

# A Refined Analysis of the Poisson Channel in the High-Photon-Efficiency Regime

Ligong Wang, *Member, IEEE*, and Gregory W. Wornell, *Fellow, IEEE*

**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  $\mathcal{E}$ . We consider the wideband, high-photon-efficiency extreme where  $\mathcal{E}$  approaches zero, and where the channel’s dark current approaches zero proportionally with  $\mathcal{E}$ . 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.

**Index Terms**—Optical communication, Poisson channel, channel capacity, wideband, low SNR.

## 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

$$\text{Pois}_\xi(y) = e^{-\xi} \frac{\xi^y}{y!}, \quad y \in \mathbb{Z}_0^+. \quad (1)$$

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

$$W(y|x) = \text{Pois}_{\lambda+x}(y), \quad x \in \mathbb{R}_0^+, \quad y \in \mathbb{Z}_0^+. \quad (2)$$

This channel models pulse-amplitude modulated optical communication where the transmitter sends light signals in *coherent states* (which are usually produced using laser devices), and where the receiver employs *direct detection* (i.e., photon counting) [1]. The channel input  $x$  describes the expected number of *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,<sup>1</sup> the pulse duration, the channel’s transmissivity, and the detector’s efficiency; see [1] for details. The channel output  $y$  is the actual number of photons that are detected in the pulse duration. The dark current  $\lambda$  is the average number of extraneous counts that appear in  $y$ . We note that, although the name “dark current” is traditionally used in the Poisson-channel literature, these extraneous counts usually include both the detector’s “dark clicks” (which are where the name “dark current” comes from) and photons in background radiation.

We impose an *average-power constraint*<sup>2</sup> on the input

$$E[X] \leq \mathcal{E} \quad (3)$$

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

In applications like free-space optical communications, the cost of producing and successfully transmitting photons is high, hence high *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 *wideband* regime, where the pulse duration of the input approaches zero and, assuming that the *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

$$\lambda = c\mathcal{E}, \quad (4)$$

where  $c$  is some nonnegative constant that does not change with  $\mathcal{E}$ . In 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 [2, Proposition 2].

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,

<sup>1</sup>We assume that the pulses are square, which is usually the case in practice. If they are not square, the light signal’s intensity should be averaged over the pulse duration.

<sup>2</sup>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.

Manuscript received November 19, 2013; revised March 13, 2014; accepted April 22, 2014. Date of publication April 29, 2014; date of current version June 12, 2014. This work was supported in part by the Defense Advanced Research Projects Agency InPho Program under Contract HR0011-10-C-0159 and in part by the Air Force Office of Scientific Research under Grant FA9550-11-1-0183. This paper was presented at the 2012 IEEE Information Theory Workshop and 2014 IEEE International Symposium on Information Theory.

The authors are with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: wlg@mit.edu; gww@mit.edu).

Communicated by R. Fischer, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2014.2320718

our results on the asymptotic behavior of  $C(\mathcal{E}, c)$  hold irrespectively of whether a *peak-power constraint*

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

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

We now formally define the maximum achievable photon efficiency  $C_{\text{PE}}(\mathcal{E}, c)$ , where the subscript notation services a reminder that this quantity is photon efficiency and not capacity:

$$C_{\text{PE}}(\mathcal{E}, c) \triangleq \frac{C(\mathcal{E}, c)}{\mathcal{E}}. \quad (7)$$

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

$$\lim_{\mathcal{E} \downarrow 0} \frac{C(\mathcal{E}, c)}{\mathcal{E} \log \frac{1}{\mathcal{E}}} = 1, \quad c \in [0, \infty). \quad (8)$$

In other words, the maximum photon efficiency satisfies

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

Furthermore, the proof of [2, Proposition 1] shows that the limit (8) is achievable using on-off keying. In the Appendix we provide a proof for the converse part of (8) that is much simpler than the original proof in [2].<sup>3</sup>

The approximation in (9) can be compared to the maximum photon efficiency achievable on the *pure-loss bosonic channel* [7]. This is a quantum optical communication channel that attenuates the input optical signal, but that does *not* add any noise to it. The transmitter and the receiver of this channel may employ any structure permitted by quantum physics. When the transmitter is restricted to sending coherent optical states, and when the receiver is restricted to ideal direct detection (with no dark counts), the pure-loss bosonic channel becomes the Poisson channel (2) with  $\lambda = 0$ . 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 [7, eq. (4)], 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.

<sup>3</sup>For the achievability part of (8), we note that its proof can be simplified by letting the decoder ignore multiple photons, rather than considering all possible values of  $Y$  as in [2]. This can be seen via the achievability proofs in the current paper, which do ignore multiple photons at the receiver, and which yield stronger results than those in [2].

In the present paper, we refine the analysis in [2] in two aspects. First, we provide a more accurate approximation for  $C_{\text{PE}}(\mathcal{E}, c)$  that contains higher-order terms, and that reflects the influence of the dark current, i.e., of 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 (9). It is noticed in [10] that the photon efficiency achievable on the Poisson channel (2) with  $\lambda = 0$  may be of the form

$$\log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1), \quad (11)$$

which means that restriction to coherent-state inputs and direct detection may induce a loss in the *second-order term* in the photon efficiency on the pure-loss bosonic channel. For practical values of  $\mathcal{E}$ , this second-order term can be significant. For example, for  $\mathcal{E} = 10^{-5}$ , the difference between the first-order term in (9) and the first- and second-order terms in (11) is about 20%.

The analysis in [10] (whose main focus is not on the Poisson channel itself) is based on certain assumptions on the input distribution.<sup>4</sup> It is therefore unclear from [10] 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  $C_{\text{PE}}(\mathcal{E}, 0)$ . 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  $C_{\text{PE}}(\mathcal{E}, c)$ , 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

<sup>4</sup>According to conversation with the authors.

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

$$\lim_{\mathcal{E} \downarrow 0} \frac{o(f(\mathcal{E}))}{f(\mathcal{E})} = 0, \quad (12)$$

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

$$\overline{\lim}_{\mathcal{E} \downarrow 0} \left| \frac{O(f(\mathcal{E}))}{f(\mathcal{E})} \right| < \infty. \quad (13)$$

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

$$0 \log 0 = 0. \quad (14)$$

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 an approximation for the photon efficiency  $C_{\text{PE}}(\mathcal{E}, c)$  in the high-photon-efficiency regime.

*Theorem 1: The maximum photon efficiency  $C_{\text{PE}}(\mathcal{E}, c)$  achievable on the Poisson channel (2) subject to constraint (3) and with dark current (4), which we define in (7), satisfies*

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

Furthermore, denote

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

then

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

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  $C_{\text{PE-PPM}}(\mathcal{E}, c)$ , satisfies

$$C_{\text{PE-PPM}}(\mathcal{E}, c) \geq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1), \quad c \in [0, \infty); \quad (18)$$

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

$$\lim_{c \rightarrow \infty} \overline{\lim}_{\mathcal{E} \downarrow 0} \frac{C_{\text{PE-PPM(SD)}}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}}}{\log c} \geq -1. \quad (19)$$

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

$$C_{\text{PE}}(\mathcal{E}, 0) \leq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1); \quad (20)$$

and second, that

$$\overline{\lim}_{c \rightarrow \infty} \overline{\lim}_{\mathcal{E} \downarrow 0} \frac{C_{\text{PE}}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}}}{\log c} \leq -1. \quad (21)$$

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

To better understand the capacity results in Theorem 1, we make the following remarks.

- 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  $-\log c$ .

- Theorem 1 suggests the approximation

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

The terms constituting the approximation error, 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 asserts 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  $C_{\text{PE}}(\mathcal{E}, c)$  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. We hence need a better approximation for the constant term, which behaves like  $-\log c$  for large  $c$ , but which does not diverge for small  $c$ . As both the nonasymptotic bounds and the numerical simulations we show later will suggest,  $-\log(1+c)$  is a good approximation:

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

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

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

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.<sup>5</sup>

- In the PPM schemes that achieve (18) and (19), which we describe in Sections IV and V, the pulse has amplitude  $1/(\log(1/\mathcal{E}))$ , which depends on  $\mathcal{E}$ , and which tends to zero as  $\mathcal{E}$  tends to zero. An on-off keying scheme having the same amplitude for its “on” signals can also achieve these lower bounds, because the dependence between the channel inputs introduced by PPM can only reduce the total mutual information between channel inputs and outputs. These (PPM and on-off keying schemes) are different from the on-off keying scheme used in [2] where the “on” signal has a fixed amplitude that does not depend on  $\mathcal{E}$ . The latter on-off keying scheme, as well as any PPM scheme with a fixed pulse amplitude (with respect to  $\mathcal{E}$ ), 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  $\mathcal{E}$  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  $\mathcal{E}$ ) 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  $\mathcal{E}$  is small enough so that

$$\mathcal{E} \log \frac{1}{\mathcal{E}} < 1 \quad (25a)$$

$$\mathcal{E} < e^{-c}. \quad (25b)$$

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  $x_1, \dots, x_b$  and  $b$  corresponding output symbols  $y_1, \dots, y_b$ . We set

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

- 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”  $\tilde{x}$  that takes value in  $\{1, \dots, b\}$ . Here  $\tilde{x} = i$ ,  $i \in \{1, \dots, b\}$ , means

$$x_i = \eta \quad (27a)$$

$$x_j = 0, \quad j \in \{1, \dots, b\}, j \neq i. \quad (27b)$$

To meet the average-power constraint (3) with equality, we require

$$\eta = b\mathcal{E}. \quad (28)$$

- The  $b$  output symbols  $y_1, \dots, y_b$  are mapped to one “super output symbol”  $\tilde{y}$  that takes value in  $\{1, \dots, b, ?\}$

<sup>5</sup>A scheme between simple and soft-decision PPM is the following. When detecting multiple pulses in a frame, the receiver randomly selects and records one of the positions (possibly together with a “quality” flag). This scheme outperforms simple PPM in photon efficiency, but its third-order term is still linear in  $c$ : it scales like  $-c/2$  instead of  $-c$ .

in the following way:  $\tilde{y} = i$ ,  $i \in \{1, \dots, b\}$ , if  $y_i$  is the unique nonzero term in  $\{y_1, \dots, y_b\}$ ; and  $\tilde{y} = ?$  if there is more than one or no nonzero term in  $\{y_1, \dots, y_b\}$ .

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*

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

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

$$\left\lfloor \frac{1}{\mathcal{E} \log \frac{1}{\mathcal{E}}} \right\rfloor = \frac{1}{\mathcal{E} \log \frac{1}{\mathcal{E}}} + O(1) \quad (30)$$

and simplify (29) to obtain

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

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

■

When proving Proposition 1, as well as Propositions 2 and 3, we frequently use the inequalities summarized in the following lemma.

*Lemma 1: For all  $a \geq 0$ ,*

$$\log(1+a) \leq a, \quad (33)$$

$$1 - e^{-a} \leq a, \quad (34)$$

$$1 - e^{-a} \geq a - \frac{1}{2}a^2. \quad (35)$$

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

$$\tilde{W}(i|i) = \Pr[\tilde{Y} = i | \tilde{X} = i] \quad (36)$$

$$= \Pr[Y_i \geq 1 | X_i = \eta] \prod_{\substack{k=1 \\ k \neq i}}^b \Pr[Y_k = 0 | X_k = 0] \quad (37)$$

$$= (1 - W(0|\eta)) (W(0|0))^{b-1} \quad (38)$$

$$= (1 - e^{-\eta-c\mathcal{E}}) e^{-(b-1)c\mathcal{E}} \quad (39)$$

$$= e^{-(b-1)c\mathcal{E}} - e^{-\eta-bc\mathcal{E}} \quad (40)$$

$$\triangleq p_0, \quad i \in \{1, \dots, b\}; \quad (41)$$

$$\tilde{W}(j|i) = \Pr[\tilde{Y} = j | \tilde{X} = i] \quad (42)$$

$$= \Pr[Y_i = 0 | X_i = \eta] \Pr[Y_j \geq 1 | X_j = 0]$$

$$\cdot \prod_{\substack{k=1 \\ k \notin \{i, j\}}}^b \Pr[Y_k = 0 | X_k = 0] \quad (43)$$

$$= W(0|\eta) (1 - W(0|0)) (W(0|0))^{b-2} \quad (44)$$

$$= e^{-\eta-c\mathcal{E}} (1 - e^{-c\mathcal{E}}) e^{-(b-2)c\mathcal{E}} \quad (45)$$

$$= e^{-\eta-(b-1)c\mathcal{E}} - e^{-\eta-bc\mathcal{E}} \quad (46)$$

$$\triangleq p_1, \quad i, j \in \{1, \dots, b\}, i \neq j; \quad (47)$$

$$\tilde{W}(?|i) = 1 - p_0 - (b-1)p_1, \quad i \in \{1, \dots, b\}. \quad (48)$$

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

$$H(\tilde{Y}|\tilde{X}) = p_0 \log \frac{1}{p_0} + (b-1)p_1 \log \frac{1}{p_1} + (1 - p_0 - (b-1)p_1) \log \frac{1}{1 - p_0 - (b-1)p_1}. \quad (49)$$

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

$$\tilde{C}(\mathcal{E}, c, b, \eta) = \max_{P_{\tilde{X}}} I(\tilde{X}; \tilde{Y}). \quad (50)$$

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  $C_{\text{PE-PPM}}(\mathcal{E}, c)$ :

$$C_{\text{PE-PPM}}(\mathcal{E}, c) \geq \frac{\tilde{C}(\mathcal{E}, c, b, \eta)}{\eta}. \quad (51)$$

It can be easily verified that the optimal input distribution for (50) is the uniform distribution

$$P_{\tilde{X}}(i) = \frac{1}{b}, \quad i \in \{1, \dots, b\}, \quad (52)$$

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

$$P_{\tilde{Y}}(i) = \frac{p_0 + (b-1)p_1}{b}, \quad i \in \{1, \dots, b\}, \quad (53a)$$

$$P_{\tilde{Y}}(?) = 1 - p_0 - (b-1)p_1. \quad (53b)$$

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

$$\tilde{C}(\mathcal{E}, c, b, \eta) = I(\tilde{X}; \tilde{Y}) \quad (54)$$

$$= H(\tilde{Y}) - H(\tilde{Y}|\tilde{X}) \quad (55)$$

$$= (1 - p_0 - (b-1)p_1) \log \frac{1}{1 - p_0 - (b-1)p_1} + (p_0 + (b-1)p_1) \log \frac{1}{p_0 + (b-1)p_1} - (1 - p_0 - (b-1)p_1) \log \frac{1}{1 - p_0 - (b-1)p_1} - p_0 \log \frac{1}{p_0} - (b-1)p_1 \log \frac{1}{p_1} \quad (56)$$

$$= p_0 \log \frac{bp_0}{p_0 + (b-1)p_1} + (b-1)p_1 \log \frac{bp_1}{p_0 + (b-1)p_1} \quad (57)$$

$$= p_0 \log b - p_0 \log \frac{p_0 + (b-1)p_1}{p_0} - (b-1)p_1 \log \frac{p_0 + (b-1)p_1}{bp_1} \quad (58)$$

$$\geq p_0 \log b - p_0 \log \left(1 + \frac{bp_1}{p_0}\right) - \underbrace{\frac{b-1}{b}}_{\leq 1} \underbrace{(p_0 - p_1)}_{\leq p_0} \quad (59)$$

$$\geq p_0 \log b - p_0 \log \left( 1 + \frac{bp_1}{p_0} \right) - p_0. \quad (60)$$

Next note that  $p_0$  can be upper-bounded as

$$p_0 = e^{-(b-1)c\mathcal{E}} - e^{-\eta-bc\mathcal{E}} \quad (61)$$

$$= e^{-(b-1)c\mathcal{E}} (1 - e^{-\eta-c\mathcal{E}}) \quad (62)$$

$$\leq 1 - e^{-\eta-c\mathcal{E}} \quad (63)$$

$$\leq \eta + c\mathcal{E}, \quad (64)$$

where the last step follows from (34). It can also be lower-bounded as

$$p_0 = e^{-(b-1)c\mathcal{E}} - e^{-\eta-bc\mathcal{E}} \quad (65)$$

$$\geq e^{-bc\mathcal{E}} - e^{-\eta-bc\mathcal{E}} \quad (66)$$

$$= \underbrace{e^{-bc\mathcal{E}}}_{\geq 1-bc\mathcal{E}} \underbrace{(1 - e^{-\eta})}_{\geq \eta - \frac{1}{2}\eta^2} \quad (67)$$

$$\geq (1 - bc\mathcal{E}) \left( \eta - \frac{1}{2}\eta^2 \right), \quad (68)$$

where the last inequality follows from (34) and (35), and from the fact that the first multiplicand on its right-hand side is (in fact, both multiplicands are) nonnegative for  $\mathcal{E}$  satisfying (25) and other parameters chosen accordingly. Also note that  $p_1$  can be upper-bounded as

$$p_1 = e^{-\eta-(b-1)c\mathcal{E}} - e^{-\eta-bc\mathcal{E}} \quad (69)$$

$$= \underbrace{e^{-\eta-(b-1)c\mathcal{E}}}_{\leq 1} \underbrace{(1 - e^{-c\mathcal{E}})}_{\leq c\mathcal{E}} \quad (70)$$

$$\leq c\mathcal{E}. \quad (71)$$

Using (64), (68) and (71) we can continue the chain of inequalities (60) to further lower-bound  $\tilde{C}(\mathcal{E}, c, b, \eta)$  as

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

$$\begin{aligned} &\geq (1 - bc\mathcal{E}) \left( \eta - \frac{1}{2}\eta^2 \right) \log b \\ &\quad - (\eta + c\mathcal{E}) \log \left( 1 + \frac{bc\mathcal{E}}{(1 - bc\mathcal{E}) \left( \eta - \frac{1}{2}\eta^2 \right)} \right) - \eta - c\mathcal{E} \quad (72) \end{aligned}$$

$$\begin{aligned} &\geq (1 - bc\mathcal{E}) \left( \eta - \frac{1}{2}\eta^2 \right) \log b - (\eta + c\mathcal{E}) \log \left( 1 + \frac{bc\mathcal{E}}{\eta} \right) \\ &\quad + (\eta + c\mathcal{E}) \left\{ \log(1 - bc\mathcal{E}) + \log \left( 1 - \frac{\eta}{2} \right) \right\} - \eta - c\mathcal{E} \quad (73) \end{aligned}$$

$$\begin{aligned} &\geq \left( \eta - \frac{1}{2}\eta^2 \right) \log b - bc\mathcal{E} \eta \log b - (\eta + c\mathcal{E}) \log \left( 1 + \frac{bc\mathcal{E}}{\eta} \right) \\ &\quad + (\eta + c\mathcal{E}) \left\{ \log(1 - bc\mathcal{E}) + \log \left( 1 - \frac{\eta}{2} \right) \right\} - \eta - c\mathcal{E}. \quad (74) \end{aligned}$$

Here, (73) follows from the fact that, for all  $\alpha \geq 0$  and  $\beta \geq 1$ ,

$$\log(1 + \alpha\beta) \leq \log(\beta + \alpha\beta) = \log \beta + \log(1 + \alpha), \quad (75)$$

with the choices

$$\alpha = \frac{bc\mathcal{E}}{\eta}, \quad (76)$$

$$\beta = \frac{1}{(1 - bc\mathcal{E}) \left( 1 - \frac{1}{2}\eta \right)}, \quad (77)$$

which indeed satisfy  $\alpha \geq 0$  and  $\beta \geq 1$ ; and (74) follows by dropping the term  $\frac{1}{2}bc\mathcal{E}\eta^2 \log b$ , which is positive.

Combining (28), (51), and (74) 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  $\{1, \dots, b, ?\} \cup \{\{i, j\} \subset \{1, \dots, b\}\}$ . The mapping rule is as follows. Take  $\hat{y} = i$  if  $y_i$  is the unique nonzero term in  $\{y_1, \dots, y_b\}$ ; take  $\hat{y} = \{i, j\}$  if  $y_i$  and  $y_j$  are the only two nonzero terms in  $\{y_1, \dots, y_b\}$ ; if there are more than two or no nonzero term in  $\{y_1, \dots, y_b\}$ , take  $\hat{y} = ?$ .*

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

$$C_{\text{PE-PPM(SD)}}(\mathcal{E}, c)$$

$$\begin{aligned} &\geq \left( 1 - \frac{\eta}{2} \right) \log b - c\eta \log b - \left( 1 + \frac{c\mathcal{E}}{\eta} \right) \log(1 + c) \\ &\quad + \left( 1 + \frac{c\mathcal{E}}{\eta} \right) \left\{ \log(1 - c\eta) + \log \left( 1 - \frac{\eta}{2} \right) \right\} - 1 - \frac{c\mathcal{E}}{\eta} \\ &\quad + \left( 1 - \frac{\eta}{2} \right) (b - 1) \left( c\mathcal{E} - \frac{c^2\mathcal{E}^2}{2} \right) (1 - c\eta) \log \frac{b}{2} \\ &\quad - \frac{c^2}{2}\eta - c(\eta + c\mathcal{E}) \quad (78) \end{aligned}$$

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

*Proof of (19) Using Proposition 2:* Simplifying (78) we obtain

$$\begin{aligned} &\lim_{\mathcal{E} \downarrow 0} \left\{ C_{\text{PE-PPM(SD)}}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}} \right\} \\ &\geq -\log(1 + c) - \frac{3}{2}. \quad (79) \end{aligned}$$

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

$$\hat{W}(i|i) = p_0, \quad (80)$$

$$\hat{W}(j|i) = p_1, \quad i, j \in \{1, \dots, b\}, i \neq j, \quad (81)$$

where  $p_0$  and  $p_1$  are given in (41) and (47), respectively. For the remaining elements of the transition matrix we have

$$\begin{aligned} \hat{W}(\{i, j\}|i) &= \Pr[Y_i \geq 1|X_i = \eta] \Pr[Y_j \geq 1|X_j = 0] \\ &\quad \cdot \prod_{\substack{\ell=1 \\ \ell \notin \{i, j\}}}^b \Pr[Y_\ell = 0|X_\ell = 0] \quad (82) \end{aligned}$$

$$= (1 - e^{-\eta-c\mathcal{E}})(1 - e^{-c\mathcal{E}}) e^{-(b-2)c\mathcal{E}} \quad (83)$$

$$\triangleq p_2, \quad \{i, j\} \subset \{1, \dots, b\}; \quad (84)$$

$$\begin{aligned} \hat{W}(\{j, k\}|i) &= \Pr[Y_i = 0|X_i = \eta] \Pr[Y_j \geq 1|X_j = 0] \\ &\quad \cdot \Pr[Y_k \geq 1|X_k = 0] \end{aligned}$$

$$\cdot \prod_{\substack{\ell=1 \\ \ell \notin \{i, j, k\}}}^b \Pr[Y_\ell = 0|X_\ell = 0] \quad (85)$$

$$= e^{-\eta-c\mathcal{E}} (1 - e^{-c\mathcal{E}})^2 e^{-(b-3)c\mathcal{E}} \quad (86)$$

$$\triangleq p_3, \quad \{i, j, k\} \subset \{1, \dots, b\}; \quad (87)$$

$$\hat{W}(?|i) = 1 - p_0 - (b - 1)p_1 - (b - 1)p_2$$

$$- \frac{(b-1)(b-2)}{2} p_3 \quad (88)$$

$$\triangleq p_4, \quad i \in \{1, \dots, b\}. \quad (89)$$

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

$$\begin{aligned} H(\hat{Y}|\tilde{X}) &= p_0 \log \frac{1}{p_0} + (b-1)p_1 \log \frac{1}{p_1} \\ &+ (b-1)p_2 \log \frac{1}{p_2} + \frac{(b-1)(b-2)}{2} p_3 \log \frac{1}{p_3} \\ &+ p_4 \log \frac{1}{p_4}. \end{aligned} \quad (90)$$

Choosing a uniform  $\tilde{X}$  (which is optimal as in Section IV), we have the following distribution on  $\hat{Y}$

$$P_{\hat{Y}}(i) = \frac{p_0 + (b-1)p_1}{b}, \quad i \in \{1, \dots, b\}; \quad (91)$$

$$P_{\hat{Y}}(\{i, j\}) = \frac{2p_2 + (b-2)p_3}{b}, \quad \{i, j\} \subset \{1, \dots, b\}; \quad (92)$$

$$P_{\hat{Y}}(?) = p_4. \quad (93)$$

Therefore

$$\begin{aligned} H(\hat{Y}) &= (p_0 + (b-1)p_1) \log \frac{b}{p_0 + (b-1)p_1} \\ &+ \left( (b-1)p_2 + \frac{(b-1)(b-2)}{2} p_3 \right) \log \frac{b}{2p_2 + (b-2)p_3} \\ &+ p_4 \log \frac{1}{p_4}. \end{aligned} \quad (94)$$

Using (90) and (94) we have

$$\begin{aligned} I(\tilde{X}; \hat{Y}) &= H(\hat{Y}) - H(\hat{Y}|\tilde{X}) \\ &= p_0 \log \frac{bp_0}{p_0 + (b-1)p_1} \\ &+ (b-1)p_1 \log \frac{bp_1}{p_0 + (b-1)p_1} \\ &+ (b-1)p_2 \log \frac{bp_2}{2p_2 + (b-2)p_3} \\ &+ \frac{(b-1)(b-2)}{2} p_3 \log \frac{bp_3}{2p_2 + (b-2)p_3}. \end{aligned} \quad (95)$$

At this point, note that the first two summands on the right-hand side of (96) constitute  $I(\tilde{X}; \hat{Y})$ , 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 (96) as

$$\begin{aligned} (b-1)p_2 \log \frac{bp_2}{2p_2 + (b-2)p_3} &= (b-1)p_2 \log \frac{b}{2} - (b-1)p_2 \underbrace{\log \frac{p_2 + \frac{b-2}{2} p_3}{p_2}}_{=\log\left(1 + \frac{(b-2)p_3}{2p_2}\right) \leq \frac{(b-2)p_3}{2p_2}} \quad (97) \\ &\geq (b-1)p_2 \log \frac{b}{2} - \frac{(b-1)(b-2)}{2} p_3 \quad (98) \\ &\geq (b-1)p_2 \log \frac{b}{2} - \frac{b^2}{2} p_3. \quad (99) \end{aligned}$$

We lower-bound the fourth term on the right-hand side of (96) as

$$\begin{aligned} \frac{(b-1)(b-2)}{2} p_3 \log \frac{bp_3}{2p_2 + (b-2)p_3} &= - \frac{(b-1)(b-2)}{2} p_3 \underbrace{\log \frac{2p_2 + (b-2)p_3}{bp_3}}_{=\log\left(1 + \frac{2(p_2 - p_3)}{bp_3}\right) \leq \frac{2(p_2 - p_3)}{bp_3} \leq \frac{2p_2}{bp_3}} \quad (100) \\ &\geq - \frac{(b-1)(b-2)}{2} \cdot \frac{2p_2}{b} \quad (101) \\ &\geq -bp_2. \quad (102) \end{aligned}$$

Using (34) and (35) we have the upper bound on  $p_2$

$$p_2 = \underbrace{(1 - e^{-\eta - c\mathcal{E}})}_{\leq \eta + c\mathcal{E}} \underbrace{(1 - e^{-c\mathcal{E}})}_{\leq c\mathcal{E}} \underbrace{e^{-(b-2)c\mathcal{E}}}_{\leq 1} \quad (103)$$

$$\leq (\eta + c\mathcal{E})c\mathcal{E}, \quad (104)$$

the lower bound on  $p_2$

$$p_2 = \underbrace{(1 - e^{-\eta - c\mathcal{E}})}_{\geq 1 - e^{-\eta} \geq \eta - \frac{\eta^2}{2}} \underbrace{(1 - e^{-c\mathcal{E}})}_{\geq c\mathcal{E} - \frac{c^2\mathcal{E}^2}{2}} \underbrace{e^{-(b-2)c\mathcal{E}}}_{\geq 1 - bc\mathcal{E}} \quad (105)$$

$$\geq \left( \eta - \frac{\eta^2}{2} \right) \left( c\mathcal{E} - \frac{c^2\mathcal{E}^2}{2} \right) (1 - bc\mathcal{E}), \quad (106)$$

and the upper bound on  $p_3$

$$p_3 = \underbrace{e^{-\eta - c\mathcal{E}}}_{\leq 1} \underbrace{(1 - e^{-c\mathcal{E}})^2}_{\leq c\mathcal{E}} \underbrace{e^{-(b-3)c\mathcal{E}}}_{\leq 1} \quad (107)$$

$$\leq c^2\mathcal{E}^2. \quad (108)$$

From (96), (99), (102), (104), (106), and (108) we obtain a lower bound on the additional mutual information that is gained by considering output frames with two detection positions:

$$\begin{aligned} I(\tilde{X}; \hat{Y}) - I(\tilde{X}; \tilde{Y}) &\geq (b-1)p_2 \log \frac{b}{2} - \frac{b^2}{2} p_3 - bp_2 \quad (109) \\ &\geq (b-1) \left( \eta - \frac{\eta^2}{2} \right) \left( c\mathcal{E} - \frac{c^2\mathcal{E}^2}{2} \right) (1 - bc\mathcal{E}) \log \frac{b}{2} \\ &\quad - \frac{b^2}{2} c^2\mathcal{E}^2 - b(\eta + c\mathcal{E})c\mathcal{E}. \quad (110) \end{aligned}$$

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

$$\begin{aligned} \frac{I(\tilde{X}; \hat{Y}) - I(\tilde{X}; \tilde{Y})}{\eta} &\geq \left(1 - \frac{\eta}{2}\right) (b-1) \left( c\mathcal{E} - \frac{c^2\mathcal{E}^2}{2} \right) (1 - c\eta) \log \frac{b}{2} \\ &\quad - \frac{c^2}{2} \eta - c(\eta + c\mathcal{E}). \quad (111) \end{aligned}$$

Adding the right-hand side of (111) to the right-hand side of (29) yields (78).  $\blacksquare$

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

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

$$\mathcal{E} < e^{-1} \quad (112a)$$

$$\mathcal{E} \log \frac{1}{\mathcal{E}} < \frac{e^{-\frac{1}{1-2e^{-1}}}}{1+c} \quad (112b)$$

$$\mathcal{E} \left( \log \frac{1}{\mathcal{E}} \right)^4 < \frac{144}{(1+c)^2}, \quad (112c)$$

then

$$\begin{aligned} C_{\text{PE}}(\mathcal{E}, c) &\leq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} - \log(1+c) + 2 + \log 13 \\ &\quad + \frac{1}{\mathcal{E}} \left( \log \frac{1}{1-(1+c)\mathcal{E}} - (1+c)\mathcal{E} \right) \\ &\quad + \frac{c^2}{2} \mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1+c) \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}}. \end{aligned} \quad (113)$$

*Proof of (20) Using Proposition 3:* The last three summands on the right-hand side of (113) are all of the form  $o(1)$ . Setting  $c = 0$  in (113) we have

$$C_{\text{PE}}(\mathcal{E}, 0) \leq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + 2 + \log 13 + o(1) \quad (114)$$

$$= \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1). \quad (115)$$

*Proof of (21) Using Proposition 3:* From (113) we have

$$\begin{aligned} \overline{\lim}_{\mathcal{E} \downarrow 0} \left\{ C_{\text{PE}}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}} \right\} \\ \leq -\log(1+c) + 2 + \log 13. \end{aligned} \quad (116)$$

Hence

$$\overline{\lim}_{c \rightarrow \infty} \overline{\lim}_{\mathcal{E} \downarrow 0} \frac{C_{\text{PE}}(\mathcal{E}, c) - \log \frac{1}{\mathcal{E}} + \log \log \frac{1}{\mathcal{E}}}{\log c} \quad (117)$$

$$\leq \lim_{c \rightarrow \infty} \frac{-\log(1+c) + 2 + \log 13}{\log c} \quad (118)$$

$$= -1. \quad (118)$$

*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

$$C \leq \sup \mathbb{E} \left[ D(W(\cdot|X) \| R(\cdot)) \right], \quad (119)$$

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

$$R(y) = \begin{cases} 1 - (1+c)\mathcal{E}, & y = 0 \\ (1+c)\mathcal{E} \left( 1 - \frac{1}{\log \frac{1}{\mathcal{E}}} \right) \left( \frac{1}{\log \frac{1}{\mathcal{E}}} \right)^{y-1}, & y \geq 1. \end{cases} \quad (120)$$

For any  $x \geq 0$  we have

$$\begin{aligned} D(W(\cdot|x) \| R(\cdot)) \\ = \sum_{y=0}^{\infty} \text{Pois}_{x+c\mathcal{E}}(y) \cdot \log \frac{1}{R(y)} - H(Y|X=x) \end{aligned} \quad (121)$$

$$\begin{aligned} = e^{-(x+c\mathcal{E})} \log \frac{1}{1 - (1+c)\mathcal{E}} \\ - \sum_{y=1}^{\infty} \text{Pois}_{x+c\mathcal{E}}(y) \\ \cdot \log \left\{ (1+c)\mathcal{E} \left( 1 - \frac{1}{\log \frac{1}{\mathcal{E}}} \right) \left( \frac{1}{\log \frac{1}{\mathcal{E}}} \right)^{y-1} \right\} \\ - H(Y|X=x) \end{aligned} \quad (122)$$

$$\begin{aligned} = e^{-(x+c\mathcal{E})} \log \frac{1}{1 - (1+c)\mathcal{E}} + \underbrace{\sum_{y=1}^{\infty} \text{Pois}_{x+c\mathcal{E}}(y) y \log \log \frac{1}{\mathcal{E}}}_{= \mathbb{E}[Y|X=x]=x+c\mathcal{E}} \\ + \underbrace{\left( \sum_{y=1}^{\infty} \text{Pois}_{x+c\mathcal{E}}(y) \right)}_{= 1 - e^{-(x+c\mathcal{E})}} \end{aligned}$$

$$\begin{aligned} \cdot \left( \log \frac{1}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}} \right) \\ - H(Y|X=x) \end{aligned} \quad (123)$$

$$\begin{aligned} = e^{-(x+c\mathcal{E})} \log \frac{1}{1 - (1+c)\mathcal{E}} + (x+c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} \\ + (1 - e^{-(x+c\mathcal{E})}) \\ \cdot \left( \log \frac{1}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}} \right) \\ - H(Y|X=x). \end{aligned} \quad (124)$$

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

$$H(Y|X=x) \geq H_b(e^{-(x+c\mathcal{E})}) \quad (125)$$

$$\begin{aligned} = e^{-(x+c\mathcal{E})} (x+c\mathcal{E}) \\ + (1 - e^{-(x+c\mathcal{E})}) \log \frac{1}{1 - e^{-(x+c\mathcal{E})}}. \end{aligned} \quad (126)$$

Using this and (124), we upper-bound  $D(W(\cdot|x) \| R(\cdot))$  as

$$\begin{aligned} D(W(\cdot|x) \| R(\cdot)) \\ \leq e^{-(x+c\mathcal{E})} \log \frac{1}{1 - (1+c)\mathcal{E}} + (x+c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} \\ + (1 - e^{-(x+c\mathcal{E})}) \\ \cdot \left( \log \frac{1}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}} \right) \\ - e^{-(x+c\mathcal{E})} (x+c\mathcal{E}) - (1 - e^{-(x+c\mathcal{E})}) \log \frac{1}{1 - e^{-(x+c\mathcal{E})}} \end{aligned} \quad (127)$$



$$\begin{aligned}
&= e^{-(x+c\mathcal{E})} \log \frac{1}{1-(1+c)\mathcal{E}} + (x+c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} \\
&\quad + \underbrace{(1-e^{-(x+c\mathcal{E})})}_{\leq x+c\mathcal{E}} \log \underbrace{\frac{1}{1-\frac{1}{\log \frac{1}{\mathcal{E}}}}}_{\geq 0} - e^{-(x+c\mathcal{E})} \underbrace{(x+c\mathcal{E})}_{\geq c\mathcal{E}} \\
&\quad + (1-e^{-(x+c\mathcal{E})}) \log \frac{1-e^{-(x+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \tag{128}
\end{aligned}$$

$$\begin{aligned}
&\leq \underbrace{e^{-(x+c\mathcal{E})}}_{\leq 1} \underbrace{\left( \log \frac{1}{1-(1+c)\mathcal{E}} - c\mathcal{E} \right)}_{\geq (1+c)\mathcal{E} - c\mathcal{E} \geq 0} \\
&\quad + (x+c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} + (x+c\mathcal{E}) \log \frac{1}{1-\frac{1}{\log \frac{1}{\mathcal{E}}}} \\
&\quad + \underbrace{(1-e^{-(x+c\mathcal{E})})}_{=(1-e^{-c\mathcal{E}})+(e^{-c\mathcal{E}}-e^{-(x+c\mathcal{E})})} \log \frac{1-e^{-(x+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \tag{129}
\end{aligned}$$

$$\begin{aligned}
&\leq \left( \log \frac{1}{1-(1+c)\mathcal{E}} - c\mathcal{E} \right) \\
&\quad + (x+c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} + (x+c\mathcal{E}) \log \frac{1}{1-\frac{1}{\log \frac{1}{\mathcal{E}}}} \\
&\quad + (e^{-c\mathcal{E}} - e^{-(x+c\mathcal{E})}) \log \frac{1-e^{-(x+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \\
&\quad + (1-e^{-c\mathcal{E}}) \log \frac{1-e^{-(x+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \tag{130}
\end{aligned}$$

$$\begin{aligned}
&= \mathcal{E} + \left( \log \frac{1}{1-(1+c)\mathcal{E}} - (1+c)\mathcal{E} \right) \\
&\quad + (x+c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} + (x+c\mathcal{E}) \log \frac{1}{1-\frac{1}{\log \frac{1}{\mathcal{E}}}} \\
&\quad + e^{-c\mathcal{E}}(1-e^{-x}) \log \frac{1-e^{-(x+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \\
&\quad + \underbrace{(1-e^{-c\mathcal{E}})}_{\geq c\mathcal{E} - \frac{c^2}{2}\mathcal{E}^2} \log \frac{c}{(1+c) \log \frac{1}{\mathcal{E}}} \\
&\quad \quad \quad \leq -\log \log \frac{1}{\mathcal{E}} \\
&\quad + \underbrace{(1-e^{-c\mathcal{E}})}_{\leq c\mathcal{E}} \log \frac{1-e^{-(x+c\mathcal{E})}}{c\mathcal{E}} \\
&\quad \quad \quad \leq \log \frac{x+c\mathcal{E}}{c\mathcal{E}} \leq \frac{x}{c\mathcal{E}}
\end{aligned}$$

$$\begin{aligned}
&\leq \mathcal{E} + \left( \log \frac{1}{1-(1+c)\mathcal{E}} - (1+c)\mathcal{E} \right) \\
&\quad + (x+c\mathcal{E}) \log \log \frac{1}{\mathcal{E}} + (x+c\mathcal{E}) \log \frac{1}{1-\frac{1}{\log \frac{1}{\mathcal{E}}}} \\
&\quad + e^{-c\mathcal{E}}(1-e^{-x}) \log \frac{1-e^{-(x+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \\
&\quad - \left( c\mathcal{E} - \frac{c^2}{2}\mathcal{E}^2 \right) \log \log \frac{1}{\mathcal{E}} + x. \tag{132}
\end{aligned}$$

Using (119) and (132) together with constraint (3) we have

$$\begin{aligned}
C(\mathcal{E}, c) &\leq \sup \mathbf{E} [D(W(\cdot|X)\|R(\cdot))] \tag{133} \\
&\leq \mathcal{E} + \left( \log \frac{1}{1-(1+c)\mathcal{E}} - (1+c)\mathcal{E} \right)
\end{aligned}$$

$$\begin{aligned}
&\quad + (1+c)\mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1+c)\mathcal{E} \log \frac{1}{1-\frac{1}{\log \frac{1}{\mathcal{E}}}} \\
&\quad - \left( c\mathcal{E} - \frac{c^2}{2}\mathcal{E}^2 \right) \log \log \frac{1}{\mathcal{E}} + \mathcal{E} \\
&\quad + e^{-c\mathcal{E}} \sup \mathbf{E} \left[ (1-e^{-X}) \log \frac{1-e^{-(X+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right] \tag{134}
\end{aligned}$$

$$\begin{aligned}
&= \mathcal{E} \log \log \frac{1}{\mathcal{E}} + 2\mathcal{E} \\
&\quad + \left( \log \frac{1}{1-(1+c)\mathcal{E}} - (1+c)\mathcal{E} \right) \\
&\quad + \frac{c^2}{2}\mathcal{E}^2 \log \log \frac{1}{\mathcal{E}} + (1+c)\mathcal{E} \log \frac{1}{1-\frac{1}{\log \frac{1}{\mathcal{E}}}} \\
&\quad + e^{-c\mathcal{E}} \sup \mathbf{E} \left[ (1-e^{-X}) \log \frac{1-e^{-(X+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right]. \tag{135}
\end{aligned}$$

Hence the maximum achievable photon efficiency satisfies

$$\begin{aligned}
C_{\text{PE}}(\mathcal{E}, c) &= \frac{C(\mathcal{E}, c)}{\mathcal{E}} \tag{136}
\end{aligned}$$

$$\begin{aligned}
&\leq \log \log \frac{1}{\mathcal{E}} + 2 + \frac{1}{\mathcal{E}} \left( \log \frac{1}{1-(1+c)\mathcal{E}} - (1+c)\mathcal{E} \right) \\
&\quad + \frac{c^2}{2}\mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1+c) \log \frac{1}{1-\frac{1}{\log \frac{1}{\mathcal{E}}}} \\
&\quad + \frac{e^{-c\mathcal{E}}}{\mathcal{E}} \sup \mathbf{E} \left[ (1-e^{-X}) \log \frac{1-e^{-(X+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right]. \tag{137}
\end{aligned}$$

We next upper-bound the expectation on the right-hand side of (137) as follows

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

$$= \mathbf{E} \left[ X \cdot \frac{1-e^{-X}}{X} \log \frac{1-e^{-(X+c\mathcal{E})}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right] \tag{138}$$

$$\leq \mathbf{E} \left[ X \cdot \frac{1-e^{-X}}{X} \log \frac{X+c\mathcal{E}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right] \tag{139}$$

$$\leq \mathcal{E} \cdot \sup \phi(x) \tag{140}$$

where

$$\phi(x) \triangleq \frac{1-e^{-x}}{x} \log \frac{x+c\mathcal{E}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}}, \quad x \geq 0. \tag{141}$$

We claim that, if  $\mathcal{E}$  is small enough to satisfy all of the conditions in (112), then

$$\sup \phi(x) = \phi(x^*) \quad \text{for some } x^* \leq \frac{12}{\log \frac{1}{\mathcal{E}}}. \tag{142}$$

To show this, we compute the derivative of  $\phi(x)$  (with respect to  $x$ ) as

$$\frac{d\phi(x)}{dx} = \frac{(1+x)e^{-x} - 1}{x^2} \log \frac{x+c\mathcal{E}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \frac{1-e^{-x}}{x(x+c\mathcal{E})} \quad (143)$$

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

$$\frac{d\phi(x)}{dx} = 0. \quad (144)$$

To prove (142), it thus suffices to show that, under the assumptions (112),

$$\frac{d\phi(x)}{dx} < 0, \quad x > \frac{12}{\log \frac{1}{\mathcal{E}}}. \quad (145)$$

We do this separately for two cases.

*Case 1: Consider*

$$\frac{12}{\log \frac{1}{\mathcal{E}}} < x < 1. \quad (146)$$

*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:

$$\log \frac{x+c\mathcal{E}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \geq \log \frac{x}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \quad (147)$$

$$> \log \frac{12}{(1+c)\mathcal{E} (\log \frac{1}{\mathcal{E}})^2} \quad (148)$$

$$> 0, \quad (149)$$

where the last step follows because (112a) and (112c) together imply

$$\mathcal{E} \left( \log \frac{1}{\mathcal{E}} \right)^2 < \frac{12e^{-\frac{1}{2}}}{1+c}. \quad (150)$$

Next, using Taylor's theorem with Cauchy remainder, we have

$$\frac{(1+x)e^{-x} - 1}{x^2} = -\frac{1}{2} + \frac{1}{3}x - \underbrace{\frac{(3-x')e^{-x'}}{24}x^2}_{\geq 0} \quad (151)$$

$$\leq -\frac{1}{2} + \frac{1}{3} \underbrace{x}_{\leq 1} \quad (152)$$

$$\leq -\frac{1}{6}, \quad (153)$$

where  $x' \in [0, x]$ . The underbrace in (151) follows because  $x' \leq x < 1$ . We also have

$$\frac{1-e^{-x}}{x(x+c\mathcal{E})} \leq \frac{x}{x(x+c\mathcal{E})} \leq \frac{1}{x}. \quad (154)$$

Plugging (149), (153) and (154) into (143) we have

$$\frac{d\phi(x)}{dx} \leq -\frac{1}{6} \log \frac{x+c\mathcal{E}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \frac{1}{x} \quad (155)$$

$$\leq -\frac{1}{6} \underbrace{\log x}_{\geq \log 12 - \log \log \frac{1}{\mathcal{E}}} + \frac{1}{6} \log(1+c) - \frac{1}{6} \log \frac{1}{\mathcal{E}} + \frac{1}{6} \log \log \frac{1}{\mathcal{E}} + \underbrace{\frac{1}{x}}_{\leq \frac{1}{12} \log \frac{1}{\mathcal{E}}} \quad (156)$$

$$\leq \frac{1}{6} \log \frac{1+c}{12} + \frac{1}{3} \log \log \frac{1}{\mathcal{E}} - \frac{1}{12} \log \frac{1}{\mathcal{E}} \quad (157)$$

$$= \frac{1}{12} \log \left( \frac{(1+c)^2}{144} \mathcal{E} \left( \log \frac{1}{\mathcal{E}} \right)^4 \right) \quad (158)$$

$$< 0, \quad (159)$$

where the last step follows from (112c). We have now shown

$$\frac{d\phi(x)}{dx} < 0, \quad \frac{12}{\log \frac{1}{\mathcal{E}}} < x < 1. \quad (160)$$

*Case 2: Consider*

$$x \geq 1. \quad (161)$$

In this case the logarithmic term in the derivative is still positive:

$$\log \frac{x+c\mathcal{E}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \geq \log \frac{1}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} > 0, \quad (162)$$

where the last step follows from (112b). We bound the other terms as

$$\frac{(1+x)e^{-x} - 1}{x^2} \leq -\frac{1-2e^{-1}}{x^2} \quad (163)$$

which is negative, and

$$\frac{1-e^{-x}}{x(x+c\mathcal{E})} \leq \frac{1}{x^2}. \quad (164)$$

Using (162), (163) and (164) together with (143) we have

$$\frac{d\phi(x)}{dx} \leq -\frac{1-2e^{-1}}{x^2} \left( \underbrace{\log(x+c\mathcal{E})}_{\geq 0} + \log \frac{1}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right) + \frac{1}{x^2} \quad (165)$$

$$\leq -\frac{1-2e^{-1}}{x^2} \log \frac{1}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} + \frac{1-2e^{-1}}{x^2} \cdot \frac{1}{1-2e^{-1}} \quad (166)$$

$$= -\frac{1-2e^{-1}}{x^2} \log \frac{e^{-\frac{1}{1-2e^{-1}}}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \quad (167)$$

$$< 0, \quad (168)$$

where the last step follows from (112b). Hence we have shown

$$\frac{d\phi(x)}{dx} < 0, \quad x \geq 1. \quad (169)$$

Combining (160) and (169) proves (142) for all  $\mathcal{E}$  satisfying (112).

We now proceed to upper-bound  $\sup \phi(x)$ :

$$\sup \phi(x) = \phi(x^*) \quad (170)$$

$$= \frac{1 - e^{-x^*}}{x^*} \log \frac{x^* + c\mathcal{E}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \quad (171)$$

$$\leq \max \left\{ 0, \log \frac{x^* + c\mathcal{E}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right\} \quad (172)$$

$$\leq \max \left\{ 0, \log \frac{\frac{12}{\log \frac{1}{\mathcal{E}}} + c\mathcal{E}}{(1+c)\mathcal{E} \log \frac{1}{\mathcal{E}}} \right\}. \quad (173)$$

Now note that, due to (112b), we have

$$c\mathcal{E} \leq \frac{1}{\log \frac{1}{\mathcal{E}}}. \quad (174)$$

We can now continue the chain of inequalities (173) as

$$\sup \phi(x) \leq \max \left\{ 0, \log \frac{13}{(1+c)\mathcal{E} (\log \frac{1}{\mathcal{E}})^2} \right\} \quad (175)$$

$$= \log \frac{13}{(1+c)\mathcal{E} (\log \frac{1}{\mathcal{E}})^2} \quad (176)$$

$$= \log \frac{1}{\mathcal{E}} - 2 \log \log \frac{1}{\mathcal{E}} - \log(1+c) + \log 13. \quad (177)$$

Here (176) follows because conditions (112a) and (112c) imply (150).

Combining (137), (140) and (177), and noting  $e^{-c\mathcal{E}} \leq 1$  (together with the fact that the right-hand side of (177) is positive), we obtain

$$\begin{aligned} C_{\text{PE}}(\mathcal{E}, c) &\leq \log \log \frac{1}{\mathcal{E}} + 2 + \frac{1}{\mathcal{E}} \left( \log \frac{1}{1 - (1+c)\mathcal{E}} - (1+c)\mathcal{E} \right) \\ &\quad + \frac{c^2}{2} \mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1+c) \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}} \\ &\quad + \log \frac{1}{\mathcal{E}} - 2 \log \log \frac{1}{\mathcal{E}} - \log(1+c) + \log 13 \end{aligned} \quad (178)$$

$$\begin{aligned} &= \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} - \log(1+c) + 2 + \log 13 \\ &\quad + \frac{1}{\mathcal{E}} \left( \log \frac{1}{1 - (1+c)\mathcal{E}} - (1+c)\mathcal{E} \right) \\ &\quad + \frac{c^2}{2} \mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1+c) \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}}. \end{aligned} \quad (179)$$

## 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 (137) and (140), and is given by

$$\begin{aligned} C_{\text{PE}}(\mathcal{E}, c) &\leq \log \log \frac{1}{\mathcal{E}} + 2 + \frac{1}{\mathcal{E}} \left( \log \frac{1}{1 - (1+c)\mathcal{E}} - (1+c)\mathcal{E} \right) \\ &\quad + \frac{c^2}{2} \mathcal{E} \log \log \frac{1}{\mathcal{E}} + (1+c) \log \frac{1}{1 - \frac{1}{\log \frac{1}{\mathcal{E}}}} \\ &\quad + e^{-c\mathcal{E}} \sup \phi(x), \end{aligned} \quad (180)$$

where  $\phi(\cdot)$  is given in (141) and  $\sup \phi(x)$  is computed numerically. The on-off-keying lower bound is obtained by computing the mutual information of the channel with ‘‘on’’ signal equaling  $\eta$ , and with the receiver ignoring multiple detected photons. This is given by

$$\begin{aligned} C_{\text{PE-OOK}}(\mathcal{E}, c) &\geq \frac{1}{\mathcal{E}} \left( H_b(q) - \frac{\mathcal{E}}{\eta} \cdot H_b(e^{-\eta-c\mathcal{E}}) - \left( 1 - \frac{\mathcal{E}}{\eta} \right) H_b(e^{-c\mathcal{E}}) \right), \end{aligned} \quad (181)$$

where  $H_b(\cdot)$  denotes the binary entropy function

$$H_b(a) = a \log \frac{1}{a} + (1-a) \log \frac{1}{1-a}, \quad a \in [0, 1], \quad (182)$$

and where

$$q \triangleq \frac{\mathcal{E}}{\eta} e^{-\eta-c\mathcal{E}} + \left( 1 - \frac{\mathcal{E}}{\eta} \right) e^{-c\mathcal{E}}. \quad (183)$$

The simple-PPM lower bound is computed using (51) and (57):

$$\begin{aligned} C_{\text{PE-PPM}}(\mathcal{E}, c) &\geq \frac{1}{\eta} \left( p_0 \log \frac{bp_0}{p_0 + (b-1)p_1} \right. \\ &\quad \left. + (b-1)p_1 \log \frac{bp_1}{p_0 + (b-1)p_1} \right). \end{aligned} \quad (184)$$

The soft-decision-PPM lower bound is computed using (96) and is given by

$$\begin{aligned} C_{\text{PE-PPM(SD)}}(\mathcal{E}, c) &\geq \frac{1}{\eta} \left( p_0 \log \frac{bp_0}{p_0 + (b-1)p_1} \right. \\ &\quad + (b-1)p_1 \log \frac{bp_1}{p_0 + (b-1)p_1} \\ &\quad + (b-1)p_2 \log \frac{bp_2}{2p_2 + (b-2)p_3} \\ &\quad \left. + \frac{(b-1)(b-2)}{2} p_3 \log \frac{bp_3}{2p_2 + (b-2)p_3} \right). \end{aligned} \quad (185)$$

In all these lower bounds, the parameters  $b$  and  $\eta$  are chosen as in (26) and (28), and  $p_0$ ,  $p_1$ ,  $p_2$ ,  $p_3$ , and  $p_4$  are given in (41), (47), (84), (87), and (89), respectively. We plot these bounds for  $c = 0.1$ ,  $c = 1$ , and  $c = 10$  in Figure 1.

The figure shows 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).

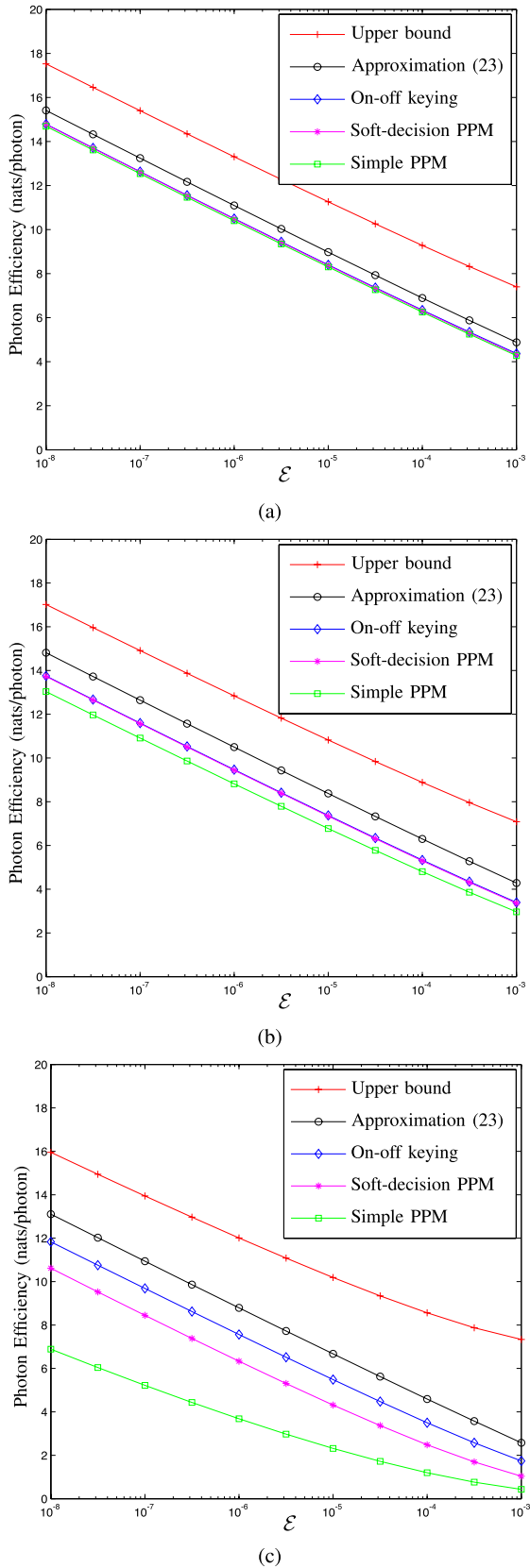


Fig. 1. Comparing the approximation (23) to nonasymptotic upper bound, and to lower bounds for simple PPM, soft-decision PPM, and on-off keying for three cases:  $c = 0.1$ ,  $c = 1$ , and  $c = 10$ . (a) Case  $c = 0.1$ . Note that the lowest three curves almost overlap. (b) Case  $c = 1$ . The on-off keying and soft-decision PPM curves overlap. (c) Case  $c = 10$ .

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

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 provide a proof for the converse part of (8) that is much simpler than the one given in [2]. Using the same arguments as in [2], we know that  $C(\mathcal{E}, c)$  is monotonically increasing in  $c$ . Hence it suffices to show

$$\lim_{\epsilon \downarrow 0} \frac{C(\mathcal{E}, 0)}{\epsilon \log \frac{1}{\epsilon}} \leq 1. \quad (186)$$

This can be verified as follows:

$$C(\mathcal{E}, 0) = \max_{\mathbb{E}[X] \leq \mathcal{E}} I(X; Y) \quad (187)$$

$$\leq \max_{\mathbb{E}[X] \leq \mathcal{E}} H(Y) \quad (188)$$

$$= \max_{\mathbb{E}[Y] \leq \mathcal{E}} H(Y) \quad (189)$$

$$= (1 + \mathcal{E}) \log(1 + \mathcal{E}) - \mathcal{E} \log \mathcal{E}, \quad (190)$$

which yields the desired asymptotic bound. Here, (190) 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 (190). Note that the right-hand side of (190) is exactly the capacity of the pure-loss bosonic channel [7].

#### ACKNOWLEDGMENT

The authors would like to thank Hye Won Chung, Yuval Kochman, Amos Lapidoth, Arya Mazumdar, and Lizhong Zheng for helpful discussions, and the anonymous reviewers for their careful reading of the manuscript.

## REFERENCES

- [1] J. H. Shapiro, "The quantum theory of optical communications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 15, no. 6, pp. 1547–1569, Nov./Dec. 2009.
- [2] A. Lapidoth, J. H. Shapiro, V. Venkatesan, and L. Wang, "The discrete-time Poisson channel at low input powers," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3260–3272, Jun. 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 a 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," *Proc. Inst. Elect. Eng. I, Commun., Speech Vis.*, vol. 137, no. 6, pp. 424–430, Dec. 1990.
- [6] A. Lapidoth and S. M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Trans. Inf. 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, pp. 027902-1–027902-4, 2004.
- [8] R. G. Gallager, "Energy limited channels: Coding, multiaccess, and spread spectrum," Laboratory for Information and Decision Systems, MIT, Cambridge, MA, USA, Tech. Rep. 1714, Nov. 1987.
- [9] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.
- [10] H. W. Chung, S. Guha, and L. Zheng, "On capacity of optical channels with coherent detection," in *Proc. 2011 49th Annu. Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, USA, Sep. 2011, pp. 879–885.
- [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. 2012, pp. 172–179.
- [12] S. Dolinar, K. M. Birnbaum, B. M. Erkmen, and B. Moision, "On approaching the ultimate limits of photon-efficient and bandwidth-efficient optical communication," in *Proc. Int. Conf. Space Opt. Syst. and Appl.*, Santa Monica, CA, USA, May 2011, pp. 269–278.
- [13] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. 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," *IEEE Trans. Inf. Theory*, submitted for publication.
- [15] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.
- [16] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

**Ligong Wang** (S'08–M'12) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2004, and the M.Sc. and Dr.Sc. degrees in electrical engineering from ETH Zurich, Switzerland, in 2006 and 2011, respectively. He is currently a Postdoctoral Associate at the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include classical and quantum information theory, and optical communication.

**Gregory W. Wornell** (S'83–M'91–SM'00–F'04) received the B.A.Sc. degree in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, and the S.M. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1985, 1987, and 1991, respectively.

Since 1991, he has been on the faculty at MIT, where he is the Sumitomo Professor of Engineering in the department of Electrical Engineering and Computer Science (EECS). He leads the Signals, Information, and Algorithms Laboratory in the Research Laboratory of Electronics, and co-chairs the EECS department graduate program. He has held visiting appointments at the former AT&T Bell Laboratories, Murray Hill, NJ, USA, the University of California, Berkeley, CA, USA, and Hewlett-Packard Laboratories, Palo Alto, CA, USA. His research interests and publications span the areas of signal processing, digital communication, and information theory, and include algorithms and architectures for wireless networks, sensing and imaging systems, digitally enhanced analog circuits and systems, multimedia applications, and aspects of computational biology and neuroscience.

Dr. Wornell has been involved in the Information Theory and Signal Processing Societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching.