COMMUNICATION OVER FADING DISPERSIVE CHANNELS

JOHN S. RICHTERS

TECHNICAL REPORT 464

NOVEMBER 30, 1967

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS 02139

The Research Laboratory of Electronics is an interdepartmental laboratory in which faculty members and graduate students from numerous academic departments conduct research.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, by the JOINT SERVICES ELECTRONICS PROGRAMS (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract No. DA 28-043-AMC-02536(E); additional support was received from the National Aeronautics and Space Administration (Grant NsG-334).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS

Technical Report 464 November 30, 1967

COMMUNICATION OVER FADING DISPERSIVE CHANNELS

John S. Richters

This report is based on a thesis submitted to the Department of Electrical Engineering, M. I. T., August, 1967, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

(Manuscript received September 1, 1967)

Abstract

The transmission of digital information over a fading dispersive channel is considered, subject to a bandwidth constraint on the input signals. A specific signaling scheme is proposed, in which information is transmitted with signals formed by coding over a set of smaller basic signals, all of which excite approximately independent and orthogonal outputs. The problem is then modeled as one of block coding over successive independent uses of a diversity channel.

Upper and lower bounds to the minimum error probability attainable by such a scheme are derived. These bounds are exponentially decreasing in terms of the time available for information transmission, and agree asymptotically for a range of rates. These bounds are used to interpret the significance of different signal and channel parameters, and the interplay between them.

Some conclusions are drawn concerning the nature of good input signals, the major one being that any basic signal should be transmitted at one of a small number of discrete voltage levels. Several numerical examples are included, to illustrate how these results may be applied in the estimation of performance levels for practical channels.

TABLE OF CONTENTS

CONTENTS

I. INTRODUCTION

Channels that exhibit both fading and dispersion are often used for communication purposes. Perhaps the best known examples of such channels are fading ionospheric radio links, and tropospheric scatter systems. More recently, communication has been achieved by reflection from the lunar surface and through plasmas, and fading-dispersive channels have been artificially created as in the case of Project West Ford.

Our main concern is the use of such a channel for digital communication when the input signals are bandwidth-constrained. We wish to obtain estimates of performance for such a communication system, and to use these to determine how the interaction between input signals and channel statistics affects the ultimate error probability. Our tools for this study consist of the techniques of communication theory and information theory, and we shall rely heavily on the use of bounds on the attainable error probability.

1.1 COMMUNICATION SYSTEM

By digital communication, we refer to transmission of one of M equiprobable input signals, at a rate $R = \frac{1}{T} \ln M$ nats/sec. The signal is corrupted by the channel, and we choose to decode at the channel output for minimum error probability. We shall use the notation P_e to stand for the minimum attainable error probability, minimized over all sets of input signals and decoders (subject to such constraints as bandwidth, average power, etc.). We wish to determine in what fashion R and the channel characteristics affect P_e.

We envisage the channel as composed of a large number of point scatterers, each with its own time delay, Doppler shift, and reflection cross section. Subject to certain restrictions on the input signals and assumptions about the scatterers, the channel characteristics may be summarized by the scattering function $\sigma(r, f)$, where $\sigma(r, f)\,dr\,df$ is the average energy scattering cross section of those scatterers with time delays between r and r + dr and Doppler shifts between f and f + df.

We consider constructing the M input signals by modulating N smaller basic signals designed so that each is independently affected by the channel (that is, no intersymbol interference or memory). Each of the M signals will consist of a set of N modulation levels, one for each basic signal, and the problem then reduces to coding over the N basic signal modulation levels. The bandwidth constraint enters into the determination of N.

Such a model describes many actual systems for communication over fading-dispersive channels, although usually no attempt is made to exploit the possibilities of coding. For this type of scheme our analysis will provide ultimate performance limits. Moreover, our results should be indicative of the manner in which the signal and channel parameters interact for more general communication systems.

1.2 HISTORY OF THE PROBLEM

There is a large body of previous work concerning this kind of problem. We make no attempt to give a complete listing of results, but try to mention earlier work that seems most pertinent to the present analysis.

For digital communication, with specified input signals, the optimum receiver is defined as the one that minimizes error probability. The determination of the optimum receiver has been considered by several authors in recent years. Price,[1,2] Price and Green,[3] and Kailath[4] have studied problems involving M-ary signaling over Gaussian multipath channels, and Middleton,[5] Turin,[6] and Bello[7] have considered the binary-signal situation. The result of these studies is that the form of the optimum receiver is known (it can be interpreted as an "estimation-correlation" operation[4]).

Although the form of the optimum receiver for an arbitrary set of M input signals is known, little can be said of the resulting system performance. For example, we would like to vary the set of input signals (subject to the constraints), consider each set with the corresponding optimum receiver, and choose the combination that results in an error probability smaller than all the others. This set of signals and receiver, or decoder, would then be defined as the optimum combination for communication over the particular channel under consideration. Unfortunately, because of the complicated form of the optimum decoder, exact error probabilities have been calculated only in special cases.

There is, however, another method for attacking this problem, one that avoids some of the complexities just mentioned. This method by-passes the exact calculation of error probabilities in favor of the determination of <u>bounds</u> to the probability of error that may be achieved by transmitting over such a channel. It is frequently possible to derive upper and lower bounds to the achievable probability of error which agree exponentially. For our purposes, this will be nearly as satisfactory as an exact error probability because our primary interest is not merely the determination of a minimum error probability but the study of how the interaction between input signals and channel statistics affects the error probability.

These bounding techniques have been widely applied to discrete memoryless channels (see for example, Shannon,[8] Fano,[9] and Gallager[10]), and have recently been used by Holsinger[11] and Ebert[12] to study the problem of communication over a deterministic colored Gaussian noise channel. Pierce[13,14] has derived bounds for the binary case with slowly fading channel and variable data rate, and for a special M-ary transmission case with linear filtering at the decoder. Jacobs[15] first showed that the capacity of an infinite bandwidth fading channel is the same as that of an equivalent Gaussian channel.

The problem considered here is quite similar to that analyzed by Kennedy,[16] who treated fading dispersive channels with M-ary orthogonal signals, and derived both upper and lower bounds to the attainable error probability. There is one most important difference, however, in that orthogonal signals require a large amount of bandwidth, and we assume that the input signals must be bandwidth-constrained. This turns out to be a major distinction, making the problem more difficult to solve, and in many cases, physically more meaningful.

The reader who is interested mainly in the nature of our results might choose at this point to glance at the first sections of Section IV, where some of the major points are summarized and discussed.

1.3 SUMMARY OF THIS RESEARCH

In Section II, we derive an approximate mathematical model of the channel, and discuss the problems involved in using a fading dispersive channel for digital communication of one of M arbitrary input signals. Then we enumerate several specific types of signaling schemes, and mention briefly the problems involved in signal design.

Section III is devoted to determination and evaluation of bounds to error probability, by using the channel statistics derived in Section II. It is highly mathematical in nature and provides the abstract results necessary for the evaluation of performance levels.

In Section IV, we discuss the physical significance of the results of Section III, and relate these results to the communication problem set forth in Section II. We illustrate possible applications through some numerical examples, and return to the question of signal design.

II. DIGITAL COMMUNICATION OVER FADING CHANNELS

2.1 CHANNEL MODEL

The problem under consideration is the transmission of information over a linear, time-continuous, fading, dispersive channel. The channel is envisioned as a collection of a large number of independent point scatterers, such as the ones that made up the West Ford belt, for example. Any input signal s(t) will be a narrow-band signal centered at some nominal high carrier frequency, ω_o, so that

$$s(t) = R_e\left\{u(t)\, e^{j\omega_o t}\right\}, \tag{1}$$

where u(t) is known as the complex lowpass modulation of s(t). By narrow-band, we mean that the bandwidth of s(t) is much less than ω_o. Consider a particular scatterer with range r sec and Doppler shift f Hz. If $2\pi f \ll \omega_o$, then the signal returned by that scatterer will be approximately

$$y(t) = R_e\left\{Au(t-r)\, e^{j[(\omega_o - 2\pi f)t - \omega_o r]}\right\}, \tag{2}$$

where A is an attenuation factor. It is unlikely that the value of r could be specified within $2\pi/\omega_o$ sec, so that it is reasonable to consider the quantity $\theta \equiv -\omega_o r$ as a random variable distributed uniformly over $(0, 2\pi)$.

Let us partition the set of possible ranges and Doppler shifts into small cells, and consider one particular cell centered on range r and Doppler shift f, containing a number of scatterers. If the dimension of the cell in r is small compared with the reciprocal bandwidth of u(t), and the dimension in f is much less than the reciprocal time duration of u(t), then the contribution to the total received process from all the scatterers in that cell will be

$$y(t, r, f) = R_e\left\{A(r, f)\, e^{j\theta(r, f)}\, u(t-r)\, e^{j(\omega_o - 2\pi f)t}\right\}, \tag{3}$$

where A and θ describe the resultant amplitude and phase from all scatterers in the cell.

With a large number of independent scatterers in the cell, each with random phase and approximately the same reflection coefficient, A(r, f) will tend to have a Rayleigh distribution, while θ(r, f) will become uniform on $(0, 2\pi)$.[17] In that case, the real and imaginary parts of $A(r, f)\, e^{j\theta(r, f)}$ will be statistically independent, zero-mean, Gaussian random variables with equal variances. Note that the number of independent scatterers in a cell does not have to be particularly large for this to be true; in fact, as few as six may give a good approximation to Rayleigh amplitude and random phase.[18]

With the previous assumptions, the total signal y(t) received from all of the cells for a given input u(t) will be a zero-mean Gaussian random process, and thus the statistics of y(t) may be completely determined from the correlation function

$$R_y(t, \tau) = \overline{y(t)\, y(\tau)}. \tag{4}$$

If it is assumed that scatterers in different cells are uncorrelated, and that the cells are small enough so that sums may be replaced by integrals, the expectation may be carried out, with the result that

$$R_y(t, \tau) = R_e\left\{e^{j\omega_o(t-\tau)} \int_0^\infty \int_0^\infty \sigma(r, f)\, u(t-r)\, u^*(\tau-r)\, e^{j2\pi f(\tau-t)}\, drdf\right\}. \tag{5}$$

The quantity $\sigma(r, f)$ equals $\frac{1}{2}\overline{A^2(r, f)}$, and is known as the channel <u>scattering</u> <u>function</u>. It represents the density of reflecting cross section, that is, $\sigma(r, f)\, drdf$ is the average energy scattering cross section of those scatterers with ranges between r and r + dr sec and Doppler shifts between f and f + df Hz. If we assume that there is no average energy loss through the channel (this is no restriction, since any average attenuation may be accounted for by a normalization of the input signal level), then

$$\int_0^\infty \int_0^\infty \sigma(r, f)\, drdf = 1. \tag{6}$$

A typical scattering function is illustrated in Fig. 1. The most important characteristics of $\sigma(r, f)$ are B, the frequency interval in f outside of which $\sigma(r, f)$ is essentially zero, and L, the time interval in r outside of which $\sigma(r, f)$ is effectively zero. The quantity B is called the <u>Doppler</u> <u>spread</u>, and represents the average amount that an input signal will be spaced in frequency, while L is known as the <u>multipath</u> <u>spread</u>, and represents the average amount in time that an input signal will be spread. Table 1 lists approximate values of B and L for some practical cases. The exact way in which B and L are defined is unimportant, since these will only be used as rough indications of channel behavior.

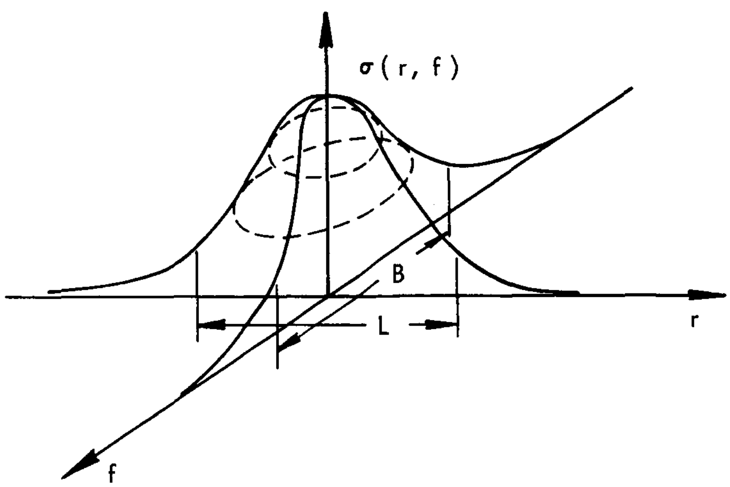

Fig. 1. A typical $\sigma(r, f)$.

Table 1. Values of B and L.

Channel	B(Hz)	L(sec)
Ionospheric Scatter	10	10^{-4}
Tropospheric Scatter	10	10^{-6}
West Ford Belt	10^3	10^{-4}
Chaff Clouds	10^2	5×10^{-6}
Moon Scatter	10	10^{-2}

If $\sigma(r, f)$ is well-behaved and unimodal as shown, then an input of a single sine wave will result in an output whose correlation time or fade duration will be approximately $1/B$ sec; that is, the fading of any two points on the received process should be essentially independent if they are separated by $1/B$ or more seconds. If the input consists of two sinusoids, they should fade independently if their frequency spacing is at least of the order of $1/L$ Hz. It has been shown elsewhere,[16] that if the input signal has duration T and bandwidth W, the received waveform will have a number of degrees of freedom (the total number of independent samples in the output process) approximately equal to

$$K = \begin{cases} (1+BT)(1+LW), & \text{if } BL \leq 1 \text{ or } TW = 1 \\ (T+L)(W+B), & \text{otherwise} \end{cases} \tag{7}$$

We emphasize the roughness of relation (7). The channel will later by treated as a diversity channel with K paths, and (7) is used only to estimate K. The validity of this channel model will not depend on the accuracy of (7), although any numerical results will be approximate to the extent that it gives a good approximation of K.

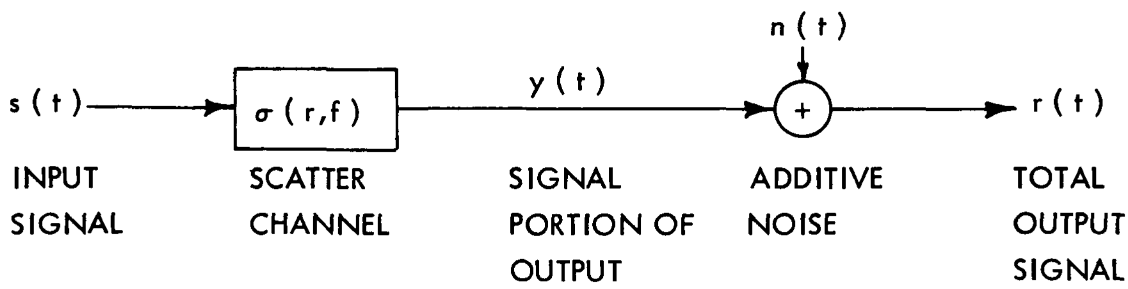

Fig. 2. The channel model.

We also assume that any signal that we receive is corrupted by the addition of additive, white, zero-mean Gaussian noise (such as front-end receiver noise) with density $N_0/2$ watts per Hz. A block diagram of the over-all channel model is shown in Fig. 2.

2.2 DIGITAL COMMUNICATION

The basic problem to be considered is digital communication over a fading, dispersive channel, accomplished by sending one of a set of M equiprobable signals and decoding at the output for minimum average error probability, P_e. These input signals are assumed to be time-limited to (0,T), and we shall be interested mainly in the asymptotic case, where M and T are large. We also require that the signals all lie essentially within a bandwidth W, and that the average transmitter power be less than or equal to P_t. Ideally, our procedure would be to choose an arbitrary set of M input signals that satisfy the constraints, determine the optimum receiver for these signals, evaluate the resulting P_e, and then choose the set of M signals that result in minimum P_e for the particular fading channel under consideration.

The first step is to find the optimum receiver for an arbitrary set of M signals. This problem has been previously considered by several authors, with the result that the form of the optimum receiver is known. Perhaps the simplest description is that given by Kailath,[4] who shows that the optimum receiver may be interpreted as a bank of "estimator-correlator" operations. That is, one element of the receiver takes the channel output, assumes that a particular input signal was sent, computes a minimum mean-square error estimate of the signal portion of the output (before the noise is added), and correlates this with the actual output signal. Unfortunately, except in a few very special cases, it is practically impossible to use the previous results for the computation of error probabilities for a given set of M signals, because of the complicated expressions that must be evaluated. This is hardly surprising, for even in the simpler case of the nonfading, additive Gaussian noise channel, exact error probabilities cannot usually be computed for large numbers of input signals.

In the case of the additive Gaussian noise channel, one approach is to use <u>bounds</u> on P_e, rather than to attempt to compute actual error probabilities. This has been done successfully by various authors,[10,11,12,19] who have found upper and lower bounds to P_e which in many cases agree very well for large M and T. This allows one to design signals that minimize the <u>bound</u> on P_e, which has proved much simpler than the minimization of P_e itself, and yet has often yielded useful results. When a fading channel is considered, however, the more severe nature of the channel variations makes even computing bounds to P_e a major problem.

In order to illustrate the difficulties, and obtain results that will be useful later, we consider an input signal, xs(t), where s(t) is a unit energy signal, and x is an amplitude modulation. It is possible to expand the output signal in a Karhunen-Loeve orthonormal expansion,[20] using the eigenfunctions of $R_y(t,\tau)$ to obtain

$$r(t) = \sum_{k=1}^{\infty} r_k \Psi_k(t) \tag{8}$$

$$r_k = \int_{-\infty}^{\infty} r(t)\ \Psi_k^*(t)\ dt \tag{9}$$

$$\mu_k \Psi_k(t) = \int_{-\infty}^{\infty} R_y(t,\tau)\ \Psi_k(\tau)\ d\tau. \tag{10}$$

Properties of the solutions to the integral equation given in (10) have been enumerated by several authors.[10, 11, 12, 19, 21] If ω_o is very large, it has been shown[16] that the eigenfunctions $\Psi_k(t)$ will approximate conjugate pairs, with both functions of any pair having the same eigenvalue. Thus, for any particular eigenvalue μ_k, if $\Psi_k(t)$ is a solution to (10), then $\Psi_k^*(t)$ is orthogonal to $\Psi_k(t)$ and is also a solution. Moreover, it may be shown that

$$\overline{r_i r_k^*} = \left(x^2 \mu_k + \tfrac{1}{2} N_o\right) \delta_{ij} \tag{11}$$

$$\overline{r_i r_k} = 0, \tag{12}$$

provided the additive noise is white. Thus each r_k is a zero-mean complex Gaussian random variable, and they are all statistically independent.

Because of our normalization of the channel and input,

$$\sum_{k=1}^{\infty} \mu_k = \int_0^{\infty} \sum_{k=1}^{\infty} \mu_k \Psi_k(t)\ \Psi_k^*(t)\ dt = \int_{-\infty}^{\infty} R_y(t,t)\ dt = 1. \tag{13}$$

For later convenience, lump each pair of μ_k together and call the result λ_k, so that $\lambda_k = 2\mu_k$; then, there will be half as many λ's as μ's, and

$$\sum_{k=1}^{\infty} \lambda_k = 1. \tag{14}$$

Each λ_k may be interpreted as the fraction of total received energy contributed by the k^{th} diversity path, on the average. The total number of significant λ_k will be approximately equal to K, as given by (7).

If one of two input signals is sent, we can make use of the fact that any two symmetric, positive-definite matrices can be simultaneously diagonalized,[22] to find one (nonorthogonal) expansion that results in independent components of $\underline{r}$, regardless of which signal was sent. The eigenvalues will no longer have the same simple interpretation as before, however, and more importantly, if both signals are simultaneously modulated and transmitted, they cannot be separated at the output because they will each excite the same output components and hence result in crosstalk. For M arbitrary signals, there is no one output expansion that results in independent signal components, regardless of what signal was sent.

Due to the fact that we cannot separate the components received at the channel output from many different arbitrary input signals, the general communication problem seems insurmountable, at this time. If, however, we start with a set of basic input signals that can be uniquely separated at the channel output, with output components that are all statistically independent, then we could consider coding over this signal set and have some hope of computing error probabilities. In particular, the problem could be formulated as one of communication over a time-discrete, amplitude-continuous memoryless channel.

Unfortunately, such a restriction to a specific basic input set would preclude any general determination of ultimate performance limits. Coding over a specific set will certainly result in an upper bound to the minimum attainable P_e. Moreover, for reasons of simplicity, many existing fading dispersive systems employ basic signal sets of the type just described, and thus analysis would yield bounds on the best possible results for systems of that class. Finally, an analysis of this type of scheme will still allow us to obtain a greater understanding of the performance limits of bandwidth-constrained fading dispersive channels than has been previously available. We shall return to some of these points.

2.3 SIGNALING SCHEME

We choose a set of input signals that have the desirable properties mentioned above, namely a set of time and frequency translates of one basic signal, with sufficient guard spaces allowed so that the output signals are independent and orthogonal. In the first subsection we consider the output statistics for an arbitrary choice of basic signal and derive an equivalent discrete memoryless channel. In the second, some aspects of signal design are discussed. The problem is so formulated that existing bounds on error probability may later be applied.

2.31 Input Signals and Output Statistics

Consider a unit energy input signal that is approximately time-limited to T_s seconds and bandlimited to W_s Hz, but is otherwise arbitrary. Time- and frequency-shifted replicas of this signal with sufficient guard spaces in time and frequency left between them will then yield approximately orthogonal and independent channel outputs. As we have said, separating the input signals by more than L seconds in time will make the output signals approximately orthogonal, and an additional 1/B will make them independent, too. Similarly, in the frequency domain, a spacing of B Hz will make the output signals approximately orthogonal, and an additional 1/L will make them independent. We take our basic input signals to be the (say) N time and frequency translates of the given (T_s, W_s) waveform that satisfy these independence and orthogonality conditions and exhaust the available time-bandwidth space (see Fig. 3). The outputs resulting from these N basic signals will be independent and orthogonal.

Fig. 3. Basic input signals.

Suppose we construct our M input signals by simultaneously amplitude-modulating the N basic signals. In this case, the m^{th} input signal may be expressed as $\underline{x}_m = (x_{m1}, x_{m2}, \ldots, x_{mN})$, where x_{mn} is the modulation on the n^{th} basic signal. Since each basic input signal is assumed to have unit energy, we require

$$\frac{1}{M} \sum_{m=1}^{M} \sum_{n=1}^{N} x_{mn}^2 \leq TP_t, \tag{15}$$

in order to satisfy the average input energy constraint.

Consider one basic signal, with scalar amplitude modulation x. This is just the situation that we considered in section 2.2, and the resulting output random process r(t) may be transformed into a vector $\underline{r}$, whose components are statistically independent, zero-mean, complex Gaussian random variables. The k^{th} component of $\underline{r}$ may be split into its real and imaginary parts,

$$r_k = r_{kre} + jr_{kim}, \tag{16}$$

where each part is a real zero-mean Gaussian variable with variance $\frac{1}{2}\left(x^2\lambda_k+N_o\right)$, and the two are independent, so that

$$p(r_{kre}, r_{kim}|x) = \frac{1}{\pi\left(x^2\lambda_k+N_o\right)} \exp\left(-\frac{r_{kre}^2 + r_{kim}^2}{x^2\lambda_k + N_o}\right). \tag{17}$$

The components of r_k enter the density in the form $r_{kre}^2 + r_{kim}^2$, so, for the purpose of estimating x, it is only necessary to record this sum, and not the individual terms.

Define

$$y_k = \frac{1}{N_o}\left(r_{kre}^2+r_{kim}^2\right), \tag{18}$$

then it can be shown[23] that

$$p(y_k|x) = \left(1+x^2\lambda_k/N_o\right)^{-1} \exp\left(-\frac{y_k}{1 + x^2\lambda_k/N_o}\right). \tag{19}$$

Recall that the properties of $\underline{\lambda} \equiv (\lambda_1, \lambda_2, \dots, \lambda_K)$ depended on the assumption that the channel was lossless, and that any average attenuation should be absorbed into the energy constraint. From (19), we see that N_o can also be absorbed into the energy constraint, provided that we redefine the constraint on the input to be

$$\frac{1}{M} \sum_{n=1}^{N} \sum_{m=1}^{M} x_{mn}^2 \leq \frac{PT}{N_o}. \tag{20}$$

The quantity P is the average output signal power, and the value of x_{mn} in (20) is $\sqrt{\frac{P}{N_oP_t}}$ times the actual channel input signal level. Now

$$p(y_k|x) = \frac{1}{1 + x^2\lambda_k} \exp\left(-\frac{y_k^2}{1 + x^2\lambda_k}\right), \tag{21}$$

so that the density

$$p_{\underline{\lambda}}(\underline{y}|x) \equiv \prod_{k=1}^{K} \left[\frac{1}{1 + x^2\lambda_k} \exp\left(-\frac{y_k}{1 + x^2\lambda_k}\right)\right] \tag{22}$$

governs the output vector $\underline{y}$, given x. Note that $\underline{y}$ is a sufficient statistic for the estimation of x, and is thus all the receiver needs to record.

Until now, except for noting the over-all power constraint in (20), we considered transmission of one amplitude modulated input signal, modeled as one use of an amplitude-continuous channel. Since the N basic signals were chosen to be independent and orthogonal, this is easily generalized to simultaneous modulation and transmission of each signal, thereby reducing the problem to N uses of a memoryless, amplitude-continuous, time-discrete channel, where on each use the input is a scalar x, and the output a vector $\underline{y}$.

Let T_g and W_g denote, respectively, the guard space between adjacent signals in time and bandwidth. Then it is easy to see that

$$N = \frac{TW}{(T_s+T_g)(W_s+W_g)}. \tag{23}$$

The rate of information transmission is usually defined as

$$R \equiv \frac{1}{T} \ln M \text{ nats/second} \tag{24}$$

but the rate that will prove most useful in Section III is

$$R_N \equiv \frac{1}{N} \ln M \text{ nats/channel use.} \tag{25}$$

The energy constraint may be written

$$\frac{1}{M} \sum_{n=1}^{N} \sum_{m=1}^{M} x_{mn}^2 \leq N\alpha \tag{26}$$

$$\alpha \equiv \frac{P}{N_o W} (T_s+T_g)(W_s+W_g). \tag{27}$$

As we have noted, the number of positive λ_k will be approximately given by

$$K = \begin{cases} (1+BT_s)(1+LW_s), & \text{if } BL \leq 1 \text{ or } T_sW_s = 1 \\ (T_s+L)(W_s+B), & \text{otherwise} \end{cases} \tag{28}$$

Note that, as W goes to infinity (the bandwidth constraint is removed), α will go to zero, provided T_s, W_s and the guard spaces are kept finite.

2.32 Signal Design

We consider now several obvious questions that arise with respect to the signaling scheme just described. The N basic signals were chosen to be time- and frequency-shifted replicas of one original signal, and we could conceivably do better by using the same type of time and frequency spacing, but allowing each of the N basic signals to be different. Before evaluation of P_e, this question cannot be answered, but it seems likely that nothing would be gained by such a scheme. For our model, the density $p_\lambda(\underline{y}|x)$ will be the same for a given input signal, no matter how it is shifted in the allotted T, W space. Thus, if a basic signal is in some sense "good," it should be "good," no matter where it is placed in the space. If we find one "good" signal, then we can make them all as "good" by replicating the original in time and frequency. This is quite vague, but in Section IV we shall present some results that reinforce these statements. In any event, it is clearly a simpler mathematical problem to consider just one $\underline{\lambda}$ in place of a set of vectors, so we shall continue to suppose that the N basic signals all have the same $p_{\underline{\lambda}}(\underline{y}|x)$.

When the basic signals are all replicas of one signal, what should that signal be? Referring to Eq. 1, we are free to specify u(t), but it appears in the analysis only through T_s, W_s, and $\underline{\lambda}$. Although T_s and W_s can usually be satisfactorily defined for reasonable signals, there is no simple relationship between u(t) and $\underline{\lambda}$; indeed, finding $\underline{\lambda}$, given u(t), consists of finding the eigenvalues of an integral equation, in general, a very difficult problem. The inverse problem, of finding a u(t) that generates a particular $\underline{\lambda}$, is even harder. Since $\underline{\lambda}$ is the important quantity in the mathematical model, one reasonable approach would be to find the "best" $\underline{\lambda}$, T_s, W_s, and ignore the problem of transforming $\underline{\lambda}$ into a u(t). This would provide a bound on the performance of this type of system, and would give systems designers a goal to aim at. For further discussion of this problem, in the context of an orthogonal signal, unconstrained bandwidth analysis, see Kennedy.[16] Even in this simpler case, however, no definitive answer to this design question could be obtained.

In Section III, when we compute bounds to P_e, it will develop that there are formidable mathematical obstacles to evaluation of the bounds for arbitrary $\underline{\lambda}$, so even this approach appears to be lost to us. In the infinite bandwidth case, however, it was found that equal-strength eigenvalues, where

$$\lambda_k = \frac{1}{K}, \qquad k = 1, 2, \ldots, K, \tag{29}$$

minimized the bound to P_e, and in that sense, were optimum. Furthermore, by varying K in an equal-strength system, a good approximation to the performance of a system with arbitrary $\underline{\lambda}$ could often be obtained. When $\underline{\lambda}$ is given by (29), we find[23] that

$$p_{\underline{\lambda}}(y|x) = (1+x^2/K)^{-K} \exp \frac{-\sum_{k=1}^{K} y_k}{1 + x^2/K} \tag{30}$$

Thus $y \equiv \sum_{k=1}^{K} y_k$ is a sufficient statistic, with conditional density

$$p_K(y|x) = \frac{y^{K-1} \exp\left(-\frac{y}{1+x^2/K}\right)}{\Gamma(K)(1+x^2/K)^K} \tag{31}$$

Therefore, in the special case of an equal-eigenvalue signal, the channel model becomes scalar input-scalar output. Note that if K is presumed to be given by (28), the specification of T_s and W_s provides a complete channel description. These are compelling reasons for restricting the analysis to equal eigenvalues, but in Section III $\underline{\lambda}$ will be kept arbitrary until the need for (and further justification of) the equal eigenvalue assumption becomes apparent.

One other question can be considered at this time. We have remarked that T_g should be at least $L + 1/B$, and W_g should be at least $B + 1/L$ to ensure orthogonality and independence at the output. In some instances, this can result in a large amount of wasted guard space. For example, if B (or L) becomes large, an input signal will be spread a large amount in frequency (or time), and a large guard space is required to separate adjacent signals at the channel output. On the other hand, if B (or L) is very small, there are no orthogonality problems, but a large guard space is required to obtain independent output signals. In the former instance, there is no obvious way of circumventing the need for guard space, in the latter there is.

In particular, one can artificially obtain independence by means of scrambling. Consider several sequential (in time) uses of a (T, W) block of signals. We could now use up any guard space resulting from the 1/B and 1/L terms by interleaving signals from another, or several other (T, W) blocks. Thus, signals could be placed in a (T, W) block with guard spaces of only B in frequency and L in time, while independence can be obtained by coding over signals in different blocks that were all independent. Thus for the same rate R, given by (24), the coding constraint length, N could be increased to

$$N_{sc} = \frac{TW}{(T_s+L)(W_s+B)}, \tag{32}$$

while

$$\frac{1}{M} \sum_{n=1}^{N_{sc}} \sum_{m=1}^{M} x_{mn}^2 \leq a_{sc} N_{sc} \tag{33}$$

$$a_{sc} = \frac{P}{N_o W} (T_s+L)(W_s+B) \tag{34}$$

$$R_{N_{sc}} = \frac{1}{N_{sc}} \ln M, \tag{35}$$

which would allow (we expect) a decrease in error probability.

Of course, we have in a sense changed the problem because we are no longer coding over a block of (T, W), but over several blocks. While this should be a practical method of lessening the guard-space requirements, it is not applicable to the problem of using <u>one</u> (T, W) block for communication. We shall outline a "rate-expanding" scheme that does not have this drawback.

In one block, there will be N (given by (23)) basic signals whose outputs are independent and orthogonal, but N_{sc} basic signals whose outputs are orthogonal. Thus if we packed in N_{sc} signals with $T_g = L$ and $W_g = B$, there would be $N_{sc}/N \equiv a$ sets of

N basic signals, where

$$a = \frac{(T_s+L+1/B)(W_s+B+1/L)}{(T_s+L)(W_s+B)}. \tag{36}$$

The elements within any one of these sets would be orthogonal to and independent of each other, but the elements of different sets would, in general, only be orthogonal, not independent. But if N is large, we expect reliable communication when any set is transmitted, so we could consider sending information over all sets simultaneously. This would increase our data rate to

$$R_{re} = a\,R, \tag{37}$$

where the constraint length N is again given by (23), and now

$$\frac{1}{M}\sum_{n=1}^{N}\sum_{m=1}^{M} x_{mn}^2 \leq N\alpha_{re} = N\left(\frac{\alpha}{a}\right). \tag{38}$$

If P_e is the average error probability for one set (with constraint length N), and P_{re} is the probability of an error occurring in at least one of the a sets, then if P_e is small,

$$P_{re} \cong a\,P_e. \tag{39}$$

Since we expect P_e to be decreasing exponentially in N, the difference between P_{re} and P_e should be negligible.

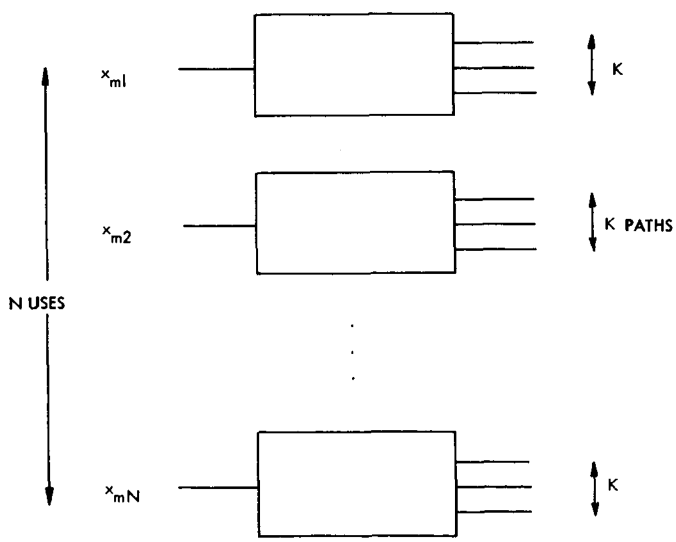

Fig. 4. Equivalent diversity channel.

The model just derived may be considered as N independent uses of a classical diversity channel, with K equal-strength paths, as illustrated in Fig. 4. When the m^{th} code word, consisting of $\underline{x}_m = (x_{m1}, x_{m2}, \ldots x_{mN})$ is transmitted as shown, each input component will excite K independent, identically distributed outputs, which may be summed to produce a random variable y, governed by the conditional density function of (31).

The description of the channel and its use just presented is, to be sure, an approximate one. This has been necessitated by the complex manner in which the channel operates on the input signals. We once again emphasize the roughness of the characterization of $\sigma(r,f)$ by just B and L. Although this may be reasonable when $\sigma(r,f)$ is smooth and concentrated as shown in Fig. 1, if $\sigma(r,f)$ is composed of several disjoint "pieces" (typical for HF channels, for example), such a gross simplification obviously omits much information about scattering function structure. We shall return to this point in Section IV.

On the other hand, this channel model is essentially a simple one (particularly when the equal-eigenvalue assumption is invoked), and for the first time, provides an explicit means of taking into account a bandwidth constraint on the input signals. We shall find that this simplified model will still provide some insight into the communication problem, and a useful means of estimating system performance.

III. BOUNDS TO ERROR PROBABILITY

We consider block coding over N uses of the amplitude-continuous, time-discrete channel model derived in Section II. We discuss both upper and lower bounds to P_e, the minimum attainable error probability, making use of many results presented in Appendices. Each bound is found to involve an arbitrary probability density function that must be chosen so as to obtain the tightest bound. The central problem considered here is the specification of the optimum density and the resulting bound. The expurgated random-coding upper bound to error probability is presented first, since it is functionally simplest and yields results that are typical of all of the bounds. Then the standard random-coding upper bound is discussed, along with the sphere-packing lower bound. These two bounds are found to agree exponentially in N for rates between R_{crit} and Capacity, and hence determine the true exponential behavior of P_e for that range of rates.

Some novel aspects of the present work should now be noted. The optimum density just discussed consists of a finite set of impulses. This corresponds to a finite number of input levels, an unexpected result for a continuous channel. The normal sphere-packing bound cannot be applied to a continuous channel because the number of possible inputs and outputs is unbounded. For this channel, however, the optimality of impulses allows derivation of the lower bound, although with extensive modification. The results presented in this section are in terms of quantities defined in Section II, plus an additional parameter (ρ or s) that simplifies the analysis. In Section IV, these parametric results will be converted into forms suitable for obtaining performance estimates for some practical systems. The reader who is mainly interested in the applications may skip to Section IV.

3.1 EXPURGATED BOUND

Our point of departure is the expurgated upper bound to error probability derived by Gallager.[10] That bound is directly applicable only to independent uses of a continuous channel whose input and output are both scalars. The generalization to a scalar input-vector output channel such as the one considered here, is straightforward (for an example, see Yudkin[24]), so the details will be omitted. When applied to this channel model, the bound states: If each code word is constrained to satisfy $\sum_{n=1}^{N} x_{mn}^2 \leq N\alpha$, then for any block length N, any number of code words $M = e^{NR_N}$, any $\rho \geq 1$, $r \geq 0$, and any probability density p(x) such that $\int_0^\infty x^2 p(x)\,dx = \alpha < \infty$, there exists a code for which

$$P_e < \exp - N\{\mathscr{E}_x[\rho, p(x), r] - \rho[R_N + \Delta_N]\} \tag{40}$$

$$\mathscr{E}_x[\rho, p(x), r] = -\rho \ln \left[\int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r(-2\alpha + x^2 + x_1^2)}\, H_{\underline{\lambda}}(x, x_1)^{1/\rho}\, dx\,dx_1\right] \tag{41}$$

$$H_{\underline{\lambda}}(x,x_1) = \int_0^\infty p_{\underline{\lambda}}(\underline{y}|x)^{1/2}\, p_{\underline{\lambda}}(\underline{y}|x_1)^{1/2}\, d\underline{y} = \prod_{k=1}^{\infty} \frac{\left(1+\lambda_k x^2\right)^{1/2} \left(1+\lambda_k x_1^2\right)^{1/2}}{1+\frac{1}{2}\lambda_k\left(x^2+x_1^2\right)} \tag{42}$$

$$p_{\underline{\lambda}}(\underline{y}|x) = \prod_{k=1}^{\infty} \left\{ \frac{1}{1+\lambda_k x^2}\, e^{-\frac{y_k}{1+\lambda_k x^2}} \right\} \tag{43}$$

where the quantity $\Delta_N \to 0$ as $N \to \infty$, provided $\int_0^\infty p(x)\, |x^2-a|^3\, dx < \infty$. All of the probability functions to be considered here will satisfy this constraint, so if N is large, the Δ_N term can be neglected. This parametric (on ρ) formulation of the bound is the usual one, and the reader unfamiliar with its properties should consult Gallager.[10] We recall that $\lambda_k \geq 0$, $\sum_{k=1}^{\infty} \lambda_k = 1$. The difficulty lies in determining the r, p(x) combination that results in the tightest bound to P_e for a given $\underline{\lambda}$, ρ, a, that is,

$$E_x(\rho, a, \underline{\lambda}) = -\rho \ln \left[\min_{r,\,p(x)} \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(-2a+x^2+x_1^2\right)} H_{\underline{\lambda}}(x,x_1)^{1/\rho}\, dx dx_1 \right], \tag{44}$$

subject to the constraints

$$r \geq 0, \quad p(x) \geq 0, \quad \int_0^\infty p(x)\, dx = 1, \quad \int_0^\infty x^2 p(x)\, dx = a \tag{45}$$

In general, it is possible for a number of local minima to exist, so that the problem of minimizing over p(x) and r is a difficult one. Fortunately, for the particular channel model under consideration here, $H_{\underline{\lambda}}(x,x_1)^{1/\rho}$ is a non-negative definite kernel (see Appendix A, Theorem A. 2, for a proof), and these possible difficulties do not arise. In Theorem A. 1 it is shown that a sufficient condition on r and p(x) to minimize

$$\int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(-2a+x^2+x_1^2\right)} H_{\underline{\lambda}}(x,x_1)^{1/\rho}\, dx dx_1,$$

subject to the constraints (45) is

$$\int_0^\infty p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_{\underline{\lambda}}(x,x_1)^{1/\rho}\, dx_1 \geq \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_{\underline{\lambda}}(x,x_1)^{1/\rho}\, dx dx_1 \tag{46}$$

for all $x \geq 0$, $0 < \rho < \infty$, with equality when $p(x) > 0$. At this point it is possible that (46) may have many solutions, or none but, because of the sufficiency, all solutions must result in the same (maximum) value of exponent for the given $\underline{\lambda}$, ρ, α. It may also be shown that condition (46) is necessary as well as sufficient, but the proof is lengthy and tedious. Since we shall later exhibit solutions to (46), we can afford to dispense with a proof of necessity.

Unfortunately, the determination of the p(x) and r that satisfies condition (46) is a very difficult matter. In practice, the usual method of "solution" is to take a particular r and p(x), and plug them into (46) to see if the inequality holds. If it does, then r, p(x) solve the minimization problem for the ρ, α, $\underline{\lambda}$ under consideration. We are now faced with the prospect of doing this for arbitrary $\underline{\lambda}$, for all values of ρ and α of interest.

Since the purpose of this study is to arrive at a better understanding of communication over fading channels, it is worth while to make use of any reasonable approximation that simplifies the analysis. Even if the most general problem could be solved, it seems likely that the essential nature of the channel would be buried in a maze of details, and one would have to resort to more easily evaluable special cases in order to gain insight into the basic factors involved.

3.11 Equal Eigenvalues

The simplest possibility to consider is that of equal eigenvalues,

$$\lambda_k = \frac{1}{K}, \quad k = 1, 2, \ldots, K. \tag{47}$$

In this case, define

$$E_{xe}(\rho, \alpha, K) \equiv -\rho \ln \left[\min_{r, p(x)} \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(-2\alpha + x^2 + x_1^2\right)} H_K(x, x_1)^{1/\rho}\, dx dx_1 \right] \tag{48}$$

$$H_K(x, x_1) = \left[\frac{\left(1 + \frac{x^2}{K}\right)^{1/2} \left(1 + \frac{x_1^2}{K}\right)^{1/2}}{1 + \frac{1}{2K}\left(x^2 + x_1^2\right)} \right]^K, \tag{49}$$

where r, p(x) satisfy constraints (45), and the subscript e denotes "equal eigenvalues." A simple change of variables proves that

$$E_{xe}(\rho, \alpha, K) = K\, E_{xe}\left[\frac{\rho}{K}, \frac{\alpha}{K}, 1\right]. \tag{50}$$

Thus, as far as the minimization is concerned, we may set K = 1, compute $E_{xe}(\rho, \alpha, 1)$, and use (50) to obtain the result for arbitrary K. The only change is that we must loosen the range $1 \leq \rho < \infty$ to $0 < \rho < \infty$ to allow for K > 1. This involves no additional effort, since our theorems on minimization are valid over the whole range $0 < \rho < \infty$. Thus, when the eigenvalues are equal, $\underline{\lambda}$ may be completely absorbed into the parameters ρ and α

for the purposes of minimization on p(x) and r, a pleasant bonus.

As we have mentioned, this simplification is a good reason for restricting the analysis to equal eigenvalues, but in addition, Theorem B. 3 in Appendix B states:

$$E_x(\rho, \alpha, \underline{\lambda}) \geq E_{xe}\left(\rho, \frac{b^2\alpha}{d}, \frac{b^3}{d^2}\right) = \alpha\left(\frac{b^2}{d}\right)\left\{\frac{E_{xe}\left(\frac{\rho d^2}{b^3}, \frac{\alpha d}{b}, 1\right)}{\frac{\alpha d}{b}}\right\}, \tag{51}$$

where

$$b \equiv \sum_{k=1}^{\infty} \lambda_k^2 \tag{52}$$

$$d \equiv \sum_{k=1}^{\infty} \lambda_k^3, \tag{53}$$

and inequality (51) is satisfied with equality for any equal eigenvalue system. Thus, an exponent that results from arbitrary $\underline{\lambda}$ may be lower-bounded in terms of an exponent with equal eigenvalues, thereby resulting in a further upper bound to P_e. This bound is very similar to one derived by Kennedy[16] for the infinite-bandwidth, orthogonal signal case. The b^2/d multiplier on the right-hand side of (51) may be interpreted as an efficiency factor, relating the performance of a system with arbitrary eigenvalues to that of an equal eigenvalue system with an energy-to-noise ratio per diversity path of $\alpha d/b$. The orthogonal signal analogy of this bound was found to be fairly good for many eigenvalue sets, thereby indicating that the equal eigenvalue assumption may not be as restrictive as it may at first appear. In any event, with equal eigenvalues we may set $K = 1$, and condition (46) may be simplified and stated as follows:

A sufficient condition for r, p(x) to minimize

$$\int_0^{\infty}\int_0^{\infty} p(x)\, p(x_1)\, e^{r\left(-2\alpha+x^2+x_1^2\right)} H_1(x, x_1)^{1/\rho}\, dx dx_1,$$

subject to constraints (45), is

$$\int_0^{\infty} p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_1(x, x_1)^{1/\rho}\, dx_1 \geq \int_0^{\infty}\int_0^{\infty} p(x)\, p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_1(x, x_1)^{1/\rho}\, dx dx_1 \tag{54}$$

for $0 < \rho < \infty$, with equality when $p(x) > 0$. In Appendix B, Theorem B. 1, it is shown that, if r, p(x) and r_1, $p_1(x)$ both satisfy (54), then $r = r_1$, and

$$\int_0^\infty [p(x)-p_1(x)]^2 \, dx = 0, \tag{55}$$

so that for all practical purposes, if a solution exists, it will be unique.

We shall digress and consider the zero-rate exponent, attained when $\rho = \infty$. The previous results are invalid at this point (although correct for any $\rho < \infty$), so this point must be considered separately. We choose to do it now because the optimization problem turns out to be easiest at $\rho = \infty$, and yet the results are indicative of those that will be obtained when $\rho < \infty$.

3.12 Zero-Rate Exponent

When the limit $\rho \to \infty$ is taken, it is easy to show that $r = 0$ is required for the optimization, and

$$E_x(\infty, a, \underline{\lambda}) = -\min_{p(x)} \int_0^\infty \int_0^\infty p(x)\, p(x_1) \ln H_{\underline{\lambda}}(x, x_1) \, dx dx_1 \tag{56}$$

$$E_x(\infty, a, \underline{\lambda}) = -\min_{p(x)} \sum_{k=1}^{\infty} \int_0^\infty \int_0^\infty p(x)\, p(x_1) \ln H_1(x\sqrt{\lambda_k}, x_1\sqrt{\lambda_k}) \, dx dx_1. \tag{57}$$

In the equal-eigenvalue case,

$$E_{xe}(\infty, a, 1) = -\min_{p(x)} \int_0^\infty \int_0^\infty p(x)\, p(x_1) \ln H_1(x, x_1) \, dx dx_1, \tag{58}$$

where now

$$E_{xe}(\infty, a, K) = K\, E_{xe}\left(\infty, \frac{a}{K}, 1\right). \tag{59}$$

We may change variables in (57) to obtain

$$E_x(\infty, a, \underline{\lambda}) = -\min_{p(x)} \sum_{k=1}^{\infty} \int_0^\infty \int_0^\infty q_k(x)\, q_k(x_1) \ln H_1(x, x_1) \, dx dx_1, \tag{60}$$

where $q_k(x) = \frac{1}{\sqrt{\lambda_k}}\, p\left(\frac{x}{\sqrt{\lambda_k}}\right)$, a probability density function with

$$\int_0^\infty x^2 q_k(x) \, dx = a\lambda_k. \tag{61}$$

The minimum in (60) may be decreased by allowing minimization over the individual $q_k(x)$, subject to (61), so that

$$E_x(\infty, \alpha, \underline{\lambda}) \leq - \sum_{k=1}^{\infty} \min_{q_k(x)} \int_0^{\infty} \int_0^{\infty} q_k(x)\, q_k(x_1) \ln H_1(x, x_1)\, dx dx_1 = \sum_{k=1}^{\infty} E_{xe}(\infty, \alpha\lambda_k, 1) \tag{62}$$

with equality when $\underline{\lambda}$ consists of equal eigenvalues. This, together with (51), shows that

$$\frac{b^3}{d^2} E_{xe}\left(\infty, \frac{\alpha d}{b}, 1\right) \leq E_x(\infty, \alpha, \underline{\lambda}) \leq \sum_{k=1}^{\infty} E_{xe}(\infty, \alpha\lambda_k, 1) \tag{63}$$

with equality on both sides when $\underline{\lambda}$ consists of equal eigenvalues. Thus at zero rate, we have upper and lower bounds to the expurgated bound exponent for an arbitrary eigenvalue channel in terms of the exponent for a channel with one eigenvalue. Of course, at this point, before evaluation of $E_{xe}(\infty, \alpha, 1)$, we do not know how tight these bounds are.

The derivation of the conditions on the p(x) that optimizes (56) is complicated by the fact that $\ln H_{\underline{\lambda}}(x, x_1)$ is not non-negative definite. The derivation, however, only requires that $\ln H_{\underline{\lambda}}(x, x_1)$ be non-negative definite with respect to all functions f(x) that can be represented as the difference of two equal-energy probability functions. In Theorem A. 2 it is proved that this is indeed the case. Utilizing this, the same theorem states a condition on p(x) sufficient for the maximization of $E_x(\infty, \alpha, \underline{\lambda})$. When simplified for one eigenvalue, this becomes a sufficient condition for p(x) to optimize $E_{xe}(\infty, \alpha, 1)$, subject to constraints (45):

$$\int_0^{\infty} p(x_1) \ln H_1(x, x_1)\, dx_1 \geq \int_0^{\infty} \int_0^{\infty} p(x)\, p(x_1) \ln H_1(x, x_1)\, dx dx_1 + \lambda_o(x^2 - \alpha) \tag{64}$$

for some λ_o, with equality when $p(x) > 0$. We dispense with the question of necessity, since we shall now exhibit a p(x) that satisfies the sufficient condition and thus maximizes $E_{xe}(\infty, \alpha, 1)$.

In Theorem B. 2, it is shown that the probability function

$$p(x) = p_1 u_o(x) + p_2 u_o(x - x_o) \tag{65}$$

satisfies condition (64) for all α when the parameters p_1, p_2, and x_o are correctly chosen. Therefore, at zero rate, the optimum p(x) consists of two impulses, one of which is at the origin. In Appendix B, expressions are presented relating p_1, p_2, and α as functions of x_o^2. This was done because it was simpler to express the results in terms of x_o^2 instead of α. Here we return to the more natural formulation, and in Figs. 5 and 6 we present graphically the optimum distribution in terms of α. The resulting exponent is presented in Figs. 7 and 8. Note that $\frac{1}{\alpha} E_{xe}(\infty, \alpha, 1)$ is a decreasing function of α, and so

$$\frac{1}{\alpha} E_{xe}(\infty, \alpha, 1) \leq \lim_{\alpha \to 0} \frac{1}{\alpha} E_{xe}(\infty, \alpha, 1) \equiv E_{\infty} \cong 0.15. \tag{66}$$

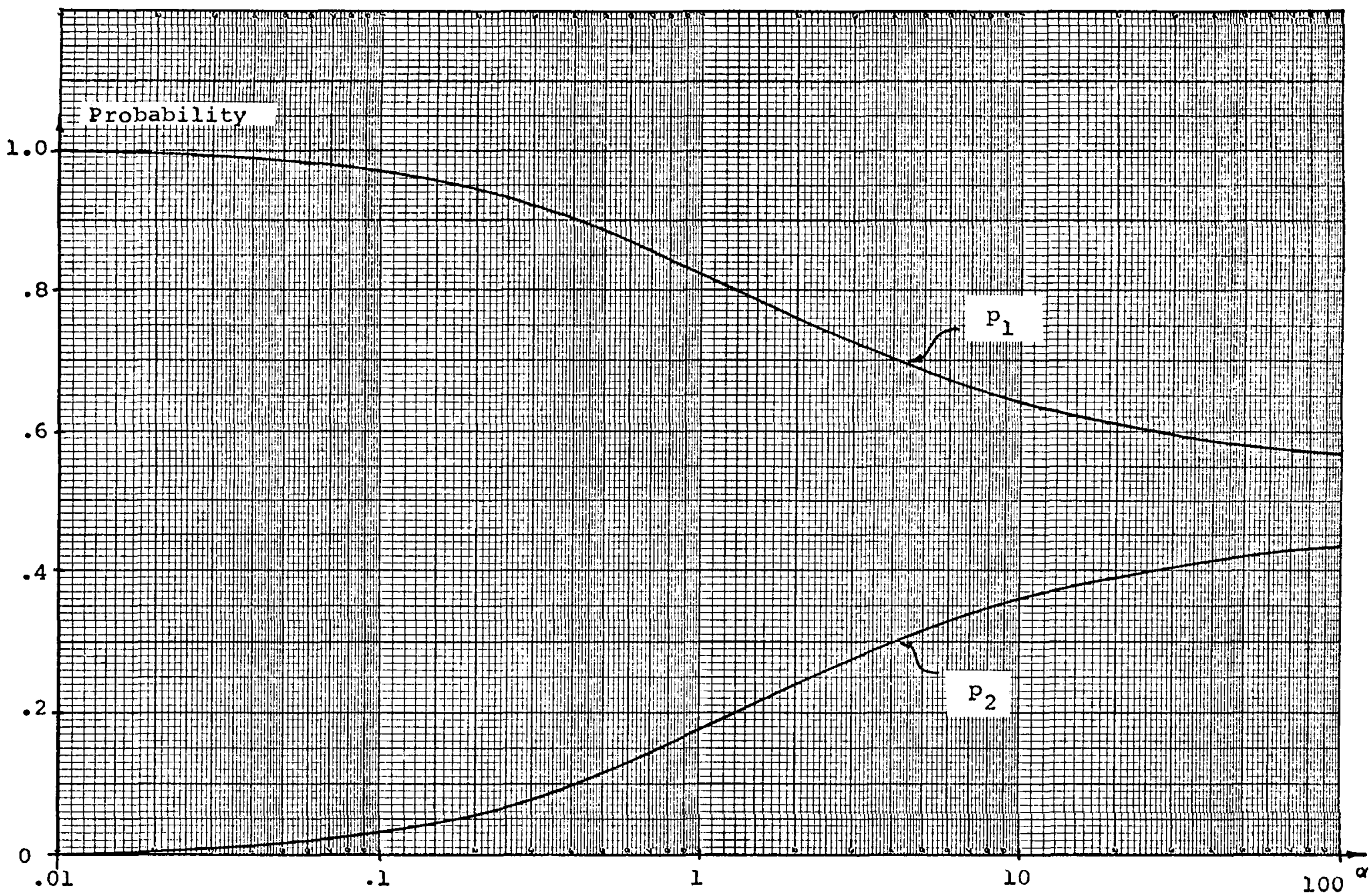

Fig. 5. p_1 and p_2 vs α.

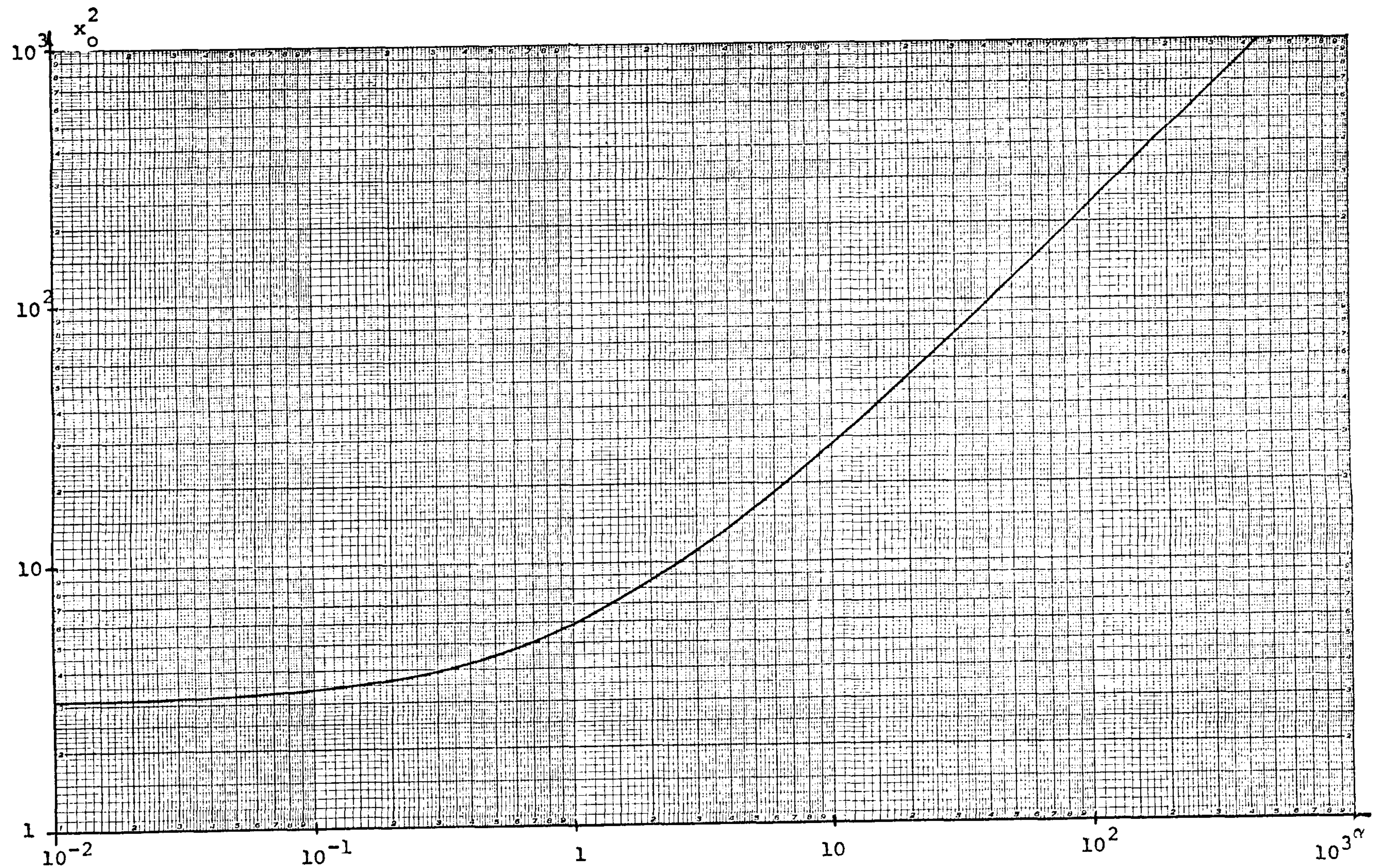

Fig. 6. x_o^2 vs α.

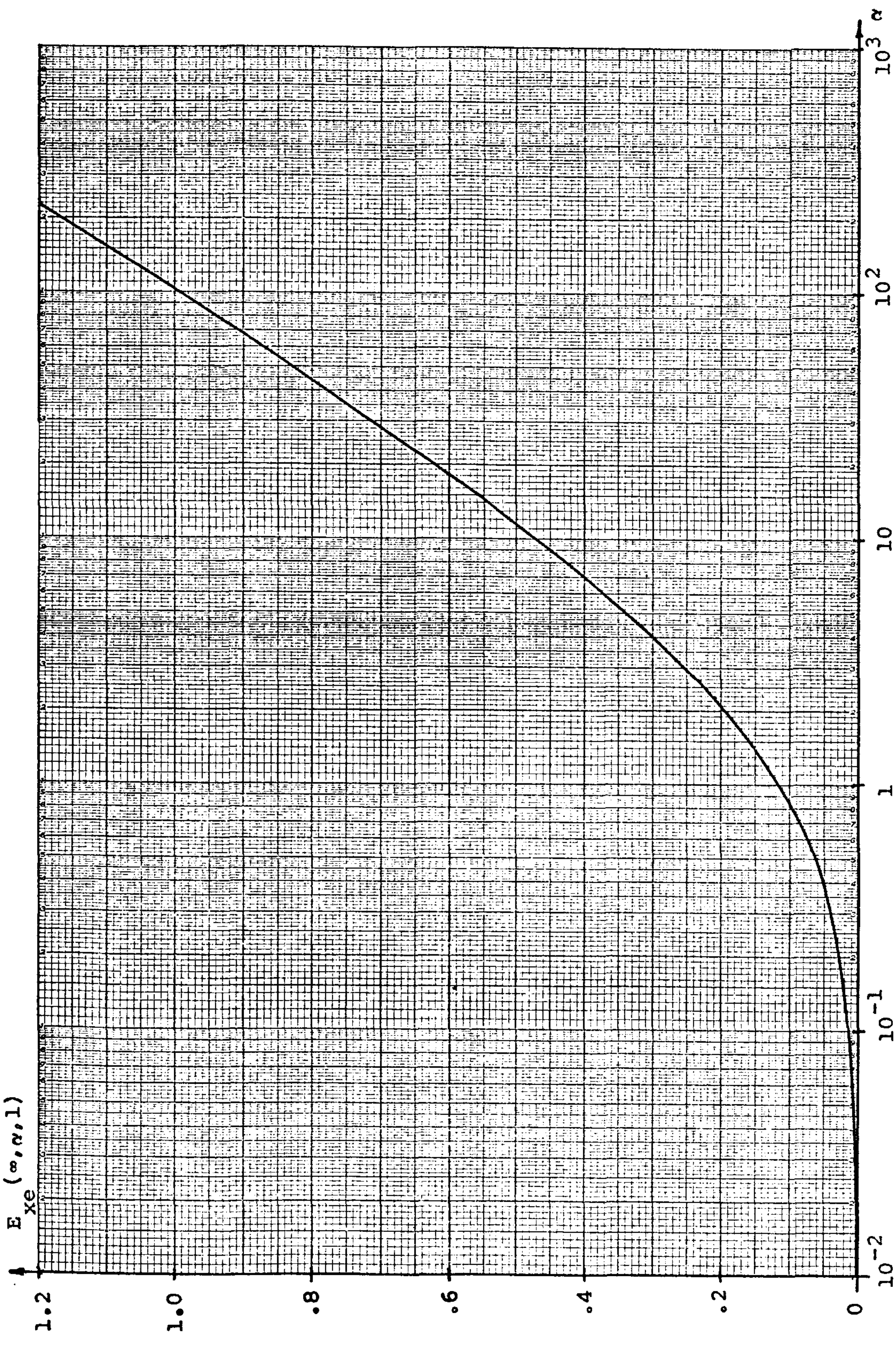

Fig. 7. $E_{xe}(\infty, \alpha, 1)$ vs α.

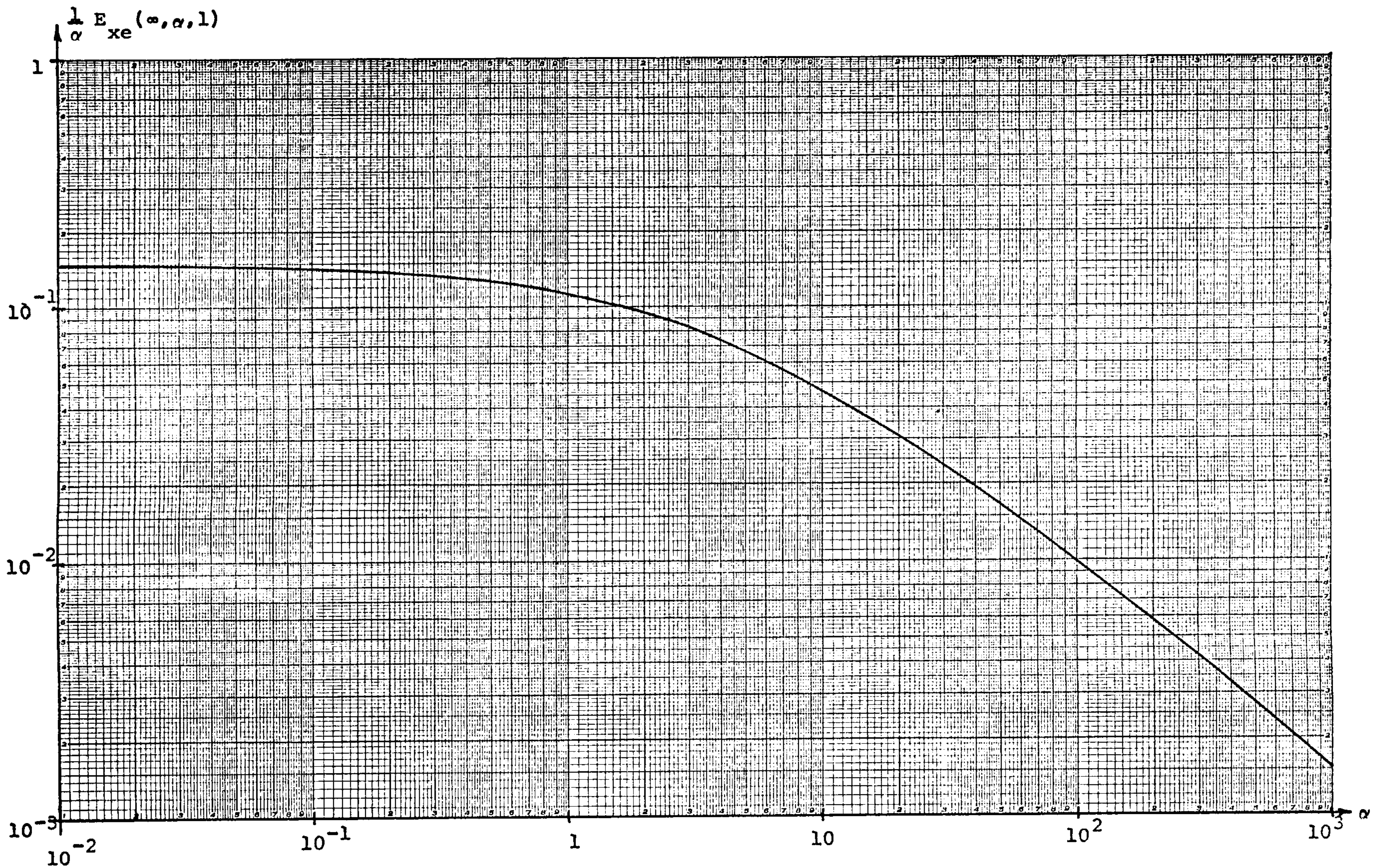

Fig. 8. $\frac{1}{a} E_{xe}(\infty, a, 1)$ vs a.

This is the same zero-rate result that Kennedy found for this channel when an infinite bandwidth was available and orthogonal signals were used. Since $\alpha \to 0$ is the equivalent of $W \to \infty$, this is the expected result. In addition, we find that, as $\alpha \to \infty$, $E_{xe}(\infty, \alpha, 1) \to \frac{1}{4} \ln \alpha$.

Returning to the bound of (63), we see that

$$\left(\frac{b^2}{d}\right)\left\{\frac{E_{xe}\left(\infty, \frac{\alpha d}{b}, 1\right)}{\frac{\alpha d}{b}}\right\} \leq \frac{E_x(\infty, \alpha, \underline{\lambda})}{\alpha} \leq \sum_{k=1}^{\infty} \lambda_k \left\{\frac{E_{xe}(\infty, \alpha\lambda_k, 1)}{\alpha\lambda_k}\right\} \leq E_\infty, \tag{67}$$

where the last inequality may be approached by a channel with K equal eigenvalues, as $K \to \infty$. We shall defer discussion and interpretation of these results, except to note that, if we consider coding over N channel uses (at zero rate), and were free to choose K independently of anything else, then K should go to infinity, and the infinite bandwidth exponent would result.

Consider the two inequalities on either side of $\frac{1}{\alpha} E_x(\infty, \alpha, \underline{\lambda})$ in (67). When $\underline{\lambda}$ consists of any number of equal eigenvalues, the inequalities will be satisfied with equality, and are thus as tight as possible. As a second example, let $\underline{\lambda} = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$ and $\alpha = 1$. Numerical evaluation of the bounds shows that $0.119 \leq E_x(\infty, \alpha, \underline{\lambda}) \leq 0.133$, and the bounds are fairly tight. We expect, however, that any system with a small number of nonzero eigenvalues should be well approximated by some equal-strength eigenvalue system, so a more severe test of our bounds should be with a system with an infinite number of positive λ_k. A channel with a Gaussian-shaped scattering function, when excited with a Gaussian modulation, can be shown to have the eigenvalues $\lambda_k = (1-c)\, c^{k-1}$. Let $c = \frac{1}{2}$ and $\alpha = 1$ for the sake of another numerical example. The lower bound is easy to evaluate, and the upper bound involves an infinite sum, which can in turn be bounded by computing the first $K - 1$ significant terms and noting that all of the rest contribute, at most, $E_\infty \sum_{k=K}^{\infty} \lambda_k$. This results in $0.102 \leq E_x(\infty, \alpha, \underline{\lambda}) \leq 0.135$, not as tight as we would like, but still a reasonable set of bounds.

3.13 Positive Rates

We now return to the more general problem of optimization of r and p(x) for values of ρ in the range $0 < \rho < \infty$. Once again, we restrict $K = 1$, since other values may be obtained by suitable trade-offs between ρ and α. Recall that condition (54) is sufficient for maximization of the exponent when $K = 1$. In Appendix D, it is shown that, for any $0 < \rho < \infty$ and $\alpha > 0$, condition (54) must be satisfied by some r, p(x) combination, where $0 \leq r \leq \frac{1}{2\rho}$ and p(x) consists of a finite number of impulses. To be precise, $p(x) = \sum_{n=1}^{N_1} p_n u_o(x-x_n)$, where N_1 is a finite integer, and $0 \leq x_n^2 \leq z_\rho$, where z_ρ is a function only of ρ and α, and is finite for $0 < \rho < \infty$ and $\alpha > 0$.

Even when armed with the knowledge that a finite set of impulses provides an optimum, it is still a very difficult problem to actually solve for the optimizing p(x) and r. There are some special cases, however, for which simplifications can be made. The first is in the limit as α goes to zero.

a. Small α

In this case, it can be shown (Theorem B. 4, Appendix B) that a two-impulse p(x), when combined with a suitable value or r, will asymptotically satisfy the sufficient condition (54) in the limit as α goes to zero. To be specific, the optimizing combination is

$$p(x) = \left(1 - \frac{\alpha}{z_o}\right) u_o(x) + \frac{\alpha}{z_o} u_o(x - \sqrt{z_o}) \tag{68}$$

$$r = -\frac{1}{\rho z_o} \ln H_1(0, \sqrt{z_o}) \tag{69}$$

where $z_o = 3.071$. The resulting exponent is the same as the low-rate infinite-bandwidth exponent, that is,

$$E_{xe}(\rho, \alpha, 1) \cong \alpha f_1(z_o) = \alpha E_\infty \tag{70}$$

$$f_1(z) \equiv \frac{2}{z}\left[\ln\left(1 + \frac{1}{2}z\right) - \frac{1}{2}\ln(1+z)\right], \tag{71}$$

where z_o is the value of z that maximizes $f_1(z)$. Thus as $\alpha \to 0$, the same exponent is obtained for all values of ρ, thereby confirming the known[16] result that when an infinite bandwidth is available, expurgation does not improve the bound.

When ρ is specified, it is also possible to show that for some small, but nonzero, α, a two-impulse p(x) will exactly satisfy condition (54), so that a p(x) consisting of two impulses is more than just asymptotically optimum. The proof amounts to specifying values for ρ and α, under the assumption of a two-impulse p(x), solving for the optimum probabilities, positions, and r, and then numerically verifying that the resulting p(x) does satisfy (54). Because of the computational difficulties, no general proof for arbitrary ρ and α has been found, but some specific examples have been verified. These same computational problems have made it impossible to analytically specify the best r, p(x) combination for given values of ρ and α, so that we are forced to consider a numerical solution to the optimization problem.

b. Numerical Solution

Let us suppose that a value of r in the range $0 \leq r \leq \frac{1}{2\rho}$ and a set of impulse positions $\{x_n\}$, n = 1, 2, ..., N_1, have been specified. We are then left with the discrete problem:

$$\min_{\{p_n\}} \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} p_i p_j \, e^{-r\left(2a-x_i^2-x_j^2\right)} H_1(x_i, x_j)^{1/\rho}, \tag{72}$$

subject to the conditions

$$p_n \geq 0, \quad \sum_{n=1}^{N_1} p_n = 1, \quad \sum_{n=1}^{N_1} p_n x_n^2 = a. \tag{73}$$

The kernel $H_1(x, x_1)^{1/\rho}$ is non-negative definite, so that the matrix whose ij^{th} entry is $e^{r\left(-2a+x_i^2+x_j^2\right)} H_1(x_i, x_j)^{1/\rho}$ is also non-negative definite, and thus the quadratic form (72) is a convex function of $\{p_n\}$. This type of minimization problem is known as a quadratic programming problem, and with the convexity property just mentioned, numerical techniques for solution are well known[25, 26] and are conceptually straightforward.

Unfortunately, if a joint minimization over $\{p_n\}$, $\{x_n\}$, r is attempted, difficulties are encountered because the function to be minimized is not a convex function of $\{x_n\}$. This means that several local minima could exist, and hence the over-all minimization is a much harder problem. Since the problem of minimizing over $\{p_n\}$, given $\{x_n\}$ and r, is readily adapted for numerical solution on a computer, however, the obvious approach is to specify a grid of $\{x_n\}$, and minimize over $\{p_n\}$ for a range of values of r. As remarked earlier, we need only consider $0 \leq r \leq \frac{1}{2\rho}$ and $0 \leq x_n^2 \leq z_\rho$, so that all variables lie within a bounded region.

Any permutation of the x_n will result in the same minimum (this is just a renumbering) and making two x_n the same is redundant (two impulses at the same position may be condensed to one). Thus the x_n may be spaced out as distinct points on the real line between 0 and $\sqrt{z_\rho}$, and for each selection of $\{x_n\}$ the minimization should be performed for a range of r between 0 and $1/2\rho$. As the grid spacing becomes very small (involving a large number of possible impulses), we expect to get very close to the true minimizing p(x) and r. If the minimizing p(x) were a continuous function of x, we would expect that, as the grid spacing was reduced, more and more impulses should be used with smaller and smaller probabilities, to approximate the continuous solution. Since it is known that the optimum p(x) consists of impulses, however, the smaller grid spacing should be used merely to find better locations for the impulses that are used, and we would expect that the optimizing $\{p_n\}$ for a given grid spacing should contain many zero probabilities. This was confirmed when the computations were performed.

In order to carry out the minimizations just described, the author modified a multipurpose FORTRAN minimization program obtained elsewhere,[27] and ran the problem on the M. I. T. Computation Center's IBM 7094 digital computer. The procedure used to approximate the minimizing p(x) and r was not quite as described above, for the following reasons: (i) The program could not accept a large number of possible x_n as inputs, and (ii) great savings in running time could be obtained if a good guess as to the optimizing $\{p_n\}$ for a given r, $\{x_n\}$ could be specified in advance. The first makes it impossible to feed in an extremely fine grid of $\{x_n\}$. This is no real restriction, since, as has been pointed out already, only a small number of x_n will have nonzero p_n after the minimization is performed, so that once the approximate locations of the best impulses are known, the grid spacing need only be reduced in the vicinity of these locations. Thus each problem was run twice, once with a coarse grid to get an approximate solution, and again with a finer grid to get a more accurate answer.

The only disadvantage to this method is that, theoretically, if the approximating grid is too coarse, we might wind up in the neighborhood of a local, but not universal, minimum. For a given value of ρ, however, the minimization was performed with α as a variable, starting near $\alpha = 0$ (where the solution is known) and gradually increasing α. As might be expected, for the same ρ, when values of α are not too far apart, the optimizing r, $\{p_n\}$ and $\{x_n\}$ were not very different. Thus the minimizing p(x) and r were "tracked," always starting with a previously known minimum. Moreover, there is always the option of plugging any p(x) and r into (54) and confirming whether or not it represents a minimum. This amounts to computing many points of a fairly complicated function, and thus was done sparingly. One other point to be noted is that the exponent is relatively insensitive to changes in $\{p_n\}$, $\{x_n\}$ and r in the vicinity of the optimum, so even if the p(x) and r are only approximately correct, the exponent will still be reasonably accurate.

c. Results

We shall now present and discuss some of the numerical results obtained from the procedure just mentioned. As a first example, the minimizing p(x) and r are presented graphically as functions of α for the case K = 1, ρ = 1, in Figs. 9, 10, and 11. With regard to Fig. 9, $x_1 = 0$ for all α (an impulse at the origin) and hence does not show up on the logarithmic scale. The other x_n are numbered in the order in which they appear, as α is increased from zero. Drawing a vertical line at any value of α allows one to read off the minimizing impulse positions for that α. Figure 10 presents the probabilities corresponding to the positions shown in Fig. 9, numbered accordingly. Figure 11 shows the optimizing value of r for K = 1 and several different values of ρ, including $\rho = 1$, as a function of α.

From a study of the solutions to condition (54) of which the previous figures represent a typical set, the following general properties have become apparent.

1. The optimum p(x) always contains an impulse at the origin, and this impulse has the largest probability associated with it. It is reasonable to expect that some probability would be concentrated at the origin because the variance of the output y is smallest when the input, x, is zero. Therefore this input results, in some sense, in a less spread output distribution than any other, and should be good for information communication. Moreover, the energy constraint tends to force probability at

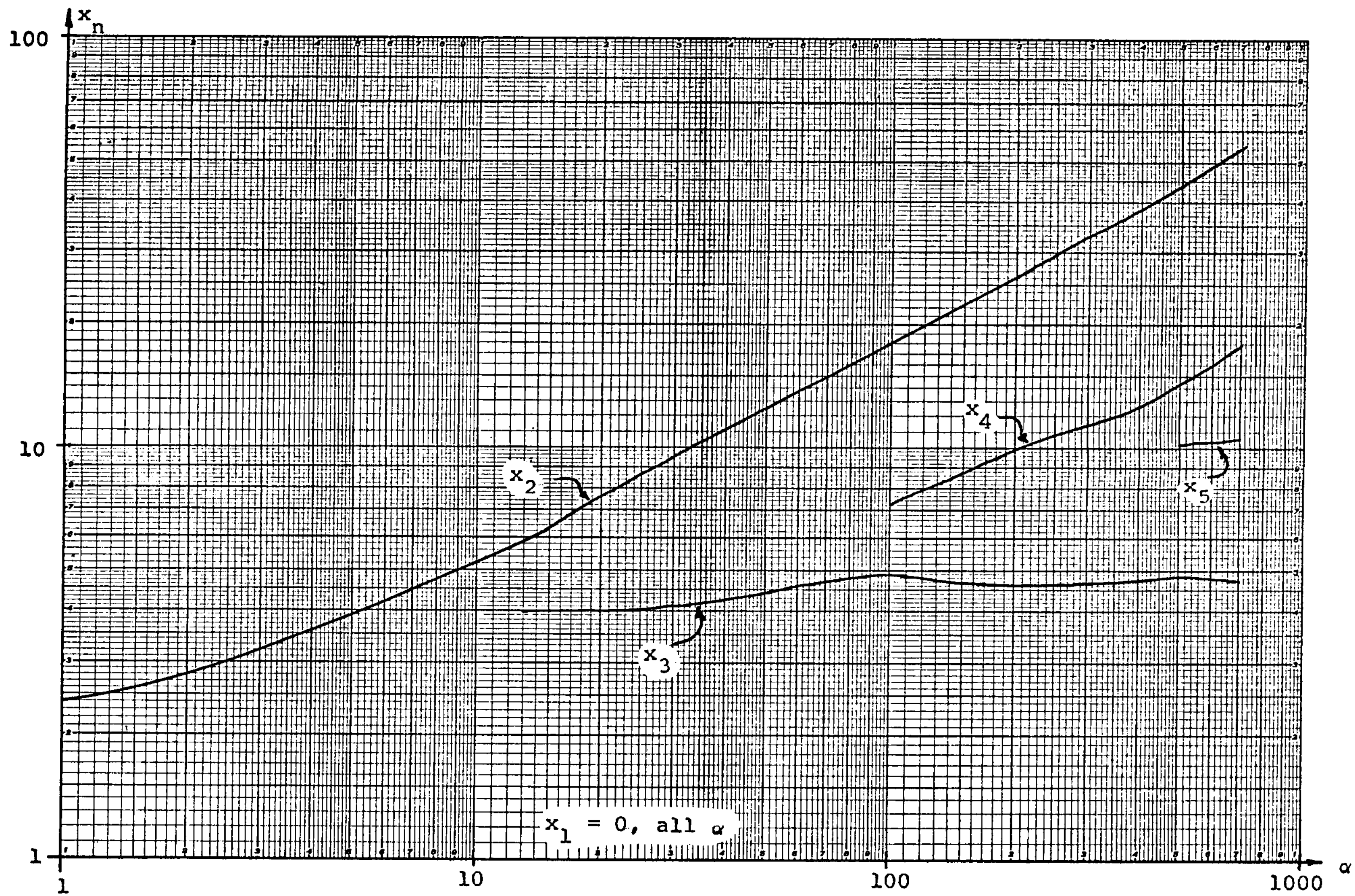

Fig. 9. Optimizing $\{x_n\}$ vs α, $K = \rho = 1$.

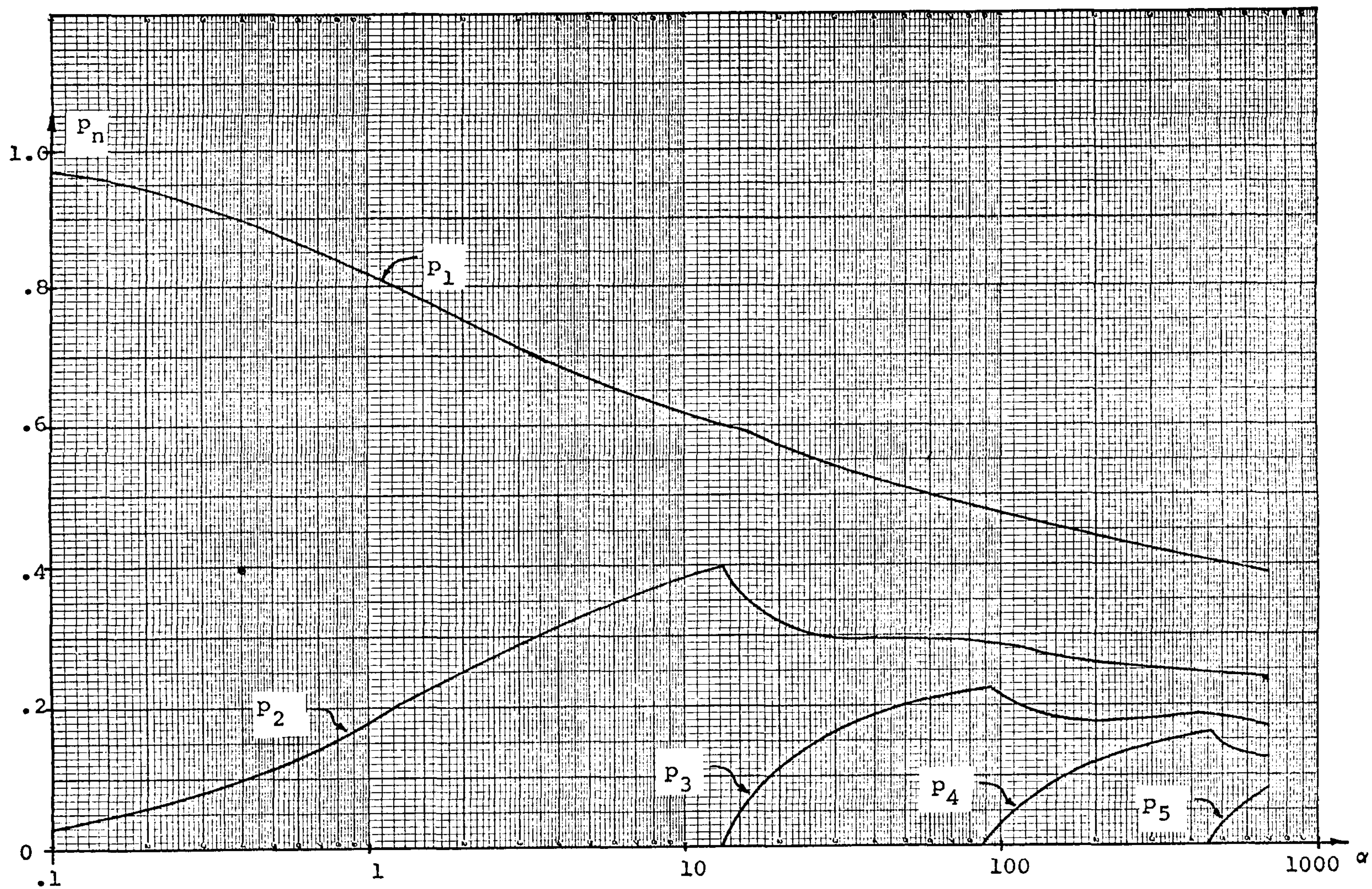

Fig. 10. Optimum $\{p_n\}$ vs α, K = ρ = 1.

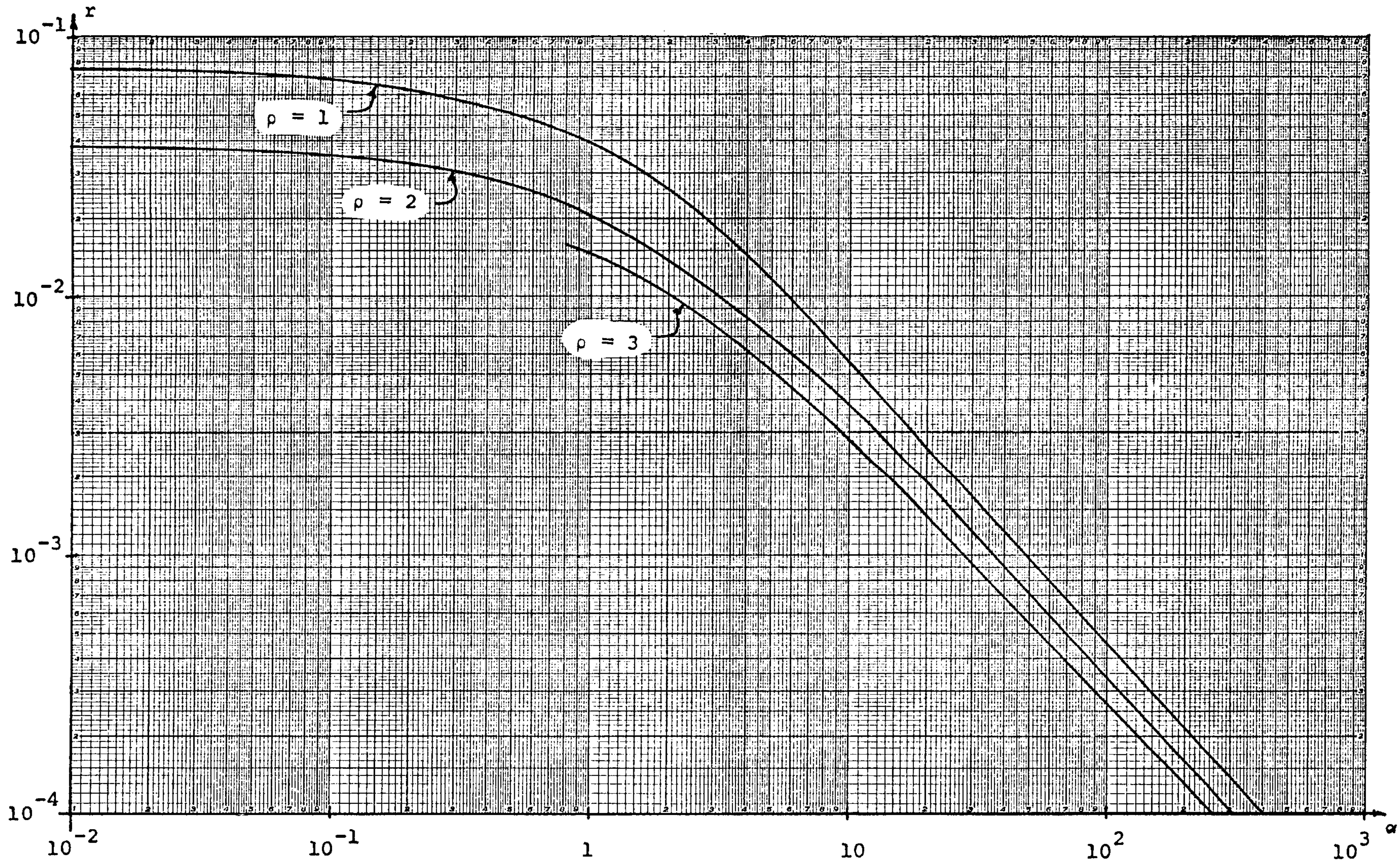

Fig. 11. Optimum r vs α, K = 1.

small values of x.

2. For any given K and ρ, the solution starts with two impulses when α is near zero. As α is increased, the impulse not at the origin gets relocated to larger values of x, while its probability decreases. Eventually, a third impulse is required, appearing between the other two, then a fourth, and so forth (see property 5 below). Thus, if N_1 is the number of impulses with nonzero probabilities used in the optimum p(x) for a given ρ and K, then N_1 is a nondecreasing function of α, and increases in steps of one.

3. As α is increased for given values of K and ρ, r decreases monotonically, and as α goes to infinity, r goes to an apparent asymptote of $r = a/\alpha^b$, where a and b are constants.

4. If α is held constant and ρ is varied, N_1 is a nonincreasing function of ρ, and decreases in steps of one.

5. As α is increased for a given ρ, it is interesting to note what happens in the neighborhood of a value of α where an extra impulse is added. Sufficient condition (54) may be rewritten for impulses as

$$F(z) \equiv \sum_{n=1}^{N_1} p_n e^{r(z+z_n)} \left[\frac{(1+z)^{1/2}(1+z_n)^{1/2}}{1+\frac{1}{2}(z+z_n)}\right]^{1/\rho}$$

$$- \sum_{n=1}^{N_1} \sum_{m=1}^{N_1} p_n p_m e^{r(z_n+z_m)} \left[\frac{(1+z_n)^{1/2}(1+z_m)^{1/2}}{1+\frac{1}{2}(z_n+z_m)}\right]^{1/\rho} \geq 0 \qquad (74)$$

where, for convenience, we have set $z = x^2$, $z_n = x_n^2$, and if (74) is satisfied for all $z \geq 0$, then $\{p_n\}$, r, $\{z_n\}$ comprise the optimizing set. For a typical value of α for which two impulses is optimum, F(z) will appear as sketched in Fig. 12, where $z_1 = 0$ and z_2 are

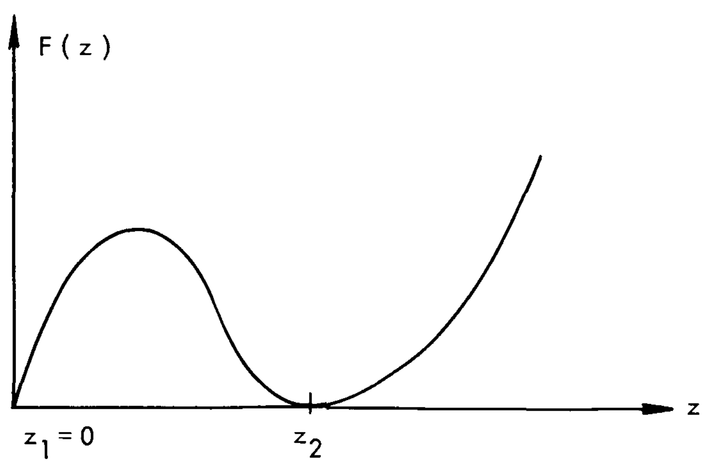

Fig. 12. Sketch of F(x) vs z when two impulses are optimum.

the impulse positions. As α is increased almost to the point where 3 impulses are needed, F(z) will begin to appear as sketched in Fig. 13a. Figure 13b represents F(z) at the breakpoint, where at this value of α, p(x) still consists of two impulses

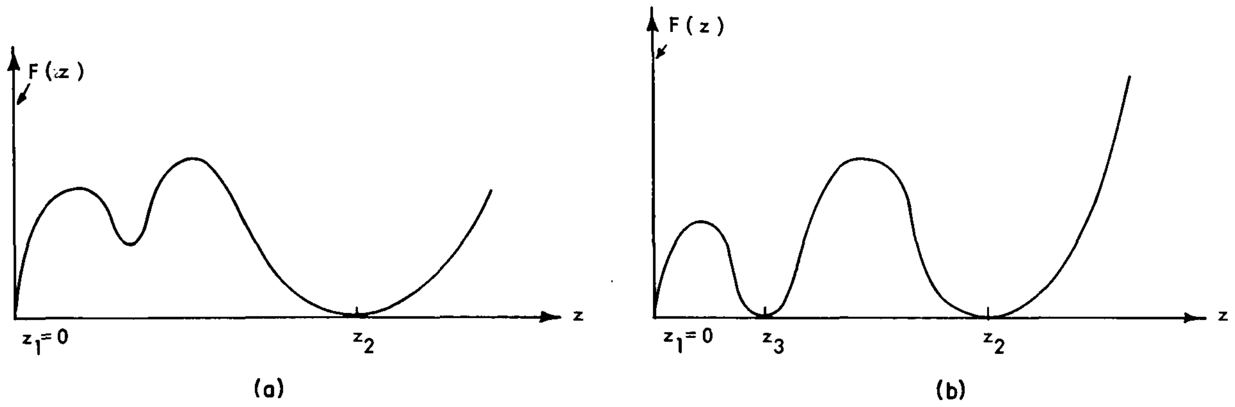

Fig. 13. (a) F(z) vs z near breakpoint.
(b) F(z) vs z at breakpoint.

(at $\sqrt{z_1}$ and $\sqrt{z_2}$). As α is increased slightly, the new optimum p(x) will contain a third impulse at $\sqrt{z_3}$, with very small probability. Of course, the positions of z_2 and z_3 will be continuously changing as α is increased. When a new impulse is needed it is evident, from Fig. 10, that most of its probability comes at the expense of the impulse that was added the time before, since that probability undergoes a sharp dip.

Consider the sets of (ρ, α) for which the optimum p(x) consists of 2 impulses. One impulse must always be at the origin, and in Fig. 14, we indicate roughly how the position of the second impulse varies with ρ and α. The line going from lower left to upper right indicates the line of breakpoints, and the region below this line is the range of ρ and α, where 2 impulses are no longer optimum. The other lines represent constant $z_2 = x_2^2$ in the region where a two-impulse p(x) is optimum.

Some exponents resulting from the optimal sets of p(x) and r are presented in Figs. 15 and 16. The first of these shows $E_{xe}(\rho, \alpha, 1)$ as a function of α for several values of ρ, while the second shows $\frac{1}{\alpha} E_{xe}(\rho, \alpha, 1)$ versus α.

Recall that, when $\rho = \infty$ and α and N were fixed, there was a monotonic improvement in the exponent as K was increased, and as K went to infinity for any value of α, the infinite bandwidth exponent was approached. This is no longer the case for $\rho < \infty$. In Fig. 17, we show $E_{xe}(\rho, 1, K)$ as a function of K for several values of ρ. When K is increased beyond a certain point the exponent starts decreasing again; this corresponds to the fact that the signal energy is being split among too many diversity paths. We shall return to this point in Section IV.

We have previously found a lower bound to $E_x(\rho, \alpha, \underline{\lambda})$ in terms of an equal eigenvalue system, as well as an upper bound when $\rho = \infty$. Unfortunately, it has not been possible to find an equally tight upper bound for $0 < \rho < \infty$. About the best

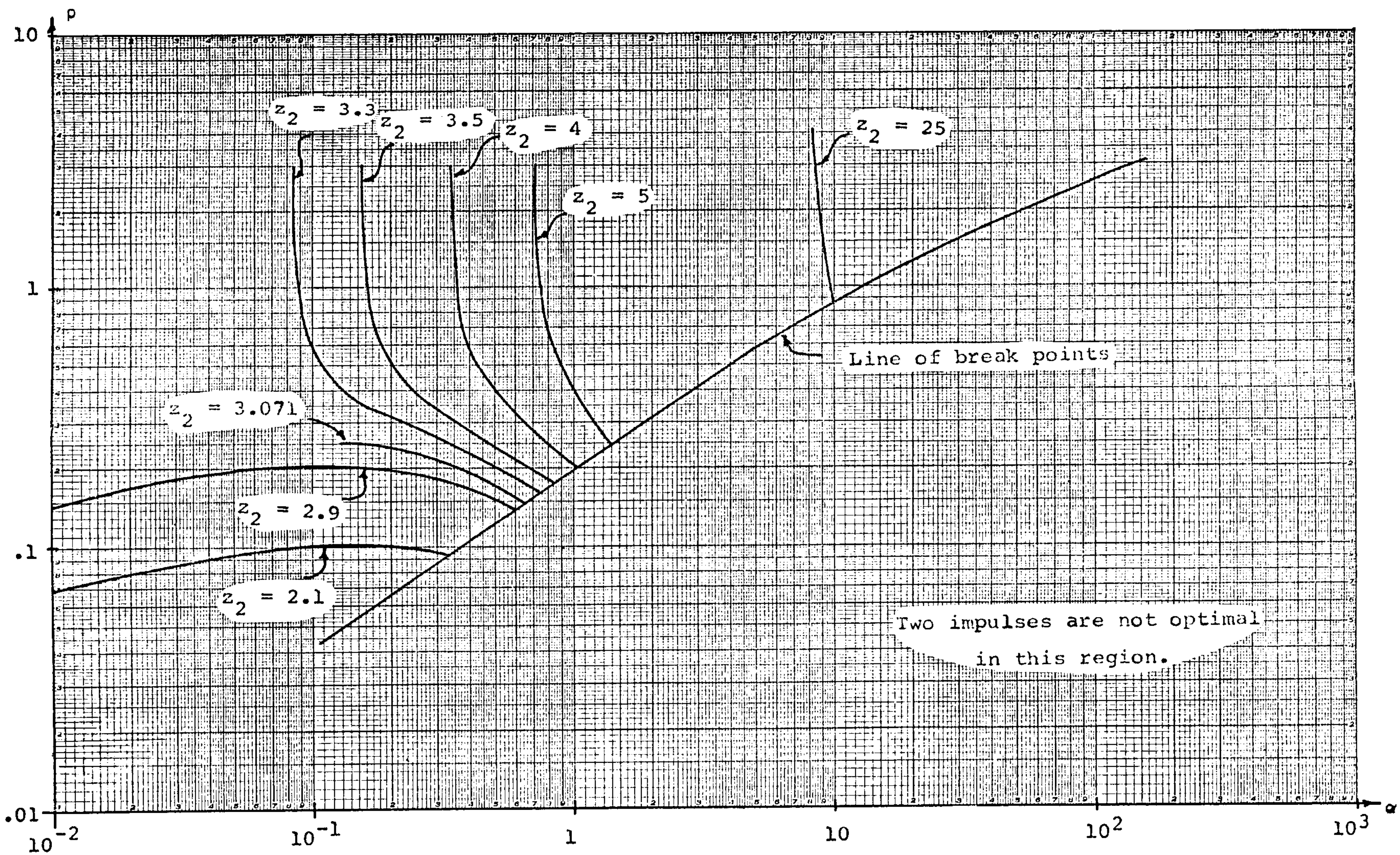

Fig. 14. Position of second impulse vs ρ and α, K = 1.

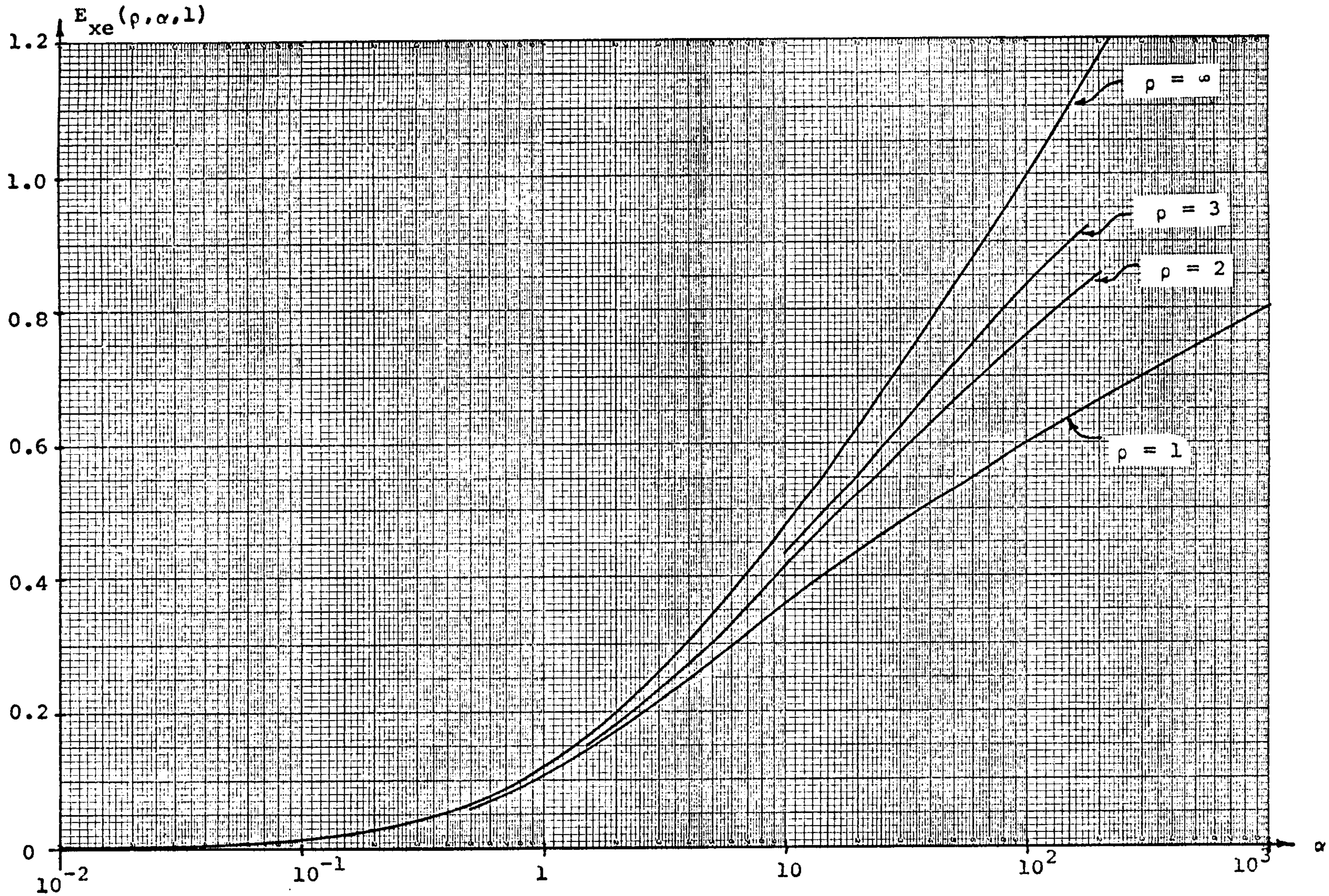

Fig. 15. $E_{xe}(\rho, \alpha, 1)$ vs α.

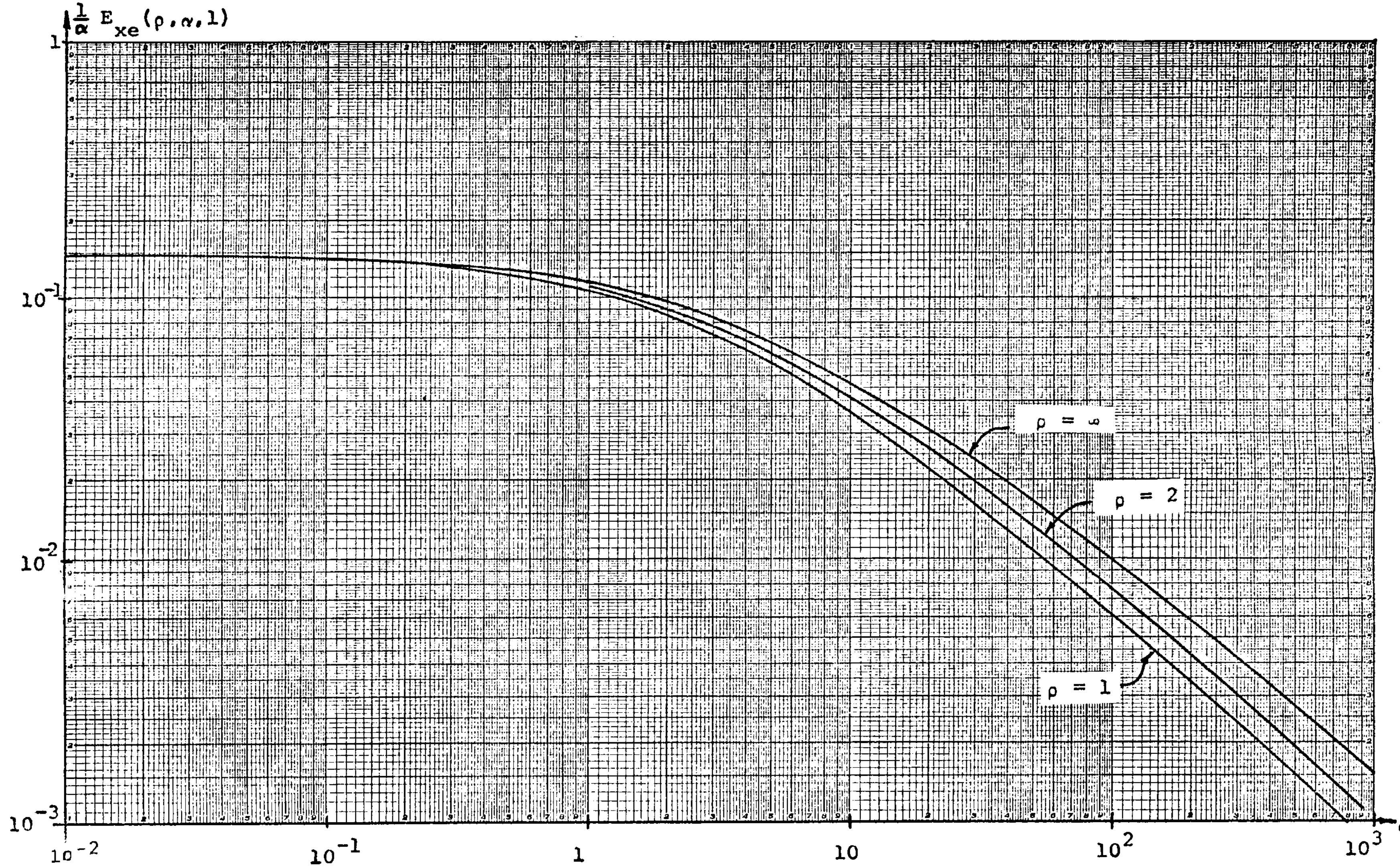

Fig. 16. $\frac{1}{\alpha} E_{xe}(\rho, \alpha, 1)$ vs α.

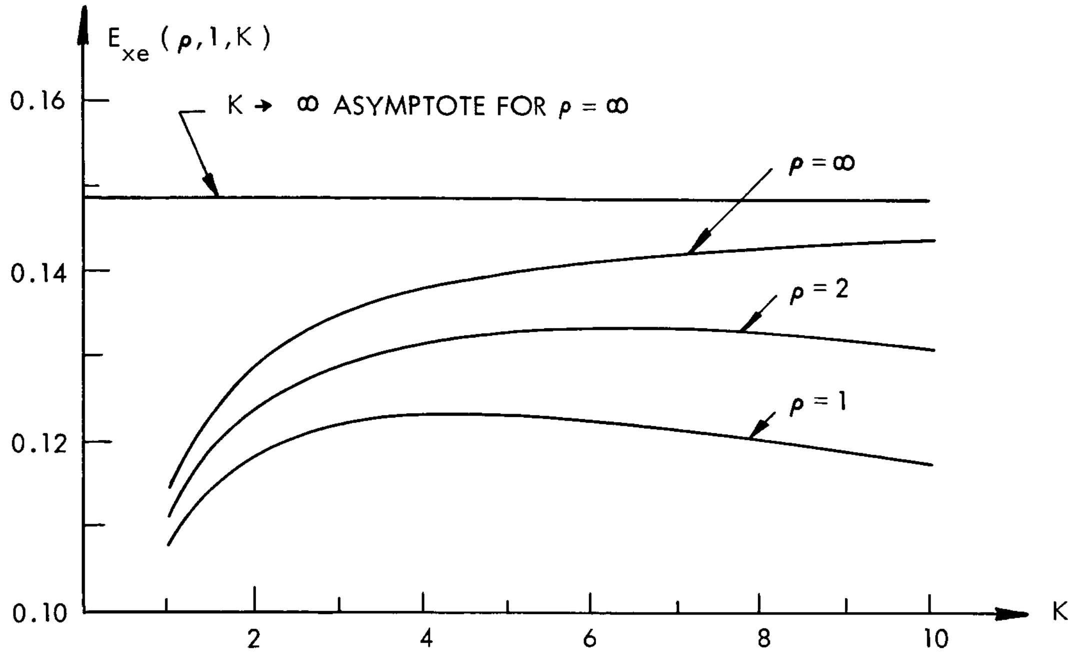

Fig. 17. $E_{xe}(\rho, 1, K)$ vs K.

we can do is to say that

$$E_{xe}(\rho, \alpha, \underline{\lambda}) \leq \alpha E_{\infty}. \tag{75}$$

This is certainly true because $E_{xe}(\rho, \alpha, \underline{\lambda})$ is an increasing function of ρ, but it can be quite loose because, even at $\rho = \infty$, the right-hand side can be approached only by letting $\alpha \to 0$ or $K \to \infty$, and for $\rho < \infty$, the only way it can be approached is through $\alpha \to 0$.

d. Large α

With the exception of the point $\rho = \infty$, our results are mainly numerical, so it is difficult to say what the behavior of the exponent will be for very large α. From the previous results, it appears that as $\alpha \to \infty$, the best p(x) will be one with many impulses spaced fairly widely apart. The proof of impulse optimality is valid for any finite α, but does not give us much information as $\alpha \to \infty$ because the ceiling on impulse positions, $\sqrt{z_\rho}$, goes to infinity with α. We can, however, obtain a lower bound to the exponent by evaluating any particular p(x) and r combination. For this purpose, let

$$p(x) = ax^b(1+x^2)^{-1/2\rho}\, e^{-cx^2}, \tag{76}$$

where a and c are chosen so that p(x) integrates to unity and $x^2p(x)$ integrates to α, and b is a free parameter that may be optimized. The factor of $(1+x^2)^{-1/2\rho}$ cancels with a similar factor in $H_1(x, x_1)^{1/\rho}$, and allows us to integrate the expression for the exponent. It may be shown that, as a goes to infinity, the resulting exponent is maximized by choosing $b = -1 + 1/\rho$, and $r = 0$. With these values we obtain a lower bound to $E_{xe}(\rho, a, 1)$ of the form

$$E_{xe}(\rho, a, 1) \geq +\rho \ln \left[\frac{U^2\left(\frac{1}{2\rho}, 1, c\right)}{2U\left(\frac{1}{\rho}, 1, c\right)} \right], \tag{77}$$

where $U(x, y, z)$ is the confluent hypergeometric function defined by

$$U(x, y, z) \equiv \frac{1}{\Gamma(x)} \int_0^{\infty} e^{-zt}\, t^{x-1} (1+t)^{y-x-1}\, dt, \tag{78}$$

and c is chosen to satisfy

$$a = \frac{1}{2\rho} \frac{U\left[1 + \frac{1}{2\rho}, 2, c\right]}{U\left[\frac{1}{2\rho}, 1, c\right]} \tag{79}$$

The U's are reasonably well-known and tabulated functions,[28] and it turns out that as $a \to \infty$, $c \to 0$ to preserve equality in (79). For small enough c, we may approximately replace the U's by their asymptotes, which are

$$U(x, 2, c) \cong \frac{1}{\Gamma(x)} c^{-1} \tag{80}$$

$$U(x, 1, c) \cong -\frac{1}{\Gamma(x)} \ln c. \tag{81}$$

This results in

$$E_{xe}(\rho, a, 1) \geq \rho \ln(-\ln c) \tag{82}$$

$$a \cong \frac{1}{-c \ln c} \tag{83}$$

as $c \to 0$. Since $-c \ln c > c$ if $c < e^{-1}$, we may write

$$E_{xe}(\rho, a, 1) > \rho \ln(\ln a) \tag{84}$$

as $a \to \infty$ for $0 < \rho < \infty$. Note that this is invalid at $\rho = \infty$.

Equation 84 implies that $E_{xe}(\rho, a, 1)$ must go to infinity as a goes to infinity. The significance is that, when all other parameters are held fixed, P_e can be forced to zero by increasing the available transmitter power, since a is proportional to P/N_o. Note that the p(x) chosen for this bound is the best of a fairly general family defined by (76), and might well be about as good as we can do with a continuous probability distribution.

There is some reason to believe that the right-hand side of (84) represents the true behavior of $E_{xe}(\rho, a, 1)$ as $a \to \infty$. It is known[10] that the expurgated-bound exponent for the additive Gaussian noise channel, $E_{xg}(\rho, a)$, behaves as

$$E_{xg}(\rho, a) \cong \rho \ln a, \quad a \to \infty, \quad 0 < \rho < \infty \tag{85}$$

$$E_{xg}(\infty, a) \cong \frac{1}{2} a, \quad a \to \infty, \tag{86}$$

or, in other words,

$$E_{xg}(\rho, a) \cong \rho \ln E_{xg}(\infty, a), \quad a \to \infty, \quad 0 < \rho < \infty. \tag{87}$$

One might conjecture that the expurgated bound for the present channel also satisfies (87). Since we have previously shown that $E_{xe}(\infty, a, 1) \cong \frac{1}{4} \ln a$ as $a \to \infty$, this would mean $E_{xe}(\rho, a, 1) \cong \rho \ln \ln a$, which is the same as the behavior of the bound just obtained.

3.2 RANDOM-CODING AND SPHERE-PACKING BOUNDS

The bounds and optimization techniques considered here are very similar to the expurgated bound just discussed, but are more complicated, so that the results obtained will be less complete. We start with the random-coding upper bound to error probability,[10] which, when applied to this channel model, states: If each code word is constrained to satisfy $\sum_{n=1}^{N} x_{mn}^2 \leq Na$, then for any block length N, any $0 \leq s \leq \frac{1}{2}$, $r \geq 0$, and any probability density p(x) such that $\int_0^\infty x^2 p(x)\, dx = a < \infty$, there exists a code for which

$$P_e < \exp - N\left\{\mathscr{E}_o\left(\frac{1}{1-s}, p(x), r\right) - \frac{s}{1-s} R_N - \zeta_N\right\} \tag{88}$$

$$\mathscr{E}_o\left(\frac{s}{1-s}, p(x), r\right) = -\ln \int_0^\infty \left[\int_0^\infty p(x)\, p_{\underline{\lambda}}(\underline{y}|x)^{1-s}\, e^{r(x^2-a)}\, dx\right]^{1/(1-s)} d\underline{y}, \tag{89}$$

where R_N was defined earlier, and $\zeta_N \to 0$ as $N \to \infty$, provided $\int_0^\infty p(x)\, |x^2-a|^3\, dx < \infty$. Readers familiar with the random-coding bound will note that instead of the more conventional parameter ρ, we use $s = \frac{\rho}{1+\rho}$. This will later simplify the derivation of the lower bound to P_e. Aside from this parameterization, the bound of (88) and (89) has the usual properties enumerated by Gallager.[10] The critical rate R_{crit}, is now defined as the rate obtained when $s = \frac{1}{2}$. As in the expurgated bound, we are faced with the problem of determining the p(x) and r that result in the tightest bound:

$$E_o\left(\frac{s}{1-s}, a, \underline{\lambda}\right) = -\ln \left[\min_{r, p(x)} \int_0^\infty \left[\int_0^\infty p(x)\, p_{\underline{\lambda}}(\underline{y}|x)^{1-s}\, e^{r(x^2-a)}\, dx\right]^{1/(1-s)} d\underline{y}\right], \tag{90}$$

subject to the constraints (45).

Because of the integral on $\underline{y}$, this minimization problem is so complex that there is little hope of obtaining results for arbitrary $\underline{\lambda}$. Even in the simpler case of the expurgated bound, when the $\underline{y}$ integral could be removed, an arbitrary $\underline{\lambda}$ was hard enough to handle so that we chose to avoid the possibility. Thus necessity dictates that we constrain the analysis to equal eigenvalues systems at the outset. Then, as we have noted, the vector $\underline{y}$ of outputs reduces to a scalar, and although we are still left with the integration of y, it is now a single integration. With this simplification, and again with the use of the subscript e for equal eigenvalues, the minimization becomes

$$E_{oe}\left(\frac{s}{1-s}, a, K\right) = -\ln\left[\min_{r, p(x)} \int_0^\infty \left[\int_0^\infty p_K(x)\, p(y|x)^{1-s}\, e^{r(x^2-a)}\, dx\right]^{1/(1-s)} dy\right], \tag{90}$$

where $p_K(y|x)$ is given by Eq. 31. To simplify things, we normalize x by dividing by $\sqrt{K}$, to obtain

$$E_{oe}\left(\frac{s}{1-s}, a, K\right) = -\ln\left[\min_{r, p(x)} \int_0^\infty \left[\int_0^\infty p(x)\, p(y|x)^{1-s}\, e^{r(x^2-a/K)}\, dx\right]^{1/(1-s)} dy\right], \tag{91}$$

where

$$p(y|x) = \frac{y^{K-1}\, e^{-y/(1+x^2)}}{\Gamma(K)(1+x^2)^K}, \tag{92}$$

and the constraints (45) now are

$$r \geq 0, \quad p(x) \geq 0, \quad \int_0^\infty p(x)\, dx = 1, \quad \int_0^\infty x^2 p(x)\, dx = a/K. \tag{93}$$

Unfortunately, it is no longer possible to remove K from the problem, as we did before.

In Theorem A.3 in Appendix A, it is shown that a sufficient condition for r, p(x) to be optimum in (91) is

$$\int_0^\infty \beta(y)^{s/(1-s)}\, p(y|x)^{1-s}\, e^{rx^2}\, dx \geq \int_0^\infty \beta(y)^{1/(1-s)}\, dy \tag{94}$$

for all x, when $0 < s < 1$, and with

$$\beta(y) \equiv \int_0^\infty p(x)\, e^{rx^2}\, p(y|x)^{1-s}\, dx. \tag{95}$$

In Theorem D.5 in Appendix D, it is shown that, if $0 < s < 1$, condition (94) must be satisfied by some r, p(x) combination, where $0 \leq r \leq K(1-s)$, and p(x) is a finite set of impulses. More exactly, $p(x) = \sum_{n=1}^{N_1} p_n u_o(x-x_n)$, where N_1 is a finite integer, and

$0 \leq x_n^2 \leq z_s$, where z_s is a function only of s, a, and K, and is finite for $0 < s \leq 1$, $0 \leq a < \infty$, and $K < \infty$. Both (94) and the impulsive solution are valid for all $0 < s < 1$, and not just $0 < s \leq \frac{1}{2}$ as required for the random-coding bound. This extension of the region of s is necessary for consideration of the lower bound, which we shall now present.

In Appendix E, it is shown that, if $\Delta_N < R_N < d_N$, then

$$P_e > \exp - N\left[\max_{0\leq s\leq 1} \left\{E_{oe}\left(\frac{s}{1-s}, a, K\right) - \frac{s}{1-s}(R_N-\Delta_N)\right\} + \delta_N\right], \tag{96}$$

where Δ_N and δ_N go to zero and d_N goes to infinity. The right-hand side of (96) differs from the random-coding upper bound to P_e, only in additive factors that go to zero as N goes to infinity, and in the difference in the range of s. When $R_N > R_{crit}$, the maximum will be in the range $0 < s < \frac{1}{2}$, so as $N \to \infty$, the upper and lower bounds exponentially agree for that range of rates, and thus describe the true asymptotic behavior of this channel model. As in the expurgated bound, the zero-rate intercept ($s = 1$) is the easiest point to consider, as far as the optimization is concerned, so we shall discuss it first.

3.21 Zero-Rate Bound

When the limit $s \to 1$ is taken, we find that $r = 0$, and

$$E_{oe}(\infty, a, K) = \min_{p(x)} -K\left[\int_0^\infty p(x) \ln (1+x^2)\, dx + \ln \left[\int_0^\infty \frac{p(x)}{1+x^2}\, dx\right]\right], \tag{97}$$

where p(x) must satisfy (93). In Theorem A.4 it is shown that a sufficient condition for p(x) to be optimum is

$$\ln (1+x^2) + \frac{1}{1+x^2}\left[\int_0^\infty \frac{p(x_1)}{1+x_1^2}\, dx_1\right]^{-1} \leq \lambda_o + \lambda_1 x^2 \tag{98}$$

for some λ_o, λ_1 and all x, with equality when $p(x) > 0$.

In Theorem C.1 in Appendix C, it is shown that a p(x) given by (65), will satisfy (98) when the parameters are chosen correctly, so that once again, at zero rate, two impulses are optimum. In Figs. 18 and 19 we present the optimum distribution in terms of a, and the exponent is shown in Fig. 20, together with the expurgated-bound zero-rate exponent for comparison. In this special case, K <u>can</u> be normalized into a. Once again, $\frac{1}{a}E_{oe}(\infty, a, K)$ is a decreasing function of a, and

$$\frac{1}{a}E_{oe}(\infty, a, K) \leq 0.2162. \tag{99}$$

Note that, as $a \to 0$, $E_{oe}(\infty, a, K) \cong 1.46\, E_{xe}(\infty, a, K)$ and as $a \to \infty$, $E_{oe}(\infty, a, K) \cong 4E_{xe}(\infty, a, K)$.

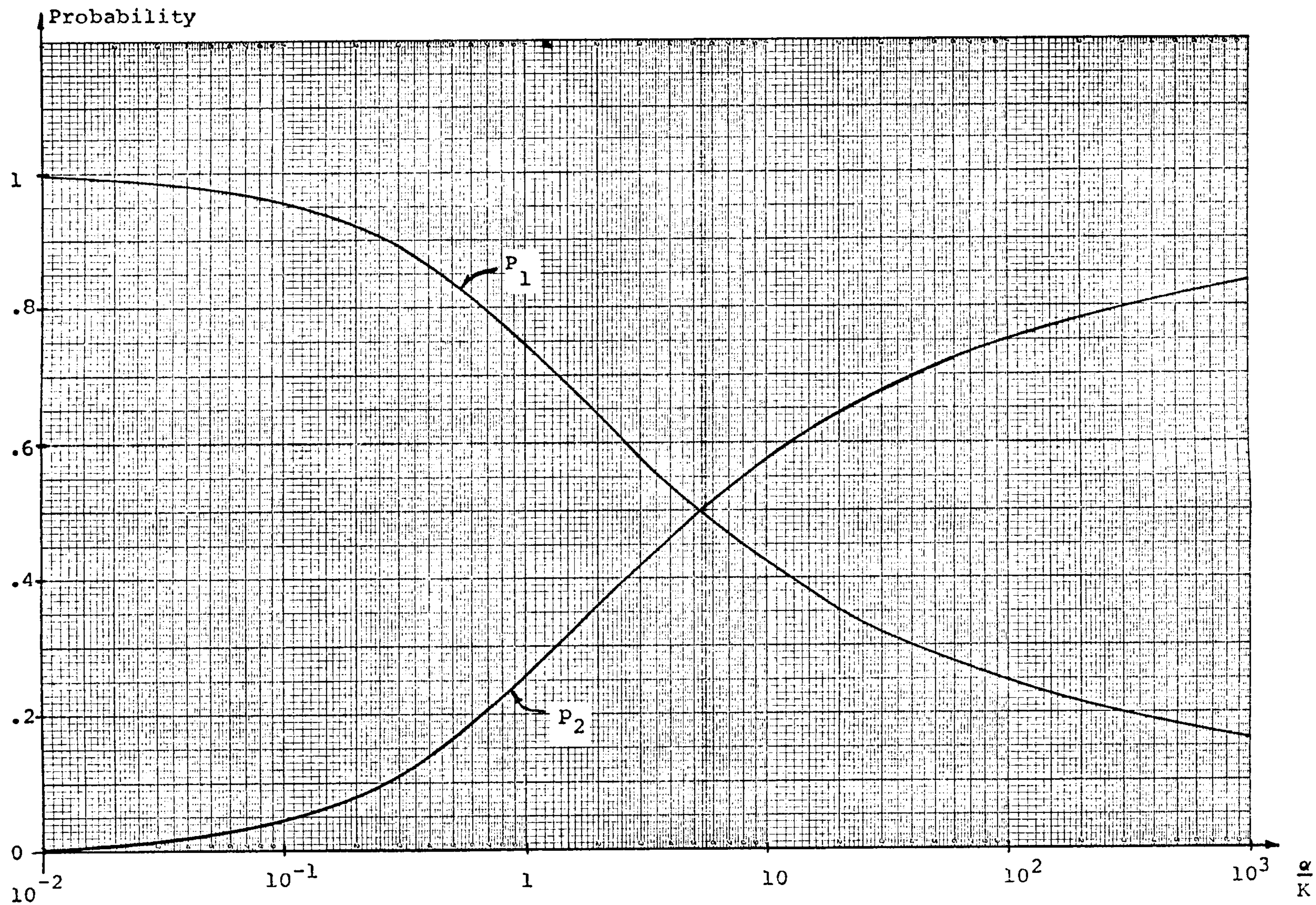

Fig. 18. p_1 and p_2 vs $\frac{\alpha}{K}$ for $E_{oe}(\infty, \alpha, K)$.

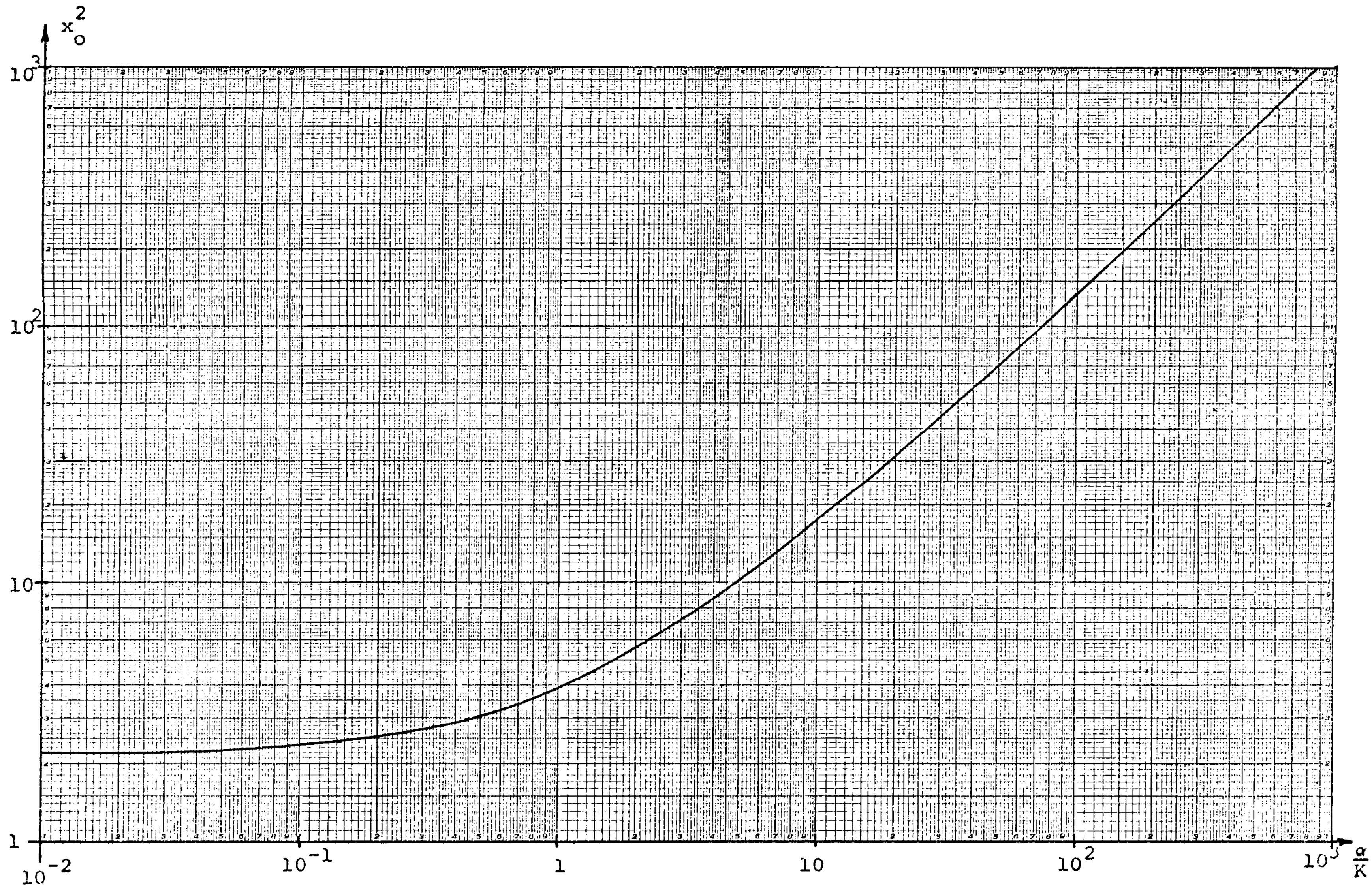

Fig. 19. x_o^2 vs $\frac{a}{K}$ for $E_{oe}(\infty, a, K)$.

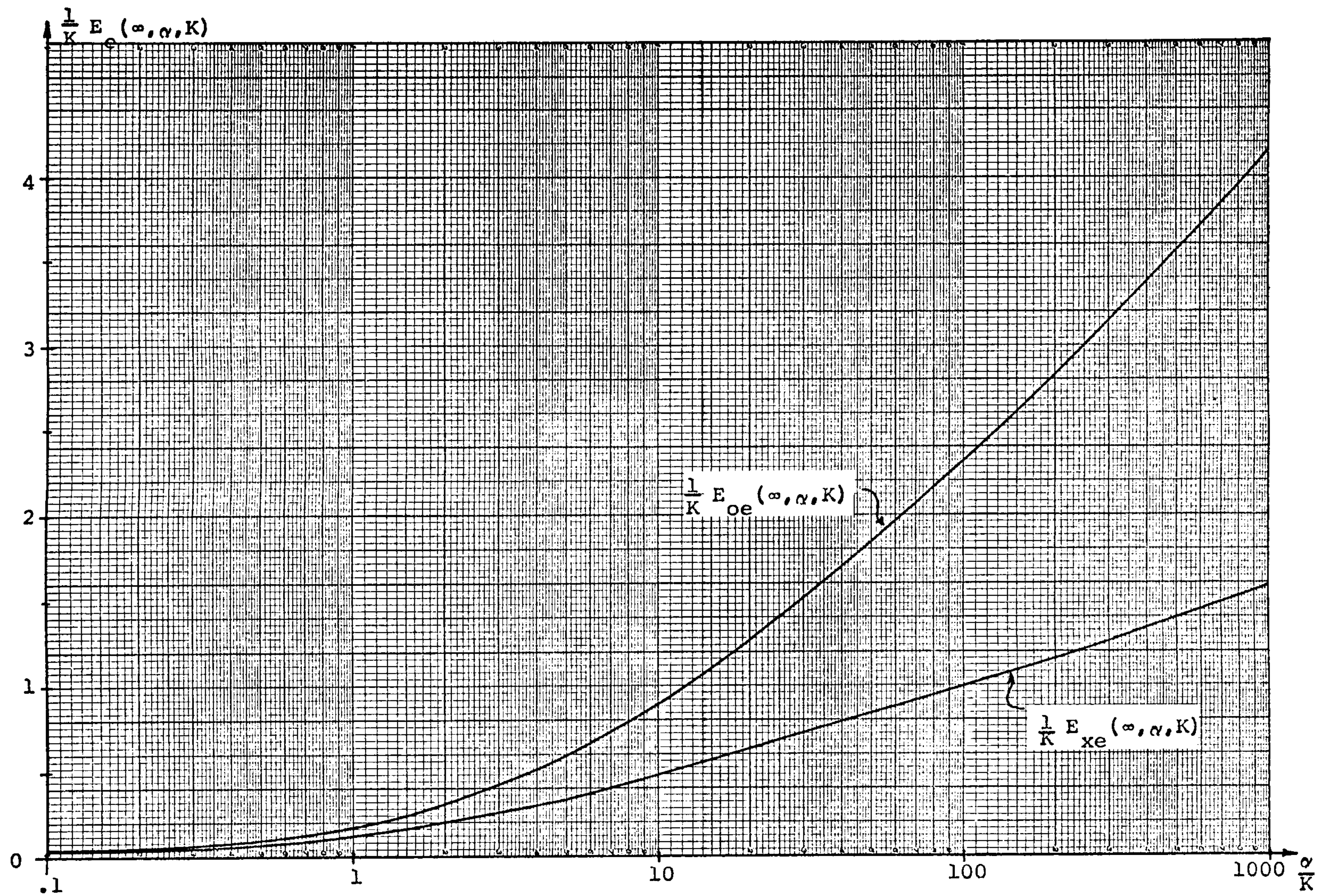

Fig. 20. $\frac{1}{K} E_{oe}(\infty, \alpha, K)$ and $\frac{1}{K} E_{xe}(\infty, \alpha, K)$ vs $\frac{\alpha}{K}$.

3.22 Positive Rates

As in the expurgated bound, when $\rho < \infty$, corresponding here to $s < 1$, results are much harder to obtain, beyond the basic fact that an impulsive p(x) is optimum. As before, there are some special cases for which results may be derived without too much effort. The first case is for $s = \frac{1}{2}$. This point reduces to the $\rho = 1$ expurgated bound by noting that the integration on y may be performed (at the expense of a double integral on x), and results in $E_{xe}(1, \alpha, K)$. This problem has already been considered, so that numerical results are already available. Moreover, it can be shown (Theorem C. 2) that a two-impulse p(x), when combined with a suitable value of r, will asymptotically satisfy (94) as α goes to zero. In detail,

$$p(x) = \left(1 - \frac{\alpha}{z_o K}\right) u_o(x) + \frac{\alpha}{z_o K} u_o(x - \sqrt{z_o}) \tag{100}$$

$$r = K(1-s)\, f_{s/(1-s)}(z_o) \tag{101}$$

$$f_{s/(1-s)}(x) \equiv \frac{1}{x(1-s)} \left[\ln(1+sx) - s \ln(1+x)\right], \tag{102}$$

for which it is known[16] that for any $0 < s < 1$, $f_{s/(1-s)}(x)$ is positive for $x > 0$, has a single maximum in x, and z_o is chosen to be that maximum. The resulting exponent is the same as the infinite-bandwidth exponent found by Kennedy:

$$E_{oe}\left(\frac{s}{1-s}, \alpha, K\right) = \alpha\, f_{s/(1-s)}(z_o), \qquad \alpha \to 0. \tag{103}$$

Aside from the cases just mentioned, only numerical results are possible. If $\{x_n\}$ and r are specified, the function to be minimized is known to be a convex function of $\{p_n\}$. This property may be used to determine conditions[29] specifying the best $\{p_n\}$, given $\{x_n\}$ and r, but since the objective function is no longer a quadratic form in $\{p_n\}$, implementation of the problem for computer solution is much more difficult. Nevertheless, this was done, using the same general program mentioned before[27] as a base. Unfortunately, because of the integral on y (which must be numerically evaluated at each step), running even a small number of $\{x_n\}$ as inputs takes a large amount of computer time, so this was done sparingly.

3.23 Capacity

There is one other case in which some simplifications are possible, and that is the determination of capacity, obtained by letting $s \to 0$. In that case, it is easily shown that $r = 0$ provides an optimum, and since the exponent goes to zero as $s \to 0$, the problem is a maximization of the rate at which this occurs rather than a maximization of the exponent. Denote this maximum rate, capacity, by $C(\alpha, K)$, where

$$C(\alpha, K) = \int_0^{\infty} \int_0^{\infty} p(x)\, p(y|x) \ln \left[\frac{p(y|x)}{\int_0^{\infty} p(x_1)\, p(y|x_1)\, dx_1} \right] dxdy. \tag{104}$$

It may be shown, by methods similar to those used before, that a sufficient condition on p(x) for maximization of $C(\alpha, K)$, subject to (93), is

$$\gamma\left(x^2 - \frac{\alpha}{K}\right) + C(\alpha, K) \geq \int_0^{\infty} p(y|x) \ln \left[\frac{p(y|x)}{\int_0^{\infty} p(x_1)\, p(y|x_1)\, dx_1} \right] dy \tag{105}$$

for all x and some γ.

Once again, if we consider the limit as $\alpha \to 0$, it can be shown that two impulses are asymptotically optimum, and result in the infinite-bandwidth capacity, which is the same as that of an additive Gaussian noise channel with the same value of P/N_0. It can also be shown numerically that two impulses are optimum for a range of small but positive α. The same program that was used before was modified for numerical maximization of $C(\alpha, K)$, and some results are shown in Fig. 21, where $C(\alpha, 1)/\alpha$ is plotted against α. Here again, even small sets of $\{x_n\}$ resulted in long running times.

When capacities were computed for channels with $K > 1$, with N and α fixed, it was found that $K > 1$ was uniformly worse than $K = 1$, for all values of α and K that were tried. This is not unexpected, since our expurgated bound work indicated that as ρ became smaller, the optimum value of K decreased, too. Once again, we defer further discussion to Section IV, in which some additional results will be presented.

3.3 COMMENTS

The results given above were quite sketchy, the major limitation being the running time of the numerical optimizations. Even so, a quite passable E(R) curve can often be obtained with a relatively small number of computed points. As an example of this, in Fig. 22 we present the E(R) curve for a channel with $K = 1$ eigenvalue, and $\alpha = 1$. For comparison, the infinite bandwidth E(R) curve is also drawn. For convenience, both axes are normalized to α.

As long as we adhere to the equal-eigenvalue assumption, it is feasible to generate numerical results. For the expurgated bound, the optimizing algorithm makes use of existing quadratic programming techniques that are basically simple and quite efficient. With the random-coding bound, the problem is no longer one of quadratic programming; furthermore, the algorithm must perform a numerical integration at each step, so that numerical results are just barely feasible.

When the eigenvalues are no longer equal, the situation changes drastically. For the expurgated bound, the problem is still one of quadratic programming, so the numerical optimization can still be performed for any finite number of λ_k, although it will take longer. The only possible drawback is that there is no longer a guarantee that optimum p(x) is impulsive, and problems might arise if a large number of x_n are

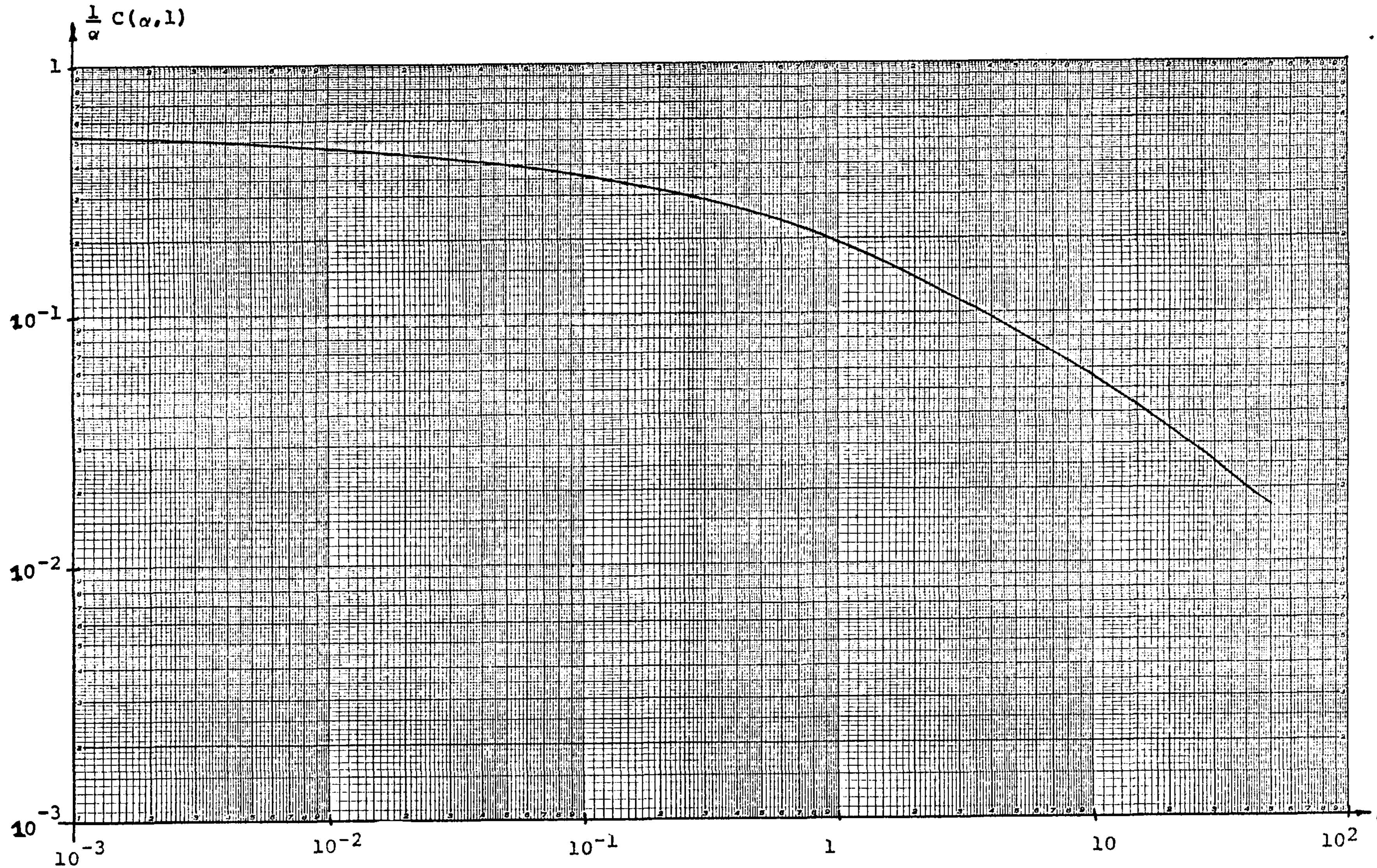

Fig. 21. $\frac{1}{\alpha}C(\alpha, 1)$ vs α.

necessary to adequately approximate a continuous p(x). For the random-coding bound, there will be one extra numerical integration per step for each eigenvalue, and that in

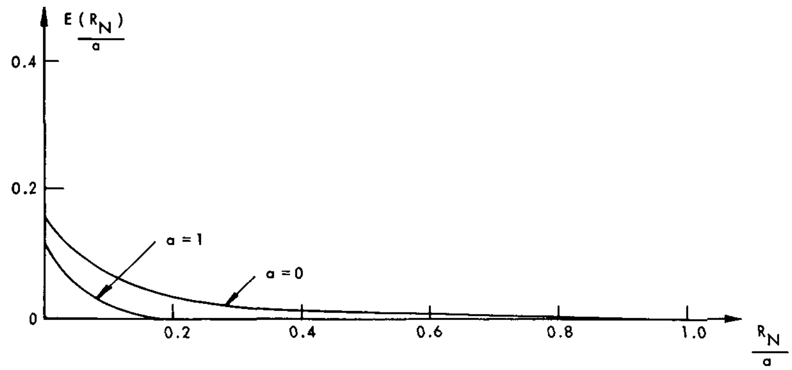

Fig. 22. Normalized E(R) curves, K = 1, α = 1, and 0.

itself should be enough to almost completely eliminate the possibility of numerical results. Also note that the lower bound presented here will not be valid for unequal eigenvalues.

Once again, we note that our purpose here was to present results that will later be applied to the questions that were raised in Section II. Thus we have purposely left much discussion, interpretation, and comparison for Section IV.

IV. RESULTS AND INTERPRETATION

4.1 DISCUSSION OF RESULTS

In Section III we evaluated both upper and lower bounds to the minimum attainable error probability, for coding over N independent uses of the channel model derived in Section II. The results are almost exclusively restricted to channels with K equal eigenvalues (although K may be arbitrary), and hereafter we shall refer only to equal-eigenvalue systems. We note, however, that (at least for low rates), a channel with an arbitrary set of eigenvalues has a minimum P_e that may be upper-bounded in terms of the P_e attainable with an equal-eigenvalue channel. Equality holds in the bound when the arbitrary eigenvalues are all equal, and a few examples indicate that the bound might be reasonably good for non-equal eigenvalues. Thus the equal-eigenvalue assumption may be regarded as an approximation to facilitate analysis and insight, rather than a restriction precluding application of the results to real channels.

The upper and lower bounds to the minimum P_e confirm that this P_e is exponentially decreasing in N, and for a range of rates the bounds agree as N becomes large. For this range of rates, then, our bounds represent the true exponential behavior of this channel model. Each bound involves an arbitrary probability density function p(x) that has to be optimized to obtain the tightest bound. We found that the optimum density always consists of a finite number of impulses, and we shall now discuss the implications.

4.11 Significance of p(x)

The density p(x) may be interpreted as a probability density from which the letters for each code word (the modulations x_{mn}) are to be chosen at random. This is not exactly correct, since to ensure that no code words have energies significantly smaller than the average, it is necessary to keep renormalizing p(x) as successive letters of each code word are chosen. For purposes of interpretation, however, this effect may be ignored. An optimum p(x) consisting of an impulsive density then corresponds to choosing the input letters from a discrete set of voltage levels.

For a small enough value of $P/N_oW = \alpha_o$, we found that the optimum p(x) consisted of only two impulses, one of which was at the origin. When this modulation level (zero) is chosen, we "use" the channel by not sending that particular basic signal. Thus on any basic signal, we may not send anything, but when we do, it is always some optimum voltage level determined from the location of the second impulse. The zero level corresponds to saving energy so that the optimum amount may be sent on some other basic signal. As α_o goes to zero (obtained, for example, when W goes to infinity) the impulse that is not at the origin is placed so that the resulting value of output energy-to-noise ratio per diversity path is identical to that found by Kennedy to be optimum for orthogonal signals when an infinite bandwidth is available. This result is independent of the number of diversity paths of the basic signals because the ensemble p(x) can be used to compensate for any value of K.

As α_0 increases from zero, two impulses eventually cease to be optimum, and p(x) goes to three impulses, then four, etc. Thus, as the signal-to-noise ratio is increased for a fixed bandwidth, there will be a set of optimum voltage levels to be used for communication. This corresponds to the fact that we must use more levels to make best use of the increase in available signal power. Furthermore, if the rate R is increased with other parameters held constant, the number of levels also increases because we require more levels to transmit the greater amount of information. As a final note, we recall that one impulse is always located at the origin, thereby reflecting the fact that this input has the smallest output variance, and, of course, the energy constraint also tends to keep signal levels low.

4.12 Comparison with the Gaussian Channel

We shall compare the fading channel to an additive Gaussian noise channel with an equivalent value of output signal-to-noise ratio per channel use (a detailed discussion of the Gaussian channel results has been given by Gallager[10]). The first major difference is in the form of the optimum p(x). Here an impulsive density was optimum, while for the Gaussian channel, the best p(x) was Gaussian. When the resulting exponents are evaluated, it is clear that performance is always better for the Gaussian channel.

As representative examples, consider Figs. 23, 24, and 25. The first shows $1/\alpha\, E_{xe}(\rho, \alpha, 1)$ vs α (proportional to P/N_0W) for the fading channel with $\rho = 1$ and ∞, and the equivalent exponents for the Gaussian channel. Without going into detail, we recall that ρ is a parameter (used in Section III) inversely related to rate, such that $\rho = \infty$ is zero rate, $\rho = 1$ is critical rate, and $\rho = 0$ is capacity. This will be elaborated on later, along with the normalization with respect to α. Note that, at zero rate for the Gaussian channel, the infinite-bandwidth exponent may be obtained for any signaling bandwidth, but not for the fading channel. This is due to the fact that the only known way to combat the probability of a deep fade is through sufficient signal diversity, and for small W, even at small rates, there is not enough diversity available.

Figures 24 and 25 show equivalent capacities for the fading and nonfading channels. It is known[15] that the capacity of the infinite-bandwidth fading channel is the same as that of an equivalent Gaussian channel, a fact which, although not apparent from the figures (as $\alpha \to 0$), was shown in Section III. Note the vast difference in the rate of approach to this asymptote for the two channels.

4.13 Comparison with Orthogonal Signals

We have already noted that, as α goes to zero, this channel model provides the orthogonal-signal, infinite-bandwidth exponent. Furthermore, the optimizing density consists of two impulses, one at the origin, and one at a level that excites the same value of output energy-to-noise ratio per diversity path that was optimum in the

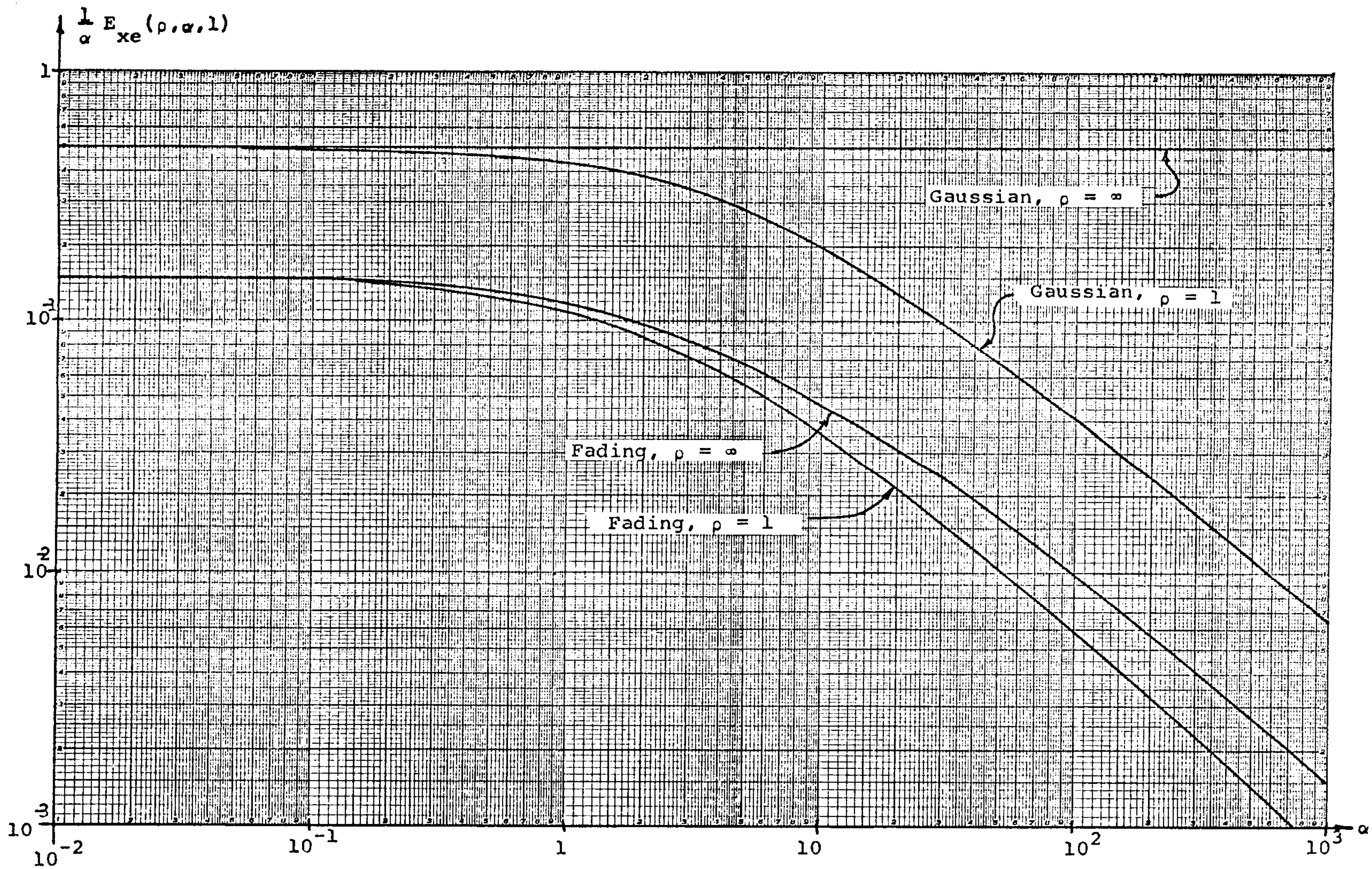

Fig. 23. Exponents for equivalent Gaussian and fading channels.

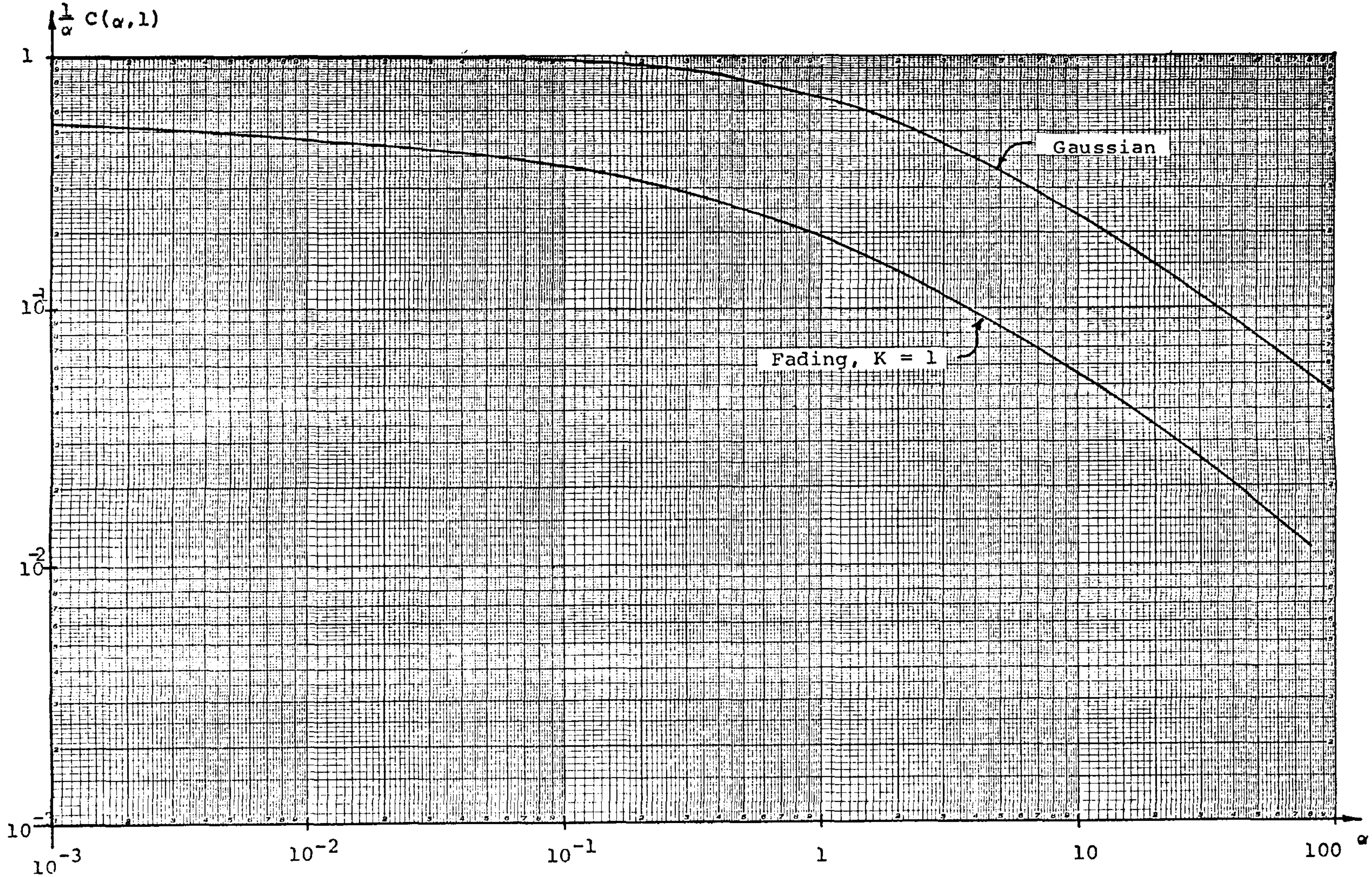

Fig. 24. Capacities for equivalent Gaussian and fading channels.

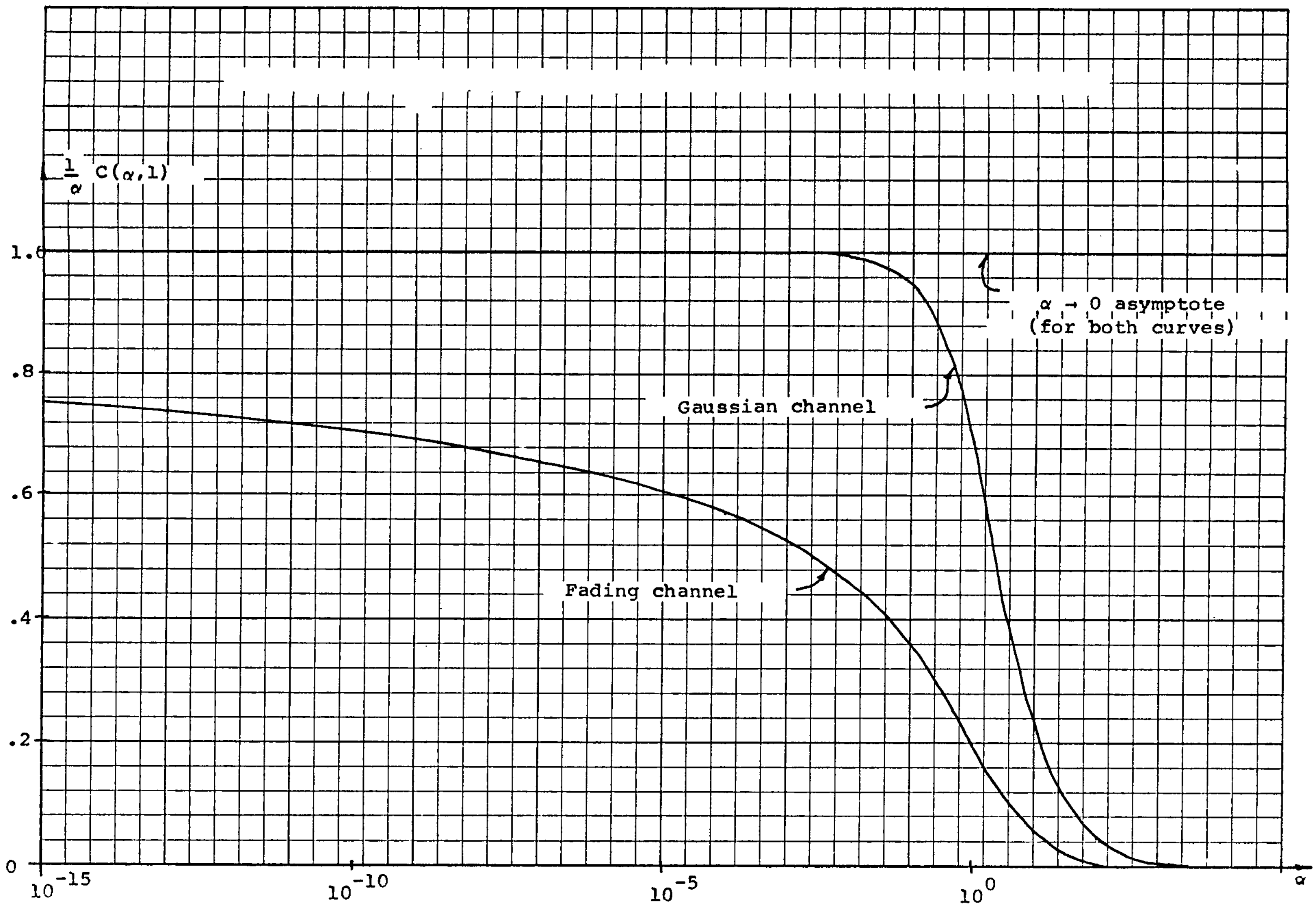

Fig. 25. $\frac{1}{a}$ C(a, 1) for small a, Gaussian and fading channels.

orthogonal-signal analysis. We note from Figs. 23-25 that as the rate is increased, the degradation in performance, owing to lack of sufficient bandwidth, becomes more severe. This degradation may be moderate at low rates, but will be quite severe at rates close to capacity.

The infinite-bandwidth analysis revealed that the optimum fading channel was one with equal eigenvalues. Unfortunately, it has not been possible to show that the same is true for finite bandwidths, although we speculate that such is the case. As a final comment, we note that, with infinite bandwidth, Kennedy was able to bound the performance of a channel with arbitrary eigenvalues in terms of the performance of an equal-eigenvalue channel, for all rates. This has been possible here only for low rates, although our bound is completely analogous to his.

4.2 APPLICATION TO THE COMMUNICATION PROBLEM

Up to this point, the results of Section III have been presented on a per channel use basis, and now we shall relate them back to the communication scheme set forth at the end of Section II.

4.21 Exponent-Rate Curves

Since the problem has been formulated as the determination of exponential bounds on the minimum P_e attainable through coding, the quantity of interest for determination of performance levels is the exponent-rate curve. We have shown that

$$P_e \leq \min \begin{cases} e^{-N \max\limits_{0<\rho\leq 1} [E_{oe}(\rho, a, K) - \rho R_N]} \\ e^{-N \max\limits_{\rho \geq 1} [E_{xe}(\rho, a, K) - \rho R_N]} \end{cases} \tag{106}$$

$$P_e \geq e^{-N \max\limits_{0<\rho<\infty} [E_{oe}(\rho, a, K) - \rho R_N]}, \tag{107}$$

with additive factors omitted in the exponents that become negligible as N goes to infinity. In Section III a parameter, s, was used to simplify derivation of the lower bound, but for uniformity of notation, we now use $\rho = s/(1-s)$.

These bounds all have the same basic form, differing only in subscripts and ranges of ρ, and we shall describe the random-coding bound (the top inequality of (106)) as typical. We consider application of this bound to the scrambling scheme outlined at the end of Section II, and then summarize the analogous results for the other schemes. Recall that

$$R_N = RT/N \tag{108}$$

$$a = PT/N_o N \tag{109}$$

$$N = TW/Q_{sc} \tag{110}$$

$$Q_{sc} = (T_s+L)(W_s+B) \tag{111}$$

$$K = \begin{cases} (1+BT_s)(1+LW_s), & BL < 1 \\ (T_s+L)(W_s+B), & BL \geq 1 \end{cases} \tag{112}$$

Plugging these into (106) yields

$$P_e \leq e^{-\frac{PT}{N_o} E_{sc}(R)} \tag{113}$$

$$E_{sc}(R) = \max_{0<\rho\leq 1} \left[\frac{1}{\alpha_o Q_{sc}} E_{oe}(\rho, \alpha_o Q_{sc}, K) - \rho\left(\frac{R}{P/N_o}\right)\right], \tag{114}$$

in which we have defined

$$\alpha_o = P/N_o W. \tag{115}$$

Given the basic signals (in this model, just a specification of T_s and W_s), together with the parameters B, L, W, and P/N_o, we may generate $E_{sc}(R)$ given parametrically in (114). As discussed elsewhere,[10] $E_{sc}(R)$ may be constructed as the upper envelope of the set of straight lines with slopes $(-\rho)$ and intercepts $E_{oe}(\rho, \alpha_o Q_{sc}, K)/\alpha_o Q_{sc}$. Note that K and Q_{sc} depend on T_s and W_s, and we are free to choose these quantities to obtain the largest exponent (subject to the rough constraint $T_s W_s \geq 1$). For any given ρ, the intercept should be maximized, so we must compute

$$\max_{T_s, W_s} \frac{1}{\alpha_o Q_{sc}} E_{oe}(\rho, \alpha_o Q_{sc}, K).$$

In general, this must be done numerically, although, as we shall see, there are many cases for which this is not necessary.

In Table 2, we present the random-coding bound for the three schemes proposed at the end of Section II, before optimization on T_s and W_s. To get equivalent results for the sphere-packing lower bound, replace $0 < \rho \leq 1$ by $0 < \rho < \infty$, and for the expurgated bound, replace E_{oe} by E_{xe} and $0 < \rho \leq 1$ by $\rho \geq 1$. Note that, since $Q_{sc} < Q$, $E_{re}(R) < E_{sc}(R)$. Also (in Section III), we found that $E_{oe}(\rho, \beta, K)/\beta$ was a decreasing function of β, so that $E(R) < E_{sc}(R)$, too. Thus, of the three schemes, scrambling gives the largest exponent.

Unfortunately, given R, the optimum value of ρ will depend on T_s and W_s, thereby making the optimization difficult. Thus a reasonable procedure is to perform the maximization on T_s and W_s for several different values of ρ, draw the resulting straight

Table 2. Exponent-rate curves for arbitrary T_s, W_s.

Scheme	Random-Coding Exponent
Scrambling	$E_{sc}(R) = \max_{0<\rho\leq 1} \left[\frac{E_{oe}(\rho, a_o Q_{sc}, K)}{a_o Q_{sc}} - \rho\left(\frac{R}{P/N_o}\right)\right]$
No Scrambling	$E(R) = \max_{0<\rho\leq 1} \left[\frac{E_{oe}(\rho, a_o Q, K)}{a_o Q} - \rho\left(\frac{R}{P/N_o}\right)\right]$
Rate-Expanding	$E_{re}(R) = \max_{0<\rho\leq 1} \frac{Q_{sc}}{Q} \left[\frac{E_{oe}(\rho, a_o Q_{sc}, K)}{a_o Q_{sc}} - \rho\left(\frac{R}{P/N_o}\right)\right]$
Parameters	$a_o = P/N_oW$ $Q = (T_s+L+1/B)(W_s+B+1/L)$ $Q_{sc} = (T_s+L)(W_s+B)$ $K = \begin{cases} (1+BT_s)(1+LW_s), & BL < 1 \\ (T_s+L)(W_s+B), & BL \geq 1 \end{cases}$

lines, and then get the resulting exponent numerically for the rate of interest. This difficulty does not apply at the end points of the curve, since $\rho = \infty$ corresponds to $R = 0$, and $\rho = 0$ corresponds to capacity, independent of all other parameters. We shall consider these points first.

4.22 Zero Rate

At zero rate, $\rho = \infty$, independent of all other parameters. Unfortunately, the upper and lower bounds do not agree at this point, so they must be considered separately. The applicable upper bound is the expurgated bound, for which the problem becomes (note that $\lim_{R\to 0} \rho R = 0$):

$$E_{sc}(0) = \max_{T_s, W_s} \frac{E_{xe}(\infty, a_o Q_{sc}, K)}{a_o Q_{sc}} = \max_{T_s, W_s} \frac{E_{xe}\left(\infty, \frac{a_o Q_{sc}}{K}, 1\right)}{a_o Q_{sc}/K}, \tag{116}$$

in which Eq. 59 has been used. Thus the problem reduces to the minimization of Q_{sc}/K. When $BL \geq 1$, $K = Q_{sc}$, independent of T_s and W_s, and for $BL < 1$, the minimization yields $Q_{sc}/K = 1$ for any $T_sW_s = 1$ signal. Thus

$$E_{sc}(0) = \frac{1}{a_o} E_{xe}(\infty, a_o, 1), \tag{117}$$

which is easily evaluated from Fig. 23.

For the nonscrambled scheme, the optimizing signals are

$$T_s = \sqrt{\frac{La}{B}}, \quad W_s = \sqrt{\frac{Ba}{L}}, \quad a \to \infty, \tag{118}$$

in which case

$$E(0) = \begin{cases} \frac{1}{a_o} E_{xe}(\infty, a_o, 1) & BL \geq 1 \\ \frac{BL}{a_o} E_{xe}\left(\infty, \frac{a_o}{BL}, 1\right), & BL < 1 \end{cases} \tag{119}$$

As a practical matter, $T_sW_s = a$ does not have to be too large, for the exponent in (119) will be approximately attained if

$$\sqrt{BL} + \sqrt{a} \gg 1/\sqrt{BL}, \quad BL \geq 1 \tag{120}$$

$$1 + \sqrt{BLa} \gg BL, \quad BL < 1. \tag{121}$$

When $BL \geq 1$, the scrambled and nonscrambled exponents are the same, while for $BL < 1$, the nonscrambling bandwidth must be increased by a factor of $1/BL$ before the exponents are equivalent. When $BL \geq 1$, both schemes are limited primarily by the guard space necessary to ensure orthogonality, but when $BL > 1$, the nonscrambled scheme is penalized by the large guard spaces necessary for providing independence between output signals. Without scrambling, the individual basic signals take up more area in the TW plane, thereby reducing the effects of the larger guard-space requirements.

For the rate-expanding scheme, when $BL \geq 1$, the basic signals of (118) are again optimum, with the result that the same exponents are obtained for all three schemes. When $BL < 1$, the optimization must be performed numerically. It can be shown, however, that if the basic signals of (118) are chosen, we arrive at the nonscrambled exponent, so that $E(0) \leq E_{re}(0) \leq E_{sc}(0)$. In Table 3 we list the optimized zero-rate expurgated-bound exponents for these signaling schemes.

Note that these exponents represent an upper bound to P_e. For many channels of interest, however, an improved zero-rate lower bound can be derived that agrees with this zero-rate expurgated upper bound,[12,30] so there is reason to believe that these exponents represent the true zero-rate channel behavior. When the sphere-packing zero-rate lower bound is evaluated, recall that

$$E_{oe}(\infty,\beta,K) = KE_{oe}(\infty,\beta/K,1), \tag{122}$$

which permits easy optimization on the input signals. In fact, the optimum signals are the same as those for the upper bound, so that the corresponding lower bound results may be read from Table 3 with E_{xe} replaced by E_{oe}. Figure 26 may be used for lower bound exponent evaluation.

Table 3. Optimized zero-rate expurgated-bound exponents.

Scheme		Exponent	Signals
Scrambling	all	$E_{sc}(0) = \frac{1}{\alpha_o} E_{xe}(\infty, \alpha_o, 1)$	$T_s W_s = 1$
No Scrambling	BL ≥ 1	$E(0) = \frac{1}{\alpha_o} E_{xe}(\infty, \alpha_o, 1)$	$T_s = \sqrt{\frac{La}{B}}$ $W_s = \sqrt{\frac{Ba}{L}}$ $a \to \infty$
	BL < 1	$E(0) = \frac{BL}{\alpha_o} E_{xe}\left(\infty, \frac{\alpha_o}{BL}, 1\right)$	
Rate-Expanding	BL ≥ 1	$E_{re}(0) = \frac{1}{\alpha_o} E_{xe}(\infty, \alpha_o, 1)$	"
	BL < 1	$E_{re}(0) = \frac{K}{\alpha_o Q} E_{xe}\left(\infty, \frac{\alpha_o Q_{sc}}{K}, 1\right)$	unknown

These upper and lower bounds are different, but the optimizing signals are the same, so whatever the actual exponent may be, the choice of input signals presented in Table 3 is probably a good one.

4.23 Capacity

Once again, because $\rho = 0$ at capacity (defined as the largest rate at which a positive exponent may be obtained), independent of other parameters, the results are simpler than in the general case. For scrambled signals,

$$\frac{C_{sc}}{P/N_o} = \frac{C(\alpha_o Q_{sc}, K)}{\alpha_o Q_{sc}}. \tag{123}$$

Recall that given $\alpha_o Q_{sc}$, K should be minimized, and given K, Q_{sc} should be minimized. It turns out that K and Q_{sc} are simultaneously minimized by

$$T_s = \sqrt{\frac{L}{B}}, \qquad W_s = \sqrt{\frac{B}{L}} \tag{124}$$

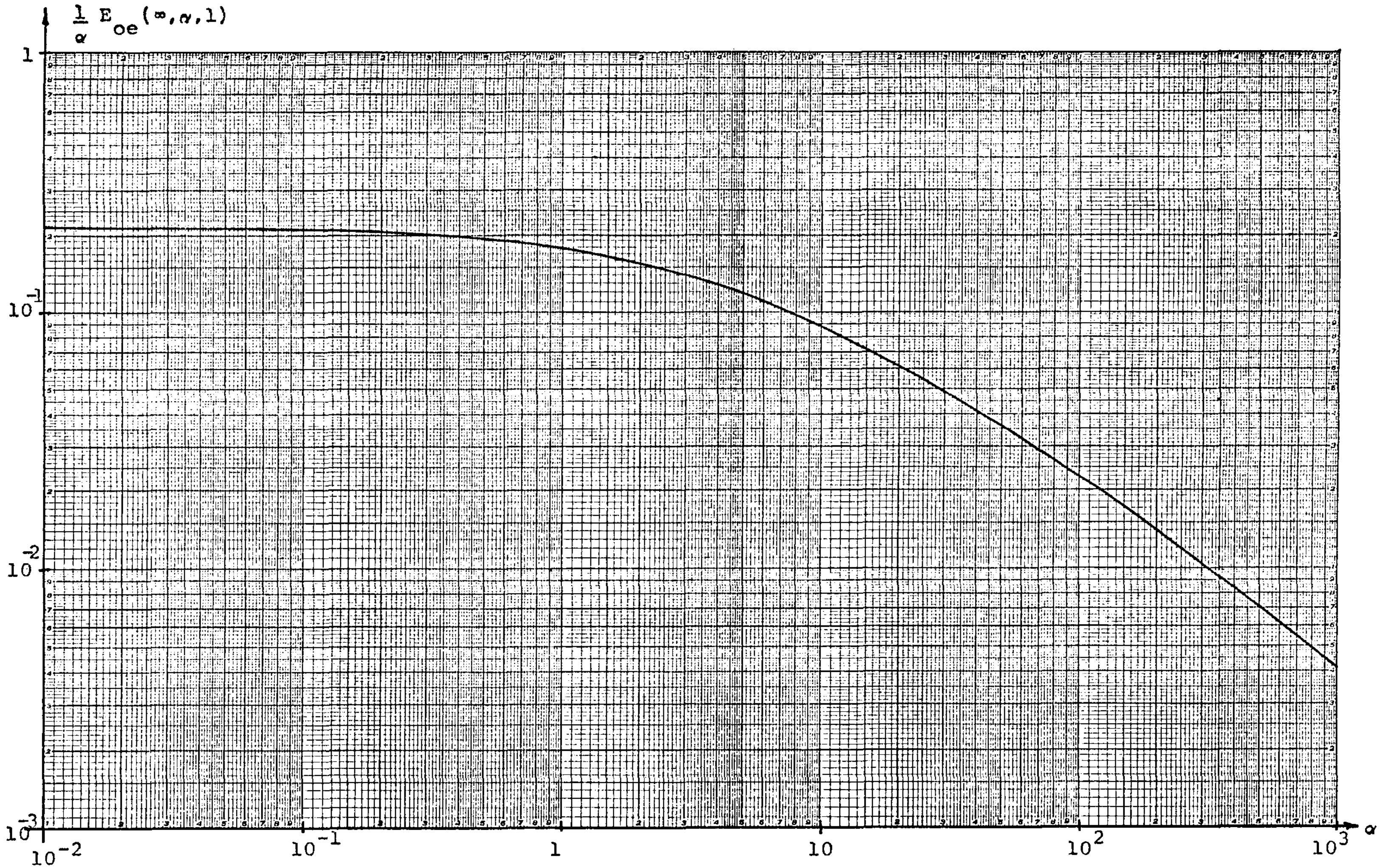

Fig. 26. $\frac{1}{\alpha} E_{oe}(\infty, \alpha, 1)$ versus α.

Table 4. Capacities for three schemes.

Scheme	Capacity	Approximation
Scrambling	$\frac{C_{sc}}{P/N_o} = \frac{C\left[\alpha_o(1+\sqrt{BL})^2,\ (1+\sqrt{BL})^2\right]}{\alpha_o(1+\sqrt{BL})^2}$	$\frac{C_{sc}}{P/N_o} \cong \begin{cases} \frac{C(\alpha_o, 1)}{\alpha_o}, & BL \leqslant 10^{-2} \\ \frac{C(\alpha_o BL, BL)}{\alpha_o BL}, & BL \geqslant 10^{2} \end{cases}$
No Scrambling	$\frac{C}{P/N_o} = \frac{C\left[\alpha_o(1+\sqrt{BL}+1/\sqrt{BL})^2,\ (1+\sqrt{BL})^2\right]}{\alpha_o(1+\sqrt{BL}+1/\sqrt{BL})^2}$	$\frac{C}{P/N_o} \cong \begin{cases} \frac{C(\alpha_o/BL, 1)}{\alpha_o/BL}, & BL \leqslant 10^{-2} \\ \frac{C(\alpha_o BL, BL)}{\alpha_o BL}, & BL \geqslant 10^{2} \end{cases}$
Rate Expanding	$C_{re} = C_{sc}$	$C_{re} = C_{sc}$

for all BL, with the resulting capacity

$$\frac{C_{sc}}{P/N_o} = \frac{C\left[\alpha_o(1+\sqrt{BL})^2,\ (1+\sqrt{BL})^2\right]}{\alpha_o(1+\sqrt{BL})^2}. \tag{125}$$

Without scrambling,

$$\frac{C}{P/N_o} = \frac{C(\alpha_o Q, K)}{\alpha_o Q} \tag{126}$$

Once again, for all BL, Q and K are simultaneously minimized by the signals of (124), with resulting capacity

$$\frac{C}{P/N_o} = \frac{C\left[\alpha_o(1+\sqrt{BL}+1/\sqrt{BL})^2,\ (1+\sqrt{BL})^2\right]}{\alpha_o(1+\sqrt{BL}+1/\sqrt{BL})^2}. \tag{127}$$

For the rate-expanding scheme, $C_{re} = C_{sc}$ for all BL, so the same input signals are optimum. These results are summarized in Table 4, together with the limiting capacities when BL gets large or small.

A graph of $\frac{1}{\alpha}C(\alpha, 1)$ was presented in Fig. 24. On account of the computational difficulties mentioned in Section III, we do not have sufficient data to draw equivalent curves for larger values of K. Many channels of interest have BL $\lesssim 10^{-1}$, however, and the variation of $\frac{1}{\alpha}C(\alpha, K)$ with K is not too large in the vicinity of K = 1, so Fig. 24 may be used in those cases for the estimation of capacities.

4.24 Other Rates

At rates between zero and capacity, the optimization is more complex. We first consider the simplest case, that of the expurgated bound.

a. Expurgated Bound

With the use of Eq. 50, we find

$$E_{sc}(R) = \max_{\rho \geq 1}\left[\frac{E_{xe}(\rho/K, \alpha_o Q_{sc}/K,\ 1)}{\alpha_o Q_{sc}/K} - \rho\left(\frac{R}{P/N_o}\right)\right]. \tag{128}$$

For a given K, Q_{sc} is minimized by choosing $T_s W_s = 1$, so that $Q_{sc}/K = 1$, and K should then be minimized. This is accomplished by using the signals of (124), with the resulting exponent

$$E_{sc}(R) = \max_{\rho \geq 1}\left[\frac{1}{\alpha_o}E_{xe}\left\{\frac{\rho}{(1+\sqrt{BL})^2},\ \alpha_o,\ 1\right\} - \rho\left(\frac{R}{P/N_o}\right)\right] \tag{129}$$

Table 5. Optimized expurgated bound exponents for positive rates.

Scheme	Exponent	Approximations
Scrambling	$E_{sc}(R) = \dfrac{E_{xe}\left\{\dfrac{\rho}{(1+\sqrt{BL})^2}, \alpha_o, 1\right\}}{\alpha_o} - \rho\left(\dfrac{R}{P/N_o}\right)$	$\cong \begin{cases} \dfrac{E_{xe}(\rho/BL, \alpha_o, 1)}{\alpha_o} - \rho\left(\dfrac{R}{P/N_o}\right), & BL \geq 10^2 \\ \dfrac{E_{xe}(\rho, \alpha_o, 1)}{\alpha_o} - \rho\left(\dfrac{R}{P/N_o}\right), & BL \leq 10^{-2} \end{cases}$
No Scrambling	$E(R) = \dfrac{E_{xe}\left\{\dfrac{\rho}{K}, \dfrac{\alpha_o Q_{sc}}{K}, 1\right\}}{\alpha_o Q_{sc}/K} - \rho\left(\dfrac{R}{P/N_o}\right)$ (numerical optimization)	$\cong \begin{cases} \left[\dfrac{E_{xe}\left(\dfrac{\rho}{BL}, \alpha_o, 1\right)}{\alpha_o} - \rho\left(\dfrac{R}{P/N_o}\right)\right], & BL \geq 10^2 \\ \left[\dfrac{E_{xe}\left(\rho, \dfrac{\alpha_o}{BL}, 1\right)}{\alpha_o/BL} - \rho\left(\dfrac{R}{P/N_o}\right)\right], & BL \leq 10^{-2} \end{cases}$
Rate Expanding	$E_{re}(R) = \dfrac{Q_{sc}}{Q}\left[\dfrac{E_{xe}\left\{\dfrac{\rho}{K}, \dfrac{\alpha_o Q_{sc}}{K}, 1\right\}}{\alpha_o Q_{sc}/K} - \rho\left(\dfrac{R}{P/N_o}\right)\right]$ (numerical optimization)	$\cong \left[\dfrac{E_{xe}(\rho/BL, \alpha_o, 1)}{\alpha_o} - \rho\left(\dfrac{R}{P/N_o}\right)\right]$, $BL \geq 10^2$

Note: All of the exponents should be maximized on ρ, $\rho \geq 1$.

for all BL.

Without scrambling, the unoptimized exponent is given by (128) with the sc subscripts removed. For this situation, the optimization must in general be numerical, with exceptions when BL is much larger or smaller than unity. In the former case, $Q \cong Q_{sc}$ and we can obtain the scrambling exponent; in the latter case, $Q \cong K/BL$, which implies that K should be minimized, and the signals of (124) are again optimum.

The rate-expanding exponent is given by

$$E_{re}(R) = \max_{\rho \geq 1} \frac{Q_{sc}}{Q} \left[\frac{E_{xe}(\rho/K, \alpha_o Q_{sc}/K, 1)}{\alpha_o Q_{sc}/K} - \rho \left(\frac{R}{P/N_o} \right) \right], \tag{130}$$

and, in this case, the maximization over basic signals is hardest to perform. Except when BL ≫ 1, when the exponent will be the same as both the scrambled and non-scrampled exponents, $E_{re}(R)$ must be evaluated numerically. These results are summarized in Table 5.

b. Random-Coding Bound

For this bound, there is little that we can do except resort to a numerical optimization, and the nonoptimized exponents have already been summarized in Table 2. There is one further computation that may easily be performed, and that is the evaluation of the bound for a particular signal set. For that purpose, we choose the signals of (4.19), which yield reasonably simple results and have proved optimum in many cases. For these signals,

$$Q_{sc} = K = (1+\sqrt{BL})^2 \tag{131}$$

$$Q = \left(1+\sqrt{BL}+\frac{1}{\sqrt{BL}}\right)^2, \tag{132}$$

which may be plugged directly into Table 2. We note the asymptotic forms for large and small BL in Table 6.

4.25 Comments

We have considered application of the results of Section III to the communication schemes outlined in Section II. Parametric exponent-rate curves were presented for basic signals of arbitrary T_s and W_s, and in many cases, the optimum signals could be determined without resorting to numerical measures. The general results appear to be the following.

1. The scrambling scheme is uniformly best, although for BL ≥ 100, all three are about equivalent.

2. When $BL \leq 10^{-2}$, the non-scrambled scheme requires an increase in bandwidth

Table 6. Approximate random-coding exponents for the signals of Eq. 124.

BL	Scheme	Approximate Exponent
$\geqslant 10^2$	All	$\max\limits_{0<\rho\leqslant 1}\left[\frac{E_{oe}(\rho, a_o BL, BL)}{a_o BL} - \rho\left(\frac{R}{P/N_o}\right)\right]$
$\leqslant 10^{-2}$	Scrambling	$E_{sc}(R) = \max\limits_{0<\rho\leqslant 1}\left[\frac{E_{oe}(\rho, a_o, 1)}{a_o} - \rho\left(\frac{R}{P/N_o}\right)\right]$
	No Scrambling	$E(R) = \max\limits_{0<\rho\leqslant 1}\left[\frac{E_{oe}\left(\rho, \frac{a_o}{BL}, 1\right)}{\frac{a_o}{BL}} - \rho\left(\frac{R}{P/N_o}\right)\right]$
	Rate Expanding	$E_{re}(R) = \max\limits_{0<\rho\leqslant 1} BL\left[\frac{E_{oe}(\rho, a_o, 1)}{a_o} - \rho\left(\frac{R}{P/N_o}\right)\right]$

by a factor of $1/BL$ to get the exponent attainable by scrambling.

(3) The rate-expanding scheme appears to lie between the other two in exponent, and has the same capacity as that attainable through scrambling.

(4) For scrambling, basic signals with $T_s = \sqrt{\frac{L}{B}}$, $W_s = \sqrt{\frac{B}{L}}$ appear to be optimum. These signals, with their associated guard spaces, take up less space in the TW plane than any others, indicating that it is better to use many basic signals, each with a small amount of diversity, than to provide more diversity per signal with a corresponding decrease in the coding constraint length.

(5) When scrambling is considered, larger values of BL result in larger error probabilities (with the exception of infinite bandwidth or zero rate operation). Without scrambling, small values of BL are also bad because of the large guard spaces required for independence.

We shall apply some of these results to the computation of numerical examples.

4.3 NUMERICAL EXAMPLES

We shall illustrate how the preceding results can be applied to the estimation of performance levels. As a first example, consider the generation of an exponent-rate curve for the scrambling scheme, used on any channel with a value $BL \leqslant 10^{-2}$. Assume that the basic signals of (124) are used, since they are optimum for the expurgated bound and capacity, and we speculate that they are at least good (and perhaps optimum) for the random-coding bound, too. With these preliminaries, and given P/N_o and W, an exponent-rate curve can be drawn, as discussed in section 4.21. In Fig. 27, we

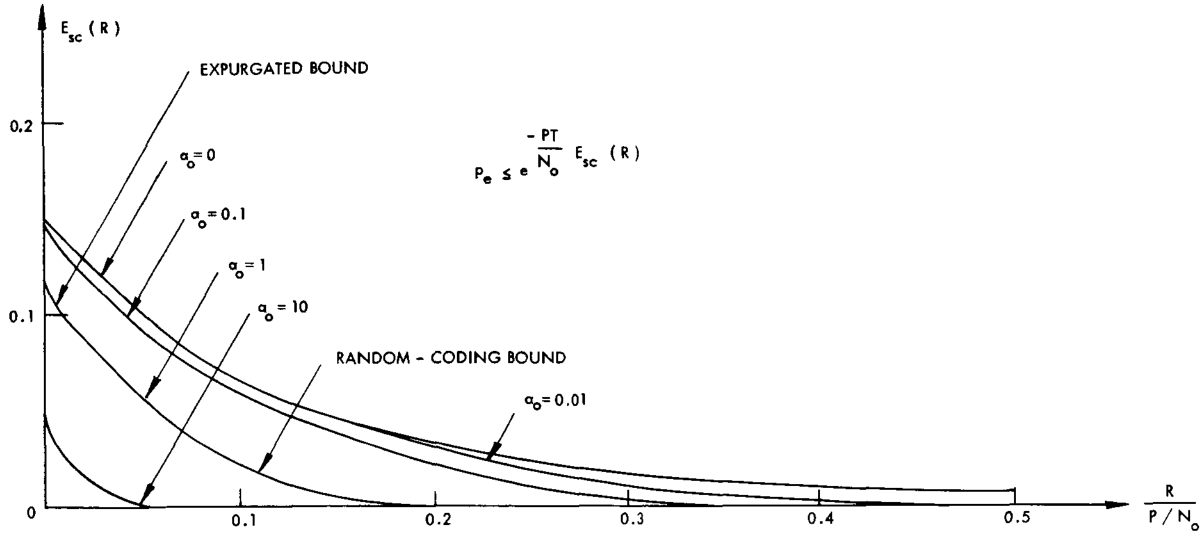

Fig. 27. Sample exponent-rate curves.

present a set of such curves (upper bounds only), parametric on $\alpha_o = P/N_oW$. The rate R (nats/sec) has been normalized by P/N_o, and may be converted to bits/sec by multiplying by 1.44.

Consider applying these curves to communication over a tropospheric scatter channel, with a value of $P/N_o = 10^5$, and $B = 10$ Hz, $L = 10^{-6}$ sec, so that $BL = 10^{-5}$. Each basic signal is then assumed to have $T_s = 300$ μsec, $W_s = 3$ kHz. A value of $\alpha_o = 0.1$ corresponds to $W = 1$ MHz, $\alpha_o = 1$ corresponds to $W_o = 100$ kHz, etc., and $\frac{R}{P/N_o} = 0.1$ on the graph corresponds to a rate of 14.4 kbits/sec. Suppose for example, that we are allowed a bandwidth of $W = 100$ kHz, and code over a constraint length $N = 300$. This corresponds to a block in time of $T = 3$ msec, but remember that we are evaluating the scrambling scheme, so that the 300 basic signals involved in one code word are actually interleaved among many blocks. For a rate of 14.4 kbits/sec, the resulting $P_e \cong e^{-6.6} \cong 10^{-3}$. The capacity of such a signal-channel combination is approximately 28 kbits/sec.

As a second example, we could consider computing the bandwidth required to attain q times the infinite bandwidth exponent $(0 < q < 1)$. Since our data are best at low rates, where we can apply the expurgated bound, we shall use this bound for exponent estimation. When this is done for communication without scrambling over the West Ford orbital dipole belt $(BL = 10^{-1})$, for a value of $P/N_o = 10^4$, the result is as shown in Fig. 28. We note that $\frac{R}{P/N_o} = 0.03$. corresponds to a bit rate of 430 bits/sec (which is indeed a low rate!), but only lack of specific computed data prevents us from extending these curves to any rates desired.

In Fig. 29, we illustrate the effect of an increase in BL on an exponent-rate curve for scrambled signals. All curves are for $\alpha_o = 1$, and the upper curve represents any channel with $BL \le 10^{-2}$, while the others show the effects of increasing

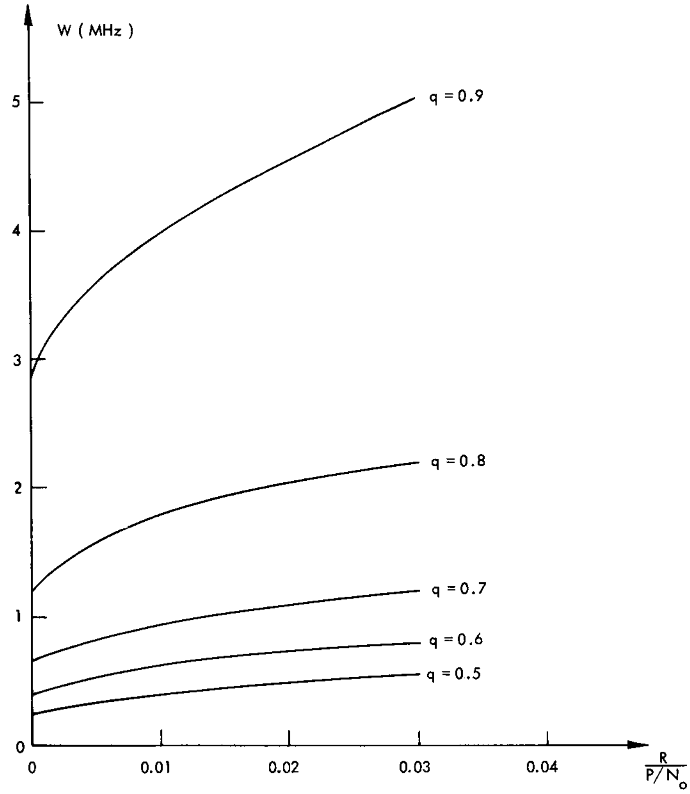

Fig. 28. Bandwidth required for an exponent of q times the infinite-bandwidth exponent.

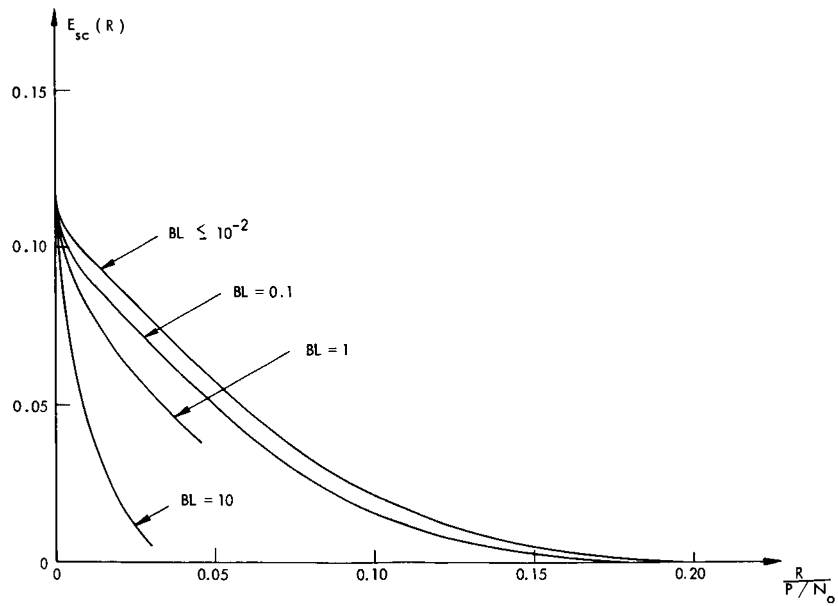

Fig. 29. Change in exponent for a change in BL.

BL in powers of ten. Although the lower two plots are incomplete because of lack of data for higher rates, it is evident that the exponent is greatly reduced as BL increases beyond unity. As remarked earlier, the zero-rate intercept is independent of BL.

In some cases, our general assumption that $\sigma(r, f)$ is a unimodal function is obviously incorrect. In Fig. 30 we show the estimated scattering function for the F layer of the ionosphere, which clearly reveals the presence of three separate paths. Each piece of the scattering function is approximately described by B = 1 Hz and $L_1 = 10^{-4}$ sec, but the total function takes up B = 1 Hz and $L = 5 \times 10^{-4}$ sec.

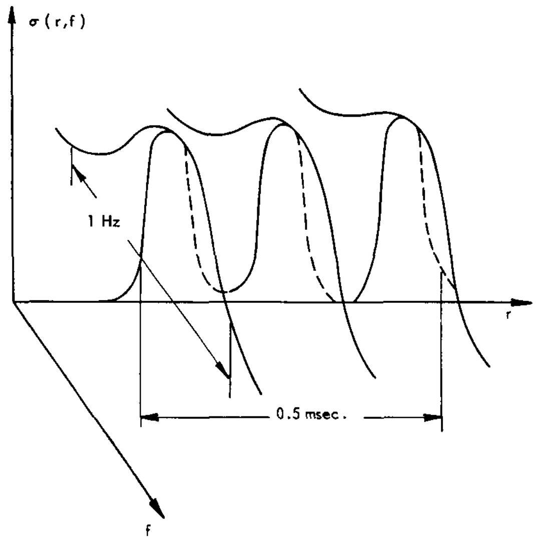

Fig. 30. Typical scattering function for HF propagation.

If a basic signal has a bandwidth greater than the reciprocal of the range difference between adjacent paths ($W_s \gtrsim 10^4$ Hz), the paths will be resolvable; that is, the contributions to the total received process from each of the three paths can be approximately separated. Since each path is made up of a different group of scatterers, these three received components should be independent, and the resulting signal diversity will be approximately

$$K_1 = 3(1+BT_s)(1+L_1W_s). \tag{133}$$

For orthogonal output signals we still require guard spaces of B and L, however. To obtain independent output signals, additional guard spaces of approximately 1/B in time and $1/L_1$ in frequency are required, owing to the path resolvability. When $W_s \lesssim 10^4$ Hz, the paths will not be resolvable, and the scattering function should have an effect comparable to that of a unimodal scattering function described by B and L.

Consider the computation of the expurgated bound for this channel. We note that BL is much less than unity, so we can apply our previous asymptotic results.

For the case $W_s < 10^4$ Hz, we can apply the results of Table 5 directly. Using the signals of Eq. 124 results in a value $W_s = 50$ Hz, well under the limit of 10^4 Hz, so that

$$E_{sc}(R) = \max_{\rho\geq 1}\left[\frac{E_{xe}(\rho, \alpha_o, 1)}{\alpha_o} - \rho\left(\frac{R}{P/N_o}\right)\right] \tag{134}$$

$$E(R) = \max_{\rho\geq 1}\left[\frac{E_{xe}\left(\rho, \frac{\alpha_o}{BL}, 1\right)}{\frac{\alpha_o}{BL}} - \rho\left(\frac{R}{P/N_o}\right)\right]. \tag{135}$$

When $W_s > 10^4$ Hz and scrambling is used, our previous discussion implies

$$E_{sc}(R) = \max_{\rho\geq 1}\left[\frac{E_{xe}\left\{\frac{\rho}{3(1+BT_s)(1+L_1W_s)}, \beta_{sc}, 1\right\}}{\beta_{sc}} - \rho\left(\frac{R}{P/N_o}\right)\right] \tag{136}$$

$$\beta_{sc} = \frac{\alpha_o(T_s+L)(W_s+B)}{3(1+BT_s)(1+L_1W_s)}, \tag{137}$$

and now numerical optimization is required to determine the best signals. Indications are that, within this resolvability framework, the basic signals should be made as much like those of (124) as possible, so we consider evaluation of (137) with $W_s = 10^4$, $T_s = 10^{-4}$. In this case,

$$E_{sc}(R) \cong \max_{\rho\geq 1}\left[\frac{E_{xe}\left(\frac{\rho}{6}, \alpha_o, 1\right)}{\alpha_o} - \rho\left(\frac{R}{P/N_o}\right)\right]. \tag{138}$$

Without scrambling,

$$E(R) = \max_{\rho\geq 1}\left[\frac{E_{xe}\left\{\frac{\rho}{3(1+BT_s)(1+L_1W_s)}, \beta, 1\right\}}{\beta} - \rho\left(\frac{R}{P/N_o}\right)\right] \tag{139}$$

$$\beta = \frac{\alpha_o(T_s+L+1/B)(W_s+B+1/L_1)}{3(1+BT_s)(1+L_1W_s)}. \tag{140}$$

Evaluation of this scheme with the same basic signals leads to

$$E(R) \cong \max_{\rho\geq 1}\left[\frac{E_{xe}\left\{\frac{\rho}{6}, \frac{\alpha_o}{6}\times 10^5, 1\right\}}{\frac{\alpha_o}{6}\times 10^5} - \rho\left(\frac{R}{P/N_o}\right)\right]. \tag{141}$$

In this case, both with and without scrambling, a smaller exponent is obtained by making the paths resolvable. This resolvability results in a smaller amount of diversity for a given input signal, but the large bandwidth required by a basic signal to obtain resolvability excessively decreases the number of basic signals available for coding.

4.4 SIGNAL DESIGN

Up to this point, all results have been for the case of basic signals that are just time and frequency translates of one another, and so are identical for the purposes of our analysis. In Section II, we remarked that, before evaluation of performance estimates, there was no way of determining whether this choice of a basic signaling set was a good one or not, although we speculated that such was the case. We now return to that question, and show that, in one situation at least, this choice of basic signals is optimum.

Since the results in Section III are only valid for equal-eigenvalue basic signals, we must restrict the analysis to that case, but we now allow different basic signals to have different orders of diversity, K, obtained by using different values of T_s and W_s on different basic signals. To make matters as simple as possible, we consider only the scrambling scheme, and use the expurgated bound for exponent estimation.

Let the n^{th} basic signal have time duration T_n and bandwidth W_n. If we allow the usual guard spaces of B and L, each basic signal may be considered to use an area

$$Q_n = (T_n+L)(W_n+B) \tag{142}$$

in the TW plane, with a diversity

$$K_n = \begin{cases} Q_n, & BL \geq 1 \\ (1+BT_n)(1+LW_n), & BL < 1 \end{cases} \tag{143}$$

If unimportant end effects and integer constraints are neglected, the bandwidth constraint may be expressed as

$$\sum_{n=1}^{N} Q_n = TW. \tag{144}$$

It may be shown, in a derivation similar to Ebert's, that

$$\frac{PT}{N_o} E_{sc}(R) = \max_{\rho\geq 1} \left\{ \sum_{n=1}^{N} - \rho \ln \left[\int_0^\infty \int_0^\infty p_n(x)\, p_n(x_1)\, e^{r\left(x^2+x_1^2-2S_n\right)} H_{K_n}(x,x_1)^{1/\rho}\, dx dx_1 \right] - \rho \left(\frac{R}{P/N_o} \right) \right\}, \tag{145}$$

where $S_n = \int_0^\infty x^2 p_n(x)\, dx$, and $H_{K_n}(x,x_1)$ is defined in Eq. 49. The right-hand side of

(145) is to be maximized on $\{p_n(x), r, T_n, W_n\}$, subject to (144) and the additional constraints that $p_n(x)$ be a probability function and

$$\sum_{n=1}^{N} S_n = \frac{PT}{N_o}, \quad T_nW_n \geq 1, \quad r \geq 0. \tag{146}$$

The maximum can obviously be increased by allowing r to vary with n, subject to the constraints $r_n \geq 0$. In that case

$$\frac{PT}{N_o} E_{sc}(R) \leq \max_{\rho \geq 1} \left[\sum_{n=1}^{N} E_{xe}(\rho, S_n, K_n) - \rho \left(\frac{R}{P/N_o} \right) \right]. \tag{147}$$

Equality will hold in (147) if S_n and K_n are independent of n, since the same value of r is then optimum for all n. For a given $\{S_n, K_n\}$, Q_n enters only in the determination of N, the number of basic signals. The larger N, the larger the exponent, so given K_n on any particular signal, Q_n should be minimized. This is achieved by choosing $T_nW_n = 1$, in which case $Q_n = K_n$. The problem has now been reduced to

$$\max_{K_n, S_n} \sum_{n=1}^{N} E_{xe}(\rho, S_n, K_n),$$

subject to the constraints

$$\sum_{n=1}^{N} K_n = TW, \quad \sum_{n=1}^{N} S_n = \frac{PT}{N_o}, \quad S_n \geq 0, \quad K_n \geq (1+\sqrt{BL})^2. \tag{148}$$

The constraint on K_n may be changed to $K_n \geq 0$, and if the solution to this new problem has $K_n \geq (1+\sqrt{BL})^2$, then it will also be a solution to the problem of interest.

In Theorem B. 5 (Appendix B) it is shown that $E_{xe}(\infty, S, K)$ is a jointly concave function of S and K. For this situation, the maximization conditions have been given by the Kuhn-Tucker theorem,[29] in which case it is easily shown that, given N, $K_n = TW/N$ and $S_n = PT/N_oN$ provide the maximum. Thus, for scrambling, identical basic signals are optimum at zero rate because the inequality of (147) becomes an equality for identical basic signals. The corresponding exponent is

$$E_{sc}(0) = \frac{1}{a_o} E_{xe}(\infty, a_o, 1) \tag{149}$$

for any set of identical basic signals with $T_sW_s = 1$, which agrees with a previous result.

When $\rho < \infty$, recall, from Section III, that a typical $E_{xe}(\rho, S, K)$ appears as sketched in Fig. 31, thereby illustrating the fact that $E_{xe}(\rho, S, K)$ is not even concave in K, much less jointly concave in K and S. For specified ρ and S, however, an optimum basic signal can never have a value of K greater than K_o, for then K could be reduced to K_o, which decreases the signal area, and allows a simultaneous increase in exponent. Thus it is possible that, for the range of K that is of interest, the function could be jointly concave, but this remains speculative, and the optimum $\{S_n, K_n\}$ is still unknown.

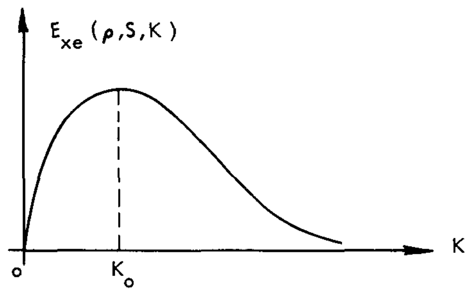

Fig. 31. Sketch of a typical $E_{xe}(\rho, S, K)$ versus K.

The preceding discussion was just meant to provide additional justification for our belief that identical basic signals provide the greatest exponent. Even at zero rate, however, we can make no conclusive statements, for two reasons:

1. We do not have an expression for the true exponent (the upper and lower bounds do not exponentially agree at zero rate).
2. We cannot evaluate exponents for arbitrary $\underline{\lambda}$.

4.41 One-Dimensional Signaling*

All of our previous work was done under the assumption that basic signals were spaced both in time and frequency. We now consider signals designed so that either $W_s = W$ or $T_s = T$, thereby eliminating the need for guard spaces in one dimension. Note that this cannot be obtained from our previous analysis because all signals were assumed to have attached guard spaces. This is intuitively attractive if, for example, B or L is very large or small, in which case a considerable amount of guard space might be eliminated without scrambling. We shall now show that little improvement is to be expected from such methods.

Consider first, basic signals that take up the entire available bandwidth, $W_s = W$. To avoid a lengthy discussion, we only consider zero-rate signaling. With scrambling, we find

*This section is tangential to the main discussion and may be skipped by the casual reader.

$$E_{sc}(0) = \frac{1}{\beta_{sc}} E_{xe}(\infty, \beta_{sc}, 1) \tag{150}$$

$$\beta_{sc} = \alpha_o \begin{cases} \dfrac{1}{1 + B/W} & BL \geq 1 \\ \dfrac{T_s + L}{(1+BT_s)(L+1/W)} & BL < 1 \end{cases} \tag{151}$$

To maximize the exponent, T_s should be minimized, so that $T_s = 1/W$ and

$$\beta_{sc} = \frac{\alpha_o}{1 + B/W}, \qquad \text{all BL.} \tag{152}$$

Thus at zero rate, for scrambled signals, we have improved results for all BL, although if B/W is small, as it often is, the improvement is negligible. For $R > 0$, a numerical optimization over T_s is necessary, and this type of signal design, depending on the parameters, may result either in improvement or degradation in performance.

Without scrambling, the comparable results are

$$E(0) = \frac{1}{\beta} E_{xe}(\infty, \beta, 1) \tag{153}$$

$$\beta = \alpha_o \begin{cases} \dfrac{1}{1 + B/W}\left[1 + \dfrac{1}{BT_s + BL}\right] & BL \geq 1 \\ \dfrac{1}{BL(1+1/LW)}\left[1 + \dfrac{BL}{BT_s + 1}\right] & BL < 1. \end{cases} \tag{154}$$

In each case, T_s should be large, with the result that

$$\beta = \alpha_o \begin{cases} \dfrac{1}{1 + B/W} & BL \geq 1 \\ \dfrac{1}{BL}\left[\dfrac{LW}{1+LW}\right] & BL < 1. \end{cases} \tag{155}$$

Once again, we find uniform improvement, which will be significant if B/W or 1/LW is at least comparable to unity, the latter of which might be reasonable for some tropospheric scatter channels. Away from zero rate, the same comments hold as for scrambling.

We can also consider the analogous situation, when $T_s = T$. In this case, at zero rate, after optimization on W_s, we find

$$\beta_{sc} = \frac{\alpha_o}{1 + L/T} \tag{156}$$

$$\beta = \alpha_o \begin{cases} \dfrac{1}{1 + L/T} & BL \geq 1 \\ \dfrac{1}{BL}\left[\dfrac{1}{1 + 1/BT}\right] & BL < 1. \end{cases} \tag{157}$$

Once again, we have an apparent increase in exponent with this type of signal design, which becomes substantial when L/T or 1/BT is unity or larger. We stress the word apparent for several reasons. First, consideration of scrambling implies sequential (in time) transmission of blocks of signals, between which we require a separation of L sec, or more, to ensure orthogonality. This has been ignored in the discussion, and inclusion of such guard spaces brings us back to the normal scrambled-signals analysis.

When considering the "one-shot" channel usage without scrambling, this difficulty does not occur, but a more fundamental question arises. In particular, E(0) is maximized by $T \to 0$, and results in the infinite-bandwidth exponent. But, for the determination of P_e, E(0) must be multiplied by T, so that in this case the resulting error probability goes to unity. This scheme may result in a substantial improvement in P_e only if there is some T comparable to (or smaller than) L or 1/B, which results in reasonable error probabilities. The exponent E(0) will be valid only for that T, and as T is increased E(0) will decrease until it approaches the lower bound given by the original signaling scheme.

Also note that the bounds used here are tight only as $N \to \infty$, and with the signals just proposed, $T \to \infty$ no longer implies $N \to \infty$, so that the bounds may not give a good indication about the true exponential behavior. For $R > 0$, the situation is again numerical.

V. SUMMARY AND CONCLUSIONS

We shall briefly summarize the major points of this research, give some conclusions about estimation of performance levels, and mention some possibilites for further work on the problem.

5.1 SUMMARY

Our central problem concerned digital communication over fading-dispersive channels, subject to a bandwidth constraint on the input signals. The channel model considered here was an extremely approximate one, yet it reflects the major channel characteristics. We considered coding over a specific set of basic input signals, namely one that resulted in independent and orthogonal outputs, a choice that reduced the problem to communication over a diversity channel. Within this framework, several different signaling schemes have been outlined.

We derived upper and lower bounds to P_e, the minimum attainable error probability, which were exponentially decreasing in the signal duration, T. These bounds agreed in exponent for a range of rates in the limit of large T; thus, they give the true asymptotic behavior of this signal-channel combination for that range of rates. We found that the signals that optimized the bounds had basic signal modulations chosen from a finite set of discrete voltage levels. This result was used to evaluate parametric exponent-rate curves in terms of the important signal and channel characteristics. We investigated the interactions of the channel with the signals and their effects upon over-all system performance.

Our results were found to agree with the previously known infinite-bandwidth, orthogonal-signal exponents in the limit as available bandwidth goes to infinity. For finite bandwidths, we illustrated the manner in which the present results might be used in the estimation of system performance, and we considered the question of design of optimum basic signals. The major stumbling block was that, even with as simple a characterization of the channel as the one used here, the problem is basically numerical, making absolute statements about the nature of parameter interactions impossible, although we can evaluate enough special cases to get some good indications.

5.2 CONCLUSIONS

The difficulties encountered in evaluation of the bounds to P_e make it impossible to draw conclusions of great generality, but some trends of channel behavior and performance may be stated.

1. With the methods and results presented here, it is possible to make some rough performance estimates for coding over reasonable signaling schemes. Also, these estimates provide ultimate performance limits for coding over many commonly used diversity schemes.

2. For this type of signaling, except at zero rate or with an infinite available bandwidth, channels with $BL > 1$ are generally inferior to those with smaller values

of BL. Unless something like scrambling is done, channels with BL « 1 will also give poor performance, because of the large guard spaces required to ensure independence between signals.

3. For BL > 1, there is little difference between scrambled and nonscrambled signals, while for BL < 1, nonscrambled signals require a bandwidth increase of a factor of approximately 1/BL to obtain a value of P_e equal to that attainable by scrambling.

We close with the speculation that, if one is interested in "one-shot" channel use, that is, the performance available within a block of time T and bandwidth W, the rate-expanding scheme proposed at the end of Section II may provide a reasonable estimate because it does not suffer quite as much from independence requirements as the scheme that was originally proposed. Unfortunately, it is also the most difficult to evaluate. Of course, if many blocks are used, and it is feasible to interleave signals among many blocks, the scrambling exponent will be better, as well as the easiest to evaluate.

5.3 FUTURE WORK

There are openings for more work on this problem along two fronts: analytical and numerical. In the analytical category, the bound presented here, relating performance of channels with arbitrary eigenvalues to that of equal eigenvalue channels, is valid only for low rates. A similar bound may be derivable for the random-coding case, too, although some time was spent in an unsuccessful search for one. Furthermore, we suspect that equal-eigenvalue basic signals are optimum, and it might be possible to prove this. Last, although it is in some respects a singular case, zero rate is by far the easiest point to evaluate, and it would be worth while to determine the true exponent for this rate. Since an improved zero-rate lower bound that agrees with the upper bound can in some cases by derived, the same may be true here.

Concerning further numerical work, there are two separate problems to be considered. For the expurgated bound, we obtained many numerical results, and any others desired can be easily computed, as long as the eigenvalues are equal. Even for nonequal eigenvalues, because of the quadratic nature of the problem, it should be possible to obtain results equivalent to those in Section III, provided there is only a small number of significant λ_k. The only possible trouble may arise because, with nonequal eigenvalues, an impulsive p(x) may no longer be optimum, and a large number of impulses may be necessary to approximate a continuous p(x). If the basic signals were specified, so that T_s and W_s could be approximately determined (T_s and W_s cannot be obtained from a specification of $\underline{\lambda}$ alone), these parameters could then be used for evaluation of system performance, exactly as in Section IV for equal-eigenvalue systems.

For the random-coding bound, without the development of some new optimization techniques, the number of numerical integrations necessary for arbitrary $\underline{\lambda}$ makes it doubtful that results can be obtained (indeed it was difficult enough for equal eigenvalues). With enough computer time, and perhaps a more specialized optimization program, it should be possible to generate a fairly complete set of curves for equal eigenvalues.

APPENDIX A

Conditions Sufficient for Exponent Maximization

THEOREM A. 1

A sufficient condition on p(x), r to minimize

$$I_x[r, p(x)] = \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(-2a+x^2+x_1^2\right)} H_{\underline{\lambda}}(x, x_1)^{1/\rho}\, dx dx_1 \tag{A. 1}$$

$$H_{\underline{\lambda}}(x, x_1) = \prod_{k=1}^{\infty} \frac{\left(1+\lambda_k x^2\right)^{1/2} \left(1+\lambda_k x_1^2\right)^{1/2}}{1 + \frac{1}{2}\lambda_k\left(x^2+x_1^2\right)}, \tag{A. 2}$$

subject to the constraints

$$r \geq 0, \quad p(x) \geq 0, \quad \int_0^\infty p(x)\, dx = 1, \quad \int_0^\infty x^2 p(x)\, dx = a \tag{A. 3}$$

is

$$\int_0^\infty p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_{\underline{\lambda}}(x, x_1)^{1/\rho}\, dx_1 \geq \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_{\underline{\lambda}}(x, x_1)^{1/\rho}\, dx dx_1 \tag{A. 4}$$

for all x, with equality when $p(x) > 0$, for any $0 < \rho < \infty$.

Proof: Assume that p(x), r satisfy constraints (A. 3) and condition (A. 4), and consider any r_1, $p_1(x)$ that satisfy constraints (A. 3).

$$I_x[r_1, p_1(x)] - I_x[r, p(x)] = \int_0^\infty \int_0^\infty H_{\underline{\lambda}}(x, x_1)^{1/\rho}$$

$$\left[p_1(x)\, p_1(x_1)\, e^{r_1\left(-2a+x^2+x_1^2\right)} - p(x)\, p(x_1)\, e^{r\left(-2a+x^2+x_1^2\right)}\right] dx dx_1. \tag{A. 5}$$

This may be rewritten

$$I_x[r_1, p_1(x)] - I_x[r, p(x)] = \int_0^\infty \int_0^\infty H_{\underline{\lambda}}(x, x_1)^{1/\rho}$$

$$\left[\left\{p_1(x)\, e^{r_1(-a+x^2)} - p(x)\, e^{r_1(-a+x^2)}\right\}\left\{p_1(x_1)\, e^{r_1\left(-a+x_1^2\right)} - p(x_1)\, e^{r(-a+x^2)}\right\}\right.$$

$$+ p(x)\, p_1(x_1)\, e^{r(-a+x^2)+r_1\left(-a+x_1^2\right)} + p_1(x)\, p(x_1)\, e^{r_1(-a+x^2)+r\left(-a+x_1^2\right)}$$

$$\left. - 2p(x)\, p(x_1)\, e^{r\left(-2a+x^2+x_1^2\right)}\right] dx dx_1. \tag{A. 6}$$

At this point, define

$$q_{\underline{\lambda}}(\underline{y}\,|x) \equiv \prod_{k=1}^{\infty} \frac{y_k^{\frac{1}{\rho}-1} e^{-\frac{y_k}{1+\lambda_k x^2}}}{\Gamma(1/\rho)\left(1+\lambda_k x^2\right)^{1/\rho}}. \tag{A. 7}$$

It can be easily verified that

$$H_{\underline{\lambda}}(x, x_1)^{1/\rho} = \int_0^{\infty} q_{\underline{\lambda}}(\underline{y}\,|x)^{1/2}\, q_{\underline{\lambda}}(\underline{y}\,|x_1)^{1/2}\, d\underline{y}, \tag{A. 8}$$

and hence

$$\int_0^{\infty}\int_0^{\infty} f(x)\, f(x_1)\, H_{\underline{\lambda}}(x, x_1)^{1/\rho}\, dx dx_1 = \int_0^{\infty}\left[\int_0^{\infty} f(x)\, q_{\underline{\lambda}}(\underline{y}\,|x)^{1/2}\, dx\right]^2 d\underline{y} \geq 0. \tag{A. 9}$$

Using this result, with $f(x) = p_1(x)\, e^{r_1(-a+x^2)} - p(x)\, e^{r(-a+x^2)}$, together with the symmetry properties of $H_{\underline{\lambda}}(x, x_1)$, we have

$$I_x[r_1, p_1(x)] - I_x[r, p(x)] \geq$$

$$2\left[\int_0^{\infty} p_1(x)\, e^{(r_1-r)(-a+x^2)-2ra} \int_0^{\infty} p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_{\underline{\lambda}}(x, x_1)^{1/\rho}\, dx_1 dx - I_x[r, p(x)]\right]. \tag{A. 10}$$

Since $p_1(x)\, e^{(r_1-r)(-a+x^2)-2ra} \geq 0$, and (A. 4) is satisfied

$$I_x[r_1, p_1(x)] - I_x[r, p(x)] \geq$$

$$2\left[\int_0^{\infty} p_1(x)\, e^{(r_1-r)(-a+x^2)}\, I_x[r, p(x)]\, dx - I_x[r, p(x)]\right] \tag{A. 11}$$

$$I_x[r_1, p_1(x)] - I_x[r, p(x)] \geq$$

$$2 I_x[r, p(x)] \int_0^{\infty} p_1(x)\left[e^{(r_1-r)(-a+x^2)} - 1\right] dx. \tag{A. 12}$$

Note that $p_1(x) \geq 0$, $I_x[r, p(x)] > 0$, $e^{(r_1-r)(-a+x^2)} \geq 1 + (r_1-r)(-a+x^2)$, so

$$I_x[r_1, p_1(x)] - I_x[r, p(x)] \geq 2 I_x[r, p(x)] \int_0^{\infty} p_1(x)(r_1-r)(-a+x^2)\, dx = 0 \tag{A. 13}$$

because $p_1(x)$ satisfies constraints (A. 3). This completes the proof.

THEOREM A. 2

A sufficient condition for p(x) to minimize

$$I_\infty[p(x)] = \int_0^\infty \int_0^\infty p(x)\, p(x_1) \ln H_{\underline{\lambda}}(x, x_1)\, dx dx_1, \tag{A. 14}$$

subject to conditions (A. 3) is

$$\int_0^\infty p(x_1) \ln H_{\underline{\lambda}}(x, x_1)\, dx_1 \geqslant \int_0^\infty \int_0^\infty p(x)\, p(x_1) \ln H_{\underline{\lambda}}(x, x_1)\, dx dx_1 + \lambda(x^2 - \alpha) \tag{A. 15}$$

for all x and some λ.

Proof: Assume that p(x) satisfies (A. 3) and (A. 15), and $p_1(x)$ satisfies (A. 3).

$$I_\infty[p_1(x)] - I_\infty[p(x)] = \int_0^\infty \int_0^\infty \ln H_{\underline{\lambda}}(x, x_1) \{p_1(x) - p(x)\}\{p_1(x_1) - p(x_1)\}\, dx dx_1$$

$$+ 2 \int_0^\infty \int_0^\infty \ln H_{\underline{\lambda}}(x, x_1) \{p_1(x) p(x_1) - p(x) p(x_1)\}\, dx dx_1. \tag{A. 16}$$

We shall now show that the first integral is non-negative.

$$\ln H_{\underline{\lambda}}(x, x_1) = \sum_{k=1}^{\infty} \left[\frac{1}{2}\ln\left(1+\lambda_k x^2\right) + \frac{1}{2}\ln\left(1+\lambda_k x_1^2\right) - \ln\left\{1 + \frac{1}{2}\lambda_k\left(x^2 + x_1^2\right)\right\}\right] \tag{A. 17}$$

$$\ln H_{\underline{\lambda}}(x, x_1) = \sum_{k=1}^{\infty} \left[\ln\left(1 + \frac{1}{2}\lambda_k x^2\right) + \ln\left(1 + \frac{1}{2}\lambda_k x_1^2\right) - \ln\left\{1 + \frac{1}{2}\lambda_k\left(x^2 + x_1^2\right)\right\}\right.$$

$$\left. + \frac{1}{2}\ln\left(1+\lambda_k x^2\right) - \ln\left(1 + \frac{1}{2}\lambda_k x^2\right) + \frac{1}{2}\ln\left(1+\lambda_k x_1^2\right) - \ln\left(1+\lambda_k x_1^2\right)\right] \tag{A. 18}$$

$$\ln H_{\underline{\lambda}}(x, x_1) = \sum_{k=1}^{\infty} \left\{ \ln\left[\frac{\left(1+\frac{1}{2}\lambda_k x^2\right)\left(1 + \frac{1}{2}\lambda_k x_1^2\right)}{1 + \frac{1}{2}\lambda_k\left(x^2 + x_1^2\right)}\right] + \ln\left[\frac{\left(1+\lambda_k x^2\right)^{1/2}}{1 + \frac{1}{2}\lambda_k x^2}\right] + \ln\left[\frac{\left(1+\lambda_k x_1^2\right)^{1/2}}{1 + \frac{1}{2}\lambda_k x_1^2}\right]\right\}. \tag{A. 19}$$

Let $f(x) = p_1(x) - p(x)$, and then

$$\int_0^\infty \int_0^\infty f(x)\, f(x_1) \ln H_{\underline{\lambda}}(x, x_1)\, dx dx_1 = \sum_{k=1}^{\infty} \int_0^\infty \int_0^\infty f(x)\, f(x_1) \ln\left[\frac{\left(1+\frac{1}{2}\lambda_k x^2\right)\left(1+\frac{1}{2}\lambda_k x_1^2\right)}{1 + \frac{1}{2}\lambda_k\left(x^2 + x_1^2\right)}\right] dx dx_1$$

$$+ 2 \sum_{k=1}^{\infty} \int_0^\infty f(x_1)\, dx_1 \int_0^\infty f(x) \ln\left[\frac{\left(1+\lambda_k x^2\right)^{1/2}}{1 + \frac{1}{2}\lambda_k x^2}\right] dx. \tag{A. 20}$$

But $\int_0^\infty f(x_1)\,dx_1 = 0$, and

$$\left|\int_0^\infty f(x)\ln\left[\frac{\left(1+\lambda_k x^2\right)^{1/2}}{1+\frac{1}{2}\lambda_k x^2}\right]dx\right| \leq \int_0^\infty \left|p_1(x)-p(x)\right| \left|\ln\left[\frac{\left(1+\lambda_k x^2\right)^{1/2}}{1+\frac{1}{2}\lambda_k x^2}\right]\right| dx. \tag{A. 21}$$

$$\left|p_1(x)-p(x)\right| \leq p_1(x)+p(x). \tag{A. 22}$$

$$\left|\ln\left[\frac{\left(1+\lambda_k x^2\right)^{1/2}}{1+\frac{1}{2}\lambda_k x^2}\right]\right| = \ln\left(1+\frac{1}{2}\lambda_k x^2\right)-\frac{1}{2}\ln\left(1+\lambda_k x^2\right) \leq \frac{1}{2}\lambda_k x^2. \tag{A. 23}$$

$$\sum_{k=1}^\infty \left|\int_0^\infty f(x)\ln\left[\frac{\left(1+\lambda_k x^2\right)^{1/2}}{1+\frac{1}{2}\lambda_k x^2}\right]dx\right| \leq$$

$$\sum_{k=1}^\infty \int_0^\infty \left(\frac{1}{2}\lambda_k x^2\right)\left[p_1(x)+p(x)\right]dx = \sum_{k=1}^\infty \lambda_k a = a < \infty. \tag{A. 24}$$

Thus the second term in (A. 20) is zero, and

$$\int_0^\infty \int_0^\infty f(x)\,f(x_1)\ln H_{\underline{\lambda}}(x,x_1)\,dx\,dx_1 =$$

$$\sum_{k=1}^\infty \int_0^\infty \int_0^\infty f(x)\,f(x_1)\ln\left[\frac{\left(1+\frac{1}{2}\lambda_k x^2\right)\left(1+\frac{1}{2}\lambda_k x_1^2\right)}{1+\frac{1}{2}\lambda_k\left(x^2+x_1^2\right)}\right]dx\,dx_1. \tag{A. 25}$$

It has been shown[31] that

$$\ln\left[\frac{\left(1+\frac{1}{2}a\right)\left(1+\frac{1}{2}a_1\right)}{1+\frac{1}{2}(a+a_1)}\right] = \int_0^1 \frac{\left(t^{\frac{1}{2}a}-1\right)\left(t^{\frac{1}{2}a_1}-1\right)}{\ln\left(\frac{1}{t}\right)}\,dt \tag{A. 26}$$

when a and a_1 are non-negative, hence

$$\ln\left[\frac{\left(1+\frac{1}{2}a\right)\left(1+\frac{1}{2}a_1\right)}{1+\frac{1}{2}(a+a_1)}\right] = \int_0^1 g(a,t)\,g(a_1,t)\,dt, \tag{A. 27}$$

where

$$g(a,t) = \frac{t^{\frac{1}{2}a} - 1}{\sqrt{\ln(1/t)}}. \tag{A. 28}$$

Therefore

$$\int_0^\infty \int_0^\infty f(x)\, f(x_1) \ln H_{\underline{\lambda}}(x, x_1)\, dx dx_1 = \sum_{k=1}^\infty \int_0^1 \left[\int_0^\infty f(x)\, g\left[\lambda_k x^2, t\right] dx\right]^2 dt \geq 0. \tag{A. 29}$$

Therefore

$$I_\infty[p_1(x)] - I_\infty[p(x)] \geq 2 \int_0^\infty p_1(x) \int_0^\infty p(x_1) \ln H_{\underline{\lambda}}(x, x_1)\, dx_1 dx - 2I_\infty[p(x)]. \tag{A. 30}$$

But $p_1(x) \geq 0$, and $p(x)$ satisfies (A. 15) by hypothesis, so

$$I_\infty[p_1(x)] - I_\infty[p(x)] \geq 2\left[\int_0^\infty p_1(x)\left\{I_\infty[p(x)] + \lambda(x^2 - a)\right\} dx - I_\infty[p(x)]\right] \tag{A. 31}$$

or $I_\infty[p_1(x)] - I_\infty[p(x)] \geq 0$, thereby proving the theorem.

THEOREM A. 3

A sufficient condition on $p(x)$, r for minimization of

$$I_0[r, p(x)] = \int_0^\infty \left[\int_0^\infty p(x)\, e^{r(x^2 - a)}\, p(y|x)^{1-s}\, dx\right]^{1/(1-s)} dy \tag{A. 32}$$

when $p(x)$ and r are subject to the constraints (A. 3) is the following:

$$\int_0^\infty \beta(y)^{s/(1-s)}\, p(y|x)^{1-s}\, e^{rx^2}\, dx \geq \int_0^\infty \beta(y)^{1/(1-s)}\, dy \tag{A. 33}$$

for all x, where $0 < s < 1$, and

$$\beta(y) \equiv \int_0^\infty p(x)\, e^{rx^2}\, p(y|x)^{1-s}\, dx. \tag{A. 34}$$

Recall that

$$p(y|x) = \frac{y^{K-1}\, e^{-y/(1+x^2)}}{\Gamma(K)(1+x^2)^K},$$

and when $p(y|x)$ is normalized in this fashion, a should be replaced by a/K in constraints (A. 3). At this point, however, a is arbitrary, so we shall leave (A. 3) unchanged, and make any necessary alterations later on.

Proof: Let r, $p(x)$ satisfy (A. 3) and (A. 33), and consider any r_1, $p_1(x)$ that satisfy (A. 3). We shall show that

$$I_o[r_1, p_1(x)] - I_o[r, p(x)] \geqslant 0. \qquad (A.35)$$

It has been shown[32] that, if a > 0, then

$$\overline{t^a} \lesseqgtr \bar{t}^a, \qquad a \lesseqgtr 1. \qquad (A.36)$$

Thus, if $0 < \lambda < 1$,

$$\int_0^\infty \left\{(1-\lambda) \int_0^\infty p(x)\, e^{r(x^2-\alpha)}\, p(y|x)^{1-s}\, dx + \lambda \int_0^\infty p_1(x)\, e^{r_1(x^2-\alpha)}\, p(y|x)^{1-s}\, dx\right\}^{1/(1-s)} dy \leqslant$$

$$(1-\lambda)\, I_o[r, p(x)] + \lambda I_o[r_1, p_1(x)] \qquad (A.37)$$

or

$$I_o[r_1, p_1(x)] - I_o[r, p(x)] \geqslant$$

$$\frac{1}{\lambda}\left[-I_o[r, p(x)] + \int_0^\infty \left\{(1-\lambda) \int_0^\infty p(x)\, e^{r(x^2-\alpha)}\, p(y|x)^{1-s}\, dx + \lambda \int_0^\infty p_1(x)\, e^{r_1(x^2-\alpha)}\, p(y|x)^{1-s}\, dx\right\}^{1/(1-s)} dy\right]. \qquad (A.38)$$

Recall that

$$p(y|x) = \frac{y^{K-1}\, e^{-y/(1+x^2)}}{\Gamma(K)(1+x^2)^K},$$

so that $\left\{\frac{y^{K-1}}{\Gamma(K)}\right\}^{1-s}$ may be factored out of the integrals on x. Also $\int_0^\infty p(x)(1+x^2)^{-K(1-s)}\, e^{r(x^2-\alpha) - (1-s)y/(1+x^2)}\, dx > 0$, all y, so that

$$I_o[r_1, p_1(x)] - I_o[r, p(x)] \geqslant$$

$$\frac{1}{\lambda}\left[-I_o[r, p(x)] + \int_0^\infty [\beta(y)\, e^{-r\alpha}]^{1/(1-s)} \left\{1 - \lambda + \lambda \frac{\int_0^\infty p_1(x)\, e^{r_1(x^2-\alpha)}\, p(y|x)^{1-s}\, dx}{\int_0^\infty p(x)\, e^{r(x^2-\alpha)}\, p(y|x)^{1-s}\, dx}\right\}^{1/(1-s)} dy\right]. \qquad (A.39)$$

Since $0 < \lambda < 1$, we may apply the inequality

$$(1+t)^{1/(1-s)} \geqslant 1 + \frac{t}{1-s}, \quad -1 \leqslant t, \quad 0 < s < 1 \tag{A. 40}$$

to obtain

$$I_o[r_1, p_1(x)] - I_o[r, p(x)] \geqslant$$

$$\frac{1}{1-s}\left[-I_o[r, p(x)] + \int_0^\infty \beta(y)^{s/(1-s)} \int_0^\infty p_1(x)\, e^{r_1(x^2-a)}\, p(y|x)^{1-s}\, dxdy\right]. \tag{A. 41}$$

The order of integration may be interchanged,[33] however, and, since p(x) satisfies (A. 33) by hypothesis,

$$I_o[r_1, p_1(x)] - I_o[r, p(x)] \geqslant$$

$$\frac{I_o[r, p(x)]}{1-s}\left[-1 + \int_0^\infty p_1(x)\, e^{(r_1-r)(x^2-a)}\, dx\right]. \tag{A. 42}$$

Note that $I_o[r, p(x)] > 0$, $e^t \geqslant 1 + t$, so

$$I_o[r_1, p_1(x)] - I_o[r, p(x)] \geqslant$$

$$\frac{I_o[r, p(x)]}{1-s}(r_1-r)\int_0^\infty p_1(x)(x^2-a)\, dx = 0. \tag{A. 43}$$

THEOREM A. 4

A sufficient condition on p(x) to minimize

$$I_1[p(x)] = K\left[-\int_0^\infty p(x) \ln (1+x^2)\, dx - \ln \int_0^\infty \left[\frac{p(x)}{1+x^2}\, dx\right]\right], \tag{A. 44}$$

subject to conditions (A. 3), is

$$\ln (1+x^2) + \frac{1}{1+x^2}\left[\int_0^\infty \frac{p(x_1)}{1+x_1^2}\, dx_1\right]^{-1} \leqslant \lambda_o + \lambda_1 x^2 \tag{A. 45}$$

for some λ_o, λ_1, and all x, with equality when $p(x) > 0$.

Proof: Assume that p(x) satisfies (A. 3) and (A. 45), and $p_1(x)$ satisfies (A. 3). Note that the function $\lambda p_1(x) + (1-\lambda)\, p_2(x)$, $0 < \lambda < 1$, will also satisfy (A. 3).

$$\frac{1}{K} I_1[\lambda p_1(x)+(1-\lambda)p(x)] =$$

$$- \int_0^\infty [\lambda p_1(x)+(1-\lambda)p(x)] \ln(1+x^2)\, dx - \ln\left[\int_0^\infty \frac{\lambda p_1(x)+(1-\lambda)p(x)}{1+x^2}\, dx\right]. \qquad (A.46)$$

It is well known that[34] $\ln \overline{x} \geq \overline{\ln x}$, and by applying this inequality to the last term in (A.46),

$$I_1[\lambda p_1(x)+(1-\lambda)p(x)] \leq \lambda I_1[p_1(x)] + (1-\lambda) I_1[p(x)] \qquad (A.47)$$

or

$$I_1[p_1(x)] - I_1[p(x)] \geq \frac{1}{\lambda}[I_1[\lambda p_1(x)+(1-\lambda)p(x)] - I_1[p(x)]] \qquad (A.48)$$

$$I_1[p_1(x)] - I_1[p(x)] \geq -\int_0^\infty \{p_1(x)-p(x)\} \ln(1+x^2)\, dx$$

$$-\frac{1}{\lambda}\ln\left[1-\lambda+\lambda\frac{\int_0^\infty \frac{p_1(x)}{1+x^2}\, dx}{\int_0^\infty \frac{p(x)}{1+x^2}\, dx}\right]. \qquad (A.49)$$

We now apply the inequality $\ln(1+t) \leq t$, $-1 \leq t$.

$$I_1[p_1(x)] - I_1[p(x)] \geq -\int_0^\infty \{p_1(x)-p(x)\}\ln(1+x^2)\, dx + 1 - \frac{\int_0^\infty \frac{p_1(x)}{1+x^2}\, dx}{\int_0^\infty \frac{p(x)}{1+x^2}\, dx} \qquad (A.50)$$

$$I_1[p_1(x)] - I_1[p(x)] \geq$$

$$\int_0^\infty \{p(x)-p_1(x)\}\left[\ln(1+x^2) + \frac{1}{1+x^2}\left[\int_0^\infty \frac{p(x_1)}{1+x_1^2}\, dx_1\right]^{-1}\right] dx. \qquad (A.51)$$

By hypothesis, $p(x)$ satisfies (A.45). When $p(x) > 0$, equality holds in (A.45), and when $p(x) = 0$, $p(x) - p(x_1) \leq 0$, so that

$$I_1[p_1(x)] - I_1[p(x)] \geq \int_0^\infty \{p(x)-p_1(x)\}\left(\lambda_0+\lambda_1 x^2\right) dx = 0. \qquad (A.52)$$

APPENDIX B

Theorems Concerning the Expurgated Bound for Equal Eigenvalues

Recall that we found, in Section III, that when all of the eigenvalues are equal, we need only compute results for one eigenvalue with value unity. In this case, the conditions sufficient for maximization of the exponent, derived in Theorems A. 1 and A. 2, specialize to

$$\int_0^\infty p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_1(x,x_1)^{1/\rho}\, dx_1 \geq \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_1(x,x_1)^{1/\rho}\, dx dx_1 \tag{B. 1}$$

for any ρ in the range $0 < \rho < \infty$, and

$$\int_0^\infty p(x_1) \ln H_1(x,x_1)\, dx \geq \int_0^\infty \int_0^\infty p(x)\, p(x_1) \ln H_1(x,x_1)\, dx dx_1 + \lambda(x^2-a) \tag{B. 2}$$

for all x and some λ when $\rho = \infty$. Note that

$$H_1(x,x_1) = \frac{(1+x^2)^{1/2}\left(1+x_1^2\right)^{1/2}}{1+\frac{1}{2}\left(x^2+x_1^2\right)}, \tag{B. 3}$$

and $r, p(x)$ are constrained to satisfy (A. 3).

THEOREM B. 1

If $r, p(x)$ and $r_1, p_1(x)$ both satisfy conditions (B.1) and (A.3), then $r_1 = r$, $p(x) = p_1(x)$ almost everywhere, and furthermore

$$\int_0^\infty [p(x)-p_1(x)]^2\, dx = 0.$$

<u>Proof</u>: If $r, p(x)$ and $r_1, p_1(x)$ satisfy the conditions, by Theorem A. 1, they result in the same value of I_x. In the proof of Theorem A.1, it is shown that for this to be the case,

$$\int_0^\infty \int_0^\infty f(x)\, f(x_1)\, H_1(x,x_1)^{1/\rho}\, dx dx_1 = 0, \tag{B. 4}$$

where

$$f(x) = p_1(x)\, e^{r_1(-a+x^2)} - p(x)\, e^{r(-a+x^2)}. \tag{B. 5}$$

Equation B.4 may also be written

$$\int_0^\infty \left[\int_0^\infty f(x) \frac{y^{\frac{1}{\rho}-1} e^{-\frac{y}{1+x^2}}}{\Gamma(1/\rho)(1+x^2)^{1/\rho}} dx\right]^2 dy = 0. \tag{B.6}$$

Change variables to $t = \frac{1}{1+x^2}$, and let

$$h(t) = \frac{t^{\frac{1}{\rho}-2} f\left[\sqrt{\frac{1}{t}-1}\right]}{\sqrt{\frac{1}{t}-1}}. \tag{B.7}$$

Then, for $0 < \rho < \infty$,

$$\int_0^\infty \left[\frac{y^{\frac{1}{\rho}-1}}{2\Gamma(1/\rho)}\right]^2 \left[\int_0^1 h(t)\, e^{-yt}\, dt\right]^2 dy = 0. \tag{B.8}$$

$$\int_0^\infty \left(y^{\frac{1}{\rho}-1}\right)^2 H^2(y)\, dy = 0. \tag{B.9}$$

$$H(y) = \int_0^1 h(t)\, e^{-yt}\, dt. \tag{B.10}$$

As the Laplace transform of a pulse function, H(y) is entire.[35] Equation B.9 implies H(y) = 0 almost everywhere, $y \geq 0$, hence H(y) = 0 in the whole complex plane, and by Parseval's law,

$$\int_0^\infty h^2(t)\, dt = 0 \tag{B.11}$$

or

$$\int_0^\infty \left[\frac{f(x)}{(1+x^2)^{1/\rho}}\right]^2 dx = 0. \tag{B.12}$$

Hence

$$\int_0^\infty \left[\frac{p_1(x)\, e^{r_1(-\alpha+x)^2} - p(x)\, e^{r(-\alpha+x^2)}}{(1+x^2)^{1/\rho}} \right]^2 dx = 0. \tag{B.13}$$

Thus $p_1(x)\, e^{r_1(-\alpha+x^2)} = p(x)\, e^{r(-\alpha+x^2)}$ almost everywhere, and any region where equality does not hold cannot contain any area. Hence $p_1(x) = p(x)\, e^{(r-r_1)(-\alpha+x^2)}$ almost everywhere, and

$$1 = \int_0^\infty p_1(x)\, dx = \int_0^\infty p(x)\, e^{(r-r_1)(-\alpha+x^2)}\, dx. \tag{B.14}$$

But $e^t \geqslant 1 + t$, equality only at $t = 0$. Hence

$$1 \geqslant \int_0^\infty p(x) \left[1+(r-r_1)(-\alpha+x^2) \right] dx = 1, \tag{B.15}$$

so that $(r-r_1)(-\alpha+x^2) = 0$, where $p(x) > 0$. It can be shown that $p(x) = u_o(x-\sqrt{\alpha})$ cannot satisfy (B.1), so $r_1 = r$, $p(x) = p_1(x)$ almost everywhere, and

$$\int_0^\infty [p(x)-p_1(x)]^2\, dx = 0. \tag{B.16}$$

THEOREM B.2

When $\rho = \infty$, $p(x) = p_1 u_o(x) + p_2 u_o(x-x_o)$ satisfies (B.2) and thus maximizes the zero-rate exponent.

Proof: Condition (A.3) requires

$$p_1, p_2 \geqslant 0, \quad p_1 + p_2 = 1, \quad p_2 x_o^2 = \alpha. \tag{B.17}$$

We shall show that (B.2) is satisifed for a given value of α, when p_1, p_2, λ, x_o are chosen properly. For simplicity, let $z = x^2$, $z_o = x_o^2$, and

$$G(z) = \int_0^\infty p(x_1) \ln H_1(\sqrt{z}, x_1)\, dx_1 - \int_0^\infty \int_0^\infty p(x)\, p(x_1) \ln H_1(x, x_1)\, dx dx_1 - \lambda(z-\alpha). \tag{B.18}$$

If we can choose the parameters so that $G(z) \geqslant 0$, $z \geqslant 0$, then the theorem is true. Plugging in $p(x)$, and noting that $H(a, a) = 1$, we have

$$G(z) = p_1 \ln H_1(\sqrt{z}, 0) + p_2 \ln H(\sqrt{z}, \sqrt{z_o}) - 2p_1 p_2 \ln H(0, \sqrt{z}) + \lambda(z-\alpha). \tag{B.19}$$

Let

$$\lambda = \frac{1}{z_o} (1-2p_1) \ln \left[\frac{(1+z_o)^{1/2}}{1+\frac{1}{2} z_o} \right]. \tag{B.20}$$

Direct evaluation confirms that $G(0) = G(z_o) = 0$. If $G(z)$ is to be non-negative for $z \geq 0$, and $G(z_o) = 0$, then $G(z)$ must have a minimum at $z = z_o$, and $G'(z_o) = 0$. This condition determines p_1 and p_2 in terms of z_o as follows:

$$p_1 = 1 - p_2 = \left[2 - \frac{1}{h(z_o)} \right]^{-1} \tag{B.21}$$

$$f_1(z) = \frac{2}{z} \left[\ln \left(1 + \frac{1}{2} z\right) - \frac{1}{2} \ln (1+z) \right] \tag{B.22}$$

$$\gamma(z) = \frac{\frac{1}{2} z}{(1+z)\left(1 + \frac{1}{2} z\right)} \tag{B.23}$$

$$h(z) = \frac{f_1(z)}{\gamma(z)}. \tag{B.24}$$

The important properties of $h(z)$ are the following.

1. $h(z_1) = 1, \quad z_1 \cong 3.071$
2. $h(z)$ is a strictly increasing function of $z, \quad z \geq z_1$
3. $h(\infty) = \infty$.

Consider property 2 first. After some computation, we find that

$$h'(z) = \frac{1}{z} \left[1 - f_1(z) \left(3 + \frac{4}{z}\right) \right]. \tag{B.25}$$

It has been shown[16] that $f_1(z)$ is positive and decreasing for $z \geq z_1 \cong 3.071$ (z_1 is the point where $f_1(z)$ is maximized, and is defined by $f_1(z_1) = \gamma(z_1)$). Also, $\left(3 + \frac{4}{z}\right)$ is decreasing in this range, so

$$h'(z) \geq \frac{1}{z} \left[1 - f_1(z_1) \left(3 + \frac{4}{z_1}\right) \right] \cong \frac{.35}{z} > 0, \tag{B.26}$$

and $h(z)$ is strictly increasing for $z \geq z_1$. Properties 1 and 3 are verified by direct evaluation.

Therefore, if z_o is increased from z_1 to ∞,

1. $h(z_o)$ goes monotonically from 1 to ∞
2. p_1 goes monotonically from 1 to $1/2$
3. p_2 goes monotonically from 0 to $1/2$
4. $p_2 z_o$ goes monotonically from 0 to ∞.

Thus, for any $0 < \alpha < \infty$, there is a one-to-one correspondence with some $z_1 < z_o < \infty$, so that we can work in terms of z_o and later find the corresponding $\alpha = p_2 z_o$. To prove that $G(z) \geq 0$, consider $G'(z)$.

$$G'(z) = \frac{1}{2}\left[\frac{1}{1+z} - \frac{p_1}{1+\frac{1}{2}z} - \frac{p_2}{1+\frac{1}{2}(z+z_o)} + 2\lambda\right]. \tag{B.27}$$

Putting $G'(z)$ over its common denominator yields

$$G'(z) = \frac{b(z)}{2(1+z)\left(1+\frac{1}{2}z\right)\left[1+\frac{1}{2}(z+z_o)\right]} \equiv a(z)\, b(z), \tag{B.28}$$

where $b(z)$ is a cubic polynomial in z. For $z \geq 0$, $a(z) > 0$, so the positive zeros of $G'(z)$ are the same as those of $b(z)$.

Note that

$$G'(z_o) = 0 = b(z_o). \tag{B.29}$$

$$G'(o) = \frac{1}{2}\left[\frac{\frac{1}{2}p_2 z_o}{1+\frac{1}{2}z_o} + 2\lambda\right] = \frac{1}{2}\left[\frac{\frac{1}{2}p_2 z_o}{1+\frac{1}{2}z_o} + \left(p_1 - \frac{1}{2}\right) f_1(z_o)\right] > 0, \tag{B.30}$$

and thus $b(0) > 0$. Also,

$$z \to \pm\infty \Longrightarrow b(z) \to \frac{\lambda}{2}z^3 \to \pm\infty. \tag{B.31}$$

Since $b(z)$ is a cubic polynomial, it can have, at most, three real roots. There must be at least one negative real root because $b(0) > 0$, $b(-\infty) < 0$. Thus there can be, at most, one other positive root in addition to the one at z_o. $G(z)$ is continuous and differentiable in (O, z_o), and since $G(0) = G(z_o) = 0$, the Mean Value theorem[36] states that $G'(z_2) = 0$, for some $0 < z_2 < z_o$, and hence $b(z_2) = 0$, too. Thus $b(z)$ must be as shown in Fig. 32a.

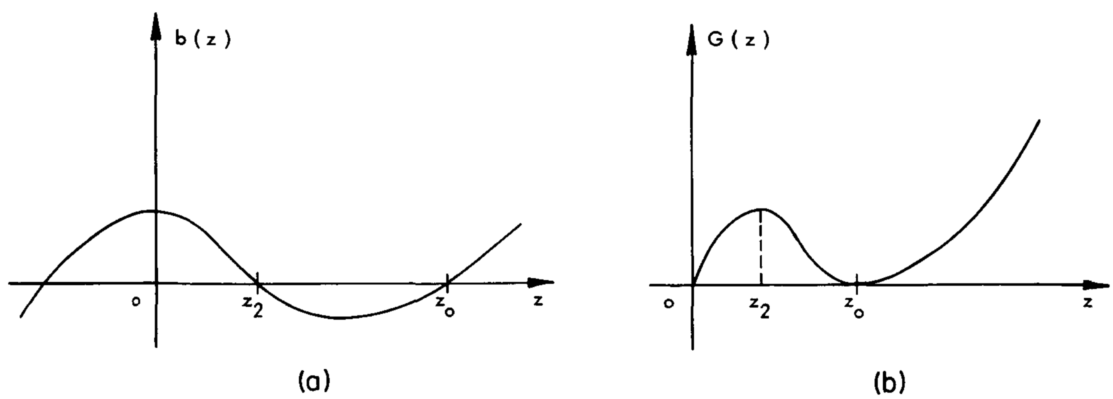

Fig. 32. (a) b(z) versus z.
(b) G(z) versus z.

Thus $\left.\begin{array}{l} b(z_o) = 0, \ b'(z_o) > 0 \\ b(z_2) = 0, \ b'(z_2) < 0 \end{array}\right\}$ the only roots of b(z) for $z \geq 0$, and $b(z) \geq 0$, $z \geq z_o$, so that $G'(z) \geq 0$, $z \geq z_o$, and thus $G(z) \geq 0$, $z \geq z_o$. Differentiating (B.28) and noting that $b(z_o) = b(z_2) = 0$ show that $G''(z_o) = a(z_o)\, b'(z_o) > 0$, and $G''(z_2) = a(z_2)\, b'(z_2) < 0$. Thus G(z) has a minimum at $z = z_o$ and a maximum at $z = z_2$. These are the only two extreme points of G(z) for $z \geq 0$, so G(z) must be as shown in Fig. 32b. Hence $G(z) \geq 0$, and two impulses, suitably chosen, optimize the zero-rate expurgated bound. The resulting exponent is

$$E_{xe}[\infty, a, 1] = a\left(1 - \frac{a}{z_o}\right) f_1(z_o) \leq a f_1(z_o) \leq a f_1(z_1) \leq a E_\infty \cong 0.15a. \tag{B.32}$$

THEOREM B.3

If we define

$$\sum_{k=1}^{\infty} \lambda_k^2 = b, \qquad \sum_{k=1}^{\infty} \lambda_k^3 = d, \tag{B.33}$$

then

$$E_x[\rho, a, \underline{\lambda}] \geq E_{xe}\left[\rho, \frac{b^2 a}{d}, \frac{b^3}{d^2}\right]. \tag{B.34}$$

Recall that $\lambda_k \geq 0$, $\sum_{k=1}^{\infty} \lambda_k = 1$, and

$$E_x[\rho, a, \underline{\lambda}] = -\rho \ln \left[\min_{r, p(x)} \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(-2a + x^2 + x_1^2\right)} \left\{ \prod_{k=1}^{\infty} \frac{\left(1 + \lambda_k x^2\right)^{1/2} \left(1 + \lambda_k x_1^2\right)^{1/2}}{1 + \frac{1}{2} \lambda_k \left(x^2 + x_1^2\right)} \right\}^{1/\rho} dx dx_1 \right] \tag{B.35}$$

$$E_{xe}[\rho, a, K] = -\rho \ln \left[\min_{r, p(x)} \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(-2a + x^2 + x_1^2\right)} \left\{ \frac{\left(1 + \frac{x^2}{K}\right)^{1/2} \left(1 + \frac{x_1^2}{K}\right)^{1/2}}{1 + \frac{1}{2K}\left(x^2 + x_1^2\right)} \right\}^{1/\rho} dx dx_1 \right], \tag{B.36}$$

where in each case, r and $p(x)$ must satisfy conditions (A. 3).

Proof:

$$E_x[\rho, \alpha, \underline{\lambda}] = -\rho \ln\left[\min_{r, p(x)} \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(-2\alpha+x^2+x_1^2\right)} G_{\underline{\lambda}}(x, x_1)\, dx dx_1\right] \tag{B. 37}$$

$$G_{\underline{\lambda}}(x, x_1) = \exp\left[-\frac{1}{\rho}\sum_{k=1}^{\infty} \lambda_k^2\left\{\frac{1}{\lambda_k^2}\left[\ln\left(1+\frac{1}{2}\left(\lambda_k x^2+\lambda_k x_1^2\right)\right)-\frac{1}{2}\ln\left(1+\lambda_k x^2\right)-\frac{1}{2}\ln\left(1+\lambda_k x_1^2\right)\right]\right\}\right]. \tag{B. 38}$$

Define

$$B(x) = \frac{1}{x^2}\left[\ln\left(1+\frac{1}{2}x(a+a_1)\right)-\frac{1}{2}\ln(1+ax)-\frac{1}{2}\ln(1+a_1x)\right]. \tag{B. 39}$$

We shall now show that $B''(x) \geq 0$, all $x \geq 0$, and any fixed $a, a_1 \geq 0$.

Evaluation of the second derivative shows that

$$B''(x) = \frac{6}{x^2}B(x) + \frac{1}{2x^3}\left[\frac{a(4+5ax)}{(1+ax)^2} + \frac{a_1(4+5a_1x)}{(1+a_1x)^2} - \frac{2\left(\frac{a+a_1}{2}\right)\left(4+5\left(\frac{a+a_1}{2}\right)x\right)}{\left[1+\left(\frac{a+a_1}{2}\right)x\right]^2}\right]. \tag{B. 40}$$

Using (B. 39), and defining

$$g(x) = -6\ln(1+x) + \frac{x(4+5x)}{(1+x)^2}, \tag{B. 41}$$

we find that

$$B''(x) = \frac{1}{x^4}\left[\frac{1}{2}\{g(ax)+g(a_1x)\} - g\left\{\left(\frac{a+a_1}{2}\right)x\right\}\right]. \tag{B. 42}$$

But

$$g''(x) = \frac{6x^2}{(1+x)^4} \geq 0, \qquad x \geq 0, \tag{B. 43}$$

so that $g(x)$ is convex, and

$$\frac{1}{2}\{g(ax)+g(a_1x)\} \geq g\left\{\left(\frac{a+a_1}{2}\right)x\right\}, \tag{B. 44}$$

thereby proving that $B''(x) \geq 0$, $x \geq 0$, and hence $B(x)$ is convex.

If $\underline{p}$ is a probability vector,

$$B\left[\sum_{k=1}^{\infty} p_k x_k\right] \leq \sum_{k=1}^{\infty} p_k B(x_k). \tag{B.45}$$

Let

$$p_k = \frac{\lambda_k^2}{b}, \qquad x_k = \lambda_k,$$

$$G_{\underline{\lambda}}(x, x_1) \leq \exp\left[-\frac{b^3}{\rho d^2}\left[\ln\left\{1 + \frac{1}{2}\frac{d}{b}\left(x^2 + x_1^2\right)\right\} - \frac{1}{2}\ln\left(1 + \frac{d}{b}x^2\right) - \frac{1}{2}\ln\left(1 + \frac{d}{b}x_1^2\right)\right]\right] \tag{B.46}$$

$$G_{\underline{\lambda}}(x, x_1) \leq \left[\frac{\left(1 + \frac{d}{b}x^2\right)^{1/2}\left(1 + \frac{d}{b}x_1^2\right)^{1/2}}{1 + \frac{1}{2}\frac{d}{b}\left(x^2 + x_1^2\right)}\right]^{b^3/\rho d^2}. \tag{B.47}$$

Plugging this into (B.37), and making use of (50) and (B.36), we have

$$E_x[\rho, a, \underline{\lambda}] \geq \frac{b^2}{d} E_{xe}\left[\frac{\rho d}{b^2}, a, \frac{b}{d}\right] = E_{xe}\left[\rho, \frac{b^2 a}{d}, \frac{b^3}{d^2}\right]. \tag{B.48}$$

THEOREM B.4

As $a \to 0$, a two-impulse p(x) asymptotically satisfies (B.1), the condition for optimization of the expurgated bound, and hence is asymptotically optimum. The resulting exponent is the same as the infinite-bandwidth, orthogonal-signal exponent found by Kennedy, for $R \leq R_{crit}$.

Proof: Define

$$F(x) = \int_0^\infty p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_1(x, x_1)^{1/\rho}\, dx_1 - \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(x^2+x_1^2\right)} H_1(x, x_1)^{1/\rho}\, dx dx_1. \tag{B.49}$$

Let

$$p(x) = p_1 u_o(x) + p_2 u_o(x - x_o) \tag{B.50}$$

$$p_2 = \frac{\alpha}{x_o^2}, \qquad p_1 = 1 - \frac{\alpha}{x_o^2} \tag{B.51}$$

where $x_o^2 \cong 3.071$ maximizes $f_1(x^2)$ given by (B.22). Also let $r = 1/2\rho\, f_1\left(x_o^2\right)$, so that $e^{-rx_o^2} = H_1(o, x_o)^{1/\rho}$. With this choice of r and p(x),

$$F(x)\, e^{-rx^2} = \left\{H_1(0,x)^{1/\rho} - e^{-rx^2}\right\} + \frac{\alpha}{x_o^2}\left\{e^{rx_o^2} H_1(x, x_o)^{1/\rho} - H_1(0,x)^{1/\rho}\right\}$$
$$- \frac{\alpha^2}{x_o^4} e^{-rx^2}\left\{e^{2rx_o^2} - 1\right\}. \tag{B.52}$$

The coefficients of α and α^2 are bounded for all x, so as $\alpha \to 0$,

$$F(x)\, e^{-rx^2} = H_1(0,x)^{1/\rho} - e^{-rx^2} + O(\alpha), \tag{B.53}$$

where we use the O notation to indicate that $\lim_{\alpha \to o} \left|\frac{o(\alpha)}{\alpha}\right| < \infty$. But

$$\ln\left[H_1(o,x)^{1/\rho}\, e^{rx^2}\right] = +\frac{x^2}{2\rho}\left[f_1\left(x_o^2\right) - f_1(x^2)\right] \geq 0. \tag{B.54}$$

Thus $H_1(o,x)^{1/\rho} \geq e^{-rx^2}$ and $F(x) \geq 0$ as $\alpha \to 0$.

The resulting exponent is

$$E_{xe}[\rho, \alpha, 1] = 2\rho r\alpha - \rho \ln\left[1 + \frac{\alpha^2}{x_o^4}\left\{e^{2rx_o^2} - 1\right\}\right] \tag{B.55}$$

$$E_{xe}[\rho, \alpha, 1] = 2\rho r\alpha - O(\alpha^2) \tag{B.56}$$

$$E_{xe}[\rho, \alpha, 1] \cong \alpha f_1\left(x_o^2\right), \tag{B.57}$$

which is the straight-line bound that Kennedy found for low rates.

THEOREM B.5

$E_{xe}(\infty, \alpha, K)$ is a jointly concave function of α and K.

Proof: Define

$$F(\alpha) \equiv E_{xe}(\infty, \alpha, 1). \tag{B.58}$$

Then

$$E_{xe}(\infty, a, K) = KF\left(\frac{a}{K}\right) \equiv \theta(a, K). \tag{B.59}$$

The conditions for $\theta(a, K)$ to be jointly concave[37] are

$$\theta_{aa} \leq 0, \quad \theta_{KK} \leq 0, \quad \theta_{aK}^2 - \theta_{aa}\theta_{KK} \leq 0, \tag{B.60}$$

where we are using standard partial derivative notation.

We first show that $F''(a) \leq 0$. From the proof of Theorem B.2,

$$F(a) = \max_{z \geq z_o} a\left(1 - \frac{a}{z}\right) f_1(z) \tag{B.61}$$

or

$$F(a) = g(a, z) \equiv a\left(1 - \frac{a}{z}\right) f_1(z), \tag{B.62}$$

subject to the constraint $g_z = 0$. Differentiating, we obtain

$$F'(a) = g_a + g_z \frac{dz}{da} = g_a = \left(1 - \frac{2a}{z}\right) f_1(z). \tag{B.63}$$

Again, in Theorem B.2, we found that $\frac{a}{z}$ and z were both increasing functions of a, and $f_1(z)$ was a decreasing function of z, and hence of a, too. Thus g_a is a decreasing function of a and hence $F''(a) \leq 0$.

Direct evaluation confirms that

$$\theta_{aa} = \frac{1}{K} F''(a/K) \leq 0 \tag{B.64}$$

$$\theta_{KK} = \frac{a^2}{K^3} F''(a/K) \leq 0 \tag{B.65}$$

$$\theta_{aK} = -\frac{a}{K^2} F''(a/K) \tag{B.66}$$

$$\theta_{aK}^2 - \theta_{aa}\theta_{KK} = 0, \tag{B.67}$$

and so (B.60) is satisfied, thereby proving the theorem.

APPENDIX C

Theorems Relating to the Random-Coding Bound for Equal Eigenvalues

THEOREM C.1

When $s = 1$, $p(x) = p_1 u_o(x) + p_2 u_o(x-x_o)$ satisfies (A.45).

Proof: Condition (A.3) requires (remember to replace a with $\frac{a}{K}$)

$$p_1, p_2 \geq 0, \quad p_1 + p_2 = 1, \quad p_2 x_o^2 = \frac{a}{K}. \tag{C.1}$$

Inequality (A.45) may be written

$$G(z) \equiv \lambda_o + \lambda_1 z - \ln (1+z) - \frac{1}{1+z}\left[\int_0^\infty \frac{p(x)}{1+x^2}\, dx\right]^{-1} \geq 0 \tag{C.2}$$

for all $z \geq 0$, with equality when $p(x) > 0$, where for convenience, we have set $z = x^2$, $z_o = x_o^2$. Equality at $z = 0$ and $z = z_o$ requires

$$\lambda_o = \frac{1+z_o}{1+p_1 z_o} \tag{C.3}$$

$$\lambda_1 = \frac{1}{z_o} \ln (1+z_o) - \frac{1}{1+p_1 z_o}. \tag{C.4}$$

If $G(z)$ is to be non-negative, and $G(z_o) = 0$, we require that $G'(z_o) = 0$. This determines p_1 as follows:

$$p_1 = \left[\frac{1+z_o}{z_o} \ln (1+z_o) - 1\right]^{-1} - \frac{1}{z_o}. \tag{C.5}$$

Note that

$$\frac{1+z_o}{z_o} \ln (1+z_o) - 1 \leq z_o, \tag{C.6}$$

so that if $z_o > 0$, $p_1 > 0$. We also require $p_1 \leq 1$. This will be ensured if

$$\ln (1+z_o) \geq \frac{z_o + 2z_o^2}{(1+z_o)^2}. \tag{C.7}$$

Let

$$H(z_o) = \ln(1+z_o) - \frac{z_o + 2z_o^2}{(1+z_o)^2}. \tag{C.8}$$

Then

$$H'(z_o) = \frac{z_o(z_o-1)}{(1+z_o)^3}, \tag{C.9}$$

and $H'(z_o) > 0$ when $z_o > 1$. Therefore, if $z_o > 1$, $H(z_o)$ is increasing. Direct computation confirms that $z_o \cong 2.163$ satisfies (C.7) with equality, and results in $p_1 = 1$. Hence if $z_o \geqslant 2.163$, $1 \geqslant p_1 > 0$. Then we could specify $z_o \geqslant 2.163$, solve for p_1, p_2, α, and have a probability function that satisfies all of the constraints for that value of α.

Recall that $G(0) = G(z_o) = G'(z_o) = 0$. By using the continuity and differentiability of $G(z)$ to apply the mean-value theorem, there exists some z_1, $0 < z_1 < z_o$, such that $G'(z_1) = 0$. Differentiation of (C.2) leads to

$$G'(z) = \frac{\lambda_1(1+z)^2 - (1+z) + \left(\frac{1+z_o}{1+p_1 z_o}\right)}{(1+z)^2}, \tag{C.10}$$

and thus $G'(z)$ can only have the two positive zeros just mentioned. Also

$$G'(0) = \lambda_1 - 1 + \left(\frac{1+z_o}{1+p_1 z_o}\right) = \left(\frac{2+z_o}{z_o}\right) \ln(1+z_o) - 2. \tag{C.11}$$

Making use of (C.7), we find that

$$G'(0) \geqslant \frac{z_o}{(1+z_o)^2} > 0. \tag{C.12}$$

Thus $G(z)$ must have a maximum at $z = z_1$, a minimum at $z = z_o$, and since these represent the only two zeros of $G'(z)$ for $z \geqslant 0$, $G(z)$ must be as shown in Fig. 33, thereby proving the theorem.

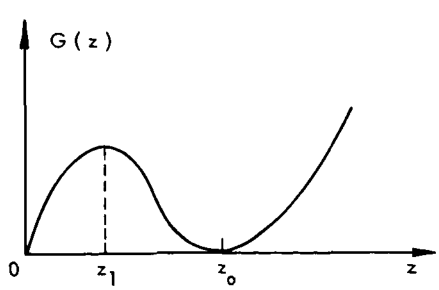

Fig. 33. G(z) versus z.

The resulting exponent is

$$E_{oe}[\infty, \alpha, K] = K\left[\frac{\alpha}{Kz_o} \ln (1+z_o) + \ln\left\{1 - \frac{\alpha/K}{1 + z_o}\right\}\right] \tag{C.12a}$$

$$E_{oe}[\infty, \alpha, K] \leqslant \alpha\left[\frac{1}{z_o} \ln (1+z_o) - \frac{1}{1 + z_o}\right]. \tag{C.12b}$$

By differentiation, it is easy to show that the right-hand side of (C.12b) is a decreasing function of z_o, for $z_o \geqslant 2.163$, so that

$$E_{oe}[\infty, \alpha, K] \leqslant 0.2162\alpha. \tag{C.12c}$$

THEOREM C.2

As $\alpha \to 0$, a two-impulse p(x) asymptotically satisfies condition (A.33), $0 < s < 1$. For $0 < s \leqslant \frac{1}{2}$ $(0 < \rho \leqslant 1$, and $R \geqslant R_{crit})$ the resulting exponent is the same as the one obtained with orthogonal signals and an infinite bandwidth.

<u>Proof:</u> Condition (A.33) states

$$F(x) \equiv \int_0^\infty \beta(y)^{s/(1-s)}\, p(y|x)^{1-s}\, e^{rx^2}\, dy - \int_0^\infty \beta(y)^{1/(1-s)}\, dy \geqslant 0 \tag{C.13}$$

is sufficient for an optimum.

Let

$$p(x) = p_1 u_o(x) + p_2 u_o(x-x_o) \tag{C.14}$$

$$p_2 = \frac{\alpha}{x_o^2 K}, \qquad p_1 = 1 - \frac{\alpha}{x_o^2 K}. \tag{C.15}$$

Define

$$f_{\frac{s}{1-s}}(x) \equiv \frac{1}{x(1-s)} [\ln (1+sx) - s \ln (1+x)]. \tag{C.16}$$

The function $f_{\frac{s}{1-s}}(x)$ is positive for $x > 0$, $0 < s < 1$, and has a single maximum in x.[16] Choose x_o^2 so that $f_{\frac{s}{1-s}}\left(x_o^2\right)$ is a maximum, and let

$$r = K(1-s)\, f_{\frac{s}{1-s}}\left(x_o^2\right). \tag{C.17}$$

Applying (A.36) gives

$$\int_0^\infty \beta(y)^{1/(1-s)}\, dy = \int_0^\infty \left[\int_0^\infty p(x)\left\{e^{rx^2}\, p(y|x)^{1-s}\right\} dx\right]^{1/(1-s)} dy$$

$$\leq \int_0^\infty \int_0^\infty p(x)\left\{e^{rx^2}\, p(y|x)^{1-s}\right\}^{1/(1-s)} dxdy \qquad \text{(C. 18)}$$

$$\int_0^\infty \beta(y)^{1/(1-s)}\, dy \leq 1 + \frac{a}{x_o^2 K}\left[e^{\frac{rx_o^2}{1-s}} - 1\right] = 1 + 0(a). \qquad \text{(C. 19)}$$

Let

$$G(x) \equiv e^{rx^2} \int_0^\infty \beta(y)^{s/(1-s)}\, p(y|x)^{1-s}\, dy \qquad \text{(C. 20)}$$

$$G(x)\, e^{-rx^2} = \left[\int_0^\infty p(y|x)^{1-s}\, p(y|o)^s\, dy\right]\left[\overline{1 + p_2\left\{-1 + e^{rx_o^2}\left[\frac{p(y|x_o)}{p(y|o)}\right]^{1-s}\right\}}\right]^{s/(1-s)}, \qquad \text{(C. 21)}$$

where the bar denotes an average with respect to

$$\frac{p(y|o)^s\, p(y|x)^{1-s}}{\int_0^\infty p(y|o)^s\, p(y|x)^{1-s}\, dy}.$$

Using (A. 36) again, and assuming that $s \geq \frac{1}{2}$, we obtain

$$G(x)\, e^{-rx^2} \geq \left[\frac{(1+x^2)^s}{1 + sx^2}\right]^K \left[1 + p_2\left\{-1 + \frac{e^{rx_o^2}\int_0^\infty p(y|o)^{-1+2s}\, p(y|x_o)^{2-2s}\, dy}{\int_0^\infty p(y|0)^s\, p(y|x_o)^{1-s}\, dy}\right\}\right]^{s/(1-s)} \qquad \text{(C. 22)}$$

$$G(x)\, e^{-rx^2} \geq \left[\frac{(1+x^2)^s}{1 + sx^2}\right]^K \left[1 + \frac{a}{x_o^2 K}\left\{-1 + e^{rx_o^2}\left[\frac{1 + sx_o^2}{\left\{1 + x_o^2(2s-1)\right\}\left\{1 + x_o^2\right\}^{1-s}}\right]^K\right\}\right]^{s/(1-s)}. \qquad \text{(C. 23)}$$

All terms in (C. 23) are bounded, so that

$$G(x)\, e^{-rx^2} \geqslant \left[\frac{(1+x^2)^s}{1+sx^2}\right]^K + O(a), \qquad 1 > s \geqslant \frac{1}{2}. \tag{C.24}$$

Now assume that $0 < s < \frac{1}{2}$, and write G(x) as

$$G(x)\, e^{-rx^2} = \int_0^\infty p(y|x)^{1-s} \left[\int_0^\infty p(x_1) \left\{p(y|x_1)^{1-s}\, e^{rx_1^2}\right\} dx_1\right]^{s/(1-s)} dy. \tag{C.25}$$

Applying (A.36) to the inner integral this time, we obtain

$$G(x)\, e^{-rx^2} \geqslant \int_0^\infty p(y|x)^{1-s} \int_0^\infty p(x_1)\, p(y|x_1)^s\, e^{rx_1^2 s/(1-s)}\, dx_1\, dy \tag{C.26}$$

$$G(x)\, e^{-rx^2} \geqslant \left[\frac{(1+x^2)^s}{1+sx^2}\right]^K + \frac{a}{x_o^2 K}\left[e^{\frac{rx_o^2 s}{1-s}}\left\{\frac{(1+x^2)^s\left(1+x_o^2\right)^{1-s}}{1+sx^2+(1-s)x_o^2}\right\}^K - \left\{\frac{(1+x^2)^s}{1+sx^2}\right\}^K\right]. \tag{C.27}$$

This result, together with (C.24), proves that

$$G(x)\, e^{-rx^2} \geqslant \left[\frac{(1+x^2)^s}{1+sx^2}\right]^K + 0(a), \qquad 0 < s < 1 \tag{C.28}$$

$$F(x)\, e^{-rx^2} \geqslant \left[\frac{(1+x^2)^s}{1+sx^2}\right]^K - e^{-rx^2} + 0(a). \tag{C.29}$$

But

$$\ln\left[e^{rx^2}\left\{\frac{(1+x^2)^s}{1+sx^2}\right\}^K\right] = x^2(1-s)\left[f_{\frac{s}{1-s}}\left(x_o^2\right) - f_{\frac{s}{1-s}}(x^2)\right] \geqslant 0. \tag{C.30}$$

Thus as $a \to 0$, $F(x) \geqslant 0$, with equality at $x = 0$, $x = \pm x_o$, and hence this probability function is asymptotically optimum.

The resulting exponent is

$$E_{oe}[\rho, a, K] = -\ln\left[e^{-ra/[K(1-s)]}\int_0^\infty \beta(y)^{1/(1-s)}\, dy\right] = af_{\frac{s}{1-s}}\left(x_o^2\right) - \ln\left[\int_0^\infty \beta(y)^{1/(1-s)}\, dy\right], \tag{C.31}$$

where $\rho = s/(1-s)$. We shall show that for $0 < s \leqslant \frac{1}{2}$, $\ln\left[\int_0^\infty \beta(y)^{1/(1-s)}\, dy\right]$ is negligible as $a \to 0$.

$$\int_0^\infty \beta(y)^{1/(1-s)}\, dy = \int_0^\infty p(y|0)\left[1 + \frac{a}{x_o^2 K} H(y)\right]^{1/(1-s)} dy \tag{C.32}$$

$$H(y) \equiv -1 + e^{rx_o^2}\left[\frac{p(y|x_o)}{p(y|0)}\right]^{1-s} = -1 + e^{rx_o^2}\left[\frac{e^{+yx_o^2/\left(1+x_o^2\right)}}{\left(1+x_o^2\right)^K}\right]^{1-s}. \tag{C.33}$$

Note that H(y) is an increasing function of y, $H(\infty) = +\infty$, and

$$0 > H(0) = -1 + \left[\frac{1 + sx_o^2}{1 + x_o^2}\right]^{K(1-s)} > -1. \tag{C.34}$$

If y_o is defined by the relation

$$1 = e^{rx_o^2}\left[\frac{e^{y_o x_o^2/\left(1+x_o^2\right)}}{\left(1+x_o^2\right)^K}\right]^{1-s}, \tag{C.35}$$

then

$$-1 + \left[\frac{1 + sx_o^2}{1 + x_o^2}\right]^{K(1-s)} \leqslant H(y) \leqslant 0, \qquad 0 \leqslant y \leqslant y_o \tag{C.36}$$

$$0 < H(y), \qquad y_o < y. \tag{C.37}$$

We will now show that, for $0 < s \leqslant \frac{1}{2}$, $a > 0$,

$$(1+t)^{1/(1-s)} \leqslant 1 + \frac{t}{1-s} + \begin{cases} t^{1/(1-s)}, & t > 0 \\ \dfrac{s}{2a(1-s)^2}\, t^2, & -1 + a \leqslant t \leqslant 0 \end{cases} \tag{C.38}$$

To show the first part, let

$$A(t) = 1 + \frac{t}{1-s} + t^{1/(1-s)} - (1+t)^{1/(1-s)} \tag{C.39}$$

$$A'(t) = \frac{1}{1-s}\left[1 + t^{s/(1-s)} - (1+t)^{s/(1-s)}\right] \tag{C.40}$$

$$A''(t) = \frac{s}{(1-s)^2} t^{(-1+2s)/(1-s)} \left[1 - \left(1 + \frac{1}{t}\right)^{-(1-2s)/(1-s)} \right], \tag{C.41}$$

so that for $t > 0$, $0 < s \leqslant \frac{1}{2}$, $A''(t) \geqslant 0$. But $A(0) = A'(0) = 0$, so that $A'(t)$ is increasing and thus non-negative, so that $A(t)$ is increasing, and $A(t) \geqslant 0$, thereby proving the assertion.

To show the second part, let

$$A_1(t) = 1 + \frac{t}{1-s} + \frac{st^2}{2a(1-s)^2} - (1+t)^{1/(1-s)} \tag{C.42}$$

$$A_1'(t) = \frac{1}{1-s} \left[1 + \frac{st}{a(1-s)} - (1+t)^{1/(1-s)} \right] \tag{C.43}$$

$$A_1''(t) = \frac{s}{(1-s)^2} \left[\frac{1}{a} - (1+t)^{-(1-2s)/(1-s)} \right]. \tag{C.44}$$

But $(1+t)^{-(1-2s)/(1-s)} \leqslant a^{-(1-2s)/(1-s)} \leqslant a^{-1}$ in this region. Hence $A_1''(t) \geqslant 0$. But $A_1'(0) = A_1(0) = 0$, so that $A_1'(t)$ is increasing and must be nonpositive, so that $A_1(t)$ is decreasing and must be non-negative in $-1 + a \leqslant t \leqslant 0$, thereby proving the second part of (C.38).

If $\alpha/\left(x_o^2 K\right) \leqslant 1$, then

$$\int_0^\infty \beta(y)^{1/(1-s)}\, dy \leqslant \int_0^\infty p(y|0) \left[1 + \frac{\alpha}{x_o^2 K(1-s)} H(y) \right] dy + \frac{s}{2(1-s)^2} \left(\frac{\alpha}{x_o^2 K} \right)^2 \left[\frac{1 + x_o^2}{1 + sx_o^2} \right]^{K(1-s)} \int_0^{y_o} p(y|0)\, H^2(y)\, dy + \left(\frac{\alpha}{x_o^2 K} \right)^{1/(1-s)} \int_{y_o}^\infty p(y|0)\, H(y)^{1/(1-s)}\, dy \tag{C.45}$$

$$\int_0^\infty \beta(y)^{1/(1-s)}\, dy \leqslant 1 + \left(\frac{\alpha}{x_o^2 K} \right)^2 \left[\frac{1 + x_o^2}{1 + sx_o^2} \right]^{K(1-s)} \int_0^{y_o} p(y|0)\, dy + \left(\frac{\alpha}{x_o^2 K} \right)^{1/(1-s)} \int_{y_o}^\infty p(y|x_o)\, e^{rx_o^2/(1-s)}\, dy \tag{C.46}$$

$$\int_0^\infty \beta(y)^{1/(1-s)}\, dy \leqslant 1 + \left(\frac{\alpha}{x_o^2 K} \right)^2 \left[\frac{1 + x_o^2}{1 + sx_o^2} \right]^{K(1-s)} + \left(\frac{\alpha}{x_o^2 K} \right)^{1/(1-s)} e^{rx_o^2/(1-s)}. \tag{C.47}$$

Also, using (A. 36), we obtain

$$\int_0^\infty \beta(y)^{1/(1-s)}\, dy \geq \left[1 + p_2\left\{-1 + e^{rx_o^2} \int_0^\infty p(y|0)^s\, p(y|x_o)^{1-s}\, dy\right\}\right]^{1/(1-s)} = 1, \tag{C.48}$$

so that

$$0 \leq \ln\left[\int_0^\infty \beta(y)^{1/(1-s)}\, dy\right] \leq 0(\alpha^{1/(1-s)}) \tag{C.49}$$

$$E_{oe}[\rho, \alpha, K] = \alpha f_{\frac{s}{1-s}}\left(x_o^2\right) + 0(\alpha^{1/(1-s)}), \tag{C.50}$$

and as $\alpha \to 0$, $0 < s \leq \frac{1}{2}$, the infinite-bandwidth, orthogonal signal exponent is obtained.

APPENDIX D

Optimality of Impulses

We shall prove that some p(x) consisting of a finite number of impulses must satisfy the sufficient conditions for optimization of the expurgated bound, $0 < \rho < \infty$, and the random coding bound, $0 < s < 1$. Let the subscript ρ relate to the expurgated bound, and s to the random-coding bound.

It has been shown (Theorems A. 1 and A. 3) that a sufficient condition for the minimization of

$$I_\rho = \int_0^\infty \int_0^\infty p(x)\, p(x_1)\, e^{r\left(x^2+x_1^2-2a\right)}\, H_1(x,x_1)^{1/\rho}\, dx dx_1 \tag{D. 1. ρ}$$

$$I_s = \int_0^\infty \left[\int_0^\infty p(x)\, e^{r(x^2-a)}\, p(y|x)^{1-s}\, dx\right]^{1/(1-s)} dy \tag{D. 1. s}$$

over r and p(x), subject to the conditions

$$r \geq 0, \quad p(x) \geq 0, \quad \int_0^\infty p(x)\, dx = 1, \quad \int_0^\infty x^2 p(x)\, dx = a \tag{D. 2}$$

is

$$\int_0^\infty p(x_1)\, e^{r\left(x^2+x_1^2\right)}\, H_1(x,x_1)^{1/\rho}\, dx_1 \geq I_\rho\, e^{2ra} \tag{D. 3. ρ}$$

$$\int_0^\infty \left[\int_0^\infty p(x_1)\, e^{rx_1^2}\, p(y|x)^{1-s}\, dx_1\right]^{s/(1-s)} e^{rx^2}\, p(y|x)^{1-s}\, dy \geq I_s\, e^{ra/(1-s)} \tag{D. 3. s}$$

for all x. Since x appears in these relations only in squared form, replace x^2 by z and rewrite the conditions:

$$\int_0^\infty p(x_1)\, e^{r\left(z+x_1^2\right)} \left[\frac{\left(1+x_1^2\right)^{1/2} (1+z)^{1/2}}{1+\frac{1}{2}\left(z+x_1^2\right)}\right]^{1/\rho} dx_1 \geq I_\rho\, e^{2ra} \tag{D. 4. ρ}$$

$$\int_0^\infty \left[\int_0^\infty p(x_1)\, e^{rx_1^2}\, p(y|x_1)^{1-s}\, dx_1\right]^{s/(1-s)} e^{rz}\, p(y|\sqrt{z})^{1-s}\, dy \geq I_s\, e^{ra/(1-s)} \tag{D. 4. s}$$

for all $z \geq 0$.

THEOREM D.1

Condition (D.4.ρ) is satisfied by some r, $p(x)$ combination that also satisfies (D.2), where $p(x)$ consists of a finite number of impulses.

Proof: Consider some $p(x)$ consisting of a finite number of impulses, that is, $p(x) = \sum_{n=1}^{N} p_n u_o(x - \sqrt{z_n})$. Label the positions so that $0 \leq z_1 < z_2 < \ldots < z_N$, and assume that $p_n > 0$ because any z_n for which $p_n = 0$ may be deleted. We start with $N = 1$ impulse, and then consider increasing N in a search for some N that satisfies (D.4.ρ).

When $N = 1$, $p(x) = u_o(x - \sqrt{\alpha})$, or constraints (D.2) will be violated. It is easy to show that the resulting I_ρ may be reduced by using $r = 0$ and any $p(x)$ with $N = 2$ that satisfies the constraints; thus we may skip $N = 1$. With $N \geq 2$, we formulate a vector minimization problem as follows: Specify the set $\{z_n\}$, and minimize

$$J_\rho \equiv \sum_{n=1}^{N} \sum_{m=1}^{N} p_n p_m e^{r(z_n + z_m - 2\alpha)} \left[\frac{(1+z_n)^{1/2} (1+z_m)^{1/2}}{1 + \frac{1}{2}(z_n + z_m)}\right]^{1/\rho} \tag{D.5}$$

over r, $\{p_n\}$, subject to the constraints

$$r \geq 0, \quad p_n \geq 0, \quad \sum_{n=1}^{N} p_n = 1, \quad \sum_{n=1}^{N} p_n z_n = \alpha. \tag{D.6}$$

This is the discrete analog of the continuous minimization problem of (D.1.ρ).

It can be shown[10] that a necessary and sufficient condition for r, $\{p_n\}$ to minimize J_ρ is

$$F_\rho(z_n) \geq 0 \tag{D.7}$$

for all n, with equality when $p_n > 0$, where

$$F_\rho(z) = \sum_{n=1}^{N} p_n e^{r(z+z_n)} \left[\frac{(1+z_n)^{1/2} (1+z)^{1/2}}{1 + \frac{1}{2}(z+z_n)}\right]^{1/\rho} - J_\rho e^{2r\alpha}. \tag{D.8}$$

If $p_n > 0$ for some $z_n > \alpha$, then r must be bounded (because $r \to \infty$ will make $J_\rho \to \infty$ — clearly not a minimum). If $N \geq 2$, some z_n must exceed α, for otherwise the constraints could not be satisfied. Thus, since J_ρ is a continuous function of r, $\{p_n\}$ in a closed, bounded region, (D.7) must have a solution. Given $\{z_n\}$, there must be some r, $\{p_n\}$ that satisfy (D.7) and result in a minimum value of J_ρ for those $\{z_n\}$ (under the assumptions that at least one $z_n < \alpha$, and at least one $> \alpha$, or else no $\{p_n\}$ can satisfy the constraints). Restrict the analysis to such impulse sets, for any set of impulses that does not first satisfy (D.7) cannot satisfy (D.4.ρ).

Theorem D.3 states that any set of N impulses for which $z_N > z_\rho$ has a value of J_ρ

that can be strictly reduced by some set of N or fewer impulses with a smaller maximum z_n, where z_ρ is a finite quantity, independent of N, defined in the theorem. Thus we need only consider sets of N impulses with all $z_n \leq z_\rho$. Theorem D. 4 states: Given any set of N impulses such that $0 \leq z_n \leq z_\rho$, there exists an $\epsilon_\rho > 0$ and a positive integer M_ρ (both independent of N, r, $\{p_n\}$) such that for any z_j in the range $0 \leq z_j \leq z_\rho$, $F_\rho^{(m_j)}(z) \neq 0$ for all z in the interval $z_j \leq z < z_j + \epsilon_\rho$, for some positive integer $m_j \leq M_\rho$. By $F_\rho^{(m)}(z)$ we mean the m^{th} derivative of $F_\rho(z)$.

Consider splitting the range $0 \leq z \leq z_\rho$ into intervals of width ϵ_ρ by letting $z_j = j\epsilon_\rho$, $j = 0, 1, \ldots, L_\rho$, where L_ρ is the smallest integer for which $\epsilon_\rho L_\rho \geq z_\rho$. Suppose there are $n_j \geq m_{j+1}$ nonzero impulses in the range $z_j \leq z \leq z_{j+1}$. Then $F_\rho(z)$ must have n_j zeros in the range if (D. 7) is to be satisfied. By the mean-value theorem, $F'_\rho(z)$ must have n_{j-1} zeros in the interval, and by repeated application to the derivatives of $F_\rho(z)$, $F_\rho^{(m_j)}(z)$ must have $n_j - m_j \geq 1$ zeros in the interval. By Theorem D. 4, this is impossible, so that at most $N_\rho = \sum_{j=0}^{L_\rho} m_j \leq (L_\rho+1)M_\rho$ nonzero impulses can lie in $(0, z_\rho)$ and still permit (D. 7) to be satisfied.

Theorem D. 2 states that any $N \geq 2$ impulses satisfying (D. 7) but not (D. 4. ρ) can be strictly improved by the addition of one more impulse. Suppose the best set of $N_1 \geq 2$ impulses does not satisfy (D. 4. ρ). Then some set (and certainly the best set) of $N_1 + 1$ impulses is strictly better. We now apply induction up to N_ρ, where either some set of $N_1 \leq N_\rho$ impulses satisfies (D. 4. ρ) or some set of $N_\rho + 1$ impulses is strictly better than the best set of N_ρ impulses. We have just shown that no set of $N_\rho + 1$ impulses restricted to $(0, z_\rho)$ can satisfy (D. 7). But (D. 7) must have a solution for any $N_\rho + 1$ z_n's, so at least one $p_n = 0$. This reduces us back to N_ρ impulses. Hence there can be no improvement in going from the best set of N_ρ to $N_\rho + 1$ impulses. Therefore, we have a contradiction, and some set of N_ρ or fewer impulses must satisfy (D. 4. ρ), thereby proving the theorem.

THEOREM D. 2

Any set of two or more impulses satisfying constraints (D. 6) and condition (D. 7) but not (D. 4. ρ) can be strictly improved by the addition of one more impulse.

Proof: Since (D. 4. ρ) is not satisfied, there exists some z_o such that $F_\rho(z_o) < 0$. Since p(x) consists of two or more impulses satisfying (D. 6) and (D. 7), there exists $z_1 < \alpha$, $z_2 > \alpha$ such that $p_1 > 0$, $p_2 > 0$, and $F_\rho(z_1) = F_\rho(z_2) = 0$.

Let

$$p_\Delta(x) = p(x) + \Delta[\beta_o u_o(x - \sqrt{z_o}) + \beta_1 u_o(x - \sqrt{z_1}) + \beta_2 u_o(x - \sqrt{z_2})] \quad \text{(D. 9)}$$

$$\beta_o = 1, \quad \beta_1 = \frac{z_o - z_2}{z_2 - z_1}, \quad \beta_2 = -\frac{z_o - z_1}{z_2 - z_1}. \quad \text{(D. 10)}$$

Note that $z_2 - z_1 > 0$, so that β_1 and β_2 are bounded. Since $p(x)$ contains the last two impulses with $p_1, p_2 > 0$, if Δ is chosen small enough and positive, $p_\Delta(x) \geq 0$, and $p_\Delta(x)$ has one more impulse than $p(x)$. Direct evaluation confirms that the other constraints in (D. 6) are also satisfied by $p_\Delta(x)$. Let $J_{\rho\Delta}$ be the new value of J_ρ resulting when $p_\Delta(x)$ is used with the same value of r as before. Then

$$J_{\rho\Delta} = J_\rho + \Delta\, e^{-2r\alpha}\left[2F_\rho(z_o) + \Delta \sum_{i=0}^{2} \sum_{j=0}^{2} \beta_i \beta_j H_1(\sqrt{z_i}, \sqrt{z_j})^{1/\rho} e^{r(z_i+z_j)}\right], \qquad \text{(D. 11)}$$

where we have used the fact that $F_\rho(z_1) = F_\rho(z_2) = 0$. All terms in the double summation are bounded, and $F_\rho(z_o) < 0$, so for some small positive Δ, $J_{\rho\Delta} < J_\rho$, and we have obtained a strict improvement by the addition of one more impulse.

THEOREM D. 3

Any set of N impulses for which some $z_n > z_\rho$, where

$$z_\rho = 2\begin{cases} e^{\frac{4(2+\alpha)}{\rho}}, & 0 < \rho < 1 \\ e^{4(2+\alpha)}, & 1 \leq \rho < \infty \end{cases} \qquad \text{(D. 12)}$$

has a value of J_ρ that is strictly greater than the J_ρ resulting from some set of $N_1 \leq N$ impulses, for which all $z_n \leq z_\rho$.

Proof: Consider a set of N impulses. If this set does not satisfy (D. 7), it may be altered by changing $\{p_n\}$, r so that (D. 7) is satisfied, with a strict reduction in J_ρ. Thus we need only consider sets for which (D. 7) is satisfied. Let z_N be the maximum z_n, by hypothesis greater than z_ρ, and

$$G_\rho(z) = e^{rz} \sum_{n=1}^{N} p_n\, e^{rz_n}\left[\frac{(1+z)^{1/2}\,(1+z_n)^{1/2}}{1+\frac{1}{2}(z+z_n)}\right]^{1/\rho}. \qquad \text{(D. 13)}$$

Consider $0 < \rho < 1$.

$$G_\rho(z_N) < e^{2rz_N} \sum_{n=1}^{N} p_n \left[\frac{(1+z_N)^{1/2}\,(1+z_n)^{1/2}}{\left\{1+\frac{1}{2}(z_N+z_n)\right\}^{(1/2+\rho/2)+(1/2-\rho/2)}}\right]^{1/\rho} \qquad \text{(D. 14)}$$

$$G_\rho(z_N) < e^{2rz_N}\left(\frac{1+z_N}{1+\frac{1}{2}z_N}\right)^{1/2\rho}\left(1+\frac{1}{2}z_N\right)^{-1/2} \sum_{n=1}^{N} p_n\left(1+\frac{z_n}{2}\right)^{1/2}\left[\frac{1+z_n}{1+\frac{1}{2}z_n}\right]^{1/2\rho}. \qquad \text{(D. 15)}$$

But $\left(1+\frac{1}{2}z_n\right)^{1/2} \leq 1 + z_n$, and $\frac{1+z_n}{1+\frac{1}{2}z_n} < 2$, so

$$G_\rho(z_N) < \frac{2^{1/\rho}(1+\alpha)}{\left(1+\frac{1}{2}z_N\right)^{1/2}} e^{2rz_N} < \frac{[2(1+\alpha)]^{1/\rho} e^{2rz_N}}{\left(1+\frac{1}{2}z_N\right)^{1/2}}. \tag{D. 16}$$

Now let $1 \leq \rho < \infty$. Making use of (A. 36) yields

$$G_\rho(z_N) < e^{2rz_N}\left[\sum_n p_n \left\{\frac{(1+z_n)^{1/2}\,(1+z_N)^{1/2}}{1+\frac{1}{2}(z_n+z_N)}\right\}\right]^{1/\rho} \tag{D. 17}$$

$$G_\rho(z_N) < e^{2rz_N}\left[\frac{(1+z_N)^{1/2}}{1+\frac{1}{2}z_N}\sum_{n=1}^{N} p_n(1+z_n)^{1/2}\right]^{1/\rho} \tag{D. 18}$$

$$G_\rho(z_N) < e^{2rz_N}\left[\frac{2^{1/2}}{\left(1+\frac{1}{2}z_N\right)^{1/2}}(1+\alpha)\right]^{1/\rho} < \frac{[2(1+\alpha)]^{1/\rho} e^{2rz_N}}{\left(1+\frac{1}{2}z_N\right)^{1/2\rho}}. \tag{D. 19}$$

By assumption, (D. 7) is satisfied, so $G_\rho(z_N) = J_\rho\, e^{2r\alpha}$, and

$$e^{2rz_N} > e^{2r(z_N-\alpha)} > \frac{J_\rho}{[2(1+\alpha)]^{1/\rho}}\begin{cases}\left(1+\frac{1}{2}z_N\right)^{1/2}, & 0 < \rho < 1\\ \left(1+\frac{1}{2}z_N\right)^{1/2\rho}, & 1 \leq \rho < \infty\end{cases} \tag{D. 20}$$

It is known[10] that $-\rho \ln J_\rho$ is an increasing function of ρ, and in Theorem B. 2 it was shown that

$$\lim_{\rho\to\infty} -\rho \ln J_\rho \leq \alpha E_\infty \cong 0.15\alpha. \tag{D. 21}$$

Hence,

$$r > \frac{1}{2z_N}\left[\frac{a_\rho}{2}\ln\left(1+\frac{1}{2}z_N\right) - \frac{1}{\rho}\left\{\alpha E_\infty + \ln 2 + \ln(1+\alpha)\right\}\right], \tag{D. 22}$$

where

$$a_\rho = \begin{cases}1, & 0 < \rho < 1\\ \frac{1}{\rho}, & 1 \leq \rho < \infty\end{cases} \tag{D. 23}$$

Differentiation of $F_\rho(z)$ yields

$$F'_\rho(z) = e^{rz}\,(1+z)^{1/2\rho} \sum_{n=1}^{N} \frac{p_n\, e^{rz_n}\,(1+z_n)^{1/2\rho}}{\left[1+\frac{1}{2}(z+z_n)\right]^{1/\rho}} \left\{ r + \frac{\frac{1}{2\rho}}{1+z} - \frac{\frac{1}{2\rho}}{1+\frac{1}{2}(z+z_n)} \right\}. \tag{D. 24}$$

If $r \geq 1/2\rho$, the bracketed term in (D. 24) is positive for all z, z_n so that $F_\rho(z)$ is a strictly increasing function. Hence $F_\rho(z)$ can have, at most, one zero, and no set of $N \geq 2$ impulses with $r \geq 1/2\rho$ can satisfy (D. 7). Thus we can restrict $r < 1/2\rho$. Using (D. 22), and dropping the second positive term, we obtain

$$F'_\rho(z_N) > e^{rz_N} \sum_{n=1}^{N} p_n\, e^{rz_n} \left[\frac{(1+z_n)^{1/2}\,(1+z_N)^{1/2}}{1+\frac{1}{2}(z_n+z_N)} \right]^{1/\rho}$$

$$\left\{ \frac{1}{2z_N}\left[\frac{a_\rho}{2} \ln\left(1+\frac{z_N}{2}\right) - \frac{1}{\rho}\left\{\alpha E_\infty + \ln 2 + \ln(1+\alpha) + 2\right\}\right]\right\}. \tag{D. 25}$$

Once again noting that $G_\rho(z_N) = J_\rho\, e^{2r\alpha}$, we have

$$F'_\rho(z_N) > \frac{J_\rho\, e^{2r\alpha}}{2z_N\rho}\left[\frac{\rho a_\rho}{2} \ln\left(1+\frac{1}{2}z_N\right) - \left[\alpha E_\infty + \ln 2 + 2 + \ln(1+\alpha)\right]\right]. \tag{D. 26}$$

Therefore, if

$$\frac{\rho a_\rho}{2} \ln\left(1+\frac{1}{2}z_N\right) > \alpha E_\infty + \ln 2 + 2 + \ln(1+\alpha) + 1, \tag{D. 27}$$

then

$$F'_\rho(z_N) > \frac{1}{2\rho z_N}\, e^{-\alpha E_\infty/\rho} > \frac{1}{2\rho z_N}\, e^{-\alpha/\rho}. \tag{D. 28}$$

Loosening inequality (D. 27), we find that, if we define z_ρ as in (D. 12), then if $z_n > z_\rho$, $F'_\rho(z_N) > \frac{1}{2\rho z_N}\, e^{-\alpha/\rho}$.

Consider $\epsilon > 0$, $z = z_N - \epsilon$. Then, by Taylor's Theorem,[38]

$$F_\rho(z) = -\epsilon F'_\rho(z_N) - \int_z^{z_N} (z-t)\, F''_\rho(t)\, dt. \tag{D. 29}$$

If $|F''_\rho(t)| \leq D$ over the region of integration, and $z_N > z_\rho$, then

$$F_\rho(z) < -\frac{\epsilon}{2z_N\rho}\, e^{-\alpha/\rho} + \frac{\epsilon^2 D}{2} < -\frac{\epsilon}{2z_N\rho}\, e^{-\alpha/\rho}\left[1 - \epsilon D z_N \rho\, e^{\alpha/\rho}\right]. \tag{D. 30}$$

Choosing

$$\epsilon_\rho = \frac{1}{2Dz_N\rho\, e^{a/\rho}},$$

we have

$$F_\rho(z_N-\epsilon_\rho) < -\frac{\epsilon_\rho}{4z_N\rho} e^{-a/\rho} < 0. \tag{D.31}$$

This depends on the properties of $F''_\rho(z)$, which we now investigate.

$$F''_\rho(z) = \left(r + \frac{\frac{1}{2\rho}}{1+z}\right) F'_\rho(z) + \frac{e^{rz}(1+z)^{1/2\rho}}{2\rho} \sum_{n=1}^{N} p_n e^{rz_n}(1+z_n)^{1/2\rho}$$

$$\left\{ \frac{-r}{\left[1+\frac{1}{2}(z+z_n)\right]^{1+1/\rho}} + \frac{\frac{1}{2}\left(\frac{1}{\rho}+1\right)}{\left[1+\frac{1}{2}(z+z_n)\right]^{2+1/\rho}} + \left[\frac{-1}{(1+z)^2\left[1+\frac{1}{2}(z+z_n)\right]^{1/\rho}} - \frac{\frac{1}{2\rho}}{(1+z)\left[1+\frac{1}{2}(z+z_n)\right]^{1+1/\rho}}\right]\right\}. \tag{D.32}$$

We recall that $r \leq 1/2\rho$, and restrict $z \leq z_N$. Then

$$|F''_\rho(z)| < \frac{1}{\rho}|F'_\rho(z)| + G_\rho(z)\left\{\frac{3(1+\rho)}{4\rho^2}\right\} \tag{D.33}$$

$$|F'_\rho(z)| < \frac{3}{2\rho} G_\rho(z) \tag{D.34}$$

$$|F''_\rho(z)| < G_\rho(z)\left\{\frac{9+3\rho}{4\rho^2}\right\} < \left(\frac{9+3\rho}{4\rho^2}\right) \sum_{n=1}^{N} p_n e^{r(z+z_n)} \tag{D.35}$$

$$|F''_\rho(z)| < \left(\frac{9+3\rho}{4\rho^2}\right) e^{z_N/\rho} \equiv D, \qquad 0 \leq z \leq z_N. \tag{D.36}$$

Note that D, and hence ϵ_ρ, depends only on ρ and z_N and not on $\{p_n, z_n\}$, r. Therefore, regardless of r and any other impulses, if $p_N > 0$, $F_\rho(z_N-\epsilon_\rho) < 0$ when $z_N > z_\rho$. By Theorem D. 2, J_ρ may be strictly reduced by adding an impulse at $z_N - \epsilon_\rho$. Consider a new set of $\{z_n\}$, consisting of the original set plus one at $z_N - \epsilon_\rho$. We must get a strict improvement if we optimize on r, $\{p_n\}$ so that (D.7) is satisfied. If $p_N > 0$, $F_\rho(z_N-\epsilon_\rho) < 0$, and (D. 7) cannot be satisfied. Therefore $p_N = 0$, and the impulses at z_N may be dropped in favor of the one at $z_N - \epsilon_\rho$, thereby resulting in a strict improvement. Therefore, unless all impulses are at values of $z_n < z_\rho$, a new set of N (or less) impulses may be found that provides a strict decrease in J_ρ, thereby proving the theorem.

THEOREM D. 4

Let M_ρ be the smallest integer greater than $\left[1 + \frac{1+z_\rho}{\rho} + \frac{2}{\rho}(1+z_\rho)^2\right]$, and $F_\rho^{(m)}(z)$ denote the m^{th} derivation of $F_\rho(z)$. Given any set of N impulses such that $0 \leq z_n \leq z_\rho$, there exists an $\epsilon > 0$, independent of N, such that, for any z_j in the range $0 \leq z_j \leq z_\rho$, $F_\rho^{(m_j)}(z) \neq 0$ for all z in the interval $z_j \leq z \leq z_j + \epsilon$, for some positive integer $m_j \leq M_\rho$.

Proof: From the form of $F_\rho(z)$ given by (D. 8), we see that $F_\rho(z)$ and all of its derivatives exist, are continuous, and can be bounded independently of N in $0 \leq z \leq z_\rho$ by methods similar to those used in (D. 32) through (D. 36). Each term of the sum can be bounded in terms of z_ρ, and $\sum_{n=1}^{N} p_n = 1$, so that any derivative can be bounded independent of N. Therefore, if we can show that there exists a positive Δ, independent of N, such that $|F_\rho^{(m_j)}(z_j)| > \Delta$ for some integer m_j in the range $1 \leq m_j \leq M_\rho$ and any z_j in the interval, then the continuity property of $F_\rho^{(m_j)}(z)$ may be invoked to prove the theorem.

Suppose the opposite is true, and there exists a value of z such that $|F_\rho^{(m)}(z)| \leq \Delta$, for all $\Delta > 0$, $1 \leq m \leq M_\rho$ and some $\{p_n\}$, r. After some manipulation, we find

$$F'_\rho(z) = a(z)\, b(z) \tag{D. 37}$$

$$a(z) = 2^{1/\rho}\, e^{rz}\, (1+z)^{\left(\frac{1}{2\rho} - 1\right)} \tag{D. 38}$$

$$b(z) = \sum_{n=1}^{N} p_n\, e^{rz_n}\, (1+z_n)^{1/2\rho} \Bigg[r(2+z+z_n)^{-\frac{1}{\rho}+1} + \left\{\frac{1}{2\rho} - r(1+z_n)\right\} (2+z+z_n)^{-1/\rho} - \frac{1}{\rho}(2+z+z_n)^{-\left(\frac{1}{\rho}+1\right)}\Bigg]. \tag{D. 39}$$

Since $0 \leq r \leq 1/2\rho$, a(z) and all of its derivatives must be bounded in the range $0 \leq z \leq z_\rho$. Note that

$$a(z) \geq 2^{1/\rho} \begin{cases} 1 & 0 < \rho < \frac{1}{2} \\ (1+z_\rho)^{1/2\rho - 1} & \frac{1}{2} \leq \rho \end{cases} \tag{D. 40}$$

If $F'_\rho(z)$ is to vanish with Δ for some z, then, since $a(z)$ is bounded away from zero, $|b(z)| \leq c_0\Delta$, where c_0 is a finite constant. Differentiating again, we have

$$F''_\rho(z) = a'(z)\, b(z) + a(z)\, b'(z). \tag{D. 41}$$

If $b(z)$ vanishes with Δ, the first term in (D. 41) will also vanish with Δ because $a'(z)$ is bounded, so if $F''_\rho(z)$ is to vanish with Δ, then $|b'(z)| \leq c_1\Delta$, where c_1 is another finite constant. Repeat this procedure, and if $F_\rho^{(m)}(z)$ and all previous derivatives are to vanish with Δ for some z, then $|b^{(j)}(z)| \leq c_j\Delta$, $j = 0, 1, \ldots, m-1$. Investigation of the derivatives of $b(z)$ reveals that

$$b^{(m)}(z) = (-1)^m (1/\rho)(1/\rho+1)\ldots(1/\rho+m-2) \sum_{n=1}^{N} p_n e^{rz_n} (1+z_n)^{1/2\rho} (2+z+z_n)^{-(1/\rho+m)} T_{mn} \tag{D. 42}$$

$$T_{mn} = \frac{1}{\rho} + m - 1 + \left\{r(1+z_n) - \frac{1}{2\rho}\right\}(2+z+z_n) - \frac{r\left(\frac{1}{\rho}-1\right)(2+z+z_n)^2}{\frac{1}{\rho}+m-2}. \tag{D. 43}$$

As m gets large, T_{mn} becomes positive; in fact,

$$T_{mn} > m - \left[1 + \frac{1}{\rho}(1+z_\rho) + \frac{2}{\rho}(1+z_\rho)^2\right]. \tag{D. 44}$$

Thus when $m = M_\rho > 1 + \frac{1}{\rho}(1+z_\rho) + \frac{2}{\rho}(1+z_\rho)^2$, $|b^{(m)}(z)|$ is strictly bounded away from zero. Thus all of the first M_ρ derivatives of $F_\rho(z)$ cannot be arbitrarily small, and the rest of the theorem follows.

THEOREM D. 5

Condition (D. 4. s) is satisfied by some r, $p(x)$ combination also satisfying conditions (D. 2), where $p(x)$ consists of a finite number of impulses.

Proof: The proof is analogous to the proof of Theorem D. 1. Once again, $N = 1$ may be disregarded, and with $N \geq 2$, specify the set $\{z_n\}$ and minimize

$$J_s = \int_0^\infty \left[\sum_{n=1}^{N} p_n e^{r(z_n-a)} p(y|\sqrt{z_n})^{1-s}\right]^{1/(1-s)} dy \tag{D. 45}$$

over r, $\{p_n\}$, subject to the constraints (D. 6). The integral on y does not affect the discrete character of the optimization, and a minor modification of Gallager's work[10] shows that a necessary and sufficient condition for r, $\{p_n\}$ to minimize J_s is

$$F_s(z_n) \geq 0 \tag{D. 46}$$

for all n, with equality when $p_n > 0$, where

$$F_s(z) = \int_0^\infty \left[\sum_{n=1}^{N} p_n e^{rz_n} p(y|\sqrt{z_n})^{1-s}\right]^{s/(1-s)} e^{rz} p(y|\sqrt{z})^{1-s}\, dy - J_s e^{ra/(1-s)}. \tag{D.47}$$

Once again we find that (D. 46) must have a solution, so we restrict the analysis to sets of impulses that satisfy (D. 46).

Theorem D. 7 states than any set of N impulses for which $z_N > z_s$ has a value of J_s that can be strictly reduced by some set of N or fewer impulses with a smaller maximum z_n, where z_s is a finite quantity, independent of N, defined in the theorem. Thus we need only consider sets of N impulses with all $z_n \leq z_s$, so that there must be some best set of N impulses. Theorem D. 8 states: Given any set of N impulses such that $0 \leq z_n \leq z_s$, and $r \geq K(1-s)/z_s$, there exists an $\epsilon_s > 0$ and a positive integer M_s (both independent of N, r, $\{p_n\}$) such that, for any z_j in the range $0 \leq z_j \leq z_s$, $F_s^{(m_j)}(z) \neq 0$ for all z in the interval $z_j \leq z < z_j + \epsilon_s$, for some positive integer $m_j \leq M_s$.

Then with reasoning identical to that used in the proof of Theorem D. 1, we can show that, at most, $N_s < \infty$ nonzero impulses can lie in $0 \leq z \leq z_s$ and still permit (D. 46) to be satisfied, if $r \geq K(1-s)/z_s$. If $r < K(1-s)/z_s$, there could conceivably be more. In the last case, however, it can be shown that $F_s(z_s) < 0$. Theorem D. 6 states that any $N \geq 2$ impulses satisfying (D. 46) but not (D. 4. s) can be strictly improved by the addition of one more impulse, at the point where $F_s(z) < 0$. Hence, if $r < K(1-s)/z_s$, we could achieve improvement by adding another impulse at z_s. If this new set is optimized, we find that $r \geq K(1-s)/z_s$, or else (D.46) could not be satisfied; hence, at most, N_s impulses can be used, with strict improvement. Once again we see that there can be no improvement in going from the best set of N_s impulses to $N_s + 1$, and this fact can be used, as in the proof of Theorem D. 1, to complete the present proof.

THEOREM D. 6

Any set of two or more impulses satisfying constraints (D. 6) and condition (D. 46) but not (D. 4. s) can be strictly improved by the addition of one more impulse.

Proof: Since (D. 4. s) is not satisfied, there exists some z_o such that $F_s(z_o) < 0$. Since p(x) consists of two or more impulses satisfying (D. 6) and (D. 46), there exist $z_1 < a$, $z_2 > a$ such that $p_1 > 0$, $p_2 > 0$, $F_s(z_1) = F_s(z_2) = 0$. Let $p_\Delta(x)$ be defined again by (D. 9) and (D. 10). Let $J_{s\Delta}$ be the new value of J_s resulting when $p_\Delta(x)$ is used with the same value of r as before. Then

$$J_{s\Delta} = e^{-ra/(1-s)} \int_0^\infty \left[\sum_{n=1}^{N} p_n p(y|\sqrt{z_n})^{1-s} e^{rz_n} + \Delta \sum_{n=0}^{2} \beta_n e^{rz_n} p(y|\sqrt{z_n})^{1-s}\right]^{1/(1-s)} dy \tag{D.48}$$

$$J_{s\Delta} = e^{-r\alpha/(1-s)} \int_0^\infty \beta(y)^{1/(1-s)} \left[1 + \Delta \frac{\sum_{n=0}^{2} \beta_n e^{rz_n} p(y|\sqrt{z_n})^{1-s}}{\beta(y)}\right]^{1/(1-s)} dy \tag{D. 49}$$

$$\beta(y) = \sum_{n=1}^{N} p_n e^{rz_n} p(y|\sqrt{z_n})^{1-s}. \tag{D. 50}$$

Consider first $0 < s \leq \frac{1}{2}$. From (C. 38), we see that

$$(1+t)^{1/(1-s)} \leq 1 + \frac{t}{1-s} + \begin{cases} t^{1/(1-s)}, & t \geq 0 \\ \frac{s}{(1-s)^2} t^2, & -\frac{1}{2} \leq t < 0 \end{cases} \tag{D. 51}$$

Let

$$t(y) \equiv \Delta \frac{\sum_{n=0}^{2} \beta_n e^{rz_n} p(y|\sqrt{z_n})^{1-s}}{\beta(y)}. \tag{D. 52}$$

Only β_1 and β_2 can possibly be negative, so that if we choose Δ small enough that $2\Delta|\beta_1| < p_1$, $2\Delta|\beta_2| < p_2$, then $t(y) > -\frac{1}{2}$ and we may apply (D. 51). Let Y be the set of $y \geq 0$ for which $t(y) \geq 0$, and $\overline{Y}$ be $y \geq 0$ for which $-\frac{1}{2} < t(y) < 0$. Then

$$J_{s\Delta} \leq e^{-r\alpha/(1-s)} \left[\int_0^\infty \beta(y)^{1/(1-s)} \left\{1 + \frac{t(y)}{1-s}\right\} dy + \int_Y \beta(y)^{1/(1-s)} t(y)^{1/(1-s)} dy + \frac{s}{(1-s)^2} \int_{\overline{Y}} \beta(y)^{1/(1-s)} t^2(y)\, dy\right] \tag{D. 53}$$

$$J_{s\Delta} \leq J_s + \frac{\Delta e^{-r\alpha/(1-s)}}{1-s} \sum_{n=0}^{2} \beta_n \int_0^\infty \beta(y)^{s/(1-s)} p(y|\sqrt{z_n})^{1-s} e^{rz_n} dy + e^{-r\alpha/(1-s)} [I_1 + I_2], \tag{D. 54}$$

where I_1 and I_2 are the last two integrals in (D. 53). Making use of the fact that the β_n sum to zero, and $F(z_1) = F(z_2) = 0$, we find

$$J_{s\Delta} - J_s \leq \frac{\Delta}{1-s} e^{-r\alpha/(1-s)} F(z_o) + e^{-r\alpha/(1-s)} [I_1 + I_2] \tag{D. 55}$$

$$I_1 = \Delta^{1/(1-s)} \int_Y \left[\sum_{n=0}^{2} \beta_n e^{rz_n} p(y|\sqrt{z_n})^{1-s}\right]^{1/(1-s)} dy$$

$$\leq \Delta^{1/(1-s)} \int_0^\infty \left[\sum_{n=0}^{2} |\beta_n| e^{rz_n} p(y|\sqrt{z_n})^{1-s}\right]^{1/(1-s)} dy. \tag{D. 56}$$

Using (A.36) it may be shown that

$$I_1 \leq \Delta^{1/(1-s)} \left[\sum_{n=0}^{2} |\beta_n| \, e^{rz_n} \right]^{1/(1-s)} . \tag{D.57}$$

Consider the last integral,

$$I_2 = \frac{\Delta^{1/(1-s)} s}{(1-s)^2} \int_{\overline{Y}} \beta(y)^{1/(1-s)} \left[\frac{\Delta^{\frac{1}{2}\left(\frac{1-2s}{1-s}\right)} \sum_{n=0}^{2} \beta_n \, e^{rz_n} \, p(y|\sqrt{z_n})^{1-s}}{\beta(y)} \right]^2 dy. \tag{D.58}$$

We can certainly choose Δ small enough that

$$\Delta^{\frac{1}{2}\left(\frac{1-2s}{1-s}\right)} |\beta_1| < \frac{p_1}{2}, \qquad \Delta^{\frac{1}{2}\left(\frac{1-2s}{1-s}\right)} |\beta_2| < \frac{p_2}{2}.$$

Then

$$I_2 \leq \frac{s}{(1-s)^2} \Delta^{1/(1-s)} \int_{\overline{Y}} \left(\frac{1}{4}\right) \beta(y)^{1/(1-s)} \, dy \leq \frac{s}{4(1-s)^2} J_s \, e^{+ra/(1-s)} \, \Delta^{1/(1-s)} \tag{D.59}$$

$$J_{s\Delta} - J_s \leq \frac{\Delta}{1-s} e^{-ra/(1-s)} \left[F(z_o) + \Delta^{s/(1-s)} \left\{ (1-s) \left[\sum_{n=0}^{2} |\beta_n| \, e^{rz_n} \right]^{1/(1-s)} + \frac{s}{4(1-s)} J_s \, e^{+ra/(1-s)} \right\} \right]. \tag{D.60}$$

The last terms are bounded, so that for some small but positive Δ, $J_{s\Delta} < J_s$, and a strict improvement has been obtained by the addition of one more impulse, when $0 < s \leq \frac{1}{2}$.

When $\frac{1}{2} < s < 1$, we can make use of the inequality

$$(1+t)^{1/(1-s)} \leq 1 + \frac{t}{1-s} + \begin{cases} \dfrac{s}{(1-s)^2} t^2 & -1 \leq t \leq t_o \\ \left[\left(\dfrac{1+t_o}{t_o} \right) t \right]^{1/(1-s)} & t > t_o \end{cases} \tag{D.61}$$

where $(1+t_o)^{(2s-1)/(1-s)} = 2$ defines t_o. The proof of (D.61) is similar to the proof of (D.51) and will be omitted. An analysis similar to the one just performed shows that

$$J_{s\Delta} - J_s \leq \frac{\Delta\, e^{-r\alpha/(1-s)}}{1-s}\left[F(z_o) + \Delta^{\frac{1}{2}}\left\{\frac{s(1+t_o)^2}{1-s}\, J_s\, e^{r\alpha/(1-s)} + (1-s)\left[\left(1+\frac{1}{t_o}\right)\left[\sum_{n=0}^{2} |\beta_n|\, e^{rz_n}\right]\right]^{1/(1-s)}\right\}\right] \tag{D. 62}$$

for some $\frac{1}{2} < s < 1$, so that for some small positive Δ, $J_{s\Delta} < J_s$, thereby proving the theorem for all $0 < s < 1$.

THEOREM D. 7

Any set of N impulses with some $z_N > z_s$, where

$$z_s = \left[e^{\frac{K(1+\alpha)}{2}}\, e^{(1+\alpha)(K+1)} + e^{(\alpha+K+1)}\,(1+\alpha)2^K\,\frac{\Gamma(2K)}{\Gamma(K)}\right]^{a_s} \tag{D. 63}$$

$$a_s = \begin{cases} 2 & \frac{1}{2} \leq s < 1 \\ \frac{1-s}{s^2} & 0 < s < \frac{1}{2} \end{cases} \tag{D. 64}$$

has a value of J_s that is strictly greater than the J_s resulting from some set of $N_1 \leq N$ impulses, for which all $z_n \leq z_s$. Note that z_s is a continuous function of s and is finite for all s, $0 < s \leq 1$.

Proof: Again, we need only consider sets for which (D. 46) is satisfied. Let z_N be the maximum z_n, by hypothesis greater than z_s, and

$$G_s(z) \equiv e^{rz} \int_0^\infty \beta(y)^{s/(1-s)}\, p(y\,|\sqrt{z}\,)^{1-s}\, dy. \tag{D. 65}$$

Consider $0 < s \leq \frac{1}{2}$. Then, by (A. 36),

$$G_s(z_N) \leq e^{rz_N}\left[\int_0^\infty p(y|\sqrt{z_N})^{1-s}\, dy\right]^{1-\frac{s}{1-s}}\left[\int_0^\infty p(y\,|\sqrt{z_N})^{1-s}\,\beta(y)\, dy\right]^{s/(1-s)}. \tag{D. 66}$$

Evaluation of the integrals leads to

$$G_s(z_N) \leq e^{rz_N}\left[\frac{\Gamma[K-sK+s](1+z_N)^s}{\Gamma(K)^{1-s}\,(1-s)^{K(1-s)+s}}\right]^{1-\frac{s}{1-s}} \left[\sum_{n=1}^{N} \frac{p_n\, e^{rz_n}\,\Gamma[1+2(K-1)(1-s)]\,\{(1+z_n)(1+z_N)\}^{1+(K-2)(1-s)}}{\Gamma(K)^{2(1-s)}\,\{(1-s)(2+z+z_n)\}^{1+2(K-1)(1-s)}}\right]^{s/(1-s)}. \tag{D. 67}$$

Recall that $K \geq 1$, so that

$$\Gamma[K+s(1-K)] \leq \Gamma(K) \tag{D.68}$$

$$\Gamma[1+2(K-1)(1-s)] < \Gamma(2K) \tag{D.69}$$

$$G_s(z_N) < e^{rz_N}\left[\frac{\Gamma(2K)}{\Gamma(K)}\right]^{s/(1-s)} \frac{(1+z_N)^{Ks}}{(1-s)^K}$$

$$\left[\sum_{n=1}^{N} \frac{p_n\, e^{rz_n}\,(1+z_n)^{1+(K-2)(1-s)}}{(2+z_n+z_N)^{(K-1)(1-s)+1+(K-1)(1-s)}}\right]^{s/(1-s)} \tag{D.70}$$

$$G_s(z_N) \leq e^{rz_N}\left[\frac{\Gamma(2K)}{\Gamma(K)}\right] \frac{(1+z_N)^{Ks}}{(1-s)^K\,(1+z_N)^{[1+(K-1)(1-s)]s/(1-s)}}$$

$$\left[\sum_{n=1}^{N} p_n\, e^{rz_n}\,(1+z_n)^s\right]^{s/(1-s)} \tag{D.71}$$

$$G_s(z_N) \leq e^{rz_N}\left[\frac{\Gamma(2K)}{\Gamma(K)}\right] \frac{(1+z_N)^{-s^2/(1-s)}}{(1-s)^K}\left[e^{rz_N}\,(1+\alpha)\right]^{s/(1-s)} \tag{D.72}$$

$$G_s(z_N) < e^{rz_N/(1-s)}\left[(1+\alpha)2^K\,\frac{\Gamma(2K)}{\Gamma(K)}\right](1+z_N)^{-s^2/(1-s)}. \tag{D.73}$$

Now consider $\frac{1}{2} < s < 1$.

$$G_s(z_N) = \int_0^\infty \left[e^{rz_N/(1-s)}\,p(y\,|\sqrt{z_N})\right]^{1-s} \beta(y)^{s/(1-s)}\,dy \equiv \int_0^\infty H_o(y)\,dy \tag{D.74}$$

$$G_s(z_N) - \int_0^\infty \beta(y)^{1/(1-s)}\,dy =$$

$$\int_0^\infty H_o(y)\left[1 - \sum_{n=1}^{N} p_n \exp(1-s)\left[\frac{rz_n}{1-s} - \frac{rz_N}{1-s} + \ln\left\{\frac{p(y\,|\sqrt{z_n})}{p(y\,|\sqrt{z_N})}\right\}\right]\right] dy. \tag{D.75}$$

But $e^t \geq 1 + t$, and, by hypothesis, $G_s(z_N) = \int_0^\infty \beta(y)^{1/(1-s)}\,dy$, so

$$0 \leq (1-s)\int_0^\infty H_o(y) \sum_n p_n\left[-\frac{rz_n}{1-s} + \frac{rz_N}{1-s} + \ln\left\{\frac{p(y\,|\sqrt{z_N})}{p(y\,|\sqrt{z_n})}\right\}\right] dy \tag{D.76}$$

$$0 \leq \frac{rz_N}{1-s} - \frac{r\alpha}{1-s} - K \ln(1+z_N) + K \sum_n p_n \ln(1+z_n)$$

$$+ \frac{\int_0^\infty H_o(y)\left[-\frac{y}{1+z_N} + y \sum_n \frac{p_n}{1+z_n}\right] dy}{\int_0^\infty H_o(y)\, dy} \tag{D. 77}$$

$$0 \leq \frac{rz_N}{1-s} - K \ln(1+z_N) + \frac{\int_0^\infty yH_o(y)\, dy}{\int_0^\infty H_o(y)\, dy}. \tag{D. 78}$$

By hypothesis, $\int_0^\infty H_o(y)\, dy = \int_0^\infty \beta(y)^{1/(1-s)}\, dy = J_s\, e^{-r\alpha/(1-s)}$. Note that $-\ln J_s$ is an increasing function of s, and recall that we have already shown (Theorem C. 1) that

$$\lim_{s\to 1} -\ln J_s \leq 0.22\alpha. \tag{D. 79}$$

But

$$\int_0^\infty yH_o(y)\, dy \leq e^{rz_N/(1-s)} \int_0^\infty y\left[\sum_n p_n \left\{p(y|\sqrt{z_n})^s\, p(y|\sqrt{z_N})^{(1-s)}\right\}^{(1-s)/s}\right]^{s/(1-s)} dy \tag{D. 80}$$

and, by applying (A. 36) once more,

$$\int_0^\infty yH_o(y)\, dy \leq e^{rz_N/(1-s)} \sum_{n=1}^{N} p_n \int_0^\infty yp(y|\sqrt{z_n})^s\, p(y|\sqrt{z_N})^{1-s}\, dy \tag{D. 81}$$

$$\int_0^\infty yH_o(y)\, dy \leq K\, e^{rz_N/(1-s)} \sum_n p_n \left[\frac{(1+z_n)(1+z_N)}{s(1+z_N)+(1-s)(1+z_n)}\right]\left[\frac{(1+z_n)^{1-s}\,(1+z_N)^s}{s(1+z_N)+(1-s)(1+z_n)}\right]^K. \tag{D. 82}$$

We note that

$$\frac{(1+z_n)(1+z_N)}{s(1+z_N)+(1-s)(1+z_n)} \leq \frac{1+z_n}{s}, \quad \frac{(1+z_n)^{1-s}\,(1+z_N)^s}{s(1+z_N)+(1-s)(1+z_n)} \leq 1, \tag{D. 83}$$

so that

$$\int_0^\infty yH_o(y)\, dy \leq \frac{K(1+\alpha)}{s}\, e^{rz_N/(1-s)} \leq 2K(1+\alpha)\, e^{rz_N/(1-s)}. \tag{D. 84}$$

Putting (D. 84) and (D. 79) into (D. 78), and loosening the inequality still further, we obtain

$$K \ln (1+z_N) < 4K(1+\alpha)\, e^{\alpha(1+K)}\, e^{rz_N/(1-s)}, \qquad \tfrac{1}{2} < s < 1 \tag{D. 85}$$

$$\frac{rz_N}{1-s} > \ln [\ln (1+z_N)] - \alpha(K+1) - \ln [K(1+\alpha)], \qquad \tfrac{1}{2} < s < 1. \tag{D. 86}$$

From (D. 73), we find

$$\frac{rz_N}{1-s} > \frac{s^2}{1-s} \ln (1+z_N) - .22\alpha - K \ln 2 - \ln (1+\alpha) - \ln \left[\frac{\Gamma(2K)}{\Gamma(K)}\right] \qquad 0 < s \leq \tfrac{1}{2}. \tag{D. 87}$$

Differentiation of $F_s(z)$ yields

$$F_s'(z) = e^{rz} \int_0^\infty \beta(y)^{s/(1-s)}\, p(y\,|\sqrt{z}\,)^{1-s} \left[r - \frac{K(1-s)}{1+z} + \frac{(1-s)y}{(1+z)^2}\right] dy. \tag{D. 88}$$

If $r \geq K(1-s)$, $F_s'(z) > 0$, and $F_s(z)$ is strictly increasing so that no set of two or more impulses can satisfy (D. 7). Thus, by restricting $r < K(1-s)$, and dropping the last term of (D. 88),

$$F_s'(z_N) > e^{rz_N} \int_0^\infty \beta(y)^{s/(1-s)}\, p(y\,|\sqrt{z_N})^{1-s} \left\{r - \frac{K(1-s)}{1+z_N}\right\} dy. \tag{D. 89}$$

We now apply the lower bounds to r, given by (D. 86) and (D. 87), in the same fashion as in the Proof of Theorem D. 3, and by loosening the inequalities somewhat, we find that, if $z_N > z_s$, $F_s'(z_N) > \frac{1-s}{z_N} e^{-.22\alpha}$, where z_s is defined by (D. 63). Note that z_s is a continuous function of s, and is strictly bounded for $0 < s \leq 1$.

As in Theorem D. 3, it may be shown that $|F_s''(t)| \leq D_s$, $0 \leq t \leq z_N$, where D_s is a finite quantity depending only on s, K, α, and z_N. Therefore, by the same reasoning that was used in Theorem D. 3, if $z_N > z_s$, there exists an ϵ_s such that

$$F_s(z_N - \epsilon_s) < \frac{-\epsilon_s(1-s)}{4z_N} e^{-.22\alpha} < 0. \tag{D. 90}$$

The same line of reasoning as before completes the proof.

THEOREM D. 8

Let M_s be the smallest positive integer m for which

$$\frac{K(1-s)}{z_s(2+z_s)^{b_s}} - m^K[K(1-s)+m]\, e^{sKz_s}\, c_s > 0, \tag{D. 91}$$

where

$$b_s = \begin{cases} K(1+s), & 0 < s \leqslant \frac{1}{2} \\ \frac{2Ks}{1-s}, & \frac{1}{2} < s < 1 \end{cases} \tag{D. 92}$$

$$c_s = \begin{cases} \left[1 + \frac{1}{1+z_s}\right]^{\frac{-ms}{1-s}}, & 0 < s \leqslant \frac{1}{2} \\ \left[1 + \frac{s}{1-s}\left(\frac{1}{1+z_s}\right)\right]^{-m}, & \frac{1}{2} < s < 1 \end{cases} \tag{D. 93}$$

Given any set of N impulses such that $0 \leqslant z_n \leqslant z_s$ and $r \geqslant K(1-s)/z_s$, there exists an $\epsilon > 0$ and independent of N, such that, for any z_j in the range $0 \leqslant z_j \leqslant z_s$, $F_s^{(m_j)} \neq 0$ for all z in the interval $z_j \leqslant z < z_j + \epsilon$, for some positive integer $m_j \leqslant M_s + 1$.

Proof:

$$F_s(z) = e^{rz} (1+z)^{-K(1-s)} \int_0^\infty g(y)\, e^{\frac{-(1-s)y}{1+z}} dy - J_s\, e^{r\alpha/(1-s)} \tag{D. 94}$$

$$g(y) = \beta(y)^{s/(1-s)} \left[\frac{y^{K-1}}{\Gamma(K)}\right]^{1-s} . \tag{D. 95}$$

Differentiating, we obtain

$$F_s'(z) = a_1(z)\, b_1(z) \tag{D. 96}$$

$$a_1(z) = e^{rz} (1+z)^{-K(1-s)-1} \tag{D. 97}$$

$$b_1(z) = \int_0^\infty g(y)\, e^{\frac{-(1-s)y}{1+z}} \left[r(1+z) - K(1-s) + \frac{y(1-s)}{1+z}\right] dy \tag{D. 98}$$

$$b_1(z) = \left[r(1+z) - K(1-s)\right] I_z(1) + \left(\frac{1-s}{1+z}\right) I_z(y), \tag{D. 99}$$

where

$$I_z(y^m) = \int_0^\infty g(y)\, e^{\frac{-(1-s)y}{1+z}} y^m\, dy. \tag{D. 100}$$

We see that $F_s^{(m)}(z)$ will consist of a sum of terms involving $I_z(y^n)$, $n = 0, 1, \ldots, m$, with bounded coefficients if $0 \leqslant z \leqslant z_s$. Thus if $I_z(y^n)$ is bounded in the region $0 \leqslant z \leqslant z_s$ for all $n \leqslant m$ (and we shall soon show that this is the case), then $F_s^{(m)}(z)$ will be bounded and continuous in the same region. Therefore, if we can show that there exists a positive Δ such that $|F_s^{(m_j)}(z_j)| > \Delta$ for some integer m_j in the range $1 \leq m_j \leq M_s$ and any z_j in the interval, then the continuity property of $F_s^{(m_j)}(z)$ may be invoked to prove the theorem.

Suppose the opposite is true, and there exists some z such that $|F_s^{(m)}(z)| \leq \Delta$, all $\Delta > 0$, $1 \leq m \leq M_s$. Since $a_1(z)$ is bounded away from zero, then $|b_1(z)| \leq c_1\Delta$, where c_1 is a bounded constant. Differentiating again, we obtain

$$F_s^{(2)}(z) = a_1'(z)\, b_1(z) + a_2(z)\, b_2(z) \tag{D.101}$$

$$a_2(z) = a_1(z)(1+z)^{-2} \tag{D.102}$$

$$b_2(z) = \int_0^\infty g(y)\, e^{\frac{-(1-s)y}{1+z}} \left[r(1+z)^2 + (1-s)y \left\{ r(1+z) - K(1-s) - 1 + \frac{(1-s)y}{1+z} \right\} \right] dy. \tag{D.103}$$

Thus, since $a_1'(z)$ is bounded, and $|b_1(z)| \leq c_1\Delta$, and $a_2(z)$ is bounded away from zero, we require $|b_2(z)| \leq c_2\Delta$, where c_2 is a bounded constant. Proceeding in the same manner, we have $|b_m(z)| \leq c_m\Delta$, where $b_m(z) = (1+z)^2\, b_{m-1}'(z)$. Thus we must investigate $b_m(z)$. Differentiation of (D.103) leads to

$$b_3(z) = \int_0^\infty g(y)\, e^{\frac{-(1-s)y}{1+z}} \left[2r(1+z)^3 + 2r(1+z)^2\,(1-s)y + [(1-s)y]^2 \left\{ r(1+z) - K(1-s) - 2 + \frac{(1-s)y}{1+z} \right\} \right] dy. \tag{D.104}$$

At each step, there will be only the two negative terms, and a little thought shows that

$$b_{m+1}(z) = m!\,r(1+z)^{m+1}\, I_z(1) - (1-s)^m\, [K(1-s)+m]\, I_z(y^m) + \text{positive terms.} \tag{D.105}$$

Consider $0 \leq s \leq \frac{1}{2}$.

$$I_z[y^m] = \int_0^\infty \frac{y^{K+m-1}}{\Gamma(K)}\, e^{\frac{-(1-s)y}{1+z}} \left[\sum_{n=1}^{N} \frac{p_n\, e^{rz_n}\, e^{\frac{-(1-s)y}{1+z_n}}}{(1+z_n)^{K(1-s)}} \right]^{s/(1-s)} dy. \tag{D.106}$$

Making use of (A.36), we find

$$I_z(y^m) \leq \frac{\Gamma(m+K)}{\Gamma(K)} \left(\frac{1+z}{1-s}\right)^{m+K} \left[\sum_{n=1}^{N} \frac{p_n\, e^{rz_n}}{(1+z_n)^{K(1-s)}} \left\{ \frac{1}{1+\left(\frac{1+z}{1+z_n}\right)} \right\}^{m+K} \right]^{s/(1-s)} \tag{D.107}$$

$$I_z(y^m) \leqslant \frac{\Gamma(m+K)}{\Gamma(K)} \left(\frac{1+z}{1-s}\right)^{m+K} e^{\frac{rsz_s}{1-s}} \left[1 + \frac{1}{1+z_s}\right]^{\frac{-ms}{1-s}}. \tag{D. 108}$$

Now consider $\frac{1}{2} < s < 1$.

$$I_z(y^m) \leqslant \frac{e^{\frac{rsz_N}{1-s}}}{\Gamma(K)} \int_0^\infty y^{K+m-1} e^{\frac{-(1-s)y}{1+z}} \left[\sum_{n=1}^{N} p_n \left\{\frac{e^{\frac{-sy}{1+z_n}}}{(1+z_n)^{Ks}}\right\}^{(1-s)/s}\right]^{s/(1-s)} dy.$$

(D. 109)

Using (A. 36) again, we find

$$I_z(y^m) \leqslant \frac{\Gamma(m+K)}{\Gamma(K)} \left(\frac{1+z}{1-s}\right)^{m+K} e^{\frac{rsz_N}{1-s}} \left[1 + \frac{s}{1-s}\left(\frac{1}{1+z_s}\right)\right]^{-m}. \tag{D. 110}$$

Thus

$$I_z(y^m) \leqslant \frac{\Gamma(m+K)}{\Gamma(K)} \left(\frac{1+z}{1-s}\right)^{m+K} e^{\frac{rsz_s}{1-s}} c_s, \tag{D. 111}$$

where c_s was defined by (D. 93).

Now consider $I_z(1)$. When $0 < s \leqslant \frac{1}{2}$, it may be shown that

$$I_z(1) \geqslant \left(\frac{1+z}{1-s}\right)^K (2+z_s)^{-K(1+s)}. \tag{D. 112}$$

Similarly, in the range $\frac{1}{2} < s < 1$, it may be shown that

$$I_z(1) \geqslant \left(\frac{1+z}{1-s}\right)^K (2+z_s)^{\frac{-2Ks}{1-s}}, \tag{D. 113}$$

and so

$$I_z(1) \geqslant \left(\frac{1+z}{1-s}\right)^K (2+z_s)^{-b_s}, \tag{D. 114}$$

where b_s was defined by (D. 92). Therefore

$$b_{m+1}(z) \geqslant \frac{(1+z)^{m+K}}{(1-s)^K} \Gamma(m+1) \Bigg[r(1+z)(2+z_s)^{-b_s}$$

$$- \frac{\Gamma(m+K)}{\Gamma(K)\,\Gamma(m+1)} [K(1-s)+m]\, e^{\frac{rsz_s}{1-s}} c_s \Bigg]. \tag{D. 115}$$

$$\frac{\Gamma(m+K)}{\Gamma(K)\,\Gamma(m+1)} = \binom{m+K-1}{K-1} < m^K. \tag{D.116}$$

Also, if we constrain $r \geq K(1-s)/z_s$, then

$$b_{m+1}(z) \geq \frac{(1+z)^{m+K}}{(1-s)^K}\,\Gamma(m+1)\left[\frac{K(1-s)}{z_s(2+z_s)^{b_s}} - m^K[K(1-s)+m]\,e^{Ksz_s}\,c_s\right]. \tag{D.117}$$

As m gets large, the exponential dependence of c_s on m will eventually dominate the last term, so for m large enough, $b_{m+1}(z)$ will be bounded away from zero. In particular, if M_s is the smallest positive integer for which (D. 91) holds, then $b_{M_s+1}(z)$ is strictly bounded away from zero. Thus all of the first $M_s + 1$ derivatives of $F_s(z)$ cannot be arbitrarily small when $r \geq K(1-s)/z_s$, and the rest of the theorem follows.

APPENDIX E

Lower Bound to Error Probability

We shall consider N uses of the amplitude-continuous, time-discrete channel derived in Section II. We shall compute a value of P_e that cannot be reduced by any coder-decoder combination using a block length N, when the channel is used with an average power constraint. This will be a lower bound to the minimum attainable P_e and, for historical reasons, it is called the "sphere-packing" lower bound. For rates $R \geq R_{crit}$, and $N \to \infty$, this bound will agree exponentially in N with the random-coding upper bound previously derived. It thus represents the true exponential behavior of the signal-channel model for this range of rates.

First, we present the body of the proof, making use of theorems whose proofs will be presented at the end of this appendix. The method will be similar to that used by Shannon, Gallager, and Berlekamp[39] for the discrete memoryless channel, although major modifications are necessary to account for the continuous amplitude of this channel model.

Let a code consist of M equiprobable code words $(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_M)$. As a preliminary, define

$$\mu(x,s) \equiv \ln\left[\int_0^{\infty} f(y)^s\, p(y|x)^{1-s}\, dy\right], \qquad 0 \leq s \leq 1 \tag{E.1}$$

$$z_m \equiv \int_{Y_m} \prod_{k=1}^{N} f(y_k)\, d\underline{y}, \tag{E.2}$$

where Y_m is the set of output vectors, $\underline{y}$, that are decoded as the m^{th} code word, and f(y) is any probability function defined on the output letter space. Note that

$$\sum_{m=1}^{M} z_m = \sum_{m=1}^{M} \int_{Y_m} \prod_{k=1}^{N} f(y_k)\, d\underline{y} = \int_{\underline{Y}} \prod_{k=1}^{N} f(y_k)\, d\underline{y} = 1. \tag{E.3}$$

Shannon, Gallager, and Berlekamp have shown that, for any code word in any code, and any s, $0 \leq s < 1$, either

$$z_m > \exp N\left[\frac{1}{N}\sum_{n=1}^{N}\{\mu(x_{mn},s)+(1-s)\mu'(x_{mn},s)\} - \frac{1-s}{N}\sqrt{2\sum_{n=1}^{N}\mu''(x_{mn},s)} - \frac{1}{N}\ln 4\right]$$

or (E.4)

$$P_{em} > \exp N\left[\frac{1}{N}\sum_{n=1}^{N}\{\mu(x_{mn},s)-s\mu'(x_{mn},s)\} - \frac{s}{N}\sqrt{2\sum_{n=1}^{N}\mu''(x_{mn},s)} - \frac{1}{N}\ln 4\right],$$

where P_{em} is the probability of error when the m^{th} code word is sent, and primes denote partial derivatives with respect to s, with f(y) held constant, that is,

$$\mu'(x,s) = \frac{\int_0^\infty f(y)^s\, p(y|x)^{1-s} \ln\left[\frac{f(y)}{p(y|x)}\right] dy}{\int_0^\infty f(y)^s\, p(y|x)^{1-s}\, dy} \tag{E.5}$$

$$\mu''(x,s) = \frac{\int_0^\infty f(y)^s\, p(y|x)^{1-s} \ln^2\left[\frac{f(y)}{p(y|x)}\right] dy}{\int_0^\infty f(y)^s\, p(y|x)^{1-s}\, dy} - [\mu'(x,s)]^2. \tag{E.6}$$

At this point, we shall choose a particular f(y) in order to simplify the bound.

$$f(y) = \frac{\beta(y)^{1/(1-s)}}{\int_0^\infty \beta(y)^{1/(1-s)}\, dy} \tag{E.7}$$

$$\beta(y) = \int_0^\infty p(x)\, e^{rx^2}\, p(y|x)^{1-s}\, dx, \tag{E.8}$$

where p(x) and r are chosen to satisfy condition (A.33), and therefore optimize the random-coding upper bound. In Appendix D, we have shown that some finite set of impulses must satisfy (A.33), so we know that a solution exists. From Theorem E.1, $s^2\mu''(x,s) \leq B(s) < \infty$, $0 < s < 1$, so that

$$s^2 \sum_{n=1}^{N} \mu''(x_{mn}, s) \leq NB(s). \tag{E.9}$$

Theorem E.2 states

$$\mu(x_{mn}, s) \geq -(1-s)E_{oe}\left[\frac{s}{1-s}, a, K\right] + r\left(a - x_{mn}^2\right), \tag{E.10}$$

Where E_{oe} is the equal eigenvalue random-coding exponent. By using (E.9) and (E.10) in (E.4), either

$$z_m > \exp N\left[-(1-s)\, E_{oe}\left[\frac{s}{1-s}, a, K\right] + ra - \frac{r}{N}\sum_{n=1}^{N} x_{mn}^2 + \frac{1-s}{N}\sum_{n=1}^{N} \mu'(x_{mn}, s) - \left(\frac{1-s}{s}\right)\sqrt{\frac{2B(s)}{N}} - \frac{1}{N}\ln 4\right]$$

or (E.11)

$$P_{em} > \exp N\left[-(1-s)E_{oe}\left[\frac{s}{1-s}, \alpha, K\right] + r\alpha - \frac{r}{N}\sum_{n=1}^{N} x_{mn}^2\right.$$

$$\left. - \frac{s}{N}\sum_{n=1}^{N} \mu'(x_{mn}, s) - \sqrt{\frac{2B(s)}{N}} - \frac{1}{N}\ln 4\right].$$

We now restrict analysis to a subset of the code words in any code; in particular, to those code words with $S_m \equiv \frac{1}{N}\sum_{n=1}^{N} x_{mn}^2 \leq \frac{\alpha}{1-\lambda}$, $0 < \lambda < 1$. Since we require $\frac{1}{M}\sum_{m=1}^{M} S_m \leq \alpha$, at most, $(1-\lambda)M$ code words can have $S_m > \frac{\alpha}{1-\lambda}$, and hence at least λM code words must lie in the subset under consideration. Take any of these code words, $\underline{x}_m$, and consider quantizing on the values of x_{mn}^2. For this code word,

$$\frac{1}{N}\sum_{n=1}^{N} x_{mn}^2 \leq \frac{\alpha}{1-\lambda}, \tag{E.12}$$

so certainly $x_{mn}^2 \leq \frac{\alpha N}{1-\lambda}$. If $j\epsilon \leq x_{mn}^2 < (j+1)\epsilon$, where ϵ is a small positive number, replace x_{mn}^2 by $x_j^2 = j\epsilon$. There are J possible values of x_j^2, $j = 0, 1, \ldots, J$, where $J = \frac{\alpha N}{\epsilon(1-\lambda)}$. From Theorem E.3, $\left|\frac{\partial}{\partial x^2}\mu'(x, s)\right| \leq D(s) = B(s)^{1/2} < \infty$, for all $x \geq 0$, $0 < s < 1$, so if $x_j^2 \leq x^2 \leq x_j^2 + \epsilon$, then

$$\mu'(x_j, s) - D(s)\,\epsilon \leq \mu'(x, s) \leq \mu'(x_j, s) + D(s)\,\epsilon. \tag{E.13}$$

Let n_{jm} be the number of x_{mn}^2 that are quantized to x_j^2 (that is, the number of x_{mn}^2 such that $j\epsilon \leq x_{mn}^2 < (j+1)\epsilon$). Then

$$\sum_{j=0}^{J} \frac{n_{jm}}{N}\mu'(x_j, s) - D(s)\,\epsilon \leq \frac{1}{N}\sum_{n=1}^{N} \mu'(x_{mn}, s) \leq \sum_{j=0}^{J} \frac{n_{jm}}{N}\mu'(x_j, s) + D(s)\,\epsilon. \tag{E.14}$$

By using (E.12) and (E.14), (E.11) becomes either

$$z_m > \exp N\left[-(1-s)E_{oe}\left[\frac{s}{1-s}, \alpha, K\right] - \frac{r\lambda\alpha}{1-\lambda} + (1-s)\left\{\sum_{j=0}^{J} \frac{n_{jm}}{N}\mu'(x_j, s) - D(s)\,\epsilon\right\}\right.$$

$$\left. -\left(\frac{1-s}{s}\right)\sqrt{\frac{2B(s)}{N}} - \frac{1}{N}\ln 4\right]$$

or (E.15)

$$P_{em} > \exp N\left[-(1-s)E_{oe}\left[\frac{s}{1-s}, \alpha, K\right] - \frac{r\lambda\alpha}{1-\lambda} - s\left\{\sum_{j=0}^{J} \frac{n_{jm}}{N}\mu'(x_j, s) + D(s)\,\epsilon\right\} - \sqrt{\frac{2B(s)}{N}} - \frac{1}{N}\ln 4\right].$$

Define the composition of any code word in the subset under consideration as the vector $(n_{om}, n_{1m}, \ldots, n_{Jm})$ that is associated with that code word. If c_o is defined as the number of possible compositions, then c_o is the number of ways in which we can choose vectors $(n_o, n_1, \ldots, n_J)$ such that $n_j \geq 0$, $\sum_{j=0}^{J} n_j = N$, and $\sum_{j=0}^{J} jn_j \leq \frac{\alpha N}{\epsilon(1-\lambda)}$. From Theorem E.4, with $\beta = \frac{\alpha}{\epsilon(1-\lambda)}$

$$c_o < 2(\beta N^2)^{\sqrt{2\beta N}+1} \tag{E.16}$$

when N and βN are greater than 100.

At least one of these c_o possible compositions must contain at least $\frac{\lambda M}{c_o}$ code words, or else the total number of words in the subset would be less than λM. We now further restrict analysis to code words in that particular composition. If P_{ec} is the total error probability for all words in this composition, certainly

$$P_e = \frac{1}{M}\sum_{m=1}^{M} P_{em} \geq \frac{\lambda}{c_o} P_{ec}. \tag{E.17}$$

Consider the $\frac{1}{2}\left(\frac{\lambda M}{c_o}\right)$ code words within this composition, that have the smallest values of P_{em}. For these words, $P_{em} \leq 2P_{ec}$, so $P_{e\,max} \leq 2P_{ec}$, where $P_{e\,max}$ is the maximum value of P_{em} in this group. Therefore, if we can derive a bound P_{eb} so that there must be some code word in any composition with at least $\frac{\lambda M}{2c_o}$ code words with $P_{em} > P_{eb}$, then

$$P_e \geq \frac{\lambda}{c_o} P_{ec} \geq \frac{\lambda}{2c_o} P_{e\,max} \geq \frac{\lambda}{2c_o} P_{eb}. \tag{E.18}$$

Choose $\epsilon = N^{-1/2}$, $\lambda = N^{-1}$, and consider a composition with at least $\frac{\lambda M}{2c_o}$ code words. For that particular composition, at least one code word must have $z_m \leq \frac{2c_o}{\lambda M}$, otherwise (E.3) will be violated. For this code word, either

$$R_N < (1-s)\,E_{oe}\left[\frac{s}{1-s}, \alpha, K\right] + \frac{r\alpha}{N-1} - (1-s)\sum_{j=0}^{J}\frac{n_{jm}}{N}\mu'(x_j, s) + (1-s)\left(1+\frac{\sqrt{2}}{s}\right)\sqrt{\frac{B(s)}{N}} + \frac{1}{N}\ln(8Nc_o) \tag{E.19}$$

or

$$P_{em} > \exp - N\left[(1-s)\, E_{oe}\left[\frac{s}{1-s}, a, K\right] + \frac{ra}{N-1} + s \sum_{j=0}^{J} \frac{n_{jm}}{N} \mu'(x_j, s) + (s+\sqrt{2})\sqrt{\frac{B(s)}{N} + \frac{1}{N}\ln 4}\right], \tag{E.20}$$

where, as before, $R_N = \frac{1}{N} \ln M$. Define I(s) to be equal to the right side of inequality (E.19). Note that B(1) is strictly bounded,

$$\lim_{s \to 1} (1-s)\, E_{oe}\left[\frac{s}{1-s}, a, K\right] = 0,$$

and consider

$$\sum_{j=0}^{J} \frac{n_{jm}}{N} \mu'(x_j, 1).$$

Calculation of $\mu'(x, 1)$ shows that

$$\mu'(x, 1) = K\left[-1 + \sum_{n=1}^{N} \frac{p_n}{1+x_n^2} + \ln(1+x^2) + \frac{1}{(1+x^2) \sum_{n=1}^{N} \frac{p_n}{1+x_n^2}}\right] \tag{E.21}$$

which is bounded by

$$|\mu'(x, 1)| \leq K\left(3+x^2+z_1\right), \tag{E.22}$$

where z_1 is given by (D.63) with s = 1. Thus

$$\left|\sum_{j=0}^{J} \frac{n_{jm}}{N} \mu'(x_j, 1)\right| \leq K\left(3 + \frac{a}{1 - 1/N} + z_1\right) < \infty, \tag{E.23}$$

and hence $I(1) = \frac{ra}{N-1} + \frac{1}{N} \ln (8Nc_o)$. Restrict s to the range $0 < \Delta \leq s < 1$, where

$$\Delta \equiv \left[\frac{4K \ln A_o}{\ln\left\{\frac{N}{(\ln N)^{1/4}}\right\}}\right]^{1/2}, \tag{E.24}$$

and A_o is a constant given later by (E.67). Clearly, for large N, Δ will be arbitrarily small.

$$I(\Delta) > \frac{\sqrt{2}(1-\Delta)}{\Delta} \sqrt{\frac{B(\Delta)}{N}} - (1-\Delta) \sum_{j=0}^{J} \frac{n_{jm}}{N} \mu'(x_j, \Delta). \tag{E.25}$$

It is known[39] that $\mu(x, s)$ is a convex function of s, so that $\mu'(x, s)$ is an increasing function of s, and thus if N is large enough so that $\Delta < 1 - \frac{1}{\sqrt{2}}$, then

$$I(\Delta) > \left[\frac{\ln\left\{\frac{N}{(\ln N)^{1/4}}\right\}}{(\ln N)^{1/2}\,(4K \ln A_o)}\right]^{1/2} - K(a+1+z_1) \equiv d_N. \tag{E.26}$$

Since $I(s)$ is continuous, if $I(1) < R_N < d_N$, then there exists an $\hat{s}$, $\Delta < \hat{s} < 1$ such that $R_N = I(\hat{s})$. Since inequality (E.19) is not satisfied for $s = \hat{s}$, (E.20) must be. By using the equality in (E.19) to eliminate $\sum_{j=0}^{J} \frac{n_{jm}}{N} \mu'(x_j, \hat{s})$ from (E.20), we have

$$P_{em} > \exp -N\left[E_{oe}\left[\frac{\hat{s}}{1-\hat{s}}, a, K\right] - \frac{\hat{s}}{1-\hat{s}}\left\{R_N - \frac{ra}{N-1} - \frac{1}{N}\ln(8Nc_o)\right\} + 2(1+\sqrt{2})\sqrt{\frac{B(\hat{s})}{N}} + \frac{1}{N}\ln 4\right] \tag{E.27}$$

for some $\hat{s}$, $\Delta < \hat{s} < 1$, and some code word in the composition, provided $I(1) < R_N < d_N$. This gives us the bound that is necessary to apply (E.18). By using (E.16) to specify c_o, and noting that $B(\hat{s}) \leq B(\Delta)$, and $0 \leq r \leq K$, (E.27) becomes

$$P_e > \exp -N\left[E_{oe}\left[\frac{\hat{s}}{1-\hat{s}}, a, K\right] - \frac{\hat{s}}{1-\hat{s}}(R_N - \Delta_N) + \delta_N\right] \tag{E.28}$$

$$\Delta_N = \frac{Ka}{N-1} + \frac{1}{N}\left[\ln(16N) + \left\{\left[\frac{2aN^{3/2}}{1-1/N}\right]^{1/2} + 1\right\} \ln\left(\frac{aN^{5/2}}{1-1/N}\right)\right] \tag{E.29}$$

$$\delta_N = \frac{2(1+\sqrt{2})}{(\ln N)^{1/4}} + \Delta_N. \tag{E.30}$$

If (E.28) is true for some $\hat{s}$, $\Delta < \hat{s} < 1$, it is true for the s that minimizes the right-hand side. Therefore, if $\Delta_N < R_N < d_N$, then

$$P_e > \exp -N\left[\max_{0<s<1}\left\{E_{oe}\left[\frac{s}{1-s}, a, K\right] - \frac{s}{1-s}(R_N - \Delta_N)\right\} + \delta_N\right], \tag{E.31}$$

where Δ_N and δ_N go to zero and d_N goes to infinity as N goes to infinity.

If we replace $\frac{s}{1-s}$ by ρ, $0 < \rho < \infty$, the right-hand side of (E.31) is the same as the random-coding upper bound to P_e, except for additive factors that go to zero as $N \to \infty$, and the difference in the range of ρ. When $R_N > R_{crit}$, the maximum is in the range $0 < \rho < 1$, and so as $N \to \infty$, the upper and lower bounds exponentially agree.

It is worth noting that this lower bound depends on the nature of the channel model in a rather limited fashion. To be specific, this bound can be applied to any time-discrete, amplitude-continuous, average energy-constrained channel $p(y|x)$, for which the following properties hold:

1. There exists $0 < r < A$, $p(x)$ satisfying constraints (A.3) such that

$$\int_{-\infty}^{\infty} \beta(y)^{s/(1-s)}\, p(y|x)^{1-s}\, e^{rx^2} dx \geq \int_{-\infty}^{\infty} \beta(y)^{1/(1-s)}\, dy,$$

where

$$\beta(y) = \int_{-\infty}^{\infty} p(x)\, e^{rx^2}\, p(y|x)^{1-s}\, dx.$$

2. $\frac{s^2}{N} \sum_{n=1}^{N} \mu''(x_{mn}, s) \leq B(s) < \infty, \qquad 0 < s \leq 1$

3. $\left| \frac{\partial}{\partial x} \mu'(x, s) \right| \leq D(s) < \infty, \qquad 0 < s \leq 1,$

where $B(s)$ and $D(s)$ are continuous and bounded for $0 < s \leq 1$.

The previous bound is not valid at the end point, $s = 1$, and since many of the complexities that arise for $0 < s < 1$ can be avoided when $s = 1$, it is worth while to consider this point as a separate case. For this point,

$$f(y) \equiv \lim_{s \to 1} \frac{\left[\int_0^{\infty} p(x)\, p(y|x)^{1-s}\, dx\right]^{1/(1-s)}}{\int_0^{\infty} \left[\int_0^{\infty} p(x)\, p(y|x)^{1-s}\, dx\right]^{1/(1-s)} dy} = \frac{y^{K-1} c_1^K e^{-c_1 y}}{\Gamma(K)}, \tag{E.32}$$

where

$$c_1 = \int_0^{\infty} \frac{p(x)}{1+x^2}\, dx. \tag{E.33}$$

Because of the simple form of $f(y)$, it is possible to evaluate $\mu'(x, 1)$ and $\mu''(x, 1)$ directly. The results are

$$\mu'(x, 1) = K\left[\ln c_1 - 1 + \ln (1+x^2) + \frac{1}{c_1(1+x^2)}\right] \tag{E.34}$$

$$\mu''(x,1) = K\left[-1 + \frac{1}{c_1(1+x^2)}\right]^2 < K\left(1+1/c_1^2\right). \tag{E.35}$$

By applying (E.4), with s = 1, for any code word in any code, either

$$z_m > \frac{1}{4}$$

or

$$P_{em} > \exp N\left[-\frac{1}{N}\sum_{n=1}^{N} \mu'(x_{mn}, 1) - \sqrt{\frac{2}{N^2}\sum_{n=1}^{N} \mu''(x_{mn}, 1)} - \frac{1}{N}\ln 4\right]. \tag{E.36}$$

By using (E.34) and (E.35), (E.36) may be replaced by

$$P_{em} > \exp -NK\left[\ln c_1 - 1 + \frac{1}{N}\sum_{n=1}^{N} \ln\left(1+x_{mn}^2\right) + \frac{1}{Nc_1}\sum_{n=1}^{N} \frac{1}{1+x_{mn}^2} + \Delta_{NK}\right], \tag{E.37}$$

where

$$\Delta_{NK} = \frac{1}{NK}\ln 4 + \sqrt{\frac{2}{NK}\left(1+1/c_1^2\right)}. \tag{E.38}$$

Choose p(x) as the one discussed in Theorem C.1. In this case,

$$\ln(1+x^2) + \frac{1}{c_1}\frac{1}{1+x^2} \leqslant \frac{1}{c_1} + \frac{x^2}{\alpha}\left[1 - \frac{1}{c_1} + \int_0^\infty p(x)\ln(1+x^2)\,dx\right]. \tag{E.39}$$

$$\frac{1}{N}\sum_{n=1}^{N}\left[\ln\left(1+x_{mn}^2\right) + \frac{1}{c_1}\frac{1}{1+x_{mn}^2}\right] \leqslant \frac{1}{c_1} + \left(\frac{H}{\alpha N}\sum_{n=1}^{N} x_{mn}^2\right). \tag{E.40}$$

$$H \equiv 1 - \frac{1}{c_1} + \int_0^\infty p(x)\ln\left(1+x^2\right)dx \geqslant 0. \tag{E.41}$$

Therefore, either

$$z_m > \frac{1}{4}$$

or

$$P_{em} > \exp -NK\left[\ln c_1 - 1 + \frac{1}{c_1} + \frac{H}{\alpha N}\sum_{n=1}^{N} x_{mn}^2 + \Delta_{NK}\right]. \tag{E.42}$$

Restrict the analysis to code words for which $\sum_{n=1}^{N} x_{mn}^2 \leq \frac{aN}{1-\lambda}$, $0 < \lambda < 1$. There must be at least λM such code words. At least one of these λM words must have $z_m \leq \frac{1}{\lambda M}$. For that word, either

$$R_N < \frac{1}{N} \ln\left(\frac{4}{\lambda}\right)$$

or

$$P_{em} > \exp -NK\left[\ln c_1 + \frac{1}{1-\lambda}\int_0^\infty p(x)\ln(1+x^2)\,dx + \frac{\lambda}{1-\lambda}\left(1 - \frac{1}{c_1}\right) + \Delta_{NK}\right]. \tag{E.43}$$

Observe that $\frac{1}{c_1} \geq 1$, so $\frac{\lambda}{1-\lambda}\left(1 - \frac{1}{c_1}\right) \leq 0$, and also $\int_0^\infty p(x)\ln(1+x^2)\,dx \leq \int_0^\infty x^2 p(x)\,dx = a$. Thus, if

$$R_N \geq \frac{1}{N} \ln\left(\frac{4}{\lambda}\right),$$

then

$$P_{em} > \exp -NK\left[\ln c_1 + \int_0^\infty p(x)\ln(1+x^2)\,dx + \frac{\lambda a}{1-\lambda} + \Delta_{NK}\right] \tag{E.44}$$

for at least one code word in any code. Let $\lambda = 1/N$. Then if we go through the same argument as before, we find that if $R_N \geq \frac{1}{N}\ln(8N)$, then

$$P_e \geq \exp -N\left[E_{oe}(\infty, a, K) + \left\{\frac{K}{N}\ln 2 + \frac{aK}{N-1} + \frac{1}{N}\ln 4\right\} + \sqrt{\frac{2K}{N}\left(1+1/c_1^2\right)}\right] \tag{E.45}$$

which is of the same form as (E.31).

THEOREM E.1

$s^2\mu''(x,s) \leq B(s) = (A_o)^{8Ka_s}$, where A_o is a bounded constant, and a_s is defined by (D.64).

Proof: Recall that

$$\mu''(x,s) = \frac{\int_0^\infty f(y)^s\, p(y|x)^{1-s} \ln^2\left[\frac{f(y)}{p(y|x)}\right] dy}{\int_0^\infty f(y)^s\, p(y|x)^{1-s}\,dy} - [\mu'(x,s)]^2, \tag{E.46}$$

where $\mu'(x,s)$ is given by (E.5), f(y) by (E.7), and now

$$\beta(y) = \sum_{n=1}^{N} p_n e^{rz_n} p(y|\sqrt{z_n})^{1-s}. \tag{E.47}$$

After some manipulation, we find

$$\frac{f(y)}{p(y|x)} = \frac{(1+x^2)^K e^{y/(1+x^2)} \left\{ \sum_{n=1}^{N} p_n e^{rz_n} \left[\frac{e^{-y/(1+z_n)}}{(1+z_n)^K} \right]^{1-s} \right\}^{1/(1-s)}}{\int_0^\infty \beta(y)^{1/(1-s)} dy}. \tag{E.48}$$

If we define

$$q_x(y) \equiv \frac{y}{1+x^2} + \frac{1}{1-s} \ln \left[\sum_{n=1}^{N} p_n e^{rz_n} \left[\frac{e^{-y/(1+z_n)}}{(1+z_n)^K} \right]^{1-s} \right] \tag{E.49}$$

and

$$\gamma(y|x) = \frac{p(y|x) e^{sq_x(y)}}{\int_0^\infty p(y|x) e^{sq_x(y)} dy}, \tag{E.50}$$

we find

$$\mu''(x,s) = \overline{\left[\ln \left\{ \frac{f(y)}{p(y|x)} \right\} - \overline{\ln \left\{ \frac{f(y)}{p(y|x)} \right\}} \right]^2}, \tag{E.51}$$

where the bars denote an average over y with respect to $\gamma(y|x)$.

In computing $\mu''(x,s)$ we can drop any multiplicative constant, independent of y, in $\frac{f(y)}{p(y|x)}$ because the logarithm will convert it to an additive constant, which will cancel when we compute the variance (E.51). Since

$$\frac{f(y)}{p(y|x)} = (1+x^2)^K \left[\int_0^\infty \beta(y)^{1/(1-s)} dy \right]^{-1} e^{q_x(y)}, \tag{E.52}$$

and the first two terms are independent of y,

$$\mu''(x,s) = \int_0^\infty \gamma(y|x) \left[q_x(y) - \overline{q_x(y)} \right]^2 dy. \tag{E.53}$$

We now make use of the fact that $\overline{(\xi-\overline{\xi})^2} \leq \overline{\xi^2}$, to obtain

$$\mu''(x, s) \leq \int_0^\infty \gamma(y|x)\, q_x^2(y)\, dy = \overline{q_x^2(y)}. \tag{E.54}$$

In Theorem D.3, it was shown that $x_n \leq \sqrt{z_s}$, with z_s as defined in that theorem. Thus

$$e^{-y(1-s)} \leq e^{\frac{-y(1-s)}{1+z_n}} \leq e^{\frac{-y(1-s)}{1+z_s}}. \tag{E.55}$$

Applying these inequalities to (E.49), we find

$$q_x(y) = a + b_x(y) \tag{E.56}$$

$$-y \leq -\frac{yx^2}{1+x^2} \leq b_x(y) \leq y\left[\frac{z_s - x^2}{(1+z_s)(1+x^2)}\right] \leq y \tag{E.57}$$

$$a = \frac{1}{1-s} \ln\left[\sum_{n=1}^{N} p_n\, e^{rz_n}\, (1+z_n)^{-K(1-s)}\right]. \tag{E.58}$$

Then

$$\overline{q_x^2(y)} = a^2 + 2a\,\overline{b_x(y)} + \overline{b_x^2(y)} \leq a^2 + 2|a|\,\overline{y} + \overline{y^2}. \tag{E.59}$$

But

$$\overline{y^m} = \frac{\int_0^\infty p(y|x)\, e^{as+sb_x(y)}\, y^m\, dy}{\int_0^\infty p(y|x)\, e^{as+sb_x(y)}\, dy} \leq \frac{\int_0^\infty p(y|x)\, y^m\, e^{sy\left[\frac{z_s - x^2}{(1+z_s)(1+x^2)}\right]} dy}{\int_0^\infty p(y|x)\, e^{-\frac{syx^2}{1+x^2}}\, dy}. \tag{E.60}$$

Evaluation of the integrals leads to

$$\overline{y^m} \leq \frac{\Gamma(m+K)}{\Gamma(K)} \left(\frac{1+sx^2}{1+x^2}\right)^K \left[\frac{(1+x^2)(1+z_s)}{1+z_s - sz_s + sx^2}\right]^{m+K} \leq \frac{\Gamma(m+K)}{\Gamma(K)} (1+z_s)^{m+K} \left(\frac{1+x^2}{1+sx^2}\right)^m \tag{E.61}$$

$$\overline{y^m} \leqslant \frac{\Gamma(m+K)}{\Gamma(K)} \frac{(1+z_s)^{m+K}}{s^m}. \quad \text{(E. 62)}$$

Also note that

$$-K \ln (1+z_s) \leqslant \frac{1}{1-s} \ln \left[\frac{1}{(1+z_s)^{K(1-s)}}\right] \leqslant a \leqslant \frac{1}{1-s} \ln \left[e^{rz_s}\right] \leqslant Kz_s, \quad \text{(E. 63)}$$

and so $|a| \leqslant Kz_s$. Therefore

$$\overline{q_x^2(y)} \leqslant (Kz_s)^2 + 2Kz_s \left\{\frac{K(1+z_s)^{K+1}}{s}\right\} + \frac{K(K+1)(1+z_s)^{K+2}}{s^2} \quad \text{(E. 64)}$$

$$s^2 \mu''(x, s) \leqslant K^2 (1+z_s)^2 + 2K^2 (1+z_s)^{K+2} + (K^2+K)(1+z_s)^{K+2} \leqslant 5K^2 (1+z_s)^{K+2}. \quad \text{(E. 65)}$$

But $z_s = A^{a_s}$, where $A > 1$ can be defined from (D. 63) and $a_s > 1$ is defined by (D. 64). Then

$$s^2 \mu''(x, s) \leqslant 2K^2 (2z_s)^{K+2} < \left(10K^2 A^{a_s}\right)^{K+2} < (14K^2 A)^{8Ka_s}, \quad \text{(E. 66)}$$

and if we let

$$A_o = 14K^2 A, \quad \text{(E. 67)}$$

$$s^2 \mu''(x, s) < (A_o)^{8Ka_s} \equiv B(s). \quad \text{(E. 68)}$$

Note that B(s) will be continuous and bounded for $0 < s \leqslant 1$, since a_s is.

THEOREM E. 2

$$\mu(x, s) \geqslant -(1-s) E_{oe}\left(\frac{s}{1-s}, \alpha, K\right) + r(\alpha - x^2). \quad \text{(E. 69)}$$

<u>Proof:</u>

$$\mu(x, s) = \ln \left[\int_0^\infty \left\{\frac{\beta(y)^{1/(1-s)}}{\int_0^\infty \beta(y)^{1/(1-s)} dy}\right\}^s p(y|x)^{1-s} dy\right] \quad \text{(E. 70)}$$

$$\mu(x,s) = -s\ \ln\left[\int_0^\infty \{\beta(y)\ e^{-r\alpha}\}^{1/(1-s)}\ dy\right] - \frac{r\alpha s}{1-s}$$

$$+ \ln\left[\int_0^\infty \beta(y)^{s/(1-s)}\ p(y|x)^{1-s}\ e^{rx^2}\ dy\ e^{-rx^2}\right], \tag{E.71}$$

and, since (A.33) is satisfied,

$$\mu(x,s) \geq -s\ \ln\left[\int_0^\infty \{\beta(y)\ e^{-r\alpha}\}^{1/(1-s)}\ dy\right] - \frac{r\alpha s}{1-s} - rx^2 + \ln\left[\int_0^\infty \beta(y)^{1/(1-s)}\ dy\right] \tag{E.72}$$

$$\mu(x,s) \geq (1-s)\ \ln\left[\int_0^\infty \{\beta(y)\ e^{-r\alpha}\}^{1/(1-s)} dy\right] + r(\alpha - x^2) \tag{E.73}$$

$$\mu(x,s) \geq -(1-s)\ E_{oe}\left[\frac{s}{1-s}, \alpha, K\right] + r(\alpha - x^2). \tag{E.74}$$

THEOREM E.3

$\left|\frac{\partial}{\partial x^2}\mu'(x,s)\right| \leq B(s)^{1/2}$ for any $x \geq 0$, $0 < s < 1$, where $B(s)$ is defined in Theorem E.1.

Proof: For convenience, define $z \equiv x^2$.

$$\mu'(\sqrt{z},s) = \frac{\int_0^\infty f(y)^s\ p(y|\sqrt{z})^{1-s}\ \ln\left[\frac{f(y)}{p(y|\sqrt{z})}\right] dy}{\int_0^\infty f(y)^s\ p(y|\sqrt{z})^{1-s}\ dy}. \tag{E.75}$$

After plugging in the functions involved, and considerable simplification, we find that

$$\mu'(\sqrt{z},s) = -\ln\left[\int_0^\infty \beta(y)^{1/(1-s)}\ dy\right] + K\ \ln\ (1+z) + \frac{\int_0^\infty p_o(u)\ e^{sh(z,u)}\ h(z,u)\ du}{\int_0^\infty p_o(u)\ e^{sh(z,u)}\ du}, \tag{E.76}$$

where we have defined

$$h(z,u) = u + \frac{1}{1-s}\ \ln\left[\sum_{n=1}^{N} p_n\ e^{rz_n}\left[\frac{e^{-u\left(\frac{1+z}{1+z_n}\right)}}{(1+z_n)^K}\right]^{1-s}\right] \tag{E.77}$$

and

$$p_o(u) = \frac{u^{K-1}\, e^{-u}}{\Gamma(K)}. \tag{E.78}$$

Differentiation of (E.76) with respect to z, and further manipulation, yields the expression

$$\frac{\partial}{\partial z}\mu'(\sqrt{z}, s) = \frac{K}{1+z} + \overline{h'(z,u)} + s[\overline{h(z,u)h'(z,u)} - \overline{h(z,u)}\ \overline{h'(z,u)}], \tag{E.79}$$

where

$$h'(z,u) = \frac{\partial}{\partial z} h(z,u) = \frac{-u \sum_{n=1}^{N} \frac{p_n\, e^{rz_n}}{1+z_n} \left[\frac{e^{-u\left(\frac{1+z}{1+z_n}\right)}}{(1+z_n)^K}\right]^{1-s}}{\sum_{n=1}^{N} p_n\, e^{rz_n} \left[\frac{e^{-u\left(\frac{1+z}{1+z_n}\right)}}{(1+z_n)^K}\right]^{1-s}}, \tag{E.80}$$

and a bar represents an average over u with respect to the density

$$\frac{p_o(u)\, e^{sh(z,u)}}{\int_0^\infty p_o(u)\, e^{sh(z,u)}\, du}.$$

$$\left|\frac{\partial}{\partial z}\mu'(\sqrt{z}, s)\right| \leq \frac{K}{1+z} + \left|\overline{h'(z,u)}\right| + s\{\left|\overline{h(z,u)h'(z,u)}\right| + \left|\overline{h(z,u)}\right|\ \left|\overline{h'(z,u)}\right|\}. \tag{E.81}$$

Note that $0 \leq \frac{1}{1+x^2} \leq 1$, so

$$-u \leq h'(z,u) \leq 0 \tag{E.82}$$

$$\left|\overline{h'(z,u)}\right| \leq \overline{u}. \tag{E.83}$$

The procedure to be followed is similar to that used in the proof of Theorem E.1. Using (E.55) and (E.58) yields

$$h(z,u) = a + c(z,u) \tag{E.84}$$

$$-u(1+z) \leq -uz \leq c(z,u) \leq u\left[\frac{z_s - z}{1+z_s}\right] \leq u \leq u(1+z) \tag{E.85}$$

$$\overline{u^m} = \frac{\int_0^\infty p_o(u)\, e^{sc(z,u)}\, u^m\, du}{\int_0^\infty p_o(u)\, e^{sc(z,u)}\, du} \leqslant \frac{\int_0^\infty u^m p_o(u)\, e^{s\left[\frac{z_s - z}{1+z_s}\right]}\, du}{\int_0^\infty p_o(u)\, e^{-usz}\, du}. \tag{E. 86}$$

Evaluation of the integrals leads to

$$\overline{u^m} \leqslant \frac{\Gamma(m+K)}{\Gamma(K)} \left[\frac{1 + z_s}{1 + sz + (1-s)z_s}\right]^{m+K} (1+sz)^K \leqslant \frac{\Gamma(m+K)}{\Gamma(K)} \frac{(1+z_s)^{m+K}}{(1+sz)^m}. \tag{E. 87}$$

Hence

$$\left|h'(z,u)\right| \leqslant K(1+z_s)^{K+1}, \tag{E. 88}$$

and also

$$\left|\overline{h(z,u)}\right| \leqslant |a| + \left|\overline{c(z,u)}\right| \leqslant |a| + (1+z)\bar{u} \tag{E. 89}$$

$$\left|\overline{h(z,u)}\right| \leqslant Kz_s + \frac{(1+z)\,K(1+z_s)^{K+1}}{1 + sz} < \frac{Kz_s + K(1+z_s)^{K+1}}{s} \tag{E. 90}$$

$$\left|\overline{h(z,u)}\right| \leqslant \frac{2K}{s}(1+z_s)^{K+1}. \tag{E. 91}$$

Note that

$$\left|\overline{h(z,u)h'(z,u)}\right| \leqslant \overline{|h(z,u)|\,|h'(z,u)|} \leqslant \overline{u|h(z,u)|} \leqslant |a|\bar{u} + (1+z)\overline{u^2} \tag{E. 92}$$

$$\left|\overline{h(z,u)h'(z,u)}\right| \leqslant Kz_s K(1+z_s)^{K+1} + \frac{(1+z)\,K(K+1)(1+z_s)^{K+2}}{(1+sz)^2} \leqslant \frac{3K^2(1+z_s)^{K+2}}{s}. \tag{E. 93}$$

Plugging (E. 88), (E. 91), and (E. 93) into (E. 81), we find

$$\left|\frac{\partial}{\partial z}\mu'(\sqrt{z}, s)\right| \leqslant 7K^2(1+z_s)^{2K+2} < (14K^2 A)^{4Ka_s} = B(s)^{1/2}, \tag{E. 94}$$

thereby proving the theorem.

THEOREM E.4

Let c_o be the number of possible vectors $(n_o, n_1, \ldots, n_J)$ such that $n_j \geqslant 0$, $\sum_{j=0}^{J} n_j = N$, $J = \beta N$, n_j an integer, and

$$\sum_{j=0}^{J} jn_j \leq \beta N. \tag{E. 95}$$

Then $c_o < 2(\beta N^2)^{\sqrt{2\beta N}+1}$ when N and $\beta N > 100$. The quantity c_o can be interpreted as the number of ways in which N indistinguishable balls may be distributed into J + 1 distinguishable boxes, subject to constraint (E. 95). When βN is large, we can change β slightly to make J an integer, without affecting the character of the result.

Proof: Constraint (E. 95) is difficult to deal with directly, but it may be loosened somewhat to compute an upper bound to c_o. We ask, How many different n_j can be positive when (E. 95) is satisfied? Clearly, the largest number of positive n_j is obtained when $n_1 = 1$, $n_2 = 1$, . . . , $n_I = 1$ until constraint (E. 95) is satisfied with equality (or as close to equality as possible), and $n_o = N - I$, provided $I < N$, of course. This means that

$$1 + 2 + 3 + \ldots + I = \frac{I(I+1)}{2} \leq \beta N \tag{E. 96}$$

$$I \leq -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 8\beta N} \leq \sqrt{2\beta N}. \tag{E. 97}$$

Therefore, we can obtain an upper bound to c_o by computing d_o, the number of possible vectors $(n_o, n_1, \ldots n_J)$ such that $n_j \geq 0$, $\sum_{j=0}^{J} n_j = N$, n_j = integer, and, at most, $\sqrt{2\beta N}$ different n_j are positive, where $J = \beta N$. This will be an upper bound because all vectors satisfying (E. 95) will also satisfy these constraints.

$$d_o = \sum_{m=1}^{\sqrt{2\beta N}} \begin{bmatrix}\text{number of ways N balls can} \\ \text{be put into m distinguishable} \\ \text{slots so that no slot is empty}\end{bmatrix} \times \begin{bmatrix}\text{number of ways the m occu-} \\ \text{pied slots can be chosen from} \\ \text{a total of J slots}\end{bmatrix} \tag{E. 98}$$

$$d_o = \sum_{m=1}^{\sqrt{2\beta N}} \binom{N-1}{m-1}\binom{\beta N}{m} \tag{E. 99}$$

$$\binom{N-1}{m-1} < \binom{N+m-1}{m-1} < N^m \tag{E. 100}$$

$$d_o < \sum_{m=1}^{\sqrt{2\beta N}} N^m \binom{\beta N}{m}. \tag{E. 101}$$

If βN is large, >100, for example, and $1 \leq m \leq \sqrt{2\beta N}$, then

$$\binom{\beta N}{m} \leq \binom{\beta N}{\sqrt{2\beta N}} \tag{E. 102}$$

$$d_o < \binom{\beta N}{\sqrt{2\beta N}} \sum_{m=1}^{\sqrt{2\beta N}} N^m = \binom{\beta N}{\sqrt{2\beta N}} \left[\frac{N}{N-1} \left(N^{\sqrt{2\beta N}} - 1 \right) \right] < 2N^{\sqrt{2\beta N}} \binom{\beta N}{\sqrt{2\beta N}},$$

(E. 103)

and if $N > 2$,

$$\binom{x}{\sqrt{2x}} = \binom{(x-\sqrt{2x}) + \sqrt{2x}}{\sqrt{2x}} < (x-\sqrt{2x})^{\sqrt{2x}+1} = x^{\sqrt{2x}+1} \left(1 - \sqrt{\frac{2}{x}}\right)^{\sqrt{2x}+1} < x^{\sqrt{2x}+1}$$

(E. 104)

$$d_o < 2N^{\sqrt{2\beta N}} (\beta N)^{\sqrt{2\beta N}+1} < 2(\beta N^2)^{\sqrt{2\beta N}+1}, \qquad \text{(E. 105)}$$

thereby proving the theorem.

Acknowledgment

I am greatly indebted to Professor Robert S. Kennedy for guidance and encouragement throughout the course of this study. Thanks are due to Professor Robert G. Gallager, who acted as a reader, for numerous helpful discussions and enlightened comments. Professor Harry L. Van Trees, who also acted as a reader, and numerous other faculty members and graduate students at M. I. T. were also most helpful.

I wish to thank the National Science Foundation and the Research Laboratory of Electronics, M.I.T., for support during my graduate studies, and the M. I. T. Computation Center for the use of its facilities.

References

1. R. Price, "Optimum Detection of Random Signals in Noise, with Application to Scatter-Multipath Communication. I," IRE Trans., Vol. IT-2, pp. 125-135, December 1956.
2. R. Price. "Notes on Ideal Receivers for Scatter Multipath," Group Report 34-39, Lincoln Laboratory, M. I. T., Lexington, Mass., May 12, 1955.
3. R. Price and P. E. Green, "Signal Processing in Radar Astronomy-Communication via Fluctuating Multipath Media," Technical Report 234, Lincoln Laboratory, M. I. T., Lexington, Mass., October 6, 1960.
4. T. Kailath, "Correlation Detection of Signals Perturbed by a Random Channel," IRE Trans., Vol. IT-6, pp. 361-366, June 1960.
5. D. Middleton, "On the Detection of Stochastic Signals in Additive Normal Noise – Part I," IRE Trans., Vol. IT-3, pp. 86-121, June 1957.
6. G. Turin, "Error Probabilities for Binary Symmetric Ideal Reception through Non-selective Slow Fading and Noise," Proc. IRE 46, 1603-1619 (1958).
7. P. Bello, "Some Results on the Problem of Discriminating between Two Gaussian Processes," IRE Trans., Vol. IT-7, pp. 224-233, October 1961.
8. C. E. Shannon, "A Mathematical Theory of Communication," Bell System Tech. J. 27, 379, 623 (1948).
9. R. M. Fano, Transmission of Information (The M. I. T. Press, Cambridge, Mass., 1961).
10. R. G. Gallager, "A Simple Derivation of the Coding Theorem and Some Applications," IEEE Trans., Vol. IT-11, pp. 3-17, January 1965.
11. J. Holsinger, "Digital Communications over Fixed Time-Continuous Channels with Memory – With Special Application to Telephone Channels," Ph. D. Thesis, Massachusetts Institute of Technology, October 1964.
12. P. M. Ebert, "Error Bounds for Parallel Communication Channels," Technical Report 448, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Mass., August 1, 1966.
13. J. N. Pierce, "Theoretical Performance Bounds for Variable Data Rate Transmissions on Fading Circuits," AFCRL Report No. 959.
14. J. N. Pierce, "M-ary Transmissions with Linear Filtering on Rayleigh Fading Circuits," AFCRL Report No. 62-150.
15. I. Jacobs, "The Asymptotic Behavior of Incoherent M-ary Communications Systems," Proc. IRE 51, 251-252 (January 1963).
16. R. S. Kennedy, Performance Limits for Fading Dispersive Channels (to be published by John Wiley and Sons, Inc., New York).
17. J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering (John Wiley and Sons, Inc., New York, 1965), p. 531.
18. M. Schwartz, W. R. Bennett, and S. Stein, Communication Systems and Techniques, Inter-University Electronics Series, Vol. 4 (McGraw-Hill Book Company, New York, 1966), p. 349.
19. C. E. Shannon, "Probability of Error for Optimal Codes in a Gaussian Channel," Bell System Tech. J. 38, 611-652 (1957).
20. W. B. Davenport and W. L. Root, An Introduction to the Theory of Random Signals and Noise (McGraw-Hill Book Company, New York, 1958), p. 96.
21. F. Riesz and B. Sz.-Nagy, Functional Anaylsis (Ungar, New York, 1955).
22. F. B. Hildebrand, Methods of Applied Mathematics (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1952), p. 77.

23. H. Cramér, _Mathematical Methods of Statistics_ (Princeton University Press, Princeton, New Jersey, 1946), p. 236.

24. H. L. Yudkin, "An Error Bound for Gaussian Signals in Gaussian Noise," Quarterly Progress Report No. 73, Research Laboratory of Electronics, M. I. T., Cambridge, Mass., April 15, 1964, pp. 149-155.

25. P. Wolfe, "The Simplex Method for Quadratic Programming," Econometrica, Vol. 27, No. 3, pp. 382-398, July 1959.

26. J. C. G. Boot, _Quadratic Programming_ (Rand-McNally, Chicago, Ill., 1964).

27. R. B. Wilson, _A Simplicial Algorithm for Concave Programming_, Doctoral Dissertation, Graduate School of Business Administration, Harvard University, Boston, Mass., 1963.

28. M. Abramowitz and I. A. Stegun (eds.), _Handbook of Mathematical Functions_, National Bureau of Standards, Washington, D. C., June 1964.

29. H. W. Kuhn and A. W. Tucker, Nonlinear Programming, in _Second Berkeley Symposium on Mathematical Statistics and Probability_, J. Neyman (ed.) (University of California Press, Berkeley, 1951), pp. 481-492.

30. C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. II," (to be published in _Information and Control_).

31. D. B. DeHaan, _Nouvelles Tables d'Intégrales Définies_ (Stechert and Company, New York, 1939), p. 177.

32. G. H. Hardy, J. E. Littlewood, and G. Pólya, _Inequalities_ (Cambridge University Press, London, 1959), p. 26.

33. E. C. Titchmarsh, _The Theory of Functions_ (Oxford University Press, London, 1932), p. 392.

34. G. H. Hardy, et al., _op. cit._, p. 72.

35. D. V. Widder, _The Laplace Transform_ (Princeton University Press, Princeton, New Jersey, 1946), p. 57.

36. G. B. Thomas, _Calculus and Analytic Geometry_ (Addison-Wesley Press, Reading, Mass., 1951), p. 136.

37. G. H. Hardy, et al., _op. cit._, p. 80.

38. G. B. Thomas, _op. cit._, p. 577.

39. C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. I," Inform. Contr. **10**, 65-103 (February 1967).

40. P. M. Ebert, _op. cit._, see section 2. 2.