THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 8, October 1975
Printed in U.S.A.

Optimum Direct Detection for Digital Fiber-Optic Communication Systems

By G. J. FOSCHINI, R. D. GITLIN, and J. SALZ

(Manuscript received January 23, 1975)

We report on optimum direct detection of digital data signals that are transmitted over optical fibers. Direct detection is provided by a photodetector whose output current is modeled as a noisy filtered Poisson stream of pulses. In this model, the time-varying pulse arrival rate is proportional to a linearly distorted version of the modulating signal. We show how the photodetector output is processed to derive the minimum probability-of-error receiver. Special attention is given to certain practical limiting cases.

When the average energy in the response of the photodetector to an individual photon is small compared to the additive thermal noise, the optimum detector is shown to be linear except for the use of precomputed bias terms. At the other extreme are the photomultiplier and the avalanche photodiode where the average energy in the response of the photodetector to a single photon is large compared with the additive noise. In this situation, we show that the optimum detector estimates the photon arrival times and then uses these estimates in a weighted counter. In both limiting cases, the detectors are specialized to one-shot M-ary and synchronous multilevel pulse-amplitude modulated (PAM) signals with intersymbol interference. For PAM signaling, we demonstrate that finite system memory allows application of dynamic programming to provide a detector implementation whose computational complexity does not increase with time.

I. INTRODUCTION

In recent years much attention has been focused on communication over optical channels.[1,2] Most early work was concerned with the physics of the electromagnetic transmission phenomena associated with various optical media and with the devices needed to change electrical signals to optical ones, and vice versa. In this paper, we are concerned with the optimum (maximum likelihood) reception of digital data transmitted over the fiber-optic channel. Our work was motivated by the many invaluable discussions we have had with S. D. Personick on this subject.

We shall not dwell on the quantum mechanical limitations imposed on the measurements of signals in the optical frequency range. Instead, we adopt a practical approach and assume at the outset that direct detection is used to convert optical energy to an electrical signal. This is accomplished by using a photodetector prior to any signal processing. Thus, we study a classical optical reception problem with the understanding that the photodetector output can be examined in every detail so as to extract all relevant information.

In a fiber-optic communication system, information is conveyed by modulating the intensity of a light source, such as a light-emitting diode. This is manifested in a photon stream whose arrival times form a Poisson process with a time-varying intensity function. The photodetector output current can then be modeled as a noisy filtered Poisson process whose intensity function is the sum of a dispersed version of the modulating wave and a background dark current. Thus, the central problem in communication systems employing a fiber-optic medium is the detection of the intensity function. Bar-David[3] and Gagliardi and Karp[4] have considered the optimal reception problem in the absence of dispersion (intersymbol interference) and additive thermal noise, while Personick[5–7] and Messerschmitt[8,9] have considered linear suboptimum receivers to combat these deleterious effects.

Section II reviews the communication theoretic model of the fiber-optic channel. Section III presents two simple examples that are intended to focus on certain system essentials and to illustrate some fundamental ideas involved in subsequent work. Section IV develops a general representation for the likelihood functional. Sections V and VI consider reception when the energy in the response of the photodetector to an individual photon is much smaller than the thermal noise, while Sections VII and VIII consider the complementary situation of large average energy per pulse-to-thermal noise.

II. A REVIEW OF THE MATHEMATICAL MODEL

In the past few years, a pragmatic communication theoretic model for data transmission over the fiber-optic channel has evolved. The papers by Personick[5–7,10] contain an up-to-date account of this model as well as provide more complete references on the physical aspects of fiber-optic communication. For the purpose of this investigation, it will suffice to think of the optical modulation process as providing a proportionate variation in the rate of photon arrivals at the photodetector. This device, of which there are several types, is a transducer that converts optical to electrical signals. The photodetector output current is illustrated in Fig. 1, and can be described as the sum of a

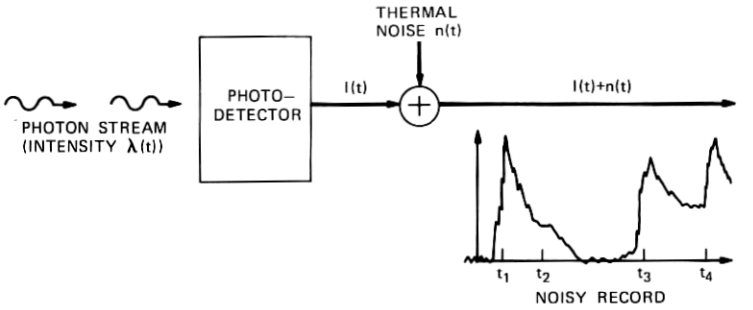

Fig. 1—Photodetection.

filtered Poisson process

$$I(t) = \sum_{k=1}^{\nu(t)} g_k w(t - t_k) \tag{1}$$

and white gaussian noise, $n(t)$, with spectral density N_0. The photon arrival times $t_1, t_2, \cdots$ are a family of independent, identically distributed, random variables, as are the positive gains $g_1, g_2, \cdots$. Moreover, these two families of random variables are independent of each other. The pulse $w(t)$ is square-integrable and is the convolution of two pulse shapes. The first pulse is the response of the photodetector circuitry to the generation of a single charge-carrier (i.e., an electron or a hole), while the second pulse is included for mathematical expediency so as to whiten the noise at the photodetector output.† We distinguish between two types of photodetectors, those that provide avalanche gain and those that do not. In the latter category is the photodiode that operates with $g_i = 1$, $i = 1, \cdots, \nu$ and results in a pulse energy-to-noise ratio $\int w^2(t)dt/N_0$, which is typically -20 dB. In other words, the response of the photodetector to an individual photon is masked by the additive background noise. This is in contrast to the photomultiplier and the avalanche photodiode where the gains possess a (discrete) probability distribution whose mean, $\bar{g}$, can be rather large and whose variance is a power ($\geqq 1$) of the mean.[11] For these devices, the average pulse energy-to-noise ratio $\overline{g^2}\int w^2(t)dt/N_0$ can be on the order of 20 dB.

The stochastic process $\nu(t)$, which is the number of pulses generated at the photodetector output in the interval $(0, t)$, is a Poisson process with intensity $\lambda(t)$, and therefore

$$\Pr[\nu(t) = N] = \exp\{-\Lambda(t)\}\frac{[\Lambda(t)]^N}{N!}, \tag{2}$$

† Note that the inclusion of a reversible operation, such as a whitening filter, does not affect the performance of an optimum detector.

where

$$\Lambda(t) = \int_0^t \lambda(t')dt'. \tag{3}$$

Moreover, each photon arrival time t_k possesses the probability density

$$p(t_k) = \frac{\lambda(t_k)}{\int \lambda(t')dt'}, \tag{4}$$

where the integral is over the observation time.†

In the digital fiber-optic communication system under discussion here, the positive intensity function $\lambda(t)$ is the information-bearing signal and is the average rate of electrons produced by the photodetector. The manner in which $\lambda(t)$ is manifest in the received optical signal (the photodetector input) is through the relation

$$\lambda(t) = k\mathcal{P}(t) + \lambda_0, \tag{5}$$

where $\mathcal{P}(t)$ is the received optical power, k is a constant conversion factor, and λ_0 is the average dark, or ambient, current in "counts" per second.‡ Thus, information is transmitted by modulating the optical power and must be recovered by processing the noisy photodetector output, $I(t) + n(t)$. As a result of transmitting the optical signal through the fiber-guide medium, the intensity function at the photodetector output will be the sum of a linearly distorted version of the transmitter intensity and the dark current. In the sequel, $\lambda(t)$ will be understood to mean the intensity function at the receiver.

Statistical averages of $I(t)$ are found by elementary calculations. For example,

$$E[I(t)] = E(g) \int_{-\infty}^{\infty} \lambda(\tau)w(t - \tau)d\tau \tag{6}$$

and

$$\sigma^2_{I(t)} = E(g^2) \int_{-\infty}^{\infty} \lambda(\tau)w^2(t - \tau)d\tau, \tag{7}$$

where $E(g)$ and $E(g^2)$ are the average and average square of the avalanche gain g. Higher moments can also be readily evaluated.

A linear channel model with additive "noise" is suggested by (6) and (7). In such a model, the desired signal is taken to be the average value of $I(t)$, namely $\lambda(t)$ passed through a filter with impulse response $E[g]w(t)$. One component of the added noise can be thought of as the signal dependent process $I(t) - E[I(t)]$, which has mean zero and

† Note that the arrival times are not assumed to be ordered.

‡ In free-space optical communication systems, $\lambda(t)$ must be regarded as having a noisy component.

variance given by (7). In addition to this noise, the gaussian noise must also be included before processing. While this linear model is a convenient approximation in some situations,[5–10] for purposes of this investigation we work with the process $I(t)$ directly.

Now that all the physical parameters have been defined, the optimum detection problem can be stated as follows:

> Given that the intensity function can assume one of M equiprobable positive functions $\lambda_m(t)$, $0 \leqq t \leqq \mathcal{T}$, $m = 1, \cdots, M$, the task of the detector is to decide which one of the M intensities has been transmitted after processing $I(t)$ plus gaussian noise for $\mathcal{T}$ seconds. Of particular interest is the synchronous pulse-amplitude modulated (PAM) signal
>
> $$\lambda(t) = \sum_k a_k f(t - kT) + \lambda_0,$$
>
> where each data bit, a_k, assumes the value 0 or 1, $1/T$ is the data rate in bits/s, and $f(t)$ is a positive time-dispersed pulse.
>
> The subject of our investigation is summarized by the question: How should the photodetector output, $I(t) + n(t)$, be processed so as to minimize the probability of error?

III. A MOTIVATING SIMPLIFIED DISCRETE MODEL—TWO EXAMPLES

To preview, in an elementary way, some ideas that are more fully developed in the sequel and also to serve as a motivation to the reader, we present a simplified version of the model discussed in the last section.

In a simplified theoretical model, the time index t is assumed to take on the discrete set of values $t_1, t_2, \cdots, t_J$, where $t_j = j\Delta$. Thus, instead of writing

$$I(t) = \sum_{k=1}^{\nu(t)} g_k w(t - t_k)$$

for the photodetector response to a photon stream, we write

$$I(t_j) = \sum_{k=1}^{J} g_k q_k w(t_j - t_k) \qquad j = 1, 2, \cdots, J. \tag{8}$$

In the above expression, $\{q_k\}_1^J$ can be regarded as an independent Bernoulli sequence with probabilities†

$$\Pr\{q_k = 1\} = \lambda_k \qquad \text{and} \qquad \Pr\{q_k = 0\} = 1 - \lambda_k,$$

† For convenience, we have taken $\Delta = 1$, and so we have written λ_k and $1 - \lambda_k$ instead of $\lambda_k\Delta$ and $1 - \lambda_k\Delta$.

where we have in mind that $0 < \lambda_k \ll 1$. Thus, $q_k = 1$ (or 0) represents the arrival (or nonarrival) of a photon at time t_k. We make the further simplifying assumption that $w(t_j - t_k) = A\delta_{jk}$ (A a positive constant), where δ_{jk} is the Kronecker delta and is nonzero only when $j = k$. This corresponds to assuming that the pulses $w(t)$ and $w(t - \Delta)$ do not overlap. Within this simplified framework, the received time-discrete signal is of the form

$$I(t_j) = g_j q_j A, \qquad j = 1, 2, \cdots, J. \tag{9}$$

We recall that $\{\lambda_j^{(m)}\}_{j=1}^{J}$ is the intensity function associated with the mth hypothesis. The particular intensity which is active is, of course, unknown at the receiver beyond the knowledge of the finite set from which it was chosen. The last ingredient of our model is to include the fact that the observation $I(t_j)$ is noisy and is given by

$$y(t_j) = g_j q_j A + n_j, \tag{10}$$

where the noise samples are assumed to be gaussian, independent, and zero-mean and have variance N_0. In relation to the more accurate model of the previous section, σ can be thought of as the standard deviation corresponding to $\int_{t_{j-1}}^{t_j} n(t)dt$. As is well known, the optimum detector computes the likelihood (the a posteriori probability density of the received signal conditioned on each hypothesis—in this case, the intensity) and selects the maximum. In statistical parlance, this is a standard multihypothesis testing problem. We now develop the form of the likelihood for two different assumptions on the nature of generation of secondary electrons:

(*i*) No avalanche gain ($g_j \equiv 1$).
(*ii*) Discrete avalanche gain (g_j takes on values 1, 2, $\cdots$, G, with probabilities $\rho_1, \rho_2, \cdots, \rho_G$).

In each case, we first obtain the likelihood for one observation. Owing to the nonoverlapping assumption on the pulses and the independent noise samples, the likelihood for J independent observations is given as a product. Our goal is to obtain a simple representation for the effective† likelihood $L^{(m)}(\lambda_1, \lambda_2, \cdots, \lambda_J; y_1, y_2, \cdots, y_J)$, where the superscript m denotes which intensity is assumed active. Given the received samples $y_1, \cdots, y_J$, the maximum likelihood (optimum) receiver selects the index m^* that maximizes $L^{(m)}$ and declares that intensity $\lambda^{(m^*)}$ is present. We shall find that, if N_0 is small, then the

† "Effective" refers to the fact that constants common to all hypotheses are dropped.

likelihood assumes an especially simple form. Specifically, in the high signal-to-noise ratio case, the likelihood is of the form

$$L^{(m)} \sim \prod_{j=1}^{J} (\lambda_j^{(m)})^{\bar{q}_j} (1 - \lambda_j^{(m)})^{1-\bar{q}_j}, \tag{11}$$

where $\bar{q}_j = 1$ if $y_j \geqq y_T$ (and zero otherwise). The quantity y_T is a threshold value that we shall derive for each example. Alternatively, the log-likelihood is expressible as the *weighted counter*

$$\sum_{j=1}^{J} \bar{q}_j \log \lambda_j^{(m)} + [(1 - \bar{q}_j) \log (1 - \lambda_j^{(m)})], \tag{12}^{\dagger}$$

where $\bar{q}_j$ is an estimated photon arrival process. In the complementary case of small signal-to-noise ratio ($N_0 \to \infty$), the detector is of the matched-filter or correlator type. The effective likelihood in this case is

$$L^{(m)} \sim c \sum_{j=1}^{J} \lambda_j^{(m)} y_j - b^{(m)}, \tag{13}$$

where c is a constant and $b^{(m)}$ is a hypothesis-sensitive bias term. We now turn to the specific examples.

(*i*) *The Photodiode (No Avalanche Gain)*

The single observation y_j is defined as

$$\begin{aligned} y_j &= n_j, && \text{with probability } 1 - \lambda \\ \text{and} & && \\ y_j &= A + n_j, && \text{with probability } \lambda. \end{aligned} \tag{14}$$

We temporarily drop the subscripts dealing with time (j) and hypothesis (m) while investigating this single observation. The likelihood is the mixture probability density

$$p(y) = (2\pi N_0)^{-\frac{1}{2}} \exp\left\{-\frac{y^2}{2N_0}\right\}\left[(1 - \lambda) + \lambda \exp\left\{\frac{Ay}{N_0} - \frac{A^2}{2N_0}\right\}\right]. \tag{15}$$

Noticing the hypothesis (λ) insensitivity of the first term, the effective likelihood becomes

$$L(y) = (1 - \lambda) + \lambda \exp\left\{\frac{Ay}{N_0} - \frac{A^2}{2N_0}\right\}. \tag{16}$$

A simple calculation shows that the two terms in (16) are equal when

---

† The reader familiar with Ref. 3 might expect an additional $-\Lambda$ term in (12). Owing to the simplified Bernoulli model employed above, this is not the case. However, the more refined analysis in the sequel will include this term.

$y = y_T$, where

$$y_T = \frac{A}{2} + \frac{N_0}{A} \log \left(\frac{1 - \lambda}{\lambda} \right). \tag{17}$$

For small N_0, $y_T \approx A/2$ and the graph of $L(y)$ converges to the solid line shown in Fig. 2. So, for $N_0/A^2 |\log \lambda|$ small, the effective likelihood can be approximated as

$$\hat{L}(y) = \begin{cases} 1 - \lambda, & y \leqq y_T \\ \lambda \exp \left\{ \dfrac{Ay}{N_0} - \dfrac{A^2}{2N_0} \right\}, & y > y_T. \end{cases} \tag{18}$$

The sense of the approximation is expressed by the following easily proven statement: For each $\delta > 0$, one can find an $N_0 > 0$ so that

$$\Pr \left\{ \left| \frac{L(y)}{\hat{L}(y)} - 1 \right| > \delta \right\} = 0. \tag{19}$$

To simplify the likelihood, note that $\exp \{(Ay/N_0) - (A^2/2N_0)\}$ and y_T are hypothesis-insensitive and can be deleted from the effective likelihood, and since we are assuming that λ is extremely small, $1 - \lambda$ can be treated as 1. The effective likelihood is then simply

$$\hat{L}^{(m)} = [\lambda_j^{(m)}]^{\bar{q}_j}, \tag{20}$$

where $\bar{q}_j = 1$ if $y_j > y_T$ and zero otherwise. Because of the independence of the noise samples and the nonoverlapping property of the

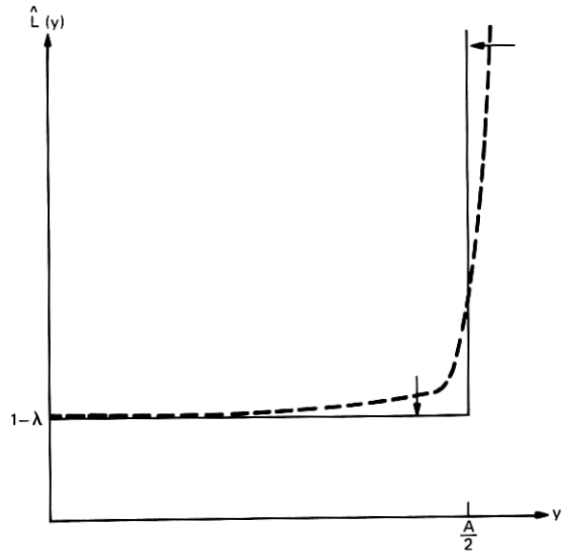

Fig. 2—Convergence of graph of $\hat{L}(y)$ to the asymptotic form ($N_0 \to 0$).

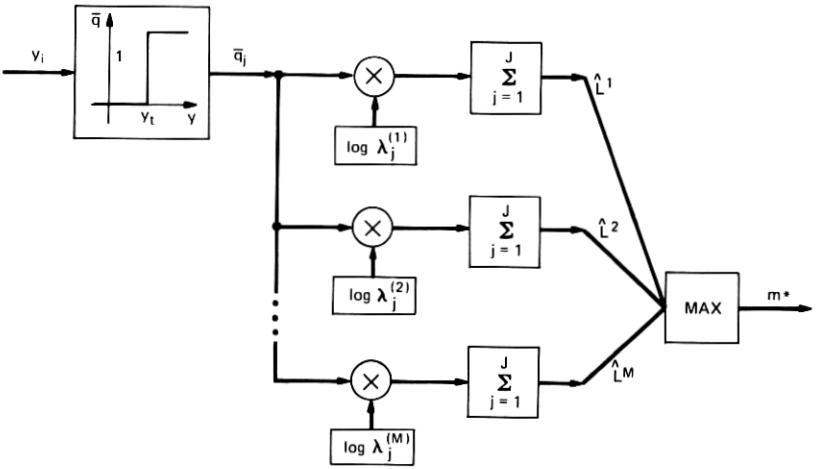

Fig. 3—Threshold-based weighted counter.

pulses, the likelihood for J observations is the product

$$\hat{L}^{(m)} = \prod_{j=1}^{J} [\lambda_j^{(m)}]^{\bar{q}_j}, \tag{21}$$

which yields the weighted counter

$$\log \hat{L}^{(m)} = \sum_{j=1}^{J} \bar{q}_j \log \lambda_j^{(m)} \tag{22}$$

shown in Fig. 3. The receiver selects the index that maximizes (22) and declares that the corresponding intensity was transmitted.

In the complementary case of low signal-to-noise ratio ($N_0 \to \infty$), we expand the likelihood function in a Taylor series and retain the dominating terms. This step must be done with care, since the numerator of the exponent has variance N_0, while N_0 also appears in the denominator. By normalizing the exponent, it is seen that the variance of the exponent is proportional to $1/N_0$; thus, the exponent will be small and a series expansion is useful. Keeping the first two terms in such an expansion of (16) gives

$$\hat{L}(y_j) = 1 - \lambda_j \left(\frac{Ay_j}{N_0} - \frac{A^2}{2N_0} \right), \tag{23}$$

and the likelihood for J observations becomes† the digital correlator

† We have used the fact that $\lambda_j/N_0[Ay_j - (A^2/2)] \ll 1$, and with $\epsilon_i \ll 1$ that $\Pi(1 + \epsilon_j) \sim 1 + \Sigma\epsilon_j$.

(matched-filter)

$$\hat{L}^{(m)} = \prod_{j=1}^{J} \left[1 - \lambda_j \left(\frac{Ay_j}{N_0} - \frac{A^2}{2N_0}\right)\right] \sim \sum_{j=1}^{J} \lambda_j^{(m)} \left(Ay_j - \frac{A^2}{2}\right), \quad (24)$$

which is shown in Fig. 4.

(ii) The Photomultiplier or Avalanche Photodiode (Discrete Avalanche)

Again, we start with the single observation case but now, because of the avalanche mechanism, a single primary gives rise to 1 or 2 or $\cdots$, G secondaries with probabilities $\rho_1, \rho_2, \cdots, \rho_G$, respectively ($\sum_1^G \rho_i = 1$). So the measurement y is modified as

$$y = \begin{cases} n, & \text{with probability } 1 - \lambda \\ A + n, & \text{with probability } \lambda\rho_1 \\ \vdots & \\ GA + n, & \text{with probability } \lambda\rho_G. \end{cases} \quad (25)$$

The likelihood is the mixture density

$$p(y) = \frac{(1-\lambda)}{\sqrt{2\pi N_0}} \exp\left\{-\frac{y^2}{2N_0}\right\} + \sum_{l=1}^{G} \frac{\lambda\rho_l}{\sqrt{2\pi N_0}} \exp\left\{-\frac{(y - lA)^2}{2N_0}\right\}. \quad (26)$$

Factoring out hypothesis-insensitive terms, the effective likelihood becomes

$$L(y) = (1 - \lambda) + \lambda \sum_{l=1}^{G} \rho_l \exp\left\{\frac{lAy}{N_0} - \frac{l^2A^2}{2N_0}\right\}. \quad (27)$$

As $N_0 \to 0$, we notice that $\hat{L}(y) \sim (1 - \lambda)$ for $y < A/2$. When $y > A/2$, let $\bar{l}A$ denote the number $A, 2A, \cdots$, or GA that is closest

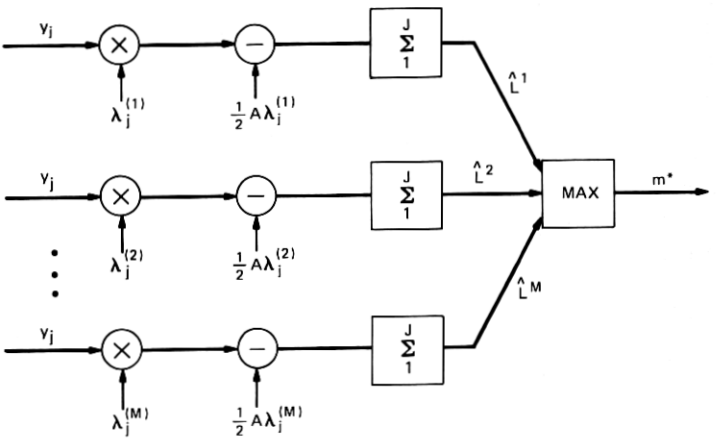

Fig. 4—Elementary version of digital correlator.

to y. Then the series appearing in (27) will be dominated by one term, and the likelihood becomes

$$\hat{L}(y) \sim \rho \bar{\imath}\, \lambda \exp\left\{ \frac{\bar{l} A y}{N_0} - \frac{\bar{l}^2 A^2}{2N_0} \right\}, \qquad \text{as } N_0 \to 0.$$

Proceeding as in the previous example, we consider both N_0 and λ small and drop hypothesis-insensitive terms from the approximate likelihood to obtain

$$\hat{L}(y) = [\lambda]^{\bar{q}}, \tag{28}$$

where $\bar{q} = 1$ when $y \geqq A/2$ and zero otherwise.† Moreover, note that the threshold is the same as in the nonavalanche case. This is because the detector is only interested in ascertaining whether or not a photon has arrived and need not estimate the magnitude of the avalanche gain. Again, for J measurements, the corresponding log-likelihood expression is simply the weighted counter

$$\log \hat{L} = \sum_{j=1}^{J} \bar{q}_j \log \lambda_j. \tag{29}$$

As $N_0 \to \infty$, we again expand the likelihood (27) in a Taylor series to obtain

$$\hat{L}(y) = 1 - \lambda \left\{ 1 - \sum_{l=1}^{G} \rho_l \left[1 + \frac{lAy}{N_0} - \frac{l^2 A^2}{2N_0} \right] \right\}, \tag{30}$$

which, for J measurements, becomes

$$\log \hat{L}^{(m)} = \sum_{j=1}^{J} \lambda_j^{(m)} \left\{ 1 - \sum_{l=1}^{G} \rho_l \left[1 + \frac{lAy_j}{N_0} - \frac{l^2 A^2}{2N_0} \right] \right\}. \tag{31}$$

The above is again interpreted as a correlator where $\lambda_j^{(m)}$ is correlated with $Ay_j/N_0 \cdot \sum_{l=1}^{G} l\rho_l = (Ay_j/N_0)E[g]$.

IV. THE MAXIMUM LIKELIHOOD DETECTOR

Here, we begin to answer the question posed at the end of Section II by presenting a derivation of the likelihood function associated with the received signal. The likelihood function is the probability measure of the photodetector output, given that a particular intensity is active. It is well known[12] that, when one of M equally likely signals $\lambda_m(t)$ is transmitted, the optimum (minimum probability error) detector computes the M values of the likelihood function evaluated at the received waveform and declares that the jth signal was sent, where the jth likelihood function is the largest.

† As expected when $N_0 \to 0$, the avalanche gain provides no essential benefit. A more interesting asymptotic evaluation and one that is more akin to reality is obtained by parameterizing the gain distribution such that $E[g]/N_0 \to \infty$.

We denote the received signal by

$$y(t) = I_m(t) + n(t), \qquad 0 \leqq t \leqq \mathcal{T}, \tag{32}$$

where $I_m(t)$ is the information-carrying, filtered, Poisson process

$$I_m(t) = \sum_{k=1}^{\nu(t)} g_k w(t - t_k), \tag{33}$$

and where the index m [corresponding to $\lambda_m(t)$] is hidden in the statistics of $\{t_k\}$ and $\nu(t)$. These statistics are described by (2) to (4) with $\lambda(t)$ replaced by $\lambda_m(t)$.

The task of the optimum receiver is thus to process the photodetector output $y(t)$ for $\mathcal{T}$ seconds and then decide which intensity function $\lambda_m(t)$, $m = 1, 2, \cdots, M$ is in effect. As we have mentioned earlier, the random variables $\{g_k\}$ represent the avalanche gains, and the pulse shape $w(t)$ is so far arbitrary with the only requirement being finite energy. Although in actual practice the noise at the output of the photodetector is not white, it can be whitened by a filter before additional processing and the effect of this filter will be manifest in the shape of $w(t)$.

The conditional likelihood function [when $I_m(t)$ is fixed] has the standard form[13]

$$L_m[y \mid I_m] = \exp\left\{\frac{1}{N_0}\int_0^{\mathcal{T}} I_m(t)y(t)dt - \frac{1}{2N_0}\int_0^{\mathcal{T}} I_m^2(t)dt\right\}. \tag{34}$$

The desired likelihood is the expectation of (34) with respect to $I_m(t)$ for fixed m, i.e.,

$$L_m(y) = E_I\{L_m[y \mid I_m]\}. \tag{35}$$

Once the intensity $\lambda_m(t)$ is specified, the above expectation is taken with respect to the number of arrivals, the arrival times, and the avalanche gain values. The detailed evaluation of this expectation and the interpretation of the resulting structures, in terms of implementable physical operations on $y(t)$, is our objective. The exact structure is sufficiently complex that many judicious approximations will have to be made to glean the essential nature of the operations.

We remark that a representation of (35) in terms of an estimator-correlator structure has recently been treated in the literature.[12,14–16] The optimum detector has been shown to be a correlation detector and the deterministic signal in the classical correlator is replaced by its least-squares estimate. This is a reformulation of the detection problem in terms of an estimation problem. Proponents of this method have taken the viewpoint that various suboptimum detectors are suggested by this formulation. A typical approach might be to replace the least-

squares estimate by the linear least-squares estimate or some other approximation, and to approximate the resulting stochastic integral by conventional integrals. While this might be reasonable, it does not indicate the direction of the approximation. We prefer an approach that, to be sure, has many approximations and makes use of estimates in place of the true quantities, but that can be explicitly related to the optimum detector under the asymptotic conditions of large and small signal-to-noise ratio.

Toward this end, we proceed by writing (35) in more detail. Neglecting edge effects on the integrals and assuming that the observation time $\mathcal{T}$ is much larger than the effective duration of a single pulse $w(t)$, we can express the inner product and the square term indicated in (34) as

$$\int_0^{\mathcal{T}} I_m(t)y(t)dt = \sum_{k=1}^{\nu} g_k P(t_k), \tag{36}$$

where

$$P(t_k) = \int_0^{\mathcal{T}} w(t - t_k)y(t)dt.$$

The square term is written as

$$\int_0^{\mathcal{T}} I_m^2(t)dt = \sum_{k,j=1}^{\nu} g_k g_j R(t_k - t_j), \tag{37}$$

where $R(t) = \int_0^{\mathcal{T}} w(\tau)w(t - \tau)d\tau$ is defined as the pulse correlation function.

Substituting (36) and (37) into (35), we obtain

$$L_m(y) = E_I \exp\left\{\frac{1}{N_0}\sum_{k=1}^{\nu} g_k P(t_k) - \frac{1}{2N_0}\sum_{k,j=1}^{\nu} g_k g_j R(t_k - t_j)\right\}. \tag{38}$$

Employing the vector notation $\mathbf{g}_\nu = (g_1, g_2, \cdots, g_\nu)$ and $\mathbf{t}_\nu = (t_1, t_2, \cdots, t_\nu)$ gives the expression

$$L_m(y) = E_{\mathbf{t}_\nu, \mathbf{g}_\nu, \nu}\left[\exp\left\{\frac{1}{N_0}\sum_{k=1}^{\nu} g_k P(t_k) - \frac{1}{2N_0}\sum_{k,j=1}^{\nu} g_k g_j R(t_k - t_j)\right\}\right], \tag{39}$$

and after performing the indicated expectations we obtain a detailed representation of the likelihood function

$$L_m(y) = \exp\left[-\Lambda_m(\mathcal{T})\right]\sum_{n=0}^{\infty}\frac{1}{n!}\sum_{\mathbf{g}_n}\int_0^{\mathcal{T}} d\mathbf{t}_n \prod_{k=1}^{n}\lambda_m(t_k)\prod_{k=1}^{n}\rho(g_k)$$
$$\times \exp\left\{\frac{1}{N_0}\sum_{k=1}^{n} g_k P(t_k) - \frac{1}{2N_0}\sum_{k,j=1}^{n}\sum_{k,j=1}^{n} g_k g_j R(t_k - t_j)\right\}, \tag{40}$$

where $\rho(g_i)$ is the (discrete) probability density function of the ava-

lanche gains and where it is understood that, when $n = 0$, the summand is taken to be unity.

To more easily interpret and/or mechanize the likelihood calculations, it will be convenient in some applications to assume that the photon arrivals can only occur at discrete instants of time $\{j\Delta\}$, where Δ is some fixed (small) interval and $j = 0, 1, 2, \cdots, J$. The integer J will be defined as the closest integer to T/Δ. This assumption is easily accommodated in (40) by replacing $\int dt_n$ with a multidimensional sum $\sum_{t_n}$ over the lattice $\{t_k = j\Delta : k = 1, 2, \cdots, n; j = 0, 1, \cdots, J\}$, and by replacing $\lambda(t_k = j\Delta)$ with the probability that† $j\Delta - \Delta/2 \leqq t_k \leqq j\Delta + \Delta/2$. The likelihood function under this set of assumptions then becomes

$$L_m(y) = \exp\left[-\lambda_m(T)\right] \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{t_n=0}^{J} \sum_{g_n} \prod_{k=1}^{n} \lambda(t_k) \prod_{i=1}^{n} \rho(g_i)$$
$$\times \exp\left\{\frac{1}{N_0}\left[\sum_{k=1}^{n} g_k P(t_k) - \tfrac{1}{2} \sum_{k,m=1}^{n} g_k g_m R(t_k - t_m)\right]\right\}, \quad (41)$$

which will be referred to as the (time) discrete likelihood function.

The two infinite functional series, (40) and (41), are not of much use as they stand. However, under a variety of physically realistic situations and by making suitable physical approximations as well as asymptotic expansions, we shall be able to deduce from these representations real-time implementable signal-processing algorithms.

By suitably normalizing the likelihood functions, (40) and (41), $1/N_0$ can be replaced by the (pulse) signal-to-noise ratio. This parameter α^2 will play a central role in our subsequent treatment, and its relative size will dictate our particular approach. The normalization entails replacing $R(t)$ by $R(t)/R(0)$, $P(t_k)$ by $P(t_k)/R(0)\bar{g}$, and the random variables g_k by $g_k/\bar{g}$, where $\bar{g} = Eg_k$; consequently,

$$\alpha^2 = \frac{\bar{g}^2 R(0)}{N_0}$$

and may be viewed as an average pulse signal-to-noise ratio. As we have discussed in the preceding section, in some applications this parameter is small, while in others it is large. Thus, our investigations in the sequel will focus on these two ranges. Additionally, different treatments of the likelihood ratio are also required, depending upon the presence or absence of avalanche gain.

It is instructive to give a still different representation for the likelihood, which will be found useful in the sequel. Towards this end, we introduce a zero-mean, stationary gaussian process $x(t)$ with correlation

† This probability is given by

$$\int_{j\Delta-\Delta/2}^{j\Delta+\Delta/2} \lambda(t)dt \approx \lambda(j\Delta)\cdot\Delta.$$

function,

$$E[x(t)x(t+\tau)] = R(\tau),$$

and can then write (39) in the form

$$L_m(y) = E_{t_\nu, g_\nu, \nu}\left[\exp\left\{\alpha^2 \sum_{k=1}^{\nu} g_k P\{t_k\}\right\} E_x \exp\left\{-i\alpha \sum_{k=1}^{\nu} g_k x(t_k)\right\}\right], \tag{42}$$

where we have used the elementary identity for gaussian processes

$$\exp\left\{-\alpha/2 \sum_{k,j=1}^{\nu} g_k g_j R(t_k - t_j)\right\} = E_x \exp\left\{i\alpha \sum_{k=1}^{\nu} g_k x(t_k)\right\}.$$

Since, over the observation interval, (42) is absolutely integrable, the expectation with respect to x and the other random variables may be interchanged. By noting that

$$E_{t_n, g_n, \nu} \exp\left\{\sum_{k=1}^{\nu} g_k x(t_k)\right\} = \exp(-\Lambda_m) \cdot \left\{\exp \sum_{g_n} \int_0^{\mathcal{T}} \rho(\mathbf{g})\lambda_m(t) \exp[i\alpha x(t)]dt\right\}, \tag{43}$$

we can write (42) in the form

$$L_m(y) = \exp(-\Lambda_m) E_x \left\{\exp\left(\sum_{\mathbf{g}} \int_0^{\mathcal{T}} \rho(\mathbf{g})\lambda_m(t) \times \exp[\alpha^2 g_j P(t) + i\alpha g_j x(t)]dt\right)\right\}. \tag{44}$$

In particular, in the absence of avalanche gain, i.e., $\rho(g) = \delta(g-1)$ (44) assumes the compact form

$$L_m(y) = \exp(-\Lambda_m) E_x \left\{\exp\left(\int_0^{\mathcal{T}} \lambda_m(t) \exp[\alpha^2 P(t) + i\alpha x(t)]dt\right)\right\}. \tag{45}$$

It may appear that the introduction of the process $x(t)$ did not simplify matters, since the explicit evaluation of the expectations again leads to an infinite functional series without adding insight into the nature of the processor. We shall nevertheless find this representation useful. As will be seen, when suitable approximations are made and asymptotic behaviors explored, a great deal of insight can be gained from the alternative representations for the likelihood† (40), (44), and (45), as well as the discrete likelihood (41).

† By normalizing the exponent, i.e., introducing α^2, we should actually use new symbols to denote $g_m/\bar{g}$ and $R/R(0)$. To avoid introducing extra notation, we retain the symbols g_m and $R(0)$, but we realize that, whenever α^2 is present, these variables have been normalized.

V. SMALL SIGNAL-TO-NOISE RATIO ($\alpha^2 \to 0$)

Here we consider the physical situation corresponding to small s/n ($\alpha^2 \to 0$). This occurs when a photodiode is used for direct detection. In this application, the response to an individual photon is masked by the background noise, and we do not expect the receiver to make explicit use of the information supplied by an individual pulse. Rather, the aggregate effect will be important. This is in contrast to the "counting" receivers (for large α^2), where individual counts contribute explicitly to the final decision. Since the avalanche gains are unity in this application, the likelihood function takes the form of (45). Two signaling situations of interest are examined next.

5.1 M-ary signaling

Since $\alpha^2 \ll 1$ (typically, $\alpha = -20$ dB), our approach will be to expand (45) in a power series in α^2 and retain the first two terms.† Consider the following Taylor series approximation to the argument of the exponent in (45). Again dropping the index m, let

$$\xi(\alpha, x) = \exp\left\{\int_0^{\mathcal{T}} \lambda(t) \exp\left[\alpha^2 P(t) + i\alpha x(t)\right] dt\right\}$$

$$\sim e^{\Lambda} + \xi'(0, x)\alpha + \xi''(0, x)\frac{\alpha^2}{2}. \quad (46)$$

Evaluating the derivatives, the asymptotic likelihood function becomes

$$L_m(y) \sim E_x\left[1 + \alpha\int_0^{\mathcal{T}} \lambda_m(t)x(t)dt + \frac{\alpha^2}{2}\left(\int_0^{\mathcal{T}} 2\lambda_m(t)P(t)dt - \int_0^{\mathcal{T}}\int_0^{\mathcal{T}} \lambda_m(t_1)\lambda_m(t_2)x(t_1)x(t_2)dt_1dt_2\right) - \frac{\alpha^2}{2}\int_0^{\mathcal{T}} \lambda_m(t)x^2(t)dt\right]. \quad (47)$$

Recalling that the exponent has been normalized such that $Ex = 0$, $Ex^2 = 1$, and $Ex(t_1)x(t_2) = R(t_1 - t_2)$, we get, after performing the averages,

$$L_m(y) \sim 1 + \alpha^2\left[\int_0^{\mathcal{T}} \lambda_m(t)P(t)dt - \frac{1}{2}\int_0^{\mathcal{T}}\int_0^{\mathcal{T}} \lambda_m(t_1)\lambda_m(t_2)R(t_1 - t_2)dt_1dt_2 - \frac{1}{2}\int_0^{\mathcal{T}} \lambda_m(t)dt\right] \quad (48)$$

or

$$\bar{L}_m(y) = \log L_m(y) \sim \int_0^{\mathcal{T}} \lambda_m(t)P(t)dt - \frac{1}{2}\int_0^{\mathcal{T}}\int_0^{\mathcal{T}} \lambda_m(t_1)\lambda_m(t_2)R(t_1 - t_2)dt_1dt_2 - \tfrac{1}{2}\Lambda_m. \quad (49)$$

The detector involves linear operations on the filtered received signal $P(t)$, addition of constants, and a maximization. As shown in Fig 5,

† Of course, the same answer would be obtained by working with (40).

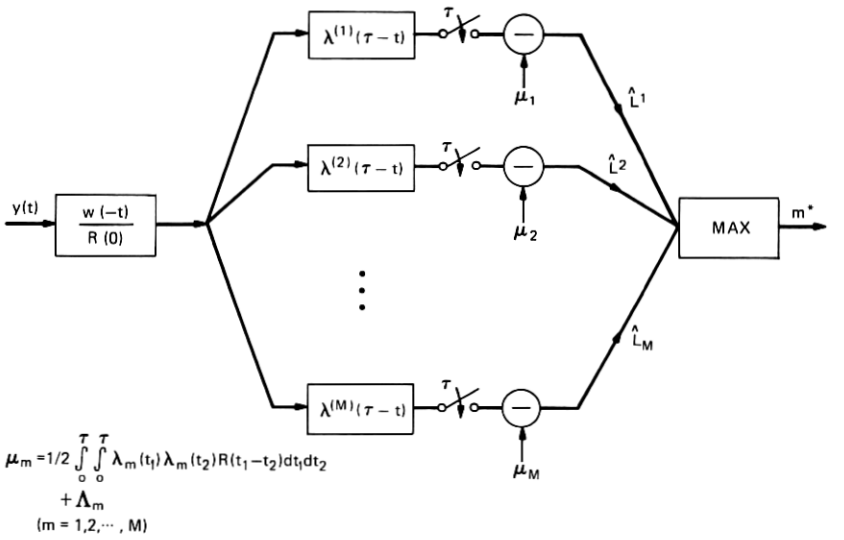

Fig. 5—Correlator filter for M-ary signaling.

a realization of the receiver is obtained by first passing the incoming signal, $y(t)$, through a filter with impulse response $w(-t)/R(0)$ to produce $P(t)$. This signal is then passed through a bank of M filters with impulse responses $\lambda_m(\mathcal{T} - t)$, $m = 1, 2, \cdots, M$ and sampled at $t = \mathcal{T}$. This is the first term in (49). The other two terms are precomputable biases. The detector then chooses the index m^*, which achieves the $\max_m \bar{L}_m(y)$, and the corresponding $\lambda_{m^*}(t)$ is declared to be the transmitted intensity.

There is a pleasing interpretation of this receiver which is reminiscent of the "linear" model discussed in Section II. If one were to consider the detection problem when the signal $I(t)$, given by (1), is replaced by its average $E[I(t)] = \bar{I}(t)$, given by (6), then the optimum detector in gaussian noise would base its decision on the likelihood function

$$\mathfrak{L} = \int_0^{\mathcal{T}} y(t)\bar{I}(t)dt - \frac{1}{2}\int_0^{\mathcal{T}} [\bar{I}(t)]^2 dt. \tag{50}$$

Substituting (6) into (50) gives

$$\begin{aligned}\mathfrak{L} &= \int_0^{\mathcal{T}} dt\, y(t) \int_0^{\mathcal{T}} w(\tau - t)\lambda(\tau)d\tau \\ &\quad - \frac{1}{2}\int_0^{\mathcal{T}} \left\{ \int_0^{\mathcal{T}} \int_0^{\mathcal{T}} w(t - t_1)\lambda(t_1)w(t - t_2)\lambda(t_2)dt_1dt_2 \right\} dt \\ &= \int_0^{\mathcal{T}} \lambda(\tau)P(\tau)d\tau - \frac{1}{2}\int_0^{\mathcal{T}} \int_0^{\mathcal{T}} \lambda(t_1)\lambda(t_2)R(t_1 - t_2)dt_1dt_2. \end{aligned} \tag{51}$$

Note that (51) differs from (49) only by the bias term Λ_m, the probability that no photons have arrived at the photodetector. We conclude, therefore, that the optimum detector structure in the case of small α^2 is thus "matched" to the average signal.

5.2 Optimum detection of PAM signals via the Viterbi algorithm

We will now develop the optimum receiver structure (still for small α^2) when the intensity is a pulse-amplitude modulated (PAM) signal†

$$\lambda(t) = \sum_{n=0}^{k} a_n f(t - nT), \qquad 0 \leqq t \leqq \mathcal{T}, \tag{52}$$

where each a_n can assume the binary values 0 or 1, $f(t)$ is a positive-valued pulse that incorporates the distortion of the optical medium, $1/T$ is the symbol rate, and $\mathcal{T} > kT$. Note that in writing (52) we have dropped the subscript m which we have used to identify the transmitted signal (intensity), since for PAM signaling it is generally more convenient to think of the receiver as finding that sequence $\{a_n\}$ which maximizes the likelihood. Substituting (52) in (49) and emphasizing that the likelihood function is now to be regarded as a function of a particular data sequence (which uniquely corresponds to a specific intensity) gives

$$\bar{L}(a_1, a_2, \cdots, a_k) = \sum_{n=1}^{k} a_n z_n - \frac{1}{2} \sum_{n,m=1}^{k} \sum a_n a_m \mathcal{K}_{n-m}, \tag{53}$$

where

$$z_n = \int_0^{\mathcal{T}} [P(t) - \tfrac{1}{2}] f(t - nT) dt \tag{54}$$

is the response at time nT of a filter matched to $f(t)$ when the input is $P(t) - \frac{1}{2}$, and the correlation-type function $\mathcal{K}$ is defined by

$$\begin{aligned} \mathcal{K}_{n-m} &= \int_0^{\mathcal{T}} d\tau \left(\int_0^{\mathcal{T}} dt f(t - nT) w(t - \tau) \right) \\ &\qquad \times \left(\int_0^{\mathcal{T}} dt' f(t' - mT) w(t' - \tau) \right) \\ &= \int_0 U(\tau - nT) U(\tau - mT) dt \\ &= \int_0^{\mathcal{T}} U(\tau) U[\tau - (n - m)T] dt, \end{aligned} \tag{55}$$

† Note that we have neglected the dark current λ_0. This obviously does not alter the final results. Also, the results can, in a straightforward manner, be extended to the multilevel case.

with

$$U(\tau - nT) = \int_0^{\mathcal{T}} f(t_1 - nT)w(\tau - t_1)dt_1,$$

and the observation time $\mathcal{T}$ is taken to be extremely large ($\mathcal{T} \gg T$).

The receiver structure indicated by (53) to (55) is similar to the maximum likelihood (ML) receiver for detecting a PAM signal distorted by a noisy linear channel.[17] The received signal is first passed through the matched filter $w(-t)$, and then (minus the bias term $\frac{1}{2}$) matched to $f(-t)$. The result is sampled at the synchronous instants nT. This produces the set of sufficient statistics $\{z_n\}$, from which the hypothesis-insensitive bias term $\frac{1}{2}\sum_{n,m=1}^{k}\sum a_n a_m \mathcal{K}_{n-m}$ is subtracted to produce the likelihood function.

The method by which the likelihood (53) is sequentially maximized in real time has become known as the Viterbi algorithm (VA), as a result of its application to the analogous problem of ML detection of linearly distorted PAM data signals.

The VA is a dynamic programming algorithm that uses the "finite memory" of $\mathcal{K}_n$, i.e., the fact that there will always be a $\bar{k}$ such that, for all practical purposes,

$$\mathcal{K}_n = 0, \qquad |n| > \bar{k}. \tag{56}$$

Because of (56), it is easy to see that the likelihood, (53), can be written in the recursive form

$$L(a_1, a_2, \cdots, a_k) = L(a_1, a_2, \cdots, a_{k-1}) + a_k z_k - \tfrac{1}{2}a_k \sum_{m=k-\bar{k}}^{k} \mathcal{K}_{k-m}. \tag{57}$$

By introducing the sequence of state vectors $\{\boldsymbol{\sigma}_n\}$, where

$$\boldsymbol{\sigma}_n = (a_{n-(\bar{k}-1)}, \cdots, a_n), \qquad n = 1, 2, \cdots, k, \tag{58}$$

the likelihood can be written in the form

$$L(\boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_k) = L(\boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_{k-1}) + h(z_k; \boldsymbol{\sigma}_k). \tag{59}$$

As is well known, the maximization of the function $L(\boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_k)$ with respect to its arguments is amenable to solution via dynamic programming since (59) is satisfied. Since this is the case, the optimum receiver now assumes the structure shown in Fig. 6.

In summary, it has been shown that the ML receiver for the limiting case of small s/n has a structure that is asymptotically approximated by the receiver designed to detect a known signal in gaussian noise (with the inclusion of certain precomputed bias terms). We remark at this point that the application of the Viterbi algorithm is, of course, only productive when intersymbol interference is the dominant im-

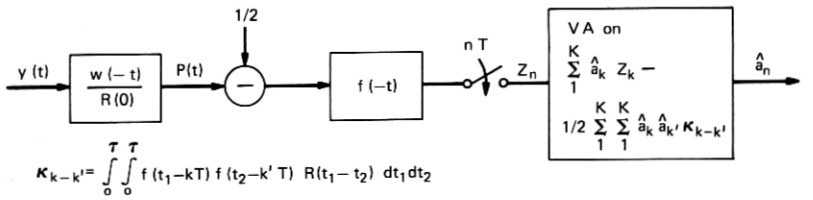

Fig. 6—Optimum detector (large noise) for PAM signaling.

pairment. In the context of the above discussion, this will be manifested in the values of $\mathcal{K}_n$ for $n \neq 0$. These values depend on the data rate relative to the channel dispersion. As in data transmission over voice-band channels, other methods of processing such as linear and decision feedback equalization should provide good results so long as the inter-symbol interference is not inordinately large. It is clear from (53) that when the distortion is small enough so that the quadratic term can be neglected, the optimization of the likelihood with respect to the data symbols can be carried out on a term-by-term or bit-by-bit basis. In other words, passing z_n through a slicer provides optimum detection. As the distortion becomes more severe, the quadratic term appearing in (53) must be retained. The linear receivers reported by Personick[5–7] and Messerschmitt[8,9] can be obtained from (53) by differentiating this expression with respect to the data symbols and then quantizing the result to the legitimate transmitted data levels. As the distortion increases still further, it becomes necessary to maximize (53), as it stands, via the Viterbi algorithm. Selecting one of these detectors in any given situation requires an evaluation of the error probability to quantify the effect of distortion on the system performance.

VI. PERFORMANCE ANALYSIS OF THE OPTIMUM DETECTOR FOR BINARY ONE-SHOT SIGNALING

6.1 An upper bound on the error rate (a simple example)

Having a description of the optimum detector structure for $\alpha^2 \to 0$, it is interesting to inquire how well it performs in certain signaling situations. Unfortunately, the M-ary mode of operation is extremely difficult to analyze, and even the general binary case poses insurmountable mathematical difficulties. We have, however, been able to analyze several special cases of interest that provide insight as to the effect of various system parameters on performance.

In the binary signaling case, information is conveyed by sending either intensity $\lambda_1(t)$ or $\lambda_2(t)$ with equal probability. From (51), the ML detector has the realization shown in Fig. 7. The detector, in this

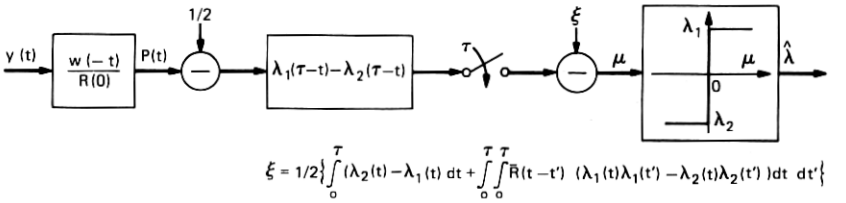

Fig. 7—Optimum detector for binary signaling ($\alpha^2 \to 0$).

situation, computes the statistic

$$\mu = \int_0^{\mathcal{T}} [\lambda_1(t) - \lambda_2(t)]P(t)dt - \frac{1}{2}\int_0^{\mathcal{T}} [\lambda_1(t) - \lambda_2(t)]dt - \frac{1}{2}\int_0^{\mathcal{T}}\int \bar{R}(t-\tau)[\lambda_1(t)\lambda_1(\tau) - \lambda_2(t)\lambda_2(\tau)]dtd\tau, \quad (60)$$

and μ is then compared to zero. When $\mu > 0$, it is decided that $\lambda_1(t)$ was sent, and when $\mu \leqq 0$, $\lambda_2(t)$ is chosen. In (60), the indicated quantities are normalized such that

$$P(t) = \frac{1}{R(0)}\int_{-\infty}^{\infty} y(\tau)w(t-\tau)d\tau$$

and $\bar{R} = R/R(0)$.

The probability of error is

$$P_e = \tfrac{1}{2}\Pr[\mu \geqq 0 | y(t) = I_1(t) + n(t), 0 \leqq t \leqq \mathcal{T}] + \tfrac{1}{2}\Pr[\mu < 0 | y(t) = I_2(t) + n(t), 0 \leqq t \leqq \mathcal{T}], \quad (61)$$

where

$$I_1(t) = \sum_1^{\nu} w(t - t_n) \qquad \text{with} \qquad E[\nu] = \int_0^t \lambda_1(\xi)d\xi,$$

and where

$$I_2(t) = \sum_1^{\nu} w(t - t_n) \qquad \text{with} \qquad E[\nu] = \int_0^t \lambda_2(\xi)d\xi.$$

It turns out that the evaluation of (61) is not mathematically tractable when λ_1 and λ_2 are arbitrary positive time functions. Even reasonable bounds on (61) are difficult to calculate in general. However, for constant intensities, exponentially tight upper bounds can be obtained. While the restriction to constant intensities might appear severe, it is shown in the appendix that in the absence of both dark current and gaussian noise the optimum choice of signals will have one intensity equal to zero while the other is arbitrary and need only satisfy a power constraint. Here we wish to illustrate a bounding approach for one special case where the upper bound can be obtained in closed form. We analyze the error rate for a system slightly modified from that depicted

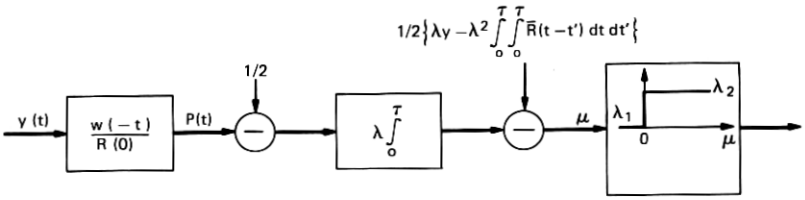

Fig. 8—Optimal detector for α^2 small and $\lambda_1 = 0$ and $\lambda_2 = \lambda$.

in Fig. 8 for $\lambda_1 = 0$ and $\lambda_2 = \lambda$. The modification will involve adjusting the threshold† so that our upper estimate of the probability of error when λ_1 is sent is equal to the estimate of probability of an error in the complementary situation.

In the binary system under consideration, the information symbol 1 is encoded into the intensity function $\lambda_1(t) = \lambda$, $0 \leqq t \leqq \mathcal{T}$ and the information symbol 0 into the intensity $\lambda_2(t) = 0$, $0 \leqq t \leqq \mathcal{T}$. Notice that the dark current is assumed to be zero. The detector structure we wish to analyze is depicted in Fig. 9. Here, the information-bearing Poisson process is passed through a matched filter $w(-t)/R(0)$, then integrated, and the result compared with a threshold set at F. If μ (refer to the block diagram) exceeds F, the symbol 1 is chosen and if $\mu \leqq F$, the symbol 0 is chosen. Our chief interest in this example is to exhibit the interplay between the various parameters in this extremely simple but informative situation.

As seen in the diagram,

$$\mu = \nu \int_0^{\mathcal{T}} R(t)dt + \int_0^{\mathcal{T}} \int_0^{\mathcal{T}} n(\tau)w(t - \tau)dtd\tau \tag{62}$$

or, equivalently, the test statistic may be written as

$$\mu_0 = \nu + n_0, \tag{63}$$

which is compared to a threshold. Note that μ_0 is just a scaled version of μ, and n_0 is a zero-mean gaussian random variable with

$$E\{n_0^2\} = N_0 \frac{\int_0^{\mathcal{T}} \int_0^{\mathcal{T}} R(t - \tau)dtd\tau}{\left[\int_0^{\mathcal{T}} R(t)dt\right]^2} \triangleq \sigma^2. \tag{64}$$

Observe that, in this situation, the receiver is just a counter since the test statistic represents the total number of photon counts observed in the entire observation interval plus an added gaussian random variable.

† By the threshold, we mean the bias terms appearing in (60), i.e., the last two terms.

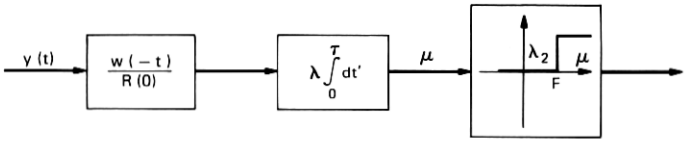

Fig. 9—Detector for α^2 small and $\lambda_1 = 0$, $\lambda_2 = \lambda$ with threshold modified so error probabilities are equal.

The integer random variable ν is Poisson-distributed with

$$E[\nu] = \lambda T \qquad \text{when} \qquad 1 \text{ is sent } (H_1)$$

and

$$E[\nu] = 0 \qquad \text{when} \qquad 0 \text{ is sent } (H_0), \tag{65}$$

where H_0 and H_1 are symbols distinguishing the two situations. The probability of error is then explicitly given by

$$P_e = \tfrac{1}{2} \Pr [\mu > F | H_0] + \tfrac{1}{2} \Pr [\mu \leqq F | H_1], \tag{66}$$

where we have made the assumption that 0s and 1s are transmitted with equal probability.

Since (66) cannot be expressed in closed form, we seek an exponentially tight upper bound. Applying the Chernoff bounding technique, we notice that the error rate under the null hypothesis, H_0, can be upper bounded immediately since under this hypothesis $\nu = 0$. Applying the bound yields

$$P_0 = \Pr [\mu > F | H_0] \leqq \exp \left\{ - \frac{F^2}{2\sigma^2} \right\}. \tag{67}$$

The second term in (66) can likewise be upper bounded since the moment generating function of ν under H_1 is known. This procedure gives

$$P_1 = \Pr [\mu \leqq F | H_1] \leqq \exp \left\{ \theta F + \frac{\theta^2 \sigma^2}{2} \right\} M_{\nu | H_1}(-\theta), \qquad \theta \geqq 0, \tag{68}$$

where

$$M_{\nu | H_1}(\theta) = E\{e^{\nu\theta} | H_1\} = \exp [\lambda T (e^{\theta} - 1)]$$

for $\theta \geqq 0$. The bound (68) then becomes

$$P_1 \leqq \exp \left\{ \theta F + \frac{\sigma^2 \theta^2}{2} + \lambda T (e^{-\theta} - 1) \right\}, \qquad \theta > 0, \tag{69}$$

and it is optimized by finding a θ^* such that

$$E(\theta^*, F) = \min_{\theta > 0} \left\{ \theta F + \frac{\sigma^2 \theta^2}{2} + \lambda T (e^{-\theta} - 1) \right\}. \tag{70}$$

To make the upper bounds on P_1 and P_2 equal, we select an $F = F_0$ such that

$$E(\theta^*, F_0) = \frac{F_0^2}{2\sigma^2}.$$

This, then, yields the final upper bound on the error rate

$$P_e \leqq \exp\,(-F_0^2/2\sigma^2). \tag{71}$$

By differentiating (70), we see that for a positive solution to exist it is required that $0 < F < \lambda\mathcal{T}$. Unfortunately, such a solution cannot be obtained in closed form. However, lower bounding $1 - e^{-\theta}$ by $\theta - \theta^2/2$, which in turn upper bounds (69), we find that

$$\theta^* = \frac{\lambda\mathcal{T} - F}{\sigma^2 + \lambda\mathcal{T}} > 0, \tag{72}$$

and consequently

$$P_1 \leqq \exp\left\{-\theta^*(\lambda\mathcal{T} - F) + \frac{\theta^{*2}}{2}(\lambda\mathcal{T} + \sigma^2)\right\}, \tag{73}$$

where θ^* has been chosen to provide the tightest bound.

Having θ^*, the threshold F_0 is obtained from

$$\frac{F_0^2}{\sigma^2} = \frac{(\lambda\mathcal{T} - F_0)^2}{\lambda\mathcal{T} + \sigma^2}.$$

Solving this quadratic equation and selecting the only reasonable root for F_0 give

$$F_0 = -\sigma^2 + \sqrt{\sigma^4 + \sigma^2\lambda\mathcal{T}}. \tag{74}$$

Substituting (74) into (72), the bound on the error rate finally becomes

$$P_e < \exp\left[-\frac{K}{2}\{\sqrt{1 + C} - \sqrt{C}\}^2\right], \tag{75}$$

where $K = \lambda\mathcal{T}$ and $C = \sigma^2/K = \dfrac{\text{Average Noise Power}}{\text{Average Shot Noise Power}}$.

It is instructive to express the bound (75) in the following alternative form

$$P_e \leqq e^{-\rho f(c)}, \tag{76}$$

where

$$\rho = \frac{1}{2}\frac{K^2}{K + \sigma^2} = \frac{\text{Average (signal)}^2}{\text{Average Total Noise Power}}$$

and where $f(c) = [1 + c - \sqrt{c^2 + c}]^2$.

As can be checked, $f(c)$ is a monotonically decreasing function of c,

and has the properties

$$\lim_{c \to 0} f(c) = 1$$

$$\lim_{c \to \infty} f(c) = \tfrac{1}{4}.$$

Thus, $P_e \leqq e^{-\rho f(c)} \to e^{-\rho}$, as $c \to 0$.

This is the situation that prevails when the shot noise dominates. On the other hand,

$$P_e \leqq e^{-\rho/4}, \qquad \text{as} \qquad c \to \infty,$$

which is the situation when the gaussian noise dominates.

6.2 Implications of the error bound

The first observation concerning (75) is that, as $C \to 0$, $P_e \leqq \exp\{-K/2\}$. This can be achieved by making $\sigma^2 \to 0$. This implies that either the gaussian noise is zero or that the number of counts is very large. However, in the absence of gaussian noise (as well as dark current), it is clear that the only way to make an error is when there are not any counts ($\nu = 0$) under H_1. The chance that $\nu = 0$ under H_1 is just $\exp\{-K\}$. In the absence of gaussian noise, this is clearly the very best performance one can hope for. Notice that the upper bound predicts an outcome which is 3-dB poorer than this ideal. The factor of 2 in the exponent of (75) is attributed to our bounding technique. What, in fact, happens as $\sigma^2 \to 0$ is that θ^* increases, and that the lower bound $\theta - \theta^2/2$ becomes loose, the upshot being the factor of 2 in the exponent. To see that this factor of 2 is indeed a quirk of the parabolic approximation to the exponential, consider the exponent in (69) as $\sigma^2 \to 0$. It is clear that the optimum threshold and θ are, respectively, zero and infinity, which when substituted in (69) does indeed give $e^{-\lambda \mathcal{T}}$ ($K = \lambda \mathcal{T}$).

Another aspect of the bound, however, is that ideal performance can be achieved with this detector structure (which is optimum for $\sigma^2 \to \infty$, the large gaussian noise situation) when the noise vanishes ($\sigma^2 \to 0$). This suggests that for the case of constant intensities the linear threshold detector is robust, i.e., it performs well over the entire range of σ^2 (or α^2).

We now use the error bound to determine the number of counts required, for reasonable operating physical parameters, to achieve a desired error rate. Note that, from (64), after a simplifying calculation on the double integral, we obtain

$$\sigma^2 = \frac{N_0}{A^2} 2 \left(\mathcal{T} A - \int_0^{\mathcal{T}} tR(t)dt \right) = \frac{2N_0}{A} \left[\mathcal{T} - \frac{\int_0^{\mathcal{T}} tR(t)dt}{\int_0^{\mathcal{T}} R(t)dt} \right], \quad (77)$$

where

$$A = \int_0^{\mathcal{T}} R(t)dt.$$

Introducing the pulse stretch factor,

$$S = \int_0^{\mathcal{T}} tR(t)dt \Big/ \int_0^{\mathcal{T}} R(t)dt < \mathcal{T}, \tag{78}$$

into (78) and recalling that $\alpha^2 = R(0)/N_0$ yields explicitly

$$\sigma^2 = \frac{2}{\alpha^2} \frac{1-r}{r} \cdot \frac{R(0)S}{A}, \tag{79}$$

where $0 < r = S/T < 1$. What, then, can be said about the choice of the parameter r? Can it be selected at will? Within a good approximation, $SR(0)/A \sim 1$. Clearly, the best choice of r appears to be unity, since $r = 1$ reduces the noise variance to zero. Recall, however, that, when the mathematical model was initially introduced, it was tacitly assumed that the observation interval was much larger than the width of the pulses emanating from the photodetector so that edge effects could be neglected. This alone would restrict the range of r to be no more than, say, 0.1, which would indicate that r does not appear to be an independent parameter. With $r = 0.1$, we may conclude from (79) that the effective gaussian variance of the scaled system is roughly

$$\sigma^2 = 20/\alpha^2. \tag{80}$$

Returning now to (75), we see that ideal performance is achieved when

$$C = \frac{\sigma^2}{K} \ll 1,$$

and when (80) is substituted in the above, we arrive at the condition that

$$\frac{20}{K\alpha^2} \ll 1 \rightarrow K\alpha^2 \gg 20. \tag{81}$$

As an example, let $\alpha^2 = 1/400$, which, according to S. Personick,† is a reasonable number for this parameter. This implies that $K \gg 8000$ is required to achieve ideal performance (i.e., the error rate in this range approaches zero like e^{-K}). On the other hand, suppose it is desired that $P_e \leqq 10^{-9}$. This would imply that

$$\frac{K}{2} \{\sqrt{(20/K\alpha^2) + 1} - \sqrt{20/K\alpha^2}\}^2 \sim 20.$$

† Private communication.

For instance, $\alpha^2 \sim 1/400$ implies that K is on the order of 1200. The above discussion quantifies the facts that to achieve good performance the total number of counts must be large or, if the background gaussian noise is small then fewer counts are needed to provide satisfactory performance.

6.3 Some conclusions concerning optimum detection for constant intensities

Note that the linear receiver, which is optimum when $\alpha^2 \to 0$, seems to be robust—at least for binary systems signaling with constant intensities. The optimum detector in the small s/n case ($\alpha^2 \to 0$) yields a decision variable based on the total number of observed counts as evidenced from (63). Of course, for the error probability bound to be tight, the average number of counts, K, must be large enough that $\sigma^2/K \ll 1$. On the other hand, we saw that the optimum detector structure in the case of large s/n ($\alpha^2 \to \infty$) combined with narrow pulses† ($r \ll 1$) is also a counter. The only difference is that the counts in the $\alpha^2 \to 0$ detector are linearly corrupted by gaussian noise, while the counts in the $\alpha^2 \to \infty$ detector are determined by quantizing the incoming signal to the nearest integer in the presence of the added gaussian noise. The latter operation is, of course, nonlinear. Nevertheless, when the added noise is small ($\alpha^2 \to \infty$), the two operations are approximately equivalent, thus explaining the robustness of the linear receiver and the results of our theory.

VII. LARGE SIGNAL-TO-NOISE RATIO ($\alpha^2 \to \infty$) AND NARROW PULSES

When a photomultiplier or avalanche photodiode is used to provide direct detection, the parameter α^2 is much larger than unity. In this application, the response of the photodetector to a single electron or hole is much larger than the background gaussian noise. In this situation, intuition dictates that the detector make use of the "estimated" arrival times of the individual photons. Here we discuss a special case that will bring out the essential structure of the optimum detector. The situation examined is when there is no avalanche gain and the individual pulses $w(t)$ are time-limited to an interval much smaller than the observation interval. The more general situation is treated in Section VIII.

The approach taken in this section is to use the gaussian process formulation (45) and attempt to approximate the indicated expectation with respect to the $x(t)$ process. For this approach to be productive, we must assume that $R(t)$ has effective duration Δ. We may then

† This was demonstrated in the examples of Section III and is reestablished in Section VII.

approximate the integral appearing in (45) by a discrete sum, i.e.,

$$\int_0^T dt \exp [\alpha P(t) + i\alpha x(t)]\lambda_m(t) \to \Delta \sum_{j=1}^{J} \exp (\alpha^2 P_j + i\alpha x_j)\lambda_m(j\Delta), \quad (82)$$

where $P_j = P(j\Delta)$ and $x_j = x(j\Delta)$.

The implication of (82) can be viewed in several ways. Of course, as $\Delta \to 0$ and $J \to \infty$, irrespective of the correlation function $R(t)$, the discrete sum is an excellent approximation to the integral. But sampling the integrand at the rate $1/\Delta$ does not necessarily guarantee that the sum is a good approximation to the integral. Yet to derive any utility from representation (45), we must sample at a rate $1/\Delta$ so that the sequence of random variables $\{x_j\}$ can be regarded as identically and independently distributed. Unfortunately, this is the only case for which we can compute the indicated averages in a useful form. What then do we mean by (82)? To make sense of this representation, we must reinterpret the distribution of the arrival times, $\{t_n\}$. Evidently, the reason we have an integral representation instead of a sum is because we have assumed that the arrival times obey a continuous distribution. However, if we assume at the outset that the arrival times $\{t_n\}$ can occur only at a set of discrete points $\{t_n = n\Delta\}$, then (45) will contain a sum instead of an integral. This procedure is equivalent to that used to obtain (41) as the discrete version of (40). Hence, a rigorous interpretation of (45) is that the Poisson arrival times can only occur at the discrete instants of time $\{j\Delta\}$, $j = 0, 1, 2, \cdots$. If we now assume that the quantization of the arrival times to units of Δ is such that $R(\Delta) \sim 0$, then the set of random variables $\{x_j\}_{j=1}^{J}$ are mutually independent. Exploring this line of reasoning, (45) can be written as

$$e^{\Delta}L(y) = \prod_{j=1}^{J} E_x[\exp \{\Delta\lambda_j \exp (\alpha^2 P_j + i\alpha x_j)\}], \quad (83)$$

where $\lambda_j = \lambda(j\Delta)$, and we have suppressed the index m denoting the particular hypothesis being tested.

Expanding (83) in a power series and carrying out the indicated expectation give

$$e^{\Delta}L(y) = \prod_{j=1}^{J} \left(\sum_{n=0}^{\infty} \frac{\Delta^n(\lambda_j)^n}{n!} \exp \left\{\alpha^2 \left(nP_j - \frac{n^2}{2}\right)\right\}\right). \quad (84)$$

We are now in a position to exploit the assumed large value of α^2. In other words, we are interested in determining the behavior of (84) as $\alpha^2 \to \infty$. Towards this goal, consider the sum

$$S_j = \sum_{n=0}^{\infty} \frac{\Delta^n(\lambda_j)^n}{n!} \exp \{\alpha^2[nP_j - n^2/2]\}. \quad (85)$$

This series is in the form

$$\sum_{0}^{\infty} \beta_n \exp(\alpha^2 \gamma_n), \tag{86}$$

where β_n and γ_n have the obvious identifications.

Let $\bar{\gamma}_j$ be the largest of the γ_n and β_j be the corresponding value of β_n. Then (86) becomes

$$\bar{\beta}_j \exp(\alpha^2 \bar{\gamma}_j) \left(\sum_{n=0}^{\infty} \frac{\beta_n}{\bar{\beta}_j} \exp[\alpha^2(\gamma_n - \bar{\gamma}_j)] \right), \tag{87}$$

where each $\gamma_n - \bar{\gamma}_j$ is negative. Since (86) converges absolutely, the infinite sum can be rearranged in such a way that the exponents are decreasing; thus, the rearranged sum is recognized to be a Dirichlet series[18] in the parameter α^2. From the elementary properties of such series, we deduce that, except for the $n = j$ term, the summation portion of (87) converges to zero as $\alpha^2 \to \infty$. So, as $\alpha^2 \to \infty$, (87) behaves like

$$S_j \sim \bar{\beta}_j \exp(\alpha^2 \bar{\gamma}_j). \tag{88}$$

Now returning to the series in (85), we let n_j denote the strictly nonnegative integer attaining

$$\max_{n \in \{0,1,2,\cdots\}} \left\{ \alpha^2 n P_j - \frac{n^2}{2} \alpha^2 \right\}, \tag{89}$$

i.e., $n_j = [P]$, where $[P]$ denotes the nonnegative integer nearest to P. The corresponding coefficient becomes

$$\bar{\beta}_j = \frac{\Delta^{n_j}(\lambda_j)^{n_j}}{n_j!}. \tag{90}$$

Thus, as $\alpha^2 \to \infty$,

$$\bar{L}(y) = \log L(y) = -\Lambda + \log \times \left\{ \sum_{j=1}^{J} \left(\frac{\Delta^{n_j}(\lambda_j)^{n_j}}{n_j!} \exp[\alpha^2(n_j P_j - n_j^2/2)] \right) \right\}. \tag{91}$$

Discarding the hypothesis-insensitive terms, (91) can be rewritten in the form

$$\bar{L}_m(y) \sim -\Lambda_m + \sum_{j=1}^{J} n_j \log \lambda_j^{(m)}, \tag{92}$$

where we recall that $n_j = 0$ whenever $P_j < \frac{1}{2}$. Note that (92) is similar to the detector described by (56); however, the different statistical model (Bernoulli as opposed to Poisson occurrences) accounts for the bias term $-\Lambda_m$ appearing in (92).

The detector structure exhibited in (92) has a simple interpretation and is similar to that depicted in Fig. 3. As shown in Fig. 10, the in-

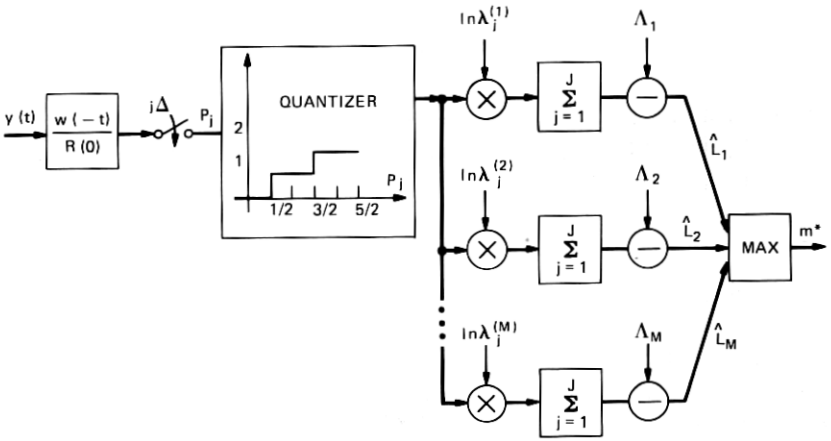

Fig. 10—Weighted counter with quantizer.

coming signal $y(t)$, having been filtered by $w(-t)/R(0)$, is sampled every Δ seconds. This is followed by quantizing the samples, P_j, to the nearest positive integer (including zero whenever $P_j < \frac{1}{2}$). The quantized samples are multiplied by the locally stored numbers $\log \lambda_j^{(m)}$ and the results summed. The sum is added to Λ_m to form the decision statistic. Since the added gaussian noise is assumed to be small and the pulse $w(t)$ is assumed to be narrow, most of the time the nearest integer at any time $t_j = j\Delta$ will be either 1 or 0, depending on whether $P_j > \frac{1}{2}$ or $P_j \leqq \frac{1}{2}$, i.e., whether the receiver determines a pulse is present or absent. Consequently, the optimum detector structure may be viewed as a weighted counter, where the decision as to which intensity was transmitted is based on selecting the largest of the weighted pulse counts.

We recognize that from an implementation point of view even this seemingly simple structure may pose practical difficulties. The indicated sampling may be difficult to carry out at this high frequency. While this is indicated mathematically, in practice the peaks of the signal at the photodetector output could be used to approximate the photon arrival times and, hence, the interrogation times.

VIII. MAXIMUM LIKELIHOOD RECEIVER FOR LARGE SIGNAL-TO-NOISE RATIO ($\alpha^2 \to \infty$)

This section extends the results of the last section by indicating a general approach to the extremely complex problem of performing optimum detection when the pulses $w(t)$ are not restricted in width or shape and when avalanche gain is provided. In the presence of avalanche gain, the average signal-to-noise ratio, α^2, is large. This implies

that the photon arrival times can be accurately estimated, and these estimates can then be used to aid the detector in making accurate decisions. One objective of this section is to indicate how the optimum detector estimates the arrival times. Heuristically, the receiver attempts to "whiten" or peak up the pulse $w(t)$. The presence of gaussian noise, however small, prevents pulse whitening via linear filtering. The nonlinear manner in which the receiver estimates the arrival times is of independent interest and will be presented in the sequel.

We begin with the most general form of the likelihood function (40). While the infinite functional series appearing in (40) is quite intimidating, it has already been shown to reduce to physically interpretable receivers in the following special cases: (*i*) small signal-to-noise ratio ($\alpha^2 \to 0$) and (*ii*) large signal-to-noise ratio ($\alpha^2 \to \infty$) combined with an extremely small decorrelation time† for $R(t)$.

Since large α^2 is a practical operating condition (photomultiplier and the avalanche photodiode), we are motivated to examine the salient features of the optimal processor under these circumstances. We also specialize our development to the PAM-Poisson intensity, or data signal,

$$\lambda_m(t) = \lambda_0 + \sum_{n=0}^{N} a_n^{(m)} f(t - nT), \qquad 0 < t \leqq \mathcal{T},$$

where $f(t)$ is a known pulse shape determined by the distortion (intersymbol interference) in the optical fiber and λ_0 is again the ambient or "dark" current. Here, the optimum receiver maximizes the likelihood function with respect to the data sequence $\{a_n^{(m)}\}_{n=0}^{N}$. As it stands, the likelihood (40) is similar‡ in form to the Volterra kernel description of a general time-varying nonlinear functional. However, such generality seems to preclude any practical value, and furthermore reveals little of the receiver's essence. To obtain a good approximation to the structure of the receiver when $\alpha^2 \to \infty$, it will again be necessary to discretize the photon arrival times.

8.1 The asymptotic ($\alpha \to \infty$) likelihood function

In this section, the basic idea is to asymptotically evaluate the multidimensional sums or integrals. Note that, when $\alpha^2 \to \infty$, the $2n$-fold integrals appearing in the likelihood become increasingly sensitive to the value of the exponent, and in the limit the integral is com-

† Note that, as $R(t) \to \delta(t)$, the gaussian noise becomes transparent to the receiver (since the integrated received signal would be discontinuous whenever an impulse arrived). The receiver then assumes the form of a counter.

‡ The difference is that, in our application, the input $P(t)$ is exponentiated rather than appearing directly.

pletely determined by the coordinates that maximize the exponent. This statement is made precise by the multidimensional version of Laplace's theorem[19] which, apart from certain hypothesis-insensitive terms, gives for each n

$$\lim_{\alpha\to\infty} \int_0^{\mathcal{T}} d\mathbf{t}_n \sum_{\mathbf{g}_n} \prod_{j=1}^{n} \rho(g_i) \prod_{j=1}^{n} \lambda(t_j) \times \exp\left\{\alpha^2\left[\sum_{m=1}^{n} g_m P(t_m) - \frac{1}{2}\sum_{m,k=1}^{n} g_m g_k R(t_m - t_k)\right]\right\} \sim \prod_{j=1}^{n} \rho(g_j^*)\lambda(t_j^*) \times \exp\left\{\alpha^2\left[\sum_{m=1}^{n} g_m^* P(t_m^*) - \frac{1}{2}\sum_{m,k=1}^{n} g_m^* g_k^* R(t_m^* - t_k^*)\right]\right\}, \quad (93)^\dagger$$

where $\{t_1^*, t_2^*, \cdots, t_n^*\}$ and $\{g_1^*, g_2^*, \cdots, g_n^*\}$ maximize the exponent,

$$\sum_{m=1}^{n} g_m P(t_m) - \frac{1}{2}\sum_{m,k=1}^{n} g_m g_k R(t_m - t_k),$$

under the constraint that $0 \leqq t_i < \mathcal{T}$, $i = 1, 2, \cdots, n$. The determination of the extremizing sets appears very difficult. For example, without avalanche gain (i.e., $g_m = 1$) and $n = 1$, it is clear that t_1^* is taken at the point where the observable $P(t)$ is a maximum. For example, when $n = 2$ the exponent becomes

$$P(t_1) + P(t_2) - R(t_1 - t_2),$$

and the choice of t_1 and t_2 is not apparent. The values of t_1^* and t_2^* tend to be near the peaks of $P(t)$, but this is not always the case.[‡] The best choice of t_1 and t_2 will, of course, depend on the interaction of the random process $P(t)$ and the correlation function $R(t)$. The problem of finding the set of points $\{t_1^*\}$ is in some sense equivalent to whitening or peaking up the pulse $w(t)$ in a nonlinear manner to minimize the noise enhancement concomitant with such an operation. Putting aside for the moment the difficulty of determining $\{t_1^*, \cdots, t_n^*\}$ and $\{g_1^*, \cdots, g_n^*\}$, we can use these values to rewrite the right-hand side of (93) as

$$\left[\prod_{j=1}^{n} \rho(g_j^*)\lambda(t_j^*)\right] \exp\left[\alpha^2\left\{\sum_{m=1}^{n} g_m^* P(t_m^*) - \frac{1}{2}\sum_{m,k=1}^{n} g_m^* g_k^* R(t_m^* - t_k^*)\right\}\right] \triangleq \gamma_n(\mathcal{T}) \exp\left[\alpha^2 \beta_n(\mathcal{T})\right], \quad (94)$$

---

† It has been assumed that there is only one set of variables $\mathbf{t} = (t_1, \cdots, t_n)$ and $\mathbf{g} = (g_1, \cdots, g_n)$, which maximize the exponent. If there are several such $\mathbf{t}^*$ and $\mathbf{g}^*$, then the right-hand side of (93) would consist of a sum of these terms. We do not pursue this approach, since the resulting structure is hopelessly complicated and appears to be impractical.

‡ This would be the case whenever $P(t)$ has equivalued maxima spaced at least a decorrelation time apart.

where we have indicated the dependence of both the coefficient and the exponent on the observation interval $\mathcal{T}$. Using (94) in (46) gives the Dirichlet series[21]

$$e^{\Lambda}L \sim \sum_{n=0}^{\infty} \gamma_n(\mathcal{T}) \cdot \exp\left[\alpha^2 \beta_n(\mathcal{T})\right]. \tag{95}$$

As $\alpha \to \infty$, it is well known that the Dirichlet series is dominated by the term with the largest exponent, i.e.,

$$\lim_{\alpha \to \infty} e^{\Lambda}L \sim \gamma_{n^*}(\mathcal{T}) \exp\left\{\alpha^2 \beta_{n^*}(\mathcal{T})\right\}, \tag{96}$$

where β_{n^*} is the largest exponent.† It is evident that n^* is an estimate of the number of (Poisson) events occurring in the interval $\mathcal{T}$ and that $t_1^*, \cdots, t_n^*$ are estimates of these occurrence times, while $g_1^*, \cdots, g_n^*$ are estimates of the avalanche gains. This is not surprising since, as $\alpha^2 \to \infty$, the vanishingly small noise implies that these estimates will be quite accurate. Hence, the receiver is intimately related to the situation considered by Bar-David,[3] where the Poisson events can be observed directly. The distinction is that estimated arrival times and avalanche gains are used rather than their true values. It is important to realize that specific estimators have been obtained for the random parameters. As we show in the sequel, the simultaneous estimation and detection described above can be recursively implemented via dynamic programming.

Since neither the exponent in (94) nor the $\prod_{j=1}^{n^*} \rho(g_i^*)$ term is hypothesis-sensitive, the relevant portion of the likelihood function is

$$L \sim e^{-\Lambda}\gamma_{n^*}(N) = e^{-\Lambda} \prod_{j=1}^{n^*} \lambda(t_j^*), \tag{97}$$

where n^* is the number of time points that maximize the exponent of (92) (which, of course, depends on $\mathcal{T}$) and $\{t_i^*\}_{i=1}^{n}$ are the values of these time points. Note that, once the exponent is jointly optimized with respect to $\mathbf{t}_n$ and $\mathbf{g}_n$, the estimate of the avalanche gain is not utilized further. This is so because the avalanche gain is a property of the photodetector and conveys no information concerning the intensity function. The asymptotic ($\alpha \to \infty$) likelihood given by (97) is exactly Bar-David's[3] likelihood formula, with the true arrival

† If the signal-to-noise ratio is not large enough so that this is not an accurate approximation, then one could designate n_* as the second largest exponent, thereby developing the more accurate series

$$L \sim \exp(-\Lambda)\gamma_{n^*} \exp(\beta_{n^*}) \left(1 + \frac{\gamma_{n_*} \exp(\beta_{n_*})}{\gamma_{n^*} \exp(\beta_{n^*})}\right).$$

times replaced by estimated arrival times. Note that the log-likelihood

$$\bar{L}_m = -\Lambda_m(\mathcal{T}) + \sum_{i=1}^{n^*} \log \lambda^{(m)}(t_i^*) \tag{98}$$

is again a weighted counter, and is similar to (98) derived in Section VII [where the pulses $w(t)$ were assumed to be narrow].

Two shortcomings are associated with the above approach, one is computational and the other involves a question of mathematical rigor. The first point is that implicit in the expression for the likelihood (97) is the ability to solve the formidable mathematical problem,

$$\max_{\substack{g_n, t_n, \text{ and } n \\ 0 \leq t_i \leq \mathcal{T},\ i=1,2,\cdots,n}} \left\{ \sum_{m=1}^{n} g_m P(t_m) - \frac{1}{2} \sum_{m,k=1}^{n} g_m g_k R(t_m - t_k) \right\}, \tag{99}$$

in real time. We are not aware of optimization techniques capable of this accomplishment. The second point involves the invocation of the large α^2 assumption in a *sequence* of operations. Recall that this assumption was used to derive (93) and then used again to obtain (96). While the validity of the preceding operations can perhaps be demonstrated (under suitable conditions), the intractable nature of (99) forces us to slightly reformulate our problem.

8.2 The optimum detector when the photon arrival times are discrete

To proceed further and obtain a physically realizable, as well as meaningful, detector, we discretize the photon arrival times. Adopting this approach, the photon arrival times are now constrained to occur at the discrete instants $j\Delta$, ($j = 0, 1, 2, \cdots, J$, where $J = \mathcal{T}/\Delta$). This gives rise to the discrete likelihood function (41), and eq. (98) then involves only sums rather than integrals. This modified expression contains a $2n + 1$ dimensional sum, which is recognized as a *bona fide* Dirichlet series. Thus, we have avoided the mathematical question concerning the validity of an asymptotic expansion by introducing a mild relaxation of the physical set-up.

Applying the asymptotic condition to the $2n + 1$ variable summation again produces the expressions (94) to (99) where it is recognized that the variables $\{t_i\}$ are now constrained to lie on the lattice, i.e., $t_i = j_i\Delta$, where $j_i = 1, 2, \cdots, J$. We now show that, using this discrete framework, the exponent appearing in (94) can be rewritten in a form readily amenable to maximization. Note that the variables $t_1^*, t_2^*, \cdots, t_n^*$ may be thought of either as specifying a single point in n-dimensional space or as specifying n points on the interval $(0, \mathcal{T})$. This latter viewpoint turns out to be more useful.

We choose the time quantization Δ so that the probability of more than one photon arrival occurring in a time interval of size Δ is vanishingly small† under each hypothesis $\lambda_m(t)$. In this framework, the set of time points $\{t_k^*\}$ specifies n points in the interval $(0, \mathcal{T})$, and the exponent can be rewritten as

$$\sum_{m=1}^{n} g_m P(t_m^*) - \frac{1}{2} \sum_{m,k=1}^{n} g_m g_k R(t_m - t_k)$$

$$= \sum_{m=1}^{J} g_m q_m P(m\Delta) - \frac{1}{2} \sum_{m,k=1}^{J} g_m g_k q_m q_k R(m\Delta - k\Delta), \quad (100)$$

where $0 \leqq t_m^* \leqq J\Delta$ and where q_m is 0 or 1. A value of $q_m = 1$ implies that the time point $m\Delta$ is "active" in the sums appearing in (100), while $q_m = 0$ implies that it is not. If one chooses Δ to provide a coarser quantization of the time axis, as might be required by practical restrictions on the sampling rate, then it is necessary to allow q_m to assume more (integer) values than 0 and 1. To see why this must be the case, recall the physical meaning‡ of the time points $\{t_i^*\}$. It is then realistic to expect that more than one photon will have arrived in a Δ interval and consequently some $t_i^* = t_j^*$ (for $i \neq j$). The increased range of q_m is necessary to accommodate this situation. Realizing that no restriction is implied, for reasons of simplicity we assume in the sequel that Δ is chosen small enough so that $q_m = 0$ or 1. At this point, it is clear that the product $g_m q_m$ is inseparable in the optimization of (100). Note that, once the optimum values of q_m and g_m are determined, only the value of q_m plays a further role in the detection procedure. With this in mind, we let $\beta_m = q_m g_m$, where β_m will range over the allowable values of g_m as well as zero. For convenience, we call this discrete set B. In the context of this new notation, the optimization problem posed in (99) becomes

$$\max_{\substack{\beta_1, \cdots, \beta_J \\ \beta \in B}} \sum_{m=1}^{J} \beta_m P(m\Delta) - \frac{1}{2} \sum_{m,k=1}^{J} \beta_m \beta_k R(m\Delta - k\Delta), \quad (101)$$

where it is important to realize that the maximization with respect to n, appearing in (99), has been removed in (101) by eliminating the restriction that only a predetermined number of q_n's be nonzero. It is also apparent that the exponent is of the required recursive form so that the exponent can be maximized via the Viterbi algorithm. With

† This probability is $1 - e^{-\lambda} - \lambda e^{-\lambda} \approx \lambda^2$.

‡ The $\{t_j^*\}$ are estimates of the pulse arrival times.

this in mind, the likelihood function can now be written as

$$L \sim e^{-\Lambda} \prod_{j=1}^{J} [\lambda(j\Delta)]^{q_j}, \tag{102}$$

and the log-likelihood again assumes the weighted-counter form

$$\bar{L} = -\int_0^{\mathcal{T}} \lambda(t)dt + \sum_{j=1}^{J} q_j \log [\lambda(j\Delta)], \tag{103}$$

which is similar to the detector described by (92) but without the restriction on the correlation function $R(t)$, i.e., $R(t)$ need not be confined to an interval Δ. The result embodied in (92) for nonoverlapping pulses can be easily derived from (101) by setting $R(m\Delta - k\Delta) = \delta_{k-m}$. The exponent then becomes $\sum_{k=1}^{J}[\beta_k P(k\Delta) - \frac{1}{2}\beta_k^2]$, which is optimized, over the integer values of β_k, by choosing β_k to be the quantized version of $P(k\Delta)$.

The structure of the optimum detector (103) is shown in Fig. 11, and is of the estimator-detector type. The arrival time indicators $\{q\}_{i=1}^{J}$ (as well as the avalanche gains) are determined by applying the Viterbi algorithm to the exponent. Once these values are available, the likelihood is computed for each hypothesis $\lambda^{(m)}(t)$ and the maximum is selected.

8.3 Optimum detection of PAM intensities

The above methodology is now applied to the optimum detection of a digital (PAM) data signal. The 2^{N+1} intensity functions in this situation are given by

$$\lambda(t) = \lambda_0 + \sum_{n=0}^{N} a_n f(t - nT), \qquad 0 \leqq t \leqq \mathcal{T},$$

where the effect of optical channel distortion (intersymbol interference) is included in $f(t)$.

To optimally detect these signals, it is convenient to rewrite the original likelihood expression so that time is directly expressed in units of Δ. Bringing out this dependence, the likelihood function then becomes

$$L_J \sim \exp\left\{-\int_0^{J\Delta} \lambda(t)dt\right\} \prod_{j=1}^{J} [\lambda(j\Delta)]^{q_j}, \tag{104}$$

where the index J designates time in units of Δ. Note that the exponent (101) is already expressed in this form.

It is important to emphasize that a simultaneous or *two-tier* real-time sequential optimization procedure is required to extract the ML estimate of the data sequence, $\{a_n\}_{n=0}^{N}$. The exponent is first maximized

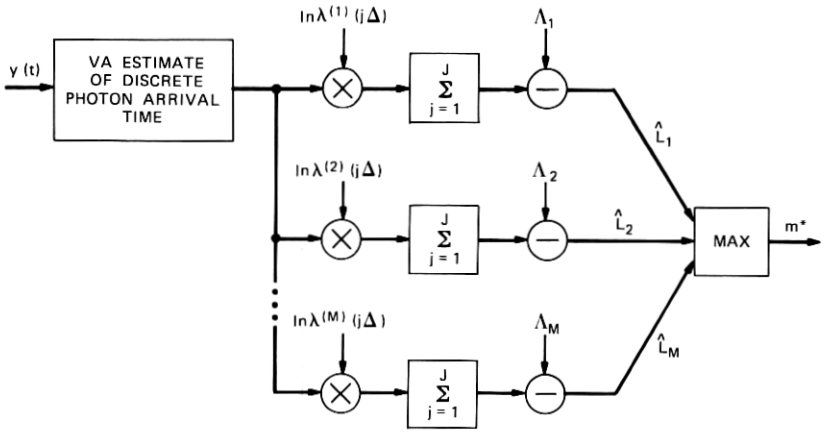

Fig. 11—Estimator detector type of weighted counter.

with respect to the $\{\beta_i\}_{j=0}^{J}$, and the corresponding q_i values are then used to maximize (104) with respect to the data symbols. The optimization of the exponent is identical to that occurring in ML data sequence estimation in the presence of intersymbol interference.[19] The maximization of the exponent will, at random intervals,† produce optimum values of $\{q_j\}$, say, $\hat{q}_0, \hat{q}_1, \cdots, \hat{q}_k$. At this instant, the optimization of the likelihood L_k can then proceed using this new information. At some later instant, $\hat{q}_{k+1}, \hat{q}_{k+2}, \cdots, \hat{q}_{k+n}$ will become available and attention again shifts to maximizing the likelihood L_{k+n}. As we shall show, the dynamic programming algorithm which maximizes the coefficient (103) is quite different from the conventional Viterbi algorithm. In fact, the application of dynamic programming to the iterative‡ maximization of this function illustrates the more general principle that dynamic programming is applicable to the iterative real-time ML sequence estimation of digital data that has undergone a wide variety of nonlinear distortion. The only requirements are that (*i*) the likelihood possesses the mathematical property of additivity and (*ii*) the nonlinearity is of finite memory so that the notion of a "state" is meaningful. In this application, both these requirements are satisfied.

To apply dynamic programming to the optimization problem exhibited in (103), we need only show that the likelihood satisfy a par-

† Owing to the merge aspect of the Viterbi algorithm.

‡ The two main virtues of dynamic programming are that (*i*) it is essentially a real-time processing scheme (although there is random signal-processing delay) and (*ii*) the number of computations is linearly proportional to time (n), as opposed to a straightforward evaluation that requires an exponentially growing number of computations.

ticular recursive form. To put the likelihood in this recursive form, we define the state vector

$$\mathbf{S}_j = (a_{j-1-\eta}, \cdots, a_j) \qquad j = \bar{f}, \bar{f}+1, \cdots, J, \tag{105}$$

where $\bar{f}$ is the memory (in units of Δ) of the dispersed pulse $f(t)$, i.e.,

$$f(n\Delta) = 0, \qquad n > \bar{f}, \tag{106}$$

and where η is the closest integer to $\bar{f}\Delta/T$.

As the optimum $\{\hat{q}_j\}$ time instants emerge from the Viterbi algorithm in a random manner (owing to the merge mechanism), they are classified according to which time segment $(0, NT)$ they belong. Once optimum time instants begin appearing that are active in the $(N+1)T$ time segment, those optimum q_n's which are in the NT time segment are available to maximize the coefficient or, equivalently, the likelihood.

By substituting the PAM signal into (104), the log-likelihood has the form

$$L_J = -\sum_{n=0}^{N} a_n F_n + \sum_{j=0}^{J} q_j \log\left(\lambda_0 + \sum_{m=0}^{N} a_m f(j\Delta - mT)\right), \tag{107}$$

where J is now interpreted as the index of the latest† merge in the Viterbi algorithm associated with the time interval $(0, NT)$ and

$$F_n = \int_0^{J\Delta} f(t - nT)dt. \tag{108}$$

It is important to keep in mind the fact that, once the decisions $(\hat{q}_1, \hat{q}_2, \cdots, \hat{q}_J)$ are available, the iterative procedure for maximizing the likelihood proceeds in units of T. The log-likelihood can be put in the required form by letting $D = T/\Delta$ and writing the likelihood as

$$L_N = -\sum_{n=0}^{N-1} a_n F_n + \sum_{j=0}^{ND-D} q_j \log\left(\lambda_0 + \sum_{m=N-j-\bar{f}/D}^{N-1} a_m f(j\Delta - mD\Delta)\right)$$
$$+ a_N F_N + \sum_{j=ND-D+1}^{ND} q_j \log\left(\lambda_0 + \sum_{m=N-j-\bar{f}/D}^{N} a_m f(j\Delta - mD\Delta)\right). \tag{109}$$

It is crucial to realize that the last term in (109) only involves $a_{j-1-\eta}$, $a_{j-1-\eta+1}, \cdots, a_j$; therefore, with the state vector defined by (105), (109) can be written as

$$L_N = L_{N-1} + h(\mathbf{q}_N; \mathbf{S}_N), \tag{110}$$

where

$$\mathbf{q}_N = (q_{ND-D+1}, \cdots, q_{ND}). \tag{111}$$

† In other words, the next segment of optimum q_n's will penetrate beyond the time instant NT.

Fig. 12—Two-tier dynamic programming algorithm.

It is well known that, through the use of the recursion (110), dynamic programming may be applied to the maximization of L_N.

The resulting receiver is depicted in Fig. 12, and is a two-tier dynamic programming algorithm that simultaneously iterates the exponent and the coefficient to obtain a sequential (or real-time) maximum likelihood sequence estimate of the transmitted sequence $\{a_n\}$. While the above detector requires sampling at a rate that could preclude practical implementation, we remark that, in the large α^2 environment, a peak detector could be used to estimate the photon arrival times. These estimated arrival times would then be used in a dynamic programming algorithm to mitigate the effect of intersymbol interference.

IX. DISCUSSION

The communication-theoretic model for the fiber-optic communication system has proven to be quite useful. Using this model, the optimum (maximum-likelihood) receiver was exhibited under a wide variety of physical circumstances for M-ary and digital PAM signaling. Whether or not the energy in the response of the photodetector to an individual photon is large or small compared to the background gaussian noise, the detector structure turned out to be a weighted counter. The details of how the weighting is carried out have been shown to be complex in some cases. Further investigation into system performance is needed before assessing whether or not such complexity is warranted in any particular application. For values of pulse energy-to-noise ratio (α^2) much less than unity, the structure of the optimum detector can be simply instrumented in terms of analog operations on the photodetector output. On the other hand, when $\alpha^2 \gg 1$, and with or without avalanche gain, we have been unable to realize the optimum detector without first sampling the photo-detector output many times per symbol interval. This procedure may impose practical limitations on the implementation. Since the digital operations are required solely to estimate the photon arrival times, it has been pointed out that certain suboptimum operations (such as peak detection) may be used to estimate these instants. The power of maximum likelihood processing can still be used to mitigate the effect of intersymbol interference.

From a communications and information theoretic point of view, there remain many important and, as yet untouched, problems asso-

ciated with the fiber-optic channel. Sharp bounds on the performance of the various detectors are extremely difficult to obtain, and very little can be said at this time. Also, questions concerned with capacity, reliability, and complexity need be addressed.

X. ACKNOWLEDGMENT

The authors are grateful to S. D. Personick for patiently explaining to them the important aspects of the communication-theoretic model of the fiber-optic channel. We would also like to express our appreciation to D. G. Messerschmidt for his helpful participation in several informative discussions on this subject.

APPENDIX

Optimum Binary Intensities in the Absence of Gaussian Noise

In this appendix, we determine the optimum binary intensities $\lambda_1(t)$ and $\lambda_2(t)$ in the absence of gaussian noise. We proceed initially by neglecting the dark current. Of course, the optimum intensities must satisfy an energy constraint†

$$\int_0^{\mathcal{T}} [\lambda_1(t) + \lambda_2(t)]dt = P. \tag{112}$$

Consider the performance of a system that uses the equiprobable intensities

$$\begin{aligned} \lambda_1(t) &= \frac{P}{\mathcal{T}} \\ \lambda_2(t) &= 0 \end{aligned} \qquad 0 \leqq t \leqq \mathcal{T}. \tag{113}$$

The only way an error can be made under (113) is when $\lambda_1(t)$ is transmitted and no photons arrive; the probability of this event is

$$P_I = \tfrac{1}{2}e^{-P}. \tag{114}$$

Consider now the performance of a system that uses the arbitrary and equiprobable intensities $\lambda_1(t)$ and $\lambda_2(t)$. The probability of error for this system is

$$P_{II} = \tfrac{1}{2}P_1 + \tfrac{1}{2}P_2, \tag{115}$$

where P_1 and P_2 denote the conditional error probabilities given that $\lambda_1(t)$ and $\lambda_2(t)$ are active. Let

$$\Lambda_i = \int_0^{\mathcal{T}} \lambda_i(t)dt, \qquad i = 1, 2, \tag{116}$$

† Since the intensity is proportional to the transmitted optical energy, the constraint is on the average energy.

and let Λ_1 be greater than Λ_2. It is clear that, when Λ_1 is transmitted, the optimum detector must make an error when there are no photon arrivals. These observations provide the following sequence of lower bounds

$$P_{II} \geqq \tfrac{1}{2}P_1 \geqq \tfrac{1}{2}e^{-\Lambda_1}, \tag{117}$$

and since $\Lambda_1 + \Lambda_2 = P$ we have

$$P_{II} \geqq \tfrac{1}{2}e^{-\Lambda_1} \geqq \tfrac{1}{2}e^{-P} = P_I. \tag{118}$$

It is thus established that the intensities described by (113) minimize the probability of error and therefore are optimum. It is also clear that any system that has one of the intensities equal to zero, and the other arbitrary (and satisfying the power constraint), will perform equally as well as (113).

The effect of dark current on the probability of error can be made arbitrarily small by choosing $\lambda_2(t) = 0$ and picking $\lambda_1(t)$ so that the set of points where $\lambda_1(t)$ is nonzero is sufficiently small.

REFERENCES

1. Special Issue on Optical Communication, Proc. IEEE, *58*, No. 10 (October 1970).
2. W. K. Pratt, *Laser Communication Systems*, New York: John Wiley, 1969.
3. I. Bar-David, "Communication Under the Poisson Regime," IEEE Trans. on Information Theory, *IT-15*, No. 1 (January 1969), pp. 31–37.
4. R. M. Gagliardi and S. Karp, "M-ary Poisson Detection and Optical Communications," *IEEE Trans. on Communications Technology, COM-17*, No. 2 (April 1969), pp. 208–216.
5. S. D. Personick, "Receiver Design for Digital Fiber Optic Communication Systems, Part I," B.S.T.J., *52*, No. 6 (July–August 1973), pp. 843–874.
6. S. D. Personick, "Receiver Design for Digital Fiber Optic Communication Systems, Part II," B.S.T.J., *52*, No. 6 (July–August 1973), pp. 875–886.
7. S. D. Personick, "Baseband Linearity and Equalization in Fiber Optic Digital Communication Systems," B.S.T.J., *52*, No. 7 (September 1973), pp. 1175–1194.
8. D. G. Messerschmitt, "Performance of Several Equalizers in a Digital Fiber Optic Receiver," unpublished work.
9. D. G. Messerschmitt, "Optimum Mean-Square Equalization for Digital Fiber Optic Systems," ICC Conference Record, 1975.
10. S. D. Personick, "New Results on Avalanche Multiplication Statistics with Applications to Optical Detection," B.S.T.J., *50*, No. 1 (January 1971), pp. 167–189.
11. H. Melchior, M. B. Fisher, and F. R. Arams, "Photodetectors for Optical Communication Systems," Proc. IEEE, *58*, No. 10 (October 1970), pp. 1466–1486.
12. T. Kailath, "A General Likelihood-Ratio Formula for Random Signals in Gaussian Noise," IEEE Trans. on Information Theory, *IT-15*, No. 3 (May 1969), pp. 350–361.
13. T. T. Kadota, "Nonsingular Detection and Likelihood Ratio for Random Signals in White Gaussian Noise," IEEE Trans. on Information Theory, *IT-16*, No. 3 (May 1970), pp. 291–298.
14. T. Kailath, "A Further Note on a General Likelihood Formula for Random Signals in Gaussian Noise," IEEE Trans. on Information Theory, *IT-16*, No. 4 (July 1970), pp. 393–396.

15. T. T. Kadota and L. A. Shepp, "Conditions for Absolute Continuity Between a Certain Pair of Probability Measures," *Z. Wahrscheilichkeitstheorie*, Feb. 16, 1970, pp. 250–260.
16. E. V. Hoversten, D. L. Snyder, R. O. Harger, and K. Kurimoto, "Direct-Detection Optical Communication Receivers," IEEE Trans. on Communications Technology, *COM-22*, No. 1 (January 1974), pp. 17–27.
17. G. D. Forney, Jr., "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," IEEE Trans. on Information Theory, *IT-18*, No. 3 (May 1972), pp. 363–378.
18. D. V. Widder, *An Introduction to Transform Theory*, New York: Academic Press, 1971, Chapter 2.
19. D. S. Jones, "Asymptotic Behavior of Integrals," SIAM Review, *14*, No. 2 (April 1972), pp. 286–317.