases as the size of the null-zone region is increased. The price paid for this increase in reliability is an increase in the length of time which it will take to transmit the message and a consequential reduction in the effective rate of data transmission. To minimise the probability of error, the null-zone region should be made as large as possible whereas to decrease the probability of retransmission, the null-zone region is required to be small. Generally, a compromise size of the null-zone region, which does not decrease the transmission rate to unacceptable value but gives the desired probability of error, can be found.[66,73]

An important improvement to the fixed null-zone system is variable null-zone reception. This is particularly useful in combating fast fading. It has been shown by Schwartz et al[73], that the variable-null system is superior to all fixed-null systems except in one case. This is when sufficiently large transmission times are allowed, in an effort to achieve very low error rates. In this case the repetition-integration (or cumulative decision feedback) scheme[73] performs better with a fixed-null than with a variable-null system. The reason for this is that when the number of rejections is very large, it is inefficient to discard decisions when requesting retransmissions.

The disadvantages of null-zone detection techniques with retransmissions as a method of counteracting fading are as follows. For the fixed-null operation, the size of the null-zone region must be large in order to improve sufficiently the performance of the system. As discussed above, this leads to a very low 'through-put' rate of data transfer. The variable-null operation solves this problem to some extent but then the complexity of the system increases with this operation because the null-zone region is varied with time.

Another postdecision feedback scheme, the ARQ system, will be described in Section 2.4 and analysed in Chapter 4. In the next sub-section, the use of predecision feedback schemes is discussed.

2.1.3 Predecision Feedback Schemes

To make use of the predecision feedback technique an estimate (preferably in the future, i.e. a prediction) of the state of the forward channel is required. There are several methods by which such a ~~stastic~~ statistic of the channel can be obtained. Two of them are described below.

Consider the piece-wise constant Rayleigh fading model shown in Fig. 1.2. Because the fading is time-flat, it is possible to estimate, in a quasi-stationary sense, the multiplicative constant, a. This estimate can then be made known to the transmitter and is used as a basis of adjusting a transmission parameter to 'match' the changes in the forward channel.

Another method of estimating characteristics of the forward channel is to monitor at the receiver a 'pilot' tone initiated at the transmitter. An estimate of the future channel gain is then made at the receiver using the received pilot tone signal. Information about this estimate is passed on to the transmitter which would then take some action based on the received feedback signal.

An example of a practical predecision feedback scheme is the 'JANET'[47] system. The 'JANET' system employs the pilot tone technique of estimating the forward channel. In the JANET system a very simple parameter is fed back to the transmitter, namely the times at which the conditions in the channel are favourable for transmission. Burst transmission of stored data then takes place during such intervals, with data piling up in storage at the transmitter between intervals. Thus the JANET scheme is also an example of an intermittent system. The disadvantage of the intermittent systems will be discussed in section 2.6.

2.2. Automatic Power Control

As was stated in Chapter 1, Yeh[93] has shown experimentally that the large amount of transmitter power required to counteract fading is necessary for only 0.1% of the transmission time. It is argued, also in Chapter 1, that to radiate continuously such excessive amount of power is undesirable and expensive. An answer to this problem is to control the transmitter power such that the large amount is radiated only when it is required.

Axelby and Osborne[4] have analysed a method of automatically controlling the transmitter power under fading conditions. In their system, the signal at the receiver is compared with a reference level. The difference between the two (the actuating signal) is amplified and used to control the power radiated from the transmitter. When there is fading in the channel, the amplified actuating signal goes up and the transmitter power also increases. Conversely, when the level of the signal at the receiver is high, i.e. relatively good conditions exist in the forward channel, the transmitter power goes down as a result of a decrease in the actuating signal. This is clearly a conventional control process by which the system automatically and continuously makes adjustments to the transmitted power to suit the conditions in the channel.

The main advantages of automatic power control described by Axelby and Osborne are the significant reduction of interference with other systems, a saving in the average transmitter power, particularly for very high reliability operation, and a reduction in the maintenance and operation costs.

The chief disadvantage of this system operation is that a powerful, and hence expensive, transmitter is required, and the full capacity of the transmitter is not exploited most of the transmission time. The interference problem is not completely resolved because there are periods when very large amounts of power must be radiated from the transmitter to counteract the fading. During these periods the interference with other systems is severe and could be intolerable.

2.3 A Review of Existing Variable Rate Schemes

Rate variation, as a method of counteracting signal fading, can be implemented by a direct or an indirect technique. Direct techniques of varying the data rate; for example, those which adjust the duration or the bandwidth of the transmitted signal are reviewed in this section. The indirect techniques which include retransmission and forward error correction are reviewed in Sections 2.1.2, 2.4 and 2.5.

The simplest method of direct rate variation is the intermittent system in which the rate is switched between a fixed value and zero. This system was suggested by Montgomery (Ref. 5, pp.1678-1684) in 1957. As mentioned in section 1.3.3, Pierce[65] has shown that the improvement in the system performance obtainable by the intermittent operation is close to that attainable by the optimum rate variation. Intermittent operation has been further analysed by Martin[60] and Cowan, et al[23].

Martin carried out a computation for an intermittent system and presented his results in such a way as to make them easily comparable with those of a multireceiver (space diversity) system. Martin shows the point at which these two systems are equivalent.

In Martin's computation, the average rate was not evaluated. Also the effect, on the data rate, of increasing the SNR used for transmission was not discussed. Thus the improvement in performance due to the intermittent operation in which the average rate is not fixed is either overrated or underrated, depending on whether the average rate is increased or decreased. An increase in the average rate makes the performance worse and a decrease in the average rate improves the performance. The average rate must, whenever possible, be either held fixed or evaluated.

Cowan, et al[23], have carried out a further analysis of the intermittent system. The analysis takes into account the constraints which would be introduced by a physical system, namely, noisy feedback, the average and peak power limitations and the round trip delay. Cowan, et al, show the effects of these physical constraints on the performance of an intermittent system.

A modified version of the continuously variable-rate system analysed by Montgomery (Ref. 47, pp. 1678-1684) has been proposed by Palmer, et al[64]. In this modified system the rate is varied continuously when the SNR lies between an upper and a lower value. Outside this range the system operates at fixed rates. Above the upper threshold, a high fixed data rate is used and below the lower threshold, a low fixed data rate is used. This system operates close to the optimum variable-rate system for high error rates but at low error rates its performance suffers from the fact that it does not stop the message transmissions during a deep fade.

In variable-rate schemes, it is necessary to know the relationship between the instantaneous SNR and the instantaneous coherent bandwidth of the system. The data rate could then be made directly proportional to the SNR. This, in fact, has been shown by Pierce[65] to be the optimum rate variation.

Unfortunately little experimental work has been done to try and find the relationship between the instantaneous SNR and the coherent bandwidth of the system, for any of the fading media. One exception is the work of Statas[79] who has found from experimental studies that the instantaneous coherent bandwidth of a tropospheric scatter system increases with the signal level. Statas, however, does not specify the exact relationship.

2.4 Retransmission Error Control

In retransmission error control, repeats of individual symbols, or a group of symbols, are made when a retransmission request is made from the receiver to the transmitter. Most retransmission techniques that have been analysed[11,12,54,85] have a common factor in that the feedback channel is used to convey a decision taken at the receiver.

When requests are made for retransmission of erroneous code words, the detection of the errors is generally done by short codes. An example of this is the fixed ratio (3- out-of-7) code for the van Duuren ARQ system. The ARQ system operation is described in Section 2.4.2. The following section deals with retransmission logics, i.e. the actual technique and procedure of requesting repetitions, which can vary considerably depending on the application.

2.4.1 Retransmission Logics

The simplest retransmission logics are the idle-RQ type[11] in which the the transmitter idles by sending dummy symbols until it receives a confirmation or a repeat-request (RQ) symbol. The disadvantages of the idle-RQ are that considerable time is wasted by the idling and that extra redundancy is required to protect the confirmation and the RQ transmitted symbols as they are transmitted independently.

The simple RQ logics have been designed[11] to overcome some of the difficulties of the idle-RQ logics. The simple-RQ logics send confirmation and RQ symbols in the same block as the data symbols. These logics, unlike the idle-RQ logics, send information continuously. The simple-RQ logics use one bit position following the data to indicate confirmation or RQ. This position is thus protected by coding along with the data. The operation of the simple-RQ[11] logics is such that undetected errors can cause accepted messages to be repeated and can cause the failure of discarded messages to be repeated. The assumption that messages containing errors contain requests for retransmissions can cause unrequested retransmissions. This results in messages being printed twice without the occurrence of undetected errors. In any case, since the next code word is being transmitted before the confirmation or RQ for the preceding word is received, code words will be often received out of proper order. Additional equipment is required to restore the proper order.

Both of the difficulties with the simple RQ logics can be overcome by using dual-RQ logics[11,85], in which more than one retransmission is made each time a request is made.

Of all retransmission schemes, probably the most widely known is van Duuren's - automatic request for retransmission - system (ARQ)[85]. The ARQ technique, which uses dual-RQ logics, is described in the next sub-section.

2.4.2 The ARQ Technique[85]

The ARQ process uses 35 out of the possible 128 combinations of seven binary digits. These 35 are those which each have exactly three 'ones', hence the terminology 3-out-of-7 fixed ratio code. The 35 code words are three more than the telegraph alphabet requires, and one of them is used as a request signal (RQ). Any errors affecting an odd number of binary digits of a code word and some errors affecting an even number can be detected since they transform the original code word into another which does not conform to the fixed ratio of the code. Errors which transform a code word into any of the other 34 valid code words are of course undetected. These particular error patterns cause transpositions.

The ARQ technique uses a two-way communications system between terminals A and B, say. It is a synchronous system and, therefore, it does not require start and stop signals. The principle of the ARQ technique is illustrated in Fig. 2.1 which represents, on two parallel time scales, the flow of information entering and leaving at the two terminals.

The detection of an error at station B when, for example, character 'C' from station A is mutilated, causes an RQ to be transmitted to request the repetition of the character which was detected to be in error. In practice, because of the delay in the propagation path and in the equipment, it is necessary to repeat more than one character. In the case shown in Fig. 2.1, four characters are repeated. Upon reception of RQ, station A sends back an identical RQ followed by the previous four characters (the repetition cycle). For reasons of symmetry, station B also repeats its previous four characters. If both stations each detect an error simultaneously, both initiate a repetition cycle and the system will still function correctly.

The ARQ system will be analysed in Chapter 4. It will be seen from its performance curve, given in that Chapter, that the chief disadvantage of the technique is the severe decrease in the average rate of transmission, particularly for low SNRs. This is because the system will keep on repeating, often the same cycle, for as long as is necessary for correct detection. A method of restricting the number of the retransmissions in the ARQ technique for the SNRs is suggested and analysed in Chapter 4.

Because of the retransmissions and coding (when used), the information, or through-put, rate will be less than the actual rate at which symbols are transmitted over the channel. An analysis of the reduction in the transmission rate due to retransmissions has been given by Kuhn[54]. Kuhn finds that the chief factor controlling the decrease in the transmission rate is the probability of detecting a code word correctly except where very long code words are used and when these are combined with the existence of a high element error rate.

It is clear that a retransmission error control technique attempts to 'match' indirectly the transmission (through-put) rate to the conditions prevailing in the forward channel. To elaborate on this point, it is evident that in a random-error channel, retransmissions would be spread out in a random fashion across the transmission time whereas in a burst-error channel the retransmissions, like the errors, will tend to cluster. Further, under extremely poor propagation conditions in the channel (as may occur in a meteor-burst channel[6] when there is no suitable meteor-trail), a retransmission technique keeps repeating the same cycle until conditions in the channel improve. This is clearly an information rate of zero. From this point of view, a retransmission becomes an intermittent type system.

2.5 Forward Error Correction

In the retransmission technique discussed above, the detection of the errors is generally accomplished by short error detecting codes and a feedback link is used to request the retransmission of the erroneous messages. Instead of asking for retransmissions, the digits in error could be corrected at the receiver by suitable error-correcting codes. It may seem that the latter is not really 'matching' the variations in the forward channel to a transmission parameter. However, as different types of codes and associated techniques such as interleaving exist for use in different types of channels, it is reasonable to select, whenever possible, a code suited to a given channel. This is, in effect, a kind of 'matching technique'.

Coding has become a large branch of Information Theory and hence no attempt will be made in this thesis to review the results in this field. Instead, some of the practical limitations of codes and some methods which have been used to overcome these difficulties are discussed. The following two paragraphs discuss the limitations of forward error correction under fading conditions.

By definition an e-error-correcting code will correct all the error patterns in a code word with e or less erroneous digits. Further, the code may correct some of the error patterns with say q, $q > e$, erroneous digits but without any guarantee. The higher q is the greater the probability of decoding error. Under fading conditions, the code words transmitted during the fades will generally contain a considerable number of erroneous digits. Clearly, in order to maintain a reasonably low probability of error, the error-correcting code will need to be very long. As seen in the following paragraph, this is not desirable. In the presence of fading, it is generally preferable to design for error detection with retransmission.

In addition, it is much simpler to implement error-detection schemes than error-correction schemes. Error detection for any parity checking code can be realised by a shift register decoder. Error correction, involving logical operations among sets of parity checks, is much more complex, and one of the major problems in the practical implementation of this scheme is that of finding codes with reasonably inexpensive decoders. Relatively few error-correcting codes that possess easily implementable decoders are known.

An important difference exists between random-error and burst-error correcting codes. Random-error correcting codes work well only for a channel, like the Gaussian channel, over which errors occur at random, i.e. there is no statistical dependence between the errors. Measurements over practical radio circuits[16], however, show that in general errors occur not at random but in bursts, i.e. in clusters. The bursts occur because, as stated in Chapter 1, some channels are characterised by fading or impulse interference as well as additive noise.

A technique, particularly useful for fading signals, is to use interleaving with random-error correcting codes[22,17]. This technique is described in Chapter 4.

2.6 A Summary of Disadvantages of Existing Variable-Parameter Systems

A brief summary of the disadvantages of existing variable-parameter systems discussed in the present Chapter is given in this Section. This is not an attempt to 'write-off' these techniques. Indeed, each technique, as has already been seen, has its advantages. However, it is very important to be aware of the disadvantages of each technique. Even though the new techniques analysed in the later part of the thesis share some disadvantages with the existing techniques, it will be seen from the discussions below that the former techniques have the advantage of overcoming some of the difficulties encountered in the established techniques.

In variable-parameter systems, such as variable-rate, re-transmission and automatic power control systems, it is absolutely necessary to have a reliable feedback link. This disadvantage should be borne in mind for all these systems in the discussion below.

The intermittent systems require the continuous use of a pilot tone channel for the prediction of fades. Further, when the feedback channel is noisy, the intermittent systems ideally require the use of another channel over which is passed a signal to confirm whether or not transmission is taking place. This channel also needs to be extremely reliable as will be seen in Chapters 5 and 7. The additional disadvantage of the intermittent system is therefore the fact that it requires the use of far too many separate channels. This clearly increases the complexity of the system. For the non-intermittent variable-rate systems, the pilot tone channel can be removed and prediction of fades formed from the received waveforms, that is, from the fading data. Further, the use of fading data to predict the channel conditions would make it possible to relax the flat-flat conditions that were imposed on the fading model and thus increase the number of physical channels that are handled.

As was discussed in Section 2.2, the automatic power control has the disadvantage of not using all the available power except for relatively short periods of time. In retransmission error control, apart from the disadvantage already given, the through-put rate will be seen in Chapter 4 to fall to extremely low values for the low SNRs. The new techniques analysed in the thesis have, therefore, the advantage over the retransmission techniques in that they each work at a fixed average rate.

Forward error-correction techniques suffer from the fact that very long burst-error correcting codes are required to handle burst of error that occur under fading conditions. It is extremely difficult, and often impossible, to implement very long burst-error correcting codes.

The other forward error control technique for fading channels, already mentioned, is interleaving. It will be seen in Chapter 4 that one disadvantage of the interleaving process is that the memory in the channel due to fading is destroyed and is, therefore, not exploited by the decoding scheme. In the case of long bursts, as may occur under fading conditions, a very high interleaving factor is necessary. This leads to another disadvantage in that considerable delay is inevitable as all the interleaved code words must be received before error correction can be effected.

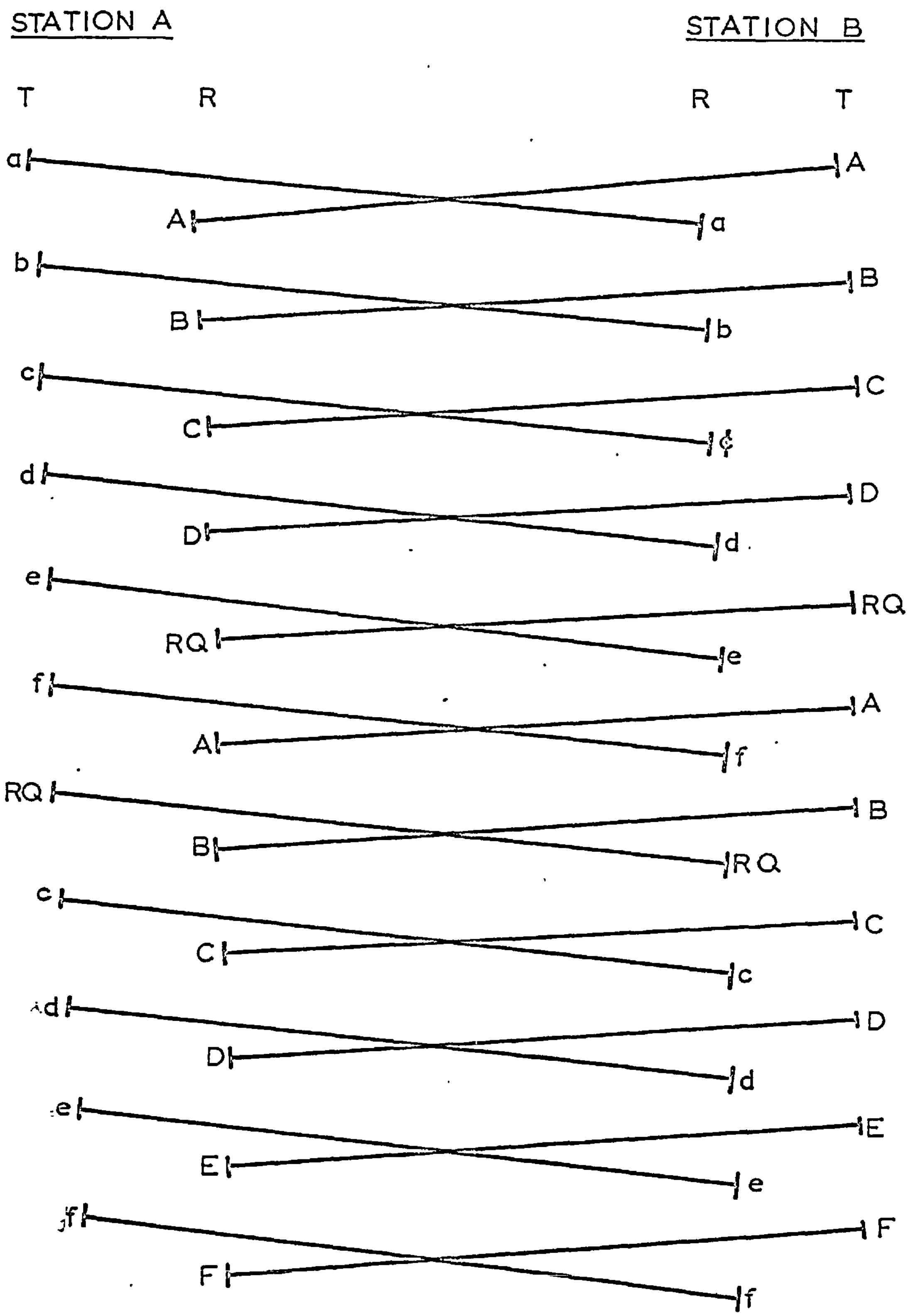


Fig. 2.1. A 4-Character Repetition Cycle of the ARQ System.

3. The Detection of Fading and Non-Fading Signals in Additive Gaussian Noise

3.1 Introduction

The data transmission system models used in this Chapter are shown in Figs. 3.1 and 3.2. In the first model[3], the medium (channel) coupling the transmitter and the receiver is assumed to add stationary white-zero-mean Gaussian noise, but is otherwise distortionless, i.e. there is no multiplicative distortion in the channel. In Fig. 3.2, the channel model is the same as that given in Fig. 1.2 and described in Section 1.2.2. Thus flat-flat Rayleigh fading is assumed.

The problem investigated in this Chapter is that of finding the optimum (minimum error rate) receiver structure and calculating the actual error probability for the optimum receiver. The problem is simplified by transforming it into one involving geometric considerations based on the Khahunen-Loeve expansion rule. This expansion rule states (Ref. 25, pp. 96-99) that a nonperiodic process x(t) may be represented over an interval a to b by the expansion.

$$x(t) = \lim_{N\to\infty} \sum_{j=1}^{N} \lambda_j \quad \phi_j(t)$$

where $$\int_a^b \phi_j(t)\ \phi_k(t)\ dt = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

and $$\lambda_j = \int_b^a x(t)\ \phi_j(t)\ dt$$

and where ϕ_j are real or complex numbers.

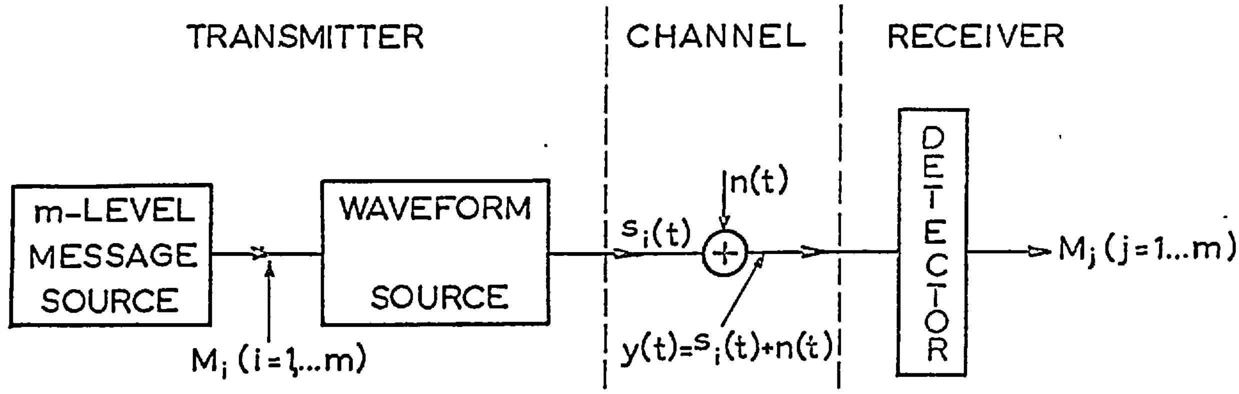


Fig. 3.1. A Data Transmission System Using a Nonfading Channel.

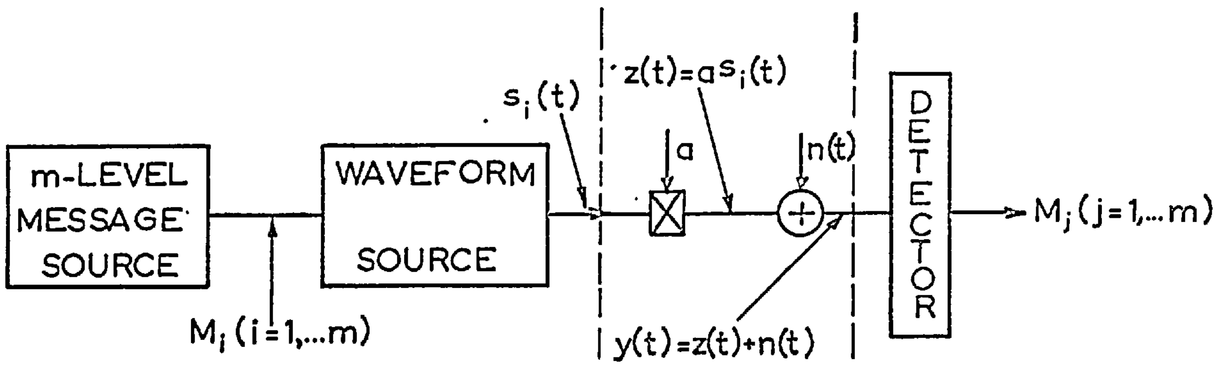


Fig. 3.2. A Data Transmission System Using a Fading Channel.

The set of waveforms $\phi_j(t)$ are known as an orthonormal set. The number N is referred to as the dimension of the signal alphabet x(t) and it depends on the interval (a, b) and the bandwidth of the signal x(t).

Each member $s_i(t)$ of the set of m transmitted waveforms can, therefore, be expressed, using Khahunen-Loeve expansion rule, as a linear combination of N orthonormal waveforms, $\phi_1(t), \phi_2(t) \ldots\ldots \phi_n(t)$, i.e.

$$s_i(t) = \sum_{j=1}^{N} \lambda_{ij} \phi_j(t) \qquad i = 1, \ldots, m$$

where $$\lambda_{ij} = \int_o^T s_i(t) \phi_j(t)\, dt$$

and $$\int_o^T \phi_j(t) \phi_n(t)\, dt = \begin{cases} 1 & j = k \\ o & j \neq k \end{cases}$$

From this expansion it is apparent that each signal waveform may actually be specified uniquely in terms of the coefficients of $\phi_j(t)$, j = 1, ..., N . Thus, the set $s_i(t)$, i = 1, ... m, can be represented by a set of N-tuples $(\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{iN})$. By using a conceptual extension of 2- and 3-dimensional Euclidean space, the numbers, λ_{i1}, λ_{i2}, ..., λ_{iN}, can be thought of as the N co-ordinate projections of the signal, when the signal is regarded as a point in an N-dimensional space.

For example, for N = 3, the point corresponding to the waveform

$$s_i(t) = \lambda_{i1}\ \phi_1(t) + \lambda_{i2}\ \phi_2(t) + \lambda_{i3}\ \phi_3(t),$$

can be plotted on a 3-dimensional Euclidean space as shown in Fig. 3.3.

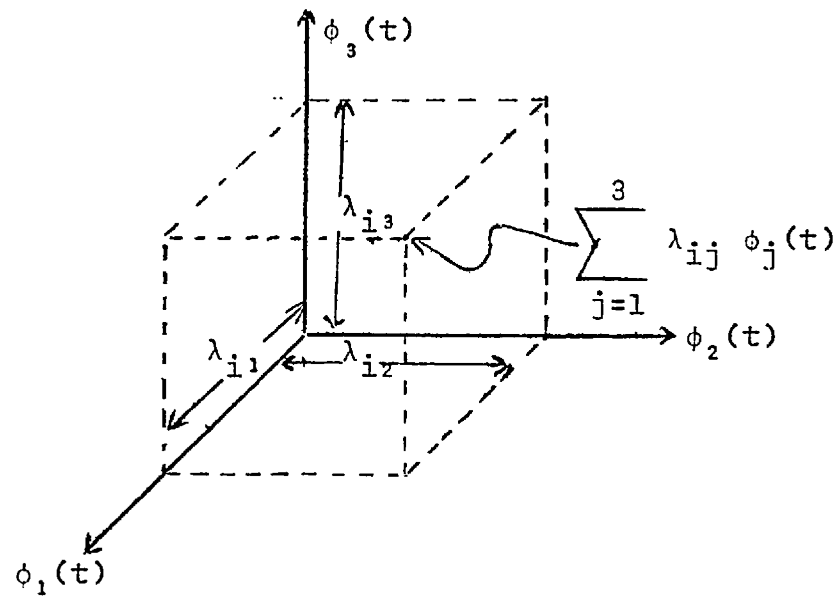


Fig. 3.3 Geometrical Representation of a Signal Waveform in a 3-Dimensional Euclidean Space

It is not difficult to show [3] that the energy content, E_i, in $s_i(t)$ is equal to the square of the distance of the signal from the origin, i.e. $E_i = d^2(s_i, o)$, where $d(s_i,o)$ is the distance of $s_i(t)$ from the origin in the N-dimensional

Euclidean space. This demonstrates the point that simple but important relationships between the signal parameters and the distance-angle configuration of the N-dimensional space can be derived.

In geometric terms, the detection problem can be seen to be one of determining the coefficients λ_{ij}. Once these coefficients are known, the original waveforms, which they determine, are also known.

In practice, the decision problem is complicated by the fact that the transmitted signal is corrupted by interference which is a random process. In the absence of multiplicative distortion, the interference is assumed to be purely additive, stationary zero-mean Gaussian noise with double-sided spectral power density, N_o.

The additive noise n(t) falling within the bandwidth of the receiver can also be expanded by the Khahunen-Loeve rule, i.e.

$$n(t) = \lim_{N\to\infty} \sum_{j=1}^{N} n_{ij} \phi_j(t)$$

Since the noise is purely additive, the received signal is $z(t) = s_i(t) + n(t)$. It can be represented by a point in a Euclidean space of appropriate dimension. The co-ordinator of this point can be determined by a series of product integrators which evaluate the sum of λ_{ij} and n_{ij}. Each coordinate is thus composed of two parts - a deterministic part λ_{ij} due to the transmitted signal and n_{ij} due to the noise which has been superimposed on the signal by the channel. The decision making stage in the receiver has to use the noise perturbed signal point in order to infer, i.e. guess, which actual signal point was transmitted.

For coherent detection (i.e., the phase of the transmitted signal is known at the receiver) and assuming equally likely messages, a detector which selects as transmitted signal that signal which is closest to the received 'noisy' point minimises the probability of error [92,61,74,3] and is, therefore, the optimum detector. It is known as the maximum likelihood detector. With incoherent detection, no provision is made to phase synchronise the receiver. With such a detection scheme, the channel is assumed to add a random phase, uniformly distributed between 0 and 2π, to the phase of the transmitted signal. The probability of error for the incoherent detection is averaged over the phase added by the channel in order to obtain the average probability of error.

For transmission systems in which fading occurs, it is not always possible to find the optimum detector. However, in the case where the fading is flat-flat Rayleigh fading (Fig. 1.2), it is possible to derive the optimum receiver.

In this case a set of appropriately weighted matched filters* constitute the optimum receiver[74, 80].

For the flat-flat fading channel, the probability of error depends on the channel gain, a, which is a random variable. The <u>average</u> probability of error can be found by integrating the probability of error (as a function of 'a') over all possible values of the channel gain.

Expressions are derived in the following two sections for the probability of error for an incoherent FSK system and a coherent ASK system. These expressions are derived in detail because they will be required in Chapter 6 and Chapter 7; where incoherent FSK and coherent ASK signalling techniques are used as examples of the application of the variable-duration and variable-level methods of varying the data rate.

Table 3.1 gives a summary of the formulae for probability of error for three basic data transmission systems, namely, m-level coherent and incoherent FSK, ASK and PSK systems when there is no fading in the channel. Table 3.2 summarises the formulae of the probability of error for these systems for the fading case. The formulae for the probability error, in the absence of fading, for coherent FSK, incoherent ASK, coherent and incoherent PSK have been quoted from Arthurs and Dym[3].

---

*A matched filter is simply a special product integrator with an impulse response which is the time inverse of the waveform to which it is matched.

3.2 The Detection of Steady-Level Signals in Additive Gaussian Noise

3.2.1. Probability of Error for the m-Tone FSK System with Incoherent Detection

In an m-tone FSK system, the keying is over m different frequencies. With reference to Fig. 3.1, if the ith message waveform is transmitted, the received signal z(t) will be

$$z(t) = \sqrt{\frac{2E}{T}} \cos(\omega_i t + \alpha) + n(t) \qquad \dots \quad 3.1$$

With incoherent detection, α, an unknown angle, is taken to be a random variable uniformly distributed between 0 and 2π.

On expanding Equation 3.1, the received signal z(t) is

$$z(t) = \sqrt{\frac{2E}{T}}\cos \omega_i(t)\cos\alpha - \sqrt{\frac{2E}{T}} \sin \omega_i t \sin \alpha + n(t) \qquad \dots \quad 3.1a$$

The received waveform z(t) can be expanded, by the use of the Khahunen-Loeve expansion rule described in Section 3.1. If this is done with

$$\phi_{xj} = \sqrt{\frac{2}{T}} \cos \omega_j t$$

and

$$\phi_{yj} = \sqrt{\frac{2}{T}} \sin \omega_j t$$

then

$$\lambda_{xj} = x_j = \int_0^T z(t)\cos\omega_j(t) = \begin{cases} n_{xj} & j \neq i \\ E\cos\alpha + n_{xi} & j = i \end{cases}$$

$$\lambda_{yj} = y_j = \int_0^T z(t)\ \sin\omega_j t = \begin{cases} n_{yj} & j \neq i \\ -\ E\sin\alpha + n_{yj} & j = i \end{cases}$$

The value of $z(t)$ used in the above integrals is taken from Equation 3.1a. The variables n_{xj} and n_{yj} are the noise perturbations falling within the signal space. They are independent random variables with zero mean and variance N_o.

If $s_i(t)$ is transmitted, then as Arthurs and Dym (Ref. 3, pp. 371-372) show a correct decision is made by the detector if the inequality

$$\sqrt{x_j^2 + y_j^2} < \sqrt{x_i^2 + y_i^2} \qquad \ldots\ 3.4$$

holds for all $j \neq i$.

Given that $\alpha = A$, the two-dimensional joint conditional probability of x_i and y_i will be Gaussian with x_i having a mean of E Cos$A$ and y_i having a mean of -E Sin A. In polar coordinates, the two-dimensional joint conditional Gaussian probability density function is (Ref. 25, pp. 147-149).

$$p(x_i y_i \mid \alpha = A) = \frac{1}{2\pi N_o} \exp\left\{\frac{1}{2N_o}\left[(x_i - \sqrt{E}\cos A)^2 + (y_i + \sqrt{E}\sin A)^2\right]\right\} \qquad \ldots\ 3.5$$

Averaging over all possible values of A

$$p(x_i,y_i) = \int_A p(x_i,y_i/A)\ P(A)dA$$

Now, $P(A) = \frac{1}{2\pi}$

Hence,

$$p(x_i y_i) = \int_0^{2\pi} p_i(x_i, y_i | A) \frac{dA}{2\pi}$$

$$= \frac{1}{2\pi N_o} \exp.\left\{ -\frac{1}{2N_o}\left[x_i^2 + y_i^2 + E\right] \right\}$$

$$\cdot \int_0^{2\pi} \exp\left\{ \frac{x_i\sqrt{E}\cos A}{N_o} - \frac{y_i\sqrt{E}\sin A}{N_o} \right\} \frac{dA}{2\pi}$$

$$= \frac{1}{2\pi N_o} \exp\left\{ -\frac{1}{2N_o}\left[x_i^2 + y_i^2 + E\right] \right\} \cdot I_o\left(\frac{\sqrt{(x_i^2 + y_i^2) E}}{N_o} \right) \qquad \ldots \quad 3.6$$

where,

$$I_o(z) = \int_0^{2\pi} e^{z\cos x} \, dx$$

is the modified Bessel function of the first kind and zeroth order.

If a change of variables operation is performed by using the substitutions

$$x_i = v_i \sqrt{N_o} \cos \phi_i$$

$$y_i = v_i \sqrt{N_o} \sin \phi_i$$

then the joint probability density function of v_i and ϕ_i becomes

$$q(v_i, \phi_i) = \frac{v_i}{2\pi} \exp\left\{ -\frac{1}{2}\left[v_i^2 + \frac{E}{N_o}\right] \right\} \cdot I_o\left(v_i \sqrt{\frac{E}{N_o}}\right) \qquad \ldots \quad 3.7$$

The probability density function $q(v_i)$ can be obtained by integrating Equation 3.7 over all possible values of ϕ_i, that is,

$$q(v_i) = \int_0^{2\pi} q(v_i, \phi_i) \, d\phi_i$$

$$= v_i \exp \left\{ -\frac{1}{2} \left[v_i^2 + \frac{E}{N_o} \right] \right\} \cdot I_o\left(v_i \sqrt{\frac{E}{N_o}}\right) \quad \ldots \quad 3.8$$

The density function in Equation 3.6 was derived assuming $j = i$. If $i \neq j$ then the associated density function $P(x_j, y_j)$ can be derived from Equation 3.6 by putting $E = 0$, (see Equations 3.2 and 3.3).

Thus

$$q(v_j) = v_j e^{-\frac{1}{2} v_j^2} \quad \ldots \quad 3.9$$

Now, since $x_i = v_i \sqrt{N_o} \cos \phi_i$ and $y_i = v_j \sqrt{N_o} \sin \phi_j$ it follows that inequality 3.2 is the same as

$$v_j < v_i \quad \ldots \quad 3.10$$

Inequality 3.10 leads to a correct detection, assuming $s_i(t)$ is transmitted. As v_j are independent random variables the probability of 3.10 being satisfied simultaneously for all j, $j \neq i$, is simply

$$P(v_j < v_i, \text{ all } j \neq i | v_i) = \prod_i P(v_j < v_i | v_i) \quad \ldots \quad 3.11$$

For any particular $j \neq i$ the probability that $v_j < v_i$ is

$$P\left[v_j < v_i | v_i\right] = \int_0^{v_i} v_j e^{-\frac{1}{2}v_j^2} \, dv_j$$

$$= 1 - e^{-\frac{1}{2}v_i^2} \qquad \ldots \quad 3.12$$

Thus Equation 3.11 becomes

$$P\left[v_j < v_i, \text{ all } j \neq i | v_i\right] = \left(1 - e^{-\frac{1}{2}v_i^2}\right)^{n-1}$$

$$\ldots \quad 3.13$$

The probability of making a correct decision when $s_i(t)$ is transmitted is obtained by averaging Equation 3.11 over all v_i, that is, the probability deciding on $s_i(t)$ is

$$P\left[v_j < v_i \text{ all } j \neq i\right] = \int_0^{\infty} P\left[v_j < v_i \text{ all } j \neq i | v_i\right] q(v_i) dv_i$$

$$= \int_0^{\infty} v_i \exp\left\{-\frac{1}{2}\left[v_i^2 + \frac{E}{N_o}\right]\right\} \cdot I_o\left(v_i \frac{E}{N_o}\right) \cdot \left[1 - e^{-v_i^2}\right]^{n-1} dv_i$$

$$\ldots \quad 3.14$$

By the bionomial expansion

$$\left[1 - e^{-\frac{1}{2}v_i^2}\right]^{n-1} = \sum_{k=o}^{m-1} {}^{(m-1)}C_k \; (-1)^k \; e^{-\frac{1}{2}v_i^{2}k}$$

$$\ldots \quad 3.15$$

Thus Equation 3.14 can be written as

$$P\left[v_j < v_i, \text{ all } j \neq i\right]$$

$$= \exp\left\{-\frac{E}{2N_o}\right\} \sum_{k=o}^{n-1} {}^{(n-1)}C_k \; (-1)^k \cdot \int_0^{\infty} \exp\left\{-\frac{v_i^2 (k+1)}{2}\right\} \cdot I_o\left(v_i \frac{E}{N_o}\right)$$

$$\ldots \quad 3.16$$

The integral on the right-hand side is a standard integral and its solution can be found in a table of integrals[37]. It is given by

$$\int_0^{\infty} v_i \exp\left\{-\frac{v_i^2(k+1)}{2}\right\} \cdot I_o\left(v_i \frac{E}{N_o}\right) dv_i$$

$$= \frac{1}{k+1} \exp \frac{E}{2N_o(k+1)} \qquad \ldots \quad 3.17$$

Therefore the probability of correct detection, P_c, when $s_i(t)$ is transmitted is,

$$P_c = P\left[v_j < v_i, \text{ all } j \neq i\right]$$

$$= \exp\left\{-\frac{E}{2N_o}\right\} \sum_{k=o}^{m-1} {}^{(m-1)}C_k \, (-1)^k \cdot \frac{\exp \frac{E}{2N_o(k+1)}}{k+1}$$

$$\ldots \quad 3.18$$

Since the right-hand side of equation 3.18 is independent of i, the average probability of error, P_e, for an orthogonal m-tone FSK system using incoherent detection is given by

$$P_e = 1 - P_c$$

$$= 1 - \exp\left\{-\frac{E}{2N_o}\right\} \cdot \sum_{k=o}^{m-1} {}^{(m-1)}C_k \, (-1)^k \cdot \frac{\exp \frac{E}{2N_o(k+1)}}{k+1}$$

$$= -\exp\left\{-\frac{E}{2N_o}\right\} \cdot \sum_{k=o}^{m-1} {}^{(m-1)}C_k (-1)^k \cdot \frac{\exp \frac{E}{2N_o(k+1)}}{k+1}$$

$$\ldots \quad 3.19$$

Now,

$$ {}^{(n-1)}C_k \cdot \frac{1}{k+1} = \frac{(n-1)!}{(n-1-k)!\,k!} \cdot \frac{1}{k+1} $$

$$ = \frac{m!}{m(n-k-1)!\,(k+1)!} = \frac{1}{n} \cdot {}^{m}C_{(k+1)} \qquad \ldots \quad 3.20 $$

On putting $r = k + 1$, Equation 3.19 becomes

$$ P_e = \frac{\exp\left\{-\frac{E}{4N_o}\right\}}{n} \cdot \sum_{r=2}^{m-1} {}^{m}C_r \,(-1)^r \cdot \exp \frac{E(2-r)}{4N_o r} \qquad \ldots \quad 3.21 $$

For the binary case, $n = 2$ and the probability of error, $P_{e,b}$, can be found from Equation 3.21 by putting $n = 2$. If this is done, it is found that

$$ P_{e,b} = \frac{1}{2} e^{-\frac{E}{4N_o}} = \frac{1}{2} e^{-\frac{1}{2}u} \qquad \ldots \quad 3.22 $$

where $u = \frac{E}{2N_o}$ is the detection SNR

3.2.1. Probability of Error for n-Amplitude ASK System with Coherent Detection

In the ASK modulation scheme the information is carried in the amplitude of the transmitted waveform. A set of ASK message waveforms can be described by the equations

$$ s_i(t) = \begin{cases} \sqrt{\frac{2E_i}{T}} \cos \omega_o t & o < t < T,\ i = 1, \ldots, m \\ o & \text{elsewhere} \end{cases} $$

where E_i is the energy content of $s_i(t)$ and $\omega_o = 2\pi f_o/T$ is the carrier frequency.

Each transmitted waveform can be expanded, by the use of the Khahunen-Loeve expansion rule described in Section 3.1. If this is done with

$$\phi_1(t) = \sqrt{\frac{2}{T}}\cos\omega_o t$$

then,

$$\lambda_{i1} = \int_0^T \sqrt{\frac{2E_i}{T}}\cos\omega_o t \cdot \sqrt{\frac{2}{T}}\cos\omega_o t = \sqrt{E_i}$$

The possible transmitted signal points can be represented geometrically on a straight line. This is shown in Fig. 3.4 for the case when n = 3.

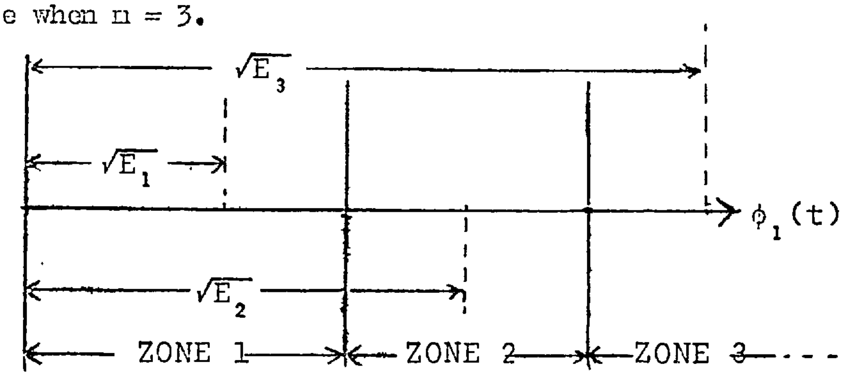


Fig. 3.4 The Detection Signal Space for 3-Level ASK System

In this detection scheme, a received signal which lies in detection zone 2 (see Fig. 3.4) is chosen as s_2 rather than s_1 or s_3. Received signals in Zones '1' and '3' are assumed to be s_1 and s_3 respectively.

A received signal is a point on the $\phi_1(t)$ coordinate. Let it be assumed that it lies at a point x on $\phi_1(t)$, then,

$$x = \int_0^T z(t)\ \phi_1(t)\ dt = \sqrt{E_1} + n$$

Clearly x is a random Gaussian distributed variable with mean $\sqrt{E_i}$ and variance N_o. Thus, the probability density function of x is given by

$$p(x) = \frac{1}{\sqrt{2\pi N_o}} \exp\left[- (x - \sqrt{E_i})^2 / 2N_o\right] \qquad \ldots \quad 3.23$$

If it is assumed for convenience, that the transmitted signal points are uniformly spaced, and that $E_1 = o$, then E_i can be expressed as

$$\sqrt{E_i} = (i - 1)\Delta \qquad \ldots \quad 3.24$$

The detection signal space is shown in Fig. 3.5.

If $i < 1 < m$ and if Equation 3.25 is used in conjunction with Figs. 3.5 and 3.6, then the probability of incorrectly detecting the signal s_i is

$$P \notin R_i/s_i = \frac{2}{\sqrt{2\pi N_o}} \int_{\Delta/2}^{\infty} e^{-\frac{x^2}{2N_o}}\, dx \qquad \ldots . \; .3.25$$

where R_i corresponds to the ith zone in the detection signal space.

Using the change of variable. $y = \frac{x}{\sqrt{N_o}}$

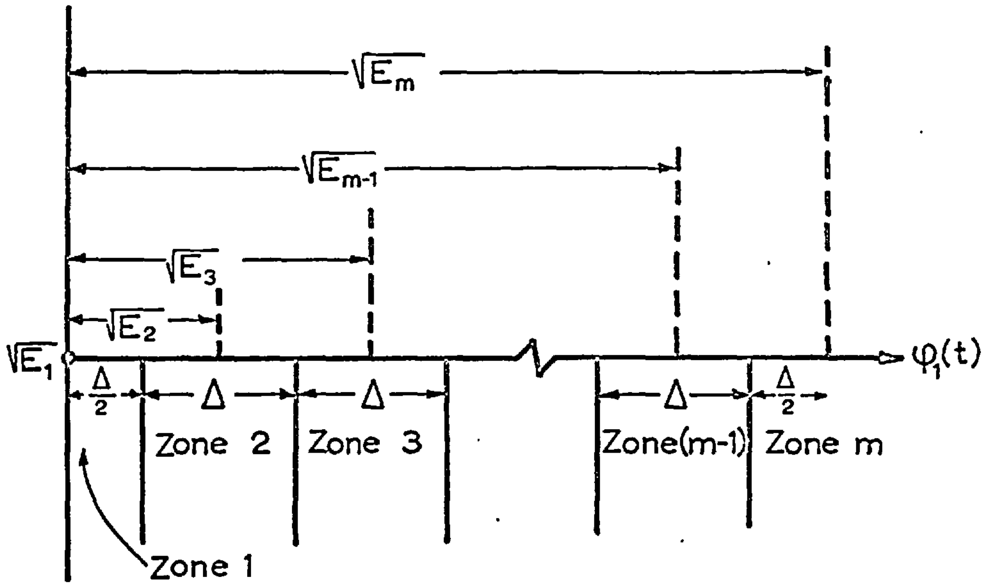


Fig. 3.5. The Detection Signal Space for m-level ASK System

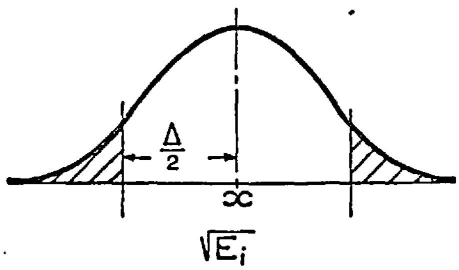


Fig. 3.6. The Probability Dentity Function of Received Signal Point.

Equation 3.25 becomes

$$P\left[z \notin R_i/s_i\right] = \frac{2}{\sqrt{2\pi}} \int_{\frac{\Delta}{2\sqrt{N_o}}}^{\infty} e^{\frac{-y^2}{2}} \, dy \qquad \ldots 3.26$$

Similarly, if i = 1 or i = m, the probability of incorrectly detecting an error is

$$P\left[z \notin R_1 \middle| s_1\right] = P\left[z \notin R_n \middle| s_n\right]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{\Delta}{2\sqrt{N_o}}}^{\infty} e^{-\frac{y^2}{2}} \, dy \qquad \ldots 3.27$$

The average probability of error for any i is

$$P_e = \sum_i P(s_i) \quad P\left[z \notin R_i/s_i\right]$$

and for equally likely messages,

$$P_e = \frac{1}{n} \sum_{i=1}^{n} P\left[z \notin R_i/s_i\right] \qquad \ldots \quad 3.28$$

Thus

$$P_e = \frac{2}{n\sqrt{2\pi}} \int_{\frac{\Delta}{2\sqrt{N_o}}}^{\infty} e^{-\frac{y^2}{2}} \, dy + \frac{2(n-2)}{n\sqrt{2\pi}} \int_{\frac{\Delta}{2\sqrt{N_o}}}^{\infty} e^{\frac{-y^2}{2}} \, dy$$

$$= \frac{2(m-1)}{m\sqrt{2\pi}} \int_{\frac{\Delta}{2\sqrt{N_o}}}^{\infty} e^{-\frac{y^2}{2}} dy \qquad \ldots \quad 3.29$$

If an average power limitation of $\frac{E}{T}$ watts is placed on the transmitter, then

$$\sum_{i=1}^{m} \left(\frac{E_i}{T}\right) \frac{1}{m} = \frac{E}{T} \qquad \ldots \quad 3.30$$

By using Equation 3.24 and the identity

$$\sum_{i=1}^{m} (i-1)^2 = \sum_{j=2}^{m-1} j^2 = \frac{m(m-1)(2m-1)}{6}$$

in Equation 3.30, it can be shown that

$$\Delta = \sqrt{\frac{6E}{(m-1)(2m-1)}} \qquad \ldots \quad 3.31$$

Thus,

$$P_e = \frac{2(m-1)}{m} \frac{1}{\sqrt{2\pi}} \int_{\sqrt{6E/4N_o(m-1)(2m-1)}}^{\infty} e^{\frac{-y^2}{2}} dy$$

$$= \frac{2(m-1)}{m}\left[1 - 2\phi\left(\sqrt{6E/4N_o(m-1)(2m-1)}\right)\right] \qquad \ldots \quad 3.32$$

where

$$\phi(y) = \frac{1}{\sqrt{2\pi}} \int_{o}^{y} e^{\frac{-t^2}{2}} dt \qquad \ldots \quad 3.33$$

For the binary case m = 2, and the probability of error, $P_{e,b}$, is

$$P_{e,b} = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\frac{E}{2N_o}}}^{\infty} e^{\frac{-y^2}{2}} \, dy$$

$$= \frac{1}{2} \left[1 - 2\phi \, (\sqrt{u}) \right] \qquad \ldots \; 3.34$$

where, as before, $u = \frac{E}{2N_o}$, in the detection SNR

In Table 3.1, the results that have been derived above are given. In addition, the expression, for the probability of error for coherent and incoherent ASK, and coherent and incoherent PSK are given. These latter results are taken from Arthurs and Dym[3].

The function $\phi(y)$ is an alternative definition of the error function which is normally defined as

$$\text{erf}x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, dt \qquad \ldots \; 3.35$$

The function ϕ is defined by Equation 3.33.

From Equations 3.33 and 3.35, the following relationships between erf and ϕ can be shown to hold.

$$\phi(x) = \frac{1}{2} \, \text{erf} \, \frac{x}{\sqrt{2}} \qquad \ldots \; 3.36$$

$$\text{erf}x = 2\phi \left(\sqrt{2}x \right) \qquad \ldots \; 3.37$$

DETECTION / MODULATION	COHERENT	INCOHERENT
F S K	$\frac{1}{2}\left[1-2\phi(\sqrt{u})\right] \leqslant P_e \leqslant \frac{1}{2}(n-1)\left[1-2\phi(\sqrt{u})\right],$ for $n > 2$	$P_e = \frac{\exp\left[-\frac{u}{2}\right]}{n}\sum_{r=2}^{n} {}^{n}C_r\,(-1)^r \exp\frac{u(2-r)}{2r}$
A S K	$P_e = \frac{n-1}{n}\left[1 - 2\,\phi\left(\sqrt{\frac{6u}{2(n-1)(2n-1)}}\right)\right]$	$\frac{1}{n}\exp\left\{\frac{-6u}{4(n-1)(2n-1)}\right\} < P_e < \exp\left\{-\frac{6u}{4(n-1)(2n-1)}\right\}$
P S K	$\frac{1}{2}\left[1-2\phi\left(\sqrt{2u}\sin\frac{\pi}{n}\right)\right] < P_e < \left[1-2\phi\left(\sqrt{2u}\sin\frac{\pi}{n}\right)\right]$ for $n > 2$	$P_e \approx \left[1-2\phi\left(\sqrt{2u}\sin\frac{\pi}{2n}\right)\right]$

Table 3.1 Formulae for the Probabilities of Error for Non-fading Signals Detected in Additive Gaussin Noise

3.3. Error Probabilities in the Presence of Idealised Flat-Flat Rayleigh Fading

As was shown in the previous section, if the received signal has a non-fading amplitude then the element probability of error, p_e, is a function of u. The presence of a random fading process gives rise to a variation in the received SNR and this can be expressed in terms of a probability density function f(u). Under these fading conditions, the average probability of error is

$$\overline{P}_e = E\left[P_e(u)\right] \qquad \ldots\ 3.38$$

where the expectation, E, is taken over the probability density function of u.

Equation 3.38 can be expressed in integral form as

$$\overline{P}_e = \int_0^{\infty} P_e(u)\, f(u)\, du \qquad \ldots\ 3.39$$

where $P_e(u)$ is the probability of error for a given SNR, u, and f(u) is the probability density function of u.

Using the integral in Equation 3.39 and the expressions for the element probability of error given in Table 3.1, formulae for the average Rayleigh fading are derived below. It is noted from an examination of Table 3.1 that $P_e(u)$ for each of the six signalling techniques can be expressed in terms of one of two generalized functions,

$$\left[1 - 2\phi\ (\alpha\sqrt{u})\right]$$

and

$$\exp\ (-\ \beta u)$$

In these generalised expressions α and β take different values for the various signalling schemes. For the derivations of the average probability of error for the six signalling techniques, the solutions of the following two integrals are needed.

$$I_1 = \int_0^{\infty} \left[1 - 2\,\phi\ (\alpha\sqrt{u})\right]\ f(u)\ du \qquad \ldots\ 3.40$$

$$I_2 = \int_0^{\infty} \exp\ (-\ \beta u)\ \ f\ (u)\ du \qquad \ldots\ 3.41$$

It is shown in Appendix I that when the amplitude of the signal has Rayleigh fading characteristics, then the probability density function, $f(u)$, of the SNR is given by

$$f(u) = \frac{1}{U_o}\ e^{-u/U_o}$$

where U_o is the average detection SNR.

Thus, for $f(u)$ of this form,

$$I_1 = \int_0^{\infty} \left[1 - 2\,\phi\ \ (\alpha\sqrt{u})\right]\ \frac{1}{U_o}.\ \ e^{-u/U_o}\ \ du$$

$$= \int_0^{\infty} \frac{1}{U_o} e^{-u/U_o} du - 2\int_0^{\infty} \phi(\alpha\sqrt{u}) \frac{1}{U_o} e^{-u/U_o} du$$

The first part of I, is directly integrable and the second can be integrated by parts. Thus,

$$I_1 = 1 - 2\left\{\left[-e^{-u/U_o} (\alpha\sqrt{u})\right]_0^{\infty} + \int_0^{\infty} e^{-u/U_o} \frac{d}{du} \phi(\alpha\sqrt{u})du\right\}$$

$$= 1 - 2\left\{0 + \int_0^{\infty} e^{-u/U_o} \cdot \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{\alpha}{\sqrt{u}} \cdot e^{\frac{-\alpha^2 u}{2}} du\right\}$$

$$= 1 - \frac{\alpha}{\sqrt{2\pi}} \int_0^{\infty} \frac{1}{\sqrt{u}} e^{-u\left[\frac{1}{U_o} + \frac{\alpha^2}{2}\right]} du$$

Let $K = \frac{1}{U_o} + \frac{\alpha^2}{2} = \frac{2 + \alpha^2 U_o}{2U_o}$

and $2Ku = y^2$

$\therefore \quad du = \frac{2\sqrt{u}}{2K} \cdot dy$

Thus,

$$I_1 = 1 - \frac{2}{\sqrt{2K}} \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-y^2/2} dy$$

$$= 1 - \frac{\alpha}{\sqrt{2K}}$$

Substituting back for K

$$I_1 = 1 - \sqrt{\frac{\alpha^2 U_o}{2 + \alpha^2 U_o}} \qquad \ldots \quad 3.42$$

Now the second integral is

$$I_1 = \int_0^{\infty} \exp\left[-\beta u\right] \frac{1}{U_o} e^{-u/U_o} du \qquad \ldots \quad 3.43$$

$$= \int_0^{\infty} \frac{1}{U_o} e^{-u\left[\beta + \frac{1}{U_o}\right]} du$$

$$= \int_0^{\infty} \frac{1}{U_o} e^{-u\left[\frac{\beta U_o + 1}{U_o}\right]} du$$

$$= \left[-\frac{1}{1 + \beta U_o} e^{-u\left(\frac{\beta U_o + 1}{U_o}\right)}\right]_0^{\infty}$$

$$= \frac{1}{1 + \beta U_o} \qquad \ldots \quad 3.44$$

Equations 3.42 and 3.44 can be used in conjunction with Table 3.1 to find the explicit expressions for the average probability of error for the six signalling techniques in the presence of flat-flat Rayleigh fading and additive Gaussian noise. Two examples are given below to illustrate this.

Example 1

For the coherent ASK case

$$P_e(u) = \frac{m - 1}{m}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m-1)(2m-1)}}\right)\right]$$

Therefore,

$$\alpha = \sqrt{\frac{6}{2(m-1)(2m-1)}}$$

Using this expression in equation 3.42, the average probability of error is seen to be

$$P_e = \frac{m-1}{m}\left[1 - \sqrt{\frac{6U_o}{4(m-1)(2m-1) + 6U_o}}\right]$$

$$= \frac{n-1}{n}\left[1 - \sqrt{\frac{3U_o}{2(m-1)(2m-1) + 6U_o}}\right] \qquad \ldots \; 3.45$$

<u>Example 2</u>

Consider the incoherent FSK case, for which

$$P_e(u) = \frac{\exp - \frac{u}{2}}{n} \sum_{r=2}^{n} {}^{m}C_r \, (-1)^r \exp\left\{ \frac{u(2-r)}{2r} \right\}$$

This expression can be re-written as

$$P_e(u) = \frac{1}{n} \sum_{r=2}^{n} {}^{m}C_r \, (-1)^r \exp \left\{ - \frac{u(r-1)}{r} \right\}$$

$$\therefore \quad \beta = \frac{r-1}{r}$$

On using this expression in Equation 3.44, the average probability of error is

$$\overline{P}_e = \frac{1}{n} \sum_{r=2}^{n} {}^{m}C_r \quad (-1)^r \frac{1}{1 + U_o(1-\frac{1}{r})} \qquad \ldots \; 3.46$$

All the other values of the average probability of error, $\overline{P}_e$, in Table 3.2 are similarly evaluated.

Using Equations 3.21, 3.33, 3.45 and 3.46, the performance curves of the m-tone incoherent FSK and the m-amplitude coherent ASK systems were evaluated for fading and non-fading conditions. These curves are given in Figs. 3.7 and 3.8 and discussed in the following Section.

3.4 The effect of Flat-flat Rayleigh Fading on the Error Probabilities of the m-Tone Incoherent FSK and the m-Amplitude ASK Systems.

It is seen from Figs.3.7 and 3.8 that in the absence of fading, the probabilities of error of the m-tone incoherent FSK and the m-amplitude coherent ASK systems decrease exponentially with increasing detection SNR. However, as seen from the same Figures, in the presence of flat-flat Rayleigh fading, the probabilities of error for both types of systems decrease only linearly with increasing detection SNR.

The performance curves of both types of systems have been evaluated for m = 2, 4, 8, 16 and 32. It is seen from Fig .3.8 that in the absence of fading, increasing m decreases the error probability for the m-tone incoherent FSK system whereas in the presence of flat-flat Rayleigh fading, increasing m increases the error probability. Thus for this system, the higher m is, the worse is the effect on fading on the probability of error.

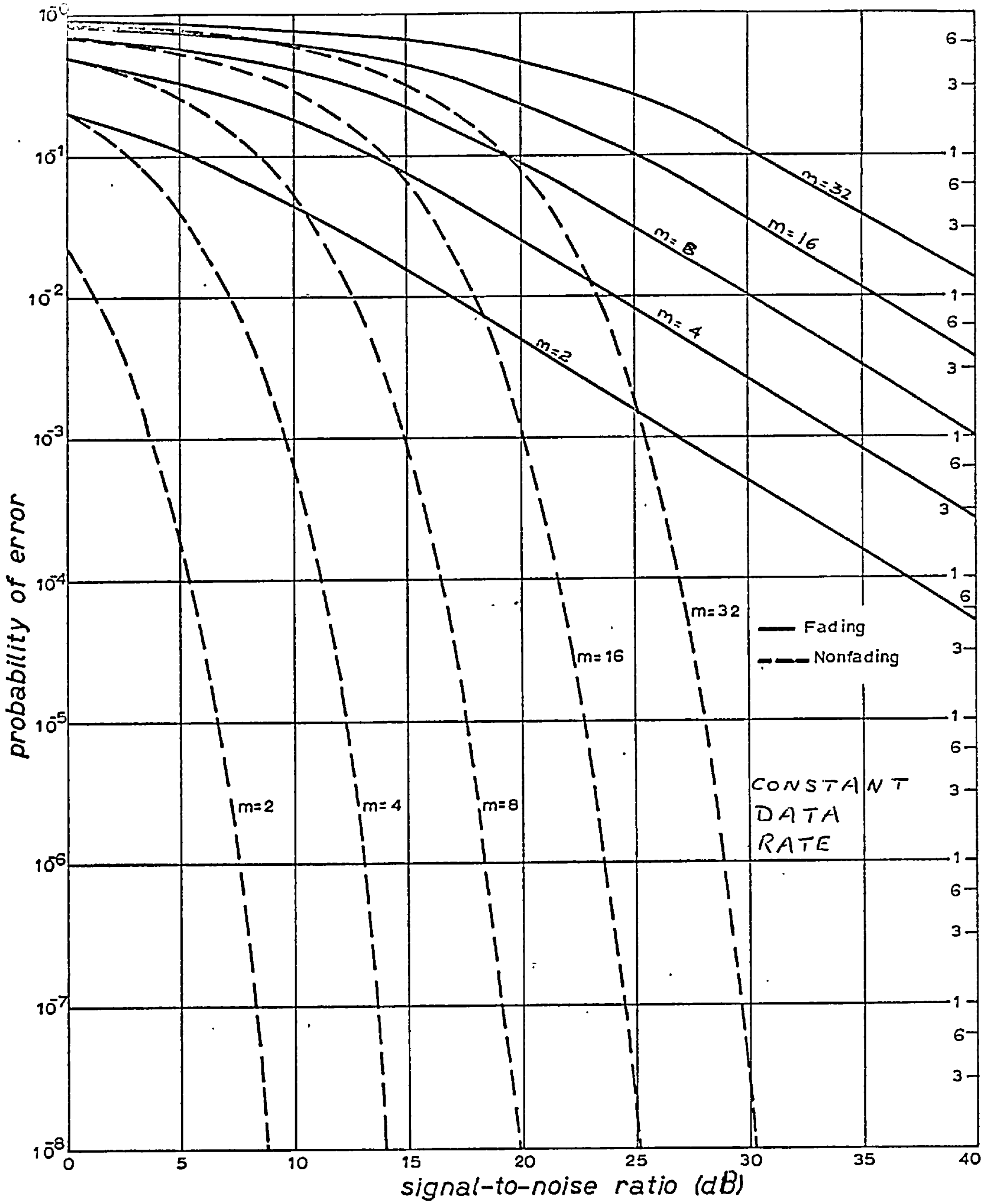


Fig. 3.7. Performance of m-Amplitude ASK Systems under Flat flat Rayleigh Fading and Nonfading Conditions; m = 2,4,8,16 & 32

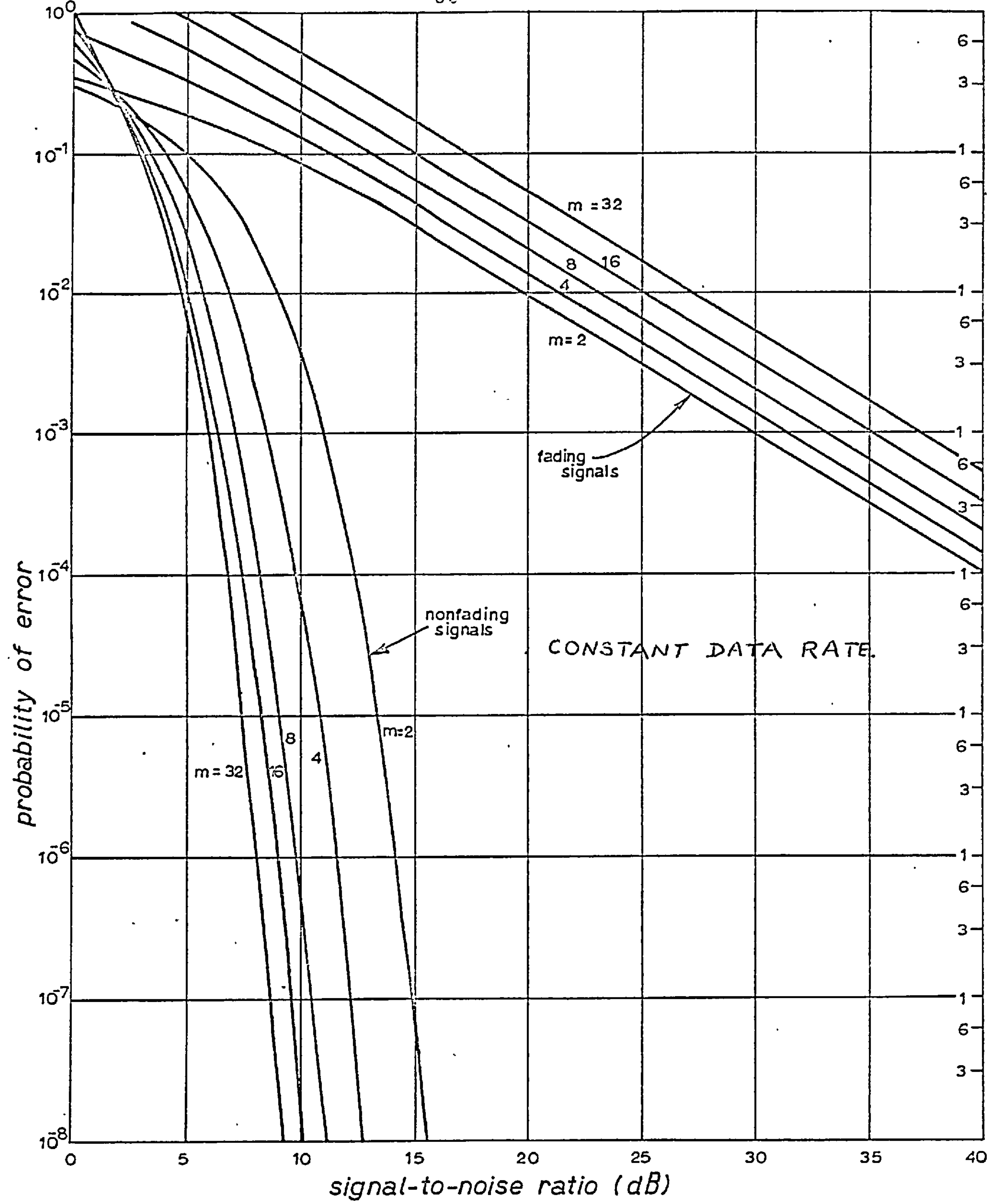


Fig. 3.8. Performance of m Tune FSK Systems under Flat-flat Rayleigh Fading and Nonfading Conditions; m=2,4,8,16&32

DETECTION \ MODULATION	COHERENT	INCOHERENT
F.S.K.	$\frac{1}{2}\left[1-\sqrt{\frac{U_o}{2+U_o}}\right] < \bar{P}_e < \frac{m-1}{2}\left[1-\sqrt{\frac{U_o}{2+U_o}}\right]$ for $m > 2$	$\bar{P}_e = \frac{1}{m}\sum_{r=2}^{m} {}^{m}C_r (-1)^r \frac{1}{1+U_o(1-\frac{1}{r})}$
A.S.K.	$\bar{P}_e = \frac{m-1}{m}\left[1-\sqrt{\frac{3U_o}{2(m-1)(2m-1)+3U_o}}\right]$	$\frac{2(m-1)(2m-1)}{2m(m-1)(2m-1)+3mU_o} < \bar{P}_e < \frac{2(m-1)(2m-1)}{2(m-1)(2m-1)+3U_o}$
P.S.K.	$\frac{1}{2}\left[1-\sqrt{\frac{U_o \sin^2\frac{\pi}{m}}{1+U_o\sin^2\frac{\pi}{m}}}\right] < P_e < \left[1-\sqrt{\frac{U_o \sin^2\frac{\pi}{m}}{1+U_o\sin^2\frac{\pi}{m}}}\right]$ for $m > 2$	$\bar{P}_e \approx \left[1-\sqrt{\frac{U_o \sin^2\frac{\pi}{2m}}{1+U_o\sin^2\frac{\pi}{2m}}}\right]$

Table 3.2 Formulae for the Probabilities of Error in the Presence of Time- and Frequency-Flat Rayleigh Fading

It is seen from Fig. 3.7 that in the absence of fading and in the presence of flat-flat Rayleigh fading, increasing m increases the error probability for the m-amplitude coherent ASK system. However, for this system, as can be seen from Fig. 3.7, the higher m is, the smaller is the decrease in the error probability due to fading.

4. ANALYSIS OF SOME EXISTING VARIABLE-RATE SCHEMES

Descriptions of some of the techniques used at present to combat the detrimental effects of fading have already been given in Chapter 2. In particular, the methods which alter one or more transmission parameters according to the state of the forward channel were stressed. Most of these, in effect, tend to match, often indirectly, the transmission rate to conditions prevailing in the channel. For example, retransmission error control adjusts indirectly the rate of transfer of information to suit propagation conditions in the forward channel.

In the first section of the present Chapter, the van Duuren ARQ system [85] is analysed. Formulae for the probability of error and for the information rate for this system are derived and put in a form suitable for comparison with the new variable rate techniques analysed in the next four Chapters. The comparison is given in Chapter 8.

In Section 4.2, a description and a simplified analysis of the 'AN/GSC-10 (KATHRYN) Variable Rate Modem' [94, 51] are given. In the last Section, some forward error control techniques are discussed. In particular the performance of an interleaved (23,12) Golay code is given. Although forward error control techniques have been used over several important fading channels [17, 16, 22] they will not be analysed in this thesis. An exception is the example given in Section 4.3 on the use of interleaving with a random-error correcting code. In addition there is the possibility of combining forward error correcting codes with the new techniques discussed in the latter Chapters of the thesis. The objective would be to try to obtain more effective error control by using a hybrid technique, i.e. a combination of several techniques. Such possibilities will be discussed a little further in Chapter 10.

4.1 The ARQ System [85, 45]

The van Duuren ARQ system [85] which uses a 3-out-of-7 constant ratio code for the detection of errors has been effectively employed over long-distance HF radio circuits for telegraph message transmissions. It has been found to provide adequate protection (Ref.45 pp. 187-200) on what would other- be unreliable circuits.

Strictly, the ARQ system works over a circuit transmitting information in both directions (duplex working). A description of the ARQ system operation was given in Chapter 2. From symmetry considerations, it is evident that only one half of the duplex circuit need be considered. In this form, a discrete model of the communication channel, Fig. 4.1, with a feedback link gives a good representation of the system. The channel is in general a binary asymmetric channel. For the flat-flat Rayleigh fading assumed in the thesis, the errors are independent of whether previous transmissions result in errors or incorrect detection.

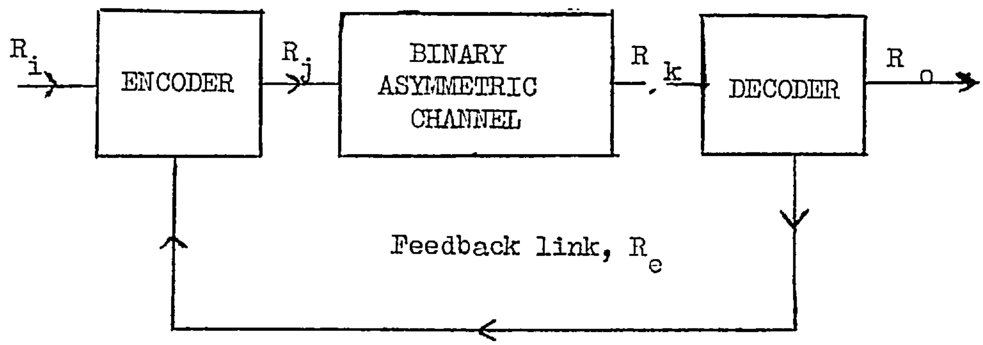


Fig. 4.1 A Discrete Communication Channel

In the van Duuren system messages are encoded in the form of code words of length n = 7. Out of 128 possible sequences only those with three 'ones' are used as information carrying sequences. There are thus 35 code words. An over-bound, R_o, to the transmission rate efficiency of this code is thus,

$$R_o \leq \frac{1}{7} \log_e 35 = 0.733 \qquad \ldots \quad 4.1$$

Now, it is evident that all patterns of 1, 3, 5, 7 errors will be detected since the number of 'ones' in the received sequence will not be three. However, not all of the patterns 2, 4 or 6 errors will be detected. The undetected error occurrences are referred to as transpositions. The simplest transposition occurs when a 'mark' and a 'space' are each incorrectly detected. In this case, the number of 'ones' remains three and the pair of errors passes undetected. Note, however, that if both errors had occurred on either 'mark' or 'space' only, then the error pattern would be detected. Similarly, four and six error occurrences can be split to yield higher order transpositions.

Let $p_n(j)$ be the probability that exactly j errors occur in an n-bit sequence. Only the undetected error sequences (transpositions) are of interest since for detected error patterns repetitions are requested as many times as necessary. The total probability of transposition P_t can be written as

$$P_t = \sum_j p_n(j)\ p(\text{transposition}|j) \quad \ldots \quad \ldots \quad 4.2$$

For the 3-out-of-7 constant ratio code, transpositions occur only for j = 2, 4 or 6. Thus

$$P_t = \sum_{j=2,4,6} p_n(j)\ p(\text{transposition}|j)$$

$$= p_7(2)\ p(\text{transposition}|2)$$

$$+ p_7(4)\ p(\text{transposition}|4)$$

$$+ p_7(6)\ p(\text{transposition}|6) \quad \ldots \quad 4.3$$

Now, the above conditional probabilities are

$$p(\text{transposition}|2) = \frac{4}{7}$$

$$p(\text{transposition}|4) = \frac{18}{35} \quad \ldots \quad 4.4$$

$$p(\text{transposition}|6) = \frac{4}{7}$$

whence 4.3 becomes

$$p_t = \frac{4}{7}\,p_7(2) + \frac{18}{35}\,p_7(4) + \frac{4}{7}\,p_7(6) \quad \ldots \quad 4.5$$

For independent errors $p_n(j)$ is given by the bionomial formula:

$$p_n(j) = {}_nC_j\, p^j\, (1 - p)^{n-j} \quad \ldots\ldots \quad 4.6$$

where p = the element probability of error

On substituting 4.6 into 4.5 it is found that

$$p_t = 12p^2(1 - p)^s + 18p^4(1 - p)^3 + 4p^6(1 - p) \quad \ldots \quad 4.7$$

Equation 4.7 is the same as equation 8 of van Duuren 85 with p instead of α..

Following van Duuren, the average transmission speed, R_o, is calculated by considering the number repetition cycles. The total number of characters in the repetition cycles is $(1 - P_o)$ and the number of the repetition cycles is $(1 - P_o)/4$, since there are four characters in a repetition cycle.

A repetition can start only when the character that is in error is outside a repetition cycle (probability P_o) or if the character in error is the last of a repetition cycle. A repetition cycle starts in the first case if an error is detected at the receiving terminal. This happens (for noiseless feedback) with probability $1 - p_t - p_c$, where

$$p_c = (1 - p)^7 \qquad \ldots \quad 4.8$$

is the probability of correctly detecting a digit.

If the character in error is the last in a repetition cycle, then (again for noiseless feedback) another repetition cycle starts with probability $1 - p_c(p_c + p_t)$

As there is no other way in which a repetition cycle can start, the following identity holds [85]

$$1 - (p_c + p_t) + \frac{1 - P_o}{4}\left[1 - p_c(p_c + p_t)\right]$$

$$= \frac{1 - P_o}{4} \qquad \ldots \quad 4.9$$

From which,

$$P_o = \frac{p_c\,(p_c + p_t)}{4 - (4 - p_c)(p_c + p_t)} \qquad \ldots \quad 4.10$$

Similarly, the probability of undetected errors for a repetition cycle of 4 can be shown [85] to be

$$P_e = \frac{p_c p_t}{4 - (4 - p_c)(p_c + p_t)} \qquad \ldots \quad 4.11$$

Hence, the number of correctly received characters, R_{oc}, is given by

$$R_{oc} = \frac{p_c^{\,2}}{4 - (4 - p_c)(p_c + p_t)} \qquad \ldots \quad 4.12$$

From equation 4.11 it is seen that when $p_c + p_t = 1$, then $p_e = p_t$. Since $p_c + p_t \leq 1$, it follows that the output probability of error for the ARQ system is at best equal to (i.e. overbounded by) the probability of undetected transpositions p_t.

Since the transmission efficiency of the 3-out-of-7 constant ratio code is at most 0.733 (equation 4.1), equation 4.10 becomes

$$P_o \leq 0.733 \frac{p_c(p_c + p_t)}{4 - (4 - p_c)(p_c + p_t)} \qquad \ldots \quad 4.13$$

In Fig. 4.2, the probability of error, P_e, and the average transmission rate, R_o, for the ARQ system are plotted as functions of the detection SNR for an incoherent FSK system. From this Fig. it is seen that the average rate, R_o, falls to a very low value for low detection SNRs, i.e. a high element error rate. Clearly for a through-put rate as low as 10^{-3} the system is virtually at a standstill and the fact that probability of error is very low is of little practical significance. In an attempt to find a compromise between the output error rate, it is proposed that the average rate be adjusted for low SNRs so that it is always greater than some fixed number α, a fraction of the limit the rate approaches for high SNRs. The fraction could, for example, be chosen to be $\frac{1}{10}$. The effect of this constraint on the average rate R_o is now determined.

The condition imposed on equation 4.13 results in the following

$$\alpha \nless \frac{P_o}{0.733} \leq \frac{p_c \; (p_c + p_t)}{4 - (4 - p_c)(p_c + p_t)} \quad ,\ldots \quad 4.14$$

It is evident that the probability of correct detection, p_c, is not altered in any way by the restriction on the rate of transmission. The constraint, however, places a limit on the number of permissible repetitions during any one cycle.

Some detected errors will thus be printed out in the output, preferably as special symbols (for example, as erasures) to indicate that the word was detected to be in error. This, in effect, divides the detected errors into two categories, those that result in correct detection after a fixed number, or less, of permitted repetitions and those that are eventually printed out with the message. To incorporate the number of detected errors that do not result in successful repetitions into the transposition, p_t, is altered to $p_{t\alpha}$ by using inequality 4.14. Thus

$$p_{t_\alpha} = \frac{p_c \left[p_c(1 - \alpha) + 4\alpha\right] - 4\alpha}{p_c(\alpha - 1) - 4\alpha} \quad \ldots \; 4.15.$$

The value of $p_{t\alpha}$ given by 4.15 guarantees the average rate will be greater than α. The procedure of calculating the probability of error is as follows. Equations 4.13 and 4.11 are used to calculate the average transmission rate and the probability of error respectively. If the average transmission rate is found to be less than α, equation 4.15 is used to find $p_{t\alpha}$ and the probability of error is now calculated using $p_{t\alpha}$ instead of p_t in equation 4.11.

Curves showing the performance of the ARQ technique under both modes of operation are plotted in Fig. 4.2.

4.2 The AN/GSC-10 (KATHRYN) Variable Rate Modem

4.2.1 Introduction

The AN/GSC-10 system[94, 51] (Fig. 4.3) is composed of transmit-receive terminal equipments that are still in the experimental stages. The modem has been built for use over HF-radio circuits. The attempt in this system is to combine two radically different techniques that have been developed in the past. The first technique uses a modem which provides a high data rate and efficient operation under favourable conditions. The second technique employs a system specifically designed to provide reliable communication in spite of heavy multipath (severe fading). An example of the latter system is the 'RAKE' receiver[68] developed at the Lincoln Laboratory. It should be noted that though the 'RAKE' receiver provides reliable communication under poor propagation conditions, its inflexibly low data rate is wasteful of bandwidth under more favourable conditions.

The AN/GSC-10 systems has been designed to operate at

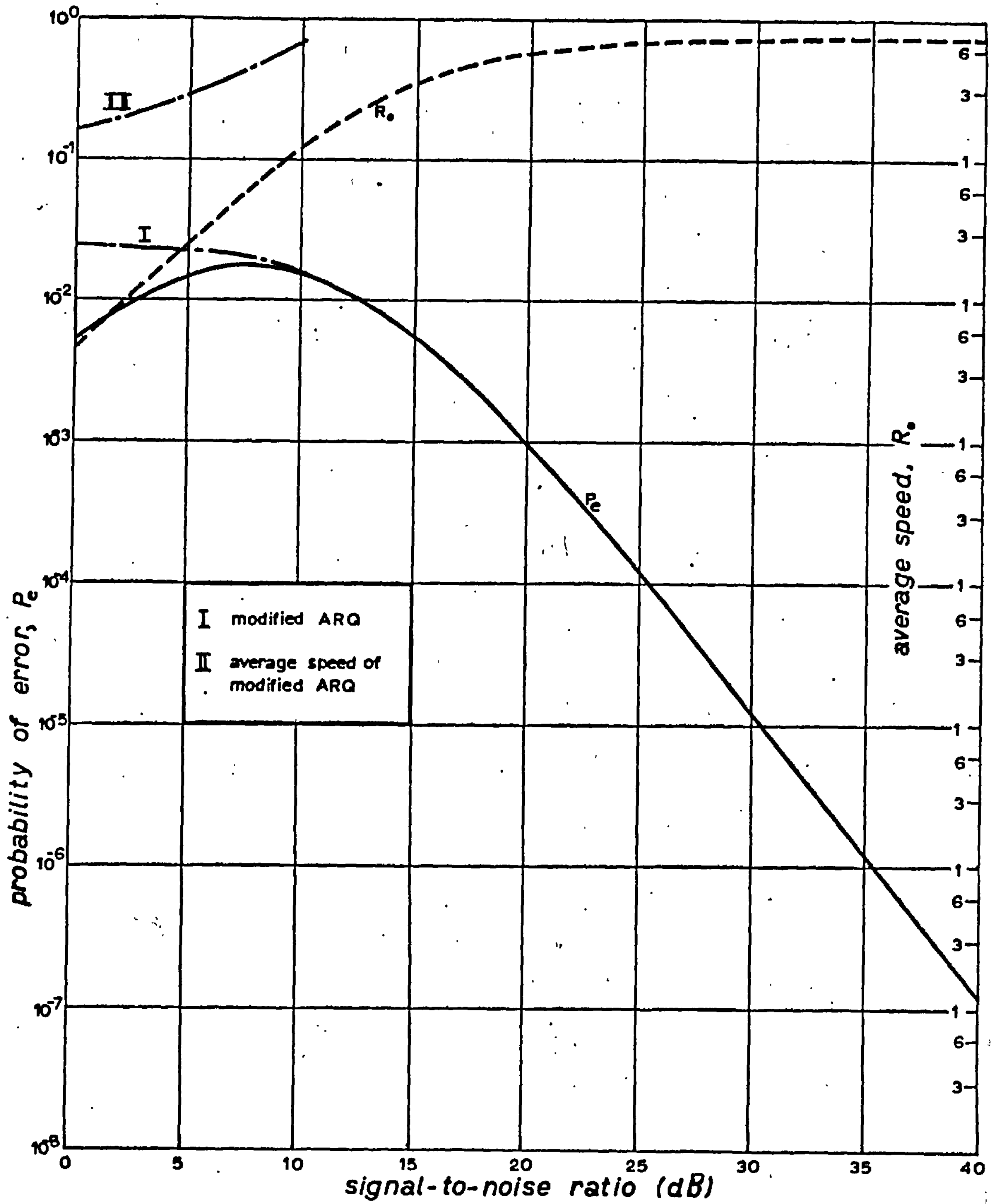
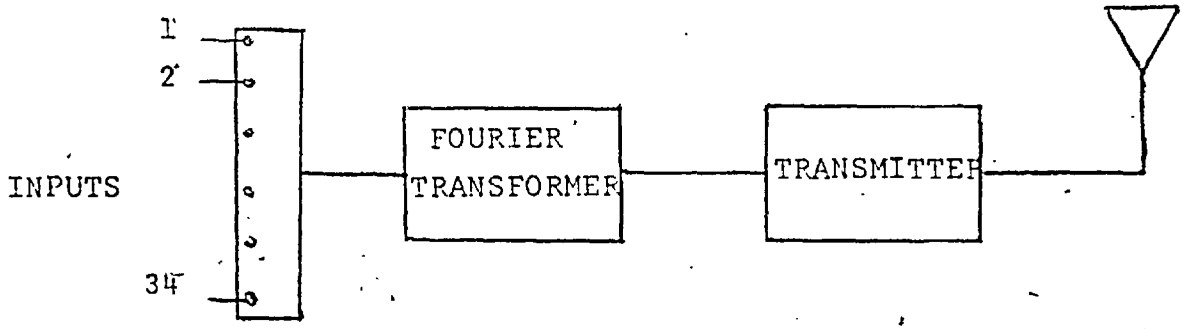
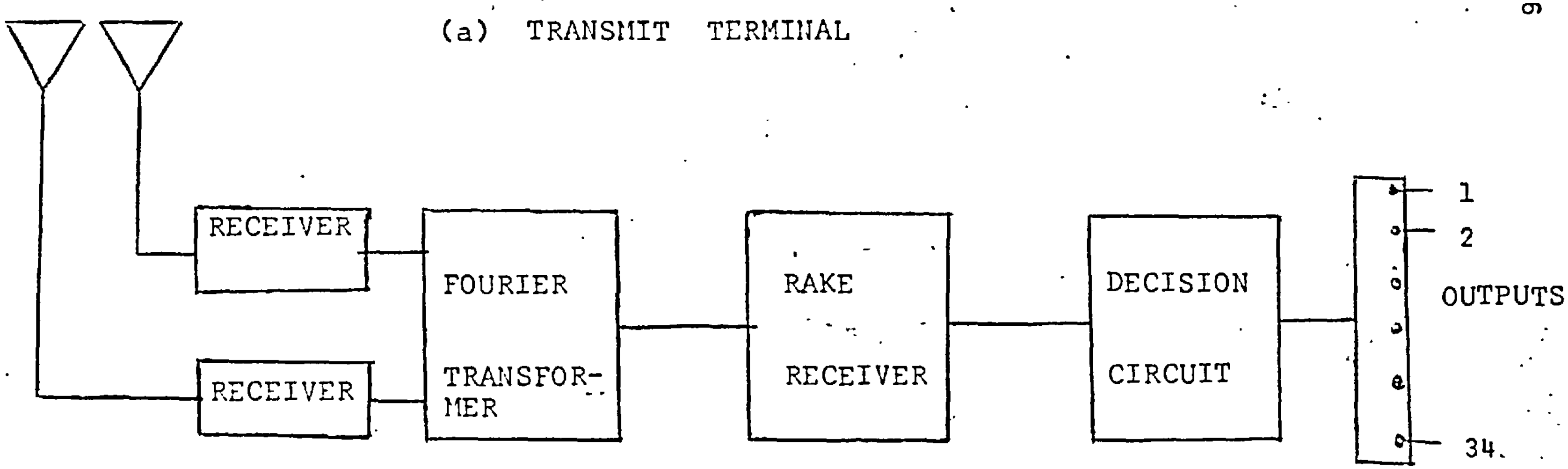


Fig. 4.2. The Performance of the ARQ System and the Modified ARQ System



(a) TRANSMIT TERMINAL



(b) RECEIVE TERMINAL

FIG. 4.3 A SIMPLIFIED DIAGRAM FOR THE AN/GSC-10 SYSTEM various levels of redundancy so as to provide a trade-off between the operating data rate and the reliability for a given transmission channel capacity.

The transmitted signal of the AN/GSC-10 system is contained in a nominal 3 - kHz channel bandwidth and consists of 34 PSK subcarriers spaced by 81.4Hz. All the 34 subcarriers are keyed simultaneously at a rate of 75 Hz; the duration of each keying frame is, therefore $13\frac{1}{3}$ms. At the receiver, the integration interval utilized is approximately 12.2 ms, the reciprocal of the subchannel spacing. The 1-ms guard band allows for errors in the frame at the receiver.

The phase and amplitude of each subcarrier is determined by three components as shown in Fig. 4.4. The I_1 phase component is the basic information bit associated with a particular frame

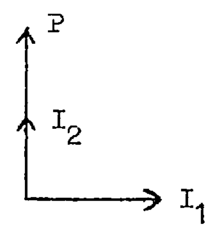


Fig. 4.4 Phasor Diagram for the AN/GSC-10 Signal

and subcarrier. The I_2 component is a redundant duplicate of I_1 component and is 17 tones away. The pilot component P is used for channel phase and amplitude measurements. The channel is thus 'probed' at each subchannel frequency in the system band, The P vector forms the basis of the channel measurement carried out for each of the 34 subchannels. The measured phase is used to phase correct each component; the measured amplitude is used to implement maximal ratio combining; and the final bit decision based on the (maximal-ratio weighted, phase corrected) sum of all of the appropriate components. At the full data rate there are four components for each bit (I_1 and I_2 for each of the space diversity receivers). This is equivalent to 4th-order diversity. At lower data rates more components are included but the same combining procedure is employed.

The changing of the data rate, accomplished by altering the number of parallel channels used for the transmission of the same bit, is done manually in the current model of the AN/GSC-10 modem. The adjustment to the data rate is made on the evidence of either the reliability of the received data or on the evidence provided by a real-time visual display of the ionospheric spectual response.

A more detailed description of the AN/GSC-10 (KATHRYN) variable rate data modem is given by Zimmerman and Kirsch[94].

4.2.2. A Theoretical Basis for the 'KATHRYN' Modem

Fully coherent PSK is used in the 'KATHRYN' modem at all the data rates. Thus, in the absence of fading or with flat-flat Rayleigh fading, the formulae presented in Chapter 3 for probabilities of error for m-phase coherent PSK signalling apply to the 'KATHRYN' modem when m is set equal to two. In this Section, the probability of error for a fixed-rate system transmitting continuously is denoted by $P_{e0}(u)$ in the absence of fading and by $\overline{P}_{e1}$ in the presence of flat-flat Rayleigh fading. The probability of error for two independently flat-flat Rayleigh fading channel, when maximal ratio combining is used, is denoted by $\overline{P}_{e2}$.

On substituting m = 2 into the formulae given in Tables 3.1 and 3.2 for the performance of m-phase coherent PSK signalling, the following are obtained as the theoretical performance of the 'KATHRYN' modem,

$$\overline{P}_{e0}(u) = \frac{1}{2}\left[1 - 2\,\varphi(\sqrt{2u})\right] \qquad \ldots \quad 4.16$$

$$\overline{P}_{e1} = \frac{1}{2}\left[1 - \frac{1}{\sqrt{1 + \frac{1}{u_o}}}\right] \qquad \ldots \quad 4.17$$

In the case of two independently fading channels, maximal-ratio combining based on channel measurements is used in the 'KATHRYN' modem. Assuming perfect measurements, the error probability for this case is

$$\overline{P}_{e2} = \int_0^{\infty} \frac{u}{u_o^{\,2}} \, e^{-u/U_o} \, \frac{1}{2}\left[1 - 2\varphi(\sqrt{2u})\right] du \qquad \ldots \quad 4.18$$

Integrating equation 4.18 by parts successively yields,

$$\overline{P}_{e2} = \frac{1}{2}\left[1 - \frac{1}{1 + \frac{1}{U_o}} - \frac{1}{2U_o\left(1 + \frac{1}{U_o}\right)}\right] \quad \ldots 4.19$$

The theoretical curves for the 'KATHRYN' system when it is operating at its maximum data rate are plotted in Fig.4.5.

4.2.3 Discussion of the AN/GSC-10 System

A description of the field tests carried out for the AN/GSC-10 (KATHRYN) system has been given by Kirsch, at al [51]. The main conclusions from the tests are summarized below. An assessment of the 'KATHRYN' system is then given.

From the results of these field tests, Kirsh, et al., reported a widespread of the probability of error for the 'KATHRYN' system. In particular, the probability of error was found to range from 10^{-4} to 10^{-6} even if SNRs above 40 dB were considered. Despite this spread in the probability of error no clear dependence on SNR was observed for SNR greater than 20 dB.

On doubling the redundancy, the theoretical improvement obtainable for flat-flat Rayleigh fading is an increase of 3 dB in the SNR. The field test results, however, also indicated a wide spread in the actual improvement due to redundancy. In general, the data collected by Kirsh, et al., for the 'KATHRYN' system fell reasonably well within the bounds suggested by the

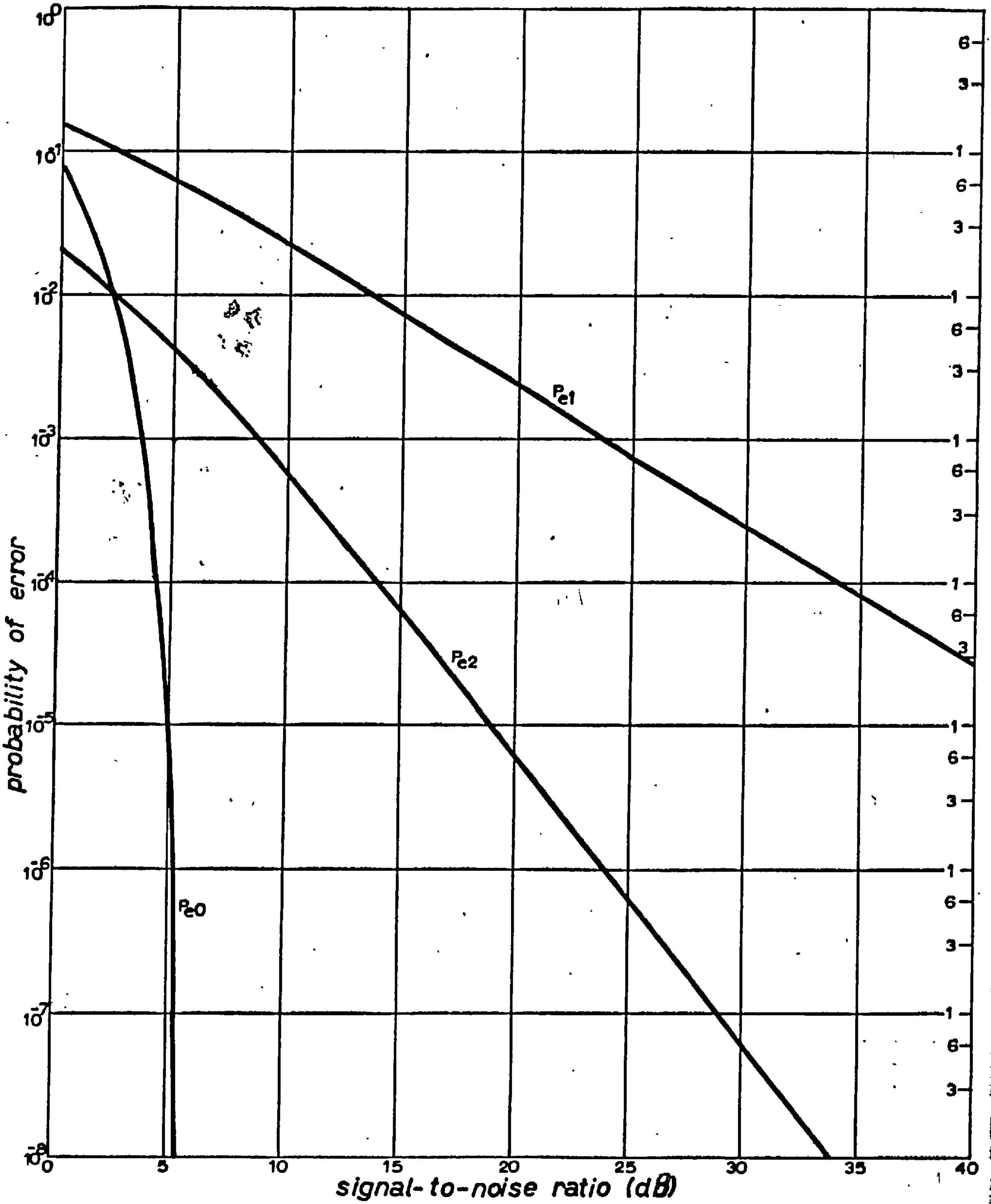


Fig. 4.5. Theoretical Performance of the AN/GSC-10 (KATHRYN) System.

simplified flat-flat Rayleigh fading model.

Some disadvantages of the AN/GSC-10 system are similar to those of diversified systems (See Section 1.3.1), since space diversity is utilised in the 'KATHRYN' system. The data rate is varied in the AN/GSC-10 system by varying the redundancy. In the current model of this system, there are 34 parallel channels, but only half of these channels are used for transmitting distinct signals because second-order space diversity is utilized. When used in its most redundant mode, the AN/GSC-10 system has a redundancy factor of 16. Because of bandwidth restrictions, the higher the redundancy factor (and hence the larger the theoretical improvement factor should be) the shorter is the frame length (i.e. symbol duration). (In the present system, the duration of the symbol is 12.2 ms.) Increasing the redundancy factor, decreases the symbol duration and, thus, leads to errors in the phase and amplitude measurements of the channel. It is clear that the improvement due to redundancy saturates after a given number of divisions of the available bandwidth to obtain parallel channels.

4.3 Discussion of Coding techniques Used Over Fading Circuits

4.3.1. Introduction

Coding can be used with or without the use of feedback. Generally, when a feedback link is available, a short code is used to detect errors and requests for retransmissions are made.

For example the ARQ system analysed in Section 4.1 uses a fixed ratio (or fixed count) code with retransmission.

As was stated in the introduction to this Chapter, forward error control (coding) techniques will not be analysed in the thesis. However, a simple example on the use of interleaving with a random-error correcting code is given in the next Section. As described in that Section, the interleaving process aims at spreading out the error clusters that occur over fading circuits. For the purposes of illustrating the effectiveness of interleaving, the performance of a (23,12) Golay Code is given at the end of the following Section. Some of the promising coding techniques for use over fading circuits that have been reported are: the use of block and convolutional codes [22,52]; and the use of codes in tandem (concatenation), see Ref. 17, part II and Ref.22.

4.3.2 Interleaved Random-Error Correcting Codes

Interleaving (also known as interlacing or scrambling) is a technique which makes possible the use of random-error correcting codes under burst-error conditions. A typical interleaver consists of a rectangular array of digital storage filled with say r code words each of length n. Two identical interleavers are used in a data transmission link, one at the transmitter and the other at the receiver. The interleaver at the transmitter is filled with the code words on a row-by-row basis. The reading out for transmission is done on a column-by-column basis (see Fig. 4.6). At the receiver's interleaver, the reverse operation is carried out to restore the code works to their original compositions.

The idea behind interleaving is seen to be that of separating out the digits of a code word by a fixed number, n, of channel unit times. Fading tends to cluster errors and is, therefore, a type of channel memory. This memory decreases with time as signal power rises and increases again as another fade is encountered. Generally, the intervals between fades are long enough* to allow the spreading of the error clusters into random errors to be completed.

With the use of interleaving, the calculation to find the improvement due to a random-error correcting code is the

$$\begin{array}{cccc}
A_{11} & A_{21} & A_{31} \cdots\cdots\cdots & A_{n1} \\
A_{12} & A_{22} & \cdots\cdots\cdots & A_{n2} \\
A_{13} & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
A_{1r} & A_{2r} & A_{3r} \cdots\cdots & A_{nr}
\end{array}$$

(a)

*and the fades are sufficiently short duration as compared with nr

A_{11} A_{12} A_{13} . . . A_{1r} A_{21} A_{22} . . . A_{2r} A_{31}. . .
. . A_{n1} A_{n2} . . . . A_{nr}

(b)

Fig. 4.6 (a) Storage of Code Words A_{ij} at the Transmitter's Interleaver (b) Reading out Code Words from Transmitter's Interleaver

same as if transmission were over a random-error channel. This is much easier than the calculations involving a burst-error channel.

Let (n, k, e) represent a code used for correcting randomly occurring errors. The number n is the length of a block code and k is the number of information digits. The code rate is k/n. To explain the parameter e, assume that exactly m errors occur in a transmission of n digits. Then, the code, often simply called an (n, k) code, is capable of correcting all the error patterns so long as $m \leq e$. From $m > e$, some, but not all of the error patterns, can be corrected. However, the (23,12) Golay code, used here for illustration of the interleaving process, will correct all error patterns for $m \leq 3$ but none of the error patterns for $m > 3$. For such a code, called a perfect code, an n-bit word is correctly decoded if and only if it contains $m \leq e$ errors. Otherwise, the probability of decoding error is given by

$$P_e = \sum_{j=e+1}^{n} P_n(j) \qquad \ldots \qquad 4.20$$

where, as before, $p_n(j)$ is the probability of j error occurring in a sequence of length n.

When interleaving is used, the errors in a code word occur independently and the equation 4.6 can be used in equation 4.20 to arrive at

$$p_e = \sum_{j=e+1}^{n} nC_j \; p^j \; (1-p)^{n-j} \quad \ldots. \; 4.21$$

For a given (n, k) code with a given e, p_e is explicitly a function of the element probability of error, p.

The results of using equation 4.21 to evaluate the performance of a (23, 12) interleaved Golay code are given in Fig. 4.7 at the end of this Section. The interleaving process has been shown [17] to redefined an (n, k, e) code as an (rn, rk, re) code, where r is the interleaving factor. If r is set equal to unity, then no interleaving is used.

The direct numerical computation of equation 4.21 was found to lead to underflow in the computer storage for small values of p. ~~Provided~~ However if $p < \frac{e}{n}$, these computational difficulties can be overcome by giving upper and lower bounds to p_e. These bounds can easily be evaluated using the following inequality due to Fano [27].

$$\frac{p(1-\rho)}{(1-p)(\rho+\frac{1}{n})} \, b_e < P_e < \frac{p(1-\rho)}{\rho-p} \, b_e$$

where $\rho = \frac{e}{n}$

$$b_e = {}^{n}C_e \; p^e \, (1 - p)^{n - e}$$

On using Equation 4.22 the performance of the (23, 12) Golay code without interleaving ($r = 1$) and with interleaving by a factor of 5 has been evaluated and is given in Fig. 4.7.

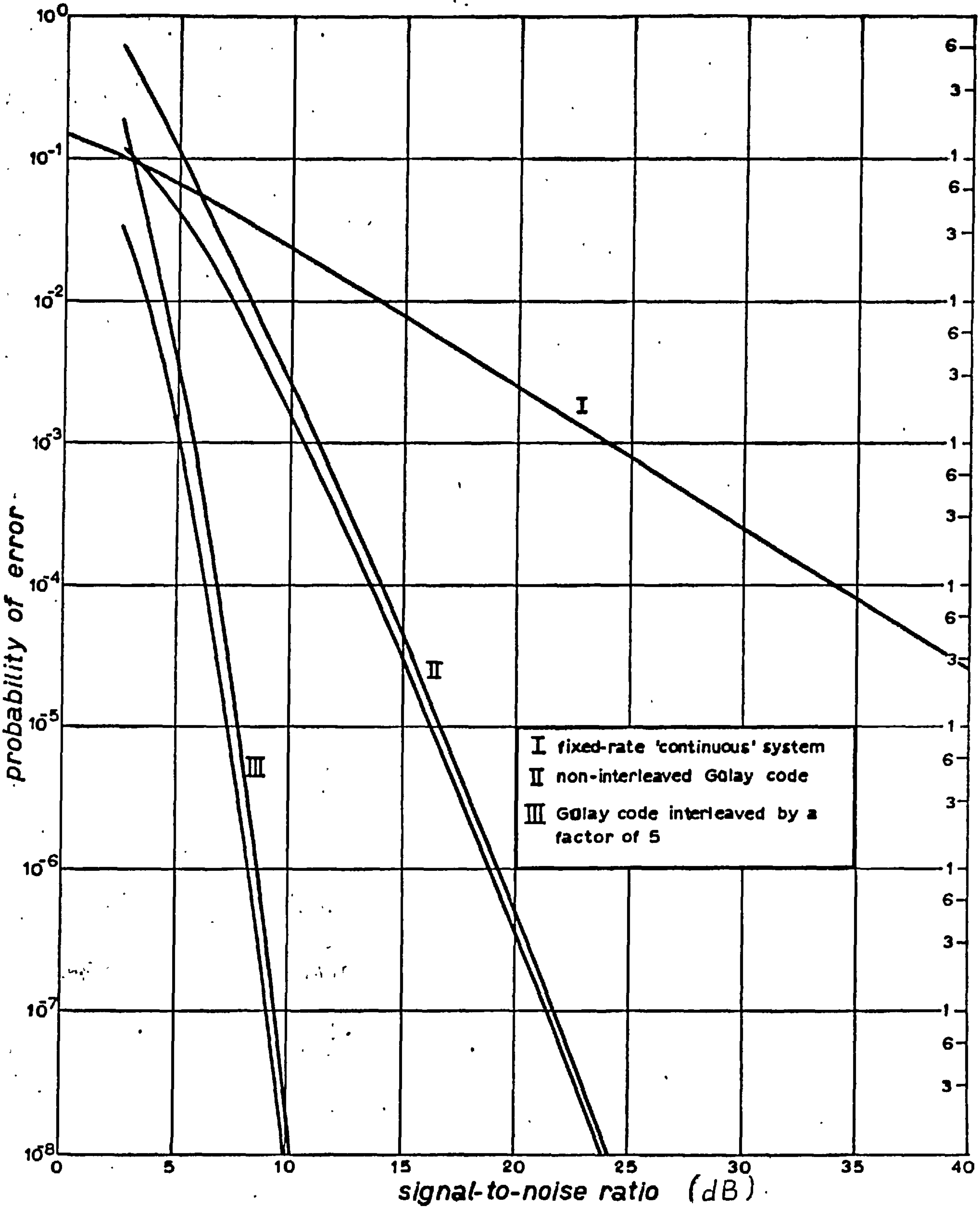


Fig. 4.7. Performance of the (23, 12) Golay Code

5. THE VARIABLE DURATION n-DATA-RATE FSK SYSTEM

5.1 Introduction

Many techniques of varying the rate of transfer of data from the transmitter to the receiver can be proposed. Some of the existing techniques, which were analysed in Chapter 4, are: (1) the use of error detection with retransmission. The van Duuren ARQ technique[85], analysed in Section 4.1, is an example of the retransmission technique. (2) the use of a coding technique. As an example of such a technique, the performance of an interleaved (23, 12) Golay code for an incoherent FSK operating in flat-flat Rayleigh fading conditions was evaluated in Section 4.3.2. (3) the variation of the redundancy of the transmitted signals. The AN/GSC-10 (KATHRYN) modem[94,51], discussed in Section 4.2, is an example of a system using the variable signal-redundancy technique.

Two new methods of varying the rate in a more direct manner were proposed in Section 1.3.1. The first method, that of varying the duration of the transmitted signal, is analysed in detail in this Chapter. The second method, that of varying the number of the signals from which the transmitted signal is chosen will be analysed in the next Chapter.

Fig. 5.1 is a general block diagram of an n-data-rate system. The Figure shows the forward tranmission channel(s) and a feed-back channel over which may be sent a signal to instruct the transmitter to alter a transmission parameter in order to counter-act the detrimental effects of fading. Such a technique was called a variable-parameter technique in Chapter 2. Fig. 5.1 shows a general variable rate scheme in which the parameter altered at the transmitter is the rate of transmission.

With a scheme of this kind, the source transmits information at one of n possible data rates and the aim is to match the data rates to the conditions prevailing in the forward channel. The attempt at the matching is carried out by switching to a lower data rate when the received signal level falls below a threshold and by switching to a higher data rate when the received signal level rises above a threshold. The underlying philosophy being that some improvement in the performance of the system should be achieved by avoiding the use of high instantaneous data rates during the periods when the channel is in a deep fade.

A number of techniques for varying the data rate are out-lined above. Clearly, these are only a few of many possible techniques. In this and the following chapter two particular methods, the variable-duration and the variable-amplitude-set

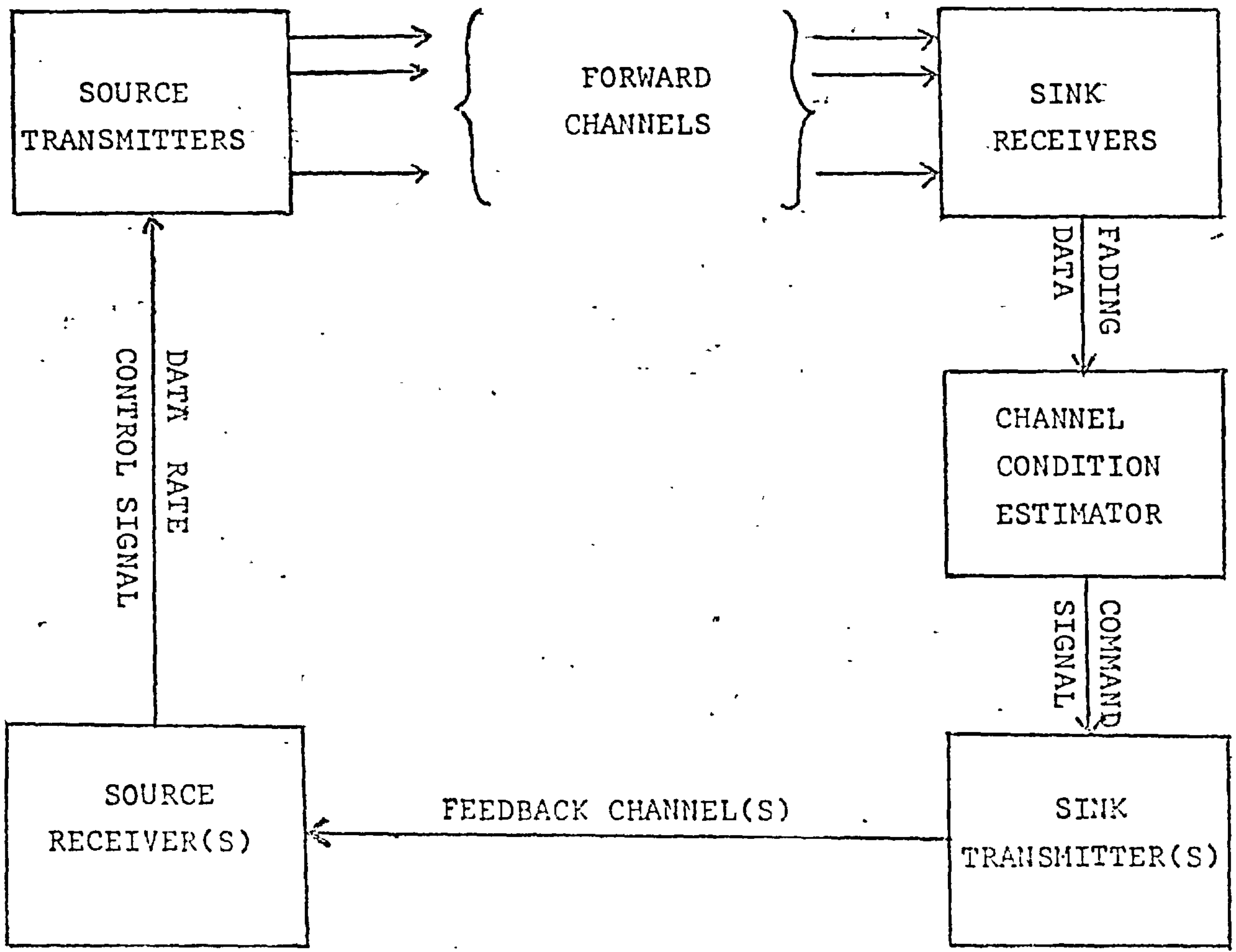


FIG. 5.1 BLOCK DIAGRAM FOR THE GENERAL n-DATA-RATE SYSTEM

techniques will be studied in detail. The reasons for selecting the first method for detailed study are discussed later in this section and the reasons for selecting the second method are discussed in Chapter 6.

The technique of varying the data rate is the first part of the problem of designing a variable-rate system. The second part of the problem is the criterion by which it is decided, usually at the receiver, when to change the data rate. This generally entails estimating the conditions in the data channel. The implication in the discussion of the n-data-rate system given above is that the received signal level is measured and used as a basis of making a decision whether or not to instruct the transmitter to change the data rate. Under certain conditions it may be simpler or more advantageous to estimate the channel conditions by other techniques. For example, under frequency-flat (i.e. non-selective) fading, a pilot tone signal can be sent over a separate pilot-tone channel from the transmitter to the receiver. The level of the pilot tone signal is taken, under the above condition, to be a good indication of the state of the channel. An estimation of the probability of error can also be made directly by using a performance monitoring unit (PMU)[35,44]. The instruction that is sent to the transmitter is then based on the estimate of the probability of error.

From the discussions above, it is evident that it is possible to propose a large number of variable-rate systems. The choice, for a detailed study, of one system from this number must, therefore, to some extent be arbitrary, but when making this choice factors such as realisability, compatibility with existing systems, equipment complexity and other costs should be taken into account.

From the implementation point of view, it is easier to vary the data rate in discrete steps than to vary the data rate continuously. Thus, in this thesis the stress is placed on the analyses of discrete-variable-rate systems. In Section 5.2, which follows, a formulation of a general n-data-rate system is given.

From Section 5.3 onwards, the system analysed is a discrete-variable-duration system, which was defined in Section 1.3.3. The general theory of the variable-duration system is given with application to the incoherent FSK system. It should be noted that it is particularly easy to implement the discrete-variable-duration technique when Frequency Shift Keying is used. To transmit longer pulses, for example, the keying over each transmitted pulse is simply maintained for a period equal to k times the duration of a basic pulse width (where k is integral).

The criterion assumed in the analysis of n-data-rate systems for changing the data rate is that of monitoring at the receiver a pilot tone signal sent from the transmitter. The decision as to which of the possible n data rates is to be used is based on the level of the pilot tone signal. The transmitter is informed accordingly over the feedback path.

5.2 Performance of the n-Data-Rate System

The performance of the n-data-rate (nDR) system can be described in terms of the following functions:

(1) The average probability of error, P_{en}

(2) The probability, P_d, of transmitting data at a rate R_i and receiving the data at a different rate R_j $(j \neq i)$.

(3) The average rate defined as follows. If over a long period of transmission time, T_L, N_L symbols are transmitted, then $R_o = N_L / T_L$ symbols/sec.

Whenever an error is made over the SYNC channel (see Fig. 5.5), the receiver processes the next group of incoming symbols using incorrect duration for them, i.e. the sampling is done at wrong instants of time. The consequence of this is that some groups of data are deleted or that some groups of data are inserted. The average probability of error, P_{en}, should be derived taking this phenomenon into account. This is in general very difficult to do. A method of circumventing this problem is to increase the power over the SYNC channel until the probability of making an error over it is so small as to be negligible. The latter technique is followed in the thesis.

For an n-data-rate system, the average data rate, R_o, can be expressed as

$$R_o = \sum R_i \Omega_i$$

where R_i is the ith data rate and Ω_i is the duty cycle for the ith rate. The duty cycle for ith rate is defined as the percentage of total transmission time that the ith rate is in use.

The object of rate variation is to minimise the probability of error while at the same time maintaining a constant average rate, R_o. Further, if several systems have to be compared, it is necessary to hold R_o at a fixed value for all the systems if the equal-rate comparison (see Section 1.4) is to be used when comparing systems. A difficulty arises, however, that for a given error rate, it is not always possible to hold R_o at a fixed value for somesystems. Consider, for example, the ARQ system.

It is seen from Fig.4.2, that for high SNRs, the rate can, to a good approximation, be assumed to be constant. At low SNRs, however, because of the then too frequent retransmissions, the average rate falls very sharply. In applications where extremely good performance is not essential, it is possible to accept some deterioration in the performance for an increased through-put rate. A method of implementing such a compromise for the ARQ system was suggested in Section 4.1. The extent to which the compromise can be carried out will, of course, depend on the particular application. It will be seen in Chapters 7 and 8 that practical limitations sometimes make it necessary to put a constraint on the highest data rate so that it is less than some maximum fixed value. This requirement also makes it necessary to accept a small decrease in the average rate, R_o, of the n-data-rate system.

In order to determine the performance of the n-data-rate system, that is, in order to determine the average probability of error and the average rate, R_o, it is necessary to define certain operating features of the nDR system. The features that are important are defined in conjunction with Figs. 5.2. and 5.3 as follows:

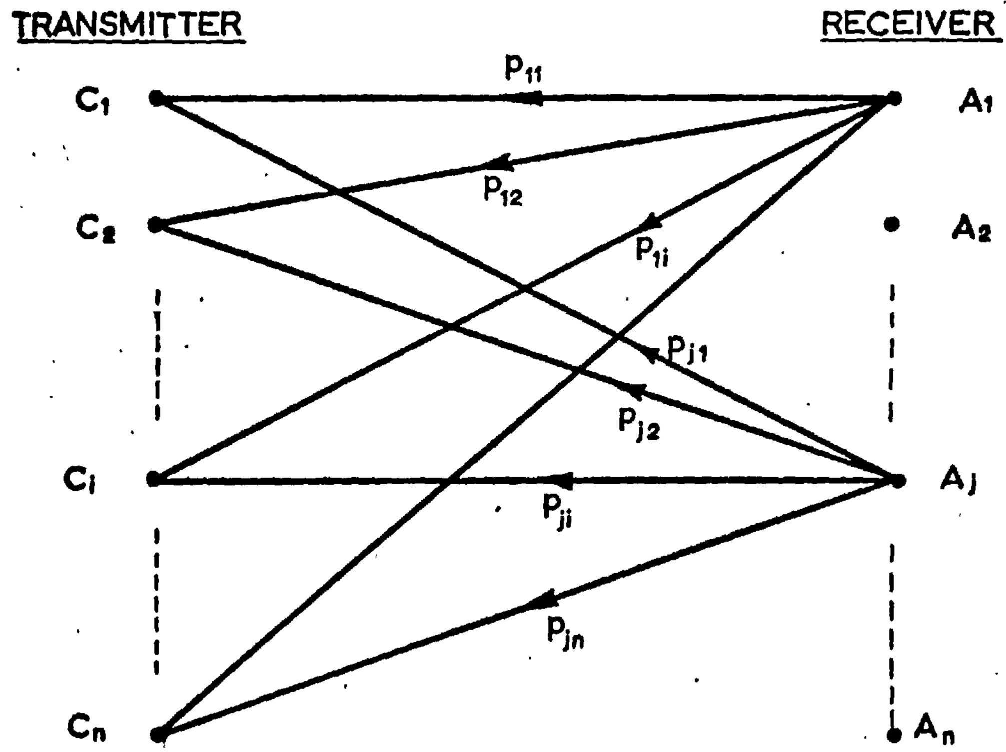


Fig. 5.2. The Probability p_{ji} of Requesting the use of R_j but actually using R_i for Transmission

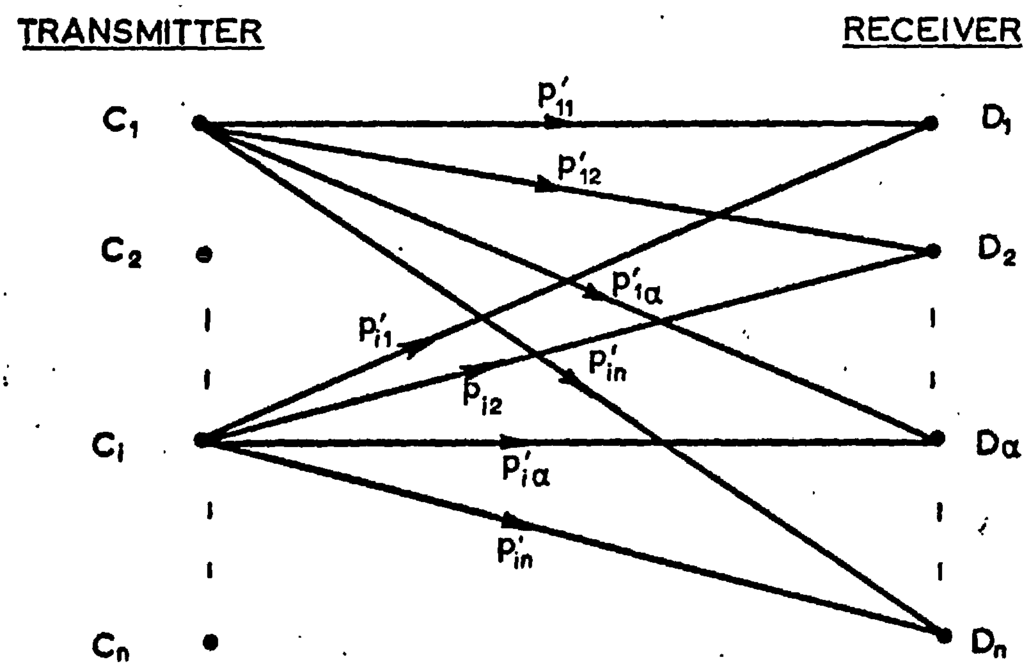


Fig. 5.3. The Probability $p'_{i\alpha}$ of the Transmitter using R_i but the Receiver using R_α

A_j - to indicate that the transmitted feedback signal (i.e. sent from the receiver to the transmitter) indicates transmit at the jth data rate, R_j.

C_i - to indicate that the received feedback signal actually initiates transmission at the ith data rate, R_i.

D_α - to indicate that the data rate control signal (sent from the transmitter to the receiver) signifies the incoming symbols were transmitted at the rate R_α.

B_i - to indicate that a data digit randomly selected from a group of digits transmitted at the rate R_i is in error, given that the event D_i has occurred, i.e. given that no error has occurred over the SYNC channel (see Fig. 5.5).

Further discussion of D_α and B_i is necessary. The 'event' D_α is caused by a signal which completes the data rate control loop (receiver-transmitter-receiver). The average probability of error P_{en} is formulated assuming that $\alpha = i$, that is, assuming that the event D_i occurs whenever C_i has occurred. This, in effect, assumes that P_d is negligible. It is for this case that the event B_i is defined above. The probability of making an error over the SYNC channel, P_d, can be made smaller and smaller by increasing the power over this channel. Because an average power constraint is placed on the transmitter, however, it is necessary to evaluate P_d in order to determine whether it is truly negligible. An expression for evaluating P_d is thus derived in Section 5.4.

Consider the effects of the errors over the feedback and SYNC channels. The subscript $i \neq j$ indicates that an error has occurred over the feedback channel and subscript $\alpha \neq i$ indicates that an error has occurred over the SYNC channel. The feedback and SYNC channel error can also be interpreted in terms of the transitional probabilities given in Figs. 5.2 and 5.3.

The errors that occur over the SYNC channel are taken into account by the probability P_d. In particular, these errors cause message deletions and insertions and are, therefore, extremely undesirable. This is a further reason for making P_d as small as possible. Suppose that an error is made over the feedback channel but not over the SYNC channel, that is, $i \neq j$ but $\alpha = i$, then the rate most suited for the current transmissions is not utilised.

The effect of this is that the instantaneous probability of making an error in the message transmission is higher than for the case when no error is made over both the feedback and SYNC channels.

As stated above, it is initially assumed that errors over the SYNC channels can be ignored. With this assumption, the probability that a digit randomly selected from a group of symbols transmitted at rate R_i will be in error is given by

$$P_{e,i} = P(C_i \; A_j) \; P(B_i \mid C_i A_j)$$

$$= P(B_i \; C_i \; A_j) \qquad \text{.....5.1}$$

NOTE: It is assumed in Equation 5.1 that the receiver had requested transmission of data at rate R_j $(j = 1, \ldots n)$.

Summing over all the possible events C_i and A_j, the average probability of error for n rates, P_{en} is

$$P_{en} = \sum_{i=1}^{n} \sum_{j=1}^{n} P(B_i \; C_i \; A_j) \qquad \text{.....5.2}$$

Now suppose that an error occurs over the SYNC channel. It is only in this case that the receiver actually uses an incorrect symbol rate to process the received waveforms. In the cases where the receiver's original instructions to the transmitter are not followed, $i \neq j$, the detection of the received signals is carried out using the rate R_i, assuming that $\alpha = i$. When $\alpha \neq i$, then deletions or insertions will take place.

With similar reasoning as above, the probability of using $R_\alpha (\alpha \neq i)$ at the receiver, which is the probability of making an error, $P_{d\alpha}$, over the SYNC channel at the αth rate, R_α, is given by

$$P_{d\alpha} = P(C_i A_j) \; P(D_\alpha | C_i A_j)$$

$$= P(D_\alpha C_i A_j) \qquad \text{.....5.3}$$

Summing over all the possible events, that is, over all the A's, C's and D's, the overall probability of making an error over the SYNC channel is

$$P_d = \sum_{\substack{\alpha=1 \\ \alpha \neq i}}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} P(D_\alpha C_i A_j) \qquad \text{.....5.4}$$

NB As stated above, using incorrect symbol durations for processing the received waveforms leads either to insertion of digits that were not sent or to deletions of some digits that were sent. Processing data at a higher data rate than the rate actually used leads to insertions (see Fig. 5.4) whereas processing data at a lower data rate than the rate actually used leads to deletions.

$\frac{1}{R_i}$ = sampling period that should be used.

$\frac{1}{R_j} = \frac{1}{2R_i}$ = sampling period actually used.

Fig. 5.4 Message Insertions Resulting from an Error Over the SYNC Channel

Finally, the average data rate R_o is given by

$$R_o = \sum_{i=1}^{n} R_i \; P(\,C_i\,) \qquad \text{.....5.5}$$

where $P\,(\,C_i\,) = \Omega_i$, the duty cycle for the rate R_i.

Now,

$$P(C_i) = \sum_{j=1}^{n} P(A_j)\; P\,(C_i \mid A_j\,)$$

$$= \sum_{j=1}^{n} P(C_i\, A_j\,) \qquad \text{.....5.6}$$

Physically, Equation 5.6 expresses the fact that the event C_i, could be caused by the event A_i (which was meant to cause it) or by another event A_j ($j \neq i$) as a result of an error over the feedback channel.

On substituting Equation 5.6 into 5.5, the expression for the average rate becomes

$$R_o = \sum_{i=1}^{n} \sum_{j=1}^{n} R_i \; P\,(C_i \; A_j) \qquad \text{.....5.7}$$

Equations 5.2, 5.4 and 5.7 are the expressions for the performance functions of a general n-data-rate system in terms of the joint probabilities of certain events that occur in the system. So far, in the analysis, no restrictions have been placed on the system. The following sections of this Chapter will particularize the results for a discrete-variable-duration incoherent FSK system.

5.3 Performance Functions for a Variable-duration FSK System

This Section and the following Sections of this Chapter are devoted to the analysis of a system in which binary Frequency-Shift-Keying (FSK) is used and in which the message waveform $m_1(t)$ represents a mark or "1" and the message waveform $m_2(t)$ represents a space or "0". The duration of these waveforms can take discrete values, that is, $t \in T_i$, $i = 1, \dots, n$. Thus, when the ith data rate is being used for transmission, the duration of each transmitted waveform, be it a "1" or a "0", is T_i. Since the ith data rate (or symbol rate) is defined as $R_i = \frac{1}{T_i}$, it is evident that switching between differing duration T_i of the transmitted is, in effect, the same as switching between different data rates, R_i ($i = 1, \dots, n$).

The SYNC and the feedback channels are in general not binary. When only two data rates are used, the SYNC and feedback channels are made binary by sending two waveforms over each of the channels and arranging for each waveform to correspond to one of the data rates. The description below sets out how a generalised n-data-rate system might operate.

The receiver continually monitors a carrier sent over the pilot tone channel. Prior to transmission of any data, the receiver informs the transmitter that the channel is suitable for transmission at the jth data rate, R_j. This is done by sending over the feedback channel a signal $f_i(t)$ of duration f sec. After a delay of d sec, a signal $f_i(t)$ is received at the transmitter. If no error has occurred over the feedback channel, then $f_i(t) = f_j(t)$. For noiseless feedback $f_i(t)$ is always equal to $f_j(t)$. When the signal $f_i(t)$ is received at the transmitter it initiates a signal $s_i(t)$ of duration s to be sent over the SYNC channel to inform the receiver that the rate R_i is being used to transmit the next group of symbols. The group of symbols of total duration m sec is then transmitted over the message channel after a total delay of d + f + s sec. The first symbol in the group arrives at the receiver d sec later, assuming equal delay in the feedback and forward transmissions.

N.B. If no error has occurred over the SYNC channel, the received data rate signal is $s_i(t)$; otherwise it is $s_\alpha(t)$, $\alpha \neq i$.

With reference to Fig. 5.5, the following quantities are defined:

$|E|^2$ = squared envelope at the output of the predictor which is used to predict the state of the channel.

$|F|^2$ = squared envelope at the output of the integrate-and-dump matched filters of the feedback channel. This is used to instruct the transmitter which of the n possible data rates to use for

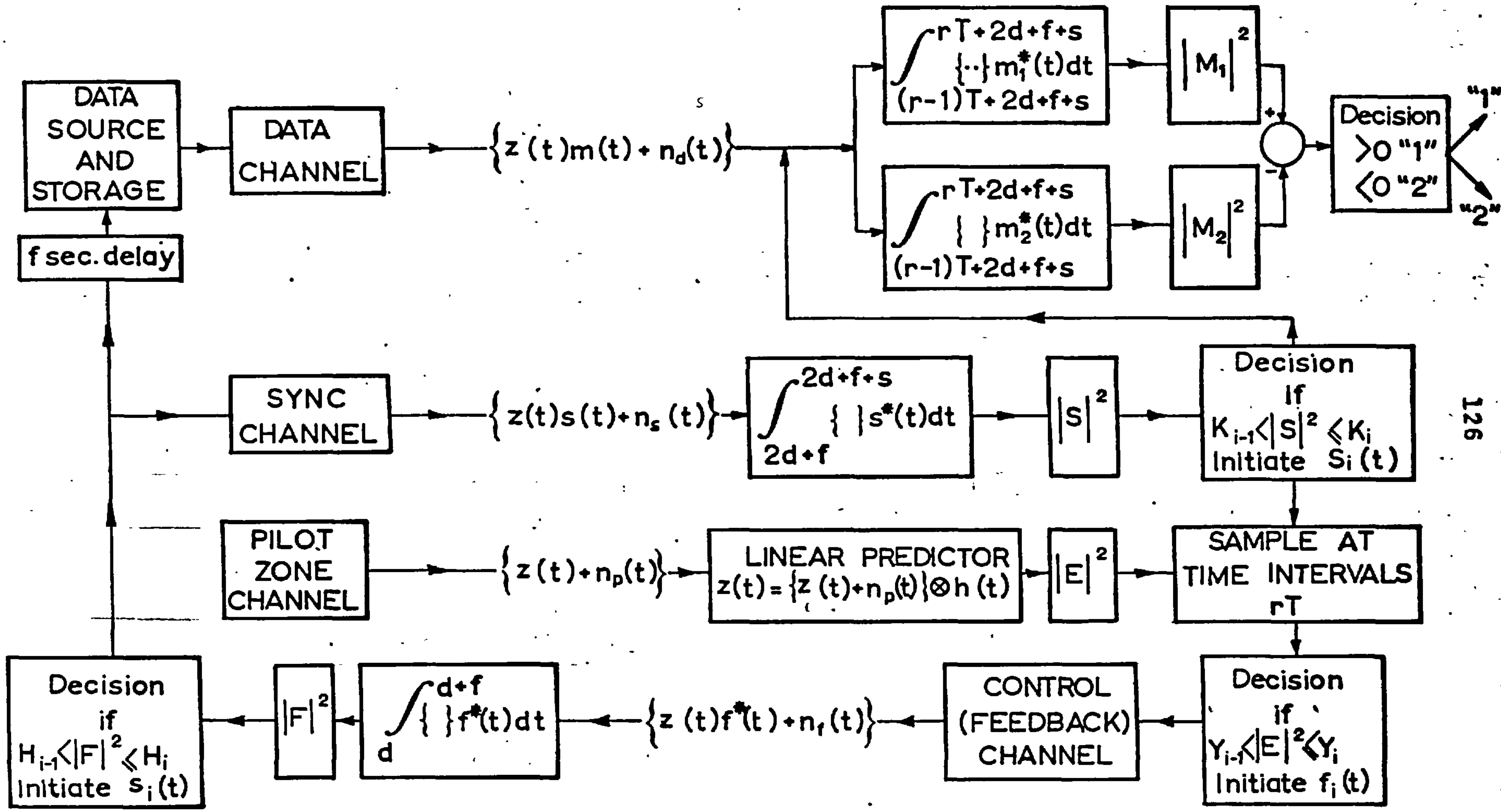


Fig. 5.5. Details of the Variable-Rate FSK Feedback Communication System in Fig. 5.1.

transmission of the next group of symbols.

$|S|^2$ = squared envelope at the output of the integrate-and-dump filters of the SYNC channel. This is used to inform the receiver which of the n possible rates the transmitter is using (strictly, which rate the transmitter is <u>supposed</u> to be using).

$|M_1|^2, |M_2|^2$ = the squared envelopes at the outputs of the integrate-and-dump matched filters of the message channel. These are then compared and output "0" or "1" obtained.

Y_i, H_i, K_i ($i = 1, \ldots, n$) are the thresholds at the outputs of the pilot tone, the feedback and the SYNC channels respectively.

The performance functions of the general n-data-rate system can now be written for the variable-duration FSK system specified above. Firstly, the events defined in Section 5.2 can be re-defined in terms of the above system variables as follows:

A_j = event to indicate that $Y_{j-1} < |E|^2 \leq Y_j$

C_i = event to indicate that $H_{i-1} < |F|^2 \leq H_i$

D_α = event to indicate that $K_{\alpha-1} < |S|^2 \leq K_\alpha$

B_i = event to indicate that $|M_1|^2 < |M_2|^2$, given that $m_1(t)$ was transmitted and that no error was made over the SYNC channel.

Also, the performance functions expressed in terms of Equations 5.2, 5.4 and 5.7 can be rewritten in terms of the above definitions of the events of the n-data-rate system. Note, firstly, that certain one-to-one correspondences exist between some of the variables because of the system operation. Examples are: the event A_j implies $f(t) = f_j(t)$; the event C_i implies $s(t) = s_i(t)$. Using the system variables and these implied events the following are seen to be true.

$$P(B_i\ C_i\ A_j) = P\left[|M_1|^2 < |M_2|^2,\ H_{i-1} < |F|^2 \leq H_i,\ Y_{j-1} < |E|^2 \leq Y_j \,\middle|\, m(t) = m_1(t)\right] \qquad \ldots 5.8$$

$$P(D_\alpha\ C_i\ A_j) = P\left[K_{\alpha-1} < |S|^2 \leq K_\alpha,\ H_{i-1} < |F|^2 \leq H_i,\ Y_{j-1} < |E|^2 \leq Y_j\right] \qquad \ldots 5.9$$

$$P(C_i\ A_j) = P\left[H_{i-1} < |F|^2 \leq H_i,\ Y_{j-1} < |E|^2 \leq Y_j\right] \qquad \ldots 5.10$$

On using Equations 5.8 - 5.10 in Equations 5.2, 5.4 and 5.7, the performance functions become

$$P_{e\ n} = \sum_{i=1}^{n} \sum_{j=1}^{n} P\left[|M_1|^2 < |M_2|^2,\ H_{i-1} < |F|^2 \leq H_i,\ Y_{j-1} < |E|^2 \leq Y_j \,\middle|\, m(t) = m_1(t)\right] \qquad \ldots 5.11$$

$$P_d = \sum_{\alpha=1}^{n} \sum_{\substack{i=1 \\ i=\alpha}}^{n} \sum_{j\neq 1}^{n} P\left[K_{\alpha-1} < |S|^2 \leq K_\alpha,\ H_{i-1} < |F|^2 < H_i\ Y_{j-1} < |E|^2 \leq Y_j\right] \qquad \ldots 5.12$$

$$R_o = \sum_{i=1}^{n} \sum_{j=1}^{n} R_i\ \ P\left[H_{i-1} < |F|^2 \leq H_i,\ Y_{j-1} < |E|^2 \leq Y_j\right] \qquad \ldots 5.13$$

5.4 Independent Fading in the Forward and Feedback Channels

In some transmission channels, the fading in the forward and the feedback paths are independent. When this is the case, then the probabilities in the expressions for the performance functions, Equations 5.11 - 5.13, can be simplified. On carrying out the simplifications, Equations 5.11 - 5.13 become

$$P_{en} = \sum_{i=1}^{n} \sum_{j=1}^{n} P\left[|M_1|^2 < |M_2|^2,\ Y_{j-1} < |E|^2 \le Y_j \middle| m(t) = m_1(t)\right] \cdot P\left[H_{i-1} < |F|^2 \le H_i\right] \qquad \ldots 5.14$$

$$P_d = \sum_{\substack{\alpha=1 \\ =i}}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} P\left[K_{\alpha-1} < |S|^2 \le K_\alpha,\ Y_{j-1} < |E|^2 \le Y_j\right] \cdot \left[|H_{i-1} < |F|^2 \le H_i\right] \qquad \ldots 5.15$$

and

$$R_o = \sum_{i=1}^{n} \sum_{j=1}^{n} R_i \ P\left[Y_{j-1} < |E|^2 \le Y_j\right] P\left[H_{i-1} < |F|^2 \le H_i\right] \qquad \ldots 5.16$$

These simplifactions need physical interpretation. The receiver conveys, via a feedback link, a request for the use of R_j (= event A_j). If the request is conveyed over a noisy feed-back, there are nonzero transitional probabilities, $P_{j\,i}$, that the request for the use of the rate R_j is incorrectly received. These transitional probabilities can be represented diagramatically as in Fig. 5.2.

The transitional probabilities can also be represented in a matrix form as follows:

$$R_{j\,i} = \begin{bmatrix} P_{11} & P_{21} & \text{------} & P_{n1} \\ P_{12} & P_{22} & \text{------} & P_{n2} \\ & & & \\ P_{1n} & P_{2n} & \text{------} & P_{nn} \end{bmatrix}$$

Fig. 5.6 Matrix for the Transitional Probabilities

For the noiseless feedback case, $R_{j\,i}$ becomes an identity matrix, that is, $P_{11} = P_{12} = \text{------} P_{nn}$

The above simplifications lead to easier evaluation of the joint probabilities as is shown in the following section.

5.5 Evaluation of Performance Functions for Rayleigh Fading

In order to evaluate the probabilities in quations 5.14 to 5.16, it is necessary to specify the fading process. A model for the fading process which has been found to agree closely with experiment(Ref. 5, pp. 129-130) is the one characterised by a complex-valued Gaussian process. The envelope of the fading signal is then governed by the Rayleigh probability law (Ref.92,p.527). The square of the envelope, which is proportional to the power in the signal, is shown in Appendix A to have an exponential probability density function, that is,

$$P(|X|^2) = \frac{1}{\overline{|X|^2}} \exp\left[-\frac{|X|^2}{\overline{|X|^2}}\right]$$

where $\overline{|X|^2}$ is the average power in the signal.

The one-dimensional probabilities which are in Equations 5.14 to 5.16 can be expressed as

$$P\left[W_{k-1} < |X|^2 \leq W_k\right]$$

where X stands for E, F or S in equations 5.14 to 5.16 and W stands for K, H and Y, that is, the thresholds on the various channels.

Now,

$$P\left[W_{k-1} < |X|^2 \leq W_k\right] = \int_{W_{k-1}}^{W_k} P(|X|^2)\, d\,|X|^2$$

With the above probability density function, the one-dimensional probabilities can be expressed as follows:

$$P\left[W_{k-1} < |X|^2 \leq W_k\right] = \int_{W_{k-1}}^{W_k} \frac{1}{\overline{|X|^2}} \exp\left[-\frac{|X|^2}{\overline{|X|^2}}\right] d|X|^2$$

$$= \exp\left[-\frac{W_{k-1}}{\overline{|X|^2}}\right] - \exp\left[-\frac{W_k}{\overline{|X|^2}}\right]$$

The following boundary values on the thresholds are defined for later use.

1) when $k = 1$, $W_{k-1} = W_o = 0$. This notation simply says that the lowest threshold, W_o, is set at zero for all channels. This is done because the square of the envelope cannot take negative values. Thus, the lowest data rate R_1 is in use for $0 < |X|^2 \leq W_1$. The probability that $|X|^2$ is in this region is

$$P\left[0 < |X|^2 \leq W_1\right] = P\left[|X|^2 \leq W_1\right]$$

$$= \int_{0}^{W_1} \frac{1}{|X|^2} \exp\left[- \frac{|X|^2}{\overline{|X|^2}}\right] d|X|^2$$

$$= 1 - \exp\left[\frac{W_1}{\overline{|X|^2}}\right]$$

2) When $k = n$ (n being the largest data rate used), $W_n = 0$. It follows that the region for which the highest data rate is in use is given by $W_{n-1} < |X|^2 \leq \infty$. The highest threshold is set at ∞ because when the highest data rate is in use, it is advantageous to keep transmitting at this rate however large the square of the envelope.

The probability that the system will be operating at the highest rate is

$$P\left[W_{n-1} < |X|^2 \leq \infty\right]$$

$$= P\left[W_{n-1} < |X|^2\right]$$

$$= \int_{W_{n-1}}^{\infty} \frac{1}{|X|^2} \exp\left[-\frac{|X|^2}{|X|^2}\right] d|X|^2 \Bigg]$$

$$= \exp\left[-\frac{W_{n-1}}{|X|^2}\right]$$

NOTE: 1) From above, it is evident that very large squares of the envelope occur with very small probabilities.

2) The set of the n data rates used is ordered, that is, $R_1 < R_2 < \ldots. < R_n$.

The two-dimensional probabilities occurring in Equations 5.14 to 5.15, namely,

(i) $P\left[|M_1|^2 - |M_1|^2 < 0,\ Y_{j-1} < |E|^2 \leq Y_j \,\middle|\, m(t) = m_1(t)\right]$

(ii) $P\left[K_{\alpha-1} < |S|^2 \leq K_\alpha,\ Y_{j-1} < |E|^2 \leq Y_j\right]$

are evaluated in Appendix B. The solutions are:

(i) $P\left[|M_1|^2 - |M_2|^2 < 0,\ Y_{j-1} < |E|^2 \leq Y_j \middle| \ m = m_1 \right]$

$$= \frac{1}{1 + \dfrac{\overline{|M_1|^2}}{\overline{|M_2|^2}}} \left\{ \exp\left[-\hat{y}_{j-1} \frac{1 + \dfrac{\overline{|M_1|^2}}{\overline{|M_2|^2}}}{1 + \dfrac{\overline{|M_2|^2}}{\overline{|M_1|^2}}(1-\rho)} \right] - \exp\left[-\hat{y}_j \frac{1 + \dfrac{\overline{|M_1|^2}}{\overline{|M_2|^2}}}{1 + \dfrac{\overline{|M_1|^2}}{\overline{|M_2|^2}}(1-\rho)} \right] \right\}$$

....5.17

(ii) $P\left[K_{\alpha-1} < |S|^2 \leq K_\alpha,\ Y_{j-1} < |E|^2 \leq Y_j \right]$

$$= e^{-\hat{y}_{j-1}} \; Q\left(\sqrt{\frac{2\hat{k}_{\alpha-1}}{1-\mu}}, \sqrt{\frac{2\hat{y}_{j-1}\mu}{1-\mu}}\right) - e^{-\hat{k}_{\alpha-1}} \; Q\left(\sqrt{\frac{2\hat{k}_{\alpha-1}\mu}{1-\mu}}, \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}}\right)$$

$$- e^{-\hat{y}_{j}} \; Q\left(\sqrt{\frac{2\hat{k}_{\alpha}}{1-\mu}}, \sqrt{\frac{2\hat{y}_{j}\mu}{1-\mu}}\right) + e^{-\hat{k}_{\alpha}} \; Q\left(\sqrt{\frac{2\hat{k}_{\alpha}\mu}{1-\mu}}, \sqrt{\frac{2\hat{y}_{j}}{1-\mu}}\right)$$

In these expressions the following abbreviations are used

$$\hat{y}_j = \frac{Y_j}{\overline{|E|^2}}$$

$$\hat{h}_i = \frac{H_i}{\overline{|S|^2}}$$

$$\hat{k}_\alpha = \frac{K_\alpha}{\overline{|F|^2}}$$

$$\rho = \frac{\overline{|M_1 \; E^*|}}{\overline{|M_1|^2} \quad \overline{|E|^2}}, \text{ given that } m(t) = m_1(t)$$

$$\mu = \frac{\overline{|S \;\; E^*|^2}}{\overline{|S|^2} \quad \overline{|E|^2}}$$

E^* is the complex conjugate of E

and Q is the Marcum Q-function.

Substituting the values of the one- and two-demensional probabilities given above into Equations 5.14 to 5.16, the performance functions of the nDR system for time- and frequency-flat Rayleigh fading become:

$$P_{e,n} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}}{1 + \frac{m_{d_1}}{m_{d_2}}(1-\rho)} \left\{ \exp\left[-\hat{y}_{j-1} \frac{1 + \frac{m_{d_1}}{m_{d_2}}}{1 + \frac{m_{d_1}}{m_{d_2}}(1-\rho} \right] - \exp\left[-\hat{y}_j \frac{1 + \frac{m_{d_1}}{m_{d_2}}}{1 + \frac{m_{d_1}}{m_{d_2}}(1-\rho)} \right] \right\} \qquad \ldots 5.19$$

where

$$m_{d_1} = \overline{|M_1|^2} \quad \text{and} \quad m_{d_2} = \overline{|M_2|^2}$$

$$P_d = \sum_{\alpha=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \left[e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i} \right] \left[e^{-\hat{y}_{j-1}} \; Q\left(\sqrt{\frac{2k_{\alpha-1}}{1-\mu}}, \sqrt{\frac{2y_{j-1}\mu}{1-\mu}} \right) \right. \right.$$

$$- e^{-\hat{k}_{\alpha-1}} \; Q\left(\sqrt{\frac{2\hat{k}_{\alpha-1}\mu}{1-\mu}}, \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}} \right) - e^{-\hat{y}_j} \; Q\left(\sqrt{\frac{2\hat{k}_\alpha}{1-\mu}}, \sqrt{\frac{2\hat{y}_j \mu}{1-\mu}} \right)$$

$$-e^{-k_\alpha} Q\left(\sqrt{\frac{2\hat{k}_\alpha'\mu}{1-\mu}}, \sqrt{\frac{2\hat{y}_j}{1-\mu}}\right)\Bigg]\Bigg\} \qquad \ldots 5.20$$

$$R_o = \sum_{i=1}^{n} \sum_{j=1}^{n} R_i \left(e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}\right)\left(e^{-\hat{y}_{j-1}} - e^{-\hat{y}_j}\right) \qquad \ldots 5.21$$

5.6 Expressing the Performance Functions in Forms Suitable for Computation

In the analysis that follows, the waveforms will be specified to be square pulses of sinusoidal signals. In this case, the performance functions can be written in terms of the duration of the waveforms and the power in each symbol.

Let it be assumed that the additive noise in the channels (assumed to be statistically independent white Gaussian processes) have the same double-sided spectral density, N_o. With reference to Fig. 5.5, the signals at the inputs of the data (message), the SYNC, the pilot tone and the feedback receiving filters are, respectively,

(1) $z(t)\, m_1(t) + n_d(t)$, given that $m(t) = m_1(t)$

(2) $z(t)\, s(t) + n_s(t)$

(3) $z(t)p(t) + n_p(t)$. In Fig. 5.5, $z(t)p(t)$ has been written as $z(t)$.

(4) $w(t)f(t) + n_f(t)$

The functions $n_d(t)$, $n_s(t)$, $n_p(t)$ and $n_f(t)$ are the independent white Gaussian processes each with double spectral density, N_o. The function $z(t)$ and $w(t)$, the fading processes in the forward and the feedback channels respectively, are the independent complex-valued Gaussian processes described above.

In writing out the outputs of the matched filters, use is made of the assumption that the fading is time-flat. The fading level is thus assumed to be constant over each transmission epoch, i.e. the period of transmission of a group of N symbols. The fading process $z(t)$ is therefore constant over the group of symbols. The value of this constant is calculated at $t = 2d + f + s + m$, where t is measured from the instant the signal $f_j(t)$ is initiated from the receiver. The fading is further assumed to vary linearly over the feedback and the SYNC pulses. Because of the symmetry of the linear variation of $z(t)$, the fading level for the feedback and the SYNC pulses are constants calculated at $t = 2d + \frac{f}{2}$ and $t = 2d + f + \frac{s}{2}$, respectively, where as before, t is measured from the instant $f_j(t)$ was initiated from the receiver.

Again, with reference to Fig. 5.5, the outputs of the matched filter detectors of the data, the SYNC, the pilot-tone and the feedback channels, respectively, can be written as below.

(1) The output of the filter matched to $m_1(t)$ is

$$M_1 = S_1 + N_1 ,$$

when the transmitted waveform is $m_1(t)$.

$$S_1 = z(zd + f + s + m) \int_0^{T_i} m_1(t)^2 \, dt$$

S_1 is calculated for the last digit in the current group of symbols being received.

N.B. From the characteristic of the prediction filter used (see Fig. 5.7) it is seen that the further into the future prediction is made, the poorer the quality of the prediction. Since the probability of error deteriorates as the quality of prediction goes down, the calculation of the probability of error is carried out for the last digit of a group of received symbols.

$$N_1 = \int_0^{T_i} n(t) \, m_1^*(t) \, dt$$

where $m_1^*(t)$ is the complex conjugate of $m_1(t)$.

Let it be assumed that $m_1(t)$ and $m_2(t)$ are orthogonal waveforms, then when $m_1(t)$ is transmitted, the output of the filter matched to $m_2(t)$ is

$$M_2 = S_2 + N_2$$

where,

$$S_2 = z(2d+f+s+m) \int_0^{Ti} m_1(t)m_2^*(t) \, dt = 0$$

and,

$$N_2 = \int_0^{Ti} n(t)m_2^*(t) = N_1 = N_o.$$

(2) The output of the filter matched to the SYNC pulse is

$$S = S_3 + N_3$$

where,

$$S_3 = z(2d+f+\frac{s}{2}) \int_0^s s_1(t)^2 \, dt$$

and

$$N_3 = \int_0^s n_s(t) S_1^*(t) \, dt$$

(3) The output of the filter matched to the feedback signal is

$$F = S_4 + N_4$$

where,

$$S_4 = z(2d+\frac{f}{2}) \int_0^f f_i(t)^2 \, dt$$

and

$$N_4 = \int_0^f n_f(t) \; i^*(t) \, dt$$

(4) If the pilot tone is specified to be sine waveform with constant amplitude then in complex notation $p(t) = p\, e^{jwt}$. Thus the magnitude of p(t) is p.

The correlation coefficients can be evaluated for the data, the SYNC, the feedback and the pilot tone channels respectively, and are found to be as follows

(1) $$m_{d_1} = \overline{|M_1|^2} = \overline{|S_1|^2} + \overline{|N_1|^2}$$

$$\overline{|S_1|^2} = R_z(o) \int_o^{T_i} |m_1(t)|^2 \, dt$$

where

$R_z(\tau)$ is the autocorrelation function of z(t)

If z(t) is defined to have unit mean power and

$$\int_0^{T_i} m_1(t)^2 \, dt = R_{mi}(0)$$

Where $R_{m_i}(0)$ is the autocorrelation function of m(t), then,

$$\overline{|S_1|^2} = R^2_{mi}(o)$$

$$\overline{|N_1|^2} = 4N_o R_{mi}(o),$$

since $\overline{n(t)\, n\ (t+\tau)} = 4\, N_o\, \delta(\tau)$

$$m_{d_2} = \overline{|M_2|^2} = \overline{|N_2|^2} = 4N_o\, R_{m_i}(o)$$

(2) For the SYNC channel,

$$\overline{|S|^2} = \overline{|S_3|^2} + \overline{|N_3|^2}$$

$$\overline{|S_3|^2} = \int_o^s |s_i(t)|^2\, dt = R_s^2\ (o)$$

where, $R_s\ (\tau)$ is the autocorrelation function of s(t)

$$\overline{|N_3|^2} = 4\, N_o\, R_s\ (o)$$

(3) For the feedback channel

$$\overline{|F|^2} = \overline{|S_4|^2} + \overline{|N_4|^2}$$

$$\overline{|S_4|^2} = \int_o^f |f_i(t)|^2\, dt = R^2_f\ (o)$$

where, $R_f(\tau)$ is the autocorrelation function of f(t)

(4) For the pilot tone channel

$$\overline{|E|^2} = \int_{-\infty}^{o}\int_{-\infty}^{o} \left[p\ z(t_1)+n_p(t)\right] h(-t_1) \left[p\, z(t_2)+n_p^*(t)\right]$$

$$.h\ (-t_2)\ dt_1\ dt_2$$

where,

$$R_h(\tau) = \int_0^\infty h(t)\, h^*(t+\tau)\, dt$$

To calculate ρ the following is required

$$\overline{|M_1 E^*|} = \left| \int_{-\infty}^{0} p^2\, z(2d+f+s+m)\, R_{m_i}(o)\, z(t)\, h^*(-t)\, dt \right|$$

$$= \left| R_{m_i}(o) \int_0^\infty R_z\left[t - (2d+f+s+m)\right] h^*(t)\, dt \right|$$

Thus,

$$\rho = \frac{\overline{|M_1 E^*|}}{m_{d_1}\, \overline{|E|^2}} = \frac{R_{m_i}^2(o) \left| \int_0^\infty p^2\, R_z(o) \left[t-(2d+f+s+m)\right] h^*(t)\, dt \right|^2}{\left[R_{m_i}^2(o) + 4N_o\, R_{m_i}(o)\right] \left[\int_{-\infty}^{\infty} p^2\, R_z(\tau)\, R_h(\tau)\, d\tau + 4N_o\, R_h(o)\right]}$$

$$= \frac{\left| \int_0^\infty p^2\, R_z\left[t - (2d+f+s+m)\right] h^*(t)\, dt \right|^2}{\left[1 + \frac{4N_o}{R_m(o)}\right] \left[\int_0^\infty p^2\, R_z(\tau)\, R_h(\tau)\, d\tau + 4N_o\, R_h(o)\right]} \quad \ldots 5.22$$

Further, to calculate μ, the following is required

$$\overline{|S\ E^*|} = \left| \int_{-\infty}^{o} p^2\ z(2d+f+\tfrac{S}{2})\ R_s(o)\ z^*(t)\ h^*(-t)\ dt \right|$$

Thus,

$$\mu = \frac{\overline{|S\ E^*|}}{\overline{|S|^2}\ \overline{|E|^2}} = \frac{\left| \int_{o}^{\infty} p^2\ R_z\left[t - (2d+f+\tfrac{S}{2})\right]\ h^*\ (t)\ dt \right|^2}{\left[1 + \frac{4N_o}{R_s(o)}\right]\left[\int_{-\infty}^{\infty} p^2\ R_z(\tau)\ R_h(\tau)\ d\tau + 4N_o\ R_h(o)\right]}$$

...5.23

Let

$$\lambda(\tau) = \frac{\left| \int_{o}^{\infty} p^2\ R_z(t-\tau)\ h^*\ (t)\ dt \right|^2}{\int_{-\infty}^{\infty} p^2\ R_z(\tau)\ R_h(\tau)\ d\tau + 4N_o\ R_h(o)}$$

...5.24

On using Equation 5.24 in Equations 5.22 and 5.23, ρ and μ become

$$\rho = \frac{\lambda(2d+f+a+m)}{1 + \frac{4N_o}{R_{mi}(o)}}$$

...5.25

$$\mu = \frac{\lambda(2d+f+\frac{s}{2})}{1 + \frac{4N_o}{R_s(o)}} \qquad \ldots 5.26$$

From the results for the correlation coefficients of the matched filters of the data channel, it follows that

$$\frac{m_{d1}}{m_{d2}} = \frac{4N_o\, R_{m_i}(o) + R^2_{m_i}(o)}{4N_o\, R_{m_i}(o)} \qquad \ldots 5.27$$

$$= 1 + \frac{R_{m_i}(o)}{4N_o}$$

If the following are defined,

$$M_p = \frac{\text{Power in the data channel}}{4N_o}$$

$$S_p = \frac{\text{Power in the SYNC channel}}{4N_o}$$

$$F_p = \frac{\text{Power in the feedback channel}}{4N_o}$$

u_i = average signal-to-noise ratio at the receiver when the ith data rate is being used

$T_i = \frac{1}{R_i}$ = duration of a symbol transmitting using the rate R_i.

then with the definitions of s and f already given, the following abbreviations hold,

$$\frac{R_{m_i}(o)}{4N_o} = M_p \, T_i = u_i$$

$$\frac{R_s\,(o)}{4N_o} = S_p s$$

$$\frac{R_f\,(o)}{4N_o} = F_p \, f.$$

With these notations Equations 5.25 and 5.26 become

$$\rho = \frac{u_i \,\lambda(2d+f+s+m)}{1 + u_i} \qquad \text{...5.28}$$

$$\mu = \frac{\lambda(2d+f+\frac{s}{2})}{1 + S_p \, s} \qquad \text{...5.29}$$

Further, note from the definitions of k and h_i that

$$\hat{k}_\alpha = \frac{K_\alpha}{1 + S_p \, s} \qquad \text{...5.30}$$

and

$$\hat{h}_i = \frac{H_i}{1 + F_p \, f} \qquad \text{...5.31}$$

With these results in mind, the probability of error given by Equation 5.19 can be rewritten as

$$P_{en} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}}{2 + u_i} \left\{ \exp\left[-\hat{y}_{j-1} \frac{2 + u_i}{2+u_i(1-\lambda)}\right] - \exp\left[-y_j \frac{2 + u_i}{2+u_i(1-\lambda)}\right]\right\} \quad ..5.32$$

On using the series expansion of the Marcum Q-function,

$$Q(\sqrt{x}, \sqrt{y}) = e^{-\frac{1}{2}(x+y)} \sum_{m=o}^{\infty} \left(\frac{x}{y}\right)^{\frac{m}{2}} I_m(\sqrt{xy})$$

where I_m is the mth order modified Bessel function,

The probability of making an error over the SYNC channel, Equation 5.20, is given by

$$P_d = \sum_{=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}\right] \left[\exp\left(-\frac{\hat{k}_{\alpha-1} + \hat{y}_{j-1}}{1- \mu}\right) \sum_{m=o}^{\infty}\right.$$

$$I_m\left(\frac{2\sqrt{2\hat{k}_{\alpha-1}\ \hat{y}_{j-1}}}{1-\mu}\right)x\left(\frac{\hat{k}_{\alpha-1}}{y_{j-1}}\right)^{\frac{m}{2}}\left(\frac{1}{\mu^{\frac{m}{2}}}-\mu^{\frac{m}{2}}\right)-\exp\ -\frac{\hat{k}_\alpha+\hat{y}_j}{1-\mu}$$

$$\left(\frac{k_\alpha}{y_j}\right)^{\frac{m}{2}}\left(\frac{1}{\mu^{\frac{m}{2}}}-\mu^{\frac{m}{2}}\right)\Bigg]$$

Lastly, the average data rate remains as given by Equation 5.21., etc.,

$$R_o=\sum_{i=1}^{n}\sum_{j=1}^{n}R_i\ (e^{-\hat{h}_{i-1}}-e^{-\hat{h}_i})\ (e^{-\hat{y}_{j-1}}-e^{-\hat{y}_j})\ \ldots 5.34$$

5.7 The Prediction of the Fades

The parameter λ appearing in some of the equations of the previous section is shown below to be a characteristic of the linear filter used to predict the fades. It is the squared normalised complex correlation between the output of the linear predictor and the complex signal being predicted. The function $\lambda(\tau)$ was defined by Equation 5.24 to be

$$\lambda(\tau) = \frac{\left| \int_{0}^{\infty} p^2 \, R_z(t-\tau) \, h^* (t) \, dt \right|}{\int_{-\infty}^{\infty} p^2 \, R_z(\tau) \, R_h(\tau) \, d\tau + 4N_o \, R_h(o)} \qquad \ldots 5.35$$

The linear predictor is shown in Fig. 5.6

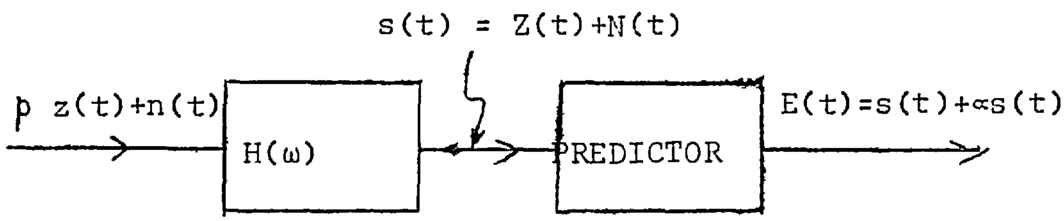


Fig. 5.6 The Fading Predictor

The parameter p is a real constant, namely, the amplitude of the sine wave (i.e. the pilot tone signal).

The object of rate variation, as was mentioned earlier, is to minimise the probability of error while at the same time maintaining a fixed average rate. The average probability will be minimised if the instantaneous rate of transmission is matched to the instantaneous probability of error, that is, to request transmission at the rate R_i if the instantaneous probability of error lies between appropriate thresholds. It is shown below that this optimum procedure is equivalent to forming the best mean-square-error (m.s.e.) linear prediction of the complex envelope of the noiseless received pilot tone. A request for transmission at the rate R_i is then made when this estimate lies between some other thresholds. In particular, it is shown that comparing the probability of error conditional on the whole part of the pilot tone with a given set of thresholds is equivalent to comparing the squared magnitude of the best linear prediction of the complex envelope of the noiseless pilot tone with some other thresholds.

Suppose that the forward channel gain g(t), assumed to be complex Gaussian process, is written as $g(t) = a(t) + j\,b(t)$. The estimate (prediction) of what the future probability of error $\hat{P}_e(t+\tau)$, conditioned on the whole past of the pilot tone, can be written as

$$\hat{P}_e(t+\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\left[a^2(t+\tau) + b^2(t+\tau)\right]\hat{P}\left[a(t) + b(t)\right] da\,db$$

where

f is the function describing the instantaneous probability of digit error

$\hat{p}$ is the joint conditional probability density of the random variables a, b and is a functional of the past of the received pilot tone.

The received pilot tone is Gaussian since it is the product of a deterministic sine wave and a complex Gaussian process (i.e. the fading) to which is added white Gaussian noise. From the properties of Gaussian distributions[25] it can be shown that

$$\hat{P}\left[a(t+\tau)+b(t+\tau)\right] = \frac{1}{2\pi\sigma^2} \exp - \frac{\left[a(t+\tau)-a_o(t+\tau)\right]^2 + \left[b(t+\tau)-b_o(t+\tau)\right]}{2\sigma^2}$$

where a_o, b_o are the averages of the variables a and b respectively, conditional on the whole past of the pilot tone and σ^2 is the conditional variance of a and b.

Now, the conditional mean of a Gaussian variable (conditional on given Gaussian variables) is the best m.s.e. estimate of the variable[64]. Thus, the quantities $a_b(t+\tau)$ and $b_o(t+\tau)$ can be obtained as the output at the time $t+\tau$ of an optimum (m.s.e.)

linear network whose input is the received pilot tone.

An optimum decision procedure is to compare $(t+\tau)$ with a preset threshold. An equivalent procedure would be to compare a monotonic function of $(t+\tau)$ with someother threshold. The advantage of the latter procedure being that the more easily measured quantities, for example, the square of the envelope of the received pilot tone, can be compared with different thresholds.

$\hat{P}(t+\tau)$ is a monotonically decreasing function of $a_o^2(t+\tau) + b^2(t+\tau)$ so long as $f\left[a^2(t+\tau) + b^2(t+\tau)\right]$ is a monotonically decreasing function of $a^2(t+\tau) + b^2(t+\tau)$. Also, $\hat{P}$ is a monotonically decreasing function of $\left[a(t+\tau) - a_o(t+\tau)\right]^2 + \left[b(t+\tau) - b_o(t+\tau)\right]^2$. Since $\hat{\rho}(t+\tau)$ is a monotonically decreasing function of $a_o(t+\tau)$ and $b_o(t+\tau)$. This completes the proof.

In the n-data-rate system, a sub-optimum filter derived in Appendix C is used. The decision to request data transmission at a given data rate will be based only on the present complex envelope of the sub-optimum filter output. The decision process of comparing the squared magnitude of the envelope of the filter output with preset thresholds would be optimum. The proof of this statement is the same as that given above for the decision process of an optimum filter.

It is shown below that the m.s.e., ε, is related to the parameter λ in Equation 5.35 by

$$\varepsilon = 1 - \lambda$$

Thus the performance of the n-data-rate system, given by Equation 5.35 improves as the m.s.e. decreases.

To derive the relationship between ε and λ, note that the normalised m.s.e. is defined as

$$\varepsilon = \frac{\overline{|(\sigma + j\omega)\ x - z)|}}{z\ z^*}$$

where

z is the random variable being predicted

x is the output of the linear filter

α, β are constants which must be set to minimise

To find σ set

$$\frac{\partial}{\partial \sigma}\left\{ \frac{\overline{|(\sigma + j\omega)\ x - z|^2}}{z\ z^*} \right\}$$

to zero.

Whence,

$$\sigma = \frac{\overline{xz^*} + \overline{x^*z}}{2\overline{xx^*}}$$

similarly

$$\omega = \frac{\overline{x^*z} - \overline{xz^*}}{2j \; \overline{xx^*}}$$

Hence

$$\varepsilon = 1 - \frac{\overline{|x^*z|^2}}{xx^* \quad zz^*} = 1 - \lambda \qquad \text{...5.36}$$

The received pilot tone signal is passed through a band-limiting filter (to limit the noise power) before going to the predictor. To a first order approximation, the predictor is assumed to have the characteristic

$$E(t) = s(t) + \alpha \dot{s}(t) \qquad \text{......5.37}$$

where,

$E(t)$ is the output of the predictor

$s(t)$ is the input of the predictor

$\dot{s}(t)$ is the first derivative of $s(t)$

α is a fixed multiplier

It is shown in Appendix C that, for reasonably large power in the pilot-tone channel,

$$\lambda(\tau) = 1 - \tfrac{1}{2} (\tau\omega_f)^4$$

where,

ω_f is the fading rate in radians per sec

Strictly, the Gaussian fading spectrum is not acceptable if the fading process is to be truly random. This follows from the Parlev-Weiner theorem which states that a wide-sense stationary process is non-deterministic if and only if

$$\int_{-\infty}^{\infty} \left| \frac{\log S(f)}{1 + f^2} \right| dt < \infty$$

If S(f) is Gaussian, the above condition is not satisfied and the corresponding fading process is deterministic. This means that the past, however remote, contains all the information about the whole development of the process. However, for the sub-optimum filter used, it is required that the Gaussian function approximate the fading spectrum only near the origin, i.e., for small values of τ.

5.8 Optimisation of the Data Rates of a n-Data-Rate System

In conclusion to this Chapter, a general discussion on the

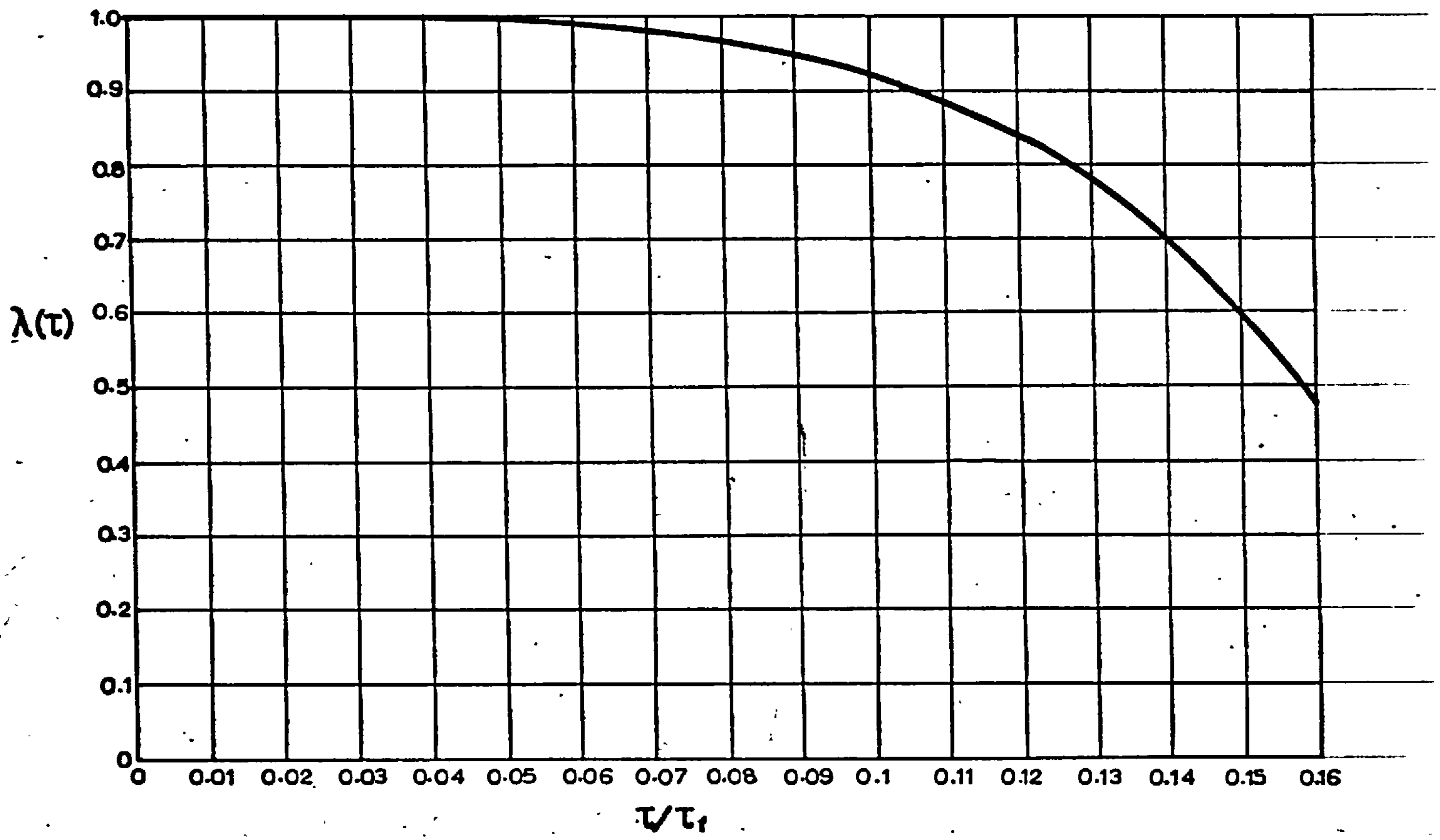


Fig. 5.7. The Characteristic of the Suboptimum Filter used for Prediction.

optimization of the data rates of an n-data-rate system in order to minimize the probability of error while maintaining a constant average data rate is given. Formulae that have been derived previously in this Chapter for the probability of error and the average rate and are required for this discussion are rewritten below.

The average probability of error of an nDR system is given by (see Equation 5.32).

$$P_{en} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}}{2 + u_i} \left\{ \exp\left[-\hat{y}_{j-1} \frac{2 + u_i}{2+u_i(1-\lambda)}\right] - \exp\left[-\hat{y}_j \frac{2 + u_i}{2+u_i(1-\lambda)}\right] \right\} \quad \ldots 5.39$$

The average rate R_o is (see Equation 5.40)

$$R_o = \sum_{i=1}^{n} \sum_{j=1}^{n} R_i \, (e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}) \, (e^{-\hat{y}_{j-1}} - e^{-\hat{y}_j}) \quad \ldots 5.40$$

Consider removing R_1 from the summation signs, i.e., writing equation 5.40 as

$$R_o = R_1 \sum_{i=1}^{n} \sum_{j=1}^{n} k_i \, (e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}) \, (e^{-\hat{y}_{j-1}} - e^{-\hat{y}_j})$$

where,

$$k_i = \frac{R_i}{R_1} \quad , \quad \text{therefore } k_1 = 1$$

In general, any, say the rth, R_i may be removed from the summation signs. Thus, writing Equation 5.40 as

$$R_o = R_r \sum_{i=1}^{n} \sum_{j=1}^{n} k_i \, (e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}) \, (e^{-\hat{y}_{j-1}} - e^{-\hat{y}_j}) \qquad \ldots 5.41$$

where

$$k_i = \frac{R_i}{R_r} \text{ and therefore } k_r = 1.$$

Equation 5.41 can be written as

$$\frac{1}{R_r} = \frac{1}{R_o} \sum_{i=1}^{n} \sum_{j=1}^{n} k_i \, (e^{-j_{i-1}} - e^{-h_i}) \, (e^{-y_{j-1}} - e^{-y_j})$$

$$T_r = \frac{1}{R_o} \sum_{i=1}^{n} \sum_{j=1}^{n} k_i \, (e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}) \, (e^{-\hat{y}_{j-1}} - e^{-\hat{y}_j})$$

$$M_p T_r = \frac{M_p}{T_r} \sum_{i=1}^{n} \sum_{j=1}^{n} k_i \, (e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}) \, (e^{-\hat{y}_{j-1}} - e^{-\hat{y}_j})$$

$$u_r = U_o \sum_{i=1}^{n} \sum_{j=1}^{n} k_i \left(e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}\right)\left(e^{-\hat{y}_{j-1}} - e^{-\hat{y}_j}\right) \quad ...5.42$$

where u_r is the detection SNR when the rth data rate is being used for transmission and U_o is the average detection SNR.

From Equation 5.31

$$\hat{h}_i = \frac{H_i}{1 + F_p\, f}$$

where F_p is the power in the feedback channel

It was shown computationaly that H_i for a two-data-rate system is related by a logarithmic function to P_d which is determined by the power over the SYNC channel. Let it be assumed that this holds for an n-data-rate system. For ease of analysis the duration of the feedback signal, f, is made equal to the duration of the SYNC signal. The optimum power over the SYNC channel, found by solving the equation,

$$\frac{d\mu}{dS_p} = o,$$

is, say S_o, and this optimum SYNC power

~~therefore~~ determines f. The function $\mu(S_p)$ is given by Equation 5.29.

It is evident that the only independant variables in u_r are F_p and $\hat{y}_i$. Thus equation 5.42 can be written as

$$u_r = g(F_p, y_i) \qquad \ldots 5.43$$

From Equation 5.39, it can be seen that in order to minimize P_{en}, u_i should be maximised. The SNR at a given rate is maximised over the feedback channel power by solving the equation

$$\frac{\partial u_r}{\partial F_p} = 0 \qquad \ldots 5.44$$

where $u_r(F_p)$ is given by Equation 5.42. The solution to Equation 5.44 is found for a given set of k_i in Equation 5.42.

The optimum feedback channel power, F_o, and the optimum instantaneous SNR, u_{or}, are used in Equation 5.39 to find the minimum probability of error which can then be further minimised over the remaining parameter y_i by a trial method.

It is clear that general multidimensional optimization problem is far too difficult and will also require too large a computation time. Thus, the problem will be tackled for only two data rates in Chapters 7 and 8.

6. THE VARIABLE-LEVEL n-DATA-RATE SYSTEM

6.1 Introduction

In the previous Chapter, a variable-duration n-data-rate FSK system was analysed. The data rate was varied by altering the duration of the transmitted symbols. A disadvantage of this scheme is that symbol synchronisation becomes very difficult since the pulse widths are not constant. The object of varying the symbol width is to vary the energy content of the symbol under fading conditions. An alternative method of doing this is to vary the amplitude of the transmitted signal. Such a scheme has an advantage as compared with the variable-duration FSK system since the symbol widths remain constant and synchronisation, which is always a difficult problem, is thereby eased.

In this Chapter, the method of varying the data rate by altering the size of the signalling alphabet, which was suggested in Chapter 2, is particularised to Amplitude Shift Keying. When ASK signalling is used, the transmitted signals are determined by amplitude levels. Thus, the transmitted signals could be represented by 2 amplitude levels, by 3 amplitude levels or in general by m amplitude levels. In a variable-level system using ASK signalling, at any particular instant in time, the transmitted signals are chosen from a fixed-size alphabet but the size of the alphabet may be varied from instant to instant.

~~a variable number of amplitude levels but the algebraic sum of the levels used at any particular instant of time remains constant~~. Though the available transmitter power is used all the time, an advantage could be expected from using the technique because a fewer, and hence more easily distinguishable, number of amplitude levels are used during the periods the signal fades. During strong signal periods (i.e. surges) the data rate is increased by increasing the number of amplitude levels, that is, the size of the alphabet, from which the transmitted signals are selected.

When m amplitude levels are used, the binary signalling rate is increased by a factor of $\log_2 m$. Thus, by using 4 amplitude levels the rate of information transmission is doubled.

It is worth noting at the outset that the number of levels necessary to give a similar change in the data rate as a variable-duration scheme increases as the power of two. The reason for this is that the rate of transmission varies as the log (to base two) of the number of the levels (i.e. as the size of the signalling alphabet), namely as $\log_2 m$, whereas in the variable-duration scheme the rate varies linearly with the duration of the symbols. For example, decreasing the level of a signal from 32 to 2 decreases the rate by a factor of 5. Clearly, a similar rate variation is obtainable by varying the duration by only a factor of 5.

The main point of this observation is that the wider the range of rate variation, the more unrealisable the variable-level system becomes as compared with the variable-duration system. For example, changing the data rate by a factor of 10 requires the size of signalling alphabet to be 2^{10} equal 1024 which, for ASK signalling, requires the use of 1024 different amplitude levels, an unmanageable number from implementation point of view. A similar change in the data rate is obtained by changing the duration of the transmitted signals by a factor of 10.

In the next section, a brief discussion of a general variable-number-of-levels ASK system is given. The rest of the Chapter gives the analysis of this general variable-level ASK system assuming perfect prediction and noiseless feedback.

The results for the variable-level ASK systems obtained under the assumptions of perfect prediction and noiseless feed-back are only directly comparable to those of the variable-duration FSK system under the same assumptions. If it is assumed that the noise over the feedback channel has a similar effect on the variable-level ASK system as it has on the variable-duration FSK system, then the comparisons of these two types of systems for the noise-less feedback is valid for the noisy feedback case.

6.2 A Variable-Level ASK System

Fig. 6.1 shows a general form of a variable-level ASK system with a capability of using up to n amplitude levels. The storage accepts information at a constant data rate from a binary source, e.g. tape, deck of binary cards or a digitalizer sampling say a speech waveform at one of n possible data rates depending on which of the coders (shown in Fig. 6.1) is being used for transmission. The storage is necessary because the low data rates (corresponding to small number of amplitude levels) are less than the average rate of the source.

In Fig. 6.1 only the 'coders' and the 'm-level keying control' will be discussed further. The channel, the receiver processor, the fading predictor, the feedback control and the actual system operation are similar to those of the variable-duration FSK system described in the previous Chapter.

The 'coders' in Fig.6.1 are used to convert binary (two-amplitude-level) digits obtained from the source into m-amplitude-level (m-level) signals for transmission over the channel. This can be done, for example, by dividing each of the existing levels into two in order to give a 4-level signalling alphabet. To increase the size of the signalling alphabet to eight each of the

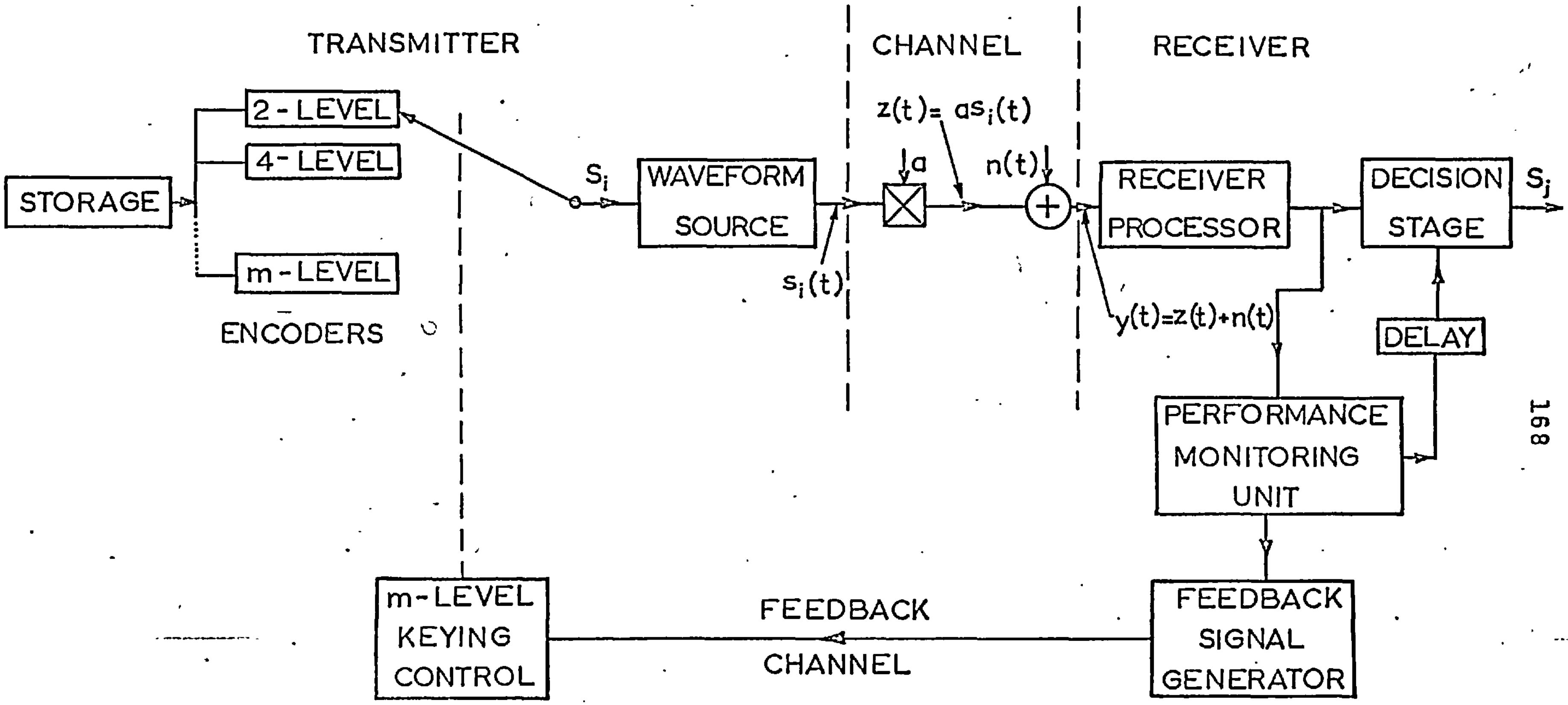


Fig. 6.1. A Variable-Level ASK Feedback Communication System.

4-level signals are further divided into two. When the information is taken in bits, i.e. the information source is binary as assumed in the thesis, it is convenient to constrain the signalling alphabets to take the values $m = 2, 4, 8, \ldots, 2^n$, where n is an integer.

In a variable-level n-data-rate system, n different sizes of signalling alphabets(that is, n different coders) are used. Which one of the n possible coders is used at any particular instant is determined by the feedback signal. The level of the received feedback signal is compared with different thresholds. The ith coder is used when the feedback signal level lies between the (i - 1)th threshold and the ith threshold. For an n-data-rate intermittent system, transmission is halted when the received feedback signal level is below the lowest nonzero threshold.

A system which switches between two different sets of amplitude levels (namely, between two discrete data rates) will be analysed in the following Chapter (with variable-duration two-data-rate FSK system) as a special case of the general variable-level n-data-rate system.

6.3 The Average Probability of Error for a General Variable-Level n-Data-Rate ASK Signal

From the results derived in Chapter 3, it is seen that the error probability, $P_{em}(u)$, for a coherent ASK system is,

$$P_{em} = \frac{(m-1)}{m}\left[1 - 2\,\phi\left(\sqrt{\frac{6u}{2(m-1)(2m-1)}}\right)\right]$$

where

$$\phi(x) = \frac{1}{2\pi}\int_0^x e^{-y^2/2}\,dy$$ and where u is the detection SNR.

When the amplitude of the received signal fades, the SNR, u, is dependent on the channel gain, a. For the channel model assumed in this thesis, 'a' is a random variable from symbol to symbol. To find the average probability of error P_{en}, it is necessary to average $P_{em}(u)$ over all possible values of u. Thus the average probability of error is given by

$$P_{en} = E\left\{P_{em}(u)\right\} \qquad \ldots 6.2$$

Before carrying the averaging process, it is important that for the variable-level n-data-rate system, the range of values which U can take $(0, \infty)$ is divided into $(n - 1)$ cells, where n is the number of separate m-levels used. In each cell the appropriate (or optimal) number of m (size of the signalling alphabet) is used.

Suppose that the thresholds placed on the received SNR, for the purposes of changing between the different sizes of the signalling alphabet are u_{T1}, u_{T2}, ... , $u_{T(n-1)}$ $\left[u_{T_1} < u_{T_2} < \ldots < u_{T(n-1)}\right]$ and let it be further assumed that an m_1 - level signal alphabet is used when $0 < u \leq u_{T_1}$ and an m_2-level signal alphabet is used when $u_{T_1} < u < u_{T_2}$ and generally m_i-level signal alphabet is used when $u_{T(i-1)} < u \leq u_{Ti}$, then

$$P_{en} = \int_{0}^{u_{T_1}} P_{e,m_1}(u)\, f(u)\, du + \int_{u_{T_1}}^{u_{T_2}} P_{e,m_2}(u)\, f(u)\, du + \ldots + \int_{u_{T(i-1)}}^{u_{Ti}} P_{e,m_i}(u)\, f(u)\, du + \ldots + \int_{u_{T(n-1)}}^{\infty} P_{e,m_n}(u)\, f(u)\, du \qquad \ldots 6.3$$

From Equation 6.1

$$P_{e,m_i} = \frac{m_i - 1}{m_i}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_i-1)(2m_i-1)}}\right)\right]$$

Thus,

$$P_{en} = \int_{0}^{u_{T1}} \frac{m_1 - 1}{m_1}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_1-1)(2m_1-1)}}\right)\right] f(u)\ du$$

$$+ \int_{u_{T1}}^{u_{T2}} \frac{m_2 - 1}{m_2}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_2-1)(2m_2-1)}}\right)\right] f(u)\ du$$

$$+ \ \ldots\ldots$$

$$+ \int_{u_{T(i-1)}}^{u_{Ti}} \frac{m_i - 1}{m_i}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_i-1)(2m_i-1)}}\right)\right] f(u)\ du$$

$$+ \ \ldots\ldots$$

$$+ \int_{u_{T(n-1)}}^{\infty} \frac{m_n - 1}{m_n}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_n-1)(2m_n-1)}}\right)\right] f(u)\ du$$

As was mentioned in Chapter 5 and as is shown in Appendix A, if the envelope of a signal is Rayleigh-faded, then

$$f(u) = \frac{1}{U_o} e^{-u/U_o}$$

where U_o is the average SNR (i.e. the mean of the random variable u).

For f(u) of this form it is possible to evaluate the integrals in Equation 6.5. Only three of these integrals, the first, the ith and the last need be evaluated in full. The other integrals can be written by inspection from the solution of the ith integral.

The first integral, I, is given by

$$I_1 = \int_0^{u_{T_1}} P_{em}(u)\, f(u)\, du$$

$$= \int_0^{u_T} \frac{m_1 - 1}{m_1} \left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m-1)(2m-1)}} \right) \right] \frac{1}{U_o} e^{-u/U_o}\, du \qquad \ldots 6.6$$

Similarly the ith integral, I_i, is

$$I_i = \int_{u_{T(i-1)}}^{u_{Ti}} \frac{m_i - 1}{m_i} \left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_i-1)((2m_i-1)}} \right) \right] \frac{1}{U_o} e^{-\frac{u}{U_o}}\, du$$

$$\ldots 6.7$$

and,

$$I_n = \int_{u(Tn-1)}^{\infty} \frac{m_n - 1}{m_n} \left[1 - 2\phi \sqrt{\frac{6u}{2(m_n-1)(2m_n-1)}} \right] \frac{1}{U_o} e^{-u/U_o} du$$

...6.8

Now,

$$I_1 = \int_0^{u_{T_1}} \frac{(m_1-1)}{m_1} \frac{1}{U_o} e^{-u/U_o} du$$

$$- \int_0^{u_{T_1}} \frac{m_1-1}{m_1} \frac{1}{U_o} 2\phi \left(\sqrt{\frac{6u}{2(m_1-1)(2m_1-1)}} \right) e^{-u/U_o} du$$

The first term of I_1 is directly integrable and the second term can be integrated by parts. On evaluating these integrals,

$$I_1 = \left[- \frac{m_1-1}{m_1} e^{-u/U_o} \right]_0^{u_{T_1}} - \left\{ - \left[\frac{2(m\ -1)}{m} e^{-u/U_o} \phi \cdot \left(\sqrt{\frac{6u}{2(m_1-1)(2m_1-1)}} \right) \right]_0^{u_{T_1}} \right.$$

$$+ \int_0^{u_T} \frac{2(m_1-1)}{m_1} \quad e^{-u/U_o} \quad \phi^1\left(\sqrt{\frac{6u}{2(m\ -1)(2m\ -1)}}\right) \ du \Bigg\}$$

where,

$$\phi' = \frac{d\phi}{du}$$

$$I_1 = \frac{m_1-1}{m_1}\left[1 - e^{-u_{T_1}/U_o}\right] + \frac{2(m\ -1)}{m} \ e^{-u_{T_1}/U_o}$$

$$\cdot \ \phi\left(\sqrt{\frac{6u_{T_1}}{2(m_1-1)(2m_1-1)}}\right)$$

$$- \int_0^{u_{T_1}} \frac{2(m_1-1)}{m_1} \quad \frac{1}{\sqrt{2\pi}} \cdot \frac{\sqrt{6}}{\sqrt{2(m_1-1)(2m_1-1)u}} \ \cdot$$

$$\frac{1}{2} \cdot e^{-\left[\frac{6u}{4(m\ -1)(2m\ -1)} + \frac{u}{U_o}\right]} du \qquad \ldots 6.9$$

Consider the integral in Equation 6.9

$$I = \int_0^{u_{T_1}} \frac{2(m_1-1)}{m_1} \quad \frac{\sqrt{6}}{\sqrt{2(m_1-1)(2m_1-1)u}} \cdot \frac{1}{\sqrt{2\pi}} \ \cdot$$

$$\frac{1}{2} \ e^{-\left[\frac{6u}{4(m\ -1)(2m\ -1)} + \frac{u}{U_o}\right]} du$$

Let

$$K_1 = \left[\frac{3}{2(m_1-1)(2m_1-1)} + \frac{1}{U_o}\right] = \frac{3U_o + 2(m_1-1)(2m_1-1)}{2U_o(m_1-1)(2m_1-1)}$$

On using the change of variables

$$x = \sqrt{2K_1 u} \quad , \qquad \therefore\ dx = \frac{\sqrt{2K_1}}{2\sqrt{u}} \cdot du$$

The integral I becomes

$$I = \int_o^{\sqrt{2K_1 u_{T_1}}} \frac{2(m_1-1)}{m_1} \sqrt{\frac{3U_o}{3U_o + 2(m_1-1)(2m_1-1)}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}\, dx$$

$$= \frac{2(m_1-1)}{m_1} \sqrt{\frac{3U_o}{3U_o + 2(m_1-1)(2m_1-1)}} \cdot \frac{1}{\sqrt{2\pi}} \int_o^{\sqrt{2K_1 u_{T_1}}} e^{-\frac{x^2}{2}}\, dx$$

$$= \frac{2(m_1-1)}{m_1} \sqrt{\frac{3U_o}{3U_o + 2(m_1-1)(2m_1-1)}} \cdot \phi(\sqrt{2K_1 u_{T_1}})$$

where,

$$\phi\left(\sqrt{2K_1 u_{T_1}}\right) = \frac{1}{\sqrt{2\pi}} \int_o^{\sqrt{2K_1 u_{T_1}}} e^{-x^2/2}\, dx$$

On substituting back for K_1 in the argument of ϕ, I becomes,

$$I = \frac{2(m_1-1)}{m_1} \cdot \sqrt{\frac{3U_o}{3U_o+2(m_1-1)(2m_1-1)}} \cdot \phi\left(\sqrt{\frac{u_{T_1}\left[3U_o+2(m_1-1)(2m_1-1)\right]}{U_o(m_1-1)(2m_1-1)}}\right)$$

...6.10

On writing $u_{c_1} = u_{T_1}/U_o$ and on substituting Equation 6.10 into 6.9,

$$I_1 = \frac{m_1-1}{m_1}\left[1 - e^{-u_{c_1}}\right] + \frac{2(m_1-1)}{m_1}\, e^{-u_{c_1}}\, \phi\left(\sqrt{\frac{6u_{c_1}U_o}{2(m_1-1)(2m_1-1)}}\right)$$

$$- \frac{2(m_1-1)}{m_1}\sqrt{\frac{3U_o}{3U_o+2(m_1-1)(2m_1-1)}}\ \phi\left(\sqrt{\frac{u_{c_1}\ 3U_o+2(m_1-1)(2m_1-1)}{(m_1-1)(2m_1-1)}}\right)$$

...6.11

The ith integral is given by Equation 6.7

$$I_i = \int_{u_{T(i-1)}}^{u_{Ti}} \frac{m_i-1}{m_i}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_i-1)(2m_i-1)}}\right)\right]\frac{1}{U_o}\, e^{-u/U_o}\, du$$

$$= \int_{u_{T(i-1)}}^{u_{Ti}} \frac{m_i-1}{m_i}\ \frac{1}{U_o}\ e^{-u/U_o}\, du$$

$$- \int_{u_{T(i-1)}}^{u_{Ti}} \frac{2(m_i-1)}{m_i} \frac{1}{U_o} \phi\left(\sqrt{\frac{6u}{2(m_i-1)(m_i-1)}}\right) e^{-u/U_o} \, du$$

The first integral of I_i above is directly integrable and second can be integrated by parts, as was done for a similar integral of I_1. Because of the similarity of the steps taken in evaluating I_1 and I_i, the solution to the latter will be written by inspection from I_1.

$$I_i = \frac{m_{i-1}}{m_i}\left[e^{-u_{c(i-1)}} - e^{-u_{ci}}\right] - \frac{2(m_i-1)}{m_i} \phi\left(\sqrt{\frac{6u_{c(i-1)}U_o}{2(m_i-1)(2m_i-1)}}\right)$$

$$+ \frac{2(m_i-1)}{m_i} e^{-u_{ci}} \phi\left(\sqrt{\frac{6u_{ci}U_o}{2(m_i-1)(2m_i-1)}}\right)$$

$$+ \frac{2(m_i-1)}{m_i} \sqrt{\frac{3U_o}{3U_o+2(m_i-1)(2m_i-1)}} \left\{ \phi\left(\sqrt{\frac{u_{ci}\left[3U_o+2(m_i-1)(2m_i-1)\right]}{(m_i-1)(2m_i-1)}}\right)\right.$$

$$\left. - \phi\left(\sqrt{\frac{u_{c(i-1)}\left[3U_o+2(m_i-1)(2m_i-1)\right]}{(m_i-1)(2m_i-1)}}\right)\right\} \qquad \ldots 6.12$$

where,

$$u_{c(i-1)} = \frac{u_{T(i-1)}}{U_o}$$

$$u_{ci} = \frac{u_{Ti}}{U_o}$$

The third and last integral, given by Equation 6.8 is now evaluated.

$$I_n = \int_{u_{T(n-1)}}^{\infty} \frac{m_{n-1}}{m_n} \frac{1}{U_o} e^{-u/U_o} du$$

$$- \int_{u_{T(n-1)}}^{\infty} \frac{2(m_n-1)}{m_n-1} \frac{1}{U_o} \phi\left(\sqrt{\frac{6U_o}{2(m_n-1)(2m_n-1)}}\right) e^{-u/U_o} du$$

The first part of I_n is integrated directly and second by parts.

$$I_n = \left[- \frac{m_{n-1}}{m_n} e^{-u/U_o} \right]_{u_{T(n-1)}}^{\infty} -$$

$$\left\{ \left[- \frac{2(m_n-1)}{m_n} e^{-u/U_o} \phi\left(\sqrt{\frac{6U_o}{2(m_n-1)(2m_n-1)}}\right) \right]_{u_{T(n-1)}}^{\infty} \right.$$

$$+ \int_{u_{T(n-1)}}^{\infty} \frac{2(m_n-1)}{m_n} e^{-u/U_o} \phi'\left(\sqrt{\frac{6u}{2(m_n-1)(2m_n-1)}}\right) du \Bigg\}$$

where as before $\phi' = \frac{d\phi}{du}$.

Thus

$$I_n = \frac{m_n-1}{m_n} e^{-u_{T(n-1)}/U_o} - \frac{2(m_n-1)}{m_n} \phi\left(\sqrt{\frac{6u_{T(n-1)}}{2(m_n-1)(2m_n-1)}}\right)$$

$$+ \int_{u_{T(n-1)}}^{\infty} \frac{2(m_n-1)}{m_n} \sqrt{\frac{6}{2(m_n-1)(2m_n-1)u}} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2} \cdot e^{-u/U_o}$$

$$\cdot e^{-\frac{6u}{4(m_n-1)(2m_n-1)}} du$$

Let

$$K_n = \frac{3}{2(m_n-1)(2m_n-1)} + \frac{1}{U_o} = \frac{3U_o+2(m_n-1)(2m_n-1)}{2U_o(m_n-1)(2m_n-1)}$$

Using the change of variables

$$x = \sqrt{2K_n u} \quad ; \qquad dx = \frac{\sqrt{2K_n}}{2\sqrt{u}} du$$

Thus

$$I_n = \frac{m_n - 1}{m_n} e^{-u_{c(n-1)}} - \frac{2(m_n-1)}{m_n} \quad \phi\left(\sqrt{\frac{6u_{c(n-1)}\,U_o}{2(m_n-1)(2m_n-1)}}\right)$$

$$- \int_{2K_n\, u_{c(n-1)}\, U_o}^{\infty} \frac{2(m_n-1)}{m_n} \sqrt{\frac{3U_o}{3U_o+2(m_n-1)(2m_n-1)}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}\, dx.$$

where,

$$u_{c(n-1)} = \frac{u_{T(n-1)}}{U_o}$$

But,

$$\int_{\beta}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{\alpha^2}{2}}\, d\alpha = \int_{o}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\alpha^2}{2}}\, d\alpha - \int_{o}^{\beta} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{\alpha^2}{2}}\, d\alpha$$

$$= \frac{1}{2} - \phi\,(\beta) = \frac{1}{2}\left[1 - 2\phi(\beta)\right]$$

Thus,

$$I_n = \frac{m_n - 1}{m_n} e^{-u_{c(n-1)}} - \frac{2(m_n-1)}{m_n} \quad \phi\left(\sqrt{\frac{6u_{c(n-1)}\,U_o}{2(m_n-1)(2m_n-1)}}\right)$$

$$- \frac{m_n-1}{m_n} \sqrt{\frac{3U_o}{3U_o+2(m_n-1)(2m_n-1}} \left\{ 1-2\phi\left(\sqrt{\frac{u_{c(n-1)}\left[3U_o+2(m_n-1)(2m_n-1\right]}{(m_n-1)(2m_n-1)}}\right)\right\}$$

On substituting Equations 6.11 - 6.13 into Equation 6.3, the expression for the average probability of error for a variable-level ASK coherent system (with noiseless feedback) is

$$P_{en} = \frac{m_1-1}{m_1}\left[1 - e^{-u_{c1}}\right] + \frac{2(m_1-1)}{m_1}\, e^{-u_{c1}}\, \phi\left(\sqrt{\frac{6u_{c1}\,U_o}{2(m_1-1)(2m_1-1)}}\right)$$

$$- \frac{2(m_1-1)}{m_1}\sqrt{\frac{3U_o}{3U_o+2(m\ -1)(2m\ -1)}}\ \phi\left(\sqrt{\frac{u_{c1}\,3U_o+2(m\ -1)(2m\ -1)}{(m_1-1)(2m_1-1)}}\right)$$

$$+ \sum_{i=2}^{n-1} \frac{m_i-1}{m_i}\left[e^{-u_{c(i-1)}} - e^{u_{ci}}\right] + \frac{2(m_i-1)}{m_i}\left\{e^{-u_{ci}}\ \phi\left(\sqrt{\frac{6u_{ci}\,U_o}{2(m_i-1)(2m_i-1)}}\right)\right.$$

$$\left. - \phi\left(\sqrt{\frac{6u_{c(i-1)}\,U_o}{2(m_i-1)(2m_i-1)}}\right)\right\} + \frac{2(m_i-1)}{m_i}\sqrt{\frac{3U_o}{3U_o+2(m_i-1)(2m_i-1)}}$$

$$\cdot\left\{\phi\left(\sqrt{\frac{u_{ci}\left[3U_o+2(m_i-1)(2m_i-1)\right]}{(m_i-1)(2m_i-1)}}\right)-\phi\left(\sqrt{\frac{u_{c(i-1)}\left[3U_o+2(m_i-1)(2m_i-1)\right]}{(m_i-1)(2m_i-1)}}\right)\right\}$$

$$+\frac{m_n-1}{m_n}e^{-u_{c(n-1)}}-\frac{2(m_n-1)}{m_n}\phi\left(\sqrt{\frac{6u_{c(n-1)}U_o}{2(m_n-1)(2m_n-1)}}\right)$$

$$-\frac{2(m_n-1)}{m_n}\sqrt{\frac{3U_o}{3U_o+2(m_n-1)(2m_n-1)}}\left\{1-2\phi\left(\sqrt{\frac{6u_{c(n-1)}U_o}{(m_n-1)(2m_n-1)}}\right)\right\}$$

...6.14

6.4 The Average Rate for the Variable-Level n-Data-Rate ASK Coherent System

The average rate R_o for an n-data-rate system was defined in previous Chapters to be

$$R_o = \sum_{i=1}^{n} R_i \Omega_i \qquad \ldots 6.15$$

where Ω_i is the duty cycle for the ith data rate.

Let the channel ~~data~~ symbol rate of the system be denoted by R symbols/sec. When m_i levels are used for transmission, that is, when the transmitted signals are selected from one of m_i possible signals, the information content of each symbol, namely, the symbol rate is $\log_2 m_i$ bits/symbol. Thus the ith rate R_i in Equation 6.15 for m_i-level signalling is

$$R_i = R \log_2 m_i \quad \text{bits/sec.}$$

Thus Equation 6.15 becomes

$$R_o = R\sum_{i=1}^{n} \Omega_i \log_2 m_i = R\left[\Omega_1 \log_2 m_1 + \Omega_2 \log_2 m_2 + \ldots + \Omega_n \log_2 m_n\right] \qquad \ldots 6.16$$

where Ω_1 is the percentage of time that u, the instantaneous SNR, is between o and u_{T_1}. Therefore for time- and frequency-flat Rayleigh fading

$$\Omega_1 = \int_o^{u_{T_1}} \frac{1}{U_o} e^{-u/U_o} du$$

$$= 1 - e^{-u_{T_1}/U_o} = 1 - e^{-u_{c_1}}$$

In general Ω_i is the percentage of time u is between $u_{T(i-1)}$ and u_{Ti}. Hence,

$$\Omega_i = \int_{u_{T(i-1)}}^{u_{Ti}} \frac{1}{U_o} e^{-u/U_o} du$$

$$= e^{-u_{T(i-1)}/U_o} - e^{-u_{Ti}/U_o}$$

$$= e^{-u_{c(i-1)}} - e^{-u_{ci}}$$

The duty cycle for the highest data rate R_n is Ω_n and it is the percentage of time u exceeds the highest finite threshold, i.e.,

$$\Omega_n = \int_{u_{T(n-1)}}^{\infty} \frac{1}{U_o} e^{-u/U_o} du = e^{-u_{T(n-1)}/U_o} = e^{-u_{c(n-1)}}$$

On substituting these values of the duty cycles into Equation 6.16 the expression for the average rate becomes

$$R_o = R\Big[(1-e^{-u_{c1}}) \log_2 m_1 + (e^{-u_{c1}} - e^{-u_{c2}}) \log_2 m_2 + \ldots\ldots\ldots$$

$$\ldots + (e^{-u_{c(i-1)}} - e^{-u_{ci}}) \log_2 m_i + \ldots + e^{-u_{c(n-1)}} \log_2 m_n\Big] \qquad \ldots 6.17$$

After rearrangement, Equation 6.17 can be written more compactly as

$$R_o = R \log_2 m_1 + R \sum_{i=1}^{n-1} \Big[\log m_{i+1} - \log_2 m\Big] e^{-u_{ci}} \qquad \ldots 6.18$$

Equations 6.14 and 6.18 can be used to evaluate the performance of a variable-level coherent ASK system.

In deriving the general performance Equations, 6.14 and 6.18, for the variable-level ASK system, it was assumed firstly that the feedback is noiseless and secondly that the fading is flat in both time and frequency. The noiseless feedback assumption means that the transmitter and the receiver know, at all times, which one of the n possible sets of signalling alphabets is currently being used, that is, how many amplitude levels are being used.

Also, in the analysis carried out in this Chapter, the numerical values of the signalling alphabets were not specified. In the following Chapter, the performance of the variable-level two-data-rate ASK systems will be considered in detail.

7. TWO-DATA-RATE SYSTEMS

7.1 Introduction

The simplest variable rate system is the two-rate system, that is, the system in which $n = 2$. The system is the simplest to implement. It requires only one threshold for switching between the two data rates and the essential logic circuitry will be simpler for the two-data-rate system than for the general n-data-rate system with $n > 2$. On account of this simplicity of practical implementation, the present chapter is devoted to a more detailed consideration of two-Data-Rate systems.

When developing the theory of general n-data-rate systems in the previous two Chapters, no restriction was placed on the actual value of the upper data rate. Similarly in all systems analysed in the literature, and these analyses are restricted to two-rate intermittent systems alone, no restriction is placed on the upper data rate. Clearly, in practice, an upper limit to the rate at which symbols can be transmitted exists and if the higher operating data rate, as calculated from the system equations, cannot be obtained because of these physical limitations, then the performance curves are not valid for practical systems. It is thus important to analyse the case in which there is an upper limit on the higher operating data rate.

The two-data-rate intermittent system (when one of the two rates is zero) has been studied previously by Montgomery (Ref.47, pp.1678-1684), Pierce[65], Martin[60] and Cowan, et al[23]. As was mentioned before, the two-data-rate non-intermittent system has not been analysed previously. This system is more general than the two-data-rate intermittent system because it does not restrict, in advance, one of the two data rates to be zero.

In Sections 7.2 and 7.3, analyses of the two-data-rate systems with an unbounded upper data rate and a fixed average rate are given for variable-duration FSK systems and for variable-level ASK systems. In Sections 7.4 and 7.5 analyses of the two-data-rate systems with an upper limit placed on the higher data rate are given, again for variable-duration FSK systems and for variable-level ASK systems.

7.2 Two-Data-Rate Intermittent Systems With an Unbounded Upper Rate and a Fixed Average Rate

As stated above, an intermittent system is one that ceases transmission at certain times, that is, a system in which one of the two data rates is zero. It will be seen for the analysis of two-data-rate systems given later in this Chapter that, for a fixed average rate, the optimum two-data-rate system is, in fact, the two-data-rate intermittent system. Evidently, for the same average rate, the intermittent system must use a higher data rate (when transmitting) than that used by a fixed-rate system transmitting continuously.

In the following Section the two-data-rate variable-duration intermittent FSK system is analysed in detail; in Section 7.2.2, the variable-amplitude-set intermittent ASK system will be analysed in detail.

7.2.1 The Variable-Duration FSK System

Although the two-data-rate variable-duration FSK system is a special case of the general n-data-rate, it is advantageous to give a complete analysis, rather than use results of Chapter 5, because the analysis of two-data-rate intermittent system is conceptually simple and should lead to a better understanding of variable rate schemes.

With the terminology of Section 5.2, C is the event to indicate that transmission is taking place (and this must be, clearly, at the rate R_2). Not C (i.e. $\bar{C}$) is, then, the event to indicate that transmission is halted. The probability of error, P_{e2}^{INT}, can only be defined when data is being transmitted. Thus the average probability of error for a two-data-rate intermittent system is

$$P_{e2}^{INT} = P(B|C)$$

$$= \frac{P(BC)}{P(C)} \qquad \text{....7.3}$$

From general probability theory

$$p(C) = P(A)\,P(C|A) + P(\bar{A})\,P(C|\bar{A})$$

$$= P(CA) + P(C\bar{A}) \qquad \text{....7.4}$$

where A is an event to indicate that the receiver has instructed that transmission should take place and $\bar{A}$ is the event to indicate that the receiver has instructed that transmission should be halted.

On using Equation 7.4 in Equation 7.3, the average probability of error becomes

$$P_{e2}^{INT} = \frac{P(BCA) + P(PC\bar{A})}{P(CA) + P(C\bar{A})} \qquad \text{....7.5}$$

For two-data-rate systems, the presence and the absence of a signal over the feedback and the SYNC channels are used to indicate which one of the two rates is being used. This allows certain useful approximations to be made. Thus, the following gating procedure could be followed in the case of two-data-rate intermittent system. If a loop signal is received over the SYNC channel, the receiver assumes that transmission is being carried out. The receiver thus opens its 'gate' for the detection of symbols that it assumes follow the received SYNC signal. When no signal is received over the SYNC channel the receiver assumes that transmission is not taking place. In this case the receiver closes its gates.

With the terminology of Section 5.2, D is an event to indicate that the receiver has decided, on the evidence of received SYNC signal, that transmission is taking place. The event $\overline{D}$ (not D) indicates that the receiver has decided that transmission is not taking place.

There are two ways in which an error can occur over the SYNC channel. First, an error occurs over the SYNC channel if the event C occurs at the transmitter, i.e. $s_i(t) = s(t)$ is sent, but the receiver decides that no signal was sent (= an event $\overline{D}$ occurs). This leads to loss or deletion of messages because the receiver's gate is closed. Second, an error occurs over the SYNC channel, if the event $\overline{C}$ occurs, i.e. $s_i(t) = 0$ is sent over the SYNC channel, but the receiver decides that a signal was sent, that is, the event D occurs. This leads to insertion of messages that were not transmitted. Expressing these statements in mathematical terms, the total probability of making an error over the SYNC channel is

$$P_d = P(D\overline{C}) + P(\overline{D}C) \qquad \ldots 7.6$$

On using Equations 7.4 and 7.6, the probability of making an error over the SYNC channel becomes

$$P_d = P(D\overline{C}A) + P(D\overline{C}\overline{A}) + P(\overline{D}CA) + P(\overline{D}\overline{C}\overline{A}) \qquad \ldots 7.7$$

From Equation 5.5 the average rate, R_o, of a two-data-rate intermittent system is

$$R_o = R_1 P(C_1) + R_2 P(C_2)$$

$$= 0.P(C_1) + R_2 P(C_2)$$

$$= R_2 P(C_2)$$

As there is only one data rate being used, the subscripts are dropped to give

$$R_o = R\ P(C) \qquad \ldots 7.8$$

On using Equation 7.4 in 7.8, the expression for the average rate becomes

$$R_o = R\left[P\,(C\,A) + P\,(C\,\bar{A})\right] \qquad \ldots 7.9$$

From this point onwards, the arguments for the simplifications and the solutions of the joint probabilities in Equation 7.5, 7.7 and 7.9 are exactly the same as those given in Chapter 5 for the n-data-rate system. The subscripts on all the thresholds are dropped as there is only one threshold. On using the results of Chapter 5 and Appendix B, the average probability of error, P_{e2}^{INT}, becomes

$$P_{e2}^{INT} = \frac{\dfrac{e^{-h}}{2+u}\left\{1-\exp\left[-\hat{y}\;\dfrac{2+u}{2+u(1-\lambda)}\right]\right\} + \dfrac{e^{\hat{h}}}{2+u}\exp\left[-\hat{y}\;\dfrac{2+u}{2+u(1-\lambda)}\right]}{e^{-h}\,(1-e^{-\hat{y}}) + e^{-\hat{h}-\hat{y}}}$$

$\ldots 7.10$

In writing Equation 7.10, the following rules have been observed. If the transmission is carried out as a result of an instruction from the receiver that has been received correctly back at the transmitter (event A leading to event C), the threshold on the feedback channel is written as $\hat{h}$. If, on the other hand, transmission is carried out as a result of an incorrectly received instruction (event $\bar{A}$ leading to event C), the threshold on the feedback channel is written simply h. This distinction is used below to arrive at some inequalities which are used to simplify equation 7.10.

The expression for the threshold on the feedback channel is given by equation 5.31 which, on dropping the subscripts because there is only one threshold, becomes,

$$\hat{h} \doteq \frac{H}{1 + F_p f}$$

where F_p is the power in the feedback channel and f is the duration of the feedback signal.

From the last two paragraphs, it is evident that

$$h = H$$

since $F_p = 0$ when no signal (i.e. $f_i(t) = 0$) is sent over the feedback signal. Also

$$\hat{h} \doteq \frac{H}{1 + F_p f} = \frac{h}{1 + F_p f} \qquad \text{....7.11}$$

Clearly, for reasonably large power over the feedback channel, the inequality

$$e^{-h} \ll e^{-\hat{h}} \qquad \text{....7.12}$$

holds. Further, the inequality

$$e^{-h} << e^{-\hat{h}-\hat{y}} \qquad \text{...7.13}$$

is true because the optimum threshold, $\hat{y}_{opt}$, on the pilot tone channel is close to unity. This will be shown computationally in Section 8.1.

On using the results of Chapter 5 and Appendix B in Equation 7.7 and following the rules set out above, the probability of making an error over the SYNC channel becomes

$$P_d = e^{-k}\left[1 - e^{-\hat{y}-\hat{h}}\right] + e^{-h}\left[1 - e^{-\hat{k}}\right] + e^{-\hat{h}}\left\{e^{-\hat{y}}\, Q\left(\sqrt{\frac{2\hat{k}}{1-\mu}}, \sqrt{\frac{2\hat{y}\mu}{1-\mu}}\right) - e^{-\hat{k}}\, Q\left(\sqrt{\frac{2\hat{k}\mu}{1-\mu}}, \sqrt{\frac{2\hat{y}}{1-\mu}}\right)\right\} \qquad \text{...7.14}$$

On substituting Equation 7.11 (and a similar equation for k since 'on-off' keying is also used over the SYNC channel) into Equation 7.14, the probability of making an error over the SYNC channel becomes

$$P_d = e^{-k} \left[1 - e^{-\hat{y}-\hat{h}}\right] + e^{-h} \left[1 - e^{-\hat{k}}\right]$$
$$+ e^{\frac{-h}{1+F_p f}} \left\{ e^{-\hat{y}} \; Q\left(\sqrt{\frac{2k}{1-\mu}}, \sqrt{\frac{2y\mu}{1-\mu}}\right) - e^{\frac{k}{1+S_p s}} \; Q\left(\sqrt{\frac{2\hat{k}\mu}{1-\mu}}, \sqrt{\frac{2\hat{y}}{1-\mu}}\right)\right\}$$

...7.15

On substituting the relevant results of Chapter 5 into Equation 7.9, the average data rate, R_o, becomes

$$R_o = \frac{1}{T}\left\{e^{-\hat{y}-\hat{h}} - e^{-h} \left(1 - e^{-\hat{y}}\right)\right\} \qquad \text{....7.16}$$

where $T = \frac{1}{R}$

Equation 7.10 can be re-written as

$$P_{e2}^{INT} = \frac{\frac{1}{2+u}\left[\left\{\exp\left[-\hat{y}\,\frac{2+u}{2+u(1-\lambda)}\right]\right\}\left\{e^{-\hat{h}} - e^{-\hat{h}}\right\} - e^{h}\right]}{e^{-h} + e^{-\hat{y}} \left(e^{-\hat{h}} - e^{-h}\right)}$$

...7.17

On using the inequalities 7.12 and 7.13 in the above Equation for P_{e2}^{INT}, the average probability of error for a two-data-rate intermittent system becomes

$$P_{e2}^{INT} = \frac{1}{2+u}\left\{\exp\left[-\hat{y}\,\frac{2+u}{2+u(1-\lambda)}\right] - e^{-h+\hat{h}+\hat{y}}\right\} \qquad \ldots 7.18$$

Now, from a careful observation of Equation 7.15, it is evident that the threshold, k, on the SYNC channel and the threshold, h, on the feedback channel have a strong influence on the value of P_d. It can be shown, by a simple analysis, that for small values of P_d, the approximation

$$P_d \simeq e^{-h} + e^{-k}$$

is valid. Thus, if h and k are set equal in magnitude, then

$$h = k \simeq \log_e\left(\frac{2}{P_d}\right) \qquad \ldots 7.19$$

It is clear from Equation 7.19 that the last term of Equation 7.18 is of the order of P_d. If it is assumed that the SYNC signal power is sufficiently large and, therefore, that $P_d \ll P_{e2}^{INT}$, i.e. Equation 7.19 is satisfied, then the last term of Equation

7.18 can be ignored and the probability of error for two-data-rate intermittent system can be written as

$$P_{e2}^{INT} = \frac{1}{2+u} \exp\left[-\hat{y} \frac{u\lambda}{2+u(1-\lambda)}\right] \qquad \ldots 7.20$$

On using Equations 7.12 and 7.13 in Equation 7.16, the average data rate R_o becomes

$$R_o = \frac{1}{T} \left(e^{-\hat{y}-\hat{h}}\right) \qquad \ldots 7.21$$

where T is the duration of the transmitted symbols.

From the definitions in Chapter 5 and Equation 7.21, the detection SNR, u, can be written as

$$u = M_p T = \frac{M_p}{R_o} \left(e^{-\hat{y}-\hat{h}} \right) \qquad \ldots 7.22$$

where, as before, M_p is the power in each symbol.

But $\frac{M_p}{R_o} = U_o$, the average detection SNR.

Therefore, substituting Equation 7.22 into 7.20 yields

$$P_{e2}^{INT} = \frac{1}{2+U_o e^{-\hat{y}-\hat{h}}} \exp\left[-\hat{y}\,\frac{U_o e^{-\hat{y}-\hat{h}_\lambda}}{2+U_o e^{-\hat{y}-\hat{h}(1-\lambda)}}\right] \qquad \ldots 7.23$$

Equation 7.23 can be used directly to compute the probability of error, P_{e2}^{INT}, for the two-data-rate intermittent system. This computation is carried out in the following Chapter for the case when the average data rate, R_o, is constrained to be constant, that is, the value of Equation 7.21 is clearly fixed.

The optimum system performance, that is, the minimization of the probability of error is obtainedby varying $\hat{y}$ and $\hat{h}$. Evidently, this is a joint minimization problem, that is, the values of $\hat{y}$ and $\hat{h}$ that give jointly the minimum probability of error are to be computed. However, as is shown below, the optimum value of $\hat{h}$ ($\hat{h}_{opt}$) is independent of the value of $\hat{y}$.

From Equation 7.11, $\hat{h}$ is given by

$$\hat{h} = \frac{h}{1 + F_p\, f}$$

where, as before, $F_p = \dfrac{\text{Power in the Feedback Channel}}{4\ N_o}$

f = the duration of the feedback signal

Since the value h is fixed by Equation 7.19 and the duration of the feedback signal f will be shown in Chapter 8 to be determined by the quality of prediction λ, only F_p in the above expression for h can be adjusted to give a maximum detection SNR, u, which results in a minimum probability of error, P_{e2}^{INT}.

The detection SNR, u, is given as a function of $\hat{y}$ and $\hat{h}$ by Equation 7.22, i.e.

$$u\ (\hat{y},\ \hat{h}\ ) = U_o\ \ e^{-\hat{y}-\hat{h}} \qquad \text{...7.24}$$

where $\qquad U_o = \dfrac{M_p}{R_o}$, the average detection SNR.

Substituting for $\hat{h}$ in Equation 7.24, the detection SNR, u, as a function of F_p and $\hat{y}$, becomes

$$u(F_p\ ,\ \hat{y}\ ) = U_o\ \ e^{-\hat{y}\ -\ \frac{h}{1\ +\ F_p\ f}}$$

Thus, solving the Equation

$$\frac{\partial u}{\partial F_p} = o$$

gives the optimum power of the feedback channel F_{opt}, that is

$$\frac{\partial u}{\partial F} = \frac{\partial}{\partial F_p}\left[U_o\, e^{-\hat{y} - \frac{h}{1+F_p f}}\right] = o$$

But the power per symbol that actually goes into the data channel is

$$U_o = \frac{M_p}{R_o} = \frac{P-F_p}{R_o}$$

where P is the total power available less the sum of the power used over the SYNC channel and the power used over the pilot tone channel.

Thus,

$$\frac{\partial u}{\partial F_p} = \frac{\partial}{\partial F_p}\left[\frac{P-F_p}{F_p}\, e^{-\hat{y} - \frac{h}{1+F_p f}}\right] = o$$

$$-\frac{e^{-\hat{y}}}{R_o}\, e^{-\frac{h}{1+F_p f}} + \frac{e^{-\hat{y}}}{R_o} \cdot \frac{P-F_p}{1} \cdot \frac{fh}{(1+F_p f)^2}\, e^{-\frac{h}{1+F_p f}} = o \qquad \ldots 7.25$$

On solving Equation 7.25 for F_p, the following is obtained for the optimum feedback power, F_{opt}.

$$F_{opt} = \frac{1}{f}\left[\sqrt{h(1+fP) + \frac{h^2}{4}} - \left(1+\frac{h}{2}\right)\right] \qquad ...7.26$$

It is seen that F_{opt} is independent of $\hat{y}$ and hence $\hat{h}_{opt}$ is also independent of $\hat{y}$.

The optimum threshold $\hat{h}_{opt}$ over the feedback channel is, therefore,

$$\hat{h}_{opt} = \frac{h}{1 + F_{opt}\, f} \qquad ...7.27$$

where F_{opt} is given by Equation 7.26.

Having found $\hat{h}_{opt}$, the value of $\hat{y}$ that minimises the probability of error, P_{e2}^{INT}, can be found by differentiating the expression $P_{e2}^{INT}(\hat{y})$ and equating the differential to zero. This is involved because the function $P_{e2}^{INT}(\hat{y})$ is complicated. The values of $\hat{y}$ that minimise the probability of error, P_{e2}^{INT}, for different values of U_o will be shown in Section 7.4 to lie between zero and unity, for the case of noiseless feedback and perfect prediction. A simpler technique than differentiating the expression $P_{e2}^{INT}(\hat{y})$ is, therefore, to use a trial method to find the value of y that minimises P_{e2}^{INT}.

The computational procedure for the minimisation of P_{e2}^{INT} and the curves for the performance of the two-data-rate system will be given in Chapter 8.

7.2.2 The Variable-Level ASK System

The general n-data-rate system in which the rate is varied by changing the size of the signalling alphabet, m, was analysed in Chapter 6 for the case of coherent ASK signalling. An n-data-rate intermittent system for this case is defined as that in which an empty signalling alphabet (i.e. halting transmission) is used as one of the n possible sizes of the alphabet. Thus, in a two-data-rate variable-level intermittent system two signalling alphabets of size zero (halting) and size m are used for transmission.

For brevity, the technique of employing variable sizes of signalling alphabet will often be called a variable-level technique.

For the two-data-rate variable-level intermittent system, the average probability of error can be written from Equation 6.3 as

$$P_{e2}^{INT} = \int_{u_T}^{\infty} P_{em}(u) \; f(u) \; du \qquad \ldots 7.28$$

where u_T is the threshold SNR for changing between zero-level signalling (halting) and m-level signalling (i.e. selecting symbols from one of possible m signals).

In m-level coherent ASK signals the m signals are represented by m amplitude levels. From Table 3.1, the probability of error, $P_{em}(u)$, for the m-level coherent ASK signalling is given by

$$P_{em}(u) = \frac{m-1}{m}\left[1-2\phi\left(\sqrt{\frac{6u}{2(m-1)(2m-1)}}\right)\right] \qquad \ldots 7.29$$

where, as before,

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-y^2/2}\,dy$$

For the flat-flat Rayleigh fading assumed in this thesis, it is shown in Appendix A that

$$f(u) = \frac{1}{U_o}\, e^{-u/U_o}$$

Thus,

$$P_{e2}^{INT} = \int_{u_T}^{\infty} \frac{m-1}{m}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m-1)(2m-1)}}\right)\right]\frac{1}{U_o}\, e^{-u/U_o}\,du \qquad \ldots 7.30$$

This integral was evaluated in Chapter 6 and the solution is given by Equation 6.13. On writing the solution below the subscript n on the parameters m and u is deleted because there is only one size of signalling alphabet and there is only one threshold. Thus, the probability of error for the two-data-rate variable-level intermittent system is

$$P_{e2}^{INT} = \frac{m-1}{m} e^{-u_c} - \frac{2(m-1)}{m} \phi\left(\sqrt{\frac{6u_c U_o}{2(m-1)(2m-1)}}\right)$$

$$- \frac{2(m-1)}{m} \sqrt{\frac{3U_o}{3U_o+2(m-1)(2m-1)}} \left\{1-\phi\left(\sqrt{\frac{u_c\left[3U_o+2(m-1)(2m-1)\right]}{(m-1)(2m-1)}}\right)\right\}$$

where, as before, $u_c = u_T/U_o$

The average rate, R_o, for the two-data-rate variable-level system is

$$R_o = R \log_2 m \int_{u_T}^{\infty} \frac{1}{U_o} e^{-\frac{u}{U_o}} du \qquad \ldots 7.32$$

where $R = \frac{1}{T}$ is the rate in symbols/sec, and hence T is the duration of the transmitted symbols.

On evaluating the integral in Equation 7.32, the average rate R_o is found to be

$$R_o = R \log_2 m \; e^{-u_c}$$

$$= \frac{1}{T} \log_2 m \; e^{-u_c} \qquad \ldots 7.33$$

From Equation 7.33, the duration of the symbols is given by

$$T = \frac{1}{R_o} \log_2 m \; e^{-u_c}$$

Thus,

$$U = M_p T = \frac{M_p}{R_o} \log_2 m \; e^{-u_c} = U_o \log_2 m \; e^{-u_c} \qquad \ldots 7.34$$

Equation 7.34 gives the detection SNR which is used in Equation 7.31 to calculate the average probability of error, P_{e2}^{INT}, under the constraint of constant average data rate, the average rate being given by Equation 7.33. Curves of probability of error for the two-data-rate intermittent systems are given in Chapter 8.

7.3 Two-Data-Rate Non-Intermittent Systems With an Unbounded Upper Rate and a Fixed Average Rate

In Section 7.3.1 which follows, the two-data-rate variable-duration non-intermittent system using incoherent FSK signalling is analysed in some detail as a second example of the general n-data-rate variable-duration system analysed in Chapter 5. In Section 7.3.2. a two-data-rate variable-level coherent ASK system is analysed for the case of noiseless feedback and perfect prediction.

7.3.1 The Variable-Duration FSK System

Formulae (Equations 5.32-5.34) for the performance functions derived in Chapter 5 for the general n-data-rate variable-duration FSK system apply directly to the two-data-rate non-intermittent system if n is set equal to 2.

Since on-off keying can also be used over the SYNC and feedback channels of a two-data-rate non-intermittent system, the expression for the probability of making an error over the SYNC channel, P_d, is exactly the same as for the variable-duration intermittent system analysed in Section 7.2.1. Equation 7.19 is valid for the two-data-rate non-intermittent system as well. In addition, the Inequalities 7.12 and 7.13 are also true for the same reason.

The expressions for the average probability of error, P_{e2}, and for the average rate, R_o, which are given below, are obtained by substituting n = 2 into Equations 5.32 and 5.34 respectively, Thus,

$$P_{e2} = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{e^{-\hat{h}_{i-1}} - e^{-h_i}}{2+u_i} \left\{ \exp\left[-\hat{y}_{j-1} \frac{2+u_i}{2+u_i(1-\lambda)} \right] - \exp\left[-\hat{y}_j \frac{2+u_i}{2+u_i(1-\lambda)} \right] \right\} \qquad \ldots 7.35$$

and

$$R_o = \sum_{i=1}^{2} \sum_{j=1}^{2} R_i \left(e^{-\hat{h}_{i-1}} - e^{-\hat{h}_i}\right) \left(e^{-\hat{y}_{j-1}} - e^{-\hat{y}_j}\right) \qquad \ldots 7.36$$

When performing the above summations, the following definitions are used

1. $\hat{y}_0 = 0 \qquad \hat{h}_0 = 0$

2. $\hat{y}_2 = \infty \qquad \hat{h}_2 = \infty$

On performing the summation of Equation 7.35 and using the Inequalities 7.12 and 7.13, the probability of error for the two-data-rate non-intermittent system becomes

$$P_{e2} = \frac{1-e^{-\hat{h}_1}}{2+u_1} \left\{ 1 - \exp\left[-\hat{y}_1 \frac{2+u_1}{2+u_1(1-\lambda)} \right] \right\}$$

$$+ \frac{e^{-\hat{h}_1}}{2+u_2} \exp\left[\hat{y}_1 \frac{2+u_2}{2+u_2(1-\lambda)} \right] \qquad \ldots 7.37$$

It is observed that the subscripts on the thresholds $\hat{y}$ and $\hat{h}$ can be dropped as no confusion arises from doing so. Thus,

$$P_{e2} = \frac{1-e^{-\hat{h}}}{2+u_1} \left\{ 1 - \exp\left[-\hat{y}\,\frac{2+u_1}{2+u_1(1-\lambda)}\right] \right\}$$

$$+ \frac{e^{-\hat{h}}}{2+u_2} \exp\left[-\hat{y}\,\frac{2+u_2}{2+u_2(1-\lambda)}\right] \qquad \ldots 7.38$$

On expanding Equation 7.36 using similar rules, the average rate, R_o, for the two-data-rate variable-duration non-intermittent system becomes

$$R_o = R_1\,(1 - e^{-\hat{h}})\,(1 - e^{-\hat{y}}) + R_2\,e^{-\hat{y}-\hat{h}} \qquad \ldots 7.39$$

From Equation 7.39, the duration of the symbols when the higher data rate, R_2, is used is given by

$$T_2 = \frac{1}{R_2} = \frac{1}{R_o}\left\{ k(1-e^{-\hat{h}})(1 - e^{-\hat{y}}) + e^{-\hat{y}-\hat{h}} \right\}$$

where $\quad k = R_1/R_2$

Therefore, the detection SNR, u_2, when the rate R_2 is being used is

$$u_2 = M_p T_2$$

$$= U_o \left\{ k(1 = e^{-\hat{h}})(1 - e^{-\hat{y}}) + e^{-\hat{y}-\hat{h}} \right\} \quad \ldots 7.40$$

where, as before, $U_o = \frac{M_p}{R_o}$

The detection SNR, u_1, when the rate R_1 is being used can be expressed in terms of u_2 as follows,

$$u_1 = M_p \frac{T_1}{T_2} \cdot T_2 = M_p T_2 \cdot \frac{T_1}{T_2} = \frac{u_2}{k} \quad \ldots 7.41$$

where $T_1 = \frac{1}{R_1}$

The probability of error for the two-data-rate non-intermittent system can be evaluated, under the condition of constant average data rate, by substituting Equation 7.40 and 7.41 in Equation 7.38.

The optimum system performance, that is, the minimisation of the probability of error is obtained, as was done for the two-data-rate variable-duration intermittent system, by varying the thresholds $\hat{y}$ and $\hat{h}$.

From Equation 7.11, $\hat{h}$ is given by

$$\hat{h} = \frac{h}{1 + F_p f}$$

As was explained in the case of the variable-duration intermittent system, h and f are fixed quantities and only F_p, the power over the feedback channel, can be adjusted to give a maximum detection SNR, u_2, which results in a minimum probability of error, P_{e2} (which is only a function of u_2 after substitution of Equations 7.40 and 7.41 into 7.38).

On substituting Equation 7.11 into Equation 7.40, $u_2\ (F_p, \hat{y})$ becomes

$$u_2(F_p, y) = U_o \left\{ k \left[1 - e^{-\frac{h}{1+F_p f}} \right] \left[1 - e^{-\hat{y}} \right] + e^{-\hat{y} - \frac{h}{1+F_p f}} \right\} \quad \ldots 7.42$$

To find the maximum value of u_2 for any value of $\hat{y}$, the partial differential equation

$$\frac{\partial u_2}{\partial F_p} = o \quad \ldots 7.43$$

must be solved.

If the function $g(\hat{F}_p\ , \hat{y})$ denotes $\frac{\partial u_2}{\partial F_p}$ then from solving Equation 7.43 using Equation 7.42

$$g(F_p,\hat{y}) = o = \left[(P-F_p)\frac{fh}{(1+fF_p)} - 1\right] e^{-\hat{y} - \frac{h}{1+F_p f}} - \frac{k}{1-k} \qquad \ldots 7.44$$

where, as before, P is the total available power less the power used over the SYNC and pilot tone channels.

The Newton-Raphson iterative method can be used to solve Equation 7.44. On doing this, the (i+1)th approximation to the optimum F_p, which is now a function of $\hat{y}$, is

$$F_{opt}^{i+1}(\hat{y}) = F_{opt}^{i}(\hat{y}) - \frac{g(F_p,\hat{y})}{g'(F_p,\hat{y})} \qquad \ldots 7.45$$

In this iterative technique F^i_{opt} is some reasonable guess at the optimum value of F_p and $g' = {}^{\partial g}/_{\partial F_p}$

The value of $F_p(\hat{y})$ which, within the required accuracy, satisfies Equation 7.44 using Equation 7.45 is designated $F_{opt}(\hat{y})$. Thus the optimum threshold on the feedback channel is

$$\hat{h}_{opt} = \frac{h}{1+F_{opt}(\hat{y})f} \qquad \ldots 7.46$$

The procedure for minimising the probability of error over the pilot-tone channel threshold, $\hat{y}$, is similar to that for the two-data-rate intermittent system except in computational detail. Since $\hat{y}$ now affects, $\hat{h}_{opt}$, as seen from Equations 7.44-7.46 $\hat{h}_{opt}$ must be evaluated for all values of $\hat{y}$ used in the trial method described for the intermittent system. In other words, the values of $\hat{h}$ and $\hat{y}$ which minimise the probability of error must be found jointly.

From computations, the results of which are given in the following Chapter, it is found that the optimum power, F_{opt}, over the feedback channel is much the same for the intermittent and the non-intermittent systems. This is fortunate for the comparison of the two types of systems. This comparison is made in Chapter 8.

7.3.2. The Variable-Level ASK System

In this Section performance formulae for the two-data-rate variable-level non-intermittent ASK system are derived for the case of noiseless feedback and perfect prediction. For such a system, switching is made between two nonzero sets of signalling alphabets, m_1 and m_2, with $m_2 > m_1$.

On using Equation 6.3, the average probability of error for the two-data-rate variable-level non-intermittent system can be written as

$$P_{e2} = \int_{o}^{u_T} P_{em_1}(u)\, f(u)\, du + \int_{u_T}^{\infty} P_{em_2}(u)\, f(u)\, du \quad \ldots.7.47$$

The expression for $P_{em_1}(u)$ for m_1-level coherent ASK signalling can be deduced from Table 3.1. It is

$$P_{em_1}(u) = \frac{m_1 - 1}{m_1}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_1-1)(2m_1-1)}}\right)\right]$$

Clearly, for m_2-level coherent ASK signalling $P_{em_2}(u)$ is

$$P_{em_2}(u) = \frac{m_2 - 1}{m_2}\left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_2-1)(2m_2-1)}}\right)\right]$$

For the flat-flat Rayleigh fading assumed in this thesis, f(u) is shown in Appendix A to be

$$f(u) = \frac{1}{U_o}\, e^{-u/U_o}$$

Thus Equation 7.47 becomes

$$P_{e2} = \int_{0}^{u_T} \frac{m_1-1}{m_1} \left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_1-1)(2m_1-1)}}\right)\right] \frac{1}{U_o} e^{-\frac{u}{U_o}} du$$

$$+ \int_{u_T}^{\infty} \frac{m_2-1}{m_2} \left[1 - 2\phi\left(\sqrt{\frac{6u}{2(m_2-1)(2m_2-1)}}\right)\right] \frac{1}{U_o} e^{-\frac{u}{U_o}} du$$

...7.48

The solution of the first integral in equation 7.48 is given by equation 6.11 and the solution to the second integral is given by equation 6.13 after substituting m_2 for m_n. In both of these cases, u_{c1} and $u_{c(n-1)}$ are replaced by u_c since there is only one threshold. Thus

$$P_{e2} = \frac{m_1-1}{m_1}(1-e^{-u_c}) + \frac{2(m_1-1)}{m_1} e^{-u_c} \phi \sqrt{\frac{6u_c U_o}{2(m_1-1)(2m_1-1)}}$$

$$- \frac{2(m_1-1)}{m_1} \sqrt{\frac{3U_o}{3U_o+2(m_1-1)(2m_1-1)}} \phi\left(\sqrt{\frac{u_c\left[3U_o+2(m_1-1)(2m_1-1)\right]}{(m_1-1)(2m_1-1)}}\right)$$

$$+\frac{m_2-1}{m_2}\, e^{-u_c} - \frac{2(m_2-1)}{m_2}\, \phi\left(\sqrt{\frac{6u_c\, U_o}{2(m_2-1)(2m_2-1)}}\right) - \frac{2(m_2-1)}{m_2}$$

$$\cdot \sqrt{\frac{3U_o}{3U_o+2(m_2-1)(2m_2-1)}}\quad 1-2\phi\left(\sqrt{\frac{u_c\left[3U_o+2(m_2-1)(2m_2-1)\right]}{(m_2-1)(2m_2-1)}}\right) \qquad \ldots 7.49$$

where, as before, $u_c = u_T/U_o$

On substituting n = 2 in Equation 6.18, the average rate for a two-data-rate variable-level non-intermittent system becomes

$$R_o = R\left[\log_2 m_1 + (\log_2 m_2 - \log_2 m_1)\, e^{-u_c}\right] \qquad \ldots 7.50$$

From Equation 7.50, the duration of the transmitted symbols is given by

$$T = \frac{1}{R} = \frac{1}{R_o}\left[\log_2 m_1 + (\log_2 m_2 - \log_2 m_1)\, e^{-u_c}\right]$$

Thus, the detection SNR, u, is

$$u = M_p T = \frac{M_p}{R_o} \left[\log_2 m_1 + (\log_2 m_2 - \log_2 m_1) e^{-u_c} \right]$$

$$= U_o \left[\log_2 m_1 + (\log_2 m_2 - \log_2 m_1) e^{-u_c} \right] \quad \text{....7.51}$$

Equation 7.51 gives the detection SNR which, when used in Equation 7.49 to calculate the average probability of error, maintains a constant average rate, R_o, given by Equation 7.50.

The performance curves for the two-data-rate non-intermittent coherent ASK system are given in Chapter 8 where comparisons with other two-data-rate systems are made.

7.4 Two-Data-Rate Systems With an Upper Limit to the Higher Data Rate

7.4.1 Introduction

In the analyses of two-data-rate systems given in Sections 7.2 and 7.3 above, no restriction was placed on the value that the upper rate, R_h, could take. Clearly R_h cannot be increased without limit since a very high R_h implies that very short pulses are used and this requires a very wide bandwidth which may not be available. In the following two Sections, 7.4.2 and 7.4.3, R_h will be limited to some maximum value R_{hmax}, where R_{hmax} is determined by practical considerations such as available bandwidth and system complexity.

To state the problem in a slightly different and in a more useful way, define the lower rate, R_ℓ, and the upper, R_h, as fractions of R_o, that is,

$$R_\ell = \eta_\ell \quad R_o \qquad \text{...7.52}$$

and

$$R_h = \eta_h \quad R_o \qquad \text{...7.53}$$

Evidently $\eta_\ell < 1$ and $\eta_h > 1$.

On using the equations for the average rate, R_o, given in the previous two Sections, 7.2 and 7.3, to calculate the upper rate R_h, it may turn out that the desired R_h is greater than R_{hmax}. In this case, in order to decrease R_h back to R_{hmax} either η_h or R_o can be decreased. Decreasing η_h is achieved in practice by increasing the proportion of time in the upper rate R_h is used transmission, that is, decreasing the threshold u_T. When R_o is allowed to deteriorate, η_h and hence the percentage of time when R_h is in use decreases. In the latter case, the optimum threshold, u_{Topt} can be used.

7.4.2 Two-Data-Rate Intermittent Systems with an Upper Limit to the Higher Data Rate

The equations for the average data rate for the two-data-rate incoherent FSK system and for the two-data-rate coherent ASK system were derived in the Sections 7.2 and and are

$$R_o = e^{-\hat{y}-\hat{h}}$$

and

$$R_o = R \log_2 n \, e^{-u_c}$$

respectively.

Both of these expressions for the average rate can be put into a single form as

$$R_o = R_h \, e^{-\alpha} \qquad \ldots 7.54$$

where

$$R_h = \begin{cases} R \text{ in the case of variable-duration system} \\ R \log_2 m \text{ in the case of variable-level system} \end{cases}$$

$$\alpha = \frac{u_T}{U_o} = \begin{cases} \hat{y} + \hat{h} \text{ in the case of variable-duration system} \\ u_c \quad \text{in the case of variable-level system} \end{cases}$$

The parameter α is the cut-off ratio, i.e. it is the ratio between the received SNR to average SNR below which the transmission is cut-off (halted). It also determines the percentage of time, the intermittent system is actually transmitting information.

From Equations 7.53 and 7.54, the operating data rate R_h, of an intermittent system can be expressed in terms of the average data rate, R_o, in the following manner,

$$\frac{R_h}{R_o} = \eta_h = e^{\alpha} \qquad \ldots 7.55$$

The average rate R_o is taken to be pre-determined by the specification of the two-data-rate system. The cut-off ratio, α, must be optimized, as shown below, to give minimum probability of error. Thus, whether or not R_h is achievable, for the optimum α, is determined by Equation 7.55, which is also given in Fig. 7.1. If R_h is not achievable, an upper limit must be placed on it by placing a limit on α. It is necessary, firstly, to evaluate the optimum α. This is done below for the two-data-rate variable-duration intermittent FSK system.

If perfect prediction and noiseless feedback are assumed, then $\lambda=1$ and $h=o$ and Equation 7.23 simplifies to

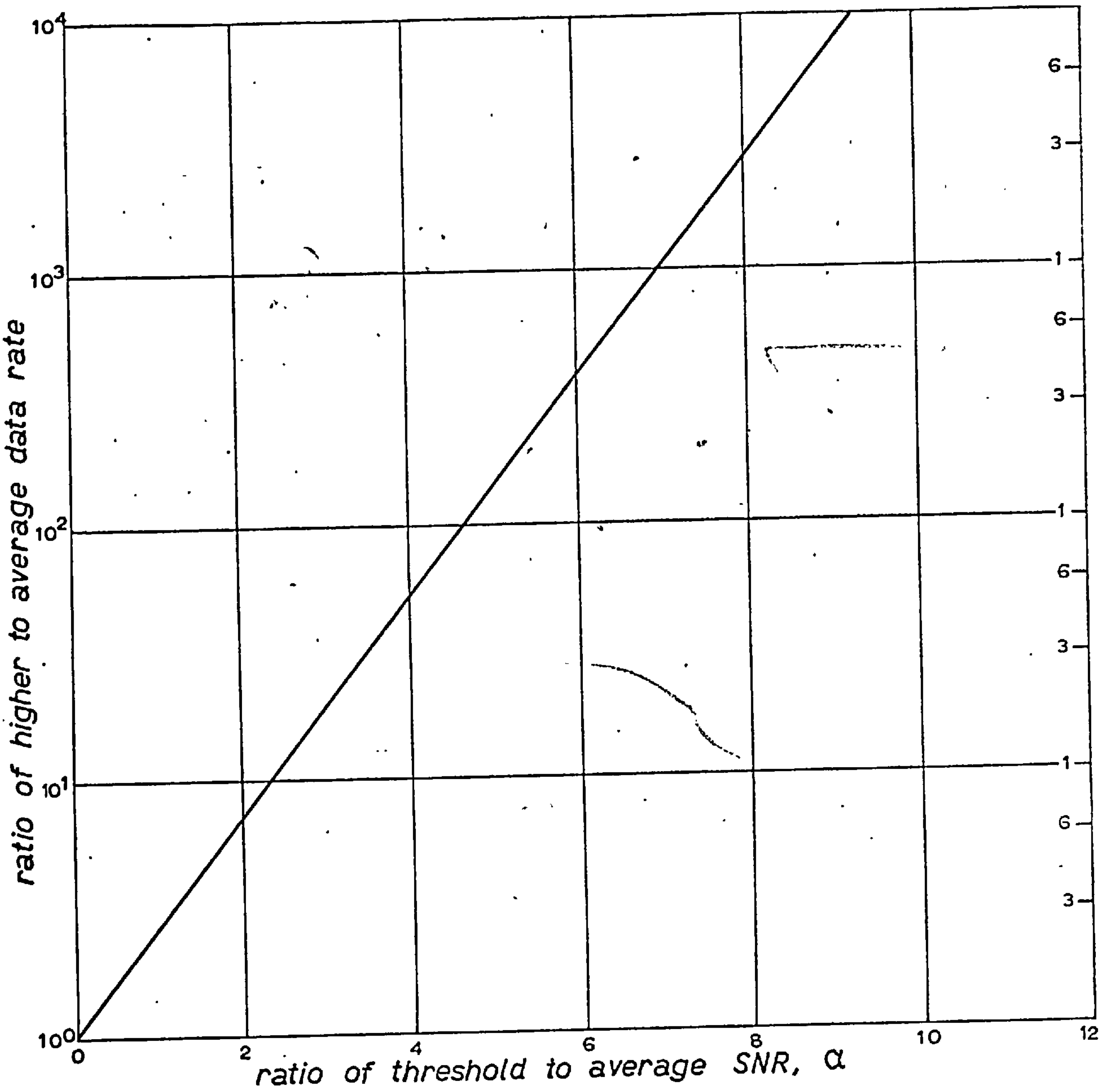


Fig. 7.1. Ratio of Higher to Average Rate Versus the Cut-Off Ratio, α

$$P_{e2}^{INT} = \frac{1}{1+U_o \bar{e}^{\alpha}} \exp\left[- \frac{\alpha U_o \bar{e}^{\alpha}}{2} \right] \qquad ...7.56$$

where α has been written for $\hat{y}$ in order to maintain the same symbolism of the present section.

On differentiating Equation 8.5 and setting the derivative equal to zero, it is found that the optimum α must satisfy the equation

$$U_o = \frac{2 \alpha e^{\alpha}}{1-\alpha} \qquad ...7.57$$

If the solution of Equation 7.57 is denoted by α_{opt} it is seen from this Equation that α_{opt} is always less than unity and that $\alpha_{opt} \to 1$ as $U_o \to \infty$. Fig. 7.2 shows the curve on which α_{opt}, which gives the minimum probability of error, must lie for a given U_o.

Since $\alpha_{opt} < 1$, it follows from Equation 7.55 that the ratio, η_h, of the operating to the average rate for an optimum intermittent system with a fixed average rate must satisfy the quality

$$\eta_h \leqslant e \qquad ...7.58$$

Thus, in order to maintain a constant average rate and minimize the probability of error, the maximum value of η_h must not exceed e and, therefore, R_h must satisfy the

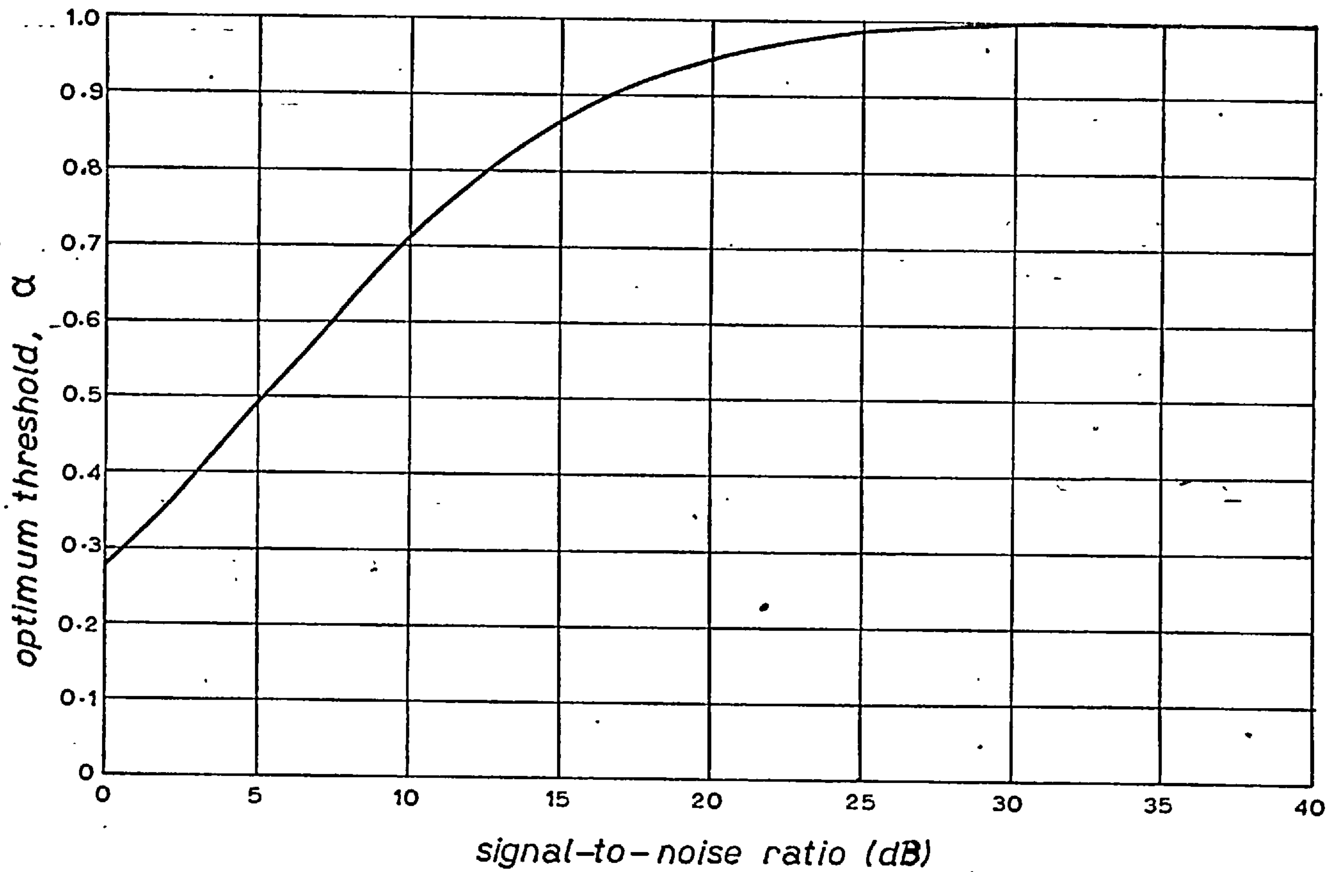


Fig. 7.2. The Optimum Threshold for the Intermittent FSK System.

inequality

$$R_h \leq e.\ R_o \qquad ...7.59$$

For channels with average SNRs greater than 25db, it is seen from Fig. 7.2 that the approximation

$$R_h \simeq e.\ R_o \qquad ...7.60$$

holds. In this case, the operating data rate is required (for constant average rate and minimum probability of error) to be approximately 2.718 times the average rate. In certain transmission situations, namely, those in which the desired R_o is high and bandwidth is restricted, the value of $R_h \simeq 2.718 \times R_o$ may not be physically attainable. It is for these cases that an upper limit say R_{max}, to the value of R_h can take is necessary. From Equation 7.55 it is seen that this can be done by placing a limit on the maximum value of α.

Using Euqations 7.55 and 7.56 the effect, on the probability of error, of placing an upper limit on the operating data rate was evaluated and the results given in the curve of Section 8.6.

Though the analysis carried out in this section is for the noiseless feedback case, from the computations of the results of Section 7.2.1, it was found that the optimum cut-off ratio, α_{opt}, for the noisy feedback case remains close to unity. The condition under which the introduction of the upper limit to the operating rate would become necessary is the same as that given above,i.e, the condition that Equation 7.60 cannot be satisfied because of physical limitations.

The discussion in Section 8.6 will thus include both noiseless and noisy feedback cases.

7.4.3. Two-Data-Rate Non-Intermittent Systems With an Upper Limit on the Higher Data Rate

The expressions for the average data rate R_o in terms of the two data rates of a 2-data-rate non-intermittent system are given by Equations 7.39 and 7.50. On rewriting these equation using the symbolism of the previous Section, the following are obtained for the average rate of variable-duration FSK and variable-level ASK systems respectively,

$$R_o = R_\ell\ (1-e^{-\hat{h}})(1-e^{-\hat{y}}) + R_h\ e^{-\hat{y}-\hat{h}} \qquad \ldots 7.61$$

$$R_o = R\ \log_2 m_1 \left[1-e^{-u_c}\right] + R\ \log_2 m_2\ e^{-u_c} \qquad \ldots 7.62$$

Both of these Equations can be expressed in one form, that is, in the form

$$R_o = R_h \left[k + e^{-\alpha}\ (1-k)\right] \qquad \ldots 7.63$$

where, as before, $\alpha = {}^{u}T/U_o$

and $k = R_\ell/R_h$

Thus,

$$\eta_h = \frac{R_h}{R_o} = \frac{1}{k + e^{-\alpha}(1-k)} \qquad \ldots 7.64$$

From Equation 7.63, the effective SNR, u_2, when the upper rate, R_h, is being used can be expressed as

$$\begin{aligned} u_2 &= M_p \cdot T_h \\ &= M_p \cdot \frac{1}{R_h} \\ &= \frac{M_p}{R_o}\left[k + e^{-\alpha}(1-k)\right] \\ &= U_o\left[k + e^{-\alpha}(1-k)\right] \qquad \ldots 7.65 \end{aligned}$$

T_h is the effective duration of the transmitted symbols when R_h is being used

M_p is the power in each symbol

$U_o = M_p/R_o$.

The effective SNR, u_1, when the lower rate R is in use can be expressed in terms of u_2 as

$$u_1 = M_pT = M_p.T_h.\frac{T_\ell}{T_h} = \frac{u_2}{k} \qquad \ldots 7.66$$

where T_{ℓ} is the effective duration of symbols when R_{ℓ} is being used for transmission.

NOTE 1. that when k=o, Equation 7.64 reduces to Equation 7.55, that is the system becomes an intermittent system using R_h

2. that when k=1, the system becomes a fixed-rate system using R_{ℓ} continuously.

On using Equation 7.64, a family of constant data rate curves for the 2-data-rate non-intermittent systems are evaluated and given in Fig. 7.3 for different values of k. The average rate, R_o, is taken to be determined by the system specification.

Strictly, to find the optimum cut-off ratio, α_{opt} for 2-data-rate non-intermittent system should be substituted into Equations 7.38 and 7.49, and the optimum α found by differentiation as was done for the intermittent system. However, it is already known from the computations of Equations 7.38 and 7.49, carried out for the case when no upper limit was placed on R_h, that the optimum threshold is close to unity for high SNRs (i.e. for SNRs greater than 25 db in this case).

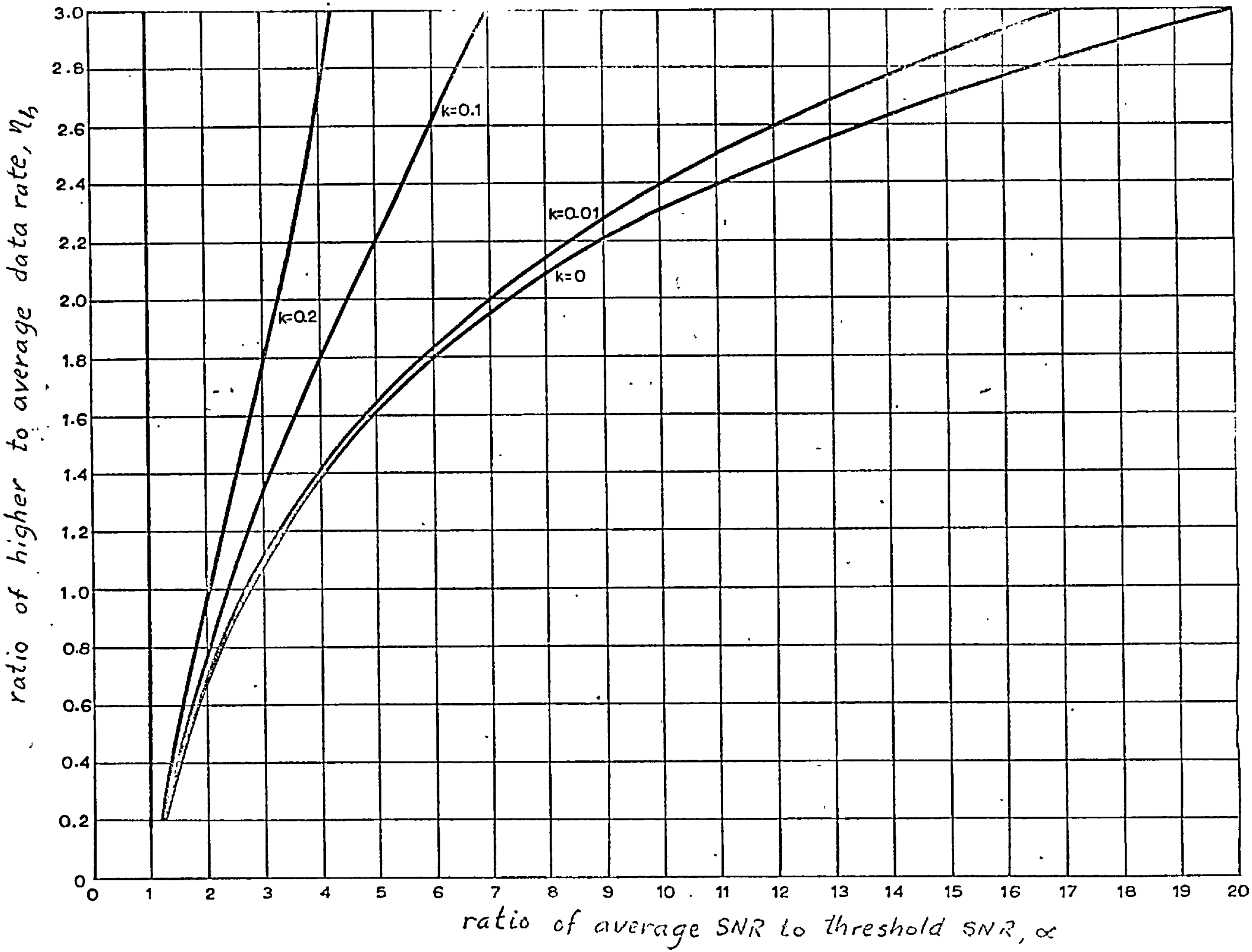


Fig. 7.3. Constant Data Rate Curves for 2-Data-Rate Non-Intermittent

Thus, a similar procedure as that already given for the intermittent system is followed when deciding whether or not to place an upper limit on the higher data rate, R_h.

If an upper limit is placed on the higher data rate, then the maximum value of α (corresponding to the R_{hmax}) is used in Equations 7.64, 7.65 and 7.66 to evaluate u_2, u_1 and k. Using these results in Equations 7.38, the probability of error, when $R_h \leq R_{hmax}$, is evaluated for the 2-data-rate variable-duration FSK system. The curves which were obtained from these calculations will be given in Section 8.6.

8. COMPUTATION AND DISCUSSION OF RESULTS

8.1 Introduction

The formulae that were obtained in the previous Chapter for the performance of two-data-rate variable-duration system are used in this Chapter to find the minimum probability of error by optimizing certain transmission parameters. The overall system, however, is a sub-optimum system because, as explained in Chapter 5, the filter used to predict the fades is a sub-optimum filter. The optimization procedure for computing the minimum probability of error is given in the following section.

The curves obtained from these computations are presented in this Chapter and a discussion of these curves is given. Comparisons are then made between the various types of two-data-rate systems and the existing variable-rate systems. Lastly, a short discussion is given on the performance of two-data-rate systems in which an upper limit is placed on the higher data rate.

As was mentioned in Chapter 5, the results given in this Chapter are only those of two-data-rate systems. Because there is a sharp increase in the complexity of the optimization problem with increasing n, the general solution was not obtained. Even with n=3, it was found that the optimization of the thresholds over the various channels became a very difficult problem. In any case, the results obtained for three-data-rate variable-duration system after lengthy computations were very close to those of the two-data-rate systems. These results have, therefore, not been given in the thesis.

8.2 The Optimization Procedure for Minimizing the Probability of Error for a Two-Data-Rate Variable-Duration System

When using Equations 7.23 and 7.38 to find the probability of error for the two-data-rate variable-duration system, it is necessary to adjust the system parameters if the probability of error is to be minimized. This section sets out in detail the procedure of obtaining the minimum probability of error for two-data-rate variable-duration system.

From physical considerations the total power, that is, the sum of the powers in the data channel, the pilot-tone channel the SYNC channel and the feedback channel must be limited to maximum value, say P. Expressed symbolically, this leads to the Equation

$$P = M_p + E_p + S_p + F_p \quad \ldots\ldots\ldots\ldots\ldots 8.1$$

where,

P is the total available power

M_p is the power over the data channel

E_p is the power over the pilot-tone channel

S_p is the power over the SYNC channel

and F_p is the power over the feedback channel

Fig. 8.1 is the flow diagram of the optimization procedure that was followed in the computation of the probability of error for two-data-rate variable-duration system. The four basic parameters of a two-data-rate system are the average rate, R_o, the total power, P, P_{e2}, and the probability, P_d, of making an error over the SYNC channel. Any three of these four parameters can be fixed and the fourth optimized, subject to the constraint of the three fixed parameters. Since the main interest in this thesis is the minimization of the probability of error, in the flow diagram in fig. 8.1, R_o, P and P_d are chosen as fixed and the probability of error, P_{e2}, minimized.

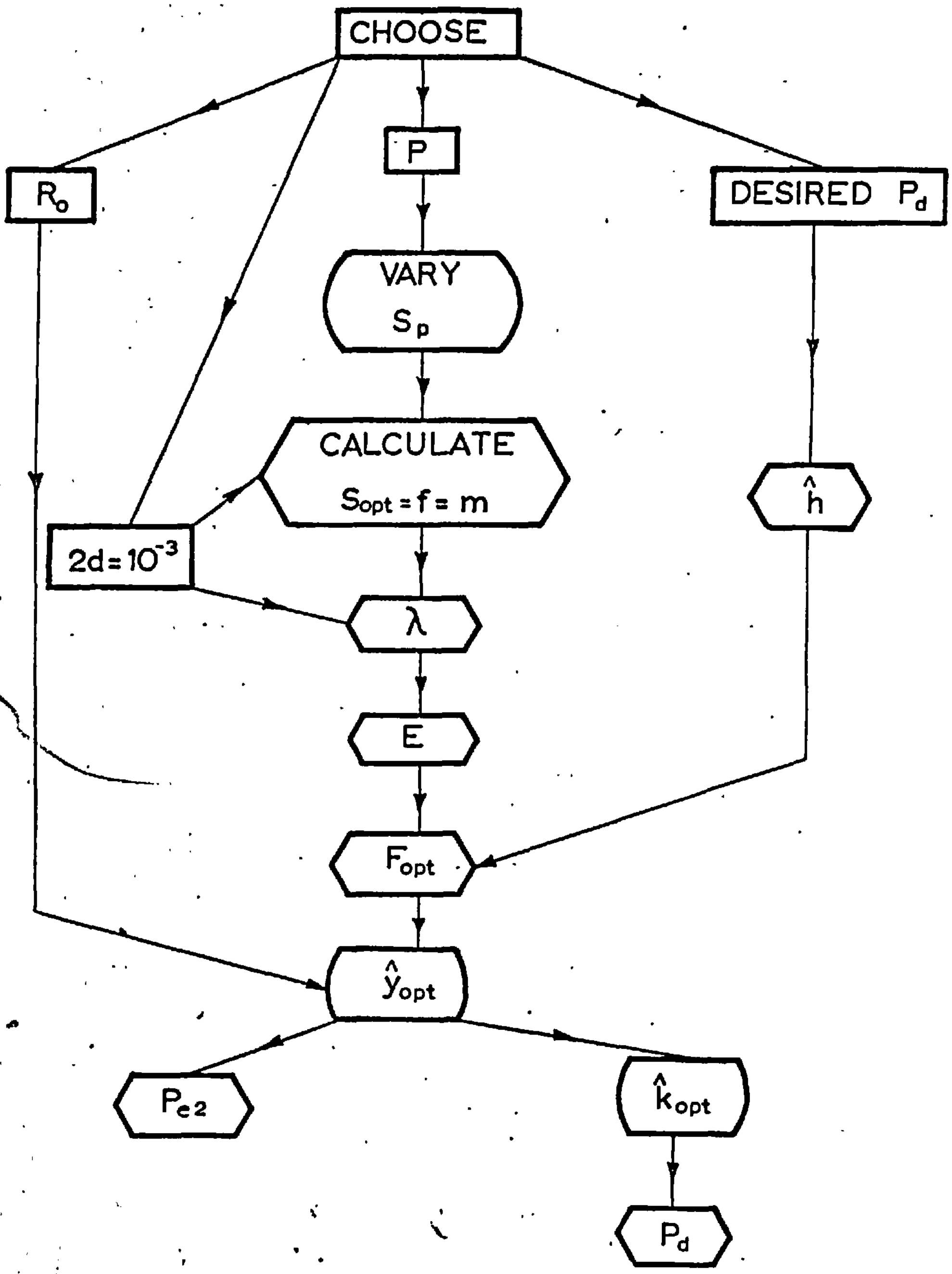


Fig. 8.1. Flow Diagram for Evaluating the Performance of the Suboptimum 2-Data-Rate Variable-Duration Systems.

It is important that the power, S_p, over the SYNC channel be sufficiently large to make the probability, P_d, of making an error over this channel less than or equal to P_D, the desired (or chosen) value of P_d. The actual value of P_d is computed using Equation 7.14.

To begin with, the optimization of the power over the SYNC channel is considered in the following section.

8.2.1 Optimization of the Duration and the Power of the SYNC Pulse

The duration of the SYNC pulse, s, is chosen so as to maximize the correlation, μ, between the output of the linear predictor and the output of the SYNC signal matched filter. From Equation 5.29, μ, can be written as

$$\mu = \frac{\lambda(2d + f + \frac{s}{2})}{1 + \frac{1}{s\,S_p}} \qquad \text{...........8.2}$$

Because λ(τ) decreases with τ, it can be seen from Equation 8.2 that too large a value of s decreases the correlation, μ. Also, two small a value of s clearly decreases the correlation because of the effect of s on the denominator of the right-hand side of Equation 8.2. As the probability, P_d, of making an error over the SYNC channel is a function of the duration of the SYNC pulse, it is evident that the optimization of s is also an approximate optimization of P_d.

On substituting Equation C.24, derived in Appendix C, into Equation 8.2, the correlation, μ, becomes

$$\mu(s) = \frac{1 - \frac{1}{2}\left(\frac{2\pi}{\tau_f}\right)^4 \left(2d + f + \frac{s}{2}\right)^4}{1 + \frac{1}{s\,S_p}} \qquad \text{......8.3}$$

Observations over practical circuits (Ref. 80, p. 343) indicate that a typical value for the fading rate is 10 Hz, i.e. $\tau_f = 0.1$. This value of τ_f was assumed throughout the calculations.

To find the optimum value of s, Equation 8.3 is differentiated with respect to s and the derivative set equal to zero, that is, the Equation

$$\frac{d\mu\ (s)}{ds} = 0 \qquad \text{........8.4}$$

is solved. For convenience the duration of the feedback signal f is set equal to the duration of the SYNC pulse, s, before obtaining the solution to Equation 8.4. This constraint has been added to simplify the calculations. On performing the differentiation in Equation 8.4, a fifth order polynomial in s is obtained. From physical considerations, the single positive zero of this fifth order polynomial is the value of s which maximizes the correlation, $\mu(s)$.

When finding the zeroes of the polynomial the value of the round trip delay, 2d, was specified to be 10^{-3} sec. This delay corresponds to a medium-range data link of length 279 miles. The effect of varying the round trip delay was investigated. It was found that longer trip delays results in the requirement that more power be available for use over the feedback and the SYNC channels. Clearly, this leads to an increase in the total power required if the same improvement in performance over the fixed-rate system which transmits continuously is to be maintained irrespective of the length of the data link.

The optimum value of the duration of the SYNC pulse which is found by the procedure specified above is for a particular value of the power, S_p, over the SYNC channel. To obtain the optimum value of S_p, that is, to find the power, S_{opt}, which minimizes the probability of error, S_p must be varied and the optimization of s described above repeated for every value of S_p used. Curves showing the average probability of error, P_e, and the probability, P_d, of making an error over the SYNC channel as functions of the power over the SYNC channel, S_p, are given in fig. 8.2. From these curves it is seen that P_d decreases monotonically with S_p whereas P_e has a minimum value.

The shape of the probability of error curve can be explained physically as follows. Using small valves of S_p allows more power to be used over the data channel and this tends to decrease the probability of error. However, the smaller the value of S_p used the longer will be the duration of the SYNC pulse. This tends to increase the total prediction time which tends to decrease the correlation. The probability of error will tend to increase because of the decrease in correlation. A minimum could thus be expected for the probability of error curve, where the two opposing factors balance.

Clearly, the best value of S_p is that which results in the minimum probability of error provided that the actual value of the probability, P_d, of making an error over the SYNC channel is at or below its desired value, that is, provided $P_d \leq P_D$ (see

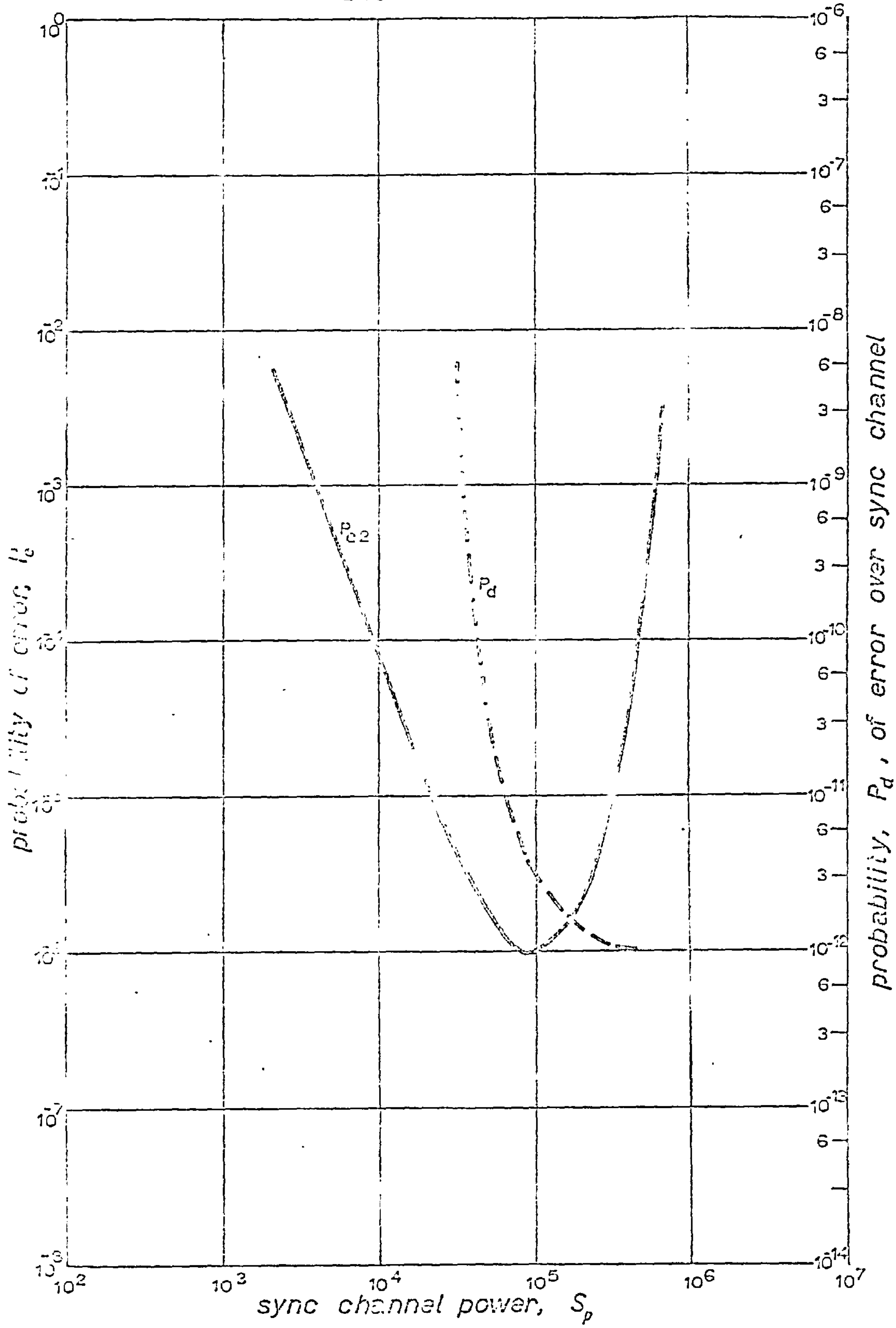


Fig. 8.2. The Probabilities of Error, P_{c2} and P_d as functions of the SYNC power, S_p

Fig. 8.2). If this condition is not satisfied, then the value of S_p for which the desired value of P_d is attained must be used.

Prediction of fades is another very important factor that controls the performance of variable-duration systems. The quality of prediction is controlled by the power over the pilot-tone channel.

8.2.2 The Pilot-Tone Channel Power

When deriving the characteristic, $\lambda(\tau)$, of the prediction filter in Appendix C, a condition was imposed on the minimum power, E_p', that is necessary for good prediction. In particular, for $\tau_f = 0.1$ and for a maximum prediction error of less than 1%, E_p' was required to be 1.18×10^4 (see Table C.1 in Appendix C). In this thesis, all the computations carried out for two-data-rate variable-duration systems are with the value of the power over the pilot tone channel ~~is~~ fixed at 1.2×10^4 times the average power.

The remainder of the total power is divided between the feedback and the data channels so as to obtain a minimum probability of error, P_{e2}.

8.2.3 The Power Over the Feedback and the Data Channels

The remainder of the total power is given by

$$P' = P - S_p - E_p \qquad \text{...........8.5}$$

From Equation 7.14 it is evident that the power, F_p, over the feedback channel does not have a strong influence on P_d. The feedback power is therefore chosen only on the basis of its effect on the average probability of error, P_{e2}, and its effect on the average rate, R_o.

The expressions (Equations 7.27 and 7.46) for the optimum power, F_{opt}, over the feedback channel, that is, the feedback power gives the minimum probability of error for a fixed average rate, R_o, were derived for the two-data-rate variable-duration intermittent and non-intermittent systems. Thus from Equations 8.5 and 8.1, the power, M_p, that is used over the data channel is given by

$$M_p = P' - F_{opt} \qquad \text{...........8.6}$$

Having determined the optimal distribution of the total available power, P, into the SYNC channel, the pilot-tone channel, the feedback channel and the data channel, the probability of error and the probability, P_d, of making an error over the SYNC channel can be evaluated using Equations 7.38, 7.23 and 7.14.

8.2.4 The Optimization of the Thresholds of the Feedback, the SYNC and the Pilot-Tone Channels

In order to minimize the probability of error using the formulae cited above, the thresholds used on the feedback channel, the SYNC channel and the pilot-tone channel have to be adjusted.

The optimization of the threshold used over the feedback channel has already been considered in the previous chapter. Equations 7.27 and 7.46 give the expressions for the optimum feedback channel thresholds.

Initially an attempt was made to optimize the choice of the threshold, k, (and hence $\hat{k}$)on the SYNC channel in the sense of obtaining minimum P_d for a chosen power, S_p, over this channel. The calculations were lengthy and the optimum k was found to be given by Equation 7.19, when the minimum P_d was close to its desired value P_D.

As was explained in Sections 7.2.1 and 7.3.1, the pilot-tone threshold, $\hat{y}$, which minimizes the probability of error is found by a trial method. This procedure ignores the effect of $\hat{y}$ on P_d but this is justified since, as can be seen from Equation 7.14, $\hat{y}$ does not have a strong influence on P_d for low values of P_d.

The optimization procedure set out in this section was used when computing the probabilities of error, P_{e2} and P_{e2}^{INT}, for two-data-rate variable-duration incoherent FSK systems. The results are given in the following section.

8.3 Discussion of Results

8.3.1 The Two-Data-Rate Variable-Duration Systems

This section is devoted to the discussion of the performance of the two-data-rate variable-duration incoherent FSK systems. Comparison of variable-duration systems with other systems are given in sections 8.3.3, 8.3.4 and 8.3.6. The curves used for the discussion in this Section are given in Figs. 8.3 - 8.6.

For two-data-rate systems, the most important parameter used in this thesis is the ratio, k, between the lower and the higher data rates, that is,

$$k = \frac{R_\ell}{R_h} \qquad \text{............8.7}$$

where R_h is the upper rate and R_ℓ is the lower rate. The value of k is always less than or equal to unity. When k is unity it is implied that transmission is carried out continuously and that $R_\ell = R_h = R_o$, where R_o is the average rate. When k is zero, the system is an intermittent system, and when data is transmitted, it is sent at a rate R_h so as to maintain an average rate of R_o.

Figs. 8.3 and 8.4 show the effect of continuously varying the ratio k from its maximum value of unity to a value approaching its lowest value of zero. These curves are given for average SNRs 10, 15 and 20 dB, and a fixed average rate R_o. The general S-shape of the curves indicates, firstly, that not much improvement can be expected from switching between two data rates that are nearly equal and, secondly, that the improvement obtained from decreasing k saturates at a given value of k depending on the average detection SNR. To illustrate the first point it is seen from Figs. 8.3 and 8.4

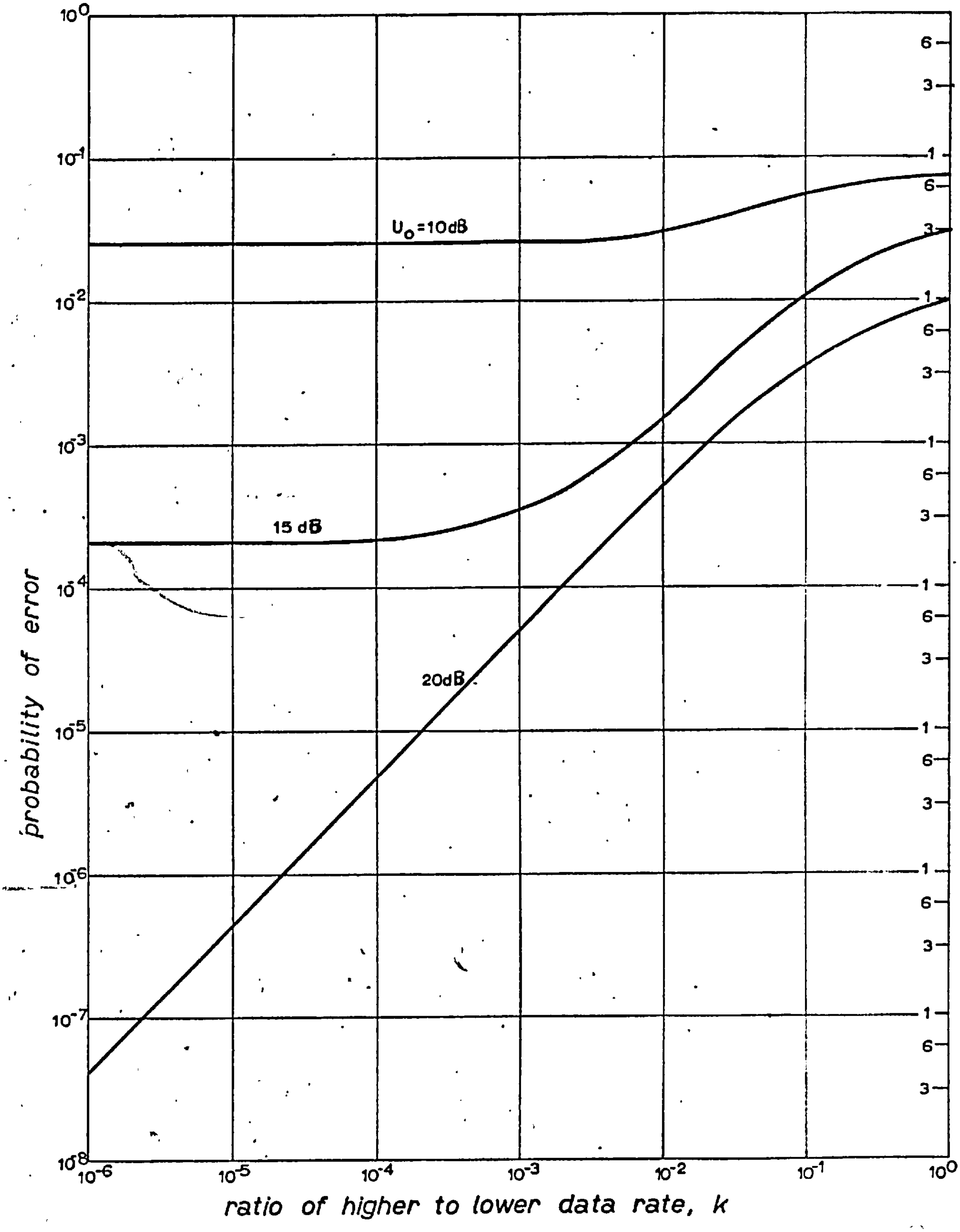


Fig. 8.3. Probability of Error Versus k = Higher Rate/Lower Rate; Binary FSK System with Noiseless Feedback and Fixed Average Rate.

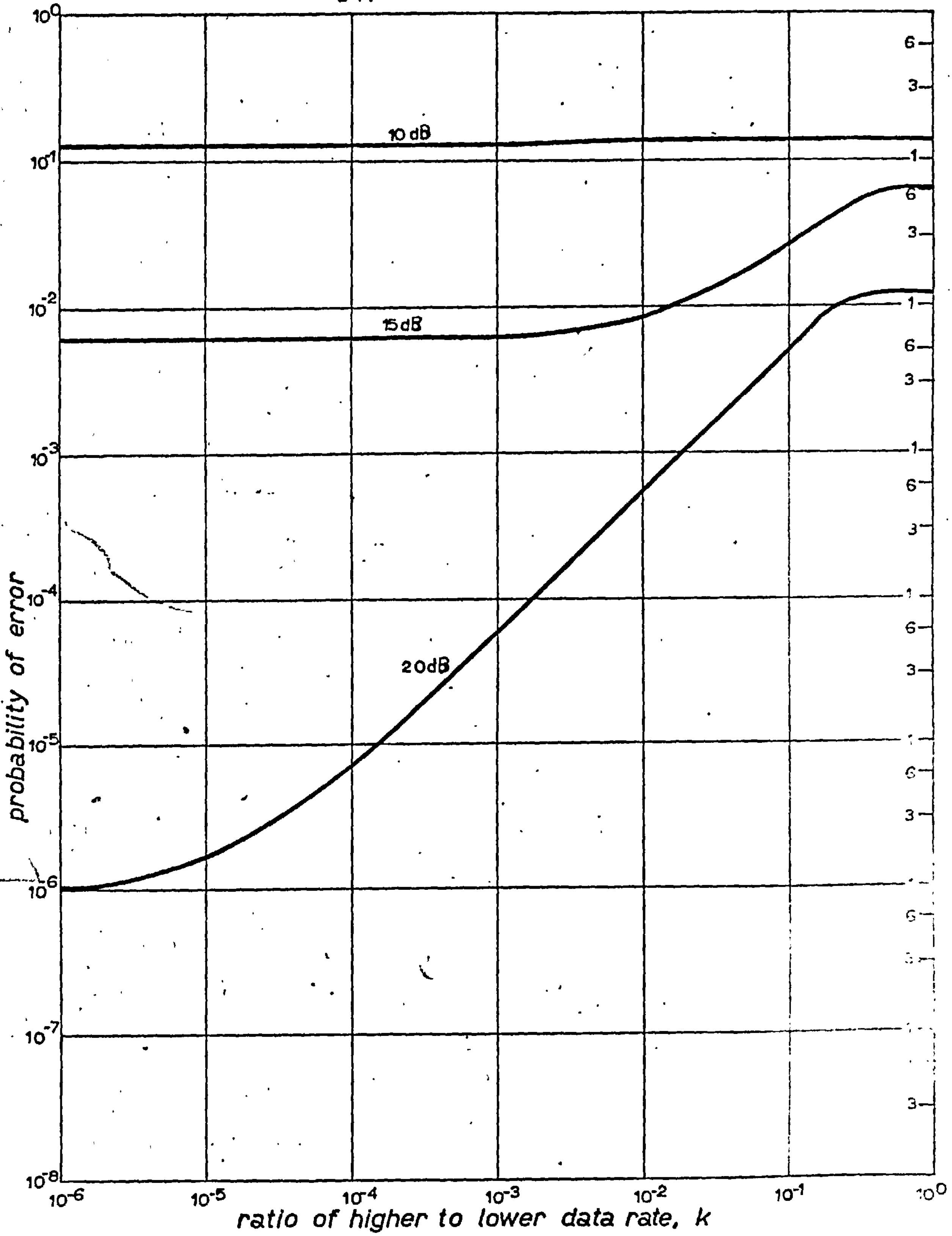


Fig. 8.4, Probability of Error Versus k = Higher Rate / Lower Rate; Binary SK System with Noisy Feedback and Fixed Average Rate

that virtually no improvement is obtained from switching between two data rates if the lower data rate is say $^3/_4$ of the higher data rate, i.e., if $k = 0.75$. To illustrate the second point it is seen from the middle curves (average SNR of 15 db) of Figs. 8.3 and 8.4 that no further improvement is obtainable when k is decreased below 10^{-4} for the noiseless feedback case, or, when k is decreased below 10^{-3} for the noisy feedback case.

Figs. 8.5 and 8.6 are curves of the average probability of error versus the average SNR for several constant values of k, assuming a fixed average rate R_o. These curves therefore show the performance of various two-data-rate variable-duration FSK systems characterised by different values of k. On a log-log scale, the general shape of the curves is an exponential shape for one extreme value of k, i.e., for $k = 0$, and a linear shape for the other extreme value of k of unity. For values of k between zero and unity the curves start off with an exponential shape and then change into a linear shape as the SNR is increased. It is evident from these general shapes that the optimum choice of the rates for a two-data-rate system is to make

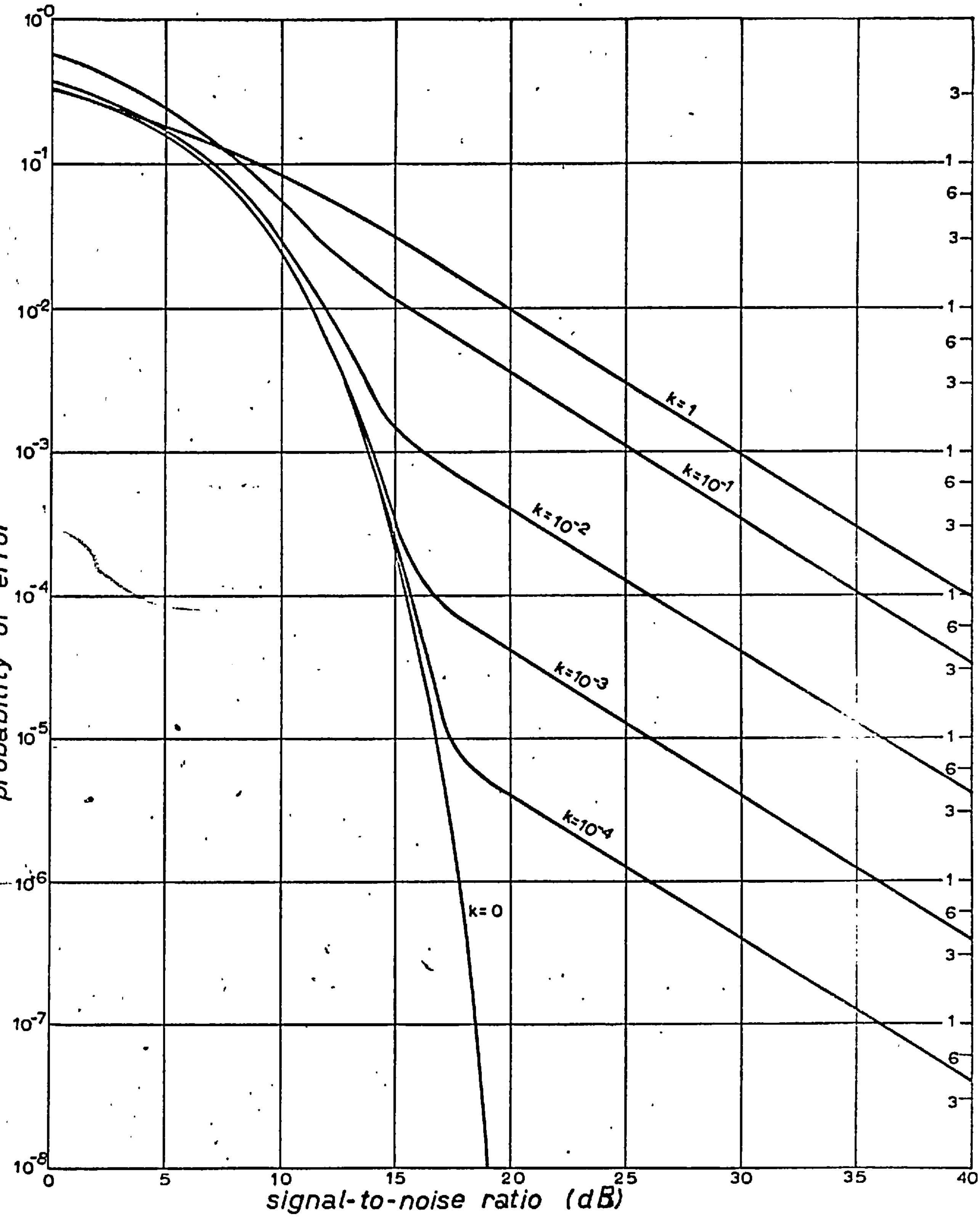


Fig. 8.5. The Performance of Binary FSK Systems with Given Ratios Between Higher and Lower Rates; Noiseless Feedback and Fixed Average Rate

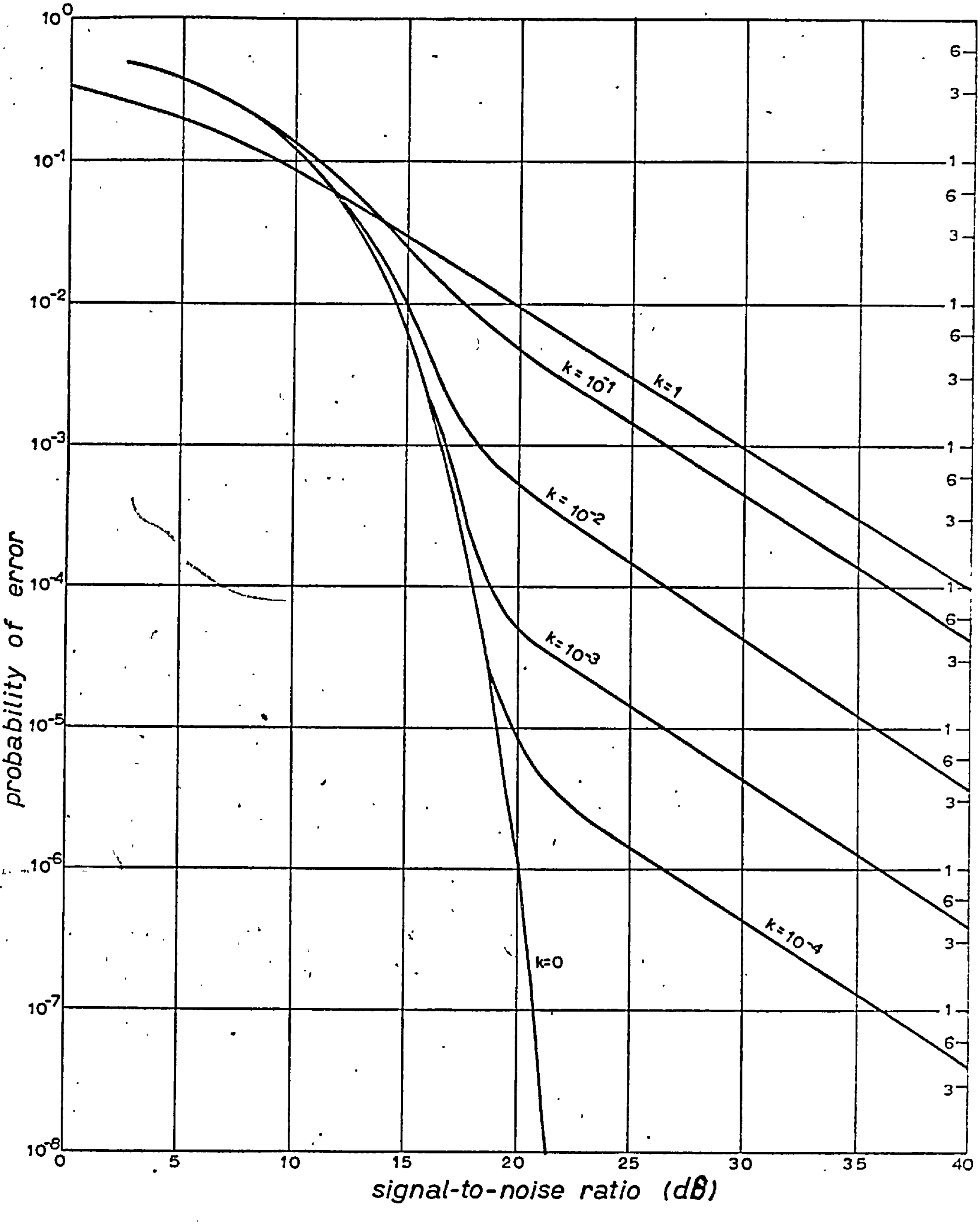


Fig. 8.6. Performance of Binary FSK Systems with Fixed Ratios between Higher and Lower Rate; Noisy Feedback and Fixed Average Rate

$R_\ell = 0$ and choose R_h so as to maintain the desired average rate, R_o. It is this choice that maintains an exponential decrease of probability of error with SNR for all values of the average SNR at the receiver. A two-data-rate system with a low k (10^{-2} or less) operates close to the optimum for high error rates but at low rates its performance suffers from the fact it does not completely stop message transmission during a deep fade.

To give quantitative results it is seen from Figs 8.5 that with noiseless feedback, the maximum attainable improvement with a value of $k = 10^{-1}$ is a saving of 4.5 dB transmitter power as compared with a fixed-rate system continuously transmitting. The corresponding saving in transmitter power with a noisy feedback is 3.7 dB. The maximum improvement is obtained for error rates below 10^{-2} for the noiseless feedback case and for error rates below 2.6×10^{-3} for the noisy feedback case. For equal to 10^{-2}, equivalent figures to those given above are a maximum transmitter power saving of 14 db for the noiseless feedback case and 12.4 dB for the noisy feedback case. The maximum improvement is obtained for error rates below 10^{-3} for the noiseless feedback case and for error rates below 5×10^{-4} for the noisy feedback case.

A major point of interest to note in these quantitative results is that Palmer, et al, obtained a maximum transmitter power saving of 15 dB (for noiseless feedback) for a continuous rate variation over a range of 1: 100. A two-data-rate system with k = 1/100 is seen from Fig. 8.5 to attain a maximum transmitter power saving of 14 dB for the noiseless feedback case. Thus discrete rate variation suggested in the thesis is seen to perform very close to a similar continuous rate variation. This is a very significant result since it is clearly much easier to realise a system using two discrete data rates than to realise a system which varies the data rate continuously between the two discrete data rates.

The effect of prediction on the performance of an intermittent system is shown in Fig. 8.7. It is evident from this Figure that the ability to form good prediction is required if worthwhile improvement in the performance above that of a fixed-rate continuous system is to be expected. It is seen, for example, that unless the prediction factor, λ, is greater than 0.6, very little improvement is obtainable.

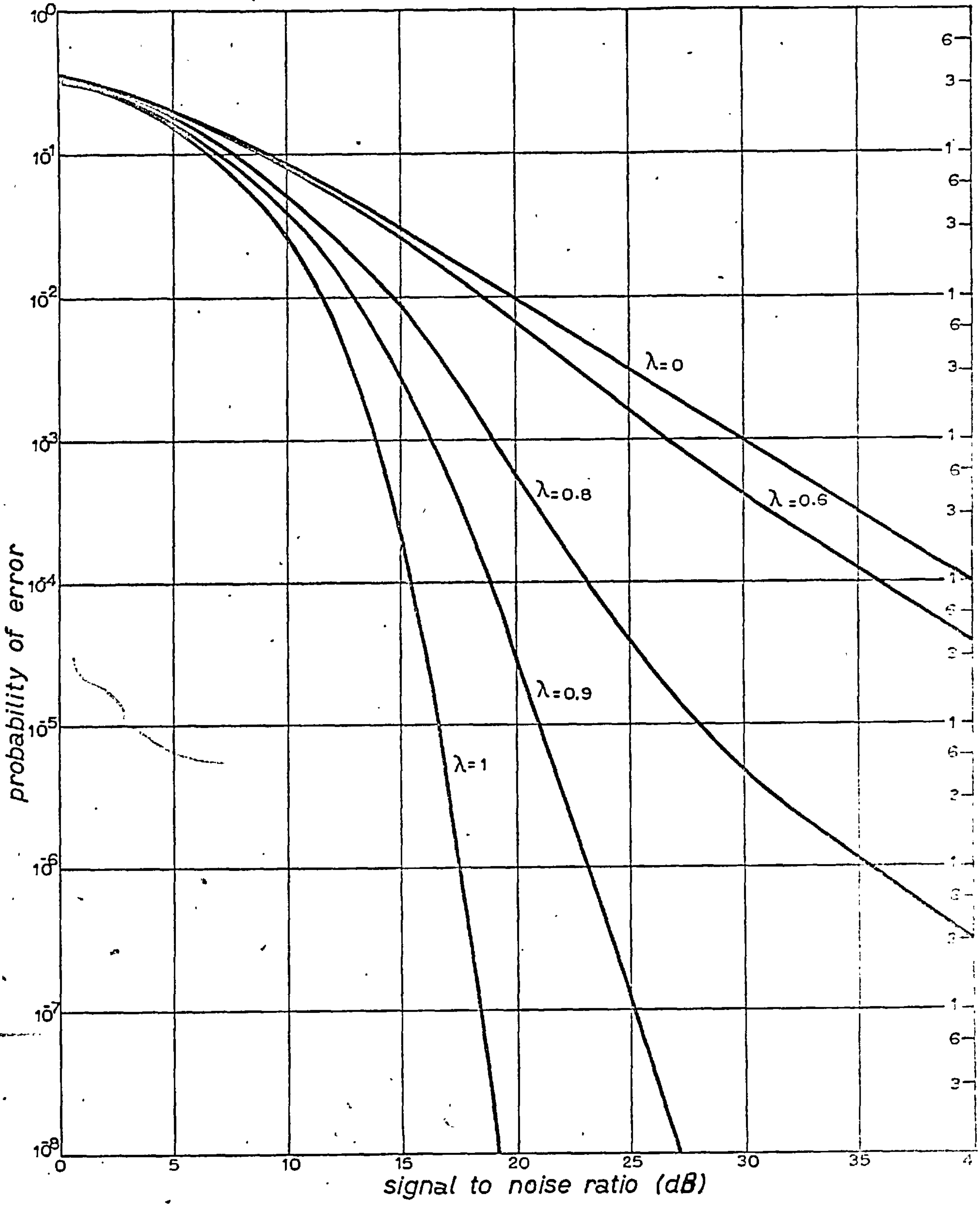


Fig. 8.7. The Effect of Prediction Factor, λ, on the Performance of Binary Intermittent FSK System with Fixed Average Rate

Lastly, the effect of noise in the feedback is seen to be that slightly more transmitter power is used than in the noiseless feedback to obtain the same performance, for values of k close to unity. For example for $k = 10^{-1}$, the extra transmitter power required is 0.8 dB. As k decreases, the extra transmitter power required because of the noise in the feedback increases. For an intermittent system (k = 0) 2.5 dB more transmitter power is necessary to achieve the same error rate as in the noiseless feedback case.

8.3.2 The Two-Data-Rate Variable-Level Systems

The results for two-data-rate variable-level systems, that is, systems in which the data rate is varied by switching between two different sets of signalling alphabets, will be considered in this section. The performance curves for these systems are shown in Figs. 8.8 - 8.12. These curves were computed from Equations 7.31 and 7.49 which were derived in the previous Chapter. The two numbers which appear on each of the curves denote the sizes of the two signalling alphabets used for the two-data-rate variable-level system. For example

(0,2) indicates that the curve refers to a binary intermittent system and (2,4) indicates that the curve refers to a system which switches between a binary signalling alphabet and a quarternary signalling alphabet.

From Fig. 8.8 it is seen that intermittent system operation improves the performance of a data transmission system from operating under flat-flat Rayleigh fading conditions from a linear decrease to an exponential decrease of the probability of error with the detection SNR. This leads to enormous saving in the transmitter power for the same error rate. For example, at an error rate of 10^{-2} the transmitter power saving obtainable due to an intermittent operation is 13 db and at an error rate of 10^{-3} the transmitter power saving increases to 20 dB. It is also seen that the saving in the transmitter power is about the same for all the intermittent systems considered, that is, for (0,2), (0,4), (0,8), (0,16) and (0,32) variable-level systems. From Fig. 8.9 it is seen that for low error rates, the exponential curves of an intermittent system operating under flat-flat Rayleigh fading conditions lies only 4.3 dB away from the exponential curve for the same system operating under nonfading conditions.

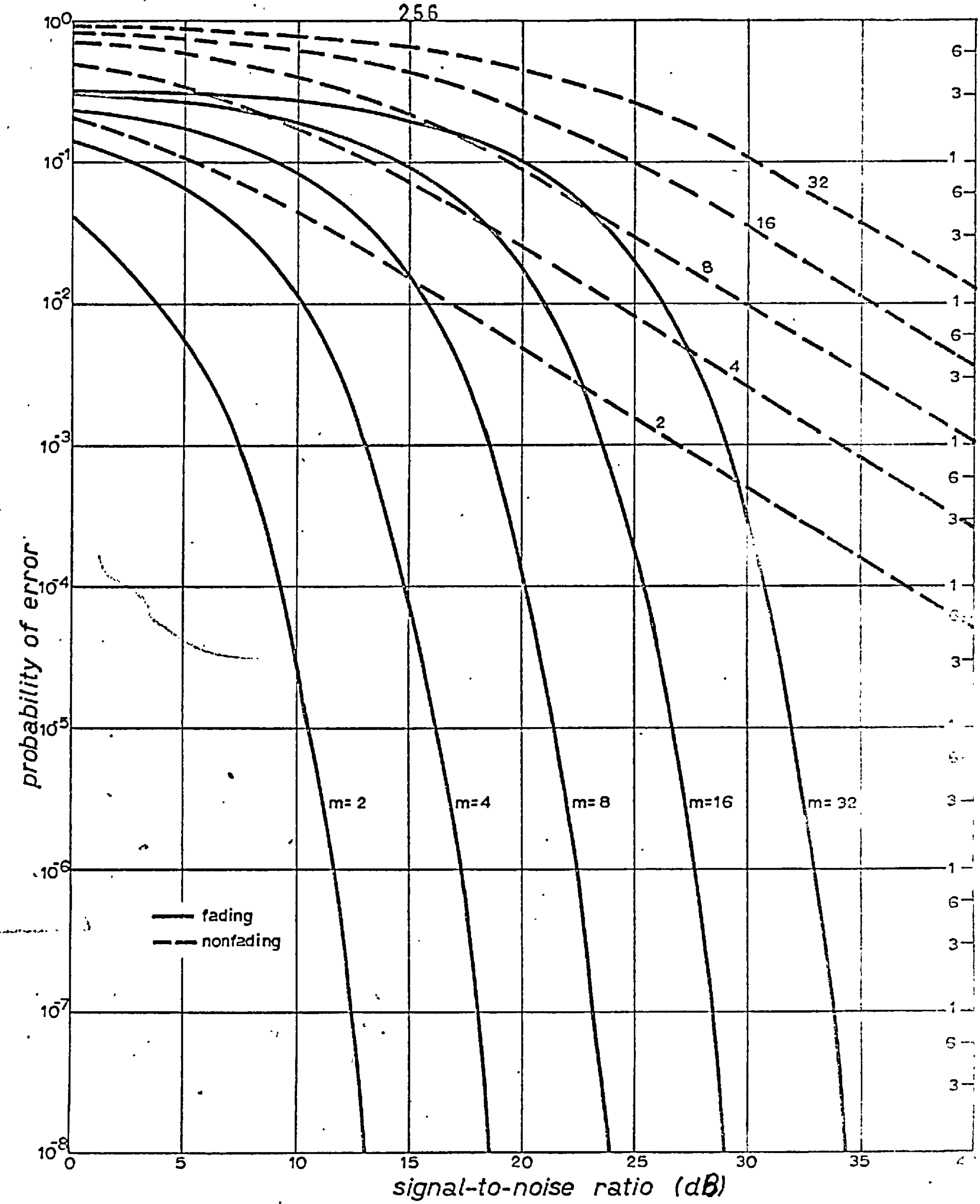


Fig. 8.8. Performance of m-Amplitude Intermittent Coherent ASK Systems with Fixed Average Rate

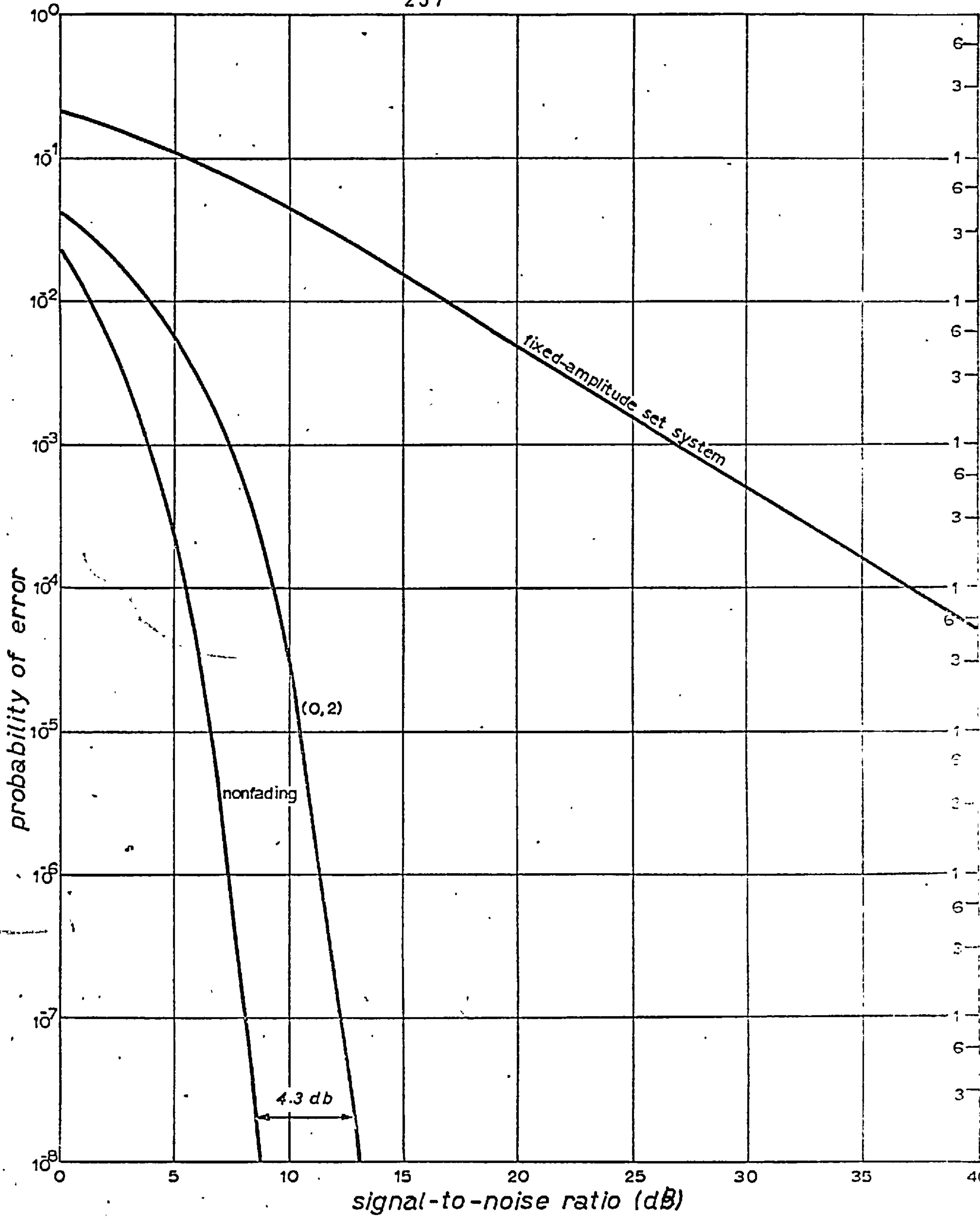


Fig. 8.9. Comparison between Nonfading, Binary Intermittent and Binary Fixed-set Coherent ASK Systems each with a Fixed Average Rate

The curves giving the performance of two-data-rate non-intermittent systems (Figs 8.10 and 8.11) show general similarity to those of variable-duration systems considered in the previous section. On a log-log scale, they start off as exponential curves and then change into linear curves. As can be seen from Fig. 8.10, these curves have interesting cross-overs. It is seen, for example, that a (2,8) variable-level system performs better than a (2,4) variable-level system only when the average detection SNR is greater than 15.5 dB; otherwise, its performance is worse than that of the (2,4) variable-level system. Below error rates of 10^{-3} the (2,8) system has a 2.7 dB power advantage over the (2,4) system, whereas at error rates higher than 10^{-2} the latter system has about the same power advantage over the former system.

Generally, the larger the signalling alphabet used with a binary alphabet in a two-data-rate system, the worse is the performance of the system at high error rates but the better the performance of the system at low error rates. It can be concluded that over channels with low average SNRs (below 15 dB), it is more advantageous to use a (2,4) variable-level system whereas over

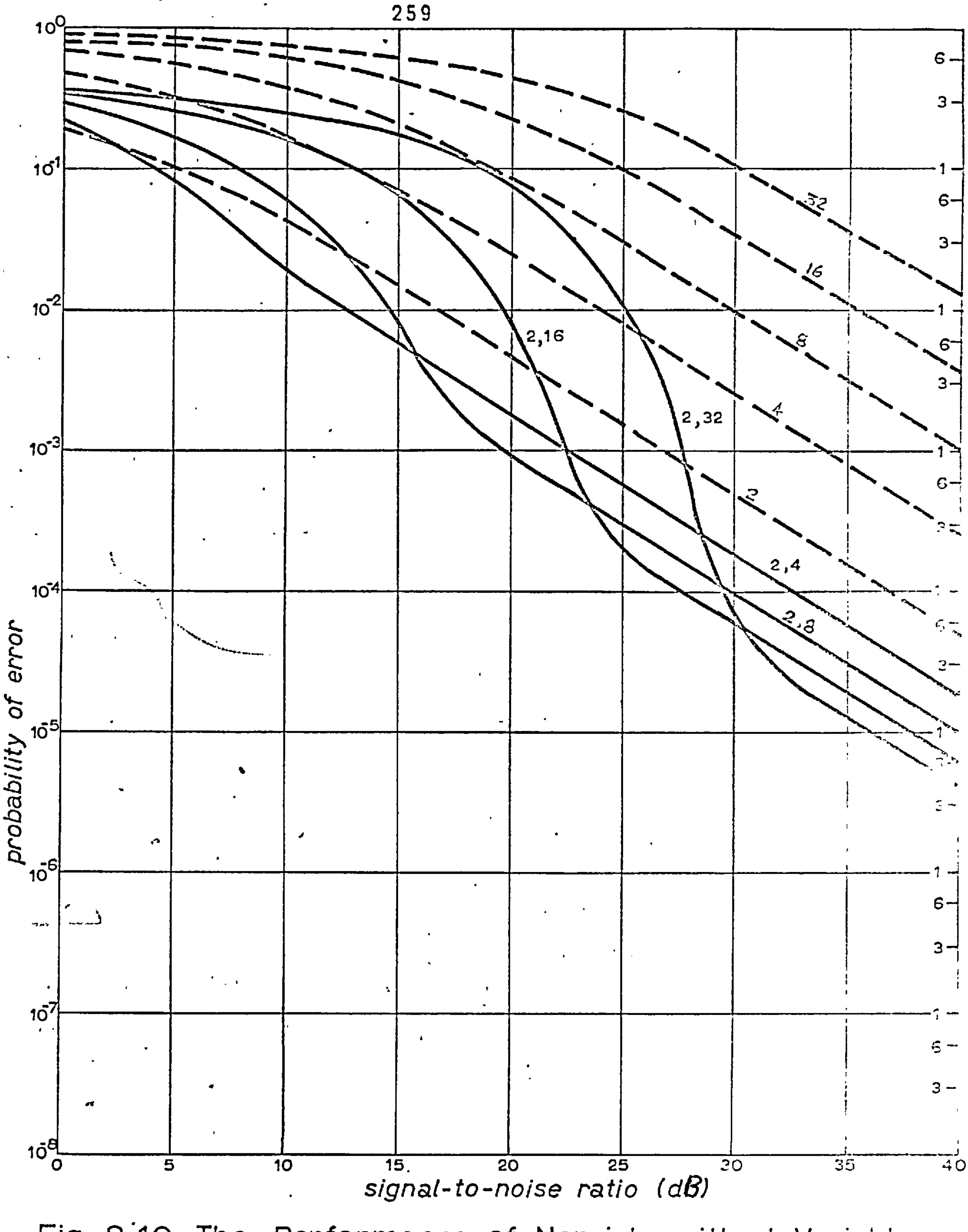


Fig. 8.10. The Performance of Non-intermittent Variable-Level ASK Systems each with a Fixed Average Rate

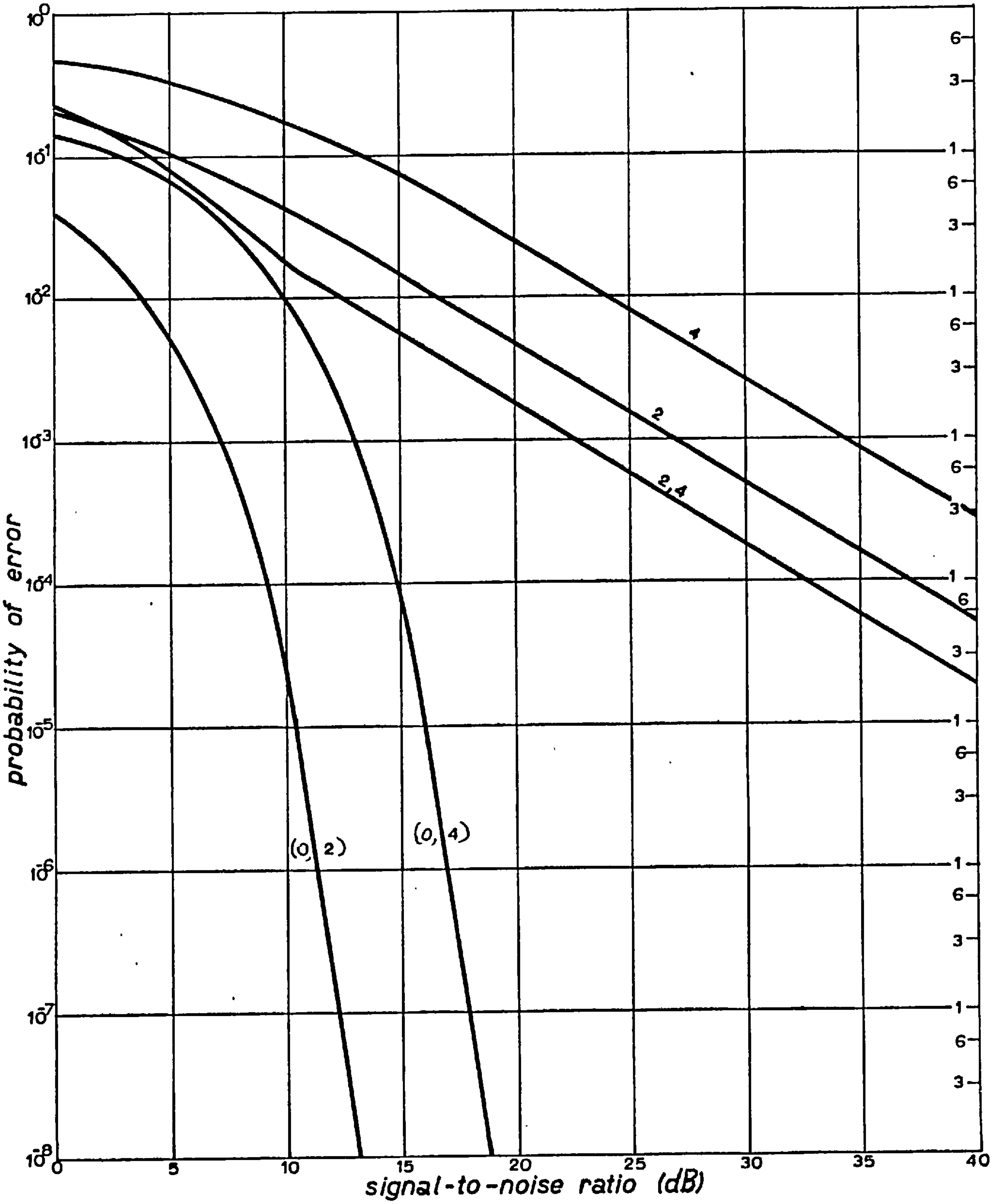


Fig. 8.11. Comparisons between simple Variable-Amplitude Systems with Fixed-Amplitude-Set Systems

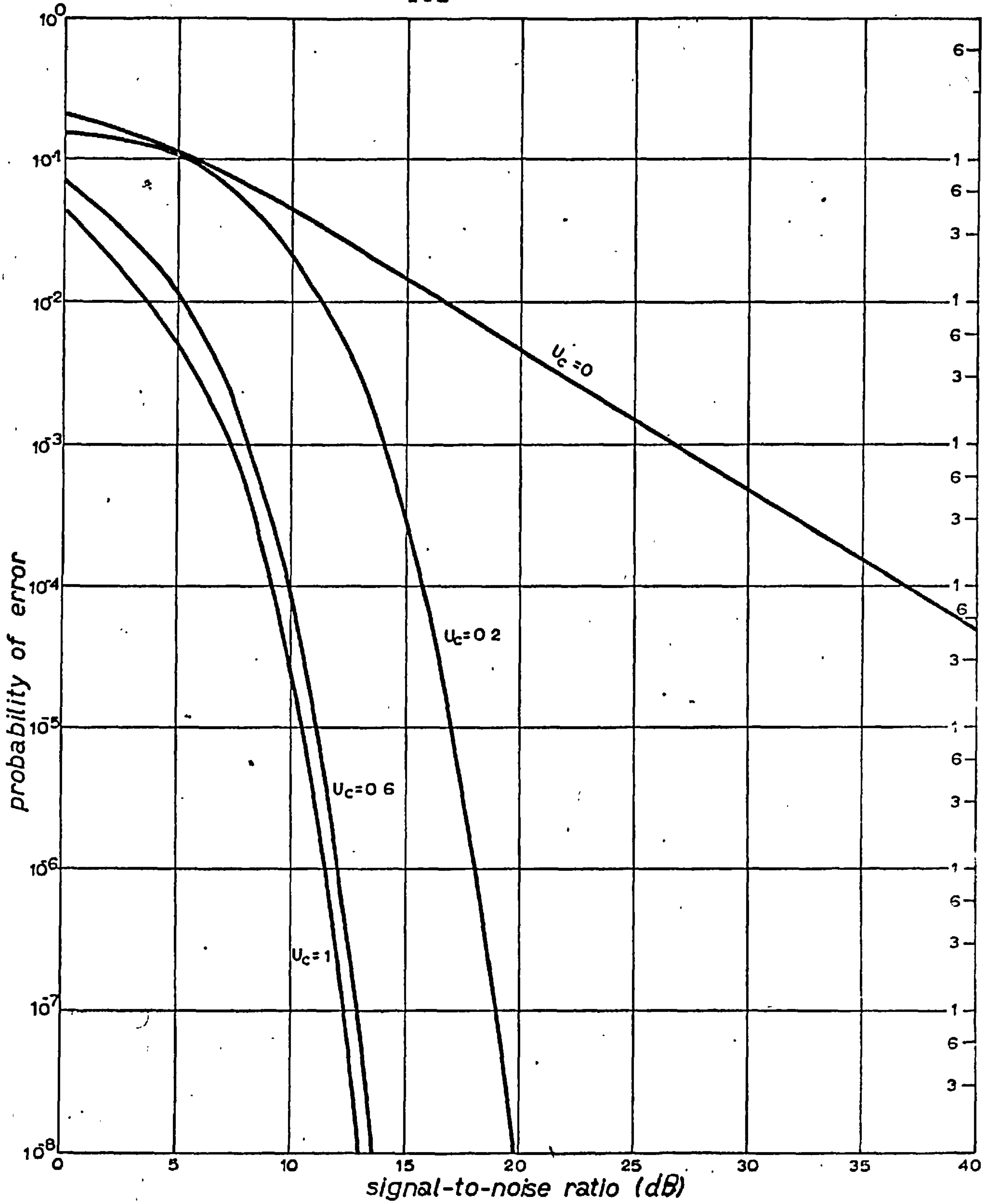


Fig. 8.12. Effect of Varying the Cut-Off Ratio, α, on the Peformance of Coherent ASK System

channels with high average SNRs, it is more advantageous to use a (2,m), m > 4, variable-level system, where the number m depends on the actual value of the average SNR of the system.

Fig. 8.12 shows the effect of adjusting the parameter $u_c = \left(\frac{\text{the threshold SNR, } u_T}{\text{the average SNR, } U_o}\right)$ which determines the percentage of the time transmission is carried out at the low and the high data rates. For an intermittent system, $u_c = o$, then clearly transmission is carried out continuously and the system becomes a fixed-rate system which transmits all the time it is in use. It is clear from Fig. 8.12 that a variable-level system is fairly insensitive to small changes in u_c so long as u_c is close to unity but for low values of u_c, namely, $u_c < .2$, any small changes in u_c is bound to have a large effect on the performance of the variable-level system.

8.3.3 Comparison Between Two-Data-Rate Variable-Duration Incoherent FSK Systems and Variable-Level Coherent ASK Systems

The two completely different techniques, namely the variable-duration and the variable-level techniques, of achieving discrete rate variation that have been analysed in the previous three chapters are compared in this section. Before carrying out this comparison, however, the advantages and disadvantages of each technique which must be borne in mind are summarized.

In order to attain good performance, the variable-duration technique requires transmission of very short pulses whenever the higher data rate is being used. It is clear from this that ~~generally~~ whenever bandwidth is at a premium, as is the case in congested bands, the variable-duration technique is at a disadvantage in comparison with the variable-level technique which does not require as large a bandwidth for the same average data rate. Further, varying symbol durations makes the problem of bit synchronisation more difficult as compared to a technique which uses fixed-duration signals.

The larger the ratio between the two data rates of a two-data-rate system, the greater is the advantage that the system can achieve. It was seen in the last section, however, that for a variable-level coherent ASK system this statement requires the qualification that the average SNR at the receiver be sufficiently high. This places a variable-level ASK system at a disadvantage in comparison with a variable-duration FSK system because in order to realise the full advantage of the former system, very high SNRs must be available at the receiver. Further, in a variable-level (selecting signals from m possible signal, where m is variable) system, the equipment and hence the complexity of the system grows as a power of two, that is as $2^{1/k}$, where k is given by equation 8.7. This, again, puts a variable-level system at a disadvantage as compared with a variable-duration system in which the equipment complexity hardly increases with increasing ratio between the two data rates.

With the discussion in the foregoing two paragraphs in mind, Figs. 8.3 and 8.13 can be used to compare the performance of a variable-level ASK system to that of a variable-duration FSK system.

A (2,4) variable-level coherent ASK system gives a maximum (achievable at error rates below 10^{-2}) power advantage of 4.38 dB over a fixed-rate binary coherent ASK system (see Fig. 8.13). The ratio of the lower to higher data rate is only 1:2 for a (2,4) variable-level system. Also as can be seen from Fig. 8.13, a two-data-rate variable-duration incoherent system requires a ratio of the

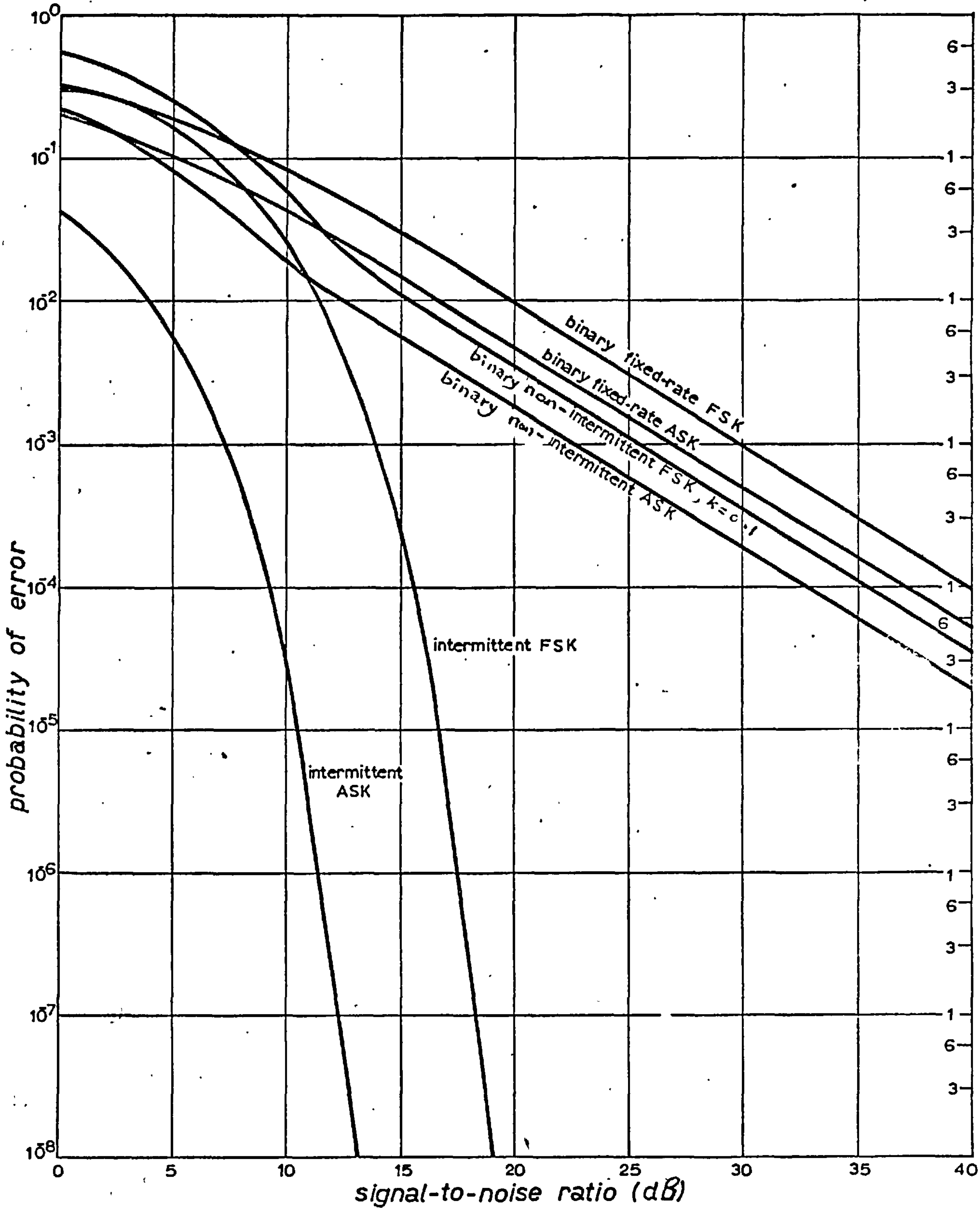


Fig. 8.13. Comparison between Variable-Duration FSK Systems and Variable-Level ASK Systems; Noiseless Feedback and Fixed Average Rate

lower to higher data rate of 1:10 to achieve a similar maximum power advantage. This is an important result in favour of the variable-level system.

It is clear, as can be seen from Figs. 8.10 and 8.13 that this considerable advantage of the variable-level ASK system cannot be extended beyond the (2,4) variable-level system without attaching a condition that good SNRs be available at the receiver. For example, eventhough a (2,8) variable-level system gives as much as 7.13 dB power advantage over the binary ASK system, this advantage is realised in full only for receiver SNRs above 20 dB. In fact, since below receiver SNRs of 15.5 dB, the (2,8) variable-level ASK system has a performance inferior to that of the (2,4) system, the power advantage of the (2,8) variable-level ASK system over that of the fixed-rate binary ASK system diminishes and hence the relative advantage of the variable-level system over the variable-duration system also diminishes.

It is seen from Fig. 8.13 that the binary intermittent coherent ASK system has a 6.25 dB power advantage over the variable-duration intermittent incoherent FSK system for similar error rates. If 3 dB is deducted from this power advantage because coherent detection is used in the ASK system, the relative advantage (due to intermittent operation) of the variable-level system over the variable-duration system is 3.25 dB saving in power for similar error rates.

Comparison between the variable-duration and the variable-level systems has been made in this section under the assumptions of noiseless feedback and perfect prediction. It is seen in section 8.3.1 that for a two-data-rate variable-duration FSK system, the effect of noisy feedback and imperfect prediction was to increase by small amount the required transmitter power for the same error rates and that this results in a slight shift of the performance curves to the right. If it is assumed that noisy feedback and imperfect prediction have similar effect on variable-level ASK systems, then the relative advantage of the variable-level ASK system over the variable duration FSK system will remain approximately the same as for the noiseless feed-back and perfect prediction case.

8.3.4 Comparison Between the Variable-Duration Incoherent FSK System and and the Van Duuren ARQ System

In this section two-data-rate variable-duration FSK systems with k = 1/10, k = 1/100 and k = 0, are compared with van Duuren ARQ system. Since in the van Duuren ARQ system the average (through put) rate of transmission decreases as the number of retransmissions increase, it is evident that for low SNRs, the average rate of transmission will in general be extremely low. The average rate of transmission for the two-data-rate systems is constant at the normalised value of unity. The average rate of transmission of the two types of systems will be taken into consideration when making the comparisons.

It can be seen from Figs. 8.14 and 8.15 that a two-data-rate variable-duration incoherent FSK system with k higher than 1/50 will perform in a manner inferior to that of the van Duuren ARQ system, that is, the former system has a higher error rate than the latter system. It can be seen from Figs. 8.16 and 8.17 that for values of k lower than 1/50, there occur cross-overs in the performance of two systems. For example, the performance curve of

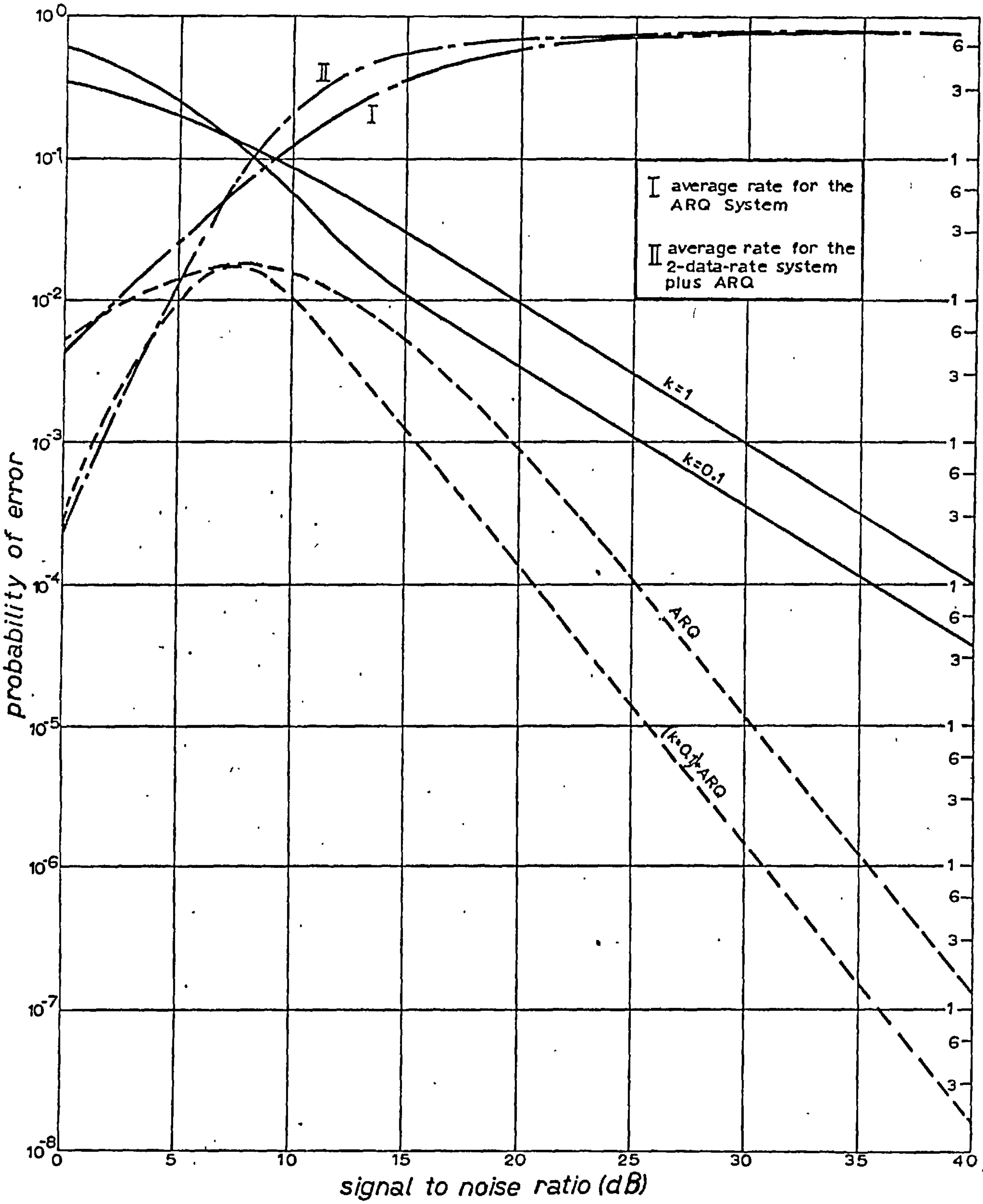


Fig.8.14. Comparisons between an ARQ System, a 2-Data-Rate Binary FSK System with k=0.1 and Fixed Average Rate and the ARQ Incorporating the 2 Data Rate System; Noiseless Feedback Case.

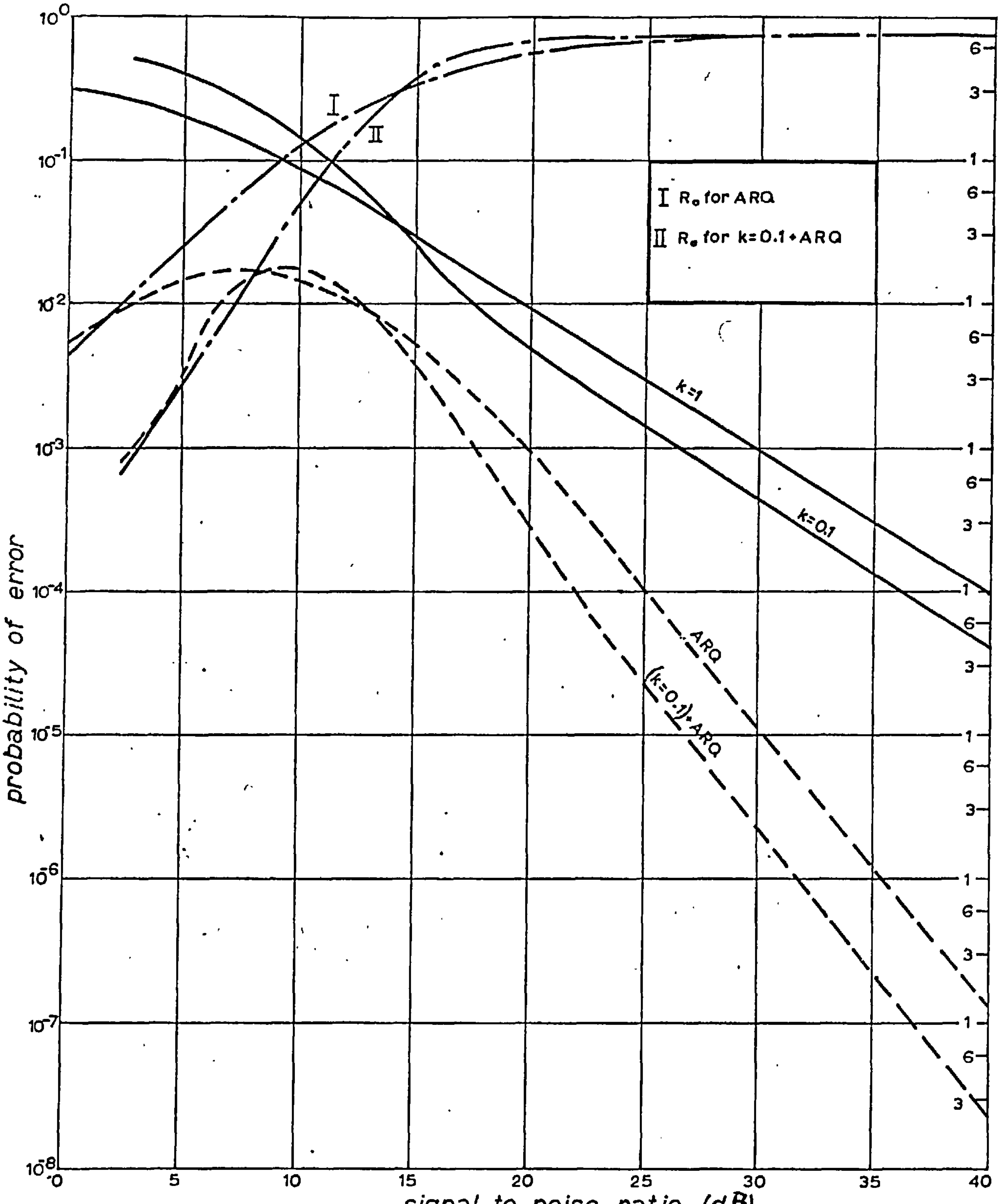


Fig. 8.15. Comparisons between an ARQ System, a 2-Data-Rate Binary FSK System with k=0.1 and Fixed Average Rate ARQ System Incorporating the 2-Data-Rate System; Noisy Feedback Case

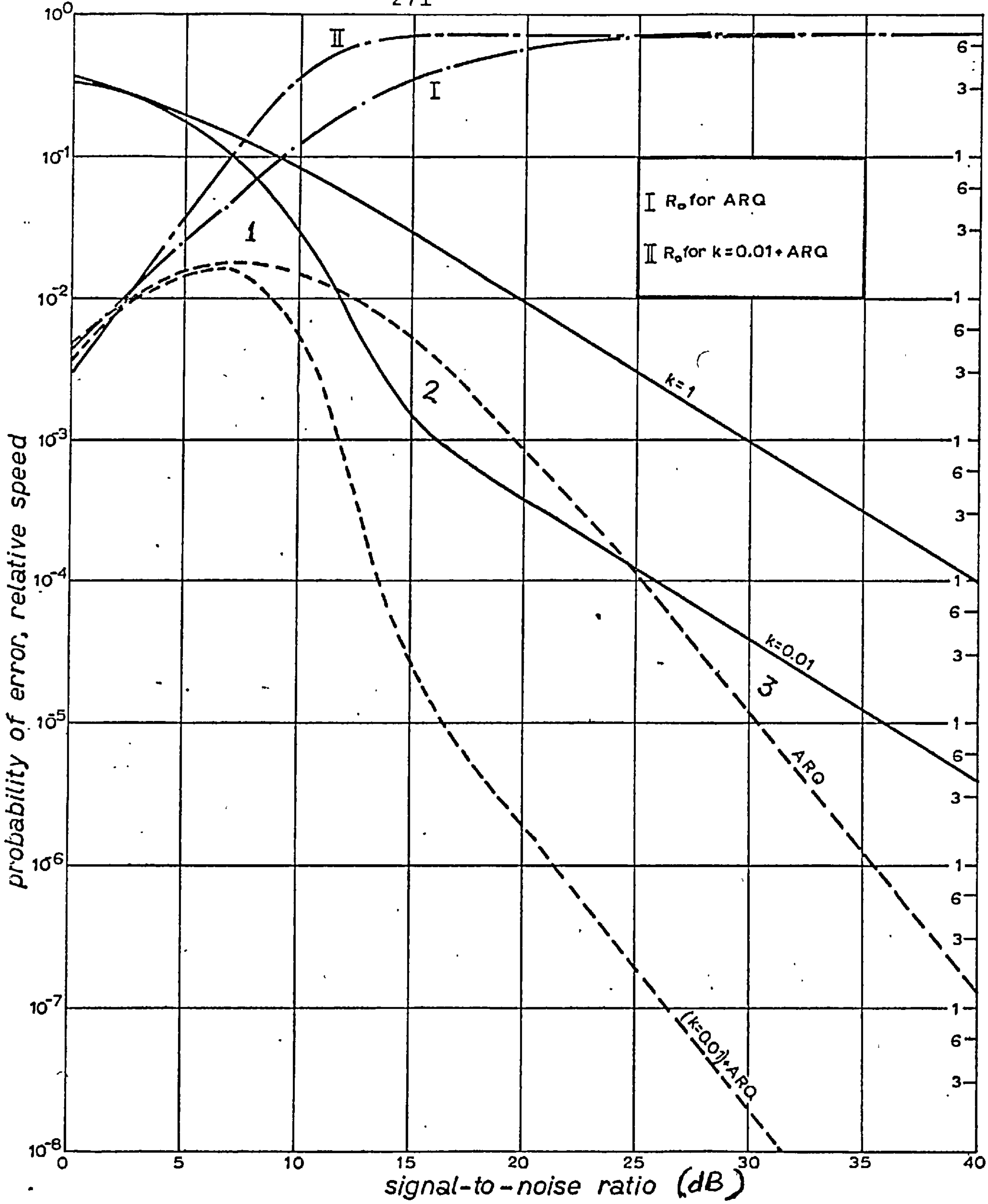


Fig. 8.16. Comparisons between an ARQ System, a 2-Data-Rate System with k=0.01 Fixed Average Rate and the ARQ System incorporating the 2-Data-Rate System; Noisy Feedback Case

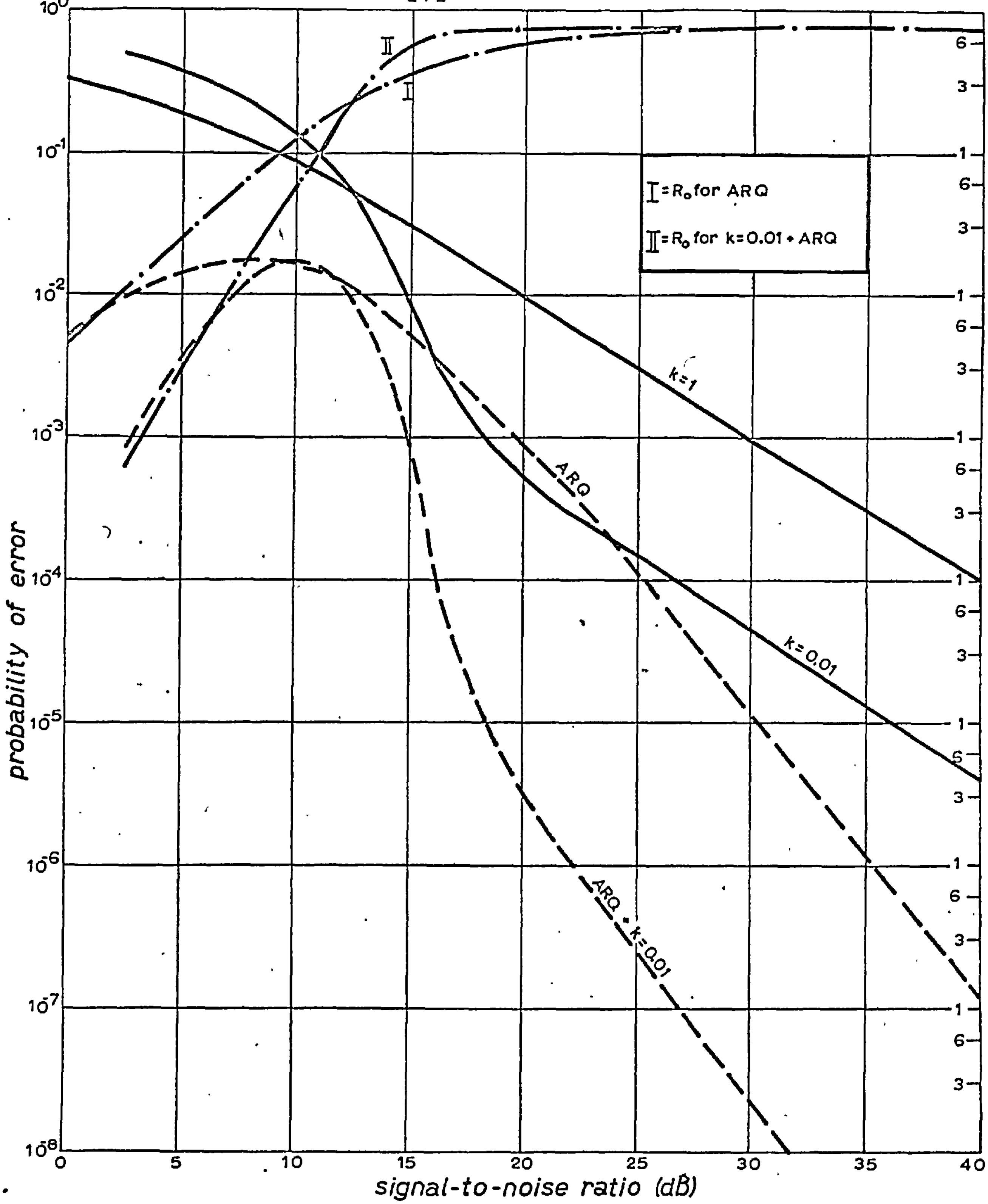


Fig. 8.17 Comparisons between an ARQ System, a 2-Data-Rate Binary FSK System with k=0.01 and fixed Average Rate and the ARQ System Incorporating the 2-Data-Rate System; Noisy Feedback Case

a two-data-rate variable-duration system with k = 1/100 has two cross-overs with that of the ARQ system. Below receiver SNR of 11.5 dB, the ARQ has better error rates than that of the two-data-rate variable-duration system. For SNRs lying between 11.5 dB and 25 dB, the variable-duration system has lower error rates as well as higher transmission rates than the van Duuren ARQ system. Above SNRs of 25 dB, the ARQ system has lower SNRs and comparable transmission rate to that of the two-data-rate variable duration FSK system.

It is necessary to elaborate on the comparison made between the ARQ system and the variable-duration incoherent FSK system. In region 1, shown in Fig. 8.16, though the ARQ has a better error rate than the variable-duration FSK system, it cannot be concluded from this that the ARQ is the better system. This is because in this region, the ARQ system has a transmission rate less than $(\frac{1}{5})$th of the average rate of the variable-duration FSK system. Thus, in the region to the immediate left of the first cross-over point, it is not easy to say which system is the better system. It will depend on the requirement at hand. The choice is between high error rates combined with high transmission rate and lower error rates combined with very low

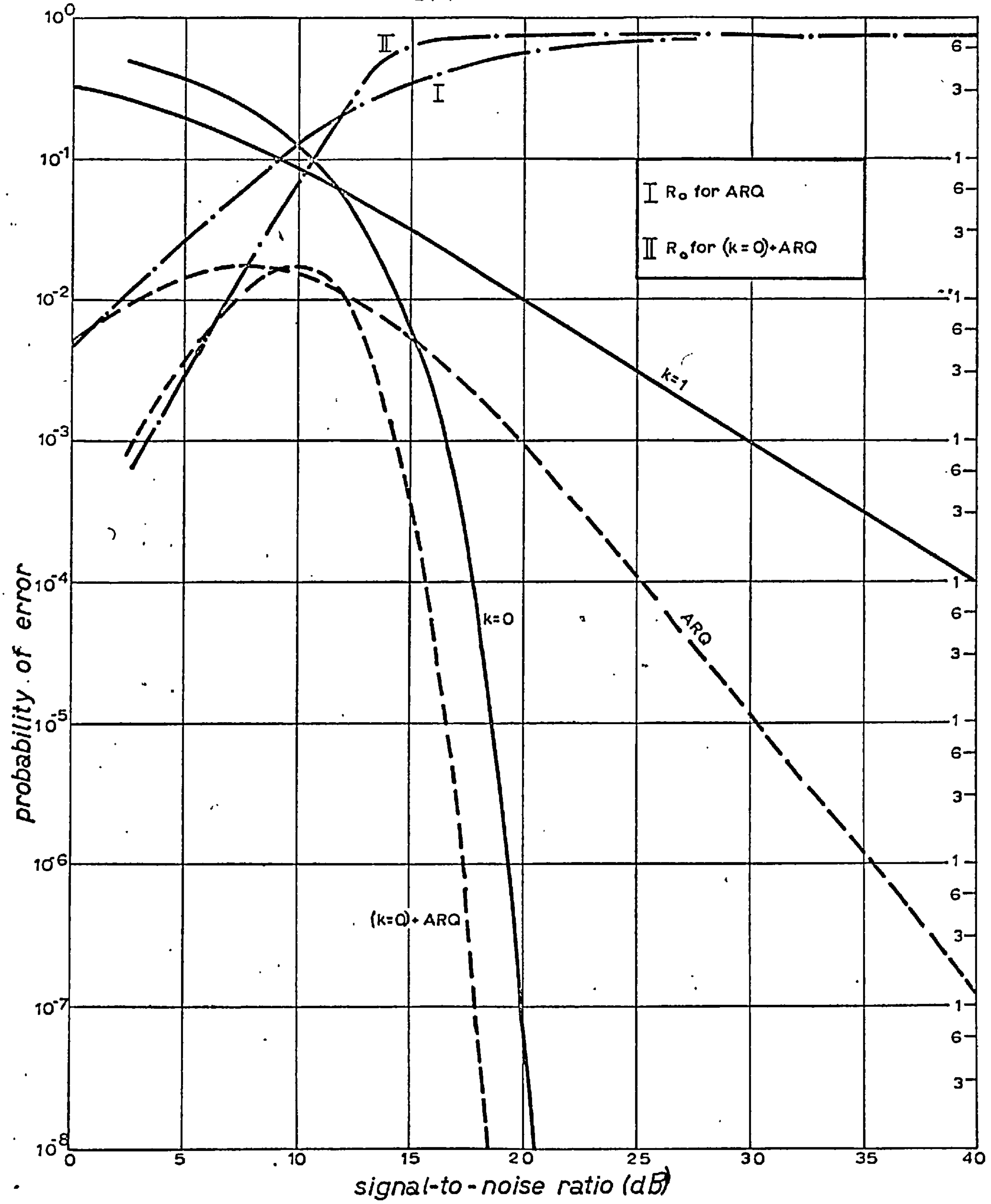


Fig. 8.18. Comparisons between an ARQ System, a Binary Intermittent FSK System and the ARQ Incorporating the Intermittent System; Noisy Feedback Case

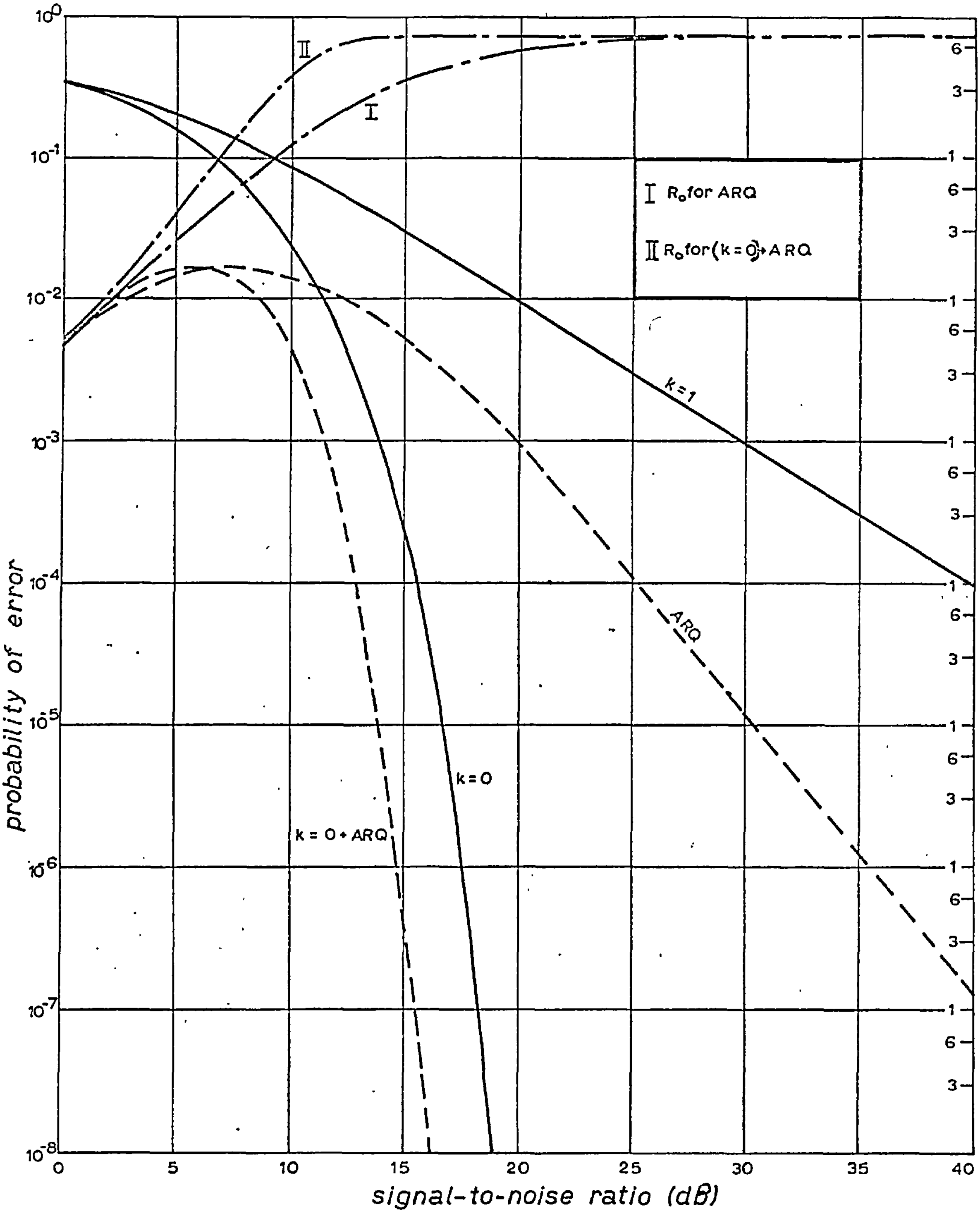


Fig. 8.19. Comparisons between an ARQ System, a Binary Intermittent FSK System and the ARQ System Incorporating the Intermittent System; Noiseless Feedback Case

transmission rates.

In region 2, see Fig. 8.16, it is clear that the two-data-rate variable-duration system performs better than the ARQ system. Not only has it lower error rates than the ARQ system in this region, but it also transmits at a higher data rate.

In region 3, the performance of the ARQ system is again better than that of the two-data-rate variable-duration system. Because of the high SNRs in this region,however, there are very few retransmissions in the ARQ system and the through-put rate of the ARQ system is very close to the average rate of the two-data-rate variable-duration incoherent FSK system. Clearly, the ARQ system is the better system in this region because it transmits at virtually the same rate as the two-data-rate variable-duration system but it has lower error rates.

From Figs. 8.18 and 8.19, it is seen that for error rates below 10^{-2}, the two-data-rate variable-duration intermittent FSK system has lower error rate than the ARQ system for the same SNR, or, alternatively the former system has a power advantage over the latter. The saving in power increases with decreasing error rates. For example, at error rate of 10^{-4}, the two-data-rate variable-duration intermittent

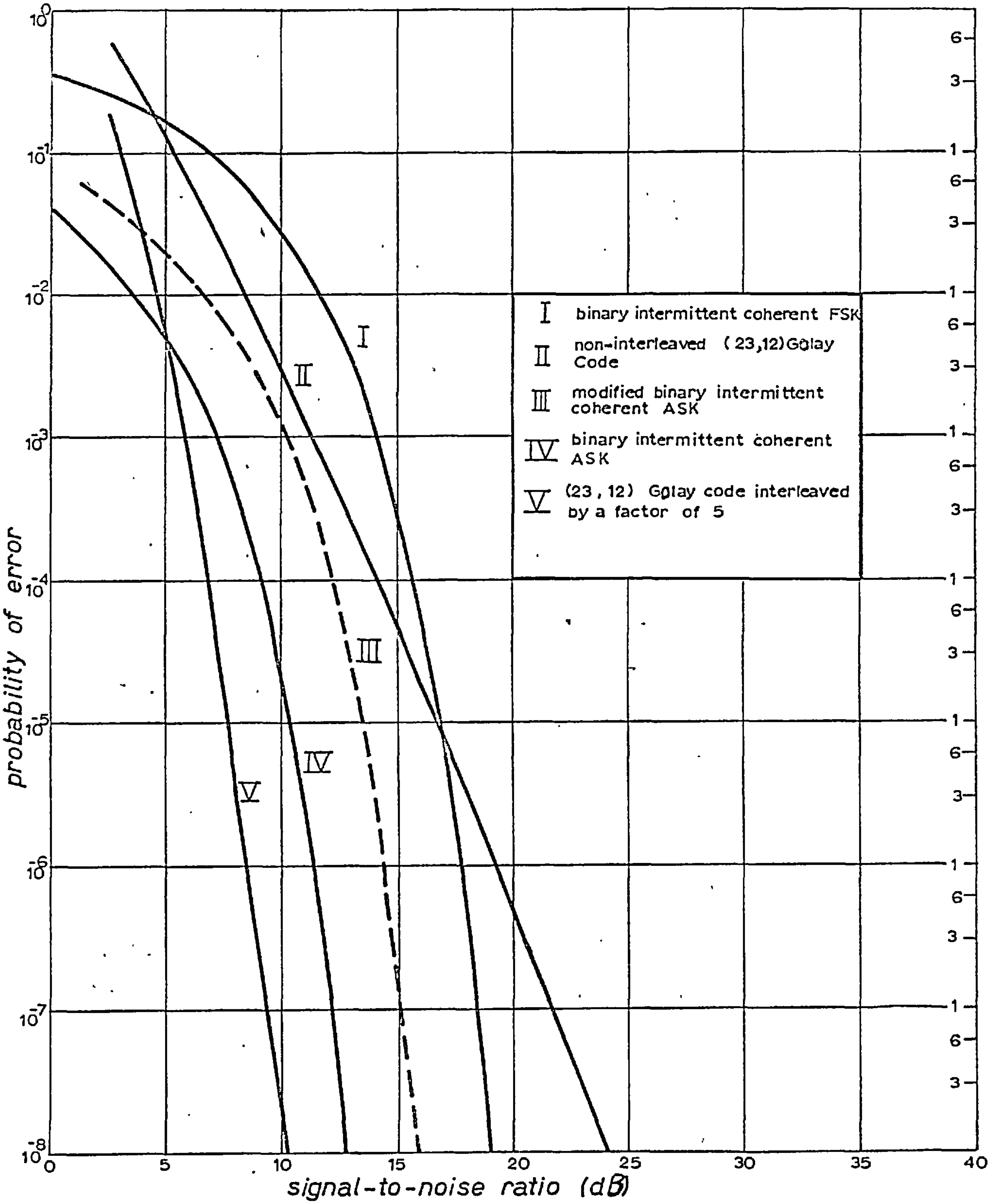


Fig. 8.20. Comparisons between the 2 Data-Rate-Binary Incoherent FSK and Coherent ASK Systems and the (23,12) Golay Code

system has a power advantage of 9.5 dB over the ARQ system, but at error rate of 10^{-5} the power advantage increases to 13.5 dB.

In Figs. 8.14 - 8.19, the dotted performance curves refer to systems which use both the ARQ technique and two-data-rate variable-duration technique. These curves will be discussed in Chapter 10.

8.3.5 Comparison Between Two-Data-Rate Systems and the (23,12) Golay Code

From Fig. 8.20 it is seen that the (23,12) Golay code interleaved by a factor of 5 performs better than the optimum two-data-rate variable-duration FSK system, showing about 9 dB at $P_e = 10^{-3}$. At very high error rates (above 3×10^{-1}), the two-data-rate variable-duration system performs better than the interleaved Golay code. Though the performance curve of the intermittent binary coherent ASK system is seen from Fig. 8.20 to lie fairly close to that of the interleaved Golay code, it should be recalled that the performance of the latter system was evaluated for an incoherent FSK system. An adjustment of 3 dB should therefore be made for the coherent ASK. If this is done, it is seen from the dotted curve that the performance of the Golay code is better than that of the variable-amplitude system, showing a 6.3 dB power saving for error rates below 10^{-3}. The interleaved Golay code, however, performs worse than the variable-level (-amplitude) system above error rates of 3×10^{-2}, for average SNRs above 3.75 dB.

The performance curve of the non-interleaved Golay code, however, has double cross-over points with that of the two-data-rate variable-duration FSK system. The intermittent FSK system performs better than the non-interleaved Golay code for error rates above 2×10^{-1} and below 5×10^{-6} equivalent to average detection SNRs of 4dB and 17 dB respectively. Between these average detection SNRs the non-interleaved Golay code performs better than the intermittent variable-duration FSK system,, showing about 3dB power advantage.

It is seen from Fig. 8.20 that the non-interleaved Golay code performs in a manner inferior to the variable-amplitude (set) system. After the adjustment, that was referred to above, to the variable-amplitude ASK system, it is seen from Fig. 8.20 that this system performs better than a system using non-interleaved Golay Code by about 1.5 dB for error rates above 10^{-5}. For error rates below 10^{-5}, the advantage of the variable-amplitude systems over the non-interleaved Golay Code increases rapidly and at an error rate of 10^{-8}, the power saving is 7.5 dB.

8.3.6 The Performance of Two-Data-Rate Systems with an Upper Limit on the Higher Data Rate

It was seen in the previous chapter that the effect of placing an upper limit on the value of the higher rate was to place another limit on the ratio, α, of the threshold to average SNR, also called the cut-off ratio for intermittent systems. For the intermittent system if the upper rate is fixed at its maximum value, then the percentage of transmission time must be increased by proportionately increasing the cut-off ratio, α. This will, however, lead to inferior performance because the optimum cut-off ratio is not used. Thus it is the effect, on the probability of error, of decreasing the cut-off ratio from its optimum value that must be evaluated. In a similar manner, for the two-data-rate non-intermittent system, if the upper rate is fixed at its maximum value, then for a fixed k(= R_ℓ/R_h), the normalised threshold, α, of switching between the data rates must be lower to maintain the same average rate. The effect on the probability of error of working at a non-optimum, α, is shown in Figs. 8.21 - 8.23.

From these Figures, it is evident that placing a limit on the upper rate of a two-data-rate system does not have a severe effect on the probability of error. For the intermittent FSK system, for example, it is seen from Fig. 8.21 that using a nonoptimum cut-off ratio of 0.4, instead of the optimum cut-off ratio, leads to a deterioration of only 1 dB, for low error rates, in the performance of the system. For the non-intermittent systems with k =0.1 and k =0.01, the extra transmitter power requirement is also seen to be about 1 dB (figs. 8.22 and 8.23) if a nonoptimum normalised threshold, α, equal to 0.4 is used instead of the optimum normalised threshold. It is seen from Fig. 7.1 that for α =0.4, the higher data rate, R_h, of the

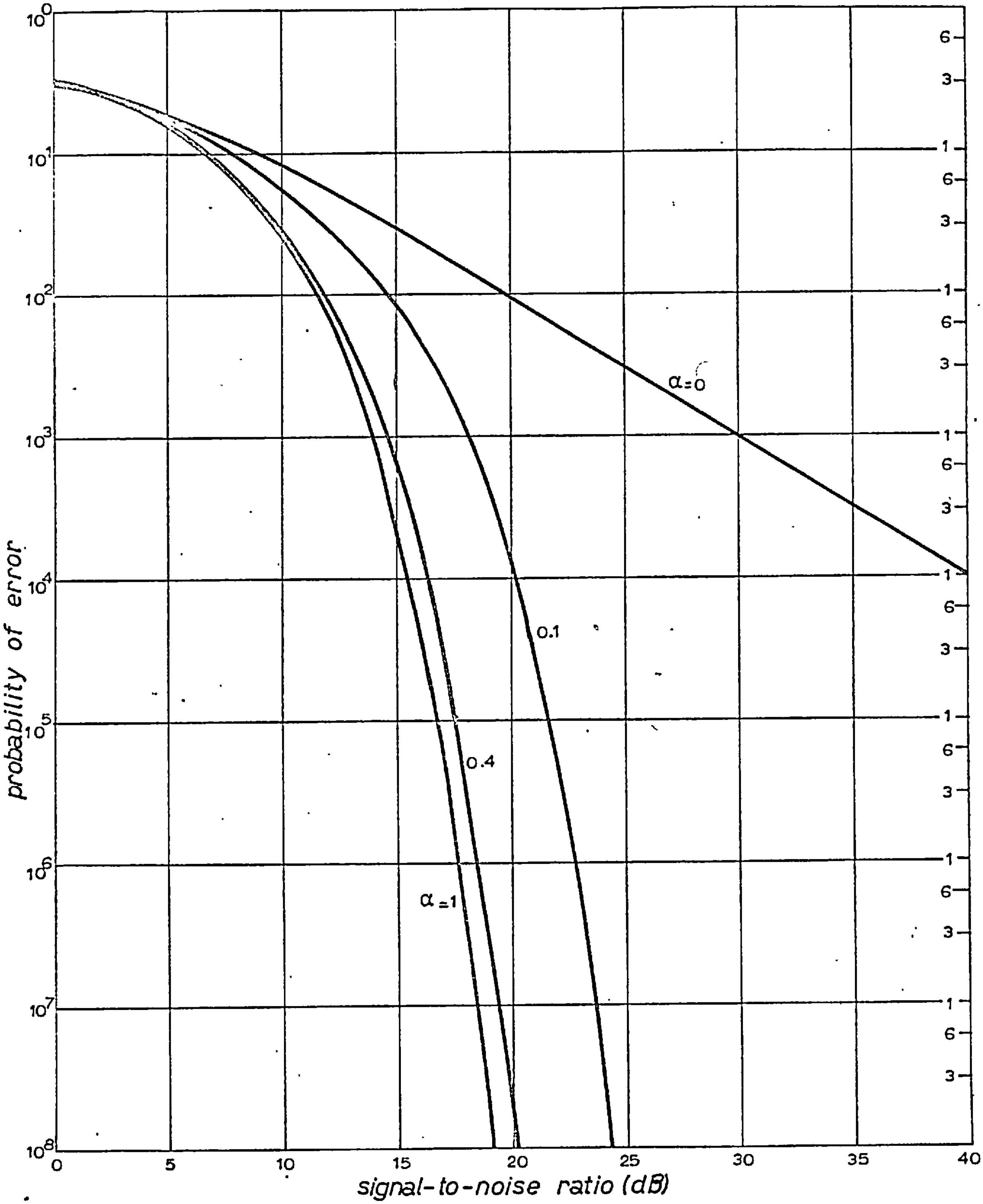


Fig. 8. 21. The Intermittent FSK System with Upper limit on the Higher Rate

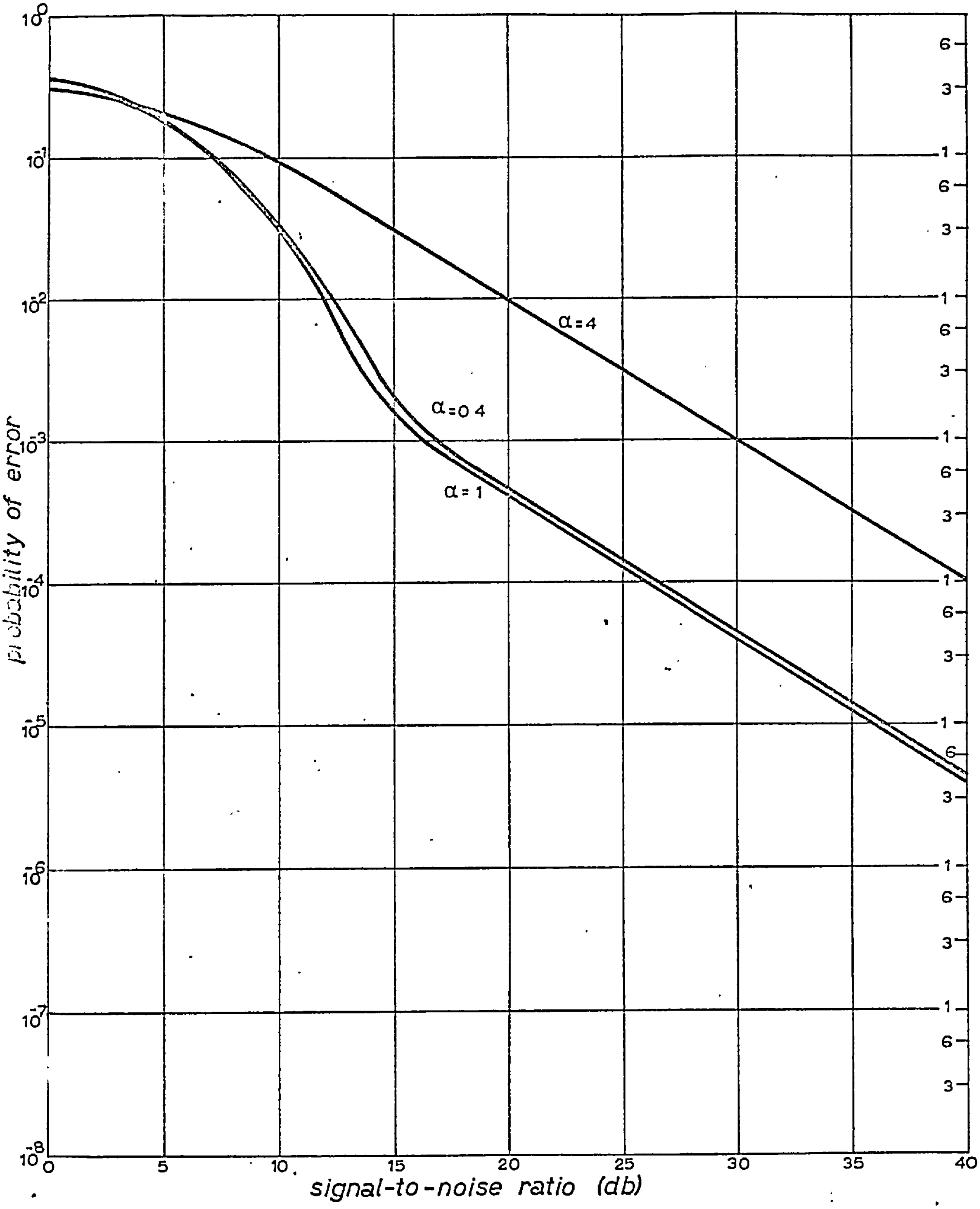


Fig. 8.22. The 2-Data-Rate Binary System with k=0.1; the Effect of Placing an Upper Limit on the Higher Rate

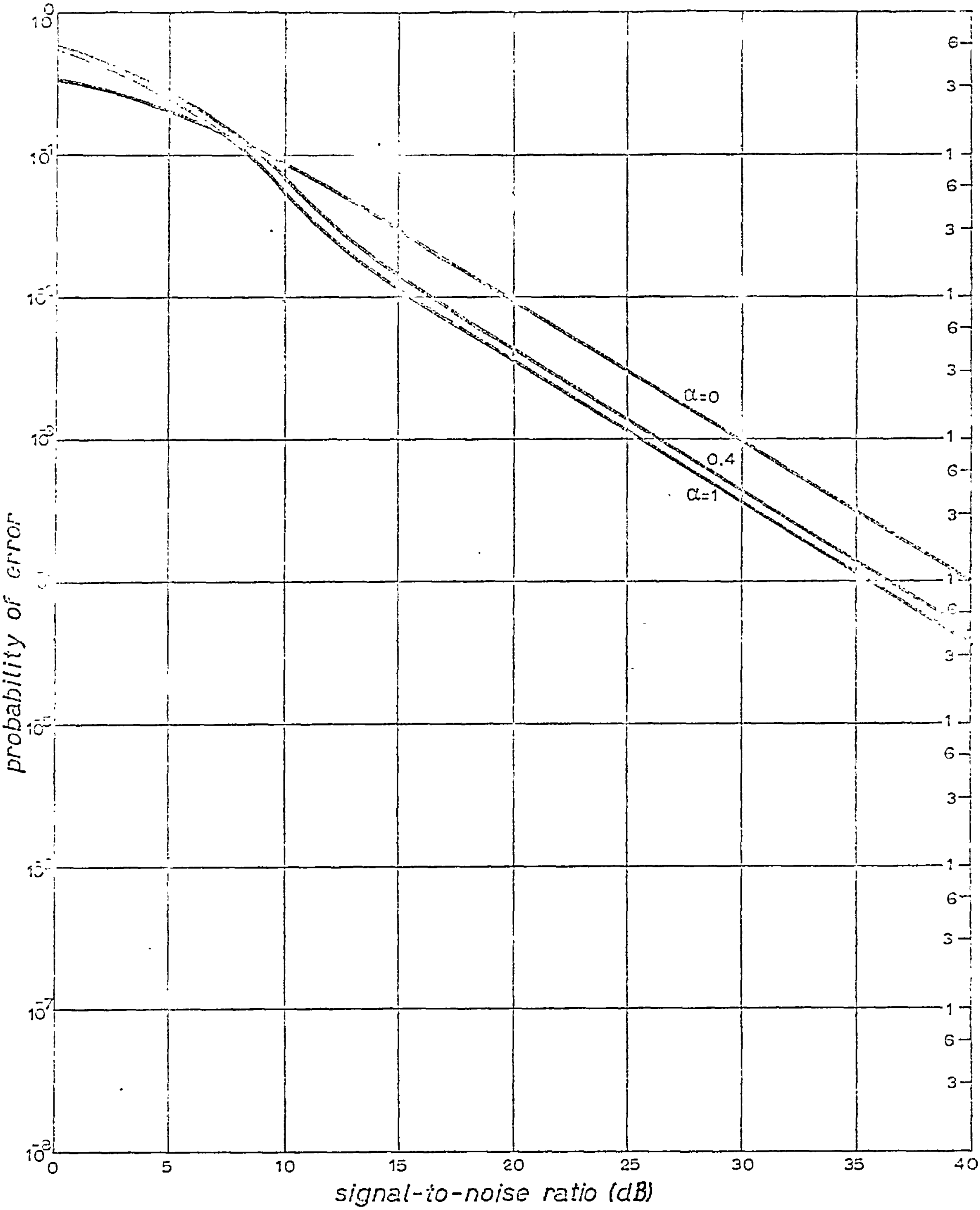


Fig. 8.23. The 2-Data-Rate Binary FSK System with k=0.01; the Effect of Placing an Upper Limit on the Higher Rate

two-data-rate system will be only 1.1 times the average rate, R_o. Thus, using a nonoptimum value of $\alpha = 0.4$ is equivalent to placing an upper limit on R_h to be at most 1.1 times R_o. It is thus concluded that an upper limit may be placed on the maximum value the upper rate can take without affecting the performance very much.

9. DISCUSSION OF THE PROBLEMS AND METHODS OF IMPLEMENTING n-DATA-RATE SYSTEMS

9.1 Introduction

The problems and some methods of impðementing n-data-rate systems were mentioned briefly in the previous four Chapters. In the present Chapter a more detailed discussion of these topics is given.

Before giving the discussion , however, two general comments on variable-rate systems need to be made. The first is that existing communication systems are composed mainly of fixed-rate equipments and as a vast sum of money is already invested in these systems, if variable-rate schemes are to find general applicability, they must be compatible with fixed-rate equipments. The second comment is that a variable-rate system should give sufficiently large saving in transmitter power. Equivalently, it should give sufficiently large improvement in the performance over the fixed-rate system transmitting continuously. From the results given in the previous Chapter, the latter condition is seen to hold for the schemes analysed in Chapter 7.

In Section 9.2, which follows, the problems of implementing n-data-rate systems are discussed. The methods, including an example, of impelementing n-data-rate systems are discussed in Section 9.3.

9.2 Problems of Implementing n-Data-Rate Systems

9.2.1. A General Problem

A general problem arises from the assumption made in some of the integrals in Chapters 5 to 7, that the SNR could take an infinite value. The Rayleigh fading probability density function used for the amplitude of the received signals is based on the assumption that the signal is the sum of infinitesimal signals of random phase. Consequently the probability, $pr(u > u_T)$, that the SNR will exceed u_T, however large, is always greater than zero. In practice this statement will not be true. Whatever the transmission medium, there will always be a finite SNR, u_m, for which $pr(u > u_m)$ is zero.

For systems which require large threshold SNR, u_T, the above limitation must be borne in mind.

9.2.2 Problem of Measuring Thresholds

An implicit assumption made in the analyses of n-data-rate systems was that it is possible to determine precisely whether the SNR is greater or less than a given threshold. Actually, this measurement can be made only with a given probability. Because of practical circuit limitations it is impossible to measure signal amplitude with as high a precision as desired within a specified period of time. As is shown below, errors due to this limitation are negligible.

Generally, the poorer the channel, the greater will be the effect of any error in the measurement of the threshold. As an example, take a channel with average SNR of 10dB and consider an intermittent incoherent FSK system working over such a channel. Further assume that the feedback channel is noiseless. The feedback channel is used to instruct the transmitter when to begin and when to cease transmission. From Fig. 7.2 the optimum threshold SNR for this system is seen to be 7.1 dB. The duty cycle is, therefore $e^{-0.71}$; and the fraction transmission should be halted is $1 - e^{-0.71}$.

The fraction of the total transmission time that is initiated by an error in the measurement of the threshold, that is, by noise rather than signal, can be expressed as

$$t_e = \frac{pr(u < u_T)}{pr(u > u_T)} \times pr(u > \text{measured } u_T) \qquad ....9.1$$

which is

$$t_e = \frac{1 - \text{Duty Cycle}}{\text{Duty Cycle}} \times pr(u > \text{measured } u_T) \qquad ....9.2$$

On substituting for the value of the duty cycle, equation 9.2 becomes

$$t_e = \frac{1 - e^{-0.71}}{e^{-0.71}} \; pr(u > \text{measured } u_T) \qquad ....9.3$$

To find the $pr(u > \text{measured } u_T)$, suppose that the maximum frequency of the fading variation does not exceed B/5, when B is the bandwidth of the system.

By simple filtering it is possible to increase the SNR at the input to the threshold detector by a factor of $\sqrt{5}$. Thus

$$\mathrm{pr}(u > \text{measured } u_T) = e^{-\sqrt{5} \times 7.1}$$

Hence,

$$t_e = \frac{1 - e^{-0.71}}{e^{-0.71}} \quad e^{-\sqrt{5} \times 7.1}$$

$$= 6.15 \times 10^{-8} \qquad \ldots.9.4$$

which represents negligible increase in error.

9.2.3. Synchronisation Problems

As was mentioned earlier, synchronization of the phase, the frequency and the duration of transmitted and received signals is always a difficult problem in a digital communication system. Solutions to this problem have, however, been found (see, for example Refs. 31 and 75) for systems using fixed-duration signals. When the duration of the transmitted signal is varied, as in the case of n-data-rate variable-duration system, the synchronization problem is clearly made more difficult. Two techniques of eliminating the increase in the problem are discussed below.

The first technique is that of sending, over a separate SYNC channel, a SYNC signal prior to transmitting the next group of symbols. The purpose of the SYNC signal is to inform the receiver which of the possible n data rates is currently in use. This, clearly, also tells the receiver what the duration of the incoming group of symbols is. The duration of the received signals is assumed to be the same until the next SYNC signal is received, that is, until the first in the next incoming group of symbols is due for detection.

Evidently, if the SYNC channel had enough power over it to make it virtually error free, the magnitude of the synchronization problem would be reduced to that of bit synchronization since the receiver would then know precisely the rate of transmission at all times. In practice, because of limitation in the available power, the requirement is that the probability, P_d, of making an error over the SYNC channel be made as small as possible while keeping enough power for use over the data channel.

In Chapter 7, p_d was evaluated for two-data-rate variable-duration systems. It was found that only 10% of the total available power was required over the SYNC channel to give error probabilities over this channel of less than 10^{-10}. Clearly, in many applications a probability of less than 10^{-10} of making an error over the SYNC channel would be acceptable and only bit synchronization problem would need to be considered.

The second technique of eliminating synchronization difficulties arising from variable-duration system operation is the use of special coding techniques. As was discussed in Section 5.2, the errors over the SYNC channel lead either to insertion or deletion of message symbols. Levenshtein[57] has found a class of binary codes for correcting spurious insertion and deletion of symbols. Clearly, using these codes the errors over the SYNC channel can be either eliminated or greatly reduced. In that case the probability of making an error over the SYNC channel is made sufficiently small and the synchronization problems arising from the errors over this channel are thus eliminated.

9.2.4 Storage Problems

In most transmission system, the storage capacity at the terminals is limited by economic and other factors. In these systems it is important to identify the problems that could arise with the storage of messages and suggest solutions to these problems. There are basically two problems.

The more serious problem that could arise is that of an overflow of the storage. This comes about when the storage is full and the rate R_o at which message is entering the storage is greater than the rate at which message is leaving the storage. See Fig. 9.1 below. The results is that some part of the message will be lost because an attempt to put it into the storage fails.

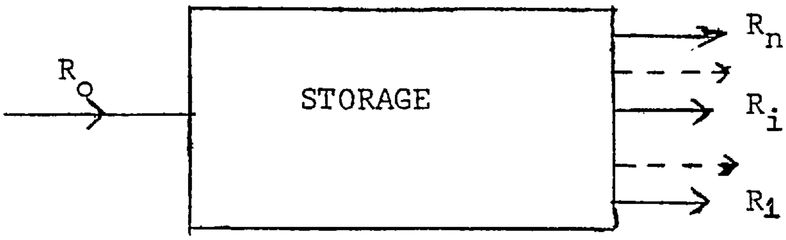


Fig. 9.1 The Data Rates at which Messages enter and leave the Storage

This problem could arise in an n-data-rate system when the average rate, R_o, is a large fraction of the highest data rate, R_n, and the channel is such that the data rates greater than R_o are used frequently.

A solution to the problem of overflow in an n-data-rate system is to adjust the threshold on the control channels so as to use the rates higher than the average rate for less percentage of transmission time. This will, in effect, lower the average rate of the system and information would then have to be taken into the storage at a reduced rate.

The other problem can be called an 'underflow' of the storage. It arises when the storage is empty and is being filled at a rate lower than the rate being used at the output. This problem could arise when the average rate, R_o, is a small fraction of the highest data rate, R_n, and the channel is such that the rates which are higher than the average rate are used more frequently than those which are lower than the average rate (see Fig. 9.1). The disadvantage of underflow is that the capability of the system is not exploited to the full and there are, therefore, periods during which the transfer of data is halted because of lack of messages in the storage.

A solution to the problem of underflow is to raise the average rate of the n-data-rate system. This can be done by adjusting the thresholds so as to use the rates higher than the average rate are used for a larger percentage of transmission time.

9.3 Methods of Implementing n-Data-Rate Systems

9.3.1 Introduction

As was mentioned in Chapter 5, there are basically two parts to the problem of designing and hence of implementing variable-rate systems. The first part of the problem is the method of estimating the conditions in the channel. The techniques which can be used for estimating the channel are discussed in Section 9.3.2, which follows. The second part of the problem is the actual technique of implementing the change of the data rates. These methods are discussed in Section 9.3.3.

Finally in Section 9.3.4, a system using discrete-variable-speed tape machines for implementing the changes of data rates is described in some detail so as to make the discussion more specific.

9.3.2 Methods of Estimating the Characteristics of the Forward Channel

For intermittent systems, the most convenient technique of measuring the forward channel is to send continuously over a separate channel a pilot tone signal and use the level of the received pilot-tone signals as a basis of forming an estimate of the channel conditions. The 'pilot-tone' technique of estimating the forward channel was described in Chapter 2 and used in the analysis of n-data-rate systems presented in Chapter 5. Two alternative techniques of estimating the forward channel are discussed in the following two paragraphs. Both techniques rely on the estimation of the probability of error.

The first technique uses the double threshold or null-zone detection described in Chapter 2. The symbols which fall in the null-zone detection region are treated as detected errors. The error symbols are then fed into a special counter which is a form of an error-rate estimator using direct counting technique.

Direct counting of error pulses suffers from the disadvantage that when the actual error rate is low, an unacceptably long period of time is required to count enough errors to give reliable estimate of the actual error rate. Chesler and Hingorani[44] have developed a performance monitoring unit (PMU) for binary systems using orthogonal signalling over a Rayleigh or Rician fading channel. It is based on estimating certain moments and joint moments of the output of the MARK and SPACE matched filters and calculating the error rate from these estimates. Gooding[14] described another general PMU technique which is stated to be much less sensitive to the statistics of the received signal and noise than the system developed by Chesler and Hingorani. Gooding employs modified decision thresholds in conjunction with a digital receiver to produce a pseudo-random error rate larger than the actual error rate of the receiver. This is then used to estimate the pseudo-random error rates corresponding to two or more modified decision thresholds. Lastly an extrapolation technique is used to determine the error rate of the receiver from the pseudo-random error rates.

9.3.3 Methods of Implementing Change of Data Rates

In a two-data-rate intermittent system, the actual technique of changing the rate is simply to halt transmission. One way of doing this physically is to use 'gates' which can be turned on or off at the transmitter and receiver. The transmitter gate is turned on for transmission when the received feedback signal indicates that the conditions in the channel are favourable. For the two-data-rate system more sophisticated techniques of switching between the rates are necessary. Two such techniques are given in the thesis. The first technique is described in this Section and the second will be described in the next Section in which an example of a complete n-data-rate system is given.

A technique which implements the change of the data rate of a transmission system by altering the duration of keying over a mark or a space symbol is referred to in this thesis as MARK-SPACE differential keying. An example of MARK-SPACE differential frequency-shift-keying is shown in Fig. 9.2. The system illustrated is a binary FSK feedback system.

For an n-data-rate variable-duration system, the higher the data rate, the shorter the time-duration of the symbols sent using the rate. Evidently, the highest data rate uses the shortest symbols. Let it be arranged by design that the duration of the symbols for any given rate of an n-data-rate system is an integral multiple of the duration of the shortest symbols, which are those sent using the highest data rate. In this case, in order to change the data rate, the keying over a mark or a space symbol is simply maintained over a period that is an integral multiple of the shortest time-duration used. Clearly, this technique is much simpler to implement than a continuously variable-duration system.

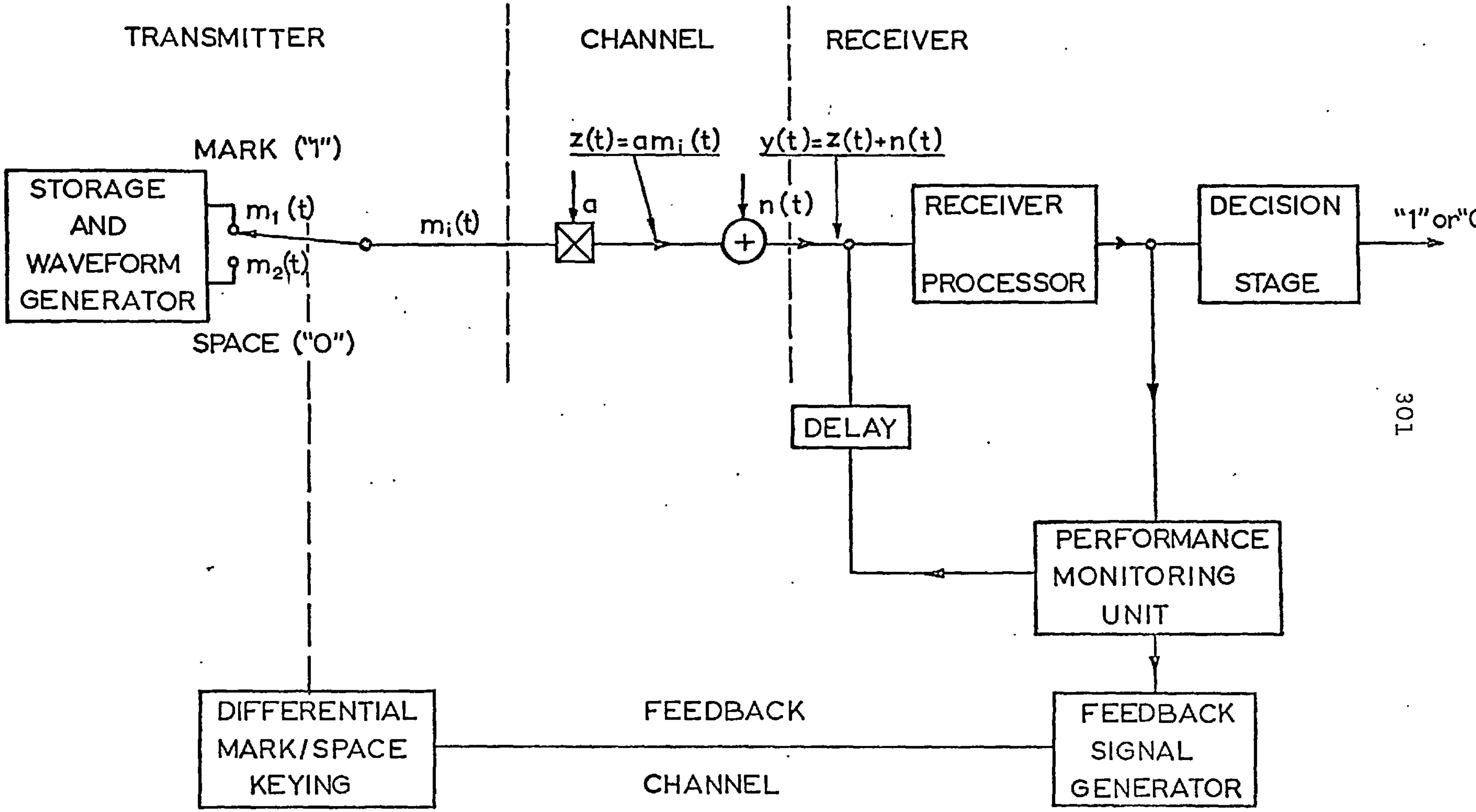


Fig. 9.2 . A Variable Rate Binary Feedback Communication System.

9.3.4 An FSK System Using Variable Speed Tape Machine

A second technique of varying the duration of the transmitted signals is that in which the duration of the signals is varied by recording and playing back the signals on a tape machine capable of working at n discrete speeds. The action of the tape machine at the transit terminal implements the change of the data rates. A description of this action is given below.

As shown in Fig. 9.3, signals of duration, D_o, are recorded at a fixed tape speed, V_o. After a slight delay, the signals are played back at one of n possible speeds of the tape V_i, $i = 1,2,..n$.

There are many possible relationships between the duration of the transmitted symbols and tape speed. As was discussed in Chapter 2, Pierce[65] has shown that the optimum rate variation is one in which the data rate is made directly proportional to the instantaneous SNR. In order to correspond to

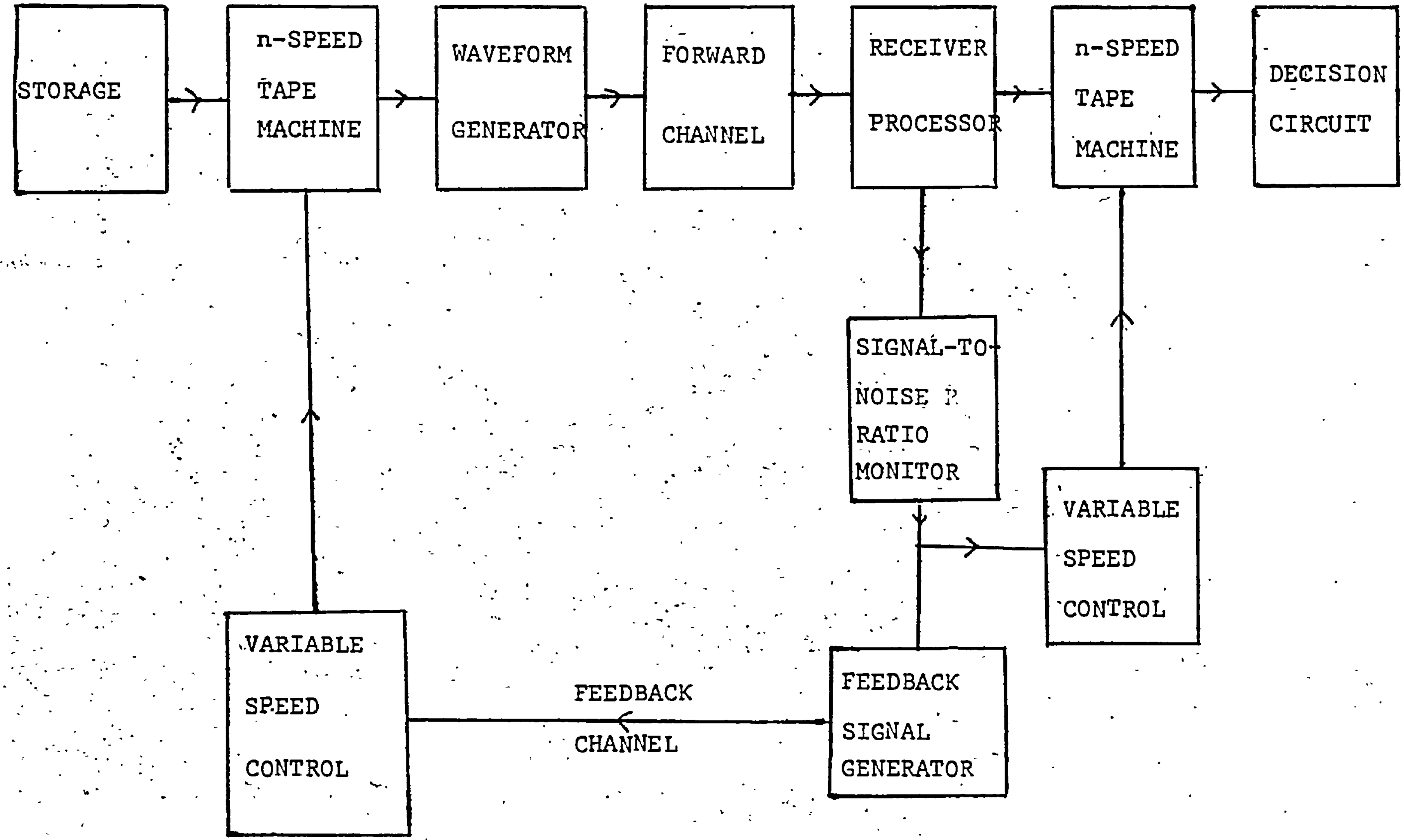


FIG. 9.3 A VARIABLE RATE FEEDBACK SYSTEM USING TAPE MACHINES

Pierce's optimum rate variation, the duration of the played back signals would need to be determined by the relation

$$D_i = \frac{V_o}{V_i} \cdot D_o \qquad \text{....9.6}$$

From the Equation 9.6 it is seen that the duration of the signal is increased if $V_i < V_o$ and is decreased if $V_i > V_o$.

At the receiver terminal, the received signals is frequency divided and translated down to near baseband in the conventional manner. It is recorded at the same speed at which it was played back for transmission at the transmit terminal. The recorded signal is then played back at a fixed speed equal to that on the recording side of the transmitter tape machine. Because of the action of the tape machines, the duration of the signal at the playback head of the receiver tape machine is equal to the original duration of the signal.

10. HYBRID SYSTEMS

10.1 Introduction

There are several techniques for counteracting the unwanted effects of signal fading in communication systems. Some of these methods were mentioned and discussed in Chapters 2, 5, 6, 7 and 8. Two new methods, the variable-duration and the variable-amplitude-set n-data-rate systems were discussed in detail in Chapters 5 - 8. In this Chapter the possibility of combining these new methods with existing methods will be investigated. A system that uses more than one method of counteracting the detrimental effects of fading can be called a 'hybrid system'.

The reason for using a hybrid system is that in many practical situations, it is not possible to control signal fluctuations adequately with only one technique. Examples which support this statement are:

1. The SNR of a received fading signal can take values which vary by very

large amounts. A consequence of this large variation is that it is necessary to vary some parameter by a correspondingly large amount. As was explained in the previous Chapter, physical limitations make this impossible in many cases.

2. When correlation between signals in the various branches is not as small as expected, an impractically large number of branches may be required to reduce signal fluctuations to an acceptable level.

3. The length of a code necessary to correct long or frequent bursts becomes unmanageable from an implementation point of view.

More acceptable signal control has been obtained by combining diversity with other ~~different~~ techniques.[66] Recently,[6,16] other types of hybrid techniques have been considered. The COMMET system,[6] for example, is a meteor-burst system incorporating ARQ and diversity reception. Because a meteor-burst communication system has an inherently low duty cycle, the ARQ technique operates in the COMET as an intermittent mode of transmission. The system thus combines an intermittent operation with ARQ and diversity. The results presented by Bartholome and Vogt[6] indicate that the combination provides a satisfactory answer to the requirements of maximum time utilization and minimum error rate. Further, there is the possibility of a trade-off between these two aspects of system performance.

In the following Section, the advantages of using a two-data-rate system incorporating the ARQ technique will be discussed in detail. An example of the use of hybrid system is given in Section 10.3.

10.2 A Two-Data-Rate System Incorporating an ARQ Technique

10.2.1 Introduction

In principle there is no reason why an n-data-rate system cannot incorporate a retransmission techniques. This would correspond physically to transmitting data at several discrete data rates (by using either discrete symbol durations or several signalling alphabets) with the rate at any instant depending on the measured signal strength of the pilot-tone signal; and requests for repeats are made after errors have been detected. The difficulty that would be encountered with such a system is not conceptual but that of implementation.

Despite the practical difficulties, a hybrid system combining an n-data-rate system with an ARQ technique can lead to an improvement in the performance as compared with that obtained when using either technique alone. The improvement is obtained in two stages. Firstly, the rate variation improves the element error rate. Generally, the better the element error rate, the better the performance of a retransmission system because of the decrease in the number of repetitions. Thus, secondly, further improvement is obtained from the retransmission technique which now performs better than it would have done for the same average detection SNR.

10.2.2 Discussion of the Results of a Two-Data-Rate System Incorporating the ARQ Technique

The performance curves of two-data-rate variable-duration systems incorporating ARQ were given in Figs. 8.14 and 8.19. The curves for the average rate for these hybrid systems were also given on these Figures. These curves are shown on the same page with those of the two-data-rate variable-duration systems and those of the ARQ system for easier comparison.

A general comment on the performance curves shown in Figs. 8.14 to 8.19 is that the overall performance of a hybrid system using the variable-duration technique is better than the ARQ system performance by a larger factor the improvement due to the variable duration technique alone. The reason for this has already been explained above.

From Figs. 8.14, 8.16 and 8.18, it is

seen that the relative speeds of hybrid systems are generally higher than those of ARQ system functioning alone except when k, the ratio of the lower to the higher data rate, is close to unity and the average SNR is low. The physical explanation of this behaviour is that when k is near unity, the lower rate which is used for transmission during poor channel conditions has a high absolute value. The consequence of this is that many retransmissions are necessary thus decreasing severely the average rate.

Clearly, the improvement is obtained at the expense of increased system complexity. It would seem in this case, however, that the complexity does not increase greatly and, in any event so long as the cost of extra equipment needed is small as compared with the total cost of the system, the saving in the transmitter power or the enhancement of the performance makes the hybrid system propositions attractive.

10.3 Examples of the Use of Hybrid Techniques

Since forward error control does not use a feedback link, a hybrid system obtained by a forward error correction scheme with a simple n-data-rate system is likely to be more practical than a system obtained by combining a retransmission technique with a simple n-data-rate technique. The coding and the decoding in a digital transmission system is independent of the actual transmission and the reception of signal waveforms. Thus, after coding a set of binary symbols, the duration of each bit (and hence the data rate) can be altered without affecting the coding.

Since coding techniques have not been studied in any depth in this thesis, a full discussion on the advantages of combining forward error correction schemes and two-data-rate systems is not possible. However, a simple example is given below to illustrate the advantage of combining such techniques.

From Fig. 8.6, it is seen that for the incoherent FSK system, a detection SNR of 19dB leads to an element error probability of 10^{-2}. Suppose that this error probability is not acceptable and, further, suppose that the minimum acceptable error rate is 5×10^{-5}. Several techniques, described below, can be used to achieve the desired minimum error rate. These are:

1) The transmitter could be increased until the element error rate of 5×10^{-3} is achieved. As seen from Fig. 8.6 this requires an increase of 3db in the SNR. This is equivalent to doubling the transmitter power. If a (15,7,2) BCH code interleaved by a factor of 89 is used over a channel with error probability of 5×10^{-3} then, as Brayer has shown,[17] an improvement of two orders of magnitude is obtained. Thus, having doubled the transmitter power, the (15,7,2) BCH code interleaved by a factor of 89 is used to obtain the desired overall performance.

2) A (255, 127, 32) BCH code interleaved by a factor of 5, which has an

improvement factor of 1000 (Ref. 17), can be used alone to obtain the desired performance.

3) From Fig. 8.6 a 2-data-rate system with k = .1 improves the element probability of error from 10^{-2} to 5×10^{-3}. From the result already quoted from Brayer's work it is clear that by (15, 7, 2) BCH code interleaved by a factor of 89 an improvement factor of 1:100 is obtained. Thus, combining a two-data-rate system with k = .1 and a (15, 7, 2) BCH code interleaved by a factor of 89 will also give the desired minimum performance.

The choice between these three methods will depend very much on prevailing circumstances. For example, if all the available power is already being used, the choice is immediately limited to the last twomethods. If a feedback link already exists in the system, then clearly use should be made of it to implement the two-data-rate scheme and the short code to achieve the error probability that is required. If there is no feedback link, then the cost of installing the feedback and the encoder and decoder for a (15, 7, 2) BCH code must be weighed against the cost of implementing the (255, 127, 32) BCH code. Clearly, for an entirely new system to be installed, the cost of each of the three possible systems above must be weighed against one another to find the cheapest. Here, cost is defined to include factors like flexibility.

Evidently, the possibility of using a simple n-data-rate system in conjunction with some of the established methods of counteracting fading increases the degree of freedom which the designer has.

11. OVERALL DISCUSSION OF RESULTS AND SUGGESTIONS FOR FURTHER WORK

Two new methods of varying the rate of transmission of a data transmission system were suggested in Chapter 1 and analysed in Chapters 5 to 7. Both techniques vary the data rate in a discrete manner. The first, the variable-duration technique, was applied to an incoherent FSK system and the second technique, the variable-level technique, was applied to a coherent ASK system. The performance of these systems was described in terms of the average probability of error and the average data rate. The results of the performance of these two new types of varying the data rate as methods of counteracting fading were discussed in Chapter 8. A summary of the major results are given below.

For two-data-rate variable-duration systems, it was seen that significant improvement in the performance was obtained only for lower to higher data rate ratios below 1:10, i.e. for $k \leq 0.1$. The optimum two-data-rate variable-duration system was found to be the intermittent system, or, when the lower data rate is made zero. Further, it was seen that the improvement obtained was subject to a maximum depending on the average detection SNR. The higher the average SNR, i.e. the better the channel, the greater the improvement obtained due to rate variation. The deterioration in the performance due to noise in the feedback path was not severe, but it increased slightly as the optimum system, namely intermittent variable-duration two-data-rate system, was approached.

In the case of variable-amplitude set (i.e. variable-level) system, it is to be noted that good detection SNR's are necessary if the improvement due to rate variation using this technique is to be realised. The only variable-level system in which the improvement is realised at reasonably low detection SNR is the (2,4) variable-level system, that is the system in which switching is made between a set of two and a set of four amplitude levels.

The performance of a (2,4) variable-level system, with a rate variation factor of only a half, was found to compare favourably with that of a variable-duration two-data-rate system having a rate variation factor of ten. For a variable-level system using a binary set of amplitude levels and another set of amplitude levels greater than four, the restriction that the improvement is obtainable only for high SNR's was found to be so severe that the variable-duration two-data-rate system is superior for a large range of detection SNR's between say 0-40 db.

In general, comparison between a variable-rate system studied in the thesis on a van Duuren system is not easy because the latter system has a varying through-put rate as a function of average detection SNR whereas the former system has a fixed average rate for all average SNR's. As was discussed in Chapter 8, however, cases exist in which the variable-rate system is definitely superior to the van Duuren ARQ system and vice versa. For example, the optimum two-data-rate (intermittent) system was found to be superior to the van Duuren ARQ system except for very low SNR's (below 5db) when the ARQ system has a better performance but has the disadvantage of having a very much lower average rate than the intermittent system.

The summary of the major results of the thesis have been given above. Suggestions for further work in this field will now be given.

As was explained in Chapter 5, insertions of symbols occur in intermittent systems when noise is interpreted as message symbols during the periods transmission is halted. By the use of coding or by choosing decision thresholds appropriately, it is possible to make the probability of insertion very small. If this is done, then the transition probabilities in the system can be represented by Fig. 11.1. The channel represented by this Figure clearly resembles ~~a binary~~ an erasure channel except for $p_{\phi\phi}$ which does not appear in the latter. This channel could form an interesting topic for theoretical study aimed at discovering further properties of the intermittent transmission technique. It would be interesting for example to evaluate the capacity of the channel and compare the result to the capacity of a binary erasure channel.

On the practical side, one technique of

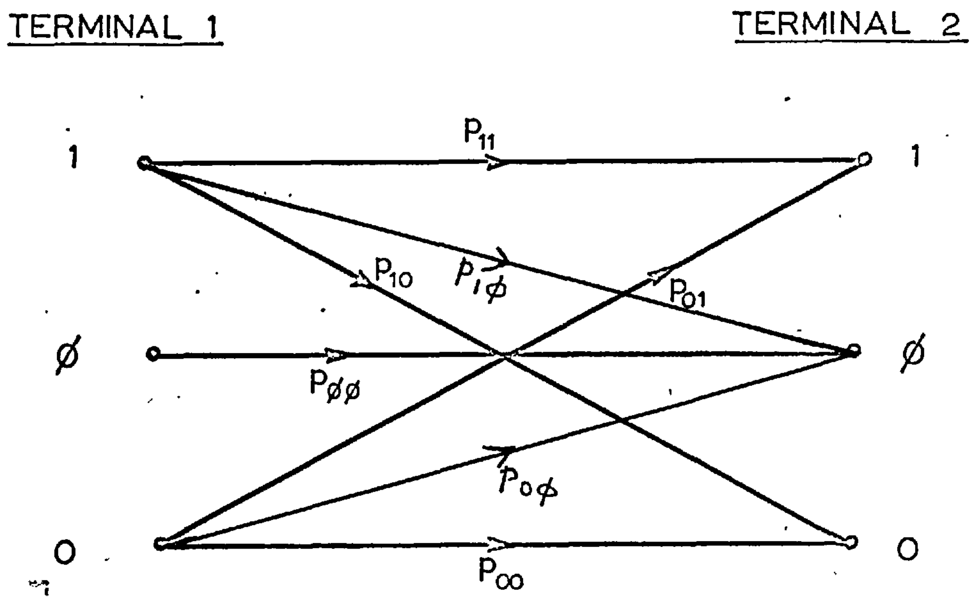


Fig. 11.1. The Transitional Probabilities of a Binary Intermittent Communications System without Insertions ~~and Deletions~~

implementing n-data-rate systems that was discussed in Chapter 9, namely, the use of tape machines, could be further developed. Chapter 9 simply discussed the feasibility of using tape machines capable of working at n different speeds in order to implement the n-data-rate system. The mechanical design and the testing of such tape machines in practical data transmission systems could form an interesting experimental study. Some problems do exist with the design of the tape machines. For example, one problem with the mechanical design is the reaction time when the tape speed is required to change by say 1:100 or more. Ideally this change should take place instantaneously. In practice, the problem is to minimize the reaction time.

The n-data-rate system was also studied in Chapters 6 and 7 with application to the variable-level (namely, selecting the signals for transmission from n different sets of amplitude levels, using one particular set at a time) ASK signalling. The block diagram in Fig. 6.1 was used as the general representation of the variable-level ASK system. The practical implementation, that is, the design and the buildings of equipment that would be capable of switching between say binary and quarternary signalling is yet another topic that could be investigated.

REFERENCES

1. Abramowitz, M. and I.A. Stegum, ed., "Handbook of Mathematical Functions," National Bureau of Standards, Washington, D.C., 1964

2. Allnatt, J.W., et al., "Frequency Diversity in Reception of Selectively Fading Binary Frequency-Modulated Signals," Proc. IEE, Vol. 104B, pp. 98 - 110, March 1968.

3. Arthurs, E. and H. Dym, "On the Optimum Detection of Digital Signals in Presence of White Gaussian Noise - A Geometric Interpretation and a Study of Three Basic Data Transmission Systems," IRE Transactions on Communications Systems, Vol. CS-10. pp. 336 - 372, Dec. 1962.

4. Axelby, G.S. and E.F. Osborne, "Automatic Control of Power in a Scatter Communications System," Proceedings of the National Symposium on Telemeting, Article 6.2, Sept. 1958.

5. Baghdady, E.T., ed., "Lectures on Communication System Theory," McGraw-Hill Book Co., 1961.

6. Bartholome, P.J. and I.M. Vogt, "COMET - A New Meteor - Burst System Incorporating ARQ and Diversity Reception," IEEE Transactions on Communication Technology, Vol. COM-16, pp. 268 - 278, April 1968.

7. Barrow, B.B., "Error Probabilities for Telegraph Signals Transmitted on a Fading FM Carrier," Proc. IRE, Vol. 48, pp. 1613 - 1629, Sept. 1960.

8. Bell, D.A., "Detection and Information Rate of Telegraphic Signals," IRE Transactions on Communications Systems, Vol. CS-9, pp. 70 - 76, March 1961.

9. Bello, P.A. and B.D. Nelin, "Influence of Fading Spectrum on Binary Error Probabilities," IRE Transactions on Communications Systems, Vol. CS-10, pp. 160 - 168, June 1962.

10. Bello, P.A., et al, "Digital Data Transmission via Troposcatter Channels," Five papers in IEEE Transactions on Communication Technology, Vol. COM-17, pp. 130 - 192, April 1969.

11. Benice, R.J. and A.H. Frey, "Analysis of Retransmission Systems," IEEE Transactions on Communication Technology, Vol. COM-12, pp. 135 - 145, Dec. 1964.

12. Benice R.J. and A.H. Frey, "Comparisons of Error Control Techniques," IEEE Transactions on Communication Technology, Vol. COM-12, pp. 146 - 154, Dec. 1964.

13. Bennett W.R. and J.R. Davey, "Data Transmission," McGraw-Hill Book Co., 1965.

14. Bloom, F.J., et al., "Improvement of Binary Transmission by Null-Zone Reception," Proc. IRE, Vol. 45, pp. 963 - 975, July 1957.

15. Brayer, K., "Error Patterns Measured On Transequatorial HF Communication Links," IEEE Transactions on Communication Technology, Vol. COM-16, pp. 215 - 221, April 1968.

16. Brayer, K., "Error Control Techniques Using Binary Symbol Burst Codes," IEEE Transactions on Communication Technology, Vol. COM-16, pp. 199 - 214, April 1968.

17. Brayer, K., "The Improvement of Digital HF Communication Through Coding: I - Interleaved Cyclic Coding; II - Tandem Interleaved Cyclic Coding," IEEE Transactions on Communication Technology, Vol. COM-16, pp. 771 - 786, Dec. 1968.

18. Chang, R.W., "Precoding for Multiple-Speed Data Transmission," B.S.T.J., Vol. 46, pp. 1631 - 1649, Sept. 1967.

19. Chang, S.S.L., "Theory of Information Feedback Systems," IRE Transactions on Information Theory, Vol. IT-2, pp. 29 - 40, Sept. 1956.

20. Cherry, E.C., "On Human Communication," Chapter 5, MIT PRess, Cambridge, Mass., 1957.

21. Costas, J.P., "Information Capacity of Fading Channels Under Conditions of Intense Interference," Proc. IRE, Vol. 51, pp. 451 -461, March 1963.

22. Cohn, D.L., et al., "Performance of Selected Block and Convolutional Codes on a Fading HF Channel," IEEE Transactions on Information Theory, Vol. IT-14, pp. 627 - 640, Sept. 1968.

23. Cowan, W.M., et al., "Theoretical Investigation of Dual Rate Transmissions Over Gaussian Multiplicative Circuits," AF 19(604) - 7488, Air Force Cambridge Research Laboratories, April 1962.

24. Cramer, H. and M.R. Leadbeater, "Stationary and Related Stochastic Processes," John Wiley & Sons INc., 1967.

25. Davenport, W.B. and W.L. Root, "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill Book Co., 1958.

26. Davies, K., "Ionosphenic Radio Propagation," National Bureau of Standards, Mono. 80, Washington D.C., 1966.

27. Fano, R.M., "Transmission of Information," MIT Press and John Wiley & Sons, Inc., 1961.

28. Fillipowsky, R.J. and E.T. Scherear, "Digital Data Transmission Systems of the Future," IRE Transactions on Communications Systems, Vol. CS-9, pp. 88 - 96, March 1961.

29. Folkestad, K. Ed., "Ionosphenic Radio Communications," Proc. of a NATO Institute on Ionosphenic Radio Communications in the Arctic, Plenum, New York, 1968.

30. Forney, G.D., "Exponential Error Bounds for Erasure, List and Decision Feedback Systems," IEEE Transactions on Information Theory, Vol. IT-15, pp. 52 1 61, Jan. 1969.

31. Gabor, D., "Theory of Communications," J. I.E.E. Vol. 93, pt. 3, pp. 429 - 457, 1946.

32. Gallager, R.G., "Information Theory and Reliable Communication," John Wiley & Sons, 1968.

33. Goldberg, B., "300 kHz -30 MHz MF/HF," IEEE Transactions on Communication Technology, Vol. COM-14, pp. 767 - 784, Dec. 1966.

34. Goldman, H.D., "Error Rate of a Cascaded Pair of Rayleigh Fading Communication Links," IEEE Transactions on Communications Systems, Vol. CS-11, pp. 354 - 355, Sept. 1963

35. Gooding, D.J., "Performance Monitor Techniques for Digital Receivers Based on Extrapolation of Error Rate," IEEE Transactions on Communication Technology, Vol. COM-16, pp. 380 - 387, June 1968.

36. Goodman, D.J., "Application of Statistical Decision Theory to the Detection of Radiotelegraph Signals," Ph.D Thesis, I.C.S.T., University of London, May 1967.

37. Gradshtein, I.S. and I.M. Ryhzhik, "Table of Integrals Series and Products," Academic Press, 1965.

38. Gumacos, C., "Analysis of Optimum Sync Search Procedure," IEEE Transactions on Communication Systems, Vol. CS-11, pp. 89 - 99, March 1963.

39. Hamsher, D.H., Ed., "Communication System Engineering Handbook," Chapter 15, McGraw-Hill Book Co., 1967.

40. Harris, B. and K.C. Morgan, "Binary Symmetric Decision Feedback," AIEE Transactions, pp. 436 - 443, Sept. 1958

41. Hayes, J.F., "Adaptive Feedback Communications," IEEE Transactions on Communication Technology, Vol. COM-16, pp. 29 - 34, Feb. 1968.

42. Hecht M. and M. Schwartz, "Suboptimal Decision Feedback Communications Systems," IEEE Transactions on Communication Technology, Vol. COM-16, pp. 822 - 824, Dec. 1968.

43. Helstrom, C.W., "The Comparison of Digital Communication Systems," IRE Transactions on Communications Systems, Vol. CS-8, pp. 141 - 150, Sept. 1960.

44. Hingorani, G.D. and D.A. Chesler, " A Performance Monitoring Scheme for FSK Transmission Over Fading Channels," IEEE Transactions on Communication Technology, Vol. COM-16, pp. 380 - 392, June 1968.

45. Institution of Electrical Engineers, "Convention on HF Communication," London, March 1963.

46. Jacobs, I., "Comparison of M-ary Modulation Systems," B.S.T.J., Vol. 46, pp. 843 - 864, May-June, 1967.

47. Janet and Related Systems; Thirteen papers in Proc. IRE, Vol. 45, pp. 1642 - 1733, Dec. 1957.

48. Johansen, D.E., "Binary Error Rates in Fading FDM-FM Communications," IRE Transactions on Communications Systems, Vol. CS-9, pp. 206 - 214, Sept. 1961.

49. Jones, J.J., "Multichannel FSK and DPSK Reception With Three-Component Multipath," IEEE Transactions on Communication Technology, Vol. COM-16, pp. 808 - 821, Dec. 1968.

50. Jones, W.B., "A Comparison of Frequency-Shift Antimultipath Signalling Technique for Digital Communication Systems," IRE Transactions on Communications System, Vol. CS-9, pp. 83 - 87, March 1961.

51. Kirsch, A.L., et al., "Field-Test Results of the AN/GSC-10 (KATHRYN) Digital Data Terminal," IEEE Transactions on Communication Technology, Vol. COM-17, pp. 118 - 128, April 1969.

52. Kohlenberg, A. and G.D. Forney, Jr., "Convolutional Coding for Channels with Memory," IEEE Transactions on Information Theory, Vol. IT-14, pp. 168 - 626, Sept. 1968.

53. Kramer, A.J., "Improving Communication Reliability by Use of Intermittent Feedback Channel," IEEE Transactions on Information Theory, Vol. IT-15, pp. 52 - 61, Jan. 1969.

54. Kuhn, T.G., "Retransmission for Error Control," IEEE Transactions on Communications Systems, Vol. CS-11, pp. 186 - 201, June 1963.

55. Lacks, G., "Optimization of Signal Waveforms," IEEE Transactions on Information Theory, Vol. IT-14, pp. 95 - 97, April 1968.

56. Law, H.B., "The Detectability of Radio-Telegraph Signals in Noise," PRoc. IEE, Vol. 104B, pp. 130 - 140, March 1957.

57. Levenshtein, V.I. "Binary Codes Capable of Correcting Spurious Insections and Deletions of Ones," Problems of Information Transmission, Vol. 1, The Faraday Press, Inc., Jan-March, 1965.

58. Marcum, J.I., "Table of Q-functions," Rand Corporation Report RM 339, Jan. 1950.

59. Marcum, J.I., "A Statistical Theory of Target Detection in Pulsed Radar," IRE Transactions on Information Theory, Vol. IT-5, pp
1960.

60. Martin, J.W., "Intermittent Binary Communications in a Rayleigh Fading Medium," IEEE Transactions on Communication Technology, Vol. COM-15, pp. 717-718, Oct. 1967.

61. Middleton, D., "An Introduction to Statistical Communication Theory," McGraw-Hill Book Co., 1960.

62. Nesenbergs, M., "Comparison of the 3-out-of-7 ARQ with Bose-Chaudhuri-Hocquenhem Coding Systems," IEEE Transactions on Communications Systems, Vol. CS-11, pp. 202-212, June 1963.

63. Nesenbergs, M., "Error Probability for Multipath Fading - The 'Slow and Flat' Idealization," IEEE Transactions on Communication Technology, Vol. COM-15, pp. 797-805, Dec. 1967.

64. Palmer, J.E., et al., "Variable Bandwidth FM Transmission System," RADC-TR-61-124, April 1961.

65. Pierce, J.N., "Theoretical Performance Bounds for Variable Data Rate Transmission on Fading Circuits," ARCRL - 959, Air Force Cambridge Research Laboratories, Nov. 1961.

66. Pfeiffer, R.R. "Coded, Diversified, Rayleigh - Faded, Binary Symmetric Threshold Channels," RLE, MIT Report No TR-384, Dec. 1960.

67. Price, R., "Optimum Detection of Random Signals in Noise, with Application to Scatter-Multipath Communications," IRE Transactions on Information Theory, Vol. IT-2, pp. 125-135, Dec. 1956.

68. Price, R. and P.E. Green, "A Communication Technique for Multipath Channels," Proc. IRE Vol. 46, pp. 555 - 569, March 1958.

69. Rice, S.O., "Distribution of the Duration of Fades in Radio Transmission," B.S.T.J., Vol. 37, pp. 581 - 636, May 1958.

70. Schneider, H.L., "Data Transmission with PSK Permutation Modulation," B.S.T.J., Vol. 47, pp. 1131 - 1138, July - August, 1968.

71. Schwartz, L.S., "Some Recent Developments in Digital Feedback Communication Systems," IRE Transactions on Communications Systems, Vol. CS-9, pp. 51 - 57, March 1961.

72. Schwartz, L.S., "Feedback for Error Control and Two-Way Communication," IEEE Transactions on Communications Systems, Vol. CS-11, pp. 45 - 56, March 1963.

73. Schwartz, L.S., et al., "Evaluation and Optimization of Digital Communication Systems," Pt.II - Third Scientific Report Performed for U.S. Air Force by New York University under Contract No. AF - 19(604) - 1964.

74. Schwartz, M. et al., "Communication Systems and Techniques," McGraw-Hill Book Co., 1966.

75. Selin, I. and F. Teteur, "Synchronization of Coherent Detectors," IEEE Transactions on Communications Systems, Vol. CS-11. pp. 100 - 109, March 1963.

76. Shannon, C.E., "The Zero-Error Capacity of a Random Channel," IRE Transactions on Information Theory, Vol. IT-2, pp. 8 - 19, Sept. 1956.

77. Shannon, C.E. and W. Weaver, "The Mathematical Theory of Communication," University of Illinois Press, 1949.

78. Slepian, D., "Permutation Modulation," Proc. IEEE, Vol. 53, pp. 228 - 236, March 1965.

79. Statas, H., "Tropospheric Scatter Propagation - A Summary of Recent Progress," RCA Review, Vol. 19, pp. 3 - 18, March 1958.

80. Stein, S. and J.J. Jones, "Modern Communications Principles," McGraw-Hill Book Co., 1967.

81. Strassman, A.J. and A.C. Chapman," A Long Range Digital Communication System," IRE Transactions on Communications Systems, Vol. CS-11, pp. 383 - 389, Dec. 1963.

82. Sussman, S.M., "A Matched Filter Communication System for Multipath Channels," IRE Transactions on Information Theory, Vol. IT-6, pp. 367 - 373, June 1960.

83. Turin, G.L., "Communication Through Noisy Random-Multipath Channel," MIT Lincoln Laboratory, IR-116, May 1956.

84. Turin, C.L., "Signal Design for Sequential Detection Systems with Feedback," IEEE Transactions on Information Theory, Vol. IT-11, pp. 401 - 408, July 1965.

85. van Duuren, H.C.A., "Error Probability and Transmission Speed on Circuits Using Error Detection and Automatic Repetition of Signals," IRE Transactions on Communications Systems, Vol. CS-9, pp. 38 - 50, March 1961.

86. Van Trees, H.L., "Detection, Estimation and Modulation Theory," Pt. 1, John Wiley & Sons, Inc., 1968.

87. Voelcker, H.B. Jr., "Simple Codes for Fading Channels," IRE Transactions on Communications Systems, Vol. CS-6, pp. 47 - 53, Dec. 1958.

88. von Baeyer, H.J., "Band Limitation and Error Rate in Digital UHF - FM Transmission," IEEE Transactions on Communications Systems, Vol. CS-11, pp. 110 - 117, March 1968.

89. Wald, A., "A Sequential Analysis," John Wiley, 1947.

90. Woodward, P.M. "Probability and Information Theory with Applications to Radar," Pergamon Press, 1952.

91. Wolfowitz, J., "Coding Theorems of Information Theory," Springer-Verlag, 1964.

92. Wozencraft, J.M. and I.M. Jacobs, "Principles of Communication Engineering," John Wiley & Sons, Inc., 1965.

93. Yeh, L.P., "Basic Analysis of Controlled Carrier Operation of Troposphenic Scatter Communications Systems," IRE National Convention Record, Pt. 8, 1958.

94. Zimmerman, M.S. and A.L. Kirsch, "The AN/GSC-10 (KATHRYN) Variable Rate Data Modern for HF Radio," IEEE Transactions on Communication Technology, Vol. COM-15, pp. 197 - 204, April 1967.

ACKNOWLEDGEMENTS

The author wishes to convey his thanks to Prof. Colin Cherry of Imperial College, London, for his help during the early stages of this work. He also wishes to thank his supervisor, Dr. L.F. Turner of Imperial College, London, for the valuable discussions and encouragement throughout the whole period of the research work. The author further appreciates the valuable suggestions of Dr. D.J. Goodman of Bell Telephone Laboratories. The inumerable arguments with his colleague, Mr. J.M. Garduno, proved to be of great value to the author and are, thus, fully acknowledged. Last, but not least, thanks are due to Miss J. Kooner for her painstaking work of typing the most difficult parts of the thesis.

APPENDIX A: The Probability Density Functions (pdf) of the Signal Amplitude and Power for Rayleigh Fading

For the time- and frequency-flat fading model assumed throughout the thesis (see Fig. 1.2), the amplitude of the transmitted signal is multiplied by a random variable 'a'. (For convenience the steady level amplitude before the multiplication is assumed to be unity.) Specifically the random variable a is assumed to follow the Rayleigh law and has a pdf given by:

$$p(a) = \frac{a}{\sigma_a^2} e^{\frac{-a^2}{2\sigma_a^2}} \qquad \text{....A.1}$$

Now, the energy-per symbol to noise spectral density ratio (which clearly also becomes a random variable) can, by observation of Table 3.1 be generally expressed as,

$$u = \frac{a^2 kE}{N_o} = a^2 U \qquad \text{....A.2}$$

where,

k is a constant which depends on the modulation-detection technique used.

E is the energy-per-symbol

N_o is the spectral density in watts/Hz of the additive noise assumed to be white Gaussian with zero mean.

The problem now is this. Given the pdf p(a) Equation A.1, and the relationship between a and another random variable u, it is required to find the pdf, f(u), of u.

In general, if the pdf, q(y), of a variable y is given, and it is required to calculate the pdf, g(x), for x = R(y), it is convenient to invert the functional relationship and write y = r(x). Then by direct substitution,

$$q(y)\ dy = q\left[r(x)\right]\ r'(x)\ dx \qquad \text{....A.3}$$

where $r'(x)=dr(x)/dx$. Hence if r(x) is single valued function of x,

$$g(x) = q\left[r(x)\right]\ r'(x) \qquad \text{....A.4}$$

If r(x) is a multivalued function of x, the expression on the right hand side is summed over all the values.

Applying Equation A.4 directly to Equations A.1 and A.2, the following is found for f(u)

$$f(u) = \frac{1}{2\sigma_a^2 U}\ e^{-\frac{u}{2\sigma_a^2 U}} \qquad \text{....A.5}$$

Let $U_o = 2\sigma_a^2 U$ in A.5,

$$f(u) = \frac{1}{U_o}\ e^{-\frac{u}{U_o}} \qquad \text{....A.6}$$

APPENDIX B: Evaluation of the Two-Dimensional Probabilities

The two-dimensional probabilities which are joint distribution functions of the envelopes of joint Gaussian processes given in Chapter 5 are now evaluated.

Consider, first, the evaluation of

(i) $P\left[|M_1|^2 - |M_2|^2 < 0,\ Y_{j-1} < |E|^2 \leq Y_j \,\middle|\, m=m_1\right]$

$$= P\left[G_1 < \gamma G_2,\ \hat{y}_{j-1} < G_3 \leq \hat{y}_j\right] \qquad \text{.....B.1}$$

where,

$$G_1 = \frac{|M_1|^2}{\overline{|M_2|^2}} \qquad \hat{y}_{j-1} = \frac{Y_j}{\overline{|E|^2}}$$

$$G_2 = \frac{|M_2|^2}{\overline{|M_2|^2}} \qquad \gamma = \frac{\overline{|M_2|^2}}{\overline{|M_1|^2}}$$

$$G_3 = \frac{|E|^2}{\overline{|E|^2}}$$

NOTE: $|M_2|$ is the envelope for noise alone for condition $m = m_1$ and $|M_1|$ is the envelope for signal plus noise. Becuase the two data signals are assumed to be orthogonal, the complex variates M_1 and M_2 are independent and the joint probability density function for G_1, G_2 and G_3 can be written as

$$p(G_1, G_2, G_3) = p(G_1, G_3)\, p(G_2) \qquad \text{.....B.2}$$

Now, the bivariate pdf associated with the power level of a Rayleigh fading signal is[25]

$$p(G_1,G_3) = \frac{1}{1-\rho} \exp\left[-\frac{G_1+G_3}{1-\rho}\right] I_o\left(\frac{2\sqrt{\rho G_1 G_3}}{1-\rho}\right) \qquad \text{.....B.3}$$

where,

$$\rho = \frac{\overline{|M_1 E|}^2}{\overline{|M_1|^2}\ \overline{|E|^2}}$$

I_o is the modified Bessel function

and,

$p(G_2) = \exp(-G_2)$, from Appendix A.

The integration over the density function required to give the probability (Equation A.1) is

$$p\left[G_1 < \gamma G_2,\ \hat{y}_{j-1} < G_3 \leq \hat{y}_j\right]$$

$$= \int_{\hat{y}_{j-1}}^{\hat{y}_j} dG_3 \int_0^\infty dG_1\ p(G_1,G_3) \int_{G_1/\gamma}^\infty dG_2\ p(G_2) \qquad \text{.....B.4}$$

On integrating over G , Equation B.4 becomes

$$p\left[G_1 < \gamma G_2,\ \hat{y}_{j-1} < G_3 \leq \hat{y}_j\right]$$

$$= \int_{\hat{y}_{j-1}}^{\hat{y}_j} dG_3 \int_0^\infty dG_1\ \frac{1}{1-\rho} \exp\left[-\frac{G_1(\gamma+1-\rho)}{\alpha(1-\rho)} - \frac{G_3}{1-\rho}\right]$$

$$I_o\left(\sqrt{\frac{4\rho G_1 G_3}{(1-\rho)^2}}\right)$$

For brevity, the following changes of variables are made.

$$y = \frac{2G_1(\gamma+1-\rho)}{\gamma(1-\rho)}$$

and

$$x = \frac{2\rho\gamma G_3}{(1-\rho)(\gamma+1-\rho)}$$

With these substitutions the inner integral of Equation B.5 becomes

$$I = \frac{\gamma}{\gamma+1-\rho} \int_0^\infty \frac{dy}{2} \exp -\left[\frac{1}{2}y - \frac{x(\gamma+1-\rho)}{2\rho\gamma}\right] I_o(\sqrt{xy})$$

This can be written as

$$I = \frac{\gamma}{\gamma+1-\rho} \exp\left[-\frac{x(\gamma+1-\rho)}{2\rho\gamma} + \frac{x}{2}\right] \int_0^\infty dy \exp\left[-\frac{1}{2}(x+y)\right] \cdot I_o(\sqrt{xy}) \quad \text{....B.6}$$

Since,

$$Q(\sqrt{x}, \sqrt{t}) = \int_t^\infty \frac{dy}{2} \exp\left[-\frac{1}{2}(x+y)\right] I_o(\sqrt{xy})$$

In terms of the Q-function, Equation B.6 is

$$I = \frac{\gamma}{\gamma+1-\rho} \exp\left[-\frac{x(\gamma+1-\rho)}{2\rho\gamma} + \frac{x}{2}\right] Q(x, o) \quad \text{....B.7}$$

Now,

$$Q(x, o) = 1.$$

On substituting back G_3 for x, Equation B.7 becomes,

$$I = \frac{\gamma}{\gamma+1-\rho} \exp\left[-\frac{\gamma+1}{\gamma+1-\rho} G_3\right] \quad .....B.8$$

Equation B.8 substituted into Equation 8.5 gives

$$P\left[G_1 < \gamma G_2, \hat{y}_{j-1} < G_3 \leq \hat{y}_j\right]$$

$$= \frac{\gamma}{\gamma+1-\rho} \int_{y_{j-1}}^{\hat{y}_j} dG_3 \exp\left[-\frac{\gamma+1}{\gamma+1-\rho} G_3\right] \quad .....B.9$$

And on integrating over G , Equation B.9 becomes

$$p\left[G_1 < \gamma G_2, \hat{y}_{j-1} < G_3 \leq \hat{y}_j\right]$$

$$= \frac{\gamma}{\gamma+1} \left\{\exp\left[-\hat{y}_{j-1} \frac{\gamma+1}{\gamma+1-\rho}\right] - \exp\left[-\hat{y}_j \frac{\gamma+1}{\gamma+1-\rho}\right]\right\} \quad .....B.10$$

Now

$$\gamma = \frac{\overline{|M_1|^2}}{\overline{|M_2|^2}}$$

Thus,

$$p\left[|M_1|^2 < |M_2|^2, \hat{y}_{j-1} < |E|^2 \leq \hat{y}_j\right]$$

$$= \frac{1}{1 + \frac{\overline{|M_1|^2}}{\overline{|M_2|^2}}} \left\{ \exp\left[-\hat{y}_{j-1} \frac{1 + \frac{\overline{|M_1|^2}}{\overline{|M_2|^2}}}{1 + \frac{\overline{|M_1|^2}}{\overline{|M_2|^2}}(1-\rho)} \right] - \exp\left[-\hat{y}_j \frac{1 + \frac{\overline{|M_1|^2}}{\overline{|M_2|^2}}}{1 + \frac{\overline{|M_1|^2}}{\overline{|M_2|^2}}(1-\rho)} \right] \right\} \qquad \text{....B.11}$$

The other two-dimensional probability is:

(ii) $p\left[K_{\alpha-1} < |S|^2 \leq K_\alpha, Y_{j-1} < |E|^2 \leq Y_j\right]$

Let

$$I_a = p\left[K_{\alpha-1} < |S|^2 \leq K_\alpha, Y_{j-1} < |E|^2 \leq Y_j\right]$$

$$= \int_{\hat{k}_{\alpha-1}}^{\hat{k}_\alpha} dG_4 \int_{\hat{y}_{j-1}}^{\hat{y}_j} dG_8 \; p(G_3, G_4) \qquad \text{....B.12}$$

where,

$$G_4 = \frac{|S|^2}{\overline{|S|^2}}$$

$$G_3 = \frac{|E|^2}{\overline{|E|^2}} \quad , \text{ as before}$$

$$\hat{k}_i = \frac{k_i}{\overline{|S|^2}}$$

and

$$\hat{y}_j = \frac{Y_j}{\overline{|S|^2}} \quad , \text{ as before.}$$

Now,

$$p(G_4, G_3) = \frac{1}{1-\mu} \exp\left[- \frac{G_4+G_3}{1-\mu} \right] I_o\left(\frac{2\sqrt{\mu G_4 \; G_3}}{1-\mu}\right)$$

where,

$$\mu = \frac{\overline{|SE^*|^2}}{\overline{|S|^2} \quad \overline{|E|^2}}$$

Thus,

$$I_a = \int_{\hat{k}_{\alpha-1}}^{\hat{k}_\alpha} dG_4 \int_{\hat{y}_{j-1}}^{\hat{y}_j} dG_3 \frac{1}{1-\mu} \exp\left[- \frac{G_4+G_3}{1-\mu} \right]$$

$$\cdot I_o\left(\frac{2\sqrt{\mu G_4 G_3}}{1-\mu} \right)$$

With the following changes of variables,

$$w = \frac{2\mu G_4}{1-\mu}$$

$$z = \frac{2G_3}{1-\mu}$$

I_a becomes

$$I_a = \frac{1-\mu}{2\mu} \int_{\frac{2\mu\hat{k}_{\alpha-1}}{1-\mu}}^{\frac{2\mu\hat{k}_\alpha}{1-\mu}} dw \int_{2\hat{y}_{j-1}}^{2\hat{y}_j} \frac{1}{2} dz \exp\left[-(\frac{w}{\mu} + z) \right]$$

$$\cdot I_o(\sqrt{wz})$$

$$= \frac{1-\mu}{2\mu} \int_{\frac{2\mu\hat{k}_{\alpha-1}}{1-\mu}}^{\frac{2\mu\hat{k}_\alpha}{1-\mu}} dw \exp\left[- \frac{w}{2\mu} + \frac{w}{z} \right] \int_{\frac{2\hat{y}_{j-1}}{1-\mu}}^{\frac{2\hat{y}_j}{1-\mu}} \frac{dz}{2}$$

$$\cdot \exp\left[- \frac{1}{2}(w+z) \right] I_o(\sqrt{wz})$$

And, with a definition of the Marcum Q-function, that is,

$$Q(\sqrt{w}, \sqrt{\alpha}) = \int_\alpha^\infty \frac{dz}{2} \exp\left[- \frac{1}{2}(w+z) \right] I_o(\sqrt{wz}),$$

the integral becomes

$$I_a = \frac{1-\mu}{2\mu} \int_{\frac{2\mu\hat{k}_{\alpha-1}}{1-\mu}}^{\frac{2\mu\hat{k}_i}{1-\mu}} dw \exp\left[-\frac{w}{2}\left(\frac{1-\mu}{\mu}\right)\right] \left\{ Q\left(\sqrt{w}, \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}}\right) - Q\left(\sqrt{w}, \sqrt{\frac{2\hat{y}_j}{1-\mu}}\right) \right\}$$

The following identity will now be used.

$$\int_\alpha^\beta = \int_0^\beta - \int_0^\alpha$$

Thus,

$$I_a = \frac{1-\mu}{2\mu} \int_0^{\frac{2\mu\hat{k}_\alpha}{1-\mu}} dw \exp\left[-\frac{w}{2}\left(\frac{1-\mu}{\mu}\right)\right] \left\{ Q\left(\sqrt{w}, \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}}\right) - Q\left(\sqrt{w}, \sqrt{\frac{2y_j}{1-\mu}}\right) \right\} - \int_0^{\frac{2\mu\hat{k}_{\alpha-1}}{1-\mu}} dw \exp\left[-\frac{w}{2}\left(\frac{1-\mu}{\mu}\right)\right] \cdot \left\{ Q\left(\sqrt{w}, \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}}\right) - Q\left(\sqrt{w}, \sqrt{\frac{2\hat{y}_j}{1-\mu}}\right) \right\}$$

$$= I_1 - I_2 - I_3 + I_4 \qquad \text{....B.13}$$

where,

$$I_1 = \frac{1-\mu}{2\mu} \int_0^{\frac{2\mu\hat{k}_\alpha}{1-\mu}} dw \exp\left[-\frac{w}{2}\left(\frac{1-\mu}{\mu}\right)\right] Q\left(\sqrt{w}, \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}}\right)$$

$$I_2 = \frac{1-\mu}{2\mu} \int_0^{2\mu\hat{k}_{\alpha-1}} dw \exp\left[-\frac{w}{2}\left(\frac{1-\mu}{\mu}\right)\right] Q\left(\sqrt{w}, \sqrt{\frac{2\hat{y}_j}{1-\mu}}\right)$$

$$I_3 = \frac{1-\mu}{2\mu} \int_0^{2\mu\hat{k}_{\alpha-1}} dw \exp\left[-\frac{w}{2}\left(\frac{1-\mu}{\mu}\right)\right] Q\left(\sqrt{w}, \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}}\right)$$

and

$$I_4 = \frac{1-\mu}{2\mu} \int_0^{2\mu\hat{k}_{\alpha-1}} dw \exp\left[-\frac{w}{2}\left(\frac{1-\mu}{\mu}\right)\right] Q\left(\sqrt{w}, \sqrt{\frac{2y_j}{1-\mu}}\right)$$

Obviously, the solution to only one of these integrals need be worked out. The remaining solutions can then be written by simply altering the subscripts. Thus, take the first integral. It can be written as:

$$I = B \int_0^A e^{-Bw} \; Q(\sqrt{w}, \sqrt{c}) \; dw \qquad \text{....B.14}$$

where,

$$A = \frac{2\mu\hat{k}_{\alpha}}{1-\mu}$$

$$B = \frac{1-\mu}{2\mu}$$

$$C = \frac{2\hat{y}_{j-1}}{1-\mu}$$

To solve I_1, the single-sided Laplace transform is used. Thus,

$$L_w\left[Q(\sqrt{w}, \sqrt{t})\right] = \int_0^{\infty} w\ e^{-pw} \int_t^{\infty} \frac{dz}{2} \exp\left[-\frac{1}{2}(w+z)\right] \cdot I_o(\sqrt{wz})$$

$$= \int_t^{\infty} \frac{dz}{2}\ e^{-z/2} \int_0^{\infty} dw\ e^{-\frac{w}{2}(2p+1)}\ I_o(\sqrt{wz}) \quad \ldots\ldots B.15$$

On making the change of variables,

$$u = w(2p+1)$$

$$v = z/(2p+1)$$

the integral of the Laplace transform can be written as

$$\frac{2e^{\frac{1}{2}v}}{2\ +1} \int_0^{\infty} \frac{du}{2}\ e^{-\frac{1}{2}(u+v)}\ I_o(\sqrt{uv})$$

$$= \frac{2\ e^{\frac{1}{2}v}}{2p+1}\ Q(\sqrt{v},\ o)$$

$$= \frac{2\ e^{\frac{1}{2}v}}{2p+1}$$

$$= \frac{2}{2p+1}\ e^{-\frac{z}{2(2p+1)}} \qquad \text{....B.16}$$

On substituting Equation B.16 into B.15,

$$L_w\left[Q(\sqrt{w},\ \sqrt{t})\right] = \int_t^{\infty} dz\ \frac{1}{2p+1}\ e^{-\frac{z}{2}\left(1-\frac{1}{2p+1}\right)}$$

....B.17

On evaluating the integral in Equation B.17, the Laplace transform is found to be

$$L\left[Q(\sqrt{w},\ \sqrt{t})\right] = \frac{1}{2p+1}\ e^{-tp/(2p+1)} \qquad \text{....B.18}$$

To find $Q(\sqrt{w}, \sqrt{t})$, the inverse of the Laplace transform of Equation B.18 is taken, i.e.,

$$Q(\sqrt{w},\ \sqrt{t}) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} dp\ \ e^{wp}\left[\frac{1}{p}\ e^{\frac{-tp}{2p+1}}\right]$$

where,

Re $C > 0$

Let

$\Omega = 2p+1$, then

$$Q(\sqrt{w}, \sqrt{t}) = -e^{-\frac{1}{2}(w+t)} \frac{1}{2\pi j} \int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega \; \frac{\exp\left[\frac{w\Omega}{2} + \frac{t}{2\Omega}\right]}{1-\Omega}$$

....B.20

From the substitution,

$\mathrm{Re}\,\Gamma > 1$

On using Equation B.20, Equation B.14 can be written as follows,

$$I_1 = -B \int_0^A dw \exp\left[-(B+\tfrac{1}{2})w - \frac{C}{2}\right] \frac{1}{2\pi j} \int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega \cdot \frac{\exp\left[\frac{w\Omega}{2} + \frac{C}{2\Omega}\right]}{1-\Omega}$$

$$= -B \exp\left[-\frac{C}{2}\right] \int_{\Gamma-j\infty}^{\Gamma+j\infty} \frac{d\Omega}{1-\Omega} \exp\left[\frac{C}{2\Omega}\right] \int_0^A dw \exp\left[\left(\frac{\Omega}{2} - B - \tfrac{1}{2}\right)w\right]$$

....B.21

On integrating over w, Equation B.21 becomes

$$I_1 = I_1' - I_1''$$

....B.22

where,

$$I_1' = B \exp\left[-\frac{C}{2}\right] \frac{1}{2\pi j} \int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega \; \frac{\exp\left[\frac{C}{2\Omega}\right]}{\frac{1}{2}(1-\Omega)(\Omega-2B-1)}$$

....B.23

$$I_1'' = B \exp\left[-\frac{C}{2}-A(B+\tfrac{1}{2})\right] \frac{1}{2\pi j} \int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega \; \frac{\exp\left[\frac{C}{2\Omega} + \frac{A\Omega}{2}\right]}{\frac{1}{2}(1-\Omega)(\Omega-2B-1)}$$

....B.24

Making use of partial fractions

$$\frac{1}{(1-\Omega)(\Omega-2B-1)} = \frac{1}{2B}\left[\frac{1}{(2B+1)-\Omega} - \frac{1}{1-\Omega}\right] \quad \text{....B.25}$$

$\therefore$

$$I_1' = \exp\left[-\frac{C}{2}\right] \quad \frac{1}{2\pi j} \int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega \; \frac{\exp\left[\frac{C}{2}\right]}{(2B+1)-\Omega} - \frac{1}{2\pi j} \int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega$$

$$\cdot \frac{\exp\left[\frac{C}{2\Omega}\right]}{1-\Omega}$$

....B.26

On substituting w=o into Equation B.20 the following is obtained.

$$Q(o, \sqrt{t}) = e^{-\frac{t}{2}} \; \frac{1}{2\pi j} \int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega \; \frac{\exp\left[\frac{t}{2\Omega}\right]}{1-\Omega}$$

which is, clearly, the same from as the second integral in Equation B.26. The first integral can be shown to be

$$\frac{\Omega}{2B+1} = D \qquad \text{....B.27}$$

$$\frac{1}{2\pi j}\int_{\Gamma-j\infty}^{\Gamma+j\infty} d\,\Omega \frac{\exp\left[\frac{C}{2\Omega}\right]}{(2B+1)-\Omega} = \frac{1}{2\pi j}\int_{\frac{\Gamma}{2B+1}-j\infty}^{\frac{\Gamma}{2B+1}-j\infty} dD \frac{\exp\left[\frac{C}{2(2B+1)D}\right]}{1-D}$$

$$= -\exp\left[\frac{C}{2(2B+1)}\right] Q\left(o, \sqrt{\frac{C}{2B+1}}\right)$$

Thus, Equation B.26 can be written as

$$I_1' = \exp\left[\frac{-C}{2}\right] \quad -\exp\left[\frac{C}{2(2B+1)}\right] Q\left(o, \sqrt{\frac{C}{2B+1}}\right)$$

$$+ \exp\left[\frac{C}{2}\right] Q(o,\sqrt{C}) \qquad \text{....B.28}$$

Now,

$$Q(\sqrt{x}, \sqrt{t}) = \int_t^\infty \frac{dy}{2} \exp\left[-\tfrac{1}{2}(x+y)\right] I_o(\sqrt{xy})$$

From which,

$$Q(o, \sqrt{t}) = e^{-\frac{t}{2}} \qquad \text{....B.29}$$

Equation B.28 substituted into Equation B.29 gives

$$I_1' = \exp\left[-\frac{C}{2}\right]\cdot\left\{-1+1\right\} = 0 \qquad \text{....B.30}$$

On substituting the partial fractions, Equation B.25, into Equation B.24, the following is obtained.

$$I_1'' = \exp\left[-\frac{C}{2} - A(B+\tfrac{1}{2})\right] \frac{1}{2\pi j}\int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega \frac{\exp\left[\frac{A\Omega}{2}+\frac{C}{2\Omega}\right]}{(2B+1)-\Omega}$$

$$-\frac{1}{2\pi j}\int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega \frac{\exp\left[\frac{A\Omega}{2}+\frac{C}{2\Omega}\right]}{1-\Omega} \qquad \text{....B.31}$$

After substituting Equation B.26 into the first integral above, the following expression is obtained for the integral:

$$\frac{1}{2\pi j}\int_{\Gamma-j\infty}^{\Gamma+j\infty} d\Omega \frac{\exp\left[\frac{A}{2}+\frac{C}{2}\right]}{(2B+1)-} = \frac{1}{2\pi j}\int_{\frac{\Gamma}{2B+1}-j\infty}^{\frac{\Gamma}{2B+1}+j\infty} dD$$

$$\cdot\frac{\exp\left[A(2B+1)\frac{D}{2}+\frac{C}{2(2B+1)D}\right]}{1-D}$$

$$= -\exp\left[-\tfrac{1}{2}\left\{A(2B+1) + \frac{C}{2B+1}\right\}\right] Q\left(\sqrt{A(2B+1}\ , \sqrt{\frac{C}{2B+1}}\right)$$

The second integral is already in the form of Equation B.20. Thus

$$I_1'' = \exp\left[-\frac{C}{2} + A(B+\tfrac{1}{2})\right] - \exp\left[\frac{A(2B+1)}{2} + \frac{C}{2(2B+1)}\right]$$

$$Q\left(\sqrt{A(2B+1)}, \sqrt{\frac{C}{2B+1}}\right) + \exp\left[\frac{A}{2} + \frac{C}{2}\right] Q(\sqrt{A}, \sqrt{C})$$

....B.32

Since I' = o, from Equation B.22,

$$I_1 = \exp\left[\frac{-CB}{2B+1}\right] Q\left(\sqrt{A(2B+1)}, \sqrt{\frac{C}{2B+1}}\right) - \exp(-AB)\ Q(\sqrt{A}, \sqrt{C})$$

....B.33

If it is recalled that

$$A = \frac{2\mu\hat{k}_\alpha}{1-\mu}$$

$$B = \frac{1-\mu}{2\mu}$$

and

$$C = \frac{2\hat{y}_{j-1}}{1-\mu}$$

then,

$$I_1 = \exp\left[-\hat{y}_{j-1}\right] Q\left(\sqrt{\frac{.k_\alpha}{1-\mu}}, \frac{2\mu\hat{y}_{j-1}}{1-\mu}\right) -$$

$$\exp\left[-\hat{k}_\alpha\right] Q\left(\sqrt{\frac{2\mu\hat{k}_\alpha}{1-\mu}}, \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}}\right)$$

....B.34

Lastly from Equations B.12, B.13 and B.34,

$$P\left[K_{\alpha-1} < |S|^2 \le K_\alpha,\ Y_{j-1} < |E|^2 \le Y_j\right]$$

$$= e^{-\hat{y}_{j-1}} \cdot Q\left(\sqrt{\frac{2\hat{k}_\alpha}{1-\mu}},\ \sqrt{\frac{2\mu\hat{y}_{j-1}}{1-\mu}}\right) - e^{-\hat{k}_\alpha} \cdot Q\left(\sqrt{\frac{2\mu\hat{k}_\alpha}{1-\mu}},\ \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}}\right)$$

$$- e^{-\hat{y}_j} \cdot Q\left(\sqrt{\frac{2\hat{k}_\alpha}{1-\mu}},\ \sqrt{\frac{2\mu\hat{y}_{j-1}}{1-\mu}}\right) + e^{-\hat{k}_\alpha} \cdot Q\left(\sqrt{\frac{2\mu\hat{k}_\alpha}{1-\mu}},\ \sqrt{\frac{2\hat{y}_j}{1-\mu}}\right)$$

$$- e^{-\hat{y}_{j-1}} \cdot Q\left(\sqrt{\frac{2\hat{k}_{\alpha-1}}{1-\mu}},\ \sqrt{\frac{2\mu\hat{y}_{j-1}}{1-\mu}}\right) + e^{-k_{\alpha-1}} \cdot Q\left(\sqrt{\frac{2\mu\hat{k}_{\alpha-1}}{1-\mu}},\ \sqrt{\frac{2\hat{y}_{j-1}}{1-\mu}}\right)$$

$$+ e^{-\hat{y}_j} \cdot Q\left(\sqrt{\frac{2\hat{k}_{\alpha-1}}{1-\mu}}\ \sqrt{\frac{2\mu\hat{y}_j}{1-\mu}}\right) - e^{-k_{\alpha-1}} \cdot Q\left(\sqrt{\frac{2\mu\hat{k}_{\alpha-1}}{1-\mu}}\ \sqrt{\frac{2\hat{y}_j}{1-\mu}}\right)$$

....B.35

APPENDIX C: Calculation of the Correlation Function of the Predictor

A block diagram of the sub-optimum prediction which was used in Chapter 5 is given in Fig. 5.1. In this Appendix, the fading pilot-tone signal, p z(t). is written simply z(t). Therefore, the autocorrelation function of z(t), $R_{zz}(o)$ (at τ=o), will give the power in the pilot tone signal, E . The quantity that is calculated here is

$$\lambda(\tau) = \frac{\left|\overline{E(t)\ z^*\ (t+\tau)}\right|^2}{\overline{EE^*}\quad \overline{zz^*}} \qquad \text{....C.1}$$

The function $\lambda(\tau)$ is a measure of the quality of the prediction made τ sec into the future. It is assumed that z(t) has a Gaussian autocorrelation, that is,

$$R_{zz}(\tau) = R_{zz}(o)\ e^{-\frac{1}{2}(\omega_f\ \tau)^2}$$

and

$$S_z(\omega) = \sqrt{2\pi}\ \frac{R_{zz}(o)}{\omega_f}\ e^{-\frac{1}{2}(\frac{\omega}{\omega_f})^2}$$

Further the noise, n(t), is assumed to have a flat spectrum, that is

$$S_n(\omega) = 4N_o$$

The band limiting filter is assumed to have frequency function,

$$H(\omega) = e^{-\frac{1}{4}(\frac{\omega}{\sigma})^2}$$

Now from Fig. 5.

$$E(t) = s(t) + \propto s(t) \qquad \text{....C.2}$$

It is intended to expand $\lambda(\tau)$ in Equation C.1 in a power series and chose $\propto$ to make the second order terms cancel. This will tend to make the prediction filter near optimum for small $w_f\ \tau$. At the output of the filter,

$$S_N(\omega) = S_n(\omega)\ |H(\omega)|^2 = 4N_o\ e^{-\frac{1}{2}(\frac{\omega}{\sigma})^2} \qquad \text{....C.3}$$

$$S_Z(\omega) = \frac{\sqrt{2\pi}\ \ R_{zz}(o)}{\omega_f}\ e^{-\frac{1}{2}\frac{\omega}{\omega_f}\left[1 + \frac{\omega_t}{\sigma^2}\right]}$$

$$= \sqrt{2\pi}\ \frac{R_{zz}(o)}{\omega_f}\ e^{-\frac{1}{2}\frac{\omega^2}{\omega_f^2}\mu^2}$$

where

$$\mu = \left[1 + \frac{\omega^2}{\mu^2}\right]$$

The first two derivatives of $R_{ZZ}(\tau)$ and $R_{NN}\ (\tau)$ will be required shortly. They are therefore tabulated below. (Note that $R_{ZZ}(\tau)$ and $R_{NN}(\tau)$ one found from the inverseFourier transforms of $S_Z(\omega)$ and $S_N(\omega)$)

$$R_{ZZ}(\tau) = \frac{R_{zz}(o)}{\mu}\ e^{-\frac{1}{2}\left[\frac{\omega_f\tau}{\mu}\right]^2}$$

$$\dot{R}_{ZZ}(\tau) = \frac{R_{zz}(o)}{\mu} \quad \frac{\omega_f^2 \tau}{\mu^2} \quad e^{-\frac{1}{2}\left[\frac{\omega_f \tau}{\mu}\right]^2}$$

$$\ddot{R}_{ZZ}(\tau) = -\frac{R_{zz}(o)}{\mu} \quad \frac{\omega_f^2}{\mu^2} \quad e^{-\frac{1}{2}\left[\frac{\omega_f \tau}{\mu}\right]^2} \left[1 - \frac{\tau^2 \ \omega_f^2}{\mu^2}\right] \qquad \text{....C.5}$$

Thus,

$$R_{ZZ}(o) = \frac{R_{zz}(o)}{\mu}$$

$$\dot{R}_{ZZ}(o) = o$$

$$\ddot{R}_{ZZ}(o) = -\frac{R_{zz}(o) \ \omega_f^2}{\mu \mu^2} \qquad \text{....C.6}$$

$$R_{NN}(\tau) = \frac{4N_o\sigma}{\sqrt{2\pi}} \quad e^{-\frac{1}{2}(\sigma\tau)^2}$$

Similarly,

$$R_{NN}(o) = \frac{4N_o\sigma}{\sqrt{2\pi}}$$

$$\dot{R}_{NN}(o) = o$$

$$\ddot{R}_{NN}(o) = -\frac{4N_o}{\sqrt{2\pi}} \quad \sigma^3 \qquad \text{....C.7}$$

Thus,

$$\overline{E(t)\ z^*(t+\tau)} = \overline{\left[s(t) + \propto \dot{s}(t)\right] z^*(t+\tau)}$$

$$= \overline{Z(t)\ z^*(t+\tau)} + \propto \overline{\dot{Z}(t)\ z^*(t+\tau)} \qquad \text{....C.8}$$

Now,

$$\overline{Z(t)\ z^*(t+\tau)} = F^{-1}\left\{H(\omega)S_z(\omega)\right\} \qquad \text{....C.9}$$

and

$$H(\omega)S_z(\omega) = \sqrt{2\pi}\ \frac{R_{zz}(o)}{\omega_f}\ e^{-\frac{1}{2}\frac{\omega^2}{\omega_f^2}\left[1 + \frac{\omega_f^2}{2\sigma^2}\right]}$$

$$= \sqrt{2\pi}\ \frac{R_{zz}(o)}{\omega_f}\ e^{-\frac{1}{2}\left[\frac{\omega}{\omega_f}\right]^2 \nu^2} \qquad \text{....C.10}$$

where,

$$\nu^2 = \left[1 + \frac{\omega_f}{2\sigma^2}\right]$$

Thus,

$$\overline{Z(t)\ z^*(t+\tau)} = \frac{R_{zz}(o)}{\nu}\ e^{-\frac{1}{2}\left[\frac{\omega_f \tau}{\nu}\right]^2} \qquad \text{....C.11}$$

Similarly,

$$\overline{\dot{Z}(t)\ z^{*}\ (t+\tau)} = \frac{R_{zz}(o)}{\nu} \quad \frac{\omega_f^2 \tau}{\nu^2}\ e^{-\frac{1}{2}\left[\frac{\omega_f \tau}{\nu}\right]}$$

....C.12

On substituting Equations C.11 and C.12 into Equation C.8, the following is obtained.

$$\overline{E(t)\ z^{*}\ (t+\tau)} = \frac{R_{zz}(o)}{\nu}\ e^{-\frac{1}{2}\left[\frac{\omega_f \tau}{\nu}\right]^2} \left\{1 + \propto \frac{\omega_f^2 \tau}{\nu^2}\right\}$$

....C.13

Also,

$$\overline{E\ E^{*}} = \overline{Z(t)\ Z^{*}\ (t)} + \propto^{*}\ \overline{Z(t)\ \dot{Z}^{*}\ (t)}$$

$$+\propto \overline{\dot{Z}(t)\ Z^{*}\ (t)} + |\propto|^2\ \overline{\dot{Z}(t)\ \dot{Z}^{*}\ (t)}$$

+ similar noise terms C.14

Derivations can be continued using

$$R_{ZZ}^{(n+m)}\ (\tau) = (-1)^n\ \overline{Z^{(n)}\ (t)\ Z^{(m)*}\ (t)}$$

and remembering that

$R_{ZZ}^{(1)}$ (o) = o for both signal and noise

Thus

$$\overline{E\ E^*} = R_{ZZ}(o) - |\alpha|^2 R_{ZZ}^{(2)}(o) + R_{NN}(o) - |\alpha|^2 R_{NN}^{(2)}(o)$$

....C.15

From the above tabulations (Equations C.6 and C.7),

$$\overline{E\ E^*} = \frac{R_{zz}(o)}{\mu}\left\{1 - |\alpha|^2 \frac{\omega_f^2}{\mu^2}\right\} + \frac{4N_o}{\sqrt{2\pi}}\left\{1 - |\alpha|^2 \sigma^2\right\}$$

From Equations C.1, C.2, C.13 and C.15, the following can now be written:

$$\lambda(\tau) = \frac{\frac{1}{\nu^2} e^{-\left[\frac{\omega_f \tau}{\nu}\right]}\left[1 + \alpha\tau \frac{\omega_f^2}{\nu^2}\right]}{\frac{1}{\mu}\left\{1 - |\alpha|^2 \frac{\omega_f^2}{\mu^2}\right\}\left\{1 + \frac{4N_o\sigma_\mu\left[1 - |\alpha|^2 \sigma^2\right]}{\sqrt{2\pi}\, R_{zz}(o)\left[1 - \alpha^2 \frac{\omega_f^2}{\mu^2}\right]}\right\}}$$

...C.16

Because the interest is mainly on good prediction, several approximations can be made will simplify optimizing α and σ. These approximations are:

$\alpha \approx \tau$

$\tau\omega_f << 1$

and

$\nu+\mu \approx 1$

Equation C.16 can then be written as

$$\lambda(\tau) = \frac{\mu}{\nu^2} e^{-\left[\frac{\omega_f \tau}{\nu}\right]^2} \frac{A}{B} \frac{1}{1 + \frac{4N_o\sigma_\mu}{\sqrt{2\mu}\, R_{zz}(o)}} \qquad \ldots.C.17$$

where,

$$A \approx \left[1 + (\tau\omega_f)^2\right]^2$$

$$B \approx 1 + (\tau\omega_f)^2$$

Thus

$$\frac{A}{B} \approx 1 + (\tau\omega_f)^2 \qquad \ldots.C.18$$

On multiplying Equation C.18 on both sides by the power series expansion of

$$e^{-\frac{1}{2}\left[\frac{\tau\omega_f}{\nu}\right]^2}$$

the following is obtained.

$$\frac{A}{B} e^{-\frac{1}{2}\left[\frac{\tau\omega_f}{\nu}\right]^2} \approx 1 - \tfrac{1}{2}(\tau\omega_f)^4 \qquad \ldots.C.19$$

From Equations C.17, C.18 and C.19

$$\lambda(\tau) = \frac{\mu}{\nu^2} \cdot \frac{1 - \frac{1}{2}(\tau\omega_f)^4}{1 + \frac{4N_o \ \sigma_\mu}{\sqrt{2\pi}\ R_{zz}(o)}}$$

Now, $\lambda(\tau)$ is required to be near unity. Therefore, from Equation C.20, it is seen that the following conditions are required to be true.

1) $\frac{\mu}{\nu^2} \approx 1$ C.21

2) $\frac{4N_o \ \sigma_\mu}{\sqrt{2\pi}\ R_{zz}(o)} \ll 1$ C.22

The first requirement is on the ability of bandpass filter to pass energy in the pilot tone signal. The second is a requirement on the SNR in the pilot tone channel.

Recall that

$$\frac{\mu}{\nu^2} = \frac{1 + \frac{\omega_f^2}{\sigma^2}}{1 + \frac{\omega_f^2}{2\sigma^2}} \qquad \text{....C.23}$$

The first requirement, Equation C.21, could be expressed as:

$\frac{\mu}{\nu^2} > 1 - \varepsilon$ where $\varepsilon \ll 1$ when an upper bound

$\frac{\omega_f^2}{\sigma^2}$ or a lower bound on $\frac{\sigma}{\omega_f}$ can be found.

From the second requirement,

$$\frac{4N_o \sigma_\mu}{\sqrt{2\pi} R_{zz}(o)} < \varepsilon$$

These bounds for chosen values of ε are tabulated in Table C.1

ε	$\frac{\sigma}{\omega_f} >$	$\frac{E_p}{\omega_f} >$	E_p for $\omega_f = 2\pi \times 10$
0.005	2.13	188	1.18×10^4
0.01	1.74	80.3	5.05×10^3
0.02	1.42	34.8	2.18×10^3

Table C.1 The Required Pilot Tone Power

From Table C.1, if E_p is made large enough so that ε is small enough, then $\lambda(\tau)$ can be written as

$$\lambda(\tau) = 1 - \tfrac{1}{2}(\tau\omega_f)^4 \qquad \text{....C.24}$$

This expression is plotted in Fig. 8.7.