Capacity of a Multimode Direct Detection Optical Communication Channel

H. H. Tan
(Consultant from University of California at Irvine)
Communications Systems Research Section

The capacity of a free-space optical channel with received background noise using a multimode direct detection receiver is derived under both peak and average signal power constraints and without a signal bandwidth constraint. A random telegraph wave type signalling scheme of Kabanov is shown to achieve capacity provided enough signalling bandwidth is available. In the absence of received background noise, an optimally coded PPM system is shown to achieve capacity with greatly reduced bandwidth as compared to Kabanov signals.

I. Introduction

There has been considerable research on optical communication systems in recent years. In particular there is considerable interest (Refs. 1-6) in determining the channel capacity or the maximum theoretically attainable information rate at which reliable communication is possible over optical channels. This article is concerned with the channel capacity of a free-space optical communication system. The reliability of such channels is affected by the quantum mechanical limitations on the measurement of the received optical field as well as the presence of noise in the received field. In addition the channel capacity depends on the particular type of receiver employed – for example, coherent linear amplifier receivers, heterodyne receivers, homodyne receivers and direct detection receivers. Furthermore, constraints imposed on the allowable transmitted signal power also affect the available channel capacity.

In order to put the results of this paper in proper perspective, we briefly review related work. Gordon's benchmark work (Ref. 1) gave the ultimate capacity of any free-space optical communication system under an average signal power constraint. This value of channel capacity determined by Gordon (Ref. 1) places no limitation on the receiver employed other than the quantum mechanical limitation on the accuracy of measurement of the received optical field and the presence of additive noise in the received field. Hence it represents the maximum reliable information rate that can be attained by any system. The importance of this result notwithstanding, the channel capacity with specified receiver structures is also of interest. Gordon (Ref. 1) has obtained the channel capacity using coherent amplification, heterodyne and homodyne receivers respectively. These channel capacities are substantially less than the ultimate channel capacity.

We are concerned here with determining the channel capacity using a direct detection or photon counting receiver. In direct detection systems, the photon counter output can be modeled by a Poisson process with a stochastic intensity rate function (Refs. 8, 10, 17). The stochastic intensity rate process describes the average rate at which photoelectrons are generated by the detector. It contains stochastic signal and stochastic noise components. The noise component in the stochastic intensity rate process arises because of the background noise observed in the receiver's field of view and can be modeled by a Gaussian white noise process (Refs. 8, 17). However, for the purpose of determining channel capacity, this Poisson model does not appear to be tractable without further simplifying assumptions. This is because of the difficulty in dealing with the Gaussian white noise component in the stochastic intensity rate process of the Poisson process. This problem is of course not present in situations of negligible background noise, thus removing the Gaussian white noise component from the stochastic intensity rate process model. This is the situation considered for example in Refs. 3-5, where the channel capacity is obtained under various signal power constraints and signal modulation constraints.

We are concerned here with the situation when the background noise cannot be neglected. Suppose also that the receiver either observes many spatial modes (i.e., a large field of view) and/or many temporal modes exist (i.e., the signal bandwidth is much smaller than the receiver's optical bandwidth) (Ref. 17). Then the stochastic photon arrival rate due to the background noise can be replaced by its expected value (Ref. 17). This is the situation considered here. Kabanov (Ref. 6) has previously derived the channel capacity in this situation under a peak signal power constraint. We shall extend his results by considering simultaneously a peak and an average signal power constraint. Finally, consider the other situation of a small number of spatial and temporal modes. This is the situation of a small field of view at the receiver and signal bandwidths approaching the optical bandwidth of the receiver. In this case Pierce et al. (Ref. 18) have determined the channel capacity.

Summarizing, this article is concerned with a free-space optical communication system using a direct detection receiver. It is assumed that the receiver either observes a large number of spatial modes and/or a large number of temporal modes. This is generally called a multimode direct detection system. The main result of this article (Theorem 1, Section II) gives the average and peak power constrained channel capacity in this situation. The derivation of this theorem is given in Section III. A random telegraph wave type signalling scheme due to Kabanov (Ref. 6) is shown to achieve capacity provided that sufficient signalling bandwidth is available. Finally, in the case of no background noise, it is shown in Section IV that the channel capacity can be achieved using M-ary pulse position modulation (PPM) along with coding, provided that sufficient signalling bandwidth is available. It is also shown that the coded PPM signalling scheme achieves capacity with a greatly reduced bandwidth as compared to the Kabanov signalling scheme.

II. Channel Capacity

Consider an optical channel using an intensity modulated light source transmitter and a multimode direct detection receiver with an ideal photodetector. The transmission medium is assumed to be free space so that no degradation other than a geometric power loss is imparted on the transmitted light beam. The receiver light power at the photodetector is assumed to be weak (Ref. 7) so that the photodetector current output can be characterized by the sequence of time instants of the photon absorption photoelectron emission process of the photodetector. We can then model the photodetector output in terms of a counting process $\{N(t): t \geqslant 0\}$ where $N(t)$ = number of photoelectron emission events in $(0, t)$. Hence this type of receiver is often called a "photon counter." It has been shown (Refs. 8, 17) that $N(t)$ can be modeled as a conditionally Poisson counting process given the intensity of the received light process at the photodetector. Let $\lambda(t)$ be the instantaneous average rate at which photoelectrons are generated at time t in units of photons per second. We shall assume that $\lambda(t)$ is given by

$$\lambda(t) = \lambda_s(t) + \bar{n} \tag{1}$$

where $\lambda_s(t)$ is the instantaneous average rate at which photoelectrons are generated as a result of the received signal field and $\bar{n}$ is the average rate of photoelectron generation due to the received background noise field and detector dark current. We shall assume that $\bar{n}$ is constant since the receiver is assumed to be multimodal (Ref. 17). Let $\{S(t)\}$ be the information bearing message stochastic process that is transmitted. Since $\lambda_s(t)$ depends on this process $\{S(t)\}$, it is also a stochastic process. In the case where there is instantaneous feedback from the photodetector output to the transmitter, $\lambda_s(t)$ can in addition also depend on $\{N(\tau): 0 \leqslant \tau < t\}$. In this case $\{N(t)\}$ is referred to as a compound regular point process by Rubin (Ref. 9), who first studied detection problems involving these processes. In the case where there is no feedback link of any kind present, $\lambda_s(t)$ depends only on the external message signal process $\{S(t)\}$. In this case $\{N(t)\}$ is often called a doubly stochastic Poisson process (Ref. 10). In either case, the stochastic process $\{\lambda(t)\}$ given by Eq. (1) is usually called the intensity rate process of the point process $\{N(t)\}$. We shall call $\{\lambda_s(t)\}$ the signal intensity rate process.

The goal of this article is to determine the channel capacity with constraints on the average and the peak received light signal power. Since $\lambda_s(t)$ is directly proportional to the instantaneous received signal power, we shall impose peak and average value constraints on the admissible $\lambda_s(t)$ processes in calculating the channel capacity. In order to define the channel capacity, denote

$$S_T = \{S(t) : 0 \leqslant t \leqslant T\}, \tag{2}$$

$$N_T = \{N(t) : 0 \leqslant t \leqslant T\}, \tag{3}$$

$$\lambda_{sT} = \{\lambda_s(t) : 0 \leqslant t \leqslant T\}, \tag{4}$$

for each $T > 0$. Also let $I(S_T; N_T)$ = average mutual information between S_T and N_T (in units of nats).

For each $0 < \bar{s} \leqslant \bar{p}$, let $\mathscr{C}(\bar{s}, \bar{p})$ denote the class of all message processes $\{S(t)\}$ and signal intensity rate processes $\{\lambda_s(t)\}$ satisfying the following conditions:

(a) λ_{sT} is a deterministic function of S_T for each $T > 0$.

(b) $0 \leqslant \lambda_s(t) \leqslant \bar{p}$ (5)

(c) For every $T > 0$,

$$G_T(\lambda_{sT}) \overset{\Delta}{=} \frac{1}{T} \int_0^T E\left[\lambda_s(t)\right] dt \leqslant \bar{s}. \tag{6}$$

We shall use $\mathscr{C}(\bar{s}, \bar{p})$ as the admissible class of message processes and signal intensity rate processes for calculating the channel capacity without feedback. In particular, for each $T > 0$ define

$$C_T(\bar{s}, \bar{p}) = \sup_{\{(S(t),\, \lambda_s(t))\} \in \mathscr{C}(\bar{s}, \bar{p})} \frac{1}{T} I(S_T; N_T). \tag{7}$$

By definition (Ref. 11),

$$C(\bar{s}, \bar{p}) = \lim_{T \to \infty} C_T(\bar{s}, \bar{p}) \tag{8}$$

is the channel capacity (in units of nats per second) without feedback under peak power constraint Eq. (5) and average power constraint Eq. (6). It is clear that the maximization in Eq. (7) is over all possible message processes $\{S(t)\}$ and over all possible modulation formats by varying $\lambda_s(t)$. Moreover Eq. (5) limits the received peak signal power and Eq. (6) limits the received average signal power.

Next let $\mathscr{C}_F(\bar{s}, \bar{p})$ denote the class of all message processes $\{S(t)\}$ and signal intensity rate processes $\{\lambda_s(t)\}$ satisfying the conditions (b) and (c) above along with the following condition (a′) in place of (a) above:

(a′) λ_{sT} is a deterministic function of S_T and N_T for each $T > 0$.

Then $\mathscr{C}_F(\bar{s}, \bar{p})$ will be the admissible class of message and signal intensity rate processes used to calculate the channel capacity with feedback. That is, define

$$C_{FT}(\bar{s}, \bar{p}) = \sup_{\{(S(t),\, \lambda_s(t))\} \in \mathscr{C}_F(\bar{s}, \bar{p})} \frac{1}{T} I(S_T; N_T), \tag{9}$$

for each $T > 0$ and let

$$C_F(\bar{s}, \bar{p}) = \lim_{T \to \infty} C_{FT}(\bar{s}, \bar{p}). \tag{10}$$

$C_F(\bar{s}, \bar{p})$ is the channel capacity (in units of nats per second) with feedback under peak power constraint Eq. (5) and average power constraint Eq. (6). The presence of instantaneous feedback arises through the possible dependence of the modulation format $\lambda_s(t)$ on the photodetector output as given by condition (a′) defining $\mathscr{C}_F(\bar{s}, \bar{p})$. The following theorem gives an expression for $C(\bar{s}, \bar{p})$ and also shows that $C(\bar{s}, \bar{p}) = C_F(\bar{s}, \bar{p})$. That is, the use of instantaneous noiseless feedback does not increase channel capacity. This feedback result is really just a special case of a similar result (Ref. 19) that is valid for all memoryless channels. The Poisson optical channel considered here is also a memoryless channel.

Theorem 1

Let

$$\sigma = \min \left\{ s, (\bar{p} + \bar{n}) \exp\left[\frac{\bar{n}}{\bar{p}} \log \left(1 + \frac{\bar{p}}{\bar{n}}\right) - 1 \right] - \bar{n} \right\} \text{ photons/sec.} \tag{11}$$

Then

$$C_F(\bar{s}, \bar{p}) = C(\bar{s}, \bar{p})$$

where

$$C(\bar{s}, \bar{p}) = \left(1 - \frac{\sigma}{\bar{p}}\right) \bar{n} \log \bar{n} + \left(\frac{\sigma}{\bar{p}}\right) (\bar{p} + \bar{n}) \log (\bar{p} + \bar{n})$$

$$- (\sigma + \bar{n}) \log (\sigma + \bar{n}) \text{ nats/second.} \tag{12}$$

Moreover, in order to achieve channel capacity with or without feedback, the received average signal power must be σ photons/second.

The case when $\bar{n} = 0$ corresponds to the situation where there are no received photons due to extraneous background radiation and when the detector dark current is zero. This situation has received considerable recent attention

(Refs. 3-5). Setting $\bar{n}=0$ in the above theorem yields the following expression for channel capacity in the no background noise case.

Corollary 1

When

$$\bar{n} = 0, C_F(\bar{s},\bar{p}) = C(\bar{s},\bar{p}),$$

where

$$C(s,p) = \sigma \log \frac{\bar{p}}{\sigma} \text{ nats/sec.} \tag{13}$$

and

$$\sigma = \min(\bar{s}, e^{-1}\bar{p}) \text{ photons/sec.} \tag{14}$$

is the received average signal power required to achieve channel capacity.

Since the average received signal power $\bar{s}$ is always less than or equal to the peak received signal power $\bar{p}$, then by setting $\bar{s}=\bar{p}$ we effectively remove the constraint on the average received signal power. That is, $C_F(\bar{p},\bar{p})$ and $C(\bar{p},\bar{p})$ are respectively the peak power constrained channel capacity with feedback and the peak power constrained channel capacity without feedback. Hence we obtain the following corollary.

Corollary 2

Under only a peak received signal power constraint of $\bar{p}$, the channel capacity with feedback is $C_F(\bar{p},\bar{p})$ and the channel capacity without feedback is $C(\bar{p},\bar{p})$. Then $C_F(\bar{p},\bar{p}) = C(\bar{p},\bar{p})$ is given by Eq. (12) with

$$\sigma = (\bar{p}+\bar{n}) \exp\left[\frac{\bar{n}}{\bar{p}}\log\left(1+\frac{\bar{p}}{\bar{n}}\right)-1\right] - \bar{n}. \tag{15}$$

Proof

We need only show that $\bar{p} \geqslant \sigma$ given by Eq. (11). This is true because it is clear that

$$\left(\frac{\bar{n}}{\bar{p}}\right)\log\left(1+\frac{\bar{p}}{\bar{n}}\right)-1 \leqslant 0 \tag{16}$$

by using the inequality $\log(1+x) \leqslant x$. Q.E.D.

We note that Kabanov's expression (Ref. 6) for the peak received signal power constrained channel capacity when $\bar{n}=1$ is a special case of Corollary 2.

In most laser communication systems, the available peak signal power is usually substantially greater than the available average signal power. The usual case is that $\bar{p} >> \bar{s}$. It is interesting to note that in contrast to additive Gaussian noise channels (Ref. 13), the primary constraint on the available channel capacity is the available peak signal power, rather than the available average signal power. This can be easily seen from Eqs. (11) and (12) where for a fixed background noise level $\bar{n}$ and fixed available average signal power $\bar{s}$, the channel capacity $C(\bar{s},\bar{p})$ can be made arbitrarily large by making the available peak signal power $\bar{p}$ arbitrarily large. We also note in this regard that constraining only the available peak signal power results in an unrealistically large estimate of the maximum achievable reliable information rate of the channel. This can be seen more readily in the no background noise case ($\bar{n}=0$), where the peak signal power constrained channel capacity is

$$C(\bar{p},\bar{p}) = e^{-1}\bar{p} \text{ nats/sec} \tag{17}$$

and the average and peak signal power constrained capacity is

$$C(\bar{s},\bar{p}) = \bar{s}\log\left(\frac{\bar{p}}{\bar{s}}\right) \text{ nats/sec} \tag{18}$$

when $\bar{p} \geqslant e\bar{s}$. So in the usual case when $\bar{p} >> \bar{s} >> 1$, $C(\bar{s},\bar{p})$ is substantially smaller than $C(\bar{p},\bar{p})$. For example, in a deep space optical channel, present technology (Ref. 5) can achieve a system with negligible background noise, $\bar{s}=10^4$ photons/sec and $\bar{p}=10^7$ photons/sec. Then $C(\bar{s},\bar{p}) = 6.9 \times 10^4$ nats/sec and $C(\bar{p},\bar{p}) = 3.7 \times 10^6$ nats/sec. We also note from Eq. (13) that in order to achieve $C(\bar{p},\bar{p})$, $\bar{s}$ must be at least $e^{-1}\bar{p}$. In a laser communication system, the available peak signal power must be reduced considerably to attain such a high average to peak signal power ratio. Similar conclusions can be reached in the case where there is background noise present.

The derivation of Eq. (11) and (12) follows Kabanov's approach (Ref. 6) and involves the following two steps:

(1) The first step establishes the formula Eq. (12) as an upper bound on $C_F(\bar{s},\bar{p})$ and so also an upper bound on $C(\bar{s},\bar{p})$.

(2) The second step gives a sequence of message processes

$$\{S_m(t) : 0 \leqslant t \leqslant T\}$$

and signal intensity rate processes

$$\{\lambda_s^{(m)}(t) : 0 \leqslant t \leqslant T\}$$

belonging to $\mathscr{C}(\bar{s},\bar{p})$ (see Eqs. (48) and (51) in Section III) with average mutual information $I(S_{mT}; N_T)$ and demonstrates that $I(S_{mT}; N_T)/T$ converges to Eq. (12) as m tends to infinity. This then proves that the upper bound in Step (1) can be achieved and is equal to $C(\bar{s},\bar{p})$ and $C_F(\bar{s},\bar{p})$.

The technical details involved in these two steps are discussed in the next section of this article.

The optical signalling bandwidth of this photon counting optical communication system can be taken to be the bandwidth of the signal intensity rate process. Section IV of this article investigates the optical signalling bandwidth required to achieve channel capacity. The bandwidth of the signal intensity rate process $\{\lambda_s^{(m)}(t)\}$ used in Step (2) above to achieve channel capacity is determined and is shown to be unbounded as m tends to infinity. Thus the formula Eq. (12) is the channel capacity without bandwidth constraint. The rate at which the bandwidth of $\{\lambda_s^{(m)}(t)\}$ increases as capacity is approached is also derived. In the case where there is no background noise, an optimally coded pulse-position modulation (PPM) system is shown to be capable of achieving channel capacity at a reduced bandwidth as compared to the signalling scheme used in Step (2). The reader who is more interested in these results than the involved technical details in Section III can skip that section and go directly to Section IV without essential loss of continuity. Section V relates the results of this article to previous work.

III. Derivation of Channel Capacity

We first carry out Step (1) to establish that Eq. (12) is an upper bound on $C_F(\bar{s},\bar{p})$ and hence also an upper bound on $C(\bar{s},\bar{p})$ since $C(\bar{s},\bar{p}) \leqslant C_F(\bar{s},\bar{p})$. In order to examine Eqs. (6) and (7) we must use the following formula for the average mutual information, which is valid for all $\{(S(t), \lambda_s(t))\}$ in $\mathscr{C}_F(\bar{s},\bar{p})$:

$$I(S_T;N_T) = \int_0^T \{E[(\lambda_s(t)+\bar{n})\log(\lambda_s(t)+\bar{n})] - E[(\hat{\lambda}_s(t)+\bar{n})\log(\hat{\lambda}_s(t)+\bar{n})]\}\,dt \tag{19}$$

where

$$\hat{\lambda}_s(t) = E[\lambda_s(t)|N_t] \tag{20}$$

is the conditional mean estimator of $\lambda_s(t)$ given observations $N_t = \{N(\tau): 0 \leqslant \tau \leqslant t\}$. The formula Eq. (19) is given in Ref. 12. We provide a formal derivation of Eq. (19) in Appendix A for completeness. Note from Eq. (20) that $I(S_T; N_T)$ depends on S_T only through λ_{sT}. So denote

$$I(S_T;N_T) \overset{\Delta}{=} I(\lambda_{sT}). \tag{21}$$

Note that since the function $f(x) = x \log x$ is convex, Jensen's inequality gives

$$\begin{aligned} &E[(\hat{\lambda}_s(t)+\bar{n})\log(\hat{\lambda}_s(t)+\bar{n})] \\ &\geqslant [E(\hat{\lambda}_s(t)+\bar{n})]\log[E(\hat{\lambda}_s(t)+\bar{n})] \\ &= [E(\lambda_s(t)+\bar{n})]\log[E(\lambda_s(t)+\bar{n})], \end{aligned} \tag{22}$$

the last equality following since $E[\hat{\lambda}_s(t)] = E[E[\lambda_s(t)|N_t]] = E[\lambda_s(t)]$. Hence Eqs. (19) and (22) yield

$$\begin{aligned} I(\lambda_{sT}) \leqslant \int_0^T &\{E[(\lambda_s(t)+\bar{n})\log(\lambda_s(t)+\bar{n})] \\ &- [E(\lambda_s(t)+\bar{n})]\log[E(\lambda_s(t)+\bar{n})]\}\,dt \\ &\overset{\Delta}{=} J(\lambda_{sT}). \end{aligned} \tag{23}$$

From Eqs. (9), (21) and (23) we have the following upper bound on $C_{FT}(\bar{s},\bar{p})$:

$$T\,C_{FT}(\bar{s},\bar{p}) \leqslant \sup_{\{(S(t),\,\lambda_s(t))\}\,\in\,\mathscr{C}_F(\bar{s},\bar{p})} J(\lambda_{sT}) \overset{\Delta}{=} J_T(\bar{s},\bar{p}). \tag{24}$$

We shall solve the optimization problem Eq. (24) to determine $J_T(\bar{s},\bar{p})$. Let us first introduce a slack variable for this optimization problem to change the inequality constraint Eq. (6) to an equality constraint. In particular, define $\mathscr{D}_F(\bar{s},\bar{p})$ to be the set of all message processes $\{S(t)\}$, signal intensity rate processes $\{\lambda_s(t)\}$ and nonnegative numbers x such that conditions (a′) and (b) in the definition of $\mathscr{C}_F(\bar{s},\bar{p})$ hold and such that

$$0 \leqslant x \leqslant \bar{s}, \tag{25}$$

$$G_T(\lambda_{sT}) + x = \bar{s}. \tag{26}$$

Then it is clear from the definitions of $\mathscr{C}_F(\bar{s},\bar{p})$ and $\mathscr{D}_F(\bar{s},\bar{p})$, and Eqs. (24) through (26) that

$$J_T(\bar{s},\bar{p}) = \sup_{(x,\{(S(t),\lambda_s(t))\})\in \mathscr{D}_F(\bar{s},\bar{p})} J(\lambda_{sT}). \qquad (27)$$

Let us next introduce the Lagrange multiplier for the equality constraint Eq. (26) in the optimization problem Eq. (27). In particular define $\hat{\mathscr{D}}_F(\bar{p})$ to be the set of all message processes $\{S(t)\}$ and signal intensity rate processes $\{\lambda_s(t)\}$ such that conditions (a′) and (b) in the definition of $\mathscr{C}_F(\bar{s},\bar{p})$ hold.

Now for each real number $\mu \geqslant 0$, consider the following optimization problem:

$$\sup_{\substack{0\leqslant x\leqslant \bar{s} \\ \{(S(t),\lambda_s(t))\}\in \hat{\mathscr{D}}_F(\bar{p})}} J(\lambda_{sT}) - \mu[G_T(\lambda_{sT}) + x]. \qquad (28)$$

The following proposition then relates (28) to (27).

Proposition 1

Suppose there exists a $x^* \epsilon\ [0,\bar{s}]$ and $\{(S^*(t),\ \lambda_s^*(t))\} \epsilon\ \hat{\mathscr{D}}_F(\bar{p})$ which achieves the supremum in Eq. (28) for some $\mu \geqslant 0$ so that $(x^*,\ \{(S^*(t),\ \lambda_s^*(t))\}) \epsilon\ \mathscr{D}_F(\bar{s},\bar{p})$. Then this $(x^*,\ \{(S^*(t),\ \lambda_s^*(t))\})$ achieves the supremum in Eq. (27).

Comment

This proposition essentially says that if we find a Lagrange multiplier $\mu \geqslant 0$ so that the solution of the optimization problem Eq. (28) is also feasible for the optimization problem Eq. (27), then this solution of Eq. (28) also solves Eq. (27). Note that since $J(\lambda_{sT})$ is not concave in λ_{sT}, generalized Kuhn-Tucker theorems valid in function space could not be invoked here to obtain this proposition.

Proof

By hypothesis,

$$(x^*,\ \{(S^*(t),\ \lambda_s^*(t))\}) \epsilon\ \mathscr{D}_F(\bar{s},\bar{p}).$$

So from Eq. (27),

$$J(\lambda_{sT}^*) \leqslant J_T(\bar{s},\bar{p}), \qquad (29)$$

where

$$\lambda_{sT}^* = \{\lambda_s^*(t) : 0 \leqslant t \leqslant T\}.$$

But for this $\mu \geqslant 0$, it follows from Eqs. (26) and (27) that

$$\begin{aligned} J_T(\bar{s},\bar{p}) &= \sup_{(x,\ \{(S(t),\lambda_s(t))\})\in \mathscr{D}_F(\bar{s},\bar{p})} J(\lambda_{sT}) \\ &= \sup_{(x,\ \{(S(t),\lambda_s(t))\})\in \mathscr{D}_F(\bar{s},\bar{p})} J(\lambda_{sT}) - \mu[G_T(\lambda_{sT}) + x - \bar{s}] \\ &\overset{①}{\leqslant} \sup_{\substack{0\leqslant x\leqslant \bar{s} \\ \{(S(t),\lambda_s(t))\}\in \hat{\mathscr{D}}_F(\bar{p})}} J(\lambda_{sT}) - \mu[G_T(\lambda_{sT}) + x - \bar{s}] \\ &\overset{②}{=} J(\lambda_{sT}^*) - \mu[G_T(\lambda_{sT}^*) + x^* - \bar{s}] \\ &\overset{③}{=} J(\lambda_{sT}^*), \end{aligned} \qquad (30)$$

where ① is because $(x,\ \{(S(t),\ \lambda_s(t))\}) \epsilon\ \mathscr{D}_F(\bar{s},\bar{p})$ implies that $0 \leqslant x \leqslant \bar{s}$ and $\{(S(t),\ \lambda_s(t))\} \epsilon\ \hat{\mathscr{D}}_F(\bar{p})$; ② is because $(x^*,\ \{(S^*(t),\ \lambda_s^*(t))\})$ achieves the supremum in Eq. (28); and ③ is because by hypothesis $G_T(\lambda_{sT}^*) + x^* = \bar{s}$ since $(x^*,\ \{(S^*(t),\ \lambda_s^*(t))\}) \epsilon\ \mathscr{D}_F(\bar{s},\bar{p})$. Then Eqs. (29) and (30) establish the proposition.

Q.E.D.

We shall solve the optimization problem Eq. (27) by finding a solution of Eq. (28) which satisfies the hypothesis of Proposition 1. Let $\mu \geqslant 0$ be arbitrary for the time being. We shall restrict μ later. The optimization problem Eq. (28) may be written as

$$\sup_{0\leqslant x\leqslant \bar{s}} \left\{ \sup_{\{(S(t),\lambda_s(t))\}\in \hat{\mathscr{D}}_F(\bar{p})} [J(\lambda_{sT}) - \mu\, G_T(\lambda_{sT})] - \mu x \right\}. \qquad (31)$$

Then from Eqs. (6) and (23) we have

$$\sup_{\{(S(t),\lambda_s(t))\}\in \hat{\mathscr{D}}_F(\bar{p})} J(\lambda_{sT}) - \mu\, G_T(\lambda_{sT})$$

$$= \sup_{\{(S(t),\, \lambda_s(t))\} \in \hat{\mathscr{D}}_F(\bar{p})} \int_0^T \Big\{ E[(\lambda_s(t) + \bar{n}) \log (\lambda_s(t) + \bar{n})]$$

$$- [E(\lambda_s(t) + \bar{n})] \log [E(\lambda_s(t) + \bar{n})]$$

$$- \left(\frac{\mu}{T}\right) E[\lambda_s(t)] \Big\} \, dt$$

$$\overset{①}{=} T \sup_{\Lambda \in \mathscr{R}(\bar{p})} \Big\{ E[(\Lambda + \bar{n}) \log (\Lambda + \bar{n})]$$

$$- [E(\Lambda + \bar{n})] \log [E(\Lambda + \bar{n})] - \left(\frac{\mu}{T}\right) E[\Lambda] \Big\}, \tag{32}$$

where $\mathscr{R}(\bar{p})$ is the set of all random variables Λ such that $0 \leqslant \Lambda \leqslant \bar{p}$. To establish ① in Eq. (32), note that for each $t \in [0, T]$, the integrand in the preceding line cannot be larger than $[1/T \times$ last line of Eq. (32)$]$. Hence the last line of Eq. (32) is an upper bound. This upper bound can be achieved if we restrict $\{(S(t),\ \lambda_s(t))\}$ to be such that for every t, $\lambda_s(t) = \Lambda \in \mathscr{R}(\bar{p})$ in the supremum in the preceding line.

In Appendix B we prove the following proposition, which gives the solution of the optimization problem in the right-hand side of Eq. (32).

Proposition 2

Suppose $\mu \geqslant 0$ is such that $k \geqslant 0$, where

$$k = (\bar{p} + \bar{n}) \exp \left[- \left(\frac{\mu}{T} + 1 \right) + \frac{\bar{n}}{\bar{p}} \log \left(1 + \frac{\bar{p}}{\bar{n}} \right) \right] - \bar{n}. \tag{33}$$

Then

$$T \sup_{\Lambda \in \mathscr{R}(\bar{p})} \Big\{ E[(\Lambda + \bar{n}) \log (\Lambda + \bar{n})] - [E(\Lambda + \bar{n})] \log [E(\Lambda + \bar{n})]$$

$$- \left(\frac{\mu}{T}\right) E[\Lambda] \Big\}$$

$$= T \left[\left(\frac{k}{\bar{p}}\right) (\bar{p} + \bar{n}) \log (\bar{p} + \bar{n}) + \left(1 - \frac{k}{\bar{p}}\right) \bar{n} \log \bar{n} \right.$$

$$\left. - (k + \bar{n}) \log (k + \bar{n}) - \left(\frac{\mu}{T}\right) k \right]. \tag{34}$$

where the $\Lambda^* \in \mathscr{R}(\bar{p})$ which achieves the supremum in Eq. (34) has distribution

$$P(\Lambda^* = \bar{p}) = 1 - P(\Lambda^* = 0) = \frac{k}{\bar{p}}. \tag{35}$$

In particular $E[\Lambda^*] = k$.

From Eqs. (31), (32) and (34) of Proposition 2, the value of this optimization problem Eq. (28) for $\mu \geqslant 0$ such that $k \geqslant 0$ is given by

$$\sup_{\substack{0 \leqslant x \leqslant \bar{s} \\ \{(S(t),\, \lambda_s(t))\} \in \hat{\mathscr{D}}_F(\bar{p})}} J(\lambda_{sT}) - \mu [G_T(\lambda_{sT}) + x]$$

$$= \sup_{0 \leqslant x \leqslant \bar{s}} \Big\{ T \left[\left(\frac{k}{\bar{p}}\right) (\bar{p} + \bar{n}) \log (\bar{p} + \bar{n}) + \left(1 - \frac{k}{\bar{p}}\right) \bar{n} \log \bar{n} \right.$$

$$\left. - (k + \bar{n}) \log (k + \bar{n}) - \left(\frac{\mu}{T}\right) k \right] - \mu x \Big\}$$

$$= T \left[\left(\frac{k}{\bar{p}}\right) (\bar{p} + \bar{n}) \log (\bar{p} + \bar{n}) + \left(1 - \frac{k}{\bar{p}}\right) \bar{n} \log \bar{n} \right.$$

$$\left. - (k + \bar{n}) \log (k + \bar{n}) - \left(\frac{\mu}{T}\right) k \right]. \tag{36}$$

Moreover, the solution $x^* \in [0, \bar{s}]$ and $\{(S^*(t),\ \lambda_s^*(t))\} \in \hat{\mathscr{D}}_F(\bar{p})$ of this optimization problem is given as follows for the two cases $\mu > 0$ and $\mu = 0$.

Case 1: $\mu > 0$ such that $k \geqslant 0$

(i) $x^* = 0$ (37)

(ii) $$\lambda_s^*(t) = \Lambda^* = \begin{cases} \bar{p}, \text{ with probability } \dfrac{k}{\bar{p}} \\ 0, \text{ with probability } \left(1 - \dfrac{k}{\bar{p}}\right) \end{cases} \tag{38}$$

Case 2: $\mu = 0$ (in which case $k \geqslant 0$) (see Appendix B).

(i) x^* arbitrary in $[0, \bar{s}]$. (39)

(ii) $\lambda_s^*(t)$ given by Eq. (38).

Also, in either Case 1 or Case 2, Eq. (38) gives

$$G_T(\lambda_{sT}^*) = k \tag{40}$$

In order to solve the optimization problem Eq. (27) we now appeal to Proposition 1. From Proposition 1, we need to find $\mu \geqslant 0$ so that the above solution $x^* \in [0, \bar{s}]$, $\{(S^*(t),\ \lambda_s^*(t))\} \in \hat{\mathscr{D}}_F(\bar{p})$ of the optimization problem Eq. (28) is also feasible for the optimization problem Eq. (27). That is, we need to find $\mu \geqslant 0$ so that $(x^*, \{(S^*(t),\ \lambda_s^*(t))\}) \in \mathscr{D}_F(\bar{s}, \bar{p})$. In examining the definition of $\hat{\mathscr{D}}_F(\bar{p})$ and $\mathscr{D}_F(\bar{s}, \bar{p})$, it is clear that we need only find $\mu \geqslant 0$ so that

$$G_T(\lambda_{sT}^*) + x^* = \bar{s}. \tag{41}$$

To do this we consider two separate cases below:

Case I: $\bar{s} < (\bar{p}+\bar{n})\exp\left[\frac{\bar{n}}{\bar{p}}\log\left(1+\frac{\bar{p}}{\bar{n}}\right)-1\right]-\bar{n}$

In this case it follows from Eq. (33) that there is a $\mu > 0$ so that

$$k = (\bar{p}+\bar{n})\exp\left[-\frac{\mu}{T}+\frac{\bar{n}}{\bar{p}}\log\left(1+\frac{\bar{p}}{\bar{n}}\right)-1\right]-\bar{n} = \bar{s}. \quad (42)$$

So here, Case I above applies for this $\mu > 0$. Then from Eqs. (37), (40) and (42) it is clear that Eq. (41) is true. So the hypothesis of Proposition 1 is satisfied. Moreover, Eqs. (23), (27), (38), (42) and Proposition 1 then give the following expression for $J_T(\bar{s},\bar{p})$:

$$J_T(\bar{s},\bar{p}) = T\left[\left(1-\frac{\sigma}{\bar{p}}\right)\bar{n}\log\bar{n}+\left(\frac{\sigma}{\bar{p}}\right)(\bar{p}+\bar{n})\log(\bar{p}+\bar{n}) - (\sigma+\bar{n})\log(\sigma+\bar{n})\right] \quad (43)$$

with

$$\sigma = \bar{s} \quad (44)$$

Case II: $\bar{s} \geqslant (\bar{p}+\bar{n})\exp\left[\frac{\bar{n}}{\bar{p}}\log\left(1+\frac{\bar{p}}{\bar{n}}\right)-1\right]-\bar{n}$

Let $\mu = 0$. It then follows from Eq. (33) with $\mu = 0$ that

$$k = (\bar{p}+\bar{n})\exp\left[\frac{\bar{n}}{\bar{p}}\log\left(1+\frac{\bar{p}}{\bar{n}}\right)-1\right]-\bar{n} \leqslant \bar{s}. \quad (45)$$

Since $\mu = 0$, Case II above applies. Now $x^* = \bar{s} - k$ satisfies Eq. (39) since $0 \leqslant k \leqslant \bar{s}$. Also it follows from Eq. (40) that Eq. (41) is true for this choice of x^*. Hence the hypothesis of Proposition 1 is satisfied. Finally Eqs. (23), (27), (38), (45), and Proposition 1 show that $J_T(\bar{s},\bar{p})$ is given by Eq. (43) with

$$\sigma = (\bar{p}+\bar{n})\exp\left(\frac{\bar{n}}{\bar{p}}\log\left(1+\frac{\bar{p}}{\bar{n}}\right)-1\right)-\bar{n}. \quad (46)$$

We can now conclude from Eqs. (10), (24), (43), (44) and (46) the following upper bound on $C_F(\bar{s},\bar{p})$:

$$C_F(\bar{s},\bar{p}) \leqslant \left(1-\frac{\sigma}{\bar{p}}\right)\bar{n}\log\bar{n}+\left(\frac{\sigma}{\bar{p}}\right)(\bar{p}+\bar{n})\log(\bar{p}+\bar{n}) - (\sigma+\bar{n})\log(\sigma+\bar{n}), \quad (47)$$

where σ is given by Eq. (11). That is, we have established that the expression Eq. (12) in the theorem is an upper bound on $C_F(\bar{s},\bar{p})$ and hence, also an upper bound on $C(\bar{s},\bar{p})$. This completes Step (1).

Let us now carry out Step (2). In his derivation (Ref. 6) of the peak power constrained channel capacity, Kabanov constructed a sequence of signal processes $\{S_m(t)\}$ and associated signal intensity rate process $\{\lambda_s^{(m)}(t)\}$ satisfying the peak power constraint, so that for each T, $(1/T)I(S_{mT};N_T)$ converges to the upper bound on channel capacity as $m\to\infty$. Our derivation above in Step (1) of the upper bound on the peak and average power constrained capacity differs considerably from Kabanov's work (Ref. 6). In the sequel, however, we shall show that Kabanov's construction can still be used to attain our upper bound.

Kabanov's construction applied here is as follows. Denote $1_A(\ )$ as the indicator function of the set A. For each integer $m \geqslant 1$, define a $\{0,1\}$- valued left-continuous stochastic process $\{S_m(t): t\geqslant 0\}$ by

$$S_m(t) = \left(\frac{1}{2}\right)-\left(\frac{1}{2}\right)[1-2\times 1_{\{1\}}(S_m(0))](-1)^{M(t)} = \begin{cases}\left(\frac{1}{2}\right)+\left(\frac{1}{2}\right)(-1)^{M(t)}, & \text{if } S_m(0) = 1\\ \left(\frac{1}{2}\right)-\left(\frac{1}{2}\right)(-1)^{M(t)}, & \text{if } S_m(0) = 0\end{cases} \quad (48)$$

for $t > 0$ where $S_m(0) = S_{m0}$ has distribution

$$\alpha \triangleq P(S_{m0} = 1) = 1 - P(S_{m0} = 0) = \frac{\sigma}{\bar{p}} \quad (49)$$

and where σ is given by Eq. (11). Here $\{M(t)\}$ is a regular point process (Ref. 9) with intensity rate process

$$\nu(t) = m\,1_{\{0\}}(S_m(t^-)) + m\left(\frac{1-\alpha}{\alpha}\right)1_{\{1\}}(S_m(t^-)) = \begin{cases} m & ,\text{if } S_m(t^-) = 0\\ m\left(\frac{1-\alpha}{\alpha}\right) & ,\text{if } S_m(t^-) = 1.\end{cases} \quad (50)$$

In other words, $\{S_m(t)\}$ takes on values 0 or 1 and switches between 0 and 1 at random times according to the occurrence times of the point process $\{M(t)\}$. The instantaneous average rate at which these point occurrences arrive at a given time t depends on the immediate past value of $S_m(t)$, being of rate m

when $S_m(t^-) = 0$ and rate $m(1-\alpha/\alpha)$ when $S_m(t^-) = 1$. Finally, set

$$\lambda_s^{(m)}(t) = \bar{p}\,|S_m(t)| = \bar{p}\,S_m(t). \tag{51}$$

It can be seen from Eqs. (48) and (51) that in the Kabanov signalling scheme $\lambda_s^{(m)}(t)$ is a random telegraph wave type process. A typical sample path of this process is shown in Fig. 2 in the case when $\alpha = \sigma/\bar{p} << 1$ (for high peak-to-average signal power ratios). When $\alpha << 1$ it can be seen from the rate process $\nu(t)$ given by Eq. (50) that $S_m(t)$ stays in the 0-state a larger percentage of time than in the 1-state. This results in a typical $\lambda_s^{(m)}(t)$ as shown in Fig. 2.

It is clear that $0 \leqslant \lambda_s^{(m)}(t) \leqslant \bar{p}$. Kabanov (Ref. 6) has shown that $E[S_m(t)] = \alpha$. An elaboration of his derivation is given in Appendix C for completeness. Since $E[S_m(t)] = \alpha$, then from Eqs. (49) and (51) we have

$$E[\lambda_s^{(m)}(t)] = \sigma, \tag{52}$$

so from Eqs. (11) and (52) we can conclude that

$$\frac{1}{T}\int_0^T E[\lambda_s^{(m)}(t)]\,dt = \sigma \leqslant \bar{s}. \tag{53}$$

Finally, it is clear from Eq. (51) that $\lambda_{sT}^{(m)} = \{\lambda_s^{(m)}(t) : 0 \leqslant t \leqslant T\}$ is a deterministic function of $S_{mT} = \{S_m(t) : 0 \leqslant t \leqslant T\}$ for each $T > 0$. So $(\{S_m(t)\}, \{\lambda_s^{(m)}(t)\}) \in \mathscr{C}(\bar{s},\bar{p})$ for each integer $m \geqslant 1$. Thus we conclude from Eq. (7) that for each $T > 0$ and each $m \geqslant 1$,

$$C_T(\bar{s},\bar{p}) \geqslant \frac{1}{T}\, I(S_{mT}; N_T). \tag{54}$$

Next, a minor modification of Kabanov's proof (Ref. 6) is given in Appendix D to show that for each $T > 0$,

$$\lim_{m\to\infty} \frac{1}{T} I(S_{mT}; N_T) = \left(1 - \frac{\sigma}{\bar{p}}\right)\bar{n}\log\bar{n} + \left(\frac{\sigma}{\bar{p}}\right)(\bar{p}+\bar{n})$$

$$\log(\bar{p}+\bar{n}) - (\sigma+\bar{n})\log(\sigma+\bar{n}). \tag{55}$$

Since $C(\bar{s},\bar{p}) \leqslant C_P(\bar{s},\bar{p})$, Eqs. (8), (47) and (55) show that $C(\bar{s},\bar{p}) = C_P(\bar{s},\bar{p})$ is given by Eq. (12), thus completing Step (2) and establishing the theorem.

IV. Signal Bandwidth and Coded PPM

Consider the sequence of signal intensity rate processes $\{\lambda_s^{(m)}(t)\}$ given by Eq. (51) used in the previous section to attain channel capacity. The bandwidth of this process can be taken to represent the optical signalling bandwidth of the channel. In order to examine the bandwidth of the $\{\lambda_s^{(m)}(t)\}$ process, consider

$$K_\lambda(t,\tau) = \mathrm{Cov}\,(\lambda_s^{(m)}(t), \lambda_s^{(m)}(\tau)). \tag{56}$$

Appendix E shows that

$$K_\lambda(t,\tau) = K_\lambda(t-\tau) = \bar{p}\sigma\left(1 - \frac{\sigma}{\bar{p}}\right)e^{-m\,\bar{p}(\sigma^{-1})|t-\tau|}. \tag{57}$$

Thus the power spectral density of this process is

$$S_\lambda(\omega) = \frac{\left(\dfrac{2\sigma^2}{m}\right)\left(1 - \dfrac{\sigma}{\bar{p}}\right)}{1 + \left(\dfrac{\omega\,\sigma}{m\,\bar{p}}\right)^2}. \tag{58}$$

So the bandwidth B of the $\{\lambda_s^{(m)}(t)\}$ process and hence also the optical signalling bandwidth can be taken to be

$$B = \frac{m\bar{p}}{\sigma}. \tag{59}$$

We see from Eqs. (55) and (59) that in order to approach capacity with this sequence of signal processes $\{\lambda_s^{(m)}(t)\}$ given by Eqs. (48) and (51), the optical signalling bandwidth B has to tend to infinity. Hence $C(\bar{s},\bar{p})$ given by Eq. (12) is the channel capacity without bandwidth constraint.

Let us examine the amount of bandwidth of $\{\lambda_s^{(m)}(t)\}$ required for the average mutual information $(1/T)I(S_{mT}; N_T)$ to approach channel capacity $C(\bar{s},\bar{p})$. It follows from Eqs. (12), (D-2), (D-7) and (D-13) that for $0 < \epsilon < 1$ and any $T > 0$,

$$C(\bar{s},\bar{p}) - (1/T)\,I(S_{mT}; N_T) = \epsilon\, C(\bar{s},\bar{p}) \tag{60}$$

implies that the bandwidth B of $\{\lambda_s^{(m)}(t)\}$ satisfies

$$B \geqslant \frac{1}{\epsilon^2}\left[\frac{A^2\,\bar{p}^2\,\sigma}{2\,C^2(\bar{s},\bar{p})}\right] \tag{61}$$

where A is given by Eq. (D-6). Consider a case where $\bar{n} = 0$, $\bar{p} = 10^7$ photons per second, $\bar{s} = 10^4$ photons per second so that $C(\bar{s}, \bar{p}) = 6.9 \times 10^4$ nats per second. Then Eq. (61) becomes

$$B \geqslant \frac{2 \times 10^{10}}{\epsilon^2} \text{ Hz} \tag{62}$$

For $\epsilon = 0.1$, we require $B \geqslant 2 \times 10^{12}$ Hz and for $\epsilon = 0.01$, we require $B \geqslant 2 \times 10^{14}$ Hz.

We now show that for the no background noise case ($\bar{n} = 0$), using coded PPM to achieve capacity is much more bandwidth-efficient than the signalling given by Eqs. (48) and (51). Consider a M-ary PPM modulation scheme shown in Fig. 1 with signal duration T, pulse duration T/M and peak power $\bar{p}$. The M possible signal intensity rate functions $\lambda_{s1}(t), \ldots, \lambda_{sM}(t)$ are given as in Fig. 1. The average power of this signal set is

$$\bar{s} = \frac{\bar{p}}{M}. \tag{63}$$

We shall assume that $M \geqslant 3$. For peak power constraint $\bar{p}$ and average power constraint $\bar{s}$ given by Eq. (63), Corollary 1 gives the channel capacity as

$$C = \bar{s} \log M \text{ nats/sec.} \tag{64}$$

Let us determine the capacity of a system that uses the PPM modulation described in Fig. 1 along with the coding. Since $\bar{n} = 0$, the demodulator then decides that the mth signal was transmitted if a photoelectron is emitted in the mth time slot in $[0, T]$, and declares an error E if no photoelectrons are emitted in the entire interval $[0, T]$. Since the Poisson process has independent increments, one use of the optical direct-detection channel with modulator and demodulator is equivalent to one use of the DMC with input alphabet $\{1, 2, \ldots, M\}$, output alphabet $\{1, 2, \ldots, M, E\}$ and transition probabilities

$$P(j|k) = \begin{cases} 1 - n, & j = k \\ \eta, & j = E, 1 \leqslant k \leqslant M \\ 0, & \text{otherwise} \end{cases} \tag{65}$$

where

$$\eta = e^{-\bar{s}T} \tag{66}$$

is the probability of having no photoelectrons emitted in the time interval $[0, T]$. If we optimally code this DMC, then the capacity of the above coded PPM channel is just the capacity of the DMC given by Eq. (65). Let C_{PPM} denote the capacity of the optimally coded PPM channel. Then it is easy to show (Ref. 11) that

$$C_{PPM} = \frac{(1 - \eta) \log M}{T} \text{ nats/sec}$$

$$= \left(\frac{1 - e^{-\bar{s}T}}{T}\right) \log M \text{ nats/sec.} \tag{67}$$

From Eqs. (64) and (67) we see that

$$\frac{C_{PPM}}{C} = \frac{1 - e^{-\bar{s}T}}{\bar{s}T} \leqslant 1 \tag{68}$$

and that

$$\lim_{T \to 0} \frac{C_{PPM}}{C} = 1.$$

Hence this optimally coded PPM system is capable of achieving channel capacity C in the limit as $T \to 0$. This also entails increasing the signalling bandwidth to infinity because the bandwidth B_{PPM} of the PPM signal set can be taken to be

$$B_{PPM} = \frac{M}{T} = \left(\frac{\bar{p}}{\bar{s}}\right) \frac{1}{T}. \tag{69}$$

Note from Eqs. (68) and (69) that for any $0 < \epsilon < 1$,

$$C - C_{PPM} = \epsilon C \tag{70}$$

implies that the bandwidth B_{PPM} satisfies

$$1 - \frac{1 - \exp(-\bar{p}/B_{PPM})}{(\bar{p}/B_{PPM})} = \epsilon. \tag{71}$$

For small $(\bar{p}/B_{PPM})$, Eq. (71) is approximately

$$B_{PPM} \approx \frac{\bar{p}}{2\epsilon}. \tag{72}$$

For $\epsilon = 0.01$ we would require $B_{PPM} \approx 5 \times 10^8$ Hz when $\bar{p} = 10^7$ photons/sec. This can be compared to the bandwidth $B \geqslant 2 \times 10^{14}$ Hz required by the signalling scheme given by Eqs. (48) and (51) to achieve the same rate. A comparison of Eqs. (72) and (61) shows the relative bandwidth advantage of coded PPM versus the signalling scheme given by Eqs. (48) and (51). This is because B increases at least inversely with ϵ^2 while B_{PPM} increases only inversely with ϵ, where ϵ is the desired proximity to channel capacity. These results apply to the case where $\bar{p}/B_{PPM}$ is small. For large values of $\bar{p}/B_{PPM}$ even less bandwidth is required for PPM modulation to approach capacity.

V. Conclusion

We have derived the capacity of a free-space optical channel using a direct detection receiver under both peak and average signal power constraints and without a signal bandwidth constraint. This result is a generalization of Kabanov's work (Ref. 6), where only a peak power constraint was imposed. In the absence of received background noise, an optimally coded PPM system was shown to achieve channel capacity in the limit as signal bandwidth approaches infinity. All of these results did not consider the effect of a signal bandwidth constraint. It would be interesting to derive the channel capacity under a fixed bandwidth constraint also.

Recent work (Ref. 3) has advocated considering the channel capacity per received signal photon. In the no background noise case ($\bar{n} = 0$) it can be seen from Eq. (13) that

$$\frac{C(\bar{s}, \bar{p})}{\sigma} = \log \frac{\bar{p}}{\sigma} \text{ nats/photon} \tag{73}$$

in the capacity per unit signal photon. It can be easily seen from Eqs. (14) and (73) that for a fixed peak signal power constraint $\bar{p}$, the capacity per unit signal photon increases to infinity as the average signal power constraint $\bar{s}$ approaches zero. However, as $\bar{s} \to 0$, the throughput channel capacity $C(\bar{s}, \bar{p}) \to 0$. Thus it does not appear meaningful to consider capacity per signal photon without fixing the throughput channel capacity. The expression Eq. (12) for $C(\bar{s}, \bar{p})$ can be used in this regard to determine the average signal capacity per unit signal photon for a fixed throughput capacity. This is another problem of interest that has been addressed by Butman, Katz and Lesh (Ref. 20) in the no background noise case.

References

1. Gordon, J. P., "Quantum Effects in Communication Systems," *Proc. Inst. Radio Eng.*, 50, pp. 1898-1908, 1962.

2. Helstrom, C. W., Liu, J. W. S., and Gordon, J. P., "Quantum Mechanical Communication Theory," *Proc. IEEE*, 58, pp. 1578-1598, 1970.

3. Pierce, J. R., "Optical Channels: Practical Limits with Photon Counting," *IEEE Trans. Comm.*, COM-26, No. 12, pp. 1819-1821, 1978.

4. McEliece, R. J., Rubin, A. L., and Rodemich, E. R., "The Practical Limits of Photon Communication," *DSN Progress Report 42-52*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 63-67, 1979.

5. Katz, J., "Information Aspects of Optical Channels," JPL Section 331 seminar.

6. Kabanov, Yu. M., "The Capacity of a Channel of the Poisson Type," *Theory of Probability and Applications*, Vol. 23, No. 1, pp. 143-147, Dec. 1978.

7. Macchi, O., and Picinbono, B. C., "Estimation and Detection of Weak Optical Signals," *IEEE Trans. Inform. Theory*, IT-18, No. 5, pp. 562-573, Sept. 1972.

8. Karp, S. and Clark, J. R., "Photon Counting: A Problem in Classical Noise Theory," *IEEE Trans. Inform. Theory*, IT-16, No. 6, pp. 672-680, Nov. 1970.

9. Rubin, I., "Regular Point Processes and Their Detection," *IEEE Trans. Inform. Theory*, IT-18, No. 5, pp. 547-557, Sept. 1972.

10. Snyder, D. L., *Random Point Processes*, Wiley, New York, 1976.

11. Gallager, R. G., *Information Theory and Reliable Communication*, Wiley, New York, 1968.

12. Lipster, R., and Shiryayev, A. N., *Statistics of Random Processes II*, Springer-Verlag, New York, 1978.

13. Smith, J. G., "The Information Capacity of Amplitude- and Variance-Constrained Scalar Gaussian Channels," *Information and Control*, 18, pp. 203-219, 1971.

14. Segall, A., and Kailath, T., "The Modelling of Randomly Modulated Jump Processes," *IEEE Trans. Inform. Theory*, IT-21, No. 2, pp. 135-143, Mar. 1975.

15. Doob, J., *Stochastic Processes*, Wiley, New York, 1953.

16. Segall, A., Davis, M. H. A., and Kailath, T., "Nonlinear Filtering with Counting Observations," *IEEE Trans. Inform. Theory*, IT-21, No. 2, pp. 143-149, Mar. 1975.

17. Gagliardi, R., and Karp, S., *Optical Communications*, Wiley, New York, 1978.

18. Pierce, J. R., Posner, E. C., and Rodemich, E. R., "The Capacity of the Photon Counting Channel," submitted for publication.

19. Kadota, T. T., Zakai, M., and Ziv, J., "Capacity of a Continuous Memoryless Channel with Feedback," *IEEE Trans. Inform. Theory*, IT-17, No. 4, pp. 372-378, July 1971.

20. Butman, S. A., Katz, J., and Lesh, J. R., "Practical Limitation on Noiseless Optical Channel Capacity," *DSN Progress Report 42-52*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 12-14, 1979.

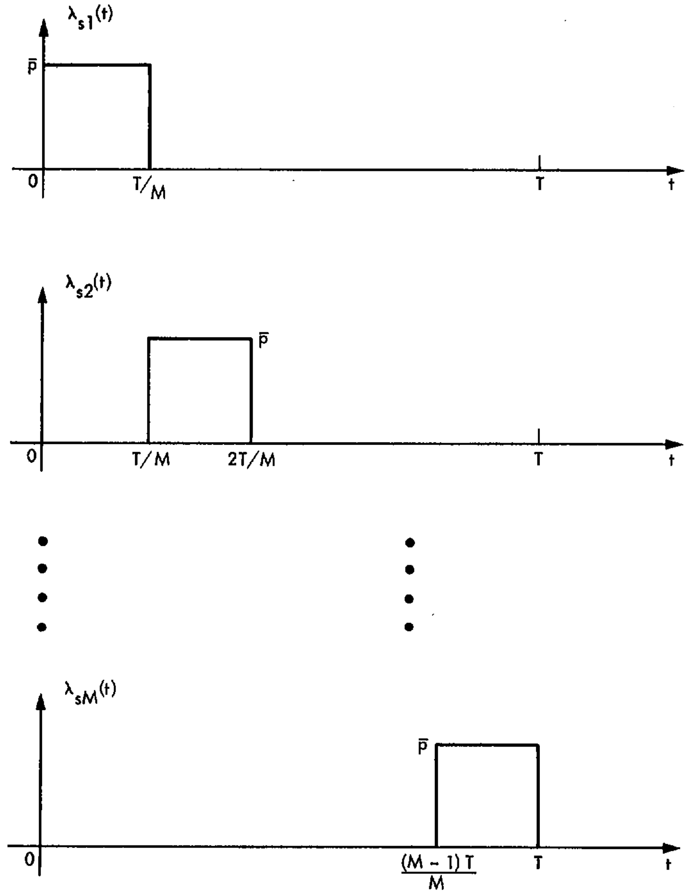

Fig. 1. PPM signal set

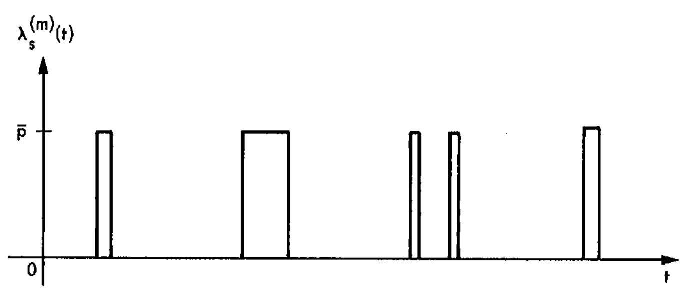

Fig. 2. Typical sample path of $\lambda_s^{(m)}(t)$ when $\alpha << 1$ for Kabanov signals

Appendix A

Derivation of Eq. (19)

Let $p(N_T)$ be the sample function density (Ref. 1) of the compound regular point process $\{N(t): 0 \leqslant t \leqslant T\}$ and $p(N_T|S_T)$ be the conditional sample function density of $\{N(t): 0 \leqslant t \leqslant T\}$ given the message signal process $\{S(t): 0 \leqslant t \leqslant T\}$. Then from Ref. 9 (Theorems 2 and 4) we have[1]

$$p(N_T|S_T) = \exp\left[-\int_0^T (\lambda_s(t) + \bar{n})\, dt + \int_0^T \log(\lambda_s(t) + \bar{n})\, dN(t)\right], \quad \text{(A-1)}$$

$$p(N_T) = \exp\left[-\int_0^T (\hat{\lambda}_s(t) + \bar{n})\, dt + \int_0^T \log(\hat{\lambda}_s(t) + \bar{n})\, dN(t)\right]. \quad \text{(A-2)}$$

Now since

$$I(S_T;N_T) \stackrel{\Delta}{=} E\left[\log \frac{p(N_T|S_T)}{p(N_T)}\right], \quad \text{(A-3)}$$

we have from Eqs. (A-1) and (A-2) that

$$\begin{aligned}
I(S_T;N_T) &= E\left[\int_0^T (\hat{\lambda}_s(t) - \lambda_s(t))\, dt + \int_0^T \log\left(\frac{\lambda_s(t)+\bar{n}}{\hat{\lambda}_s(t)+\bar{n}}\right) dN(t)\right] \\
&\stackrel{①}{=} E\left[\int_0^T \log\left(\frac{\lambda_s(t)+\bar{n}}{\hat{\lambda}_s(t)+\bar{n}}\right) dN(t)\right] \\
&= E\left[\int_0^T (\lambda_s(t)+\bar{n}) \log\left(\frac{\lambda_s(t)+\bar{n}}{\hat{\lambda}_s(t)+\bar{n}}\right) dt\right] \\
&\quad + E\left[\int_0^T \log\left(\frac{\lambda_s(t)+\bar{n}}{\hat{\lambda}_s(t)+\bar{n}}\right) (dN(t) - (\lambda_s(t)+\bar{n})\, dt)\right],
\end{aligned} \quad \text{(A-4)}$$

where ① is because

$$E[\hat{\lambda}_s(t)] = E[E[\lambda_s(t)|N_t]] = E[\lambda_s(t)].$$

Next, since

$$N(t) - \int_0^t (\lambda_s(\tau) + \bar{n})\, d\tau$$

is a martingale (Ref. 14, (3.20)), then from a theorem on stochastic integrals (Ref. 15, p. 437), the second expectation in Eq. (A-4) is zero. Hence Eq. (A-4) can be rewritten as

$$\begin{aligned}
I(S_T;N_T) &= \int_0^T \{E[(\lambda_s(t)+\bar{n})\log(\lambda_s(t)+\bar{n})] - E[(\lambda_s(t)+\bar{n})\log(\hat{\lambda}_s(t)+\bar{n})]\}\, dt \\
&\stackrel{①}{=} \int_0^T \{E[(\lambda_s(t)+\bar{n})\log(\lambda_s(t)+\bar{n})] - E[(\hat{\lambda}_s(t)+\bar{n})\log(\hat{\lambda}_s(t)+\bar{n})]\}\, dt,
\end{aligned} \quad \text{(A-5)}$$

which establishes Eq. (19). In Eq. (A-5) ① is because

$$\begin{aligned}
&E[(\lambda_s(t)+\bar{n})\log(\hat{\lambda}_s(t)+\bar{n})] \\
&= E[E[(\lambda_s(t)+\bar{n})\log(\hat{\lambda}_s(t)+\bar{n})|N_t]] \\
&= E[E[(\lambda_s(t)+\bar{n})|N_t]\log(\hat{\lambda}_s(t)+\bar{n})] \\
&= E[(\hat{\lambda}_s(t)+\bar{n})\log(\hat{\lambda}_s(t)+\bar{n})].
\end{aligned} \quad \text{(A-6)}$$

The above derivation is formal and not rigorous. We have assumed interchanges of integration and expectation without rigorous justifications. A rigorous derivation of Eq. (19) can be found in Ref. 12.

[1]The integrals with respect to $N(t)$ in Eqs. (A-1) and (A-2) and henceforth in the remainder of this paper is interpreted in the Ito sense (Refs. 9, 10).

Appendix B
Proof of Proposition 2

Let $\mathscr{R}(\bar{p}, k)$ be the set of all random variables Λ such that $0 \leqslant \Lambda \leqslant \bar{p}$ and $E[\Lambda] = k$ where $0 \leqslant k \leqslant \bar{p}$. Then

$$\sup_{\Lambda \epsilon \mathscr{R}(\bar{p})} \left\{ E[(\Lambda + \bar{n}) \log (\Lambda + \bar{n})] - [E(\Lambda + \bar{n})] \log [E(\Lambda + \bar{n})] - \left(\frac{\mu}{T}\right) E[\Lambda] \right\}$$

$$= \sup_{0 \leqslant k \leqslant \bar{p}} \left\{ \sup_{\Lambda \epsilon \mathscr{R}(\bar{p}, k)} E[(\Lambda + \bar{n}) \log (\Lambda + \bar{n})] - (k + \bar{n}) \log (k + \bar{n}) + \left(\frac{\mu}{T}\right) k \right\}. \tag{B-1}$$

Note from Fig. B-1 that if $0 \leqslant \Lambda \leqslant \bar{p}$, then the possible values of $E[(\Lambda + \bar{n}) \log (\Lambda + \bar{n})]$ must lie in the set of all y-coordinates of the closed convex hull of the graph of $y = (x + \bar{n}) \log (x + \bar{n})$ for $0 \leqslant x \leqslant \bar{p}$. Hence the largest possible values lie on the cord AB. These values can be achieved using a random variable Λ with the following distribution

$$P(\Lambda = \bar{p}) = 1 - P(\Lambda = 0) = \alpha \tag{B-2}$$

where $\alpha \epsilon [0, 1]$ must be chosen so that $E[\Lambda] = k$ in order for Λ to be in $\mathscr{R}(\bar{p}, k)$. In order for $E[\Lambda] = k$, we must have

$$\alpha = k/\bar{p}\ . \tag{B-3}$$

Hence

$$\sup_{\Lambda \epsilon \mathscr{R}(\bar{p}, k)} E[(\Lambda + \bar{n}) \log (\Lambda + \bar{n})] = \left(\frac{k}{\bar{p}}\right) (\bar{p} + \bar{n}) \log (\bar{p} + \bar{n}) + \left(1 - \frac{k}{\bar{p}}\right) \bar{n} \log \bar{n}, \tag{B-4}$$

where the Λ achieving the supremum is given by Eqs. (B-2) and (B-3). Hence the optimization problem in Eq. (B-1) can be written as

$$\sup_{0 \leqslant k \leqslant \bar{p}} g(k), \tag{B-5}$$

where $g(k)$ is given by

$$g(k) = \left(\frac{k}{\bar{p}}\right)(\bar{p} + \bar{n}) \log (\bar{p} + \bar{n}) + \left(1 - \frac{k}{\bar{p}}\right) \bar{n} \log \bar{n} - (k + \bar{n}) \log (k + \bar{n}) - \left(\frac{\mu}{T}\right) k\ . \tag{B-6}$$

Since $g(k)$ is concave in k, the supremum in Eq. (B-5) is achieved by a $k \in [0, \bar{p}]$ such that $g'(k) = 0$, provided that such a k exists. Setting $g'(k) = 0$ gets

$$k = (\bar{p} + \bar{n}) \exp \left(- \left(\frac{\mu}{T} + 1\right) + \frac{\bar{n}}{p} \log \left(1 + \frac{\bar{p}}{\bar{n}}\right)\right) - \bar{n}\ . \tag{B-7}$$

It follows immediately from Eq. (16) since $\mu \geqslant 0$ that $k \leqslant \bar{p}$ in Eq. (B-7). Since by hypothesis of the proposition, $k \geqslant 0$, we can conclude that k given by Eq. (B-7) achieves the supremum in Eq. (B-5). This establishes the proposition.

Also note from Eq. (B-6) that when $\mu = 0$, $g(0) = 0$ and $g'(0) \geqslant 0$. This means that when $\mu = 0$, the solution k to $g'(k) = 0$ must be nonnegative. So when $\mu = 0$, the k given by Eq. (B-7) is nonnegative.

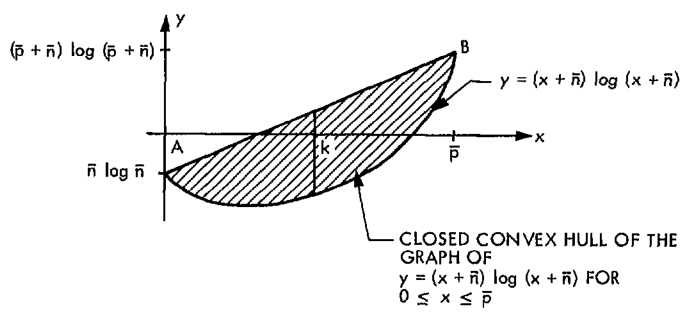

Fig. B-1. Geometry of optimization

Appendix C
Proof that $E[S_m(t)] = \alpha$

Since $M(t)$ is nonnegative integer-valued, $(-1)^{M(t)} = \cos(\pi M(t))$. So Eq. (48) can be rewritten as

$$S_m(t) = \left(\frac{1}{2}\right) + \left[1_{\{1\}}(S_m(0)) - \frac{1}{2}\right]\cos(\pi M(t)), \quad \text{(C-1)}$$

for $t > 0$. So for each t, $S_m(t)$ is a function of $M(t)$ and hence we may use the Stochastic Differential Rule (e.g., Ref. 10, Theorem 4.2.2) to obtain

$$\begin{aligned}
d\,S_m(t) &= \left\{\left(\frac{1}{2}\right) + \left[1_{\{1\}}(S_m(0)) - \frac{1}{2}\right]\cos(\pi M(t) + 1)\right. \\
&\quad \left. - \left(\left(\frac{1}{2}\right) + \left[1_{\{1\}}(S_m(0)) - \frac{1}{2}\right]\cos(\pi M(t))\right)\right\} dM(t) \\
&\overset{①}{=} [1 - 2S_m(t)]\; dM(t) \\
&= [1 - 2S_m(t)]\; \nu(t)\,dt \\
&\quad + [1 - 2S_m(t)]\; [dM(t) - \nu(t)\,dt] \\
&\overset{②}{=} [1 - 2S_m(t)]\left[m\,1_{\{0\}}(S_m(t^-))\right. \\
&\quad \left. + m\left(\frac{1-\alpha}{\alpha}\right)1_{\{1\}}(S_m(t^-))\right] dt \\
&\quad + [1 - 2S_m(t)]\; [dM(t) - \nu(t)\,dt], \qquad \text{(C-2)}
\end{aligned}$$

where ① is obtained using Eq. (C-1) and ② is because of Eq. (50). Rewriting Eq. (C-2) in integral form gets

$$\begin{aligned}
S_m(t) &= S_m(0) + \int_0^t [1 - 2S_m(u)]\; [m\,1_{\{0\}}(S_m(u)) \\
&\quad + m\left(\frac{1-\alpha}{\alpha}\right)1_{\{1\}}(S_m(u))]\; du \\
&\quad + \int_0^t [1 - 2S_m(u)]\; [dM(u) - \nu(u)\,du] \\
&\overset{①}{=} S_m(0) + \int_0^t m\alpha^{-1}(\alpha - S_m(u))\,du \\
&\quad + \int_0^t [1 - 2S_m(u)]\; [dM(u) - \nu(u)\,du], \qquad \text{(C-3)}
\end{aligned}$$

where ① follows since $S_m(t)$ is either 0 or 1. Now it follows from Eq. (49) that $E[S_m(0)] = \alpha$. Next, since

$$M(t) - \int_0^t \nu(u)\,du$$

is a martingale (Ref. 14, (3.20)), it follows that the expected value of the last integral in Eq. (C-3) iz zero (Ref. 15, p. 437). So, taking expected values in Eq. (C-3) gets

$$E[S_m(t)] = \alpha + m\alpha^{-1}\int_0^t [\alpha - E[S_m(u)]]\; du. \quad \text{(C-4)}$$

The unique solution of this initial value problem Eq. (C-4) is

$$E[S_m(t)] = \alpha. \quad \text{(C-5)}$$

Appendix D
Derivation of Eq. (55)

Note that since $S_m(t) = 0$ or 1, it follows from Eq. (C-5) that $P(S_m(t) = 1) = \alpha$. So from Eqs. (49) and (51) it follows that

$$E[(\lambda_s^{(m)}(t) + \bar{n}) \log (\lambda_s^{(m)}(t) + \bar{n})]$$

$$= \left(\frac{\sigma}{\bar{p}}\right)(\bar{p} + \bar{n}) \log (\bar{p} + \bar{n}) + \left(1 - \frac{\sigma}{\bar{p}}\right)\bar{n} \log \bar{n} . \quad \text{(D-1)}$$

So from Eqs. (19) and (D-1),

$$\frac{1}{T} I (S_{mT}; N_T) - \left(1 - \frac{\sigma}{\bar{p}}\right)\bar{n} \log \bar{n} - \left(\frac{\sigma}{\bar{p}}\right)(\bar{p} + \bar{n}) \log (\bar{p} + \bar{n})$$

$$+ (\sigma + \bar{n}) \log (\sigma + \bar{n})$$

$$= \frac{1}{T}\int_0^T E\{f(\hat{\lambda}_s^{(m)}(t)) - f(\sigma)\}\, dt \quad \text{(D-2)}$$

where

$$f(x) = (x + \bar{n}) \log (x + \bar{n}), \quad \text{(D-3)}$$

and

$$\hat{\lambda}_s^{(m)}(t) \stackrel{\Delta}{=} E[\lambda_s^{(m)}(t) | N_t)$$

$$= \bar{p}\, E[S_m(t) | N_t]$$

$$\stackrel{\Delta}{=} \bar{p}\, \hat{S}_m(t) . \quad \text{(D-4)}$$

Now it can be easily shown that for $0 \leqslant x \leqslant \bar{p}$,

$$|f(x) - f(\sigma)| \leqslant A\, |x - \sigma| , \quad \text{(D-5)}$$

where

$$A = \max \left[\frac{f(\bar{p}) - f(\sigma)}{\bar{p} - \sigma} , \left|\frac{f(0) - f(\sigma)}{\sigma}\right|\right] , \quad \text{(D-6)}$$

since $0 \leqslant \sigma \leqslant \bar{p}$. So from Eqs. (D-4), (D-5) and (49), we have

$$\frac{1}{T}\int_0^T E\{f(\hat{\lambda}_s^{(m)}(t)) - f(\sigma)\}\, dt$$

$$\leqslant \frac{A\bar{p}}{T}\int_0^T E\{|\hat{S}_m(t) - \alpha|\}\, dt$$

$$\stackrel{①}{\leqslant} \frac{A\bar{p}}{T}\int_0^T (E[(\hat{S}_m(t) - \alpha)^2])^{1/2}\, dt$$

$$\stackrel{②}{=} \frac{A\bar{p}}{T}\int_0^T (E[(\hat{S}_m(t))^2] - \alpha^2)^{1/2}\, dt. \quad \text{(D-7)}$$

Here ① is from Jensen's inequality and ② is because

$$E[\hat{S}_m(t)] = E[E[S_m(t) | N_t]] = E[S_m(t)] = \alpha$$

from Eq. (C-5). The remainder of the derivation now follows Kabanov's proof exactly to show that $E[\hat{S}_m(t))^2] - \alpha^2$ converges to zero as $m \to \infty$ uniformly in t. Specifically, from Eq. (C-3) and (Ref. 16, Eq, 1.6a) we can write

$$d\hat{S}_m(t) = m\alpha^{-1} (\alpha - \hat{S}_m(t))\, dt$$

$$+ \frac{\bar{p}E[S_m(t)\, (S_m(t) - \hat{S}_m(t)) | N_t]}{\bar{p}\,\hat{S}_m(t) + \bar{n}}$$

$$\cdot\ [dN(t) - (\bar{p}\,\hat{S}_m(t) + \bar{n})\, dt]$$

$$\stackrel{①}{=} m\alpha^{-1} (\alpha - \hat{S}_m(t))\, dt$$

$$+ \frac{\bar{p}\,\hat{S}_m(t)\, (1 - \hat{S}_m(t))}{\bar{p}\,\hat{S}_m(t) + \bar{n}}$$

$$\cdot\ [dN(t) - (\bar{p}\,\hat{S}_m(t) + \bar{n})\, dt] , \quad \text{(D-8)}$$

where ① is because $S_m(t) = S_m^2(t)$. Next, using the Stochastic Differential Rule (Ref. 10, Theorem 4.2.2), we obtain

$$d(\hat{S}_m(t))^2 = \left[2m\,\alpha^{-1}\,\hat{S}_m(t)\,(\alpha - \hat{S}_m(t)) + \frac{[\bar{p}\,\hat{S}_m(t)\,(1-\hat{S}_m(t))]^2}{\bar{p}\,\hat{S}_m(t)+\bar{n}}\right]dt$$

$$+ \left[\frac{2\bar{p}\,(\hat{S}_m(t))^2\,(1-\hat{S}_m(t))}{\bar{p}\,\hat{S}_m(t)+\bar{n}} + \left(\frac{\bar{p}\,\hat{S}_m(t)\,(1-\hat{S}_m(t))}{\bar{p}\,\hat{S}_m(t)+\bar{n}}\right)^2\right]$$

$$\cdot\,[dN(t) - (\bar{p}\,\hat{S}_m(t)+\bar{n})\,dt]. \quad \text{(D-9)}$$

Now using the martingale property of

$$N(t) - \int_0^t (\bar{p}\,\hat{S}_m(u)+\bar{n})\,du$$

as in Eqs. (A-4), (A-5) and in (C-4), we can take the expected value of the integral of Eq. (D-9) to obtain

$$E[(\hat{S}_m(t))^2] = \alpha^2 + 2m\alpha^{-1}\int_0^t \{\alpha^2 - E[\hat{S}_m(u)]\}\,du$$

$$+ \int_0^t E\left[\frac{[\bar{p}\,\hat{S}_m(u)\,(1-\hat{S}_m(u))]^2}{\bar{p}\,\hat{S}_m(u)+\bar{n}}\right]du. \quad \text{(D-10)}$$

Define

$$g(t) = E\left[\frac{[\bar{p}\,\hat{S}_m(t)\,(1-\hat{S}_m(t))]^2}{\bar{p}\,\hat{S}_m(t)+\bar{n}}\right]$$

$$\leqslant \bar{p}\,E[\hat{S}_m(t)\,(1-\hat{S}_m(t))^2]$$

$$\leqslant \bar{p}\,E[\hat{S}_m(t)]$$

$$= \sigma. \quad \text{(D-11)}$$

Now since $(\hat{S}_m(0))^2 = (E[S_m(0)])^2 = \alpha^2$, the unique solution to the initial value problem Eq. (D-10) is

$$E[(\hat{S}_m(t))^2] = \alpha^2\,e^{-2m\alpha^{-1}t}$$

$$+ e^{-2m\alpha^{-1}t}\int_0^t (g(u) + 2m\alpha)$$

$$\cdot e^{2m\alpha^{-1}u}\,du. \quad \text{(D-12)}$$

Since from Eq. (D-11), $0 \leqslant g(u) \leqslant \sigma$, we have from Eq. (D-12) that

$$E[(\hat{S}_m(t))^2] - \alpha^2$$

$$\leqslant \frac{\sigma^2}{2m\bar{p}}\,(1 - e^{-2m\alpha^{-1}t})$$

$$\leqslant \frac{\sigma^2}{2m\bar{p}}. \quad \text{(D-13)}$$

Therefore Eqs. (D-2), (D-7) and (D-13) establish Eq. (55).

Appendix E
Derivation of Eq. (57)

It follows from Eqs. (51) and (C-5) that

$$K_\lambda(t,\tau) = \bar{p}^2 \{E[S_m(t)\,S_m(\tau)] - \alpha^2\}. \qquad \text{(E-1)}$$

Since $P(S_m(t) = 1) = \alpha$ from Eq. (C-5), it follows that for $t \geqslant \tau$,

$$E[S_m(t)\,S_m(\tau)] = E[S_m(\tau)\,E[S_m(t)|S_m(\tau)]]$$

$$= \alpha\,E[S_m(t)|S_m(\tau) = 1]. \qquad \text{(E-2)}$$

Then we may write, using Eq. (C-3), for $t \geqslant \tau$

$$S_m(t) = S_m(\tau) + m\alpha^{-1}\int_\tau^t (\alpha - S_m(u))\,du$$

$$+ \int_\tau^t (1 - 2\,S_m(u))\,[dM(u) - \nu(u)\,du]. \qquad \text{(E-3)}$$

Hence for $t \geqslant \tau$,

$$E[S_m(t)|S_m(\tau) = 1]$$

$$= 1 + m\alpha^{-1}\int_\tau^t (\alpha - E[S_m(u)|S_m(\tau) = 1])\,du\,, \qquad \text{(E-4)}$$

because

$$E\left\{\int_\tau^t (1 - 2\,S_m(u))[dM(u - \nu(u)\,du]\,|S_m(\tau) = 1\right\}$$

$$= E\left\{\int_\tau^t (1 - 2\,S_m(u))[d(M(u) - M(\tau)) - \nu(u)\,du]\,|S_m(\tau) = 1\right\} \overset{①}{=} 0, \qquad \text{(E-5)}$$

where ① follows since

$$\left\{M(s) - M(\tau) - \int_\tau^s \nu(u)\;du : s \geqslant \tau\right\}$$

is a martingale given that $S_m(\tau) = 1$.

Since $E[S_m(\tau)|S_m(\tau) = 1] = 1$, the solution of the initial value problem Eq. (E-4) can be shown to be

$$E[S_m(t)|S_m(\tau) = 1] = (1 - \alpha)\;e^{-m\alpha^{-1}(t-\tau)} + \alpha, \qquad \text{(E-6)}$$

when $t \geqslant \tau$. Thus Eqs. (E-1), (E-2) and (E-6) gets

$$K_\lambda(t,\tau) = p^{-2}\,\alpha(1 - \alpha)\,e^{-m\alpha^{-1}(t-\tau)} \qquad \text{(E-7)}$$

for $t \geqslant \tau$. Since $K_\lambda(t,\tau) = K_\lambda(\tau,t)$ and since $\alpha = \sigma/p$, Eq. (57) follows from Eq. (E-7).