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

Ligong Wang
Dept. EECS, MIT
Cambridge, MA, USA
wlg@mit.edu

Gregory W. Wornell
Dept. EECS, MIT
Cambridge, MA, USA
gww@mit.edu

***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 also includes the second-order term. We also show that pulse-position modulation is optimal on this channel up to the second-order term in capacity.**

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 which does not depend on the input x. We denote the Poisson distribution of mean ξ by $\mathcal{P}_\xi(\cdot)$ so

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

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

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

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, 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 λ is the average number of extraneous counts that appear in y due to background radiation or to the detector's "dark clicks".

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.

We impose an *average-power constraint*[1] on the input

$$\mathsf{E}[X] \leq \mathcal{E} \tag{3}$$

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

In applications like free-space or outer-space optical communications, the cost of producing and successfully transmitting photons is high, hence high *photon efficiency* (information transmitted per photon) 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

$$\lambda = c\mathcal{E}, \tag{4}$$

where c is some nonnegative constant that does not change with $\mathcal{E}$. Asymptotic results in this regime are relevant in scenarios where $\mathcal{E}$ is small and where λ is comparable to or much smaller than $\mathcal{E}$. Scenarios where $\mathcal{E}$ is small but λ is large is better captured by the model where λ stays constant while $\mathcal{E}$ tends to zero, see [2].

We denote the capacity (in nats[2]) of the channel (2) under power constraint (3) with dark current (4) by $C(\mathcal{E}, c)$, then

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

where the mutual information is computed from the channel law (2) and is maximized over input distributions satisfying (3), with dark current λ 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*

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

is imposed or not, as long as $\mathcal{A}$ 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

[1]Here "power" is in discrete time, means expected number of detected photons per channel use, and is proportional to the continuous-time physical power times the pulse duration.

[2]All logarithms in this paper are natural logarithms.

978-1-4673-0223-4/12/$31.00 ©2012 IEEE

the same scenario as the present paper and shows that [2, Proposition 1]

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

In other words, the photon efficiency in nats per photon, which we henceforth denote by $C_{\mathrm{PE}}(\mathcal{E},c)$, satisfies[3]

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

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

The approximation in (9) can be compared to the photon efficiency of 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 $C_{\mathrm{PE-bosonic}}(\mathcal{E})$ the photon efficiency of the pure-loss bosonic channel under an average-power constraint that is equivalent to (3). The value of $C_{\mathrm{PE-bosonic}}(\mathcal{E})$ can be easily computed using the explicit capacity formula derived in [7], which yields

$$C_{\mathrm{PE-bosonic}}(\mathcal{E}) = \log\frac{1}{\mathcal{E}} + 1 + o(1). \tag{10}$$

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.

Later, it is observed that restriction to 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 [10], [11]. It is argued in [10] that the maximum photon efficiency achievable using on-off signaling on the Poisson channel (2) with $\lambda = 0$ is given by

$$\log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(1). \tag{11}$$

Comparing the first-order term in (9) and the first- and second-order terms in (11) for $\mathcal{E} = 10^{-5}$, which is a realistic value in practice, yields a difference of about 20%, showing the importance of the refined approximation (11). However, the analysis in [10] is restricted to on-off signaling. 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 indeed the expression for $C_{\mathrm{PE}}(\mathcal{E},0)$.

It is already observed in [3], [4] that infinite photon efficiency on the Poisson channel with zero dark current can be achieved using *pulse-position modulation (PPM)*, and [11] further shows that PPM can achieve (11). Here PPM means a signaling scheme that satisfies the following:

- The channel inputs are divided into blocks of length b;
- In each block, there is only one input that is positive, which we call the "pulse", while all the other $(b-1)$ inputs are zeros;
- The pulses in all blocks have the same amplitude.

PPM signaling 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 signaling scheme that achieves (11) has a highly skewed input distribution and is hence difficult to code.

The goal of the present paper is to answer the following questions:

1) Is (11) the correct asymptotic expression up to the second-order term for $C_{\mathrm{PE}}(\mathcal{E},0)$?
2) Does the value of c affect the second-order term in $C_{\mathrm{PE}}(\mathcal{E},c)$?
3) Is PPM *second-order optimal* when $c > 0$? Here the notion "second-order optimal" has a slightly different meaning than, e.g., in additive Gaussian noise channels [12]. It means that, when $\mathcal{E}$ tends to zero, the signaling scheme asymptotically achieves $C_{\mathrm{PE}}(\mathcal{E},c) - \log\frac{1}{\mathcal{E}}$.

All these questions will be answered in the affirmative in our main result, Theorem 1.

The rest of this paper is arranged as follows: we state and discuss our main result in Section II; we then prove the achievability part of this result in Section III, and sketch the proof of the converse part in Section IV.

II. Main Result

The main result of this paper is the following:

Theorem 1. *The photon efficiency of the Poisson channel $C_{\mathrm{PE}}(\mathcal{E},c)$ as defined in (8) satisfies*

$$C_{\mathrm{PE}}(\mathcal{E},c) = \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(1), \quad c\in[0,\infty). \tag{12}$$

Furthermore, this asymptotic expression can be achieved using a PPM signaling scheme for all $c \in [0,\infty)$.

Since $C(\mathcal{E},c)$ and hence also $C_{\mathrm{PE}}(\mathcal{E},c)$ is monotonically decreasing in c [2], to prove Theorem 1 it suffices to show two things: first, that the largest photon efficiency achievable with PPM, which we henceforth denote by $C_{\mathrm{PE-PPM}}(\mathcal{E},c)$, satisfies

$$C_{\mathrm{PE-PPM}}(\mathcal{E},c) \geq \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(1), \quad c\in(0,\infty); \tag{13}$$

and second, that

$$C_{\mathrm{PE}}(\mathcal{E},0) \leq \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(1). \tag{14}$$

[3]Throughout this paper we use the usual $o(\cdot)$ and $O(\cdot)$ notations to describe behaviors of functions of $\mathcal{E}$ in the limit where $\mathcal{E}$ approaches zero. 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

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

We prove (13) in Section III, and, due to space limitations, sketch the proof of (14) in Section IV.

Answers to the questions in Section I:

- *Answer to Question 1):* Choosing $c = 0$ in (12) 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, 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, but not due to coherent input states. Attempts to overcome this loss by employing other feasible detection techniques have so far been unsuccessful [10], [13].
- *Answer to Question 2):* Perhaps surprisingly, the right-hand side (RHS) of (12) 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.
- *Answer to Question 3):* The fact that the RHS of (12) is achievable using PPM shows that PPM is indeed second-order optimal for all $c \in [0, \infty)$.

Further remarks about Theorem 1:

- In fact, in Section III we show that

$$C_{\text{PE}-\text{PPM}}(\mathcal{E}, c) \geq \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} - 2(1+c) + o(1), \tag{15}$$

  providing a lower bound on the $O(1)$ term on the RHS of (12). However, the bound (15) might not be tight.
- In the PPM scheme that achieves (12) and (15), which we describe in Section III, the pulse has amplitude $1/\big(\log(1/\mathcal{E})\big)$, which depends on $\mathcal{E}$, and which tends to zero as $\mathcal{E}$ tends to zero. This is in contrast to the on-off signaling scheme used in [2] where the "on" signal has a fixed amplitude that does not depend on $\mathcal{E}$. The latter on-off signaling scheme, as well as any PPM scheme with a fixed pulse amplitude, is not second-order optimal on the Poisson channel.
- Because in the second-order-optimal PPM scheme the pulse tends to zero as $\mathcal{E}$ tends to zero, we know that, as claimed in Section I, (12) still holds 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).
- In order to see how well the first and second terms on the RHS of (12) approximate the low-$\mathcal{E}$ photon efficiency, we numerically compare (12) with the nonasymptotic upper bound, and with the PPM-achieved lower bounds on C_{PE} in Figure 1. The formulas used to plot the PPM lower bounds are shown in Section III, and the formula used to plot the upper bound is shown in Section IV. Comparison shows that, for moderate values of c and small enough $\mathcal{E}$, the first- and second-order expression (12) provides a good approximation to the maximum achievable photon efficiency.

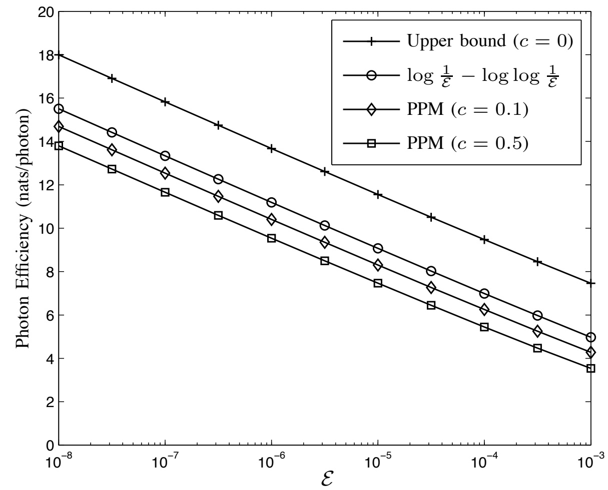

Fig. 1. Comparing the approximation (12) to nonasymptotic upper bound and PPM lower bounds.

III. Proof of the Lower Bound (13)

We show that the RHS of (13) is achievable with the following PPM scheme:

- The channel-uses are divided into blocks, with each block having b input symbols $x_1, \ldots, x_b$ and b corresponding output symbols $y_1, \ldots, y_b$. Later we set

$$b = \left\lfloor \frac{1}{\mathcal{E}\log\frac{1}{\mathcal{E}}} \right\rfloor. \tag{16}$$

- Within each length-b block, there is always one input that equals η, and all the other $(b-1)$ inputs are zeros. Each block is then fully specified by the position of its unique nonzero symbol, i.e., its pulse position. We can thus consider each block as a "super input symbol" $\tilde{x}$ which takes value in $\{1, \ldots, b\}$. Here $\tilde{x} = i$ means

$$x_i = \eta \tag{17a}$$

$$x_j = 0, \quad j \neq i. \tag{17b}$$

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

$$\eta = b\mathcal{E}. \tag{18}$$

- We further simplify the receiver by mapping the b output symbols $y_1, \ldots, y_b$ to one "super output symbol" $\tilde{y}$ which takes value in $\{1, \ldots, b, ?\}$ in the following way: let $\tilde{y} = i$, $i \in \{1, \ldots, b\}$, if y_i is the *unique* nonzero term in $\{y_1, \ldots, y_b\}$; and let $\tilde{y} = ?$ if there is more than one or no nonzero term in $\{y_1, \ldots, y_b\}$.

We can now compute the transition matrix of this PPM "super channel" as follows:

$$\tilde{W}(i|i) \triangleq \Pr\left[\tilde{Y} = i \middle| \tilde{X} = i\right] \tag{19}$$

$$= \Pr\left[Y_i \geq 1 \middle| X_i = \eta\right] \prod_{k \neq i} \Pr\left[Y_k = 0 \,|X_k = 0\right] \tag{20}$$

$$
\begin{aligned}
&= \big(1 - W(0|\eta)\big)\big(W(0|0)\big)^{b-1} && (21)\\
&= (1 - e^{-\eta - c\mathcal{E}})e^{-(b-1)c\mathcal{E}} && (22)\\
&= e^{-(b-1)c\mathcal{E}} - e^{-\eta - bc\mathcal{E}} && (23)\\
&\triangleq p_0, \qquad i \in \{1,\dots,b\}; && (24)
\end{aligned}
$$

$$
\begin{aligned}
\tilde{W}(j|i) &\triangleq \Pr\left[\tilde{Y} = j \middle| \tilde{X} = i\right] && (25)\\
&= \Pr\left[Y_i = 0 | X_i = \eta\right] \Pr\left[Y_j \geq 1 | X_j = 0\right]\\
&\quad \cdot \prod_{k \notin \{i,j\}} \Pr\left[Y_k = 0 | X_k = 0\right] && (26)\\
&= W(0|\eta)\big(1 - W(0|0)\big)\big(W(0|0)\big)^{b-2} && (27)\\
&= e^{-\eta - c\mathcal{E}}(1 - e^{-c\mathcal{E}})e^{-(b-2)c\mathcal{E}} && (28)\\
&= e^{-\eta-(b-1)c\mathcal{E}} - e^{-\eta - bc\mathcal{E}} && (29)\\
&\triangleq p_1, \qquad i,j \in \{1,\dots,b\},\ i \neq j; && (30)
\end{aligned}
$$

$$
\tilde{W}(?|i) = 1 - p_0 - (b-1)p_1, \qquad i \in \{1,\dots,b\}. \tag{31}
$$

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

$$
\tilde{C}(\mathcal{E}, c, b, \eta) = \max_{P_{\tilde{X}}} I(\tilde{X}; \tilde{Y}). \tag{32}
$$

Note that the total input power (i.e., expected number of detected signal-photons) in each block equals η. Therefore, for every admissible choice of b and η, we have the following lower bound on $C_{\text{PE-PPM}}(\mathcal{E}, c)$:

$$
C_{\text{PE-PPM}}(\mathcal{E}, c) \geq \frac{\tilde{C}(\mathcal{E}, c, b, \eta)}{\eta}. \tag{33}
$$

It can be easily verified that the optimal input distribution for (32) is the uniform distribution

$$
P_{\tilde{X}}(i) = \frac{1}{b}, \quad i \in \{1,\dots,b\}, \tag{34}
$$

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

$$
P_{\tilde{Y}}(i) = \frac{p_0 + (b-1)p_1}{b}, \quad i \in \{1,\dots,b\} \tag{35a}
$$

$$
P_{\tilde{Y}}(?) = 1 - p_0 - (b-1)p_1. \tag{35b}
$$

We can now use the above joint distribution on $(\tilde{X}, \tilde{Y})$ to explicitly compute $\tilde{C}(\mathcal{E}, c, b, \eta)$ as follows:[4]

$$
\begin{aligned}
&\tilde{C}(\mathcal{E}, c, b, \eta)\\
&\quad = I\left(\tilde{X}; \tilde{Y}\right) && (36)\\
&\quad = H\left(\tilde{Y}\right) - H\left(\tilde{Y}\middle|\tilde{X}\right) && (37)\\
&\quad = (1 - p_0 - (b-1)p_1)\log\frac{1}{1 - p_0 - (b-1)p_1}\\
&\qquad + (p_0 + (b-1)p_1)\log\frac{b}{p_0 + (b-1)p_1}\\
&\qquad - (1 - p_0 - (b-1)p_1)\log\frac{1}{1 - p_0 - (b-1)p_1}\\
&\qquad - p_0\log\frac{1}{p_0} - (b-1)p_1\log\frac{1}{p_1} && (38)\\
&\quad = p_0\log b + p_0\log\frac{p_0}{p_0 + (b-1)p_1}\\
&\qquad + (b-1)p_1\log\frac{bp_1}{p_0 + (b-1)p_1}. && (39)
\end{aligned}
$$

[4]Throughout this paper we adopt the convention $0 \log 0 = 0$.

At this point, we note that the PPM curves in Figure 1 are obtained using (33) with (39), together with the choices (16) and (18).

Using the fact that $\log(1 + a) \leq a$ for all $a \in \mathbb{R}$, we can continue (39) to lower-bound $\tilde{C}(\mathcal{E}, c, b, \eta)$ as

$$
\begin{aligned}
\tilde{C}(\mathcal{E}, c, b, \eta) &= p_0\log b - p_0\underbrace{\log\frac{p_0 + (b-1)p_1}{p_0}}_{\leq\frac{(b-1)p_1}{p_0}}\\
&\quad - (b-1)p_1\underbrace{\log\frac{p_0 + (b-1)p_1}{bp_1}}_{\leq\frac{p_0 - p_1}{bp_1}} && (40)\\
&\geq p_0\log b - (b-1)p_1 - \underbrace{\frac{b-1}{b}}_{\leq 1}\underbrace{(p_0 - p_1)}_{\leq p_0} && (41)\\
&\geq p_0\log b - (b-1)p_1 - p_0. && (42)
\end{aligned}
$$

Next note that p_0 can be upper-bounded as

$$
\begin{aligned}
p_0 &= e^{-(b-1)c\mathcal{E}} - e^{-\eta - bc\mathcal{E}} && (43)\\
&= \underbrace{e^{-(b-1)c\mathcal{E}}}_{\leq 1}\underbrace{\left(1 - e^{-\eta - c\mathcal{E}}\right)}_{\leq \eta + c\mathcal{E}} && (44)\\
&\leq \eta + c\mathcal{E}, && (45)
\end{aligned}
$$

and can also be lower-bounded as

$$
\begin{aligned}
p_0 &= e^{-(b-1)c\mathcal{E}} - e^{-\eta - bc\mathcal{E}} && (46)\\
&\geq e^{-bc\mathcal{E}} - e^{-\eta - bc\mathcal{E}} && (47)\\
&= \underbrace{e^{-bc\mathcal{E}}}_{\geq 1 - bc\mathcal{E}}\underbrace{(1 - e^{-\eta})}_{\geq \eta - \eta^2} && (48)\\
&\geq (1 - bc\mathcal{E})(\eta - \eta^2), && (49)
\end{aligned}
$$

where the last inequality holds if both multiplicands on its RHS are positive, which, for our choices of b and η in (16) and (18), is true for small enough $\mathcal{E}$. Also note that p_1 can be upper-bounded as

$$
\begin{aligned}
p_1 &= e^{-\eta-(b-1)c\mathcal{E}} - e^{-\eta - bc\mathcal{E}} && (50)\\
&= \underbrace{e^{-\eta-(b-1)c\mathcal{E}}}_{\leq 1}\underbrace{\left(1 - e^{-c\mathcal{E}}\right)}_{\leq c\mathcal{E}} && (51)\\
&\leq c\mathcal{E}. && (52)
\end{aligned}
$$

Using (45), (49) and (52) we can continue the chain of inequalities (42) to further lower-bound $\tilde{C}(\mathcal{E}, c, b, \eta)$ as

$$
\begin{aligned}
&\tilde{C}(\mathcal{E}, c, b, \eta)\\
&\quad \geq (1 - bc\mathcal{E})(\eta - \eta^2)\log b - (b-1)c\mathcal{E} - \eta - c\mathcal{E} && (53)\\
&\quad = (1 - bc\mathcal{E})(\eta - \eta^2)\log b - bc\mathcal{E} - \eta && (54)\\
&\quad = (\eta - \eta^2)\log b - bc\mathcal{E}\underbrace{(\eta - \eta^2)}_{\leq \eta}\log b - bc\mathcal{E} - \eta && (55)\\
&\quad \geq (\eta - \eta^2)\log b - bc\mathcal{E}\eta\log b - bc\mathcal{E} - \eta. && (56)
\end{aligned}
$$

Combining (33) and (56) yields

$$C_{\text{PE-PPM}}(\mathcal{E},c) \geq \frac{(\eta-\eta^2)\log b - bc\mathcal{E}\eta\log b - bc\mathcal{E} - \eta}{\eta} \quad (57)$$

$$= (1-\eta)\log b - bc\mathcal{E}\log b - \frac{bc\mathcal{E}}{\eta} - 1. \quad (58)$$

We now set the values of b and η to be as in (16) and (18). Note that, as $\mathcal{E}$ approaches zero, the RHS of (16) tends to infinity, so we may drop the $\lfloor\cdot\rfloor$ operation without affecting the asymptotic results. Hence we shall use the following:

$$b = \frac{1}{\mathcal{E}\log\frac{1}{\mathcal{E}}} \quad (59a)$$

$$\eta = \frac{1}{\log\frac{1}{\mathcal{E}}}. \quad (59b)$$

Plugging (59) into (58) we obtain the following lower bound on $C_{\text{PE-PPM}}(\mathcal{E},c)$:

$$C_{\text{PE-PPM}}(\mathcal{E},c) \geq \left(1-\frac{1}{\log\frac{1}{\mathcal{E}}}\right)\log\left(\frac{1}{\mathcal{E}\log\frac{1}{\mathcal{E}}}\right) - \frac{c}{\log\frac{1}{\mathcal{E}}}\log\left(\frac{1}{\mathcal{E}\log\frac{1}{\mathcal{E}}}\right) - c - 1 \quad (60)$$

$$= \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} - 1 + \frac{\log\log\frac{1}{\mathcal{E}}}{\log\frac{1}{\mathcal{E}}} - c + \frac{c\log\log\frac{1}{\mathcal{E}}}{\log\frac{1}{\mathcal{E}}} - c - 1 \quad (61)$$

$$= \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} - 2(1+c) + o(1) \quad (62)$$

$$= \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(1). \quad (63)$$

This establishes both (13) and (15).

IV. Proof Sketch of the Upper Bound (14)

To prove (14), like in [2], we use the duality bound [14] which states that, for any distribution $R(\cdot)$ on the output, the channel capacity satisfies

$$C \leq \sup \mathsf{E}\left[D\big(W(\cdot|X)\|R(\cdot)\big)\right], \quad (64)$$

where the supremum is taken over all allowed input distributions. Since we are interested in the limit where $\mathcal{E}$ tends to zero, we may assume that $\mathcal{E} < 1$. In this case, we choose $R(\cdot)$ to be the following distribution:

$$R(y) = \begin{cases} 1-\mathcal{E}, & y=0 \\ \mathcal{E}(1-a)a^{y-1}, & y=1,2,\ldots \end{cases} \quad (65)$$

where a is a positive constant whose exact value is not important for our analysis.

Using (64) with (65), after some calculations we obtain that, for small enough $\mathcal{E}$,

$$C(\mathcal{E},0) \leq \mathcal{E}\log\frac{1}{\mathcal{E}} - \mathcal{E}\log\log\frac{1}{\mathcal{E}} + \underbrace{\mathcal{E}\log 12 + \log\frac{1}{1-\mathcal{E}} + \mathcal{E}\log\frac{1}{1-a}}_{=O(\mathcal{E})} \quad (66)$$

$$= \mathcal{E}\left(\log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(1)\right). \quad (67)$$

At this point, we note that the upper bound in Figure 1 is computed using (66) in the limit where a tends to zero.

Dividing both sides of (67) by $\mathcal{E}$ yields the desired upper bound (14).

Acknowledgments

The authors wish to thank Hye Won Chung, Yuval Kochman, Amos Lapidoth 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. I (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. Giovanetti, 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. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
[13] 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.
[14] 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.