Technical Report | 344

T. Hagfors

Information Capacity and Quantum Effects in Propagation Circuits

24 January 1964

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

INFORMATION CAPACITY AND QUANTUM EFFECTS IN PROPAGATION CIRCUITS

T. HAGFORS

Group 31

TECHNICAL REPORT 344

24 JANUARY 1964

LEXINGTON   MASSACHUSETTS

ABSTRACT

In this report we try to establish an upper bound on the information capacity of an electromagnetic wave propagation circuit between two apertures when quantum effects come into play. We first discuss the specifications of a signal, and find that we can associate approximately BT quantities with a signal of bandwidth B and duration T, and that the energy associated with these quantities must differ by at least $\hbar\omega$. We then consider radiative attenuation loss within the framework of quantum theory and find that the radiative attenuation introduces a random noise in the circuit which is analogous to partition tube noise.

With the channel matrix determined, we discuss the problem of evaluating the channel capacity under average and peak power constraints. Due to mathematical difficulties we are only able to establish the channel capacity explicitly when the radiative losses become extreme. In this limiting case — as a by-product — we also find that the best encoding procedure is a binary on — off system.

This technical documentary report is approved for distribution.
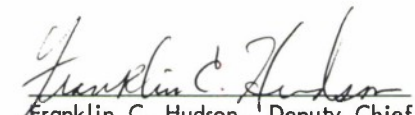
Franklin C. Hudson, Deputy Chief
Air Force Lincoln Laboratory Office

iii

TABLE OF CONTENTS

INFORMATION CAPACITY AND QUANTUM EFFECTS
IN PROPAGATION CIRCUITS

I. INTRODUCTION

The recent development of various types of coherent oscillators and amplifiers in the infrared and the optical frequency regions seems to bear great promise for certain communication and radar applications. It is immediately obvious that quantum effects will become of considerable importance in these frequency regions because of the proportionality of quantum energy and frequency. Any discussion of communication and information capacity at these frequencies must therefore take into account the full effects of the quantized energy of the waves.

Some work has been done on quantum effects in communication systems, usually with emphasis on the behavior of the various elements at the terminals of the system. In particular, the problems relating to the receiving end and the relative merits of quantum counters, linear amplifiers, etc., have been widely discussed.[1-3]

Effects relating to the transmission path itself have received much less attention. Recent works by Heffner[4] and Haus[5] have reported certain progress along this line. However, we want to consider further the fundamental problem of the transmission of information on quantized electromagnetic waves from a transmitting aperture A to a receiving aperture B as shown in Fig. 1. Because the actual propagation path is our main interest, we shall only discuss the relation between the information which is encoded on the electromagnetic wave passing out of the transmitting aperture and the information which is present on the wave passing into the receiving aperture. Questions relating to the encoding and decoding procedures will be left aside in this report. Our problem is one of information transmission from aperture to aperture.

The first question that arises is concerned with the amount of information which can be carried by an electromagnetic wave of given bandwidth and center frequency for a single polarization. In an early paper Stern[6] assumed that a message of bandwidth B and duration T could be specified by 2TB numbers, as in the classical picture. In a later work Gordon[7] argued that only half
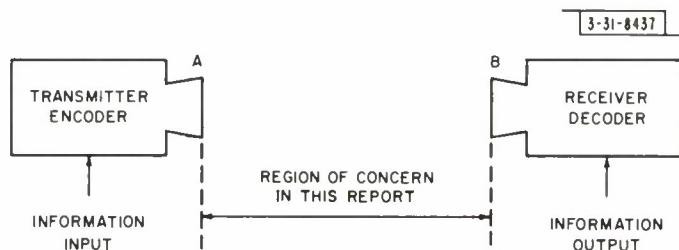

Fig. 1. Communication system.

1

as many numbers could be associated with a message. The reason for this is intuitively clear: The classical specification of a bandpass limited signal involves the in-phase and the quadrature components at a rate of B. In a quantum mechanical formulation, however, these two components must be regarded as noncommuting variables and only one of them can be specified exactly. In this report it is pointed out that even Gordon's ideas about the number of degrees of freedom of a message might be somewhat optimistic.

In previous work the "wave capacity" was derived for a plane wave only. This derivation does not leave room for considerations of radiative attenuation effects and therefore cannot be applied directly to practical communication situations. A transmitting aperture can be made to excite several independent spatial modes, none of which are plane waves. The distinction between plane and spherical waves is unimportant in certain respects. But it must be remembered that the receiving antenna usually couples weakly into one of the spatial modes set up by the transmitting antenna — unless the transmitter-receiver separation is very small. This weak coupling between the two apertures gives rise to a phenomenon which can be interpreted as partition noise. It will be shown that this noise has a rather profound effect on the capacity of the quantum circuit, particularly when the received power level is exceedingly low. In Gordon's work, the sequence of quantum numbers specified at the transmitting end reappears as the same sequence of numbers at the receiving end because of the lack of partition noise. When partition noise is important, however, the specification of a sequence of quantum numbers at the transmitter can only serve to define a set of distribution functions for the quantum numbers at the receiving end, and hence, even in the absence of additive noise, the channel must be described in terms of a matrix which exhibits noise properties.

In what follows, we first discuss the specification of a signal or a message fed into a single spatial mode (single mode includes only one polarization). We then go on to discuss the concept of spatial modes and evaluate the number of such modes which can be excited by a given transmitting aperture. The coupling between the transmitting and the receiving apertures is then established in such a way that quantization of the system can be carried out rather easily. This quantization procedure leads to a specification of the communication channel in terms of a channel matrix. The form of this channel matrix is also worked out for the case of additive noise being present.

With the form of the channel matrix established, the channel capacity is determined through a familiar variational procedure. It is shown that this procedure breaks down when partition noise is present. In fact, the solution of the variational problem leads to certain negative input probabilities for the optimum encoding procedure. This situation has previously been considered by Muroga[8] who has shown that certain of the input "symbols" (here quantum states) must be omitted. As there does not seem to be other than a trial-and-error procedure to the solution of the capacity problem in this case, we are unable to establish the capacity of the channel except in the case of extreme radiative attenuation when the channel becomes binary. In spite of the lack of a general result for the channel capacity, it is possible to draw certain conclusions about the type of encoding procedure which must be employed to optimize the transfer of information.

II. SPECIFICATION OF A SIGNAL

In information theory, a continuous signal is shown to be equivalent to a certain number of degrees of freedom determined by the duration, T, and the bandwidth, B, of the signal. The number of degrees of freedom is shown to be approximately equal to 2TB.[9,10] It is understood

that either B or T can be sharply defined, not both simultaneously. There are several ways in which 2TB quantities can be associated with a signal, the most familiar ones being by means of sampled values either in the time domain or in the frequency domain.

A bandpass limited time function, f(t), can be expressed as

$$f(t) = f_c(t) \cdot \cos \omega_o t + f_s(t) \sin \omega_o t \quad , \tag{1}$$

where ω_o is the center frequency of the pass band, and where $f_c(t)$ and $f_s(t)$ are time functions whose spectra differ from zero only when $|\omega| < \pi B$, where B is the signal bandwidth. The two functions $f_c(t)$ and $f_s(t)$ may be represented as

$$\left.\begin{array}{l} f_c(t) = \sum_k f_{ck} \dfrac{\sin \pi B(t - t_k)}{\pi B(t - t_k)} \\[2em] f_s(t) = \sum_k f_{sk} \dfrac{\sin \pi B(t - t_k)}{\pi B(t - t_k)} \end{array}\right\} \quad \text{where } t_k = \dfrac{k}{B} \quad , \tag{2}$$

and where f_{ck} and f_{sk} are the values of $f_c(t)$ and of $f_s(t)$, respectively, at the times $t = t_k$. At each sampling point it is therefore possible to specify two quantities, and hence 2B quantities per unit time. For a message of approximate duration T, the number of quantities required to specify the complete signal is therefore 2TB. It should be noted that the particular sampling functions used in (2) ensure that the sampling functions in the sums are completely independent at the sampling points.

A similar situation exists when we consider sampling in the frequency domain. A strict limitation of the signal in time leads to a sampling theorem for the frequency domain. In this case it turns out to be possible to specify 2T independent quantities (i.e., amplitude and phase) per unit bandwidth and therefore 2TB quantities in all.

When we turn to a quantum approach to the signal description, we have to make sure that we are actually dealing with observable quantities. In the classical theory it is possible to devise "thought experiments" whereby the sampled values actually appear as observables. This, however, requires measuring quantities with infinitesimal energy, and the sampling procedure therefore does not appear to be directly applicable in quantum theory. When making an observation in the quantum sense, there must always be a finite energy associated with the observed quantity. For this reason we have to replace the sampled values by some sort of average over a space or a time interval.

One way to overcome this difficulty is to interpret the time-bandwidth product in a more lax way than is usual in the classical signal theory, and assume that a signal of approximate duration T is represented in the following manner

$$f(t) = \frac{\sin \frac{\pi}{T} t}{\frac{\pi}{T} t} \sum_{\ell_1}^{\ell_2} (a_\ell \cos \omega_\ell t + b_\ell \sin \omega_\ell t) \quad , \tag{3}$$

where the increments in ω_ℓ are given by $(2\pi/T)$. With this particular choice of signal representation, we see that the spectra of the various elementary signals corresponding to different values of ℓ no longer overlap, and we have therefore been able to produce a set of functions which,

even when observed in a quantum sense, do not have to interfere. The time function represented by (3) is sharply defined in the frequency domain where the upper and the lower frequency limits are

$$f_{max} = \frac{1}{T}(\ell_2 + \frac{1}{2}) \quad ,$$

$$f_{min} = \frac{1}{T}(\ell_1 - \frac{1}{2}) \quad . \tag{4}$$

Introducing the signal bandwidth through $B = f_{max} - f_{min}$, we find that the number of independent elementary signals which can be associated with the total signal is

$$N = B \cdot T \quad .$$

Again each of the elementary signals must be specified by two quantities, such as a_ℓ and b_ℓ in (3), and the number of degrees of freedom of the total signal is the same as in the conventional theory. In the type of representation in (3), however, the fraction of the power which lies outside of the nominal time-bandwidth region BT is given by

$$\eta = \frac{2}{\pi} \int_{\pi}^{\infty} \frac{\sin^2 x}{x^2} \, dx = 0.097 \quad . \tag{5}$$

In the conventional sampling procedure, this factor would be reduced by another factor of the order of $1/BT$. The definition of the signal in the way suggested by (3) therefore makes elementary signals ill-defined in the time domain. A slight improvement appears to be possible by using instead of the $\sin x/x$ modulation a prolate spheroidal wave function.[11] This refinement will make very little difference in our conclusions and we shall avoid introducing this complication.

One of several alternative ways of defining noninteracting elementary signals is to make them sharply defined in time and then construct a composite signal from a sequence of elementary signals. In this case the frequency definition would be poor and the energy associated with elementary signals would be uncertain. In this report we shall find (3) most convenient as a description of an elementary signal.

Having defined a set of elementary signals, we must next inquire about their description in the language of quantum theory. In particular we must ask about the accuracy with which the a_ℓ and the b_ℓ can be specified, and whether they can be specified independently. For this purpose we imagine that the time function (3) has excited a single spatial mode which, for simplicity, we take to be a plane wave mode with direction of propagation along the z-axis. Because of our choice of elementary signals, all quantities such as Lagrangians and Hamiltonians can be expressed as sums over contributions from the elementary signals. From now on, therefore, we omit the subscripts ℓ, and use ω_o for the center frequency of the particular elementary signal under consideration. The detailed quantization of elementary signals is carried out in Appendix A. An elementary signal is assumed to be modulated onto a plane wave traveling in the z-direction and polarized along the x-axis. In Appendix A the wave is represented by the x-component of the vector potential in the following form

$$A_x(z,t) = \frac{\sin \frac{\pi}{T}(\frac{z}{c} - t)}{\frac{\pi}{T}(\frac{z}{c} - t)} \, [\alpha \, \cos \omega_o(\frac{z}{c} - t) - \beta \, \sin \omega_o(\frac{z}{c} - t)] \quad . \tag{6}$$

4

In spite of the fact that this is a nonharmonic signal, it is shown that the two amplitude factors α and β, which contain all the information about an elementary signal, can be combined into complex amplitudes a and a^* which obey the same commutation relations as the complex amplitudes of an harmonic oscillator. Provided the detection of the elementary signals is carried out the right way, it is shown that α and β can be specified only to an accuracy corresponding to one quantum of energy $\hbar\omega_o$.

We have therefore shown that a signal of approximate time-bandwidth product BT = N can be specified by N degrees of freedom, each with a minimum energy difference corresponding to one quantum $\hbar\omega_o$. This conclusion appears to be in complete agreement with the assumption made by Gordon[7] concerning the number of states which can be associated with a signal.

III. SPATIAL MODES AND PARTITION EFFECTS

In the previous section we discussed the quantization of a signal impressed on a single plane wave mode. In practice it is not possible to excite plane wave modes from an antenna of finite extent, and we must examine the nature of the spatial modes which can be excited by an aperture of finite extent.

Again we imagine that the fields are described by a vector potential satisfying the wave equation:

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \vec{A}(\vec{r}, t) = 0 \quad , \tag{7}$$

and

$$\nabla \vec{A} = 0 \quad . \tag{8}$$

It follows that a complete specification of the transverse components of $\vec{A}$ and their normal derivatives over an aperture will completely specify the fields in front of the aperture plane. For simplicity, we take the transmitting aperture to be in the xy-plane and to be rectangular (Fig. 2).
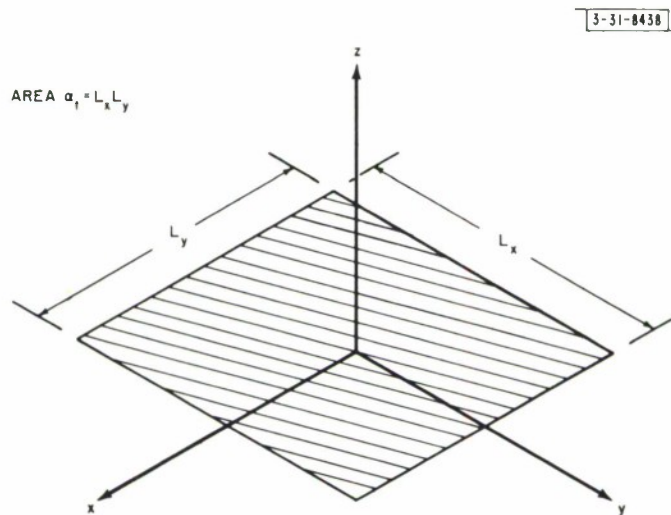

Fig. 2. Transmitting aperture over which fields are prescribed.

We now imagine that the transverse components of the vector potential are specified over the aperture, and that the time variation of the components is in agreement with that of one of the elementary signals discussed in the previous section. A general representation of the x and the y components of the vector potential in the half-space $z > 0$ is

$$A_{x,y}(\vec{r}, t) = \left(\frac{1}{2\pi}\right)^2 \frac{1}{\Delta k} \int_{k_o-(\Delta k/2)}^{k_o+(\Delta k/2)} dk \iint dk_x\, dk_y \{A_{x,y}(k_x k_y k)\, \exp[i(\vec{k}\vec{r} - kct)]$$

$$+ A^*_{x,y}(k_x k_y k)\, \exp[-i(\vec{k}\vec{r} - kct)]\} \quad . \tag{9}$$

The field on the aperture is found from this by putting $z = 0$.

$$A_{x,y}(x, y, o, t) = \frac{1}{\Delta k} \int_{k_o-(\Delta k/2)}^{k_o+(\Delta k/2)} dk \{B_{x,y}(x, y, k)\, \exp[-ikct]$$

$$+ B^*_{x,y}(x, y, k)\, \exp[+ikct]\} \quad , \tag{10}$$

where the quantities $A_{x,y}(k_x k_y k)$ are related to the $B_{x,y}(x, y, k)$ through the relations

$$A_{x,y}(k_x k_y k) = \iint_{\text{aperture}} dx\, dy\, B_{x,y}(x, y, k)\, \exp[-i(xk_x + yk_y)] \quad . \tag{11}$$

Because the fields over the transmitting aperture contain no structure larger than ℓ_x in the x-direction and ℓ_y in the y-direction, it follows that $A_{x,y}(k_x k_y k)$ is completely specified provided it is given at certain particular values of k_x and k_y given by

$$k_x = \frac{2\pi r}{\ell_x} \qquad r = 0, \pm 1, \pm 2, \ldots \quad ,$$

$$k_y = \frac{2\pi s}{\ell_y} \qquad s = 0, \pm 1, \pm 2, \ldots \quad .$$

In analogy with the sampling theorem of Shannon[9] or the aperture synthesis procedure introduced by Woodward and Lawson,[15] we therefore may put

$$A_{x,y}(k_x k_y) = \sum_r \sum_s A_{x,y}\left(\frac{2\pi r}{\ell_x}, \frac{2\pi s}{\ell_y}\right) \frac{\sin \frac{1}{2}\ell_x\left(k_x - \frac{2\pi}{\ell_x}r\right)}{\frac{1}{2}\ell_x\left(k_x - \frac{2\pi}{\ell_x}r\right)} \cdot \frac{\sin \frac{1}{2}\ell_y\left(k_y - \frac{2\pi}{\ell_y}s\right)}{\frac{1}{2}\ell_y\left(k_y - \frac{2\pi}{\ell_y}s\right)} \quad . \tag{12}$$

where we have omitted k in the argument of $A_{x,y}$ because the field amplitudes are independent of frequency within the frequency range of an elementary signal. Only an enumerably infinite set of modes can therefore be excited by an aperture of limited extent. Of these, the modes corresponding to

$$\left(\frac{2\pi r}{\ell_x}\right)^2 + \left(\frac{2\pi s}{\ell_y}\right)^2 > k^2 \tag{13}$$

will, however, not be able to propagate away from the aperture. They correspond to evanescent waves. This leaves only a finite number of possible propagating modes. These will be the only ones of interest when it comes to discussion of information transfer between distant apertures.

It is instructive to determine the number of independent spatial modes corresponding to traveling waves which can be excited by the transmitting aperture. The solid angle occupied by each spatial mode is given by

$$\Delta\Omega_{mode} = \frac{4\pi^2}{\ell_x \ell_y k_o^2 \cos\Theta_{r,s}} \quad , \tag{14}$$

where $\Theta_{r,s}$ is the angle between $\vec{k}_{r,s}$ and the z-direction. The number of propagating modes per solid angle is the inverse of this, and the total number of propagating modes therefore becomes

$$N = \int_{semi\text{-}space} \frac{d(\vec{\Omega})}{\Delta\Omega_{mode}} = \frac{\ell_x \ell_y}{4\pi} \cdot k_o^2 \quad , \tag{15}$$

which is seen to be one-quarter of the maximum gain of the aperture. Each of these spatial modes can of course support two orthogonal polarizations of the waves (circular, linear, or other types).

On the transmitting end we can therefore excite only a limited number of spatial modes, each in two different orthogonal polarizations. The elementary signals impressed on the various spatial modes can be specified in exactly the same manner as was done for the plane waves of the previous section. The artificial limiting of the cross section of the plane wave is now replaced by a natural limitation caused by the finite lateral extent of each of the spatial modes.

Next we must discuss the coupling of the receiving aperture into the fields excited by the transmitting antenna. We first carry through a discussion along completely classical lines before we go on to a quantum description of the phenomena.

Denote the Poynting vector set up by a particular transmitting mode specified by (r, s, μ) by $\vec{S}_{r,s}^{(\mu)}$. The parameter μ signifies the state of polarization, and the parameters r, s the direction. The maximum power which can be absorbed by the receiving antenna becomes

$$P_{(r,s)}^{(\mu)} = \iint_{aperture} dx'dy' \, \vec{n} \cdot \vec{S}_{r,s}^{(\mu)} \, g^{(\mu)}(x',y') \quad . \tag{16}$$

The integration extends over the whole of the receiving aperture; $\vec{n}$ is the normal to the aperture surface and $g^{(\mu)}(x',y')$ is a weight factor relating to polarization and to the particular type of antenna employed. The total power put into the (r, s, μ) mode by the transmitter (index t) is of the form

$$P_{(r,s)\,t}^{(\mu)} = \int \vec{S}_{(r,s)}^{(\mu)} \, d\vec{a} \quad . \tag{17}$$

We can now introduce the ratio of aperture power to total power for each spatial mode (r, s, μ)

$$\nu_{r,s}^{(\mu)} = \frac{P_{rs}^{(\mu)}}{P_{rs}^{(\mu)\,t}} \quad . \tag{18}$$

This number of course is always less than or equal to unity. When one or more of the r's is equal to unity, power may be transferred from the transmitting to the receiving aperture without loss. This is essentially what is implied in most earlier work on quantum communication.[7]

In order to see when such close coupling between transmitting and receiving apertures is possible, we may argue as follows: The opening angle corresponding to each spatial mode is given approximately by

$$\Delta\Omega = \frac{\lambda^2}{\alpha_t \cos\Theta_t} \quad , \tag{19}$$

where λ is the wavelength, α_t the transmitting aperture area and Θ_t the angle between the beam direction and the normal to the aperture. The area of the beam at the receiver, located a distance R away from the transmitter, becomes

$$a_{beam} = \Delta\Omega \cdot R^2 \quad . \tag{20}$$

If this area is smaller than the projected area of the receiving aperture, i.e., $\alpha_r \cos\Theta_r$, then a complete power transfer can take place. The condition for this is seen to be

$$R < \sqrt{\frac{\alpha_t \alpha_r \cos\Theta_r \cos\Theta_t}{\lambda^2}} \quad . \tag{21}$$

As an illustration, consider the following numerical example: $\alpha_r = \alpha_t = 1\,\mathrm{cm}^2$, $\lambda = 6000\,\mathring{A}$. Substitution into the inequality shows that R is less than $166\,\mathrm{m}$. In any practical communication circuit, therefore, we must be prepared to use only one of the possible spatial modes which can be excited at the transmitter, and we must also take full account of radiative attenuation effects in the quantum description of the circuit. As we shall see, this leads us to a partition noise effect which will cause a serious decrease in information capacity of a communication channel when the received power is extremely low. Such partition effects have been described previously by Haus[5] and it is implicit in some work by Heffner.[4]

Let us next try to study the relation between the transmitted and the received signal in terms of the ideas developed above. We consider only one particular transmitted mode, namely that corresponding to constant phase across the aperture, only one of the two possible polarizations, and one of the elementary signals corresponding to the center frequency ω_o.

The fields over a distant plane $z = R$ in front of the transmitting antenna, and $\vec{R}$ denotes a point on this plane. The fields over this reference plane, or receiving aperture plane, are

$$A(\vec{R}, t) = (\frac{1}{2\pi})^2 \frac{1}{\Delta k} \int_{k_o - (\Delta k/2)}^{k_o + (\Delta k/2)} dk \iint dk_x dk_y\, F_t(k_x k_y)\, \{B_t \exp[i(\vec{k}\vec{R} - kct)]$$

$$+ B_t^* \exp[-i(\vec{k}\vec{R} - kct)]\} = B_t g_t(R, t) + B_t^* g_t^*(\vec{R}, t) \quad . \tag{22}$$

Here B_t is the mode amplitude; see Eqs. (11) and (12). $F_t(k_x k_y)$ is defined by

$$F_t(k_x k_y) = \frac{\sin\frac{1}{2}k_x \ell_x}{\frac{1}{2}k_x \ell_x} \frac{\sin\frac{1}{2}k_y \ell_y}{\frac{1}{2}k_y \ell_y} \alpha_t \quad , \tag{23}$$

and the functions $g_t(\vec{R}, t)$ and $g_t^*(\vec{R}, t)$ are just shorthand for the integrals in (22). Now the fields in all space can also be derived from a knowledge of the fields over the receiving aperture plane $z = R$.

However, we divide the receiving aperture plane into two parts: one part coinciding with the receiving aperture, and the other part consisting of the rest of the aperture plane (Fig. 3). The new description of the fields therefore contains two terms, one corresponding to either of the two
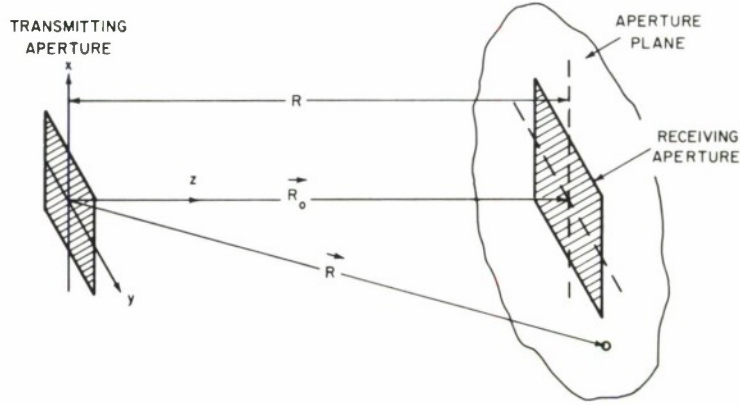
Fig. 3. Transmitting and receiving apertures with receiving aperture plane.

parts of the receiving aperture plane. In the new representation with two terms, we therefore obtain expressions of the form

$$A_r(\vec{r}, t) = (\frac{1}{2\pi})^2 \frac{1}{\Delta k} \int_{\Delta k} dk \iint dk_x dk_y \{ F_r(k_x k_y) B_r \exp[-i\vec{k}\vec{R}_o] \exp[i(\vec{k}\vec{r} - kct)]$$

$$+ F_r^*(k_x k_y) B_r^* \exp[i\vec{k}\vec{R}_o] \exp[-i(\vec{k}\vec{r} - kct)] \} \quad , \tag{24}$$

which corresponds to the receiving aperture mode, and a similar expression for $A_\ell(\vec{r}, t)$ corresponding to the mode derived from the rest of the receiving aperture plane. We are now in the position to quantize the system and we interpret the parameters B_t, B_r and B_ℓ [corresponding to $A_\ell(\vec{r}, t)$] as operators; and, furthermore, introduce creation and annihilation operators as in Appendix A. The following equations between the various operators are obtained:

$$a_t = P_r a_r + P_\ell a_\ell \quad ,$$

$$a_t^* = P_r^* a_r^* + P_\ell^* a_\ell^* \quad . \tag{25}$$

The parameters P_r and P_ℓ can be derived directly from the above equations (Appendix B). They can, however, also be derived from physical arguments as we shall show shortly. In what follows, we refer to the transmitted signal system as the t-system, the system derived from the receiving aperture field distribution as the r-system, and the system representing the lost radiation as the ℓ-system.

The t-system is specified by means of a set of eigenvectors of the operator $a_t^* a_t$, and the r- and ℓ-systems by means of eigenvectors of the operators $a_r^* a_r$ and $a_\ell^* a_\ell$, respectively. Because the r- and ℓ-systems are independent, a state of the t-system can quite generally be expressed as follows[16]

$$|N\rangle_t = \sum_n \sum_m C_{n,m}^{(N)} |n\rangle_r |m\rangle_\ell \quad , \tag{26}$$

9

where the parameters $C_{n,m}^{(N)}$ are to be determined. It is well known that the various state vectors can be expressed by means of the state vector $|0\rangle_t$ through the relation

$$|N\rangle_t = \frac{a_t^{*N}}{\sqrt{N!}} \, |0\rangle_t \quad , \tag{27}$$

where a_t^* is the creation operator of the system, here the t-system. If the same expansion is used for the r- and the ℓ-systems as well, the following relation is obtained

$$|N\rangle_t = \frac{a_t^{*N}}{\sqrt{N!}} \, |0\rangle_t = \sum \sum C_{n,m}^{(N)} \frac{a_r^{*n} a_\ell^{*m}}{\sqrt{n!m!}} \, |0\rangle_r \, |0\rangle_\ell \quad . \tag{28}$$

Because the empty state of the t-system must correspond to both r- and ℓ-systems being in the ground state (no additive noise), we conclude that

$$|0\rangle_t = |0\rangle_r \, |0\rangle_\ell \quad . \tag{29}$$

This means that the following identity must be fulfilled by the operators

$$a_t^{*N} = \sum_n \sum_m C_{n,m}^{(N)} \frac{\sqrt{N!}}{\sqrt{n!m!}} \, a_r^{*n} a_\ell^{*m} \quad . \tag{30}$$

A similar relation between the operators can, however, be derived directly from the latter of Eq. (25).

$$a_t^{*N} = \sum_n \sum_m \frac{N!}{n!m!} \, P_r^{*n} P_\ell^{*m} \, \delta_{m,N-n} \, a_r^{*n} a_\ell^{*m} \quad . \tag{31}$$

Because the two operator equations must be identical, we conclude that

$$C_{n,m}^{(N)} = \sqrt{\frac{N!}{n!m!}} \, P_r^{*n} P_\ell^{*m} \, \delta_{m,N-n} \quad . \tag{32}$$

We notice that $m + n = N$, which expresses the energy conservation of the system. Hence, the expansion of (26) simplifies to

$$|N\rangle_t = \sum_{n=0}^{N} \sqrt{\frac{N!}{n!(N-n)!}} \, P_r^{*n} P_\ell^{*N-n} |n\rangle_r \, |N-n\rangle_\ell \quad . \tag{33}$$

From the normalization condition, it is obvious that

$$1 = (|P_r|^2 + |P_\ell|^2)^N \quad ,$$

hence,

$$|P_r|^2 + |P_\ell|^2 = 1 \quad . \tag{34}$$

The transition probabilities, i.e., the probabilities of reception of state $|n\rangle_r \, |N-n\rangle_\ell$ when the transmitted signal is in state $|N\rangle_t$, is given by the square of the projection of $|n\rangle_r \, |N-n\rangle_\ell$ upon $|N\rangle_t$,

$$p(n|N) = |\,_r\langle n|\,_\ell\langle N-n|N\rangle_t|^2 = \binom{N}{n} |P_r|^{2n} |P_\ell|^{(N-n)2} \quad . \tag{35}$$

The significance of $|P_r|^2$ is seen by considering the transition

$$p(1|1) = |P_r|^2 = \nu \quad . \tag{36}$$

Hence, ν must be interpreted as the probability of receiving one photon when one was sent. This, of course, is equal to the transmission loss of the propagation circuit, or the ratio of received to transmitted power. Because of (34), we then have for the transition probability of the propagation circuit

$$p(n|N) = \binom{N}{n} \nu^n (1-\nu)^{N-n} \quad , \tag{37}$$

which can also be derived from purely intuitive arguments.[†] As is seen from Appendix B, the above interpretation of the P_r and P_ℓ is completely borne out by the more laborious calculations presented there. The transition probability $p(n|N)$ is seen to have the appearance of containing partition effects. The fact that a distinct transmitted signal might be received as a number of different signals means that the partition effects make the propagation circuit noisy even in the absence of additive noise.

IV. PROPAGATION CIRCUIT AS A COMMUNICATION CHANNEL

In the previous two sections we have discussed the breaking up of a signal into independent elementary signals, and we have established a transition probability for elementary signals. The complete transition probability is identical to the product of the elementary signal probabilities. In what follows, the quality of the propagation circuit will be evaluated in terms of the mutual information between transmitted and received signals. In the optimization problem it will be assumed that there is an average power constraint on the total signal. If the mutual information of elementary signals does not increase more rapidly than linearly with the signal energy, then the average energy should be distributed evenly among the elementary signals (assuming the relative bandwidth of the total signal to be small). It is therefore sufficient to solve the optimization problem for a single elementary signal and afterwards make sure that the mutual information associated with an elementary signal does not increase more strongly than linearly with the elementary signal mean energy. We begin by defining the channel matrix for the elementary signal channel first without additive noise, then with noise. We go on to optimize the mutual information under average power constraint both with and without additive noise, and we show that the results become identical to those of Gordon[7] for the case of $\nu \to 1$, i.e., when there is no partition noise. It is then shown that the results are improper whenever $\nu < 1$ because the familiar method with Lagrange's multipliers employed in the general analysis does not ensure that all probabilities are positive. In order to understand what happens, we construct certain simple channels and discuss them in detail. From this particular discussion, it is possible to generalize the results and draw conclusions about the channel capacity in the case of extremely weak received signals.

[†] This is also of the same form as the result derived by Shimodo, et ol.,[1] for attenuotion through obsorption processes.

A. Definition of Elementary Signal Channel Matrix

As derived in the previous section, the transition probability between signal state, N, at the transmitter and signal state, n, at the receiver is

$$p(n|N) = \binom{N}{n} \nu^n (1 - \nu)^{N-n} \quad , \tag{38}$$

i.e., a binomial or Bernoulli distribution. We note here that the number of received quanta is always less than or equal to the number transmitted. If the transition probability is expressed in matrix form, therefore, the matrix becomes triangular. The assumption that the transmitted signal can be specified to an exact number of quanta might seem rather unrealistic. In practice, however, when partition effects are important, the conditional probability (38) is rather insensitive to changes in N, so it is sufficient to specify N approximately. None of the conclusions in this paper will be affected by a "diffuse" specification of the transmitted signal.

We now turn to the question of additive noise in the path. The existence of additive noise will no longer guarantee that fewer quanta arrive than were transmitted and the transition matrix becomes less simple. If the noise is of the equilibrium type associated with the temperature, T, then matters are rather straightforward. The probability of j quanta being present in a mode becomes

$$p_0(j) = (1 - e^{-\Theta}) \exp[-j\Theta] \quad \text{with } \Theta = \frac{\hbar\omega}{KT} \quad , \tag{39}$$

and where K is Boltzmann's constant. When noise is included, the transition probability becomes

$$p'(n|N) = \sum_{z=0}^{n} p(z|N) \, p_0(n - z) \quad . \tag{40}$$

For later purposes, we shall express this in terms of a matrix product. The noiseless matrix is then termed A, with elements a_{ij}. The noise matrix is termed B, with elements b_{ij}.[†] For the noiseless case, $p(j|i) = a_{ij}$,

$$A = \begin{Bmatrix} p(0|0) & 0 & 0 & 0\ldots \\ p(0|1) & p(1|1) & 0 & 0\ldots \\ p(0|2) & p(1|2) & p(2|2) & 0\ldots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{Bmatrix} \cdot$$

If, in the noise case, the same meaning is to be attached to the various elements of the resultant transition matrix A', we must arrange the noise terms as follows:

---

† Note thot in oll our motrices we count lines and columns not from unity, os usuol, but from zero.

$$B = \begin{Bmatrix} p_o(0) & p_o(1) & p_o(2)\ldots \\ \\ 0 & p_o(0) & p_o(1)\ldots \\ \\ 0 & 0 & p_o(0)\ldots \\ . & . & . \\ . & . & . \\ . & . & . \end{Bmatrix} \quad ,$$

and

$$A' = AB \quad .$$

In passing, we observe that the inverse matrix can be expressed in terms of the inverse of the elementary matrices

$$(A')^{-1} = B^{-1}A^{-1} \quad .$$

B. Optimization with Lagrange Multipliers

In this section we first carry out an optimization of the mutual information as defined by Fano[10] for a general transition probability $p(n|N)$ which may or may not include noise. The calculation is carried out with an average power constraint at the transmitting aperture. In this situation our choice lies in the selection of a proper distribution function for the occupation number, N, at the transmitter, $p_t(N)$, or a distribution function $p_r(n)$ at the receiving end. After the general solution to this problem has been found, we go on to evaluate the noiseless capacity and the capacity with noise present (formally only!).

The mutual information for an elementary signal is defined as

$$I(N;n) = H(n) - H(n|N)$$

$$= \sum_{n=0}^{\infty} p_r(n) \log p_r(n) + \sum_{N=0}^{\infty} \sum_{n=0}^{\infty} p_t(N) \, p(n|N) \log p(n|N)$$

$$= - \sum_{n=0}^{\infty} \sum_{N=0}^{\infty} p_t(N) \, p(n|N) \, \log \sum_{N=0}^{\infty} p_t(N) \, p(n|N)$$

$$+ \sum_{n=0}^{\infty} \sum_{N=0}^{\infty} p_t(N) \, p(n|N) \, \log p(n|N) \quad . \tag{41}$$

This is to be maximized subject to the two constraints on the unknown distribution function $p_t(N)$

$$\sum_{N=0}^{\infty} p_t(N) = 1 \qquad \text{normalization} \quad ,$$

$$\sum_{N=0}^{\infty} N p_t(N) = \overline{N} \qquad \text{average power constraint} \quad .$$

Using the familiar methods of Lagrange, with the constraints included after multiplication by indeterminate factors, we have to form the expression

$$\frac{\partial}{\partial p_t(N)} \left[I(N; n) - \alpha' \sum_N p_t(N) - \beta \sum_N p_t(N) \right]$$

and equate it to zero. This leads us to the equations

$$\sum_{n=0}^{\infty} \left\{ p(n|N) \left[1 + \log p_r(n) - \log p(n|N) \right] \right\} + \alpha' + \beta \cdot N = 0 \quad . \tag{42}$$

Both sides of the equation are now multiplied by $p_t(N)$ and a summation is carried out over N. This gives

$$- \sum_{n=0}^{\infty} p_r(n) \log p_r(n) + \sum_{N=0}^{\infty} \sum_{n=0}^{\infty} p_t(N) p(n|N) \log p(n|N) = 1 + \alpha' + \beta \overline{N} = \alpha + \beta \overline{N} \quad . \tag{43}$$

But this is, at least formally, the maximum mutual information

$$C_M = \alpha + \beta \overline{N} \quad . \tag{44}$$

The real problem now lies in the determination of the Lagrange multipliers α and β. To do this, we have to evaluate the distribution function $p_r(n)$ or $p_t(N)$ and use the constraint equations. To carry out this, we proceed as follows. Rearranging Eq. (42), we have

$$\sum_{n=0}^{\infty} p(n|N) \log p_r(n) = \sum_{n=0}^{\infty} p(n|N) \left[\log p(n|N) - \alpha - \beta \right] \quad , \tag{45}$$

since

$$\sum_{n=0}^{\infty} p(n|N) = 1 \quad .$$

We therefore have an infinite set of linear equations in the unknowns $x_n = \log p_r(n)$, one equation for each value of N. Regarding x_n as the nth component of a column vector, and also defining a column vector by the components

$$d_N = \sum_{n=0}^{\infty} p(n|N) \left[\log p(n|N) - \alpha - \beta N \right] \quad , \tag{46}$$

we may write the set of equations as follows

$$A'\vec{x} = \vec{d} \quad , \tag{47}$$

where A' is the matrix form of the transition probability introduced previously. The formal solution to the problem therefore is

$$\vec{x} = A'^{-1}\vec{d} \quad \text{or} \quad x_n = \sum_{N=0}^{\infty} (A'^{-1})_{n,N} d_N \quad . \tag{48}$$

For the probability distribution at the receiving end we therefore have

$$p_r(n) = \exp\left[\sum_{N=0}^{\infty} (A'^{-1})_{n,N}\, d_N\right] \quad . \tag{49}$$

There are still the two unknowns α and β to be determined, and two equations are needed. These are found from the constraint equations in a slightly modified form, <u>viz</u>.,

$$\sum_{n=0}^{\infty} p_r(n) = 1 \quad , \quad \sum_{n=0}^{\infty} n p_r(n) = \overline{n} \quad . \tag{50}$$

In the noiseless case, we are first of all faced with the problem of inverting the lower tri-angular infinite matrix

$$A \;=\; \begin{Bmatrix} 1 & 0 & 0 & \cdots \\[2ex] (1-\nu) & \nu & 0 & \cdots \\[2ex] (1-\nu)^2 & 2\nu(1-\nu) & \nu^2 & \cdots \\[2ex] (1-\nu)^3 & 3\nu(1-\nu)^2 & 3\nu^2(1-\nu) & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ (1-\nu)^m & m(1-\nu)^{m-1} & \binom{m}{2}\nu^2(1-\nu)^{m-2} & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{Bmatrix}$$

To avoid going into detail, we just state the result. The general term is of the form

$$(A^{-1})_{i,j} = \binom{i}{j}(1-\nu)^{i-j}(-1)^{i-j}\nu^{-i} \quad \left[\text{note that } \binom{i}{j} = 0 \text{ for } j > i\right] \quad , \tag{51}$$

or explicitly

$$A^{-1} \;=\; \begin{Bmatrix} 1 & 0 & 0 & \cdots \\[2ex] -(1-\nu)\,\nu^{-1} & +\nu^{-1} & 0 & \cdots \\[2ex] +(1-\nu)^2\,\nu^{-2} & -2(1-\nu)\,\nu^{-2} & +\nu^{-2} & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{Bmatrix}$$

With these expressions, we obtain

$$x_n = \log p_r(n) = \sum_{N=0}^{\infty} (A^{-1})_{n,N} \sum_{i=0}^{\infty} p(i|N) \, [\log p(i|N) - \alpha - \beta N]$$

$$= -\alpha - \frac{n}{\nu} \, [\beta - \nu \log \nu - (1 - \nu) \log (1 - \nu)] + \sum_B \sum_i (A^{-1})_{n,N} \, p(i|N) \log \binom{N}{i} \quad . \qquad (52)$$

For the particular case of $\nu = 1$, this reduces to a result identical to that of Gordon[7]

$$p_r(n) = \exp [-\alpha - n\beta] \quad . \qquad (53)$$

Substitution into (50) gives

$$\alpha = \log (1 + \bar{n}) \quad ,$$

$$\beta = \log (1 + \bar{n}) - \log \bar{n} \quad , \qquad (54)$$

and the elementary signal entropy becomes

$$C_M = (1 + \bar{n}) \log (1 + \bar{n}) - \bar{n} \log \bar{n} \quad . \qquad (55)$$

This can be recognized as the mode entropy of a system of particles obeying Bose-Einstein statistics.[17]

Before discussing the case of $\nu < 1$ in detail, we also write down the formal solutions for the noisy case and compare the capacity for the case $\nu = 1$ with previous results. For equilibrium noise, the matrix B becomes, in agreement with (39),

$$B = (1 - e^{-\Theta}) \begin{Bmatrix} 1 & e^{-\Theta} & e^{-2\Theta} & e^{-3\Theta} & \dots \\ 0 & 1 & e^{-\Theta} & e^{-2\Theta} & \dots \\ 0 & 0 & 1 & e^{-\Theta} & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{Bmatrix} \quad . \qquad (56)$$

This has to be inverted in order to give the inverse matrix (A'^{-1}) required for the solution for $p_r(n)$. The inverse of this noise matrix is of the peculiarly simple form

$$B^{-1} = \frac{1}{1 - e^{-\Theta}} \begin{Bmatrix} 1 & -e^{-\Theta} & 0 & 0 & \dots \\ 0 & 1 & -e^{-\Theta} & 0 & \dots \\ 0 & 0 & 1 & -e^{-\Theta} & \dots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{Bmatrix} \quad , \qquad (57)$$

which can be checked by forming BB^{-1}. Instead of the $(A^{-1})_{n,m}$ used above, we must now substitute

$$(A'^{-1})_{n,m} = \sum_j (B^{-1})_{n,j} (A^{-1})_{j,m} = \frac{1}{1 - e^{-\Theta}} [(A^{-1})_{n,m} - e^{-\Theta}(A^{-1})_{n+1,m}] \quad . \tag{58}$$

The transition probabilities must also be modified in accordance with Eq. (40) and now become

$$p(n|N) = e^{-n\Theta} \sum_{j=0}^n e^{j\Theta}\binom{N}{j} \nu^j (1 - \nu)^{N-j} \cdot (1 - e^{-\Theta}) \quad . \tag{59}$$

This can be checked for two particular cases. First, let the noise disappear, i.e., $T \to 0$ and hence $\Theta \to \infty$. Only one term survives in the above sum, namely that for which $j = n$, and we have

$$\lim_{T \to 0} p(n|N) = \binom{N}{n} \nu^n (1 - \nu)^{N-n} \quad , \tag{60}$$

which is the noiseless case. The other particular case is that for which $\nu \to 1$, i.e., the intrinsic lack of determinism of the channel in the absence of noise is removed, and we obtain

$$\lim_{\nu \to 1} p(n|N) = \begin{cases} \exp[-(n - N)\Theta](1 - e^{-\Theta}) & n \geqslant N \\ 0 & n < N \end{cases} \quad . \tag{61}$$

In this case the noiseless channel matrix, A, diagonalizes and we have, instead of (58),

$$(A'^{-1})_{n,m} = (B^{-1})_{n,m} \quad . \tag{62}$$

With these simplifications, we have

$$\sum_m (A'^{-1})_{n,m} \sum_i p(i|m) \log p(i|m)$$

$$= \sum_m (A'^{-1})_{n,m} (1 - e^{-\Theta}) \sum_{i=m}^\infty \exp[-(i - m)\Theta][\log(1 - e^{-\Theta}) - (i - m)\Theta]$$

$$= -\sum_m (B^{-1})_{n,m} [(1 + \bar{L}) \log(1 + \bar{L}) - \bar{L} \log \bar{L}]$$

$$= -[(1 + \bar{L}) \log(1 + \bar{L}) - \bar{L} \log \bar{L}] = -\Gamma \tag{63}$$

where $\bar{L}$ is the mean noise occupation number defined by

$$\bar{L} = \frac{e^{-\Theta}}{1 - e^{-\Theta}} \quad . \tag{64}$$

The distribution at the receiver hence becomes

$$p_r(n) = \exp[-\alpha + \bar{L}\beta - \beta n - \Gamma] \quad . \tag{65}$$

This distribution is substituted into the two constraint equations

17

$$\sum_{n} p_r(n) = 1 \quad \text{and} \quad \sum_{n} n p_r(n) = \overline{L} + \overline{S} \quad , \tag{66}$$

where $\overline{S}$ is the mean signal occupation number. The results for α and β are

$$\alpha = (1 + \overline{L}) \log(1 + \overline{L} + \overline{S}) - \overline{L} \log(\overline{L} + \overline{S}) - \Gamma \quad ,$$

$$\beta = \log(1 + \overline{L} + \overline{S}) - \log(\overline{L} + \overline{S}) \quad , \tag{67}$$

and the capacity becomes

$$C_M = \alpha + \overline{S}\beta = [(1 + \overline{L} + \overline{S}) \log(1 + \overline{L} + \overline{S}) - (\overline{L} + \overline{S}) \log(\overline{L} + \overline{S})]$$

$$- [(1 + \overline{L}) \log(1 + \overline{L}) - \overline{L} \log \overline{L}] \quad , \tag{68}$$

which is the difference of the entropy of signal plus noise minus the entropy of the noise alone, which was to be expected, and which is in agreement with Gordon's result.[7]

It should be noted that the function entering the expressions for the capacity in (68) and in (55), namely,

$$f(x) = (1 + x) \log(1 + x) - x \log x \quad ,$$

has a negative second derivative anywhere as long as $x \geqslant 0$. This means that the elementary signal capacity increases less rapidly than linearly with energy, and hence the total signal energy should be distributed on the average evenly among the various elementary signals.

C. Mutual Information When Partition Effects Are Present

In order to evaluate the capacity in agreement with the previous section, the parameters α and β have to be solved from the set of equations

$$\sum_{n=0}^{\infty} p_r(n) = 1 \quad ,$$

$$\sum_{n=0}^{\infty} n p_r(n) = \overline{L} + \overline{S} \quad , \tag{69}$$

with

$$p_r(n) = \exp\left[-\alpha - \frac{n}{\nu}\beta + G_n(\overline{L}, \nu)\right] \quad , \tag{70}$$

where

$$G_n(\overline{L}, \nu) = \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} (A'^{-1})_{n,m} \, p(i|m) \log p(i|m) \quad , \tag{71}$$

with

$$p(i|m) = \frac{1}{1 + \overline{L}} \left(\frac{\overline{L}}{1 + \overline{L}}\right)^i \sum_{j=0}^{i} \left(\frac{1 + \overline{L}}{\overline{L}}\right)^j \binom{m}{j} \nu^j (1 - \nu)^{m-j} \quad , \tag{72}$$

18

and with

$$(A'^{-1})_{n,m} = (1 + \overline{L}) (A^{-1})_{n,m} - \overline{L}(A^{-1})_{n+1,m} \quad , \tag{73}$$

where

$$(A^{-1})_{n,m} = \binom{n}{m} (1 - \nu)^{n-m} (-1)^{n-m} \nu^{-n} \quad . \tag{74}$$

It was at first thought that these equations could be solved by fixing the parameters $\overline{L}$ and ν, evaluating $G_n(\overline{L}, \nu)$ on an electronic computer, and then proceeding to solve the equations for β and α numerically. As an illustration of the difficulties then encountered, Table I presents one of the resulting G_n sequences obtained.

TABLE I				
NUMERICAL EVALUATION OF $G_n(\overline{L}, \nu)$				
n	$G_n(\overline{L}, \nu)$	n	$G_n(\overline{L}, \nu)$	
0	0	4	335.2726------	
1	−3.2508297---	5	−2524.671-------	
2	5.9749897---	6	19537.6---------	
3	−49.996700----	7	−154812.----------	
$\overline{L} = 0$ (no noise), $\nu = 1/10$.				

It is immediately obvious that such a sequence when substituted into $p_r(n)$ in Eq. (70) does not correspond to a valid physical situation, the main reason being that a distinction occurs between the probability of reception of even and odd numbers of quanta. These difficulties arose at all the combinations of parameters chosen.

Although we were not able to prove that the following explanation is correct, it is thought that the subsequent discussion and considerations of certain models make this explanation very plausible. In the calculations to evaluate the maximum mutual information presented in Sec. IV-B there is no constraint on the transmitter probabilities $p_t(N)$ to ensure that they are all positive. What is believed to happen in the case considered above is that the occurrence of certain negative transmitting probabilities (input probabilities) explains the "unphysical" behavior of the receiving probabilities $p_r(n)$ (output probabilities). This explanation will become more plausible after our presentation of certain simplified models below. The described difficulty is, by the way, not unknown in information theory and has been described in detail by Muroga.[8] To the knowledge of the author, no satisfactory method has yet been developed in information theory to handle the situation. The trial-and-error method suggested by Muroga and by Fano[10] is clearly completely impracticable in our case where extremely large numbers of possible signals are available at the transmitter. It is possible that the extra constraints required to keep $p_t(N) \geqslant 0$ could be ensured by some linear programming technique.[18] Unfortunately, we have not been able to develop an adequate method here. In the next section we therefore consider a simple model where negative

input probabilities do occur, but where the system is so simple that we can follow in detail what happens.

D. Ternary Model Channel

As the simplest possible example, consider the case of the elementary signals only being allowed to assume the signal states $N = 0$, 1 and 2, and let there be no additive noise. As a channel matrix, we use

$$
A = \begin{Bmatrix} 1 & 0 & 0 \\ 1-\nu & \nu & 0 \\ (1-\nu)^2 & 2\nu(1-\nu) & \nu^2 \end{Bmatrix} , \tag{75}
$$

which is seen to correspond to the upper-left corner of the infinite channel matrix A of Sec. IV-A. We can use the same procedure as there to evaluate the input and output probabilities. Because there is no average power constraint in our simplified model, Eq. (44) reduces to

$$
C = \alpha \quad , \tag{76}
$$

and Eq. (46) for the column vector $\vec{d}$ in this case becomes

$$
\vec{d} = \begin{Bmatrix} -C & & \\ -C & -H_1 & \\ -C & -2H_1 & +2\nu(1-\nu)\log 2 \end{Bmatrix} , \tag{77}
$$

with the following definition for H_1

$$
H_1 = -\nu \log \nu - (1-\nu) \log(1-\nu) \quad . \tag{78}
$$

If we note that the inverse of our present particular channel matrix corresponds to the upper three by three corners of the general inverted channel matrix, we obtain immediately for the output probabilities from Eq. (49)

$$
\vec{p}_r = \begin{Bmatrix} p_r(0) \\ p_r(1) \\ p_r(2) \end{Bmatrix} = \begin{Bmatrix} \exp[-C] \\ \exp[-C-(1/\nu)H_1] \\ \exp\{-C-(2/\nu)H_1 + 2\,[(1-\nu)/\nu]\log 2\} \end{Bmatrix} \tag{79}
$$

We can immediately see that these results are not valid for large transmission losses, i.e., $\nu \to 0$, because for small ν it turns out that

$$
p_r(2) \gg p_r(1) \quad ,
$$

and this does not appear to be physically possible, whatever the input probability $p_t(N)$. The channel capacity is found directly from the normalization condition on p_r to be

$$C = \log\left\{1 + \exp\left[-\frac{1}{\nu} H_1\right] + \exp\left[-\frac{2}{\nu} H_1 + 2\,\frac{1-\nu}{\nu}\,\log 2\right]\right\} \quad . \tag{80}$$

For the transmitted probabilities we obtain

$$p_t(0) = e^{-C}\left\{1 - \frac{1-\nu}{\nu}\,\exp\left[-\frac{1}{\nu} H_1\right] + \left(\frac{1-\nu}{\nu}\right)^2 \exp\left[-\frac{2}{\nu} H_1 + 2\,\frac{1-\nu}{\nu}\,\log 2\right]\right\} \quad ,$$

$$p_t(1) = e^{-C}\left\{\frac{1}{\nu}\,\exp\left[-\frac{1}{\nu} H_1\right] - 2\,\frac{1}{\nu^2}\,(1-\nu)\,\exp\left[-\frac{2}{\nu} H_1 + 2\,\frac{1-\nu}{\nu}\,\log 2\right]\right\} \quad ,$$

$$p_t(2) = e^{-C}\left\{\frac{1}{\nu^2}\,\exp\left[-\frac{2}{\nu} H_1 + 2\,\frac{1-\nu}{\nu}\,\log 2\right]\right\} \quad . \tag{81}$$

It turns out that the limiting values for these probabilities when $\nu \to 0$ are

$$\lim\; p_t(0) = +\infty$$
$$\lim\; p_t(1) = -\infty$$
$$\lim\; p_t(2) = +\infty$$

Upon closer examination of $p_t(1)$, it is found that this probability goes negative when

$$\nu = \nu_c \approx 0.62 \quad .$$

The p_r, p_t and C are all plotted against ν in Figs. 4 and 5. For this model circuit, we therefore experience a breakdown of the validity of the general capacity formula as soon as ν becomes less than a $\nu_c \approx 0.62$.

When one of the input probabilities goes negative, as just occurred, Muroga proves that the procedure to be followed consists of putting one of the input probabilities equal to zero and evaluating the capacity again. If solutions can be obtained with one of several of the input probabilities equal to zero, the capacity is the largest among these solutions. If no solution is found, one further input probability must be equated to zero and the procedure repeated. It is obvious that such a procedure cannot be used in our general case where we might begin by having enormously large quantum numbers at the transmitter.

In the elimination process just described there will be more output signals than input signals, and it becomes necessary to introduce constraints on the output signals in order to be able to get definite relations between output and input probabilities. Muroga[8] gives a general procedure to be followed in order to introduce the constraints among the output probabilities. Because of the simplicity of the present channel matrix, we can argue that when the ternary channel breaks down, we obviously have to omit the input (transmitted) symbol N = 1 and the channel matrix reduces to

$$A = \left\{\begin{array}{ccc} 1 & 0 & 0 \\ (1-\nu)^2 & 2\nu(1-\nu) & \nu^2 \end{array}\right\} \quad . \tag{82}$$

This channel is shown in Fig. 6. By simple inspection it is obvious that the two output symbols, n = 1 and n = 2, can be combined into one symbol without reducing our capabilities of inference about the transmitted symbols. The properly reduced channel matrix thus becomes

$$A = \left\{\begin{array}{cc} 1 & 0 \\ (1-\nu)^2 & 1-(1-\nu)^2 \end{array}\right\} \tag{83}$$
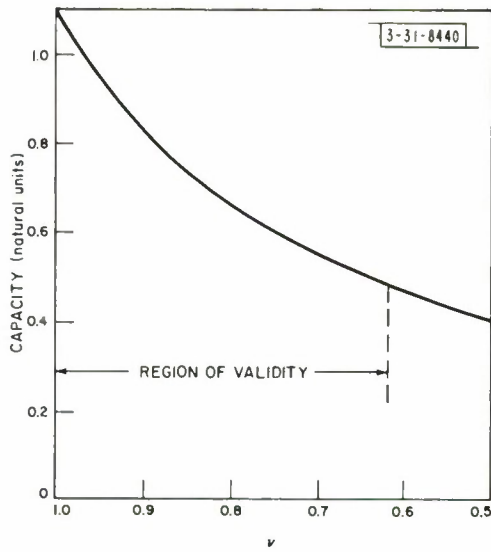
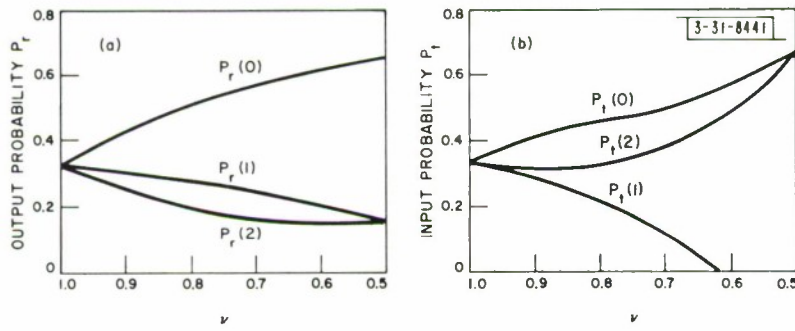Fig. 4. Capacity (natural units) of the ternary channel platted against ν.

Fig. 5(a-b). Output and input prabobilities far maximum infarmatian transfer in a ternary channel platted against ν.
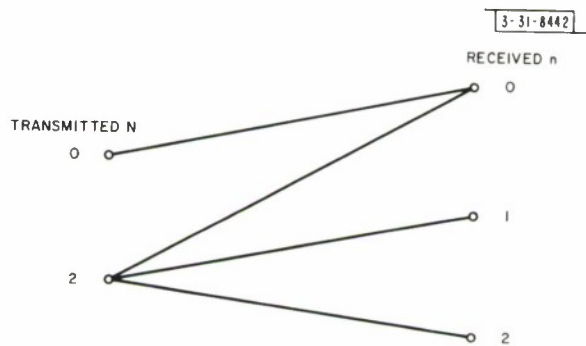

Fig. 6. Graphic representatian of the channel [Eq. (82)] .

A straightforward standard analysis shows that

$$C = \log \left\{ 1 + \exp \left[\log \nu (2 - \nu) + 2 \frac{(1 - \nu)^2}{\nu (2 - \nu)} \log (1 - \nu) \right] \right\} \quad . \tag{84}$$

The output probabilities become

$$p_r(0) = \exp [-C] \quad ,$$

$$p_r(1) = \exp [-C + \log 2 + \frac{1}{2 - \nu} \log (1 - \nu)] \quad ,$$

$$p_r(2) = \exp \left\{ -C + 2 \left[\log \nu + \frac{(1 - \nu)^2}{2 - \nu} \log (1 - \nu) \right] \right\} \quad , \tag{85}$$

and for the two non-zero input probabilities, we obtain

$$p_t(2) = \exp \left[-C + 2 \frac{(1 - \nu)^2}{2 - \nu} \log (1 - \nu) \right] \quad ,$$

$$p_t(0) = 1 - p_t(2) \quad . \tag{86}$$

The capacity for this binary scheme is plotted against ν in Fig. 7, and the different probabilities in Fig. 8. In the binary case, we do get valid results for all values of ν, but for $\nu > \nu_c = 0.62$
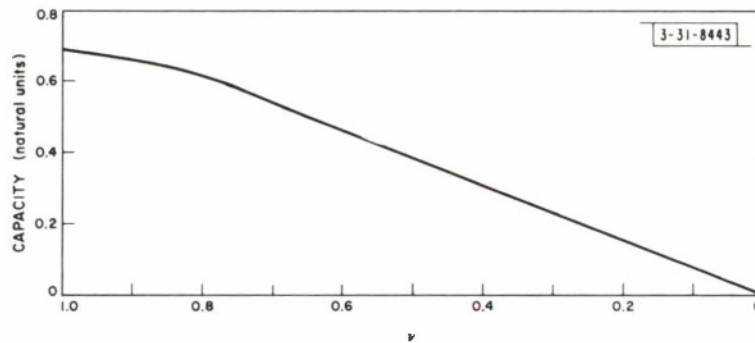
Fig. 7. Capacity (natural units) of the binary channel platted against ν.
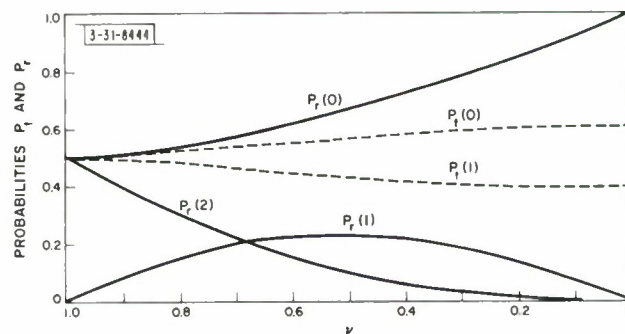
Fig. 8. Output and input probabilities far maximum infarmation transfer in a binary channel platted against ν.
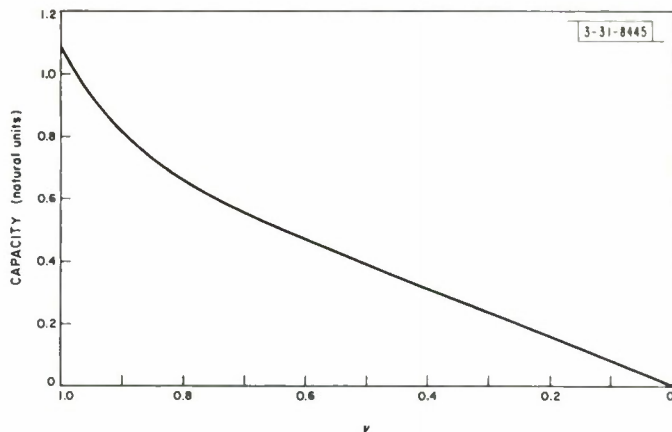
23

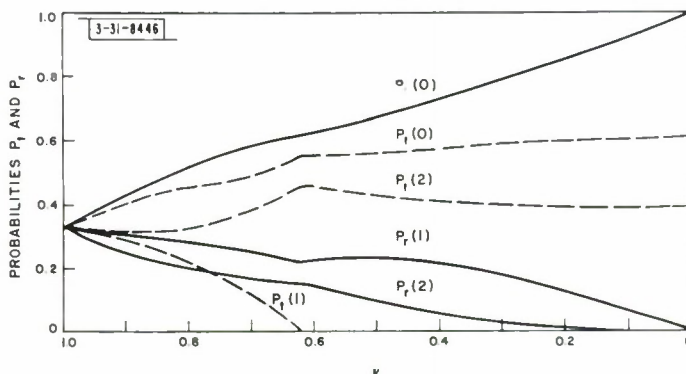Fig. 9. Capacity (natural units) of model channel plotted against ν.

Fig. 10. Signal probabilities required to achieve capacity in model channel plotted against ν.

the ternary encoding scheme gives the greater capacity. In Fig. 9 we show the maximum channel capacity for all values of ν for the particular channel under consideration. Figure 10 gives the probabilities as a function of ν for the optimum encoding scheme. It is reassuring to observe that all the curves are continuous through the value where the change between binary and ternary systems takes place.

E. Capacity for Extreme Transmission Losses When No Additive Noise Is Present

We now return to the more general case where a very large number of possible levels can be used both at the input and at the output of the channel. For the time being, we assume that there is a peak power constraint at the input, i.e., $N \leqslant M$. The general solution of the capacity problem, as we have seen, cannot be solved by any of the methods known to the author. It is possible, however, to obtain the capacity in the extreme case when $\nu \rightarrow 0$.

The effect of increasing the transmission loss, i.e., decreasing ν, is to decrease the number of input signals that can be used in order to achieve channel capacity. This process of decrease in the number of input symbols will continue until, in the extreme case, only two input symbols are present, namely, $N = 0$ and $N = M$, or until the channel has become binary. We

24

therefore conclude that for extreme transmission losses, the best we can do is to use an on − off system. When this stage has been reached, the channel matrix assumes the form

$$
A \ = \ \left\{ \begin{array}{cccccc} 1 & 0 & & 0 & \ldots & 0 \\ \\ (1-\nu)^M & \binom{M}{1}(1-\nu)^{M-1}\nu & \binom{M}{1}(1-\nu)^{M-2}\nu^2 & \ldots & \binom{M}{M}\nu^M \end{array} \right\} \quad . \tag{87}
$$

In exactly the same way as in the ternary channel discussed in Sec. IV-D, we can conclude here that $n = 0$ is one of the output symbols and $n \geqslant 1$ is the other one. The channel matrix therefore reduces to

$$
A \ = \ \left\{ \begin{array}{cc} 1 & 0 \\ \\ (1-\nu)^M & 1-(1-\nu)^M \end{array} \right\} \quad . \tag{88}
$$

Putting

$$
H_M = -M(1-\nu)^M \log(1-\nu) - [1-(1-\nu)^M] \log[1-(1-\nu)^M] \quad , \tag{89}
$$

$$
\begin{aligned}
C &= \log\left\{1 + \exp\left[-\frac{1}{1-(1-\nu)^M} H_M\right]\right\} \\
&= \log\left\{1 + [1-(1-\nu)^M] \exp\left[\frac{M(1-\nu)^M}{1-(1-\nu)^M} \log(1-\nu)\right]\right\} \quad .
\end{aligned} \tag{90}
$$

For numerical computations it is convenient to introduce the approximations:

$$
\log(1-\nu) \approx -\nu \quad ,
$$

$$
(1-\nu)^M \approx e^{-\nu M} \quad .
$$

This means that we replace the binomial transition probabilities with Poisson probabilities, which is an extremely good approximation under the particular conditions considered here. The capacity now simplifies to

$$
C \approx \log\left\{1 + (1-e^{-\nu M}) \exp\left[-\frac{\nu M}{e^{\nu M}-1}\right]\right\} \quad . \tag{91}
$$

This type of channel has previously been considered by R. C. Jones[19] who evaluated information transfer for particular input probabilities but did not optimize the information transfer with respect to the input probabilities as we have done here. The output probabilities implied in (91) are

$$
p_r(0) = e^{-C} \quad ,
$$

$$
p_r(n > 0) = 1 - e^{-C} \quad , \tag{92}
$$

and the input probabilities are

$$
p_t(M) = \frac{p_r(n > 0)}{p(n > 0/M)} = \frac{1 - p_r(0)}{1 - \exp[-\nu M]} \quad ,
$$

$$
p_t(0) = 1 - p_t(M) \quad . \tag{93}
$$

25

It would be of some interest to evaluate the probabilities of receiving a certain number of quanta, n. In the Poisson approximation, this becomes

$$p_r(n) = p_t(M) \frac{(\nu M)^n}{n!} e^{-\nu M} \qquad n \geqslant 1 \qquad .$$ (94)

Figure 11 shows the capacity as a function of νM for the binary case just considered. As νM increases, an asymptotic value is reached corresponding to the binary channel noiseless case. Note that it is only for νM less than some number near unity that the curve shown gives the actual peak power limited capacity. When νM is slightly larger than this (undetermined!) number, it would presumably be advantageous to use a ternary channel. For even larger νM the number of input symbols used should be even further increased to optimize the capacity. Unfortunately, mathematical difficulties have prevented us from establishing the exact values where the transitions between the various types of encoding procedures should take place.
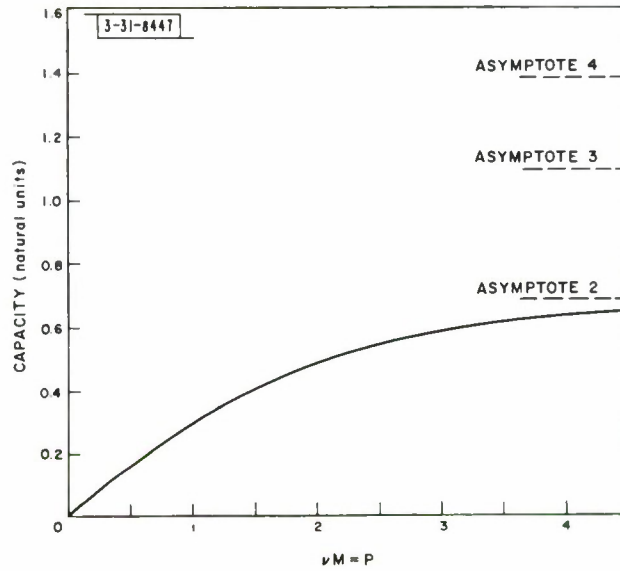
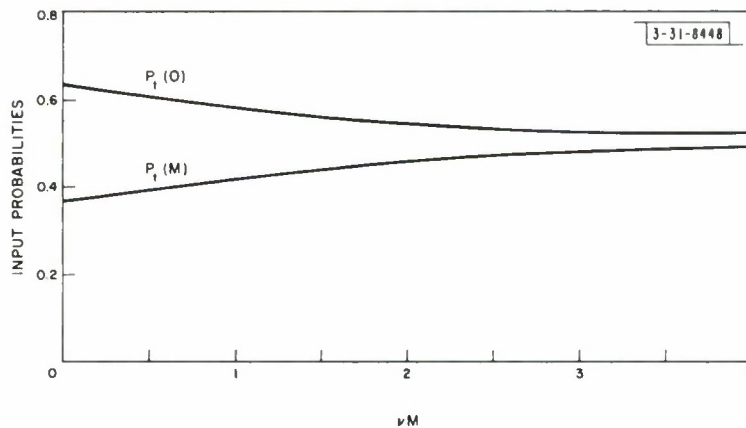Fig. 11. Binary channel capacity for peak power limitation.

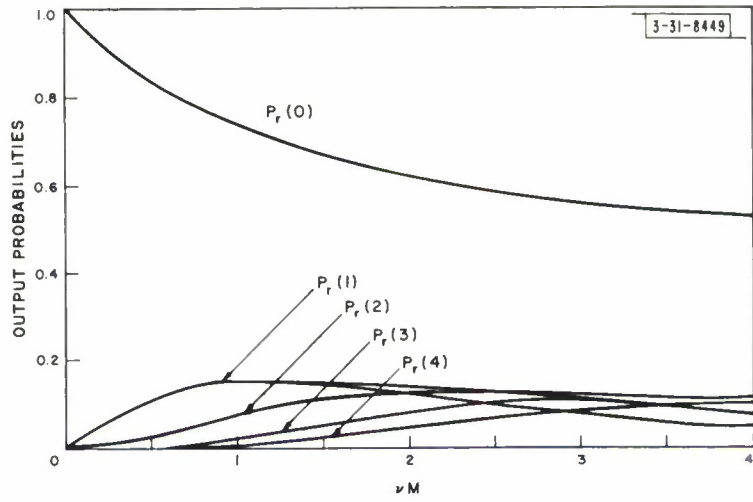Fig. 12. Input probabilities for binary channel.
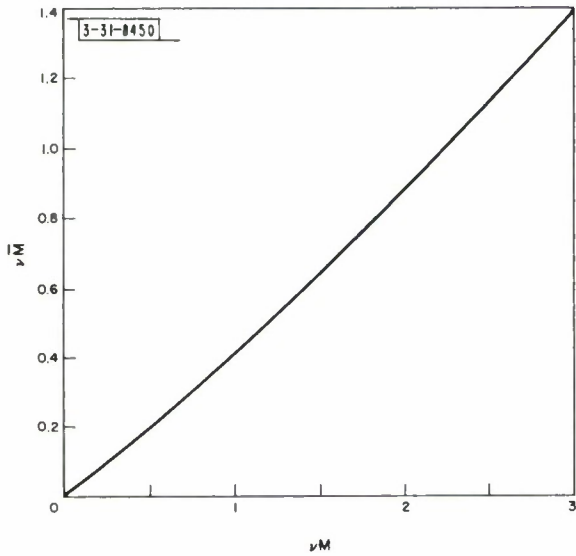
26

Fig. 13. Output probabilities for binary channel.

Fig. 14. Relation between νM and $\nu \overline{M}$.
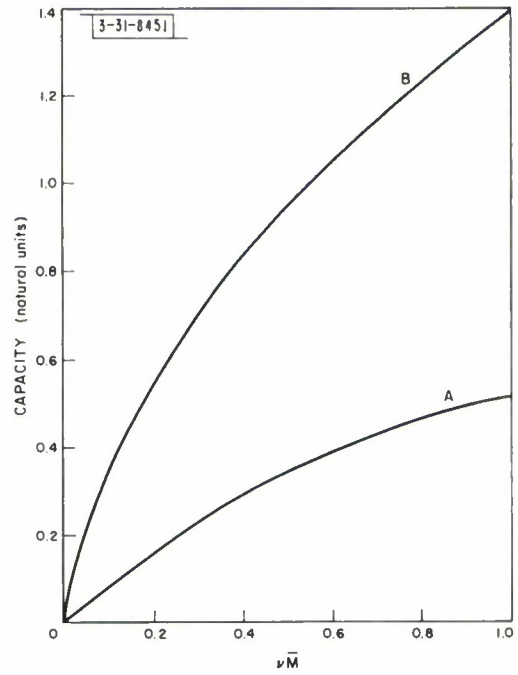

Fig. 15. Curve A shows the channel capacity against $\nu \overline{M}$ when the channel is binary. Curve B shows the no-partition noise capacity.

Finally, in order to provide a basis for the comparison of the quantized channel capacity with partition noise with the quantized channel capacity without partition noise as considered by Gordon,[7] we have to establish the binary capacity under average power constraints. Rather than going through the whole optimization procedure again, we use the results of the above peak power limited case. When M is the "on-level" of the circuit, we have already established the capacity [see Eq. (91)]. Let us denote this capacity by C(M). For this particular value of M we have for $p_t(M)$

$$p_t(M) = \frac{1 - \exp[-C(M)]}{1 - \exp[-\nu M]} \quad . \tag{95}$$

The average power constraint requires that

$$M \cdot p_t(M) = \overline{M} \quad , \tag{96}$$

where $\overline{M}$ is the average number of quanta transmitted in an elementary signal. Between Eqs. (95) and (96), we can solve for $\nu M = g(\nu\overline{M})$, and hence we can find the binary capacity C as a function of $(\nu\overline{M})$. The relation between νM and $\nu\overline{M}$ is shown in Fig. 14. In Fig. 15 is shown the capacity as a function of $\nu\overline{M}$, that is the mean number of received signal quanta. For comparison, we also show Gordon's "noiseless" capacity curve [see Eq. (55)]. It is immediately obvious that the partition noise reduces the channel capacity very considerably. In order to put this statement on a more quantitative basis, we inquire about the amount of additive noise which would make the no-partition noise capacity curve corresponding to (68) have the same slope at the origin as curve A in Fig. 15. The equivalent noise occupation number turns out to be

$$\overline{N} = 0.745 \quad . \tag{97}$$

The asymptotic value of the partition noise as $\nu \to 0$ therefore is equivalent to a rather substantial additive noise.

V. CONCLUSIONS

In this report, we have discussed several problems arising in the communication by means of electromagnetic waves when quantum effects come into play. In Sec. II the problem of describing a signal was discussed. The familiar idea of sampling as used in the classical theories was rejected as a basis for a signal description. Instead, the signal was divided into elementary signals which can be studied by means of a cross-correlation detection procedure which makes use of the total elementary signal energy. Certain quantum mechanical considerations showed that the elementary signals can be specified to an integer number of quanta of energy. It was also shown that a composite signal can be subdivided into approximately BT independent elementary signals, B being the bandwidth and T the duration of the total signal.

In Sec. III a relation between transmitted and received signals was established in agreement with quantum theory, and it was shown that a partition noise effect shows up in a proper derivation of the signal channel matrix.

In Sec. IV we first attempted to derive the channel capacity by using the familiar Lagrange multiplier optimization scheme. This procedure was shown to break down because the occurrence of radiative losses requires the number of signal possibilities to be reduced at the transmitter in order to achieve maximum capacity. It was furthermore established that under extreme transmission losses the optimum encoding procedure tends to become purely binary. For this

extreme case the channel capacity was established with both peak and average power constraints. It was shown that partition noise effects become rather important under these extreme conditions and an equivalent additive noise power was established.

It should be emphasized that the results described in this report are not complete because in several places general conclusions have been drawn from particular examples. All the problems in connection with the occurrence of negative input probabilities should be put on a firmer basis, and the transition regions between the various transmission encoding schemes should be established in order to evaluate the true capacity under more general conditions. Another point which deserves more careful attention is the meaning and nature of the measuring procedure involved in establishing the commutation relation for the signal parameters B and B* (see Appendix A).

APPENDIX A
QUANTIZATION OF AN ELEMENTARY PLANE WAVE SIGNAL

Starting with an elementary plane wave as in (6), we obtain for the Lagrangian density of an elementary signal

$$L = \frac{\epsilon_o}{2} \left(\frac{\partial A_x}{\partial t}\right)^2 - \frac{1}{2\mu_o} \left(\frac{\partial A_x}{\partial z}\right)^2 \quad , \tag{A-1}$$

the variable canonically conjugate to A_x is found to be[12,13]

$$P_x = \frac{\partial L}{\partial(\frac{\partial A}{\partial t})} = \epsilon_o \frac{\partial A_x}{\partial t} \quad , \tag{A-2}$$

and the Hamiltonian density becomes

$$H = \frac{1}{2\epsilon_o} P^2 + \frac{1}{2\mu_o} (\frac{\partial A}{\partial z})^2 \quad . \tag{A-3}$$

In order to evaluate the total Hamiltonian, it is convenient to express an elementary signal in the following way

$$A_x(z, t) = \frac{1}{\Delta k} \int_{k_o-(\Delta k/2)}^{k_o+(\Delta k/2)} \left\{ B \exp[ik(z - et)] + B^* \exp[-ik(z - et)] \right\} dk \quad , \tag{A-4}$$

where

$$k_o = \frac{\omega_o}{e} \quad ,$$

$$\Delta k = \frac{2\pi}{T} \quad ,$$

$$B = \frac{1}{2} (\alpha + i\beta) \quad .$$

In this representation, the canonically conjugate variable becomes

$$P_x(z, t) = -ie \frac{\epsilon_o}{\Delta k} \int_{k_o-(\Delta k/2)}^{k_o+(\Delta k/2)} k \left\{ B \exp[ik(z - ct)] - B^* \exp[-ik(z - ct)] \right\} dk \quad . \tag{A-5}$$

When computing the total Hamiltonian, we have to limit the wave in the plane perpendicular to the z-axis because of the infinite extent of a plane wave. Let us imagine a rectangular shape of area $L_x L_y$. This somewhat artificial limiting of the wave can be avoided if other types of spatial modes are considered. The total Hamiltonian is evaluated as follows

$$H = L_x L_y \int_{-\infty}^{\infty} H dz \quad ,$$

$$= -c^2 \frac{\epsilon_o}{\Delta k^2} L_x L_y \iint_{k_o-(\Delta k/2)}^{k_o+(\Delta k/2)} dk dk' kk' \int dz \left\{ -BB^* \exp[i(k - k')(z - ct)] \right.$$

$$\left. - B^*B \exp[-i(k - k')(z - ct)] \right\}$$

$$H = c^2 \frac{\epsilon_o}{\Delta k^2} L_x L_y 2\pi \int\!\!\int_{k_o-(\Delta k/2)}^{k_o+(\Delta k/2)} dk\,dk'\, k k' \delta(k-k')\, (BB* + B*B)$$

$$= c^2 \frac{\epsilon_o}{\Delta k^2} L_x L_y \cdot 2\pi(BB* + B*B) \int_{k_o-(\Delta k/2)}^{k_o+(\Delta k/2)} dk \cdot k^2$$

$$= L_x L_y\, cT\epsilon_o \omega_o^2 \left[1 + \frac{1}{3}\left(\frac{\pi}{\omega_o T}\right)^2\right] (BB* + B*B) \quad . \tag{A-6}$$

Next we have to establish a commutation rule for the B and B* from general quantum principles. We base the quantization procedure on the following commutator relation (see for instance Ref. 13)

$$[A_x(z,t), P_x(z',t)] = \frac{1}{L_x L_y}\,\delta(z-z')\,i\hbar \qquad ([\ ] = \text{Poisson bracket}) \quad . \tag{A-7}$$

In order to establish the desired commutation relation, we have to express B and B* in terms of A and P. This solution for B and B* cannot be made in an arbitrary way. It turns out that the commutator of B and B* depends on the way in which they are related to the field represented by A and P. If this seems strange, we must realize that the quantities B and B* only acquire a physical meaning provided we devise an experiment to measure the quantities. When we relate B and B* to A and P, we must regard this as a specification of an experiment for their observation. The dependence of the commutator upon the representation of B and B* therefore means that all experiments designed to evaluate B and B* are not equally good.

It seems that the best one can do is to evaluate B and B* by the following formulas

$$B = \frac{1}{4\pi} \int_{k_o-(\Delta k/2)}^{k_o+(\Delta k/2)} dk \int_{-\infty}^{\infty} dz \left[A_x(z,t) + \frac{ik}{\epsilon_o ck_o^2 g} P_x(z,t)\right] \exp[-ik(z-ct)] \quad ,$$

$$B* = \frac{1}{4\pi} \int_{k_o-(\Delta k/2)}^{k_o+(\Delta k/2)} dk \int_{-\infty}^{\infty} dz \left[A_x(z,t) - \frac{ik}{\epsilon_o ck_o^2 g} P_x(z,t)\right] \exp[ik(z-ct)] \quad , \tag{A-8}$$

where $g = \left\{1 + \frac{1}{3}(\pi/T\omega_o)^2\right\}$. It is obvious that these equations can be derived directly from the representations (A-4) and (A-5). The detailed nature of the corresponding observations will not be discussed here, but it appears that the representations for B and B* correspond to a cross-correlation experiment where the cross-correlative waveform is of exactly the same nature as the signal waveform. For the commutator of B and B* we obtain

$$[B, B*] = \frac{1}{16\pi^2} \int\!\!\int\!\!\int\!\!\int dk\,dk'\,dz\,dz' \left(-\frac{i}{\epsilon_o ck_o^2 g}\right) \Big\{ k\, [A_x(z,t), P_x(z',t)]$$

$$+ k'\, [A_x(z',t), P_x(z,t)]\Big\} \exp[ik'(z'-ct) - ik(z-ct)] \quad ,$$

$$= \frac{i}{16\pi^2 \epsilon_o ck_o^2 g L_x L_y} \int\!\!\int\!\!\int\!\!\int dz\,dk\,dk'\,dz' \, [ki\hbar\delta(z-z') + k'i\hbar\delta(z'-z)]$$

$$\times \exp[ik'(z'-ct) - ik(z-ct)] \quad ,$$

32

$$[B, B^*] = \frac{\hbar}{L_x L_y 16\pi^2 \epsilon_o ck^2 g} \iiint dkdk'(k + k') \, dz \, \exp[i(k' - k)(z - ct)] \quad ,$$

$$= \frac{\hbar}{L_x L_y 8\pi \epsilon_o ck^2 g} \iint dkdk'(k + k') \, \delta(k' - k) = \frac{1}{L_x L_y cT} \frac{\hbar}{2\epsilon_o \omega_o g} \quad . \tag{A-9}$$

To simplify even further, we put $V = L_x L_y cT$ which equals the volume occupied by the signal, and obtain

$$a = \sqrt{V \cdot \frac{2\epsilon_o \omega_o g}{\hbar}} \, B \quad ,$$

$$a^* = \sqrt{V \cdot \frac{2\epsilon_o \omega_o g}{\hbar}} \, B^* \quad . \tag{A-10}$$

Substituting these new operators into (A-6), we obtain the familiar expression for the elementary signal Hamiltonian

$$H = \hbar\omega_o \frac{1}{2} (aa^* + a^*a) = \hbar\omega_o (a^*a + \frac{1}{2}) \quad . \tag{A-11}$$

It is shown in many standard texts on quantum mechanics[14] that a convenient set of basis vectors for the description of such a system is formed by the eigenvectors of the operator a^*a, viz., $|0\rangle, |1\rangle, \ldots |n\rangle$. It is also shown that the corresponding eigenvalues are the integers $0, 1 \ldots n$, respectively. The creation operator, a^*, and the annihilation operator, a, are completely specified by the relations

$$a^* |n\rangle = \sqrt{n + 1} \, |n + 1\rangle \quad ,$$

$$a |n\rangle = \sqrt{n} \, |n - 1\rangle \quad ,$$

$$a |0\rangle = 0 \quad .$$

With this we have reached a valid quantum description of the elementary signals and hence of a composite signal. We observe that only a discrete set of states of the elementary signals is possible; in fact, an elementary signal at center frequency ω_o can be specified to an accuracy in energy of $\hbar\omega_o$. Similar conclusions can be reached for other types of elementary signals.

APPENDIX B
DIRECT EVALUATION OF P_r

In order to evaluate P_r, we compare the two signal representations at the transmitting aperture plane. To avoid unnecessary complications, we compare terms with identical time-variation.

From Eqs. (9) and (11) we have in the t-representation

$$A(x, y, k) = (\frac{1}{2\pi})^2 \iint dk_x dk_y F_t(k_x k_y) \, B_t \, \exp[i(k_x x + k_y y)]$$

$$+ \text{ complex conjugate} \quad . \tag{B-1}$$

The function $F_t(k_x k_y)$ is defined in (23). In the $r - l$ representation, the terms with similar time-variation are [see Eq. (24)]

$$(\frac{1}{2\pi})^2 \iint dk_x dk_y F_r(k_x k_y) \, B_r \cdot \exp\left\{i\left[k_x x + k_y y - z_0 \sqrt{k^2 - (k_x^2 + k_y^2)}\right]\right\}$$

$$+ \text{ similar } l\text{-term} \quad . \tag{B-2}$$

Integration of both (B-1) and (B-2) over the transmitting aperture gives the equation

$$\iint dk_x dk_y F_t(k_x k_y)^2 \, B_t = \iint dk_x dk_y F_r(k_x k_y) \, F_t(k_x k_y) \exp\left[-iz_0 \sqrt{k^2 - (k_x^2 + k_y^2)}\right] B_r$$

$$+ \text{ similar } l\text{-term} \quad . \tag{B-3}$$

The left side can be integrated over k_x and k_y directly, giving

$$\alpha_t (2\pi)^2 \, B_t \quad . \tag{B-4}$$

In order to make the right-hand side easy to evaluate, we make the assumption that the exponential factor varies much more rapidly than the $F_r F_t$ factor. Only contributions near $k_x = k_y = 0$ will be important, and we obtain the approximate equation

$$B_t \approx (\frac{1}{2\pi})^2 \alpha_r \iint dk_x dk_y \cdot \exp\left[-iz_0 \sqrt{k^2 - (k_x^2 + k_y^2)}\right] B_r \quad ,$$

$$+ \text{ similar } l\text{-term} \quad . \tag{B-5}$$

Because k_x and k_y can deviate but little from zero, the square root can be expanded to give

$$B_t \approx (\frac{1}{2\pi})^2 \alpha_r \cdot \exp[-ikz_0] \iint dk_x dk_y \cdot \exp\left[iz_0\left(\frac{k_x^2 + k_y^2}{2k}\right)\right] B_r \quad ,$$

$$+ \text{ similar } l\text{-term} = \frac{1}{2\pi} \alpha_r \cdot i \exp[-ikz_0] \cdot \frac{k}{z_0} B_r \quad ,$$

$$+ \text{ similar } l\text{-term} \quad . \tag{B-6}$$

Finally, we introduce the operators a and a* through the Eq. (A-10)

$$B_t = \sqrt{\frac{\hbar}{cT\alpha_t 2\epsilon_0 \omega_0 g}} \cdot a_t \quad ,$$

$$B_r = \sqrt{\frac{\hbar}{cT\alpha_r{}^2\epsilon_o\omega_o g}} \cdot a_r \quad ,$$

and obtain

$$a_t = \frac{i}{2\pi} \sqrt{\alpha_t\alpha_r} \cdot \exp[-ikz_o] \frac{k}{z_o} \cdot a_r + \ell\text{-term} \quad . \tag{B-7}$$

We therefore conclude that

$$|P_r|^2 \doteq (\frac{k}{2\pi})^2 \frac{\alpha_t\alpha_r}{z_o{}^2} \quad , \tag{B-8}$$

which is just the transmission loss in a propagation circuit (see for instance Ref. 20).

REFERENCES

1. K. Shimodo, H. Jokohoshi ond C. H. Townes, "Fluctuotions in Amplificotion of Quonto with Applicotion to Moser Amplifiers," J. Phys. Soc. Jopon 12, 686 (June 1957).

2. T. E. Stern, "Informotion Rotes in Photon Chonnels ond Photon Amplifiers," IRE Conv. Rec. 8, Port 4, 182 (1960).

3. W. H. Louisell, A. Yoriv ond A. E. Siegmon, "Quontum Fluctuotions ond Noise in Porometric Processes," Phys. Rev. 124, 1646 (December 1961).

4. H. Heffner ond C. Y. She, "Utilizotion of Opticol Mosers," Progress Report, Project 0415/215, Stonford Electronics Loborotories (1963), 1-68.

5. H. A. Hous ond J. A. Mullen, "Photon Noise," 4th Internotionol Congress on Microwove Tubes, September 1962 (Centrex Publishing Co., Eindhoven, 1963).

6. T. E. Stern, "Some Quontum Effects in Informotion Chonnels," Trons. IRE, PGIT IT-6, 435 (September 1960).

7. J. P. Gordon, "Quontum Effects in Communicotions Systems," Proc. IRE 50, 1898 (September 1962).

8. S. Murogo, "On the Copocity of o Discrete Chonnel," J. Phys. Soc. Jopon 8, 484 (1953).

9. S. Goldmon, Informotion Theory (Prentice-Holl, New York, 1953), Chopter 2.

10. R. M. Fono, Tronsmission of Informotion (M. I. T. Press, Combridge, 1961), Chopter 5.

11. D. Slepion, H. O. Pollok ond H. J. Londou, "Prolote Spheroidol Wove Functions, Fourier Anolysis, ond Uncertointy," BSTJ 40, 43 ond 65 (Jonuory 1961).

12. W. Heisenberg, Physicol Properties of the Quontum Theory (Dover, New York, n.d.).

13. L. I. Schiff, Quontum Mechonics (McGraw-Hill, New York, 1949), 362-3.

14. A. Messiah, Quontum Mechanics I (Interscience, New York, 1961), 434-5.

15. P. M. Woodword ond J. D. Lowson, "The Theoreticol Precision with Which on Arbitrary Pottern Moy Be Obtoined with o Source of Finite Size," J. Inst. Elec. Engrs. (London) 95, Port III, 362 (1948).

16. W. Heitler, The Quontum Theory of Rodiotion (Clorendon Press, Oxford, 1957).

17. L. D. Londou ond E. M. Lifshitz, Stotisticol Physics (Addison-Wesley, Reoding, 1958), 155.

18. R. E. Bellmon, Dynomic Progromming (Princeton University Press, Princeton, 1957).

19. R. C. Jones, "Informotion Copacity of o Beom of Light," J. Opt. Soc. Am. 52, 493 (Moy 1962).

20. S. Romo ond J. R. Whinnery, Fields ond Woves in Modern Rodio (Wiley, New York, 1953), 560.