DOCUMENT OFFICE DOCUMENT ROOM 36-412
RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

#1

ERROR BOUNDS FOR DIGITAL COMMUNICATION OVER SPATIALLY MODULATED CHANNELS

RICHARD LEWIS GREENSPAN

LOAN COPY ONLY

TECHNICAL REPORT 470

MAY 30, 1969

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), and by the National Aeronautics and Space Administration (Grant NGL 22-009-013).

Requestors having DOD contracts or grants should apply for copies of technical reports to the Defense Documentation Center, Cameron Station, Alexandria, Virginia 22314; all others should apply to the Clearinghouse for Federal Scientific and Technical Information, Sills Building, 5285 Port Royal Road, Springfield, Virginia 22151.

THIS DOCUMENT HAS BEEN APPROVED FOR PUBLIC RELEASE AND SALE; ITS DISTRIBUTION IS UNLIMITED.

D[illegible] OFFICE 26-327
RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 470 May 30, 1969

ERROR BOUNDS FOR DIGITAL COMMUNICATION OVER SPATIALLY MODULATED CHANNELS

Richard Lewis Greenspan

This report is based on a thesis submitted to the Department of Electrical Engineering, M.I.T., August 1968, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

(Revised manuscript received November 6, 1968)

Abstract

The transmission of digital information by means of electromagnetic wave propagation between terminals is reconsidered in order to understand the influence of the radiation channel, including the terminal antennas, on the quality of communication that can be maintained. In particular, we investigate the potential to improve performance by modulating the messages onto a set of spatially varying patterns across the source antenna.

We show that the radiation channel can be modeled by the parallel combination of independent spatially and temporally modulated subchannels when the background noise at the receiver has Gaussian statistics and is uncorrelated at points separated by more than a few wavelengths of the carrier frequency. Error bounds are evaluated for digital transmission with optimum distribution of signal power to the subchannels. These bounds are used to interpret the significance of the signal and channel parameters and the interplay among them.

The principal conclusion is that spatial modes have the same function as time-variant signal modes. They can be viewed as independent subchannels that extend the effective bandwidth of the communication link. The number of effective spatial modes depends on the dimension of the terminal antennas and on the input signal-to-noise ratio. For a significant application of this modulation technique, the receiving antenna must be large enough to register simultaneous spatial variations in the incident electromagnetic fields.

TABLE OF CONTENTS

CONTENTS

I. INTRODUCTION

We shall develop performance bounds for digital communication systems that transmit information from one point to another by means of electromagnetic wave propagation. We focus attention on the influence of the radiation channel, including the terminal antennas, on the quality of communication that can be established. We give special attention to modeling the channel for this communications analysis.

Figure 1 is a block diagram of a communication system with the following components:

1. Message Source. This produces one of M equally likely messages in each T-second time interval. We have chosen to work with the class of time-discrete and amplitude-discrete sources.

2. Data Processor. This transforms sequences of messages into a coded format especially designed for reliable transmission.

3. Modulator. This assigns a signal $j(\vec{r}, t)$ to each of the coder outputs for transmission over the transmitting antenna.

4. Antenna System. This system comprises transmitting and receiving antennas and the intervening free-space propagation path, including additive noise sources.

5. Demodulator. This maps each received signal $0(\vec{r}', t)$ into a set of decision variables.

6. Decoder. This processes the demodulator output to provide a sequence of messages to the User.

We are especially concerned with items 3, 4, and 5 in this chain and, in particular, with optimizing system performance by appropriate selection of the transmitted signals $j(\vec{r}, t)$. In conventional radiating systems waveforms are distinguished only on the basis of temporal variation of their characteristics, such as amplitude, phase, and frequency. The antenna system is modeled by an attenuator that influences the system reliability only by scaling the received signal-to-noise ratio (SNR). In contrast, we investigate the consequences of modulating with both temporal and spatial variations of the "complex amplitude" of the signal illumination across the antenna aperture.

The problem that we pose here is essentially independent of the radiation frequency. There is no benefit, however, from spatial modulation when the transmitting antenna looks like an unresolvable point source to the receiver. Thus, application of our results in microwave communications is unlikely because conventional antennas operating at microwave frequencies are too small to sense spatial variations of the intensity of the input signal over their surface areas. Application to optical communication systems looks promising for two reasons: (i) antennas (lenses) that are very large compared with wavelength of optical signals (a factor of 10^4 or higher) can be easily fabricated; and (ii) there may be useful applications of optical technology to communication over distances of a few tens of miles or less. Although the analysis shows that the greatest benefits are to be expected at optical frequencies, it is useful

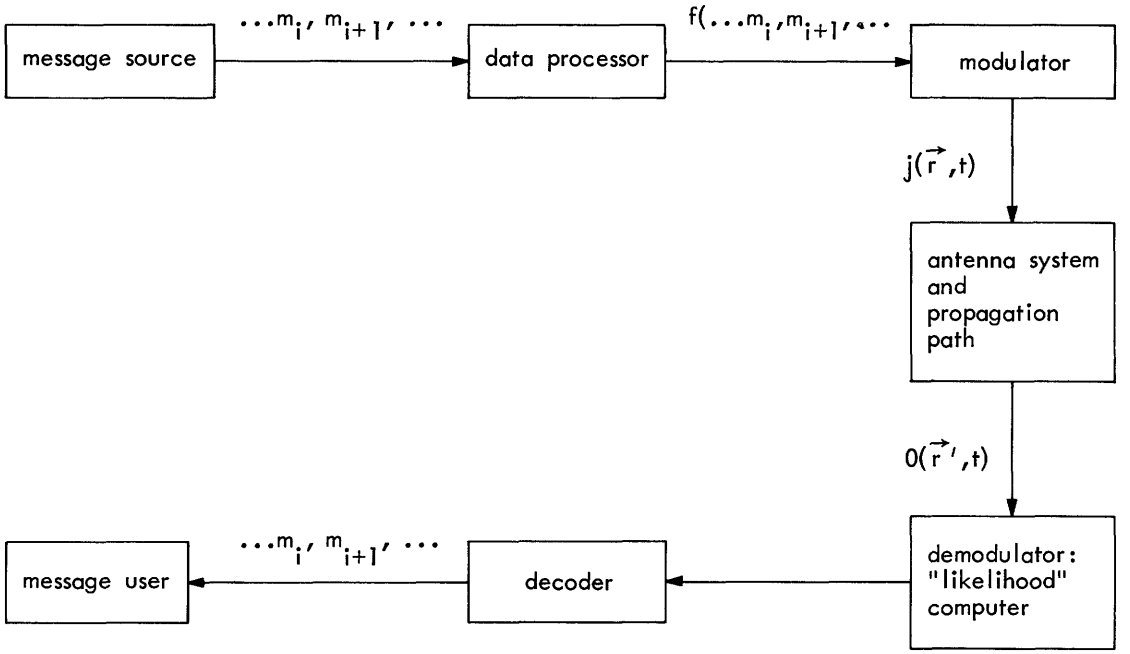


Fig. 1. Model of a communication system.

to have analytical results that confirm our intuition about the situation at lower frequencies.

Twenty years ago, Shannon[1] showed that for channels with additive Gaussian noise, there is a maximum information transmission rate called the "channel capacity" (bits/second) such that discrete bits of information can be transmitted with vanishingly small probability of error at all rates less than capacity. Subsequently, this theorem has been generalized, and estimates on the rate at which the error probability decreases to zero have been sharpened.[2-5] Our analysis is essentially an extension of the class of channels to which Shannon's theory of information transmission has been applied. We shall evaluate the "channel capacity" and the "rate-reliability" function that obtains when antennas having fixed physical dimensions are driven by the best possible spatially and temporally modulated signals. The detailed structure of these optimum signals will vary with the source information rate, the antenna system geometry, and the available signal energy and the characteristics of associated noise sources. We shall show that the influence of the antenna system is abstracted by a single gross parameter

$$D = \frac{A_1 A_2}{(\lambda Z)^2},$$

where A_1 and A_2 are the areas of the transmitting and receiving antennas, respectively; Z is the separation between antennas, and λ is the radiation wavelength. The larger D is, the more communications can be improved by use of spatial modulation.

1.1 BACKGROUND

The specific extension of coding theory to the analysis of spatially modulated signals appears to be novel. There are, however, significant mathematical roots to this work both in coding theory and in optics.

During the twenty years since Shannon's original paper on random coding, there have been many refinements of his results. Several authors, following Shannon, have shown that for each pair of filter channel impulse response and additive Gaussian noise power density spectrum there is an optimum set of average power-limited message waveforms.[6-7] For the optimum signal set the average probability of receiving a message in error is bounded by expressions of the form

$$A(R)\exp(-NE_L(R)) \leq P_e \leq \exp(-NE_U(R)),$$

where R is the information rate per channel use (relative to capacity), and N is the number of channel uses per message. The exponential terms dominate these bounds for low error probability; therefore, $E_L(R)$ and $E_U(R)$ are used for comparing communication systems. They are called the "rate-reliability" functions. We compute $E_U(R)$ for the optical channel in Section III.

Holsinger and Ebert have extended these basic results. Holsinger[8] observed that there is a unique set of orthogonal functions that are complete for representing signals at the input to a filter channel and transform into an orthogonal set at the channel output, although their functional dependence on time is altered. This result is used to represent the filter channel by the parallel combination of independent channels. Ebert[9] has studied the problem of optimum coding for such parallel channel combinations, and his results summarize the tightest available exponential bounds to the probability of error.

The particular integral equation whose solutions are the set of orthogonal basis functions we use to model the radiation channel has been studied extensively by Slepian and Pollak.[10] We draw heavily on their analysis for numerical computation and qualitative discussion of our results.

The representation of signals that propagate through free space is fundamental to the wave theory of optical instruments. Gabor has stated most clearly the classical answer to the number of degrees of freedom available from an optical instrument.[11] This number is also the antenna parameter D, but we must distinguish between the premises of communication system design and design of optical instruments to properly interpret the result. Keeping this distinction in mind, we shall discuss several related topics in optical theory and antenna theory.

1.2 SUMMARY OF THE REPORT

The central analysis of this research is reported in Sections II, III, and IV, with supporting material in three appendices. In Section II the concept of the radiation "channel" having a linear time-invariant impulse response is developed. The transmitting and

receiving antennas are seen to perform operations analogous to time-limiting and bandwidth-limiting, respectively. The integral equation whose solutions are the optimum set of orthogonal channel input waveforms is derived, and explicit solutions are given for one case of particular interest. In the rest of Section II the concept of "degrees of freedom" for the optical channel is introduced, and estimates of this number are included for several applications.

Section III details the steps for modeling the radiation channel for a communication analysis. This includes a study of additive noise sources, signal design, and the design of optimum demodulators for coded signaling. The performance of this system with the use of optimum modulation is evaluated by computing an upper bound to the rate-reliability function. Specific interpretations of the role of spatial modulation in communication systems are given. The analysis includes a discussion of the applicability of these results when quantum measurement noise is the principal source of message distortion.

In Sections II and III we develop an approach to optimum processing of spatial information in a noisy environment. In Section IV we apply this approach to some topics related to the communication problem. Perhaps the most interesting of these is the attempt to predict the dependence of visual acuity on the visual signal-to-noise ratio.

Section V recapitulates the results and includes our conclusions. Recommendations are made for extensions of the analysis that has begun here.

II. NOISE-FREE ANALYSIS

2.1 ANTENNA INPUT-OUTPUT RELATIONS

We first present a classical evaluation of the electromagnetic field set up by a source distribution, discussing certain assumptions and modeling the radiation channel by a linear filter. The field is represented by solutions to Maxwell's equations with appropriate boundary conditions imposed by sources, dielectric boundaries or conducting surfaces. These solutions are vector fields. We use a reduced, scalar analysis, because of the simple geometry of the radiating systems illustrated in Fig. 2. The components of the radiation channel are sources, diffracting apertures, and free space. The source is modeled by a linearly polarized current sheet, oriented along the x axis in the source plane R_1. The amplitude of the current density is given by $j(\vec{\rho}, t)$. We can also set up

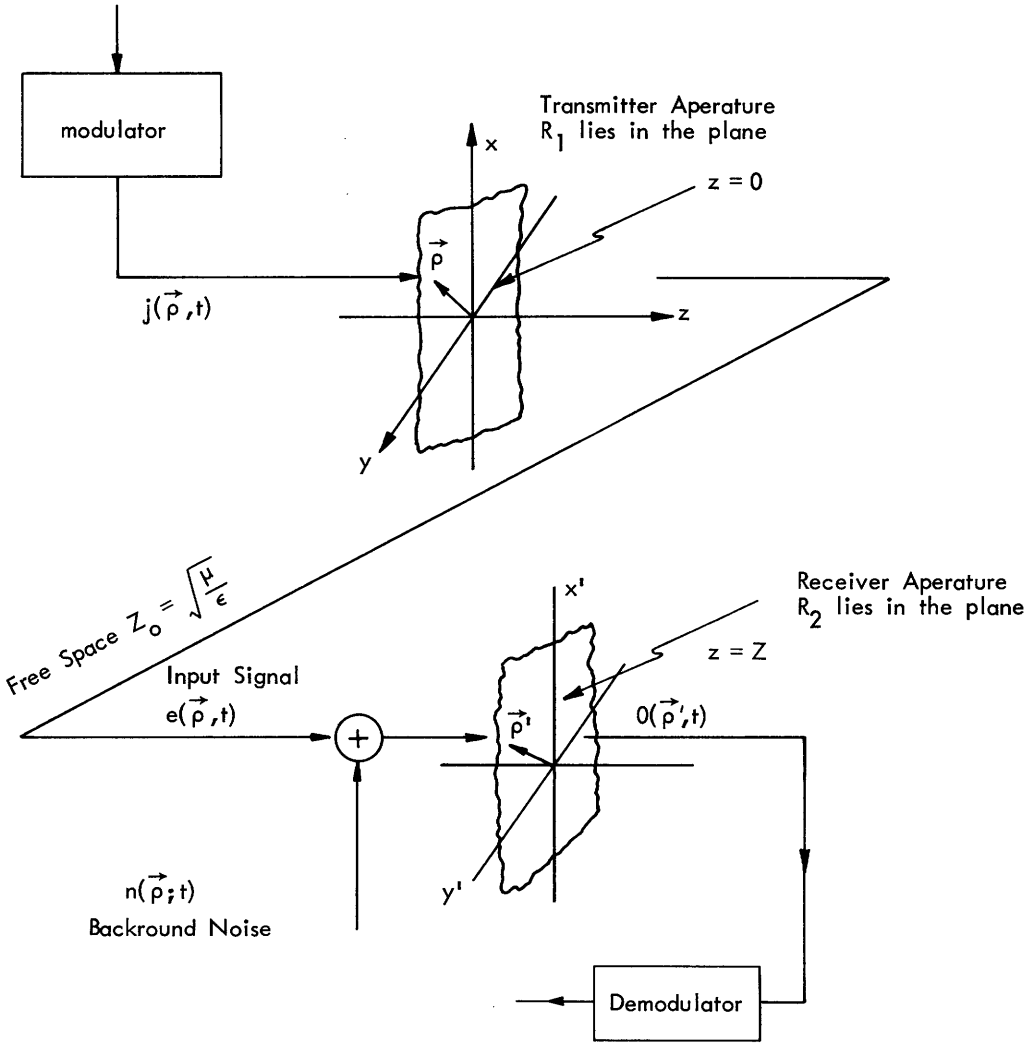


Fig. 2. Model of the antenna system geometry.

an independent field linearly polarized in the y direction. We have assumed that the field in the vicinity of the receiving aperture has an essentially plane wave front so that field measurements do not significantly couple x and y polarizations. Therefore the addition of a second, independently polarized signal field does not raise any new issues in channel modeling that limit the applicability of the scalar analysis.

We apply the Kirchhoff approximation[12] for the field behind a diffracting screen to simplify the analysis. We assume that the field behind an aperture in the screen is the same as the free-space field, and the field on the rear surface of the screen is zero. Although this approximation is mathematically inconsistent, theoretical predictions are consistent with experimental evidence at optical frequencies.

2.1.1 Electromagnetic Field in Free Space

An electromagnetic field propagating in free space satisfies Maxwell's equations with boundary conditions that characterize the source of the field. We assume that the fields are narrow-band processes. Therefore, an arbitrary field vector $\vec{U}(\vec{\rho}, t)$ is written

$$\vec{U}(\vec{\rho}, t) = \frac{1}{2} \mathrm{Re} \left[U(\vec{\rho}, t)\, e^{+i\omega_o t} \right] \vec{U},$$

where $\vec{U}$ is a unit vector in the direction of the field, and

$$U(\vec{\rho}, t) = \int_0^\infty U(\vec{\rho}, \omega)\, e^{-i\omega t}\, \frac{d\omega}{2\pi}.$$

Let $E(\vec{\rho}, \omega)$ and $H(\vec{\rho}, \omega)$ be Fourier transforms of the complex envelopes of the electric field and the magnetic field excited by a current source $J(\vec{u}, \omega)$ distributed over a finite planar input aperture. The field solutions are written in terms of spherical coordinates (r, θ, ϕ) illustrated by Fig. 3. Details of the analysis are presented in Appendix A.

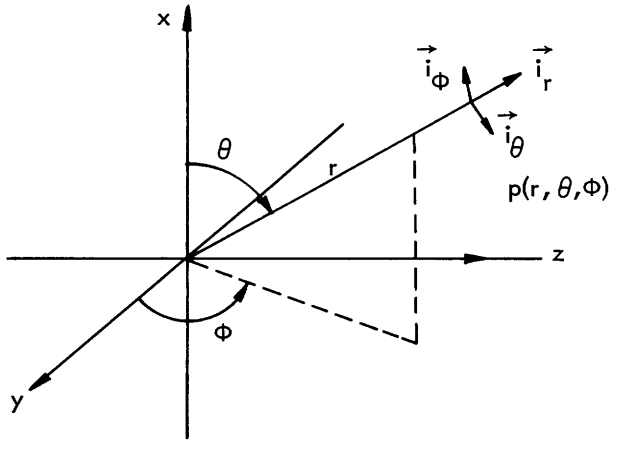


Fig. 3. Spherical coordinate system.

The field vectors are[12, 13]

$$\vec{E}(\vec{\rho},\omega) = \frac{k^2}{2\pi}\sqrt{\frac{\mu}{\epsilon}}\left[\vec{i}_r \int_{\text{sources}} J(\vec{u},\omega)\, e^{ikr}\left[\frac{1}{(ikr)^2}+\frac{1}{(ikr)^3}\right]\cos\theta\, dv\right.$$

$$\left. + \frac{\vec{i}_\theta}{2}\int_{\text{sources}} J(\vec{u},\omega)\, e^{ikr}\left[\frac{1}{ikr}+\frac{1}{(ikr)^2}+\frac{1}{(ikr)^3}\right]\sin\theta\, dV\right] \tag{1}$$

$$\vec{H}(\vec{\rho},\omega) = \frac{k^2}{4\pi}\vec{i}_\phi \int_{\text{sources}} J(\vec{u},\omega)\, e^{ikr}\left[\frac{1}{ikr}+\frac{1}{(ikr)^2}\right]\sin\theta\, dV, \tag{2}$$

where λ is the wavelength, r is the distance from the point $\vec{u}$ at the source to the point $\vec{\rho}$ in space, and

$$k = \frac{2\pi}{\lambda}$$

θ = azimuthal angle from the source to the observation point $\vec{\rho}$.

It is shown in Appendix A that the electromagnetic field diffracted by an aperture can also be approximated by Eqs. 1 and 2. In this case the source is taken to be the field in the aperture, and is assumed <u>not</u> to differ from its free-space value.

2.1.2 Radiation Channel

Let us continue the analysis, treating the electric field vector arising from a current distribution. First, simplify Eq. 1 by including only the term that varies as $1/r$.

$$\vec{E}(\vec{\rho},\omega) = \vec{i}_\theta E(\vec{\rho},\omega),$$

where

$$E(\vec{\rho},\omega) = \frac{k^2}{4\pi}\sqrt{\frac{\mu}{\epsilon}}\int_{\text{sources}} \frac{J(\vec{u},\omega)\, e^{ikr}}{ikr}\sin\theta\, dV. \tag{3}$$

This relationship has the form of a linear operator that transforms $J(\vec{u},\omega)$ into $E(\vec{\rho},\omega)$. We can write this transformation

$$E(\vec{\rho},\omega) = \int_{\text{sources}} J(\vec{u},\omega)\, h(\vec{u},\vec{\rho})\, dS,$$

where

$$h(\vec{u}, \vec{\rho}) = \frac{1}{2i\lambda r}\sqrt{\frac{u}{\epsilon}}\ \sin\theta\ e^{ikr}$$

$$r^2 = Z^2 + \|\vec{u}-\vec{\rho}\|^2. \tag{4}$$

We can interpret Eq. 4 as the defining relation for a linear filter whose impulse response is $h(\vec{u}, \vec{\rho})$. To proceed, we make the following approximations. Referring to the notation of Fig. 2, we set

$Z \gg (x^2+y^2)^{1/2}$ maximum

$\sin\theta \approx 1$: in the broadside direction $\theta \approx \pi/2$.

These approximations are attributed to Fresnel.[12] Substituting them in Eq. 4 shows that for all points $\vec{\rho}$ that are close to the optical axis (that is, within approximately 18°)

$$r \approx Z\left(1 + \frac{(x-x')^2}{2Z^2} + \frac{(y-y')^2}{2Z^2}\right).$$

Substituting this approximate value for r in the exponent in (4), and setting $r \approx Z$ in the denominator yields

$$E(\vec{\rho}, \omega) = \frac{e^{ikZ}}{2i\lambda Z}\sqrt{\frac{\mu}{\epsilon}} \int_{\text{source plane}} J(\vec{u}, \omega)\ \exp\left\{\frac{ik}{2Z}\left[(x-x')^2+(y-y')^2\right]\right\}\ dxdy. \tag{5}$$

We identify $E(\vec{\rho}, \omega)$ as the output of a linear, spatially invariant filter that has the impulse response $h_\lambda(\vec{u}, \vec{\rho})$.

$$h_\lambda(\vec{u}, \vec{\rho}) = h_\lambda(\vec{u}-\vec{\rho}) = \frac{e^{ikZ}}{2i\lambda Z}\sqrt{\frac{u}{\epsilon}}\ \exp\left[\frac{ik}{2Z}\ \|\vec{u}-\vec{\rho}\|^2\right], \tag{6}$$

where $\vec{u}$ and $\vec{\rho}$ are vectors in the input and output apertures, respectively.[14] This interpretation of (5) is significant because the well-developed techniques of communication theory for linear time-invariant channels can now be applied to evaluate the ultimate performance of the radiation channel.

Now $h_\lambda(\vec{u}, \vec{\rho})$ is defined at the single frequency $\omega = \frac{2\pi c}{\lambda}$. For a general input we write

$$h(\vec{u}, \vec{\rho}, \omega) = \frac{\mu\omega\ \exp\left[i\ \frac{\omega}{c}\ Z\right]\ \exp\left[\frac{i\omega}{2cZ}\ \|\vec{u}-\vec{\rho}\|^2\right]}{i4\pi Z}.$$

The effect of this filter on an arbitrary time function is the combination of differentiation, delay by $\left(\frac{Z}{c} + \frac{\|\vec{u}-\vec{\rho}\|^2}{2Zc}\right)$ sec, and scaling by $\left(\frac{i4\pi Z}{\mu}\right)^{-1}$. The highpass ($\omega$) variation of $|h|$ recalls the familiar result that the apparent beamwidth of an antenna is inversely proportional to the carrier frequency of the transmitted signal.

2.2 SIGNAL REPRESENTATION

The transmitted signals that represent the various source messages will be formed by taking an appropriately weighted linear sum over a set of waveforms $j_n(\vec{r}, t)$ having the following properties.

(i) The $j_n(\vec{r}, t)$ are a complete orthonormal set over the input aperture.

(ii) The $j_n(\vec{r}, t)$ give rise to an orthogonal set of functions on the receiving aperture.

Orthonormality and completeness at the input ensure a convergent representation for arbitrary inputs having finite energy. There are many complete orthonormal sets on the input aperture. We choose the unique set that transforms into orthogonal signals on the output in anticipation of the result that projections of the background noise at the receiver onto a set of orthogonal functions are statistically independent random variables. This will be discussed further in Section III.

2.2.1 Mathematical Analysis for Defining an Optimum Signal Set

As discussed above we wish to find an orthonormal set of complex functions $\{j_n(\vec{r}, t)\}$, with the following properties:

i. $j(\vec{r}, t)$ is defined on the (x, y) plane, (z = 0), illustrated in Fig. 2, $\vec{r}$ is an arbitrary vector in the region R_1, and t is limited to the interval $|t| \leq T/2$.

ii. The $j_n(\vec{r}, t)$ are orthonormal and complete on R_1, for $|t| \leq T/2$.

$$\int_{T/2}^{T/2} \int_{R_1} j_n(\vec{r}, t)\, j_k^*(\vec{r}, t)\, d\vec{r}\, dt = 0; \quad n \neq k$$

$$= 1; \quad n = k.$$

iii. Each $j_n(\vec{r}, t)$ induces an output $e_n(\vec{r}', t)$ on the (x', y') plane, (z = Z); $\vec{r}'$ is an arbitrary vector in the region R_2, and t is time-limited to the interval $|t-t_d| \leq T_o/2$, where $t_d = Z/c$ (c being the speed of light in vacuo), and T_o is an arbitrary duration. Most commonly we choose $T_o = T$.

iv. The $e_n(\vec{r}, t)$ are orthogonal on R_2 for $|t-t_d| \leq T_o/2$

$$a_{nk} = \int_{t_d-T_o/2}^{t_d+T_o/2} \int_{R_2} e_n(\vec{r}', t)\, e_k^*(\vec{r}', t)\, d\vec{r}'\, dt = 0; \quad k \neq n$$

$$= \lambda_n; \quad k = n.$$

It is convenient to present the relationship between $j_n(\vec{r}, t)$ and $e_n(\vec{r}, t)$ in terms of their Fourier transforms $J_n(\vec{r}, \omega)$ and $E_n(\vec{r}', \omega)$, respectively.

$$J_n(\vec{r},\omega) = \int_{\infty}^{\infty} j_n(\vec{r},t)\, e^{+i\omega t}\, dt$$

$$E_n(\vec{r},\omega) = \int_{\infty}^{\infty} e_n(\vec{r}',t)\, e^{+i\omega t}\, dt.$$

By substituting these transform relations in d, we find that the orthogonality condition becomes

$$a_{nk} = \int_{t_d-T_o/2}^{t_d+T_o/2} \int_{R_2} \iint_{-\infty}^{\infty} E_n(\vec{r}',\omega_1)\, E_k^*(\vec{r}',\omega_2)\, e^{-i(\omega_1-\omega_2)t}\, \frac{d\omega_1}{2\pi}\, \frac{d\omega_2}{2\pi}\, d\vec{r}'\, dt = 0; \qquad k \neq n$$

$$= \lambda_n; \qquad k = n.$$

Equation 5 gives the following relation between $E_n(\vec{r}',\omega)$ and the source distribution $J_n(\vec{r},\omega)$:

$$E_n(\vec{r}',\omega) = \frac{\omega\mu\, e^{(i\omega Z)/c}}{4\pi i Z} \int_{R_1} J_n(\vec{r},\omega)\, \exp[(i\omega)/(2cZ)\, \|\vec{r}'-\vec{r}\|^2]\, d\vec{r}.$$

Substituting this expression back into the orthogonality conditions leads to the following form for a_{nk}.

$$a_{nk} = \int_{R_2} d\vec{r}' \int_{t_d-T_o/2}^{t_d+T_o/2} dt \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega_2}{2\pi}\, e^{-i(\omega_1-\omega_2)(t-t_d)}\, \omega_1\omega_2 \left(\frac{\mu}{4\pi Z}\right)^2 \int_{R_1} d\vec{r}_1 \int_{R_1} d\vec{r}_2$$

$$\times J_n(\vec{r}_1,\omega_1)\, J_k^*(\vec{r}_2,\omega_2)\, \exp\left[(i\omega_1)/(2cZ)\, \|\vec{r}_1-\vec{r}'\|^2\right] \exp\left[(-i\omega_2)/(2cZ)\, \|\vec{r}_2-\vec{r}'\|^2\right].$$

Now assume that $i\omega J(\vec{r},\omega)\, \exp\left[+i\omega\left(t_d + \frac{\|\vec{r}-\vec{r}'\|^2}{2cZ}\right)\right]$ is the Fourier transform of a time function that is zero for $|t-t_d| \leq T_o/2$. Observe that $\omega J(\vec{r},\omega)$ is the Fourier transform of $dj(\vec{r},t)/dt$. Therefore the assumption applies whenever $\max \|\vec{r}-\vec{r}'\|^2/2cZ \leq T_o/2$. This condition requires that the received signal be confined within the processing time, with variations in delay across the receiver aperture taken into account. In other words, it is an intersymbol-interference criterion in systems involving sequential use of the channel. Assume that this condition is satisfied, then the integration of t and ω_2 can be easily done, since

$$\int_{t_d-T_o/2}^{t_d+T_o/2} \frac{dt}{2\pi}\, e^{-i(\omega_1-\omega_2)(t-t_d)} \Rightarrow \int_{-\infty}^{\infty} e^{-i(\omega_1-\omega_2)t}\, \frac{dt}{2\pi} = \delta(\omega_1-\omega_2).$$

It follows that

$$a_{nk} = \int_{R_2} d\vec{r}' \int_{R_1} d\vec{r}_1 \int_{R_1} d\vec{r}_2 \int_{-\infty}^{\infty} \left(\frac{\mu}{4\pi Z}\right)^2 \frac{d\omega}{2\pi} J_n(\vec{r}_1, \omega)\, J_K^*(\vec{r}_2, \omega)\, \omega^2$$

$$\cdot \exp\left\{+i \frac{\omega}{2cZ}\left[\|\vec{r}_1 - \vec{r}'\|^2 - \|\vec{r}_2 - \vec{r}'\|^2\right]\right\}. \tag{7}$$

Expand the exponential and regroup terms in the integral.

$$a_{nk} = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left(\frac{\omega\mu}{4\pi Z}\right)^2 \int_{R_1} d\vec{r}_1\, J_n(\vec{r}_1, \omega) \exp\left[+\frac{i\omega}{2cZ}\|\vec{r}_1\|^2\right]$$

$$\cdot \int_{R_1} d\vec{r}_2\, J_k^*(\vec{r}_2, \omega) \exp\left[-\frac{i\omega}{2cZ}\|\vec{r}_2\|^2\right] \int_{R_2} \exp\left[-\frac{i\omega}{cZ}\vec{r}'(\vec{r}_1 - \vec{r}_2)\right] d\vec{r}'.$$

Now identify the following functions

$$K_{R2}(\vec{r}_1 - \vec{r}_2) = \left(\frac{\omega}{2\pi cZ}\right)^2 \int_{R_2} \exp\left[-\frac{i\omega}{cZ}\vec{r}'(\vec{r}_1 - \vec{r}_2)\right] d\vec{r}'$$

$$J'(\vec{r}, \omega) = J(\vec{r}, \omega) \exp\left[+\frac{i\omega}{2cZ}\|\vec{r}\|^2\right].$$

It follows that

$$a_{nk} = \frac{\mu}{4\epsilon} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \int_{R_1} d\vec{r}_1\, J_n'(\vec{r}_1, \omega) \int_{R_1} d\vec{r}_2\, J_k^*(\vec{r}_2, \omega)\, K_{R_2}(\vec{r}_1 - \vec{r}_2).$$

Let us suppose that there are real functions, $J'(\vec{r}, \omega)$, that are orthonormal on R_1 and for which

$$\int_{R_1} d\vec{r}_2\, J_k'(\vec{r}_2, \omega)\, K_{R_2}(\vec{r}_1 - \vec{r}_2) = \beta_k(\omega)\, J_k'(\vec{r}_1, \omega). \tag{8}$$

Let $J_k'(\vec{r}_1, \omega)$ be a solution to Eq. 8 with the eigenvalue β_k. We shall see that the eigenvalues cluster around one and zero with few falling in between. Substitute this solution into the expansion for a_{nk}. Then

$$a_{nk} = \frac{\mu}{4\epsilon} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \int_{R_1} d\vec{r}_1\, J_n'(\vec{r}_1, \omega)\, \beta_k(\omega)\, J_k'^*(\vec{r}_1, \omega)$$

$$= \frac{\mu}{4\epsilon} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \beta_k(\omega) \quad \text{if } n = k$$

$$= 0 \quad \text{otherwise.}$$

Therefore, our original problem has been reduced to studying a certain linear integral equation. Observe that if $J'(\vec{r},\omega)$ satisfies (8), then $C(\omega)\, J'(\vec{r},\omega)$ also satisfies this equation, as long as the intersymbol-interference condition is satisfied. We now derive a useful property of the general solutions.

If $J'(\vec{r},\omega)$ satisfies (8), then an input $J(\vec{r},\omega) = C(\omega)\, J'(\vec{r},\omega) \exp[-i\omega/2cZ \,\|r\|^2]$ also gives rise to an $E(\vec{r},\omega)$ that satisfies the output orthogonality conditions. Let us check for orthogonality at the input. Define the source correlation function b_{nk}, by analogy to a_{nk}.

$$b_{nk} = \int_{-T_o/2}^{T_o/2} \int_{R_1} \iint_{-\infty}^{\infty} J_n(\vec{r},\omega)\, J_k^*(\vec{r},\omega_2)\, e^{-(\omega_1-\omega_2)t}\, \frac{d\omega_1}{2\pi}\, \frac{d\omega_2}{2\pi}\, d\vec{r}\, dt.$$

Again, observe that if $j_n(\vec{r},t)$ is essentially time-limited to $|t| \leq T_o/2$, then the integration on t can be extended to infinity. Then,

$$b_{nk} = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \int_{R_1} J_n(\vec{r},\omega)\, J_k^*(\vec{r},\omega)\, d\vec{r}$$

$$= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, C_n(\omega)\, C_k^*(\omega) \int_{R_1} J_n'(\vec{r},\omega)\, J_k'^*(\vec{r},\omega)\, d\vec{r}$$

$$= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} |C_n(\omega)|^2 = b_{nn} \qquad \text{if } k = n$$

$$= 0 \qquad \text{if } k \neq n.$$

Clearly, the weighting function, $C_n(\omega)$, can be chosen to ensure that the $\{J_n(\vec{r},\omega)\}$ have unit energy if we choose not to normalize the spatially dependent factor. We can construct a set of orthonormal input functions $\{J_{nl}(\vec{r},\omega)\}$ for each mode as follows.

$$J_{nl}(\vec{r},\omega) = C_{nl}(\omega)\, J_n'(\vec{r},\omega) \exp\left[-\frac{i\omega}{2cZ} \|\vec{r}\|^2\right] \qquad l = 1, 2, \ldots M, \tag{9}$$

where

$$\int_{-\infty}^{\infty} C_{nl}(\omega)\, C_{nl'}^*(\omega)\, \frac{d\omega}{2\pi} = 1 \qquad l = l'$$

$$= 0 \qquad \text{otherwise.}$$

Therefore, the appropriate expansion functions have a frequency-domain representation that is the product of a function of frequency and space by a function of frequency only. We indicate below that the spatial function will be only weakly dependent on frequency for small fractional bandwidth. In that case one can interpret the expansion functions

as spatial "carriers" modulated by a set of orthogonal time functions, or vice versa. At optical frequencies a 10% bandwidth is approximately 500 Å, and the interpretation of $J(\vec{r}, \omega)$ as a spatial carrier is realistic. If, however, one proposes to signal with a wide enough bandwidth that color variations are significant, then we must recognize that each $j_n(\vec{r}, t)$ is the convolution of a time function c(t) with $j'_n(\vec{r}, t)$.

Observe that the same set of time functions can be used to modulate each of the orthogonal spatial carriers. It follows that there are available approximately D' coordinates, where D' is the product of the "number" of time functions and the number of spatial carriers. We return to this point in section 2.4a, although a precise formulation of the degrees-of-freedom concept must wait until we consider the effects of additive noise on the performance of the system. This noise is the ultimate limit on the useful degrees of freedom.

In order to proceed with our interpretation we have to study Eq. 8. It turns out that it can be solved for some interesting geometries.

2.2.2 Solutions for Special Coordinate Systems

We are looking for real functions, $J'(\vec{u}, \omega)$, which have the property that

$$\int_{R_1} d\vec{u}_2 \, J'_k(\vec{u}_2, \omega) \, K_{R2}(\vec{u}_1 - \vec{u}_2) = \beta_k(\omega) \, J'_k(\vec{u}_1, \omega),$$

where

$$K_{R2}(\vec{u}_1 - \vec{u}_2) = \left(\frac{\omega}{2\pi c Z}\right)^2 \int_{R_2} \exp\left[-\frac{i\omega}{cZ} \vec{r}' \cdot (\vec{u}_1 - \vec{u}_2)\right] d\vec{r}'.$$

Observe, first, that K_{R2} is a Hermitian kernel; that is, $K_{R2}(\vec{u}_2 - \vec{u}_1) = K^*_{R2}(\vec{u}_1 - \vec{u}_2)$, the conjugate kernel. Second, $\int_{R_1} |K_{R2}(\vec{u})|^2 \, d\vec{u}$ is bounded for finite apertures R_1 and R_2. In fact,

$$\int_{R_1} |K_{R2}(\vec{u})|^2 \, d\vec{u} = \left(\frac{\omega}{2\pi c Z}\right)^2 \int_{R_1} d\vec{u} \int_{R_2} \int_{R_2} \exp\left[-\frac{i\omega}{cZ} \vec{u} \cdot (\vec{r}' - \vec{r}'')\right] d\vec{r}' \, d\vec{r}''$$

$$\int_{R_1} |K_{R2}(\vec{u})|^2 \, d\vec{u} \leq \left(\frac{\omega}{2\pi c Z}\right)^2 \int_{R_1} \int_{R_2} \int_{R_2} \left| e^{-\frac{i\omega \vec{u}}{cZ} \cdot (\vec{r}' - \vec{r}'')} \right|^2 d\vec{r}' \, d\vec{r}''$$

$$= \left(\frac{\omega}{2\pi c Z}\right)^2 A_1 A_2^2,$$

where A_1 and A_2 are the areas of R_1 and R_2, respectively, and K_{R2} is a positive definite kernel. For all $f(\vec{u})$ on R_1 such that

$$\int_{R_1} |f(\vec{u})|^2 \, dx \geq 0$$

$$\int_{R_1}\int_{R_1} f(\vec{u}_1)\, f^*(u_2)\, K_{R2}(\vec{u}_1-\vec{u}_2)\, d\vec{u}_1 d\vec{u}_2 = \left(\frac{\omega}{2\pi cZ}\right)^2 \int_{R_2} \left| \int_{R_1} f(\vec{u}) \exp\left[-\frac{i\omega}{cZ}\vec{r}'\cdot\vec{u}\right] d\vec{u} \right|^2 d\vec{r}'$$

$$\geq 0.$$

It follows that the eigenvalues of Eq. 8 are real and positive, and the eigenfunctions are orthogonal on R_1 and also complete in the class of square-integrable functions on R_1.[15] Therefore, any arbitrary function of spatial variables can be represented within arbitrarily small error by a series expansion in these eigenfunctions.

Equation 8 has been solved for only a few simple geometries. Slepian has been most intimately involved with this problem during recent years, and in the following exposition we rely heavily on his results.[16] He shows that (8) simplifies if R_2 is a scaled version of R_1; that is, if $\vec{X}$ is in R_1, then $\delta\vec{X}$ is in R_2, and if R_1 is symmetric, $\vec{X}$ in R_1 implies $-\vec{X}$ in R_1. Finally, if $\psi(\vec{X})$ is a solution to Eq. 8 in R_1 with the eigenvalue β, then

$$a\psi(\vec{X}) = \int_{R_1} \exp\left(-i\frac{\delta\omega}{cZ}\vec{x}\cdot\vec{y}\right)\psi(\vec{y})\, d\vec{y}, \tag{10}$$

with

$$a = \left(\frac{2\pi}{\delta\omega}cZ\right)\beta^{1/2}.$$

Equation 10 shows that the solutions to (8) are eigenfunctions of the finite Fourier transform. Slepian has solved this problem for circular apertures, R_1 and R_2, but we refer to his earlier results for rectangular apertures[10] for a simple illustrative example.

In particular, let R_1 be a rectangular aperture in the input plane $z = 0$ with $|x| \leq L_x$ and $|y| \leq L_y$, and R_2 be a rectangular aperture in the output plane $z = Z$ with $|x'| \leq L'_x$ and $|y'| \leq L'_y$.

$$K_{R2}(\vec{u}_1-\vec{u}_2) = \left(\frac{\omega}{2\pi cZ}\right)^2 \int_{-L'_x}^{L'_x}\int_{-L'_y}^{L'_y} dx'dy' \exp\left\{-\frac{i\omega}{cZ}\left[x'(u_{1x}-u_{2x})+y'(u_{1y}-u_{2y})\right]\right\}$$

$$= \frac{\sin\frac{\omega L'_x}{cZ}(u_{1x}-u_{2x})}{\pi(u_{1x}-u_{2x})} \cdot \frac{\sin\frac{\omega L'_y}{cZ}(u_{1y}-u_{2y})}{\pi(u_{1y}-u_{2y})}.$$

Inasmuch as K_{R2} factors into product of functions along the x and y axes, we find that the solutions to (8) also separate in these variables: $J'(\vec{u},\omega) = J'_1(x,\omega)\, J'_2(y,\omega)$

where $\vec{u} = x\vec{i}_x + y\vec{i}_y$.

$$\int_{-L_x}^{L_x} J_1'(x', \omega) \frac{\sin \frac{\omega L_x'}{cZ}(x-x')}{\pi(x-x')} = \beta_1(\omega)\, J_1'(x, \omega) \tag{11a}$$

$$\int_{-L_y}^{L_y} J_2'(y', \omega) \frac{\sin \frac{\omega L_y'}{cZ}(y-y')\, dy'}{\pi(y-y')} = \beta_2(\omega)\, J_2'(y, \omega). \tag{11b}$$

Integral equations of the form (11) are well known. The solutions are prolate-spheroidal wave functions of zero-order (m = 0).[17-19] Some properties of these eigenfunctions are summarized in Appendix B. At this point we observe only that their dependence on frequency ω is carried by the parameters D_x and D_y.

$D_x = \frac{2\omega}{\pi cZ} L_x L_x'$ for variations in the x direction

$D_y = \frac{2\omega}{\pi cZ} L_y L_y'$ for variations in the y direction.

The eigenvalues are monotonically increasing functions of D. For a fixed value of D they are monotonically decreasing with the order of their associated eigenfunction. Roughly speaking, the n^{th} eigenvalue is unity for $n \leq D$ and falls rapidly to zero for larger n. This behavior is illustrated in Fig. 4. We shall investigate the variation of $\beta_n(D)$ more carefully in conjunction with the performance analysis in section 3.3.

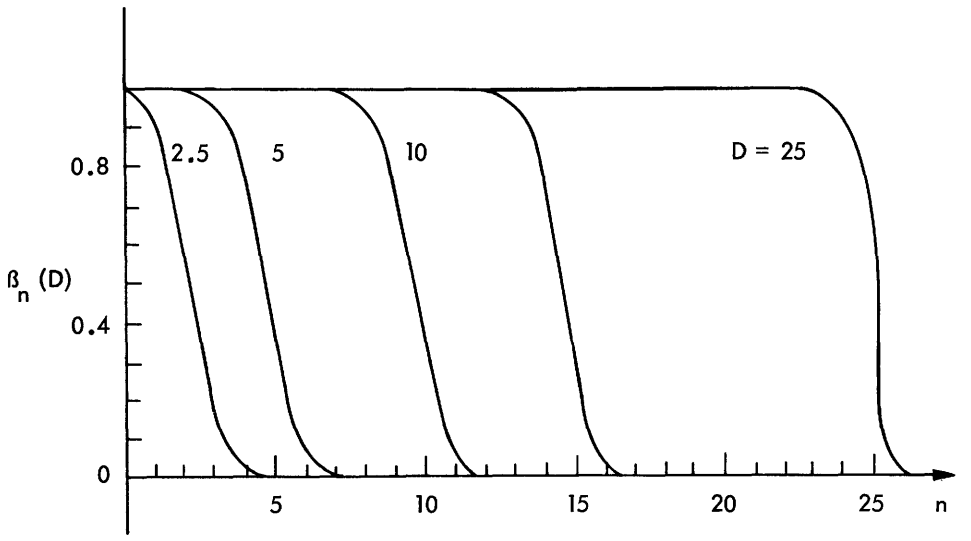


Fig. 4. Eigenvalues of the prolate spheroidal wave equation.

2.2.3 Discussion

Equation 8 arises in connection with other applications. For example, the maximum power is coupled between apertures sited at R_1 and R_2 if the spatial variation of the input signal is chosen to be the eigenfunction of (8) having the largest eigenvalue.[20-23] The net power transmitted in each mode is proportional to the eigenvalue. Also, Gamo[24] observed that the transmission kernel K could be expanded as a series of products of orthogonal functions that are solutions of Eq. 8. Although he does not explicitly make the connection, it follows directly that the functions can be paired as input and corresponding output to the optical system.

The work that is closest in spirit to our research has been done recently on the computation of resonant mode patterns within a cavity terminated by confocal reflectors.[25-27] Boyd and Gordon characterize a mode as a "field distribution that reproduces itself in spatial distribution and phase, though not in amplitude, as the wave bounces back and forth between the reflectors."[25] This definition recalls our observation that the solutions to (8) reproduce themselves, in scaled version, when processed by the finite Fourier transform operator. In fact, the results of Boyd and Kogelnik[26] do show that the solutions to (8) are the modes when the reflectors are of similar shape but different size. Therefore any progress in solving that equation for new geometries will be useful for both their analysis and ours. Moreover, this exact relationship implies a technique for synthesizing the various spatial coordinates if the cavity modes can be separated and their relative amplitudes properly controlled. With respect to waveform synthesis recall that

$$J(\vec{u},\omega) = C(\omega)\, J'(\vec{u},\omega)\, \exp\left[\frac{i\omega}{2cZ}\,\|\vec{u}\|^2\right].$$

The quadratic phase factor is equivalent to the phase added by a thin lens of focal length Z. Therefore, the optimum source is realized by focusing the eigenfunction solution $C(\omega)\, J'(\vec{u},\omega)$ onto the receiver aperture.

A second interesting group of observations concerns the distribution of eigenvalues as a function of the antenna system parameters. The number of spatial eigenfunctions that correspond to eigenvalues that are greater than 1/2 is approximately $D_s = D_x D_y = 16\, L_x L_y L'_x L'_y/(\lambda Z)^2 = A_1 A_2/(\lambda Z)^2$, where A_1 and A_2 are the areas of R_1 and R_2, respectively. This dependence on the aperture areas also obtains for circular apertures, and we conjecture that it is approximately true for arbitrarily shaped apertures.

Can the analysis be extended to other aperture geometries? Although we have no further exact results, the following arguments are useful. If one has two input apertures R_1 and R'_1, and R'_1 is contained within R_1, then the optimum performance obtained by using R'_1 is at best equal to the performance by using R_1. This follows because one can constrain the optimization over R_1 to use functions that are nonzero only over R'_1. The constrained optimum cannot exceed the unconstrained optimum. This argument clearly applies directly to comparing the performance of different-sized receiving

apertures. Note then that one can use this technique to get upper and lower bounds on the performance of arbitrary apertures, using known results for circular and rectangular apertures.

When the transmitted signal is limited to the narrow temporal bandwidth W, then the total number of significant eigenfunctions is $2TWD_s$. Therefore, we call D_s the "space-bandwidth" product, by analogy with 2TW, which is the "time-bandwidth" product. Observe that the apparent beamwidth of the source antenna is approximately λ^2/A_1 steradians. Therefore, D_s is approximately equal to the number of "beamwidths" of aperture 1 intercepted by aperture 2, and vice versa.

2.3 SPATIAL DEGREES OF FREEDOM

We use "degrees of freedom" to designate the coefficients that can be independently adjusted in the series expansion of each member of an arbitrary set of functions. It is useful to ask how many degrees of freedom are required to approximate the class to within a specified tolerance. We discuss this question as applied to the representation of two-dimensional waveforms across the source and receiver apertures of the communication link described in section 2.1.

Observe that this discussion is concerned with finding an efficient representation for arbitrary signals at the source aperture and their transformations at the receiving aperture. We shall eventually return to our principal investigation of how one designs optimum signals to convey information through the channel when the received signal is corrupted by additive noise. That analysis shows that one should spread the available average signal energy among the i_b solutions of Eq. 8 which have the largest eigenvalues. For moderate values of the signal-energy to noise-power density ratio, we find that i_b is approximately equal to D, where D is a parameter in (8) that is computed from the antenna system geometry. We shall show that there is good reason to claim that D can also be interpreted as the number of degrees of freedom in the propagation channel. For rectangular apertures $D = D_s$, as defined above. It follows that D is an important parameter for optimum signal design. For larger values, i_b increases somewhat slower than increases in the logarithm of the signal-to-noise ratio.

The coordinate functions of the expansion developed in section 2.2 are orthogonal over the source aperture and complete in the space of square-integrable functions defined on the aperture. They have two additional useful properties.

(i) The propagation channel uniquely transforms each orthogonal function at the source into one of a complete set of orthogonal functions at the receiver.

(ii) The received energy in the i^{th} orthogonal function is scaled by a factor of $\{\beta_i\}$. D of the $\{\beta_i\}$ are approximately unity; the rest are approximately zero.

We take the quality of a series approximation to a function to be measured by the energy in the remainder. This energy is easily shown to be a weighted sum of the $\{\beta_i\}$. This is one additional reason that the solutions to Eq. 8 have convenient properties for

a discussion of spatial degrees of freedom.

Recall that the eigenfunctions of (8) are also eigenfunctions of the finite Fourier transform. If one correctly handles quadratic phase terms, the channel output is the Fourier transform of the source distribution. The output voltage at a point X'_o on the x' axis is proportional to the Fourier transform component of the source at the spatial frequency $X'_o/\lambda Z$ (lines/inch). The fact that the output aperture is limited in extent means that the receiver can detect only frequency components of the input smaller than $L'_x/\lambda Z$ in the x direction and $L'_y/\lambda Z$ in the y direction. Therefore, the cascade of free-space propagation plus a finite-size receiving aperture functions like a lowpass band-limited filter for spatial modulation at the channel input. Furthermore, the finite size of the input aperture has an effect that is analogous to "time-limiting" the extent of the source inputs. In fact, it is helpful to think of the channel as the spatial analog to a bandlimited channel with "time limiting" at the input.

Having this insight, we can apply results recently derived by Landau and Pollak, to make a precise statement about the number of degrees of freedom of the channel. We quote from a paper in which they discuss the degrees of freedom in a one-dimensional, bandlimited channel.[29]

> "The purpose of this paper is to examine the mathematical truth in the engineering intuition that there are approximately 2TW independent signals (ϕ_i) of bandwidth W, concentrated in an interval of length T. Roughly speaking the result is true for the best choice of the ϕ_i (prolate-spheroidal wave function) but not for the sampling functions (of the form sin t/t). Some typical conclusions are:
>
> Let f(t), of total energy 1, be bandlimited to bandwidth W, and let
>
> $$\int_{-T/2}^{T/2} |f(t)|^2 \, dt = 1 - \epsilon_T^2.$$
>
> Then
>
> $$\inf_{(a_i)} \int_{-\infty}^{\infty} \left| f(t) - \sum_{n=0}^{[2TW]+N} a_n \phi_n \right|^2 dt \leq A\epsilon_T^2$$
>
> is
>
> (a) true for all such f with N = 0, A = 12 if the (ϕ_n) are prolate spheroidal wave functions.
>
> (b) false for some f, for any finite constants N and A if the (ϕ_n) are sampling functions."

Now, apply the analogy between spatial and temporal variables to this theorem to get the following results for a one-dimensional source. Let f(x) be of total energy 1 and limited by the input aperture to $|x| \leq L_x/2$. If all but ϵ_T^2 of the energy in f(x) is confined to spatial frequencies smaller than W_s, where $W_s = L'_x/\lambda Z$, then the error in approximating f(x) by a linear combination of the first $2L_xW_s$ prolate spheroidal wave functions is less than $12\,\epsilon_T^2$. Recall that since f(x) is the field at the source antenna, then according

to our previous arguments, $1 - \epsilon_T^2$ of the channel output energy is confined to a beam that intersects the receiving antenna over $2L'_x$ (meters).

We shall identify $D = 2L_x W_s = 2L_x L'_x/\lambda Z$, the scaled product of antenna dimensions, with the number of degrees of freedom in an antenna system. For two-dimensional, planar apertures, D is the scaled product of aperture areas.

The observation that the channel apertures impose limitations on the size (duration) and detail (bandwidth) of the source object that can be faithfully reproduced suggests that there is a potential trade between these two quantities. In fact, it is possible to scale up the object so that one can image a small area with high resolution; conversely, one can transmit a reduced version of a large area at the cost of poorer resolution because high-frequency components are attenuated.

There is another possible trade, namely between temporal and spatial variables. The factor $C(\omega)$ in Eq. 10 implies a temporal degeneracy to each spatial mode. We conclude that the spatial functions can be used as extra orthogonal "carriers" for signals that cannot be transmitted within a limited temporal bandwidth. Similarly, one can transmit pictorial (spatial) data sequentially via time modulation.

2.3.1 Comparison of Analyses of Degrees of Freedom

The most insightful analysis of the degrees of freedom in an electromagnetic wave has been given by Gabor.[11] Although he credits von Laue[30] with the essential result, this work has not been as accessible or as stimulating as Gabor's expositions. He formulated the question as follows: How many independent variables are necessary to express as much of the function $t(x, y)$, the complex amplitude transmission of an object in the source plane, as we can learn from an optical image (of finite extent)."

Gabor's analysis proceeds from the observation that the source can be represented in terms of its Fourier components, which stand for periodic plane waves propagating at all directions to the source plane. The intermediate step is to show that by proper selection of the Fourier components one can construct "elementary beams" that satisfy the following relation with equality.

$$\frac{A_1 \text{ (smallest effect beam area)} \times \Omega \text{ (solid angle of divergence)}}{\lambda^2 \text{ (wavelength)}^2} \geq 1.$$

If A_2 is the area of the receiving aperture, then each elementary beam is limited in extent to A_2 only if

$$D_S = \frac{A_1 A_2}{(\lambda Z)^2} \geq 1.$$

As D_S increases we can, in principle, cover the output aperture with D_S "nonoverlapping" elementary beams. It is clear that we cannot get completely disjoint beams with only a limited range of Fourier components. Gabor notes, however, that "physicists have their own standards in these matters," and that the accuracy of the

approximation improves as D_S increases.

It is precisely this point that is clarified by the analyses of Slepian and Pollak[10] and by Landau and Pollak.[29] Their results show that there are D_S orthogonal functions having essentially unit energy in the space of functions defined on the source and receiver apertures. The difference in their approach lies in the definition of independent variables. Gabor counts the complex amplitudes of "nonoverlapping" beams, which are essentially orthogonal, whereas Pollak allows beams to overlap but assumes that they can be separated by means of strictly orthogonal operations. Therefore, the orthogonal functions occupy the entire accessible area of the source aperture, in contrast to occupying $1/D_S$ of the area.

The number of degrees of freedom does not depend upon which aperture is considered to be the source and which is to be the receiver. If the smaller aperture happens to be the source, one can characterize the receiver processing as an attempt to use the larger aperture to focus down to get details of the source. Similarly, the source focuses multiple beams on the receiver if it has a larger transmitting antenna.

2.3.2 Further Application of Spheroidal Wave Functions in Optics

Three properties of the solutions to Eq. 10 furnish the mathematical basis for new approaches to familiar problems in applied electromagnetic theory. We discuss these results for the special case of rectangular apertures, but they can be generalized to a wider class of geometric constraints. Let $\phi_i(x)$ be the i^{th} prolate spheroidal function. Then

(i) The functions $\{\phi_i(x)\}$ form a complete orthogonal set on the finite interval $|x| \leq L/2$.

(ii) The functions are also orthogonal on the infinite interval.

(iii) The finite transform of $\phi_i(x)$ is again ϕ_i with an appropriately scaled argument. Among sets of orthogonal functions on the interval $|x| \leq L/2$, the $\{\phi_i\}$ have the largest energy concentrated in the range of frequencies $|\omega| \leq W_s$.

The first application of these properties is to antenna pattern synthesis.[31-33] Property (i) ensures that one can approximate any desired radiation pattern by spheroidal wave functions over a finite interval. When a transmitting antenna is focused at infinity its far-field radiation pattern is the Fourier transform of the field across the source aperture. Therefore, we can invoke the self-transform property (iii) to show that the source for the desired radiation pattern is synthesized by a linear combination of spheroidal wave functions. The coefficients in this expansion are the same coefficients as those in the expansion of the radiation pattern.

In addition to simplifying (computationally) the pattern synthesis problem, the analysis can also be applied to illuminate the issues surrounding "super resolution." Woodward and Lawson[34] showed that the difficulty in getting a radiation pattern much narrower than that of a uniformly illuminated aperture is "practical" rather than theoretical in nature, namely (a) it requires very accurate control of large currents in the

aperture, and (b) a relatively large amount of energy is stored in inductive fields around the aperture. This energy is not radiated.

These conclusions are easily seen from the representation of a radiation pattern in terms of spheroidal wave functions. Thus, if the desired beam is relatively narrow, the approximation requires more high-frequency spatial components. But these undergo the largest attenuation by propagation. Therefore, the corresponding functions used to synthesize the source will be given relatively large amplitude. Moreover, these high-frequency functions oscillate rapidly within the source aperture. Then too, spatial frequencies in the source that are higher than 1/wavelength do not propagate. Their energy is stored in evanescent waves in the vicinity of the source. The fraction of source energy tied up in these frequencies must increase as one requires increasingly better approximations to a narrow-beam pattern.

The second application is called apodization,[20-22] and deals with maximizing the transfer of energy between two antennas. In particular, we look for the source distribution that yields the most concentrated beam in the Fourier transform domain. From property (iii), the spheroidal wave functions are solutions to this maximization problem. Slepian[20] has published the apodization result for optical systems, in which the light distribution on the rear focal plane of a lens is the Fourier transform of the object distribution. Borgiotti[23] observed the same property with regard to power transfer between antennas. In a similar vein, Walther[35] relates the energy flux in an optical image plane to the degrees of freedom of the optical system and radiance of the source in the object plane, using the spheroidal wave function expansion. Finally, several authors have proposed that one can take advantage of the double-orthogonal property (ii) to get "super-resolution" in an imaging system.[36-40] We shall review this application in section 4.1.

2.4 TEMPORAL DEGREES OF FREEDOM

In the preceding discussion we suppressed the role of temporal modulation of the signal carriers. This enabled us to concentrate on characteristics of the spatial modulation, but it also reflects the fact that in most cases these processes can be considered separately. It is only when one seeks to determine ultimate performance limits imposed by the channel that the interaction between these modulation modes becomes prominent.

Recall that the channel is modeled by a space and time-invariant linear filter, having the impulse response function $h(r, \omega)$, where

$$h(\vec{r}, \omega) = \frac{\omega\mu \exp\left\{-\frac{i\omega}{c}\left[Z - \frac{\|\vec{r}\|^2}{2Z}\right]\right\}}{4\pi i Z}.$$

We see that the impulse response has the characteristics of a highpass filter in cascade with a variable propagation delay. The percentage of variation in either the amplitude or phase of $h(\vec{r}, \omega)$ caused by varying ω by $\delta\omega$ is

$$\delta h = 100 \frac{\delta\omega}{\omega}.$$

It follows that distortion of the temporal structure of the signal by the channel filter is negligible unless the modulation approaches a 10% bandwidth; that is, 3×10^{11} Hz for a CO_2 laser and 5×10^{12} Hz for a Helium-Neon laser. We conclude that the optical channel imposes no practical constraint on the bandwidth of the radiation field. Such constraints that exist will reflect limits on contemporary technology of wideband modulation of coherent sources.

This conclusion is supported by analysis of the functional dependence of the spheroidal wave functions on frequency variations. The frequency dependence shows up only through dependence on the degrees of freedom, D. As shown in Appendix B, this dependence is relatively weak; for example, for moderate values of D.

$$\psi(x) \approx \frac{\cos\left[\frac{\pi}{2} Dx + \frac{\delta}{2} \ln \frac{1+x}{|1-x|} - \frac{n\pi}{2}\right]}{|1-x^2|^{1/2}} \qquad 0 < x \leq 1 - \left(\frac{\pi}{2} D\right)^{-1/2}.$$

Nevertheless, some interesting issues arise in the case of large signal bandwidths. We ask the following question: Is there an interaction between the spatial characteristics of an antenna pattern and the temporal degrees of freedom associated with that mode? The answer is "Yes," for two reasons, both based on the fact that the spatial bandwidth of the n^{th} spheroidal wave function is monotonically increasing with n.

First, observe that there is a delay differential between signals arriving at the center of the receiving aperture and signals arriving at the edges. It takes a finite time for each successive spatial pattern to be built up across the receiving antenna. Inasmuch as the field incident at the receiver is the Fourier transform of the source, it follows that this "transit-time" constraint is relatively more restrictive for higher order modes.

The more interesting effect is that the effective temporal bandwidth of the spatial modes varies inversely with the mode number. That is, the n^{th} spatially dependent spheroidal function is independent of frequency (wavelength) in a bandwidth that decreases as n increases. In effect, these higher order modes tend to act more like resonant circuits having a relatively large part of the source energy stored in inductive fields around the antenna. [In a practical system one would not force more detail onto the transmitting aperture than corresponds to an optical wavelength. This argument is included for completeness of the mathematical analysis.] This phenomenon is caused by the relatively rapid rates of change of the field strength across the aperture in the higher order modes. It is the same effect that occurs with "super-gain" radiation patterns, which have a relatively narrow modulation bandwidth and are quite sensitive to deviations of the source excitation from nominal design specifications.

We estimate the bandwidth over which the i^{th} spatial mode distribution is independent

of frequency. In principle, one can compute the number of effective degrees of freedom for wider bandwidths, but that turns out not to be worth the effort.

2.4.1 Effective Bandwidth of the Spatial Modes

Our estimate of the number of temporal degrees of freedom for narrow-band modulation is based on the following property of the optimum spatial modes.

Let $F_k(\vec{r}, \omega)$ be defined as one of the set of orthonormal solutions to (8) on the source aperture. Consider the functions $F_{ik}(\vec{r}, \omega)$, where $F_{ik}(\vec{r}, \omega) = C_{ik}(\omega)\, F_k(\vec{r}, \omega)$, and the $C_{ik}(\omega)$ are arbitrary functions of frequency that represent the temporal modulation. The $\{F_{ik}(\vec{r}, \omega)\}$ corresponding to different spheroidal modes are orthogonal over the source aperture.

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \int_{R_1} F_{ik}(\vec{r}, \omega)\, F^*_{i'k'}(\vec{r}, \omega) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} C_{ik}(\omega)\, C^*_{i'k'}(\omega) \qquad k' = k$$

$$= 0 \qquad k \neq k.$$

Therefore, if the $C_{i,k}(\omega)$ are properly selected, each spatial mode can be separated into a number of orthogonal modes on the basis of time dependence. The number of useful orthogonal modes is limited, in turn, by the condition that the channel output be orthogonal over the receiving aperture. This requires that

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} C_{ik}(\omega)\, C^*_{i'k'}(\omega)\, \beta_k(\omega) = 0 \quad \text{unless } i' = i,$$

where $\beta_k(\omega)$ is the functional dependence of the eigenvalue of the k^{th} spheroidal mode. It is characteristic of narrow-band operation that $\beta_k(\omega)$ does not vary over the frequency range where $C_{ik}(\omega)$ is nonzero. Therefore, if the temporal channel is limited to a band of W Hz, each spatial mode can serve as the carrier for 2TW orthogonal temporal modes. As W increases, however, the variation of β_k with ω must be considered in counting orthogonal modes. We do know that $\beta_k(\omega)$ is more sensitive to variations in ω for larger k, but it is not feasible to continue further because analytical expressions for β_k are too awkward to manipulate. We proceed along an alternative route to estimate the bandwidth of the k^{th} mode, using results from the theory of radiated antenna modes.

It is a familiar result that the bandwidth of a resonant circuit is inversely proportional to the ratio of stored energy to energy that is radiated by the circuit. Many authors have shown that an antenna presents a reactive load to its source at the input terminals. We can derive the parameters of an equivalent circuit for this load in terms of the radiation pattern.[41] It is claimed that the quality factor, Q, for this equivalent circuit gives a good estimate of the relative bandwidth of the radiation pattern.[42] The relative bandwidth is proportional to Q^{-1}, with the approximation improving as Q increases beyond unity. Let us assume that this assertion is valid.

Since the source is limited in extent by the transmitting aperture, it can be expanded in a Fourier series whose fundamental frequency is 1/L, where L is a dimension of the aperture. It can be shown that spatial frequencies whose period is less than a wavelength do not propagate. Energy at these frequencies goes into "evanescent" waves that die out within a few wavelengths of the aperture. This energy is usually identified with the reactive energy stored in the fields around the aperture. Energy at lower frequencies is identified with dissipation in the "radiation" resistance. Therefore, Q is computed as follows.[43] In one-dimension,

$$Q = \frac{\int_{|u|\geq k} |F(u,\omega)|^2 \frac{du}{2\pi}}{\int_{|u|\leq k} |F(u,\omega)|^2 \frac{du}{2\pi}} \tag{12}$$

$$k = \frac{2\pi}{\lambda},$$

where

$$F(u,\omega) = \int_{-L/2}^{L/2} f(x,\omega)\, e^{-ixu}\, dx,$$

and $f(x,\omega)$ is the field strength of a monochromatic (narrow-band) source in the transmitting aperture; $F(u,\omega)$ is the Fourier spectrum of $f(x,\omega)$.

The computation of Q is simplified by the following manipulation.

$$Q = \frac{\int_{-\infty}^{\infty} |F(u,\omega)|^2\, du}{\int_{-k}^{k} |F(u,\omega)|\, du} - 1.$$

We can assume that $f(x,\omega)$ is of unit energy. Therefore,

$$Q = \frac{1}{\int_{-k}^{k} |F(u,\omega)|^2\, du} - 1. \tag{13}$$

The denominator in (13) is the energy in the bandlimited version of a function defined on the finite interval $|x| \leq L/2$. Now if we set $F(u,\omega) = \phi_i(u,\omega)$ with the i^{th} spheroidal wave function having a parameter $D = 2LW_s/\lambda Z$, then

$$\int_{-k}^{k} |F(u,\omega)|^2\, du \geq \int_{-W_s}^{W_s} |\phi_i(u,\omega)|^2\, du = \beta_i\left(\frac{2LW_s}{\lambda Z}\right). \tag{14}$$

Therefore we can set an upper bound on Q for the i^{th} spheroidal wave function,

$$Q_i \leq \frac{1}{\beta_i\left(\frac{2LW_s}{\lambda Z}\right)} - 1.$$

It follows that we can set a lower bound on the effective bandwidth for the i^{th} mode in terms of the distribution of eigenvalues for the spheroidal wave functions.

$$\frac{W_{eff}}{W_o} \approx Q^{-1} \geq \frac{\beta_i}{1 - \beta_i} \quad \text{for } Q \gg 1, \tag{15}$$

W_o being the maximum modulation bandwidth for a narrow-band carrier at c/λ Hz, which is arbitrarily taken to be $c/2\lambda$ Hz.

This indicates that the relative bandwidth of the higher order modes decreases with β_i as β_i approaches zero with increasing i. If we assume that the relative bandwidth is essentially unity for the lower order modes having $\beta_i \approx 1$, then we get the following intuitively satisfying result. Let W_e be the equivalent bandwidth available from narrow-band temporal modulation of all of the spatial modes. Then

$$W_e \approx W_o \sum_i \beta_i = W_o\left(\frac{2L_x W_s}{\lambda Z}\right).$$

That is, the available bandwidth is W_o times the number of spatial degrees of freedom. Therefore, for narrow-band temporal modulation, the count of total available degrees of freedom is given by the product of the number of spatial and temporal degrees of freedom.

2.5 DISCUSSION OF RESULTS

It proves quite useful to develop the analogy between spatial variables for wave propagation and temporal variables for linear time-invariant systems. Thus position across the input and output antennas is analogous to time at the input and output of a linear time-invariant filter. The source antenna constrains the domain (duration) of input signals, and the receiving antenna limits the range of spatial bandwidths of the source that is received. The net effect simulates a linear filter with a time-limited input and a bandwidth-limited output. Therefore this research can be viewed as applying principles of communication system design to analyze a special class of filter channels. The results are particularly interesting for two reasons.

1. The propagation channel has not yet been carefully studied. Recent developments in optical technology suggest the potential for using the channel more intensively than ever before.

2. The analysis can be carried through to specific conclusions that also help us to understand several other issues in addition to the original communication question.

Our results show that it is possible to separate the filter channel into a set of parallel noninteracting subchannels (under the assumption that any additive noise has Gaussian statistics and a frequency- and angle-independent power spectrum) by proper selection of signaling waveforms. This mathematical analysis also permits us to answer a related

question, How many degrees of freedom are available in the channel? The number of degrees of freedom computed by conventional means is essentially the number of eigenvalues of Eq. 8 that are approximately unity. This implies that all measurements at the receiver are equally affected by spatially white additive noise. In contrast, we shall see that in a communication application relatively weak signal modes are used if the over-all signal-to-noise ratio is high enough. Information is coded redundantly across all subchannels so that even the weak modes contribute to making a reliable estimate of the transmitted message.

Let us look further at Eq. 9 which expresses the solutions for the set of optimum signaling waveforms. The quadratic phase term is the same amount of phase advance that is imparted by a spherical converging lens of focal length Z. This means that for optimum operation the transmitting aperture should be focused on the receiving antenna. Furthermore, the field in the focal plane of a lens is the spatial Fourier transform of the field in the source plane. Therefore, the receiving aperture functions as a bandlimiting, lowpass filter. This makes more plausible the result that the $J'_n(\vec{r}, \omega)$ are prolate spheroidal wave functions that remain orthogonal when bandlimited.

Prolate spheroidal waveforms have a complicated functional dependence on spatial variables. It is reasonable to ask if there is a practical way to generate them for convenient use. Two synthesis techniques are most attractive at this time. The first and most straightforward method is to illuminate a set of transparencies from behind with a plane-wave source. Each transparency would carry the appropriate amplitude variations in density variations. Spheroidal wave functions are real-valued, so the only phase variation required is either 0° or 180°. Each message is to be encoded as a linear weighting of the basic spheroidal wave functions. This operation can be realized by splitting the beam from a laser source so that each of a basic set of transparencies can be illuminated by a plane wave having the proper strength. Alternatively, one can imagine the sequence of messages as being similar to a film strip that is passed in front of the laser source.

The second technique has already been discussed. In principle, we can devise an optical cavity, terminated by spherical mirrors that will have resonant modes whose amplitude distributions are the spheroidal wave functions. To the extent that these modes can be separated and made individually available, the synthesis problem is solved. Progress in the technology of gaseous lasers suggests that this mode separation will be feasible, for at least a few modes.[44]

Let us observe in conclusion that the source is equally well synthesized by a current distribution across an aperture or by diffraction of a plane wave by a transparency in the transmitting aperture. The optical cavity is one means of generating a current distribution in a particular region of space, at the required carrier frequency.

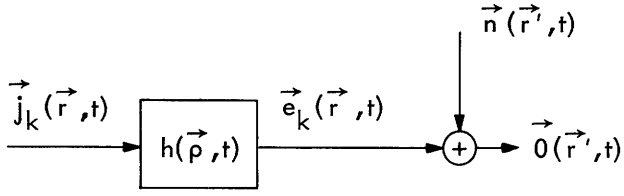
III. COMMUNICATION IN THE PRESENCE OF NOISE

We have ignored the effects of noise in order to emphasize the signal transforming properties of the channel, and the "degrees of freedom" concept. The principal difference between the classical analysis as represented by Gabor and the communication approach that will be presented here is an assumption made about our ability to control the source process that is to be propagated through the channel. In classical analysis the source cannot be modified by the instrument designer. His task is to devise a processor to reproduce the source from a distorted channel output with maximum fidelity. He counts as degrees of freedom the minimum number of independent parameters that are needed to represent the source to within his fidelity criterion. The classical analysis does not consider the effect of additive noise in its estimate of the number of degrees of freedom.

The communicator assumes that he can design the message source to optimize a performance criterion. Therefore an arbitrary source is likely to be transformed (coded) into a new source that is better matched to the channel characteristics. The parameters of the modified source are then used to amplitude-modulate the independent parallel subchannels. The number of subchannels to be used in any instance depends specifically upon the noise environment at the receiving station. For communication purposes, the number of useful degrees of freedom is limited only by noise and by equipment complexity.

Noise sources are the ultimate factor that prevent us from achieving arbitrarily good communication performance. We next apply the noise-free signal representation to analyze the effects of certain noise disturbances. We refer here to additive noise sources, including background noise, quantum noise from some measurements, and thermal shot noise in the receiver circuits. In the analysis here it is assumed that the signal propagates through free space with no turbulence. [We have excluded from our model noise-like effects caused by the random, turbulent structure of the propagation medium. It appears that the effect of turbulence is adequately modeled by a slowly time-variant, spatially dependent random attenuation and phase shift of propagating waves. Communication through atmospheric turbulence is receiving close attention at this time.[45-47]] Our approach to modeling the effects of additive noise in the free-space channel is a consequence of the analysis of section 2.1. In particular, we have shown that the channel performs a linear spatially invariant filtering on the input signals, as indicated in Fig. 5. It follows that the results of conventional noise analysis for linear temporal-filter channels can be extended to include effects of spatial variation.

We shall assume that the additive noise process $\vec{n}(\vec{r}, t)$ is a sample function from a zero-mean vector Gaussian random process. This seems to be a valid approximation for optical background noise (from incoherent sources) and for certain quantum measurement noise. We will justify this assumption eventually, as part of the statistical characterization of the noise.



$$e_k(\vec{r}',t') = \int\int h(\vec{r}'-\vec{r},t'-t)\, \vec{j}_k(\vec{r},t)\, d\vec{r}\, dt$$

$$h(\vec{\rho},t) = \int H(\vec{\rho},\omega)\, e^{-i\omega t}\, d\omega/2\pi$$

$$H(\vec{\rho},\omega) = \frac{\omega}{i4\pi Z}\, e^{\frac{i\omega}{c}\left[Z + \frac{\|\vec{\rho}\|^2}{2Z}\right]}$$

Fig. 5. Linear-filter model for the radiation channel.

First, we shall present an extension of optimum receiver principles to include spatially dependent received signals. Then the noise statistics will be derived and combined with these results to specify the optimum receiver. These results are illustrated by evaluating bounds to the performance obtained with a signaling scheme that uses random coding over a set of spatially modulated waveforms.

3.1 OPTIMUM RECEIVER PRINCIPLES

We wish to determine the structure of the signal processing performed by the receiver that makes a minimum error probability decision on what message has been transmitted. Referring to Fig. 5, assume that the transmitted signal, $\vec{j}_k(\vec{r}, t)$, is one of M waveforms (messages) excited at the transmitting aperture. A vector field $\vec{0}(\vec{r}, t) = \vec{e}_k(\vec{r}, t) + \vec{n}(\vec{r}, t)$ is incident at the receiving aperture when the k^{th} message is transmitted. The receiver is restricted to make its decision on the basis of $\vec{0}(\vec{r}, t)$ over the receiving aperture, during a T-sec detection interval, $|t| \leq T/2$.

We shall assume that the distance between the transmit and receive apertures is sufficient to ensure that the signal component of the received field is essentially a nonuniform plane wave at normal incidence to the receiver. Therefore $\vec{e}_k(\vec{r}, t)$ is a vector in the receiving plane, whose amplitude and orientation (polarization) are known at every point. We shall ignore any components of $\vec{e}_k$ that are normal to the aperture.

It is assumed that the noise $\vec{n}(\vec{r}, t)$ is a sample function from a zero-mean, vector, Gaussian random process. We also assume that at any point on the receiver plane orthogonally polarized components of the noise field are uncorrelated, and hence statistically independent. Therefore, we need consider only a scalar (not vector) problem

if the input signals $\{\vec{j}_k(\vec{r}, t)\}$ are linearly polarized; that is, they lie along the (arbitrary) vector $\vec{e}_s$.

It can be shown that the receiver that makes a minimum-error probability decision chooses the transmitted waveform that is a posteriori most likely.[6] We simplify the analysis by assuming that all inputs are a priori equally likely. The decision rule reduces to: Maximize the "likelihood," L.

L_k = Probability that the channel output $\vec{0}(\vec{r}, t)$ is received, given that the k^{th} message was transmitted.

$$L_k = P[0(\vec{r}, t) \,|\, j_k(\vec{r}, t')]. \tag{16}$$

Decision rule: Choose k if $\max_j L_j = L_k$.

It is necessary to introduce a mathematical artifice into the computation of the L_k in order to correctly model the random-noise process component of the channel output. In particular, let $\phi_k(\vec{r}, t)$ be a set of orthonormal and complete functions over the receiver aperture, then

$$\text{(a)} \quad \int_{-T/2}^{T/2} \int_{R_1} \phi_k(\vec{r}, t)\, \phi_j^*(\vec{r}, t)\, d\vec{r}\, dt = 1 \qquad \text{if } j = k$$

$$= 0 \qquad \text{otherwise.}$$

(b) If there is an $f(\vec{r}, t)$ such that

$$\int_{-T/2}^{T/2} \int_{R_1} \phi_k(\vec{r}, t)\, f^*(\vec{r}, t)\, d\vec{r}\, dt = 0,$$

then

$$\int_{-T/2}^{T/2} \int_{R_1} |f(\vec{r}, t)|^2\, d\vec{r}\, dt = 0.$$

We can expand $0(\vec{r}, t)$ in the series form:

$$0(\vec{r}, t) = \sum_{i=1}^{\infty} 0_i \phi_i(\vec{r}, t) \tag{17}$$

where

$$0_i = \int_{-T/2}^{T/2} \int_{R_1} 0(\vec{r}, t)\, \phi_i^*(\vec{r}, t)\, d\vec{r}\, dt.$$

We have left open the possibility that an infinite number of $\phi_i(\vec{r}, t)$ is used to represent the received signal. Certainly no more than M functions are required for the message components. If, however, the ϕ_i are chosen to represent the messages most efficiently, it may happen that the remaining $\{0_i\}$ cannot be ignored because they are correlated with the noise components of the first M coefficients.

Call the set of 0_i the vector $\underline{0}$. We rewrite the likelihood as

$$L_k = P[\underline{0} \,|\, j_k(\vec{r}, t)]. \tag{18}$$

Now, if the k^{th} message is transmitted, the $\{0_i\}$ are Gaussian random variables, and

$$\langle \underline{0}_i \rangle = E[0_i] = E \int \phi_i^*(\vec{r}, t)[e_k(\vec{r}, t)+n(\vec{r}, t)] \, d\vec{r}\, dt$$

$$= \int \phi_i^*(\vec{r}, t)\, e_k(\vec{r}, t)\, d\vec{r}\, dt. \tag{19}$$

Define the covariances $\langle 0_{ij} \rangle$, where

$$0_{ij} = (\underline{0}_i - \langle \underline{0}_i \rangle)(\underline{0}_j - \langle \underline{0}_j \rangle)^*. \tag{20}$$

Then

$$\langle 0_{ij} \rangle = \iiiint \phi_i^*(\vec{r}, t)\, \phi_j(\vec{r}', t')\, R(\vec{r}, t, \vec{r}', t')\, d\vec{r}\, d\vec{r}'\, dt dt',$$

where

$$R(\vec{r}, t, \vec{r}', t') = E[n(\vec{r}, t) n^*(\vec{r}', t')].$$

Define the covariance matrix R, whose $(ij)^{th}$ element is $\langle 0_{ij} \rangle$. R is seen to be independent of the transmitted signal. If we recall the form of the multivariate Gaussian distribution,[6] we see that

$$L_k = \frac{\exp -1/2(\underline{0} - \langle \underline{0} \rangle)^{*T} \cdot R^{-1} \cdot (\underline{0} - \langle \underline{0} \rangle)}{(2\pi)^{N/2} \, |R|}, \tag{21}$$

where $(\;)^{*T}$ is a vector transposed and conjugated, R^{-1} is a matrix inverse, $|R|$ is a matrix determinant, and we assume an arbitrary dimension, N to the sample size.

The selection of a finite dimensional matrix was made to avoid mathematical difficulties with $|R|$ as N increases. In the rest of this analysis we shall not explicitly treat the mathematical difficulties that arise in "singular detection" situations. In all cases series expansions are assumed to converge.

We can further simplify the decision rule indicated by the likelihood expression in

Eq. 21. Recall that R depends only on the noise and does not vary over the various source hypotheses. Therefore, $\underline{0}^{t*}R^{-1}\underline{0}$, and $|R|$ are not explicit functions of the transmitted signal. All information that depends on the transmitted signal is contained in the exponent. Observe that the likelihood is a monotonically increasing function of its logarithm. Therefore, maximizing the likelihood is equivalent to choosing the largest exponent among the expressions for L_k.

Decision Rule: Choose hypothesis k to attain

$$\max_k \operatorname{Re}\left(\langle \underline{0}^{*T}\rangle R^{-1}\underline{0} - \frac{1}{2}\langle \underline{0}\rangle^{*T}R^{-1}\langle \underline{0}\rangle\right). \tag{22}$$

Equation 22 has been simplified by deleting the term Re $(\underline{0}^{t*}R^{-1}\underline{0})$ which is shared in common by all of the likelihoods. Observe that $\langle \underline{0}^{*T}\rangle R^{-1}\langle \underline{0}\rangle$ does not depend on the channel output. It is a bias term that can be computed in advance for each k. Call it b_k. We can describe the operations performed by this receiver as follows. The receiver computes N complex numbers, each of which is the projection of the field over the aperture during the decision interval onto one of a set of orthonormal functions. For each hypothesis these data are combined in the bilinear form $\langle \underline{0}^{*t}\rangle R^{-1}\underline{0}$, and an appropriate bias is subtracted. The hypothesis selected corresponds to the largest resulting number. Observe that the essential data-gathering operation is the projection of the input onto a set of orthonormal functions. The rest of the processing is simply a manipulation of these projections. To proceed further, we must specify the particular set of expansion functions. We have delayed this choice because it does not affect optimality of the receiver structure.

3.1.1 Optimum Receiver with Uncorrelated Noise Samples

Recall that the elements of R, the $\langle 0_{ij}\rangle$, are quadratic functionals of the kernel $R(t, \vec{r}, t', \vec{r}')$. If R is diagonal, that is, $R_{ij} = \Delta_i \delta_{ij}$, then the decision rule takes a particularly convenient form, and Eq. 22 reduces to

choose the hypothesis k to attain

$$\max_k \operatorname{Re}\sum_i \left(\frac{\langle 0_i^*\rangle 0_i}{\Delta_i} - \frac{1}{2}\,\frac{|\langle 0_i\rangle|^2}{\Delta_i}\right), \tag{23}$$

where $\langle 0_i\rangle$ depends on the hypothesis k. When R is diagonal, the projections of the noise process on the expansion functions $\phi_i(\vec{r}, t)$ are statistically independent random variables. Looking ahead, we find this to be a useful condition for two reasons:

1. The receiver processing is simplified.

2. It suggests that we might design the transmitted signals so that they project onto disjoint sets of expansion functions at the receiver. When this condition obtains we can model the channel by a set of N independent parallel subchannels. This is a particularly

convenient situation for computing system performance, and it will be discussed further.

The requirement that R be diagonal is equivalent to requiring that each $\phi(r, t)$ be a solution to the following integral equation.

$$\int_{R_2} \int_{-T/2}^{T/2} R(t, \vec{r}, t', \vec{r}')\, \phi_i(\vec{r}', t')\, dt' d\vec{r}' = \Delta_i \phi_i(\vec{r}, t), \tag{24}$$

with $|t| \leq T/2$, and $\vec{r}$ in R_2 aperture. We shall assume that $R(t, \vec{r}, t', \vec{r}')$ is such that the solutions to (24), having nonzero eigenvalues, are a complete orthogonal set, and that series expansions using these functions converge appropriately. It can be shown then, that if we use a finite but increasing number of expansion functions, the receiver processing converges to a simple operation:

choose:

$$\max_k \left(\iint 0(t, \vec{r})\, g_k(t, \vec{r})\, dr dt - \frac{1}{2} \iint e_k(t, \vec{r})\, g_k(t, \vec{r})\, d\vec{r}\, dt \right),$$

where g_k is a solution to the integral equation

$$\iint R(t, \vec{r}, t', \vec{r}')\, g_k^*(t', \vec{r}')\, d\vec{r}\, dt = e_k(t, \vec{r}). \tag{25}$$

In other words, for each hypothesis, we project the input onto a single coordinate function and subtract a bias term. It is instructive to see what solutions to (25) look like.

In case $R(t, \vec{r}, t', \vec{r}') = \alpha \delta(t-\tau)\, \delta(\vec{r}-\vec{r}')$, Eq. 25 reduces to

$$\alpha g_k(t, \vec{r}) = e_k^*(t, \vec{r}).$$

Therefore, if the additive noise is spatially and temporally "white," we simply correlate the channel output with the hypothesized signal component of the output. This result is not surprising as a generalization from the analysis of temporal modulation with white noise. Moreover, we expect that this result should apply whenever the noise is "essentially" white. That is, whenever $R(\vec{r}, \vec{r}', t, t')$ is concentrated to $|t-t'| \leq \tau$ and $|\vec{r}-\vec{r}'| \leq \rho$, where τ is much less than the duration in which $e_k(\vec{r}, t)$ changes significantly (for example, the reciprocal bandwidth), and ρ is less than the distance over which $e_k(\vec{r}, t)$ changes. We shall find that these conditions are satisfied by background noise for the optical channel.

This analysis is a straightforward extension of the analysis for spatially independent signals, with the exception that it is more difficult to argue the "whiteness" of the additive noise. It is approximately true that noise enters the channel model by adding statistically independent quantities to each of the "degrees of freedom" that are utilized by

the transmitter. Within this approximation, the receiver processing can be viewed as the spatial analog of "matched-filtering." This is the process in which the channel output is projected onto stored replicas of the noise-free channel output for each message.

If we recall the discussion about the structure of the transmitted signals, and note that the output reproduces the input with a change in spatial scale, we can interpret the operations of the matched filtering as follows. There is a quadratic phase term that corresponds to a lens that focuses the receiving aperture on the transmitting aperture. A spatial mask (one for each message) is placed behind the lens (under the assumption that the signals are narrow-band so that this filter is frequency-independent). The output of this mask is collected (at the focus of a lens) at the input to a bandpass filter that is matched to the temporal frequency characteristics of the message. The decision variable is the sampled output of this filter.

Finally, observe that we have been treating complex quantities that represent the instantaneous envelope of the incident electromagnetic field vectors. The analysis requires a "coherent" sampling of the processor output at the end of the decision interval. We can take advantage, however, of the relatively narrow-band signal modulation to avoid the necessity of making direct measurements of the field vectors. The detection process can be implemented by a heterodyne stage, including optical mixing with a modulated light beam, followed by baseband (or i-f) processing in the information bandwidth.

3.2 RECEIVER PERFORMANCE

Starting from the assumption that the background noise is adequately modeled by a zero-mean Gaussian random process, it follows that we need know only the space-time autocovariance function to completely characterize this noise. We shall derive an expression for the covariance function of background noise, and then combine this result with the receiver principles presented above to evaluate the performance of one class of communication system.

There are three principal sources of additive noise in the communication link. These are background noise, thermal shot noise in the receiver processor circuits, and quantum-measurement noise. Quantum noise arising from optical amplification and heterodyne operations can be treated within the framework developed here. If this was the major noise source, however, we would have been better advised to formulate the optimum receiver principles in terms of quantum theory. The analysis in that instance does seem to lead to receiver structures that are not derivable from classical analysis.[48] Here we ignore the effects of quantum noise; we shall pick up the issue in more detail in section 3.4.

We also neglect thermal shot noise in the receiver circuitry compared with background noise. This approximation is based on the principle that the signal-to-noise ratio of an amplifier is determined primarily by the noise at the input to the first high-gain stage; noise added in subsequent processing has comparatively little effect. For the

optical case, the heterodyne operation, in which the input signal is moved down to baseband for measurement, is equivalent to a high-gain amplification if a strong local oscillator is used.

We focus our attention on background noise-field sources that are statistically independent of the transmitted signal and are collected through the receiver aperture. The sources of background noise at optical frequencies vary radically from day to night, and to some extent also with the center frequency.[49] Direct and scattered sunlight is the principal source during the day. At night these sources include starlight, reflected sunlight (moonlight and planetary reflection), thermal radiation from the earth, night glow, and man-made noise (city lights).

Solar radiation is closely modeled by the statistics of "black-body" radiation, which is a zero-mean Gaussian random process in which the power spectral density is given by

$$Rdf = \frac{4\pi}{\lambda^2} \frac{hfdf}{e^{hf/kT} - 1},$$

where

R = radiated power per unit area, per unit frequency interval

T = effective temperature °K

h = Planck's constant

k = Boltzmann's constant

$\lambda f = c = 3 \times 10^8$ m/sec.

Curves illustrating the variation of R(f, T) are given in Fig. 6. The effective temperature of the Sun is taken to be 6000°K, as illustrated in Fig. 7. The solar spectrum peaks at 0.5 μ in the visible range. Although the radiation decreases sharply at infrared frequencies, Fig. 6 indicates that radiation from direct sunlight remains from one to two orders of magnitude greater than thermal radiation from the earth, even at 10 μ (CO_2 laser) where the latter attains a maximum. Scattered sunlight is a less intense source. If the laser system is shielded from direct sunlight, the principal source of noise at 10 μ is radiation from the earth. At night (no sunlight) the earth temperature is the principal source of noise at 10 μ, but is negligible at 0.5 μ. At the shorter wavelengths, at night, the principal noise sources are man-made noise, starlight, nightglow, and reflected sunlight.

Astronomers have calculated the effective brightness of astronomical bodies and objects. In general, one might model them by a distribution of independently radiating point sources, each having black-body statistics appropriate to its temperature. Similarly, it seems that one ought to define a distribution of extended sources to characterize man-made noise. Measurements confirm our intuitive expectation that the background sky noise is highly variable, depending on terrain, cloud conditions, and city features in the field of view of the receiving aperture. In view of the potential richness

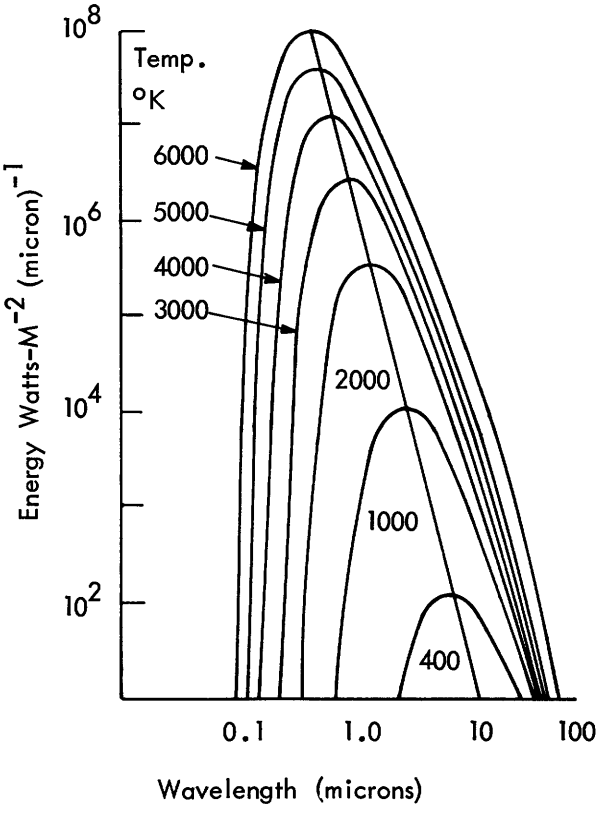


Fig. 6. Energy spectrum of black-body radiation.

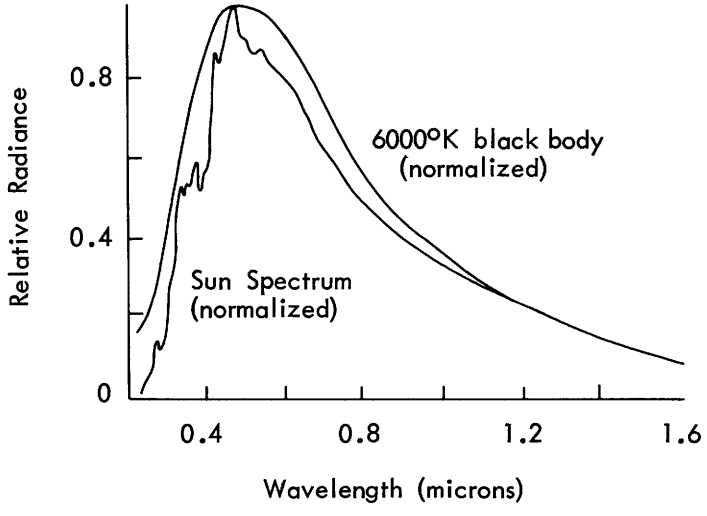


Fig. 7. Energy spectrum of scalar radiation.

of detail needed to characterize the noise, it is well to step back to see just how much detail is indicated in an adequate model.

In communications applications we are satisfied with statistical characterizations of the incident noise processes. We have seen that the natural sources of background noise are conventionally modeled as Gaussian random processes (black-body). Similarly that portion of man-made noise arising from electric light bulbs is also Gaussian in character. Although other sources, such as advertising signs, moving beacons, and stadium floodlights may not individually behave as diffuse radiators of Gaussian noise, it is tempting to claim that their net distribution is approximately Gaussian. (Apply the Central Limit theorem to a sum of independent zero-mean processes radiating with finite power.) We shall adopt this assertion without proof.

3.2.1 Covariance Function for Gaussian Noise Sources

Let $\vec{n}(\vec{r}, t)$ be the vector noise field incident on the receiver aperture. We wish to compute the covariance function $R(t, \vec{r}, t', \vec{r}')$, where

$$R(t, \vec{r}, t', \vec{r}') = E[\vec{n}(\vec{r}, t) \cdot \vec{n}^*(\vec{r}', t')]. \tag{26}$$

We have assumed that $\vec{n}(\vec{r}, t)$ is a sample function from a zero-mean vector valued complex Gaussian random process. We appeal to Huyghen's principle[18] to represent the incident noise sources by the superposition of equivalent sources distributed over a sphere surrounding the receiving antenna. Each point on this sphere of radius Z_o radiates a spherical wave of appropriate amplitude, as illustrated in Fig. 8. Formally, we write the net input noise as $\vec{n}(\vec{r}, t)$ where

$$\vec{n}(\vec{r}, t) = \int_{\text{hemisphere}} d\,\vec{N}(\vec{\beta}, t^o), \tag{27}$$

where $\vec{\beta}$ is a unit vector from the origin in the direction of (Z_o, θ, ψ), t^o is the time a wave had to leave (Z_o, θ, ψ) to reach $\vec{r}$ at time t, and $d\vec{N}(\vec{\beta}, t^o)$ is a Stieltjes differential. Choose Z_o to be large enough that the following approximation is valid

$$t^o = t - \frac{Z_o}{c} + \frac{\vec{\beta} \cdot \vec{r}}{c}.$$

Also, take Z_o large enough that the spherical wave that propagates in the direction $\vec{\beta}$ is essentially a plane wave at the receiving aperture with its E vector randomly polarized in the plane of incidence.

$$\vec{n}(\vec{r}, t) = \int \frac{e^{i2\pi f t^o}}{Z_o} d\,\vec{Z}(f, \vec{\beta}),$$

Z_o^{-1} = attenuation of the spherical wave amplitude with distance,

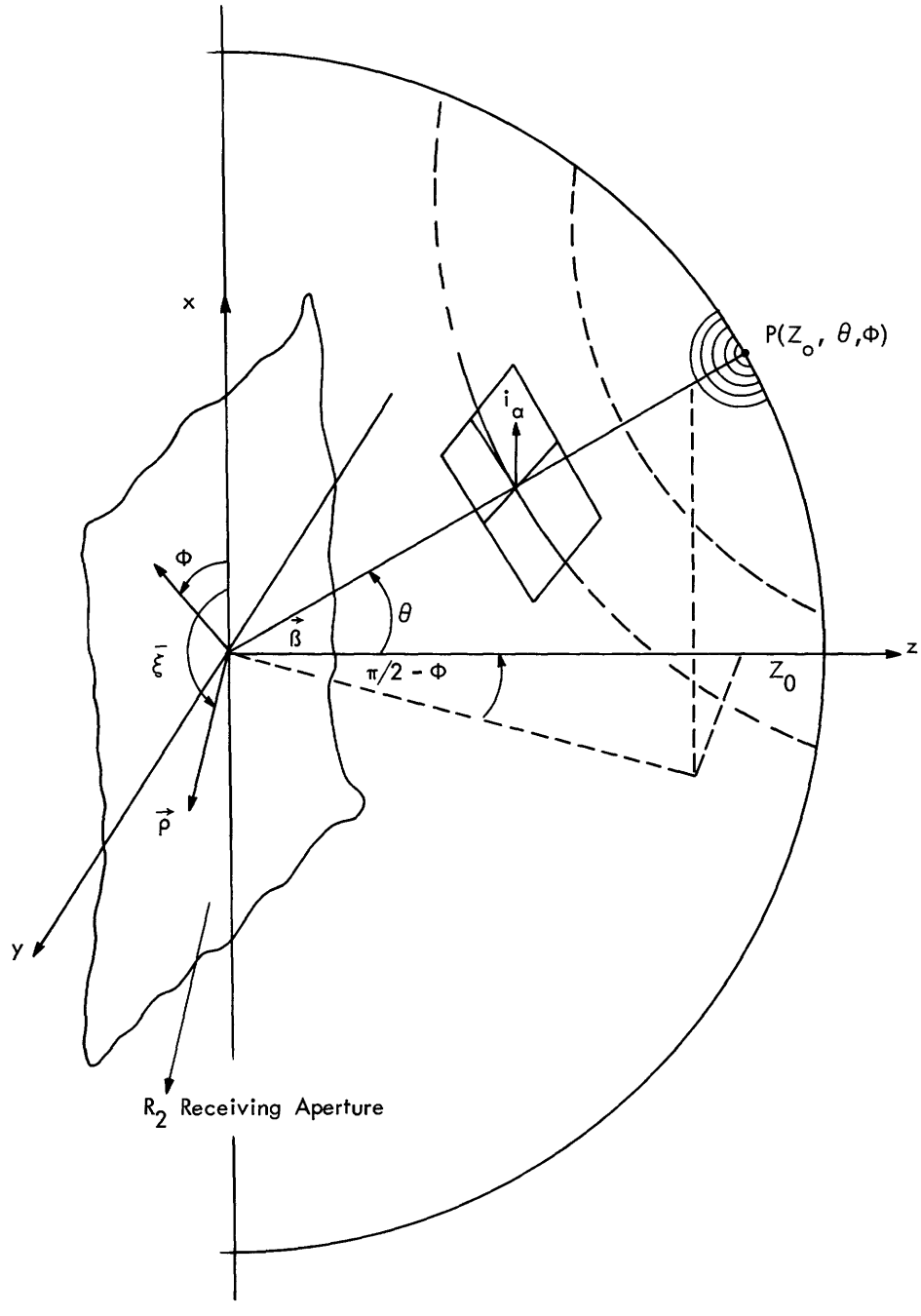


Fig. 8. Equivalent source distribution for background noise, using Huyghen's principle.

$d\vec{Z}(f, \vec{\beta}) = \vec{i}_{\alpha}(\vec{\beta})\, dZ(f, \vec{\beta})$, and

$\vec{i}_{\alpha}(\vec{\beta})$ is a unit vector in the direction of the electric field of the noise component incident from the direction $\vec{\beta}$.

The covariance function is evaluated by the following integral:

$$R_n(t, \vec{r}, t', \vec{r}') = \iint \frac{e^{i2\pi(ft^o - f't^{o\prime})}}{Z_o^2} E[d\vec{Z}(f, \vec{\beta}) \cdot (d\vec{Z}^*(f', \vec{\beta}'))]$$

$$= \iint \frac{e^{i2\pi(ft^o - f't^{o\prime})}}{Z_o^2} E\left[\vec{i}_{\alpha}(\vec{\beta}) \cdot \vec{i}_{\alpha'}(\vec{\beta}')\, dZ(f, \vec{\beta}) dZ^*(f', \vec{\beta}')\right]. \tag{28}$$

Now to work out some geometry. If $\vec{\rho}$ is a vector in R_2 of magnitude ρ at an angle ξ with respect to the x-axis,

$$\begin{aligned} &\vec{\rho} \cdot \vec{\beta} = \rho \sin\theta \cos(\xi - \psi) \qquad \text{for } \vec{\beta}(Z_o, \theta, \psi) \\ &\vec{\rho} \cdot \vec{i}_{\alpha}(\vec{\beta}) = \rho \cos\theta \cos(\xi + \alpha - \psi). \end{aligned} \tag{29}$$

Equation 28 gives an expression for the covariance function between vector functions at any point in space. It is meaningful because we assume that orthogonal noise polarizations are statistically independent. But we wish to modify these results so that the covariance reflects only that part of the noise process that propagates at normal incidence to the receiver aperture. For that reason, we project the noise vector onto the arbitrary vector $\vec{\rho}$ in the receiver plane. The vector can be arbitrarily chosen because the noise sources are assumed to be randomly polarized in the propagation direction.

$$R_n(t, \vec{r}, t', \vec{r}') = \iint e^{i2\pi(ft^o - f't^{o\prime})} E\left[\frac{\vec{\rho}}{\rho Z_o} \cdot \vec{i}_{\alpha}(\vec{\beta}) \left(\frac{\vec{\rho}}{\rho Z_o} \cdot \vec{i}_{\alpha'}(\vec{\beta}')\right) dZ(f, \vec{\beta})\, dZ^*(f', \vec{\beta}')\right]$$

$$= \iint \frac{e^{i2\pi(ft^o - f't^{o\prime})}}{Z_o^2} E[\cos\theta \cos\theta' \cos(\xi + \alpha - \psi) \cos(\xi + \alpha' - \psi')\, dZ(f, \vec{\beta})\, dZ^*(f', \vec{\beta}')]. \tag{30}$$

Now make the following assumptions:

1. The expectation over the random polarization angles α, α' is independent of the other averaging; α and α' are independent and uniformly distributed over an interval length 2π.

2. $dZ(f, \beta)$ is uncorrelated with $dZ^*(f'\beta')$, except for $f' = f$ and $\beta' = \beta$. In other words, there is no correlation between noise components at different points on the source sphere, or at different frequencies.

Accordingly, we set

(a) $E[dZ(f, \vec{\beta})dZ^*(f', \vec{\beta}')] = S(f, \vec{\beta})\ \delta(f-f')\ \delta(\beta-\beta')\ d\Omega d\Omega'\ dfdf'$,

where $d\Omega d\Omega'$ are differential spherical angles, and $S(f, \vec{\beta})$ is the noise power spectral density at each angle β.

(b) $E[\cos^2(\xi+\alpha-\psi)] = \frac{1}{2}$,

where the average is over the random polarization angle α.

Substituting these expressions in Eq. 31, and performing the integration over Ω', f', and ψ, we obtain

$$R_n(t, \vec{r}, t', \vec{r}') = \frac{1}{2} Z_o^2 \iint \exp\left[i2\pi f\left(t-t'+\frac{\vec{\beta}}{c}\cdot(\vec{r}-\vec{r}')\right)\right] S(f, \vec{\beta}) \cos^2\theta\ drd\Omega$$

$$d\Omega = Z_o^2 \sin\theta\, d\theta d\psi. \tag{31}$$

We must assume some functional form for the dependence of $S(f, \vec{\beta})$ on the direction $\vec{\beta}$. In this analysis we make the simplest choice, that is, set $S(f, \vec{\beta}) = S(f)$, independent of $\vec{\beta}$. By using Eq. 29, Eq. 31 reduces to

$R_n(t, \vec{r}, t', \vec{r}') =$

$$\frac{1}{2} \int df\, e^{i2\pi f(t-t')} S(f) \int_0^{\pi/2} \cos^2\theta \sin\theta \int_0^{2\pi} \exp\left[\frac{i2\pi f}{c}(\vec{r}-\vec{r}')\sin\theta\cos(\xi-\psi)\right] d\psi d\theta$$

$$= \pi \int df\, e^{i2\pi f(t-t')} S(f) \int_0^{\pi/2} \cos^2\theta \sin\theta\, J_o\left(\left(\frac{2\pi f}{c}\right)|\vec{r}-\vec{r}'|\right) d\theta. \tag{32}$$

Now observe the following identities

$$\int_0^{\pi/2} J_n(z\sin\theta) \sin\theta^{u+1} \cos^{2\nu+1}\theta\, d\theta = \frac{2^\nu \Gamma(\nu+1)}{z^{\nu+1}} J_{u+\nu+1}(z).$$

Substituting in Eq. 32 $\mu = 0$, $\nu = 1/2$ yields

$$R_n(t, \vec{r}, t', \vec{r}') = \pi \int S(f)\, e^{j2\pi(t-t')} \sqrt{\frac{\pi}{2}} \frac{J_{3/2}\left(\frac{2\pi f}{c}|\vec{r}-\vec{r}'|\right)}{\left(\frac{2\pi f}{c}|\vec{r}-\vec{r}'|\right)^{3/2}} df.$$

Finally, the bracketed term can be rewritten in terms of elementary functions, so that $R_n(t, \vec{r}, t', \vec{r}')$ becomes

$$R_n(t-t', \vec{r}-\vec{r}') = R(\tau, d) = \pi \int_0^\infty df\, S(f) e^{j2\pi f(t-t')} \times \left\{\frac{\sin\frac{2\pi f}{c}|\vec{r}-\vec{r}'|}{\frac{2\pi f}{c}|\vec{r}-\vec{r}'|^3} - \frac{\cos\frac{2\pi f}{c}|\vec{r}-\vec{r}'|}{\frac{2\pi f}{c}|\vec{r}-\vec{r}'|^2}\right\}. \tag{33}$$

Equation 33 can be integrated for certain interesting noise spectra, S(f). If $S(f) = N_o/4\pi$, $0 \leq f \leq \infty$, then

$$R(\tau, d) = (N_o/(d/c))\left[1 - \left(\frac{\tau}{d/c}\right)^2\right] \qquad |\tau| < d/c$$

$$= 0 \qquad \text{otherwise.}$$

This illustrates the point that $R(\tau, 0) \to \infty$ if we assume that the distributed noise sources are temporally white. This singularity is removed if we assume that the input noise is bandlimited. Let $S(f) = N_o/4\pi$ for $0 \leq f \leq f_o$, and be zero elsewhere. Then

$$R(\tau, d) = \frac{f_o N_o}{x}\left[\frac{\cos x \cos 2\pi f_o \tau}{x} - \frac{\sin x \cos 2\pi f_o \tau}{x^2} + \frac{\tau c \sin x \sin 2\pi f_o \tau}{dx}\right.$$

$$\left. + \frac{1}{2}\left(1 - \left[\frac{\tau c}{d}\right]^2\right)\left(S_i x\left(1 + \frac{\tau c}{d}\right) + S_i x\left(1 - \frac{\tau c}{d}\right)\right)\right] \qquad \text{for } |\tau| \leq d/c, \tag{34}$$

where $x = 2\pi f_o d/c = 2\pi f_o d/f\lambda$, and

$$S_i(y) = \int_0^y \sin x/x \, dx.$$

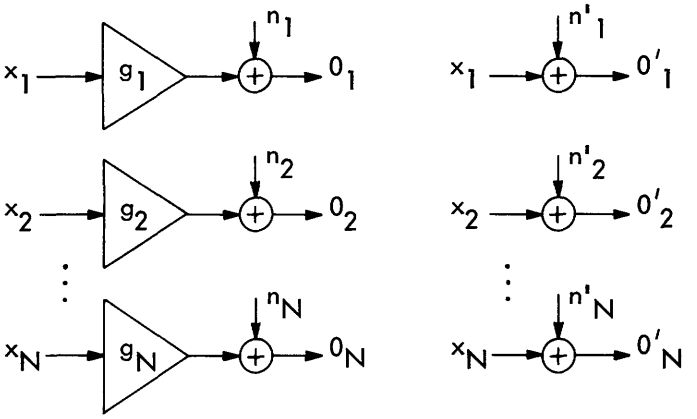
Childers and Reed[50] have studied the function $R(\tau, d)$ that is defined by (34). They point out that $R(\tau, 0)$ remains finite for all τ, and $R(0, d)$ decreases monotonically, with d going ultimately as $1/d$. Therefore, for broadband optical noise, with f_o/f of the order of unity, $R(0, d)$ decreases rapidly as d gets large compared with the wavelength.

This is the justification for modeling the noise as white, in comparison to the signal process.

3.2.2 Random Coding Bounds for Parallel Channels

Recall that we can obtain a parallel subchannel representation of the communication channel when $R_n(t, \vec{r}, t', \vec{r}')$ is of the form given by (34). It is natural to ask just how one may design the channel inputs to take advantage of this property.

We have shown that there exists a set of waveforms $j_N(\vec{r}, t)$ at the transmitter with the following properties: (a) each $j_N(\vec{r}, t)$ is transformed by the channel into one of an orthogonal set of functions $e_N(\vec{r}, t)$ on the receiver aperture, and (b) the projections of the additive noise field on the $e_N(\vec{r}, t)$ are statistically independent random variables. These properties suggest that one transmit information through the channel by amplitude modulation of the $j_N(\vec{r}, t)$. This leads to the schematic channel representation shown in Fig. 9a, in which x_i is the modulating amplitude for $j_i(\vec{r}, t)$, g_i is equivalent to a channel gain, in that $g_i^2/2$ is the energy in the signal component of the i^{th} channel output when $j_i(\vec{r}, t)$ is the input, and n_i is the additive noise projection on the i^{th} receiver coordinate. We revised the model to the equivalent form of Fig. 9b in which the channel gains are



$0_i = x_i g_i + n_i$ $\qquad$ $0_i' = x_i + n'_i$

$\langle |n_i|^2 \rangle = N_i$ $\qquad$ $\langle |n'_1|^2 \rangle = N_i/g_i$

$\langle n_i \rangle = 0$ $\qquad$ $\langle n_i \rangle = 0$

(a) $\qquad$ (b)

Fig. 9. Model of parallel channels.

included in the noise variances.

Each one of M sources messages is to be represented by an N-tuple of amplitudes (x_{ij}, x_{Nj}) that modulate their respective subchannels. We can answer two questions: If one has only a limited amount of signal energy, how should it be divided among the N subchannels?, and What is the performance with this optimum division?

It can be shown that for best performance one should choose a set of M code words in the following manner[9]: For each message m_j choose the amplitudes x_{ij} from independent zero-mean Gaussian distributions with variances σ_i^2. The appropriate values of σ_i^2, and the resulting performance are given in parametric form below.

$$P_e \leq B\, e^{-E[\rho, N_b, S, R]} \tag{35a}$$

$$\sigma_i^2 + N_i = N_b \qquad i \text{ in the set } i_b$$
$$\sigma_i^2 = 0 \qquad i \text{ not in } i_b. \tag{35b}$$

Let us explain these results. First, observe that P_e means the probability that one will incorrectly decide which of the M possible messages has been transmitted. Ebert

presents an upper bound to this error probability. This reflects the random coding approach, in that we are assured that there is at least one code set for which the average error probability is less than the upper bound, which is the average error probability over the class of codes.

We can also derive a lower bound to P_e which is exponentially the same as E for high rates R, and close to E for lower rates. Inasmuch as the error probability is dominated by the exponential, these bounds pin down optimum performance tightly. The exponent is a function of several parameters that are defined as follows:

R = the rate at which information is produced by the source. If the source output is one of M equally likely messages, then R = ln M (nats/channel use).

S = energy available for signaling in the T-sec interval.

S = TP, where P is the average power produced by the source.

ρ = a free parameter, in the interval $[0, 1]$. We choose ρ to maximize the exponent E, for each value of R.

N_b = an energy level that is computed as a function of ρ, S and the variances of the noise on each subchannel.

The relations among these parameters are complex: for a fixed value of ρ,

$$S = \sum_{N_i \leq N_b} \frac{(1+\rho)^2 (N_b - N_i)}{1 + \rho - \rho \frac{N_i}{N_b}}, \tag{36a}$$

with

$$N_i = E\left[|n_i|^2 \right].$$

From (36a) we compute N_b for each value of ρ. The set $\{N_i\}$ includes the N_i in order of increasing magnitude up to the largest $N_{i_{max}} \leq N_b$. Having a value for N_b, we compute the rate R, according to

$$R(N_b, \rho) = \frac{1}{2} \sum_{N_i \leq N_b} \ln \frac{N_b}{N_i}. \tag{36b}$$

Having these quantities, we define the exponent by

$$E(\rho, N_b, S, R) = \frac{S}{2(1+\rho) N_b} - \sum_{N_i \leq N_b} \ln \left(1 + \rho - \rho \frac{N_i}{N_b} \right) + \frac{1}{2} \sum_{N_i \leq N_b} \ln \frac{N_b}{N_i} - \rho R. \tag{36c}$$

It is possible to maximize E over ρ. If this optimum occurs for $0 \leq \rho \leq 1$,

then E takes the form

$$E[N_b, \rho] = \frac{S}{2(1+\rho)\, N_b} - \frac{1}{2} \sum_{N_i \leq N_b} \ln \left(1 + \rho - \rho \frac{N_i}{N_b}\right). \tag{36d}$$

If the optimization requires $\rho \geq 1$, we set $\rho = 1$, and

$$E(N_b) = \frac{S}{4N_b} - \frac{1}{2} \sum_{N_i \leq N_b} \ln \left(2 - \frac{N_i}{N_b}\right) + \frac{1}{2} \sum_{N_i \leq N_b} \ln \left(\frac{N_b}{N_i}\right) - R \tag{36e}$$

$$S = 4 \sum_{N_i \leq N_b} \frac{(N_b - N_i)}{2 - N_i/N_b}. \tag{36f}$$

We can represent these results graphically as shown in Fig. 10. The value of E[0] is given by (36e) with R = 0. The exponent (as a function of R) decreases linearly to a

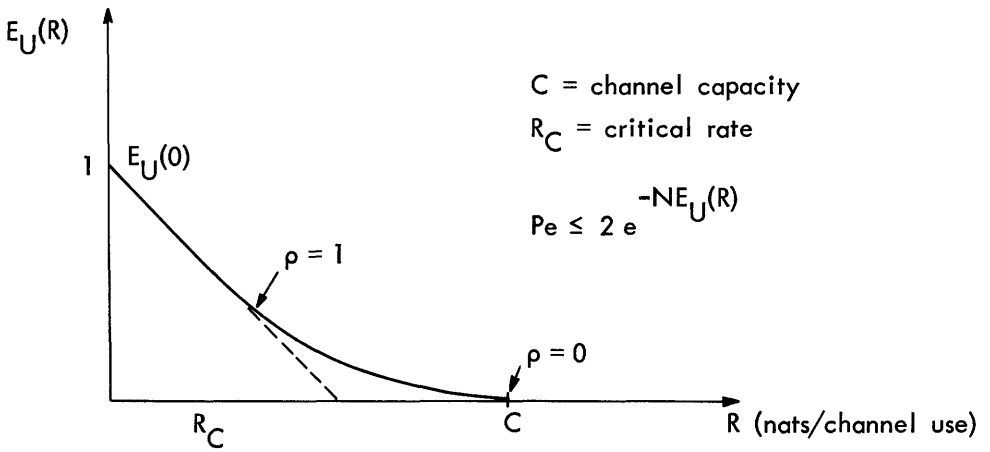


Fig. 10. Exponent $E_u(R)$ for random-coding upper bound.

value R_{CRIT}, where it follows (36c) as $\rho \to 0$. The maximum value of R for which E[R] is positive is the channel capacity, C, obtained by setting $\rho = 0$ in (36a and 36b)

$$S = \sum_{i \leq i_b} (N_b - N_i) \tag{37a}$$

$$C = \frac{1}{2} \sum_{i \le i_b} \ln \left\{ \frac{S + \sum_{i \le i_b} N_i}{i_b N_i} \right\}. \tag{37b}$$

Thus, if all N_i are equal, we have the familiar result

$$C = \frac{D_o}{2} \ln \left(1 + \frac{S}{N D_o} \right),$$

where D_o is the number of available subchannels. We shall continue to apply these results to our model of the optical channel.

3.3 EXAMPLE

We shall work one specific example to illustrate the computation of the bounds presented in section 3.2.

Suppose the transmitting and receiving antennas are rectangular apertures oriented perpendicular to the z axis (see Fig. 11). Let the apertures be separated by a distance Z, and have the dimensions $(2L_x, 2L_y)$ and $(2L'_x, 2L'_y)$ in the (x, y) coordinates,

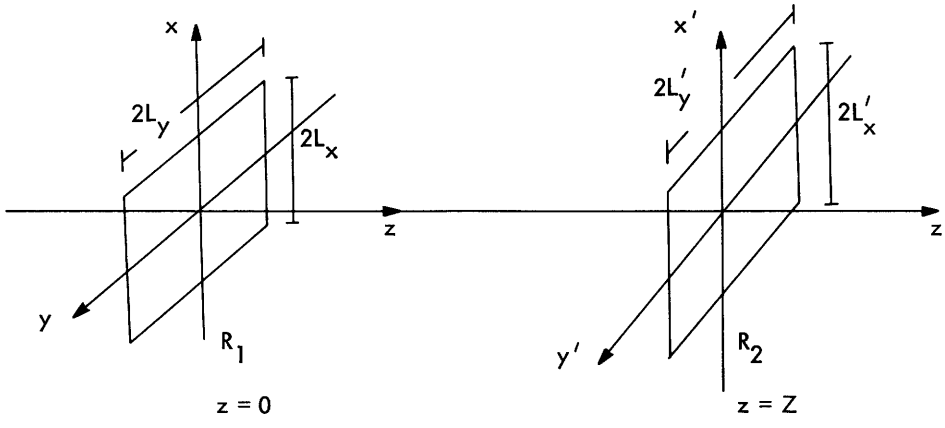


Fig. 11. Rectangular-aperture antenna system.

respectively. Radiation from the transmitters propagates along a "free-space" path to the receiver. We assume that all noise sources are adequately modeled by the "essentially white" process discussed above. It follows that we can select the transmitted signals to provide for parallel use of independent subchannels. These fundamental time-space modulation patterns are eigenfunctions of Eq. 8, as discussed in section 2.2. For rectangular apertures each eigenfunction solution separates into a product of functions that vary along one of the orthogonal directions (x,y). These one-dimensional waveforms are eigenfunctions of the prolate spheroidal equation.

For convenience, we shall evaluate performance for a system with modulation along only one spatial axis, for example, the x-axis. This simplifies the computation of error bounds; however, the computation procedure is directly applicable to the two-dimensional case.

Performance evaluation of the optimum signaling scheme does not require knowledge of the functional dependence of the transmitted messages on time and spatial variables. What one does need is the set of eigenvalues associated with this set of eigenfunctions. Therefore, a significant part of the analysis is to find a convenient analytic representation for the eigenvalues. This is not usually possible for arbitrary integral equations. A result is possible in this case, however, because of Slepian's intensive study of the prolate-spheroidal wave functions.[17]

The eigenvalues of the prolate-spheroidal equation are specified by a single parameter, D, where in our notation

$$D = \frac{4L_x L'_x}{\lambda Z}.$$

We wish to estimate the n^{th} eigenvalue $\beta_n(D)$. For $n \leq D$, $\beta_n(D)$ is approximately unity; for $n > D$, $\beta_n(D)$ approaches zero. As D increases, the transition between these extreme values becomes narrower, as indicated in Fig. 4. Inasmuch as n takes on only integral values, this figure indicates the envelope of spatial eigenvalues. Let us proceed to find an analytic approximation to the envelope function.

Slepian[17] has shown that for D of the order of 10 or greater we can write

$$\beta_n = \frac{1}{1 + e^{\pi b_n}},$$

where b_n is obtained by a graphical process. This result, together with the sharp decrease of β_n in the vicinity of n = D, suggests the following approximation

$$\beta_n \approx \frac{1}{1 + e^{\alpha(n-D)}}, \tag{38}$$

where α is to be determined as a function of D. One property of the β_n is that

$$\sum_{n=0}^{\infty} \beta_n = D. \tag{39}$$

Now substitute Eq. 38 in Eq. 39 and observe that

$$\int_0^{\infty} \frac{dn}{1 + e^{\alpha(n-D+1)}} \leq \sum_{n=0}^{\infty} \beta_n \leq \int_0^{\infty} \frac{dn}{1 + e^{\alpha(n-D)}},$$

which follows by bounding the envelope by a "staircase" function whose value is $1/1 + e^{\alpha(n_i - D)}$, in the interval $n_i \leq n \leq n_i + 1$. Performing the integrations yields

$$D - 1 + \frac{\ln (1 + e^{-\alpha(D-1)})}{\alpha} \leq \sum \beta_n \leq D + \frac{\ln (1 + e^{-\alpha D})}{\alpha}.$$

As D increases, the sum approaches closer to D, as required. Graphical work based on Slepian's analysis, and unpublished notes by Gallager[51] suggest that we take

$$\alpha = \frac{\pi^2}{\ln 2\pi D}.$$

The final approximation is

$$\beta_n \approx \frac{1}{1 + \exp \frac{\pi^2 (n-D)}{\ln 2\pi D}}.$$

We can then proceed to evaluate the rate reliability curve presented in section 3.2.2.

3.3.1 Computation of Rate-Reliability Curve

Recall that the first step in evaluating the rate-reliability curve is to compute i_b, the number of subchannels that will be used at any particular rate. Let

S = available energy (J)

$\frac{N_o}{2}$ = background noise power density (W/Hz/spatial degree of freedom)

$N_i = N_o/2\beta_i$ = effective noise power on the i^{th} subchannel

$D = 4L_x L'_x/\lambda Z$.

From Ebert[9] we obtain i_b as the value of the index i such that

$$S = \sum_{N_i \leq N_b} \frac{(1+\rho)^2 (N_b - N_i)}{1 + \rho - \rho \frac{N_i}{N_b}}.$$

Define

$$\epsilon_b = N_b/N_o \cdot \left(1 + e^{\alpha(i_b - D)}\right).$$

Then

$$\frac{2S}{N_o(1+\rho)^2} = \sum \frac{\epsilon_b - (1+e^{a(i-D)})}{1 + \rho - \frac{\rho}{\epsilon_b}(1+e^{a(i-D)})}. \tag{40}$$

As before, we approximate the sum by an integral to get the following result (see Appendix C).

$$\frac{2S}{N_o(1+\rho)^2} = \frac{1}{\left(1+\rho-\frac{\rho}{\epsilon_b}\right)}\left[i_b(\epsilon_b-1) - \frac{\epsilon_b}{a\rho}\ln\left(1+\rho-\frac{\rho}{\epsilon_b}(1+e^{-aD})\right)\right]. \tag{41}$$

As an example of how one solves this equation consider the limit as $\rho \to 0$. The equation reduces to

$$\frac{2S}{N_o} = i_b(\epsilon_b-1) + \frac{\epsilon_b - 1 - e^{-aD}}{a}.$$

Now e^{-aD} is a small number compared with $\epsilon_b - 1$. Thus

$$\frac{2S}{N_o} = (\epsilon_b-1)\left(i_b - \frac{1}{a}\right). \tag{42}$$

Now $a^{-1} = \ln 2\pi D/\pi^2$, which should be small compared with i_b so we shall ignore it also. It remains to solve for i_b. We recast (42) as follows

$$\frac{2aS\, e^{aD}}{N_o} = a i_b\, e^{a i_b} = t\, e^t, \tag{43}$$

where $t = ai_b$. When x is large, the equation $t\, e^t = x$ has the approximate solution

$$t = \ln x - \ln(\ln x)$$

$$i_b = \frac{1}{a}\left[\ln\frac{2aS}{N_o}e^{aD} - \ln\ln\frac{2aS}{N_o}e^{aD}\right] \tag{44}$$

$$= D + \frac{1}{a}\ln\left(\frac{2aS/N_o}{aD + \ln\frac{2aS}{N_o}}\right).$$

The point $\rho = 0$ is especially interesting because it is associated with the channel capacity. We can use the approximate solution for $\rho = 0$, given by (44) for a wider range of ρ and ϵ_b. We show (Appendix C) that for small ρ, Eq. 41 can be reduced to

$$\frac{2S}{(1+\rho)^2 N_o} = \frac{\epsilon_b - 1}{1 + \rho\left(\frac{\epsilon_b - 1}{\epsilon_b}\right)} \left(i_b - \frac{1}{a}\right). \tag{45}$$

Then

$$0 \leq \rho\left(\frac{\epsilon_b - 1}{\epsilon_b}\right) \leq 1,$$

where ϵ_b is large compared with unity, so Eq. 45 reduces to

$$\frac{2S}{(1+\rho)N_o} = (\epsilon_b - 1)\left(i_b - \frac{1}{a}\right), \tag{46}$$

which is in the same form as (42) with an increased noise power, $N_o' = (1+\rho)N_o$. Similarly, when $\rho(\epsilon_b - 1)/\epsilon_b$ is small, we can use (42) with $N_o' = (1+\rho)^2 N_o$.

Having N_b, we can compute the three numbers that characterize the typical E(R) curve. These are

R(ρ=0) "Capacity"

R(ρ=1) Rate at which E(R) enters the straight-line region

E(0) Exponent at zero-rate.

We find that there are two interesting conditions for computing these numbers according to whether i_b is greater than, or less than, D. We shall defer discussion of this point, in order to make the results available immediately. The detailed computation is included in Appendix C.

Observe that $i_b \geq D$ when

$$2S/N_o \geq D + [1/a] \ln 2aS/N_o.$$

Thus, for a given value of $2S/N_o$, i_b is larger the more slowly the eigenvalues decrease towards zero.

(a) $i_b \geq D$

$$R(0) = \ln\left(\frac{2aS/N_o}{aD + \ln\frac{2aS}{N_o}}\right)\left[D + .5 + \frac{1}{a}\ln\frac{2aS/N_o}{aD + \ln\frac{2aS}{N_o}}\right] \text{ nats/channel use}$$

$$R(1) = \ln\left(\frac{aS/2N_o}{aD + \ln\frac{aS}{2N_o}}\right)\left[\frac{D + .5}{2} + \frac{1}{a}\ln\frac{aS}{\frac{2N_o}{aD + \ln\frac{aS}{2N_o}}}\right]$$

$$E(0) = S/2N_o \frac{1}{1 + \frac{S}{N_o D\left[1 + \frac{1}{aD} \ln \frac{aS}{N_o}\right]}}$$

$$+ \frac{1}{4} \ln \left(\frac{S}{N_o D + \frac{\ln \frac{aS}{N_o}}{aD}} \right) \left[1 + \frac{1}{a} \ln \frac{S}{N_o D \left(1 + \frac{\ln \frac{aS}{N_o}}{aD} \right)} \right] - (i_b + 2) \ln 2.$$

(b) $i_b \leq D$

$$R(0) = \frac{1}{2} \left(D + 1 + \frac{1}{a} \ln \frac{2S}{N_o D \left(1 + \frac{1}{aD} \ln \frac{2Sa}{N_o} \right)} \right) \ln \left(1 + \frac{2S}{N_o D \left(1 + \frac{1}{aD} \ln \frac{2Sa}{N_o} \right)} \right)$$

$$R(1) = \frac{1}{2} \left(D + 1 + \frac{1}{a} \ln \frac{S}{2N_o D \left(1 + \frac{1}{aD} \ln \frac{2Sa}{N_o} \right)} \right) \ln \left(1 + \frac{S}{2N_o D \left(1 + \frac{1}{aD} \ln \frac{2Sa}{N_o} \right)} \right)$$

$$E(0) = \frac{S}{2N_o} \frac{1}{1 + \frac{S}{N_o D + \frac{1}{aD} \ln \frac{aS}{N_o}}}.$$

3.3.2 Interpretation of Results

Our analysis shows that the optical channel can be modeled by a set of independent subchannels, indexed in order of increasing noisiness. Inasmuch as the n^{th} channel is noisier than the $(n-1)^{th}$, it is not surprising that, at any rate R, only a finite number of subchannels is used. In particular, the noise increases rapidly for $n \geq D$, which suggests that it takes increasingly larger increments of available signal energy to make it worthwhile to add an additional channel in that region. In certain laser applications, however, the available energy is large enough to suggest using more than D subchannels. This then bears directly on the issue of defining the concept of "degrees of freedom." Consider the following example: Suppose one is operating a communication link via a CO_2 laser beam over a 2-km path. Day or night, the principal noise source at 10 μ is thermal radiation from the earth at an effective temperature of 283°K. [Note that we are ignoring quantum noise inherent in heterodyne detection or linear amplification. If this noise is included, the signal-to-noise ratio decreases by approximately two orders of magnitude.] Let the following parameters define the system.

$S = 10^2$ W $\times 10^{-6}$ sec $= 10^{-4}$ J

$\lambda = 10\ \mu$

$$T = 283°K$$

$$Z_o = 2 \text{ km}$$

$$2L_x = 35 \text{ cm}$$

$$2L'_x = 35 \text{ cm}$$

$$h = 6.6 \times 10^{-34} \text{ J/sec}$$

$$k = 1.38 \times 10^{-23}$$

$$c = 3 \times 10^{8} \text{ m/sec.}$$

We substitute these values in the following equations to compute $N_o/2$, D, and α.

$$N_o/2 = hf/e^{hf/kT} - 1 = 1.22 \times 10^{-22} \text{ J,}$$

where

$$\lambda f = c$$

$$D = 4L_x L'_x/\lambda Z = 6.1 \approx 6$$

$$\alpha = \pi^2/\ln 2\pi D = 2.71.$$

Observe that $2S/N_o$, the signal-to-noise ratio, is of the order of 10^{18}. This is so high that we expect to be able to use many additional degrees of freedom beyond D = 6. Therefore, we solve for i_b using Eq. 46, which is valid when e_b is large. The solution for i_b is given by modifying Eq. 44.

$$i_b = D + \frac{1}{\alpha} \ln \frac{2\alpha S}{\dfrac{(1+\rho)N_o}{\alpha D + \ln \dfrac{2\alpha S}{N_o}}}.$$

Let us compute i_b for the capacity rate, $\rho = 0$.

$$i_b = D + 14 = 3.3\ D.$$

If we change D by varying any one of the parameters, L_x, L'_x, or Z, holding all other parameters constant, we find that

(a) for D = 10 $\qquad i_b = D + 15 = 2.5\ D$

(b) for D = 64 $\qquad i_b = D + 22.6 = 1.3\ D.$

These computations illustrate the fact that the transition region between large and small eigenvalues becomes relatively narrower as D increases. Thus, as D increases for a fixed signal-to-noise ratio there is a decrease in the number of dimensions used, expressed as a percentage of D.

We wish to emphasize that there is a significant potential for using additional spatial degrees of freedom when D is moderate, say, of the order of 10. The example above illustrates one reasonable set of system parameters for which the technique is promising.

We have discussed a situation in which the signal-to-noise rates are high enough that the use of $i_b > D$ degrees of freedom is indicated. Let us look back to interpret the meaning of the signal-to-noise ratio for which $i_b = D$. Let us look at the signal distribution at channel capacity, $\rho = 0$. From Eq. 44, $i_b = D$ if

$$\frac{\frac{S}{N_o D}}{\frac{2}{}} = 1 + \frac{1}{\alpha D} \ln \frac{\alpha S}{\frac{N_o}{2}}. \tag{47}$$

From (47) we see that for large D we distribute signal over the first D coordinates when S is large enough that the signal-to-noise ratio per dimension is unity. As S increases further, we use more dimensions. For moderate values of D, more energy per dimension is necessary because the noise power begins to increase rapidly in the vicinity of n = D.

We expect that the radiation channel behaves like the bandlimited white Gaussian channel whenever $i_b \leq D$ and D gets large, since the noise power is essentially equal on the first D channels. Refer to the results at the end of section 3.3.1 and observe that for small values of $2S/N_oD$

$$R(0) \longrightarrow S/N_o \text{ nats/channel use}$$

$$R(1) \longrightarrow S/4N_o \text{ nats/channel use}$$

$$E(0) \longrightarrow S/2N_o$$

which confirms our expectations.

When $i_b > D$, the signal-to-noise ratio on the less noisy subchannels will exceed unity. This suggests that there is not enough "bandwidth" to use these channels efficiently; consequently, R(0) should be less than the "infinite-bandwidth" channel capacity which is S/N_o. Refer back to our expression for R(0), and note that

$$R(0) \leq \left(\frac{2\alpha S}{N_o \left(\alpha D + \ln \frac{2\alpha S}{N_o} \right)} \right)^{\frac{1}{2}} \left(\frac{D + .5}{2} + \frac{1}{\alpha} \left[\frac{2\alpha S}{N_o \left(\alpha D + \ln \frac{2\alpha S}{N_o} \right)} \right]^{\frac{1}{2}} \right),$$

DOCUMENT OFFICE 26-327
RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

where we have used the inequality $\ln x \leq x^{1/2}$, for $x \geq 1$.

$$R(0) \leq \frac{S}{N_o}\left[\frac{2 + (D+.5)\ a(N_o(aD+\ln 2aS/N_o)/2aS)^{1/2}}{aD + \ln\frac{2aS}{N_o}}\right]$$

$$\leq S/N_o, \qquad \text{as expected.}$$

Up to this point in our discussion we have assumed that the available signal energy is distributed among a set of subchannels characterized by the spatial dependence along one polarization component. This argument can be modified to account for two factors: (i) the potential to specify the subchannels along the perpendicular polarization, and (ii) the use of temporal degrees of freedom associated with signaling for a T-sec interval with a bandwidth of W Hz.

Each of these factors acts to make available extra degrees of freedom for each original spatial dimension. By analogy with our study of the variation of i_b with D, it follows that when TW is large (a factor of 1000 is not exceptional), there are essentially 2 TW independent temporal subchannels, each having the same noise power. This suggests that the power in each spatial degree of freedom should be equally divided among 2 TW temporal modes. The net effect is to reduce the signal-to-noise ratio per dimension in a manner to more closely approximate the infinte-bandwidth signal energy distribution. This point will be treated in greater detail.

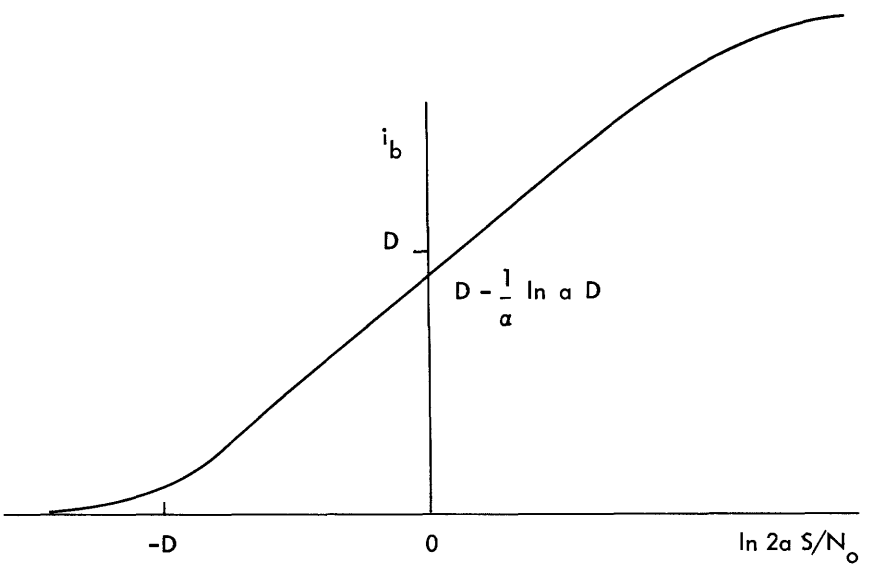


Fig. 12. Variation of i_b with signal-to-noise ratio.

Finally, let us illustrate how i_b varies with the signal-to-noise ratio. We shall use this result in our discussion of visual acuity in Section IV. We have

$$i_b = D + \frac{1}{a}\left[\ln 2aS/N_o - \ln \ln \frac{2aS}{N_o}\right]; \quad \text{for } 2aS/N_o \gg e^{aD}$$

$$= D - \frac{1}{a} \ln aD + \frac{1}{a} \ln \frac{aS}{N_o}; \quad \text{for } 2aS/N_o \ll e^{aD}$$

$$= e^{a2S/N_o} \quad \text{as } S/N_o \to 0.$$

If we plot i_b against $\ln 2aS/N_o$, we obtain the "sigmoid" shape illustrated in Fig. 12.

3.4 DISCUSSION OF RESULTS

Spatial modulation is a means to obtain independent subchannels, where each one can be modulated by orthogonal temporal waveforms. The key question is to determine how a fixed amount of signal power should be distributed in the channel for minimum error probability in signaling. The theory shows that the best result obtains when signal energy is distributed to equalize the net signal plus noise energy in each subchannel used. The channel capacity is reached if this distribution is achieved with a minimum signal power in each subchannel. For these results to hold it does not matter whether the subchannels arise from temporal modulation, spatial modulation, or both. Inasmuch as there will typically be more temporal degrees of freedom than spatial degrees of freedom, conventional practice is to optimize signal distribution by using temporal modulation. If temporal bandwidth is limited, however, the spatial degrees of freedom can be used in effect to multiply the bandwidth.

When is spatial modulation useful? When the temporal channel is bandlimited with the result that one cannot conveniently meet required communication performance with respect to either rate or reliability.

Refer back to the example discussed above. We did not include the effect of distributing the signal power among 2 TW temporal subchannels in the computation of i_b. This is easily handled by replacing S with S/2 TW. For example, if 2 TW = 1000, as in

$$T = 10^{-6} \text{ sec}$$

$$2W = 10^9 \text{ Hz},$$

and we recompute i_b for D = 6, we find $i_b = 6 + 11.4 \approx 3D$. In this instance, the signal-to-noise ratio is so high that the additional bandwidth made available by using spatial subchannels is not enough to attain the "infinite-bandwidth" channel capacity. More "bandwidth" is available by using the orthogonal polarization component. The net result is that the temporal bandwidth can be multiplied by the product of the degrees of freedom in the x and y directions.

In the course of our analysis we have found that "degrees of freedom" is a central

concept. We have observed that the conventional definition concentrates only on representation of signals, and does not consider additive noise processes; the number of degrees of freedom depends only on the geometry of the aperture system and the signal wavelength. In contrast, we have followed a statistical approach that recognizes the effects of noise on signal processing. We have illustrated, by example, that when the signal-to-noise ratio is high enough, we may use from one to ten times the conventionally computed numbers of degrees of freedom to achieve the most reliable transmission of data.

These divergent views on the number of degrees of freedom arise in part from the way degrees of freedom are originally defined. Gabor, for instance, counts the maximum number of physically disjoint spots on a receiving aperture which are images of physically disjoint spots on the transmitting aperture. His notion of disjoint is that there is only an ϵ amount of signal energy in regions where spots overlap. In our analysis we have counted the number of orthogonal patterns at the transmitter that give rise to orthogonal patterns at the receiver. This is immediately appealing because it allows us to use the whole input aperture to shape each spatial mode. Analysis shows how each mode is attenuated by the channel. The utility of a mode depends on our ability to measure its strength in an environment. Therefore, the number of useful modes depends on available signal energy and noise power density.

3.4.1 Quantum Noise

The additive noise in our channel model arises from sources that are independent of the communication process. At optical frequencies, however, we must also consider quantum noise that is apparently inherent in measurements made at the receiver.

The way that quantum noise is produced depends upon the details of the measurement process. Quantum processes produce the effect of additive Gaussian noise at the output of heterodyne detectors, and coherent wave amplifiers.[52] In both instances the power spectrum of this noise is given by hf/2 J/Hz. For a CO_2 laser operating at 10 μ, hf/2 is approximately 10^{-20} J, thereby exceeding thermal background noise by two orders of magnitude. Clearly then, we must make explicit the sense in which our analysis can be interpreted to predict bounds on the performance of communication systems.

Our model is appropriate when the quantum noise is relatively small. If background noise is of thermal origin at 283°K, this is the case for frequencies of the order of 4×10^{12} Hz, or lower. For higher frequencies, where quantum noise is significant, we can make a simple modification of the analysis to predict the performance of one interesting class of receiver processing. We showed that the optimum processing for signals in thermal noise could be implemented by a baseband correlation on the output of a heterodyne detector. The quantum noise in heterodyne detectors has Gaussian statistics.[52] Inasmuch as it is created in the measurement process, it is essentially "white" with respect to spatial variations. The quantum noise also appears white over the narrow bandwidth (not more than 10% of center frequency) that is characteristic

of the applications we have described. Therefore, all results derived for thermal noise apply to quantum noise arising in heterodyne detection, with the modification that N_o is replaced by $N_o + hf$.

There happens to be an additional benefit from recognizing the presence of quantum noise in this system. Observe that spatial modulation becomes more attractive as D, the number of available degrees of freedom, increases. D is directly proportional to the frequency; quantum noise power also rises in direct proportion to frequency. It follows that the analysis will not predict continued improvements in performance as one goes to arbitrarily high carrier frequencies.

In one respect our analysis is limited. We cannot assert that the heterodyne-correlation receiver performs optimum processing of signals corrupted by quantum noise. The specification of optimum receivers, however, may require measurement of variables that have no tangible physical meaning.[48] Among the set of receivers that make physically realizable measurements, receivers that make linear field measurements (heterodyne plus correlation type), or quadratic (energy) measurements, or that combine these operations, have been most thoroughly explored. It has been shown that the field-strength measurement is optimum when the input signal is strong; energy measurement is superior when signal levels are of the order of tens of quanta per event.[53]

Let us look at the distribution of signal energy in the various subchannels for spatially modulated signals. Let the average energy used on the i^{th} channel be σ_i^2, where

$$\sigma_i^2 = N_b - N_i = \frac{N_o}{2}\left[e^{\alpha(i_b - D)} - e^{\alpha(i-D)}\right].$$

We have $\sigma_o^2 \geq \sigma_i^2$ for all i; hence, for most typical values of αD we have

$$\sigma_i^2 \leq \frac{N_o}{2}\left[\frac{2S}{N_o D}\,\frac{1}{1 + \frac{1}{\alpha D}\ln\frac{2\alpha S}{N_o}} - 1\right]. \tag{48}$$

For a fixed value of S and D it follows that

$$\lim_{N_o \to 0} \sigma_i^2 \to \frac{S}{D}\,\frac{1}{1 + \frac{1}{\alpha D}\ln\frac{2\alpha S}{N_o}} \to 0.$$

Similarly, for fixed values of S and N_o

$$\lim_{D \to \infty} \sigma_i^2 = \frac{S}{D} \to 0 \qquad i \leq D.$$

Both of these limits indicate that optimum performance is characterized by distributing an ϵ amount of signal energy in each of a large number of subchannels. In this limit the signal energy per subchannel is small enough that energy detection may be superior to a field-strength measurement. The computations in section 3.3.2 show that with realistic parameter values, there is not enough bandwidth to achieve the limit distribution. Therefore we have been principally concerned with optimizing performance under conditions of strong signal on each mode. In this case the linear receiver is probably the realizable receiver that performs as well as possible in the given signal and noise environment. We cannot claim now that it is optimum. Indeed, if one chooses system parameters to satisfy performance specifications using the minimum signal power per bit of transmitted information, then the optimum receiver may use an energy detector.

The crosscorrelation performed on the channel output is in effect a point-by-point operation performed across the receiver aperture, that is, integrated to produce the decision variable. In some circumstances it may be desirable to implement a sampled version of this operation; that is, to use samples of the input field strength averaged over small disjoint areas of the aperture. The signal energy per sample is reduced by this operation and may become low enough that it is no longer optimum to form a linear combination of the measurements.

To summarize the argument, we claim that heterodyne detection is optimum for carrier signals at frequencies of 10^{12} Hz or less. For higher frequency carriers, in particular, the near infrared and optical heterodyning appears to be the best realizable operation when the signal energy per measurement exceeds tens of quanta. Finally, quantum noise arising from the heterodyne measurement can be treated within our analytical framework by changing the size of one parameter, namely the average noise power density spectrum.

IV. OPTICAL INFORMATION PROCESSING SYSTEMS

The signal representation developed in Section II is well suited for analyzing the application of spatial modulation to digital data transmission. The error bounds discussed in Section III are computed for the parallel-channel model that is derived from this signal expansion. We shall now show how these results can be interpreted to assess the performance of three optical information processing systems, including communication links that use spatial modulation, super-resolving microscopes, and the visual acuity function for humans.

We shall discuss the application to information transmission systems first. Let us summarize the significant characteristics of signals that use spatial modulation.

(a) Each spatial mode is equivalent to a carrier that can be independently modulated over the entire available transmission bandwidth. For present technology the signaling bandwidth is not practically limited by characteristics of the propagation channel.

(b) For moderate values of the signal-to-noise ratio (SNR) the number of useful spatial modes is D, the classical number of degrees of freedom.

(c) For higher values of the SNR additional spatial modes can be used effectively. The larger is D, the smaller is the percentage increase in useful spatial modes. The relation between D and the physical parameters of the antenna system was shown to be

$$D = \frac{A_1 A_2}{(\lambda Z)^2}.$$

For typical values of A_1 and A_2, D exceeds unity only for signaling at optical frequencies and for antenna separation (Z) of the order of a few tens of miles or less.

In order to apply this analysis to the use of spatial modulation at optical frequencies, we assume that the signal transforming property of the channel is adequately modeled by a linear time-invariant filter. A communication system based on coherent detection is feasible only in turbulence-free environments such as deep space, the lunar surface, waveguides (for example, fiber-optics) and enclosed volumes (for example, internal to a computer). This excludes application of signal design as discussed here to communication through the Earth's atmosphere at optical frequencies. Atmospheric turbulence imparts random variations to the phasefront of an optical signal that destroy the assumed spatial coherence over large areas. The significant consequence for our analysis is that orthogonality of transmitted signals is not guaranteed at the output antenna. In the presence of turbulence this orthogonality can be obtained only by producing nonoverlapping beams at the receiver. To the extent that turbulence does not substantially diffuse these beams, the number of orthogonal beams attainable by this alternative technique is given by the number of degrees of freedom as defined by Gabor.[11] This is a sharp upper bound, independent of the prevailing receiver SNR.

Returning now to signaling over nonturbulent channels, we can cast out deep-space communication as a likely area for sophisticated application of this analysis. The

distance between source and receivers is likely to be so large that each one looks like a point source to the other. For shorter ranges one might consider simple low-rate systems, for example, transmission of patterns between an astronaut and his vehicle.

Communication across the lunar surface seems to be a natural application for this technique. We expect that radio propagation beyond the horizon is limited to long surface waves that are inherently narrow-bandwidth carriers. This makes optical line-of-sight propagation an attractive alternative for light-weight low-power signaling at high data rates. In typical operation the number of useful modes will be small compared with the temporal bandwidth. In this circumstance the spatial modes are used most effectively to extend the temporal bandwidth. This may be especially useful if temporal bandwidth is limited by the necessity for providing light-weight, low-power, communication equipment. For example, several television signals might be multiplexed onto optical carriers.

The applications discussed above are characterized by a small value of D compared with the number of temporal degrees of freedom. On the scale of small separations typical of a computer, however, D becomes very large. For example, let R_1 and R_2 be square arrays and

$$A_1 = .25\ m^2$$

$$A_2 = .09\ m^2$$

$$Z = 1\ m$$

$$\lambda = .63 \times 10^{-6}\ m.$$

Then

$$D = 5.5 \times 10^{10}.$$

In this instance, D is substantially larger than the TW product for signaling in current computers, where nsec logic is in the state-of-the-art stage of technology. We do not believe that there is a realistic application of this information-carrying potential to increase the speed of machines for sequential computation because electrical pulses propagate through the computer at nearly the speed of light. Internal communication is slowed principally by stray capacitance that limits the rate at which circuits can be switched. It seems clear that the switching networks for optical pulses would also be affected to the same extent by capacitive delays. There may be, however, a practical application of optical signaling to machines that are organized to do parallel computation. It also appears to be logical to consider using optical methods to create a large, efficient, "read-only" memory suitable for storing tables and other data that change infrequently.

Recall from property (c) that for large values of D there is a relatively small

incentive to take advantage of large SNR to use more than D spatial modes. This suggests that we use the nonoverlapping beam approach to synthesize the storage system. The optical readout would then be scanned over an image plane containing the projection of these beams. In principle, information can also be stored in the intensity of each elementary beam, but there are two reasons for believing that binary ON-OFF modulation would be most suitable.

(a) Binary arithmetic is the fundamental operation in most computers.

(b) D may be large enough to approximate the "infinite-bandwidth" condition. In that case, each independent beam (parallel channel) can be modulated by short binary pulses that are coded to achieve essentially optimum performance, measured by bounds on the digital error probability.

These examples indicate our feeling that there are only limited applications of spatial modulation to the design of information transmission systems. Furthermore, in these most promising applications (lunar surface communication and computer memories) we have not proposed that spatial modulation be used in a manner that is more sophisticated than prior techniques. We believe that the value of this analysis is to set standards for optimum performance against which, as it turns out, conventional systems may perform quite satisfactorily. This should give us more confidence to proceed with those approaches.

We now consider two imaging systems in which the desired output is a faithful reproduction of an arbitrary source image.

4.1 SUPER-RESOLUTION IN OPTICAL MICROSCOPES

The classical statement of the resolving power of a microscope has been given by Abbe.[55] Consider the geometry (after Jenkins and White[55]) illustrated in Fig. 13. Two point sources are imaged by the objective lens into annular ringed diffraction patterns centered on I and I', respectively. The diffraction pattern centered at I' is said to be

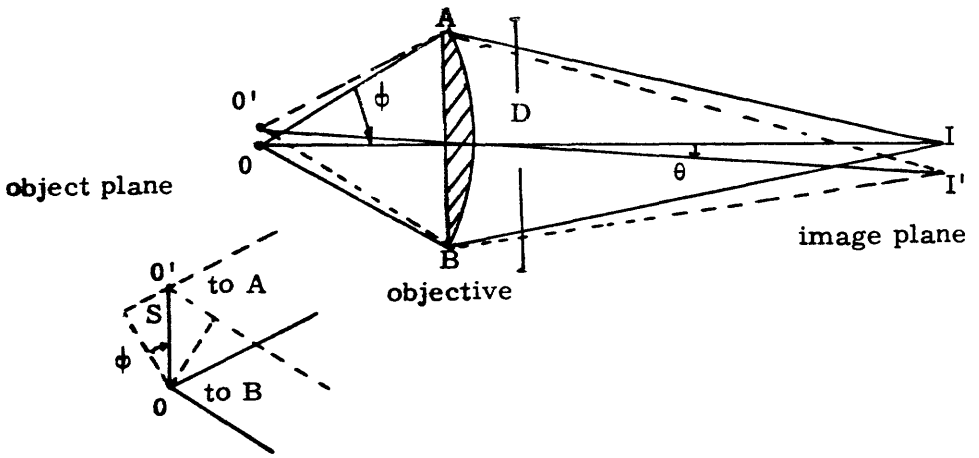


Fig. 13. Resolving power of a microscope.

resolved if its first null is located at I. This null occurs at a distance of 1.22λ for spherical lenses. If the distance between 0 and 0' is S and ϕ is the angle that the objective subtends from the object plane, the path difference between rays 0'AI and 0'BI is approximately $2S \sin \phi$. Now setting $1.22\lambda = 2S \sin \phi$, we get the condition for minimum resolvable points as

$$S \geq \frac{1.22\lambda}{2 \sin \phi}.$$

We presume that 0' and 0 are incoherent self-luminous point sources. Abbe's work showed that for arbitrary illumination, a good estimate is

$$S \geq \frac{\lambda}{2 n \sin \phi},$$

where n is the index of refraction of the medium between the object and the objective lens. A maximum practical value of $n \sin \phi$ is given in Jenkins and White[55] to be 1.6. Therefore, classical analysis suggests that the smallest resolvable distances are of the order of $\lambda/2$; in other words, the Abbe limit states that for incoherent sources, spatial frequency components higher than $2/\lambda$ cannot be resolved. For coherently illuminated sources the frequency limit is $1/\lambda$.

There have recently been several proposals for signal-processing techniques to exceed this limit. All of these are based on using a priori knowledge that the source object has known finite dimensions. The Fourier transform of a finite source is an entire function and, in principle, can be reconstructed from measurements on any finite interval. This property has led to proposals for "supergain" optics based on extraction of these expansion coefficients. The essential task here is the estimation of a parameter from a noisy signal. If the noise is adequately characterized as Gaussian and spatially "white" compared with the signal, then each coefficient can be estimated independently of the others, according to familiar techniques. Having no advance knowledge about the source, the optimum procedure is to correlate the image against a replica of the coordinate in question, averaging as long as possible to minimize the variance of the estimate. This approach has been suggested, but it seems that implicit assumptions have not been recognized. Two points require comment.

1. Typical detectors for optical imaging systems respond only to positive inputs, such as incident energy. Therefore conclusions based on a system that assumes the ability to operate on field strength rather than on power are not directly applicable. These analyses may, however, suggest new approaches in processing optical data.

2. In any specific situation there may be special characteristics of the source object or the additive noise that can be used as a priori information to apply more effective methods of parameter estimation. In particular, the noise added in an optical instrument may not be spatially white compared with the source object.

With these considerations in mind, let us look at the suggestion put forward by

several authors who observe that masking the source with an aperture is equivalent to multiplying the source distribution by a binary (0, 1) transmission function. The light distribution in the back focal plane of the objective lens is the Fourier transform of the spatially modulated source. This, in turn, is the convolution of the transforms of the source and the modulation. The latter transformation is of infinite extent; hence, even at frequencies less than $1/\lambda$, the convolution contains detail about spatial frequencies in the source for frequencies higher than $1/\lambda$. Some of this information, in principle, can be recovered by sophisticated processing of the light at the rear focal plane. The technique is essentially the same as applying the inverse filtering to recover the input to a linear filter. (In this case, this is the Fourier transform of the source.) Our experience with signal processing makes it clear that this is a nonoptimum estimation procedure, and one that is particularly sensitive to distortion by additive measurement noise.

Another approach, which also turns out to be limited by its sensitivity to additive noise, is based on representing the source object by an expansion in prolate spheroidal wave functions. Recall that the prolate spheroidal wave functions are orthogonal and complete on the space of bandlimited functions, and that their extensions to infinite limits are complete on the space of time-limited functions. It follows that a spatially limited object can be reconstructed with arbitrary precision, from expansion coefficients that are computed by projecting its image on the rear focal plane of the objective lens onto the prolate spheroidal modes. This technique requires no pupils or other apparatus that intervenes between the object and its Fourier image. Its particular attraction is simplicity in estimating the expansion coefficients. Its history is brief, being confined to the work of Walther[35] and Frieden.[38] Although Frieden has presented the basic mathematical analysis, it appears that he has not adequately distinguished between intensity (which is always positive-valued) and field strength (which may take on negative values). Inasmuch as the technique is on solid ground analytically, we proceed to evaluate its potential to provide super-resolution.

Let us look along one spatial dimension, say, along the x-coordinate. The number of spatial degrees of freedom is D_x.

$$D_x = L_x \frac{2\,L'_x}{\lambda Z},$$

where L_x and $2\,L'_x$ are the x dimensions of the source and the image, respectively. Now, the first D_x spatial modes have nearly all of their energy concentrated within the frequency passband $|W_s| \leq L'_x/\lambda Z$ that is intercepted by the receiving aperture. Higher order modes have an increasing amount of energy at higher spatial frequencies but a decreasing amount, equal to the eigenvalue β_n, within $|W_s|$.

From their defining equation we know that the spheroidal wave function is its own transform.

$$\int_{-T/2}^{T/2} \psi_n(x)\, e^{-i\omega x}\, dx = i^{-n} \left(\frac{\pi T^{1/2}}{\Omega}\right) \psi_n\left(\frac{T\omega}{2\Omega}\right) \qquad \text{for all } \omega.$$

Also, from Eq. B.6 (Appendix B), we see that $\psi_n(t)$ is approximately a decaying sinusoidal with the frequency $D/2\pi$. This means that each ψ_n has significant amplitude for a wide range of frequencies above the aperture cutoff. Therefore we expect that convergence of spheroidal expansions for spatial functions having substantial energy in high-frequency components is quite slow. Qualitatively, we have the following results.

1. A large number of high-order modes are required to get adequate convergence to the high-frequency (spatial) components of a source object.

2. The higher order modes are attenuated by a factor β_n, where β_n is going to zero exponentially fast with n. Under the assumption of a fairly uniform distribution of energy in the spatial frequencies of the source, the SNR for estimating the n^{th} expansion coefficient goes to zero along with β_n.

It does not seem worthwhile to make a computational analysis of the estimation-reconstruction technique at this time. The general conclusion, however, is inescapable. Any attempt to reconstruct the high-frequency components of the source by a series expansion using spheroidal wave functions is bound to fail because the series does not converge to the source in the presence of additive noise. It is true that the quality of the reconstruction improves as the signal-to-noise ratio increases. We can illustrate the effect of additive noise by the following example.

Suppose the source is represented as

$$S(x) = \sum_{i=0}^{\infty} a_i \psi_i(x) \qquad |x| \leq x_o/2$$

$$= 0 \qquad |x| \geq x_o/2,$$

where the $\{a_i\}$ are expansion coefficients for the source. If the noise on the input aperture is spatially white, the net signal there is given by

$$r(y) = \sum_{i=0}^{\infty} \xi_i \psi_i\left(\frac{x_o y}{2y_o}\right) \qquad |y| \leq y_o$$

where

$$\xi_i = a_i \gamma \sqrt{\beta_i} + n_i.$$

Now the $\{n_i\}$ are independently distributed zero-mean Gaussian random variables with the common variance $N_o/2$, arising from additive noise; the $\{\beta_i\}$ are eigenvalues associated with the spheroidal wave function; and γ is a propagation constant independent of i. The ξ_i are sufficient statistics for estimating the $\{a_i\}$, and we find that the

minimum mean-square estimates[56] are $\{\hat{a}_i\}$, where

$$\hat{a}_i = \frac{\xi_i}{\gamma\sqrt{\beta_i}}.$$

Let

$$\hat{S}(x) = \sum_{i=0}^{N} \hat{a}_i \psi_i(x)$$

be an N-term estimate of the source. Then ξ_i is a Gaussian random variable whose mean value is $\gamma a_i\sqrt{\beta_i}$ and variance is $N_o/2$. Therefore

$$E(\hat{a}_i) = \frac{E[\xi_i]}{\gamma\sqrt{\beta_i}} = a_i$$

$$\text{Var}\,(\hat{a}_i) = \frac{1}{\gamma^2\beta_i}\,\text{Var}\,(\xi_i) = \frac{N_o}{2\gamma^2\beta_i},$$

and, for any value of x,

$$E[\hat{S}(x)] = S(x)$$

and

$$\text{Var}\,[\hat{S}(x)] = \frac{N_o}{2\gamma^2}\sum_{i=0}^{N}\frac{\psi_i^2(x)}{\beta_i}.$$

Because the β_i are exponentially decreasing with i, it follows that the variance becomes unbounded as the number of terms in the expansion is allowed to increase toward infinity. A more complete discussion of the variance in estimates of these expansion coefficients has been given by Rushforth and Harris.[40]

4.2 VISUAL ACUITY

Visual acuity is the capacity to discriminate the fine details of objects in the field of view. Experimental measurements of acuity can be classified into categories based on four essentially different visual tasks, namely resolution, detection, recognition, and localization.[57]

The basic resolution measurement is the minimum distance between objects for the discrimination of separateness. The acuity grating is a typical test object. The grating is a row of alternating dark and light stripes which is the optical analog of a square wave.

Each bar in the pattern could be detected if presented alone, but, when taken together, it becomes difficult to distinguish detail in the pattern because of pupil diffraction. Visual acuity, in the sense of resolution, is the reciprocal of the angular separation between two elements of the test pattern when their images are barely resolved. It is comparable to the "resolving power" of a telescope.

The task of resolution has been regarded as the most critical aspect of acuity since Helmholtz' experiments, in 1866.[58] One can closely correlate experimental measurements with theoretical predictions of the pattern of retinal illuminance computed from diffraction theory and also with the separation of individual cones in the retinal mosaic.

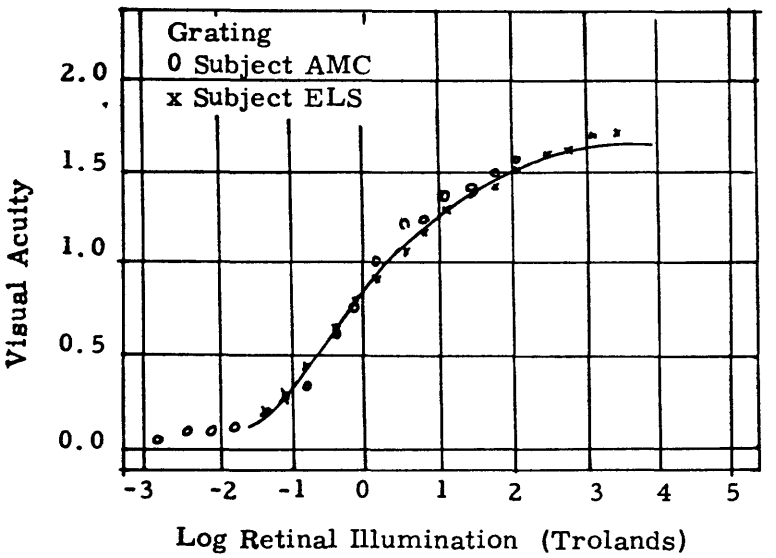


Fig. 14. Variation of acuity with illumination (from Shlaer[59]).

We are familiar with the effect that visual acuity improves monotonically with the intensity of incident illumination. The fundamental experimental data that document this effect are illustrated in Fig. 14, which is taken from a paper by Shlaer.[59] The ordinate of this graph is measured in reciprocal seconds and the abscissa in trolands, which is a measure of power density.

troland = luminance of 1 candle/m^2 on a surface viewed through an artificial pupil of area S = 1 mm^2.

candle = $4\pi/685$ W at $\lambda = .555\ \mu$.

It is rare in the literature on visual measurements for estimates of the background noise illumination to be included. The test subject is usually shielded from stray light, but the test object is not shielded.

In Shlaer's experiment the source is an aperture that subtends a 4° angle at a distance of 1 m from an artificial pupil that is a circular aperture, 2 mm in diameter.

If θ is the angle subtended by the bar pattern and D_o is the number of experimentally measured degrees of freedom in the optical system, then

$$D_o = \frac{4^\circ}{\theta^\circ} = 240 \times \text{visual acuity.}$$

We see that for high illumination, the resolution saturates at a value of 1.7, which corresponds to a minimum angular separation of .58 min. This checks closely with the prediction for angular separation based on the Rayleigh criterion, with diffraction from the artificial pupil being the limiting factor.

$$\theta = \frac{1.22\lambda}{2d} = .585 \text{ min,}$$

where d = 2 mm and $\lambda = .555$ μ, the wavelength for light having maximum luminosity. For smaller values of illuminance, Fig. 14 illustrates the S-shaped "sigmoid" dependence of acuity plotted against a logarithmic scale. The shape of this curve reminds us of Fig. 12, in which we plotted the variation of i_b (the number of spatial modes used in an optimum system) with the logarithm of the signal-to-noise ratio at the receiving antenna.

We do not wish to imply that the eye processes visual inputs in the same fashion as an optimum image processor. Nevertheless, it is valid to compare the performance attained when the eye and a spatial mode processor are connected up to equivalent optical systems. We wish to see how closely measured values of visual acuity approach the best values attained by optimum processing.

It seems natural to make this comparison for the task of resolving an optical grating. The similarity of Figs. 14 and 12 suggests that we compare D_o with the effective number of spatial modes in the optical channel as a function of the log of the signal energy. Recall our definition of spatial degrees of freedom:

$$D = \frac{L_x}{\lambda} \frac{L'_x}{Z}.$$

To compute this number for Shlaer's experiment, observe that L'_x/Z is approximately the angle (rad) subtended by the source; hence,

$$\frac{L'_x}{Z} = 4\pi/180$$

and, therefore,

$$D = \frac{2 \times 10^{-3} \text{ (m)}}{5 \times 10^{-7} \text{ (m)}} \times 4\pi/180 = 285$$

$$a = \frac{\pi^2}{\ln 2\pi D} = 1.3,$$

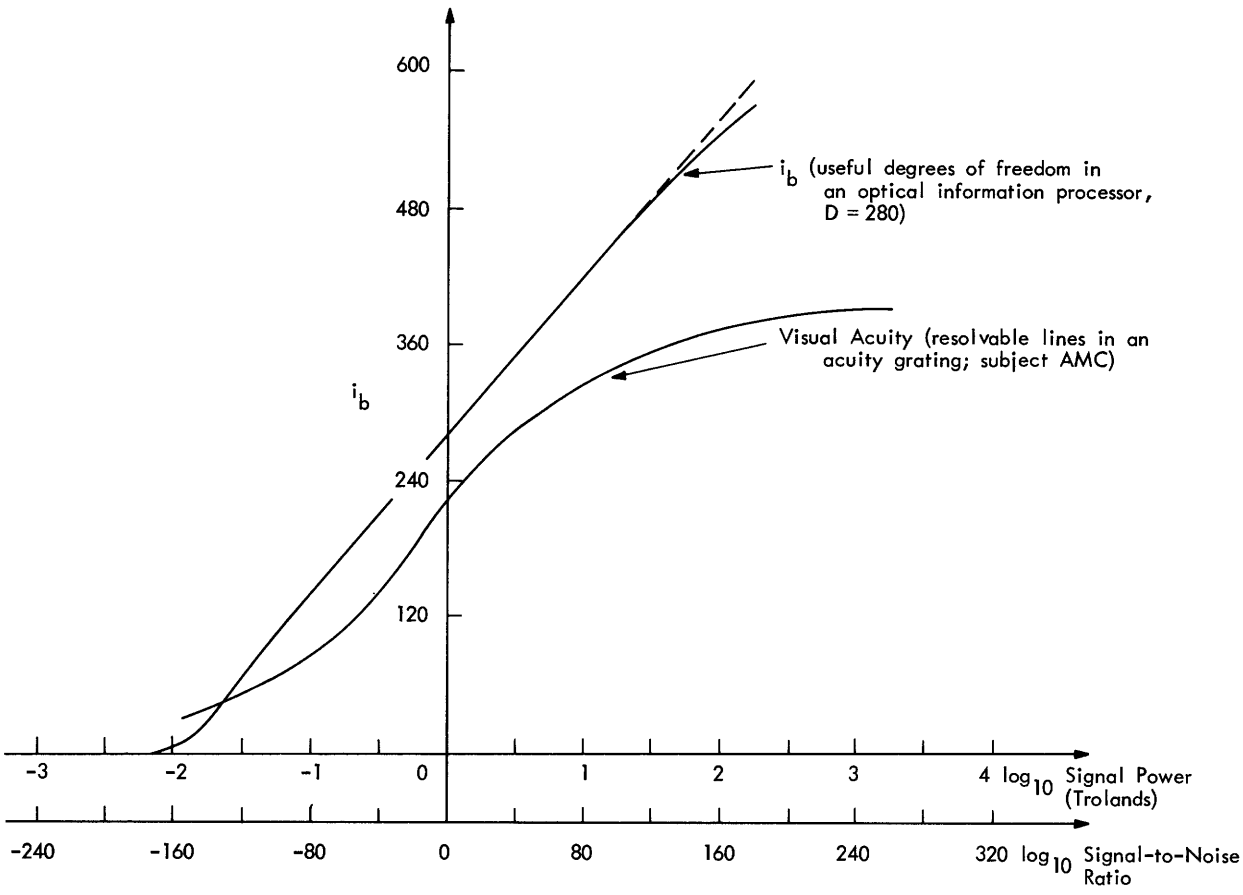


Fig. 15. Comparison between visual acuity and i_b.

where we have chosen a value for the wavelength to approximate maximum illuminance for average brightness conditions (photopic illumination[57]). We have computed the variation of i_b with input SNR for a system having D = 285. This computation is presented in Fig. 15, together with the variation of D_o taken from Fig. 14. The curves cannot be conveniently plotted on a common abscissa because the absolute noise level for the experimental data is not known. We draw the following conclusions from this comparison.

1. The variation of visual acuity with illumination is qualitatively similar to the variation of i_b, however,

2. Visual acuity saturates at a level consistent with the Rayleigh criterion, whereas i_b continues to increase (at a slightly decreasing rate) with increasing SNR, and visual acuity is far more sensitive to variations in SNR than i_b; therefore, the analogy with an optical signal-processing receiver does not bring any new insight into the quantitative prediction of visual acuity.

3. It may be that a closer correspondence can be developed between visual acuity and the optical bandwidth rather than the mode count i_b, which is a simpler estimate to manipulate.

V. CONCLUSION

We shall briefly review the principal issues treated in this research, give our conclusions about the role of spatial modulation in communication systems, and make some suggestions concerning further studies.

5.1 SUMMARY

We have explored digital communication, using spatial modulation of transmitted waveforms, modeling the radiation channel by a combined time-and-space analog of a linear time-invariant filter. We choose from among the orthogonal basis systems for representing waveforms on the input aperture the unique set that leads to an orthogonal basis on the receiving aperture. The additive noise incident on this aperture is shown to be spatially "white," as compared with anticipated signal variations. Therefore, the doubly orthogonal basis is a natural selection for evaluating the performance of subsequent receiver processing.

The signal representation shows how the radiation channel can be modeled by the parallel combination of independent subchannels. We have evaluated an exponential upper bound on the error probability for digital signaling over this set of parallel channels. This bound is interpreted to show how ultimate communication performance is influenced by antenna dimensions, signal frequency, and antenna separation. We have drawn the following conclusions.

5.2 CONCLUSIONS

1. Spatial degrees of freedom have the same function as temporal degrees of freedom, and can be viewed as equivalent to independent channels bearing additional bandwidth.

2. The physical characteristics of the antenna system are condensed into a single parameter, D, which is also the number of spatial degrees of freedom in conventional optical estimation theory. The first D spatial modes are essentially independent of temporal frequency over a 10% bandwidth range. Higher order modes are independent of temporal frequency over a bandwidth that decreases as fast as the modal eigenvalue.

3. The effective number of spatial degrees of freedom is a monotonically increasing function of the signal-to-noise ratio at the receiver antenna.

4. For most ordinary applications, spatial modes should be used only when system performance requirements cannot be met, because of limited temporal bandwidth. For typical system parameters, spatial modulation increases equivalent bandwidth by a factor not greater than ten.

5. To achieve optimum communication performance, the transmitter and receiver apertures are focussed on each other. The transmitted signals are selected from among eigenfunctions of the finite Fourier transform. These signals may be generated by combining the characteristic modes of an optical resonant cavity.

6. Additive noise limits the effective application of spatial eigenfunction modes to super-resolving optical imaging systems. The quality of the reconstruction does improve as the signal-to-noise ratio increases.

7. Spatial modulation may be useful in large computer storage memories. It appears that optimum eigenfunction modulation is not practical.

5.3 SUGGESTIONS FOR FUTURE RESEARCH

We have evaluated the performance of the optimum communication link designed for signaling over the optical channel that includes independent Gaussian background noise. It turns out that a signal-dependent quantum noise produced by the receiver processing may actually be the strongest additive noise source at optical frequencies. We can evaluate error bounds including quantum noise effects for receivers that have high-gain input amplifiers and for heterodyne detectors that use a strong local oscillator. We suggest that the present investigation of communication systems analysis, including quantum measurement noise, be extended to study optimum receivers for spatially modulated signals.

The analysis can also be extended by including atmospheric turbulence in the channel model. Unless the wave-front distortions caused by turbulence can be tracked and estimated precisely, there appears to be no possibility of separating orthogonal signal modes at the receiver except by forming nonoverlapping radiation patterns. It would be advantageous to know how the number of useful spatial modes diminishes because of turbulent scattering of signal energy outside of transmitted beams. John H. Richters' analysis of time- and bandwidth-limited signalling over a dispersive channel may be useful in evaluating receiver performance in this case.[64] The results can be applied to study image degradation caused by turbulence. The principal goal would be to get quantitative estimates for these effects.

We think that it would be interesting and useful to conduct an experimental program on the variation of visual acuity with source illumination. The goal would be to compare visual acuity with the performance level of the analogous function for an optimum image processor. Toward this end, the experiment should include a measurement of background illumination noise. For ease in interpreting results, the acuity task should be directly comparable to electronic circuit functions; for example, one should measure the spatial frequency response of the eye-brain system with a variable frequency sinusoidally modulated acuity grating. Furthermore, the results should be corrected for variation of pupil width with the incident illumination.

APPENDIX A

Electromagnetic Field Solutions

A. 1 ELECTROMAGNETIC FIELD OF A PLANAR CURRENT DISTRIBUTION

We shall treat, first, the electromagnetic field excited by a current sheet across a finite aperture. We assume that the sheet is thin compared with a wavelength, and that the current is directed along the x-axis (see Fig. A-1).

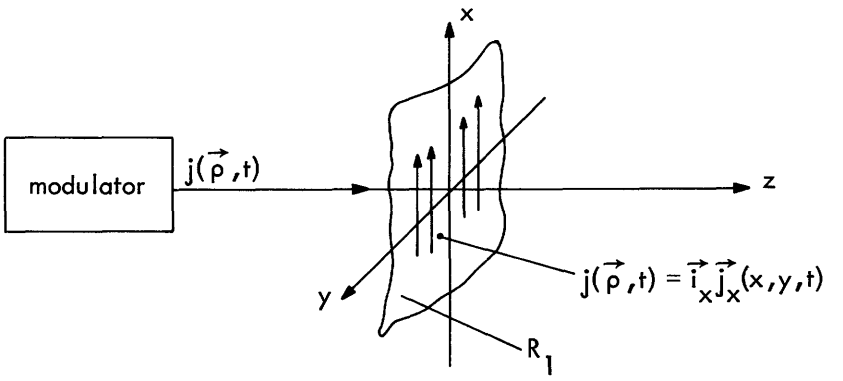


Fig. A-1. Source current distribution across the transmitting antenna.

Let us begin with Maxwell's equations in free space, written in rectangular coordinates. This presentation follows the discussion in Stratton.[12]

$$\nabla \times \vec{E} = -\mu \frac{d\vec{H}}{dt} \qquad \nabla \cdot \vec{H} = 0$$

$$\nabla \times \vec{H} = \epsilon \frac{\partial \vec{E}}{\partial t} + \frac{\partial \vec{D}}{dt} \qquad \nabla \cdot \vec{E} = -\frac{1}{\epsilon} \nabla \cdot \vec{P},$$

where $\vec{E}$, $\vec{H}$, and $\vec{P}$ are vector functions of position and time, and the current and charge densities are expressed in terms of the polarization vector $\vec{P}$, defined by

$$\vec{J} = \frac{\partial \vec{P}}{\partial t} \qquad \rho = -\nabla \cdot \vec{P}.$$

Maxwell's equations have solutions in terms of the vector $\vec{\pi}$,

$$\vec{E} = \nabla\nabla \cdot \vec{\pi} - \mu\epsilon \frac{\partial^2 \pi}{\partial t^2},$$

$$\vec{H} = \epsilon \nabla \times \frac{\partial \vec{\pi}}{\partial t},$$

where $\vec{\pi}$ is a solution of

$$\nabla \times \nabla \times \vec{\pi} - \nabla\nabla \cdot \vec{\pi} + \mu\epsilon \frac{\partial^2 \vec{\pi}}{\partial t^2} = \frac{1}{\epsilon} \vec{P}. \tag{A. 1}$$

Now since the current source $\vec{J}$ is directed along the x direction, Eq. A. 1 reduces to the scalar form

$$\nabla^2 \pi_x - \mu\epsilon \frac{\partial^2 \pi_x}{\partial t^2} = -\frac{1}{\epsilon} P_x. \tag{A. 2}$$

This equation for π_x in terms of its source distribution P_x can be solved by use of Kirchhoff integration,[12] to yield

$$\pi_x(x', y', Z', t) = \frac{1}{4\pi\epsilon} \int_{\text{sources}} \frac{P_x\left(x, y, z, t - \frac{r}{c}\right)}{r} \, dV, \tag{A. 3}$$

where the integration is over the source distribution and

$$r = \left[(x-x')^2 + (y-y')^2 + (z-Z')^2\right]^{1/2}.$$

Equation A. 3 is the basic result from which we can compute all required field solutions. For example,

$$\vec{H} = \epsilon \nabla \times \frac{\partial \vec{\pi}}{\partial t} = \nabla \times \frac{1}{4\pi} \int_{\text{sources}} \frac{1}{r} \frac{\partial}{\partial t} P\left(\vec{u}, t - \frac{r}{c}\right) dV$$

$$= \frac{1}{4\pi} \int_{\text{sources}} \nabla \times \frac{\vec{J}\left(\vec{u}, t - \frac{r}{c}\right) dV}{r},$$

where the operator, ∇, is computed at the point of observation. We can also write the field solutions in terms of steady-state quantities, that is, the Fourier transform on the time variable. We have

$$\vec{\pi}(x, \omega) = \frac{1}{4\pi\epsilon i\omega} \int_{\text{sources}} \frac{\vec{J}(\vec{u}, \omega)\, e^{\frac{+i\omega}{c} r} \, dV}{r}$$

$$\vec{H} = \frac{1}{4\pi} \int_{\text{sources}} \vec{J}(\vec{u}, \omega) \nabla \times \left(\frac{e^{\frac{+i\omega}{c} r}}{r} \right) dV \tag{A. 4a}$$

$$\vec{E} = \frac{1}{4\pi\epsilon i\omega} \int_{\text{sources}} \vec{J}(\vec{u}, \omega)(\nabla\nabla \cdot - \mu\epsilon(i\omega)^2) \frac{e^{\frac{+i\omega}{c} r}}{r} \, dV. \tag{A. 4b}$$

It is convenient to expand the vector operations in spherical coordinates (r, θ, φ) as illustrated in Fig. 3. The result of performing the operations indicated in (A. 4) is

$$\vec{H}(\vec{\rho}, \omega) = \frac{k^2}{4\pi} \sin\theta \vec{i}_\phi \int_{\text{source}} J(\vec{u}, \omega) \, e^{+ikr} \left(\frac{1}{ikr} + \frac{1}{(ikr)^2}\right) dV \tag{A. 5a}$$

$$\vec{E}(\vec{\rho}, \omega) = \frac{k^2}{2\pi}\sqrt{\frac{\mu}{\epsilon}} \left[\vec{i}_r \int_{\text{sources}} J(\vec{u}, \omega) \, e^{-ikr} \times \left(\frac{1}{(ikr)^2} + \frac{1}{(ikr)^3}\right) \cos\theta \, dV \right.$$

$$\left. + \frac{\vec{i}_\theta}{2} \int_{\text{sources}} J(\vec{u}, \omega) \, e^{-ikr} \left(\frac{1}{ikr} + \frac{1}{(ikr)^2} + \frac{1}{(ikr)^3}\right) \sin\theta \right] dV. \tag{A. 5b}$$

A. 2 ELECTROMAGNETIC FIELD DIFFRACTED BY AN APERTURE

The problem is to determine the amplitude distribution of the diffracted field in the space behind an aperture. Let $U(P_o)$ be the amplitude at an arbitrary point behind the aperture, and $U(P_1)$ be the amplitude in the aperture plane. We compute the steady-state amplitude of a monochromatic field

$$U(\vec{r}, t) = R_e[U(\vec{r}) \, e^{-i2\pi ft}],$$

where in free space $U(\vec{r})$ satisfies the scalar wave equation $(\nabla^2 + k^2) \, U(\vec{r}) = 0$. The analysis proceeds from Green's theorem, which states

$$\int_V [G(\vec{r})\nabla^2 U(\vec{r}) - U(\vec{r})\nabla^2 G(\vec{r})] \, dV = \int_S [G(\vec{r})\nabla U(\vec{r}) - U(\vec{r})\nabla G(\vec{r})] \cdot n \, ds,$$

where S is the surface enclosing a volume; n is the outward unit normal at each point on S; $U(\vec{r})$ and $G(\vec{r})$ are any two complex-valued functions of position which together with their first and second partial derivatives are single-valued and continuous within V and on S.[60]

The apparent utility of Green's theorem depends on the proper choice of the function G. First, choose G so that it too satisfies the wave equation

$$(\nabla^2 + k^2) \, G(\vec{r}) = 0.$$

It follows that

$$\int_{V'} [G(\vec{r})\nabla^2 U(\vec{r}) - U(\vec{r})\nabla^2 G(\vec{r})]\, dV = \int_{V'} G(\vec{r})[\nabla^2 U(\vec{r}) + k^2 U(\vec{r})]\, dV \equiv 0$$

independently of the volume V'. Therefore,

$$\int_{S'} [G(\vec{r})\nabla U(\vec{r}) - U(\vec{r})\nabla G(\vec{r})] \cdot n\, dS = 0,$$

where S' is the surface enclosing the volume V'. It can be shown that V' and G can be selected so that the field at any point in the interior of V is given by

$$U(P_o) = \frac{1}{4\pi} \int_{S'} \left(\frac{\partial U}{\partial n} G - \frac{U \partial G}{\partial n} \right) dS. \qquad \text{(A. 6)}$$

Now consider the geometry illustrated in Fig. A-2. Let $S = S_1 + S_2$, and choose G to vanish on S_1.

$$G = \frac{e^{ikr_{01}}}{r_{01}} - \frac{e^{+ik\tilde{r}_{01}}}{\tilde{r}_{01}}.$$

Also, on S_1

$$\frac{\partial G(P_1)}{\partial n} = 2 \cos(\vec{n}, r_{01}) \left[ik - \frac{1}{r_{01}} \right] \frac{e^{ikr_{01}}}{r_{01}}.$$

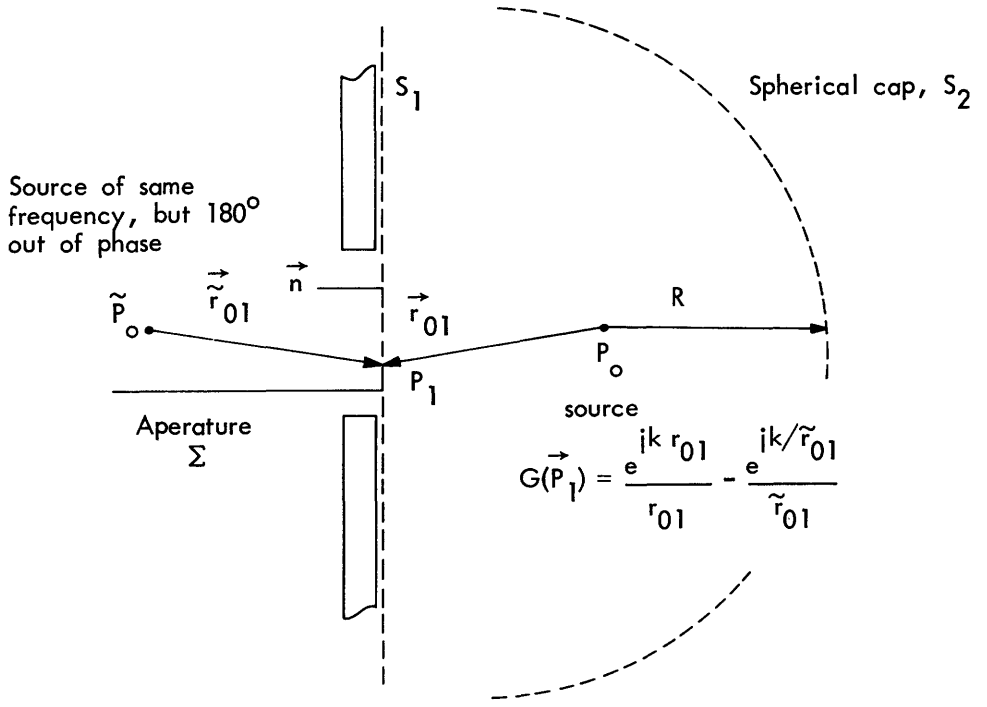


Fig. A-2. Green's function for Rayleigh-Sommerfeld integration of the wave function.

It is possible to show that the contribution of the integral over S_2 vanishes for potentials, U, that vanish as fast as a diverging spherical wave. This condition is satisfied in practice because we can represent the illumination across the aperture by a linear combination of spherical waves. The net result of these arguments is that

$$U(P_o) = \frac{-1}{4\pi} \int_{S_1} U(P_1) \frac{dG}{dn} dS_1.$$

Finally, we invoke the Kirchhoff approximation: $U(P_1) = 0$ in the immediate shadow of the aperture, and is unchanged in the aperture, Σ. Therefore

$$U(P_o) = \frac{k^2}{2\pi} \int_{\substack{\text{aperture} \\ \Sigma}} U(P_1) \left[\frac{1}{ikr_{01}} + \frac{1}{(ikr_{01})^2} \right] e^{ikr_{01}} \cos(\vec{n}_1, \vec{r}_{01}) \, d\Sigma. \qquad (A.7)$$

Let us first compare Eqs. A.5 and A.7. Note that $\cos(\vec{n}, \vec{r}_{01}) = \sin\theta$. Except for the factor 1/2, $U(P_o)$ has the same form as the magnetic field $H(P_o)$. The factor 1/2 arises because $U(P_o)$ is a complex envelope, having twice the amplitude of H_ϕ. Therefore, to the extent that the Kirchhoff approximation is valid, our results will be identical whether we use the picture of elementary sources distributed across the transmitting aperture or the picture of the aperture diffracting an incident source wave.

APPENDIX B

Properties of Prolate Spheroidal Wave Functions

It is known that the scalar wave equation

$$\nabla^2 W + k^2 W = 0 \tag{B. 1}$$

is separable in spheroidal coordinates (ξ, η, ϕ). Prolate spheroidal coordinates are formed by rotating the two-dimensional elliptical coordinate system of confocal ellipses and hyperbolas around the Z-axis, as illustrated in Fig. B-1.[19] The coordinates (ξ, η, ϕ) are related to rectangular coordinates by the transformation

$$x = \frac{d}{2}\left[(1-\eta^2)(\xi^2-1)\right]^{1/2} \cos\phi$$

$$y = \frac{d}{2}\left[(1-\eta^2)(\xi^2+1)\right]^{1/2} \sin\phi$$

$$Z = \frac{d}{2}\eta\,\xi,$$

where d is the interfocal distance, and

$$|\eta| < 1$$

$$1 \leq \xi \leq \infty$$

$$0 \leq \phi \leq 2\pi.$$

The component of W that carries the η dependence is usually called $S_{mn}(C, \eta)$, and satisfies the following differential equation.

$$\frac{d}{d\eta}(1-\eta^2)\frac{d}{d\eta} S_{mn}(C, \eta) + \left[X_{mn} - C^2\eta^2 - \frac{m^2}{1-\eta^2}\right] S_{mn}(C, \eta) = 0, \tag{B. 2}$$

where

$$C = \frac{1}{2}kd.$$

For special values of X_{mn}, Eq. B. 2 has a real continuous solution, $S_{mn}(\eta)$, that is bounded and unique to within an arbitrary constant. We are particularly interested in the functions of zero order, namely $S_{on}(C, \eta)$.

It can be shown that the S_{on} are the eigenfunctions of the finite Fourier transform operator; that is, they satisfy the integral equation

$$a_n S_{on}(t) = \int_{-1}^{1} e^{iC\eta t} S_{on}(\eta)\, d\eta \tag{B. 3}$$

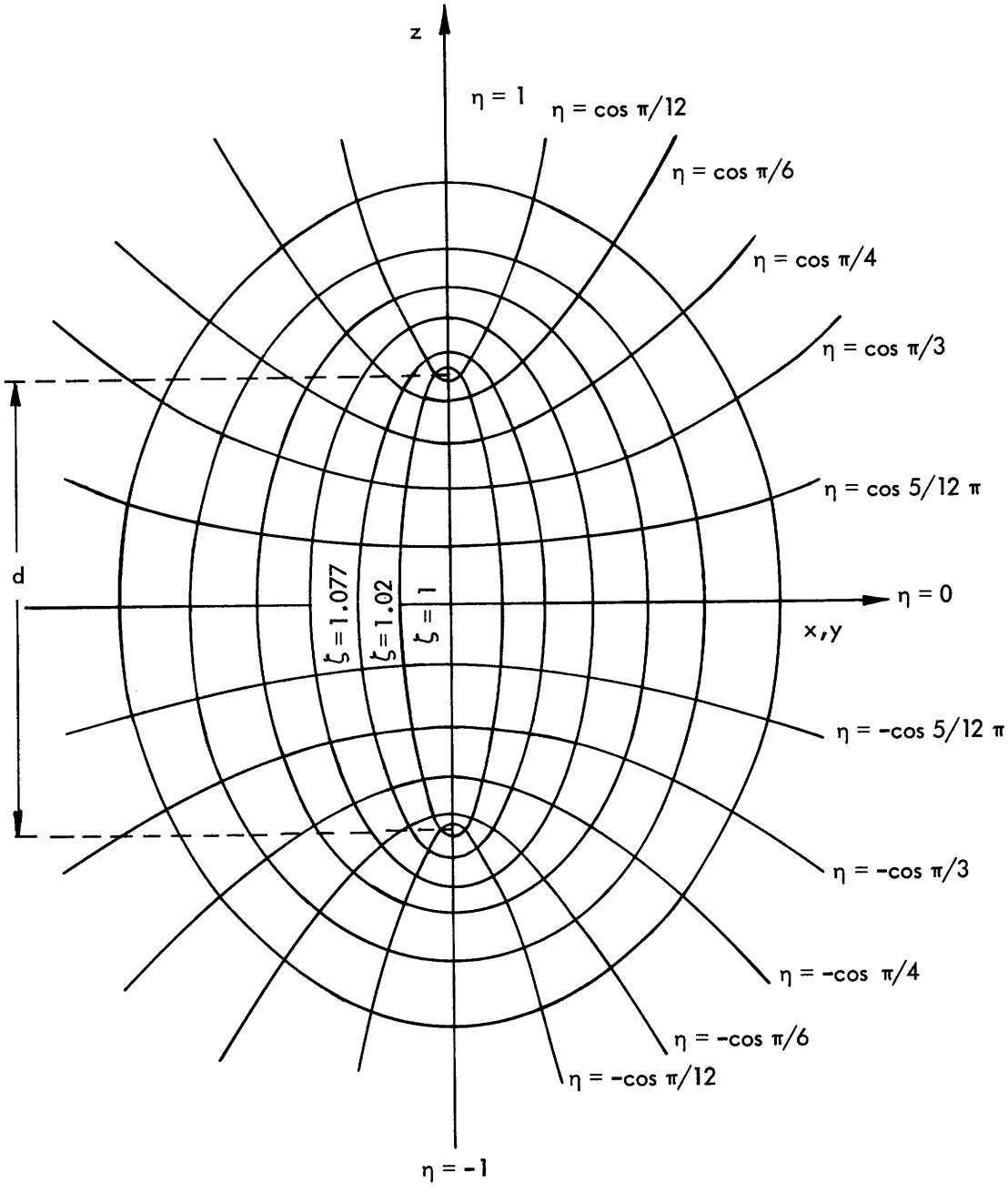


Fig. B-1. Prolate spheroidal coordinate system.

and also its first iterate

$$\beta_n S_{on}(t) = \int_{-1}^{1} \frac{\sin C(t-\eta)}{\pi(t-\eta)} S_{on}(\eta)\, d\eta,$$

with

$$\beta_n = \frac{C}{2\pi} |a_n|^2.$$

We summarize some useful properties of the functions $S_{on}(C, t)$.[17-19]

1. They are continuous in the interval $-1 \leq t \leq 1$, and real for all real t.
2. They may be extended to entire functions of the complex variable t.
3. They are continuous functions of C for $C \geq 0$.
4. They are orthogonal in the interval $-1 \leq t \leq 1$ and complete in the class of square integrable functions on that interval.
5. They are orthogonal in the interval $-\infty \leq t \leq \infty$, and complete over the class of bandlimited functions on that interval.
6. They have exactly n zeros in $-1 \leq t \leq 1$.
7. They reduce uniformly to the Legendre polynomials, $P_n(t)$, in $-1 \leq t \leq 1$ as $C \to 0$.
8. They have been normalized by Flammer,[19] and by Stratton and Chu[18] such that

$$S_{on}(C, 0) = P_n(0) \qquad n \text{ even}$$

$$S'_{on}(C, 0) = P'_n(0) \qquad n \text{ odd.}$$

B. 1 REPRESENTATION OF BANDLIMITED FUNCTIONS

In a series of articles, Landau and Pollak,[29, 43] and Slepian and Pollak[10] have investigated the special properties of the prolate spheroidal functions with special emphasis on the representation of signals. One of their fundamental theorems states[10]:

"Given any $T > 0$ and any $\Omega > 0$, we can find a countably infinite set of real functions $\psi_0(t)$, $\psi_1(t)$, ..., and a set of real positive numbers β_0, β_1, $\beta_2 \ldots$, with the following properties

i. The $\psi_i(t)$ are bandlimited, orthonormal on the real line and complete in B (the class of functions which are integrable in absolute square, and whose Fourier transforms vanish for frequencies higher than $\Omega = 2\pi W$)

$$\int_{-\infty}^{\infty} \psi_i(t)\, \psi_j(t)\, dt = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \qquad i, j = 0, 1, 2, \ldots$$

ii. In the interval $|t| \leq \frac{T}{2}$, the $\psi_i(t)$ are orthogonal and complete in the class of functions that are integrable in absolute square in the interval $|t| \leq \frac{T}{2}$.

$$\int_{-T/2}^{T/2} \psi_i(t)\, \psi_j(t)\, dt = \begin{cases} 0 & i = j \\ 1 & i = j \end{cases} \qquad i, j = 0, 1, 2, \ldots$$

iii. For all values of t, real or complex,

$$\beta_i \psi_i(t) = \int_{-T/2}^{T/2} \frac{\sin \Omega(t-s)}{\pi(t-s)} \psi_i(s)\, ds \qquad i = 0, 1, 2, \ldots"$$

Clearly, the ψ's and the β's are functions of the product $\Omega T/2$, which corresponds to the parameter C in Eq. B. 3. The ψ's are orthonormal on the infinite interval and are complete for representing bandlimited functions having finite energy. That portion of the ψ's on the interval $|t| \leq T/2$ is an orthogonal set that is complete for representing time-limited functions having finite energy. When ψ_i is time-limited, the amount of energy remaining is β_i. If this time-limited function is filtered to pass only frequencies less than Ω, we obtain a set of bandlimited orthogonal functions on the infinite time interval having energy β_i^2. Each successive time-limiting and bandlimiting operation reduces the signal energy by a factor of β_i. The bandlimited function that loses the least portion of its energy upon being time-limited is $\psi_o(t)$, and the time-limited function that loses the least portion of its energy upon being bandlimited is the restriction of $\psi_o(t)$ to $|t| \leq T/2$. Furthermore, $\psi_{n+1}(t)$ loses the least energy by these limiting operations among the class of functions that are orthogonal to the first n ψ's.

We have observed that the spheroidal wave functions are eigenfunctions of the finite Fourier transform. The formal statement of this relation is given by

$$\int_{-\infty}^{\infty} \psi_n(t)\, e^{-j\omega t}\, dt = \frac{j^{-n}}{\sqrt{\beta_n}} \sqrt{\frac{\pi T}{\Omega}}\, \psi_n\left(\frac{T\omega}{2\Omega}\right) \qquad |\omega| < \Omega \qquad \text{(B. 4)}$$

$$= 0 \qquad |\omega| \geq \Omega$$

$$\int_{-T/2}^{T/2} \psi_n(t)\, e^{-j\omega t}\, dt = j^{-n} \sqrt{\frac{\pi T}{\Omega}}\, \psi_n\left(\frac{T\omega}{2\Omega}\right) \qquad \text{for all } \omega. \qquad \text{(B. 5)}$$

B. 2 FUNCTIONAL DEPENDENCE OF THE $\psi_n(t)$

Slepian has published the most complete and detailed expressions for the variation of $\psi_n(t)$ with t for large values of C. These expressions are quite complicated, especially because many different ranges of parameters must be treated separately. We present here his result for $\psi_n(t)$ valid for large C and for large values of n.

$$\psi_n(t) = \begin{cases} f_n^{(1)}(t) & 0 \leq t \leq 1 - C^{-1/2} \\ k_2 f_n^{(2)}(t) & 1 - C^{-1/2} \leq t \leq 1 + C^{-1/2} \\ k_3 f_n^{(3)}(t) & 1 + C^{-1/2} \leq t \end{cases}$$

Here

$$f_n^{(1)}(t) = \mathrm{Re}\left\{\frac{\exp i\left(Ct + \frac{\delta}{2}\ln\frac{1+t}{|1-t|} - n\frac{\pi}{2}\right)}{|1-t^2|^{1/2}}\right\} \times \left\{1 + \sum_{j=1}^{\infty}\frac{1}{C^j}\sum_{k=1}^{\infty}\frac{M_k^j}{(1+t)^k} + \frac{N_k^j}{(1-t)^k}\right\}$$

$$f_n^{(2)}(t) = e^{iC(1-t)}\ \phi(\beta, 1; \xi) + \sum_{j=1}^{\infty}\frac{1}{C^j}V_j(\xi)$$

$$f_n^{(3)}(t) = \frac{1}{C}f_n^{(1)}(t) \quad \text{except that } n+1 \text{ replaces } n \text{ in the complex exponential.}$$

Also, $\phi(a, C;t)$ is the confluent hypergeometric function;

$$\xi = -2iC(1-t), \quad \beta = 1/2(1-i\delta)$$

and

$$V_j(\xi) = \phi(\beta, 1;\xi)\sum_{k=1}^{j}K_k^j\xi^k + \xi\phi(\beta+1,\beta;\xi)\sum_{k=1}^{j}L_k^j\xi^k.$$

The quantities M_k^j, N_k^j, L_k^j, K_k^j are polynomials in δ.

$$k_2 = \frac{CR_o(\delta)\, e^{\delta\pi/4}}{2}\left[1 + 0\left(\frac{1}{C}\right)\right]$$

$$k_3 = C\, e^{\delta\pi/2}\left[1 + 0\left(\frac{1}{C}\right)\right].$$

Finally, δ is a particular solution of

$$C + \delta\log(2\sqrt{C}) - \phi_o(\delta) = \frac{\pi}{2}\left(n + \frac{1}{2}\right) + 2\pi q$$

$$R_o(\delta)\, e^{i\phi_o(\delta)} = \Gamma\left(\frac{1+i\delta}{2}\right).$$

We shall look more closely at solutions for δ in connection with computing asymptotic expressions for the eigenvalues.

Observe that for large values of C, $\psi_n(t)$ is given by $f_n^{(1)}(t)$ [or by a suitable

modification of $f_n^{(1)}(t)$] over all but a vanishingly small region of t in the vicinity of t = 1. This suggests that the lead term in $f_n^{(1)}(t)$ be adopted as a useful approximation for numerical computations.

$$\psi_n(t) \approx \mathrm{Re}\left\{\frac{\exp i\left(Ct+\frac{\delta}{2}\ln\frac{1+t}{|1-t|}-\begin{pmatrix}n\\ \text{or}\\ n+1\end{pmatrix}\frac{\pi}{2}\right)}{\begin{pmatrix}1\\ \text{or}\\ C\end{pmatrix}|1-t^2|^{1/2}}\right\} \qquad t \neq 1.$$

B. 3 EIGENVALUES OF PROLATE SPHEROIDAL WAVE FUNCTIONS

The prolate spheroidal wave functions satisfy the following equation

$$\beta_i \psi_i(t) = \int_{-T/2}^{T/2} \psi_i(s) \sin\frac{2\pi W(t-s)}{\pi(t-s)}\,ds. \tag{B. 6}$$

It can be shown that

$$\beta_i = \int_{-T/2}^{T/2} \psi_i^2(t)\,dt,$$

and

$$\frac{\sin 2\pi W(t-s)}{\pi(t-s)} = \sum_{j=0}^{\infty} \psi_j(s)\,\psi_j(t). \tag{B. 7}$$

Therefore, setting s = t, and integrating Eq. B. 7 over t for $|t| \leq T/2$, we find

$$2TW = \sum_{i=0}^{\infty} \int_{-T/2}^{T/2} \psi_i^2(t)\,dt = \sum_{i=0}^{\infty} \beta_i.$$

Therefore

$$\sum_{i=0}^{\infty} \beta_i = \frac{2}{\pi} C. \tag{B. 8}$$

We can also get an estimate for the sum of β_i^2, as follows:

$$\sum_{i=0}^{\infty} \beta_i^2 = \iint_{T/2}^{T/2} dtds \sin\frac{2\pi W(t-s)}{\pi(t-s)} \sum_{i=0}^{\infty} \psi_i(t)\,\psi_i(s)$$

$$= \iint_{-T/2}^{T/2} \left[\frac{\sin 2\pi W(t-s)}{\pi(t-s)}\right]^2 dtds$$

$$= \frac{4C}{\pi^2}\int_0^{2C} \frac{\sin^2 x}{x^2}dx - \frac{2}{\pi^2}\int_0^{2C} \frac{\sin^2 x}{x}.$$

The following asymptotic expression, valid for large C, is given by Landau and Pollak[29]:

$$\sum_{i=0}^{\infty} \beta_i^2 = \frac{2}{\pi}C - \frac{1}{\pi^2}\log C + 0(1), \tag{B. 9}$$

and also the following lower bound which is true for all C,

$$\sum_{i=0}^{\infty} \beta_i^2 \geq \frac{2C}{\pi} - \frac{1}{\pi^2}\log C - 1.$$

Since $0 \leq \beta_i \leq 1$, it follows that for large C the first $\frac{2}{\pi}$ C of the $\{\beta_i\}$ are approximately 1, and the remainder are approximately zero. This variation of β_i with C is illustrated in Fig. 4.

It is convenient to have analytical expressions for the functional dependence of the β_n on n and C. The following results are taken from Slepian.[17] For small values of C,

$$\beta_n = \frac{2}{\pi}\left[\frac{2^{2n}(n!)^3}{(2n)!\ (2n+1)!}\right]^2 C^{2n+1} \exp\left[\frac{-(2n+1)C^2}{(2n-1)^2\ (2n+3)^2}\right](1+0(C^2)),$$

while for fixed n, and large values of C (large compared with n and unity)

$$1 - \beta_n = \frac{4\pi^{1/2}2^{3n}C^{n+1/2}\ e^{-2C}}{n!}\left[1 - \frac{6n^2 - 2n + 3}{32C} + 0(C^{-2})\right].$$

Of more immediate application to our research is the following approximation, valid when both n and C are large.

$$\beta_n = \frac{1}{1 + e^{\pi\delta}}\left[1+0(C^{-1})\right]$$

$$\approx \frac{1}{1 + e^{\pi\delta}}, \tag{B. 10}$$

where δ is the root of the smallest absolute value of

$$C + \delta \log (2C^{1/2}) - \phi_o(\delta) = \frac{n\pi}{2} + \frac{\pi}{4}. \qquad \text{(B. 11)}$$

Now $\phi_o(\delta)$ is the phase of the complex quantity $\Gamma\left[\frac{1}{2}(1+i\delta)\right]$.

$$\Gamma\left(\frac{1+i\delta}{2}\right) = R_o(\delta)\, e^{i\phi_o(\delta)}.$$

The modulus and phase of the complex gamma function $\Gamma(1+iy)$ has been well tabulated.[61] Let $\theta(1+i\delta)$ be the phase angle of $\Gamma(1+i\delta)$; $\phi_o(\delta)$ can be computed from the phase terms as follows.

$$\phi_o(\delta) = \theta(1+i\delta) - \theta\left(1 + \frac{i\delta}{2}\right) - \delta \ln 2.$$

Figure B-2 is a sketch of $\phi_o(\delta)$. The root of Eq. B. 11 is computed graphically for each n, as the intersection of the straight line, $y = \delta \log 2\sqrt{C} + C - \frac{n\pi}{2} - \frac{\pi}{4}$ with the curve $y = \phi_o(\delta)$. There is always at least one root of this system of equations.

It turns out that there are ancillary conditions, arising from the details of the approximation procedure, which must also be satisfied to get the correct value of δ.[17] For this reason, an extra factor of $2\pi q$ should be included on the right-hand side

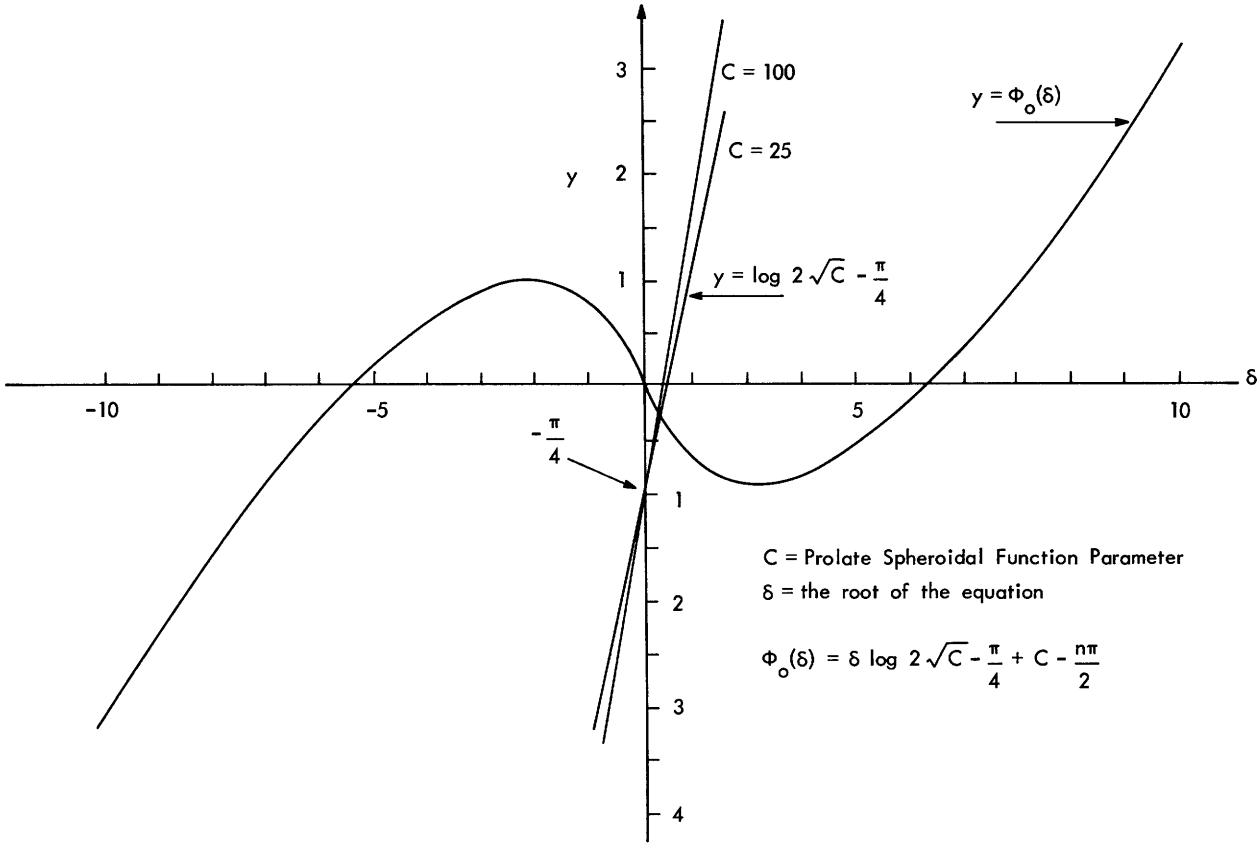


Fig. B-2. Graphical computation of δ.

of (B. 11). (q is an arbitrary integer that is adjusted in the approximation process.) In most interesting cases we can ignore this factor.

For large values of $|\delta|$, $\phi_o(\delta)$ is asymptotically $\frac{\delta}{2} \ln |\delta|/2e$. Analysis of (B. 11) is simplified by this approximation. The result is that we can set $\delta = -\pi\left(n - \frac{2}{\pi} C\right)/\ln 4C$, and

$$\beta_n \approx \frac{1}{1 + e^{-\frac{\pi^2}{\ln 4C}\left(n - \frac{2}{\pi} C\right)}},$$

which is the form used in Section III, with $D = \frac{2}{\pi} C$.

B. 4 GENERALIZATIONS OF THE PROLATE SPHEROIDAL WAVE FUNCTIONS

The prolate spheroidal wave functions are defined along a line in a one-dimensional space. Their extensive utility has prompted analysis to look for analogous functions defined on more general regions of space. Slepian[16] has discussed one formulation of the general problem. That is, let $\vec{x}$ be a vector in a region R of Euclidean space. Let S be a region in the corresponding Fourier transform space, and let $\vec{y}$ be a vector in S. If R is symmetric and S is a scaled version of R, the analog of (B. 3) is

$$\alpha\psi(\vec{x}) = \int_R e^{iC\vec{x}\cdot\vec{y}}\, \psi(\vec{y})\, dy \qquad \text{for } \vec{x} \text{ in R.} \tag{B. 12}$$

Equation B. 12 separates into a product of one-dimensional prolate spheroidal wave equations over rectangular regions. Slepian[16] has also solved the equation for 2-dimensional circular geometry. He shows that the solutions are eigenfunctions of

$$\gamma\phi(x) = \int_0^1 J_N(ixy)(Cxy)^{1/2}\, \phi(y)\, dy,$$

where $J_N(\cdot)$ is the Bessel function of order N. Furthermore, $\phi(x)$ is also the solution of a differential equation that is a generalization of (B. 2).

$$\frac{d}{dx}\left[(1-x^2)\frac{d}{dx}\,\phi(x)\right] + \left(X_{mn} - C^2x^2 + \frac{\frac{1}{4} - M^2}{x^2}\right)\phi(x) = 0.$$

Both numerical and analytical details of the solution are available in his paper. Additional computations are available in papers by Huertly.[62, 63] At the present time, I am not aware of work on other geometric systems that has been carried out to a point that is convenient for either analysis or computation.

APPENDIX C

Computation of Rate-Reliability Curves

We shall evaluate bounds on the probability of error for communication over parallel channels. By parallel channels we mean a communication system or model in which the distortion introduced by each channel is independent of the signal and the distortion in all other channels.[9] In particular, we consider the case of orthogonal signal channels and independent additive noise sources.

The basic bound is in the form

$$P_e \leq B\, e^{-E[\rho, N_o, S, R]},$$

where $E(\rho, N_o, S, R)$ is called the "reliability." When it is presented as a function of the rate R (R is the rate at which information is produced by the source), it is called the "rate-reliability curve." If the source puts out one of M equally likely messages every T sec, R = ln M, measured in nats/channel use. S is the signal energy available in the T-sec interval. S = TP, where P is the average power produced by the source. N_b is the largest variance of the noises on the subchannels that are used for signaling. N_b is computed as the first step in the process of computing the exponent $E(\rho, N_b, S, R)$.

The rate-reliability curve is computed from the following sequence of operations. First, for arbitrary values of the parameter ρ, compute N_b as a function of S and ρ from

$$S = \sum_{N_i \leq N_b} \frac{(1+\rho)^2 (N_b - N_i)}{1 + \rho - \rho \frac{N_i}{N_b}}, \tag{C.1}$$

where the subchannels are ordered in terms of increasing noise variance, $N_i \leq N_{i+1}$. Equation C.1 has a unique solution for each S, ρ. This solution also defines the number of subchannels into which signal energy is put. For, if Q_n is the average signal energy used on the n^{th} subchannel,

$$Q_n = (1+\rho)^2 \frac{N_b - N_i}{1 + \rho - \rho \frac{N_i}{N_b}} \quad \text{for } N_i \leq N_b$$

$$= 0 \qquad \text{for } N_i > N_b. \tag{C.2}$$

Having computed N_b, the effective rate at which one transmits information, using the signal energy distribution defined above, is $R(\rho, N_b)$, where

$$R(\rho, N_b) = \frac{1}{2} \sum_{N_i \leq N_b} \ln \frac{N_b}{N_i}. \tag{C.3}$$

Now ρ is a positive parameter that can be adjusted to maximize the exponent E. It turns out that there are two regions of interest. It is optimum to set $\rho = 1$ for all rates less than $R(1, N_b)$. In that case, we find

$$E[\rho=1, N_b, S, R] = \frac{S}{4N_b} - \frac{1}{2} \sum_{N_i \leq N_b} \ln \frac{N_b}{N_i\left(2 - \frac{N_i}{N_b}\right)} - R \tag{C.4}$$

for $0 < R \leq R(1, N_b)$, where N_b is computed for $\rho = 1$.

For higher rates, the optimization requires $0 \leq \rho \leq 1$, and it can be shown that E is positive only if $R \leq R(0, N_b)$.[9] In this region we have

$$E[\rho, N_b, S, R] = \frac{\rho S}{2(1+\rho)\, N_b} - \frac{1}{2} \sum_{N_i < N_b} \ln \left(1 + \rho - \rho \frac{N_i}{N_b}\right) \tag{C.5}$$

for $R(1, N_b) \leq R \leq R(0, N_b)$. The canonic form of rate-reliability curve is illustrated in Fig. 10. Let us proceed to compute these functions for the parallel-channel model developed in Section III.

C.1 COMPUTATION OF N_b

The noise variance in the i^{th} subchannel is N_i, where from section 3.3a

$$N_i = \frac{N_o}{2\beta_i} = \frac{N_o}{2}\left[1 + e^{\alpha(i-D)}\right]. \tag{C.6}$$

N_b is computed from (C.1), with N_i given by (C.6)

$$\frac{2S}{(1+\rho)^2 N_o} = \sum_{\beta_i \geq \beta_b} \frac{\frac{N_b}{N_o} - \frac{1}{\beta_i}}{1 + \rho - \rho \frac{N_o}{N_b \beta_i}}. \tag{C.7}$$

Define a new parameter

$$\epsilon_b = \frac{N_b}{N_o} = \frac{1}{\beta_b} = 1 + e^{\alpha(i_b - D)}, \tag{C.8}$$

which in effect defines i_b, the number of subchannels selected by the optimization procedure. Equation C.7 is difficult to handle as a sum, but we can approximate the sum

over the index i by an integral. This same approximation was used to bound $\Sigma \beta_i$ in Eq. 38. Then, we find

$$\frac{2S}{(1+\rho)^2 N_o} = \int_0^{i_b} \frac{(\epsilon_b - 1) - e^{\alpha(i-D)}\, di}{1 + \rho - \frac{\rho}{\epsilon_b} - \frac{\rho}{\epsilon_b} e^{\alpha(i-D)}} . \tag{C. 9}$$

For $i_b > D$ we integrate in (C. 9) over two separate regions, that is, over $0 \leq i \leq D$, and $D \leq i \leq i_b$.

$$\frac{2S}{(1+\rho)^2 N_o} = (\epsilon_b - 1) \int_0^D \frac{di}{1 + \rho - \frac{\rho}{\epsilon_b} - \frac{\rho}{\epsilon_b} e^{-\alpha i}} - \int_0^D \frac{e^{-\alpha i}\, di}{1 + \rho - \frac{\rho}{\epsilon_b} - \frac{\rho}{\epsilon_b} e^{-\alpha i}}$$

$$+ (\epsilon_b - 1) \int_0^{i_b - D} \frac{di}{1 + \rho - \frac{\rho}{\epsilon_b} - \frac{\rho}{\epsilon_b} e^{\alpha i}} - \int_0^{i_b - D} \frac{e^{\alpha i}\, di}{1 + \rho - \frac{\rho}{\epsilon_b} - \frac{\rho}{\epsilon_b} e^{\alpha i}} . \tag{C. 10}$$

Now,

$$\int \frac{dx}{a + b\, e^{px}} = \frac{x}{a} - \frac{1}{ap} \ln (a + b\, e^{px}),$$

and

$$\int_0^D \frac{e^{px}\, dx}{a + b\, e^{px}} = \frac{1}{p} \int_1^{e^{pD}} \frac{du}{a + bu} = \frac{1}{pb} \ln (a+bu) \Big|_1^{e^{pD}} .$$

The computation of the integrals in (C. 10) leads to

$$\frac{2S}{N_o (1+\rho)^2} = \frac{i_b(\epsilon_b - 1)}{1 + \rho - \frac{\rho}{\epsilon_b}} + \left[\frac{\epsilon_b - 1}{\alpha\left(1 + \rho - \frac{\rho}{\epsilon_b}\right)} - \frac{\epsilon_b}{\alpha\rho} \right] \ln \frac{1 + \rho - \frac{\rho}{\epsilon_b} - \frac{\rho}{\epsilon_b} e^{-\alpha D}}{1 + \rho - \frac{\rho}{\epsilon_b} - \frac{\rho}{\epsilon_b} e^{\alpha(i_b - D)}},$$

which simplifies to

$$\frac{2S}{N_o (1+\rho)^2} = \frac{i_b(\epsilon_b - 1)}{1 + \rho - \frac{\rho}{\epsilon_b}} - \left[\frac{\epsilon_b}{\alpha\rho\left(1 + \rho - \frac{\rho}{\epsilon_b}\right)} \right] \ln \left(1 + \rho - \frac{\rho}{\epsilon_b} (1 + e^{-\alpha D}) \right). \tag{C. 11}$$

It can be shown that (C. 11) is valid for all values of i_b, including $i_b < D$. Let us further simplify (C. 11) by asserting that $e^{-\alpha D}$ is negligible in the logarithmic argument. We

regroup the expression as

$$\frac{2S}{N_o(1+\rho)^2} = \frac{\epsilon_b - 1}{1 + \rho\dfrac{\epsilon_b - 1}{\epsilon_b}}\left\{i_b - \frac{\epsilon_b}{a(\epsilon_b-1)}\ln\left(1 + \rho\left(\frac{\epsilon_b - 1}{\epsilon_b}\right)\right)\right\}. \tag{C.12}$$

Now look at solutions for (C. 12) in the limit of small ρ; and observe that, for $\rho \to 0$, the optimum signaling rate approaches the channel capacity. We find

$$\frac{2S}{N_o} = (\epsilon_b-1)\left(i_b - \frac{1}{a}\right), \tag{C.13}$$

where the term $1/a$ is negligible compared with i_b. Observe that

$$0 \leq \rho\left(\frac{\epsilon_b - 1}{\epsilon_b}\right) < \rho \leq 1.$$

Therefore, when $\rho\dfrac{(\epsilon_b-1)}{\epsilon_b}$ is approximately .5 or less, we can expand the logarithmic term as follows:

$$\ln\left(1 + \rho\left(\frac{\epsilon_b - 1}{\epsilon_b}\right)\right) \approx \rho\frac{(\epsilon_b-1)}{\epsilon_b},$$

so that (C. 12) becomes

$$\frac{2S}{N_o(1+\rho)^2} = \frac{(\epsilon_b-1)}{1 + \rho\left(\dfrac{\epsilon_b - 1}{\epsilon_b}\right)}\left(i_b - \frac{1}{a}\right). \tag{C.14}$$

Therefore Eq. C. 13 is a special case of Eq. C. 14. If the signal-to-noise ratio is large, so that $i_b > D$, then i_b will be large compared with unity, and $(\epsilon_b-1)/\epsilon_b$ approaches 1. In that case, for small ρ, (C. 14) becomes

$$\frac{2S}{N_o(1+\rho)} = (\epsilon_b-1)\left(i_b - \frac{1}{a}\right).$$

Also, for ρ very small, (C. 14) becomes

$$\frac{2S}{N_o(1+\rho)^2} = (\epsilon_b-1)\left(i_b - \frac{1}{a}\right).$$

Therefore the solutions to (C. 13) can be modified to obtain solutions of (C. 12) for several useful limiting cases by scaling the signal energy. Equation C. 13 can be written

$$\frac{2S\,a}{N_o}e^{aD} = e^{ai_b} = t\,e^t.$$

where $t = \alpha i_b$.

For large values of t, an asymptotically correct solution for t is

$$t = \ln u - \ln [\ln u],$$

where

$$u = 2S\ \alpha / N_o\ e^{\alpha D}.$$

We can rewrite the solution as

$$i_b = D + \frac{1}{\alpha} \ln \frac{\frac{2S\ \alpha}{N_o}}{N_o\left(\alpha D + \ln \frac{2S\ \alpha}{N_o}\right)} \tag{C.15a}$$

$$N_b = \frac{N_o}{2} \left\{1 + \frac{2S\ \alpha}{N_o\left(\alpha D + \ln \frac{2S\ \alpha}{N_o}\right)}\right\}. \tag{C.15b}$$

These solutions are good to better than a few percent for $u > 15$. When the signal-to-noise ratio becomes very small, then the solution becomes approximately $t = u$,

$$i_b = \frac{2S}{N_o} e^{\alpha D},$$

which goes to zero linearly with the signal-to-noise ratio, as expected. This is expected because β_i varies linearly with i, for small values of i.

C.2 COMPUTATION OF $R(\rho, N_b)$

Each value of ρ specifies an N_b which defines the upper limit in the expression for the rate $R(\rho, N_b)$. We recall that

$$R = \frac{1}{2} \sum_{N_i \leq N_b} \ln \frac{N_b}{N_i}$$

$$= \frac{i_b + 1}{2} \ln \epsilon_b - \frac{1}{2} \sum_{i \leq i_b} \ln (1 + e^{\alpha(i-D)}). \tag{C.16}$$

Now consider, again, the two possibilities, $i_b \leq D$, and $i_b > D$. If $i_b \leq D$, then the sum reduces to

$$\sum_{i=D-i_b}^{D} \ln \left(1 + e^{-\alpha i}\right) \approx \sum_{i=D-i_b}^{D} e^{-\alpha i} = e^{-\alpha(D-i_b)} \frac{1 - e^{-\alpha(i_b+1)}}{1 - e^{-\alpha}}.$$

This term is negligible compared with

$$\frac{(i_b+1)\ln \epsilon_b}{2};$$

hence, $R = \frac{1}{2}(i_b+1)\ln \epsilon_b$.

$$R = \frac{1}{2}\left[1 + D + \frac{1}{a}\ln \frac{2S}{N_o D\left(1 + \frac{1}{aD}\ln \frac{2S\,a}{N_o}\right)}\right] \ln 1 + \frac{2S}{N_o D\left(1 + \frac{1}{a}\ln \frac{2S\,a}{N_o}\right)}. \qquad \text{(C. 17)}$$

For $i_b \geq D$, we have approximately

$$R \approx \frac{i_b + 1}{2}\ln \epsilon_b - \frac{1}{2}\left[\ln 2 + \sum_{i=1}^{i_b-D} \ln (1+e^{ai})\right].$$

Again approximate the logarithmic terms, this time assuming $1 + e^{ai} \gg 1$.

$$\sum_{i=1}^{i_b-D} \ln (1+e^{ai}) \approx \sum_{i=1}^{i_b-D} ai = \frac{a}{2}(i_b-D)(i_b-D+1).$$

From Eq. C. 15a, we have

$$\frac{a}{2}(i_b-D)(i_b-D+1) = \frac{1}{2}\left[1 + \frac{1}{a}\ln \frac{2S}{N_o D\left(1 + \frac{\ln \frac{2S\,a}{N_o}}{aD}\right)}\right] \ln \frac{2S}{N_o D\left(1 + \frac{\ln \frac{2S\,a}{N_o}}{aD}\right)}. \qquad \text{(C. 18)}$$

Substituting (C. 18) in (C. 16) we get

$$R = \frac{1}{2}\left[1 + D + \frac{1}{a}\ln \frac{2S}{N_o D\left(1 + \frac{1}{aD}\ln \frac{2S\,a}{N_o}\right)}\right] \ln \left[1 + \frac{2S}{N_o D\left(1 + \frac{1}{aD}\ln \frac{2S\,a}{N_o}\right)}\right]$$

$$- \frac{1}{4}\left[1 + \frac{1}{a}\ln \frac{2S}{N_o D\left(1 + \frac{1}{aD}\ln \frac{2S\,a}{N_o}\right)}\right] \ln \frac{2S}{N_o D\left(1 + \frac{\ln \frac{2S\,a}{N_o}}{aD}\right)}. \qquad \text{(C. 19)}$$

C. 3 COMPUTATION OF E(0)

The exponent for zero rate is computed from the value of N_b for $\rho = 1$.

$$E(0) = \frac{S}{4N_b} + \frac{1}{2} \sum_{N_i \leq N_b} \ln \frac{N_b}{N_i\left(2 - \frac{N_i}{N_b}\right)}$$

$$= \frac{S}{4N_b} + (i_b+1) \ln \left(1 + e^{\alpha(i_b-D)}\right) - \frac{1}{2} \sum_{i \leq i_b} \ln (1+e^{\alpha(i-D)}) \left(1 + 2e^{\alpha(i_b-D)} - e^{\alpha(i-D)}\right). \tag{C. 20}$$

For $i_b \leq D$, the arguments in each term of the logarithmic sum can be expanded as

$$\ln (1+e^{\alpha(i-D)}) + \ln \left(1 + 2e^{\alpha(i_b-D)} - e^{\alpha(i-D)}\right) \approx e^{\alpha(i-D)} + 2e^{\alpha(i_b-D)} - e^{\alpha(i-D)} = 2e^{\alpha(i_b-D)}.$$

Therefore, to first order

$$E(0) = \frac{S}{4N_b} + (i_b+1)\, e^{\alpha(i_b-D)} - \frac{1}{2}(i_b+1)\, 2e^{\alpha(i_b-D)} = \frac{S}{4N_b}$$

$$E(0) = \frac{S}{2N_o\left(1 + \frac{S}{N_o D} \frac{1}{1 + \frac{1}{\alpha D} \ln \frac{S\alpha}{N_o}}\right)}. \tag{C. 21}$$

For $i_b \geq D$, we break up the sum in (C. 20) into three terms

$$\sum_{i \leq i_b} = \sum_{i=0}^{D} \ln (1+e^{-\alpha i}) + \sum_{i=1}^{i_b-D} \ln (1+e^{\alpha i}) + \sum_{i=0}^{i_b} \ln 2e^{\alpha(i_b-D)} \left[1 + \frac{1 - e^{\alpha(i-D)}}{2e^{\alpha(i_b-D)}}\right]$$

$$\approx \ln 2 + \sum_{i=1}^{D} e^{-\alpha i} + \alpha \sum_{i=1}^{i_b-D} i + (i_b+1) \ln 2e^{\alpha(i_b-D)} + \sum_{i=0}^{i_b} \ln \left[1 + \frac{1 - e^{\alpha(i-D)}}{2e^{\alpha(i_b-D)}}\right].$$

Therefore

$$\sum_{i \leq i_b} \approx (i_b+2) \ln 2 + e^{-\alpha} \frac{1-e^{-\alpha D}}{1-e^{-\alpha}} + \frac{\alpha}{2}(i_b-D)(i_b-D+1) + (i_b+1)\, \alpha(i_b-D) + \sum_{i=0}^{i_b} \frac{1-e^{\alpha(i-D)}}{2e^{\alpha(i_b-D)}}.$$

Now, combine similar terms and neglect small quantities. We find

$$\sum = (i_b+1)\left[\ln 2 + \frac{3}{2}\alpha(i_b-D)\right] - \frac{D}{2}\alpha(i_b-D) - \ln 2. \tag{C. 22}$$

Substituting (C. 22) in (C. 20), we find

$$E(0) = \frac{S}{4N_b} + (i_b+1)\left[\ln\left(1+e^{\alpha(i_b-D)}\right) - \frac{1}{2}\left(\ln 2 + \frac{3}{2}\alpha(i_b-D)\right)\right] - \left(\frac{D}{4}\alpha(i_b-D) + \frac{\ln 2}{2}\right).$$

Now, with $i_b > D$, we can set $\ln\left(1+e^{\alpha(i_b-D)}\right) = \alpha(i_b-D)$. Therefore

$$E(0) = \frac{S}{4N_b} + \frac{\alpha(i_b-D)}{4}(i_b-D+1) - (i_b+2)\frac{\ln 2}{2}.$$

Finally, substituting for i_b, we find

$$E(0) = \frac{S}{2N_o\left[1 + \frac{S}{N_o D}\,\frac{1}{1+\frac{1}{\alpha D}\ln\frac{\alpha S}{N_o}}\right]} + \frac{1}{4}\left[\ln\frac{S}{N_o D\left(1+\frac{1}{\alpha D}\ln\frac{\alpha S}{N_o}\right)}\right]\left[1+\frac{1}{\alpha}\ln\frac{S}{N_o D\left[1+\frac{1}{\alpha D}\ln\frac{S\alpha}{N_o}\right]}\right] - (i_b+2)\frac{\ln 2}{2}.$$

Acknowledgment

I am pleased to have this opportunity to express sincere appreciation to the people who have contributed to the successful completion of my graduate studies. My supervisor, Professor Robert S. Kennedy, has given timely advice in keeping the appropriate focus in my research. Professor Peter Elias made acute comments on draft versions of this report. Professor Hermann A. Haus, who also acted as a reader, and several other faculty members and graduate students who have been most helpful deserve thanks. Mr. John Avery Tucker and Professor Leonard A. Gould have been gracious counsellors and good friends during my career at M. I. T.

I appreciate the support extended to me during my graduate studies by the Communication Systems Laboratory of Sylvania Electronic Systems, and especially the friendship of Dr. David A. Chesler who embodies the tradition of scientific investigation. I also wish to thank the Research Laboratory of Electronics, M. I. T., for support during the period of my thesis research.

References

1. C. E. Shannon, "A Mathematical Theory of Communication," Bell System Tech. J. 27, 379, 623 (1948).
2. A. Feinstein, "Error Bounds in Noisy Channels without Memory," IRE Trans., Vol. IT-1, pp. 13-14, September 1955.
3. J. Wolfowitz, Coding Theorems of Information Theory (Prentice-Hall, Inc., New York, 1965).
4. R. M. Fano, Transmission of Information (The M. I. T. Press, Cambridge, Mass., 1961).
5. R. G. Gallager, "A Simple Derivation of the Coding Theorem and Some Applications," IEEE Trans., Vol. IT-11, pp. 3-17, January 1965.
6. J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering (John Wiley and Sons, Inc., New York, 1965), Chaps. 4 and 5.
7. H. J. Landau and D. Slepian, "On the Optimality of Regular Simplex Codes," Bell System Tech. J. 45, 1247-1272 (1966).
8. J. L. Holsinger, "Digital Communication over Fixed Time-Continuous Channels with Memory – With Special Application to Telephone Channels," Technical Report 430, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, October 20, 1964.
9. P. M. Ebert, "Error Bounds for Parallel Communication Channels," Technical Report 448, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, August 1, 1966.
10. D. Slepian and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty-I," Bell System Tech. J. 40, 43-64 (1961).
11. D. Gabor, "Light and Information," in Progress in Optics, Vol. I., E. Wolf (ed.) (North-Holland Publishing Company, Amsterdam, 1961), pp. 111-153.
12. J. A. Stratton, Electromagnetic Theory (McGraw-Hill Book Company, New York, 1941).
13. R. B. Adler, L. J. Chu, and R. M. Fano, Electromagnetic Energy Transmission and Radiation (John Wiley and Sons, Inc., New York, 1960).
14. J. Winthrop and C. Worthington, "Convolution Formulation of Fresnel Diffraction," J. Opt. Soc. Am. 56, 588-591 (1966).
15. F. Riesz and B. Sz-Nagy, Functional Analysis (F. Ungar Publishing Company, New York, 1955).
16. D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-IV: Extensions to Many Dimensions; Generalized Prolate Spheroidal Functions," Bell System Tech. J. 43, 3009-3057 (1964).
17. D. Slepian, "Some Asymptotic Expansions for Prolate Spheroidal Wave Functions," J. Math. Phys. 44, 99-140 (June 1965).
18. J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and P. J. Corbató, Spheroidal Wave Functions (John Wiley and Sons, Inc., New York, 1956).
19. C. Flammer, Spheroidal Wave Functions (Stanford University Press, Stanford, California, 1957).
20. D. Slepian, "Analytic Solution of Two Apodization Problems," J. Opt. Soc. Am. 55, 1110-1115 (1965).
21. P. Jacquinot and B. Roizen Dossier, "Apodization," in Progress in Optics, Vol. III, E. Wolf (ed.) (North-Holland Publishing Company, Amsterdam, 1964), pp. 29-186.
22. R. Barakat, "Solution of the Luneberg Apodization Problems," J. Opt. Soc. Am. 52, 264-275 (1962).

23. G. V. Borgiotti, "Maximum Power Transfer between Two Planar Antennas in the Fresnel Zone," IEEE Trans., Vol. AP-14, pp. 158-163, March 1966.

24. H. Gamo, "Matrix Treatment of Partial Coherence," in Progress in Optics, Vol. III, E. Wolf (ed.) (North-Holland Publishing Company, Amsterdam, 1964), pp. 187-382.

25. G. D. Boyd and J. P. Gordon, "Confocal Multimode Resonator for Millimeter through Optical Wavelength Masers," Bell System Tech. J. 40, 489-508 (1961).

26. G. D. Boyd and H. Kogelnik, "Generalized Confocal Resonator Theory," Bell System Tech. J. 41, 1347-1371 (July 1962).

27. A. G. Fox and T. Li, "Resonant Modes in a Maser Interferometer," Bell System Tech. J. 40, 453-489 (1961).

28. A. W. Lohman, "The Space-Bandwidth Product," Notes for a Lecture presented at the 1966 Summer School on Optical Data Processing, University of Michigan, Ann Arbor, Michigan.

29. H. J. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty-III," Bell System Tech. J. 41, 1295-1336 (1962).

30. M. von Laue, "Die Freiheitsgrade von Strahlenbündelen," Ann. Physik 44, 1197-1212 (August 1914).

31. T. T. Taylor, "Design of Line-Source Antennas for Narrow Beamwidth and Low Sidelobes," IRE Trans., Vol. AP-3, pp. 16-28, January 1955.

32. D. R. Rhodes, "The Optimum Line Source for the Best Mean-Square Approximation to a Given Radiation Pattern," IEEE Trans., Vol. AP-11, pp. 440-446, July 1963.

33. W. P. Harthill, "Prolate Spheroidal Wave Functions and Their Applications to Planar Aperture Antenna Pattern Synthesis," Ph. D. Thesis, University of Washington, Seattle, Washington, 1966.

34. P. M. Woodward and J. Lawson, "Theoretical Precision with which an Arbitrary Radiation-Pattern May Be Obtained," Proc. IEEE, Part III, pp. 363-370, May 1949.

35. A. Walther, "Gabor's Theorem and Energy Transfer through Lenses," J. Opt. Soc. Am. 57, 639-644 (1967).

36. J. L. Harris, "Diffraction and Resolving Power," J. Opt. Soc. Am. 54, 931-936 (1964).

37. C. W. Barnes, "Object Restoration in a Diffraction-Limited Imaging System," J. Opt. Soc. Am. 56, 575-578 (1966).

38. B. Frieden, "Band Unlimited Reconstruction of Optical Objects and Spectra," J. Opt. Soc. Am. 57, 1013-1019 (1967).

39. C. W. McCutchen, "Superresolution in Microscopy and the Abbe Resolution Limit," J. Opt. Soc. Am. 57, 1190-1192 (1967).

40. C. K. Rushforth and R. W. Harris, "Restoration, Resolution, and Noise," J. Opt. Soc. Am. 58, 539-544 (1968).

41. L. J. Chu, "Physical Limitations on the Gain and Q of Omni-directional Antennas," J. Appl. Phys. 19, 1163-1175 (1948).

42. D. R. Rhodes, "On the Stored Energy of Planar Apertures," IEEE Trans., Vol. AP-14, pp. 676-683, November 1966.

43. H. S. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-II," Bell System Tech. J. 40, pp. 65-84 (1961).

44. D. Auston, "Transverse Mode Locking," Paper No. 17-Q2, presented at the 1968 International Quantum Electronics Conference, Miami Beach, Florida, May 1968.

45. E. V. Hoversten, "The Atmosphere as an Optical Communication Channel," IEEE International Convention Record, Vol. 15, Part 2, pp. 137-145, March 1967.

46. J. Moldon, "Imaging of Objects Viewed through a Turbulent Atmosphere," Ph. D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, September 1968; to be published as Technical Report 469, Research Laboratory of Electronics, M. I. T.

47. S. J. Halme, "On Optimum Reception through a Turbulent Atmosphere," Quarterly Progress Report No. 88, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, January 15, 1968, pp. 247-254.

48. Jane W. S. Liu, "Reliability of Quantum-Mechanical Communication Systems," Technical Report 468, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, December 31, 1968.

49. M. Ross, Laser Receivers (John Wiley and Sons, Inc., New York, 1966).

50. D. G. Childers and I. S. Reed, "Space-Time Cross-Correlation Functions for Antenna Array Elements in a Noise Field," IEEE Trans., Vol. IT-11, pp. 182-190, April 1965.

51. R. G. Gallager, Information Theory and Reliable Communication (John Wiley and Sons, Inc., New York, in press).

52. C. W. Helstrom, "Detectability of Coherent Optical Signals in a Heterodyne Receiver," J. Opt. Soc. Am. 57, 353-361 (1967).

53. Jane W. S. Liu, "Comparison of the Performance of the Photon Counter, and the Classical Optimum Receiver," Quarterly Progress Report No. 88, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, January 15, 1968, pp. 255-259.

54. C. W. Helstrom, "Detection Theory and Quantum Mechanics," Inform. Contr. 10, 254-291 (1967).

55. F. A. Jenkins and H. E. White, Physical Optics (McGraw-Hill Book Company, New York, 1957).

56. H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part I. (John Wiley and Sons, Inc., New York, 1968).

57. C. H. Graham (ed.), Vision and Visual Perception (John Wiley and Sons, Inc., New York, 1965).

58. H. von Helmholz, Handbuch der Physiologischen Öptik (Voss, Hamburg and Leipzig, 1886). English translation by J. R. C. Southall, "Helmholz-Physiological Optics," Vol. 1-3 (Optical Society of America, Rochester, N. Y., 1924-1925).

59. S. Shlaer, "Visual Acuity and Illumination," J. Gen. Physiol. 21, 165-188 (November 1937).

60. M. Born and E. Wolf, Principles of Optics (Pergamon Press, London, 1959).

61. National Bureau of Standards, "Tables of Coulomb Wave Functions," Vol. I, Applied Mathematics Series No. 17 (U. S. Department of Commerce, Washington, D. C., 1952), pp. 130-135.

62. J. Huertly, "Hyperspheroidal Functions – Optical Resonators with Circular Mirrors," in Proc. Symposium on Quasi-Optics (Polytechnic Press, New York, 1964), pp. 367-375.

63. J. Huertly, "Maximum Power Transfer between Finite Antennas," IEEE Trans., Vol. AP-15, pp. 298-299, March 1967.

64. J. S. Richters, "Communication over Fading Dispersive Channels," Technical Report 464, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, November 30, 1967.

JOINT SERVICES ELECTRONICS PROGRAM
REPORTS DISTRIBUTION LIST

Department of Defense

Dr. A. A. Dougal
Asst Director (Research)
Ofc of Defense Res & Eng
Department of Defense
Washington, D. C. 20301

Office of Deputy Director
(Research and Information, Rm 3D1037)
Department of Defense
The Pentagon
Washington, D. C. 20301

Director
Advanced Research Projects Agency
Department of Defense
Washington, D. C. 20301

Director for Materials Sciences
Advanced Research Projects Agency
Department of Defense
Washington, D. C. 20301

Headquarters
Defense Communications Agency (340)
Washington, D. C. 20305

Defense Documentation Center
Attn: DDC-TCA
Cameron Station
Alexandria, Virginia 22314

Director
National Security Agency
Attn: TDL
Fort George G. Meade, Maryland 20755

Weapons Systems Evaluation Group
Attn: Colonel Blaine O. Vogt
400 Army-Navy Drive
Arlington, Virginia 22202

Central Intelligence Agency
Attn: OCR/DD Publications
Washington, D. C. 20505

Department of the Air Force

Hq USAF (AFRDDD)
The Pentagon
Washington, D. C. 20330

Hq USAF (AFRDDG)
The Pentagon
Washington, D. C. 20330

Hq USAF (AFRDSD)
The Pentagon
Washington, D. C. 20330

Colonel E. P. Gaines, Jr.
ACDA/FO
1901 Pennsylvania Avenue N. W.
Washington, D. C. 20451

Lt Col R. B. Kalisch (SREE)
Chief, Electronics Division
Directorate of Engineering Sciences
Air Force Office of Scientific Research
Arlington, Virginia 22209

Dr. I. R. Mirman
AFSC (SCT)
Andrews Air Force Base, Maryland 20331

AFSC (SCTSE)
Andrews Air Force Base, Maryland 20331

Mr. Morton M. Pavane, Chief
AFSC Scientific and Technical Liaison
Office
26 Federal Plaza, Suite 1313
New York, New York 10007

Rome Air Development Center
Attn: Documents Library (EMTLD)
Griffiss Air Force Base, New York 13440

Mr. H. E. Webb (EMIIS)
Rome Air Development Center
Griffiss Air Force Base, New York 13440

Dr. L. M. Hollingsworth
AFCRL (CRN)
L. G. Hanscom Field
Bedford, Massachusetts 01730

AFCRL (CRMPLR), Stop 29
AFCRL Research Library
L. G. Hanscom Field
Bedford, Massachusetts 01730

Hq ESD (ESTI)
L. G. Hanscom Field
Bedford, Massachusetts 01730

Professor J. J. D'Azzo
Dept of Electrical Engineering
Air Force Institute of Technology,
Wright-Patterson Air Force Base,
Ohio 45433

Dr. H. V. Noble (CAVT)
Air Force Avionics Laboratory
Wright-Patterson Air Force Base,
Ohio 45433

Director
Air Force Avionics Laboratory
Wright-Patterson Air Force Base,
Ohio 45433

AFAL (AVTA/R. D. Larson)
Wright-Patterson Air Force Base,
Ohio 45433

Director of Faculty Research
Department of the Air Force
U.S. Air Force Academy
Colorado Springs, Colorado 80840

Academy Library (DFSLB)
USAF Academy
Colorado Springs, Colorado 80840

Director
Aerospace Mechanics Division
Frank J. Seiler Research Laboratory (OAR)
USAF Academy
Colorado Springs, Colorado 80840

Director, USAF PROJECT RAND
Via: Air Force Liaison Office
The RAND Corporation
Attn: Library D
1700 Main Street
Santa Monica, California 90406

Hq SAMSO (SMTTA/Lt Nelson)
Air Force Unit Post Office
Los Angeles, California 90045

Det 6, Hq OAR
Air Force Unit Post Office
Los Angeles, California 90045

AUL3T-9663
Maxwell Air Force Base, Alabama 36112

AFETR Technical Library
(ETV, MU-135)
Patrick Air Force Base, Florida 32925

ADTC (ADBPS-12)
Eglin Air Force Base, Florida 32542

Mr. B. R. Locke
Technical Adviser, Requirements
USAF Security Service
Kelly Air Force Base, Texas 78241

Hq AMD (AMR)
Brooks Air Force Base, Texas 78235

USAFSAM (SMKOR)
Brooks Air Force Base, Texas 78235

Commanding General
Attn: STEWS-RE-L, Technical Library
White Sands Missile Range,
New Mexico 88002

Hq AEDC (AETS)
Attn: Library/Documents
Arnold Air Force Station, Tennessee 37389

European Office of Aerospace Research
APO New York 09667

Department of the Army

Physical & Engineering Sciences Division
U.S. Army Research Office
3045 Columbia Pike
Arlington, Virginia 22204

Commanding General
U.S. Army Security Agency
Attn: IARD-T
Arlington Hall Station
Arlington, Virginia 22212

Commanding General
U.S. Army Materiel Command
Attn: AMCRD-TP
Washington, D. C. 20315

Commanding Officer
Harry Diamond Laboratories
Attn: Dr. Berthold Altman (AMXDO-TI)
Connecticut Avenue and
Van Ness Street N. W.
Washington, D. C. 20438

Director
Walter Reed Army Institute of Research
Walter Reed Army Medical Center
Washington, D. C. 20012

Commanding Officer (AMXRD-BAT)
U.S. Army Ballistics Research Laboratory
Aberdeen Proving Ground
Aberdeen, Maryland 21005

Technical Director
U.S. Army Limited War Laboratory
Aberdeen Proving Ground
Aberdeen, Maryland 21005

Commanding Officer
Human Engineering Laboratories
Aberdeen Proving Ground
Aberdeen, Maryland 21005

U.S. Army Munitions Command
Attn: Science & Technology Information Branch, Bldg 59
Picatinny Arsenal, SMUPA-VA6
Dover, New Jersey 07801

U.S. Army Mobility Equipment Research and Development Center
Attn: Technical Document Center, Bldg 315
Fort Belvoir, Virginia 22060

Director
U.S. Army Engineer Geodesy, Intelligence & Mapping Research and Development Agency
Fort Belvoir, Virginia 22060

Dr. Herman Robl
Deputy Chief Scientist
U.S. Army Research Office (Durham)
Box CM, Duke Station
Durham, North Carolina 27706

Richard O. Ulsh (CRDARD-IPO)
U.S. Army Research Office (Durham)
Box CM, Duke Station
Durham, North Carolina 27706

Technical Director (SMUFA-A2000-107-1)
Frankford Arsenal
Philadelphia, Pennsylvania 19137

Redstone Scientific Information Center
Attn: Chief Document Section
U.S. Army Missile Command
Redstone Arsenal, Alabama 35809

Commanding General
U.S. Army Missile Command
Attn: AMSMI-REX
Redstone Arsenal, Alabama 35809

Commanding General
U.S. Army Strategic Communications Command
Attn: SCC-CG-SAE
Fort Huachuca, Arizona 85613

Commanding Officer
Army Materials and Mechanics Research Center
Attn: Dr. H. Priest
Watertown Arsenal
Watertown, Massachusetts 02172

Commandant
U.S. Army Air Defense School
Attn: Missile Science Division, C&S Dept,
P. O. Box 9390
Fort Bliss, Texas 79916

Commandant
U.S. Army Command and General Staff College
Attn: Acquisitions, Lib Div
Fort Leavenworth, Kansas 66027

Commanding Officer
U.S. Army Electronics R&D Activity
White Sands Missile Range,
New Mexico 88002

Mr. Norman J. Field, AMSEL-RD-S
Chief, Office of Science & Technology
Research and Development Directorate
U.S. Army Electronics Command
Fort Monmouth, New Jersey 07703

Mr. Robert O. Parker, AMSEL-RD-S
Executive Secretary, JSTAC
U. S. Army Electronics Command
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Fort Monmouth, New Jersey 07703
Attn: AMSEL-SC
RD-GF
RD-MT
XL-D
XL-E
XL-C
XL-S (Dr. R. Buser)
HL-CT-DD
HL-CT-R
HL-CT-L (Dr. W.S.McAfee)
HL-CT-O
HL-CT-I
HL-CT-A
NL-D
NL-A
NL-P
NL-P-2 (Mr. D. Haratz)
NL-R (Mr. R. Kulinyi)
NL-S
KL-D
KL-E
KL-S (Dr. H. Jacobs)
KL-SM (Drs. Schiel/Hieslmair)
KL-T
VL-D
VL-F (Mr. R.J. Niemela)
WL-D

Dr. A. D. Schnitzler, AMSEL-HL-NVII
Night Vision Laboratory, USAECOM
Fort Belvoir, Virginia 22060

Dr. G. M. Janney, AMSEL-HL-NVOR
Night Vision Laboratory, USAECOM
Fort Belvoir, Virginia 22060

Atmospheric Sciences Office
Atmospheric Sciences Laboratory
White Sands Missile Range,
New Mexico 88002

Missile Electronic Warfare Technical
Area, (AMSEL-WT-MT)
White Sands Missile Range,
New Mexico 88002

Deputy for Research and Engineering
(AMSWE-DRE)
U.S. Army Weapons Command
Rock Island Arsenal
Rock Island, Illinois 61201

Project Manager
Common Positioning & Navigation Systems
Attn: Harold H. Bahr (AMCPM-NS-TM),
Bldg 439
U.S. Army Electronics Command
Fort Monmouth, New Jersey 07703

Director
U.S. Army Advanced Materiel
Concepts Agency
Washington, D. C. 20315

Department of the Navy

Director, Electronic Programs
Attn: Code 427
Department of the Navy
Washington, D. C. 20360

Commander
U.S. Naval Security Group Command
Attn: G43
3801 Nebraska Avenue
Washington, D. C. 20390

Director
Naval Research Laboratory
Washington, D. C. 20390
Attn: Code 2027
Dr. W. C. Hall, Code 7000
Dr. A. Brodzinsky, Supt. Elec. Div.

Dr. G. M. R. Winkler
Director, Time Service Division
U.S. Naval Observatory
Washington, D. C. 20390

Naval Air Systems Command
AIR 03
Washington, D. C. 20360

Naval Ship Systems Command
Ship 031
Washington, D. C. 20360

Naval Ship Systems Command
Ship 035
Washington, D. C. 20360

U. S. Naval Weapons Laboratory
Dahlgren, Virginia 22448

Naval Electronic Systems Command
ELEX 03, Room 2046 Munitions Building
Department of the Navy
Washington, D. C. 20360

Head, Technical Services Division
Naval Investigative Service Headquarters
4420 North Fairfax Drive
Arlington, Virginia 22203

Commander
U.S. Naval Ordnance Laboratory
Attn: Librarian
White Oak, Maryland 21502

Commanding Officer
Office of Naval Research Branch Office
Box 39 FPO
New York, New York 09510

Commanding Officer
Office of Naval Research Branch Office
219 South Dearborn Street
Chicago, Illinois 60604

Commanding Officer
Office of Naval Research Branch Office
495 Summer Street
Boston, Massachusetts 02210

Commander (ADL)
Naval Air Development Center
Johnsville, Warminster,
Pennsylvania 18974

Commanding Officer
Naval Training Device Center
Orlando, Florida 32813

Commander (Code 753)
Naval Weapons Center
Attn: Technical Library
China Lake, California 93555

Commanding Officer
Naval Weapons Center
Corona Laboratories
Attn: Library
Corona, California 91720

Commander
U.S. Naval Missile Center
Point Mugu, California 93041

W. A. Eberspacher, Associate Head
Systems Integration Division
Code 5340A, Box 15
U.S. Naval Missile Center
Point Mugu, California 93041

Commander
Naval Electronics Laboratory Center
Attn: Library
San Diego, California 92152

Deputy Director and Chief Scientist
Office of Naval Research Branch Office
1030 East Green Street
Pasadena, California 91101

Library (Code 2124)
Technical Report Section
Naval Postgraduate School
Monterey, California 93940

Glen A. Myers (Code 52 Mv)
Assoc. Prof. of Electrical Engineering
Naval Postgraduate School
Monterey, California 93940

Commanding Officer and Director
U.S. Naval Underwater Sound Laboratory
Fort Trumbull
New London, Connecticut 06840

Commanding Officer
Naval Avionics Facility
Indianapolis, Indiana 46241

Other Government Agencies

Dr. H. Harrison, Code RRE
Chief, Electrophysics Branch
National Aeronautics and
Space Administration
Washington, D. C. 20546

NASA Lewis Research Center
Attn: Library
21000 Brookpark Road
Cleveland, Ohio 44135

Los Alamos Scientific Laboratory
Attn: Reports Library
P. O. Box 1663
Los Alamos, New Mexico 87544

Federal Aviation Administration
Attn: Admin Stds Div (MS-110)
800 Independence Avenue S. W.
Washington, D. C. 20590

Mr. M. Zane Thornton, Chief
Network Engineering, Communications
and Operations Branch
Lister Hill National Center for
Biomedical Communications
8600 Rockville Pike
Bethesda, Maryland 20014

U. S. Post Office Department
Library - Room 6012
12th & Pennsylvania Avenue, N. W.
Washington, D. C. 20260

Non-Government Agencies

Director
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Mr. Jerome Fox, Research Coordinator
Polytechnic Institute of Brooklyn
333 Jay Street
Brooklyn, New York 11201

Director
Columbia Radiation Laboratory
Columbia University
538 West 120th Street
New York, New York 10027

Director
Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Director
Stanford Electronics Laboratories
Stanford University
Stanford, California 94305

Director
Microwave Physics Laboratory
Stanford University
Stanford, California 94305

Director
Electronics Research Laboratory
University of California
Berkeley, California 94720

Director
Electronic Sciences Laboratory
University of Southern California
Los Angeles, California 90007

Director
Electronics Research Center
The University of Texas at Austin
Austin, Texas 78712

Division of Engineering and
Applied Physics
Harvard University
Cambridge, Massachusetts 02138

Dr. G. J. Murphy
The Technological Institute
Northwestern University
Evanston, Illinois 60201

Dr. John C. Hancock, Head
School of Electrical Engineering
Purdue University
Lafayette, Indiana 47907

Department of Electrical Engineering
Texas Technological College
Lubbock, Texas 79409

Aerospace Corporation
P. O. Box 95085
Los Angeles, California 90045
Attn: Library Acquisition Group

Prof. Nicholas George
California Institute of Technology
Pasadena, California 91109

Aeronautics Library
Graduate Aeronautical Laboratories
California Institute of Technology
1201 E. California Blvd.
Pasadena, California 91109

The Johns Hopkins University
Applied Physics Laboratory
Attn: Document Librarian
8621 Georgia Avenue
Silver Spring, Maryland 20910

Hunt Library
Carnegie-Mellon University
Schenley Park
Pittsburgh, Pennsylvania 15213

Dr. Leo Young
Stanford Research Institute
Menlo Park, California 94025

School of Engineering Sciences
Arizona State University
Tempe, Arizona 85281

Engineering and Mathematical
Sciences Library
University of California at Los Angeles
405 Hilgard Avenue
Los Angeles, California 90024

The Library
Government Publications Section
University of California
Santa Barbara, California 93106

Carnegie-Mellon University
Electrical Engineering Department
Pittsburgh, Pennsylvania 15213

Prof. Joseph E. Rowe
Chairman, Dept of Electrical Engineering
The University of Michigan
Ann Arbor, Michigan 48104

New York University
College of Engineering
New York, New York 10019

Syracuse University
Department of Electrical Engineering
Syracuse, New York 13210

Yale University
Engineering Department
New Haven, Connecticut 06520

Airborne Instruments Laboratory
Deerpark, New York 11729

Raytheon Company
Attn: Librarian
Bedford, Massachusetts 01730

Raytheon Company
Research Division Library
28 Seyon Street
Waltham, Massachusetts 02154

Dr. Sheldon J. Welles
Electronic Properties Information Center
Mail Station E-175
Hughes Aircraft Company
Culver City, California 90230

Dr. Robert E. Fontana
Systems Research Laboratories Inc.
7001 Indian Ripple Road
Dayton, Ohio 45440

Nuclear Instrumentation Group
Bldg 29, Room 101
Lawrence Radiation Laboratory
University of California
Berkeley, California 94720

Sylvania Electronic Systems
Applied Research Laboratory
Attn: Documents Librarian
40 Sylvan Road
Waltham, Massachusetts 02154

Hollander Associates
P. O. Box 2276
Fullerton, California 92633

Illinois Institute of Technology
Department of Electrical Engineering
Chicago, Illinois 60616

The University of Arizona
Department of Electrical Engineering
Tucson, Arizona 85721

Utah State University
Department of Electrical Engineering
Logan, Utah 84321

Case Western Reserve University
Engineering Division
University Circle
Cleveland, Ohio 44106

Lincoln Laboratory
Massachusetts Institute of Technology
Lexington, Massachusetts 02173

The University of Iowa
The University Libraries
Iowa City, Iowa 52240

Lenkurt Electric Co., Inc.
1105 County Road
San Carlos, California 94070
Attn: Mr. E. K. Peterson

Philco Ford Corporation
Communications & Electronics Division
Union Meeting and Jolly Roads
Blue Bell, Pennsylvania 19422

Union Carbide Corporation
Electronic Division
P. O. Box 1209
Mountain View, California 94041

Department of Electrical Engineering
Rice University
Houston, Texas 77001

Research Laboratories for the
Engineering Sciences
School of Engineering and Applied Science
University of Virginia
Charlottesville, Virginia 22903

Department of Electrical Engineering
College of Engineering and Technology
Ohio University
Athens, Ohio 45701

Project MAC
Document Room
Massachusetts Institute of Technology
545 Technology Square
Cambridge, Massachusetts 02139

Department of Electrical Engineering
Lehigh University
Bethlehem, Pennsylvania 18015

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
Error Bounds for Digital Communication over Spatially Modulated Channels

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*
Richard L. Greenspan

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 30, 1969	108	64

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
Da 28-043-AMC-02536(E)	Technical Report 468
b. PROJECT NO. 200-14501-B31F	
c. NASA Grant NGL 22-009-013	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* None
d.	

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Joint Services Electronics Program through U.S. Army Electronics Command

13. ABSTRACT

The transmission of digital information by means of electromagnetic wave propagation between terminals is reconsidered in order to understand the influence of the radiation channel, including the terminal antennas, on the quality of communication that can be maintained. In particular, we investigate the potential to improve performance by modulating the messages onto a set of spatially varying patterns across the source antenna.

We show that the radiation channel can be modeled by the parallel combination of independent spatially and temporally modulated subchannels when the background noise at the receiver has Gaussian statistics and is uncorrelated at points separated by more than a few wavelengths of the carrier frequency. Error bounds are evaluated for digital transmission with optimum distribution of signal power to the subchannels. These bounds are used to interpret the significance of the signal and channel parameters and the interplay among them.

The principal conclusion is that spatial modes have the same function as time-variant signal modes. They can be viewed as independent subchannels that extend the effective bandwidth of the communication link. The number of effective spatial modes depends on the dimension of the terminal antennas and on the input signal-to-noise ratio. For a significant application of this modulation technique, the receiving antenna must be large enough to register simultaneous spatial variations in the incident electromagnetic fields.

DD FORM 1 NOV 65 1473 (PAGE 1)

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
information theory						
digital communication						
parallel communication channels, error bounds and block coding						
antenna radiation patterns						
optical communication						
resolution						
degrees of freedom, spatial						