A refined analysis of the Poisson channel in the high-photon-efficiency regime

Ligong Wang and Gregory W. Wornell

Abstract

We study the discrete-time Poisson channel under the constraint that its average input power (in photons per channel use) must not exceed some constant $E$. We consider the wideband, high-photon-efficiency extreme where $E$ approaches zero, and where the channel’s “dark current” approaches zero proportionally with $E$. Improving over a previously obtained first-order capacity approximation, we derive a refined approximation, which includes the exact characterization of the second-order term, as well as an asymptotic characterization of the third-order term with respect to the dark current. We also show that pulse-position modulation is nearly optimal in this regime.

I. INTRODUCTION

We consider the discrete-time memoryless Poisson channel whose input $x$ is in the set $\mathbb{R}_0^+$ of nonnegative reals and whose output $y$ is in the set $\mathbb{Z}_0^+$ of nonnegative integers. Conditional on the input $X = x$, the output $Y$ has a Poisson distribution of mean $(\lambda + x)$, where $\lambda \geq 0$ is called the “dark current” and is a constant that does not depend on the input $X$. We denote the Poisson distribution of mean $\xi$ by $\text{Pois}_\xi(\cdot)$ so

$$
\text{Pois}_\xi(y) = e^{-\xi} \frac{\xi^y}{y!}, \quad y \in \mathbb{Z}_0^+.
$$

(1)

With this notation the channel law $W(\cdot|\cdot)$ is

$$
W(y|x) = \text{Pois}_{\lambda+x}(y), \quad x \in \mathbb{R}_0^+, y \in \mathbb{Z}_0^+.
$$

(2)

This work was supported in part by the DARPA InPho program under Contract No. HR0011-10-C-0159, and by AFOSR under Grant No. FA9550-11-1-0183. This paper was presented in part at the 2012 IEEE Information Theory Workshop.

The authors are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA.
This channel models pulse-amplitude modulated optical communication where the transmitter sends light signals in \textit{coherent states} (which are usually produced using laser devices), and where the receiver employs \textit{direct detection} (i.e., photon counting) \cite{1}. The channel input $x$ describes the expected number of \textit{signal-photons} (i.e., photons that come from the input light signal rather than noise) to be detected in the pulse-duration, and is proportional to the light signal’s intensity, the pulse-duration, the channel’s transmissivity, and the detector’s efficiency; the channel output $y$ is the actual number of photons that are detected in the pulse-duration; and \(\lambda\) is the average number of extraneous counts that appear in \(y\) due to background radiation or to the detector’s “dark clicks”.

We impose an \textit{average-power constraint}\footnote{Here “power” is in discrete time, means expected number of detected signal-photons per channel use, and is proportional to the continuous-time physical power times the pulse duration.} on the input

$$\mathbb{E}[X] \leq \mathcal{E}$$

for some $\mathcal{E} > 0$.

In applications like free-space optical communications, the cost of producing and successfully transmitting photons is high, hence high \textit{photon-information efficiency}—amount of information transmitted per photon, which we henceforth call simply “photon efficiency”—is desirable. As we later demonstrate, this can be achieved in the \textit{wideband} regime, where the pulse-duration of the input approaches zero and, assuming that the \textit{continuous-time} average input power is fixed, where \(\mathcal{E}\) approaches zero proportionally with the pulse-duration. Note that in this regime the average number of detected background photons or dark clicks also tends to zero proportionally with the pulse-duration. Hence we have the linear relation

$$\lambda = c\mathcal{E},$$

where $c$ is some nonnegative constant that does not change with $\mathcal{E}$. For practice, asymptotic results in this regime are useful in scenarios where $\mathcal{E}$ is small and where $\lambda$ is comparable to or much smaller than $\mathcal{E}$. Scenarios where $\mathcal{E}$ is small but $\lambda$ is much larger is better captured by the model where $\lambda$ stays constant while $\mathcal{E}$ tends to zero; see \cite{2}.

January 23, 2014 DRAFT
We denote the capacity (in nats per channel use) of the channel (2) under power constraint (3) with dark current (4) by $C(\mathcal{E}, c)$, then

$$C(\mathcal{E}, c) = \sup_{E[X] \leq \mathcal{E}} I(X; Y), \quad (5)$$

where the mutual information is computed from the channel law (2) and is maximized over input distributions satisfying (3), with dark current $\lambda$ given by (4). As we shall see, our results on the asymptotic behavior of $C(\mathcal{E}, c)$ hold irrespectively of whether a peak-power constraint

$$X \leq A \quad \text{with probability 1} \quad (6)$$

is imposed or not, as long as $A$ is positive and does not approach zero together with $\mathcal{E}$.

Various capacity results for the discrete-time Poisson channel have been obtained [2]–[6]. Among them, [2] considers the same scenario as the present paper and shows that [2, Proposition 1]

$$\lim_{\mathcal{E} \downarrow 0} \frac{C(\mathcal{E}, c)}{\mathcal{E} \log \frac{1}{\mathcal{E}}} = 1, \quad c \in [0, \infty). \quad (7)$$

In other words, the maximum achievable photon efficiency in nats per photon, which we henceforth denote by $C_{\text{PE}}(\mathcal{E}, c)$, satisfies

$$C_{\text{PE}}(\mathcal{E}, c) \triangleq \frac{C(\mathcal{E}, c)}{\mathcal{E}} \quad (8)$$

$$= \log \frac{1}{\mathcal{E}} + o \left( \log \frac{1}{\mathcal{E}} \right), \quad c \in [0, \infty). \quad (9)$$

It is also shown in [2] that the limit (7) is achievable using on-off keying. In the Appendix we sketch a proof for (7) that is simpler than the original proof in [2], especially in the converse part.

The approximation in (9) can be compared to the maximum photon efficiency achievable on the pure-loss bosonic channel, which describes an optical communication channel where the transmitter can send any quantum state, where the receiver can employ any quantum detector, and where no background radiation is present. We denote by $C_{\text{PE-bosonic}}(\mathcal{E})$ the maximum photon efficiency of the pure-loss bosonic channel under an average-power constraint that is equivalent to (3). The value of $C_{\text{PE-bosonic}}(\mathcal{E})$ can be easily computed using the explicit capacity formula derived in [7], which yields

$$C_{\text{PE-bosonic}}(\mathcal{E}) = \log \frac{1}{\mathcal{E}} + 1 + o(1). \quad (10)$$
Comparing (9) and (10) shows the following:

- For the pure-loss bosonic channel in the wideband regime, coherent-state inputs and direct detection are optimal up to the first-order term in photon efficiency (or, equivalently, in capacity). For example, they achieve infinite capacity per unit cost [8], [9].
- The dark current does not affect this first-order term.

In the present paper, we refine the analysis in [2] in two aspects. First, we provide a more accurate approximation for $C_{PE}(E, c)$ that contains higher-order terms, and that reflects the influence from the dark current, i.e., from the constant $c$. Second, we identify near-optimal modulation schemes that facilitate code design for this channel.

Some progress has been made in improving the approximation for $C_{PE}(E, 0)$. It is argued in [10] that the maximum photon efficiency achievable on the Poisson channel (2) with $\lambda = 0$ using on-off keying is of the form

$$\log \frac{1}{E} - \log \log \frac{1}{E} + O(1),$$

which means that restriction to on-off coherent-state inputs and direct detection does induce a loss in the second-order term in the photon efficiency on the pure-loss bosonic channel. Comparing the first-order term in (9) and the first- and second-order terms in (11) for $E = 10^{-5}$ yields a difference of about 20%, showing the importance of the refined approximation (11). However, the analysis in [10] is restricted to on-off keying. It is therefore unclear if (11) is the maximum photon efficiency achievable on the Poisson channel (2) with $\lambda = 0$ subject to constraint (3) alone, i.e., if (11) is the correct expression for $C_{PE}(E, 0)$. In the present paper, we prove that this is indeed the case.

Recent works such as [10], [11] often ignore the dark current in the Poisson channel. It has been unknown to us whether the dark current, i.e., whether the constant $c$ influences the second term in (11) and, if not, whether it influences the next term in photon efficiency. We show that the constant $c$ does not influence the second-order term in $C_{PE}(E, c)$, but does influence the third-order, constant term.

It has long been known that that infinite photon efficiency on the Poisson channel with zero dark current can be achieved using pulse-position modulation (PPM) combined with an outer code [3], [4]. Further, [11] observes that PPM can achieve (11) on such a channel. PPM greatly simplifies the coding task for this channel, since one can easily apply existing codes, such as a
Reed-Solomon code, to the PPM “super symbols”; while the on-off keying scheme that achieves (11) has a highly skewed input distribution and is hence difficult to code. The question then arises: how useful is PPM when there is a positive dark current? This question has two parts. First, is PPM still near optimal in terms of capacity (or photon efficiency) when there is a positive dark current? Second, does PPM still simplify coding when there is a dark current? We answer the first part of the question in the affirmative in our main results. We cannot fully answer the second question in this paper, but we shall discuss it in the concluding remarks in Section VII.

The rest of this paper is arranged as follows. We introduce some notation and formally define PPM in Section II. We state and discuss our main results in Section III. We then prove the achievability parts of these results in Sections IV and V, and prove the converse parts in Section VI. We conclude the paper with numerical comparison of the bounds and some remarks in Section VII.

II. NOTATION AND PPM

We usually use a lower-case letter like $x$ to denote a constant, and an upper-case letter like $X$ to denote a random variable.

We use natural logarithms, and measure information in nats.

We use the usual $o(\cdot)$ and $O(\cdot)$ notation to describe behaviors of functions of $\mathcal{E}$ in the limit where $\mathcal{E}$ approaches zero with other parameters, if any, held fixed. Specifically, given a reference function $f(\cdot)$ (which might be the constant $1$), a function described as $o(f(\mathcal{E}))$ satisfies

$$\lim_{\mathcal{E} \downarrow 0} \frac{o(f(\mathcal{E}))}{f(\mathcal{E})} = 0,$$

and a function described as $O(f(\mathcal{E}))$ satisfies

$$\lim_{\mathcal{E} \downarrow 0} \left| \frac{O(f(\mathcal{E}))}{f(\mathcal{E})} \right| < \infty.$$

We emphasize that, in particular, we do not use $o(\cdot)$ and $O(\cdot)$ to describe how functions behave with respect to $c$.

We adopt the convention

$$0 \log 0 = 0.$$ 

We next formally describe what we mean by PPM. On the transmitter side:

- The channel-uses are divided into frames of equal lengths;
In each frame, there is only one channel input that is positive, which we call the “pulse”, while all the other inputs are zeros;

- The pulses in all frames have the same amplitude.

On the receiver side, we distinguish between two cases, which we call simple PPM and soft-decision PPM, respectively. In simple PPM, the receiver records at most one pulse in each frame; if more than one pulse is detected in a frame, then that frame is recorded as an “erasure”, as if no pulse is detected at all. In soft-decision PPM, the receiver records up to two pulses in each frame; frames containing no or more than two detected pulses are treated as erasures.

III. MAIN RESULTS AND DISCUSSIONS

The following theorem provides approximation for the photon efficiency $C_{PE}(\mathcal{E}, c)$ in the high-photon-efficiency regime:

**Theorem 1.** The maximum photon efficiency $C_{PE}(\mathcal{E}, c)$ achievable on the Poisson channel (2) subject to constraint (3) and with dark current (4), which we define in (8), satisfies

$$C_{PE}(\mathcal{E}, c) = \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1), \quad c \in [0, \infty). \quad (15)$$

Furthermore, denote

$$K(\mathcal{E}, c) \triangleq C_{PE}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}}, \quad (16)$$

then

$$\lim_{c \to \infty} \lim_{\mathcal{E} \downarrow 0} \frac{K(\mathcal{E}, c)}{\log c} = \lim_{c \to \infty} \lim_{\mathcal{E} \downarrow 0} \frac{K(\mathcal{E}, c)}{\log c} = -1. \quad (17)$$

Our second theorem shows that PPM is nearly optimal in the regime of interest:

**Theorem 2.** The asymptotic expression on the right-hand side of (15) is achievable with simple PPM. The limits in (17) are achievable with soft-decision PPM.

The achievability part of Theorem 1 follows from Theorem 2. To prove the latter, we need to show two things: first, that the largest photon efficiency that is achievable with simple PPM, which we henceforth denote by $C_{PE-PPM}(\mathcal{E}, c)$, satisfies

$$C_{PE-PPM}(\mathcal{E}, c) \geq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1), \quad c \in [0, \infty); \quad (18)$$
and second, that the maximum photon efficiency that is achievable with soft-decision PPM, which we henceforth denote by $C_{\text{PE-PPM(SD)}}(\mathcal{E}, c)$, satisfies

$$
\lim_{c \to \infty} \lim_{\mathcal{E} \downarrow 0} \frac{C_{\text{PE-PPM(SD)}}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}}}{\log c} \geq -1.
$$

(19)

To prove the converse part of Theorem 1, we note that the capacity of the channel is monotonically decreasing in $c$ [2], and hence so is the photon efficiency. It thus suffices to show two things: first, that, in the absence of dark current, the largest photon efficiency achievable with any scheme satisfies

$$
C_{\text{PE}}(\mathcal{E}, 0) \leq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1);
$$

(20)

and second, that

$$
\lim_{c \to \infty} \lim_{\mathcal{E} \downarrow 0} \frac{C_{\text{PE}}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}}}{\log c} \leq -1.
$$

(21)

We prove (18) in Section IV, (19) in Section V, and (20) and (21) in Section VI.

Some remarks concerning channel capacity follow Theorem 1.

- Choosing $c = 0$ in (15) confirms that (11) is the correct asymptotic expression up to the second-order term for $C_{\text{PE}}(\mathcal{E}, 0)$. Compared to (10) this means that, on the pure-loss bosonic channel, for small $\mathcal{E}$, restricting the receiver to using direct detection induces a loss in photon efficiency of about $\log \log \frac{1}{\mathcal{E}}$ nats per photon. Note that the capacity of the pure-loss bosonic channel can be achieved using coherent input states only [7], so this loss is indeed due to direct detection, and not due to coherent input states. Attempts to overcome this loss by employing other feasible detection techniques have so far been unsuccessful [10], [12].

- The expression on the right-hand side of (15) does not depend on $c$, so the value of $c$ affects neither the first-order term nor the second-order term in $C_{\text{PE}}(\mathcal{E}, c)$. In particular, these two terms do not depend on whether $c$ is zero or positive.

- The first term in $C_{\text{PE}}(\mathcal{E}, c)$ that $c$ does affect is the third, constant term. Indeed, though we have not given an exact expression for the constant term, the asymptotic property (17) shows that $c$ affects the constant term in such a way that, for large $c$, the constant term is approximately $- \log c$. 

January 23, 2014 DRAFT
Theorem 1 suggests the approximation

$$C_{PE}(\mathcal{E}, c) \approx \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} - \log c.$$  (22)

The correction terms, roughly speaking, are either vanishing for small $\mathcal{E}$, or small compared to $\log c$ for large $c$. Note that if we fix the dark current, i.e., if we fix the product $c\mathcal{E}$, then the first and third terms on the right-hand side of (22) cancel. This is intuitively in agreement with [2, Proposition 2], which states that, for fixed dark current, photon efficiency scales like some constant times $\log \log(1/\mathcal{E})$ and hence not like $\log(1/\mathcal{E})$. We note, however, that (17) cannot be derived directly using (15) together with [2, Proposition 2], because we cannot change the order of the limits. In (17) we do not let $\mathcal{E}$ tend to zero and $c$ tend to infinity simultaneously. Instead, we first let $\mathcal{E}$ tend to zero to close down onto the constant term in $C_{PE}(\mathcal{E}, c)$ with respect to $\mathcal{E}$, and then let $c$ tend to infinity to study the asymptotic behavior of this constant term with respect to $c$.

The approximation (22) is good for large $c$, but diverges as $c$ tends to zero. As the numerical simulations we show later will suggest, $-\log(1+c)$ is a better approximation for the constant term for moderate values of $c$:

$$C_{PE}(\mathcal{E}, c) \approx \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} - \log(1 + c).$$  (23)

For modulation and coding considerations, we make some remarks following Theorem 2.

Theorem 2 shows that PPM is optimal up to the second-order term in photon efficiency when $c = 0$. Hence, compared to the restriction to the receiver using direct detection, the further restriction to PPM induces only a small additional loss in photon efficiency.

Furthermore, for the third-order, constant term, (soft-decision) PPM is also not far from optimal, in the sense that it achieves the optimal asymptotic behavior of this term for large $c$.

In Section IV we show that

$$C_{PE-PPM}(\mathcal{E}, c) \geq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} - c - \log(1 + c) - \frac{3}{2} + o(1).$$  (24)

A careful analysis will confirm that the bound (24) is tight in the regime of interest, in the sense that simple PPM cannot achieve a constant term that is better than linear in $c$ (while being second-order optimal). This is in contrast to soft-decision PPM, which can achieve a
constant term that is logarithmic in \( c \). In particular, simple PPM is clearly not third-order optimal.

- In the PPM schemes that achieve (18) and (19), which we describe in Sections IV and V, the pulse has amplitude \( 1/(\log(1/E)) \), which depends on \( E \), and which tends to zero as \( E \) tends to zero. Clearly, an on-off keying scheme having the same amplitude for its “on” signals can also achieve these lower bounds. These are different from the on-off keying scheme used in [2] where the “on” signal has a fixed amplitude that does not depend on \( E \). The latter on-off keying scheme, as well as any PPM scheme with a fixed pulse amplitude (with respect to \( E \)), is not second-order optimal on the Poisson channel.

- Because in the PPM schemes to achieve (18) and (19) the pulse tends to zero as \( E \) tends to zero, we know that, as claimed in Section I, both our theorems hold if a constant (i.e., not approaching zero together with \( E \)) peak-power constraint as in (6) is imposed on \( X \) in addition to (3).

We next proceed to prove our main results, and leave further discussions to Section VII.

IV. PROOF OF THE SIMPLE-PPM LOWER BOUND (18)

Assume that \( E \) is small enough so that

\[
E \log \frac{1}{E} < 1 \quad \quad (25a)
\]

\[
E < e^{-c}. \quad \quad (25b)
\]

Consider the following simple PPM scheme:

**Scheme 1.** • The channel-uses are divided into frames of length \( b \), so each frame contains \( b \) input symbols \( x_1, \ldots, x_b \) and \( b \) corresponding output symbols \( y_1, \ldots, y_b \). We set

\[
b = \left\lfloor \frac{1}{E \log \frac{1}{E}} \right\rfloor. \quad \quad (26)
\]

• Within each length-\( b \) frame, there is always one input that equals \( \eta \), and all the other \( (b - 1) \) inputs are zeros. Each frame is then fully specified by the position of its unique nonzero symbol, i.e., its pulse position. We consider each frame as a “super input symbol” \( \bar{x} \) that takes value in \( \{1, \ldots, b\} \). Here \( \bar{x} = i \) means

\[
x_i = \eta \quad \quad (27a)
\]

\[
x_j = 0, \quad j \neq i. \quad \quad (27b)
\]
To meet the average-power constraint (3) with equality, we require

$$\eta = b\mathcal{E}. \tag{28}$$

• The $b$ output symbols $y_1, \ldots, y_b$ are mapped to one “super output symbol” $\bar{y}$ that takes value in $\{1, \ldots, b, ?\}$ in the following way: $\bar{y} = i$, $i \in \{1, \ldots, b\}$, if $y_i$ is the unique nonzero term in $\{y_1, \ldots, y_b\}$; and $\bar{y} = ?$ if there is more than one or no nonzero term in $\{y_1, \ldots, y_b\}$.

We have the following lower bound on the photon efficiency achieved by the above scheme.

**Proposition 1.** Under the conditions (25), Scheme 1 achieves photon efficiency

$$C_{\text{PE-PPM}}(\mathcal{E}, c) \geq \left(1 - \frac{\eta}{2}\right) \log b - c\eta \log b - \left(1 + \frac{c\mathcal{E}}{\eta}\right) \log(1 + c) - \left(1 + \frac{c\mathcal{E}}{\eta}\right) \left\{\log(1 - c\eta) + \log \left(1 - \frac{\eta}{2}\right)\right\} - 1 - \frac{c\mathcal{E}}{\eta} \tag{29}$$

where $b$ and $\eta$ are given in (26) and (28), respectively.

**Proof of (18) using Proposition 1:** We plug (26) and (28) into (29), note that

$$\left|\frac{1}{\mathcal{E} \log \frac{1}{\xi}}\right| = \frac{1}{\mathcal{E} \log \frac{1}{\xi}} + O(1) \tag{30}$$

and simplify (29) to obtain

$$C_{\text{PE-PPM}}(\mathcal{E}, c) \geq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} - c - \log(1 + c) - \frac{3}{2} + o(1) \tag{31}$$

$$= \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1). \tag{32}$$

**Proof of Proposition 1:** We compute the transition matrix of the PPM “super channel” as follows:

$$\tilde{W}(i|i) = \Pr \left[ \tilde{Y} = i \mid \tilde{X} = i \right] \tag{33}$$

$$= \Pr [Y_i \geq 1 \mid X_i = \eta] \prod_{k \neq i} \Pr [Y_k = 0 \mid X_k = 0] \tag{34}$$

$$= \left(1 - W(0|\eta)\right) \left(W(0|0)\right)^{b-1} \tag{35}$$

$$= (1 - e^{-\eta - c\mathcal{E}}) e^{-(b-1)c\mathcal{E}} \tag{36}$$

$$= e^{-(b-1)c\mathcal{E}} - e^{-\eta - bc\mathcal{E}} \tag{37}$$

January 23, 2014 DRAFT
\[ \tilde{p}_0, \quad i \in \{1, \ldots, b\}; \]  

\[ \tilde{W}(j|i) = \Pr \left[ \tilde{Y} = j \mid \tilde{X} = i \right] = \Pr \left[ Y_i = 0 \mid X_i = \eta \right] \Pr \left[ Y_j \geq 1 \mid X_j = 0 \right] \cdot \prod_{k \notin \{i,j\}} \Pr \left[ Y_k = 0 \mid X_k = 0 \right] \]  

\[ = W(0|\eta)(1 - W(0|0))(W(0|0))^{b-2} = e^{-\eta - cE}(1 - e^{-cE})e^{-(b-2)cE} = e^{-\eta - (b-1)cE} - e^{-\eta - bcE} \]  

\[ \tilde{W}(?|i) = 1 - \tilde{p}_0 - (b - 1)p_1, \quad i \in \{1, \ldots, b\}. \]  

Denote the capacity of this super channel by \( \bar{C}(\mathcal{E}, c, b, \eta) \), then

\[ \bar{C}(\mathcal{E}, c, b, \eta) = \max_{P_{\tilde{X}}} I(\tilde{X}; \tilde{Y}). \]  

Note that the total input power (i.e., expected number of detected signal-photons) in each frame equals \( \eta \). Therefore we have the following lower bound on \( C_{PE-PPM}(\mathcal{E}, c) \):

\[ C_{PE-PPM}(\mathcal{E}, c) \geq \frac{\bar{C}(\mathcal{E}, c, b, \eta)}{\eta}. \]  

It can be easily verified that the optimal input distribution for (46) is the uniform distribution

\[ P_{\tilde{X}}(i) = \frac{1}{b}, \quad i \in \{1, \ldots, b\}, \]  

which induces the following marginal distribution on \( \tilde{Y} \):

\[ P_{\tilde{Y}}(i) = \frac{p_0 + (b - 1)p_1}{b}, \quad i \in \{1, \ldots, b\} \]  

\[ P_{\tilde{Y}}(?) = 1 - p_0 - (b - 1)p_1. \]  

We can now use the above joint distribution on \( (\tilde{X}, \tilde{Y}) \) to explicitly compute \( \bar{C}(\mathcal{E}, c, b, \eta) \) as
follows:

\[ \tilde{C}(\mathcal{E}, c, b, \eta) = I(\tilde{X}; \tilde{Y}) \]
\[ = H(\tilde{Y}) - H(\tilde{Y}|\tilde{X}) \]
\[ = (1 - p_0 - (b - 1)p_1) \log \frac{1}{1 - p_0 - (b - 1)p_1} \]
\[ + (p_0 + (b - 1)p_1) \log \frac{b}{p_0 + (b - 1)p_1} \]
\[ - (1 - p_0 - (b - 1)p_1) \log \frac{1}{1 - p_0 - (b - 1)p_1} \]
\[ - p_0 \log \frac{1}{p_0} - (b - 1)p_1 \log \frac{1}{p_1} \]
\[ = p_0 \log \frac{bp_0}{p_0 + (b - 1)p_1} + (b - 1)p_1 \log \frac{bp_1}{p_0 + (b - 1)p_1}. \]  

(50)

We continue (53) to lower-bound \( \tilde{C}(\mathcal{E}, c, b, \eta) \) as

\[ \tilde{C}(\mathcal{E}, c, b, \eta) = p_0 \log b - p_0 \log \frac{p_0 + (b - 1)p_1}{p_0} \]
\[ \leq \log \left(1 + \frac{bp_1}{p_0}\right) \]
\[ - (b - 1)p_1 \log \frac{p_0 + (b - 1)p_1}{bp_1} \]
\[ \leq \frac{p_0 - p_1}{bp_1} \]
\[ \geq p_0 \log b - p_0 \log \left(1 + \frac{bp_1}{p_0}\right) - \frac{b - 1}{b} \]
\[ \leq \frac{p_0 - p_1}{bp_1} \leq p_0. \]  

(54)

Next note that \( p_0 \) can be upper-bounded as

\[ p_0 = e^{-(b-1)c\mathcal{E}} - e^{-\eta - bc\mathcal{E}} \]
\[ = e^{-(b-1)c\mathcal{E}} \left(1 - e^{-\eta - c\mathcal{E}}\right) \]
\[ \leq \eta + c\mathcal{E}, \]  

(57)
and can also be lower-bounded as

\[ p_0 = e^{-(b-1)c\mathcal{E}} - e^{-\eta-bc\mathcal{E}} \]

\[ \geq e^{-bc\mathcal{E}} - e^{-\eta-bc\mathcal{E}} \]  

\[ \geq e^{-bc\mathcal{E}} \left( 1 - e^{-\eta} \right) \]  

\[ \geq 1 - be^{\mathcal{E}} \geq 1 - \frac{1}{2}\eta^2 \]

\[ \geq \left( 1 - bc\mathcal{E} \right) \left( \eta - \frac{1}{2}\eta^2 \right), \]  

where the last inequality holds if both multiplicands on its right-hand side are positive, which, for our choices of \( b \) and \( \eta \) in (26) and (28), is true for all \( \mathcal{E} \) satisfying (25). Also note that \( p_1 \) can be upper-bounded as

\[ p_1 = e^{-\eta-(b-1)c\mathcal{E}} - e^{-\eta-bc\mathcal{E}} \]

\[ = e^{-\eta-(b-1)c\mathcal{E}} \left( 1 - e^{-c\mathcal{E}} \right) \]  

\[ \leq c\mathcal{E}. \]  

Using (59), (63) and (66) we can continue the chain of inequalities (56) to further lower-bound \( \tilde{C}(\mathcal{E}, c, b, \eta) \) as

\[ \tilde{C}(\mathcal{E}, c, b, \eta) \]

\[ \geq \left( 1 - bc\mathcal{E} \right) \left( \eta - \frac{1}{2}\eta^2 \right) \log b \]

\[ - \left( \eta + c\mathcal{E} \right) \log \left( 1 + \frac{bc\mathcal{E}}{(1 - bc\mathcal{E}) \left( \eta - \frac{1}{2}\eta^2 \right)} \right) - \eta - c\mathcal{E} \]

\[ \geq \left( 1 - bc\mathcal{E} \right) \left( \eta - \frac{1}{2}\eta^2 \right) \log b - \left( \eta + c\mathcal{E} \right) \log \left( 1 + \frac{bc\mathcal{E}}{\eta} \right) \]

\[ + \left( \eta + c\mathcal{E} \right) \left\{ \log(1 - bc\mathcal{E}) + \log \left( 1 - \frac{\eta}{2} \right) \right\} - \eta - c\mathcal{E} \]

\[ \geq \left( \eta - \frac{1}{2}\eta^2 \right) \log b - bc\mathcal{E} \eta \log b - \left( \eta + c\mathcal{E} \right) \log \left( 1 + \frac{bc\mathcal{E}}{\eta} \right) \]

\[ - \left( \eta + c\mathcal{E} \right) \left\{ \log(1 - bc\mathcal{E}) + \log \left( 1 - \frac{\eta}{2} \right) \right\} - \eta - c\mathcal{E}. \]

Combining (28), (47), and (69) yields (29).
V. PROOF OF THE SOFT-DECISION-PPM LOWER BOUND (19)

Consider the following soft-decision PPM scheme:

**Scheme 2.** The transmitter performs the same PPM as in Scheme 1. The receiver maps the b output symbols to a “super symbol” $\hat{y}$ that takes value in $\{1, \ldots, b, ?\} \cup \{(i,j): i, j \in \{1, \ldots, b\}, i \neq j\}$. The mapping rule is as follows. Take $\hat{y} = i$ if $y_i$ is the unique nonzero term in $\{y_1, \ldots, y_b\}$; take $\hat{y} = (i, j)$ if $y_i$ and $y_j$ are the only two nonzero terms in $\{y_1, \ldots, y_b\}$; if there are more than two or no nonzero term in $\{y_1, \ldots, y_b\}$, take $\hat{y} = ?$.

**Proposition 2.** Under the conditions (25), Scheme 2 achieves photon efficiency

$$C_{\text{PE-PPM(\text{SD})}}(\mathcal{E}, c) \geq \left(1 - \frac{\eta}{2}\right) \log b - c\eta \log b - \left(1 + \frac{c\mathcal{E}}{\eta}\right) \log(1 + c)$$

$$- \left(1 + \frac{c\mathcal{E}}{\eta}\right) \left\{ \log(1 - c\eta) + \log \left(1 - \frac{\eta}{2}\right) \right\} - 1 - \frac{c\mathcal{E}}{\eta}$$

$$+ \left(1 - \frac{\eta}{2}\right) (b - 1) \left( c\mathcal{E} - \frac{c^2\mathcal{E}^2}{2} \right) (1 - c\eta) \log \frac{b}{2}$$

$$- \frac{c^2}{2} \eta - c(\eta + c\mathcal{E}) \right)$$

(70)

where $b$ and $\eta$ are given in (26) and (28), respectively.

Proof of (19) using Proposition 2: Simplifying (70) we obtain

$$\lim_{\mathcal{E} \downarrow 0} \left\{ C_{\text{PE-PPM(\text{SD})}}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}} \right\}$$

$$\geq - \log(1 + c) - \frac{3}{2}.$$

(71)

This immediately yields (19).

Proof of Proposition 2: We compute the transition matrix of the super channel that results from Scheme 2. We first note

$$\hat{W}(i|i) = p_0,$$

(72)

$$\hat{W}(j|i) = p_1, \quad i, j \in \{1, \ldots, b\}, \ i \neq j,$$

(73)

where $p_0$ and $p_1$ are given in (38) and (44), respectively. For the remaining elements of the
transition matrix we have

\[ \hat{W}(i, j | i) = \Pr[Y_i \geq 1 | X_i = \eta] \Pr[Y_j \geq 1 | X_j = 0] \]

\[ \cdot \prod_{\ell \notin \{i, j\}} \Pr[Y_\ell = 0 | X_\ell = 0] \]  

(74)

\[ = (1 - e^{-\eta - cE})(1 - e^{-cE})e^{-(b-2)cE} \]  

(75)

\[ \triangleq p_2, \quad \{i, j\} \subset \{1, \ldots, b\}; \]  

(76)

\[ \hat{W}(j, k | i) = \Pr[Y_i = 0 | X_i = \eta] \Pr[Y_j \geq 1 | X_j = 0] \]  

\[ \cdot \Pr[Y_k \geq 1 | X_k = 0] \cdot \prod_{\ell \notin \{i, j, k\}} \Pr[Y_\ell = 0 | X_\ell = 0] \]  

(77)

\[ = e^{-\eta - cE}(1 - e^{-cE})^2 e^{-(b-3)cE} \]  

(78)

\[ \triangleq p_3, \quad \{i, j, k\} \subset \{1, \ldots, b\}; \]  

(79)

\[ \hat{W}(? | i) = 1 - p_0 - (b - 1)p_1 - (b - 1)p_2 - \frac{(b - 1)(b - 2)}{2} p_3 \]  

(80)

\[ \triangleq p_4, \quad i \in \{1, \ldots, b\}. \]  

(81)

We now have, irrespective of the distribution of \( \tilde{X} \),

\[ H(\hat{Y} | \tilde{X}) = p_0 \log \frac{1}{p_0} + (b - 1)p_1 \log \frac{1}{p_1} \]

\[ = (b - 1)p_2 \log \frac{1}{p_2} + \frac{(b - 1)(b - 2)}{2} p_3 \log \frac{1}{p_3} + p_4 \log \frac{1}{p_4}. \]  

(82)

Choosing a uniform \( \tilde{X} \), we have the following distribution on \( \hat{Y} \)

\[ P_{\hat{Y}}(i) = \frac{p_0 + (b - 1)p_1}{b} \]  

(83)

\[ P_{\hat{Y}}((i, j)) = \frac{2p_2 + (b - 2)p_3}{b} \]  

(84)

\[ P_{\hat{Y}}(?) = p_4, \quad \{i, j\} \subset \{1, \ldots, b\}, \]  

(85)

therefore

\[ H(\hat{Y}) = \frac{p_0 + (b - 1)p_1}{b} \log \frac{b}{p_0 + (b - 1)p_1} \]

\[ + \left( (b - 1)p_2 + \frac{(b - 1)(b - 2)}{2} p_3 \right) \log \frac{b}{2p_2 + (b - 2)p_3} \]

\[ + p_4 \log \frac{1}{p_4}. \]  

(86)
Using (82) and (86) we have

\[
I(\tilde{X}; \tilde{Y}) = H(\tilde{Y}) - H(\tilde{Y}|\tilde{X})
\]

\[
= p_0 \log \frac{b p_0}{p_0 + (b - 1)p_1}
+ (b - 1)p_1 \log \frac{b p_1}{p_0 + (b - 1)p_1}
+ (b - 1)p_2 \log \frac{b p_2}{2p_2 + (b - 2)p_3}
+ \frac{(b - 1)(b - 2)}{2}p_3 \log \frac{b p_3}{2p_2 + (b - 2)p_3}.
\]

\[
(87)
\]

At this point, note that the first two summands on the right-hand side of (88) constitute
\(I(\tilde{X}; \tilde{Y})\), which we analyzed in Section IV. It remains to find lower bounds on the last two summands.

We lower-bound the third term on the right-hand side of (88) as

\[
(b - 1)p_2 \log \frac{b p_2}{2p_2 + (b - 2)p_3}
= (b - 1)p_2 \log \frac{b}{2} - (b - 1)p_2 \log \frac{p_2 + \frac{b - 2}{2}p_3}{p_2}
\leq \frac{(b - 2)p_3}{2p_2}
\]

\[
\geq (b - 1)p_2 \log \frac{b}{2} - \frac{b^2}{2} p_3.
\]

\[
(89)
\]

We lower-bound the fourth term on the right-hand side of (88) as

\[
\frac{(b - 1)(b - 2)}{2}p_3 \log \frac{b p_3}{2p_2 + (b - 2)p_3}
= - \frac{(b - 1)(b - 2)}{2}p_3 \log \frac{2p_2 + (b - 2)p_3}{b p_3}
\leq \frac{2p_2}{b p_3}
\]

\[
\geq - \frac{(b - 1)(b - 2)}{2} \cdot \frac{2p_2}{b}
\]

\[
\geq -bp_2.
\]

\[
(90)
\]

\[
(91)
\]

\[
(92)
\]

\[
(93)
\]

We then observe the upper bound on \(p_2\)

\[
p_2 = \frac{1}{(1 - e^{-\eta - c\mathcal{E}})(1 - e^{-c\mathcal{E}}) e^{-(b - 2)c\mathcal{E}}} \leq (\eta + c\mathcal{E})c\mathcal{E}.
\]

\[
\leq (\eta + c\mathcal{E})c\mathcal{E}.
\]

\[
(94)
\]

\[
(95)
\]
the lower bound on $p_2$

$$p_2 = \left( \frac{1 - e^{-\eta - cE}}{1 - e^{-cE}} \right) \left( \frac{1 - e^{-cE}}{1 - e^{-2cE}} \right) \geq 1 - e^{\eta - \frac{cE}{2}} \geq \frac{1 - e^{-cE}}{1 - e^{-2cE}} \geq 1 - e^{\eta - \frac{cE}{2}}$$

(96)

$$\geq \left( \frac{\eta - \frac{c^2 E^2}{2}}{2} \right) \left( cE - \frac{c^2 E^2}{2} \right) (1 - bcE),$$

(97)

and the upper bound on $p_3$

$$p_3 = e^{-\eta - cE} \left( \frac{1 - e^{-cE}}{1 - e^{-cE}} \right)^2 e^{-(b-3)cE} \leq e^{-\eta - cE} \leq \frac{1 - e^{-cE}}{1 - e^{-cE}} \leq 1.$$ 

(98)

$$\leq c^2 E^2.$$ 

(99)

From (88), (90), (93), (95), (97), and (99) we obtain a lower bound on the additional mutual information that is gained by considering output frames with two detection positions:

$$I(\tilde{X}; \hat{Y}) - I(\tilde{X}; \tilde{Y}) \geq (b - 1)p_2 \log \frac{b}{2} - \frac{b^2}{2}p_3 - bp_2 - 1$$

(100)

$$\geq (b - 1) \left( \frac{\eta - \frac{c^2 E^2}{2}}{2} \right) \left( cE - \frac{c^2 E^2}{2} \right) (1 - bcE) \log \frac{b}{2}$$

$$- \frac{b^2}{2}c^2 E^2 - b(\eta + cE)cE.$$ 

(101)

Dividing the above by $\eta$ and plugging in (28) we obtain a lower bound on the additional photon efficiency:

$$\frac{I(\tilde{X}; \hat{Y}) - I(\tilde{X}; \tilde{Y})}{\eta} \geq \left( 1 - \frac{\eta}{2} \right) (b - 1) \left( cE - \frac{c^2 E^2}{2} \right) \left( 1 - c\eta \right) \log \frac{b}{2}$$

$$- \frac{c^2}{2} \eta - c(\eta + cE).$$ 

(102)

Adding the right-hand side of (102) to the right-hand side of (29) yields (70).

VI. PROOF OF THE UPPER BOUNDS (20) AND (21)

**Proposition 3.** Assume that $\mathcal{E}$ is small enough so that

$$\mathcal{E} < e^{-1}$$

(103a)

$$\mathcal{E} \log \frac{1}{\mathcal{E}} < \frac{1}{1 + c}$$

(103b)

$$\mathcal{E} \left( \log \frac{1}{\mathcal{E}} \right)^4 < \frac{144}{(1 + c)^2}.$$ 

(103c)
then

\[
C_{PE}(\mathcal{E}, c) \leq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} - \log(1 + c) + 2 + \log 13 \\
+ \frac{1}{\mathcal{E}} \left( \log \frac{1}{1 - (1 + c)\mathcal{E}} - (1 + c)\mathcal{E} \right) \\
+ \frac{c^2}{2} \mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1 + c) \log \frac{1}{1 - \log \frac{1}{\mathcal{E}}}.
\]  

(104)

**Proof of (20) using Proposition 3:** The last three summands on the right-hand side of (104) are all of the form \(o(1)\). Setting \(c = 0\) in (104) we have

\[
C_{PE}(\mathcal{E}, 0) \leq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + 2 + \log 13 + o(1)
\]  

(105)

\[= \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1).
\]  

(106)

\[\blacksquare\]

**Proof of (21) using Proposition 3:** From (104) we have

\[
\lim_{\mathcal{E} \downarrow 0} \left\{ C_{PE}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}} \right\} \\
\leq - \log(1 + c) + 2 + \log 13.
\]  

(107)

Hence

\[
\lim_{\mathcal{E} \downarrow 0} \lim_{c \to \infty} \frac{C_{PE}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}}}{\log c} \\
\leq \lim_{c \to \infty} \frac{- \log(1 + c) + 2 + \log 13}{\log c}
\]  

(108)

\[= -1.
\]  

(109)

\[\blacksquare\]

**Proof of Proposition 3:** Like in [2], we use the duality bound [13] which states that, for any distribution \(R(\cdot)\) on the output, the channel capacity satisfies

\[
C \leq \sup \mathbb{E} \left[ D(W(\cdot|X)\| R(\cdot)) \right],
\]  

(110)

where the supremum is taken over all allowed input distributions. We choose \(R(\cdot)\) to be the following distribution:

\[
R(y) = \begin{cases} 
1 - (1 + c)\mathcal{E}, & y = 0 \\
(1 + c)\mathcal{E} \left( 1 - \frac{1}{\log \frac{1}{\mathcal{E}}} \right) \left( \frac{1}{\log \frac{1}{\mathcal{E}}} \right)^{y-1}, & y = 1, 2, \ldots.
\end{cases}
\]  

(111)
For any $x \geq 0$ we have

\[
D(W(\cdot|x)\|R(\cdot)) = \sum_{y=0}^{\infty} \text{Pois}_{x+c\mathcal{E}}(y) \cdot \log \frac{1}{R(y)} - H(Y|X = x) 
\]

\[
= e^{-(x+c\mathcal{E})} \log \frac{1}{1 - (1 + c)\mathcal{E}} 
\]
\[
- \sum_{y=1}^{\infty} \text{Pois}_{x+c\mathcal{E}}(y) \log \left\{ (1 + c)\mathcal{E} \left( 1 - \frac{1}{\log \frac{1}{\mathcal{E}}} \right) \left( \frac{1}{\log \frac{1}{\mathcal{E}}} \right)^{y-1} \right\} 
\]

\[
= e^{-(x+c\mathcal{E})} \log \frac{1}{1 - (1 + c)\mathcal{E}} + \sum_{y=1}^{\infty} \text{Pois}_{x+c\mathcal{E}}(y) \log \log \frac{1}{\mathcal{E}} 
\]

\[
+ \left( \sum_{y=1}^{\infty} \text{Pois}_{x+c\mathcal{E}}(y) \right) \left( \log \frac{1}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \log \frac{1}{1 - \log \frac{1}{\mathcal{E}}} \right) 
\]

\[
- H(Y|X = x) 
\]

\[
= e^{-(x+c\mathcal{E})} \log \frac{1}{1 - (1 + c)\mathcal{E}} + (x + c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} 
\]

\[
+ (1 - e^{-(x+c\mathcal{E})}) \left( \log \frac{1}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \log \frac{1}{1 - \log \frac{1}{\mathcal{E}}} \right) 
\]

\[
- H(Y|X = x). 
\]

We can lower-bound $H(Y|X = x)$ as

\[
H(Y|X = x) \geq H_b(e^{-(x+c\mathcal{E})}) 
\]

\[
= e^{-(x+c\mathcal{E})} (x + c\mathcal{E}) + (1 - e^{-(x+c\mathcal{E})}) \log \frac{1}{1 - e^{-(x+c\mathcal{E})}}. 
\]

Now we can continue (115) to upper-bound $D(W(\cdot|x)\|R(\cdot))$ as

\[
D(W(\cdot|x)\|R(\cdot)) \leq e^{-(x+c\mathcal{E})} \log \frac{1}{1 - (1 + c)\mathcal{E}} + (x + c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} 
\]

\[
+ (1 - e^{-(x+c\mathcal{E})}) \left( \log \frac{1}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \log \frac{1}{1 - \log \frac{1}{\mathcal{E}}} \right) 
\]

\[
- e^{-(x+c\mathcal{E})} (x + c\mathcal{E}) - (1 - e^{-(x+c\mathcal{E})}) \log \frac{1}{1 - e^{-(x+c\mathcal{E})}} 
\]
\[= e^{-(x+c\mathcal{E})} \log \frac{1}{1 - (1 + c)\mathcal{E}} + (x + c\mathcal{E}) \log \log \frac{1}{\mathcal{E}}\]

\[+ (1 - e^{-(x+c\mathcal{E})}) \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}} \leq_{x+c\mathcal{E}} 0\]

\[- e^{-(x+c\mathcal{E})} (x + c\mathcal{E}) \geq_{c\mathcal{E}}\]

\[+ (1 - e^{-(x+c\mathcal{E})}) \log \frac{1 - e^{-(x+c\mathcal{E})}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \leq_{1} \left( \log \frac{1}{1 - (1 + c)\mathcal{E} - c\mathcal{E}} \right) \geq_{0}\]

\[+ (x + c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} + (x + c\mathcal{E}) \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}}\]

\[+ (e^{-c\mathcal{E}} - e^{-(x+c\mathcal{E})}) \log \frac{1 - e^{-(x+c\mathcal{E})}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + (1 - e^{-c\mathcal{E}}) \log \frac{1 - e^{-(x+c\mathcal{E})}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}}\]

\[= \mathcal{E} + \left( \log \frac{1}{1 - (1 + c)\mathcal{E}} - (1 + c)\mathcal{E} \right)\]

\[+ (x + c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} + (x + c\mathcal{E}) \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}}\]

\[+ e^{-c\mathcal{E}} (1 - e^{-x}) \log \frac{1 - e^{-(x+c\mathcal{E})}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + (1 - e^{-c\mathcal{E}}) \log \frac{c}{(1 + c) \log \frac{1}{\mathcal{E}}} \geq_{c\mathcal{E} - \mathcal{E}^{2}c^{2}} \leq_{\log \log \frac{1}{\mathcal{E}}}\]
\[
\begin{align*}
&+ (1 - e^{-\epsilon}) \log \frac{x + c\epsilon}{c\epsilon} \\
&\leq \epsilon + \left( \log \frac{1}{1 - (1 + c)\epsilon} - (1 + c)\epsilon \right) \\
&+ (x + c\epsilon) \log \frac{1}{\epsilon} + (x + c\epsilon) \log \frac{1}{1 - \frac{1}{\log \frac{1}{\epsilon}}} \\
&+ e^{-\epsilon} (1 - e^{-x}) \log \frac{1 - e^{-(x+c\epsilon)}}{(1 + c)\epsilon \log \frac{1}{\epsilon}} \\
&- \left( c\epsilon - \frac{c^2}{2} \epsilon^2 \right) \log \log \frac{1}{\epsilon} + x.
\end{align*}
\]

(122)

Using (110) and (123) together with constraint (3) we have

\[
C(\epsilon, c) \leq \sup \mathbb{E} \left[ D(W(\cdot | X) \| R(\cdot)) \right] \\
\leq \epsilon + \left( \log \frac{1}{1 - (1 + c)\epsilon} - (1 + c)\epsilon \right) \\
+ (1 + c)\epsilon \log \log \frac{1}{\epsilon} + (1 + c)\epsilon \log \frac{1}{1 - \frac{1}{\log \frac{1}{\epsilon}}} \\
- \left( c\epsilon - \frac{c^2}{2} \epsilon^2 \right) \log \log \frac{1}{\epsilon} + \epsilon \\
+ e^{-\epsilon} \sup \mathbb{E} \left[ (1 - e^{-X}) \log \frac{1 - e^{-(X+c\epsilon)}}{(1 + c)\epsilon \log \frac{1}{\epsilon}} \right] \\
= \epsilon \log \log \frac{1}{\epsilon} + 2\epsilon \\
+ \left( \log \frac{1}{1 - (1 + c)\epsilon} - (1 + c)\epsilon \right) \\
+ \frac{c^2}{2} \epsilon^2 \log \log \frac{1}{\epsilon} + (1 + c)\epsilon \log \frac{1}{1 - \frac{1}{\log \frac{1}{\epsilon}}} \\
+ e^{-\epsilon} \sup \mathbb{E} \left[ (1 - e^{-X}) \log \frac{1 - e^{-(X+c\epsilon)}}{(1 + c)\epsilon \log \frac{1}{\epsilon}} \right].
\]

(125)

Hence the maximum achievable photon efficiency satisfies

\[
C_{PE}(\epsilon, c) = \frac{C(\epsilon, c)}{\epsilon} \\
\leq \log \log \frac{1}{\epsilon} + 2 + \frac{1}{\epsilon} \left( \log \frac{1}{1 - (1 + c)\epsilon} - (1 + c)\epsilon \right)
\]

(127)
\[ + \frac{e^2}{2} \mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1 + c) \log \frac{1}{1 - \log \frac{1}{\mathcal{E}}} \]
\[ + \frac{e^{-c\mathcal{E}}}{\mathcal{E}} \sup_{x} \mathbb{E} \left[ (1 - e^{-X}) \log \frac{1 - e^{-(X+c\mathcal{E})}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right]. \quad (128) \]

We next upper-bound the expectation on the right-hand side of (128) as follows
\[ \mathbb{E} \left[ (1 - e^{-X}) \log \frac{1 - e^{-(X+c\mathcal{E})}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right] \]
\[ = \mathbb{E} \left[ X \cdot \frac{1 - e^{-X}}{X} \log \frac{1 - e^{-(X+c\mathcal{E})}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right] \quad (129) \]
\[ \leq \mathbb{E} \left[ X \cdot \frac{1 - e^{-X}}{X} \log \frac{X + c\mathcal{E}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right] \quad (130) \]
\[ \leq \mathcal{E} \cdot \sup_{x} \phi(x) \quad (131) \]

where
\[ \phi(x) \triangleq \frac{1 - e^{-x}}{x} \log \frac{x + c\mathcal{E}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}}, \quad x \geq 0. \quad (132) \]

We claim that, if \( \mathcal{E} \) is small enough to satisfy all of the conditions in (103), then
\[ \sup_{x} \phi(x) = \phi(x^*) \quad \text{for some} \quad x^* \leq \frac{12}{\log \frac{1}{\mathcal{E}}}. \quad (133) \]

To show this, we compute the derivative of \( \phi(x) \) (with respect to \( x \)) as
\[ \frac{d\phi(x)}{dx} = \frac{(1 + x)e^{-x} - 1}{x^2} \log \frac{x + c\mathcal{E}}{(1 + c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \frac{1 - e^{-x}}{x(x + c\mathcal{E})} \quad (134) \]

which exists and is continuous for all \( x > 0 \). Further note that it is negative for large enough \( x \), and is positive for small enough \( x \). Hence \( x^* \) must be achieved at a point where
\[ \frac{d\phi(x)}{dx} = 0. \quad (135) \]

To prove (133), it thus suffices to show that, under the assumptions (103),
\[ \frac{d\phi(x)}{dx} < 0, \quad x > \frac{12}{\log \frac{1}{\mathcal{E}}}. \quad (136) \]

We do this separately for two cases.

**Case 1. Consider**
\[ \frac{12}{\log \frac{1}{\mathcal{E}}} < x < 1. \quad (137) \]

*Note that this case need not be considered if \( \mathcal{E} \geq e^{-12} \).*
In this case the logarithmic term in the derivative is positive:

\[
\log \frac{x + c\mathcal{E}}{(1 + c)\mathcal{E} \log \frac{1}{x}} \geq \log \frac{x}{(1 + c)\mathcal{E} \log \frac{1}{x}} \quad (138)
\]
\[
> \log \frac{12}{(1 + c)\mathcal{E} \left(\log \frac{1}{x}\right)^2} \quad (139)
\]
\[
> 0, \quad (140)
\]

where the last step follows because (103a) and (103c) together imply

\[
\mathcal{E} \left(\log \frac{1}{\mathcal{E}}\right)^2 < \frac{12e^{-\frac{1}{2}}}{1+c}. \quad (141)
\]

Next note

\[
\frac{(1 + x)e^{-x} - 1}{x^2} \leq -\frac{1}{2} + \frac{1}{3} \leq \frac{1}{6} \quad (142)
\]

and

\[
\frac{1 - e^{-x}}{x(x+c\mathcal{E})} \leq \frac{1}{x}. \quad (143)
\]

Plugging (140), (142) and (143) into (134) we have

\[
\frac{d\phi(x)}{dx} \leq -\frac{1}{6} \log \frac{x + c\mathcal{E}}{(1 + c)\mathcal{E} \log \frac{1}{x}} + \frac{1}{x}
\]
\[
\leq -\frac{1}{6} \log x + \frac{1}{6} \log(1 + c) - \frac{1}{6} \log \frac{1}{\mathcal{E}}
\]
\[
\quad + \frac{1}{6} \log \log \frac{1}{\mathcal{E}} + \frac{1}{x} \quad (145)
\]
\[
\leq \frac{1}{6} \log \frac{1 + c}{12} + \frac{1}{3} \log \log \frac{1}{\mathcal{E}} - \frac{1}{12} \log \frac{1}{\mathcal{E}}
\]
\[
= \frac{1}{12} \log \left(\frac{(1 + c)^2}{144} \mathcal{E} \left(\log \frac{1}{\mathcal{E}}\right)^4\right) \quad (147)
\]
\[
< 0, \quad (148)
\]

where the last step follows from (103c). We have now shown

\[
\frac{d\phi(x)}{dx} < 0, \quad \frac{12}{\log \frac{1}{x}} < x < 1. \quad (149)
\]

**Case 2. Consider**

\[
x \geq 1. \quad (150)
\]
In this case the logarithmic term in the derivative is still positive:
\[
\log \frac{x + cE}{(1 + c)E \log \frac{1}{E}} \geq \log \frac{1}{(1 + c)E \log \frac{1}{E}} > 0,
\]
where the last step follows from (103b). We bound the other terms as
\[
\frac{(1 + x)e^{-x} - 1}{x^2} \leq -\frac{1 - 2e^{-1}}{x^2}
\]
which is negative, and
\[
\frac{1 - e^{-x}}{x(x + cE)} \leq \frac{1}{x^2}.
\]
Using (151), (152) and (153) together with (134) we have
\[
\frac{d\phi(x)}{dx} \leq -\frac{1 - 2e^{-1}}{x^2} \left( \log(x + cE) + \log \frac{1}{(1 + c)E \log \frac{1}{E}} \right) + \frac{1}{x^2}
\]
\[
\leq -\frac{1 - 2e^{-1}}{x^2} \log \frac{e^{-1 - 2e^{-1}}}{(1 + c)E \log \frac{1}{E}}
\]
\[
< 0,
\]
where the last step follows from (103b). Hence we have shown
\[
\frac{d\phi(x)}{dx} < 0, \quad x \geq 1.
\]
Combining (149) and (157) proves (133) for all $E$ satisfying (103).

We now proceed to upper-bound $\sup \phi(x)$:
\[
\sup \phi(x) = \phi(x^*)
\]
\[
= \frac{1 - e^{-x^*}}{x^*} \log \frac{x^* + cE}{(1 + c)E \log \frac{1}{E}}
\]
\[
\leq \max \left\{ 0, \log \frac{x^* + cE}{(1 + c)E \log \frac{1}{E}} \right\}
\]
\[
\leq \max \left\{ 0, \log \frac{12}{\log \frac{1}{E}} + cE \right\}.
\]
Now note that, due to (103b), we have
\[
cE \leq \frac{1}{\log \frac{1}{E}}.
\]
We can now continue the chain of inequalities (161) as
\[
\sup \phi(x) \leq \max \left\{ 0, \log \frac{13}{(1 + c) \mathcal{E} \left( \log \frac{1}{\mathcal{E}} \right)^2} \right\}
\]  
(163)

\[
= \log \frac{13}{(1 + c) \mathcal{E} \left( \log \frac{1}{\mathcal{E}} \right)^2}
\]  
(164)

\[
= \log \frac{1}{\mathcal{E}} - 2 \log \log \frac{1}{\mathcal{E}} - \log(1 + c) + \log 13.
\]  
(165)

Here (164) follows because conditions (103a) and (103c) imply (141).

Combining (128), (131) and (165), and noting \(e^{-c\mathcal{E}} \leq 1\), we obtain
\[
C_{PE}(\mathcal{E}, c) \leq \log \log \frac{1}{\mathcal{E}} + 2 + \frac{1}{\mathcal{E}} \left( \log \frac{1}{1 - (1 + c) \mathcal{E}} - (1 + c) \mathcal{E} \right)
\]
\[
+ \frac{c^2}{2} \mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1 + c) \log \frac{1}{1 - \log \frac{1}{\mathcal{E}}}
\]
\[
+ \log \frac{1}{\mathcal{E}} - 2 \log \log \frac{1}{\mathcal{E}} - \log(1 + c) + \log 13
\]  
(166)

\[
= \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} - \log(1 + c) + 2 + \log 13
\]
\[
+ \frac{1}{\mathcal{E}} \left( \log \frac{1}{1 - (1 + c) \mathcal{E}} - (1 + c) \mathcal{E} \right)
\]
\[
+ \frac{c^2}{2} \mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1 + c) \log \frac{1}{1 - \log \frac{1}{\mathcal{E}}}
\]  
(167)

\[\blacksquare\]

VII. Numerical Comparison and Concluding Remarks

We numerically compare the approximation (23) with nonasymptotic upper and lower bounds on photon efficiency. Specifically, the plotted upper bound is based on (110), (122), (132), and (165). The on-off-keying lower bound is obtained by computing the mutual information of the channel with “on” signal equaling \(1/\left(\log(1/\mathcal{E})\right)\), and with the receiver ignoring multiple detected photons. The simple-PPM lower bound is computed using (47) and (53). The soft-decision-PPM lower bound is computed using (88). We plot these bounds for \(c = 0.1, c = 1,\) and \(c = 10\) in Figures 1, 2, and 3, respectively. The figures show the following.

- The approximation (23) is reasonably accurate for small enough \(\mathcal{E}\).
• The on-off-keying and the soft-decision-PPM bounds are consistently close to the approximation (23).

• As \( c \) increases, the simple-PPM bound deviates significantly from the other lower bounds, and hence also from the actual value of \( C_{PE}(E, c) \).

• The upper bound is always about 3 nats per photon above the the approximation (23). This is largely due to the \( \log 13 \) term in the upper bound (165). This gap could be narrowed by a more careful study of \( \sup_{x} \phi(x) \), in particular, by improving the bound on \( x^* \) in (133), which we do not attempt in this paper.

The capacity bounds as well as the asymptotic results in this paper provide insights to how technical restrictions and device or channel imperfections influence the communication rates in optical channels in the wideband regime. Our results and techniques can also be applied to secret communication and key distribution over optical channels [14]. It would be interesting to see if these are further extendable, e.g., to multiple-mode optical channels and optical networks.

As we have demonstrated, PPM is nearly optimal on the Poisson channel in the high-photon-efficiency regime. When \( c \) is small, the simple-PPM super channel has high erasure probability but low “error” (by which we mean the receiver detects a single pulse at a position that is different from the transmitted signal) probability. In this case, Reed-Solomon codes can perform rather close to the theoretical limit. However, when \( c > 1 \), Reed-Solomon codes can no longer achieve any positive rates on this channel. Nevertheless, we believe that, for \( c > 1 \), PPM still has its advantages over on-off keying in terms of code design. This is because the optimal input distribution for (both simple and soft-decision) PPM is uniform, whereas the optimal input distribution for on-off keying for this channel is highly skewed. The uniformity of PPM inputs allows the usage of more structured codes, in particular linear codes. One possible direction, for instance, is to employ the idea of multilevel codes [15], [16] on this channel.

**Appendix**

In this appendix we sketch a proof of (7). We start with the upper bound, which is much simpler than the one given in [2]. Using the same arguments as in [2], we know that \( C(E, c) \) is monotonically increasing in \( c \). Hence it suffices to show

\[
\lim_{E \downarrow 0} \frac{C(E, 0)}{E \log \frac{1}{E}} \leq 1. \tag{168}
\]
This can be verified as follows:

\[ C(\mathcal{E}, 0) = \max_{\mathbb{E}[X] \leq \mathcal{E}} I(X; Y) \] (169)

\[ \leq \max_{\mathbb{E}[X] \leq \mathcal{E}} H(Y) \] (170)

\[ = \max_{\mathbb{E}[Y] \leq \mathcal{E}} H(Y) \] (171)

\[ = (1 + \mathcal{E}) \log(1 + \mathcal{E}) - \mathcal{E} \log \mathcal{E} \] (172)

which yields the desired asymptotic bound. Here, (172) follows from the well-known fact that, for a nonnegative discrete random variable under a first-moment constraint, maximum entropy is achieved by the geometric distribution, and is given by the right-hand side of (172). Note that the right-hand side of (172) is exactly the capacity of the pure-loss bosonic channel [7].

For the lower bound, we let the encoder employ on-off keying, choosing the “on” signal to be some constant \( \eta \). We let the decoder ignore multiple photons: all nonzero values of \( Y \) are mapped to \( Y' = 1 \). This induces an asymmetric binary channel:

\[ W'(1|0) = e^{-c \mathcal{E}} \] (173a)

\[ W'(1|1) = e^{-\eta - c \mathcal{E}}. \] (173b)

We skip the remaining calculations and simply claim that, as \( \eta \) approaches zero, the mutual information thus obtained gives the desired lower bound. The main difference between this proof procedure and the one in [2] is that the latter considers all values of \( Y \), rather than simplifying it to binary.

ACKNOWLEDGMENTS

The authors thank Hye Won Chung, Yuval Kochman, Amos Lapidoth, Arya Mazumdar, and Lizhong Zheng for helpful discussions.

REFERENCES

[1] J. H. Shapiro, “The quantum theory of optical communications,” IEEE J. Select. Top. Quantum Electron., vol. 15, no. 6, pp. 1547–1569, 2009.

[2] A. Lapidoth, J. H. Shapiro, V. Venkatesan, and L. Wang, “The discrete-time Poisson channel at low input powers,” IEEE Trans. Inform. Theory, vol. 57, no. 6, pp. 3260–3272, June 2011.

[3] J. Pierce, “Optical channels: practical limits with photon counting,” IEEE Trans. Commun., vol. 26, no. 12, pp. 1819–1821, Dec. 1978.
[4] J. L. Massey, “Capacity, cutoff rate, and coding for direct-detection optical channel,” *IEEE Trans. Commun.*, vol. 29, no. 11, pp. 1615–1621, Nov. 1981.

[5] S. Shamai (Shitz), “Capacity of a pulse amplitude modulated direct detection photon channel,” in *Proc. IEE*, vol. 137, pt. 1 (Communications, Speech and Vision), no. 6, Dec. 1990, pp. 424–430.

[6] A. Lapidoth and S. M. Moser, “On the capacity of the discrete-time Poisson channel,” *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 303–322, Jan. 2009.

[7] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, “Classical capacity of the lossy bosonic channel: the exact solution,” *Phys. Rev. Lett.*, vol. 92, no. 2, p. 027902(4), 2004.

[8] R. G. Gallager, “Energy limited channels: Coding, multiaccess, and spread spectrum,” Laboratory for Information and Decision Systems, MIT, Tech. Rep. 1714, Nov. 1987.

[9] S. Verdú, “On channel capacity per unit cost,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.

[10] H. W. Chung, S. Guha, and L. Zheng, “On capacity of optical channels with coherent detection,” in *Proc. IEEE Int. Symp. Inform. Theory*, Saint Petersburg, Russia, July 31–August 5 2011.

[11] Y. Kochman and G. W. Wornell, “On high-efficiency optical communication and key distribution,” in *Proc. Inf. Theory and Appl. Workshop*, San Diego, CA, USA, Feb. 5–10 2012.

[12] S. Dolinar, K. M. Birnbaum, B. M. Erkmen, and B. Moision, “On approaching the ultimate limit of photon-efficient and bandwidth efficient optical communication,” in *Int. Conf. Space Optical Sys.and App.*, Santa Monica, CA, USA, May 11—13 2011.

[13] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[14] Y. Kochman, L. Wang, and G. W. Wornell, “Toward photon-efficient key distribution over optical channels,” 2013, subm. to *IEEE Trans. Inform. Theory*.

[15] H. Imai and S. Hirakawa, “A new multilevel coding method using error correcting codes,” *IEEE Trans. Inform. Theory*, vol. 23, pp. 371–377, 1977.

[16] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, “Multilevel codes: Theoretical concepts and practical design rules,” *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.
Fig. 1. Comparing the approximation (23) to nonasymptotic upper bound, and to lower bounds for simple PPM, soft-decision PPM, and on-off keying for $c = 0.1$. Note that the last three curves almost overlap.
Fig. 2. Comparing the approximation (23) to nonasymptotic upper bound, and to lower bounds for simple PPM, soft-decision PPM, and on-off keying for $c = 1$. Note that the last two curves almost overlap.
Fig. 3. Comparing the approximation (23) to nonasymptotic upper bound, and to lower bounds for simple PPM, soft-decision PPM, and on-off keying for $c = 10$. 