Deep-Space Optical Communications Downlink Budget: Modulation and Coding

B. Moision[1] and J. Hamkins[1]

A link budget for a deep-space optical channel depends in part on the choice of modulation format and error-control coding scheme. This article describes several properties of the channel capacity that lead to an appropriate selection of modulation format, pulse-position modulation (PPM) order, and error-control code rate. It also describes performance limits when additional constraints—such as bounds on average power, peak power, and uncoded symbol-error rate—are imposed. We compare these limits to the performance of Reed–Solomon codes and convolutional codes concatenated with PPM, and show that, when iteratively decoded, the concatenated convolutional codes operate approximately 0.5 dB from capacity over a wide range of signal levels, about 2.5 dB better than Reed–Solomon codes.

I. Introduction

NASA is developing optical links to support deep-space communication to a satellite orbiting Mars at data rates on the order of 10 to 100 Mbits/second. The maximum supportable data rate changes by an order of magnitude as the range, atmospheric conditions, Sun-Earth-probe angle and Sun-probe-Earth angle vary. A characterization of these conditions for the Mars link is detailed in a companion article [1]. For optimum data throughput, the modulation and coding should be correspondingly changed during a mission. In this article, we describe the selection of coding and modulation for various operating points and illustrate achievable user data rates.

As in [1], the transmitted signal is divided into slots of duration T_s seconds, during which a pulse may be transmitted. For the purpose of coding and modulation design, the link may be parameterized by the maximum average detected signal photons/pulse, P_{pk}, the average detected signal photons/slot, P_{av}, and the average detected noise photons/slot, n_b. P_{pk} and P_{av} are proportional to the peak and average received optical signal power, respectively, and may be thought of as peak and average power constraints. The laser transmitter may have additional constraints on the minimum and maximum separation between transmitted pulses. Efficient signaling under these constraints is discussed in [2] and will not be considered here.

Hence we may reduce each operating point to the triple (P_{av}, P_{pk}, n_b) and determine coding and modulation that maximize the achievable user bits/slot. Units are normalized by the slot width and

[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.

results stated in bits/slot, which may be divided by the slot width in seconds to give bits/second. The analysis may be repeated in a similar manner to maximize bits/(signal)photon, a measure of energy efficiency. We choose to maximize bits/slot, a measure of time, or bandwidth, efficiency, as this is the parameter of primary importance for links of interest—we typically have a fixed average power budget and desire to maximize the channel throughput. Although the design procedure is analogous, a link design that maximizes bits/photon will not necessarily maximize bits/second.

We model the channel as Poisson to simplify presentation and facilitate comparisons with known results. More accurate models exist for particular detectors, e.g., the Webb+Gaussian model for avalanche photodiode (APD) detectors [3] and the Polya model for photomultiplier tubes (PMTs) [4], and may be used to provide more accurate link estimates. Much of the analysis here extends in a straightforward manner to other models.

The optical detector output is quantized and synchronized, providing either slot counts or symbol estimates to the channel decoder. Using this information, the decoder, which may also perform the function of demodulation, generates estimates of the transmitted user data. The fidelity of the system may be measured in terms of the bit-error rate (BER) at the output of the decoder, with a target typically on the order of 10^{-5} to 10^{-6}. Where relevant, link budgets in this article require a BER $\leq 10^{-6}$.

It is well-known that the deep-space optical link operates efficiently at high peak-to-average power ratios, e.g., [5,6], which may be achieved by modulating the data using M-ary pulse-position modulation (PPM). In PPM, each $\log_2 M$ user bits map to the location of a single pulsed slot in an M-slot frame. For most of this article, we assume the data are modulated with PPM and focus on choosing the optimum order. In Section II.E, we illustrate the losses incurred by restricting the modulation to PPM.

The error-control code (ECC) rate is determined by the order and target bits/second. The baseline candidate coding and modulation for the deep-space optical channel is Reed–Solomon (RS) coded pulse-position modulation (PPM), denoted RSPPM [7]. This baseline is compared with an iteratively decoded serial concatenation of a convolutional code and coded PPM, denoted SCPPM. We illustrate gains of $\approx$2.5 dB in average power using SCPPM relative to RSPPM for $n_b = 1.0$.

The article is organized as follows. In Section II, we discuss the behavior of optical channel capacity for a Poisson channel using PPM. We show the impact of peak and average power constraints and dead time on capacity. In Section III, we illustrate a method to choose a modulation order and ECC code rate. In Section IV, we demonstrate the performance of specific coded modulation schemes. In Section V, we show how a symbol-error rate (SER) constraint affects capacity and how to trade off SER and throughput. In Section VI, we work through a sample link design and link budget.

II. Capacity of the PPM Channel

Let X_i be the ith binary input to the optical channel, corresponding to the ith slot, where a 1 denotes the transmission of a pulse in the ith slot and a 0 denotes no pulse. Let Y_i be the corresponding output, the number of photons detected in the ith slot. Channel uses are assumed to be conditionally independent, i.e., for each i, j, $i \neq j$,

$$f_{Y_i Y_j|X_i X_j}(y_i, y_j|x_i, x_j) = f_{Y_i|X_i}(y_i|x_i) f_{Y_j|X_j}(y_j|x_j) \tag{1}$$

where $f_{Y|X}(y|x)$ is the conditional density function of Y given X. For channel models in consideration, $f_{Y_i|X_i}(y_i|0)$ and $f_{Y_i|X_i}(y_i|1)$ do not depend on i; hence, we will use the shorthand $p_0(k) = f_{Y_i|X_i}(k|0)$, $p_1(k) = f_{Y_i|X_i}(k|1)$, to denote the densities for signal and non-signal slots.

The capacity of M-ary PPM on a soft-output channel satisfying Eq. (1) is shown in Appendix A to be

$$C(M) = \frac{1}{M} E_{Y_1,\cdots,Y_M} \log_2 \left[\frac{ML(Y_1)}{\sum_{j=1}^{M} L(Y_j)} \right] \text{ bits/slot} \tag{2}$$

where Y_1 has distribution $p_1(\cdot)$, Y_i has distribution $p_0(\cdot)$ for all $i > 1$, and $L(y) = p_1(y)/p_0(y)$ is the likelihood ratio for y.

A. Poisson Channel

Let λ_s denote the average number of signal photons incident on the detector per second (including both pulsed and non-pulsed periods in the average), λ_b the average number of noise photons incident per second, η the quantum efficiency of the detector, T_s the slot time, and $1/M$ the duty cycle, the average number of pulses transmitted per T_s. For a PPM channel, the average number of signal photons per pulse is $n_s = \eta\lambda_s M T_s$, and the average number of noise photons per slot time is $n_b = \eta\lambda_b T_s$. We assume throughout that the pulse energy is captured in a slot, so that n_s also denotes the mean number of signal photons per pulsed slot.

For the Poisson channel

$$p_0(k) = \frac{e^{-n_b} n_b^k}{k!} \tag{3}$$

$$p_1(k) = \frac{e^{-(n_s+n_b)} (n_s + n_b)^k}{k!} \tag{4}$$

Let $C_{\text{PPM}}(M, n_s, n_b)$ be the capacity of the M-ary PPM Poisson channel. The behavior of the case $n_b = 0$ is distinct, so we will carry analysis for the two cases $n_b = 0$ and $n_b > 0$ in parallel. When $n_b = 0$, we have

$$p_0(k) = \begin{cases} 1 & k = 0 \\ 0 & k > 0 \end{cases} \tag{5}$$

$$L(k) = \begin{cases} e^{-n_s} & k = 0 \\ \infty & k > 0 \end{cases} \tag{6}$$

and Eq. (2) collapses to

$$C_{\text{PPM}}(M, n_s, 0) = \frac{\log_2 M}{M} (1 - e^{-n_s}) \text{ bits/slot} \tag{7}$$

When $n_b > 0$, we have $L(k) = e^{-n_s} \left(1 + [n_s/n_b]\right)^k$, and Eq. (2) becomes

$$C_{\text{PPM}}(M, n_s, n_b) = \frac{\log_2 M}{M} \left(1 - \frac{1}{\log_2 M} E_{Y_1,\cdots,Y_M} \log_2 \left[\sum_{i=1}^{M} \left(1 + \frac{n_s}{n_b} \right)^{(Y_j - Y_1)} \right] \right) \text{ bits/slot} \tag{8}$$

B. Properties of $C_{\mathrm{PPM}}(M,n_s,n_b)$

1. Asymptotic Slope and Upper Shell. Figures 1 and 2 illustrate C_{PPM} as a function of $P_{av} = n_s/M$ for backgrounds $n_b = 0$, $n_b = 1$, and orders $M \in \{2, 4, \cdots, 2048\}$. Each curve has a horizontal asymptote at $\log_2 M/M$, the limit imposed by restricting the input to M-ary PPM. In Appendix B, we show the asymptotic slope for small n_s/M in the log-log domain is 1 for $n_b = 0$ and 2 for $n_b > 0$, as shown in the figures.

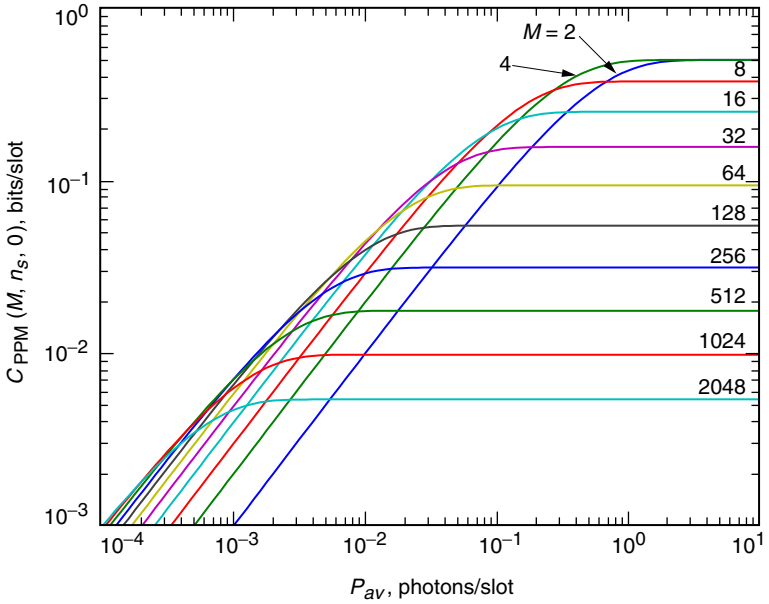

Fig. 1. Capacity for Poisson PPM channel,
$C_{\mathrm{PPM}}(M, n_s, n_b)$, $n_b = 0$, $M \in \{2, 4, \cdots, 2048\}$.

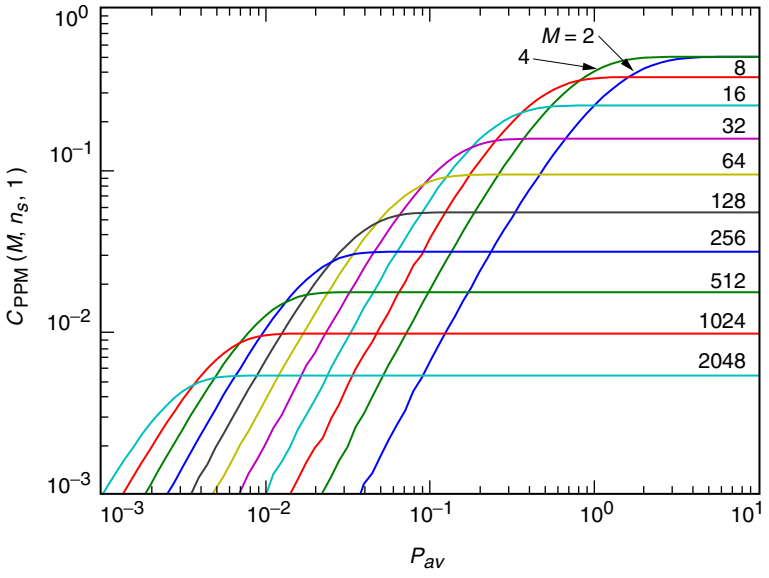

Fig. 2. Capacity for Poisson PPM channel,
$C_{\mathrm{PPM}}(M, n_s, n_b)$, $n_b = 1.0$, $M \in \{2, 4, \cdots, 2048\}$.

Let

$$C_{\mathrm{PPM}}(P_{av}, n_b) = \max_m C_{\mathrm{PPM}}(2^m, P_{av} \cdot 2^m, n_b)$$

the capacity maximized over order $M = 2^m$. $C_{\mathrm{PPM}}(P_{av}, 0)$ and $C_{\mathrm{PPM}}(P_{av}, 1)$ are the upper shells of the functions in Figs. 1 and 2 and are illustrated in Fig. 3 along with $C_{\mathrm{PPM}}(P_{av}, n_b)$ for $n_b \in \{0.01, 0.1, 10\}$.

2. Concavity. For $n_b = 0$, we have, from Eq. (7),

$$\frac{\partial^2}{\partial n_s^2} C_{\mathrm{PPM}}(M, n_s, 0) = -\frac{\log M}{M} e^{-n_s} < 0$$

hence, $C_{\mathrm{PPM}}(M, n_s, 0)$ is concave in n_s. In Appendix B, we show that $\lim_{n_s \to 0}(\partial/\partial n_s) C_{\mathrm{PPM}}(M, n_s, n_b) = 0$ for $n_b > 0$. Since C_{PPM} is non-decreasing in n_s and bounded, $\lim_{n_s \to \infty}(\partial/\partial n_s) C_{\mathrm{PPM}}(M, n_s, n_b) = 0$. Since the derivatives go to zero at the boundaries and C_{PPM} is nonzero for $0 < n_s < \infty$, it follows that $C_{\mathrm{PPM}}(M, n_s, n_b)$ is not concave in n_s for $n_b > 0$. This has implications for time-sharing the channel, discussed in Section II.D.

3. Small n_s Behavior. We may obtain an estimate of C_{PPM} for small n_s by expanding $C_{\mathrm{PPM}}(M, n_s, n_b)$ in a Taylor series about $n_s = 0$,

$$\begin{aligned} & C_{\mathrm{PPM}}(M, n_s, n_b) \\ & = C_{\mathrm{PPM}}(M, 0, n_b) + n_s \lim_{n_s \to 0} \frac{\partial}{\partial n_s} C_{\mathrm{PPM}}(M, n_s, n_b) + \frac{n_s^2}{2} \lim_{n_s \to 0} \frac{\partial^2}{\partial n_s^2} C_{\mathrm{PPM}}(M, n_s, n_b) + O(n_s^3) \\ & = \left(\frac{M-1}{2M^2 \ln 2} \right) \frac{n_s^2}{n_b} + O(n_s^3) \qquad (9) \\ & \approx \left(\frac{M-1}{2M^2 \ln 2} \right) \frac{n_s^2}{n_b} \text{ bits/slot} \qquad (10) \end{aligned}$$

for small n_s. In Eq. (9), we used the derivation of $\lim_{n_s \to 0}(\partial^2/\partial n_s^2) C_{\mathrm{PPM}}(M, n_s, n_b)]$ given in Appendix B, Eq. (B-1), being careful to divide by M to convert bits/symbol to bits/slot. We see from Eq. (10) that the capacity goes as n_s^2/n_b. In this region, each 1-dB increase in signal power compensates for a 2-dB increase in noise power. The dependence on slot width is shown by substituting for n_s, n_b in Eq. (10) and converting units to bits/second,

$$\begin{aligned} C_{\mathrm{PPM}}(M, n_s, n_b) & \approx \left(\frac{M-1}{2M^2 \ln 2} \right) \frac{(\eta \lambda_s M T_s)^2}{\eta \lambda_b T_s} \frac{1}{T_s} \text{ bits/second} \\ & = \left(\frac{M-1}{2 \ln 2} \right) \frac{\eta \lambda_s^2}{\lambda_b} \text{ bits/second} \end{aligned}$$

Hence, the capacity is invariant to slot width and linear in the quantum efficiency for small n_s. In this region, we may increase the slot width for fixed M to simplify receiver implementation while paying a negligible penalty in capacity.

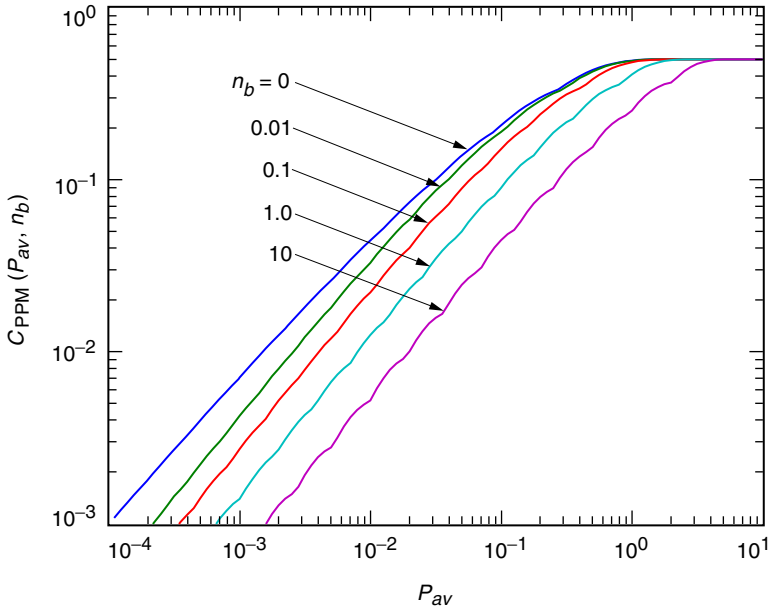

Fig. 3. Capacity maximized over order, $C_{\text{PPM}}(P_{av}, n_b)$, $n_b \in \{0, 0.01, 0.1, 1, 10\}$.

C. Peak and Average Power Constraints

The laser transmitter will have peak and average power constraints imposed by physical limitations and available resources. Assume the losses between the transmitter and receiver are not a function of the pulse energy so that the peak and average transmitter power constraints may be translated into the peak and average detected power constraints P_{pk}, P_{av}, and the peak-to-average power constraint $P_{pa} = P_{pk}/P_{av}$ is constant. Recall n_s denotes the average detected photons per pulse, and $1/M$ the average number of pulses per slot. For the constraints to be satisfied, we must have

$$\frac{n_s}{M} \leq P_{av}$$

$$n_s \leq P_{pk}$$

which imply $n_s \leq \min\{MP_{av}, P_{pk}\}$. P_{pa} divides PPM orders into two regions. If $M \leq P_{pa}$, then $n_s \leq MP_{av} \leq P_{pk}$, and the channel is average power constrained. If $M > P_{pa}$, then $n_s \leq P_{pk} \leq MP_{av}$, and the channel is peak power constrained. Since the capacity is non-decreasing in n_s, we choose $n_s = MP_{av}$ for $M \leq P_{pa}$ and $n_s = P_{pk}$ for $M > P_{pa}$. We argue that the capacity maximizing PPM order for a given (P_{av}, P_{pk}, n_b) always satisfies $M \leq P_{pa}$, i.e., that we always operate in the average-power-constrained region.

Assume P_{pa} is a power of 2, let $M^* = P_{pa}$, and choose $M > M^*$. Then for both M and M^* we have $n_s = P_{pk}$. The following result shows that the capacity is strictly smaller by choosing the larger order in a certain region.

Result 1. If $C(M) \geq 1/M$, then $C(M) \geq C(2M)$.

Proof. Suppose $C(M) \geq 1/M$. For the memoryless PPM channel, Eq. (2) yields

$$C(2M) = \frac{1}{2M} E \log_2 \frac{2ML(Y_1)}{\sum_{j=1}^{2M} L(Y_j)}$$

$$\leq \frac{1}{2M} E \log_2 \frac{2ML(Y_1)}{\sum_{j=1}^{M} L(Y_j)}$$

$$= \frac{1}{2}\left(\frac{1}{M} + C(M)\right)$$

$$\leq C(M) \qquad \square$$

If the above result holds for the entire range of $C(M)$, then the capacity maximizing M is always no larger than the peak-to-average power ratio. We conjecture (as confirmed by numerical results, and easily shown to hold for $n_b = 0$) this holds for $C_{\mathrm{PPM}}(2^k, n_s, n_b)$.

Conjecture 1. For integer $k > 2$, $C_{\mathrm{PPM}}(2^k, n_s, n_b)$ is monotonically decreasing in k.

Assuming the conjecture holds, peak and average transmitter constraints impose an effective order constraint $M \leq P_{pa}$. This can be seen as follows. If there is a signaling scheme with average photons per pulse n_s and duty cycle $1/M$ that meets the peak and average power constraints, then $n_s \leq \min\{MP_{av}, P_{pk}\}$. If $M > P_{pa}$ (and P_{pa} is a power of 2), then by the conjecture we can increase the capacity by reducing M to $M = P_{pa}$, without violating the power constraints.

D. The Impact of Dead Time

A Q-switched laser works well with the PPM format because it can successfully confine a large pulse energy to a narrow slot, e.g., [8,9]. One side effect of Q-switched lasers, however, is a required delay, or dead time, between pulses, during which the laser is recharged. We denote the dead-time duration by the integer d, which represents the number of slots of dead time. PPM may be modified to satisfy a dead-time constraint by following each frame with a period during which no pulses are transmitted.

Adding dead time is a special case of time-sharing the channel by transmitting more than one pulse energy. For fixed M and average power, one can transmit pulses with $n_s = n_{s1}$ a fraction $(1-\alpha)$ of the time and $n_s = n_{s2}$ a fraction α of the time, so long as $\alpha n_{s2} + (1-\alpha)n_{s1} = P_{av}M$. Dead time represents the special case $n_{s1} = P_{av}(M+d)$, $n_{s2} = 0$, $\alpha = d/(M+d)$. If the capacity is not concave in P_{av}, then there are regions where time-sharing will gain over using a single value of n_s.

For fixed M and n_s, adding a dead time of d slots decreases the capacity by a factor of $M/(M+d)$. However, for the same average power P_{av}, adding a dead time of d slots allows an increase in n_s from $P_{av}M$ to $P_{av}(M+d)$—assuming this does not violate a peak power constraint. Alternatively, with n_s fixed, the average power is decreased by a factor of $M/(M+d)$. Hence, for each point $(\text{photons/slot}, \text{bits/slot}) = (P_{av}, C_{\mathrm{PPM}}(M, P_{av}M, n_b))$ achievable with PPM, there exists a family of points $(\text{photons/slot}, \text{bits/slot}) = (P_{av}M/(M+d), C_{\mathrm{PPM}}(M, P_{av}M, n_b)M/(M+d))$ achievable by adding dead time. In a log-log domain plot of average power versus capacity, this is represented by extending each point on the C_{PPM} versus P_{av} curve down and to the left with a line of slope 1.

As shown in Section II, $C_{\mathrm{PPM}}(M, n_s, n_b)$ is concave in n_s for $n_b = 0$, but not for $n_b > 0$. It follows that there are regions where time-sharing will increase the capacity for $n_b > 0$, for which dead time is one case (time-sharing will not increase capacity when $n_b = 0$). We will see that adding dead time will increase the capacity for moderate-to-low average power when $n_b > 0$. Let

$$C_d(M, P_{av}, n_b) = \max_d \left\{ C_{\mathrm{PPM}}(M, n_s, n_b) \left. \frac{M}{M+d} \right| n_s = P_{av}(M+d) \right\}$$

the capacity maximized over dead time d, illustrated in Fig. 4 for $M = 64$. C_d is equal to C_{PPM} for P_{av} above the point where the tangent of C_{PPM} has slope 1, and equal to that tangent below that point. We conjecture that C_{PPM} has a single inflection point as a function of n_s. The implication of this is that the optimum capacity under time-sharing is always achieved by using dead time as one of the time-shared operating points (or no time-sharing).

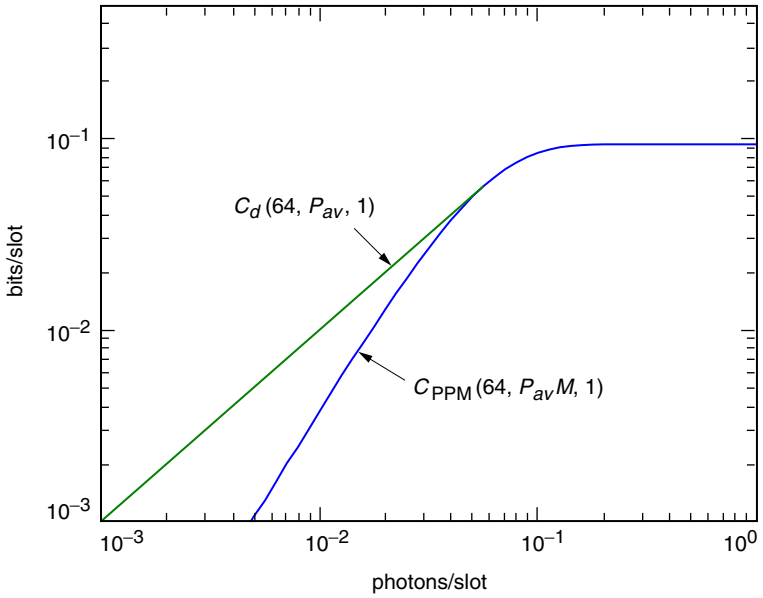

Fig. 4. Achievable rates for $M = 64$, $n_b = 1$, with and without dead time.

E. Suboptimality of PPM

What loss is incurred by restricting the modulation to PPM? PPM is essentially a binary modulation code with duty cycle $1/M$ and a single pulse (binary 1) in each (synchronized) window of M slots. Suppose we were to replace PPM with a binary modulation code with duty cycle $1/M$ but no constraint on the distribution of pulses. What gains are available by allowing an arbitrary pulse distribution?

The capacity of a memoryless Poisson channel with input restricted to duty cycle $1/M$ is

$$C_{\mathrm{OOK}}(M, n_s, n_b) = \frac{1}{M} E_{Y|1} \log \frac{f_{Y|X}(Y|1)}{f_Y(Y)} + \frac{M-1}{M} E_{Y|0} \log \frac{f_{Y|X}(Y|0)}{f_Y(Y)}$$

where $f_{Y|X}(y|0), f_{Y|X}(y|1)$ are given by Eqs. (3) and (4), and

$$f_Y(y) = \frac{1}{M} f_{Y|X}(y|1) + \frac{M-1}{M} f_{Y|X}(y|0)$$

is the probability mass function for a randomly chosen slot.

Figures 5 and 6 illustrate $C_{\mathrm{OOK}}(M, n_s, n_b)$ for backgrounds $n_b = 0$, $n_b = 1$ and M a power of 2. Let $h(p)$ be the entropy function, $h(p) = p\log_2(1/p) + (1-p)\log_2(1/(1-p))$. $C_{\mathrm{OOK}}(M, n_s, n_b)$ has a horizontal asymptote at $h(1/M)$, the limit imposed by restricting the input to duty cycle $1/M$. The duty cycle may be any positive real value, the restriction to powers of 2 being an artifice for comparison with PPM orders. Let

$$C_{\mathrm{OOK}}(P_{av}, n_b) = \max_M C_{\mathrm{OOK}}(M, P_{av}, n_b)$$

the capacity maximized over real-valued order M, illustrated in Fig. 7 for $n_b \in \{0, 0.01, 0.1, 1, 10\}$.

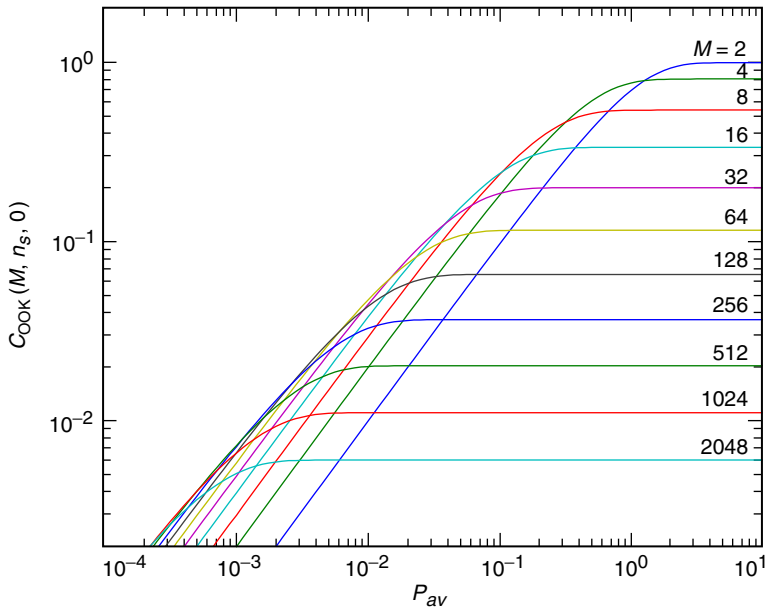

Fig. 5. Capacity for duty-cycle constraint, $C_{\mathrm{OOK}}(M, n_s, n_b)$, $n_b = 0$, $M \in \{2, 4, \cdots, 2048\}$.

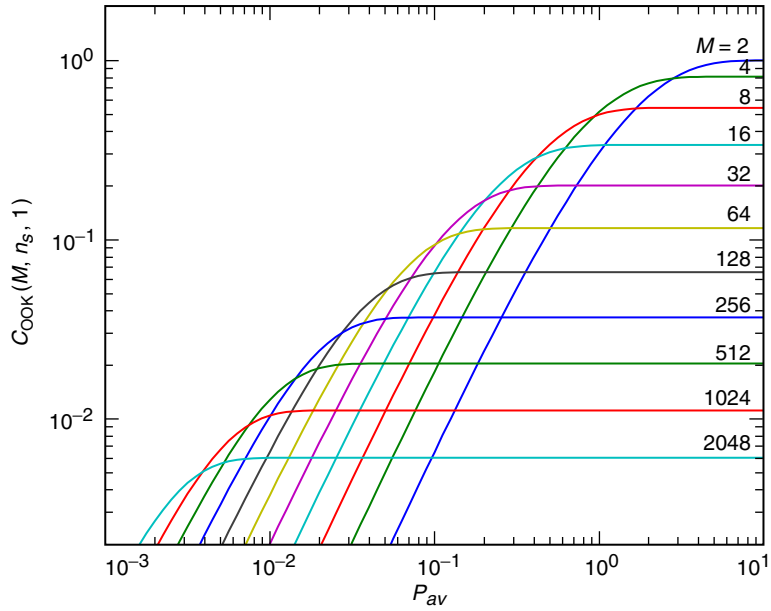

Fig. 6. Capacity for duty-cycle constraint, $C_{\mathrm{OOK}}(M, n_s, n_b)$, $n_b = 1.0$, $M \in \{2, 4, \cdots, 2048\}$.

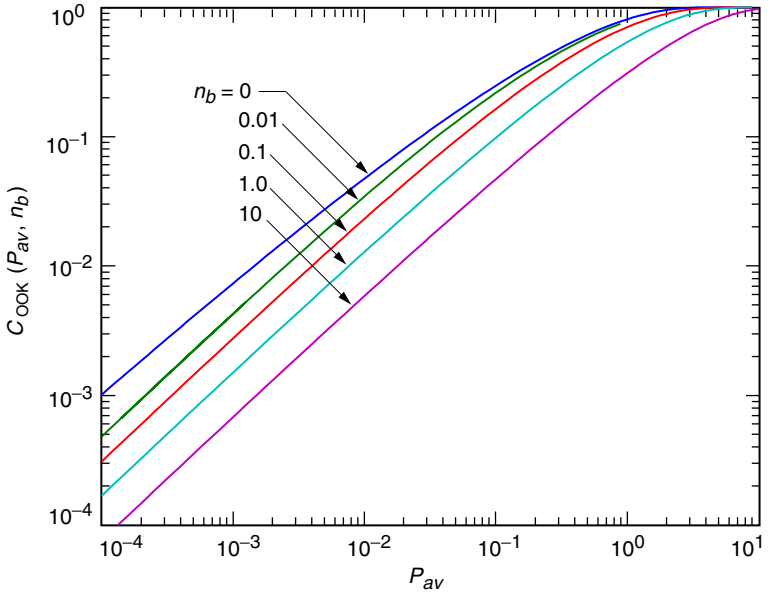

Fig. 7. Duty-cycle-constrained capacity maximized over order, $C_{OOK}(P_{av}, n_b)$, $n_b \in \{0, 0.01, 0.1, 1, 10\}$.

Figure 8 illustrates $C_{\mathrm{OOK}}(P_{av}, n_b)/C_{\mathrm{PPM}}(P_{av}, n_b)$, the potential gain in using an arbitrary duty-cycle constraint relative to PPM. The gains are larger for high average power, corresponding to small PPM orders, and for smaller background noise levels. We can potentially double the capacity for moderate-to-high average power. We note, however, that in this discussion we have not specified codes that achieve arbitrary duty cycles. There are systematic methods to construct such codes, e.g., [10], but we will not explore their use here. We illustrate results of Fig. 8 to demonstrate regions where their use should be explored. In the remainder of the article, we discuss results for PPM.

III. Parameter Selection

We may operate close to capacity by choosing the optimum PPM order for the desired bits/slot and concatenating the PPM mapping with an error-correction code (ECC) so that the decoded BER satisfies some threshold. The optimal order and ECC rate vary with n_s and n_b. To optimize the throughput over the course of a mission requires choosing a set of modulation orders and corresponding ECCs.

A. PPM Order

In the absence of a peak power constraint, we choose to use the capacity-maximizing PPM order,

$$M^*(P_{av}, n_b) = 2^{\arg\max_m C_{\mathrm{PPM}}(2^m, MP_{av}, n_b)}$$

illustrated in Fig. 9 for $n_b \in \{0, 0.01, 0.1, 1.0, 10\}$. Discontinuities correspond to switching the order and are shown as vertical lines. M^* increases with increasing n_b, showing that as the channel becomes worse, an efficient use of the channel increases the peak pulse power. Similarly, M^* decreases with P_{av}.

We may also characterize the optimal duty cycle for an input not constrained to use PPM, as discussed in Section II.E. The corresponding optimal duty cycle,

$$M^*_{\mathrm{OOK}}(P_{av}, n_b) = \arg\max_M C_{\mathrm{OOK}}(M, MP_{av}, n_b)$$

is illustrated in Fig. 10.

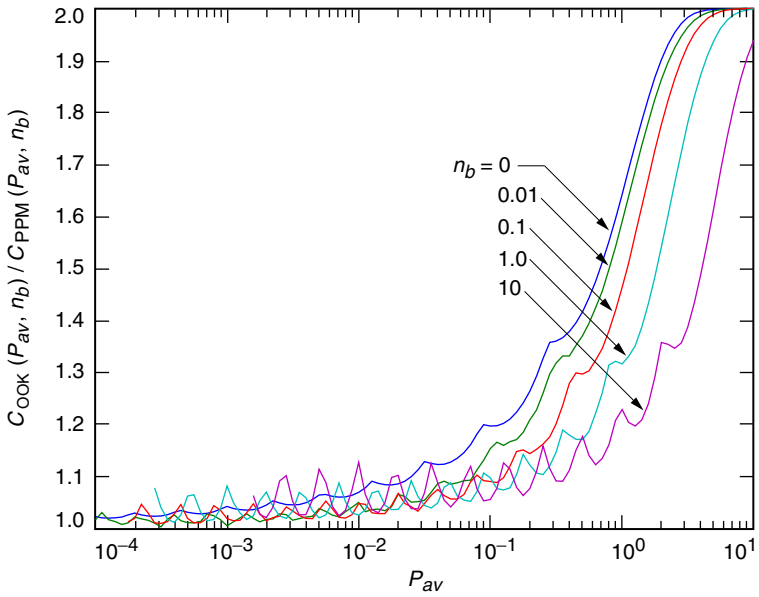

Fig. 8. Relative loss due to using PPM,
$C_{OOK}(P_{av}, n_b) / C_{PPM}(P_{av}, n_b)$, $n_b \in \{0, 0.01, 0.1, 1.0, 10\}$.

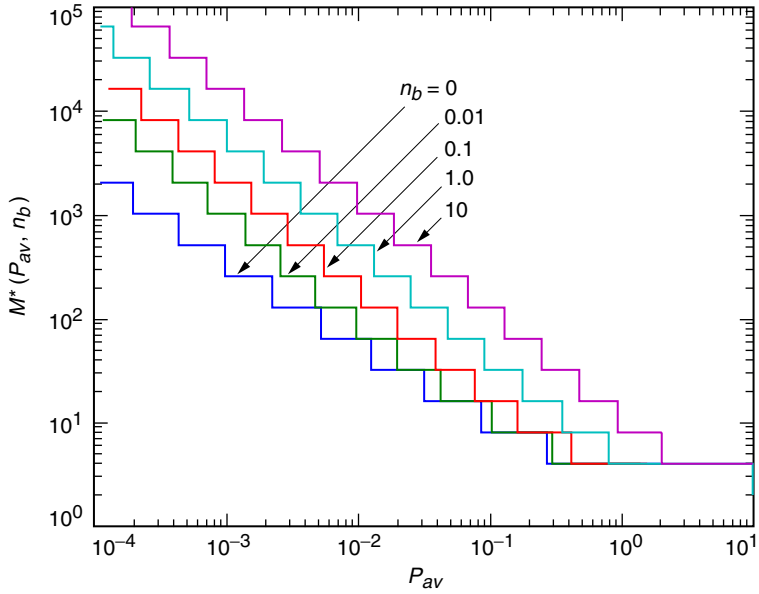

Fig. 9. Optimal PPM order,
$M^*(P_{av}, n_b)$, $n_b \in \{0, 0.01, 0.1, 1.0, 10\}$.

B. Error-Control Code Rate

We choose the ECC code rate to satisfy

$$R_{\mathrm{PPM}}(P_{av}, n_b) \approx \frac{C_{\mathrm{PPM}}(P_{av}, n_b) M^*(P_{av}, n_b)}{\log_2 M^*(P_{av}, n_b)}$$

Figure 11 illustrates the ECC rate as a function of average power for several background levels. The discontinuities, corresponding to switching the order, obscure the general behavior, which is more clearly

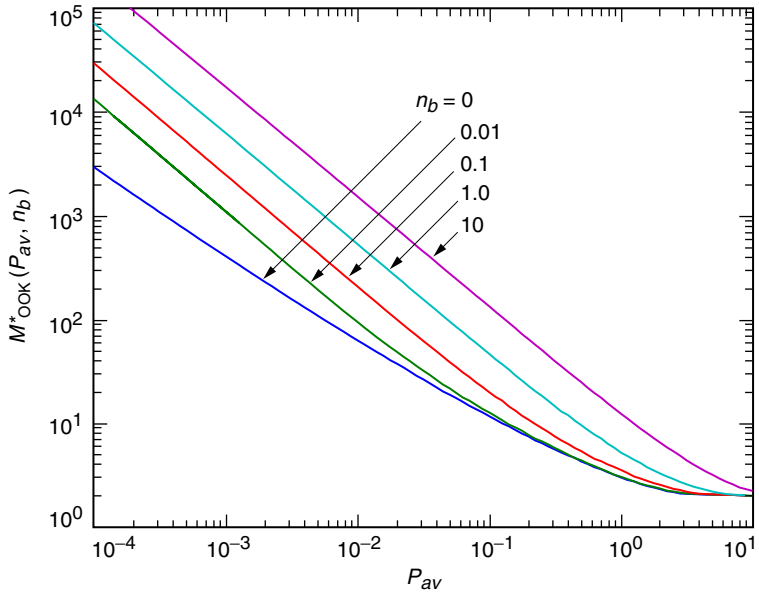

Fig. 10. Optimal real-valued duty cycle, $M^*_{\mathrm{OOK}}(P_{av}, n_b)$, $n_b \in \{0, 0.01, 0.1, 1.0, 10\}$.

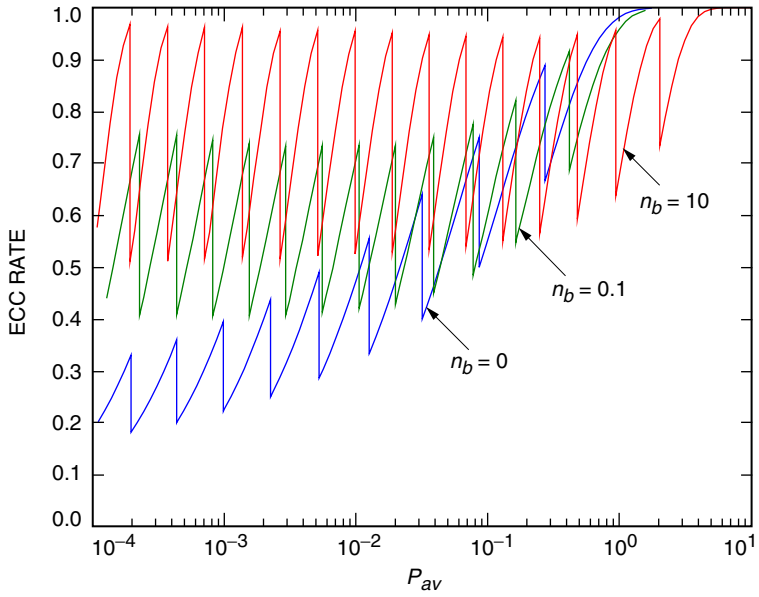

Fig. 11. PPM ECC rates, $R_{\mathrm{PPM}}(P_{av}, n_b)$, $n_b \in \{0, 0.1, 10\}$.

illustrated in Fig. 12 for the duty-cycle-constrained case. The average ECC rate generally increases as the background level increases.

As the average power available increases, the optimal PPM order decreases, switching from order $2M$ to order M at the intercept of the corresponding capacity curves. In the noiseless case, this intercept, where $C_{\mathrm{PPM}}(M, n_s, 0) = C_{\mathrm{PPM}}(2M, 2n_s, 0)$, occurs at $\tilde{n}_s = \log\left(\log_2(2M)/\log_2(M/2)\right)$, $C_{\mathrm{PPM}}(\tilde{n}_s M, 0) = 2\log_2 M/(M\log_2(2M))$, with a corresponding discontinuity in the ECC rates, $R_{\mathrm{PPM}}(\tilde{n}_s M, 0) = C_{\mathrm{PPM}}(\tilde{n}_s M, 0)M/\log_2(M) = 2/(1+\log_2 M)$, $R_{\mathrm{PPM}}(\tilde{n}_s 2M, 0) = C_{\mathrm{PPM}}(\tilde{n}_s 2M, 0)2M/\log_2(2M) = 4\log_2 M/(\log_2(2M))^2$, which indicates the range of rates used in the noiseless case.

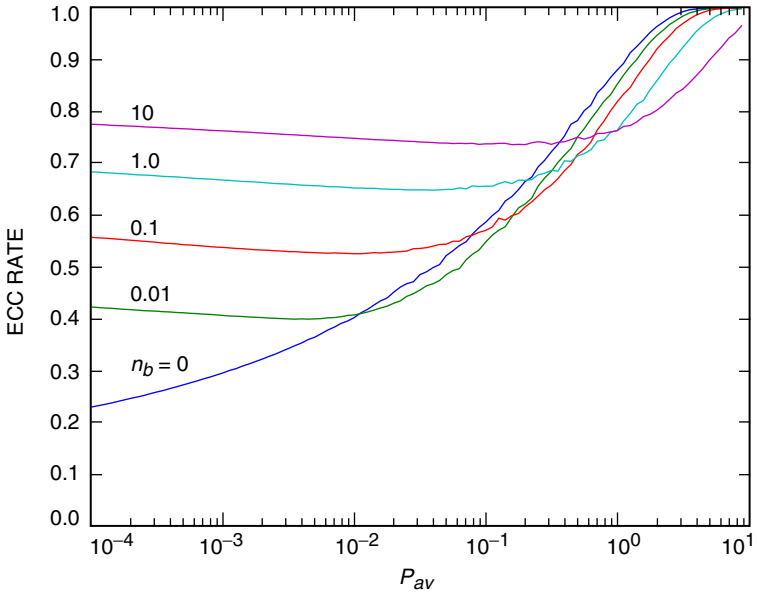

Fig. 12. Duty-cycle-constrained ECC rates, $R_{\mathrm{OOK}}(P_{av}, n_b)$, $n_b \in \{0, 0.01, 0.1, 1.0, 10\}$.

The behavior of the ECC rate is obscured in Fig. 11 due to switching of the order. The behavior of the rate can be seen more clearly by considering the duty-cycle-constrained case, which allows any real-valued duty cycle. The capacity and duty cycle may be used to specify an ECC data rate with

$$R_{\mathrm{OOK}}(P_{av}, n_b) = \frac{C_{\mathrm{OOK}}(P_{av}, n_b)}{h\left(\dfrac{1}{M^*_{OOK}(P_{av}, n_b)}\right)}$$

illustrated in Fig. 12 for $n_b \in \{0, , 0.01, 0.1, 1, 10\}$.

IV. Coded Performance—Gap to Capacity

Figure 11 shows that a broad range of ECC rates may be required to support various operating points. We evaluated two classes of codes to support the required rates. A class of soft-decision iterative codes and a class of hard-decision Reed–Solomon codes.

The two classes of decoders require different receiver/demodulators. Soft-decision decoders nominally require slot counts for each slot, although the complexity may be mitigated by using a subset of the slot counts [11]. The receiver does not make preliminary PPM symbol decisions, but passes on slot counts to the decoder. The soft-decision algorithm is initialized by using the slot counts to determine the probability of each candidate PPM symbol. A receiver for a hard-decision decoder is initialized with estimates of each PPM symbol, which may be determined by the receiver, but no explicit slot counts. Hard-decision receiver/decoders are generally less complex than their soft-decision counterparts, but have worse performance. To quantify the performance–complexity trade-off, we consider both classes.

A. Hard-Decision Decoded Codes

With hard-decision decoders, the minimum distance and, to a lesser degree, the number of nearest neighbors are the critical parameters. Reed–Solomon (RS) codes are optimal hard-decision codes in that

they have the largest minimum distance (Hamming distance between code symbols) for any code with the same rate, block length, and field order.

An (n, k) RS code (rate $R = k/n$) is conventionally tailored to fit an M-ary PPM channel by choosing RS code symbols from $GF(M)$ and using $n = M-1$, so that there is a one-to-one correspondence between PPM symbol errors and code-word symbol errors [7]. However, by following this convention, small PPM orders imply short block lengths, and codes for small n perform poorly and have less flexibility in choosing the rate (with $M = 4$, this convention would yield the small class of $(3, k)$RS codes). Instead, we design an RS code to be used in conjunction with M-PPM by grouping together β M-PPM symbols to form an element of $GF(M^\beta)$. The RS code is then taken to be an $(n, k) = (M^\beta - 1, k)$ code. The optimum choice of β will be a function of the target bit-error rate. We refer to the concatenation of an RS code with PPM formed in this manner as an RSPPM(n, k, M) code.

Figure 13 shows the performance of rate $\approx 3/5$ RSPPM with $M = 64$ for $\beta \in \{1, 2, 3\}$, along with soft-decision capacity of rate 3/5 coded 64-PPM. We would conventionally use the RSPPM$(63, 37, 64)$ code, which matches 64-PPM, but the RSPPM$(26143, 15685, 64)$ code does 0.41 dB better at a BER of 10^{-5}. As the block length increases, the waterfall region is seen to move slightly to the right and to get steeper, so that the optimal β is a function of the target BER. For example, with 64-PPM and a target BER of 10^{-5}, RSPPM$(26143, 15685, 64)$ outperforms the smaller RSPPM$(4095, 2457, 64)$ and RSPPM$(63, 37, 64)$ codes as well as the larger RSPPM$(16777215, 10066329, 64)$ code. For small orders of PPM, using RSPPM codes with long block lengths can result in multiple-dB gains over RS codes matched to the PPM size.

B. Iterative Soft-Decision Decoders

Two sub-classes of iterative codes should be used for the optical channel, one for the region of very small n_b (approximate range $0 \leq n_b \leq 0.1$) and one for moderate-to-large n_b (approximate range $n_b > 0.1$). For moderate-to-large n_b, the performance of the codes is dominated by the number of nearest neighbors—the distance distribution. In this region, we propose to use the serially concatenated codes described here. For very small n_b, the channel behaves more like an erasure channel (for $n_b = 0$, it reduces to an erasure channel) and is dominated by shot noise. In this region, the

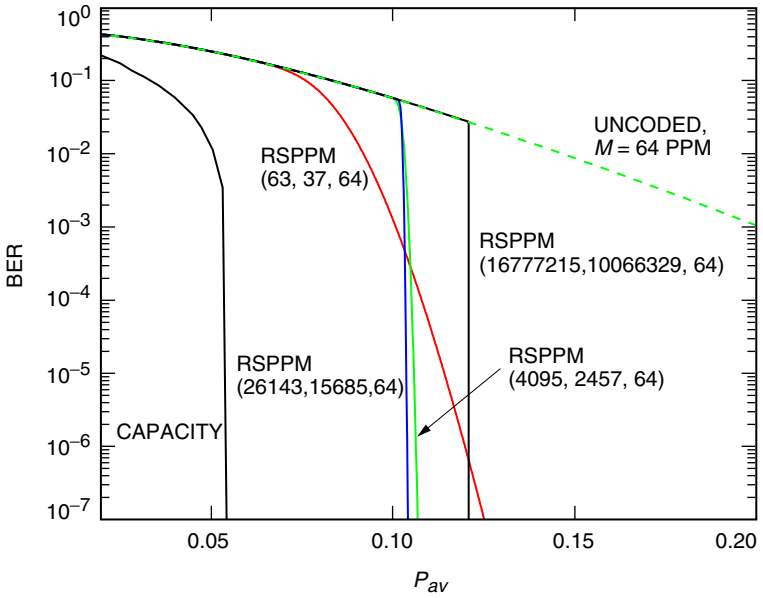

Fig. 13. Performance of rate ≈ 3/5 RSPPM, $M = 64$.

minimum distance of the code dominates performance, and we propose to use low-density parity-check codes with good distance properties, e.g., [12].

Our current case of interest is in the moderate-to-high background noise region. We propose the use of iterative codes for this region that are composed of a serial concatenation of a small memory convolutional code, a bit interleaver, a recursive rate-1 accumulator, and mapping to PPM symbols. The iterative codes are described in more detail in [2,11]. We refer to the codes as SCPPM(R, m, M, Π), where R, m are the convolutional code rate and memory, M is the PPM order, and Π is the bit-interleaver length. The convolutional code is treated as an outer code and the accumulator and PPM mapping as an inner code, the pair being decoded iteratively following the description in [13]. For convolutional codes with rates above 1/2, we decode using the dual of the code, e.g., [14], to reduce complexity.

The BER performance of an SCPPM code may be improved up to a point by increasing the interleaver size, using a designed interleaver and optimizing over choice of convolutional code. We note that the convolutional codes considered have small memory—typically 2 or 3—and, hence, low-complexity encoders and decoders.

V. Symbol-Error Rates, Receiver Losses

The performance of the iterative error-correcting codes described in Section IV.B theoretically allow operation at signal powers within a fraction of a dB of capacity. However, this assumes no losses in the receiver—i.e., perfect synchronization and no loss due to quantization. In a link-budget analysis, we need to incorporate losses due to using a non-ideal receiver. Quantization loss may be kept negligibly small by choosing a sufficient number of levels, albeit at increasing complexity costs. We anticipate 5- to 6-bit quantization of likelihood ratios will be sufficient to ensure negligible degradation [15]. However, synchronization at operating points within a fraction of a dB of capacity may be infeasible and become the weak link in the receiver. Synchronization depends on the ability to correctly resolve a signal pulse location, which we may characterize in terms of the uncoded symbol-error rate (SER). If the synchronization loss is excessive, we may switch to a PPM order larger than optimal, sacrificing bits/slot for a higher energy pulse and, hence, more reliable estimates of pulse positions. In this section, we address methods to decrease synchronization losses by changing the modulation, trading off data rate for lower SERs.

The maximum-likelihood symbol decision chooses the PPM symbol corresponding to the maximum slot count, with probability of error [16]

$$\mathrm{SER}(M, n_s, n_b) = 1 - \frac{1}{M} e^{-(n_s + M n_b)} - \sum_{j=1}^{\infty} \frac{(n_s + n_b)^j}{j!} e^{-(n_s+n_b)} \left[\sum_{m=0}^{j-1} \frac{n_b^m}{m!} e^{-n_b} \right]^{M-1} \frac{1}{aM} \left[(1+a)^M - 1 \right]$$

where

$$a_j = \frac{\dfrac{n_b^j}{j!}}{\sum_{m=0}^{j-1} \dfrac{n_b^m}{m!}}$$

This can be simplified to

$$\mathrm{SER}(M, n_s, n_b) = 1 - \sum_{j=0}^{\infty} \frac{e^{-n_s}}{M} \left(1 + \frac{n_s}{n_b} \right)^j \left(b_{j+1}^M - b_j^M \right)$$

where

$$b_j = \sum_{m=0}^{j-1} \frac{n_b^m}{m!} e^{-n_b}$$

Recall $M^*(P_{av}, n_b)$ denotes the optimal order and let

$$n_s^*(P_{av}, n_b) = P_{av} M^*(P_{av}, n_b)$$

the corresponding mean signal photons per signal slot. Figure 14 illustrates $\mathrm{SER}(M^*, n_s^*, n_b)$ as a function of P_{av} for $n_b = 0$, 0.1, and 10. Discontinuities due to switching orders are illustrated as vertical lines. For high P_{av}, the uncoded performance is very good—so good in fact that little or no coding is required. However, over a broad range of interest, we would, without other constraints, choose to operate with uncoded SERs above 0.1. It is the low P_{av} regime that causes timing losses. The uncoded SER at low average powers may be as high as 0.6. The high SER may be significantly aggravated by peak power constraints, which restrict the maximum achievable order.

A. Trading SER for Throughput, Unconstrained Order

What loss is incurred by adding a constraint on the SER? Assume there is no order constraint—no peak power transmitter constraint, so we may vary n_s and M as long as $n_s/M = P_{av}$. In this case, we may lower the uncoded SER by increasing M and n_s relative to M^* and n_s^*, trading off SER for data rate. Let

$$C_{\mathrm{SER}}(P_{av}, n_b, \alpha) = \max_m \{C_{\mathrm{PPM}}(2^m, n_s, n_b) | n_s = P_{av} 2^m, \mathrm{SER}(2^m, n_s, n_b) < \alpha\}$$

the capacity of the PPM channel constrained to have SER less than α. Figure 15 illustrates C_{SER} $(P_{av}, 1.0, \alpha)$ for $\alpha \in \{1, 0.1, 0.01\}$, and Fig. 16 illustrates the corresponding SERs.

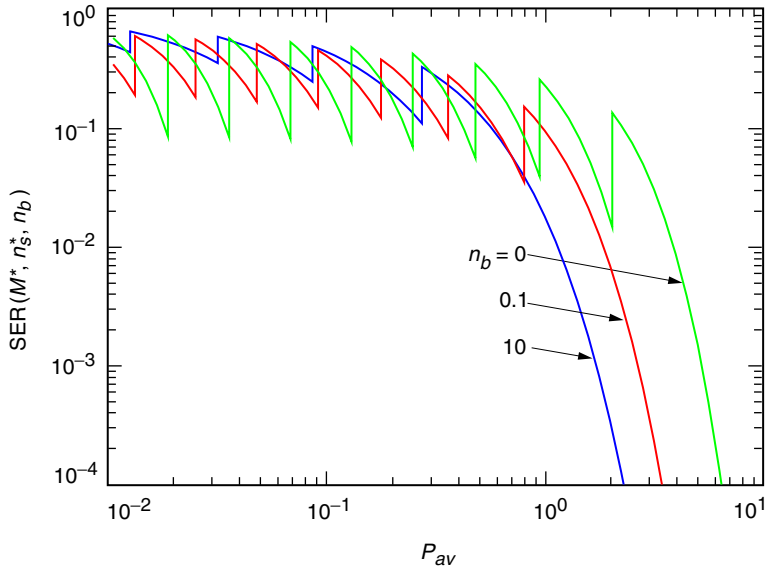

Fig. 14. SER for optimum operating points.

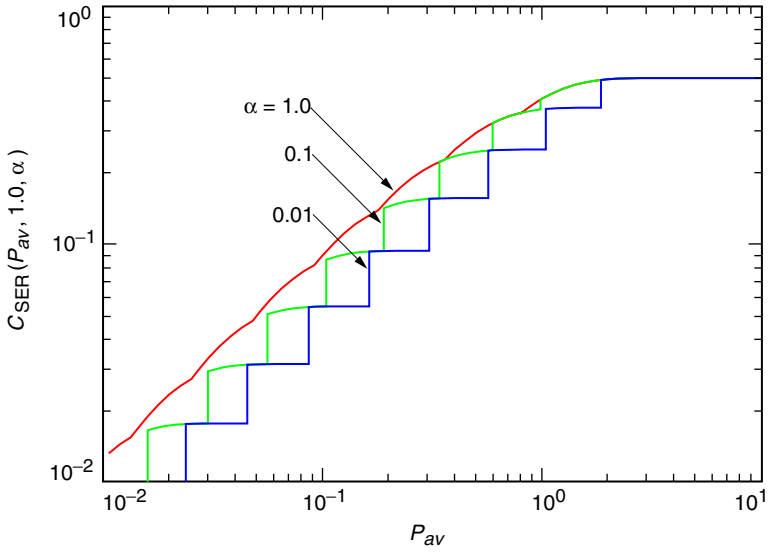

Fig. 15. SER-constrained PPM capacity,
$C_{\mathrm{SER}}(P_{av}, n_b, \alpha)$, $n_b = 1.0$, $\alpha \in \{1, 0.1, 0.01\}$.

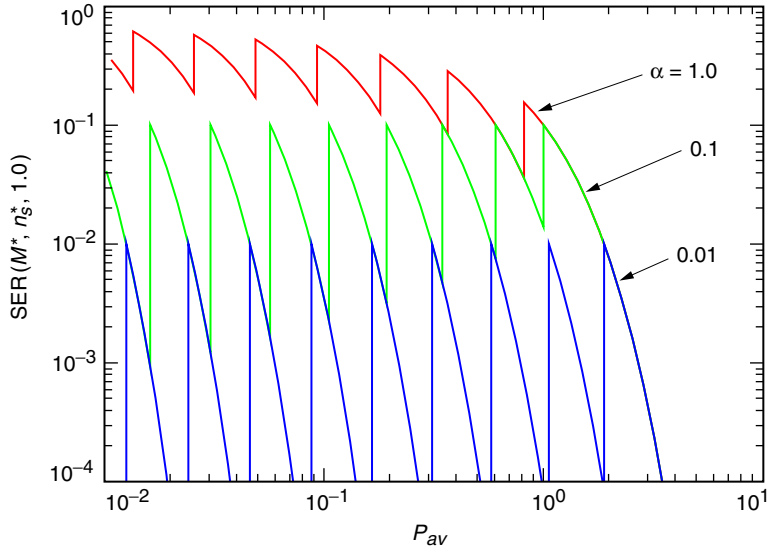

Fig. 16. SER corresponding to
$C_{\mathrm{SER}}(P_{av}, n_b, \alpha)$, $n_b = 1.0$, $\alpha \in \{1, 0.1, 0.01\}$.

Discontinuities occur at the point where we switch orders and are illustrated as vertical lines. The curve may be smoothed by time-sharing PPM orders, i.e., we may use order M_1 a fraction β of the time and order M_2 a fraction $(1-\beta)$ of the time. This yields achievable capacities

$$C^*_{\mathrm{SER}}(P_{av}, n_b, \alpha) = \max_{M_1, M_2, \beta} \{\beta C_{\mathrm{PPM}}(M_1, P_{av,1} M_1, n_b) + (1-\beta) C_{\mathrm{PPM}}(M_2, P_{av,2} M_2, n_b)|$$

$$P_{av} = \beta P_{av,1} + (1-\beta) P_{av,2}, \mathrm{SER}(M_1, P_{av,1} M_1, n_b) < \alpha, \mathrm{SER}(M_2, P_{av,2} M_2, n_b) < \alpha\} \tag{11}$$

the concave hull of Fig. 15, illustrated in Fig. 17. Figure 18 shows the corresponding SERs. Time-sharing PPM orders allows all real-valued duty cycles, filling in the discontinuities due to switching orders. In

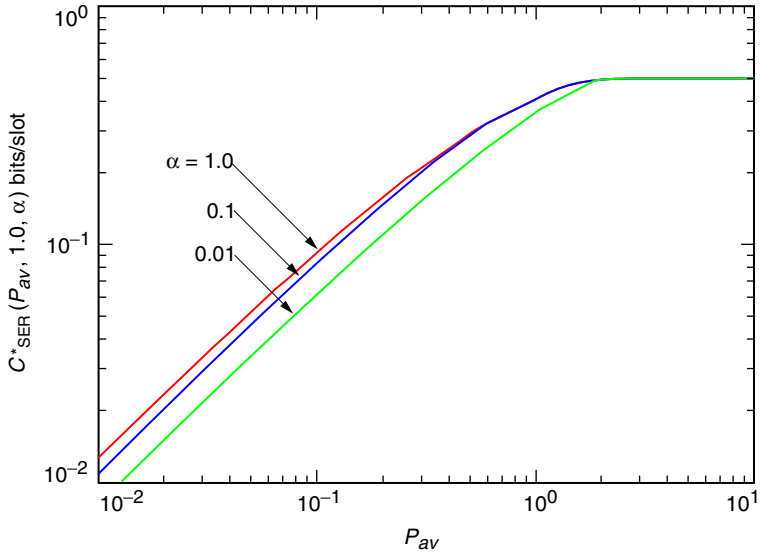

Fig. 17. SER-constrained time-sharing PPM capacity, $C^*_{SER}(P_{av}, n_b, \alpha)$, $n_b = 1.0$, $\alpha \in \{1, 0.1, 0.01\}$.

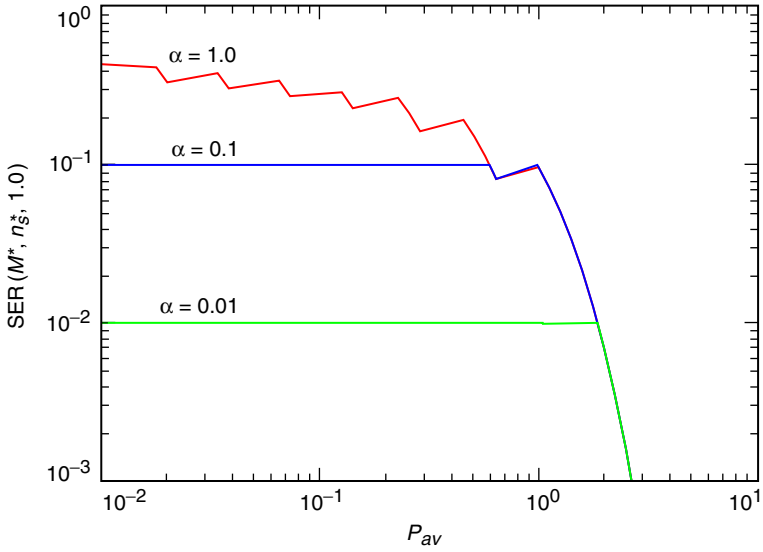

Fig. 18. SER corresponding to, $C^*_{SER}(P_{av}, n_b, \alpha)$, $n_b = 1.0$, $\alpha \in \{1, 0.1, 0.01\}$.

Eq. (11), we time-share points that individually satisfy the SER constraint. This may be extended by time-sharing points such that the mean SER satisfies the constraint. We do not treat that problem here.

Data rates for $n_b = 1.0$ are up to 16 percent lower over the range illustrated than the unconstrained channel for SER < 0.1, and up to 38 percent lower for SER < 0.01. The losses incurred by satisfying an SER constraint generally will be smaller for small n_b, as the optimum operating points at small n_b correspond to generally lower SERs. Figures 11 and 12 show the optimum ECC rate is generally higher for larger n_b, corresponding to lower SERs for the same BER.

Time-sharing PPM orders may be impractical to implement. Similar performance may be obtained by using integer PPM orders, or time-sharing with dead time. Adding dead time also provides a periodic non-pulsed synchronization marker that may be used to aid in symbol-timing recovery. Pulses may be located in the dead-time windows to denote frame or packet timing.

B. Trading SER for Throughput, Order Constrained

If the system is peak power constrained, so that using a higher order and the same average power does not lower the SER, we may choose a smaller PPM order and add a dead time, a period during which no pulse may be fired, between frames. This will shorten the window over which noise is observed, increasing the probability of resolving a pulse. This is less efficient than increasing the peak power and results in more substantial losses.

VI. Required Power for Specified Data Rate

In this section, we run through a sample code and modulation design and link budget calculation.

Suppose we have a system with a slot width of 1 ns and background noise $n_b = 1.0$. We would like to find the power required to achieve 56 Mb/s and choose appropriate coding and modulation. From Fig. 2, we find the optimum PPM order to achieve this data rate is $M = 64$, and the minimum required $n_s/M = 0.0541$. To achieve 56 Mb/s, we choose a rate $R = 0.6 \approx 0.056/(\log_2(M)/M)$ ECC and concatenate it with 64-PPM. The performance of two candidate ECCs for this operating point, SCPPM$(3/5, 2, 64, 16410)$ and RSPPM$(4095, 2457, 64)$, are illustrated in Fig. 19. Their performance may be compared with capacity for 56 Mb/s and the uncoded $M = 64$ performance, which, since it carries no coding redundancy, yields 94 Mb/s. The SCPPM code operates 0.5 dB from capacity, the RS code operates 2.5 dB from capacity, and uncoded PPM 7.2 dB from capacity (at 56 Mb/s). An appropriate comparison for uncoded 64-PPM is with capacity for 94 Mb/s, from which uncoded performance is 4.7 dB. (It would be more efficient to achieve 94 Mb/s with a rate 3/5 code mapped to 32-PPM).

These comparisons may be extended over a range of desired rates. Figure 20 illustrates achievable rates for $n_b = 1$ populated by points corresponding to the class of SCPPM codes, the class of RSPPM codes, and uncoded PPM. The coded and uncoded channels are evaluated at a finite number of rates, which we connect in a line for illustration—or by allowing time-sharing. The point RSPPM corresponding to $M = 64$ shows a cluster of points corresponding to the codes illustrated in Fig. 13. Other RSPPM

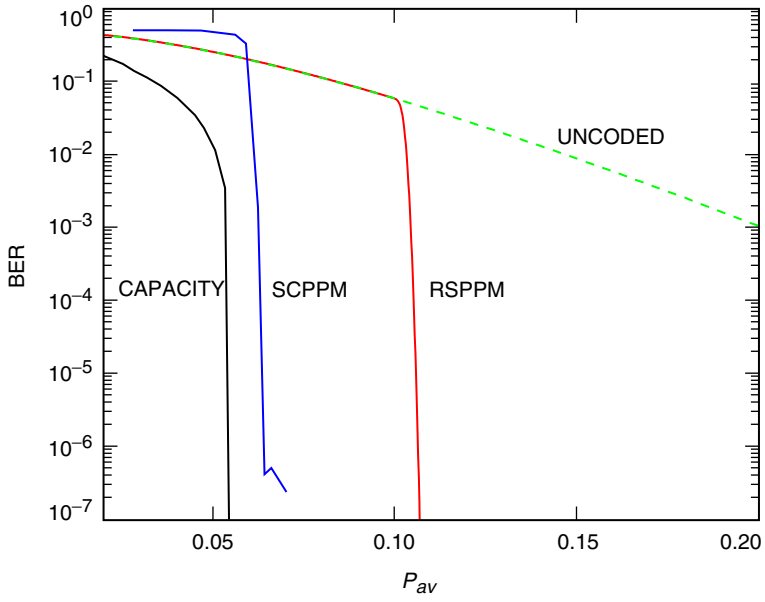

Fig. 19. Performance of SCPPM (3/5, 2, 64, 16410), RSPPM (4095, 2457, 64), $n_b = 1.0$.

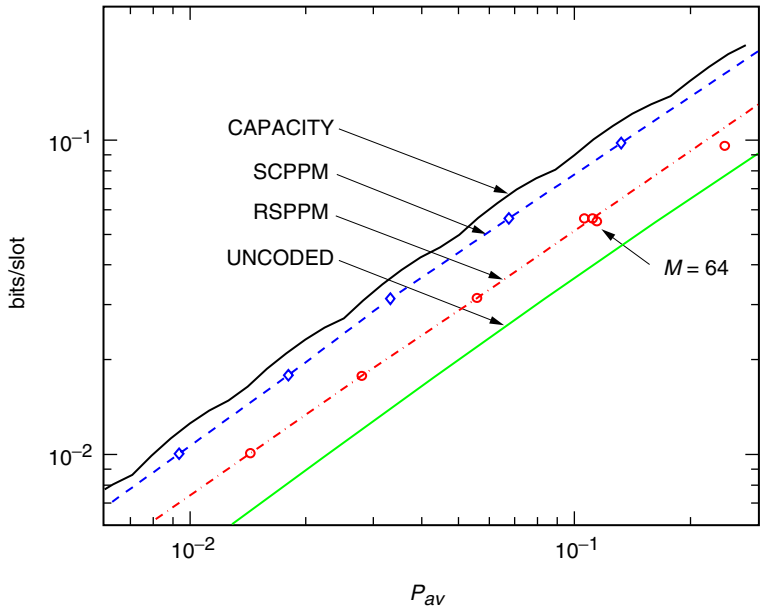

Fig. 20. Sample operating points, $n_b = 1$.

points use the convention $n = M - 1$ to illustrate the degradation in performance relative to capacity by using this convention for small M. Points correspond to the average power at which the BER is 10^{-5}. We exclude iterative codes that exhibit error floors at BERs greater than 10^{-6}.

The class of SCPPM codes lies approximately 0.5 dB from capacity, while the class of RSPPM codes lies approximately 2.75 dB from capacity, and uncoded performance is 4.7 dB from capacity. These gaps will vary with n_b but provide a good approximation over a range of expected background noise levels. For a conservative link-budget calculation that does not require the design and evaluation of a specific code, we assess a loss of 0.75 dB relative to capacity for iterative codes, 3.0 dB for RS codes, and 5.0 dB for uncoded. Although large SERs may make conventional data-aided timing recovery difficult, we anticipate synchronization losses may be kept sufficiently small with non-data-aided timing-recovery algorithms, even at average powers close to capacity. For the purpose of constructing link budgets, we assess a conservative 1-dB loss for all operating points due to quantization and synchronization losses. A robust system design should also incorporate a margin above the minimum requirement, which we will take to be 3.0 dB. From these approximations, we may obtain achievable data rates as a function of average power and background noise, or the required average power for a specified data rate. To a rough approximation, the achievable data-rate-versus-average-power curve (including receiver losses and margin) would be given by the capacity curve shifted by 4.75 dB for iterative codes, 7.0 dB for RS codes, and 9.0 dB for the uncoded case.

Acknowledgments

Many thanks to Bob McEliece, who contributed to discussions of characterizing the channel capacity function and extending Reed–Solomon codes. The content of the article was also influenced by work conducted in parallel on a deep-space optical communications link at Lincoln Laboratory, Lexington, Massachusetts.

References

[1] A. Biswas and S. Piazzolla, "Deep-Space Optical Communications Downlink Budget from Mars: System Parameters," *The Interplanetary Network Progress Report 42-154, April–June 2003,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–38, August 15, 2003.
http://ipnpr/progress_report/42-154/154L.pdf

[2] B. Moision and J. Hamkins, "Constrained Coding for the Deep-Space Optical Channel," *The Interplanetary Network Progress Report 42-149, January–March 2002,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–29, May 15, 2002.
http://ipnpr.jpl.nasa.gov/tmo/progress_report/42-149/149E.pdf

[3] S. Dolinar, D. Divsalar, J. Hamkins, and F. Pollara, "Capacity of Pulse-Position Modulation (PPM) on Gaussian and Webb Channels," *The Telecommunications and Mission Operations Progress Report 42-142, April–June 2000,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–31, August 15, 2000.
http://tmo.jpl.nasa.gov/tmo/progress_report/42-142/142H.pdf

[4] A. Biswas and W. H. Farr, "Laboratory Characterization and Modeling of a Near-Infrared Enhanced Photomultiplier Tube," *The Interplanetary Network Progress Report 42-152, October–December 2002,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–14, February 15, 2003.
http://ipnpr.jpl.nasa.gov/tmo/progress_report/42-152/152E.pdf

[5] R. G. Lipes, "Pulse-Position-Modulation Coding as Near-Optimum Utilization of Photon Counting Channel with Bandwidth and Power Constraints," *The Deep Space Network Progress Report 42-56, January and February 1980,* Jet Propulsion Laboratory, Pasadena, California, pp. 108–113, April 15, 1980.
http://tmo.jpl.nasa.gov/tmo/progress_report2/42-56/56N.PDF

[6] A. D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel—Part I," *IEEE Transactions on Information Theory*, vol. 34, pp. 1449–1461, November 1988.

[7] R. J. McEliece, "Practical Codes for Photon Communication," *IEEE Transactions on Information Theory*, vol. IT-27, pp. 393–398, July 1981.

[8] G. G. Ortiz, J. V. Sandusky, and A. Biswas, "Design of an Opto-Electronic Receiver for Deep-Space Optical Communications," *The Telecommunications and Mission Operations Progress Report 42-142, April–June 2000,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–17, August 15, 2000.
http://tmo.jpl.nasa.gov/tmo/progress_report/42-142/142I.pdf

[9] M. Srinivasan, J. Hamkins, B. Madden-Woods, A. Biswas, and J. Beebe, "Laboratory Characterization of Silicon Avalanche Photodiodes (APDs) for Pulse-Position Modulation (PPM) Detection," *The InterPlanetary Network Progress Report 42-146, April–June 2001,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–14, August 15, 2001.
http://ipnpr.jpl.nasa.gov/tmo/progress_report/42-146/146F.pdf

[10] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-State Modulation Codes for Data Storage," *IEEE Journal Selected Areas Communications*, vol. 10, pp. 5–37, January 1992.

[11] B. Moision and J. Hamkins, "Low Complexity Serially Concatenated Codes for the Deep Space Optical Channel," *Proceedings of IEEE International Symposium on Information Theory*, Yokohama, Japan, June 2003.

[12] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient Erasure Correcting Codes," *IEEE Transactions on Information Theory*, vol. 47, pp. 569–584, February 2001.

[13] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A Soft-Input Soft-Output Maximum A Posteriori (MAP) Module to Decode Parallel and Serial Concatenated Codes," *The Telecommunications and Data Acquisition Progress Report 42-127, July–September 1996,* Jet Propulsion Laboratory, Pasadena, California, pp. 1–20, November 15, 1996.
http://tmo.jpl.nasa.gov/tmo/progress_report/42-127/127H.pdf

[14] J. Berkmann and C. Weiss, "On Dualizing Trellis-Based APP Decoding Algorithms," *IEEE Transactions on Communications,* vol. 50, pp. 1743–1757, November 2002.

[15] G. Montorsi and S. Benedetto, "Design of Fixed-Point Iterative Decoders for Concatenated Codes with Interleavers," *IEEE Journal Selected Areas Communications*, vol. 19, pp. 871–882, May 2001.

[16] C.-C. Chen, "Figure of Merit for Direct-Detection Optical Channels," *The Telecommunications and Data Acquisition Progress Report 42-109, January–March 1992,* Jet Propulsion Laboratory, Pasadena, California, pp. 136–151, May 15, 1992.
http://tmo.jpl.nasa.gov/tmo/progress_report/42-109/109L.PDF

Appendix A

Capacity of the PPM Channel: General Form

The PPM modulator can be viewed as an encoder producing the $M = 2^k$ codewords $\{\mathbf{x}_1, \cdots, \mathbf{x}_M\}$ of a $(2^k, k)$ orthogonal code, where $\mathbf{x}_j$ is a vector of length M containing a one in the jth position. Let $\mathbf{X} = (X_1, \cdots, X_M)$ be a PPM symbol drawn according to density $f_{\mathbf{X}}$, and $\mathbf{Y} = (Y_1, \cdots, Y_M)$ the corresponding vector received over a discrete-time channel. We assume the optical channel is memoryless, so that

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}\,|\mathbf{x}) = \prod_{i=1}^{M} f_{Y_i|X_i}(y_i|x_i)$$

We assume $f_{Y_i|X_i}(y_i|0)$ and $f_{Y_i|X_i}(y_i|1)$ do not depend on i; hence, we will use the shorthand $p_0(k) = f_{Y_i|X_i}(k|0)$, $p_1(k) = f_{Y_i|X_i}(k|1)$, to denote the densities for signal and non-signal slots. It follows that

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}\,|\mathbf{x}) = p_1(y_j) \prod_{i=1, i\neq j}^{M} p_0(y_i) \tag{A-1}$$

where j is the index of the non-zero element of $\mathbf{x}$.

The capacity of a channel transmitting $\mathbf{X}$ and receiving $\mathbf{Y}$ is given by the maximum mutual information between $\mathbf{X}$ and $\mathbf{Y}$,

$$\begin{aligned} C &= \max_{f_{\mathbf{X}}} I(\mathbf{X};\mathbf{Y}) \quad \text{bits per PPM symbol} \\ &= \max_{f_{\mathbf{X}}} H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \\ &= \max_{f_{\mathbf{X}}} H(\mathbf{Y}) - \sum_{j=1}^{M} f_{\mathbf{X}}(\mathbf{x}_j) H(\mathbf{Y}|\mathbf{X} = \mathbf{x}_j) \end{aligned} \tag{A-2}$$

where $H(\cdot)$ is the entropy function. From Eq. (A-1), it follows that $H(\mathbf{Y}|\mathbf{X} = \mathbf{x}_j)$ is not a function of $\mathbf{x}_j$ and that $H(\mathbf{Y})$ is maximized with an equiprobable distribution on $\mathbf{X}$, i.e., $f_{\mathbf{X}}(\mathbf{x}_j) = 1/M$ for all j. Thus,

$$C = \int_{\mathbb{R}^M} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_1) \log_2 \left(\frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_1)}{\frac{1}{M}\sum_{k=1}^{M} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_k)} \right) d\mathbf{y} \text{ bits per PPM symbol} \tag{A-3}$$

The term in parentheses in Eq. (A-3) can be expressed as

$$\frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_1)}{\frac{1}{M}\sum_{k=1}^{M} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_k)} = \frac{M p_1(y_1) \prod_{i=1, i\neq 1}^{M} p_0(y_i)}{\sum_{j=1}^{M} p_1(y_j) \prod_{i=1, i\neq j}^{M} p_0(y_i)} = \frac{M L(y_1)}{\sum_{j=1}^{M} L(y_j)}$$

where $L(y) \triangleq p_1(y)/p_0(y)$ is the likelihood ratio of receiving statistic y. Thus, the capacity in Eq. (A-3) can be rewritten in terms of the likelihood ratios of the M slot statistics:

$$C = E \log_2 \left[\frac{ML(Y_1)}{\sum_{j=1}^{M} L(Y_j)} \right] \text{ bits per PPM symbol} \tag{A-4}$$

where the expectation is taken over $\mathbf{Y}$, with Y_1 having pdf $p_1(\cdot)$ and Y_j having pdf $p_0(\cdot)$, $j > 1$. Note that, as expected, Eq. (A-4) gives a capacity near $\log_2 M$ bits per PPM symbol when the channel is high quality and the likelihood ratio of Y_1 dominates the sum of the other likelihood ratios.

In principle, the capacity formula in Eq. (A-4) can be computed as long as the statistics $p_0(\cdot)$ and $p_1(\cdot)$ governing the binary-input channel are known. However, an uninspired computation of the expectation over $\mathbf{Y}$ requires evaluation of an M-dimensional integral. A low-complexity method to compute the expectation is to generate a random vector sample $\mathbf{Y} = (Y_1, \cdots, Y_M)$, where Y_1 has pdf $p_1(\cdot)$ and $Y_2, \cdots, Y_M$ each has pdf $p_0(\cdot)$; then evaluate the logarithmic function inside the expectation for the vector sample thus generated, and average the computed logarithm over many such samples. This Monte Carlo method was the most practical method to calculate numerical capacities.

Appendix B
Asymptotic Behavior of $C_{\mathrm{PPM}}(M,n_s,n_b)$

This appendix derives the asymptotic slope of the capacity of M-ary PPM on a Poisson channel. On a log-log plot of capacity versus average signal photons per pulse, the slope is shown to be 1 when no background is present and 2 when background is present.

Equations (7) and (8) correctly indicate that capacity approaches zero as $n_s \to 0$. We aim to determine the rate of its approach toward zero, as seen on a log-log plot. To simplify notation in determining the behavior of C_{PPM} as a function of n_s, write $C(n_s) = MC_{\mathrm{PPM}}(M, n_s, n_b)$ for fixed M and n_b.

We wish to determine $\lim_{n_s \to 0}(d \log C/du)$, where $u = \log n_s$. Stating explicitly the dependence of capacity on n_s, we may write

$$\frac{d \log C(n_s)}{du} = \frac{C'(n_s)}{C(n_s)} \cdot \frac{dn_s}{du} = \frac{C'(n_s)n_s}{C(n_s)}$$

This leads immediately to the following observation.

Observation B-1. Let $C(n_s) = a_0 + a_1 n_s + \cdots + a_n n_s^n + o(n_s^n)$ be the Taylor series expansion of $C(n_s)$ about $n_s = 0$, and let $u = \log n_s$. If k is the smallest non-negative integer k such that $a_k \neq 0$, then $\lim_{n_s \to 0}[d \log C(n_s)/du] = k$.

When $n_b = 0$, from Eq. (7) we have $C(n_s) = 0 + (\log_2 M)n_s + o(n_s)$, and the following result.

Theorem B-1. *For M-PPM signaling on a Poisson channel with non-signal slot-count average $n_b = 0$ and signal slot-count average n_s, the asymptotic log-log slope of capacity, $\lim_{n_s \to 0}[d \log C(n_s)/d(\log n_s)]$, equals 1.*

Next, we consider the case $n_b > 0$. To simplify the taking of derivatives, we define

$$f(n_s) \equiv \sum_{i=1}^{M} \left(1 + \frac{n_s}{n_b}\right)^{(Y_i - Y_1)}$$

$$g(n_s) \equiv \log_2 f(n_s)$$

$$h(n_s) \equiv p_1(y_1) = \frac{e^{-(n_s+n_b)}(n_s + n_b)^{y_1}}{y_1!}$$

It follows that

$$f'(n_s) = \sum_{i=1}^{M} (Y_i - Y_1) \left(1 + \frac{n_s}{n_b}\right)^{(Y_i - Y_1 - 1)} \frac{1}{n_b}$$

$$f''(n_s) = \sum_{i=1}^{M} (Y_i - Y_1)(Y_i - Y_1 - 1) \left(1 + \frac{n_s}{n_b}\right)^{(Y_i - Y_1 - 2)} \frac{1}{n_b^2}$$

Thus, $f(0) = M$, $f'(0) = (1/n_b)\sum_{i=2}^{M}(Y_i - Y_1)$, $E_{Y_1,\cdots,Y_M} f'(0) = 0$, $f''(0) = (1/n_b^2)\sum_{i=2}^{M}(Y_i - Y_1)(Y_i - Y_1 - 1)$, and

$$E_{Y_1,\cdots,Y_M} f''(0) = E_{Y_1,\cdots,Y_M} \frac{1}{n_b^2} \sum_{i=2}^{M} [Y_i^2 - 2Y_i Y_1 + Y_1^2 - (Y_i - Y_1)]$$

$$= \frac{M-1}{n_b^2}\big((n_b + n_b^2) - 2n_b n_b + (n_b + n_b^2) - 0\big)$$

$$= \frac{2(M-1)}{n_b}$$

Also, $g'(n_s) = f'(n_s)/(f(n_s)\ln 2)$, and $g''(n_s) = (f(n_s)f''(n_s) - f'(n_s)^2)/(f(n_s)^2 \ln 2)$. Finally, we note that

$$h'(n_s) = \left(\frac{y_1}{n_s + n_b} - 1\right) h(n_s)$$

$$h''(n_s) = \left(\frac{-y_1}{(n_s + n_b)^2} + \left(\frac{y_1}{n_s + n_b} - 1\right)^2\right) h(n_s)$$

Using these definitions, we may write

$$C(n_s) = \log_2 M - E_{Y_1,\cdots,Y_M} g(n_s)$$

$$= \log_2 M - E_{Y_2,\cdots,Y_M}\big\{E_{Y_1} g(n_s)\big\}$$

$$= \log_2 M - E_{Y_2,\cdots,Y_M} \sum_{y_1=0}^{\infty} g(n_s)\Big|_{Y_1=y_1} h(n_s)$$

where in the second step we used the fact that Y_1 is independent of $Y_2, \ldots, Y_M$, and in the last step $g(n_s)$ is evaluated at $Y_1 = y_1$ (notation is dropped in the following). It follows that

$$C'(n_s) = -E_{Y_2,\cdots,Y_M} \sum_{y_1=0}^{\infty} g'(n_s)h(n_s) + g(n_s)h'(n_s)$$

$$= -E_{Y_2,\cdots,Y_M} \sum_{y_1=0}^{\infty} \left(g'(n_s) + g(n_s)\left(\frac{y_1}{n_s + n_b} - 1\right)\right) h(n_s)$$

$$= -E_{Y_1,\cdots,Y_M} \left[g'(n_s) + g(n_s)\left(\frac{Y_1}{n_s + n_b} - 1\right)\right]$$

$$= -E_{Y_1,\cdots,Y_M} \left[\frac{f'(n_s)}{f(n_s)\ln 2} + \log_2 f(n_s)\left(\frac{Y_1}{n_s + n_b} - 1\right)\right]$$

From the above, it follows that $\lim_{n_s \to 0} C'(n_s) = 0$. Proceeding to the second derivative,

$$\begin{aligned} C''(n_s) = & -E_{Y_2,\ldots,Y_M} \sum_{y_1=0}^{\infty} \left(g''(n_s)h(n_s) + 2g'(n_s)h'(n_s) + g(n_s)h''(n_s) \right) \\ = & -E_{Y_1,\cdots,Y_M} \left[\frac{f(n_s)f''(n_s) - f'(n_s)^2}{f(n_s)^2 \ln 2} + 2\frac{f'(n_s)}{f(n_s)\ln 2} \left(\frac{Y_1}{n_s + n_b} - 1 \right) \right. \\ & \left. + \log_2 f(n_s) \left(\frac{-Y_1}{(n_s+n_b)^2} + \left(\frac{Y_1}{n_s+n_b} - 1 \right)^2 \right) \right] \end{aligned}$$

Note that

$$\lim_{n_s \to 0} E\left[\frac{-Y_1}{(n_s+n_b)^2} + \left(\frac{Y_1}{n_s+n_b} - 1 \right)^2 \right] = \frac{-n_b}{n_b^2} + E\left[\frac{Y_1^2}{n_b^2} - \frac{2Y_1}{n_b} + 1 \right] = -\frac{1}{n_b} + \frac{n_b + n_b^2}{n_b^2} - 2 + 1 = 0$$

Also,

$$\begin{aligned} E\left[f'(0)^2\right] &= \frac{1}{n_b^2} E\left[\left(\sum_{i=2}^{M} (Y_i - Y_1) \right)^2 \right] = \frac{1}{n_b^2} E\left[\sum_{i=2}^{M} \sum_{j=2}^{M} (Y_i - Y_1)(Y_j - Y_1) \right] \\ &= \frac{1}{n_b^2} E\left[\sum_{i=2}^{M} \sum_{j=2}^{M} \left(Y_i Y_j - Y_i Y_1 - Y_j Y_1 + Y_1^2 \right) \right] = \frac{1}{n_b^2} \left[(M-1)2n_b + (M-1)(M-2)n_b \right] \\ &= \frac{M(M-1)}{n_b} \end{aligned}$$

where in the second to last step we partitioned the double summation into the part where $i = j$ and where $i \neq j$. From the above, it follows that

$$\begin{aligned} \lim_{n_s \to 0} C''(n_s) &= -\frac{M\frac{2(M-1)}{n_b} - \frac{M(M-1)}{n_b}}{M^2 \ln 2} - \lim_{n_s \to 0} \frac{2}{Mn_b \ln 2} E_{Y_1,\cdots,Y_M} \left[\sum_{i=2}^{M} (Y_i - Y_1) \left(\frac{Y_1}{n_b} - 1 \right) \right] \\ &= -\frac{M-1}{Mn_b \ln 2} - \frac{2}{Mn_b \ln 2} \cdot (M-1) \left(n_b - n_b - \frac{n_b + n_b^2}{n_b} + n_b \right) \\ &= \frac{M-1}{Mn_b \ln 2} \qquad \text{(B-1)} \\ &\neq 0 \end{aligned}$$

This establishes the following.

Theorem B-2. *For M-PPM signaling on a Poisson channel with non-signal slot-count average $n_b > 0$ and signal slot-count average $n_b + n_s$, the asymptotic log-log slope of capacity, $\lim_{n_s \to 0}[d \log C(n_s)/ d(\log n_s)]$, equals 2.*