UNIVERSITY OF SOUTHERN CALIFORNIA

COMMUNICATION THEORY FOR THE FREE SPACE OPTICAL CHANNEL

Interim Technical Report
August 1970

R. Gagliardi
S. Karp
E. O'Neill

This work was sponsored in part by the National Aeronautics and Space Administration, under NASA Contract NGR-05-018-104. This grant is part of the research program at Goddard Space Flight Center, Greenbelt, Maryland.

ELECTRONIC SCIENCES LABORATORY

USC Engineering

USCEE Report 397

August, 1970

Interim Technical Report

August 1970

Communication Theory for the Free Space Optical Channel

R. Gagliardi
S. Karp
E. O'Neill

Department of Electrical Engineering
University of Southern California
Los Angeles, California 90007

This work was sponsored by the National Aeronautics and Space Administration, under NASA Contract NGR-05-018-104. This grant is part of the research program at Goddard Space Flight Center, Greenbelt, Maryland.

This document represents a technical report of a NASA Research Program to study synchronization techniques for optical communication systems. The work is being carried out at the Electrical Engineering Department at the University of Southern California, under NASA Contract No. NGR-05-018-104, with Professor R. M. Gagliardi as principal investigator. This grant is part of the research program being conducted through NASA's Goddard Space Flight Center, Greenbelt, Maryland.

TABLE OF CONTENTS

ABSTRACT

The objective of this paper is to summarize the current understanding of quantum detectors, the noise mechanisms which are basic to their operation, and the application to optical communication theory. In this context we are considering channels in which the Electromagnetic field is not subjected to any propagation effects other than a geometric loss. (Such a channel would exist between satellites.) Consequently, we will concentrate on optimum time processing using the tools of statistical communication theory.

Fundamental to the study of a detection process is the need to develop a good mathematical model to describe it.[1-6] Therefore, approximately one-fifth of the paper is devoted to establishing in a semi-classical analysis the quantum detector output electron number as a conditional Poisson process with the conditioning variable being the modulus of the electromagnetic field. Once this has been established, these results are used to derive various limiting probability densities related to actual practice. Although the mathematical details are omitted, this will be presented from the viewpoint of orthogonal function expansions and interpreted in terms of an eigen-space.

The resulting current flow is next analyzed as a shot noise process and the power density spectrum is calculated. Attention is

focused on isolating the signal components from the noise in terms of both the current probability density and the power density spectrum. Examples are given where appropriate. At this point an understanding of the underlying noise processes will have been presented and attention will shift to analog and digital communications.

The analog communication will be presented primarily in terms of the signal-to-noise ratio although some attention will be given to continuous estimation. The S/N ratio in direct detection will be presented both as a ratio of the integrals of two separate portions of the spectrum and as a ratio of two moments of the probability density describing the current. These calculations will be extended to include heterodyne detection.

Digital communications will be discussed in the context of detection theory. It will be shown that the likelihood ratio is often a monotonic function of the random variable representing the number of electrons flowing. Hence optimum processing will consist of a weighted count of electrons from various counting modes. Digital design will be presented in terms of M-ary signalling, error probabilitites, and information rates.

I. INTRODUCTION

We begin with a classical description for the energy and momentum densities of a radiation field for both the single- and multi-mode cases. Confining our treatment to the semi-classical theory, we sketch briefly the argument that the probability of ejecting an electron from a photo-cathode surface in a short time interval is proportional to the light intensity. From this point of view, we deduce an expression for the probability of releasing n photoelectrons in a time T in terms of a weighted Poisson distribution. The weight factor is the probability distribution for the accumulated energy received on the photodetective surface in equal times.

Semi-Classical Theory

(A) Normal Mode Decomposition of the Field.- We begin our description of the semi-classical theory of radiation and matter by writing down the free space wave equation for the vector potential $\vec{A}(\vec{r},t)$,

$$\nabla^2\vec{A} - \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2} = 0 \tag{1}$$

Electing to work in the Coulomb gauge, div $\vec{A} = 0$, the electric and magnetic field vectors are now given by:

$$\begin{aligned} \vec{E} &= -\frac{\partial\vec{A}}{\partial t} \\ \vec{B} &= \text{curl}\ \vec{A} \end{aligned} \tag{2}$$

Concentrating first on a single mode of the radiation field, a plane wave is characterized by the components of the wave vector $\vec{k} = (k_x, k_y, k_z)$ where $\omega = |k|c$. However, even after specifying the direction and frequency of a plane electromagnetic wave, there still exists the posibility of two, independent, orthogonal polarization directions, $\hat{\sigma}_1$ and $\hat{\sigma}_2$.

A plane wave, then, at frequency ω propagating in the direction $\hat{k}$ can be written as:

$$\text{(a)} \quad \vec{A}(\vec{r},t) = \vec{a}(t)e^{i\vec{k}\cdot\vec{r}} + \vec{a}^*(t)e^{-i\vec{k}\cdot\vec{r}}$$

$$\text{(b)} \quad \vec{E} = i\omega\left(\vec{a}e^{i\vec{k}\cdot\vec{r}} - \vec{a}^*e^{-i\vec{k}\cdot\vec{r}}\right) \qquad (3)$$

$$\text{(c)} \quad \vec{B} = i\left[\left(\vec{k} \times \vec{a}\right)e^{i\vec{k}\cdot\vec{r}} - \left(\vec{k} \times \vec{a}^*\right)e^{-i\vec{k}\cdot\vec{r}}\right]$$

where $\quad \vec{a} = (a_1\hat{\sigma}_1 + a_2\hat{\sigma}_2)e^{-i\omega t}$

It will also turn out to be useful to list the energy density μ plus the linear and angular momentum densities $\vec{g}$ and $\vec{m}$ associated with this wave.

$$\mu = \frac{\vec{E}\cdot\vec{D} + \vec{B}\cdot\vec{H}}{2} = 2\omega^2\varepsilon_0|a|^2$$

$$\vec{g} = \frac{\vec{E} \times \vec{H}}{c^2} = -\frac{1}{\mu_0 c^2}(\dot{\vec{A}} \times \text{curl } \vec{A}) = \frac{2\omega^2\varepsilon_0}{c}|a|^2\hat{k} \qquad (4)$$

$$\vec{m} = \frac{1}{\mu_0 c^2}(\vec{A} \times \dot{\vec{A}}) = 2\omega\varepsilon_0(|b_+|^2 - |b_-|^2)\hat{k}$$

where

$$b_\pm = \frac{\sqrt{2}}{2}(a_1 \pm ia_2)$$

We are following here the notation of Louisell (ref. 7). The ambiguity in sign in the last expression is removed when we choose either right- or left-handed circularly polarized light. Of course, for linearly polarized light, a_1 and a_2 are in phase so that with $|b_+|^2 = |b_-|^2$ no net angular momentum is propagated. We also add in passing that the second term in Eq. (3a) is added to ensure the reality of $\vec{A} = \vec{A}^*$. A plane wave traveling in the opposite direction $(-\hat{k})$ is obtained by changing the sign of $\vec{k}$. Finally, a standing wave is described by taking a linear combination of the expression with $+\vec{k}$ and $-\vec{k}$. Before moving on to the multimode description of the radiation field, we will now select a single polarization component of the field and decompose this complex quantity in the form:

$$
\begin{aligned}
a_j &= \frac{1}{\sqrt{4\epsilon_0\omega^2}}(\omega q_j + ip_j) \\
a_j^* &= \frac{1}{\sqrt{4\epsilon_0\omega^2}}(\omega q_j - ip_j)
\end{aligned}
\tag{5}
$$

Under this transformation of variables, the energy and momentum densities become:

$$
\begin{aligned}
\mu_j &= \frac{p_j^2 + \omega^2 q_j^2}{2} = H_j \\
g_j &= \frac{H_j}{c}\hat{k}
\end{aligned}
\tag{6}
$$

so that as far as energy and momentum considerations are concerned, the radiation field can be treated as a simple harmonic oscillator obeying Hamilton's canonical equations of motion:

$$\dot{q}_j = \frac{\partial H_j}{\partial p_j}$$
$$\dot{p}_j = -\frac{\partial H_j}{\partial q_j} \qquad (7)$$

Turning now to the multimode description of the field, we impose periodic boundary conditions by introducing the triad of integers ℓ_1, ℓ_2, and ℓ_3 into the relation:

$$\vec{k} = \left(k_x,\ k_y,\ k_z\right) = \frac{2\pi}{L}\left(\ell_1,\ \ell_2,\ \ell_3\right) \qquad (8)$$

For economy in notation, we will henceforth use the symbol ℓ to imply this triad, and all Fourier sums will be treated as:

$$\sum_{\ell_1,}\ \sum_{\ell_2,}\ \sum_{\ell_3} \Rightarrow \sum_{\ell} \qquad (9)$$

Moreover, the orthogonality relation:

$$\int_0^L \int_0^L \int_0^L e^{i(\vec{k}_\ell - \vec{k}_{\ell'})\cdot\vec{r}}\, dxdydx = V\delta_{\ell\ell'} \qquad (10)$$

taken over a cube of volume $V = L^3$ will guarantee that each mode will contribute independently† to the total energy and momentum of the field.

†Of course, this lack of cross terms in adding up the total energy of a system is the whole idea behind normal mode decomposition. Also, choosing plane wave eigenfunctions, orthogonal over cubic geometry, is merely the simplest way to proceed. Ultimately, we will work with the mode density, in which case the size and shape of the cavity will not appear.

We are now in position to put all these pieces together. Starting with the multi-mode description of the vector potential:

$$\vec{A}(\vec{r},t) = \sum_{\ell,\sigma}\sum a_{\ell\sigma} e^{i\vec{k}_\ell \cdot \vec{r}} + \text{complex conjugate} \tag{11}$$

and introducing the canonical variables $q_{\ell\sigma}$ and $p_{\ell\sigma}$ through the relation:

$$a_{\ell\sigma} = \frac{1}{\sqrt{4\varepsilon_0 V\omega_\ell^2}} \left(\omega_\ell q_{\ell\sigma} + ip_{\ell\sigma}\right) \tag{12}$$

we may now list the expressions for the total energy and momentum of the field in the form:

$$U = \sum_{\ell,\sigma}\sum 2\varepsilon_0 V\omega_\ell^2 a_{\ell\sigma}^* a_{\ell\sigma} = \sum_{\ell,\sigma}\sum \frac{p_{\ell r}^2 + \omega_\ell^2 q_{\ell r}^2}{2}$$

$$= \sum_{\ell,\sigma}\sum H_{\ell\sigma} = H \tag{13}$$

$$\vec{G} = \sum_{\ell,\sigma}\sum \frac{2\varepsilon_0 V\omega_\ell^2}{c} a_{\ell\sigma}^* a_{\ell\sigma}\hat{k}_\ell = \sum_{\ell,\sigma}\sum \frac{H_{\ell\sigma}}{c}\hat{k}_\ell$$

These equations indicate that so far as energy and momentum are concerned, the radiation field may be considered as a collection of oscillators, each contributing (per mode) to the total energy and momentum. We point out here that a quantum oscillator's level of excitation is given by $H_{\ell\sigma} = n_{\ell\sigma} \hbar_{\omega\ell}$, and when this condition is inserted into Eq. (13), there results the conclusion that a radiation field may be treated as a superposition of discrete photons, $n_{\ell\sigma}$ per mode, each possessing energy $\hbar\omega_\ell$, momentum $\hbar\omega_\ell/c$ and angular momentum $\pm\hbar$.

(B) Interaction Between an Atom and a Radiation Field.- A complete description of the emission and absorption of light by an atom influenced by a radiation field is well beyond the scope of this paper. The reader, interested in the details of the process, is urged to consult references (7-10). We present here only a bare outline of the approach insofar as it related to the photon counting distribution.

Starting with the complete Hamiltonian for a charged particle in an electromagnetic field:

$$H = \frac{(\vec{p} - e\vec{A})^2}{2m} + H_R + eV \qquad (14)$$

we neglect the term in e^2 and use the gauge condition $\mathrm{div}A = 0$ to reduce this to:

$$H = H_A + H_R + H_I \qquad (15)$$

where $H_A = \frac{p^2}{2m} + eV$ is the Hamiltonian of the atom, $H_R = \sum_\ell \sum_\sigma H_{\ell\sigma}$ is the Hamiltonian of the radiation field, and $H_I = -\frac{e}{m}\vec{A}\cdot\vec{p}$ is the interaction Hamiltonian. Combining the first two terms into the unperturbed Hamiltonian $H_0 = H_A + H_R$, we next treat H_I as a perturbation and attempt to solve the Schrodinger equation:

$$(H_0 + H_I)|\psi\rangle = i\hbar \frac{\partial|\psi\rangle}{\partial t} \qquad (16)$$

Using the method of first order perturbation theory, we attempt an expansion of $|\psi\rangle$ into a linear combination (with time varying coefficients) of the eigenstate $|\psi_n^0\rangle$ of the unperturbed Hamiltonian, known to satisfy the equation:

$$H_0|\psi_n^0\rangle = i\hbar \frac{\partial|\psi_n^0\rangle}{\partial t} \tag{17}$$

In this expansion we have

$$|\psi\rangle = \sum_n C_n(t) e^{-\frac{i}{\hbar}E_n t} |\psi_n^0\rangle \tag{18}$$

and the probability of finding the system in a state $|\psi_n^0\rangle$ is:

$$|C_n(t)|^2 = |\langle\psi_n^0|\psi\rangle|^2 \tag{19}$$

Assuming then that the combined system, atom plus radiation field, begins in some initial state, $|i\rangle$, Eq. (18) implies a set of coupled equations for the probability amplitude ($C_n(t)$) from which one can determine $|C_f(t)|^2$, the probability of finding the combined system in the final state $|f\rangle$. Summing over all final states, and making a number of simplifying assumptions (refs. 8, 9, 10), one ends up with Fermi's "golden rule" for the probability per second for a transition in the form:

$$\frac{dP}{dt} \cong \frac{2\pi}{\hbar} |\langle f|H_I|i\rangle|^2 \rho(E_f) \tag{20}$$

Here, $\rho(E_f)$ is the density-in-energy of the final states, and $\langle f|H_I|i\rangle$ represents the matrix element of the perturbation Hamiltonian between the initial and final states. When applied to the problem of an atom in a radiation field, one must distinguish between the cases when only the atom and both the atom and the radiation field are treated as quantized systems. In the former, the semi-classical treatment, one can correctly deduce Einstein's B coefficient for stimulated emission and absorption in terms of the electric dipole moment taken between the initial and final wave functions of the atom. On the other hand, when one also quantizes the field including the zero point fluctuation, then Eq. (20) also predicts the existence of Einstein's A coefficient for spontaneous emission.

(C) Photon Counting Statistics.- The consequence of Eq. (20) that is of importance to us is that it leads (ref. 8) to the result that in a short time, Δt, the probability of ejecting an electron from an atom on the surface of a photocathode is proportional to the incident intensity of the light, $I(t)$. That is,

$$P_1(t, t + \Delta t) = \beta I(t)\, \Delta t \tag{21}$$

For sufficiently short times $P_0(t, t + \Delta t) \cong 1-\beta\, I(t)\, \Delta t$ so that in an interval $(0, t + \Delta t)$ there are but two ways of releasing n photo-electrons, given by:

$$P_n(0, t + \Delta t) = P_{n-1}(0,t)\beta I(t)\Delta t + P_n(0,t)(1-\beta I(t)\Delta t) \tag{22}$$

Subtracting $P_n(0,t)$ from both sides and dividing by Δt before passing to the limit, we can write:

$$\frac{dP_n}{dt} = \beta I(t) P_{n-1}(t) - \beta I(t) P_n(t) \tag{23}$$

The solution to this differential-difference equation is:

$$P_n(t) = \frac{[\beta\int_0^t I(t')dt']^n e^{-\beta\int_0^t I(t')dt'}}{n!} \tag{24}$$

Now, if this process were carried out a number of times over similarly prepared realizations of the field, the average over this ensemble would lead to

$$P_n(t,T) = \int_0^\infty \frac{(\beta w)^n e^{-\beta w}}{n!} P(w)dw \tag{25}$$

where

$$w = \int_t^{t+T} I(t')\, dt'$$

and P(w)dw is the probability for w to lie in the range (w, w+dw).

(D) Mode Density.- So far as the question of density of radiation modes is concerned, we can start from one of several points of view. From the viewpoint of wave optics, light of wavelength λ emerging from a slit of width Δx can be expected to produce interference and diffraction effects over an angle $\Delta\alpha$ such that $\Delta x\, \Delta\alpha \sim \lambda$. Extending this notion to the elemental area $\Delta s = \Delta x \Delta y$ we see that:

$$\Delta\alpha\, \Delta\beta \sim \frac{\lambda^2}{\Delta s} = \frac{\Delta A}{R^2} \tag{26}$$

In terms of the "coherence area", this can be written as:

$$\Delta A \sim \frac{\lambda^2}{\frac{\Delta s}{R^2}} = \frac{\lambda^2}{\Delta\Omega} \tag{27}$$

Further, if light proceeding from Δs has a bandwidth $\Delta\nu$ then there exists a "coherence time", $\Delta t \sim 1/\Delta\nu$, corresponding to a "coherence length", $\Delta\ell = c\Delta t \sim c/\Delta\nu$. Dividing by two to take into account the two independent polarization states, we now write for the "coherence volume":

$$\Delta V = \frac{\Delta A \Delta\ell}{2} = \frac{(c\Delta t)\lambda^2}{2\Delta\Omega} = \frac{1}{2\Delta\Omega \frac{\nu^2}{c^3} \Delta\nu} \tag{28}$$

In a volume V, we expect to find $\Delta N = V/\Delta V$ modes, or in terms of mode density:

$$N_\nu = \frac{\Delta N}{V\Delta\nu} = (2)(\Delta\Omega) \frac{\nu^2}{c^3} \tag{29}$$

For isotropic radiation, this reduces to the familiar expression:

$$N_\nu = \frac{8\pi\nu^2}{c^3} \tag{30}$$

From a purely quantum statistical point of view, the elementary cell size in phase space is given by:

$$\Delta x \Delta p_x \Delta y \Delta p_y \Delta z \Delta p_z \sim h^3, \tag{31}$$

so that for a beam of photons of momentum $p = \frac{h\nu}{c}$ in a solid angle $\Delta\Omega$ about p we have:

$$\Delta x \Delta y \Delta z \sim \frac{h^3}{p^2 \Delta p \Delta\Omega} = \frac{h^3}{\frac{h^3 \nu^2}{c^3} \Delta\nu\Delta\Omega} \tag{32}$$

Dividing by two to account for the two orthogonal polarization states, we end up with, again:

$$\Delta V = \frac{1}{2\Delta\Omega \frac{\nu^2}{c^3} \Delta\nu} \tag{33}$$

It is important therefore to know how many spatial and temporal modes of the radiation field interact with the photodetector. We shall see that a single mode of chaotic, thermal radiation, and stabilized laser radiation lead, respectively, to Bose-Einstein and Poisson photocount distributions. For the case of several radiation modes, one needs to calculate the probability distribution for the sum of several random variables leading to multiple convolutions.

II. COMPOUND PHOTOCOUNT DISTRIBUTIONS

It is clear in view of the preceeding discussion that when using a quantum detector, one always has a Poisson process governing the current flow. That is, the number N_t of electrons flowing in any interval (0,t) is a random variable. If the time-space envelope of the projected EM field $|V(\tau,\underline{r})|$, (0,t) is given, then the probability density for $N_t = k$ electrons to flow in this interval is

$$P_{N_t}(k) = \frac{\left[\int_0^t\int_A \beta|V(\tau,\underline{r})|^2\, d\tau\, d\underline{r}\right]^k}{k!}\, e^{-\int_0^t\int_A \beta|V(\tau,\underline{r})|^2\, d\tau\, d\underline{r}} \tag{34}$$

If on the other hand the quantity $|V(\tau,\underline{r})|$ is random or has a random component, then equation (1) is a conditional density and must be written as $P_{N_t}(k/|V(\tau,\underline{r})|)$. To find $P_{N_t}(k)$ requires the additional averaging

$$P_{N_t}(k) = \langle P_{N_t}[k/|V(\tau,\underline{r})|]\rangle_{|V(\tau,\underline{r})|} \tag{35}$$

For the purpose of this discussion we will assume that the integration over the detector surface merely yields a constant (e.g., a point detector) and that we can write

$$\int_A\int_0^t \beta|V(\tau,\underline{r})|^2\, d\tau\, d\underline{r} = \alpha\int_0^t |a(\tau)|^2\, d\tau$$

with $\alpha = \eta/h\nu$, η the quantum efficiency, and $|a(\tau)|^2$ the instantaneous power in the received process. Notice that $|a(\tau)|$ is the envelope of the received process and that Eq. (35) really amounts to performing the final average over the statistics of the envelope.

In most communication problems (and the ones which we will consider), the function $a(\tau)$ can be expressed as the linear sum of a known signal $s(\tau)$ and a noise process $n(\tau)$. The signal may also contain a stochastic parameter, σ, to represent a channel disturbance such as fading. As is common at lower frequencies, the component $n(\tau)$ can be accurately modelled as a Gaussian noise process.

Hence, we will assume that $a(\tau)$ can be written as:

$$a(\tau,\sigma) = s(\tau,\sigma) + n(\tau)$$

which is the complex envelope of a deterministic signal plus a narrow-band Gaussian noise process, $\alpha(\tau)$, centered at some high frequency f_0;

$$\alpha(\tau,\sigma) = \mathrm{Re}[a(\tau,\sigma)e^{i2\pi f_0 \tau}].$$

It is also meaningful to expand $a(\tau,\sigma)$ in a complete orthonormal Karhunen-Loeve series (ref. 11):

$$a(\tau,\sigma) = \sum_{i=0}^{\infty} a_i(\sigma)\ \phi_i(\tau)$$

$$= \sum_{i=0}^{\infty} (s_i(\sigma) + n_i)\ \phi_i(\tau)$$

having the following properties:

(1) The $\{\phi_i(\tau)\}$ are solutions to the integral equation:

$$\lambda_i \phi_i(\mu) = \int_0^t K_n(\mu,v)\phi_i(v)dv$$

where

$$K_n(\mu,v) = E[n(\mu)n^*(v)]$$

is the covariance function of the noise and is a real function.

(2) $a_i(\sigma) = \int_0^t a(\tau,\sigma)\phi_i^*(\tau)d\tau = (a,\phi_i)$

$= (s,\phi_i) + (n,\phi_i)$.

(3) The equality is in the sense of "limit-in-the-mean".

(4) $(\phi_i,\phi_j) = \delta_{ij}$

(5) The $a_i(\sigma)$ are independent Gaussian random variables, when conditioned on σ.

The generating function of this process N_t can then be written as (ref. 12):

$$M_{N_t}(s) = E\left[e^{\alpha\int_0^t |a(t,\sigma)|^2 dt[e^\mu-1]}\right] = E\left[e^{\alpha\sum_{i=0}^{\infty}|a_i(\sigma)|^2(e^\mu-1)}\right]$$

which, using property (5) and reference (13) reduces to:

$$M_{N_t}(s) = \prod_{i=0}^{\infty} E\left[e^{\alpha|a_i(\sigma)|^2[e^\mu-1]}\right]$$

$$= \prod_{i=0}^{\infty} \frac{e^{\alpha|s_i(\sigma)|^2(e^\mu-1)/[1-\alpha\lambda_i(e^\mu-1)]}}{1-\alpha\lambda_i(e^\mu-1)} \tag{36}$$

At this point, the variable σ will be suppressed, although it must be considered as a conditioning variable when encountered in practice.

Notice that $M_{N_t}(s)$ is a product of identically distributed functions. The inverse transform of the i^{th} component is:

$$P_{N_{t_i}}(k_i) = \frac{(\alpha\lambda_i)^{k_i}}{(1 + \alpha\lambda_i)^{1 + k_i}} \, e^{\frac{-\alpha|s_i|^2}{1 + \alpha\lambda_i}} \, L_{k_i}\left(\frac{-\alpha|s_i|^2}{\alpha\lambda_i(1 + \alpha\lambda_i)}\right) \tag{37}$$

where $L_x(y)$ is the Laguerre polynomial.

(A) No Additive Noise.- In the limit as $\lambda_i \to 0$ Eq. (37) approaches:

$$\lim_{\lambda_i \to 0} P_{N_{t_i}}(k_i) = \frac{[\alpha|s_i|^2]^{k_i}}{k_i!} \, e^{-\alpha|s_i|^2}$$

and

$$P_{N_t}(k) = \frac{[\alpha \sum_{i=0}^{\infty} |s_i|^2]^k}{k!} \, e^{-\alpha \sum_{i=0}^{\infty} |s_i|^2} = \frac{(\alpha E_S)^k}{k!} \, e^{-\alpha E_S} \tag{38}$$

where $k = \sum_{i=0}^{\infty} k_i$ is the total count and $E_S = \sum_{i=0}^{\infty} |s_i|^2$ is the total signal energy in the $(0,t)$ interval. Thus the deterministic signal alone yields a Poisson distributed count. This, of course, could have been deduced immediately from Eq. (34). Notice, however, that when $|s_i|^2 = 0$,

$$P_{N_{t_i}}(k_i) = \frac{(\alpha\lambda_i)^{k_i}}{(1 + \alpha\lambda_i)^{1 + k_i}} \tag{39}$$

and each of the coordinate components is Bose-Einstein distributed (ref. 4).

In summary we see that; the signal alone can be considered to be Poisson distributed along each of its coordinate axes in Hilbert Space; Gaussian noise alone is Bose-Einstein distributed along a particular set of coordinate axes in Hilbert Space; when signal is added to the noise the resultant process is distributed according to Eq. (37) along each of the coordinate axes determined by the noise alone.

(B) Band-Limited White Gaussian Noise.- An important case occurs in communication theory when the signal and noise are passes through a filter before detection. We will consider the case where the process $a(\tau)$ is band limited by a rectangular filter with bandwidth 2B. We will also assume that the noise was initially white, with spectral density N_o.

It has been shown (refs. 12, 13, 14) that when a process is band limited and then observed over a time interval (0,t) the Eigenfunctions are prolate spheroidal wavefunctions. It has also been shown that the first (2Bt+1) of these functions accurately approximate the original function. This appears valid for values of 2Bt as low as 3 and 5 (ref. 11). Therefore, it is a good engineering approximation to assume that the eigenvalues associated with the first (2Bt+1) coordinates are each N_o with the remaining ones being zero. The generating function, $M_{N_t}(s)$ in Eq. (3) then becomes:

$$M_{N_t}(s) \simeq \frac{\exp\left[\dfrac{\alpha(s,s)(e^{\mu}-1)}{1 - \alpha N_o(e^{\mu}-1)}\right]}{[1 - \alpha N_o(e^{\mu}-1)]^{2Bt + 1}} \tag{40}$$

with the corresponding probability density being

$$P_{N_t}(k) = \frac{(\alpha N_o)^k}{(1 + \alpha N_o)^{k + 2Bt + 1}} e^{\frac{-\alpha(s,s)}{1 + \alpha N_o}} L_k^{2Bt}\left[\frac{-\alpha(s,s)}{\alpha N_o(1 + \alpha N_o)}\right] \tag{41}$$

where L_k^{2Bt} is the Laguerre function.
We will now consider some limiting forms of Eq. (41).

(C) No Signal.- In the absence of signal, Eq. (41) reduces to:

$$P_{N_t}(k) = \binom{2Bt + k}{k}\left(\frac{1}{1 + \alpha N_o}\right)^{2Bt + 1}\left(\frac{\alpha N_o}{1 + \alpha N_o}\right)^k$$

which is a negative binomial distribution. There are two important limiting cases for this distribution:

(1) $$\lim_{2Bt \to 0} P_{N_t}(k) = \frac{(\alpha N_o)^k}{(1 + \alpha N_o)^{k + 1}}$$

For 2Bt << 1, there is only one significant eigenvalue, the average value. Since this occurs when $t << \frac{1}{2B}$, it can clearly be related to the approximation

$$\int_0^t |a(\tau)|^2 d\tau \simeq |a(0)|^2 t = \lambda_o t \tag{42}$$

using the mean value theorem for integrals. This latter approximation is commonly used to obtain this result but lacks the insight as to the meaning or the range of validity.

(2) $$\lim_{\substack{2Bt \text{ large} \\ \alpha N_o \ll 1}} P_{N_t}(k) = \frac{[\alpha\, 2Bt\, N_o]^k}{k!} e^{-\alpha 2BtN_o}$$

Notice that since $2N_oB$ is the noise power, $2Bt\alpha N_o$ is the total noise energy in the $(0,t)$ interval. If we write this as $\bar{I}t$, we have:

$$\int_0^t \alpha|a(t)|^2\, dt = \bar{I}t$$

where $\bar{I}$ is in fact the time-averaged noise power

$$\bar{I} = \frac{1}{t}\int_0^t \alpha|a(t)|^2\, dt$$

Thus, for large 2Bt, there is a smoothing of the fluctuations in the noise process, and Poisson statistics prevail. The condition $\alpha N_o \ll 1$ is a little difficult to interpret, except that it implies there be much less than one noise count per degree of freedom, which is easily obtained in practice. If one recognizes that a narrow optical filter has a bandwidth on the order of 1Å at visible wavelengths, or about 100 GHz, it is clear that large 2Bt is the most common form of operation. 2Bt will be comparable in magnitude to the ratio of the optical filter and system bandwidths. Further, since almost all noise has a thermal origin,

$$\alpha N_o = \frac{\eta}{e^{\frac{h\nu}{kt}} - 1} \ll 1$$

is satisfied at optical frequencies. Actually, this is true assuming one mode of operation. However, for the purposes of this discussion we have considered a plane wave, or one spatial mode.

(D) Signal Plus Noise.- For this case, there are also two limiting conditions for Eq. (41):

$$\text{(1)} \quad \lim_{2Bt \to 0} P_{N_t}(k) = \frac{(\alpha N_o)^k}{(1 + \alpha N_o)^{1+k}} \, e^{\frac{-\alpha(s,s)}{1 + \alpha N_o}} \, L_k\left(\frac{-\alpha(s,s)}{\alpha N_o(1 + \alpha N_o)}\right)$$

As in the case for no signal, the probability density reduces to that of an individual coordinate, Eq. (37). Again, this can be interpreted as the zero order eigenvalue or average value, as in Eq. (42):

$$\text{(2)} \quad \lim_{\substack{2Bt \text{ large} \\ \alpha N_o \ll 1}} P_{N_t}(k) = \frac{[\alpha\{2BN_o t + (s,s)\}]^k}{k!} \, e^{-\alpha\{2BN_o t+(s,s)\}}$$

As might be expected from condition (C-2) and Eq. (38), the limiting condition for large 2Bt and $\alpha N_o \ll 1$ corresponds to a Poisson-distributed signal plus independent Poisson-distributed noise. Since this is the most common condition that one encounters in practice considerable effort has gone into exploring this approximation (refs. 17-20).

(E) An Equivalent Eigenspace.- Let us re-examine Eq. (37) and (41). Equation (41) is obtained as a (2Bt + 1)-fold convolution of probability densities in Eq. (37), where all the λ_i's are equal to N_o. This can be written as:

$$P_{N_t}(k) = \bigotimes_{i=0}^{2Bt} \frac{(\alpha N_o)^{k_i}}{(1 + \alpha N_o)^{1 + k_i}} e^{\frac{-\alpha |s_i|^2}{1 + \alpha N_o}} L_{k_i}\left(\frac{-\alpha |s_i|^2}{\alpha N_o (1 + \alpha N_o)}\right) \qquad (43)$$

where $\bigotimes_{i=0}^{2Bt}$ denotes a (2Bt + 1) fold convolution. Notice that the only way in which the signal enters is through the energy (s,s). Now

$$(s,s) = \int_0^t |s(\tau)|^2 \, d\tau$$

and since the signal is band limited to ±B, we can partition the (0,t) interval into (2Bt + 1) equal Δt intervals where (2Bt + 1) ΔT = t. We can then closely approximate (s,s) as:

$$(s,s) \approx \sum_{j=0}^{2Bt} |s_j(j\Delta T)|^2 \Delta T; \quad j = 0, 1, 2, \cdots, 2Bt$$

We can also write k as $k = \sum_{j=0}^{2Bt} k_j$ where k_j is the contribution of the j^{th} interval to the total count k. Equation (41) can then be decomposed into a (2Bt + 1)-fold convolution of the form:

$$P_{N_t}(k) = \bigotimes_{j=0}^{2Bt} \frac{(\alpha N_o)^{k_j}}{(1 + \alpha N_o)^{k_j + 1}} e^{\frac{-\alpha |s_j|^2 \Delta T}{1 + \alpha N_o}} L_{k_j}\left(\frac{-\alpha |s_j|^2 \Delta T}{\alpha N_o (1 + \alpha N_o)}\right) \qquad (44)$$

Notice that Eq. (44) is equivalent to Eq. (43) and would be identical if $|s_i|^2 = |s_j|^2 \, \Delta T$ for all i = j. On the other hand, Eq. (44) is meaningful as representing a processable signal formed from independent samples as opposed to an abstract eigenspace.

For the particular case where the noise process is wide sense stationary and 2Bt is large, (See, for example, reference (11)), one can approximate the eigenfunctions by harmonically related cissoids, and $|s_i|^2$ and N_o represent the Fourier coefficients of the power density spectrum. Equations (43) and (44) then express the duality of signal processing and design in both time and frequency.

We can elaborate on this duality using the time-frequency representation first considered by Gabor (ref. 21), Figure 1. The received process $a(\tau)$ considered, exists over the interval (0,t), with frequency components primarily contained in the interval (-B, +B). This is a Hilbert space of (2Bt + 1) dimensions which can be considered either as intervals of bandwidth $\frac{1}{t}$ in frequency or duration $\frac{1}{2B}$ in time. Hence, we can observe the count k_j by looking in the time interval $\left(\frac{j}{2B}, \frac{j+1}{2B}\right)$ with a filter of bandwidth 2B or we can observe the count k_i by looking in the frequency band $\left(-B + \frac{i}{t}, -B + \frac{i+1}{t}\right)$ for a time t. The first measurement is a sum of all the squares in the j^{th} column, while the latter is a sum of all the squares in the i^{th} row.

If the process is not wide sense stationary, we can still use Parseval's Theorem to write (s,s) as:

$$(s,s) = \int_0^\infty P(f)\, df \simeq \sum_{\ell=0}^{2Bt} P(\ell\, \Delta f)\, \Delta f$$

and write a density similar to Eq. (44). K_ℓ would be the total count in the band Δf in the interval $(0,t)$. However, one cannot assign the rigorous definition of power density spectrum to the noise and the noise coefficients.

We note, finally, that the most common statistical behavior encountered in practice yields $2B\Delta T >> 1$. Hence, condition (D-2) applies to any measurement interval of length ΔT. Thus, the observance of counts over many independent ΔT intervals is a sum of independent Poisson variables. This interpretation was first proposed by Reiffen and Sherman (ref. 17) on the heuristic basis, but can clearly be shown to have a solid foundation.

III. SHOT NOISE PROCESSES

We have shown that a linear relation exists between the average power $\bar{I}$ of the radiation (over some finite aperture) and the rate of flow of photons $\bar{n}$. Thus, if n is a function of time, we can write:

$$I(t) = h\nu n(t) \tag{45}$$

where h is Planck's constant and ν is the photon frequency. Thus, the detector of optical radiation can be represented either as an instantaneous power detector or as an instantaneous rate detector. This relationship is generally explained by postulating that each incident particle independently releases an electron with probability η upon arrival at the photo-detector surface, the electron in turn traveling to a cathode surface yielding a current "impulse" effect at the detector output. Thus, the total output current i(t) is due to the motion of a collection of electrons, proportional in number to the arriving particles. We can, therefore, write for the output current flow, i(t):

$$i(t) = \sum_{m=1}^{N_t} h(t - t_m) \tag{46}$$

where h(t) is the current "impulse" effect, t_m is the time of release of the m^{th} photo-electron, and N_t is the number of such electrons occurring in the total time interval $(-\infty, t)$. The function h(t) has area equal to the charge of an electron, while N_t is the counting statistic, discussed in Section II, of the photo-electron emissions. Note that if we neglect space-charge effects in the photo-detector, the travel time of each released photo-electron is finite, which means that the function h(t) must be time limited to some interval τ. That is, $h(t) = 0$ for $t < 0$ and $t > \tau$. In this case, N_t becomes the counting statistic over the finite interval $(t-\tau, t)$. Since τ is inversely related to the detector bandwidth, τ is relatively short (10^{-10} - 10^{-7} sec), and can be considered a "delta function" with respect to most modulation waveforms. It perhaps should be pointed out that if h(t) is assumed to be a flat rectangular function over $(0, \tau)$, then $i(t) = N_t - N_{t-\tau}$ and the detector output is precisely the counting process of the received optical radiation. If, instead, a nonrectangular impulse waveshape is to be accounted for, then one is forced into a closer examination of the processes described by Eq. (46). This class of processes can loosely be defined as "shot noise" processes (although the exact definition of the latter tends to vary at different points in the literature).

As discussed in Section II, the parameter N_t is a random variable depending upon the intensity of the received field. Recall that if n(t) is a deterministic function, N_t is a Poisson random variable, with mean value given by the integral of $\eta n(t)$ over $(t-\tau,t)$, and is a conditional Poisson random variable if the intensity n(t) is a sample function of a continuous stochastic process. That is, given any intensity function of the ensemble, the counting process N_t is Poisson. With Poisson counting processes the resulting shot noise processes are referred to as Poisson shot noise (PSN). Some excellent discussions of PSN processes are given by Rice (ref. 11), Middleton (ref. 22), and Papoulis (ref. 24). In essence, first- and second-order statistics, such as probability densities, moments, power spectra, and correlation functions have been well developed. For the conditional PSN, the foregoing statistical characteristics can be formally attained by taking subsequent averages over the PSN results. For example, consider the power spectrum of the conditional PSN process in Eq. (46), where the intensity n(t) is a sample function of an ensemble of positive, random, stationary process N defined over $(-\infty,\infty)$. We formally define the time averaged power density spectrum (ref. 25) of the shot noise process i(t) by:

$$\underline{\overline{S}}_i(\omega) = \lim_{T\to\infty} \frac{1}{2T} E[|I_T(\omega)|^2] \tag{47}$$

where E is the expectation operator and $I_T(\omega)$ is the Fourier Transform of i(t) over (-T,T). For the PSN processes, Eq. (47) can be readily determined as:

$$\underline{\overline{S}}_{PSN}(\omega) = \lim_{T\to\infty} \frac{1}{2T} [\overline{N}_T + |\Phi_T(\omega)|^2]|H(\omega)|^2 \tag{48}$$

where:

$$\overline{N}_T = \int_{-T}^{T} n(t)\ dt \tag{49}$$

and $H(\omega)$ and $\Phi_T(\omega)$ are the Fourier Transforms of h(t) and n(t), $-T < t < T$, respectively. The subsequent statistical average over N, and time average over T via the limiting operation, yield the power density spectrum for conditional PSN processes:

$$\underline{\overline{S}}_i(\omega) = |H(\omega)|^2 [E(N) + S_N(\omega)] \tag{50}$$

where $S_N(\omega)$ is now the time averaged power density spectrum of the stochastic intensity n(t). The foregoing results are somewhat significant, since it is valid for any counting statistic generated from conditional poisson statistics and, therefore, includes those discussed in Section III. Note that the spectrum always takes the form of the intensity spectrum immersed in a background of "noise" of spectral shape $E(N)|H(\omega)|^2$. (For infinite bandwidth detectors, $H(\omega) \approx 1$ for all ω, and the above represents basically "white" noise.) This noise constitutes the "shot noise" of the detector, and is due to the discreteness of the photoelectron model. The intensity spectrum $S_N(\omega)$, in general, contains portions due to desired intensity modulation, portions due to background effects, and associated cross-spectral terms. These latter two components constitute the "fluctuation" noise of the photo-detector output. Since the spectrum in Eq. (50) has the form of a "signal noise", there is a tendency to view the photo-detected output as signal plus *additive* noise. The difficulty, of course, is that the signal and noise are not independent, and usual "signal plus noise" interpretations, familiar to communication engineers, often lead to false conclusions (e.g., see Section IV.)

It is often instructive to examine the "instantaneous" or "short-term averaged" power spectrum of the detector output, which can be viewed basically as the conditional spectrum of Eq. (48) before the time averaging limit is taken. If we interpret the 2T interval to be the interval $(t-\tau,t)$, instead of $(-T,T)$, we see that the bracketed term in (48) will contain terms dependent on t. Furthermore, if we include the fact that the electron functions h(t) have time widths τ much shorter than the time variations in n(t), then the intensity $n_\tau(t)$ is approximately constant over $(t-\tau,t)$. Its "power spectrum" is then a delta function and the bracketed terms in Eq. (48) take the form $[k_\tau(t) + n^2(t)(\frac{\sin \omega\tau/2}{\omega\tau/2})^2]|H(\omega)|^2$. That is, the "instantaneous" spectrum (power spectrum before the time average is taken) has the appearance of a background shot noise whose level varies with time, and whose "average" value varies according to the instantaneous value of n(t). In this sense, the detector acts as an instantaneous "power" detector, which is the accepted classical definition of photo-detectors. The true frequency content of the shot noise is not exhibited, however, until the time averaging is invoked.

The foregoing discussed raises an interesting query that cannot be answered from a spectral density point of view. If the shot noise process is to represent a true intensity detector, even when n(t) is a stochastic process, then the statistical properties of the shot noise in Eq. (46) must be related to those of the intensity process n(t). When the intensity is a deterministic time function, the relations between the shot noise and its intensity are well known. However, when the intensity is itself stochastic, the manner in which the statistics of the intensity and the conditional PSN are related is somewhat vague. For example, although the first-order probability density of i(t) is difficult to write in closed form, its characteristic function is immediately available by making use of the known characteristic function of PSN (refs. 12, 23, 24). Thus:

$$\phi_{i_t}(\omega) = E_N\left[\exp \sum_{r=1}^{\infty} \frac{(j\omega)^r}{r!}\int_{t-\tau}^{t} n(\rho)h^r(t-\rho)\, d\rho\right] \qquad (51)$$

where E_N is the average over the process N. One way to interpret Eq. (51) is to assume infinite bandwidth detectors, and factor the first term of the exponential summation. Thus:

$$\phi_{i_t}(\omega) = E_N\{e^{j\omega n(t)}G[\omega,n(t)]\} \qquad (52)$$

where the G function represents the remaining factors. The average of the first term alone is precisely the characteristic function of the intensity process N at any time t. Thus, the effect of the function G is to cause a departure of the first-order probability density of i(t) from that of n(t). The conditions under which the latter effect is negligible, and the shot noise process has approximately the first order density of N, has been studied by Karp and Gagliardi (ref. 26). In this latter instance, we can say that the shot noise represents (statistically) the intensity process. This representation can be related to the "denseness" of the photon arrivals; i.e., the average number of photons per second. In fact, when the latter parameter is large, it can be shown that the bracketed term in Eq. (52) is approximately the characteristic function of a Gaussian random variable, with mean n(t) and variance n(t). This infers that the conditional (on N) probability density of i(t) at any t approaches asymptotically a Gaussian density, which again may be loosely interpreted as an instantaneous "signal" n(t), immersed in additive, non-stationary, Gaussian noise of variance n(t).

The relation between shot noise and its intensity can be further investigated by consideration of the individual moments of the two processes. The moments of the process i(t) can be obtained from its semi-invariants, which are, for PSN processes:

$$\lambda_n(t) = \int_{t-\tau}^{t} h^n(t-\rho)\,n(\rho)\,d\rho \tag{53}$$

The moments can then be obtained by the sequence of relations $E(i) = \lambda$, $E(i^2) = \lambda_2 + \lambda_1^2$, $E(i^3) = \lambda_3 + \lambda_1\lambda_2 + \lambda_1^3$, etc. For conditional PSN processes, the λ_n are themselves random processes, and the moments of i(t) depend upon the higher-order moments of the process n(t). However, if the intensities are continuous, or if the detector bandwidth is much larger than the bandwidth of the intensities, the r^{th} moments are related by:

$$E(i^r) = E(N^r) + D(r) \tag{54}$$

where D(1) = 0 and D(r), r > 2, is function depending upon the higher-order statistics of n(t) and upon the function h(t). The above relationship was investigated in reference (26). It was shown, for example, that if the function h(t) was rectangular over $(0,\tau)$ the r^{th} moment of i(t) was approximately equal to the r^{th} moment of the intensity process N if:

$$\begin{bmatrix}\text{average number of} \\ \text{photon arrivals} \\ \text{in } \tau \text{ sec}\end{bmatrix} >> \frac{r(r-1)}{2} \tag{55}$$

Equation (55) essentially states that the denseness of the shot events (i.e., the average number of h(t) functions overlapping the time interval of one function) must be sufficiently large for moment representation. The right side of Eq. (55) serves as a rough rule of thumb for determining how large this denseness must be for approximate equality of the r^{th} moment. It may be recalled (ref. 20) that for PSN processes (deterministic intensities) a condition of large number of shot occurrences is required before the PSN loses its discrete nature. Equation (55) can therefore be interpreted as the statistical equivalent of this statement; i.e., the condition under which the conditional PSN begins to take on the statistics of its intensity.

By using Eq. (54), it is also possible to relate the fluctuations in the detector output i(t) to those of the intensity n(t). Specifically, if we define the signal to noise ratio (SNR) of a positive process as the ratio of its mean value squared to the variance, then Eq. (54) leads to the fact:

$$\text{SNR of } i(t) \le \text{SNR of } n(t) \qquad (56)$$

which implies that the percent fluctuations in the shot noise are always at least as great as those of the intensity itself. We make this point mainly because the foregoing definition of SNR is commonly used in assessing signal quality in communication system analysis.

It may be noted that the conditions for which the intensity is represented by a shot noise process are also useful in "building up" intensity models as shot noise. This type of shot noise modeling has been used for studying radiation scattering and perturbation effects, (refs. 27, 28) in which the impulse functions h(t) were reinterpreted as wave packets.

With the statistics of the conditional shot noise process identified (at least in first- and second-order statistics), the problem of optimal processing procedures at the photo-detector output can now be properly formulated, and in some instances, solved. For example, the problem of optimal linear filtering of the process i(t), so as to minimize the mean squared error from the desired intensity, was considered in reference (26). For certain types of pulsed intensities, as in PCM communications optimal operations maximizing output signal to noise ratios have also been considered (ref. 29). The application of estimation theory (ref. 30), tracking operations (ref. 31), and detection procedures (refs. 17, 18, 20) to the photo-detector shot noise output has been under study, and appears to be a problem area of considerable interest from both a practical and theoretical point of view.

IV. DIGITAL COMMUNICATIONS AND OPTICAL SYSTEMS

The availability of an easily generated, extremely narrow pulse in the optical region of the spectrum suggests a natural application to communication by digital methods. This notion, in turn, has fostered an increasing interest in the application of both classical detection theory and information to optical systems. Since the output of a photodetector is a sequence of electron counts, the detection problem is formally one of decisioning in the presence of generalized poisson statistics. While early approaches to the problem basically were confined to pure Poisson counting (refs. 32, 33), more recent attention has included the generalized Laguerre counting processes in Section III.

The formulation of the general M-ary detection problem involving counting statistics proceeds as follows. The transmitter sends a signal whose intensity is modulated with one of a set of M possible intensities, each T sec. long. The received signal is corrupted by background radiation, which we assume here is white

Gaussian noise of level N_o watts per hertz per unit area, and optical bandwidth B. The output of the photodetector at the receiver is then a time varying process of electron counts, obeying a generalized Poisson distribution, as in Section III. The receiver observes the counting process for (0,t) and decides which of the M possible intensities is being received. Since K binary digits can be uniquely encoded into $2^K = M$ possible intensity waveforms, a correct decision effectively decodes K data bits. The foregoing model can be cast into a discrete format by subdividing the interval T into ΔT sec intervals ($\Delta T \approx 1/$ information bandwidth) and associating a signal energy component s_{ji} for the j^{th} intensity and i^{th} interval. (That is, s_{ji} is the total energy associated with the $2B\Delta T$ samples, or modes, of the Karhunen-Loeve expansion of the j^{th} intensity during the i^{th} ΔT interval.) Under a fixed energy constraint, we require $\sum_i s_{qi} = E$ for all q. The discrete problem then is to detect which of the possible intensity vectors $\underline{s}_q = \{s_{qi}\}$ is controlling the counting process by observing the sequence of independent counts $\underline{k} = \{k_i\}$, $i = 1,2,\cdots,M(= T/\Delta T)$. Under a maximum likelihood detection criterion, and a priori equally likely signals, the optimal test is to form the likelihood functionals $\Lambda_q(\underline{k})$ and select $\underline{s}_q$ as being transmitted if no other $\Lambda_i(\underline{k})$ exceeds $\Lambda_q(\underline{k})$. If a likelihood draw occurs (more than one $\Lambda_q(\underline{k})$ is maximum) any randomized choice among the maxima can be used. From Eq. (37), the likelihood test is therefore equivalent to comparing:

$$\Lambda_q(\underline{k}) = \prod_{i=1}^{M} L_{k_i}^{2B\Delta T} \left(\frac{(-s_{qi})}{N_o(1 + \alpha N_o)} \right) \tag{57}$$

for all q, where s_{qi} is now a normalized signal intensity obeying the constraint $\Sigma s_{qi} = E \equiv N$. In typical operation, $2B\Delta T >> 1$ (i.e., the optical bandwidth is much greater than the information bandwidth) and Eq. (1) is approximately:

$$\Lambda_q(\underline{k}) \cong \prod_{i=1}^{M} \left[2B\Delta T + \frac{s_{qi}}{N_o(1 + \alpha N_o)}\right]^{k_i} \Big/ k_i! \tag{58}$$

After observing $\underline{k}$, examination of the above set of $\{\Lambda_q\}$ for maxima is equivalent to the comparison of the set of functions $\Sigma k_i \log\left(1 + \frac{s_{qi}}{K}\right)$, where $K = 2BN_o\Delta T$ represents the noise energy per counting interval per unit area. (Recall it was previously shown in Section II that under the condition $2B\Delta T >> 1$ the counts k_i are Poisson variates so that complete statistics of the foregoing test can be determined.)

An indication of the performance of a detection test is given by the divergence, or "expected distance between hypothesis". The divergence is formally defined as:

$$D_{jq} = E_{\underline{k}}(\Lambda_{jq}|j) - E_{\underline{k}}(\Lambda_{jq}|q) \tag{59}$$

where $\Lambda_{jq} = \Lambda_j(\underline{k}) - \Lambda_q(\underline{k})$ and $E_{\underline{k}}(\Lambda|j)$ is the conditional average of Λ over $\underline{k}$ given the intensity $\underline{s}_j$. Abend (ref. 18) had shown that for Poisson counting, using the functions of (58) and $M = 2$ (binary detection), the divergence normalized by the variance of Λ, is maximized by a "pulsed" type of intensity, in which the available signal energy is wholly concentrated in a single counting interval. That is, an intensity set defined by:

$$\underline{s}_q = \{N\delta_{qi}\} \tag{60}$$

where δ_{qi} is the Kronecker delta function. Kailath (ref. 19) extended this result by showing that under a total energy constraint, other suitable forms of distance are maximized by similar pulsed intensities. Gagliardi and Karp (ref. 20) applied an average divergence criterion to the general M-ary poisson detection problem and again showed the optimality of the intensity set of Eq. (60). In the latter reference, the intensity set that maximized the probability of correctly detecting the true intensity, rather than maximizing divergence, was also considered, and shown to correspond to the pulsed set in two special cases, (1) $M = 2$ with symmetric intensity sets, and (2) any M and low intensity to noise energy ratio. However, the determination of global optimal intensity sets in the pure Poisson case, based upon detection probability still remains a difficult task. It has been conjectured by many that the pulsed set of Eq. (60) is, in fact, a global optimal set, but to the authors' knowledge a rigorous proof has not been shown. The optimality of the pulsed set, even under this special crition, is somewhat significant, since it indicates the importance of intensity waveshape in digital system design. This, of course, is partly due to the general advantage of orthogonal signals in detectability, a property afforded by the disjointness of the pulsed set in Eq. (60). The use of signals placed in adjacent time slots is in essence a pulse position modulated system in which each position corresponds to a digital word. The dual of such a system (a frequency keyed system), which also retains the orthogonality property, can similarly be generated by redefining the expansion functions of the received field (ref. 34).

It should be pointed out that if the condition $2B\Delta T >> 1$ is not valid, care must be used in accepting the pulsed set of Eq. (60) as an optimal intensity set. In particular, the Poisson assumption and the use of Eq. (58) is violated. For the case of $2B\Delta T << 1$, the divergence in Eq. (59) must be obtained by averaging terms as in Eq. (57) over the Laguerre densities. If this averaging is carried out, Eq. (59) takes the form:

$$D_{qj} \cong C \left\{ \prod_{i=1}^{M} I_o\left(\frac{s_{qi}}{N_o}\right) + \prod_{i=1}^{M} I_o\left(\frac{s_{ji}}{N_o}\right) - 2 \prod_{i=1}^{M} I_o\left(\frac{s_{qi}\, s_{ji}}{N_o}\right)\right\} \tag{61}$$

where $I_o(x)$ is the imaginary Bessel function of zero order and C contains terms common to all q and j. Now it is no longer immediately evident that the pulsed set of Eq. (60) maximized D_{qj}. The last term, however, is minimized if either $s_{qi} = 0$ or $s_{ji} = 0$ for all i, which suggests a disjoint intensity set, but it is not evident that the first terms are maximized under the same condition. The difficulties of this problem are quite reminiscent of similar difficulties in attempting to find optimal signal sets in non-coherent additive Gaussian noise channels.

When the pulsed set of Eq. (60) is used, and the general Laguerre counting is assumed, the analysis procedures are similar to the Poisson case. It is easy to show the monotonicity of

Laguerre functions with respect to their indices. It then follows from Eq. (57) that $\Lambda_q \gtrless \Lambda_i$ if $L^{\alpha}_{k_q}\left(\frac{N}{N_o}\right) \gtrless L^{\alpha}_{k_i}\left(\frac{N}{N_o}\right)$ which, in turn, is true if $k_q \gtrless k_i$. Hence, the maximum likelihood test need only count over each interval, selecting the signal corresponding to the interval with the largest count.

Error Probabilities with Pulsed Intensity Sets and Poisson Counting

The performance of the pulsed intensity set in M-ary detection can be evaluated by considering the error probability when Poisson counting statistics are assumed. This can be obtained by noting that for the pulsed intensity set of Eq. (60) the log of the likelihood functions for each k_i constitutes a set of independent Poisson random variables. The variable for k_q has intensity $(N + 2B\alpha N_o \Delta T)$ if the q^{th} intensity was sent, and has intensity $K = 2B\alpha N_o \Delta T$ otherwise. Recall that if the q^{th} intensity is sent a correct decision will be made with probability $1/(r + 1)$ of the log likelihood equals r others and exceeds the remaining $M-(r + 1)$. Therefore, upon considering all possibilities, the error probability can be derived as (ref. 20):

$$P_E(E,K,M) = 1- \frac{e^{-(N + MK)}}{M} - \sum_{x=1}^{\infty} \left[\frac{(N + K)^x \, e^{-(N + K)}}{x!}\right] \bullet \left[\sum_{t=0}^{x-1} \frac{K^t \, e^{-K}}{t!}\right]^{M-1} \bullet \left[\frac{(1 + a)^M - 1}{aM}\right] \qquad (62)$$

where:

$$a = \left[\frac{K^x}{x! \sum_{t=0}^{x-1} \frac{K^t}{t!}}\right]$$

The function $P_E(N,K,M)$ has been plotted by Pratt (ref. 32) for $M = 2$, and recently a digital computation has been generated (ref. 23) for a complete plot of the function. An exemplary plot is shown in Figure 2 in which $P_E(N,3,M)$ is plotted for various M as a function of N. It is important to note that P_E depends on both the normalized signal energy N and the normalized noise energy K in the counting interval, and not simply on their ratio. This fact is emphasized in Figure 3, in which $P_E(N,K,2)$ is plotted as a function of K for 2 fixed ratios N/K. This dependence on both signal and noise energies distinguishes the Poisson detection problem from the analogous coherent Gaussian channel problem. Note that the interfering noise energy K depends only upon the background energy in the interval ΔT, which is the width of the transmitted intensity pulse. The prime advantage of Poisson systems is precisely their ability to remove the effect of background noise by making ΔT small, and has been emphasized in previous reportings (refs. 36, 37).

The actual dependence of P_E on the parameter ΔT has been considered (ref. 38), and the improvement in error probability with decreasing ΔT has been demonstrated. The improvement, of course, is made at the expense of information bandwidth and peak power, both inversely proportional to ΔT. Surprisingly, the improvement is quite small at low values of N, and the increase in bandwidth may not be worth the decrease obtained in error probability. The effect on error probability of additive extraneous thermal noise in the decisioning system and statistical characteristics of photomultipliers has also been considered (ref. 38).

For Laguerre counts, Eq. (62) must be rederived using the Laguerre densities discussed in Section III. Recently, general bounds or the error probability in this latter case, using the orthogonal (disjoint) signal intensity sets, have been reported (ref. 34).

Information Rate of a Poisson PPM System

We have so far analyzed only one aspect of system performance, i.e., error probabilities. The actual information rate that the link achieves is another important design consideration. As stated, the transmitter sends optical energy in one of M time intervals, which is ΔT seconds wide, thereby transmitting one of M possible signals in $M\Delta T$ seconds, or at a rate $\log_2 M/M\Delta T$ bit/s. The receiver correctly determines the true signal with probability $1-P_E$ and is in error with probability P_E. Because of symmetry, the erroneous signal may be equally likely interpreted as any of the M-1 incorrect signals. Thus, the overall channel may be depicted as an M-ary symmetric channel, in which each of the M possible transmitter signals is converted to itself with probability $1-P_E$ and converted to each other signal with probability $P_E/(M-1)$. The information rate for such a channel is known to be:

$$H = \frac{\log_2 M + P_E \log_2 [P_E/(M - 1)] + (1 - P_E) \log_2 (1 - P_E)}{M\Delta T} \qquad (63)$$

For convenience we shall denote this as:

$$H = C(N,K,M)/M\Delta T \tag{64}$$

to emphasize the dependence of the numerator on the stated parameters. By using Eq. (63) and the families of error probability curves as in Figure 2, the rate H can be evaluated by straightforward substitution. Although specific curves for such a computation are not shown here, if suffices to note that if N and K are such that $P_E < 10^{-1}$, then Eq. (63) is, to a good approximation:

$$\begin{aligned} H &\approx (1 - P_E)\ [\log_2 M)/M\Delta T] \\ &= (\log_2 M)/M\Delta T - P_E[(\log_2 M)/M\Delta T]. \end{aligned} \tag{65}$$

If we interpret the rate H as the source rate minus the equivocation of the channel, then the PPM optical system behaves approximately as if a source rate of log M/MΔT is passed into a channel of equivocation P_E log M/MΔT. As noted in Eq. (62), even if $K \to 0$ (no background interference), $P_E \to \exp(-N)/2$, so that the equivocation is not due entirely to the background noise.

The use of Eq. (63) and the previous equations are helpful in determining the rate, given operating parameters. However, the converse design problem, which is to determine particular parameter values that achieve a desired rate, is not so straightforward. This is due to the fact that the rate is a somewhat complicated function of the parameters. We shall consider here two aspects of this design problem that have practical application under certain operating conditions. First, the word period T = MΔT is held fixed while the information bandwidth $1/\Delta T$ is allowed to vary, and second, the bandwidth is held fixed while the word period is allowed to vary. In both cases, we are interested in the relationship between the rate H and the transmitter parameters N and M, assuming that the noise power is held fixed.

Fixed Work Period

We assume here that ΔT is allowed to vary with M so as to maintain $T = M\Delta T$ constant. Thus, the system "squeezes" more signals into the T-second period as M increases. The resulting rate is then:

$$H = C(N, K_T/M, M)T \tag{66}$$

where K_T is the noise energy in T. Thus, the rate depends only upon the numerator of Eq. (63). With N fixed, increasing M increases the source rate, but the error probability also increases and eventually reaches an asymptotic value of:

$$P_E = \left(1 + \frac{N[K_T - 1 + \exp(-K_T)]}{K_T}\right) \exp(-N)$$

for large M. The resulting system rate increases, to within a constant of the entropy of the alphabet, $\log_2 M/T$. Therefore, it is clear that if the bandwidth is expendable, one will always increase the system rate for large M by increasing M. In a practical system, this implies that one should operate with as wide a bandwidth as possible to fully exploit the capability of the PPM system. We are, therefore, led naturally to consider the design of a system for an arbitrary rate H, when the full bandwidth ($1/\Delta T$) of the system is limited.

Fixed Bandwidth

In this case, ΔT is held constant (thereby fixing the noise energy K in ΔT) so that both the numerator and denominator in Eq. (63) depend upon M, and the rate degrades quickly as M increases due to the log M/M dependence. A given rate, e.g., H_0, may be obtained by many different combinations of N and M. Analytically, these equivalent operating points may be obtained graphically by noting that they are the values for which the numerator C(N,T,M), considered as a function of M, intersects the straight line $H_0 \Delta T M$. By plotting these functions for various N, their intersection will identify (N,M) pairs which achieve the rate H_0. One may then decide on a particular operating point by invoking suitable design criteria. For example, one may select the smallest M from among the candidate pairs, which then minimizes the word period $T = M\Delta T$. Alternatively, one may choose to minimize the average transmitter power per information bit, which is proportional to N/C. In the latter case, therefore, one would select the operating pair (N,M) for which N/C is minimal. The latter parameter is recognized as the β-efficiency parameter (energy per data bit) of a communication system (ref. 39). If the value of N/C, corresponding to the optimal (N,M) pair is tabulated, the results can be compared to previously derived performance based upon the same parameter. This type of comparison was considered (ref. 40) and it was shown that the PPM system outperformed an optical heterodyne system for sufficiently large M, approaching in fact the minimum β generated by the Gordon bound for quantum systems. This type of result further emphasizes the importance of expending system bandwidth (increasing M also implies increasing information bandwidth) to improve overall performance. The effect of Laguerre statistics (when the information bandwidth approaches the optical bandwidth) and the effect of additive noise can be accounted for by modifying these Poisson results (ref. 40).

The extension of the discrete model for optical detection, assumed almost entirely in the aforementioned references, to the continuous model has received little attention. In usual procedures, the continuous case is generated from the discrete by taking limits of infinitely small intervals. Although this procedure can be properly structured to generate the continuous version of the counting process, the continuous process representing the photodetector output must be viewed entirely as a shot noise process (see Section III). Unfortunately, such processes have first order densities that are expressible only through transforms of their characteristic function. Hence, the building up of a general detection model based upon shot noise, rather than discrete, processes would be severely hampered by the inability to express observable statistics. It would appear, however, that shot noise detectability cannot continue to be avoided when consideration is given to operation with information bandwidths on the order of optical detector bandwidths. This aspect of detection deserves more attention in future research studies.

V. ANALOG COMMUNICATIONS

The major portion of work in the area of analog communications for optical systems, has centered on first- and second-moment theory, spectral analysis, and signal-to-noise ratios. We have already discussed spectral analysis for shot noise processes with emphasis on signal representation. For the remainder of the paper we will concentrate on trying to bring together some of these ideas in a unified way, leaning heavily on physical motivation.

Before turning to the analyses required, it is very instructive to reconsider the behavior of a photodetector from a phenomenological point of view. As we have already seen an important parameter in a photodetector is the time ΔT over which the intensity fluctuations remain relatively constant. This is related to the bandwidth, B, of the optical signal by $\Delta T \leq \frac{1}{2B}$. When an electron is released from the detecting surface and flows through the ensuing circuitry, there is always the fixed electron charge e. This fixes the area of the resulting current pulse. Hence, higher energy electrons will flow faster, the current pulses will be narrower in time resulting in an increased frequency response of the detector.

Generally, one thinks of counting circuitry as literally counting each of these events. On the other hand, one can also consider the following viewpoint: suppose we "match" the detector response, B_d, to the incident radiation bandwidth, $\Delta T \simeq \frac{1}{2B} = \frac{1}{2B_d}$. Then, each current pulse created will be approximately ΔT seconds wide. Hence, at any time t, the effects of all pulses from the previous ΔT seconds will still be present. Therefore, if k_i

electrons flow in the interval $(t_i - \Delta T, t_i)$ than at the time t_i the value of the current can be approximated by $k_i \frac{e}{\Delta T}$, or since $\frac{k_i}{\Delta T} \simeq \alpha\tilde{I}(t_i)$, $i(t_i) \simeq \alpha e\tilde{I}(t_i)$, which was shown earlier. If the response of the detector were square pulses, this description would be exact. On the other hand, the distortions occurring due to end effects are the normal effects of filtering. The so-called shot noise represents the fact that K_i is an integer, making $\tilde{I}(t_i)$ take on discrete values, whereas the true $I(t)$ would be continuous.

The previous argument was intended to justify consideration of the (2Bt + 1) Nyquist samples for analog processes also. It was shown in Eq. (55) that these samples can also be considered statistically independent.

(A) Maximizing Signal-to-Noise Ratio for Direct Detection.- For maximum likelihood detection, the optimum form of processing consisted of weighting the counts on each of the (2Bt + 1) intervals. We will, therefore, consider the form of processing where each k_j is weighted by the number β_j. The processed signal then becomes v, where:

$$v = \sum_{j=0}^{2Bt} \beta_j K_j . \tag{67}$$

As a criterion for signal processing, we will use the signal-to-noise ratio defined as:

$$\frac{S}{N} = \frac{E^2[v]\,|_{N_o}}{\text{var}\ [v]} = 0 \tag{68}$$

Thus, the mean of v in the absence of noise can be obtained from Eq. (37) and is:

$$E[v]|_{N_o = 0} = \alpha \sum_{j=0}^{2Bt} \beta_j |s_j(j\Delta T)|^2 \, \Delta T \tag{69}$$

with the variance being:

$$\mathrm{var}[v] = \alpha \sum_{j=0}^{2Bt} \beta_j^2 \left\{ \left(|s_j(j\Delta T)|^2 + N_o' \right) + \alpha \left(N_o'^2 + 2N_o' |s_j(j\Delta T)|^2 \right) \Delta T \right\} \Delta T \tag{70}$$

Thus, the signal-to-noise ratio becomes:

$$\frac{S}{N} = \frac{\left\{ \alpha \sum_{j=0}^{2Bt} \beta_j |s_j(j\Delta T)|^2 \, \Delta T \right\}^2}{\alpha \sum_{j=0}^{2Bt} \beta_j^2 \left\{ \left(|s_j(j\Delta T)|^2 + N_o' \right) + \alpha \left(N_o'^2 + 2N_o' |s_j(j\Delta T)|^2 \right) \Delta T \right\} \Delta T} \tag{71}$$

which can be bounded by using the Schwarz inequality. Hence:

$$\frac{S}{N} \leq \sum_{j=0}^{2Bt} \frac{\alpha \{|s_j(j\Delta T)|^2\}^2 \, \Delta T}{|s_j(j\Delta T)|^2 + N_o' + \alpha (N_o'^2 + 2N_o' |s_j(j\Delta T)|^2) \, \Delta T} \tag{72}$$

with the equality holding when:

$$\beta_j = \frac{|s_j(j\Delta T)|^2}{|s_j(j\Delta T)|^2 + N_o' + \alpha (N_o'^2 + 2N_o' |s_j(j\Delta T)|^2) \, \Delta T} \tag{73}$$

Notice that in the absence of noise $N_o = 0$:

$$\left(\frac{S}{N}\right) \leq \alpha \sum_{j=0}^{2Bt} |s_j(j\Delta T)|^2 \, \Delta T = \alpha E_s = \frac{\eta E_s}{h\nu} \tag{74}$$

This however, is the average number of photoelectron counts in the (0,t) interval and is generally referred to as the quantum-limited signal-to-noise ratio.

Let us now rewrite the right-hand side of Eq. (72) as:

$$\left(\frac{S}{N}\right) \leq \sum_{j=0}^{2Bt} \frac{\alpha |s_j(j\Delta T)|^2 \, \Delta T}{1 + \alpha N_o + \dfrac{1}{\alpha |s_j(\Delta T)|^2 \, \Delta T} \{\alpha N_o' + \alpha^2 N_o^2\}} \leq \alpha E_s \tag{75}$$

Recall now that αN_o is the number of noise counts per ΔT interval and for thermal noise sources is much less than one. In addition, $\alpha |s_j(j\Delta T)|^2 \, \Delta T$ is the average number of signal counts in the j^{th} ΔT interval. Suppose, therefore, that we construct a signal:

$$|s_j(\Delta T)|^2 = \frac{E_s}{\Delta T} \quad \text{for one value of } j$$

$$= 0 \quad \text{for all other values of } j.$$

Then clearly:

$$\sum_{j=0}^{2Bt} |s_j(\Delta T)|^2 \, \Delta T = E_s$$

is not violated, and in addition:

$$\left(\frac{S}{N}\right) = \frac{\alpha E_s}{1 + \alpha N_o + \left[\frac{\alpha N_o + (\alpha N_o)^2}{\alpha E_s}\right]} \approx \alpha E_s \tag{76}$$

for all values of $\alpha E_s > \alpha N_o$. Thus, low duty-cycle operation is preferable when maximizing the signal-to-noise ratio of detected radiation in the absence of detector noise.

The addition of independent thermal noise with temperature T at the detector output changes the variance in Eq. (70). After some manipulation to take into account the electron charge e, the bandwidth and the load R, the signal to noise ratio in Eq. (76) can be written as:

$$\frac{S}{N} = \frac{\alpha E_s}{1 + \alpha N_o + \left[\frac{\alpha N_o + (\alpha N_o)^2}{\alpha E_s}\right] + \frac{kT}{e^2 \ \alpha E_s \ RB_d}} \tag{77}$$

The quantity $kT/e^2 \ \alpha E_s \ RB_d$ is, in general, much greater than one. Therefore, except under extreme conditions of temperature, impedance bandwidth, and signal level, a normal detector will be "thermal noise limited" in operation and S/N will be much less than αE_s.

We have been considering the case where each sampling interval represented one mode. If, in fact, each interval contained L modes, then clearly we need only replace αN_o by L αN_o everywhere.

(B) Direct Detection with Photomultiplication.- We have just shown that most detectors are inherently thermal noise limited except under extreme temperature conditions. This was true because the current that was released by the surface immediately encountered a thermal environment. There are devices, presently limited to the visible region of the spectrum which impart a preamplification to the photocurrent before the thermal environment is met. The most common device, a photomultiplier tube, consists of a cascade of stages through which each emitted electron passes and is amplified many thousands of times. When the effect of an electron emitted at the cathode reaches the anode, it appears as an actual current pulse well above the anode thermal environment. It is, therefore, possible to view the effects of individual electrons. These devices are commonly referred to as "photon counters".

To first order, one can account for this amplification A, by assuming an electron charge equal to Ae. Then we can see from (Eq. (77) that the term which previously made the device thermal noise limited becomes:

$$\frac{kT}{A^2 e^2 \alpha E_s \; RB_d}$$

Thus, if the gain of the device is such that the inequality:

$$A > \sqrt{\frac{kT}{e^2 (\alpha E_s) RB_d}} \approx 3 \times 10^6 \sqrt{\frac{T}{(\alpha E_s) B_d}}$$

is satisfied, it is again shot-noise limited. In practice, the gain is a random variable and an "excess noise" appears because of the finite variance of A. This, however, only causes changes on the order of 20% or about 1 db, and for the purposes of this discussion can be ignored.

(C) Heterodyne Detection.- If the electric field of a local oscillator is aligned coincident with the received signal over the detector surface, then one can directly add the two electric fields. Thus, if we designate the signal by $E_1 e^{j\omega_1 t + \phi(t)}$ and the local oscillator by $E_{LO} e^{j\omega_2 t}$ then:

$$s_j(j\Delta T) = E_1 e^{j\omega_1 (j\Delta T) + \phi(j\Delta T)} + E_{LO} e^{i\omega_2 (j\Delta T)}$$

and:

$$|s_j(j\Delta T)|^2 = |E_1|^2 + |E_{LO}|^2 + 2|E_1||E_{LO}|\cos\{(\omega_1-\omega_2)j\Delta T+\phi(j\Delta T)\} \tag{78}$$

If the local oscillator is made large, then it can be shown that, under these conditions, the density in Eq. (41) approaches a Gaussian density with a mean value of $2\alpha|E_1||E_{LO}|\cos\{(\omega_1-\omega_2)j\Delta T + \phi(j\Delta T)\}$ (excluding the dc component) and variance $\alpha|E_{LO}|$ multiplied by the bandwidth considered. Then, if the bandwidth of the signal in Eq. (78) is 2W, and the bandpass of the detector is greater than $(\omega_1-\omega_2) + W$, one can pass the detected signal through a bandpass filter centered at $(\omega_1-\omega_2)$ with bandpass 2W and recreate the signal $2\alpha E_1 E_{LO} \cos\{(\omega_1-\omega_2)j\Delta T + \phi(j\Delta T)\}$. The resulting carrier signal to noise ratio will be:

$$\left(\frac{S}{N}\right)_{het} = \frac{\frac{1}{2}(2\alpha|E_1||E_{LO}|)^2}{\alpha|E_{LO}|2W} = \frac{\alpha|E_1|^2}{W} = \frac{\eta|E_1|^2}{h\nu W}$$

which can again be recognized as the quantum limited condition.

(D) Power Spectrum Analysis.- In Section III, it was shown that the time-averaged power density spectrum of the current could be written as:

$$\bar{S}_i(\omega) = |H(\omega)|^2 \; [E(N) + S_N(\omega)].$$

Since $S_N(\omega)$ is the spectrum of a non-negative definite function (the normalized power), it can be written in terms of a dc and an ac component. The ac component is, $S_{AC}(\omega)$ where:

$$S_{AC}(\omega) = (\eta\bar{n}e)^2 \; \Phi_M(\omega)$$

and $n(t)$ has been normalized to:

$$n(t) = \bar{n}(1 + m(t)); \; m(t) \geq -1$$

with:

$$\int_{-T}^{T} m(t) \; dt = 0,$$

and $\Phi_M(\omega)$ the time-average power density spectrum of $m(t)$. Notice that the modulation index is included in $m(t)$. For an unmodulated source, such as noise, $m(t) \equiv 0$, and only the shot noise term and the dc remain. Thus, if we have a signal plus additive noise impinging on the detector, where the average noise rate is designated $\bar{n}_n$, the power density spectrum, minus the dc terms, can be written as:

$$S_T(\omega) = e^2|H(\omega)|^2 \; [\eta(\bar{n}_n + \bar{n}) + (\eta\bar{n})^2 \; \Phi_M(\omega)] + \frac{2kT}{R}$$

where we have also included the thermal noise contribution. If we define the signal-to-noise ratio as the ratio of the total signal power:

$$\frac{(e\eta\bar{n})^2}{2\pi} \int_{2W'} \Phi_M(\omega)\, d\omega$$

over the bandwidth of the signal, divided by the total non-signal power over the same bandwidth:

$$\frac{1}{2\pi} \int_{2W'} \left[e^2 |H(\omega)|^2 \eta(\bar{n} + \bar{n}_n) + \frac{2kT}{R}\right] d\omega$$

then, assuming that $|H(\omega)|^2$ is "flat" over the $2W'$ region of interest:

$$\frac{S}{N} = \frac{(\eta\bar{n}e)^2 \{\frac{1}{2\pi}\int \Phi_M(\omega)\, d\omega\}}{[e^2\eta(\bar{n} + \bar{n}_n) + \frac{2kT}{R}]\ 2W} \leq \frac{(\eta\bar{n})}{\left[1 + \frac{\bar{n}_n}{\bar{n}} + \frac{2kT}{e^2 R\eta\bar{n}}\right] 2W} \qquad (79)$$

where W is now the cyclic frequency. Notice again that if e is replaced by Ae and the shot noise term $2\eta A^2e^2\bar{n}W = 2\eta AeI_{DC}W > 4kTW$, the device will be again shot-noise limited. The term:

$$\frac{\eta\bar{n}}{2W} = \frac{\eta P}{2h\nu W}$$

can again be recognized as being related to the quantum limited condition.

SUMMARY REMARKS

We have tried to present in this paper a review of the basic concepts in optical communications viewed strictly from a classical point of view, in the absence of any channel effects. In this vein, we have viewed the received signal as an electromagnetic field and described its interaction with a photodetector. We then described some of the fundamental properties of the resulting current flow as seen by the communications engineer.

The treatment in this paper is not complete, since the study of this problem has not finished. Consequently, some portions have been given more emphasis than others, while some have been omitted entirely. For example, in the literature the topic of continuous estimation for shot noise processes has barely been touched (ref. 13). The same is true for synchronization in a shot noise environment (ref. 31), although this will be fundamental to any sophisticated optical communications system.

What has been attempted, rather, was a presentation which answered the questions concerning the physical modelling of the system and a reduction to the terms most useful for analysis. Where such analysis had reached a level of conveying a reasonably complete understanding of an aspect of the problem, it was also presented. It is hoped that this paper is thorough enough to motivate additional research in this area.

REFERENCES

(1) Glauber, R. J.: The Quantum Theory of Optical Coherence. Physical Review, vol. 130, no. 6, June 1963, pp. 2529-2539.

(2) Glauber, R. J.: Coherent and Incoherent States of the Radiation Field. Physical Review, vol. 131, no. 6, September 1963.

(3) Glauber, R. J.: Optical Coherence and Photon Statistics. Quantum Optics and Electronics, C. de Witt, et al., eds., Gordon and Breach, New York, 1965, pp. 65-185.

(4) Mandel, L., and Wolf, E.: Coherence Properties of Optical Fields. Reviews of Modern and Physics, vol. 37, no. 2, April 1965, pp. 231-287.

(5) Klauder, J. R., and Sudarshan, E. C. G.: Fundamentals of Quantum Optics. W. A. Benjamin, Inc., New York, 1968.

(6) Troup, G. J.: Optical Coherence Theory. Methuen, 1967.

(7) Louisell, W. H.: Radiation and Noise in Quantum Electronics. McGraw-Hill, 1964.

(8) Mandel, L., Sudarshan, E. C. G., and Wolf, E.: Proc. Phys. Soc., vol. 84, no. 435, 1964.

(9) Fermi, E.: Rev. Mod. Phys., vol. 4, no. 87, 1932.

(10) Fano, V.: Am. J. Phys., vol. 29, no. 539, 1961.

(11) Van Trees, H. L.: Detection, Estimation, and Modulation Theory. Part I, John Wiley and Sons, Inc., New York, 1968.

(12) Rice, S. O.: Mathematical Analysis of Random Noise. Bell System Technical Journal, vol. 23, 1944, pp. 282-332.

(13) Karp, S., and Clark, J. R.: Photon Counting: A Problem in Classical Noise Theory. IEEE Trans. on Information Theory, January 1970.

(14) Slepian, D., and Pollak, H. O.: Prolate Spheroidal Wave Function, Fourier Analysis, and Uncertainty, I. Bell Systems Technical Journal, vol. 40, 1961, pp. 43-64.

(15) Landau, H. J., and Pollak, H. O.: Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty, II. Bell Systems Technical Journal, vol. 40, 1961, pp. 65-84.

(16) Landau, H. J., and Pollak, H. O.: Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty, III. Bell Systems Technical Journal, vol. 41, 1962, pp. 1295-1336.

(17) Reiffen, B., and Sherman, H.: An Optimum Demodulator for Poisson Processes. Proceedings of the IEEE, vol. 51, no. 10, October 1963, p. 1319.

(18) Abend, K.: Optimum Photon Detection. IEEE Transactions on Information Theory, vol. 12, no. 1, January 1966, pp. 64-65.

(19) Kailath: The divergence and Bhattacharyya distance measures in signal selection. IEEE Transactions Communication Technology, vol. COM-15, February 1967, pp. 52-60.

(20) Gagliardi, R. M., and Karp, S.: M-ary Poisson Detection and Optical Communications. IEEE Transactions on Communication Technology, COM-17, no. 2, April 1969, pp. 208-216.

(21) Gabor, D.: Theory of Communication. JIEE, vol. 93(III), November 1946, pp. 429-457.

(22) Middleton, D.: Introduction to Statistical Communication Theory. McGraw-Hill Book Co., Inc., New York, 1960.

(23) Parzen, E.: Stochastic Processes. Holden-Day, Inc., San Francisco, California, 1962.

(24) Papoulis, A.: Probability, Random Variables, and Stochastic Processes. McGraw-Hill Book Co., Inc., New York, 1965.

(25) Ibid: Chapter 10.

(26) Karp, S., and Gagliardi, R. M.: On the Representation of a Continuous Stochastic Intensity by Poisson Shot Noise. IEEE Transactions on Information Theory, March, 1970.

(27) Karp, S., Gagliardi, R. M., and Reed, I. S.: Radiation Models Using Discrete Radiation Ensembles. Proceedings of the IEEE, vol. 56, no. 10, October 1968, pp. 1704-1711.

(28) Middleton, D.: A Statistical Theory of Reverberation and First Order Scattered Fields. Part I and Part II, IEEE Trans. Information Theory, vol. IT-13, July 1967, pp. 372-414.

(29) Chang, R.: Photon Detection for Optical PCM Systems. IEEE Trans. on Information Theory, vol. IT-15, no. 6, November 1969, pp. 725-728.

(30) Bar-David, I.: Communicating Under the Poisson Regime. IEEE Trans. on Information Theory, vol. IT-15, January 1969, pp. 31-38.

(31) Gagliardi, R. M.,

(32) Pratt, W. K.: Binary detection in an optical polarization modulation communication channel. IEEE Trans. Communication Technology (Concise Papers), vol. COM-14, October 1966, pp. 664-665.

(33) Ross, M.: Laser Receivers. John Wiley & Sons, Inc., New York, 1966.

(34) Liu, J.: Reliability of Quantum Mechanical Communication Systems. IEEE Trans. Information Theory, May 1970.

(35) Karp, S., et al.: Error Probabilities for Poisson Detection. NASA Tech. Note TN-D-4721, October 1968.

(36) Ross, M.: Pulse interval modulation laser communications. Presented at the Eastcon Convention, Washington, D. C., October 1967.

(37) Karp, S., and Gagliardi, R. M.: A low duty cycle optical communication system. Presented at the Eastcon Convention, Washington, D. C., October 1967.

(38) Karp, S., and Gagliardi, R. M.: The Design of a PPM Optical System. IEEE Trans. on Comm. Tech., vol. COM-17, December 1969, p. 670.

(39) Sanders, R. W.: Communication Efficiency Comparison of Several Communication Systems. Proc. IRE, vol. 48, no. 4, April 1960, pp. 575-588.

(40) Karp, S.: Communication Efficiency of Quantum Systems. NASA Tech. Report TR-R-R-320, September 1969.

USC Engineering