Capacity of Pulse-Position Modulation (PPM) on Gaussian and Webb Channels

S. Dolinar,[1] D. Divsalar,[1] J. Hamkins,[1] and F. Pollara[1]

This article computes the capacity of various idealized soft-decision channels modeling an optical channel using an avalanche photodiode detector (APD) and pulse-position modulation (PPM). The capacity of this optical channel depends in a complicated way on the physical parameters of the APD and the constraints imposed by the PPM orthogonal signaling set. This article attempts to identify and separate the effects of several fundamental parameters on the capacity of the APD-detected optical PPM channel. First, an overall signal-to-noise ratio (SNR) parameter is defined such that the capacity as a function of a bit-normalized version of this SNR drops precipitously toward zero at quasi-brick-wall limits on bit SNR that are numerically the same as the well-understood brick-wall limits for the standard additive white Gaussian noise (AWGN) channel. A second parameter is used to quantify the effects on capacity of one unique facet of the optical PPM channel (as compared with the standard AWGN channel) that causes the noise variance to be higher in signal slots than in nonsignal slots. This nonuniform noise variance yields interesting capacity effects even when the channel model is AWGN. A third parameter is used to measure the effects on capacity of the difference between an AWGN model and a non-Gaussian model proposed by Webb (see reference in [2]) for approximating the statistics of the APD-detected optical channel. Finally, a fourth parameter is used to quantify the blending of a Webb model with a pure AWGN model to account for thermal noise.

Numerical results show that the capacity of M-ary orthogonal signaling on the Webb channel exhibits the same brick-wall Shannon limit, $(M \ln 2)/(M-1)$, as on the AWGN channel (≈ -1.59 dB for large M). Results also compare the capacity obtained by hard- and soft-output channels and indicate that soft-output channels offer a 3-dB advantage.

I. Introduction

Our objective in this article is to compute and compare the capacities of various idealized channels that might be used to approximate the optical communication system using M-ary pulse-position modulation (PPM) and an avalanche photodiode (APD) detector: (1) the standard additive white Gaussian noise

[1] Communications Systems and Research Section.

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

channel (here called AWGN-1); (2) a more general AWGN channel (AWGN-2) allowing different variances in signal and noise slots; (3) a Webb-distributed channel (Webb-2) modeling the output of an APD in the absence of thermal noise and surface leakage current; and (4) a blended Webb and Gaussian channel (Webb+Gaussian) channel modeling Gaussian thermal noise and surface leakage current added to Webb-2 channel outputs.

In this article, we analyze in detail the capacities of these idealized channels in terms of certain fundamental channel parameters defined in such a way as to accentuate the similarities between the more complicated optical channel models and the familiar AWGN-1 channel for which capacity limits are well understood. We also compare the capacities achievable with soft- and hard-decision channel outputs. The appeal of using soft decisions lies in the ability to take advantage of better performing codes (e.g., turbo codes), which admit soft decoding algorithms.

Our ultimate goal is to develop an understanding of the role of various optical parameters on the capacity of an optical communication system. A future article will attempt to utilize the conclusions drawn from the idealized models to develop an understanding of the role of various physical optical parameters on the capacity of an optical communication system.

II. Modeling the APD with Webb and Gaussian Approximations

In an optical communication system using an APD detector, the number of absorbed photons, n, is a Poisson random variable with mean $\bar{n}$ related to the total optical power, $P(t)$, and to the observation time, T_s, by $\bar{n} = (\eta/h\nu)\int_0^{T_s} P(t)dt$, where η is the detector's quantum efficiency and $h\nu$ is the photon energy. In response to n absorbed photons, the APD generates an "avalanche" of q electrons with a complicated conditional probability distribution derived by McIntyre (see reference in [2]), and the probability mass function $p(q)$ is obtained by averaging this conditional probability over the Poisson-distributed n. Alternatively, $p(q)$ can be approximated by a simpler continuous probability density derived by Webb (see reference in [2]). Furthermore, it is known [2] that the Webb density is well approximated by a Gaussian away from its tails, and that the approximation accuracy improves as $\bar{n}$ gets large.

We begin by briefly discussing the McIntyre distribution and the various approximations to it that form the basis of the idealized channels to be studied in this article.

A. McIntyre–Conradi Model of the APD

The actual number of photons absorbed is a Poisson-distributed random variable, n, with $p(n) = (\bar{n}^n/n!)e^{-\bar{n}}$. The conditional probability $p(q|n)$, $q \geq n, n \geq 0$, that the APD generates q electrons in response to n absorbed photons was derived by McIntyre and Conradi and can be found in [11]. This conditional probability distribution depends on several parameters of the APD detector, including its gain, G, and ionization ratio, k_{eff}. See Appendix A for physical quantities and notation. The McIntyre–Conradi probability density function of q, given an average $\bar{n}$ absorbed photons, is obtained by averaging the conditional probability distribution $p(q|n)$ over the Poisson-distributed n:

$$p_M(q) = \sum_{n=0}^{q} p(q|n)\frac{\bar{n}^n}{n!}e^{-\bar{n}}, \quad q \geq 0 \tag{1}$$

The full APD output also contains additive Gaussian noise contributed by thermal noise and surface leakage current. Thus, the total electron count $x = q + q'$ at the output of the APD has a probability density function that is a convolution of the McIntyre–Conradi density, Eq. (1), for q with a Gaussian probability density function for q':

$$p(x) = \sum_{q=0}^{\infty} \frac{1}{\sigma'} \phi\left(\frac{x - q - m'}{\sigma'}\right) p_M(q) \tag{2}$$

where $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ denotes the standardized (zero-mean, unit-variance) Gaussian probability density, and where [2] the mean m' and variance σ'^2 of the convolving Gaussian density depend on the surface leakage current, I_s, the thermal noise temperature, T, and other physical parameters. Again, see Appendix A.

$$m' = \frac{I_s T_s}{e_-}$$

$$\sigma'^2 = 2BT_s\left(\frac{I_s T_s}{e_-} + \kappa T \frac{2T_s}{R_L e_-^2}\right) \tag{3}$$

B. The Webb Approximation

The McIntyre–Conradi distribution $p_M(q)$ in Eq. (1) can be approximated by a continuous probability density, $p_W(q)$, derived by Webb (see reference in [2]):

$$p_W(q) = \frac{1}{\sqrt{2\pi\bar{n}G^2F}}\left(1 + \frac{(q - G\bar{n})(F-1)}{GF\bar{n}}\right)^{-3/2} \exp\left(\frac{-(q-G\bar{n})^2}{2\bar{n}G^2F\left(1 + \frac{(q-G\bar{n})(F-1)}{GF\bar{n}}\right)}\right), \quad q > \frac{-G\bar{n}}{F-1} \tag{4}$$

where G is the APD gain and $F = k_{eff}G + (2 - 1/G)(1 - k_{eff})$ is an excess noise factor, given in terms of the gain and the ionization ratio, k_{eff}.

The Webb-distributed electron count, q, is conveniently represented in terms of a standardized, scaled-and-translated Webb random variable, w. Defining $q = m + w\sigma$, where $m = G\bar{n}$ and $\sigma = \sqrt{\bar{n}G^2F}$, the probability density for the standardized Webb random variable w simplifies to

$$\phi(w; \delta^2) = \frac{1}{\sqrt{2\pi}}\left(1 + \frac{w}{\delta}\right)^{-3/2} e^{-w^2/[2(1+w/\delta)]}, \quad w > -\delta \tag{5}$$

where $\delta^2 = \bar{n}F/(F-1)^2$. The standardized Webb probability density $\phi(w; \delta^2)$ is plotted in Fig. 1 for various values of the parameter δ^2, which determines the skewness of the density function around the origin. Note that this standardized Webb probability reduces exactly to a standardized Gaussian probability density when the parameter $\delta^2 \to \infty$. In other words, the standardized Gaussian density $\phi(x)$ can be regarded as the limiting case $\phi(x; \infty)$ of the family of standardized Webb densities $\phi(w; \delta^2)$.

If w is a standardized Webb random variable with skewness parameter δ^2 and probability density given by Eq. (5), then $q = m + w\sigma$ is a Webb random variable with mean m, variance σ^2, and skewness δ^2, and is denoted as $W(m, \sigma^2, \delta^2)$. The standardized Webb random variable w has zero mean and unit variance and can be denoted as $W(0, 1, \delta^2)$; higher moments of the standardized Webb probability density can be computed as polynomials in the parameter $1/\delta^2$, as shown in Appendix B. The Webb random variable q in Eq. (4), approximating the APD output without thermal noise and surface leakage current, is $W(m, \sigma^2, \delta^2)$ with $m = G\bar{n}$, $\sigma^2 = \bar{n}G^2F$, and $\delta^2 = \bar{n}F/(F-1)^2$.

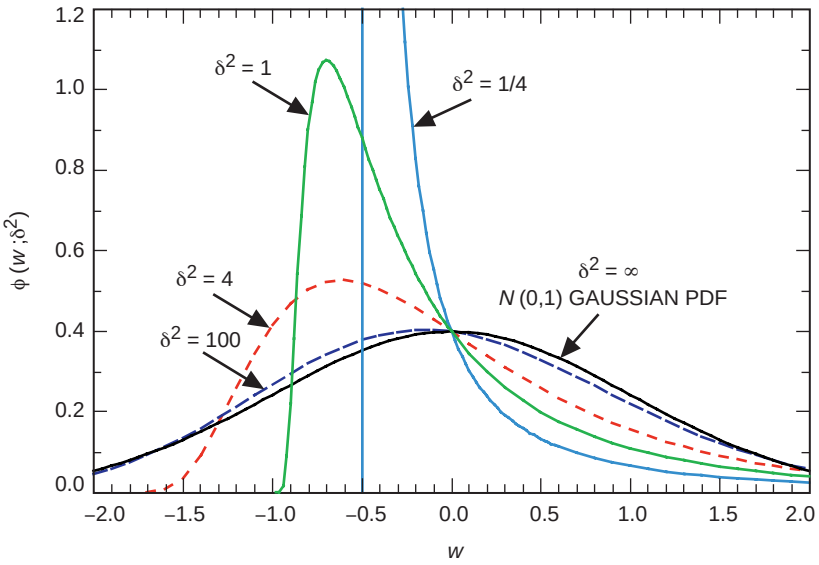

Fig. 1. Family of standardized Webb probability density functions [see Eq. (5)].

C. The Webb+Gaussian Approximation

With the Webb approximation to the McIntyre–Conradi random variable q, the full output $x = q + q'$ of the APD is then modeled as a sum of an electrical current due to the Webb-distributed electron count q, plus an independent Gaussian count q' due to thermal and surface leakage currents. It is convenient to define a standardized, zero-mean, unit-variance Webb+Gaussian random variable, $WG(0, 1, \delta^2, \beta) = \sqrt{\beta}\, W(0, 1, \delta^2) + \sqrt{1-\beta}\, N(0, 1)$, that represents a blending of a pure Webb random variable, $W(0, 1, \delta^2)$, with a pure Gaussian random variable, $N(0, 1)$. The blending parameter β measures the proportion of the total variance contributed by the Webb component. The standardized Webb+Gaussian random variable $WG(0, 1, \delta^2, \beta)$ has a probability density $\phi(x; \delta^2, \beta)$ obtained as a convolution of scaled Webb and Gaussian densities:

$$\phi(x; \delta^2, \beta) = \int_{-\delta}^{\infty} \frac{1}{\sqrt{\beta}}\, \phi\left(\frac{w}{\sqrt{\beta}}; \delta^2\right) \frac{1}{\sqrt{1-\beta}}\, \phi\left(\frac{x-w}{\sqrt{1-\beta}}\right) dw \tag{6}$$

The standardized Webb+Gaussian probability density $\phi(x; \delta^2; \beta)$ is plotted in Fig. 2 for $\delta^2 = 1$ and various values of the blending parameter β. We note that the pure Webb and pure Gaussian densities can be obtained as special cases of the Webb+Gaussian density: $\phi(w; \delta^2) = \phi(w; \delta^2, 1)$, $\phi(x) = \phi(x; \delta^2, 0)$ for any δ^2, and $\phi(x) = \phi(x; \infty, \beta)$ for any β.

The full output $x = q + q'$ of the APD is written in terms of the standardized Webb+Gaussian random variable as $x = m + m' + \sqrt{\sigma^2 + \sigma'^2}\, WG(0, 1, \delta^2, \beta)$, where $\beta = \sigma^2/(\sigma^2 + \sigma'^2)$, and m' and σ'^2 are given by Eq. (3), as before. The probability density $p(x)$ for the full APD output is then

$$p(x) = \frac{1}{\sqrt{\sigma^2 + \sigma'^2}}\, \phi\left(\frac{x - m - m'}{\sqrt{\sigma^2 + \sigma'^2}}; \delta^2, \beta\right)$$

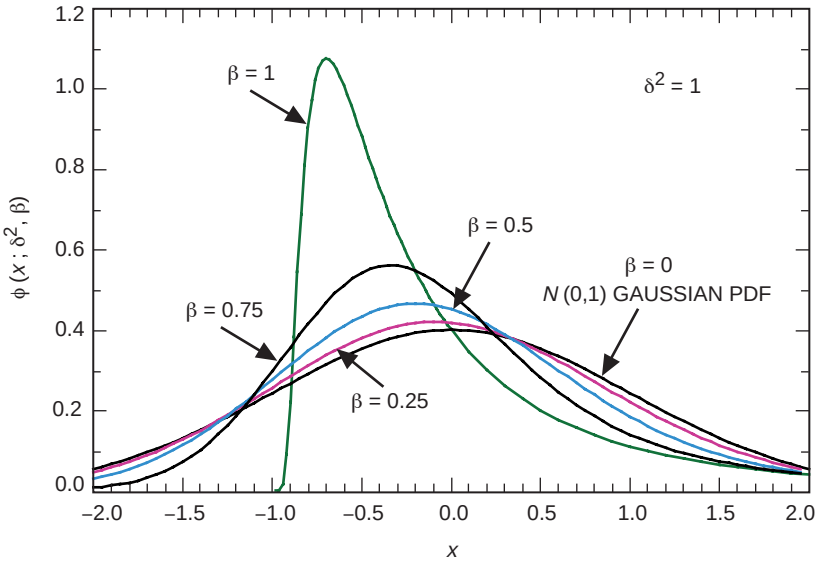

Fig. 2. Family of standardized Webb+Gaussian probability density functions [see Eq. (6)].

D. The AWGN Approximation

When the skewness parameter δ^2 is large, the Webb random variable q in Eq. (4) can be approximated by a Gaussian random variable $N(m, \sigma^2)$ with the same mean, $m = G\bar{n}$, and variance, $\sigma^2 = \bar{n}G^2F$. The corresponding Gaussian probability density for q is

$$p_G(q) \approx \frac{1}{\sigma}\,\phi\left(\frac{x-m}{\sigma}\right) \tag{7}$$

Furthermore, the full APD output $x = q + q'$ is also Gaussian, $N(m+m', \sigma^2+\sigma'^2)$, and Eq. (2) simplifies to a single Gaussian probability density function:

$$p(x) \approx \frac{1}{\sqrt{\sigma^2+\sigma'^2}}\,\phi\left(\frac{x-m-m'}{\sqrt{\sigma^2+\sigma'^2}}\right) \tag{8}$$

where [2]

$$m + m' = G\bar{n} + \frac{I_s T_s}{e_-}$$

$$\sigma^2 + \sigma'^2 = G^2F\bar{n} + 2BT_s\left(\frac{I_s T_s}{e_-} + \kappa T\frac{2T_s}{Re_-^2}\right)$$

E. The APD Models with PPM Signaling

For PPM signaling, the mean number of photons absorbed in a given PPM slot depends on whether the signal is present or absent in that slot. The average number of photons, $\bar{n}$, absorbed by an APD during a PPM slot of duration T_s is given by

$$\bar{n} = \frac{\eta}{h\nu}\int_0^{T_s} P(t)dt = \begin{cases} \bar{n}_1 \approx \eta(\bar{n}_b + \bar{n}_s) & \text{signal present} \\ \bar{n}_0 \approx \eta\bar{n}_b & \text{signal absent} \end{cases} \tag{9}$$

where $\bar{n}_b$ and $\bar{n}_s$, are, respectively, the average number of background photons incident on the detector per PPM slot and the average number of signal photons incident on the detector per signal pulse. Here the approximations in relating $\bar{n}_1$ and $\bar{n}_0$ to the physical parameters ignore small contributions from bulk leakage current and incomplete modulation extinction. See Appendix A for physical quantities and notation. A more complete analysis of the dependence of the optical capacity on the physical parameters will be given in a future article. In the current article, the general capacity expressions are derived without explicitly relating $\bar{n}_1$ and $\bar{n}_0$ to the physical quantities, except for relying on the fact that $\bar{n}_1 \geq \bar{n}_0$. The approximations in Eq. (9) are only used later to relate our numerical evaluations of the general capacity expressions to realistic physical parameter values.

The Webb model for PPM signaling (here called Webb-2) uses the density in Eq. (4) with two different values of $\bar{n}$: once with the average number $\bar{n}_1$ of photons in the signal slot and a second time with the average number $\bar{n}_0$ of photons in the $M-1$ non-signal slots. Similarly, the AWGN approximation for PPM signaling (here called AWGN-2) uses the density in Eq. (7) or Eq. (8) twice, with $\bar{n} = \bar{n}_1$ in the signal slot and $\bar{n} = \bar{n}_0$ in the $M-1$ nonsignal slots.

Since the mean m and variance σ^2 of the Webb-distributed electron count q are both proportional to $\bar{n}$, and $\bar{n}_1 \geq \bar{n}_0$, it follows that the mean m_1 and the variance σ_1^2 of q in the signal slots are no smaller than the corresponding mean m_0 and variance σ_0^2 in the nonsignal slots. The additional mean m' and the additional variance σ'^2 contributed by thermal noise and surface leakage current is the same for both signal and nonsignal slots. Just as a single Webb density with variance σ^2 approaches a Gaussian density with the same variance when $\delta^2 \to \infty$, the two different Webb densities characterizing the signal and nonsignal slots can each be approximated by Gaussian densities if their individual skewness parameters δ_1^2 and δ_0^2 become large. Thus, a Gaussian approximation for the optical PPM channel leads to a somewhat more general Gaussian channel than the familiar additive white Gaussian noise (AWGN) channel. In this article, we examine two separate versions of the Gaussian problem, the familiar Gaussian model with a common variance in all of the slots (here called AWGN-1) and also a "double Gaussian" model (AWGN-2) with greater variance in the signal slot than the common variance in the nonsignal slots.

F. Fundamental Parameter Definitions for PPM Capacity Calculations

In our PPM capacity derivations in the next section for the various idealized channel models, we find that the following four fundamental parameters give useful insights into the capacity limits derived for these models. A signal-to-noise ratio (SNR) parameter, $\rho_0 = (m_1 - m_0)^2/(\sigma_0^2 + \sigma'^2)$, measuring the ratio of the energy in the difference of the mean counts in the signal and nonsignal slots to the variance in the nonsignal slots, plays a role similar to that of the channel symbol SNR, E_s/N_0, for the standard AWGN-1 channel. Next, a second SNR parameter, $\rho_+ = (m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$, measured with reference to the *excess* noise variance in the signal slots above that in the nonsignal slots, determines how closely the AWGN-2 model resembles the classic AWGN-1 model. A third parameter, $\Delta = \delta_1^2 - \delta_0^2$, measures the degree to which the Webb-2 model differs from the AWGN-2 model. A fourth parameter, $\beta_0 = \sigma_0^2/(\sigma_0^2 + \sigma'^2)$, measures the blending of the Webb and Gaussian components in the nonsignal slots. Finally, a derived parameter, ρ_b, obtained from ρ_0 and the computed capacity, plays a role similar to that of the bit SNR, E_b/N_0, for the standard AWGN-1 channel.

The usefulness of these parameter definitions is that they tend to isolate three types of different effects on capacity due to using the more complicated Webb-2 model versus the familiar AWGN-1 model: (1) the normal monotonic variation of capacity with the signal-to-noise ratio supplied to the channel; (2) the sometimes strange effects on capacity that have nothing to do with the Webb density but are solely due to the unequal variances in the signal and nonsignal slots; and (3) effects on capacity when

the Webb probability density is not adequately approximated by a Gaussian, or when the total density blends Webb and Gaussian components.

By plugging in $\bar{n}$ for the signal and nonsignal slots from Eq. (9), we find that the three parameters for a pure Webb-2 channel, modeling the APD output without thermal noise and surface leakage current, are given in terms of the physical optical system parameters by

$$\left.\begin{aligned} \rho_0 &= \frac{(m_1 - m_0)^2}{\sigma_0^2} = \frac{(\bar{n}_1 - \bar{n}_0)^2}{\bar{n}_0 F} \approx \frac{\eta \bar{n}_s^2}{\bar{n}_b F} \\ \rho_+ &= \frac{(m_1 - m_0)^2}{\sigma_1^2 - \sigma_0^2} = \frac{\bar{n}_1 - \bar{n}_0}{F} \approx \frac{\eta \bar{n}_s}{F} \\ \Delta &= \delta_1^2 - \delta_0^2 = \frac{(\bar{n}_1 - \bar{n}_0)F}{(F-1)^2} \approx \frac{\eta \bar{n}_s F}{(F-1)^2} \end{aligned}\right\} \tag{10}$$

We note from these equations that, for the optical channel, the parameter Δ, measuring the deviation of a Webb density from a Gaussian approximation, and the parameter ρ_+, measuring the effects of unequal variances, are different by a factor depending only on the excess noise factor F, i.e., $\Delta = \rho_+ F^2/(F-1)^2$. Furthermore, the fundamental SNR parameter ρ_0 is approximately equal to ρ_+ times a factor $\bar{n}_s/\bar{n}_b$ that is often routinely considered to be the proper SNR measure for the optical channel. Because the variance σ^2 and the skewness parameter δ^2 on the optical APD channel are each linearly proportional to the number of absorbed photons, this imposes an additional constraint that

$$\frac{\sigma_0^2}{\sigma_1^2} = \frac{\delta_0^2}{\delta_1^2} = \frac{\bar{n}_0}{\bar{n}_1} = \frac{\rho_+}{\rho_0 + \rho_+} \tag{11}$$

Finally, this allows us to express the Webb-2 skewness parameters δ_0^2 and δ_1^2 in terms of the fundamental capacity parameters ρ_0, ρ_+, and Δ as

$$\left.\begin{aligned} \delta_0^2 &= \frac{\rho_+}{\rho_0}\Delta \\ \delta_1^2 &= \frac{\rho_0 + \rho_+}{\rho_0}\Delta \end{aligned}\right\} \tag{12}$$

For the Webb+Gaussian channel, modeling the full APD output, including the effects of thermal noise and surface leakage current, we find that the fundamental parameters ρ_+ and Δ are exactly as given above, while the definition of the primary SNR parameter ρ_0 takes into account the additional common variance σ'^2 contributed by the Gaussian thermal noise and surface leakage current, and the blending parameter β_0 measures the proportion of the total variance contributed by the Webb component:

$$\left.\begin{aligned} \rho_0 &= \frac{(m_1 - m_0)^2}{\sigma_0^2 + \sigma'^2} = \frac{\beta_0(\bar{n}_1 - \bar{n}_0)^2}{\bar{n}_0 F} \\ \beta_0 &= \frac{\sigma_0^2}{\sigma_0^2 + \sigma'^2} = \frac{1}{1 + \dfrac{2BT_s}{G^2 F \bar{n}_0}\left(\dfrac{I_s T_s}{e_-} + \kappa T \dfrac{2T_s}{R e_-^2}\right)} \end{aligned}\right\} \tag{13}$$

The parameter relationships noted in Eq. (11) for the pure Webb-2 channel remain valid for the Webb+Gaussian channel if ρ_0 is replaced by ρ_0/β_0. In this case, the analog to Eq. (11) is

$$\frac{\sigma_0^2}{\sigma_1^2} = \frac{\delta_0^2}{\delta_1^2} = \frac{\beta_0}{\beta_1} = \frac{\bar{n}_0}{\bar{n}_1} = \frac{\beta_0\rho_+}{\rho_0 + \beta_0\rho_+} \tag{14}$$

and we can express δ_0^2, δ_1^2, and β_1 in terms of the other parameters as

$$\left.\begin{aligned} \delta_0^2 &= \frac{\beta_0\rho_+}{\rho_0}\Delta \\ \delta_1^2 &= \frac{\rho_0 + \beta_0\rho_+}{\rho_0}\Delta \\ \beta_1 &= \beta_0 + \frac{\rho_0}{\rho_+} \end{aligned}\right\} \tag{15}$$

When the AWGN-2 model is used as in Eq. (7) to model the pure Webb-2 channel, then the parameters for this channel are ρ_0 and ρ_+, defined by Eq. (10). When the AWGN-2 model is used as in Eq. (8) to model the full APD output, the parameters are ρ_0, given by Eq. (13), and ρ_+, given by Eq. (10).

Notice that the gain, G, of the APD does not enter into the definitions of the parameters ρ_+ or Δ or the primary SNR parameter ρ_0 for the pure Webb-2 model. The APD gain G only enters into the relative weighting of the σ_0^2 and the σ'^2 variances (i.e., the parameter β_0) in the definition of ρ_0 for the Webb+Gaussian model (or for the AWGN-2 model that approximates the Webb+Gaussian model).

III. Capacity of *M*-ary PPM on Channels with Soft Outputs

For each channel model, we consider the communication system shown in Fig. 3. The output $\mathbf{U} = (U_1, U_2, \cdots, U_k)$ of a k-bit source is modulated with $(M = 2^k)$-ary PPM to yield a signal $\mathbf{X} = (X_1, X_2, \cdots, X_M)$. A PPM modulator is equivalent to an encoder producing the $M = 2^k$ codewords of a $(2^k, k)$ orthogonal code. Thus, the capacity of M-PPM is the same as that of an orthogonal signal set with M codewords, and of M-ary frequency shift keying (M-FSK).

Since $\mathbf{X}$ is an invertible function of $\mathbf{U}$, the observation of $\mathbf{Y}$ provides an average $I(\mathbf{Y};\mathbf{X})$ bits of information about the input $\mathbf{U}$, where $I(\mathbf{Y};\mathbf{X})$ is the mutual information between $\mathbf{Y}$ and $\mathbf{X}$. The capacity of PPM modulation on the channel is the maximum amount of information that can be transmitted reliably and is given by $C = \max_{\mathbf{p}(\mathbf{X})} I(\mathbf{Y};\mathbf{X})$. Because of the symmetry of orthogonal signals and of the channels considered, capacity is achieved with an equiprobable M-ary source distribution $p(\mathbf{X} = \mathbf{x}) = 1/M$, $\forall \mathbf{x} \in \{\mathbf{x_1}, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$.

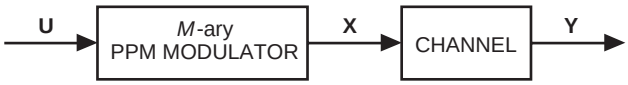

Fig. 3. Model for PPM capacity calculations.

The channel capacity with input signals restricted to an M-ary orthogonal constellation $\mathcal{C}$, and no restriction on the channel output, is given by

$$C = \frac{1}{M} \sum_{\mathbf{x} \in \mathcal{C}} \int_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \log_2 \left(\frac{p(\mathbf{y}|\mathbf{x})}{\frac{1}{M} \sum_{\hat{\mathbf{x}} \in \mathcal{C}} p(\mathbf{y}|\hat{\mathbf{x}})} \right) d\mathbf{y} \tag{16}$$

where $\mathbf{y} = (y_1, \cdots, y_M)$ is the received vector. Because of the symmetry of orthogonal codes with respect to the inputs $\{\mathbf{x}_1, \cdots, \mathbf{x}_M\}$, this expression reduces to

$$\begin{aligned} C &= \int_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}_1) \log_2 \left(\frac{p(\mathbf{y}|\mathbf{x}_1)}{\frac{1}{M} \sum_{\hat{\mathbf{x}} \in \mathcal{C}} p(\mathbf{y}|\hat{\mathbf{x}})} \right) d\mathbf{y} \\ &= \log_2 M - E_{\mathbf{y}|\mathbf{x_1}} \log_2 \left[\frac{\sum_{j=1}^{M} p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_1)} \right] \qquad (17) \\ &= \log_2 M - E_{\mathbf{v}|\mathbf{x_1}} \log_2 \left[\frac{\sum_{j=1}^{M} p(\mathbf{v}|\mathbf{x}_j)}{p(\mathbf{v}|\mathbf{x}_1)} \right] \qquad (18) \end{aligned}$$

where $\mathbf{v}$ is a random vector obtained from $\mathbf{y}$ via an arbitrary invertible transformation.

The capacity formula in Eq. (17) can be computed in principle, as long as $p(\mathbf{y}|\mathbf{x}_j)$ is known for all $\mathbf{x}_j$. However, uninspired computation of the expectation of the right side of Eq. (17) requires evaluation of an M-dimensional integral. Fortunately, this M-dimensional expectation can be accurately estimated via Monte Carlo simulation for all of the channels considered. The method is simply to generate pseudorandom M-vectors $\mathbf{y}$ according to the probability density, $p(\mathbf{y}|\mathbf{x}_1)$, then to evaluate the logarithmic function inside the expectation for each sample $\mathbf{y}$ thus generated, and finally to average the computed logarithm over many pseudorandom samples. This Monte Carlo method was the most practical method to calculate numerical capacities for all of the channels considered below, whenever M was larger than about 4.

The subsections below give the explicit forms of the channel transition probability function $p(\mathbf{y}|\mathbf{x}_j)$ for each of the channels considered, along with some channel-specific simplifications in the capacity formulas.

A. Capacity of *M*-ary PPM on the Standard AWGN Channel (AWGN-1)

In this case, the possible signals $\mathbf{X}$ are of the form $\mathbf{x}_j = (x_{j1}, \cdots, x_{jM}) = (0, 0, 0, \cdots, 0, m, 0, \cdots, 0)$, where the nonzero signal value m is in position j. The transmitted vector $\mathbf{x}$ is corrupted by additive white Gaussian noise with zero mean and variance σ^2 in each component. Thus, given a transmitted signal $\mathbf{x} = \mathbf{x}_j$, the components of the received vector $\mathbf{y}$ are conditionally independent Gaussian random variables, identically distributed except for y_j:

$$\left.\begin{array}{llll} y_j & \text{is} & N(m, \sigma^2) & \text{(signal present)} \\ y_i & \text{is} & N(0, \sigma^2), \quad i \neq j & \text{(signal absent)} \end{array}\right\} \tag{19}$$

This is the model for any set of M-ary orthogonal signals, with energy per M-dimensional symbol $E_s = m^2$, transmitted on an AWGN channel with two-sided noise spectral density $N_0/2 = \sigma^2$. Note that,

in the case of PPM signaling, E_s refers to the energy per PPM symbol, not per PPM slot. A symbol signal-to-noise ratio can be defined by $\rho = m^2/\sigma^2 = 2E_s/N_0$.[2]

For the distributions in Eq. (19), we have

$$p(\mathbf{y}|\mathbf{x}_j) = \left(\frac{1}{2\pi\sigma^2}\right)^{M/2} \exp\left[-\frac{(y_j - m)^2}{\sigma^2}\right] \prod_{\substack{i=1 \\ i\neq j}}^{M} \exp\left[-\frac{y_i^2}{2\sigma^2}\right] \tag{20}$$

Defining $v_j = y_j/\sigma$, we obtain

$$\frac{p(\mathbf{v}|\mathbf{x}_j)}{p(\mathbf{v}|\mathbf{x}_1)} = \exp\left[\frac{1}{2}\left[v_j^2 - (v_j - \sqrt{\rho})^2 - v_1^2 + (v_1 - \sqrt{\rho})^2\right]\right] \tag{21}$$

The random variables $\{v_j\}$ are conditionally independent, given $\{\mathbf{x_1}\}$, and distributed as

$$\left.\begin{array}{llll} v_1 & \text{is} & N(\sqrt{\rho}, 1) & \\ v_j & \text{is} & N(0,1), & j \neq 1 \end{array}\right\} \tag{22}$$

Using Eq. (21) in Eq. (18), we get an expression for the capacity of orthogonal signaling on the AWGN-1 channel:

$$C(\rho) = \log_2 M - E_{\mathbf{v}|\mathbf{x_1}} \log_2 \sum_{j=1}^{M} \exp\left[\sqrt{\rho}(v_j - v_1)\right] \tag{23}$$

For M-PPM on the AWGN-1 channel, both the capacity in Eq. (23) and the uncoded probability of symbol error (see Section V) are functions of the single parameter $\rho = m^2/\sigma^2$, not of m and σ separately. This is a statement of the well-known fact that the AWGN-1 channel is fully characterized by its SNR and does not depend on an overall scaling of the signal and noise.

For large M, the law of large numbers allows us to approximately replace the last $M-1$ terms in Eq. (23) by $M-1$ times their mean (see derivation in Appendix C). Then the AWGN capacity can be approximated as[3]

$$C(\rho) \approx \log_2 M - E_{v_1|\mathbf{x_1}} \log_2 \left[1 + (M-1)e^{\rho/2} e^{-\sqrt{\rho} v_1}\right] \tag{24}$$

where v_1 is $N(\sqrt{\rho}, 1)$ given $\mathbf{x_1}$.

B. Capacity of *M*-ary PPM on a More General Gaussian Channel (AWGN-2)

Now we extend the analysis to cover a "double Gaussian" problem (here called AWGN-2) related more directly to the PPM optical model and characterized by different means and variances depending on whether the signal is present or absent.

---

[2] Note that ρ defined in this manner is 3 dB higher than the usual symbol SNR defined as E_s/N_0.

[3] A similar result for noncoherent M-FSK was derived in [1].

Given a transmitted signal $\mathbf{x} = \mathbf{x}_j$, the components of the received vector $\mathbf{y}$ are conditionally independent Gaussian random variables, identically distributed except for y_j:

$$\left.\begin{array}{llll} y_j & \text{is} & N(m_1, \sigma_1^2) & \text{(signal present)} \\ y_i & \text{is} & N(m_0, \sigma_0^2), \quad i \neq j & \text{(signal absent)} \end{array}\right\} \tag{25}$$

where $m_1 > m_0$ and $\sigma_1^2 > \sigma_0^2$. By symmetry of the orthogonal PPM signal constellation, capacity can be evaluated by Eq. (17). For the distribution in Eq. (25), we have

$$p(\mathbf{y}|\mathbf{x}_j) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{(M-1)/2} \left(\frac{1}{2\pi\sigma_1^2}\right)^{1/2} \exp\left[-\frac{(y_j - m_1)^2}{2\sigma_1^2}\right] \prod_{\substack{i=1 \\ i\neq j}}^{M} \exp\left[-\frac{(y_i - m_0)^2}{2\sigma_0^2}\right] \tag{26}$$

Thus,

$$\frac{p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_1)} = \exp\left[\frac{v_j^2 - u_j^2 - v_1^2 + u_1^2}{2}\right] \tag{27}$$

where $v_j = (y_j - m_0)/\sigma_0$ and $u_j = (y_j - m_1)/\sigma_1$.

We define a symbol SNR parameter, $\rho_0 = (m_1 - m_0)^2/\sigma_0^2$, analogously to the SNR parameter ρ for the AWGN-1 channel, where the reference noise level in the AWGN-2 case is the variance σ_0^2 in the noise slots. We also define a second SNR parameter, $\rho_+ = (m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$, where the reference noise level is the *excess* noise variance in the signal slot compared with that in the noise slots (guaranteed to be nonnegative since $\sigma_1^2 \geq \sigma_0^2$). The AWGN-2 problem reduces to the AWGN-1 problem in the limit as this second SNR parameter $\rho_+ \to \infty$. With these definitions, the $\{v_j\}$ are conditionally independent given $\mathbf{x}_1$, and distributed as

$$\left.\begin{array}{llll} v_1 & \text{is} & N\left(\sqrt{\rho_0},\ \dfrac{\rho_0 + \rho_+}{\rho_+}\right) & \\ v_j & \text{is} & N(0,1), \quad j \neq 1 & \end{array}\right\} \tag{28}$$

In terms of the $\{v_j\}$, the $\{u_j\}$ are determined by the invertible transformation:

$$u_j = \sqrt{\frac{\rho_+}{\rho_0 + \rho_+}}\,(v_j - \sqrt{\rho_0}) \tag{29}$$

Thus, we have

$$u_j \pm v_j = v_j \left(\sqrt{\frac{\rho_+}{\rho_0 + \rho_+}} \pm 1\right) - \sqrt{\frac{\rho_0 \rho_+}{\rho_0 + \rho_+}} \tag{30}$$

Noting that $u_j^2 - v_j^2 = (u_j + v_j)(u_j - v_j)$, we use the expressions for $u_j \pm v_j$ to calculate the exponent in Eq. (27), and obtain the capacity from Eq. (18) in the form

$$C(\rho_0, \rho_+) = \log_2 M - E_{\mathbf{v}|\mathbf{x}_1} \log_2 \sum_{j=1}^{M} \exp\left[\frac{\rho_+}{\rho_0+\rho_+}\sqrt{\rho_0}(v_j - v_1) + \frac{\rho_0}{\rho_0+\rho_+}\ \frac{v_j^2 - v_1^2}{2}\right] \tag{31}$$

This equation reduces to the standard AWGN-1 capacity for orthogonal signals, Eq. (23), in the limit as $\rho_+ \to \infty$ for a given ρ_0.

For M-PPM on an AWGN-2 channel, both the capacity in Eq. (31) and the uncoded probability of symbol error (see Section V) are functions of the parameters $\rho_0 = (m_1 - m_0)^2/\sigma_0^2$ and $\rho_+ = (m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$. This is a statement that the AWGN-2 channel is fully characterized by just two parameters, an overall SNR parameter, ρ_0, and an additional SNR parameter, ρ_+, measured relative to the *excess* variance in the signal slots versus that in the nonsignal slots. The four individual parameters m_0, m_1, σ_0^2, and σ_1^2 defining the AWGN-2 channel do not affect the capacity independently. The reduction from four to two parameters in this case implies that the capacity is not affected by either an overall scale factor or a common displacement of the means.

In the limit as $\rho_+ \to \infty$ for a given ρ_0, the problem reduces to the familiar AWGN-1 channel. At the other extreme, in the limit as $\rho_+/\rho_0 \to 0$, (i.e., the average background count, $\bar{n}_b$, is insignificant compared with the average signal count, $\bar{n}_s$), the conclusions resemble those derived for a pure Poisson quantum-limited model, and there is no fundamental lower limit on the ratio of signal energy to the variance contributed by the signal alone.

For large M, the AWGN-2 capacity can be approximated as (see derivation in Appendix C)

$$C(\rho_0, \rho_+) = \log_2 M - E_{v_1|\mathbf{x}_1} \log_2 \left[1 + (M-1)\sqrt{\frac{\rho_0+\rho_+}{\rho_+}} \exp\left(\frac{\rho_0\rho_+ - 2\rho_+\sqrt{\rho_0}v_1 - \rho_0 v_1^2}{2\,(\rho_0+\rho_+)}\right)\right]$$

where v_1 is $N\left(\sqrt{\rho_0}, [\rho_0+\rho_+]/\rho_+\right)$ given $\mathbf{x}_1$.

C. Capacity of *M*-ary PPM on Webb-Distributed Channels (Webb-1 and Webb-2)

The Webb-1 channel model simply substitutes Webb random variables $W(\cdot,\cdot,\cdot)$ for the Gaussian random variables $N(\cdot,\cdot)$ in Eq. (19) for the AWGN-1 channel model. Given a transmitted signal $\mathbf{x} = \mathbf{x}_j$, the components of the received vector $\mathbf{y}$ are conditionally independent Webb random variables, identically distributed except for y_j:

$$\left.\begin{array}{llll} y_j & \text{is} & W(m, \sigma^2, \delta^2) & \text{(signal present)} \\ y_i & \text{is} & W(0, \sigma^2, \delta^2), \quad i \neq j & \text{(signal absent)} \end{array}\right\} \tag{32}$$

That is, the conditional probability density functions are

$$p(y_j|\mathbf{x}_j) = \frac{1}{\sigma}\ \phi\left(\frac{y_j - m}{\sigma}; \delta^2\right)$$

$$p(y_i|\mathbf{x}_j) = \frac{1}{\sigma}\ \phi\left(\frac{y_i}{\sigma}; \delta^2\right),\ i \neq j$$

where $\phi(\cdot;\cdot)$ is given in Eq. (5). Thus, for the Webb-1 model,

$$\frac{p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_1)} = \frac{\sigma^{-M}\ \phi\left(\frac{y_j - m}{\sigma}; \delta^2\right) \prod_{\substack{i=1 \\ i\neq j}}^{M} \phi\left(\frac{y_i}{\sigma}; \delta^2\right)}{\sigma^{-M}\ \phi\left(\frac{y_1 - m}{\sigma}; \delta^2\right) \prod_{i=2}^{M} \phi\left(\frac{y_i}{\sigma}; \delta^2\right)} = \frac{\phi(v_j - \sqrt{\rho}; \delta^2)\phi(v_1; \delta^2)}{\phi(v_1 - \sqrt{\rho}; \delta^2)\phi(v_j; \delta^2)} \tag{33}$$

where, as in the AWGN-1 channel, $v_j = y_j/\sigma$ and $\rho = m^2/\sigma^2$. These variables $\{v_j\}$ are conditionally independent, given $\{\mathbf{x_1}\}$, and distributed as

$$\left.\begin{array}{lll} v_1 & \text{is} & W(\sqrt{\rho}, 1, \delta^2) \\ v_j & \text{is} & W(0, 1, \delta^2), \quad j \neq 1 \end{array}\right\} \tag{34}$$

The capacity of the Webb-1 channel is given by plugging into Eq. (18):

$$C(\rho, \delta) = \log_2 M - E_{\mathbf{v}|\mathbf{x_1}} \log_2 \sum_{j=1}^{M} \frac{\phi\left(v_j - \sqrt{\rho}; \delta^2\right)\ \phi\left(v_1; \delta^2\right)}{\phi\left(v_1 - \sqrt{\rho}; \delta^2\right)\ \phi\left(v_j; \delta^2\right)} \tag{35}$$

The Webb-2 channel model substitutes Webb random variables $W(\cdot, \cdot, \cdot)$ for the Gaussian random variables $N(\cdot, \cdot)$ in Eq. (25) for the AWGN-2 channel model. Given a transmitted signal $\mathbf{x} = \mathbf{x}_j$, the components of the received vector $\mathbf{y}$ are conditionally independent Webb random variables, identically distributed except for y_j:

$$\left.\begin{array}{lllll} y_j & \text{is} & W(m_1, \sigma_1^2, \delta_1^2) & & \text{(signal present)} \\ y_i & \text{is} & W(m_0, \sigma_0^2, \delta_0^2), & i \neq j & \text{(signal absent)} \end{array}\right\} \tag{36}$$

The conditional probability density functions are

$$p(y_j|\mathbf{x}_j) = \frac{1}{\sigma_1}\ \phi\left(\frac{y_j - m_1}{\sigma_1}; \delta_1^2\right)$$

$$p(y_i|\mathbf{x}_j) = \frac{1}{\sigma_0}\ \phi\left(\frac{y_i - m_0}{\sigma_0}; \delta_0^2\right), i \neq j$$

where $\phi(\cdot;\cdot)$ is given by Eq. (5). Following the same method as in Eq. (33), we have

$$\frac{p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_1)} = \frac{\phi(u_j; \delta_1^2)\phi(v_1; \delta_0^2)}{\phi(u_1; \delta_1^2)\phi(v_j; \delta_0^2)} \tag{37}$$

where, as in the AWGN-2 channel, $v_j = (y_j - m_0)/\sigma_0$ and $u_j = (y_j - m_1)/\sigma_1$. Using Eq. (29), this becomes

$$\frac{p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_1)} = \frac{\phi\left(\sqrt{\frac{\rho_+}{\rho_0 + \rho_+}}\left(v_j - \sqrt{\rho_0}\right); \delta_1^2\right) \phi\left(v_1; \delta_0^2\right)}{\phi\left(\sqrt{\frac{\rho_+}{\rho_0 + \rho_+}}\left(v_1 - \sqrt{\rho_0}\right); \delta_1^2\right) \phi\left(v_j; \delta_0^2\right)} \tag{38}$$

where again $\rho_0 = (m_1 - m_0)^2/\sigma_0^2$ and $\rho_+ = (m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$. Defining $\Delta = \delta_1^2 - \delta_0^2$ and using Eq. (12) to eliminate δ_1^2 and δ_0^2, we obtain the capacity of the Webb-2 channel by plugging into Eq. (18):

$$C(\rho_0, \rho_+, \Delta) = \log_2 M - E_{\mathbf{v}|\mathbf{x_1}} \log_2 \sum_{j=1}^{M} \frac{\phi\left(\sqrt{\frac{\rho_+}{\rho_0+\rho_+}}\left(v_j - \sqrt{\rho_0}\right); \frac{\rho_0+\rho_+}{\rho_0}\Delta\right)\ \phi\left(v_1; \frac{\rho_+}{\rho_0}\Delta\right)}{\phi\left(\sqrt{\frac{\rho_+}{\rho_0+\rho_+}}\left(v_1 - \sqrt{\rho_0}\right); \frac{\rho_0+\rho_+}{\rho_0}\Delta\right)\ \phi\left(v_j; \frac{\rho_+}{\rho_0}\Delta\right)} \tag{39}$$

In these expectations, the $\{v_j\}$ are conditionally independent, given $\{\mathbf{x_1}\}$, and distributed as

$$\left.\begin{array}{lll} v_1 & \text{is} & W\left(\sqrt{\rho_0}, \frac{\rho_0+\rho_+}{\rho_+}, \frac{\rho_0+\rho_+}{\rho_0}\Delta\right) \\ v_j & \text{is} & W\left(0, 1, \frac{\rho_+}{\rho_0}\Delta\right), \quad j \neq 1 \end{array}\right\} \tag{40}$$

D. Capacity of *M*-ary PPM on a Webb+Gaussian Channel

The Webb+Gaussian channel model substitutes Webb+Gaussian random variables $WG(\cdot,\cdot,\cdot,\cdot)$ for the Webb random variables $W(\cdot,\cdot,\cdot)$ in Eq. (36) for the Webb-2 channel model. Given a transmitted signal $\mathbf{x} = \mathbf{x}_j$, the components of the received vector $\mathbf{y}$ are conditionally independent Webb+Gaussian random variables, identically distributed except for y_j:

$$\left.\begin{array}{lll} y_j & \text{is} & WG(m_1, \sigma_1^2, \delta_1^2, \beta_1) \qquad\qquad \text{(signal present)} \\ y_i & \text{is} & WG(m_0, \sigma_0^2, \delta_0^2, \beta_0), \quad i \neq j \quad \text{(signal absent)} \end{array}\right\} \tag{41}$$

The conditional probability density functions are

$$p(y_j|\mathbf{x}_j) = \frac{1}{\sigma_1}\ \phi\left(\frac{y_j - m_1}{\sigma_1}; \delta_1^2, \beta_1\right)$$

$$p(y_i|\mathbf{x}_j) = \frac{1}{\sigma_0}\ \phi\left(\frac{y_i - m_0}{\sigma_0}; \delta_0^2, \beta_0\right), \quad i \neq j$$

where $\phi(\cdot;\cdot,\cdot)$ is given by Eq. (6). Following the same method as in Eq. (37), we have

$$\frac{p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_1)} = \frac{\phi(u_j; \delta_1^2, \beta_1)\phi(v_1; \delta_0^2, \beta_0)}{\phi(u_1; \delta_1^2, \beta_1)\phi(v_j; \delta_0^2, \beta_0)} \tag{42}$$

where, as for the AWGN-2 channel, $v_j = (y_j - m_0)/\sigma_0$ and $u_j = (y_j - m_1)/\sigma_1$. Using Eq. (29), this becomes

$$\frac{p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_1)} = \frac{\phi\left(\sqrt{\frac{\rho_+}{\rho_0+\rho_+}}(v_j - \sqrt{\rho_0}); \delta_1^2, \beta_1\right)\phi(v_1; \delta_0^2, \beta_0)}{\phi\left(\sqrt{\frac{\rho_+}{\rho_0+\rho_+}}(v_1 - \sqrt{\rho_0}); \delta_1^2, \beta_1\right)\phi(v_j; \delta_0^2, \beta_0)} \tag{43}$$

where $\rho_0 = (m_1 - m_0)^2/(\sigma_0^2 + \sigma'^2)$ and $\rho_+ = (m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$. Again defining $\Delta = \delta_1^2 - \delta_0^2$, and using Eq. (15) to eliminate δ_1^2, δ_0^2, and β_1, we obtain the capacity of the Webb+Gaussian channel by plugging into Eq. (18):

$$C(\rho_0, \rho_+, \Delta, \beta_0) =$$

$$\log_2 M - E_{\mathbf{v}|\mathbf{x_1}} \log_2 \sum_{j=1}^{M} \frac{\phi\left(\sqrt{\frac{\rho_+}{\rho_0+\rho_+}}(v_j - \sqrt{\rho_0}); \frac{\rho_0+\rho_+}{\rho_0}\Delta, \beta_0 + \frac{\rho_0}{\rho_+}\right) \; \phi\left(v_1; \frac{\rho_+}{\rho_0}\Delta, \beta_0\right)}{\phi\left(\sqrt{\frac{\rho_+}{\rho_0+\rho_+}}(v_1 - \sqrt{\rho_0}); \frac{\rho_0+\rho_+}{\rho_0}\Delta, \beta_0 + \frac{\rho_0}{\rho_+}\right) \; \phi\left(v_j; \frac{\rho_+}{\rho_0}\Delta, \beta_0\right)} \tag{44}$$

In these expectations, the $\{v_j\}$ are conditionally independent, given $\{\mathbf{x_1}\}$, and distributed as

$$\left.\begin{array}{lll} v_1 & \text{is} & WG(\sqrt{\rho_0}, \frac{\rho_0+\rho_+}{\rho_+}, \frac{\rho_0+\rho_+}{\rho_0}\Delta, \beta_0 + \frac{\rho_0}{\rho_+}) \\ v_j & \text{is} & WG(0, 1, \frac{\rho_+}{\rho_0}\Delta, \beta_0), \quad j \neq 1 \end{array}\right\} \tag{45}$$

IV. Bit-SNR Capacity Limits of *M*-ary PPM on Channels with Soft Outputs

In the case of the classic AWGN-1 channel, the capacity formulas imply a well-known threshold on the minimum required signal-to-noise ratio (SNR) *per information bit* communicated over the channel. If the AWGN-1 channel SNR is E_s/N_0 (per channel symbol), the corresponding bit SNR is computed as $E_b/N_0 = (E_s/N_0)/R$ (per information bit), where R (information bits/channel symbol) is the rate of the overall code applied to the channel. If the rate is at the capacity limit, then $R = C$, and the formula for the minimum possible bit SNR is $(E_b/N_0)_{\min} = (E_s/N_0)/C$. For the AWGN-1 channel, it is well known that reliable communication is possible with appropriate coding if and only if $(E_b/N_0)_{\min}$ does not drop below a brick-wall threshold of -1.59 dB. Higher thresholds apply if the code rate is constrained away from zero or if there are constraints on the modulation format.

In this section, we find numerically that all of the channels considered in the previous section have quasi-brick-wall bit-SNR thresholds analogous to the well-known thresholds for the classic AWGN-1 channel. To unify the treatment of these channels, we define a minimum bit-SNR parameter $\rho_b = \rho/(2C)$ or $\rho_b = \rho_0/(2C)$ for each channel. Note that this definition for the AWGN-1 channel reduces to $\rho_b = (E_s/N_0)/C = (E_b/N_0)_{\min}$, so graphs of capacity versus ρ_b have exactly the same interpretation in this case as classic AWGN-channel graphs of capacity versus $(E_b/N_0)_{\min}$.

A. The AWGN-1 Channel

Exact computation of C for large dimensions is extremely complex, and it is necessary to resort to Monte Carlo methods as described in [5, Appendix I]. Figure 4 is an evaluation of PPM capacity on the AWGN-1 channel obtained by Monte Carlo simulation of Eq. (23) for $M = 2, 4, 8, 16, 32, 64, 256$. The figure shows AWGN-1 capacity as a function of the minimum required bit SNR $\rho_b = \rho/(2C) = E_b/N_0$. We see from this figure that the restriction to an M-ary signal set causes the capacity curves for M-PPM to saturate at $\log_2 M$ bits, no matter how much SNR is supplied. On the other end of the scale, there is a characteristic minimum value of E_b/N_0 for each M below which the capacity drops to zero. These minimum values result directly from restricting the M-dimensional signaling set to be the orthogonal set, and are discussed more in Section VI. When these capacity results are plotted on a (vertical) log scale (e.g., see Fig. 7 later), each capacity curve appears to approach a vertical brick wall at the minimum attainable value of E_b/N_0.

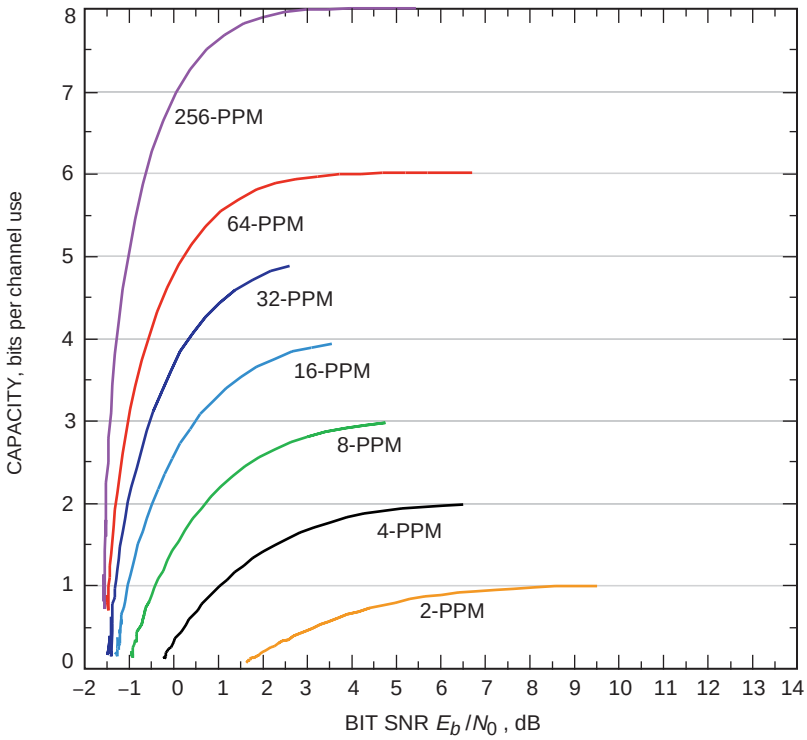

Fig. 4. PPM capacity on the AWGN-1 channel, determined by Monte Carlo simulation.

Figure 5 gives a subset of the same capacity curves in Fig. 4, except that in this case the curves for $M = 2, 4$ were computed exactly from Eq. (23) using *Mathematica*. Figure 5 also overlays several related performance curves for comparison. First, in the case of 256-PPM, the analytic Approximation (24) is compared with the more exact evaluation of Eq. (23) using Monte Carlo simulation. Next, the figure superimposes a set of points (artificially connected by a curve) showing the actual signaling rate (not capacity) in bits per channel use versus the required E_b/N_0 for uncoded M-PPM to achieve a bit-error probability (BER) of $P_b = 10^{-6}$, where $P_b = (1/2)(M/[M-1])P_M$, and the probability of uncoded symbol error, P_M, is given later in Eq. (48). The horizontal displacements between this curve and the family of capacity curves show how much room there is for improvement over an uncoded system at this BER. Finally, the figure shows (as dashed curves) the capacity formulas $C = (k/2)\log_2(1+\rho) = (k/2)\log_2(1+2RE_b/N_0)$ for k-dimensional Gaussian channels unrestricted as to modulation type, which can be explicitly solved at the capacity limit $R = C$ for $(E_b/N_0)_{\min} = (2^{2C/k}-1)/2C$.

B. The AWGN-2 Channel

Figure 6 shows the capacity of 256-PPM obtained by Monte Carlo integration of Eq. (31) for the AWGN-2 model, plotted for different values of ρ_+ versus the bit-normalized SNR parameter defined by $\rho_b = \rho_0/(2C)$. Plotted (on a log scale) versus ρ_b, the AWGN-2 capacity curves for different values of ρ_+ all saturate at 8 bits per channel use (since this is for 256-PPM), and they all appear to approach the same brick-wall threshold of -1.57 dB on the minimum attainable value of ρ_b for 256-PPM, at least down to capacity levels below about 1 bit per channel use. At low values of ρ_+ (e.g., $\rho_+ = 5$ in the figure, which corresponds to a physical parameter $\eta\bar{n}_s = 10.8$ with an APD excess noise factor of $F = 2.16$), the capacity curves appear to approach the same asymptotes but the curvature in between is less pronounced and the capacity for fixed ρ_b is less than for the pure AWGN-1 model. When ρ_+ is made really small, typically corresponding to an average of less than one absorbed photon per slot, we have noticed that the capacity curves do not stop at the brick-wall limit of -1.57 dB, i.e., there is a very small nonzero capacity for ρ_b below -1.57 dB. This effect probably also shows up for any finite ρ_+, but at capacity levels too

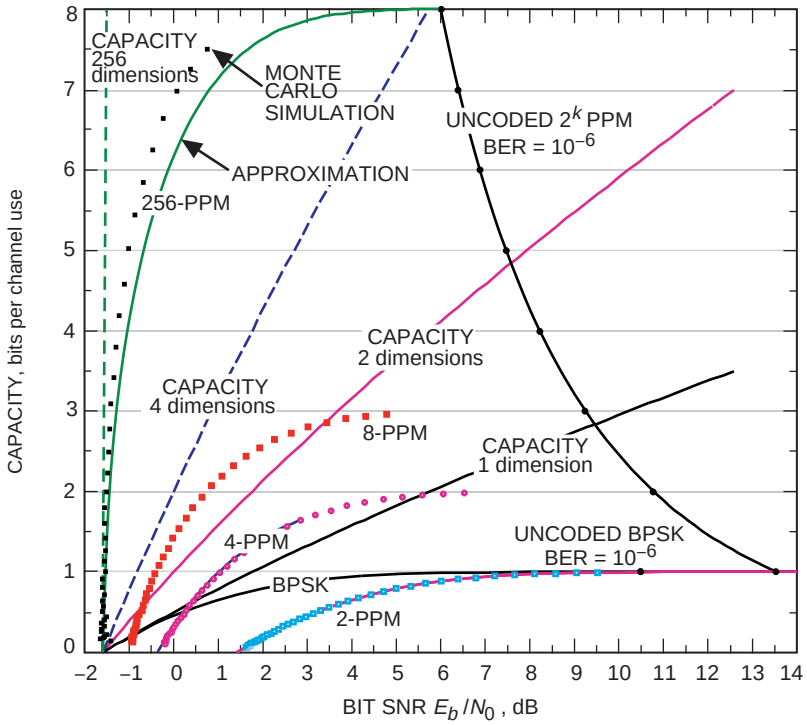

Fig. 5. PPM capacity on the AWGN-1 channel, compared with other performance curves.

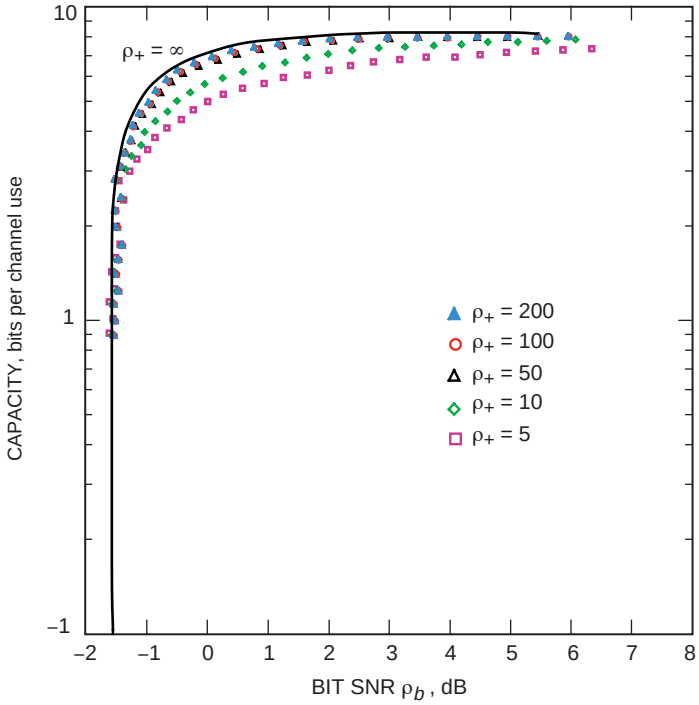

Fig. 6. 256-PPM capacity on the AWGN-2 channel, determined by Monte Carlo simulation. The case $\rho_+ = \infty$ is equivalent to the AWGN-1 channel.

low to be simulated accurately (much less than 1 bit per channel use). This breaching of the brick-wall threshold for very low values of capacity per channel use needs to be studied in more depth whenever we extend our results to the Poisson limit (corresponding to small ρ_+ as compared with ρ_0).

C. The Webb Channels

We evaluated the M-dimensional expectations in Eq. (39) accurately via Monte Carlo simulation. Some results are plotted in Fig. 7 for the Webb-2 channels for different PPM orders M. Corresponding curves for the AWGN-1 channel are shown for comparison. The abscissa in this figure is the normalized bit SNR, $\rho_b = \rho/(2C)$ for AWGN-1 and $\rho_b = \rho_0/(2C)$ for Webb-2. Along each Webb-2 curve, the two independent variables held constant are $\Delta = 60.8$ and $\rho_+ = 17.6$, which correspond to a representative optical APD problem with $\eta\bar{n}_s = 38$ detected signal photons per PPM word and an excess noise factor $F = 2.16$. The x-axis in this figure also shows "simplex-to-orthogonal" penalty factors, which are discussed later in Section VI.

We see from this figure that the capacity curves for the more complicated Webb-2 model again resemble those for the well-known AWGN-1 model, as they bend around from horizontal asymptotes at $\log_2 M$ bits per channel use to an apparent vertical asymptote at $\rho_b = 10\log_{10}[(M\ln 2)/(M-1)]$ dB. Again, as with the AWGN-2 capacity curves, there is evidence that the vertical asymptote is breached at very low values of capacity per channel use, but this is not noticeable in simulations except when ρ_+ is also very small.

Figure 8 gives the Webb-2 capacity curves for the same set of physical parameters considered in Fig. 7, except that in this case the capacity is plotted versus the average background count, $\bar{n}_b$, instead of versus ρ_b. In each case, the capacity curves start at their asymptotic limit, $\log_2 M$ for small values of $\bar{n}_b$ and then plummet toward zero as $\bar{n}_b$ is increased above $\bar{n}_s$.

Figure 9 plots the capacity of 256-PPM on a Webb-2 channel for various values of $\bar{n}_s = F\rho_+/\eta$. Again the abscissa is the bit-normalized SNR parameter $\rho_b = \rho_0/(2C)$. Note that the Webb-2 capacity curves

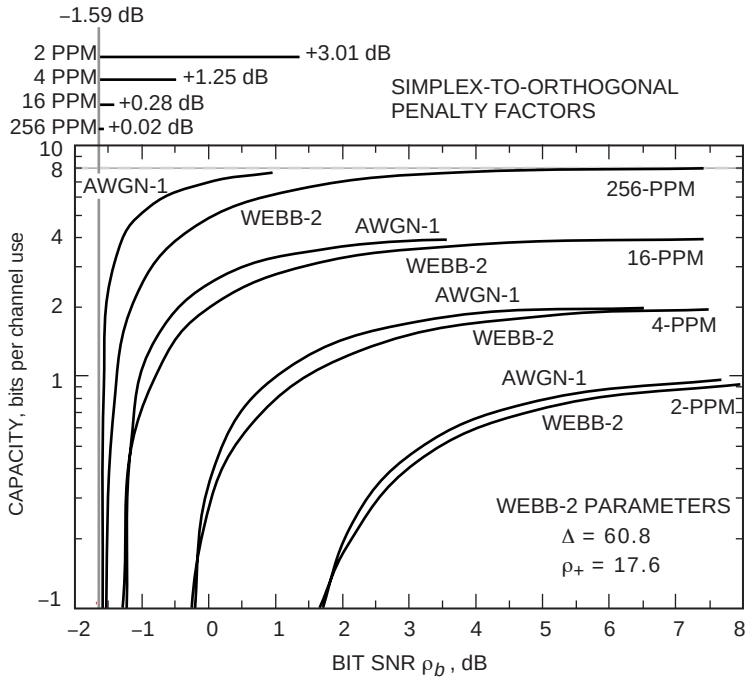

Fig. 7. Capacity of AWGN-1 and Webb-2 channels for different PPM sizes.

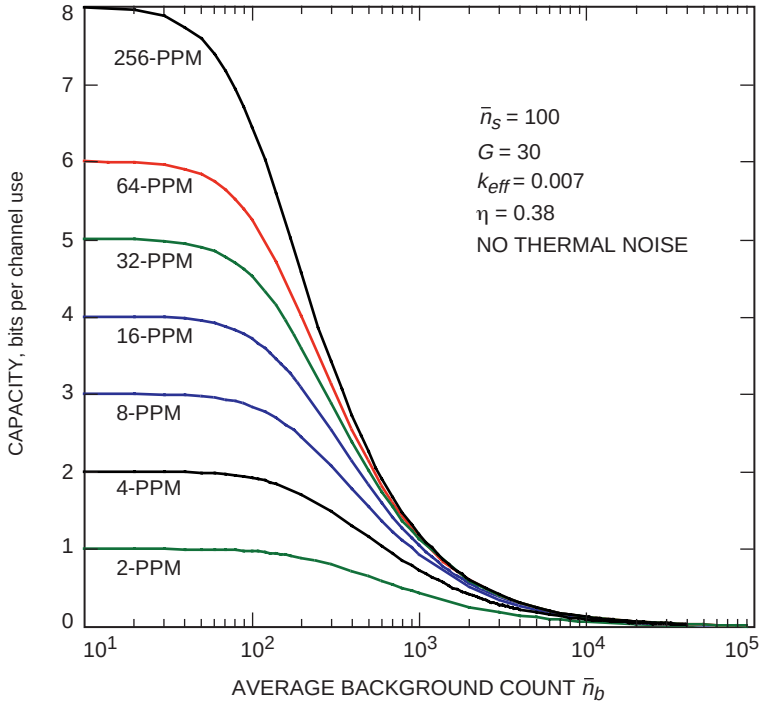

Fig. 8. Capacity of *M*-PPM on a Webb-2 channel.

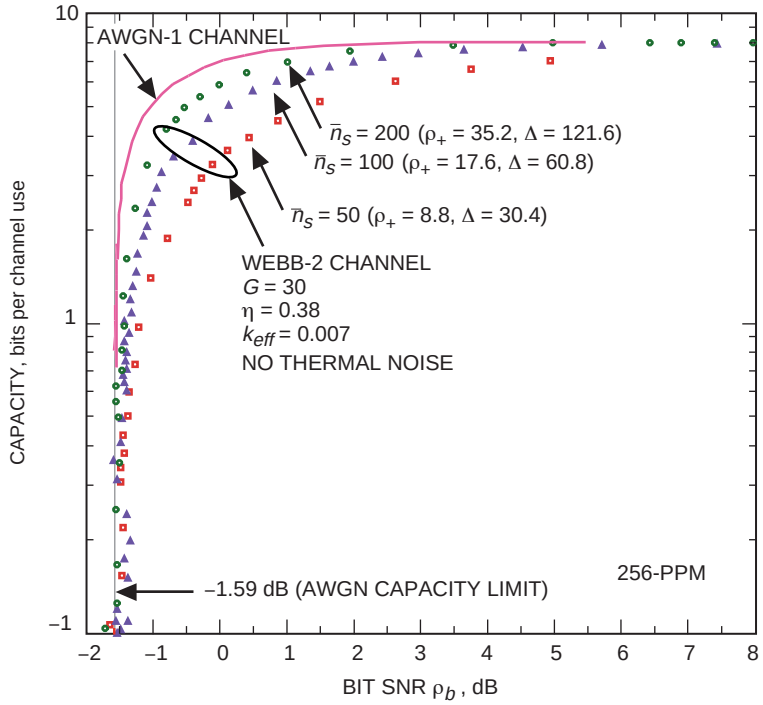

Fig. 9. Capacity of 256-PPM for AWGN-2 and Webb-2 channels.

in Fig. 9 approach the capacity curve of the AWGN-2 channel (shown for reference) more closely as $\bar{n}_s$, or, equivalently, ρ_+, gets large.

V. Capacity of *M*-ary PPM on Channels with Hard Outputs

For the AWGN-2 and Webb-2 channels, we also computed the capacity with hard-decision channel outputs. When a hard-decision detection scheme is used, the decoder operates on PPM symbol decisions from the demodulator, not individual soft counts. The modulator output is the PPM symbol having the maximum slot count [10]. This hard-decision channel is an M-ary input, M-ary output, symmetric channel with capacity given by [3]

$$C = \log_2 M + (1 - P_M)\log_2(1 - P_M) + P_M \log_2\left(\frac{P_M}{M-1}\right) \tag{46}$$

bits per channel use, where P_M is the probability of incorrect symbol detection.

The probability of incorrect symbol detection is calculated as $P_M = 1 - \Pr[y_j < y_1, \forall j \neq 1 | \mathbf{x_1}] = \Pr[v_j < v_1, \forall j \neq 1 | \mathbf{x_1}]$ for the various definitions of $\{y_j\}$ and $\{v_j\}$ used in determining the capacities of the AWGN and Webb channels. Given the conditional independence of the observations in the $M-1$ nonsignal slots, this reduces to

$$P_M = 1 - E_{v_1|\mathbf{x_1}} P_0(v_1)^{M-1} \tag{47}$$

where $P_0(v)$ is the cumulative probability distribution function for the normalized channel symbols $\{v_j\}$ in the nonsignal slots. Using the conditional statistics of v_1 given $\mathbf{x_1}$ in Eqs. (22), (28), (34), (40), and (45), we can find the probability of uncoded M-PPM symbol error, P_M, for the various channels considered in Section III.

For the AWGN-1 channel, P_M is given by

$$P_M = 1 - \int_{-\infty}^{\infty} \phi\left(v - \sqrt{\rho}\right) \Phi(v)^{M-1} dv \tag{48}$$

where $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ is the standard normalized Gaussian probability density function and $\Phi(x) = \int_{-\infty}^{x} \phi(u)du$ is the corresponding normalized Gaussian probability distribution function. For the AWGN-2 channel, we have

$$P_M = 1 - \int_{-\infty}^{\infty} \sqrt{\frac{\rho_+}{\rho_0 + \rho_+}}\, \phi\left(\sqrt{\frac{\rho_+}{\rho_0 + \rho_+}}\left(v - \sqrt{\rho_0}\right)\right) \Phi(v)^{M-1} dv \tag{49}$$

Note that, when $\rho_+ \to \infty$ and $\rho_0 = \rho$, the AWGN-2 channel becomes the AWGN-1 channel, and Eq. (49) reduces to Eq. (48).

For the Webb channels, first define $\Phi(x; \delta^2) = \int_{-\delta}^{x} \phi(w; \delta^2)dw$ to be the cumulative probability distribution corresponding to the normalized Webb probability density function defined in Eq. (5). Then P_M for the Webb-1 channel is computed as

$$P_M = 1 - \int_{-\infty}^{\infty} \phi\left(w - \sqrt{\rho}; \delta^2\right) \Phi(w; \delta^2)^{M-1} dw \tag{50}$$

Similarly, for the Webb-2 channel,

$$P_M = 1 - \int_{-\infty}^{\infty} \sqrt{\frac{\rho_+}{\rho_0 + \rho_+}}\ \phi\left(\sqrt{\frac{\rho_+}{\rho_0 + \rho_+}}\,(w - \sqrt{\rho_0})\,;\delta_1^2\right) \Phi\left(w;\delta_0^2\right)^{M-1}\,dw \tag{51}$$

Finally, for the Webb+Gaussian channel,

$$P_M = 1 - \int_{-\infty}^{\infty} \sqrt{\frac{\rho_+}{\rho_0 + \rho_+}}\ \phi\left(\sqrt{\frac{\rho_+}{\rho_0 + \rho_+}}\,(x - \sqrt{\rho_0})\,;\delta_1^2,\beta_1\right) \Phi(x;\delta_0^2,\beta_0)^{M-1}dx \tag{52}$$

where $\Phi(x;\delta^2,\beta) = \int_{-\delta}^{x} \phi(y;\delta^2,\beta)dy$ is the cumulative probability distribution corresponding to the standardized Webb+Gaussian probability density function defined in Eq. (6).

The soft-decision capacities obtained in Section III can be compared with the hard-decision capacities obtained by plugging the various formulas for P_M into the general hard-decision capacity expression of Eq. (46). Figure 10 compares capacities for the hard-output and soft-output AWGN-2 channels for the case of $M = 256$. A similar comparison of capacities is shown in Fig. 11 for the hard-output and soft-output Webb-2 channels. The hard-output Webb-2 capacity was computed in [6].

The capacity curves for both the AWGN-2 and the Webb-2 channels show that a minimum value of ρ_b is reached at a nonzero code rate. Unlike the soft-output channels, which exhibit monotonically better efficiency in terms of the bit-normalized SNR parameter ρ_b as the code rate (and hence the capacity per channel use) is reduced toward zero, the bit-normalized SNR efficiency of the hard-output channel

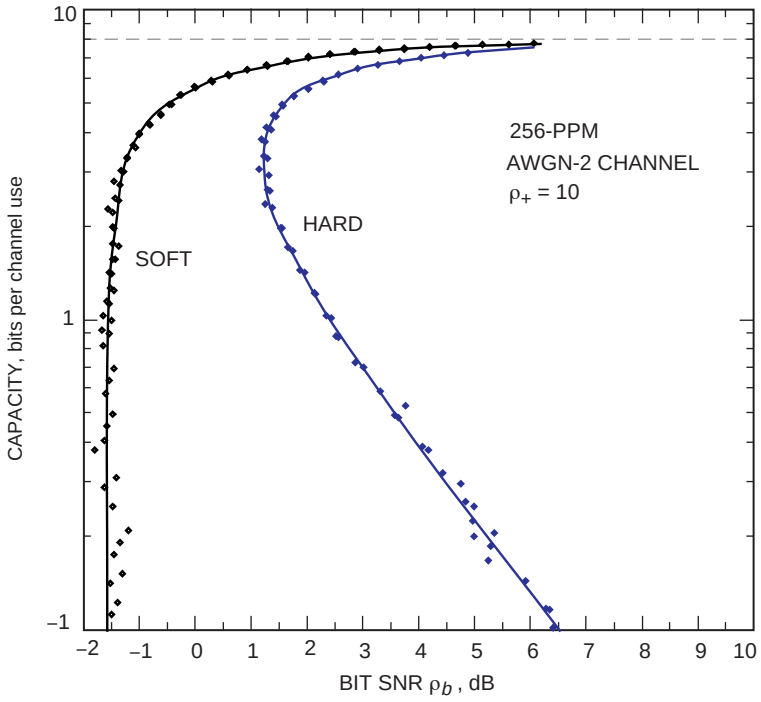

Fig. 10. Capacity of 256-PPM on the hard- and soft-output AWGN-2 channels.

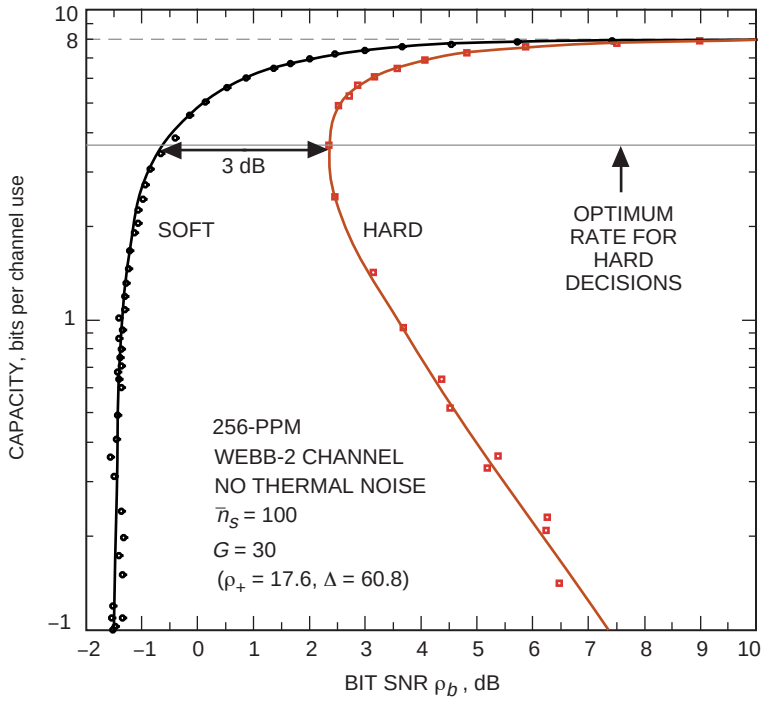

Fig. 11. Capacity of 256-PPM on the hard- and soft-output Webb-2 channels.

worsens if the capacity is lowered below about 4 bits per channel use. This implies that an optimum code rate of about 1/2 will achieve the lowest ρ_b for the hard-output channel, while the soft-output channel achieves lowest ρ_b in the limit as the code rate goes to 0.

Another comparison of capacity for the hard-output and soft-output Webb models is shown in Fig. 12, this time plotted versus $\bar{n}_b$. The hard-output capacity in this figure was computed in [6] for the general Webb+Gaussian channel that also models the effects of thermal noise.

VI. Capacity Results for Simplex Signaling

In this section, we demonstrate a similarity between AWGN and Webb channels in the fundamental brick-wall limits for M-ary orthogonal signaling and simplex signaling.

A. AWGN Channels

Among all M-ary signaling schemes on the AWGN-1 channel, the simplex set maximizes the capacity C at low E_b/N_0 [7]. The simplex set achieves the same capacity as an M-ary orthogonal signaling scheme that uses a factor of $M/(M-1)$ more energy per bit. Applying the Shannon limit of $E_b/N_0 > \ln 2 \approx -1.59$ dB, it follows that any M-ary orthogonal signaling scheme requires $E_b/N_0 > (M \ln 2)/(M-1)$ for reliable communication to occur. Thus, each M-ary orthogonal signal has an associated E_b/N_0 penalty of $10 \log_{10} M/(M-1)$ dB. This is illustrated in Fig. 13(a), where each M-ary orthogonal signaling capacity curve has a vertical asymptote at $E_b/N_0 > (M \ln 2)/(M-1)$. When translated to simplex signal sets by shifting left by $10 \log_{10} M/(M-1)$ dB, the vertical asymptotes for each of the curves coincide at $E_b/N_0 = -1.59$ dB, as shown in Fig. 13(b).

B. Webb Channels

On the Webb-2 channel, we observe the same behavior. In Fig. 14(a), capacity curves for M-ary orthogonal signaling on the Webb-2 channel have varying vertical asymptotes. When translated by

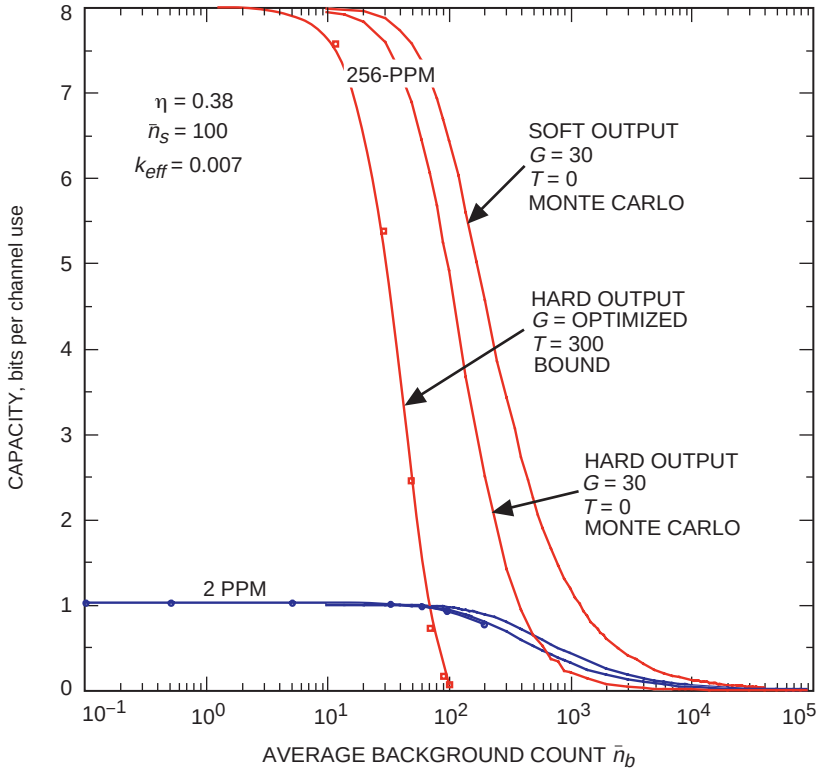

Fig. 12. Comparison of capacity for the hard- and soft-output Webb models.

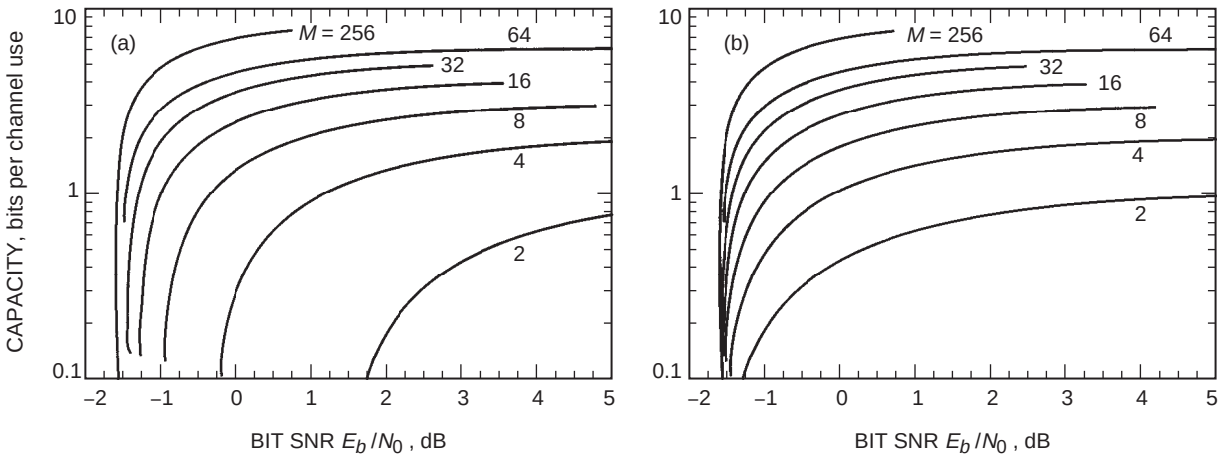

Fig. 13. AWGN-1 capacity versus bit SNR for (a) *M*-ary orthogonal signaling and (b) *M*-ary simplex signaling.

$10 \log_{10} M/(M-1)$ dB, as in Fig. 14(b), these vertical asymptotes coincide at -1.59 dB in the same way they did on the AWGN-1 channel. Although simplex signaling on the Webb-2 channel may not be practical because it would necessitate coherent detection instead of simple photon detection, this finding illustrates an important similarity between the brick-wall limits on the AWGN and Webb channels. In both cases, the restriction to a finite orthogonal signal set imposes the same fundamental penalty on the minimum achievable bit SNR.

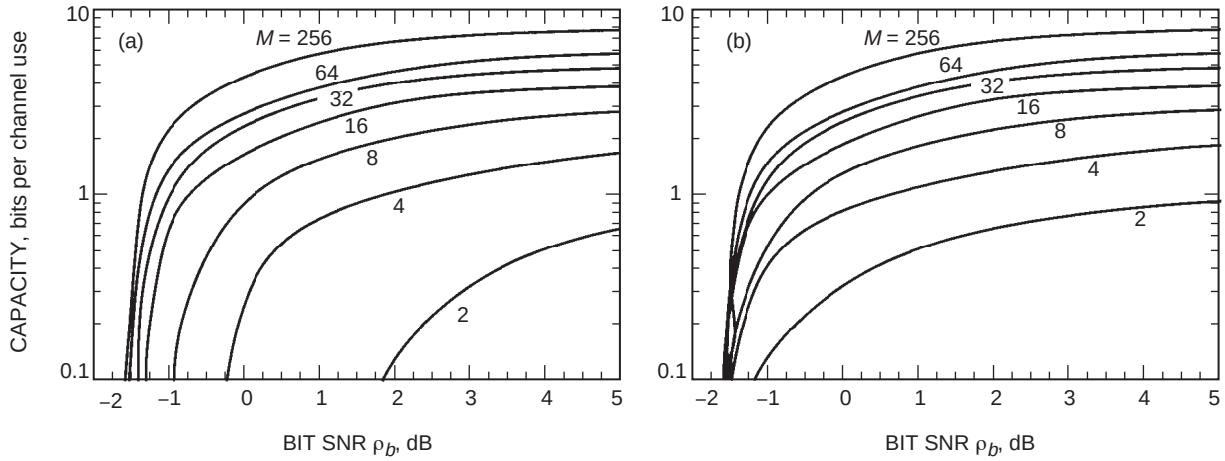

Fig. 14. Webb-2 capacity versus bit SNR for (a) *M*-ary orthogonal signaling and (b) *M*-ary simplex signaling.

VII. Comparison of Capacity Results with Actual Code Performance

Figure 15 shows the performance of Reed–Solomon (RS) codes on GF(2^k) applied to 2^k-PPM for $P_b = 10^{-6}$. For each curve, the alphabet size is fixed and, therefore, the Reed–Solomon codeword size is fixed at $(2^k - 1)$ k-bit symbols. The curves are obtained by varying the code rate within each RS code family. For these curves, the RS decoder is assumed to correct only errors (i.e., no erasures), and the uncoded symbol-error probability is given by Eq. (48) with $M = 2^k$. As an example, the performance of the (255,223) RS code, with code rate approximately 7/8, is plotted at approximately 7 bits per channel use on the 8-bit RS curve in Fig. 15. This code requires 2.6 dB for $P_b = 10^{-6}$ and is only 1.8 dB worse than the capacity limit achievable by arbitrary codes of the same rate for 256-PPM (see the gap marked with "B" in Fig. 15). The additional gap, marked as "A," is due to constraining the 256-dimensional signal set to be orthogonal. Note that the comparison between the performance of RS codes and the two capacity limits does not account for the fact that the RS decoder uses hard quantized inputs while both capacity limits are computed for unconstrained channel output. This limitation must account for some portion of the nonoptimality of RS codes. Another interesting observation from Fig. 15 is that RS codes appear to be optimum approximately at rate 3/4 for all alphabet sizes. Lower-rate RS codes have progressively worse performance. However, Fig. 15 suggests the possibility of finding codes of rate 1/2 or lower that can achieve $P_b = 10^{-6}$ at $E_b/N_0 \leq 0$ dB. This would be a gain of 2.6 dB or more over the (255,223) RS code with 256-PPM. Thus, it is natural to look for codes of rate 1/2 and lower to attempt to approach capacity more closely than RS codes.

Some results are available on simple binary turbo codes of rate 1/2 and 1/3 as compared with RS codes of the same rate.[4] (See also [4].) The performance of these codes is *estimated* in Fig. 15 by using the following approximation. The required input symbol-error probabilities to achieve $P_b = 10^{-6}$ with RS codes or turbo codes are related to the E_b/N_0 required for uncoded 256-PPM to obtain those symbol-error probabilities. The difference in required E_b/N_0 between RS and turbo codes is used to plot the turbo-code performance relative to the RS performance. This is an approximate comparison because the turbo codes were not simulated exactly on a PPM system with Gaussian noise. These results indicate that, while these binary turbo codes do outperform RS codes of the same rate, there still remains a gap of several dB to the capacity limit.

[4] *Data Compression and Channel Coding*, X2000 Report (internal document), Jet Propulsion Laboratory, Pasadena, California, September 15, 1997.

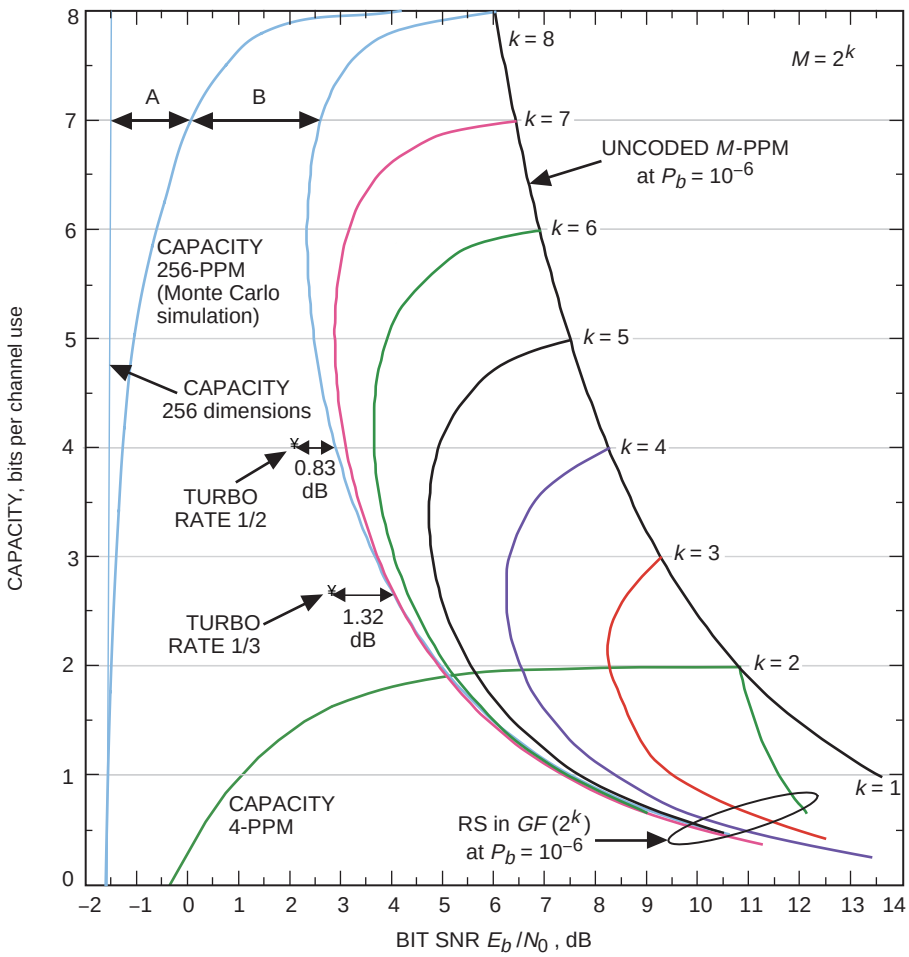

Fig. 15. RS code performance compared with PPM capacity.

VIII. Conclusions and Future Work

This article has analyzed several idealized channel models that can be used to approximate PPM signaling over an APD-detected optical communication channel. We were able to define a suitable bit-normalized SNR parameter, ρ_b, such that all of these channels with soft outputs yield quasi-brick-wall thresholds on the minimum acceptable value of ρ_b above which reliable communication is theoretically possible and below which it is not possible except at extremely low channel rates. Furthermore, under all of these models with soft channel outputs, the bit-SNR thresholds for different values of M differ from each other by the same "simplex-to-orthogonal penalty," $(M-1)/M$. Under both the AWGN-2 and Webb-2 models, the gap between the capacities of hard- and soft-output channels is about 3 dB at the code rate giving the optimum hard-output bit SNR. On the hard-output channels, there is an optimum M beyond which capacity is diminished [6] because much of the small incremental information available from each slot is destroyed when all of that information must be summarized as a single decision among an increasing number of candidate slots. This contrasts sharply with the results for soft decisions, for which larger M gives uniformly better capacity under each model.

The capacities computed in this article assume that there are no time gaps between successive channel uses. However, laser transmitters require a sufficient recharging time between pulses, which is guaranteed by introducing a dead time, T_d, between successive channel uses, as shown in Fig. 16. This reduces the capacity computed above by a factor $(T_w + T_d)/T_w$, where $T_w = MT_s$ is the PPM word duration. Since

this factor can be large, it is important to optimize the trade-off between achievable peak power and dead time. A more efficient and more information-theoretically sound method to guarantee a minimum spacing, T_d, between pulses is to design a code that inherently satisfies such a constraint. Codes of this kind, usually called (d, k) codes, have been used extensively in the magnetic recording field. In these codes, at least d 0s must follow every 1, and not more than k 0s can follow a 1. Here a 1 corresponds to a pulse and a 0 to no pulse. For laser transmission, the k constraint is not applicable, but the d constraint directly reflects the need for dead time. In a future article, we will investigate how (d, k) codes can be designed to maximize the amount of information transmitted subject to this type of constraint.

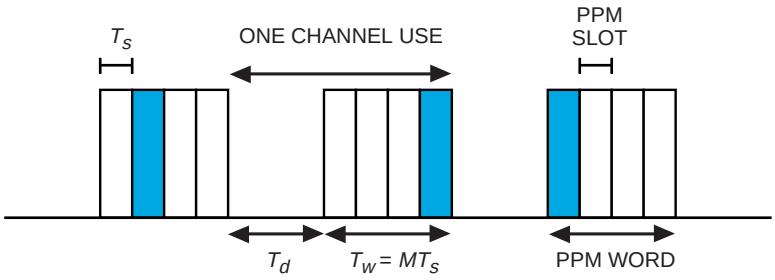

Fig. 16. PPM signaling ($M = 4$, $k = 2$).

References

[1] S. Butman and M. J. Klass, "Capacity of Noncoherent Channels," *The Deep Space Network Progress Report, September and October 1973*, JPL Technical Report 32-1526, vol. XVIII, Jet Propulsion Laboratory, Pasadena, California pp. 85–93, December 15, 1973.

[2] F. M. Davidson and X. Sun, "Gaussian Approximation Versus Nearly Exact Performance Analysis of Optical Communication Systems with PPM Signaling and APD Receivers," *IEEE Trans. Commun.*, vol. 36, no. 11, pp. 1185–1192, November 1988.

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

[4] J. Hamkins and M. Srinivasan, "Turbo Codes for APD-Detected PPM," Allerton Conference, Monticello, Illinois, October 1998.

[5] S. J. MacMullan and O. M. Collins, "The Capacity of Orthogonal and Bi-Orthogonal Codes on the Gaussian Channel," *IEEE Trans. on Information Theory*, vol. 44, no. 3, May 1998.

[6] J. Hamkins, "The Capacity of Avalanche Photodiode-Detected Pulse-Position Modulation," *The Telecommunications and Mission Operations Progress Report 42-138, April–June 1999*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–19, August 15, 1999.
http://tmo.jpl.nasa.gov/tmo/progress_report/42-138/138A.pdf

[7] J. L. Massey, "All Signal Sets Centered About the Origin Are Optimal at Low Energy-to-Noise Ratios on the AWGN channel," *Proc. Int. Symp. Information Theory*, Ronneby, Sweden, pp. 80–81, June 1976.

[8] K. R. Baker, "On the WMC Density as an Inverse Gaussian Probability Density," *IEEE Trans. Commun.*, vol. 44, no. 1, pp. 15–17, January 1996.

[9] J. T. K. Tang and K. B. Letaief, "The Use of WMC Distribution for Performance Evaluation of APD Optical Communication Systems," *IEEE Trans Commun.*, vol. 46, no. 2, pp. 279–285, February 1998.

[10] V. Vilnrotter, M. Simon, and M. Srinivasan, "Maximum Likelihood Detection of PPM Signals Governed by an Arbitrary Point Process Plus Additive Gaussian Noise," JPL Publication 98-7, Jet Propulsion Laboratory, Pasadena, California, April 1998.

[11] P. P. Webb, R. J. McIntyre, and J. Conradi, "Properties of Avalanche Photodiodes," *RCA Review*, vol. 35, pp. 234–278, June 1974.

Appendix A
Parameters and Notation

The following is a description of parameters and notation in this article, along with nominal values used in the numerical results.

	Laser and modulator parameters	
ν	1064	Optical frequency, nm
M	2-256	PPM order
T_s	3.125×10^{-8}	Width of the PPM slot required by the laser, s
T_d	4.32×10^{-4}	Dead time between PPM symbols required by the laser, s
α_{er}	10^6	Modulation extinction ratio
	APD detector parameters	
η	38 percent	Quantum efficiency
k_{eff}	0.007	Ionization ratio
G	50–200	Gain
F	2.2–3.4	Excess noise factor, $F = k_{eff}G + (2 - 1/G)(1 - k_{eff})$
T	300	Noise temperature, K
R_L	179,700	Load resistance (transimpedance model), $5.75 \times 10^{12} \times T_s$, Ω
B	$1/2T_s$	Noise equivalent one-sided bandwidth, Hz
I_b	4×10^{-14}	Bulk leakage current, A
I_s	2×10^{-9}	Surface leakage current, A
$\bar{n}$		Mean photons incident on the photodetector, per pulse ($= \bar{n}_b$ or $\bar{n}_s$)
$\bar{n}_b$	0.001–10,000	Mean background photons incident on the photodetector, per slot
$\bar{n}_s$	100	Mean signal photons incident on the photodetector, per pulse
	Physical constants	
e_-	1.38×10^{-23}	Electron charge, C
h	6.7×10^{-34}	Planck's constant, J/Hz
κ	1.6×10^{-19}	Boltzmann's constant, J/K
	Mathematical notation	
C		Capacity, bits per channel use
m, m_0, m_1		Mean of Gaussian or Webb distribution
$\sigma^2, \sigma_0^2, \sigma_1^2$		Variance of Gaussian or Webb distribution
$\delta^2, \delta_0^2, \delta_1^2$		Skewness parameter of Webb distribution
β, β_0, β_1		Blending fraction for Webb+Gaussian distribution
ρ		Equals m^2/σ^2 (AWGN-1 or Webb-1 channel)
ρ_0		Equals $(m_1 - m_0)^2/\sigma_0^2$ (AWGN-2 or Webb-2 channel)
ρ_0		Equals $(m_1 - m_0^2)/(\sigma_0^2 + \sigma'^2)$ (Webb+Gaussian channel)

Mathematical notation (cont'd)	
ρ_b	The minimum bit SNR at the capacity limit (equals $(E_b/N_0)_{/min}$ on AWGN-1)
ρ_b	Equals $\begin{cases} \rho/(2C) & \text{AWGN-1 or Webb-1 channel} \\ \rho_0/(2C) & \text{AWGN-2 or Webb-2 or Webb+Gaussian channel} \end{cases}$
ρ_+	Equals $(m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$
Δ	Equals $\delta_1^2 - \delta_0^2$
q	Number of electrons generated at the output of the APD
P_M	Probability of uncoded M-PPM symbol error
$\phi(x)$	Gaussian probability density function, $1/(\sqrt{2\pi})e^{-x^2/2}$
$\Phi(x)$	Gaussian cumulative distribution function, $\int_{-\infty}^{x} \phi(y)dy$
$\phi(w;\delta^2)$	Webb probability density function, see Eq. (5)
$\Phi(w;\delta^2)$	Webb cumulative distribution function, $\int_{\infty}^{w} \phi(u;\delta^2)du$
$\phi(x;\delta^2,\beta)$	Webb+Gaussian probability density function, see Eq. (6)
$\Phi(x;\delta^2,\beta)$	Webb+Gaussian cumulative distribution function, $\int_{-\infty}^{x} \phi(y;\delta^2,\beta)dy$

Appendix B

Moments of the Standardized Webb Probability Distribution

Let W be distributed according to the standardized Webb density

$$\phi(w;\delta^2) = \frac{1}{\sqrt{2\pi}}\left(1+\frac{w}{\delta}\right)^{-3/2}\exp\left[\frac{-w^2}{2\left(1+\frac{w}{\delta}\right)}\right], \quad w > -\delta$$

Let $X = 1 + W/\delta$. Then the probability density function of X is related to the inverse Gaussian density [8,9] by

$$f_X(x) = \frac{\delta}{\sqrt{2\pi}}\, x^{-3/2}\exp\left[-\frac{\delta^2(x-1)^2}{2x}\right] = f(x;1,\delta^2)$$

It follows that the moment generator function for W is given by

$$\begin{aligned}
M(s) &= \int_{-\delta}^{\infty} e^{sw}p(w)dw \\
&= \int_0^{\infty} e^{s\delta(x-1)}f(x;1,\delta^2)dx \\
&= e^{-s\delta}\int_0^{\infty}\frac{\delta x^{-3/2}}{\sqrt{2\pi}}\exp\left[-\frac{\delta^2(x-1)^2}{2x}+s\delta x\right]dx \\
&= e^{-s\delta}\int_0^{\infty}\underbrace{\frac{\delta x^{-3/2}}{\sqrt{2\pi}}\exp\left[-\frac{\delta^2\left(1-\frac{2s}{\delta}\right)\left(x-\left(1-\frac{2s}{\delta}\right)^{-1/2}\right)^2}{2x}\right]}_{f(x;(1-2s/\delta)^{-1/2},\delta^2)}\exp\left[\delta^2\left(1-\sqrt{1-\frac{2s}{\delta}}\right)\right]dx \\
&= \exp\left[\delta^2\left(1-\sqrt{1-\frac{2s}{\delta}}-\frac{s}{\delta}\right)\right]
\end{aligned}$$

Thus, the mean of W is $M'(0) = 0$, the variance is $M''(0) = 1$, and each moment of W is a polynomial in the parameter $1/\delta^2$.

Appendix C

Approximation to *M*-ary PPM AWGN Capacity for Large *M*

To evaluate the capacity of the AWGN-2 channel for large M, we rewrite Eq. (31) as

$$C = \log_2 M - E_{\mathbf{v}|\mathbf{x}_1} \log_2 \left[1 + \sum_{j=2}^{M} e^{z_j - z_1} \right] \tag{C-1}$$

where $z_j = [\rho_0/(\rho_0+\rho_+)](v_j^2/2) + [\rho_+/(\rho_0+\rho_+)]\sqrt{\rho_0} v_j$. When M is large, the sum of random variables $\sum_{j=2}^{M} e^{z_j}$ converges to $(M-1)$ times the mean of each e^{z_j}, because $\{z_j,\ j=2,\cdots,M\}$ are independent and identically distributed, given $\mathbf{x_1}$. Thus, the capacity expression can be approximated by

$$C \approx \log_2 M - E_{v_1|\mathbf{x}_1} \log_2 \left[1 + e^{-z_1} (M-1) \overline{e^{z_j}} \right] \tag{C-2}$$

where $\overline{e^{z_j}}$ denotes the mean of e^{z_j}. It remains to evaluate $\overline{e^{z_j}}$.

For any $j > 1$, z_j is a quadratic function of an $N(0,1)$ random variable v_j. To evaluate $E_v\left[e^{av^2/2+bv}\right]$ when v is $N(0,1)$, we write out the integral

$$\begin{aligned}
E_v\left[e^{av^2/2+bv}\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-v^2/2} e^{av^2/2} e^{bv} dv \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1-a)[v^2/2 - bv/(1-a)]} dv \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-([1-a]/2)(v-[b/(1-a)])^2} \quad e^{([1-a]/2)(b/[1-a])^2} dv \\
&= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2\pi}{1-a}}\ e^{(b^2/2)/(1-a)} \quad \text{if} \quad a < 1 \\
&= \frac{1}{\sqrt{1-a}} \quad e^{(b^2/2)/(1-a)}
\end{aligned} \tag{C-3}$$

Since z_j is of the above quadratic form with $a = \rho_0/(\rho_0+\rho_+) < 1$ and $b = \sqrt{\rho_0}\,\rho_+/(\rho_0+\rho_+)$, we have $\overline{e^{z_j}} = \sqrt{(\rho_0+\rho_+)/\rho_+}\exp\left([\rho_0\rho_+/2]/[\rho_0+\rho_+]\right)$. Thus, the approximate capacity expression is

$$\begin{aligned}
C \approx\ & \log_2 M - E_{v_1|\mathbf{x}_1} \log_2 \left[1 + (M-1)\sqrt{\frac{\rho_0+\rho_+}{\rho_+}} \exp\left(\frac{\frac{\rho_0\rho_+}{2}}{\rho_0+\rho_+} \right) e^{-z_1} \right] \\
=\ & \log_2 M - E_{v_1|\mathbf{x}_1} \log_2 \left[1 + (M-1)\sqrt{\frac{\rho_0+\rho_+}{\rho_+}} \exp\left(\frac{\rho_0\rho_+ - 2\rho_+\sqrt{\rho_0} v_1 - \rho_0 v_1^2}{2(\rho_0+\rho_+)} \right) \right]
\end{aligned} \tag{C-4}$$

where v_1 given $\mathbf{x}_1$ is $N(\sqrt{\rho_0}, [\rho_0+\rho_+]/\rho_+)$, as in Eq. (28). Equation (C-4) is valid for the AWGN-2 channel and reduces to the AWGN-1 channel expression in Eq. (24) in the limit as $\rho_+ \to \infty$, with ρ_0 replaced by ρ.