Single-threshold detection of a random signal in noise with multiple independent observations. 1: Discrete case with application to optical communications

Paul R. Prucnal and Malvin Carl Teich

A single-threshold processor is derived for a wide class of classical binary decision problems involving the likelihood-ratio detection of a signal embedded in noise. The class of problems we consider encompasses the case of multiple independent (but not necessarily identically distributed) observations of a nonnegative (nonpositive) signal, embedded in additive, independent, and noninterfering noise, where the range of the signal and noise is discrete. We show that a comparison of the sum of the observations with a unique threshold comprises optimum processing, if a weak condition on the noise is satisfied, independent of the signal. Examples of noise densities that satisfy and violate our condition are presented. The results are applied to a generalized photocounting optical communication system, and it is shown that most components of the system can be incorporated into our model. The continuous case is treated elsewhere [IEEE Trans. Inf. Theory **IT-25**, (March, 1979)].

I. Introduction

The likelihood-ratio detection of a signal embedded in noise constitutes an important class of classical binary decision problems that has found widespread applicability in the synthesis and analysis of many types of systems.[1] These applications range from optical communications[2–19] and radar systems[18–21] to sensory detection in visual[18,22–25] and auditory[18,26,27] psychophysics. For complex signal and noise statistics, it is sometimes difficult or impossible to express the likelihood ratio in closed form, however. Even for simple signal and noise statistics, direct implementation of the likelihood ratio as an optimum processor may be rather difficult. It may be possible to reduce the likelihood ratio to a simpler, but equivalent processor by using various *ad hoc* geometric arguments or lengthy algebraic manipulations.

It is the purpose of this paper to derive a remarkably simple processor that is optimum for a broad range of classical binary decision problems involving the likelihood-ratio detection of a signal embedded in noise. The class of problems we consider encompasses the case of N independent (but not necessarily identically distributed) observations of a nonnegative (or nonpositive) signal random variable embedded in an additive, independent, and noninterfering noise random variable, where the range of the signal and noise is discrete. We show that a comparison of the sum of the N observations with a unique threshold comprises optimum processing, provided that the logarithm of the noise probability density does not contain a point of inflection. This condition on the noise probability density is not necessary, but is sufficient, to imply our single-threshold processor and does not depend on the signal probability density. The results are applicable to a spatial array of detectors exposed to a temporal sequence of observations. We show by example that in many cases it is not difficult to test the log of the noise density for a point of inflection analytically. In more difficult cases, a graphical representation of the noise density with a logarithmic ordinate scale may be useful in revealing a point of inflection. We apply the results to a generalized photocounting optical communication system and show that background noise, dark noise, modulation, avalanche multiplication, and channel distortions are easily included in our model.

We have previously[18] derived a limited version of the results presented here for a single observation ($N = 1$) of a nonnegative signal embedded in noise, when the logarithm of the noise density is concave downward. The proof was based on the existence of a nonunique continuous extension of the noise density, so that implementation of the result depended on a proper choice of this continuous extension. In the present paper, no such ambiguity exists. We have eliminated the need

The authors are with Columbia University, Department of Electrical Engineering & Computer Science, New York, New York 10027.

Received 30 May 1978.
0003-6935/78/1115-3576$0.50/0.

for a continuous extension by applying a finite-difference condition directly to the discrete noise density. The continuous case of N observations is considered elsewhere[28] since it differs substantially from the discrete case presented here.

We consider the following general classical binary detection problem. Each of two source outputs corresponds to a hypothesis, H_0 or H_1. To decide which hypothesis is true, based on the Bayes or Neyman-Pearson criterion, optimum processing of the observation vector $\mathbf{n}$ is the well known likelihood-ratio test[1]

$$\Lambda(\mathbf{n}) = \frac{p(\mathbf{n}|H_1)}{p(\mathbf{n}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda, \tag{1}$$

where $\Lambda(\mathbf{n})$ represents the likelihood ratio, $p(\mathbf{n}|H_i)$ is the probability density of $\mathbf{n}$ given that H_i is true, and λ is a constant dependent on the choice of decision criterion. The observation vector $\mathbf{n} = (n_1, \ldots, n_N)$ consists of N independent observations, which may arise from a spatial array of N_s detectors sampled during a sequence of N_t time intervals, in which case $N = N_s N_t$.

In the simplest situation $N = N_s = N_t = 1$, so that a single detector samples a single observation n_1. In this case $\Lambda(n_1)$ may be graphically represented by a curve in a 2-D Cartesian coordinate system. In Sec. II we consider a condition on the noise density which implies that $\Lambda(n_1)$ is monotonic with respect to n_1. The monotonicity of $\Lambda(n_1)$ implies, in turn, that Eq. (1) is equivalent to the single-threshold processor

$$n_1 \underset{H_0}{\overset{H_1}{\gtrless}} \lambda', \tag{2}$$

with threshold λ'. Equation (2) completely specifies the optimum processing of n_1.

For the case of multiple observations ($N > 1$), we visualize $\Lambda(\mathbf{n})$ as an N-dimensional surface in $N + 1$ space. An N-dimensional hyperplane, orthogonal to the Λ axis at λ, cuts through the surface $\Lambda(\mathbf{n})$. This is illustrated in Fig. 1 for $N = 2$. Given an observation $\hat{\mathbf{n}}$, the test given by Eq. (1) is equivalent to determining whether $\Lambda(\hat{\mathbf{n}})$ is located above or below the hyperplane: if it is located above, H_1 is chosen; if it is located below, H_0 is chosen. The projections of the intersections of the hyperplane and $\Lambda(\mathbf{n})$ partition the remaining N coordinates into N-dimensional decision regions R_1 and R_0, corresponding to the regions where $\Lambda(\mathbf{n})$ is above the hyperplane (H_1 is chosen) and below the hyperplane (H_0 is chosen), respectively. The decision is then based upon the region in which the tip of the observation vector $\hat{\mathbf{n}}$ falls. In Fig. 1, R_0 is represented by the cross-hatched region and R_1 by the unshaded region. If there are multiple intersections of the surface and the hyperplane, as in Fig. 1, multiple boundaries divide the decision regions R_0 and R_1.

In Sec. II, we prove that if the same condition on the noise density considered for $N = 1$ applies to each component of the N-dimensional noise density, $\Lambda(\mathbf{m})$ is monotonic with respect to

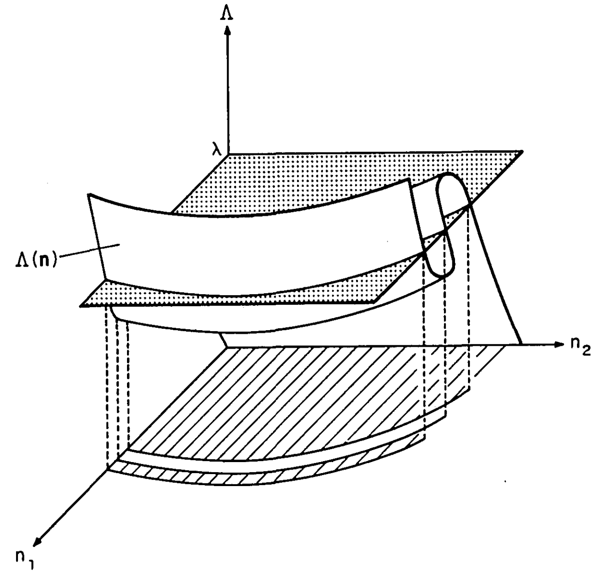

Fig. 1. The likelihood ratio $\Lambda(\mathbf{n})$ vs the observations n_i for the case $N = 2$. The solution $\Lambda(\mathbf{n}) = \lambda$ is represented by the multiple curved intersections of $\Lambda(\mathbf{n})$ with the dotted plane. The decision regions R_0 are cross-hatched and represent the coordinates (n_1, n_2) for which $\Lambda(\mathbf{n}) < \lambda$. The decision regions R_1 are unshaded and represent the coordinates (n_1, n_2) for which $\Lambda(\mathbf{n}) \geq \lambda$. This case exhibits multiple curved decision boundaries.

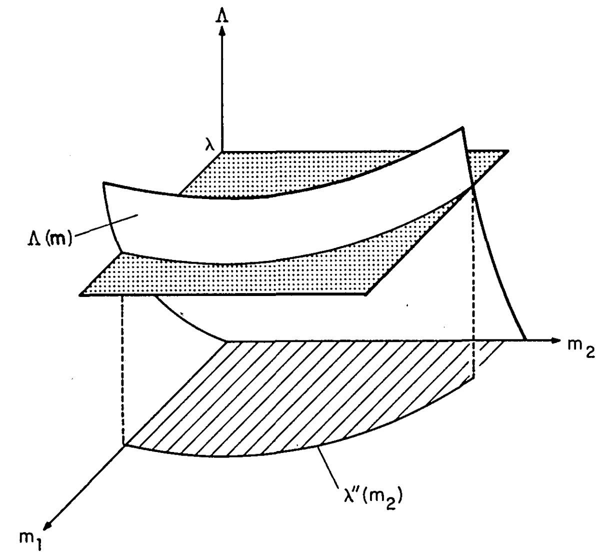

Fig. 2. The transformed likelihood ratio $\Lambda(\mathbf{m})$ for the case $N = 2$ where $\Lambda(\mathbf{m})$ is monotonic with respect to m_1. The solution $\Lambda(\mathbf{m}) = \lambda$ is represented by a single curved intersection of $\Lambda(\mathbf{m})$ with the dotted plane. The region R_0 is cross-hatched and represents the coordinates (m_1, m_2) for which $\Lambda(\mathbf{m}) < \lambda$. The decision region R_1 is unshaded and represents the coordinates (m_1, m_2) for which $\Lambda(\mathbf{m}) \geq \lambda$. This case exhibits a decision boundary $\lambda''(m_2)$ which is single valued, and therefore single-threshold processing.

$$m_1 = \sum_{i=1}^{N} n_i,$$

as illustrated in Fig. 2. (Here the likelihood ratio has been transformed to the coordinate system $m_1, \ldots, m_N$.) This implies that the decision regions R_0 and R_1 are partitioned by a boundary λ'', which is a single-valued function of $m_2, \ldots, m_N$, as in Fig. 2. In this case, therefore, Eq. (1) is equivalent to the single-threshold processor

$$m_1 \underset{H_0}{\overset{H_1}{\gtrless}} \lambda''(m_2, \ldots, m_N) = \lambda''(\mathbf{n}), \tag{3}$$

where $\lambda''(\mathbf{n})$ is single valued.

This single-threshold processor does not completely specify optimum processing as does the single-threshold processor in the $N = 1$ case, since λ'' is now a function of $\mathbf{n}$. However, Eq. (3) does assure the uniqueness of the threshold in contrast to the nonmonotonic case of Fig. 1. Note that if $N = 1$, Eq. (3) reduces to Eq. (2). In Sec. III, we examine a number of noise densities to determine whether single-threshold processing is optimum.

The transformed likelihood ratio $\Lambda(\mathbf{m})$ may depend explicitly only on the coordinate m_1 (in which case the decision boundaries in 3-D space would be straight lines). The quantity m_1 then contains all the information necessary to make a decision and is therefore a sufficient statistic. If, in addition, the conditions discussed in Sec. II are satisfied, optimum detection is completely specified by the comparison

$$\sum_{i=1}^{N} n_i \underset{H_0}{\overset{H_1}{\gtrless}} \lambda', \tag{4}$$

with threshold λ'.

A sufficient condition on $\Lambda(\mathbf{n})$, which implies that m_1 is a sufficient statistic, has been considered for the continuous case of N observations.[28] Extension of the results presented here from two to M hypotheses does not appear to be straightforward.

II. Single-Threshold Processing for Discrete Distributions with *N* Observations

Let H_1 represent the presence of a signal with probability density $p_{Si}(s_i)$, embedded in noise with probability density $p_{Di}(d_i)$, and let H_0 represent the absence of a signal (noise alone). The noise is within the discrete range $b \le d_i \le c$, and the signal is within the discrete range $g \le s_i \le h$. We assume that the signal and noise random variables are additive, independent, and noninterfering. The probability density of $n_i = s_i + d_i$ under each hypothesis is then

$$p_i(n_i|H_1) = \sum_{k=u_0}^{u_1} p_{Di}(n_i - k) p_{Si}(k) \tag{5}$$

and

$$p_i(n_i|H_0) = p_{Di}(n_i), \tag{6}$$

where $u_0 = \max(n_i - c, g)$ and $u_1 = \min(n_i - b, h)$. We further assume that the n_i are statistically independent, though not necessarily identically distributed, so that the likelihood-ratio test in Eq. (1) becomes

$$\Lambda(\mathbf{n}) = \prod_{i=1}^{N} \Lambda_i(n_i) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda, \tag{7}$$

with

$$\Lambda_i(n_i) = p_i(n_i|H_1)/p_i(n_i|H_0). \tag{8}$$

Substituting Eqs. (5) and (6) into Eq. (8), we obtain

$$\Lambda_i(n_i) = \left[\sum_{k=u_0}^{u_1} p_{Di}(n_i - k) p_{Si}(k)\right] \Big/ p_{Di}(n_i). \tag{9}$$

We now prove that if the noise distribution satisfies either the simple finite difference[29] condition

$$\Delta^2[\log p_{Di}(d_i)]_k \le 0 \qquad \forall k \ge 0, \forall d_i, \forall i, \tag{10}$$

or

$$\Delta^2 [\log p_{Di}(d_i)]_k \ge 0 \quad \forall k \ge 0, \forall d_i, \forall i, \tag{11}$$

the test

$$\sum_{i=1}^{N} n_i \underset{H_0}{\overset{H_1}{\gtrless}} \lambda''(\mathbf{n}), \tag{12}$$

is optimum. Thus, if the logarithm of the noise distribution does not contain a point of inflection (in the discrete sense), single-threshold processing is optimum. It must be kept in mind, however, that Eqs. (10) and (11) represent only a sufficient condition. Therefore, we cannot rule out the possibility of a situation in which Eq. (12) may hold even when neither Eq. (10) nor Eq. (11) is satisfied. We have reason to suspect, however, that Eqs. (10) and (11) may represent necessary conditions as well.[28]

Using the definition of the second finite difference[29] in Eq. (10), we obtain

$$[\log p_{Di}(n_i + 1 - k) - \log p_{Di}(n_i + 1)] - [\log p_{Di}(n_i - k) - \log p_{Di}(n_i)] \underset{\le}{\ge} 0 \qquad \forall k \underset{\le}{\ge} 0, \forall n_i, \forall i, \tag{13}$$

where the equation is ≥ 0 if $k \ge 0$, and the equation is ≤ 0 if $k \le 0$, and where $p_{Di}(d_i)$ has been evaluated at n_i. The difference of logarithms can be reexpressed as a ratio of their arguments, and since the logarithm is monotonic, Eq. (13) is equivalent to a comparison of the ratios of these arguments, i.e.,

$$\frac{p_{Di}(n_i + 1 - k)}{p_{Di}(n_i + 1)} - \frac{p_{Di}(n_i - k)}{p_{Di}(n_i)} \underset{\le}{\ge} 0 \qquad \forall k \underset{\le}{\ge} 0, \forall n_i, \forall i. \tag{14}$$

Equation (11) leads to an expression that is identical to Eq. (14) with one set of inequalities reversed (e.g., $\forall k \lesseqgtr 0$).

Forming the finite difference of the likelihood ratio in Eq. (9), we obtain

$$\Delta[\Lambda_i(n_i)] = \sum_{k=u_0'}^{u_1'} \frac{p_{Di}(n_i + 1 - k) p_{Si}(k)}{p_{Di}(n_i + 1)} - \sum_{k=u_0}^{u_1} \frac{p_{Di}(n_i - k) p_{Si}(k)}{p_{Di}(n_i)}, \tag{15}$$

where $u_0' = \max(n_i + 1 - c, g)$ and $u_1' = \min(n_i + 1 - b, h)$. Rearranging terms yields

$$\Delta[\Lambda_i(n_i)] = -\frac{(u_0' - u_0)p_{Di}(n_i - u_0)p_{Si}(u_0)}{p_{Di}(n_i)} + \frac{(u_1' - u_1)p_{Di}(n_i + 1 - u_1')p_{Si}(u_1')}{p_{Di}(n_i + 1)} + \sum_{k=u_0'}^{u_1} p_{Si}(k)\left[\frac{p_{Di}(n_i + 1 - k)}{p_{Di}(n_i + 1)} - \frac{p_{Di}(n_i - k)}{p_{Di}(n_i)}\right]. \quad (16)$$

The first term on the right-hand side (RHS) of Eq. (16) disappears if $n_i < b$, $n_i > c$, or $n_i \leq c + g - 1$ (so that $u_0 = u_0' = g$). Therefore, the first term on the RHS of Eq. (16) disappears $\forall n_i$ if $g \geq 1$ or $c = \infty$. The second term on the RHS of Eq. (16) disappears if $n_i < b - 1$, $n_i > c - 1$, or $n_i \geq b + h - 1$ (so that $u_1 = u_1' = h$). Therefore, the second term on the RHS of Eq. (16) disappears $\forall n_i$ if $h \leq 0$ or $b = -\infty$. According to Eq. (14) the integral on the RHS of Eq. (16) is nonnegative if Eq. (10) is satisfied and $k \geq 0$, or Eq. (11) is satisfied and $k \leq 0$, and is nonpositive if Eq. (10) is satisfied and $k \leq 0$, or Eq. (11) is satisfied and $k \geq 0$. The restriction $k \geq 0$ requires that $u_0' \geq 0$ (this is true if $g \geq 0$), and the restriction $k \leq 0$ requires that $u_1 \leq 0$ (this is true if $h \leq 0$).

Combining the above requirements, we find that

$$\Delta[\Lambda_i(n_i)] \geq 0 \qquad \forall n_i, \forall i \quad (17)$$

if Eq. (10) is satisfied, and the signal is nonnegative, or if Eq. (11) is satisfied, the signal is nonpositive, and the upper limit of the noise is infinity. If the upper limit of the noise $c \neq \infty$, Eq. (17) holds $\forall n_i < c$. Similarly, we find that

$$\Delta[\Lambda_i(n_i)] \leq 0 \qquad \forall n_i, \forall i \quad (18)$$

if Eq. (10) is satisfied and the signal is nonpositive, or if Eq. (11) is satisfied, the signal is nonnegative, and the lower limit of the noise is minus infinity. If the lower limit of the noise $b \neq -\infty$, Eq. (18) holds $\forall n_i > b$. In accordance with the discussion preceding Eq. (2), Eqs. (17) and (18) imply that single-threshold processing is optimum for the case $N = 1$.

We now consider an N-dimensional discrete coordinate system defined by the orthonormal basis $\mathbf{E}_1, \ldots, \mathbf{E}_N$, where $\mathbf{E}_1 = N^{-1/2}(1, \ldots, 1)$ and $\mathbf{E}_i = (e_{i1}, \ldots, e_{iN})$. The basis vectors $\mathbf{E}_2, \ldots, \mathbf{E}_N$ are selected by an orthonormalization procedure.[30] The proof is carried out for a general set of such basis vectors since the result is independent of the particular choice of $\mathbf{E}_2, \ldots, \mathbf{E}_N$. An arbitrary vector $\mathbf{m}$ in this coordinate system is expressible as a linear combination of the $\mathbf{E}_i$. Transformation of the $\mathbf{n}$-coordinate system into the $\mathbf{m}$-coordinate system is defined by

$$\mathbf{m} = A\mathbf{n}, \quad (19)$$

where

$$A = \begin{bmatrix} N^{-1/2} & N^{-1/2} & \cdots & N^{-1/2} \\ e_{21} & e_{22} & \cdots & e_{2N} \\ \vdots & \vdots & & \vdots \\ e_{N1} & e_{N2} & \cdots & e_{NN} \end{bmatrix}. \quad (20)$$

In particular,

$$m_1 = N^{-1/2} \sum_{i=1}^{N} n_i. \quad (21)$$

Since the rows of A form a basis, they are linearly independent, and A^{-1} exists. The orthonormality of the rows of A insures that $A^{-1} = A^T$, so that the inverse transformation is

$$\mathbf{n}(\mathbf{m}) = A^T\mathbf{m}, \quad (22)$$

with components

$$n_i(\mathbf{m}) = \sum_{k=1}^{N} e_{ki}\, m_k. \quad (23)$$

The partial difference of a multivariable finite valued function $f(\mathbf{n})$ with respect to a single variable n_i will be represented by the notation

$$\Delta_{n_i}[f(\mathbf{n})] \equiv f(n_1, \ldots, n_i + 1, \ldots, n_N) - f(n_1, \ldots, n_i, \ldots, n_N). \quad (24)$$

Using Eq. (7), the partial difference of $\Lambda[\mathbf{n}(\mathbf{m})]$ with respect to m_j is

$$\Delta_{m_j}\{\Lambda[\mathbf{n}(\mathbf{m})]\} = \Delta_{m_j}\left\{\prod_{i=1}^{N} \Lambda_i[n_i(\mathbf{m})]\right\}. \quad (25)$$

Using the product rule for finite differences,[31] Eq. (25) may be rewritten as

$$\Delta_{m_j}\{\Lambda[\mathbf{n}(\mathbf{m})]\} = \left\{\prod_{k=1}^{N-1} \Lambda_k[n_k(\mathbf{m})]\right\} \Delta_{m_j}\{\Lambda_N[n_N(\mathbf{m})]\} + \sum_{i=2}^{N-1}\left(\left\{\prod_{k=1}^{i-1} \Lambda_k[n_k(\mathbf{m})]\right\} \Delta_{m_j}\{\Lambda_i[n_i(\mathbf{m})]\}\left\{\prod_{k=i+1}^{N} \Lambda_k[n_k(\mathbf{m}) + 1]\right\}\right) + \Delta_{m_j}\{\Lambda_1[n_1(\mathbf{m})]\}\left\{\prod_{k=2}^{N} \Lambda_k[n_k(\mathbf{m}) + 1]\right\}. \quad (26)$$

From Eq. (23), the partial difference of $n_i(\mathbf{m})$ with respect to m_j is

$$\Delta_{m_j}[n_i(\mathbf{m})] = \Delta_{m_j}\left(\sum_{k=1}^{N} e_{ki} m_k\right), \quad (27)$$

which can be expanded, using Eq. (24), yielding

$$\Delta_{m_j}[n_i(\mathbf{m})] = e_{ji}(m_j + 1) + \sum_{\substack{k=1 \\ k\neq j}}^{N} e_{ki} m_k - \sum_{k=1}^{N} e_{ki} m_k. \quad (28)$$

Combining terms, Eq. (28) becomes

$$\Delta_{m_j}[n_i(\mathbf{m})] = e_{ji}. \quad (29)$$

Using the chain rule for finite differences (see Appendix), the partial difference of $\Lambda_i[n_i(\mathbf{m})]$ with respect to m_j is

$$\Delta_{m_j}\{\Lambda_i[n_i(\mathbf{m})]\} = \Delta_{n_i}[\Lambda_i(n_i)]\, \Delta_{m_j}[n_i(\mathbf{m})]. \quad (30)$$

Substituting Eq. (29) into Eq. (30), we obtain

$$\Delta_{m_j}\{\Lambda_i[n_i(\mathbf{m})]\} = e_{ji}\Delta_{n_i}[\Lambda_i(n_i)]. \quad (31)$$

Setting $j = 1$ and using Eq. (20), Eq. (31) becomes

$$\Delta_{m_1}\{\Lambda_i[n_i(\mathbf{m})]\} = N^{-1/2}\, \Delta_{n_i}[\Lambda_i(n_i)]. \quad (32)$$

From Eq. (21), the partial difference of m_1 with respect to

$$\sum_{i=1}^{N} n_i$$

is

$$\Delta_{\sum n_i}[m_1(n_i)] = \Delta_{\sum n_i}(N^{-1/2}\textstyle\sum n_i), \quad (33)$$

where it is understood that the summations extend from $i = 1$ through $i = N$. From Eq. (24), it is apparent that the difference operator Δ is linear, so that the constant $N^{-1/2}$ can be brought outside of the operation. Equation (33) then simplifies to

$$\Delta_{\Sigma n_i}[m_1(n_i)] = N^{-1/2}. \tag{34}$$

We now use the chain rule (see Appendix) to provide

$$\Delta_{\Sigma n_i}\{\Lambda_i[n_i(\mathbf{m})]\} = \Delta_{m_1}\{\Lambda_i[n_i(\mathbf{m})]\}\,\Delta_{\Sigma n_i}[m_1(n_i)] \tag{35}$$

and combine Eqs. (32), (34), and Eq. (35) to yield

$$\Delta_{N^{1/2}m_1}[\Lambda_i(n_i)] = N^{-1}\Delta[\Lambda_i(n_i)]. \tag{36}$$

In accordance with Eq. (17) $\Delta[\Lambda_i(n_i)]$ is nonnegative $\forall n_i$, $\forall i$; since N is also nonnegative it follows that

$$\Delta_{N^{1/2}m_1}[\Lambda_i(n_i)] \geq 0 \qquad \forall n_i,\ \forall i. \tag{37}$$

Since $\Lambda_k[n_k(\mathbf{m})]$ is nonnegative, using Eqs. (37) and (26) it is apparent that

$$\Delta_{\Sigma n_i}[\Lambda(\mathbf{n})] \geq 0 \quad \forall n_i. \tag{38}$$

If, instead, Eqs. (11) and (18) are satisfied, $\Delta[\Lambda_i(n_i)]$ is nonpositive, $\forall n_i$, $\forall i$. Since N is nonnegative it follows that

$$\Delta_{N^{1/2}m_1}[\Lambda_i(n_i)] \leq 0 \ \forall n_i,\ \forall i. \tag{39}$$

Since $\Lambda[n_k(\mathbf{m})]$ is nonnegative, using Eqs. (39) and (26) it is apparent that

$$\Delta_{\Sigma n_i}[\Lambda(\mathbf{n})] \leq 0 \quad \forall n_i. \tag{40}$$

Equations (38) and (40) indicate that $\Lambda(\mathbf{n})$ is either entirely monotonic nondecreasing or entirely monotonic nonincreasing, with respect to Σn_i, so that the test

$$\sum_{i=1}^{N} n_i \underset{H_0}{\overset{H_1}{\gtrless}} \lambda''(\mathbf{n}) \tag{41}$$

is optimum, in accordance with the discussion preceding Eq. (3). Therefore, if Eq. (10) or Eq. (11) is satisfied and if the conditions stated at the beginning of Sec. II are adhered to, single-threshold processing is optimum.

III. Discussion

In this section, we consider optimum processing of the observation vector $\mathbf{n}$ for a number of different noise densities. If $p_{Di}(d_i)$ satisfies Eq. (10) or (11) and the conditions stated in Sec. II are adhered to, single-threshold processing is optimum. As indicated in Sec. II, the N components of the noise density need not be identically distributed. Though single-threshold processing is optimum for most of the usual noise densities encountered, we also cite counterexamples for which our single-threshold processor does not necessarily apply. For convenience, we use the natural logarithm in Eqs. (10) and (11), though the logarithm with an arbitrary base may be used.

The hypergeometric noise density,[32] for $d_i < r$ and $d_i < a_1$, is

$$p_{Di}(d_i) = \mathrm{u}(d_i)\frac{\binom{a_1}{d_i}\binom{a-a_1}{r-d_i}}{\binom{a}{r}}, \tag{42}$$

with mean $\langle d_i\rangle = ra_1/a$ and variance $\langle(\Delta d_i)^2\rangle = \langle d_i\rangle(1 - a_1/a)(a-r)/(a-1)$, where the Heavyside unit step function $\mathrm{u}(d_i)$ is unity for $d_i \geq 0$ and zero otherwise.

In the limit where r^2/a, d_i^2/a, and $(r-d_i)^2/(a-a_1)^2$ all approach zero in such a manner that $a_1/a = p$, where $0 < p < 1$, the hypergeometric density reduces to the binomial density.[32] The binomial density has mean $\langle d_i\rangle = rp$ and variance $\langle(\Delta d_i)^2\rangle = \langle d_i\rangle(1-p)$, corresponding to the mean and variance of the hypergeometric density with $a_1/a = p$ and $r^2/a \to 0$. Using the hypergeometric density in Eq. (42)

$$\Delta^2[\ln p_{Di}(d_i)]_k = \ln\left[\frac{(d_i+1-k)(a_1-d_i)(r-d_i)(a-a_1-r+d_i+1-k)}{(d_i+1)(a_1-d_i+k)(r-d_i+k)(a-a_1-r+d_i+1)}\right] < 0 \quad \forall d_i \geq 0, \tag{43}$$

which satisfies Eq. (10), so that single-threshold processing is optimum for both the hypergeometric and binomial noise densities.

The Polya noise density,[32] with arbitrary real constants $q > 0$ and $0 < p < 1$ and arbitrary integer constant $r > 0$, is

$$p_{Di}(d_i) = [\mathrm{u}(d_i) - \mathrm{u}(d_i - r)]\frac{\binom{-p/q}{d_i}\binom{-(1-p)/q}{r-d_i}}{\binom{-1/q}{r}}, \tag{44}$$

where $0 \leq d_i \leq r$. In the limit where $r \to \infty$, $p \to 0$, and $q \to 0$, in such a manner that $rp \to \langle d_i\rangle$ and $rq \to \delta = \langle d_i\rangle/M$, for real $M \geq 1$, the Polya density reduces to the negative binomial noise density[33]

$$p_D(d_i) = \mathrm{u}(d_i)\binom{M+d_i-1}{d_i}\left(\frac{M}{M+\langle d_i\rangle}\right)^M\left(\frac{\langle d_i\rangle}{M+\langle d_i\rangle}\right)^{d_i}, \tag{45}$$

with mean $\langle d_i\rangle$ and variance $\langle d_i\rangle + \langle d_i\rangle^2/M$. If we define $a_1 = M/(M+\langle d_i\rangle)$ and $a_2 = M + d_i$, and let $a_1 \to 0$, and $a_2 \to \infty$, with $a_1a_2 \to d_i$, a_1 is the constant of proportionality between the negative binomial and its continuous analog, the gamma density[28] with $\beta = 1$. When the negative binomial is used to represent the photon-counting distribution for chaotic light, the degeneracy parameter δ represents the average number of photons per cell of phase space, the parameter M represents the number of modes, or degrees of freedom, and contains information relative to the spatio-temporal coherence and polarization properties of the light, the flash duration and area, and the detector integration time and area.[33] In general, the parameter M is real (and ≥ 1), but when M is restricted to positive integers only, Eq. (45) is known as the Pascal[32] density. When $M = 1$, the negative binomial becomes the Bose-Einstein[33] (geometric) density with mean $\langle d_i\rangle$ and variance $\langle d_i\rangle + \langle d_i\rangle^2$. For $M \gg 1$, and $\delta \ll 1$, the negative binomial reduces to the Poisson density,[33,34] with mean and variance $\langle d_i\rangle$. (The Poisson can also be obtained from the binomial discussed earlier, if $a_1 = 1$, $r \to \infty$, and $a \to \infty$, so that $p = 1/a$ and $\langle d_i \rangle = r/a$.) Alternatively, for $M \gg 1$, and $\delta > 1$, the negative binomial reduces to

$$p_{Di}(d_i) = u(d_i)(2\pi\delta)^{-1/2} d_i^{-3/2} \langle d_i \rangle \times \exp[-(2\delta d_i)^{-1}(d_i - \langle d_i \rangle)^2] \qquad \delta \geq 1, \quad (46)$$

which was obtained by Glauber,[35] and which we refer to as the Glauber density.

Using the Polya noise density, given in Eq. (44), we obtain

$$\Delta^2[\ln p_{Di}(d_i)]_k = \ln \frac{d_i + 1 - k}{d_i + 1} + \ln \frac{(p/q) + d_i}{(p/q) + d_i - k} + \ln \frac{r - d_i}{r - d_i + k} + \ln \frac{[(1-p)/q] + r - d_i - 1 + k}{[(1-p)/q] + r - d_i - 1}. \quad (47)$$

From Eq. (47) we determine that if

$$1 - q \lessgtr p \lessgtr q, \quad (48)$$

then

$$\Delta^2[\ln p_{Di}(n_i)]_k \gtrless 0 \qquad \forall d_i \qquad 0 \leq d_i \leq r. \quad (49)$$

For $q \geq 1$, Eq. (48) has the less-than sense, so that Eq. (49) has the greater-than sense, and, according to Eq. (11), single-threshold processing is optimum. If $q \leq \min(p, 1-p) < 1$, Eq. (49) has the less-than sense, so Eq. (10) is satisfied. If $\max(p, 1-p) \leq q \leq 1$, Eq. (49) has the greater-than sense, Eq. (11) is satisfied. Therefore, single-threshold detection is optimum for the Polya noise density provided that q does not lie within the interval $(p, 1-p)$. In the limit where the Polya reduces to the negative binomial, $rp \to \langle d_i \rangle$ and $rq \to \langle d_i \rangle / M$, where $M \geq 1$, so that $q \to p/M \leq p$. Since $p \to 0$, $q \leq \min(p, 1-p)$, so that our single-threshold processor applies. Although the proof of the single-threshold processor could be carried out independently for each limiting case of the negative binomial, it is not necessary, since no restriction has been placed on M or on $\langle d_i \rangle / M$ in the derivation of the general result. To summarize, single-threshold processing is optimum for the negative binomial, Pascal, Bose-Einstein (geometric), Poisson, and Glauber noise densities and for the Polya noise density provided that q does not lie within the interval $(p, 1-p)$. If q lies within the interval $(p, 1-p)$ the single threshold detector does not necessarily apply for the Polya density.

As an example of a noise density that is always concave upward, consider the photon-counting distribution arising from a sinusoidally modulated chaotic light source,[36] with unity modulation depth, and mean $\langle d_i \rangle$,

$$p_{Di}(d_i) = u(d_i)(2d_i)! \, \langle d_i \rangle^{d_i} / [2^{d_i}(d_i!)^2(1 + 2\langle d_i \rangle)^{d_i + 1/2}] \quad (50)$$

from which

$$\Delta^2[\ln p_{Di}(d_i)]_k = \ln \frac{(2d_i + 1 + 2k)(d_i + 1)}{(2d_i + 1)(d_i + 1 + k)} > 0 \qquad \forall d_i. \quad (51)$$

Equation (51) satisfies Eq. (11), so single-threshold processing is optimum.

Finally, a noise density function, which never satisfies Eq. (10) or Eq. (11), is the photon-counting distribution arising from a sinusoidally modulated single-mode laser source.[36] This can be concluded from a plot of the logarithm of the density, which has an inflection point, so that the single-threshold processor does not necessarily apply to this case.

Noise densities that cannot be expressed in closed form are difficult to test analytically for a point of inflection. The Neyman Type-A density,[26] for example, contains an infinite sum. In such cases it may be possible to inspect a plot of the noise density with a logarithmic ordinate for a point of inflection.

IV. Optical Communication System

Our method is most powerful for the $N = 1$ case, for which the optimum receiver structure is completely specified by the single-threshold processor [Eq. (2)], if the logarithm of the noise density does not contain an inflection point. For example, if the Neyman-Pearson criterion[1] is used, λ' is the unique fixed solution to

$$P_F = \sum_{n_i = \lambda'}^{\infty} p_{Di}(n_i) \leq \alpha, \quad (52)$$

where the false-alarm rate P_F is constrained to be less than the constant α. Since λ' is a fixed solution to Eq. (52), Eq. (2) completely specifies optimum processing.

For $N > 1$, however, λ' is not necessarily fixed, so that Eq. (3) alone does not completely specify optimum processing. Here Eq. (52) becomes

$$P_F = \sum_{m_1 = \lambda''(\mathbf{n})}^{\infty} p_{Di}(\mathbf{m}) \leq \alpha, \quad (53)$$

where λ'' is now dependent on the observation $\mathbf{n}$. Therefore, although λ'' is a unique solution to Eq. (53), it is not fixed for different values of $\mathbf{n}$, so that m_1 is not a sufficient statistic.

We now apply our single-threshold processor to a generalized photocounting optical communication system. For our model to be applicable, the essential requirement is that the signal and noise be additive, independent, and noninterfering, so that $p(n_i | H_1)$ may be represented by the convolution sum in Eq. (5). The single-threshold processor is then optimum if the logarithm of the noise density does not have a point of inflection. In Fig. 3 we present a block diagram of such a system. A sequence of 1's and 0's (representing binary information) is used to gate an optical source, which may be modulated in an arbitrary manner.[36] A 1 corresponds to the light being transmitted for T sec (H_1), whereas a 0 corresponds to the light being blocked for T sec (H_0). In general, each T-sec bit may be repeated N_t times in order to improve system performance.

In the simplest situation, there is no multiplicative or scattering channel, no avalanche multiplication, no dead time, and the discrete signal [with probability density $p_S(s_i)$] is embedded in additive, independent, noninterfering discrete background noise [with probability density $p_B(b_i)$]. Dark noise [with probability density $p_D(d_i)$] arises within the detector and results in additive counts, which are independent of both the

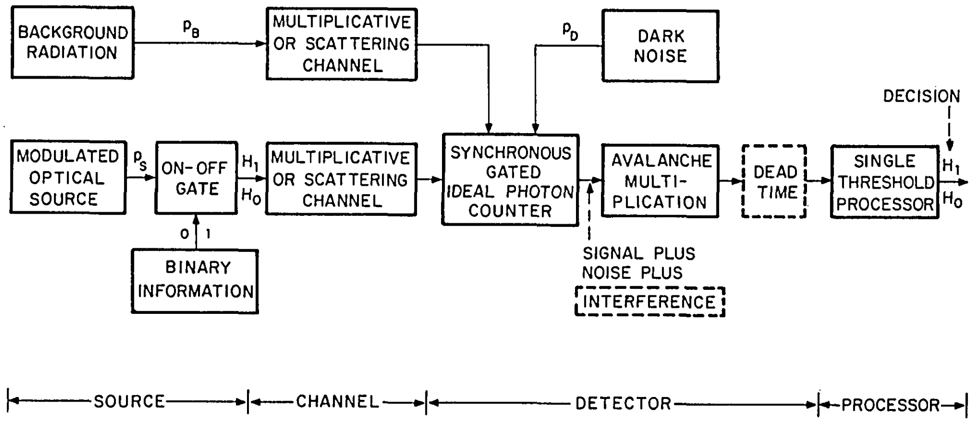

Fig. 3. A generalized photocounting optical communication system. The solid boxes represent components that can be included in our model, whereas the dashed boxes represent components that cannot be included in our model.

signal and the background noise. The dark noise and the background noise may therefore be convolved to provide the over-all noise density $p_H = p_D * p_B$. The signal plus noise are detected by a synchronously gated ideal photon counter with arbitrary quantum efficiency. In general, N_s ideal photon counters form a spatial array. The $N = N_t N_s$ observations are then processed to decide optimally whether a 1 or a 0 was transmitted. Since in this case the signal and noise are additive, independent, and noninterfering, our model applies immediately: if the logarithm of the noise density does not have a point of inflection, the single-threshold processor is optimum.

It is clear that this result remains unchanged if the optical source is modulated, since the source statistics alone are affected. This may come about, for example, as a result of laser output fluctuations, multimode operation, or modulation of the laser output. Modulation, however, broadens the source probability density $p_S(s_i)$, causing a degradation of system performance.[37]

Teich and Yen[7] evaluated the performance of just such a system, without modulation, for the $N = 1$ case, implicitly assuming that the single-threshold processor was optimum. In their model, the source was a multimode laser so that $p_S(s_i)$ was taken to be Bose-Einstein, the dark current density $p_D(d_i)$ was assumed to be Poisson, and $p_B(b_i)$ was taken to be either Poisson or Bose-Einstein. The validity of their single-threshold processor assumption for Poisson background noise (and arbitrary signal) is verified in Eq. (49).

Multiplicative[38,39] or scattering channels[4,5] that do not invalidate the additivity, independence, and noninterference assumptions are clearly also admissable to our model, as is optical communication through the clear-air turbulent atmosphere,[8–10] provided that the intensity fluctuations imparted to the signal and background radiation are independent.

Random multiplication in an avalanche detector is the result of each primary current pulse giving rise to a random distribution $p_G(g_i)$ of secondary pulses. The distribution of primary pulses v is $p(v,\langle v\rangle)$ with mean $\langle v\rangle$. To obtain the total counting distribution, $p_G(g_i)$ must be convolved with itself v times and averaged over $p(v,\langle v\rangle)$. The result is given by Eq. (3) of Personick *et al.*[17] Since the secondary pulses arising from independent primary pulses are themselves additive and independent, signal and noise remain additive and independent even after avalanche multiplication. Consequently, avalanche multiplication is admissable to our model.

Several effects cannot be included in our model, however, and are indicated by the dashed boxes in Fig. 3. These include the situation in which interference between the signal and noise results in cross-mixing terms that prevent $p(n_i|H_1)$ from being expressed as a simple convolution. In the special situation where the dark noise alone represents H_0, however, the interfering background radiation is lumped with the signal, so that the over-all signal and the dark noise are independent and additive, in which case the model does apply. Interference can also be ignored if it lies outside the bandwidth of the detector in which case it is averaged out in time, or if the background radiation enters the detector from a broad range of angles in which case it is averaged out in space.

Dead time,[40] being a nonlinear effect, destroys the independence of the signal and noise and cannot be included in our model. In the limit of large mean noise count, however, for Poisson signal and Poisson noise in the presence of dead time, it has been shown[19] by direct calculation that the single-threshold processor is optimum. Dead time effects are negligible when $\lambda\tau \ll 1$, where λ is the rate of the underlying Poisson process, and τ is the dead time.

This work was supported by the National Science Foundation with supplemental support from the Joint Services Electronics Program.

Appendix: Chain Rule for Finite Differences[41]

We consider a function $v = \Lambda(u)$, which is a mapping from an integer domain u into a real range v. The function u, in turn, is a mapping from a countable subset of the real numbers $\mathbf{m}$ into an integer range, $u =$

$n_i(\mathbf{m})$. The equation of the straight line u' through the points

$$n_i(m_1,\ldots,m_{j-1},a,m_{j+1},\ldots,m_N) \equiv n_i(\ldots a\ldots) \text{ and}$$

$$n_i(m_1,\ldots,m_{j-1},a+1,m_{j+1},\ldots,m_N) \equiv n_i(\ldots a+1\ldots) \text{ is}$$

$$u' = n_i(\ldots a\ldots) + (m_j' - a)\Delta_{m_j}[n_i(\ldots a\ldots)], \tag{A1}$$

where m_j' is some intermediate value. Similarly, the equation of the straight line v' through $\Lambda[n_i(\ldots a\ldots)]$ and $\Lambda[n_i(\ldots a+1\ldots)]$ is

$$v' = \Lambda[n_i(\ldots a\ldots)] + [n_i' - n_i(\ldots a\ldots)]\Delta_{n_i}\{\Lambda[n_i(\ldots a\ldots)]\}, \tag{A2}$$

where n_i' is some intermediate value. Substituting Eq. (A1) into Eq. (A2) we obtain

$$v' = \Lambda[n_i(\ldots a\ldots)] + (m_j' - a)\Delta_{m_j}[n_i(\ldots a\ldots)]\Delta_{n_i}\{\Lambda[n_i(\ldots a\ldots)]\}. \tag{A3}$$

Comparing Eq. (A3) with Eq. (A1) yields the chain rule for finite differences.

$$\Delta_{m_j}\{\Lambda[n_i(\mathbf{m})]\} = \Delta_{n_i}\{\Lambda[n_i(\mathbf{m})]\}\Delta_{m_j}[n_i(\mathbf{m})]. \tag{A4}$$

References

1. H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part 1* (Wiley, New York, 1968).
2. B. Reiffen and H. Sherman, Proc. IEEE **51,** 1316 (1963).
3. C. W. Helstrom, IEEE Trans. Inf. Theory **IT-10,** 275 (1964).
4. W. K. Pratt, *Laser Communication Systems* (Wiley, New York, 1969).
5. Proc. IEEE (Special Issue on Optical Communications), **58** (1970).
6. E. Hoversten, "Optical Communication Theory," in *Laser Handbook,* F. T. Arrechi and E. O. Schulz-DuBois, Eds. (North-Holland, Amsterdam, 1972), Vol. 2.
7. M. C. Teich and R. Y. Yen, IEEE Trans. Aerosp. Electron. Syst. **AES-8,** 13 (1972).
8. R. Y. Yen, P. Diament, and M. C. Teich, IEEE Trans. Inf. Theory **IT-18,** 302 (1972).
9. M. C. Teich and S. Rosenberg, Appl. Opt. **12,** 2616 (1973).
10. S. Rosenberg and M. C. Teich, Appl. Opt. **12,** 2625 (1973).
11. E. V. Hoversten, D. L. Snyder, R. O. Harger, and K. Kurimoto, IEEE Trans. Commun. **COM-22,** 17 (1974).
12. D. L. Snyder, *Random Point Processes* (Wiley, New York, 1975).
13. R. M. Gagliardi and S. Karp, *Optical Communications* (Wiley, New York, 1976).
14. C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
15. S. D. Personick, Bell Syst. Tech. J. **50,** 3075 (1971).
16. S. D. Personick, Proc. IEEE **65,** 1670 (1977).
17. S. D. Personick, P. Balaban, J. H. Bobsin, and P. R. Kumar, IEEE Trans. Commun **COM-25,** 541 (1977).
18. M. C. Teich, P. R. Prucnal, and G. Vannucci, Opt. Lett. **1,** 208 (1977).
19. M. C. Teich and B. I. Cantor, IEEE J. Quantum Electron. **QE-14** (December, 1978).
20. G. W. Flint, IEEE Trans. Mil. Electron. **MIL-8,** 22 (1964).
21. J. W. Goodman, Proc. IEEE **53,** 1688 (1965).
22. W. W. Peterson, T. G. Birdsall, and W. C. Fox, Trans. IRE Prof. Group Inf. Theory **PGIT-4,** 171 (1954).
23. W. P. Tanner and J. A. Swets, Psychol. Rev. **61,** 401 (1954).
24. H. B. Barlow, J. Opt. Soc. Am. **46,** 634 (1956).
25. M. C. Teich and P. R. Prucnal, J. Opt. Soc. Am. **67,** 1426 (1977).
26. W. J. McGill, J. Math. Psychol. **4,** 351 (1967).
27. M. C. Teich and W. J. McGill, Phys. Rev. Lett. **36,** 754 (1976).
28. P. R. Prucnal and M. C. Teich, IEEE Trans. Inf. Theory **IT-25,** (March, 1979).
29. The finite difference is $\Delta[f(n)]_k \equiv f(n+k) - f(n)$. If no subscript is used, k is assumed to be unity. We employ the second difference $\Delta^2[f(n)]_{k,j} \equiv [f(n+j) - f(n)] - [f(n+j-k) - f(n-k)]$ with $j = 1$ to obtain $\Delta^2[f(n)]_k$.
30. D. C. Murdoch, *Linear Algebra* (Wiley, New York, 1970), p. 165.
31. The product rule can be seen from

$$\begin{aligned}\Delta[f(n)g(n)h(n)] &= f(n+1)g(n+1)h(n+1) - f(n+1)g(n+1)h(n) \\ &\quad + f(n+1)g(n+1)h(n) - f(n+1)g(n)h(n) \\ &\quad + f(n+1)g(n)h(n) - f(n)g(n)h(n) \\ &= f(n+1)g(n+1)\Delta[h(n)] + f(n+1)\Delta[g(n)]h(n) \\ &\quad + \Delta[f(n)]g(n)h(n),\end{aligned}$$

which extends easily to the case of an N-fold product.
32. W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1957).
33. M. C. Teich and P. R. Prucnal, unpublished.
34. M. Woodroofe, *Probability with Applications* (McGraw-Hill, New York, 1975).
35. R. J. Glauber, "Photon Counting and Field Correlations," in *Physics of Quantum Electronics,* P. L. Kelley, B. Lax, and P. E. Tannenwald, Eds. (McGraw-Hill, New York, 1966), pp. 788–811.
36. P. Diament and M. C. Teich, J. Opt. Soc. Am. **60,** 682 (1970).
37. P. R. Prucnal and M. C. Teich, J. Opt. Soc. Am., to be published.
38. P. Diament and M. C. Teich, J. Opt. Soc. Am. **60,** 1489 (1970).
39. P. Diament and M. C. Teich, Appl. Opt. **10,** 1664 (1971).
40. B. I. Cantor and M. C. Teich, J. Opt. Soc. Am. **65,** 786 (1975).
41. P. R. Prucnal, Proc. IEEE, to be published.