Capacity and Error Exponent for the Direct Detection Photon Channel—Part I

AARON D. WYNER, FELLOW, IEEE

Abstract—The capacity and error exponent of the direct detection optical channel are considered. The channel input in a T-second interval is a waveform $\lambda(t)$, $0 \le t \le T$, which satisfies $0 \le \lambda(t) \le A$, and $(1/T)\int_0^T \lambda(t)\,dt \le \sigma A$, $0 < \sigma \le 1$. The channel output is a Poisson process with intensity parameter $\lambda(t) + \lambda_0$. The quantities A and σA represent the peak and average power, respectively, of the optical signal, and λ_0 represents the "dark current." In Part I the channel capacity of this channel and a lower bound on the error exponent are calculated. An explicit construction for an exponentially optimum family of codes is also exhibited. In Part II we obtain an upper bound on the error exponent which coincides with the lower bound. Thus this channel is one of the very few for which the error exponent is known exactly.

Dedication

These papers are dedicated to the memory of Stephen O. Rice, an extraordinary mentor, supervisor, and friend. He was a master of numerical methods and asymptotics and was very much at home with the nineteenth-century menagerie of special functions. The generous and easy way in which he shared his genius with his colleagues is legendary, and I was fortunate to have been a beneficiary of his advice and expertise during the first decade of my career at Bell Laboratories. As did all of Steve's colleagues, I learned much from this gentle and talented man. We will remember him always.

I. Introduction

THIS IS THE first of a two-part series on the capacity and error exponent of the direct-detection optical channel. Specifically, in the model we consider, information modulates an optical signal for transmission over the channel, and the receiver is able to determine the arrival time of the individual photons which occur with a Poisson distribution. Systems based on this channel have been discussed widely in the literature [1]–[5] and are of importance in applications.

The channel capacity of our channel was found by Kabanov [3] and Davis [2] using martingale techniques. In the present paper we obtain their capacity formula using an elementary and intuitively appealing method. We also obtain a "random coding" exponential upper bound on the probability of error for transmission at rates less than capacity. In Part II [8], we obtain a lower bound on the error probability which has the same asymptotic exponential behavior (as the delay becomes large with the transmission rate held fixed) as the upper bound. Thus this channel joins the infinite bandwidth additive Gaussian noise channel as the only channel for which the "error exponent" is known exactly for all rates below capacity. In Section IV of the present paper we also give an explicit construction of a family of codes for use on our channel, the error probability of which has the optimal exponent. Here too our channel and the infinite bandwidth additive Gaussian noise channel are the only two channels for which an explicit construction of exponentially optimal codes is known.

Manuscript received March 10, 1988. This work was originally presented at the IEEE International Symposium on Information Theory, Brighton, England, June 1985.

The author is with AT&T Bell Laboratories, Murray Hill, NJ 07974.

IEEE Log Number 8824874.

Precise Statement of the Problem and Results

The channel input is a waveform $\lambda(t)$, $0 \le t < \infty$, which satisfies

$$0 \le \lambda(t) \le A, \tag{1.1}$$

where the parameter A is the peak power. The waveform $\lambda(\cdot)$ defines a Poisson counting process $\nu(t)$ with "intensity" or ("rate") equal to $\lambda(t) + \lambda_0$, where $\lambda_0 \ge 0$ is a background noise level (sometimes called "dark current"). Thus the process $\nu(t)$, $0 \le t < \infty$, is the independent-increments process such that

$$\nu(0) = 0, \tag{1.2a}$$

and, for $0 \le \tau,\ t < \infty$,

$$\Pr\{\nu(t+\tau) - \nu(t) = j\} = \frac{e^{-\Lambda}\Lambda^j}{j!}, \qquad j = 0,1,2,\cdots \tag{1.2b}$$

where

$$\Lambda = \int_t^{t+\tau} (\lambda(t') + \lambda_0)\,dt'. \tag{1.2c}$$

Physically, we think of the jumps in $\nu(\cdot)$ as corresponding to photon arrivals at the receiver. We assume that the receiver has knowledge of $\nu(t)$, which it would obtain using a photon-detector.

For any function $g(t)$, $0 \le t < \infty$, let g_a^b denote $\{g(t):\ a \le t \le b\}$. Let $S(\tau)$ denote the space of (step) functions $g(t)$, $0 \le t \le \tau$, such that $g(0) = 0$, $g(t) \in \{0,1,2,\cdots\}$, $g(t)\uparrow$. Therefore, ν_0^T, the Poisson counting process defined above, takes values in $S(T)$.

A code with parameters (M, T, σ, P_e) is defined by the following:

a) a set of M waveforms $\lambda_m(t)$, $0 \le t \le T$, which satisfy the "peak power constraint" (1.1) and the "average power constraint"

$$\frac{1}{T}\int_0^T \lambda_m(t)\,dt \le \sigma A \tag{1.3}$$

(of course, $0 \le \sigma \le 1$);

b) a "decoder" mapping D: $S(T) \to \{1, 2, \cdots, M\}$.

The overall error probability is

$$P_e = \frac{1}{M}\sum_{m=1}^{M} \Pr\{D(\nu_0^T) \ne m \mid \lambda_m(\cdot)\}, \tag{1.4}$$

where the conditional probabilities in (1.4) are computed using (1.2) with $\lambda(\cdot) = \lambda_m(\cdot)$.

A code as defined above can be used in a communication system in the usual way to transmit one of M messages. Thus when $\lambda_m(t)$ corresponding to message m, $1 \le m \le M$, is transmitted, the waveform $\nu(t)$, $0 \le t \le T$, is received, and is decoded as $D(\nu_0^T)$. Equation (1.4) gives the "word error probability," the probability that $D(\nu_0^T) \ne m$ when message m is transmitted, averaged over the M messages (which are assumed to be equally likely). The rate of the code (in nats per second) is $(1/T)\ln M$.

Let A, λ_0, σ be given. A rate $R \ge 0$ is said to be *achievable* if, for all $\epsilon > 0$, there exists (for T sufficiently large) a code with parameters (M, T, σ, P_e) with $M \ge e^{RT}$ and $P_e \le \epsilon$. The channel capacity C is the supremum of achievable rates. In Section II we establish the following theorem, which was found earlier by Kabanov [3] and Davis [2] using less elementary methods.

Theorem 1: For $A, \lambda_0, \sigma \ge 0$,

$$C = A[q^*(1+s)\ln(1+s)) + (1-q^*)s\ln s - (q^*+s)\ln(q^*+s)] \tag{1.5a}$$

where

$$s = \lambda_0/A, \tag{1.5b}$$

$$q^* = \min(\sigma, q_0(s)), \tag{1.5c}$$

and

$$q_0(s) = \frac{(1+s)^{1+s}}{s^s e} - s. \tag{1.5d}$$

For the interesting case where $s = \lambda_0 = 0$ (no dark current), (1.5) yields

$$C = Aq^*\ln\frac{1}{q^*}, \tag{1.6}$$

where $q^* = \min(\sigma, e^{-1})$. Further, we show in Appendix I (Proposition A.3) that when $s \to \infty$ (i.e., high noise), $q_0(s) = (1/2) + O(1/s)$, and the capacity is

$$C = \frac{Aq^*(1-q^*)}{2s} + O\left(\frac{1}{s^2}\right), \tag{1.7a}$$

$$q^* = \min(\sigma, 1/2). \tag{1.7b}$$

Equation (1.7) was also obtained by Davis [2].

The quantity

$$q_0(s) \triangleq \frac{(1+s)^{1+s}}{s^s e} - s$$

turns out to be the optimum ratio of signal energy ($\int\lambda_m(t)\,dt$) to AT (the maximum allowable signal energy) to achieve the maximum transmission rate. Should $q_0(s) \le \sigma$, then code signals $\lambda_m(t)$ which satisfy $\int\lambda_m(t)\,dt = q_0(s)AT$ will satisfy constraint (1.3). Should $q(s) > \sigma$, then we chose signals for which $\int\lambda_m(t)\,dt = \sigma AT$. Thus for codes which achieve capacity, the average number of received photons per second is q^*AT.

Next, let A, λ_0, σ be given. Define $P_e^*(M, T)$ as the infimum of those P_e for which a code with parameters (M, T, σ, P_e) is achievable. For $0 \le R < C$, define the optimal error-exponent by

$$E(R) = \limsup_{T\to\infty} \frac{1}{T}\ln P_e^*\lfloor e^{RT}\rfloor, T). \tag{1.8}$$

Thus we can write $P_e^*(\lfloor e^{RT}\rfloor, T) = \exp\{-E(R)T + o(T)\}$ for large T. In Section III, we establish an upper bound on P_e^* using "random code" techniques which yields a lower bound on $E(R)$. In Part II [8] we establish an upper bound on $E(R)$ which agrees with the lower bound for all R, $0 \le R < C$. Finally, in Section IV of the present paper we give an explicit construction for a family of codes with parameter $M \ge e^{RT}$, such that

$$P_e = \exp\{-E(R)T + o(T)\}, \qquad \text{as } T \to \infty.$$

Thus this family of codes is essentially optimal.

We now give the formula for the optimal error exponent $E(R)$:

$$E(R) = \max[AE_1(\rho, q) - \rho R], \tag{1.9a}$$

where the maximization is over $\rho \in [0,1]$, and $q \in [0, \sigma]$. $E_1(\cdot)$ is defined by

$$E_1(\rho, q) = (q+s) - s[1+\tau q]^{1+\rho}, \tag{1.9b}$$

where $s = \lambda_0/A$ and

$$\tau = \left(1 + \frac{1}{s}\right)^{1/(1+\rho)} - 1. \tag{1.9c}$$

The maximization in (1.9a) with respect to ρ and q is done in Section III. As a result of this maximization, we obtain a convenient way to represent $E(R)$ in which we express R and $E(R)$ parametrically in the variable $\rho \in [0,1]$. Thus set

$$R^*(\rho) = As\left[\frac{(1+q^*\tau)^\rho}{(1+\rho)} q^*\left(1+\frac{1}{s}\right)^{1/1+\rho}\ln\left(1+\frac{1}{s}\right) - (1+q^*\tau)^{1+\rho}\ln(1+q^*\tau)\right], \qquad 0 < \rho \le 1 \tag{1.10a}$$

where $s = \lambda_0/A$ and $\tau = \tau(\rho)$ is given by (1.9c), and

$$q^* = q^*(\rho) = \min\left(\sigma, \frac{1}{\tau}\left\{\left[\frac{1}{s(1+\rho)\tau}\right]^{1/\rho} - 1\right\}\right). \tag{1.10b}$$

We show in Appendix I (Proposition A.2) that

$$\lim_{\rho \to 0} R^*(\rho) = C, \tag{1.11}$$

the channel capacity given by (1.5). Furthermore, as ρ increases from zero to one, $R^*(\rho)$ strictly decreases from C to $R^*(1) > 0$. Thus for $R^*(1) \le R < C$, there is a unique ρ, $0 < \rho \le 1$, such that $R = R^*(\rho)$. The error exponent can be written, for $R = R^*(\rho) \in [R^*(1), C]$, as

$$E(R) = AE_1(\rho, q^*(\rho)) - \rho R. \tag{1.12}$$

As in the expression for channel capacity,

$$q_1(\rho) \triangleq \frac{1}{\tau(\rho)} \left\{ \left[\frac{1}{s(1+\rho)\tau(\rho)} \right]^{1/\rho} - 1 \right\}, \qquad 0 < \rho \le 1 \tag{1.13}$$

is the optimum ratio of average signal energy $\int_0^T \lambda_m(t)\, dt$ to AT. It can be shown that for $0 < \rho \le 1$, $q_1(\rho) \le 1/2$ and $q_1(1) = 1/2$. Furthermore, we show in Appendix I (Proposition A.1) that, as $\rho \to 0$,

$$\lim q_1(\rho) = q_0(s),$$

where $q_0(s)$ is given by (1.5d).

The quantity $R^*(1)$ is sometimes called the "critical rate." For $0 \le R \le R^*(1)$, the expression for $E(R)$ is

$$E(R) = R_0 - R, \tag{1.14a}$$

where

$$R_0 = Aq^*(1-q^*)\left(\sqrt{1+s} - \sqrt{s}\right)^2 \tag{1.14b}$$

and

$$q^* = \min(\sigma, 1/2). \tag{1.14c}$$

Thus for $0 \le R \le R^*(1)$, $E(R)$ is a straight line with slope -1. The quantity R_0, called the "cutoff rate" is, by (1.14a), equal to $E(0)$.

We can get a better idea of the form of the $E(R)$ curve by looking at the special cases $s \to 0$, $s \to \infty$. When $s = 0$ (no dark current), we show in Appendix I (Proposition A.4) that

$$\tau(\rho) \approx s^{-(1/(1+\rho))}, \qquad q^*(\rho) = \min\left(\sigma, \frac{1}{(1+\rho)^{1/\rho}}\right), \tag{1.15a}$$

and

$$R^*(\rho) = -A[q^*(\rho)]^{1+\rho} \ln(q^*(\rho)), \tag{1.15b}$$

and with $R = R^*(\rho)$,

$$E(R) = Aq^* - Aq^{*1+\rho} - \rho R. \tag{1.15c}$$

Thus the critical rate is

$$R(1) = Aq^{*2} \ln 2$$

where $q^* = q^*(1) = \min(\sigma, 1/2)$. The cutoff rate $R_0 = Aq^*(1-q^*)$, so that for $0 \le R \le R^*(1)$,

$$E(R) = Aq^*(1-q^*) - R.$$

For the high-noise case, $s \to \infty$, the capacity C is given by (1.7). We show in Appendix I that the error exponent has the form

$$E(R) \sim \begin{cases} \dfrac{C}{2} - R, & 0 \le R \le C/4 \\ \left(\sqrt{C} - \sqrt{R}\right)^2, & C/4 \le R \le C \end{cases}. \tag{1.16}$$

This is identical to the error exponent for the so-called "very noisy channel."

II. Direct (Existence) Theorems I—Channel Capacity

In this section we make an *ad hoc* assumption on the structure of the channel input signal $\lambda(t)$ and the receiver. Under this assumption, we compute lower bounds on the channel capacity C and the error exponent $E(R)$. In the companion paper [8] we show that this *ad hoc* assumption degrades performance in a negligible way and that the bounds obtained here on C and $E(R)$ are, in fact, tight. Here are the assumptions. Let $\Delta > 0$ be given. Then assume the following.

a) The channel input waveform $\lambda(t)$ is constant for $(n-1)\Delta < t \le n\Delta$, $n = 1, 2, 3, \cdots$, and $\lambda(t)$ takes only the values 0 or A. For $n = 1, 2, \cdots$, let $x_n = 0$ or 1 according as $\lambda(t) = 0$ or A in the interval $((n-1)\Delta, n\Delta]$.

b) The receiver observes only the samples $\nu(n\Delta)$, $n = 1, 2, \cdots$, or alternatively the increments

$$\hat{y}_n = \nu(n\Delta) - \nu((n-1)\Delta). \tag{2.1}$$

(Recall that $\nu(0) = 0$.)

c) Further, the receiver interprets $\hat{y}_n \ge 2$ (a rare event when Δ is small) as being the same as $\hat{y}_n = 0$. Thus the receiver has available

$$y_n \triangleq \begin{cases} 1, & \hat{y}_n = 1 \\ 0, & \hat{y}_n \ne 1 \end{cases}. \tag{2.2}$$

Subject to assumptions a, b, c, the channel reduces to a two-input two-output discrete memoryless channel (DMC) with transition probability $W(j|k) = \Pr\{y_n = j | x_n = k\}$ given by

$$\begin{aligned} W(1|0) &= \lambda_0 \Delta e^{-\lambda_0 \Delta} = sA\Delta e^{-sA\Delta}, \\ W(1|1) &= (A + \lambda_0)\, \Delta e^{-(A + \lambda_0 \Delta)} \\ &= (1+s) A\, \Delta e^{-(1+s)A\Delta}. \end{aligned} \tag{2.3}$$

We now apply the standard formulas for channel capacity and random coding error-exponent to find lower bounds on C and $E(R)$. For $\Delta > 0$ given, let $T = N\Delta$. We will hold Δ fixed and let $N \to \infty$. The average power constraint (1.3) is equivalent to

$$\frac{1}{N} \sum_{n=1}^{N} x_{mn} \le \sigma, \tag{2.4}$$

where $(x_{m1}, x_{m2}, \cdots, x_{mN})$ corresponds to $\lambda_m(\cdot)$ as in assumption a). Thus the lower bound on C is $\max I(X; Y)/\Delta$ nats per second, where X, Y are binary random variables connected by the channel $W(\cdot|\cdot)$, and the maximum is taken with respect to all input distributions which satisfy

$$E(X) \le \sigma. \tag{2.5}$$

Now set

$$q=\Pr\{X=1\},\qquad a=sA\Delta e^{-sA\Delta},$$
$$b=(1+s)A\Delta e^{-(1+s)A\Delta}. \tag{2.6}$$

We have

$$\begin{aligned}I(X;Y)&=H(Y)-H(Y|X)\\&=h(qb+(1-q)a)-qh(b)-(1-q)h(a)\\&\triangleq f(q)\end{aligned} \tag{2.7}$$

where $(h(u)=-u\ln u-(1-u)\ln(1-u)$, $0\le u\le 1$, is the binary entropy function. Thus

$$\Delta\cdot C\ge \max_{EX\le\sigma} I(X;Y)=\max_{0\le q\le\sigma} f(q). \tag{2.8}$$

So far, the parameter Δ has been arbitrary. We now assume that Δ is very small and estimate $f(q)$. Using $h(u)=-u\ln u+u+O(u^2)$, as $u\to 0$, and $qb+(1-q)a\sim\Delta A(q+s)$, we have

$$\begin{aligned}f(q)&=h(qb+(1-q)a)-qh(b)-(1-q)h(a)\\&=\Delta A[-(q+s)\ln(q+s)+q(1+s)\ln(1+s)\\&\quad+(1-q)s\ln s]+o(\Delta).\end{aligned}$$

Thus (2.8) yields, as $\Delta\to 0$,

$$C\ge\sup_{0\le q\le\sigma} A[-(q+s)\ln(q+s)+q(1+s)\ln(1+s)+(1-q)s\ln s]. \tag{2.9}$$

Since the term in brackets in (2.9) is concave in q, its unconstrained maximum with respect to q occurs when its derivative is equal to zero. This occurs when

$$\ln(q+s)=(1+s)\ln(1+s)-s\ln s-1, \tag{2.10a}$$

or

$$q=\frac{(1+s)^{1+s}}{s^s}-s\triangleq q_0(s). \tag{2.10b}$$

From (2.9) and (2.10) we conclude that $\max_{0\le q\le\sigma} f(q)$ is achieved for $q=q_0(s)$, provided $\sigma\ge q_0(s)$. When $\sigma\le q_0(s)$ (from the concavity of the term in brackets in (2.9)) this maximum is achieved with $q=\sigma$. Thus we have

$$C\ge A[-(q^*+s)\ln(q^*+s)+q^*(1+s)\ln(1+s)+(1-q^*)s\ln s] \tag{2.11a}$$

where

$$q^*=\min(\sigma,q_0(s)). \tag{2.11b}$$

Let us remind the reader at this point that (2.11) is a *lower bound* on C because we have not as yet shown that we can make assumptions a, b, c with negligible loss in performance. We will do this in Part II [8]. In the next section, we turn to the random code error-exponent.

III. Direct Theorems II—Error Exponent

For an arbitrary DMC with input constraint, the random code exponent is given in [6, ch. 7, eqs. (7.3.19), (7.3.20)]. When specialized to our channel with transition probability $W(\cdot|\cdot)$ given by (2.3) and constraint function $=x$, [6, theorem 7.3.2] asserts the existence of a code with block length N, average cost q, with e^{R_1N} code words, and error probability

$$P_e\le\exp\{-N[E_0(\rho,q,r)-\rho R_1]+o(N)\},\quad \text{as } N\to\infty, \tag{3.1a}$$

where

$$E_0(\rho,q,r)=-\log\sum_{j=0}^{1}\left\{\sum_{k=0}^{1}Q(k)e^{r(k-q)}W(j|k)^{1/1+\rho}\right\}^{1+\rho}, \tag{3.1b}$$

and where $0\le\rho\le1$, $0\le q\le 1$, $0\le r<\infty$ are arbitrary, and $Q(1)=q$, $Q(0)=1-q$. To obtain the tightest bound, we maximize $E_0(\rho,q,r)$ over $0\le\rho\le1$, $0\le r<\infty$, and $0\le q\le\sigma$.

We now substitute $W(j|k)$ given by (2.3) into the expression for $E_0(\cdot)$ in (3.1b). Write

$$E_0(\rho,q,r)=-\log\sum_{j=0}^{1}V_j^{1+\rho}+r(1+\rho)q \tag{3.2a}$$

where

$$V_j=\sum_{k=0}^{1}Q(k)e^{rk}W(j|k)^{1/(1+\rho)},\qquad j=0,1. \tag{3.2b}$$

Making use of

$$(1+x)^t=1+tx+O(x^2),\qquad \text{as } x\to 0, \tag{3.3}$$

we can write

$$\begin{aligned}V_0&=(1-q)e^0(1-sA\Delta e^{-sA\Delta})^{1/1+\rho}\\&\quad+qe^r(1-(1+s)A\Delta e^{-(1+s)A\Delta})^{1/1+\rho}\\&=(1-q)\left(1-\frac{sA\Delta}{1+\rho}\right)+qe^r\left(1-\left(\frac{1+s}{1+\rho}\right)A\Delta\right)+O(\Delta^2)\\&=(1-q+qe^r)\\&\quad\cdot\left\{1-\frac{A\Delta}{1+\rho}\left[\frac{(1-q)s+(1+s)qe^r}{1-q+qe^r}\right]+O(\Delta^2)\right\},\end{aligned}$$

and

$$\begin{aligned}V_1&=(1-q)e^0(sA\Delta e^{-sA\Delta})^{1/(1+\rho)}\\&\quad+qe^r[(1+s)A\Delta e^{-(1+s)A\Delta}]^{1/(1+\rho)}\\&=(A\Delta)^{1/(1+\rho)}[s^{1/(1+\rho)}(1-q)+qe^r(1+s)^{1/(1+\rho)}]\\&\quad\cdot[1+O(\Delta)]\\&=(A\Delta)^{1/(1+\rho)}(1-q+qe^r)\\&\quad\cdot\left[\frac{s^{1/(1+\rho)}(1-q)+qe^r(1+s)^{1/(1+\rho)}}{(1-q+qe^r)}\right][1+O(\Delta)].\end{aligned}$$

Substituting into (3.2) and using

$$\ln(1+x)=x-\frac{x^2}{2}+O(x^3),\qquad \text{as } x\to 0, \tag{3.4}$$

we have

$$\begin{aligned}E_0(\rho,q,r) &= r(1+\rho)q-(1+\rho)\ln(1-q+qe^r)\\&\quad + A\Delta\left[\frac{(1-q)s+(1+s)qe^r}{(1-q)+qe^r}\right]\\&\quad - A\Delta\left[\frac{s^{1/(1+\rho)}(1-q)+qe^r(1+s)^{1/(1+\rho)}}{(1-q)+qe^r}\right]^{1+\rho}\\&\quad + O(\Delta^2). \qquad (3.5)\end{aligned}$$

We now maximize $E_0(\rho,q,r)$ with respect to ρ,q,r. Let us first maximize with respect to r. With q,ρ held fixed let

$$g(r)=r(1+\rho)q-(1+\rho)\ln(1-q+qe^r), \qquad (3.6a)$$

so that[1]

$$E_0(\rho,q,r)=g(r)+\Delta O_1(1). \qquad (3.6b)$$

Now $g(r)$ is a concave function of r, and $g(0)=g'(0)=0$. Thus the r which maximizes $E_0(\rho,q,r)$ must tend to zero as $\Delta\to 0$. In fact, the maximizing r satisfies

$$\begin{aligned}\frac{\partial E_0}{\partial r}=0&=\frac{dg(r)}{dr}+O(\Delta)\\&=(1+\rho)q+\frac{(1+\rho)qe^r}{1-q+qe^r}+O(\Delta).\end{aligned}$$

Thus using $e^r=1+r+O(r^2)$ as $(r\to 0)$, we have

$$r=O(\Delta), \qquad \text{as } \Delta\to 0.$$

Writing a Taylor series for $g(r)$, we have as $\Delta\to 0$,

$$\begin{aligned}g(r)&=g(0)+g'(0)r+g''(0)r^2+O(r^3)\\&=O(\Delta^2).\end{aligned}$$

Thus (3.5) yields, using $r=O(\Delta)$,

$$\max_r E_0(\rho,q,r)=A\Delta E_1(\rho,q)+O(\Delta^2) \qquad (3.7a)$$

where

$$\begin{aligned}E_1(\rho,q)&=s(1-q)+(1+s)q\\&\quad-\left[s^{1/(1+\rho)}(1-q)+q(1+s)^{1/(1+\rho)}\right]^{1+\rho}\\&=q+s-s[1+\tau q]^{1+\rho} \qquad (3.7b)\end{aligned}$$

and

$$\tau=\left(1+\frac{1}{s}\right)^{1/1+\rho}-1. \qquad (3.7c)$$

Further, since $T=\Delta N$, we have from (3.1)

$$P_e\le \exp\{-T(AE_1(\rho,q)-\rho R)+o(T)\} \qquad (3.8)$$

where we have passed to the limit $\Delta\to 0$, and $R=R_1/\Delta$ is the rate in nats per second. Taking the maximum of the exponent in (3.8) with respect to $0\le\rho\le 1$, and $0\le q\le\sigma$, we obtain (1.9). We next perform this maximization.

[1]It can be verified that $|O_1(1)|$ in (3.6b) and $|\partial O_1/\partial r|$ are $\le B<\infty$, for all ρ,q,r.

With R held fixed, set

$$E_2(\rho,q)=AE_1(\rho,q)-\rho R. \qquad (3.9)$$

It is easy to show that $\partial^2E_1(\rho,q)/\partial q^2\le 0$, so that with R,ρ held fixed, $E_2(\rho,q,R)$ is maximized with respect to q for $q=q_1(\rho)$ such that

$$\left.\frac{\partial E_1}{\partial q}\right|_{q=q_1}=0.$$

Since

$$\frac{\partial E_1}{\partial q}=A\left[1-s(1+\rho)(1+\tau q)^\rho\tau\right],$$

we have

$$q_1(\rho)=\frac{1}{\tau}\left[(s(1+\rho)\tau)^{-1/\rho}-1\right]. \qquad (3.10)$$

Furthermore, since, with ρ held fixed, $E_1(q,\rho)$ increases with q for $q\le q_1(\rho)$, we conclude that

$$\begin{aligned}\max_{0\le q\le\sigma} E_2(\rho,q)&=E_2(\rho,q^*(\rho))\\&\triangleq E_3(\rho) \qquad (3.11a)\end{aligned}$$

where

$$q^*(\rho)=\min(\sigma,q_1(\rho)). \qquad (3.11b)$$

Finally we must maximize $E_3(\rho)$ with respect to ρ. We begin by taking the derivative:

$$\frac{dE_3(\rho)}{d\rho}=\frac{\partial E_2}{\partial\rho}(\rho,q^*(\rho))+\frac{\partial E_2}{\partial q}(\rho,q^*(\rho))\frac{dq^*(\rho)}{d\rho}. \qquad (3.12)$$

Now if ρ is such that $q_1(\rho)\le\sigma$ (so that $q^*(\rho)=q_1(\rho)$), then

$$\left.\frac{\partial E_2}{\partial q}\right|_{q=q^*(\rho)}=A\left.\frac{\partial E_1}{\partial q}\right|_{q=q_1(\rho)}=0.$$

On the other hand, if $\rho=\rho^*$ such that $q_1(\rho^*)>\sigma$, then, since $q_1(\rho)$ is continuous in ρ, $q^*(\rho)=\sigma$ for ρ in a neighborhood of ρ^*, and therefore $dq^*(\rho)/d\rho|_{\rho=\rho^*}=0$. Thus in either case, the second term in the right member of (3.12) is zero. We conclude that $E_3(\rho)$ has a stationary point (with respect to ρ) when

$$\frac{dE_3(\rho)}{d\rho}=\frac{\partial E_2}{\partial\rho}=A\frac{\partial E_1}{\partial\rho}-R=0. \qquad (3.13)$$

Using the fact that $\partial^2E_1/\partial^2\rho\le 0$, a similar argument shows that $(d^2/d\rho^2)E_3(\rho)\le 0$, so that $E_3(\rho)$ is concave and the solution to (3.13) maximizes $E_3(\rho)$. Thus $E_3(\rho)$ is maximized when ρ satisfies

$$\begin{aligned}R=A\frac{\partial E_1}{\partial\rho}=As\Bigg[q^*\left(1+\frac{1}{s}\right)^{1/(1+\rho)}\left(\frac{1+\tau q^*}{1+\rho}\right)^\rho\log\left(1+\frac{1}{s}\right)\\-(1+\tau q^*)^{1+\rho}\log(1+\tau q^*)\Bigg]\triangleq R^*(\rho)\end{aligned}$$

with $q^*=q^*(\rho)$. We express R parametrically as a function of ρ, i.e., $R = R^*(\rho)$. Hence for $R = R^*(\rho_1)$, the error exponent $E_2(\rho, q)$ is maximized for $\rho = \rho_1$, and $q = q^*(\rho_1)$. Thus we have shown that the optimal exponent $E(R)$ defined in (1.8) is at least equal to the right member of (1.10) for $R^*(1) \le R \le C$.

Let us remark at this point that for purposes of establishing a lower bound on the error exponent, it would have sufficed simply to guess the optimizing r, ρ, q, since the bound of (3.1) holds for arbitrary r, ρ, q. However, as we shall see in the Part II, this optimization with respect to r, ρ, q is necessary when applying the sphere-packing lower bound on P_e. Since this optimization fits naturally into this section, we performed it here.

It remains to establish the lower bound on $E(R)$ for $0 \le R \le R^*(1)$. We begin by applying the general bound on $E(R)$ given by (3.1), for *ad hoc* r, ρ, q. Since our lower bound on P_e for $R \in [0, R^*(1)]$ does not depend on the optimization over these parameters, there is no need to perform the optimization.

Let us apply (3.1) with $r = 0$, $\rho = 1$, and $q = q^* \triangleq \min(\sigma, 1/2)$. Then, as in the derivation of (3.7), when $r = 0$,

$$\frac{1}{A\Delta} E_0(\rho, q, 0) = q + s - s[1 + \tau q]^{1+\rho} + O(\Delta)$$

where

$$\tau = \left(1 + \frac{1}{s}\right)^{1/(1+\rho)} - 1.$$

Setting $\rho = 1$, we have

$$\begin{aligned}
\lim_{\Delta \to 0} \frac{1}{A\Delta} E_0(1, q, 0) &= q + s - s\left[1 + q\left(\frac{\sqrt{s+1} - \sqrt{s}}{\sqrt{s}}\right)\right]^2 \\
&= q + s - \left[\sqrt{s} + q(\sqrt{s+1} - \sqrt{s})\right]^2 \\
&= q + s - \left[(1-q)\sqrt{s} + q\sqrt{s+1}\right]^2 \\
&= q + s - \left[(1-q)^2 s + 2q(1-q)\sqrt{s}\sqrt{s+1} + q^2(s+1)\right] \\
&= q(1-q)\left[1 + 2s - 2\sqrt{s}\sqrt{s+1}\right] \\
&= q(1-q)(\sqrt{s+1} - \sqrt{s})^2. \qquad (3.14)
\end{aligned}$$

Substituting into (3.1) and using $N = T/\Delta$, $R = R_1/\Delta$, we obtain

$$P_e \le \exp\{-T(R_0 - R)\}$$

where

$$R_0 = Aq(1-q)(\sqrt{s+1} - \sqrt{s})^2.$$

IV. Construction of Exponentially Optimal Codes

We begin by giving, in Section IV-A the construction of the codes. The estimation of the error probability follows in Section IV-B.

A. Code Construction

We describe a family of codes with parameters T, M with each code waveform $\lambda_m(\cdot)$ satisfying

$$\frac{1}{T}\int_0^T \lambda_m(t)\, dt = \frac{k}{M} A, \qquad 1 \le m \le M \qquad (4.1)$$

where k, $1 \le k \le M$ is an arbitrary integer. This family of codes is identical to the signal sets given in [7] in a different though related context. In fact this family was first discovered for the present application. Here is the code construction.

For T, M, k given, let $\mathscr{A}$ be the $M \times \binom{M}{k}$ binary matrix, the columns of which are the $\binom{M}{k}$ binary M-vectors with exactly k ones (and $(M-k)$ zeros). For example, if $M = 5$, $k = 2$, then $\binom{M}{k} = 10$, and $\mathscr{A}$ is the 5×10 matrix

$$\mathscr{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Note that the total number of nonzero entries in $\mathscr{A}$ is $k\binom{M}{k}$, so that (by symmetry) the number of nonzero entries in each row of a is $(k/M)\binom{M}{k}$. Let the (m, j)th entry of $\mathscr{A}$ be denoted by a_{mj}.

We now construct the code waveforms $\{\lambda_m(\cdot)\}_{m=1}^M$. Divide the interval $[0, T]$ into $\binom{M}{k}$ subintervals, each of length $T/\binom{M}{k}$. Then for t in the jth subinterval, set $\lambda_m(t) = Aa_{mj}$. Then

$$\frac{1}{T}\int_0^T \lambda_m(t)\, dt = \frac{1}{T}\sum_{j=1}^{\binom{M}{k}} \frac{TAa_{mj}}{\binom{M}{k}} = \frac{A}{\binom{M}{k}} \sum_j a_{mj} = \frac{Ak}{M}, \qquad (4.2)$$

which is (4.1).

We will also need to compute the Euclidean distance between distinct code waveforms. Thus for $m \ne m'$,

$$\begin{aligned}
\frac{1}{T}\int_0^T [\lambda_m(t) - \lambda_{m'}(t)]^2\, dt &= \sum_{j=1}^{\binom{M}{k}} \frac{A^2}{\binom{M}{k}} (a_{mj} - a_{m'j})^2 \\
&= \frac{2A^2}{\binom{M}{k}} \left\{ \begin{matrix} \text{number of } j \text{ such that} \\ a_{mj} = 1, a_{m'j} = 0 \end{matrix} \right\} \\
&= \frac{2A^2}{\binom{M}{k}} \binom{M-2}{k-1}.
\end{aligned}$$

The last step follows from the fact that the columns of $\mathscr{A}$ are precisely those M vectors with exactly k ones, so that, if we specify $a_{mk}=1$ and $a_{m'j}=0$, then the remaining $(M-2)$ entries in the jth column of A can be chosen in $\binom{M-2}{k-1}$ ways. Continuing, we have for $m \neq m'$

$$\frac{1}{T}\int_0^T [\lambda_m(t)-\lambda_{m'}(t)]^2\,dt$$
$$=\frac{2A^2}{\binom{M}{k}}\binom{M-2}{k-1}=\frac{2(M-k)(k)A^2}{M(M-1)}$$
$$=\frac{A^2 2M}{(M-1)}\left(\frac{k}{M}\right)\left(1-\frac{k}{M}\right). \tag{4.3}$$

Let us set $q=k/M$. Then let $M=\lfloor e^{RT}\rfloor$ and $T\to\infty$, while q is held fixed. Then since $\lambda_m(t)=0$ or A from (4.2),

$$\frac{1}{T}\mu\{t\colon \lambda_m(t)=A\}=q \tag{4.4a}$$

and from (4.3),

$$\frac{1}{T}\mu\{t\colon \lambda_m(t)=A \quad \lambda_{m'}(t)=0\}=\frac{M}{(M-1)}q(1-q)$$
$$\approx q(1-q) \tag{4.4b}$$

where μ denotes Lebesgue measure. Thus the pairwise the distribution of the code waveforms is nearly that which we would expect if the waveforms were chosen independently; and for each t, the probability that $\lambda_m(t)=A$ was q. This encourages us to hope that the resulting error probability is close to the random code bound, and in fact that turns out to be the case.

Decoder: We next define a decoder mapping D for our code. For $1\le m\le M$, let

$$S_m=\{t\in[0,1]\colon \lambda_m(t)=A\}. \tag{4.5}$$

Our decoder observes ν_0^T, and computes

$$\psi_m=\int_{S_m} d\nu(t)=\begin{Bmatrix}\text{number of arrivals}\\ \text{in } S_m\end{Bmatrix}, \qquad 1\le m\le M. \tag{4.6}$$

Then $D(\nu_0^T)=m^*$ if

$$\psi_m<\psi_{m^*}, \qquad 1\le m<m^*,$$
$$\psi_m\le\psi_{m^*}, \qquad m^*<m\le M. \tag{4.7}$$

Thus ν_0^T is decoded as that m which maximizes ψ_m, with ties resolved in favor of the smallest m. Although this decoding rule is the maximum likelihood decoding rule, we do not exploit this fact here. In the following section we overbound the error probability which results when this decoding rule is applied to our code.

In Fig. 1 we give a schematic diagram showing the graphs of two typical code waveforms when M is large. Of course, for nearly all the code waveforms $\lambda_m(\cdot)$, the support sets S_m would not be connected as they are in the figure.

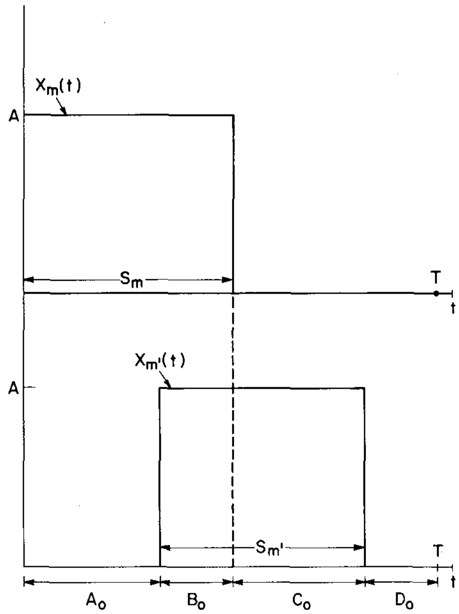

Fig. 1.

Let the intervals A_0, B_0, C_0, D_0 be as shown in the figure. Thus

$$A_0=S_m\cap S_{m'}^c \qquad C_0=S_m^c\cap S_{m'},$$
$$B_0=S_m\cap S_{m'} \qquad D_0=S_m^c\cap S_{m'}^c. \tag{4.8}$$

Also let W_A, W_B, W_C, W_D be the number of arrivals in the intervals A_0, B_0, C_0, D_0, respectively. Then $\psi_m=W_A+W_B$, and $\psi_{m'}=W_C+W_B$. Assuming that λ_m is transmitted, the decoder will prefer m' over m (thereby making an error) only if $\psi_{m'}\ge\psi_m$, or equivalently $W_C\ge W_A$. We assume that M is large so that the factor $M/(M-1)\approx 1$, and (4.4) holds. Thus, in particular,

$$\mu(A_0)=q(1-q)T \qquad \mu(C_0)=q(1-q)T,$$
$$\mu(B_0)=q^2T \qquad \mu(D_0)=(1-q)^2T, \tag{4.9}$$

B. Error Probability when $\lambda_0=0$

It turns out that the bounding process for the special case where there is no dark current ($\lambda_0=0$) is far easier than for the general case. For this reason we will bound P_e for this special case separately, leaving the general case for the most hardy spirits.

Let us begin by taking a look at what we have to prove. Refer to (1.9). When the dark current intensity $\lambda_0=0$, then $s=\lambda_0/A=0$. As $s\to 0$, τ (as given in (1.9c)) satisfies

$$\tau\approx s^{-1/(1+\rho)}$$

and from (1.9b)

$$E_1(\rho,q)=q-q^{1+\rho}.$$

We will show that for any $R\ge 0$, $q\in[0,\sigma]$ and $\rho\in[0,1]$, for T sufficiently large, there is a code in our family with parameters ($\lfloor e^{RT} \rfloor, T, \sigma, P_e$), where

$$-\frac{1}{T}\ln P_e \geq AE_1(\rho, q) - \rho R$$
$$= Aq - Aq^{(1+\rho)} - \rho R. \tag{4.10}$$

To do this we set $M = e^{RT}$ and $k = \sigma M$ (ignoring the constraint that M, k must be integers), and construct the code as specified in Section IV-A. Note that (1.3) is satisfied and the code has average power σ. We now estimate the error probability.

Given the code $\{\lambda_m(\cdot)\}_{m=1}^{M}$ as specified in Section IV-A, define, for $1 \leq m \leq M$,

$$P_{em} = \Pr\{D(\nu_0^T) \neq m | \lambda_m(\cdot) \text{ is transmitted}\}. \tag{4.11}$$

The decoder D is defined by (4.7), and of course

$$P_e = \frac{1}{M}\sum_{m=1}^{M} P_{em}. \tag{4.12}$$

Let m be given ($1 \leq m \leq M$). For $m' \neq m$, define the event

$$\boldsymbol{E}_{m'} = \{\psi_{m'} \geq \psi_m\}. \tag{4.13}$$

Thus the decoder D "prefers" m' over m only if $\boldsymbol{E}_{m'}$ occurs. Furthermore,

$$P_{em} \leq \Pr\Big(\bigcup_{m' \neq m} \boldsymbol{E}_{m'} | \lambda_m(\cdot)\Big). \tag{4.14}$$

Now, without loss of generality, we can assume that the support set S_m of $\lambda_m(\cdot)$ is the interval $[0, qT]$. Thus, given that $\lambda_m(\cdot)$ is transmitted, the random variable ψ_m, the number of arrivals in S_m, is Poisson distributed with parameter qAT, i.e.,

$$\Pr\{\psi_m = n | \lambda_m(\cdot)\} = e^{-qAT}\frac{(qAT)^n}{n!}.$$

Thus we can write (4.14) as

$$\begin{aligned} P_{em} &\leq \sum_{n=0}^{\infty} \Pr\{\psi_m = n|\lambda_m(\cdot)\} \\ &\quad \cdot \Pr\Big(\bigcup_{m'\neq m} \boldsymbol{E}_{m'}|\lambda_m(\cdot), \psi_m = n\Big) \\ &= \sum_{n=0}^{\infty} \frac{e^{-qAT}(qAT)^n}{n!} \Pr\Big(\bigcup_{m'\neq m} \boldsymbol{E}_{m'}|\lambda_m(\cdot), \psi_m = n\Big) \\ &\leq \sum_{n=0}^{\infty} \frac{e^{-qAT}(qAT)^n}{n!}\Big[\sum_{m'\neq m} \Pr(\boldsymbol{E}_{m'}|\lambda_m(\cdot), \psi_m = n)\Big]^{\rho} \end{aligned} \tag{4.15}$$

where $\rho \in [0,1]$ is the arbitrary parameter in (4.10). The last step in (4.15) follows from the fact that, for any set of events, say $\{A_i\}$,

$$\Pr \bigcup_i A_i \leq \Big[\sum_i \Pr(A_i)\Big]^{\rho}, \qquad 0 \leq \rho \leq 1. \tag{4.16}$$

For $\rho = 1$, this is the familiar union bound. For $\sum \Pr(A_i) < 1$, raising the sum to the ρth power only weakens the union bound. For $\sum \Pr(A_i) \geq 1$, the right member of (4.16) is ≥ 1, so that the bound holds trivially.

Let us now look at $\Pr\{E_{m'}|\lambda_m(\cdot), \psi_m = n\}$. The code waveforms $\lambda_m(\cdot)$ and $\lambda_{m'}(\cdot)$ can be represented schematically as in Fig. 1. Since there is no dark current, there cannot be any arrivals on the interval $S_m^c = C_0 + D_0$, when $\lambda_m(t) = 0$. Thus, given that $\lambda_m(\cdot)$ is transmitted, $\psi_{m'} = W_B$ = the number of arrivals in interval B_0. Furthermore, given that $\psi_m = n$, the event $\{\psi_{m'} \geq \psi_m\}$ occurs if and only if the n arrivals on $S_m = A_0 + B_0$ all fall on interval B_0. Since these n arrivals are uniformly and independently distributed on S_m, we have

$$\Pr(\boldsymbol{E}_m|\lambda_m(\cdot), \psi_m = n) = [\mu(B_0) \quad \mu(A_0 + B_0)]^n = q^n.$$

Substituting into (4.15), we obtain

$$\begin{aligned} P_{em} &\leq \sum_{n=0}^{\infty} \frac{e^{-qAT}(qAT)^n}{n!}\Big[\sum_{m'\neq m} q^n\Big]^{\rho} \\ &= (M-1)^{\rho}\sum_{n=0}^{\infty} \frac{e^{-qAT}(qAT)^n}{n!} q^{\rho n} \\ &\leq M^{\rho}e^{-qAT}\sum_{n=0}^{\infty}\frac{[q^{(1+\rho)}AT]^n}{n!} \\ &= M^{\rho}\exp\{-AT(q - q^{1+\rho})\}. \end{aligned}$$

Setting $M = e^{RT}$ yields

$$P_{em} \leq \exp\{-T(Aq - Aq^{(1+\rho)} - \rho R)\}, \tag{4.17}$$

which, combined with (4.12), yields (4.10) which is what we have to establish.

C. Error Probability for Positive λ_0

The bounding process for the case of positive λ_0 parallels the process for $\lambda_0 = 0$. We bound the error probability P_{em} (defined by (4.11)) conditioned on both the total number of arrivals on $[0, T]$, i.e., $\nu(T)$, and on ψ_m. We then apply the generalized union bound of (4.16) to obtain the desired bound on P_{em} and P_e.

We will show that for any $R \geq 0$, $q \in [0, \sigma]$ and $\rho \in [0,1]$, for T sufficiently large, there is a code in our family with parameters ($\lfloor e^{RT} \rfloor, T, \sigma, P_e$) where

$$-\frac{1}{T}\ln P_e \geq AE_1(\rho, q) - \rho R$$

where $E_1(\rho, q)$ is given in (1.9). As in the case $\lambda_0 = 0$, we do this by setting $M = e^{RT}$ and $k = \sigma M$, and construct the code as specified in Section IV-A.

Given the code $\{\lambda_m(\cdot)\}_{m=1}^{M}$ as specified in Section IV-A, define P_{em} by (4.11), so that P_e is given by (4.12). For a given m and $m' \neq m$, define the event $\boldsymbol{E}_{m'}$ by (4.13), and observe that (4.14) also holds in the general case, $\lambda_0 > 0$, i.e.,

$$P_{em} \leq \Pr\Big(\bigcup_{m'\neq m} \boldsymbol{E}_{m'}|\lambda_m(\cdot)\Big). \tag{4.18}$$

Again as in the above discussion, assume that the support set S_m of $\lambda_m(\cdot)$ is the interval $[0, qT]$. Now when $\lambda_m(\cdot)$ is transmitted, $\nu(T)$, the number of arrivals in $[0,T]$, is the sum of two independent Poisson distributed random variables $\nu(qT)$ and $(\nu(T)-\nu(qT))$, with parameters Λ_1 and Λ_0, respectively, where

$$\Lambda_1 = (A+\lambda_0)qT = qAT(1+s)$$
$$\Lambda_0 = \lambda_0(1-q)T = (1-q)ATs. \quad (4.19)$$

Thus $\nu(T)$ is Poisson distributed with parameter

$$\Lambda = \Lambda_1 + \Lambda_0 = (q+s)AT. \quad (4.20)$$

Furthermore, given that $\nu(T)=n$, $\psi_m = \nu(qT)$ has the binomial distribution

$$\Pr\{\psi_m = n_1 | \nu(T) = n, \quad \lambda_m(\cdot)\} = \binom{n}{n_1}\pi^{n_1}(1-\pi)^{n-n_1} \quad (4.21a)$$

where

$$\pi = \frac{\Lambda_1}{\Lambda_1+\Lambda_0} = \frac{q(1+s)}{q+s}. \quad (4.21b)$$

Thus the joint probability

$$\Pr\{\psi_m = n_1, \nu(T) = n | \lambda_m(\cdot)\}$$
$$= \frac{e^{-\Lambda}\Lambda^n}{n!}\left[\binom{n}{n_1}\pi^{n_1}(1-\pi)^{n-n_1}\right]$$
$$\triangleq Q(n_1, n). \quad (4.22)$$

We will bound P_{em} starting from (4.18) using a technique similar to that used for the case $\lambda_0 = 0$. Specifically, we will condition on $\psi_m = n_1$ and $\nu(T) = n$. This will lead us to consider terms like

$$\Pr(\boldsymbol{E}_{m'} | \lambda_m(\cdot), \psi_m = n_1, \nu(T) = n), \quad (4.23a)$$

where $m' \neq m$. Whenever it is unambiguous, we will write such conditional probabilities as

$$\Pr(\boldsymbol{E}_{m'} | \lambda_m, n_1, n). \quad (4.23b)$$

Now for a given m, m' $(m \neq m')$, we can, as we did before, assume that the waveforms $\lambda_m(\cdot)$ and $\lambda_{m'}(\cdot)$ are represented as in Fig. 1. As we remarked following (4.8), $\psi_{m'} \geq \psi_m$ if and only if $W_C - W_A \geq 0$. Let us define

$$n_0 = n - n_1 \quad (4.24)$$

so that the conditions $\nu(T) = n$, $\psi_m = n_1$ are equivalent to

$$W_A + W_B = n_1 \qquad W_C + W_D = n_0. \quad (4.25)$$

Now, given $\lambda_m(\cdot)$ transmitted and conditions (4.25), W_C, W_A are independent random variables with

$$\Pr\{W_A = k_1 | \lambda_m, n_1, n\}$$
$$= \binom{n_1}{k_1}\left[\frac{\mu(A_0)}{\mu(A_0)+\mu(B_0)}\right]^{k_1}\left[1 - \frac{\mu(A_0)}{\mu(A_0)+\mu(B_0)}\right]^{n_1-k_1}$$
$$= \binom{n_1}{k_1}(1-q)^{k_1}q^{n_1-k_1}, \qquad 0 \leq k_1 \leq n_1$$

$$\Pr\{W_C = k_0 | \lambda_m, n_1, n\} = \binom{n_0}{k_0}q^{k_0}(1-q)^{n_0-k_0},$$
$$0 \leq k_0 \leq n_0. \quad (4.26)$$

Of course,

$$\Pr(\boldsymbol{E}_{m'} | \lambda_m, n_1, n) = \Pr\{W_C - W_A \geq 0 | \lambda_m, n, n_1). \quad (4.27)$$

Now, from (4.26)

$$E(W_C - W_A | \lambda_m, n_1, n) = n_0 q - n_1(1-q)$$
$$= (n-n_1)q - n_1(1-q). \quad (4.28)$$

Should this expectation be negative, we might expect the probability in (4.27) to be small, and in fact this is the case. This motivates us to define the set

$$A = \{(n_1, n): 0 \leq n_1 \leq n$$
$$n_0 q - n_1(1-q) = (n-n_1)q - n_1(1-q) < 0\}. \quad (4.29)$$

Then, returning to (4.18), we have

$$P_{em} \leq \Pr\left(\bigcup_{m' \neq m} \boldsymbol{E}_{m'} | \lambda_m\right)$$
$$= \sum_{n_1, n} Q(n_1, n)\Pr\left(\bigcup_{m' \neq m} \boldsymbol{E}_{m'} | \lambda_m, \psi_m = n_1, \nu(T) = n\right)$$
$$\leq \sum_{(n_1, n) \notin A} Q(n_1, n)$$
$$+ \sum_{(n_1, n) \in A} Q(n_1, n)\left[\sum_{m' \neq m} \Pr(\boldsymbol{E}_{m'} | \lambda_m, n_1, n)\right]^{\rho} \quad (4.30)$$

where $\rho \in [0,1]$ is arbitrary. The first term in (4.30) is bounded by Lemma B.1 in Appendix II as

$$\sum_{(n_1, n) \notin A} Q(n_1, n) = \Pr\{(\psi_m, \nu(T)) \notin A | \lambda_m\}$$
$$\leq \exp\{-\hat{E}T) \quad (4.31a)$$

where

$$\hat{E} = A\left[q(1+s) + (1-q)s - (1+s)^q s^{1-q}\right]. \quad (4.31b)$$

This leads us to consider the second term in (4.30). Specifically, let us look at $\Pr(\boldsymbol{E}_{m'} | \lambda_m, n_1, n)$, for $m' \neq m$ and $(n_1, n) \in \boldsymbol{A}$. From (4.27),

$$\Pr\{\boldsymbol{E}_{m'} | \lambda_m, n_1, n\} = \Pr\{W_C - W_A \geq 0 | \lambda_m, n, n_1\}$$
$$\leq E\left(e^{\tau(W_C - W_A)} | \lambda_m, n_1, n\right) \quad (4.32)$$

where $\tau \geq 0$ is arbitrary and $(n_1, n) \in \boldsymbol{A}$.[2] The expectation in (4.32) under the indicated conditions can be found directly using the distributions for W_A and W_B in (4.26). Thus

$$E\left(e^{-\tau W_A} | \lambda_m, n_1, n\right) = \sum_{k_1=0}^{n_1}\binom{n}{k_1}(1-q)^{k_1}q^{n_1-k_1}e^{-\tau k_1}$$
$$= \left(q + (1-q)e^{-\tau}\right)^{n_1} \quad (4.33a)$$

$$E\left(e^{\tau W_C} | \lambda_m, n_1, n\right) = \sum_{k_0=0}^{n_0}\binom{n}{k_0}q^{k_0}(1-q)^{n_0-k_0}e^{\tau k_0}$$
$$= \left(1 - q + qe^{\tau}\right)^{n_0}. \quad (4.33b)$$

[2] We have made use of the well-known inequality $\Pr(U \geq 0) \leq Ee^{\tau U}$, for any random variable U and $\tau \geq 0$.

Using the conditional independence of W_A, W_C and (4.33), (4.32) becomes, for $(n_1, n) \in A$, $\tau \geq 0$,

$$\Pr(E_m, |\lambda_m, n_1, n) \leq \exp\{\gamma(\tau)\} \tag{4.34a}$$

where

$$\gamma(\tau) = n_0 \ln(1-q+qe^{\tau}) + n_1 \ln(q+(1-q)e^{-\tau}). \tag{4.34b}$$

To get value of τ which yields the best bound, set the derivative of $\gamma(\tau)$ equal to zero:

$$\gamma'(\tau) = \frac{n_0}{1-q+qe^{\tau}} qe^{\tau} - \frac{n_1(1-q)e^{-\tau}}{q+(1-q)e^{-\tau}} = 0. \tag{4.35}$$

Equation (4.35) is a quadratic equation in (e^{τ}):

$$(e^{2\tau})(n_0 q^2) + e^{\tau}[n_0 q(1-q) - n_1(1-q)q] - n_1(1-q)^2 = 0,$$

which factors as

$$[e^{\tau} n_0 q - n_1(1-q)][e^{\tau} q + (1-q)] = 0. \tag{4.36}$$

Only one of the solutions of (4.36) for e^{τ} is positive:

$$e^{\tau} = \frac{n_1(1-q)}{n_0 q}. \tag{4.37}$$

Note that $(n_1, n) \in A$ implies that

$$\tau = \ln\left[\frac{n_1(1-q)}{n_0 q}\right] \geq 0,$$

so that (4.32) will hold. Substituting (4.37) into (4.34) yields after a bit of manipulation

$$P(E_{m'}|\lambda_m, n_1, n) \leq (1-q)^{n_0} q^{n_1} \frac{n^n}{n_1^{n_1} n_0^{n_0}} \triangleq \Gamma(n_1, n) \tag{4.38}$$

for $m \neq m'$, $(n_1, n) \in A$, $n_0 = n - n_1$. Substituting (4.38) and (4.31) into (4.30), we obtain

$$P_{em} \leq e^{-\hat{E}T} + \sum_{(n_1, n) \in A} Q(n_1, n)\left[\sum_{m' \neq m} \Gamma\right]^{\rho}$$

$$\leq e^{-\hat{E}T} + M^{\rho} \sum_{n=1}^{\infty} \sum_{n_1=0}^{n} Q(n_1, n)[\Gamma(n_1, n)]^{\rho} \tag{4.39}$$

where $\hat{E}$ is given (4.31b), Γ by (4.38), and Q by (4.22).

Now for a given n,

$$\sum_{n_1=0}^{n} Q(n_1, n)\Gamma^{\rho}(n_1, n) = \frac{e^{-\Lambda}\Lambda^n}{n!} \sum_{n_1=0}^{n} \binom{n}{n_1} \pi^{n_1}(1-\pi)^{n-n_1}\left[\frac{(1-q)^{n_0} q^{n_1} n^n}{n_1^{n_1} n_0^{n_0}}\right]^{\rho} \tag{4.40}$$

where π is given by (4.21b), and Λ by (4.20). Let

$$\xi = n_1/n \tag{4.41}$$

so that

$$\sum_{n_1=0}^{n} Q(n_1, n)\Gamma^{\rho} = \frac{e^{-\Lambda}\Lambda^n}{n!} \sum_{n_1=0}^{n} \binom{n}{\xi n} \pi^{\xi n}(1-\pi)^{(1-\xi)n} \frac{(1-q)^{\rho(1-\xi)n} q^{\rho\xi n}}{\xi^{\rho\xi n}(1-\xi)^{\rho(1-\xi)n}}. \tag{4.42}$$

Using

$$\binom{n}{\xi n} \leq e^{nh(\xi)} = \exp\{-n\xi \ln \xi - n(1-\xi)\ln(1-\xi)\},$$

(4.42) becomes

$$\sum_{n_1=0}^{n} Q(n_1, n)\Gamma^{\rho} = \frac{e^{-\Lambda}\Lambda^n}{n!} \sum_{n_1=0}^{n} \exp\left\{n\left[\xi \ln \frac{\pi q^{\rho}}{\xi^{(1+\rho)}} + (1-\xi)\ln \frac{(1-\pi)(1-q)^{\rho}}{(1-\xi)^{1+\rho}}\right]\right\}$$

$$\leq \frac{e^{-\Lambda}\Lambda^n}{n!} n \exp\left\{n \max_{0 \leq \xi \leq 1}\left[\xi \ln \frac{q^{\rho}\pi}{\xi^{(1+\rho)}} + (1-\xi)\ln \frac{(1-\pi)(1-q)^{\rho}}{(1-\xi)^{1+\rho}}\right]\right\}. \tag{4.43}$$

Set

$$\alpha = (q^{\rho}\pi)^{1/(1+\rho)} \qquad \beta = [(1-\pi)(1-q)^{\rho}]^{1/(1+\rho)}. \tag{4.44}$$

Then the term in square brackets in (4.43) is

$$(1+\rho)\left[\xi \ln \frac{\alpha}{\xi} + (1-\xi)\ln \frac{\beta}{(1-\xi)}\right]. \tag{4.45}$$

To maximize this concave function of ξ with respect to ξ, we set its derivative equal to zero which yields $\xi = \alpha/(\alpha+\beta)$. Therefore, the maximum of the square-bracketed term in (4.45) is $(1+\rho)\ln(\alpha+\beta)$. Substituting into (4.43) yields

$$\sum_{n_1=0}^{n} Q(n_1, n)\Gamma^{\rho} \leq \frac{e^{-\Lambda}\Lambda^n}{(n-1)!}(\alpha+\beta)^{n(1+\rho)}. \tag{4.46}$$

Now substitute (4.46) into (4.39) to obtain

$$P_{em} \leq e^{-\hat{E}T} + M^{\rho} \sum_{n=1}^{\infty} \frac{e^{-\Lambda}\Lambda^n}{(n-1)!}(\alpha+\beta)^{n(1+\rho)}$$

$$= e^{-\hat{E}T} + M^{\rho}\Lambda(\alpha+\beta)^{1+\rho} \cdot \sum_{n=1}^{\infty} e^{-\Lambda} \frac{(\alpha+\beta)^{(1+\rho)(n-1)}\Lambda^{n-1}}{(n-1)!}$$

$$= e^{-\hat{E}T} + M^{\rho}\Lambda(\alpha+\beta)^{1+\rho} \exp\{-\Lambda + \Lambda(\alpha+\beta)^{1+\rho}\} \tag{4.47}$$

where $\hat{E}$ is given by (4.31b), Λ by (4.20), and α, β by (4.44) with π given by (4.21b). Thus

$$\alpha = (q^\rho \pi)^{1/(1+\rho)} = \left[\frac{q^{1+\rho}(1+s)}{q+s}\right]^{1/(1+\rho)} = q\left(\frac{1+s}{q+s}\right)^{1/(1+\rho)}$$

$$\beta = \left((1-q)^\rho(1-\pi)\right)^{1/(1+\rho)} = (1-q)\left[\frac{s}{q+s}\right]^{1/(1+\rho)}$$

and the exponent in the right member of (4.47) is

$$\begin{aligned}
-\Lambda + \Lambda(\alpha+\beta)^{1+\rho} &= -TA(q+s)\left\{1 - \left[q\left(\frac{1+s}{q+s}\right)^{1/(1+\rho)} + (1-q)\left(\frac{s}{q+s}\right)^{1/(1+\rho)}\right]^{1+\rho}\right\}\\
&= -TA\left\{q+s-\left[q(1+s)^{1/(1+\rho)}+(1-q)s^{1/1+\rho}\right]^{1+\rho}\right\}\\
&= -TA\left\{q+s-s(1+\tau q)^{1+\rho}\right\}\\
&= -TAE_1(\rho,q)
\end{aligned}$$

where $\tau = (1+(1/s))^{1/(1+\rho)} - 1$, and $E_1(\rho,q)$ is given by (1.9). Substituting into (4.47) and setting $M = e^{RT}$, we have

$$P_{em} \le e^{-\hat{E}T} + \exp\{-T(AE_1(\rho,q)-\rho R)+o(T)\},$$

as $T \to \infty$. We show in Appendix II that for all q, ρ, R, the exponent $\hat{E} \ge E_1(\rho,q) - \rho R$. Thus

$$P_e = \frac{1}{M}\sum_{m=1}^{M} P_{em} \le \exp\{-T[AE_1(\rho,q)-\rho R]+o(T)\},$$

which is what we have to show.

Appendix I

In this Appendix we will verify the limiting and asymptotic formulas given in Section 1. We begin by verifying (1.13).

Proposition A.1: For fixed $s \ge 0$,

$$\lim_{\rho\to 0} q_1(\rho) = q_0(s) \triangleq \frac{(1+s)^{1+s}}{s^s e} - s \qquad \text{(A.1a)}$$

where $q_1(\rho)$ is given by (1.13), i.e.,

$$q_1(\rho) = \frac{1}{\tau(\rho)}\left\{\left[\frac{1}{s(1+\rho)\tau(\rho)}\right]^{1/\rho} - 1\right\}, \qquad \text{(A.1b)}$$

and $\tau(\rho)$ is given by (1.9c), i.e.,

$$\tau(\rho) = \left(1+\frac{1}{s}\right)^{1/(1+\rho)} - 1. \qquad \text{(A.1c)}$$

Proof: We begin by expanding $\tau(\rho)$ in a Taylor series about $\rho = 0$. We have

$$\tau(\rho) = \tau(0) + \rho\tau'(0) + O(\rho^2) = \frac{1}{s} - \rho\left(1+\frac{1}{s}\right)\ln\left(1+\frac{1}{s}\right) + O(\rho^2).$$

Thus $s\tau = 1 - \rho(1+s)\ln(1+(1/s)) + O(\rho^2)$ and

$$\left(\frac{1}{s\tau(1+\rho)}\right)^{1/\rho} = \frac{1}{(1+\rho)^{1/\rho}}\left[\frac{1}{1-\rho(1+s)\ln\left(1+\frac{1}{s}\right)+O(\rho^2)}\right]^{1/\rho} \underset{\rho=0}{\to} e^{-1}\exp\left\{(1+s)\ln\left(1+\frac{1}{s}\right)\right\} = e^{-1}\left(1+\frac{1}{s}\right)^{1+s}.$$

Thus, since $\tau(0) = 1/s$, substitution into (A.1b) yields

$$q_1(\rho) \to q_0(s), \qquad \text{as } \rho \to 0,$$

which is (A.1a) and the proposition.

We next verify (1.11).

Proposition A.2: $\lim_{\rho\to 0} R^*(\rho) = C$, where $R^*(\rho)$ is given by (1.10a).

Proof: Since, from (1.10b),

$$q^*(\rho) = \min(\sigma, q_1(\rho)),$$

we have from Proposition A.1 that

$$q^*(0) = \min(\sigma, q_0(s)).$$

Further, since $\tau(0) = 1/s$, (1.10a) yields

$$R^*(0) = As\left[q^*\left(1+\frac{1}{s}\right)\ln\left(1+\frac{1}{s}\right) - \left(1+\frac{q^*}{s}\right)\ln\left(1+\frac{q^*}{s}\right)\right]$$

where $q^* = q^*(0)$. Rearranging yields

$$R^*(0) = A[q^*(1+s)\ln(1+s) - (s+q^*)\ln(s+q^*) + s(1-q^*)\ln s].$$

Comparison with (1.5), establishes the proposition.

We now turn to the limiting formulas for channel capacity C when $s = 0$ (no dark current) and $s = \infty$. The formula for C when $s = 0$ follows immediately from the general formula and is given by (1.6). For $s = \infty$, the asymptotic formula for C is given by (1.7) which is Proposition A.3.

Proposition A.3: As $s \to \infty$, the capacity

$$C = \frac{Aq^*(1-q^*)}{2s} + O\left(\frac{1}{s^2}\right) \qquad \text{(A.2a)}$$

where

$$q^* = \min(\sigma, 1/2). \qquad \text{(A.2b)}$$

Proof: Using the Taylor series $\ln(1+x) = x - (x^2/2) + (x^3/3) + \cdots$, and $e^x = 1 + x + (x^2/2!) + (x^3/3!) + \cdots$, we get the following expansion for $q_0(s)$ (given by (1.5d)) for large s:

$$\begin{aligned} q_0 &= \frac{(1+s)^{1+s}}{s^s e} - s = \frac{(1+s)}{e}\left(1+\frac{1}{s}\right)^s - s \\ &= \frac{(1+s)}{e}\exp\left\{s\ln\left(1+\frac{1}{s}\right)\right\} - s \\ &= \frac{(1+s)}{e}\exp\left\{s\left(\frac{1}{s}-\frac{1}{2s^2}+O\left(\frac{1}{s^3}\right)\right)\right\} - s \\ &= \frac{(1+s)}{e}\exp\left\{1-\frac{1}{2s}+O\left(\frac{1}{s^2}\right)\right\} - s \\ &= \frac{(1+s)}{e}e\left(1-\frac{1}{2s}+O\left(\frac{1}{s^2}\right)\right) - s = \frac{1}{2}+O\left(\frac{1}{s}\right). \end{aligned} \tag{A.3}$$

We now approximate C as given by (1.5a) for large s:

$$\begin{aligned} C &= A[q^*(1+s)\ln(1+s)+(1-q^*)s\ln s-(q^*+s)\ln(q^*+s)] \\ &= A\left[q^*(1+s)\ln\left(1+\frac{1}{s}\right)-(q^*+s)\ln\left(1+\frac{q^*}{s}\right)\right] \\ &= A\left[q^*(1+s)\left[\frac{1}{s}-\frac{1}{2s^2}+O\left(\frac{1}{s^3}\right)\right]\right. \\ &\quad \left. -(s+q^*)\left[\frac{q^*}{s}-\frac{1}{2}\frac{q^{*2}}{s^2}+O\left(\frac{1}{s^3}\right)\right]\right] \\ &= A\left[q^*\left(1+\frac{1}{2s}+O\left(\frac{1}{s^2}\right)\right)-q^*-\frac{1}{2}\frac{q^{*2}}{s}+O\left(\frac{1}{s^2}\right)\right] \\ &= \frac{Aq^*(1-q^*)}{2s}+O\left(\frac{1}{s^2}\right). \end{aligned} \tag{A.4}$$

Comparison of (A.3) and (A.4) with (A.2) yields the proposition.

We now look at the error exponents for the cases $s=0,\infty$. We first verify (1.15).

Proposition A.4: For $s=0$, let

$$R^*(\rho) = A[q^*(\rho)]^{1+\rho}\ln(q^*(\rho)), \tag{A.5a}$$

where

$$q^*(\rho) = \min\left(\sigma, \frac{1}{(1+\rho)^{1/\rho}}\right). \tag{A.5b}$$

Then, for $R=R^*(\rho)$, $0\le\rho\le 1$,

$$E(R) = Aq^*(\rho) - A[q^*(\rho)]^{1+\rho} - \rho R. \tag{A.6}$$

Proof: As $s\to 0$, from (1.9c) (with ρ held fixed)

$$\tau(\rho) = \left(1+\frac{1}{s}\right)^{1/(1+\rho)} - 1 = \left(\frac{1}{s}\right)^{1/(1+\rho)}(1+o(1)).$$

Thus $s\tau(\rho) = s^{\rho/(1+\rho)}$, and (1.13) yields

$$q_1(\rho) \to (1+\rho)^{-1/\rho},$$

as $s\to 0$. Thus from (1.10b), q^* is as given in (A.5b). Substitution of this q^* and $\tau(\rho)$ into (1.10a) yields (A.6) and the proposition.

Finally, we turn to the case of large background noise, $s\to\infty$. Here the asymptotics are a bit tricky, and we must proceed with care.

We start with (1.9) which is

$$E(R) = \max_{\substack{0\le\rho\le 1\\ 0\le q\le\sigma}} [AE_1(\rho,q)-\rho R] \tag{A.7}$$

where $E_1(\rho,q)$ is given by (1.9b). Let us expand $E_1(\rho,q)$ and $\tau(\rho)$ in powers of $(1/s)$, as $s\to\infty$ with ρ, q held fixed. Using the binomial formula, we have from (1.9c)

$$\begin{aligned} \tau &= \left(1+\frac{1}{s}\right)^{1/(1+\rho)} - 1 \\ &= 1+\frac{1}{(1+\rho)s}+\left(\frac{1}{1+\rho}\right)\left(\frac{1}{1+\rho}-1\right)\frac{1}{s^2}\frac{1}{2}+O\left(\frac{1}{s^3}\right)-1 \\ &= \frac{1}{(1+\rho)s}\left[1-\frac{\rho}{2(1+\rho)s}+O\left(\frac{1}{s^2}\right)\right]. \end{aligned} \tag{A.8}$$

Also, applying the binomial formula to (1.9b), we obtain, as $s\to\infty$ (so that $\tau = O(1/s)\to 0$),

$$\begin{aligned} E_1(\rho,q) &= q+s-s[1+\tau q]^{1+\rho} \\ &= q+s-s\left[1+(1+\rho)\tau q+\frac{(1+\rho)\rho}{2}(\tau q)^2+O(\tau^3 q^3)\right]. \end{aligned}$$

Substituting (A.8), we have

$$\begin{aligned} E_1(\rho,q) &= q - s\tau(1+\rho)q - \frac{s(1+\rho)\rho}{2}\tau^2 q^2 + sO(\tau^3 q^3) \\ &= q - q\left[1-\frac{\rho}{2(1+\rho)s}+O\left(\frac{1}{s^2}\right)\right] \\ &\quad -\frac{\rho q^2}{2(1+\rho)s}+O\left(\frac{1}{s^2}\right) \\ &= \frac{q(1-q)\rho}{2s(1+\rho)}+O\left(\frac{1}{s^2}\right). \end{aligned} \tag{A.9}$$

Finally, substituting (A.9) into (A.7) we obtain

$$E(R) \sim \max_{\substack{0\le\rho\le 1\\ 0\le q\le\sigma}} \frac{Aq(1-q)\rho}{2s(1+\rho)} - \rho R, \tag{A.10}$$

as $s\to\infty$. The right member of (A.10) is maximized with respect to q when $q=\min(\sigma,1/2)\triangleq q^*$. Furthermore, from (1.7), the channel capacity $C\sim Aq^*(1-q^*)/2s$, as $s\to\infty$. Thus we can rewrite (A.7) as

$$E(R) \sim \max_{0\le\rho\le 1}\left[\frac{\rho}{1+\rho}C-\rho R\right]. \tag{A.11}$$

Differentiating the term in brackets in (A.11) with respect to ρ and setting the result equal to zero, we conclude that, for $0\le R < C$, the maximizing ρ is

$$\rho = \min\left(1, \sqrt{\frac{C}{R}}-1\right).$$

Since $\sqrt{C/R}-1\le 1$, for $C/4\le R\le C$, we have

$$E(R) \sim \begin{cases} C/2-R, & 0\le R\le C/4 \\ (\sqrt{C}-\sqrt{R})^2, & C/4\le R\le C \end{cases},$$

which is (1.16).

Appendix II

In this appendix we establish two lemmas which we needed in Section IV. The first lemma is needed to verify (4.31). Let V_0 and V_1 be independent Poisson random variables with

$$\begin{aligned} E(V_0) &= \Lambda_0 = (1-q)ATs \\ E(V_1) &= \Lambda_1 = qAT(1+s). \end{aligned} \tag{B.1}$$

When code waveform $\lambda_m(\cdot)$ is transmitted, $(\psi_m, \nu(T))$ has the same distribution as $(V_1, V_1 + V_0)$. Thus, with set A defined by (4.29),

$$\Pr\{(\psi_m, \nu(T)) \notin A\} = \sum_{(n_1,n)\notin A} Q(n_1, n) = \Pr\{V_0 q - V_1(1-q) > 0\}. \quad \text{(B.2)}$$

Thus (4.31) follows from Lemma B1.

Lemma B1: With V_0, V_1 defined as above, let $Z = qV_0 - (1-q)V_1$. Then

$$\Pr\{Z > 0\} \le e^{-\hat{E}T} \quad \text{(B.3a)}$$

where

$$\hat{E} = A\left[q(1+s) + (1-q)s - (1+s)^q s^{1-q}\right]. \quad \text{(B.3b)}$$

Proof: We use the same (Chernoff) bounding technique as in Section IV. For all $\tau \ge 0$, we have

$$\Pr\{Z > 0\} \le Ee^{\tau Z}. \quad \text{(B.4)}$$

We now compute this expectation and optimize with respect to $\tau \ge 0$ to obtain the tightest bound.

$$Ee^{\tau(qV_0)} = \sum_{k=0}^{\infty} \frac{e^{-\Lambda_0}\Lambda_0^k}{k!} e^{\tau q k} = \exp\{-\Lambda_0 + \Lambda_0 e^{\tau q}\}$$

$$Ee^{-\tau(1-q)V_1} = \sum_{k=0}^{\infty} \frac{e^{-\Lambda_1}\Lambda_1^k}{k!} e^{-\tau(1-q)k} = \exp\{-\Lambda_1 + \Lambda_1 e^{-\tau(1-q)}\}.$$

Thus

$$Ee^{\tau Z} = \exp\{-(\Lambda_0 + \Lambda_1) + \Lambda_0 e^{\tau q} + \Lambda_1 e^{-\tau(1-q)}\}. \quad \text{(B.5)}$$

To get the tightest bound, set the derivative of the exponent with respect to τ equal to zero yielding

$$\Lambda_0 q e^{\tau q} - \Lambda_1(1-q)e^{-\tau(1-q)} = 0$$

or

$$e^{\tau} = \frac{\Lambda_1(1-q)}{\Lambda_0 q} = \frac{1+s}{s}. \quad \text{(B.6)}$$

Note that $\tau = \log[(1+s)/s] \ge 0$, as required. Combining (B.6), (B.5), and (B.4) we have

$$\Pr\{Z > 0\} \le \exp\left\{-(\Lambda_0 + \Lambda_1) + \Lambda_0\left(\frac{1+s}{s}\right)^q + \Lambda_1\left(\frac{s}{1+s}\right)^{1-q}\right\} = e^{-\hat{E}T}$$

where $\hat{E}$ is given by (B.3b). Hence the lemma.

It remains to show that $\hat{E}$ is not less than the error exponent $E_1(\rho, q) - \rho R$, where $E_1()$ is given by (1.9). In fact,

$$E_1(\rho, q) - \rho R \overset{1}{\le} E_1(\rho, q) \overset{2}{\le} E_1(1, q) \overset{3}{=} Aq(1-q)(\sqrt{1+s} - \sqrt{s})^2.$$

Step 1 follows from $\rho, R \ge 0$, step 2 from $\partial E_1/\partial\rho \ge 0$ (which follows from [6, theorem 5.6.3], for DMC's), and step 3 by using the same steps as in (3.14). That $\hat{E} \ge E_1(\rho, q) - \rho R$ follows from the following proposition which was proved by Reeds.

Proposition B.2 (Reeds): Let $\hat{E}$ be defined by (B.3b). Let

$$\tilde{E} = Aq(1-q)[\sqrt{1+s} - \sqrt{s}]^2.$$

Then $\hat{E} \ge \tilde{E}$.

Proof: Let $t = (s/1+s)^{1/2}$, so that t increases from 0 to 1 as s increases from 0 to ∞. We have

$$\frac{\hat{E} - \tilde{E}}{A(1+s)} = \frac{q(1+s) + (1-q)s - s^{1-q}(1+s)^q - q(1-q)(\sqrt{1+s} - \sqrt{s})^2}{1+s}$$
$$= q + t^2(1-q) - t^{2(1-q)} - q(1-q)(1-t)^2$$
$$= q^2 + 2q(1-q)t + (1-q)^2 t^2 - t^{2(1-q)}$$
$$= [q + (1-q)t]^2 - t^{2(1-q)}.$$

Therefore, it will suffice to prove that

$$q + (1-q)t \ge t^{(1-q)}, \qquad 0 \le t \le 1.$$

Since $0 \le q \le 1$, the function $f_1(t) = t^{(1-q)}$ is concave and $f_1(1) = 1$, $f_1'(1) = (1-q)$. Thus the graph of $f_1(t)$ versus t lies below its tangent line at $(1, f(1))$: $q + (1-q)t$. Hence the proposition.

References

[1] I. Bar-David and G. Kaplan, "Information rates of photon-limited overlapping pulse position modulation channels," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 455–464, 1984.

[2] M. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 710–715, 1980.

[3] Y. M. Kabanov, "The capacity of a channel of the Poisson type," *Theory Probab. Appl.*, vol. 23, pp. 143–147, 1978.

[4] J. E. Mazo and J. Salz, "On optical data communication via direct detection of light pulses," *Bell Syst. Tech. J.*, vol. 55, pp. 347–369, 1976.

[5] D. L. Snyder and I. B. Rhodes, "Some implications of the cutoff-rate criterion for coded direct-detection optical communication channels," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 327–338, 1980.

[6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[7] H. J. Landau and A. D. Wyner, "Optimum waveform signal sets with amplitude and energy constraints," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 615–622, 1984.

[8] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Part II," *IEEE Trans. Inform. Theory*, this issue, pp. 1462–1471.