Covert Communication Over the Poisson Channel

Ligong Wang, *Member, IEEE*

***Abstract*—We consider the problem of communication over a continuous-time Poisson channel subject to a covertness constraint: The relative entropy between the output distributions when a message is transmitted and when no input is provided must be small. In the absence of both bandwidth and peak-power constraints, we show the covert communication capacity of this channel, in nats per second, to be infinity. When a peak-power constraint is imposed on the input, the covert communication capacity becomes zero, and the "square-root scaling law" applies. When a bandwidth constraint but no peak-power constraint is imposed, the covert communication capacity is again shown to be zero, but we have not determined whether the square-root law holds or not.**

***Index Terms*—Covert communication, low probability of detection, Poisson channel.**

I. Introduction

COVERT communication, or communication with low probability of detection [2]–[5], refers to applications where the transmitter and the receiver must keep an eavesdropper from discovering the fact that they are using the channel to communicate. Specifically, the signals observed by the eavesdropper when the channel is used must be statistically close to the signals observed when the channel is not used. For additive white Gaussian noise (AWGN) channels, the channel not being used is modeled by the transmitter always sending zero; for a discrete memoryless channel (DMC), this is modeled by the transmitter sending an "innocent" symbol—a specific input symbol, the choice of which is given as part of the channel model. For a DMC, if the output distribution at the eavesdropper generated by the innocent symbol is a convex combination of the output distributions generated by the other, non-innocent input symbols, then a positive covert communication rate is achievable; otherwise the maximum amount of information that can be covertly communicated grows proportionally to the square root of the total number of channel uses [4]. For the AWGN channel, the latter, square-root scaling law applies [2], [4].

As most practically relevant channel models appear to follow the square-root scaling law and thus have zero capacity (in nats per channel use)[1] for covert communication, some works have considered variations of the DMC and AWGN channel models. For example, [6]–[8] consider channels with parameters that are unknown to the eavesdropper, and [9] considers channels with channel-state information (CSI) available to the transmitter. These works show that, under such additional assumptions, positive covert communication rates are sometimes achievable.

The present paper is concerned with covert communication over the Poisson channel [10]–[12]. The Poisson channel is a standard model for optical communication employing intensity modulation at the transmitter and photodetection at the receiver, hence covert communication over such a channel is of practical importance.

On a continuous-time Poisson channel, the input is a nonnegative waveform, and the output is a doubly stochastic Poisson process whose conditional intensity is given by the input waveform plus the "dark current"—a constant describing background noise. We assume that, when the transmitter is not sending a message, it sends the all-zero waveform.

Like [4], we adopt a setup where the intended receiver and the eavesdropper observe the same channel output, and where the transmitter and the receiver share a sufficiently long secret key. The key allows the transmitter and the receiver to use a random codebook whose realization is unknown to the eavesdropper. In all our achievability proofs, we shall generate the random codebook by choosing all input symbols to be independent and identically distributed (IID). This renders the output also IID, so the covertness analysis can be performed by computing relative entropies between distributions of a single output random variable. Using IID inputs is obviously inefficient in the usage of the secret key. For bounds on the number of secret key nats that need to be used, or for covert communication without using a secret key, see, e.g., [3], [5]. We note that, when one adopts a wiretap channel model [13] with different channel coefficients (gain and dark current in the Poisson case) for the intended receiver and the eavesdropper, the scaling behavior (e.g., linear or square-root) that we derive in this paper will remain unchanged, but the required number of key nats will depend heavily on the difference between the two channels.

The first interest of this work lies in the case where there is neither a peak constraint nor a bandwidth constraint on the input waveform. We show that, in this case, the Poisson channel has an infinite per-second covert communication capacity. Thus, the continuous-time Poisson channel follows no square-root scaling law that is analogous to the one for discrete-time AWGN channels and many DMCs; instead, it allows the number of covert information nats to grow super-linearly with total communication time.

Compared to most previously studied channel models that allow positive-rate covert communication, our model is of a different nature. First, the Poisson channel is memoryless, but is dissimilar to those DMCs that allow positive-rate covert communication [4], in the sense that its output distribution

This work was presented in part at the 2018 IEEE International Symposium on Information Theory (ISIT) in Vail, CO, USA [1].

The author is with ETIS, UMR 8051, CY Cergy Paris Université, ENSEA, CNRS, Cergy-Pontoise, 95000 France (email: ligong.wang@ensea.fr).

[1]In some works, the term "covert capacity" refers to the scaling factor of information nats with respect to the square root of the total number of channel uses. In the present work, we use "capacity" is its traditional sense, i.e., nats per channel use or per second.

conditional on the all-zero input is *not* a convex combination of the output distributions conditional on any input waveforms that are not almost everywhere zero. To see this, simply note that any input waveform with a positive Lebesgue integral will increase the expectation of the total number of output arrivals compared to the all-zero input waveform. Second, our model is dissimilar to the queueing channel considered in [14]. In [14] it is assumed that, when no message is being communicated via the queueing channel, packets arrive in a random, Poisson process, making the problem similar to that of "stealth" [15]. Third, we do not assume any unknown parameter or channel state as in [6]–[9]. Such additional assumptions would typically allow the transmitter to send a non-vanishing average signal power or a non-vanishing percentage of non-innocent symbols. As we shall see, our Poisson model requires that the average input power tend to zero as total communication time grows large, but even this vanishing input power provides unbounded communication capacity.

In some sense, this Poisson channel has infinite covert communication capacity because it is a continuous-time channel with no bandwidth limit. Indeed, as we show later on in this paper, once we restrict the communication bandwidth to a certain value to reduce the Poisson channel to discrete time, the maximum covert communication rate will again vanish as total communication time grows large. Also crucial to the infinite-capacity result is the absence of a peak-power constraint. It is well known that a peak-constrained infinite-bandwidth Poisson channel has a finite per-second communication capacity [10]–[12], whereas, without a peak constraint, its capacity for standard, non-covert communication is infinity. We shall show that the peak-constrained Poisson channel obeys the square-root scaling law for covert communication, i.e., that the maximum number of nats that can be communicated covertly grows proportionally to the square root of total communication time.

The rest of this paper is organized as follows. Section II studies the Poisson channel with neither peak-power nor bandwidth constraint, and shows its per-second covert communication capacity to be infinity. Section III treats the peak-limited case and derives the exact scaling law. Section IV treats the band-limited case and derives an upper bound. Section V then concludes the paper with some remarks.

II. The No-Constraint Case

Consider a continuous-time Poisson channel with constant dark current $\lambda > 0$. The input to the channel must be a Lebesgue-measurable, nonnegative signal $x(t) \geq 0$, $t \in [0, \mathsf{T}]$, where $\mathsf{T} > 0$ is the total communication time. Conditional on this input signal, the channel output $Y(t)$, $t \in [0, \mathsf{T}]$, is a Poisson process whose time-t intensity is $x(t) + \lambda$. That means $Y(\cdot)$ is a random counting function with $Y(0) = 0$ with probability one and, for $t_1, t_2 \in [0, \mathsf{T}]$, $t_1 < t_2$,

$$\Pr[Y(t_2) - Y(t_1) = k | X(\cdot) = x(\cdot)] = e^{-s} \frac{s^k}{k!}, \quad k \in \mathbb{Z}_0^+, \tag{1}$$

where

$$s = \int_{t_1}^{t_2} \big(x(t) + \lambda\big) \, \mathrm{d}t. \tag{2}$$

A codebook on $[0, \mathsf{T}]$ at rate R nats per second for communication over this channel contains e^{TR} input signals $x_m(t)$, $t \in [0, \mathsf{T}]$, $m \in \{1, \ldots, e^{\mathsf{TR}}\}$. An encoder maps a message $m \in \{1, \ldots, e^{\mathsf{TR}}\}$ to the corresponding input signal $x_m(t)$, $t \in [0, \mathsf{T}]$, and sends this signal into the channel. A corresponding decoder maps the output signal $y(t)$, $t \in [0, \mathsf{T}]$, to the decoded message $\hat{m} \in \{1, \ldots, e^{\mathsf{TR}}\}$. With the help of a sufficiently long secret key, the transmitter and the receiver use a random code to communicate their message. The eavesdropper, which is colocated with the receiver, is assumed to know the distribution according to which the random code is chosen, but not the actual choice (because it does not know the secret key).

In this section, we do not impose a peak constraint on the input signal, hence the value of $x_m(t)$ at a specific time t may be arbitrarily large. The peak-limited case is discussed in the next section. As we shall see, our result does not depend on whether there is an average-power constraint on the input or not.

Let Q^{T} denote the distribution of $Y(t)$, $t \in [0, \mathsf{T}]$, when there is no input signal, i.e., when $x(t) = 0$ for all $t \in [0, \mathsf{T}]$. Thus, Q^{T} describes the homogeneous Poisson process of intensity λ on $[0, \mathsf{T}]$. Further, let P^{T} denote the distribution of $Y(t)$, $t \in [0, \mathsf{T}]$, induced by the random codebook and a uniformly chosen message M. Thus, P^{T} is the unconditional output distribution when the channel is used to communicate a message.

We say that a rate R is *achievable covertly* if, for every $\mathsf{T} > 0$, we can construct a random code on $[0, \mathsf{T}]$ of rate at least R such that the probability of a decoding error, i.e., of $\hat{M} \neq M$, tends to zero as T tends to infinity, and that

$$\lim_{\mathsf{T} \to \infty} \mathscr{D}(P^{\mathsf{T}} \| Q^{\mathsf{T}}) = 0, \tag{3}$$

where $\mathscr{D}(\cdot \| \cdot)$ denotes the relative entropy: For two probability measures P and Q over the same set, if P is absolutely continuous with respect to Q, then

$$\mathscr{D}(P \| Q) = \mathsf{E}_P \left[\log \frac{\mathrm{d}P}{\mathrm{d}Q} \right], \tag{4}$$

where $\frac{\mathrm{d}P}{\mathrm{d}Q}$ is the Radon-Nikodym derivative of P with respect to Q. (Alternatively, one can define $\mathscr{D}(P \| Q)$ as the supremum over all finite partitions of the sample space of the relative entropy between the probability mass functions resulting from P and Q by the partition; see [16].) Note that, by Pinsker's inequality [17], the condition (3) implies that the total variation distance between P^{T} and Q^{T} must also tend to zero as $\mathsf{T} \to \infty$.

We define the *covert communication capacity* for this channel as the supremum of all rates that are achievable covertly.

Theorem 1: The covert communication capacity of the continuous-time Poisson channel described above is infinity, i.e., all positive rates are achievable covertly.

Clearly, only the achievability part of Theorem 1 needs proof. Before proceeding with the proof, we describe a random coding scheme and analyze its covertness. These will be used in the proof of Theorem 1, and again in the next section.

Scheme 1: Fix $\tau, \eta > 0$ and $q \in (0,1)$. Divide the interval $[0, \mathsf{T}]$ into slots of τ seconds: $[0,\tau), [\tau, 2\tau), \ldots$. Randomly generate a codebook[2] by choosing each codeword independently according to the following distribution: Within the kth slot, $k \in \{1, \ldots, \lfloor \frac{\mathsf{T}}{\tau} \rfloor\}$, $X(t) = X'_k$ with probability one for all $t \in [(k-1)\tau, k\tau)$, where the random variables X'_k, $k = 1, \ldots, \lfloor \frac{\mathsf{T}}{\tau} \rfloor$, are IID with

$$X'_k = \begin{cases} \eta, & \text{with probability } q, \\ 0, & \text{with probability } 1-q. \end{cases} \tag{5}$$

Proposition 1: Suppose the transmitter uses Scheme 1 to transmit a message. If $(\lambda + q\eta)\tau < 1$, then

$$\mathscr{D}(P^{\mathsf{T}} \| Q^{\mathsf{T}}) \le \left(\frac{q^2\eta^2}{2\lambda} + \frac{q^3\eta^3}{3\lambda^2} + \frac{q^4\eta^4}{3\lambda^3} + \lambda\eta\tau + q^2\eta^2\tau + \frac{q\eta^2\tau}{2} + \frac{q\eta}{\lambda}(\eta+\lambda)^2\tau \right) \cdot \mathsf{T}. \tag{6}$$

Proof: For the kth τ-second slot, let Y'_k denote the total number of arrivals in $Y(\cdot)$ within this slot, that is $Y'_k = Y(k\tau) - Y((k-1)\tau)$, $k \in \{1, \ldots, \lfloor \frac{\mathsf{T}}{\tau} \rfloor\}$. Note that, since with probability one $X(t)$ is constant within the slot (both when the transmitter is sending and when it is not sending a message), $Y'_1, \ldots, Y'_{\lfloor \frac{\mathsf{T}}{\tau} \rfloor}$ form a sufficient statistic for the eavesdropper. Under both P^{T} and Q^{T}, $Y'_1, \ldots, Y'_{\lfloor \frac{\mathsf{T}}{\tau} \rfloor}$ are IID. When no message is being communicated (i.e., under Q^{T}), each Y'_k has the Poisson distribution of mean $\lambda\tau$, which we denote by Q'. When the transmitter is sending a message (under P^{T}), for each k, with probability $1-q$ the random variable Y'_k has the Poisson distribution of mean $\lambda\tau$, and with probability q it has the Poisson distribution of mean $(\eta+\lambda)\tau$; we denote this mixture of two Poisson distributions by P'. Then we have

$$\mathscr{D}(P^{\mathsf{T}} \| Q^{\mathsf{T}}) = \left\lfloor \frac{\mathsf{T}}{\tau} \right\rfloor \mathscr{D}(P' \| Q'). \tag{7}$$

We next bound $\mathscr{D}(P' \| Q')$. For each $i \in \mathbb{Z}_0^+$, we have

$$Q'(i) = \frac{e^{-\lambda\tau}(\lambda\tau)^i}{i!} \tag{8a}$$

$$P'(i) = (1-q)\frac{e^{-\lambda\tau}(\lambda\tau)^i}{i!} + q\frac{e^{-(\eta+\lambda)\tau}(\eta\tau+\lambda\tau)^i}{i!}. \tag{8b}$$

Hence

$$\mathscr{D}(P' \| Q') = \sum_{i=0}^{\infty} P'(i) \log \frac{P'(i)}{Q'(i)} \tag{9}$$

$$= \sum_{i=0}^{\infty} P'(i) \log\left(1 - q + qe^{-\eta\tau}\left(1+\frac{\eta}{\lambda}\right)^i\right) \tag{10}$$

$$= \sum_{i=0}^{\infty} \sigma_i, \tag{11}$$

where in the last line we defined

$$\sigma_i \triangleq P'(i) \log\left(1 - q + qe^{-\eta\tau}\left(1+\frac{\eta}{\lambda}\right)^i\right), \quad i \in \mathbb{Z}_0^+. \tag{12}$$

[2]We emphasize that this random code is what the transmitter and the receiver will use, and not an intermediate step in proving the existence of a good deterministic (or "less random") code, as in, e.g., [17].

We separately bound σ_0, σ_1, and the remaining terms.

For σ_0 we have

$$P'(0) = (1-q)e^{-\lambda\tau} + qe^{-(\eta+\lambda)\tau} \tag{13}$$

$$\ge (1-q)(1-\lambda\tau) + q(1-(\eta+\lambda)\tau) \tag{14}$$

$$= 1 - (\lambda + q\eta)\tau, \tag{15}$$

where we used the fact that $e^{-a} \ge 1 - a$ for all $a \in \mathbb{R}$. We also have

$$\log(1 - q + qe^{-\eta\tau}) \le -q(1 - e^{-\eta\tau}) \tag{16}$$

$$\le -q\left(\eta\tau - \frac{\eta^2\tau^2}{2}\right), \tag{17}$$

where the first inequality follows because $\log(1+a) \le a$ for all $a \in (-1, \infty)$, and the second inequality because $e^{-a} \le 1 - a + \frac{a^2}{2}$ for $a \ge 0$. Note that the left-hand side of (16) is negative, whereas, when $(\lambda + q\eta)\tau < 1$, the right-hand side of (15) is positive. We can hence combine (15) and (17) to obtain, whenever $(\lambda + q\eta)\tau < 1$,

$$\sigma_0 \le -(1 - (\lambda + q\eta)\tau) \cdot q\left(\eta\tau - \frac{\eta^2\tau^2}{2}\right) \tag{18}$$

$$\le -q\eta\tau\left(1 - \left(\lambda + q\eta + \frac{\eta}{2}\right)\tau\right), \tag{19}$$

where the last inequality follows by dropping a negative term.

For σ_1 we have

$$P'(1) = (1-q)\lambda\tau \underbrace{e^{-\lambda\tau}}_{\le 1} + q(\eta+\lambda)\tau \underbrace{e^{-(\eta+\lambda)\tau}}_{\le 1} \tag{20}$$

$$\le (\lambda + q\eta)\tau \tag{21}$$

and

$$\log\left(1 - q + q\underbrace{e^{-\eta\tau}}_{\le 1}\left(1+\frac{\eta}{\lambda}\right)\right) \le \log\left(1+\frac{q\eta}{\lambda}\right) \tag{22}$$

$$\le \frac{q\eta}{\lambda} - \frac{q^2\eta^2}{2\lambda^2} + \frac{q^3\eta^3}{3\lambda^3}. \tag{23}$$

Combining (21) and (23) and dropping a negative term we obtain

$$\sigma_1 \le q\eta\tau\left(1 + \frac{q\eta}{2\lambda} + \frac{q^2\eta^2}{3\lambda^2} + \frac{q^3\eta^3}{3\lambda^3}\right). \tag{24}$$

For the remaining $\{\sigma_i \colon i \ge 2\}$, we have

$$\log\left(1 - q + q\underbrace{e^{-\eta\tau}}_{\le 1}\left(1+\frac{\eta}{\lambda}\right)^i\right)$$

$$\le \log\left(1 + \underbrace{q}_{\le 1}\left(\left(1+\frac{\eta}{\lambda}\right)^i - 1\right)\right) \tag{25}$$

$$\le \log\left(1 + \left(1+\frac{\eta}{\lambda}\right)^i - 1\right) \tag{26}$$

$$= i \cdot \log\left(1+\frac{\eta}{\lambda}\right) \tag{27}$$

$$\le i \cdot \frac{\eta}{\lambda}. \tag{28}$$

Thus

$$\sum_{i=2}^{\infty} \sigma_i \le \frac{\eta}{\lambda} \sum_{i=2}^{\infty} P'(i) \cdot i, \tag{29}$$

where

$$\begin{aligned}
&\sum_{i=2}^{\infty} P'(i)\cdot i \\
&\quad= \sum_{i=1}^{\infty} P'(i)\cdot i - P'(1)\cdot 1 && (30)\\
&\quad= \mathsf{E}_{P'}[Y'] - P'(1) && (31)\\
&\quad= \big((1-q)\lambda\tau + q(\eta+\lambda)\tau\big) \\
&\qquad - \big((1-q)\underbrace{e^{-\lambda\tau}}_{\geq 1-\lambda\tau}\lambda\tau + q\underbrace{e^{-(\eta+\lambda)\tau}}_{\geq 1-(\eta+\lambda)\tau}(\eta+\lambda)\tau\big) && (32)\\
&\quad\leq (1-q)\lambda\tau + q(\eta+\lambda)\tau - (1-q)(1-\lambda\tau)\lambda\tau \\
&\qquad - q(1-(\eta+\lambda)\tau)(\eta+\lambda)\tau && (33)\\
&\quad= (1-q)\lambda^2\tau^2 + q(\eta+\lambda)^2\tau^2. && (34)
\end{aligned}$$

Hence we obtain

$$\sum_{i=2}^{\infty}\sigma_i \leq \frac{\eta}{\lambda}\big((1-q)\lambda^2\tau^2 + q(\eta+\lambda)^2\tau^2\big). \tag{35}$$

Combining (7), (11), (19), (24), and (35), we conclude that (6) holds whenever $(\lambda+q\eta)\tau<1$. ∎

Proof of Theorem 1: Let the transmitter use the random code in Scheme 1, where τ, η, and q are chosen as (deterministic) functions of T:

$$\tau = \mathsf{T}^{-1}e^{-\mathsf{T}} \tag{36a}$$
$$\eta = e^{\mathsf{T}/2} \tag{36b}$$
$$q = \mathsf{T}^{-3/4}e^{-\mathsf{T}/2}. \tag{36c}$$

Note that the average input power by the above choice equals $q\eta = \mathsf{T}^{-3/4}$, which tends to zero as T tends to infinity. Hence any nonzero average-power constraint on the input will be satisfied when T grows sufficiently large.

Applying the choices (36) to Proposition 1 confirms that (3) is satisfied, i.e., that the random code is covert. It remains to derive a lower bound on the communication rates that are achievable using the proposed scheme. To this end, let the decoder map the output waveform to a sequence $\hat{Y}_1,\ldots,\hat{Y}_{\lfloor\frac{\mathsf{T}}{\tau}\rfloor}$ as follows:

$$\hat{Y}_k = \begin{cases} 0, & \text{if } Y(k\tau) = Y((k-1)\tau) \\ 1, & \text{otherwise.}\end{cases} \tag{37}$$

Then we have a discrete memoryless channel with

$$\Pr(\hat{Y}=1|X'=0) = 1-e^{-\lambda\tau} \tag{38a}$$
$$\Pr(\hat{Y}=1|X'=\eta) = 1-e^{-(\eta+\lambda)\tau}. \tag{38b}$$

We then use the *information-spectrum method* [18], [19] to lower-bound the maximum achievable rate. Let the input vector of length $\lfloor\frac{\mathsf{T}}{\tau}\rfloor$ be denoted by $\mathbf{X}'$ and the corresponding binary output vector be denoted by $\hat{\mathbf{Y}}$, then the following rate is achievable:

$$P\text{-}\liminf_{\mathsf{T}\to\infty}\left\{\frac{1}{\mathsf{T}}\log\frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}\big(\hat{\mathbf{Y}}\big|\mathbf{X}'\big)}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})}\right\}, \tag{39}$$

where P-$\liminf$ denotes the *limit infimum in probability*; see [18], [19]. To bound (39), we bound the mean and the variance of the random variable inside the P-$\liminf$. Henceforth we ignore the $\lfloor\cdot\rfloor$ operator, whose effect vanishes as $\mathsf{T}\to\infty$.

For any fixed T, $(\mathbf{X}',\hat{\mathbf{Y}})$ is IID across different time slots, hence

$$\mathsf{E}\left[\frac{1}{\mathsf{T}}\log\frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}\big(\hat{\mathbf{Y}}\big|\mathbf{X}'\big)}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})}\right] = \frac{1}{\tau}\,\mathrm{I}(X';\hat{Y}). \tag{40}$$

By (5) and (38), $\mathrm{I}(X';\hat{Y})$ can be written as

$$\begin{aligned}
\mathrm{I}(X';\hat{Y}) = \mathsf{H}_{\mathrm{b}}\Big((1-q)(1-e^{-\lambda\tau}) + q(1-e^{-(\lambda+\eta)\tau})\Big) \\
- (1-q)\,\mathsf{H}_{\mathrm{b}}(1-e^{-\lambda\tau}) - q\,\mathsf{H}_{\mathrm{b}}(1-e^{-(\lambda+\eta)\tau}),
\end{aligned} \tag{41}$$

where $\mathsf{H}_{\mathrm{b}}(\cdot)$ denotes the binary entropy function:

$$\mathsf{H}_{\mathrm{b}}(p) = -p\log p - (1-p)\log(1-p), \quad p\in[0,1]. \tag{42}$$

We bound the three terms on the right-hand side of (41) separately.

For the first term on the right-hand side of (41), we drop a nonnegative term to get

$$\begin{aligned}
&\mathsf{H}_{\mathrm{b}}\Big((1-q)(1-e^{-\lambda\tau}) + q(1-e^{-(\lambda+\eta)\tau})\Big) \\
&\quad\geq \Big((1-q)(1-e^{-\lambda\tau}) + q(1-e^{-(\lambda+\eta)\tau})\Big) \\
&\qquad\cdot\log\frac{1}{(1-q)(1-e^{-\lambda\tau}) + q(1-e^{-(\lambda+\eta)\tau})}.
\end{aligned} \tag{43}$$

Then we bound

$$\begin{aligned}
&(1-q)\underbrace{(1-e^{-\lambda\tau})}_{\geq\lambda\tau-\frac{\lambda^2\tau^2}{2}} + q\underbrace{(1-e^{-(\lambda+\eta)\tau})}_{\geq(\lambda+\eta)\tau-\frac{(\lambda+\eta)^2\tau^2}{2}} \\
&\qquad\geq (1-q)\left(\lambda\tau - \frac{\lambda^2\tau^2}{2}\right) \\
&\qquad\quad + q\left((\lambda+\eta)\tau - \frac{(\lambda+\eta)^2\tau^2}{2}\right) && (44)\\
&\qquad\geq \left((\lambda+q\eta) - \frac{q(\lambda+\eta)^2\tau}{2} - \frac{\lambda^2\tau}{2}\right)\tau, && (45)
\end{aligned}$$

where the last inequality follows by dropping a term $\frac{q\lambda^2\tau^2}{2}$. We also have

$$\begin{aligned}
&(1-q)\underbrace{(1-e^{-\lambda\tau})}_{\leq\lambda\tau} + q\underbrace{(1-e^{-(\lambda+\eta)\tau})}_{\leq(\lambda+\eta)\tau} \\
&\qquad\leq (1-q)\lambda\tau + q(\lambda+\eta)\tau = (\lambda+q\eta)\tau.
\end{aligned} \tag{46}$$

Combining (43), (45), and (46) we have, for large enough T,

$$\begin{aligned}
&\mathsf{H}_{\mathrm{b}}\Big((1-q)(1-e^{-\lambda\tau}) + q(1-e^{-(\lambda+\eta)\tau})\Big) \\
&\quad\geq \left((\lambda+q\eta) - \frac{q(\lambda+\eta)^2\tau}{2} - \frac{\lambda^2\tau}{2}\right)\tau\log\frac{1}{(\lambda+q\eta)\tau}.
\end{aligned} \tag{47}$$

Now consider the other two terms on the right-hand side of (41). By our choices (36), when T is sufficiently large, $\lambda\tau<\frac{1}{2}$

and $(\lambda+\eta)\tau < \frac{1}{2}$. Since $\mathrm{H_b}(\cdot)$ is monotonically increasing on $(0, \frac{1}{2})$, we can bound

$$\mathrm{H_b}(1-e^{-\lambda\tau}) \le \mathrm{H_b}(\lambda\tau) \tag{48}$$

$$= \lambda\tau \log\frac{1}{\lambda\tau} + (1-\lambda\tau)\underbrace{\log\frac{1}{1-\lambda\tau}}_{\le \frac{\lambda\tau}{1-\lambda\tau}} \tag{49}$$

$$\le \lambda\tau\log\frac{1}{\lambda\tau} + \lambda\tau. \tag{50}$$

Similarly,

$$\mathrm{H_b}(1-e^{-(\eta+\lambda)\tau}) \le (\eta+\lambda)\tau\log\frac{1}{(\eta+\lambda)\tau} + (\eta+\lambda)\tau. \tag{51}$$

Combining (40), (41), (47), (50), and (51) yields, for large enough T,

$$\begin{aligned}
&\mathsf{E}\left[\frac{1}{\mathsf{T}}\log\frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}\big(\hat{\mathbf{Y}}|\mathbf{X}'\big)}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})}\right] \\
&\ge \left((\lambda+q\eta) - \frac{q(\lambda+\eta)^2\tau}{2} - \frac{\lambda^2\tau}{2}\right)\log\frac{1}{(\lambda+q\eta)\tau} \\
&\quad - (1-q)\lambda\log\frac{1}{\lambda\tau} - (1-q)\lambda \\
&\quad - q(\eta+\lambda)\log\frac{1}{(\eta+\lambda)\tau} - q(\eta+\lambda) \qquad (52)\\
&= (1-q)\lambda\log\frac{\lambda}{\lambda+q\eta} + q(\lambda+\eta)\log\frac{\lambda+\eta}{\lambda+q\eta} \\
&\quad - (\lambda+q\eta) - \frac{q(\lambda+\eta)^2+\lambda^2}{2}\tau\log\frac{1}{(\lambda+q\eta)\tau}. \qquad (53)
\end{aligned}$$

For our choices (36), for large enough T, the right-hand side of (53) is dominated by its second term, which can be bounded as

$$q(\lambda+\eta)\log\frac{\lambda+\eta}{\lambda+q\eta} \ge q\eta\log\frac{\eta}{\lambda+q\eta} \tag{54}$$

$$= \mathsf{T}^{-3/4}\log\frac{e^{\mathsf{T}/2}}{\lambda+\mathsf{T}^{-3/4}} \tag{55}$$

$$= \frac{\mathsf{T}^{1/4}}{2} - \mathsf{T}^{-3/4}\log(\lambda+\mathsf{T}^{-3/4}). \tag{56}$$

One can verify that the sum of the remaining terms on the right-hand side of (53) tends to $-\lambda$ as T grows to infinity; we omit the details. We thus have

$$\liminf_{\mathsf{T}\to\infty} \frac{\mathsf{E}\left[\frac{1}{\mathsf{T}}\log\frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}\big(\hat{\mathbf{Y}}|\mathbf{X}'\big)}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})}\right]}{\mathsf{T}^{1/4}} \ge \frac{1}{2}. \tag{57}$$

We next bound the variance of the random variable inside the P-$\liminf$ in (39). Since $(\mathbf{X}', \hat{\mathbf{Y}})$ is IID across different slots, we have

$$\begin{aligned}
&\mathsf{Var}\left[\frac{1}{\mathsf{T}}\log\frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}\big(\hat{\mathbf{Y}}|\mathbf{X}'\big)}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})}\right] \\
&= \frac{1}{\mathsf{T}^2}\cdot\frac{\mathsf{T}}{\tau}\mathsf{Var}\left[\log\frac{P_{\hat{Y}|X'}(\hat{Y}|X')}{P_{\hat{Y}}(\hat{Y})}\right] \qquad (58)\\
&\le \frac{1}{\mathsf{T}\tau}\mathsf{E}\left[\left(\log\frac{P_{\hat{Y}|X'}(\hat{Y}|X')}{P_{\hat{Y}}(\hat{Y})}\right)^2\right] \qquad (59)\\
&= \frac{1}{\mathsf{T}\tau}\sum_{\substack{x'\in\{0,\eta\}\\ \hat{y}\in\{0,1\}}} P_{X'\hat{Y}}(x',\hat{y})\left(\log\frac{P_{\hat{Y}|X'}(\hat{y}|x')}{P_{\hat{Y}}(\hat{y})}\right)^2. \qquad (60)
\end{aligned}$$

We simplify each summand on the right-hand side of (60) and provide an upper bound on it. For the summand with $x'=0$ and $\hat{y}=0$ we have

$$\begin{aligned}
&P_{X'\hat{Y}}(0,0)\left(\log\frac{P_{\hat{Y}|X'}(0|0)}{P_{\hat{Y}}(0)}\right)^2 \\
&\le 1\cdot\left(\log\frac{e^{-\lambda\tau}}{(1-q)e^{-\lambda\tau}+qe^{-(\lambda+\eta)\tau}}\right)^2 \qquad (61)\\
&= \big(\log(1-q(1-e^{-\eta\tau}))\big)^2. \qquad (62)
\end{aligned}$$

For the summand with $x'=0$ and $\hat{y}=1$ we have

$$\begin{aligned}
&P_{X'\hat{Y}}(0,1)\left(\log\frac{P_{\hat{Y}|X'}(1|0)}{P_{\hat{Y}}(1)}\right)^2 \\
&= (1-q)(1-e^{-\lambda\tau}) \\
&\quad\cdot\left(\log\frac{1-e^{-\lambda\tau}}{(1-q)(1-e^{-\lambda\tau})+q(1-e^{-(\lambda+\eta)\tau})}\right)^2 \qquad (63)\\
&\le \lambda\tau\left(\log\frac{\lambda\tau+q\eta\tau}{1-e^{-\lambda\tau}}\right)^2. \qquad (64)
\end{aligned}$$

For the summand with $x'=\eta$ and $\hat{y}=0$ we have

$$\begin{aligned}
&P_{X'\hat{Y}}(\eta,0)\left(\log\frac{P_{\hat{Y}|X'}(0|\eta)}{P_{\hat{Y}}(0)}\right)^2 \\
&= qe^{-(\lambda+\eta)\tau}\left(\log\frac{e^{-(\lambda+\eta)\tau}}{(1-q)e^{-\lambda\tau}+qe^{-(\lambda+\eta)\tau}}\right)^2 \qquad (65)\\
&\le q(\eta\tau)^2. \qquad (66)
\end{aligned}$$

For the summand with $x'=\eta$ and $\hat{y}=1$ we have

$$\begin{aligned}
&P_{X'\hat{Y}}(\eta,1)\left(\log\frac{P_{\hat{Y}|X'}(1|\eta)}{P_{\hat{Y}}(1)}\right)^2 \\
&= q(1-e^{-(\lambda+\eta)\tau}) \\
&\quad\cdot\left(\log\frac{1-e^{-(\lambda+\eta)\tau}}{(1-q)(1-e^{-\lambda\tau})+q(1-e^{-(\lambda+\eta)\tau})}\right)^2 \qquad (67)\\
&\le q(\lambda+\eta)\tau\left(\log\frac{(\lambda+\eta)\tau}{1-e^{-\lambda\tau}}\right)^2. \qquad (68)
\end{aligned}$$

Plugging our choices (36) into (62), (64), (66), and (68), one can verify that, as T grows large, the right-hand side of (68) satisfies

$$\lim_{\mathsf{T}\to\infty}\frac{q(\lambda+\eta)\tau\left(\log\frac{(\lambda+\eta)\tau}{1-e^{-\lambda\tau}}\right)^2}{\mathsf{T}^{5/4}\tau} = \frac{1}{4}, \tag{69}$$

and dominates all the other summands. We then have

$$\limsup_{\mathsf{T}\to\infty}\frac{\mathsf{Var}\left[\frac{1}{\mathsf{T}}\log\frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}\big(\hat{\mathbf{Y}}|\mathbf{X}'\big)}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})}\right]}{\mathsf{T}^{1/4}} \le \frac{1}{4}. \tag{70}$$

Combining (57) and (70) and applying Chebyshev's inequality we have, for any finite a,

$$\lim_{\mathsf{T}\to\infty} \Pr\left(\frac{1}{\mathsf{T}}\log\frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}\big(\hat{\mathbf{Y}}\big|\mathbf{X}'\big)}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})}\le a\right)$$
$$\le \lim_{\mathsf{T}\to\infty} \frac{\mathsf{Var}\left[\frac{1}{\mathsf{T}}\log\frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}\big(\hat{\mathbf{Y}}\big|\mathbf{X}'\big)}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})}\right]}{\left(\mathsf{E}\left[\frac{1}{\mathsf{T}}\log\frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}\big(\hat{\mathbf{Y}}\big|\mathbf{X}'\big)}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})}\right]-a\right)^2} \tag{71}$$
$$\le \lim_{\mathsf{T}\to\infty} \frac{\frac{\mathsf{T}^{\frac{1}{4}}}{4}}{\left(\frac{\mathsf{T}^{\frac{1}{4}}}{2}-a\right)^2} = 0. \tag{72}$$

Hence (39) is infinity and, consequently, the proposed random coding scheme can achieve an arbitrarily large per-second rate. ∎

III. The Peak-Limited Case

Now consider the same continuous-time Poisson channel with dark current λ as in the previous section, but now we impose a peak-power constraint on the input: At every time t, the channel input must satisfy, with probability one,

$$X(t)\le \mathsf{A}, \tag{73}$$

where $\mathsf{A}>0$ is a given constant.

We modify the covertness constraint. Instead of (3), we now require, for all $\mathsf{T}>0$,

$$\mathscr{D}(P^{\mathsf{T}}\|Q^{\mathsf{T}})\le\delta \tag{74}$$

for some given constant $\delta>0$. This is because, as we shall see, we are now in the square-root-scaling scenario, and the exact scaling law depends on δ; adopting the constraint (3) will not allow us to demonstrate this dependence.

As in the previous section, we allow the transmitter and the receiver to use a random codebook by exploiting a shared secret key. Denote by $\mathsf{M}(\mathsf{T},\delta,\epsilon)$ the largest possible cardinality of the message set that can be communicated over the channel in T seconds such that (74) is satisfied, and that the average probability of a decoding error is no larger than ϵ.

Theorem 2: Over the above peak-limited Poisson channel,

$$\lim_{\epsilon\downarrow 0}\lim_{\mathsf{T}\to\infty}\frac{\log\mathsf{M}(\mathsf{T},\delta,\epsilon)}{\sqrt{\mathsf{T}}} = \sqrt{2\lambda\delta}\left(\frac{\mathsf{A}+\lambda}{\mathsf{A}}\log\frac{\mathsf{A}+\lambda}{\lambda}-1\right). \tag{75}$$

Proof: We first prove the converse part. The capacity of a peak- and average-power-limited Poisson channel is known. To use this capacity result to derive an upper bound on the left-hand side of (75), we derive an upper bound on the average input power from the covertness constraint.

Given a random code over $[0,\mathsf{T}]$ for a message set of cardinality M, let ρ denote its average input power:

$$\rho\triangleq\frac{1}{\mathsf{T}}\mathsf{E}\left[\int_0^{\mathsf{T}}X(t)\,\mathrm{d}t\right]. \tag{76}$$

By Fano's Inequality and the Data-Processing Inequality [17] we have

$$(1-\epsilon')\log\mathsf{M}\le \mathrm{I}(X_0^{\mathsf{T}};Y_0^{\mathsf{T}}), \tag{77}$$

where ϵ' tends to zero when ϵ tends to zero. In the above, X_0^{T} denotes the waveform $X(t)$, $t\in[0,\mathsf{T}]$, and similarly for Y_0^{T}. The maximum possible value for $\mathrm{I}(X_0^{\mathsf{T}};Y_0^{\mathsf{T}})$ under peak-power constraint A and average-power constraint ρ is a classic result [11], [12]:

$$\mathrm{I}(X_0^{\mathsf{T}};Y_0^{\mathsf{T}})$$
$$\le \mathsf{T}\cdot\max_{p\in[0,\frac{\rho}{\mathsf{A}}]}\big\{p(\mathsf{A}+\lambda)\log(\mathsf{A}+\lambda)+(1-p)\lambda\log\lambda$$
$$-(p\mathsf{A}+\lambda)\log(p\mathsf{A}+\lambda)\big\}, \tag{78}$$

which can be weakened to

$$\mathrm{I}(X_0^{\mathsf{T}};Y_0^{\mathsf{T}})\le\mathsf{T}\rho\cdot\left(\frac{\mathsf{A}+\lambda}{\mathsf{A}}\log\frac{\mathsf{A}+\lambda}{\lambda}-1\right). \tag{79}$$

We next derive a bound on ρ from the covertness criterion (74). The relative entropy $\mathscr{D}(P^{\mathsf{T}}\|Q^{\mathsf{T}})$ can be computed as follows. For the distribution P^{T}, one can define the *conditional intensity function* at time $t\in[0,\mathsf{T}]$ as

$$\alpha_t(y_0^t)\triangleq\lim_{\Delta\downarrow 0}\frac{\Pr(Y_{t+\Delta}-Y_t\ge 1|Y_0^t=y_0^t)}{\Delta}. \tag{80}$$

At any y_0^{T}, the Radon-Nikodym derivative between P^{T} and Q^{T} is given by [20, Eq. (19.125)]

$$\frac{\mathrm{d}P^{\mathsf{T}}}{\mathrm{d}Q^{\mathsf{T}}}\big(y_0^{\mathsf{T}}\big)$$
$$=\exp\left\{\int_0^{\mathsf{T}}\log\frac{\alpha_t(y_0^t)}{\lambda}\,\mathrm{d}y_0^t-\int_0^{\mathsf{T}}(\alpha_t(y_0^t)-\lambda)\,\mathrm{d}t\right\}, \tag{81}$$

so

$$\mathscr{D}\big(P^{\mathsf{T}}\big\|Q^{\mathsf{T}}\big)$$
$$=\mathsf{E}_{P^{\mathsf{T}}}\left[\int_0^{\mathsf{T}}\left(\alpha_t(Y_0^t)\log\frac{\alpha_t(Y_0^t)}{\lambda}-\alpha_t(Y_0^t)+\lambda\right)\mathrm{d}t\right]. \tag{82}$$

Using Jensen's Inequality [17] and noting that

$$\frac{1}{\mathsf{T}}\mathsf{E}\left[\int_0^{\mathsf{T}}\alpha_t(Y_0^t)\,\mathrm{d}t\right]=\frac{1}{\mathsf{T}}\mathsf{E}\left[\int_0^{\mathsf{T}}\big(X(t)+\lambda\big)\,\mathrm{d}t\right]=\rho+\lambda, \tag{83}$$

we obtain from (74) and (82):

$$(\lambda+\rho)\log\frac{\lambda+\rho}{\lambda}-\rho\le\frac{\delta}{\mathsf{T}}. \tag{84}$$

This requires that, as $\mathsf{T}\to\infty$, ρ must approach zero; more specifically, since the left-hand side of (84) behaves like $\frac{\rho^2}{2\lambda}$ for small ρ, we need

$$\limsup_{\mathsf{T}\to\infty}\frac{\mathsf{T}\rho^2}{2\lambda}\le\delta. \tag{85}$$

Combining (77), (79), and (85) proves the converse part of the theorem.

We next prove the achievability part. To this end, we use Scheme 1 with the parameters

$$\tau = \mathsf{T}^{-2} \tag{86a}$$

$$\eta = \mathsf{A} \tag{86b}$$

$$q = \frac{1-\gamma}{\mathsf{A}}\sqrt{\frac{2\lambda\delta}{\mathsf{T}}}, \tag{86c}$$

where $\gamma > 0$ will be chosen to approach zero later on.

To show that this random code is covert, we apply Proposition 1 with the parameters (86) to obtain

$$\limsup_{\mathsf{T}\to\infty} \mathscr{D}(P^{\mathsf{T}} \| Q^{\mathsf{T}}) \le (1-\gamma)^2\delta. \tag{87}$$

Therefore, for any $\gamma > 0$, (74) is satisfied for large enough T.

To characterize the rates achievable using the above scheme, we again map the output within each slot to a binary random variable as in (37), yielding the transition probabilities

$$\Pr(\hat{Y} = 1 | X' = 0) = 1 - e^{-\lambda\tau} \tag{88a}$$

$$\Pr(\hat{Y} = 1 | X' = \mathsf{A}) = 1 - e^{-(\mathsf{A}+\lambda)\tau}. \tag{88b}$$

Again using the information-spectrum method [18], [19] we can lower-bound the left-hand side of (75) by

$$P\text{-}\liminf_{\mathsf{T}\to\infty} \left\{ \frac{1}{\sqrt{\mathsf{T}}} \log \frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}(\hat{\mathbf{Y}}|\mathbf{X}')}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})} \right\}. \tag{89}$$

For the random variable inside the P-lim inf we have

$$\mathsf{E}\left[\frac{1}{\sqrt{\mathsf{T}}} \log \frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}(\hat{\mathbf{Y}}|\mathbf{X}')}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})} \right] = \frac{\sqrt{\mathsf{T}}}{\tau} \mathrm{I}(X'; \hat{Y}). \tag{90}$$

We omit some details to claim

$$\lim_{\mathsf{T}\to\infty} \frac{\sqrt{\mathsf{T}}}{\tau} \mathrm{I}(X'; \hat{Y}) = (1-\gamma)\sqrt{2\lambda\delta}\left(\frac{\mathsf{A}+\lambda}{\mathsf{A}} \log \frac{\mathsf{A}+\lambda}{\lambda} - 1 \right). \tag{91}$$

By bounding the variance of the random variable inside the P-lim inf in (89), where we again omit the details, we can then show

$$\begin{aligned} &P\text{-}\liminf_{\mathsf{T}\to\infty} \left\{ \frac{1}{\sqrt{\mathsf{T}}} \log \frac{P_{\hat{\mathbf{Y}}|\mathbf{X}'}(\hat{\mathbf{Y}}|\mathbf{X}')}{P_{\hat{\mathbf{Y}}}(\hat{\mathbf{Y}})} \right\} \\ &\quad = (1-\gamma)\sqrt{2\lambda\delta}\left(\frac{\mathsf{A}+\lambda}{\mathsf{A}} \log \frac{\mathsf{A}+\lambda}{\lambda} - 1 \right). \end{aligned} \tag{92}$$

The proof is completed by letting γ approach zero. ∎

Remark 1: The term $\left(\frac{\mathsf{A}+\lambda}{\mathsf{A}} \log \frac{\mathsf{A}+\lambda}{\lambda} - 1\right)$ on the right-hand side of (75) approaches zero as $\mathsf{A} \downarrow 0$, and behaves like $\log \mathsf{A}$ when A is large. Further note that the converse proof of Theorem 2 continues to hold when A is a function of T. Therefore, if we allow A to grow large with T (while keeping λ and δ constant), then, for large T,

$$\log \mathsf{M}(\mathsf{T}, \delta, \epsilon) \lesssim \sqrt{2\lambda\delta}\sqrt{\mathsf{T}} \log \mathsf{A}. \tag{93}$$

This implies that, in order to achieve a positive per-second rate, the peak input power must at least grow exponentially with $\sqrt{\mathsf{T}}$. Recall that, in our achievability scheme for the no-constraint case, we chose the peak input power to grow exponentially with T; see (36).

IV. The Band-Limited Case

In intensity-modulated optical communication, the input usually consists of rectangular pulses, and the "bandwidth" refers to the reciprocal of the smallest possible duration of an input rectangular pulse. Thus, to model a band-limited scenario, we require that the transmitter must send a sequence of rectangular pulses each of τ seconds, where τ is now a given constant and cannot be chosen to be arbitrarily small as in previous sections. We hence have a *discrete-time* Poisson channel with input alphabet $\mathbb{R}_0^+$, output alphabet $\mathbb{Z}_0^+$, and transition law

$$W(y|x) = e^{-(\mu+x)} \frac{(\mu+x)^y}{y!}, \qquad x \in \mathbb{R}_0^+, y \in \mathbb{Z}_0^+. \tag{94}$$

Here $\mu > 0$ is the discrete-time dark current, which equals $\lambda\tau$, with λ being the continuous-time dark current. We do *not* impose a peak constraint on the input X. The standard (non-covert) communication capacity of the discrete-time Poisson channel under a peak- or average-power constraint, or both, does not have a known closed-form expression; some bounds and asymptotic results can be found in [21]–[24].

Suppose the channel (94) is used n times (so the total communication time is $n\tau$ seconds). Let Q^n denote the output distribution when there is no input, i.e., Q^n describes n random variables that are IID according to the Poisson distribution of mean μ. Let P^n denote the output distribution when the transmitter is active: P^n is averaged over the message and the randomly generated codebook. The covertness criterion is

$$\mathscr{D}(P^n \| Q^n) \le \delta, \tag{95}$$

where $\delta > 0$ is a given constant. Let $\mathsf{M}(n, \delta, \epsilon)$ denote the largest possible cardinality of the message set for a (random) codebook over n channel uses such that (95) is satisfied, and that the average probability of a decoding error is no larger than ϵ. The following asymptotic upper bound shows that the number of nats that can be communicated covertly grows at most proportionally to $\sqrt{n}\log\log n$. Thus, in particular, the covert communication capacity of this channel is zero nats per channel use or, equivalently, zero nats per second in continuous time.

Theorem 3: For the above discrete-time Poisson channel,

$$\lim_{\epsilon\downarrow 0} \limsup_{n\to\infty} \frac{\log \mathsf{M}(n, \delta, \epsilon)}{\sqrt{n}\log\log n} \le \sqrt{8\mu\delta}. \tag{96}$$

Proof: As in the proof of the converse for Theorem 2, we use the covertness criterion to derive an upper bound on the average input power. We then apply this average-power bound to a previously obtained capacity upper bound for the discrete-time Poisson channel with an average-power constraint.

Consider any random code over n channel uses and for a message set of cardinality M. Following the same steps as [4, Eq. (13)], we have

$$\mathscr{D}(P^n \| Q^n) \ge n\, \mathscr{D}(\bar{P} \| \bar{Q}), \tag{97}$$

where $\bar{P}$ denotes the average per-letter output distribution when the transmitter is sending a message, and $\bar{Q}$ that when

the transmitter is sending only zeros, i.e., $\bar{Q}$ is the Poisson distribution of mean μ. Let

$$\rho \triangleq \mathsf{E}_{\bar{P}}[Y] - \mu. \tag{98}$$

(Note that ρ is nonnegative.) For a fixed $\mathsf{E}_{\bar{P}}[Y]$, $\mathscr{D}(\bar{P}\|\bar{Q})$ is minimized by $\bar{P}$ also being a Poisson distribution; see [17, Problem 12.2]. Hence we have

$$\mathscr{D}(\bar{P}\|\bar{Q}) \geq \sum_{y=0}^{\infty} e^{-(\mu+\rho)} \frac{(\mu+\rho)^y}{y!} \log \frac{e^{-(\mu+\rho)} \frac{(\mu+\rho)^y}{y!}}{e^{-\mu} \frac{\mu^y}{y!}} \tag{99}$$

$$= \sum_{y=0}^{\infty} e^{-(\mu+\rho)} \frac{(\mu+\rho)^y}{y!} \left(y \log \frac{\mu+\rho}{\mu} - \rho \right) \tag{100}$$

$$= (\mu+\rho) \log \frac{\mu+\rho}{\mu} - \rho. \tag{101}$$

Recalling (97), the right-hand side of (101) must be less than or equal to $\frac{\delta}{n}$. This implies that ρ must tend to zero as $n \to \infty$. Further note that the right-hand side of (101) behaves like $\frac{\rho^2}{2\mu}$ for small ρ, so ρ must satisfy

$$\limsup_{n\to\infty} \frac{n\rho^2}{2\mu} \leq \delta. \tag{102}$$

By a standard argument using Fano's Inequality, the Data-Processing Inequality, and the chain rule, we have, for some ϵ' that tends to zero when $\epsilon \downarrow 0$,

$$(1-\epsilon') \log \mathsf{M} \leq \mathrm{I}(X^n; Y^n) \leq n\, \mathrm{I}(\bar{X}; \bar{Y}), \tag{103}$$

where, by slightly abusing notation, $\bar{X}$ denotes a random variable that has the average per-letter input distribution computed from the given random codebook, and $\bar{Y}$ denotes the corresponding output. Note that $\bar{Y}$ has distribution $\bar{P}$ and, by (94) and (98), $\mathsf{E}\left[\bar{X}\right] = \rho$. In the limit where ρ tends to zero, we have the following upper bound [23, Proposition 2]:

$$\limsup_{\rho\downarrow 0} \frac{\sup_{\mathsf{E}[\bar{X}]\leq\rho} \mathrm{I}(\bar{X}; \bar{Y})}{\rho \log\log \frac{1}{\rho}} \leq 2. \tag{104}$$

Combining (102), (103), and (104) and letting $n \to \infty$ and $\epsilon \downarrow 0$ yield the desired bound. ∎

Remark 2: We have not fully determined the scaling behavior of $\log \mathsf{M}(n, \delta, \epsilon)$, but it must lie between $\sqrt{n}$ and $\sqrt{n} \log\log n$. To see that it should grow at least proportionally to $\sqrt{n}$, we observe the following: If one imposes a peak constraint on X, then using the methods in [4] one can show that $\log \mathsf{M}(n, \delta, \epsilon)$ grows proportionally to $\sqrt{n}$, i.e., the square-root scaling law holds in the band- and peak-limited case.

We have not been able to find a scheme that allows $\log \mathsf{M}$ to grow faster than $\sqrt{n}$, let alone like $\sqrt{n} \log\log n$. Although the bound (104) is order-tight, to achieve it, [23] uses flash-signaling, where the transmitter sends large signals sparsely. These large signals are easy to detect, hence not suitable for covert communication. (Note that (102) is a necessary but not sufficient condition for covertness.)

V. Concluding Remarks

We have shown that the continuous-time Poisson channel with neither peak nor bandwidth constraint allows unbounded data rates for covert communication. The proposed scheme is however not very practical. Notably, the peak input power that we use grows exponentially with total communication time T. As we pointed out in Remark 1, in order to achieve a positive (not necessarily infinite) rate, the peak input power must at least grow exponentially with $\sqrt{\mathsf{T}}$.

The Poisson channel, being a model for optical communication, is related to the *bosonic channel*, a quantum model for optical communication. Covert communication over the bosonic channel was studied in [25]. Results on these two channels are however not directly comparable. First, when studying the bosonic channel, one usually considers a finite number of optical modes, so the channel is effectively in discrete time and hence not comparable to the continuous-time Poisson channel models. Second, in discrete time, the background noise on the bosonic channel is usually assumed to be in a thermal state, in which the number of photons follows a geometric distribution, whereas for the discrete-time Poisson channel, the background noise is modeled as having a Poisson distribution.

Some recent works [26], [27] have studied the Gaussian channel in continuous time. In [26] we have shown that, like the Poisson channel, the continuous-time Gaussian channel without bandwidth constraint also permits positive-rate covert communication. The mathematical models of the Gaussian channel and the Poisson channel are however fundamentally different, as are the optimal covert communication schemes for them. In the Poisson case, our chosen input signal consists of sparse narrow pulses; in the Gaussian case, we have shown it to be optimal to spread the input power evenly over both time and frequency. It is also interesting to note that an average-power constraint on the input would be inactive in the Poisson case, but active in the Gaussian case.

References

[1] L. Wang, "The continuous-time Poisson channel has infinite covert communication capacity," in *Proc. IEEE Int. Symp. Inform. Theory*, (Vail, CO, USA), June 17–22 2018.

[2] B. A. Bash, D. Goekel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Select. Areas Commun.*, vol. 31, pp. 1921–1930, Sept. 2013.

[3] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. IEEE Int. Symp. Inform. Theory*, (Istanbul, Turkey), July 10–15 2013.

[4] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inform. Theory*, vol. 62, pp. 3493–3503, June 2016.

[5] M. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inform. Theory*, vol. 62, pp. 2334–2354, May 2016.

[6] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. Inform. Theory Workshop (ITW)*, (Hobart, Australia), Nov. 2–5, 2014.

[7] S. Lee, R. Baxley, M. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE J. Select. Topics. Signal Processing.*, vol. 9, pp. 1195–1205, Oct. 2015.

[8] T. Sobers, B. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Trans. Wireless Comm.*, vol. 16, no. 9, pp. 6193–6206, 2017.

[9] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with channel-state information at the transmitter," *IEEE Trans. Inform. Forensics and Security*, vol. 13, no. 9, pp. 2310–2319, 2018.

[10] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and Its Appl.*, vol. 23, pp. 143–147, 1978.

[11] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inform. Theory*, vol. 26, pp. 710–715, Nov. 1980.

[12] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel — parts I and II," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1462–1471, Nov. 1988.

[13] A. D. Wyner, "The wiretap channel," *Bell System Techn. J.*, vol. 54, pp. 1355–1387, 1975.

[14] P. Mukherjee and S. Ulukus, "Covert bits through queues," in *IEEE Conf. Comm. and Network Security*, (Philadelphia, PA, USA), Oct. 2016.

[15] J. Hou and G. Kramer, "Effective secrecy: reliability, confusion and stealth," in *Proc. IEEE Int. Symp. Inform. Theory*, (Honolulu, HI, USA), June 29–July 4 2014.

[16] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco: Holden-Day, 1964.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, second ed., 2006.

[18] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.

[19] T. S. Han, *Information Spectrum Methods in Information Theory*. Springer Verlag, 2003.

[20] R. S. Lipster and A. N. Shiryaev, *Statistics of Random Processes II: Applications*. Berlin: Springer Verlag, 2nd ed., 2001.

[21] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," in *Proc. IEE*, vol. 137, pt. I (Communications, Speech and Vision), pp. 424–430, Dec. 1990.

[22] A. Lapidoth and S. M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Trans. Inform. Theory*, vol. 55, pp. 303–322, Jan. 2009.

[23] 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, pp. 3260–3272, June 2011.

[24] L. Wang and G. W. Wornell, "A refined analysis of the Poisson channel in the high-photon-efficiency regime," *IEEE Trans. Inform. Theory*, vol. 60, pp. 4299–4311, July 2014.

[25] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, "Quantum-secure covert communication on bosonic channels," *Nature Communications*, vol. 6, no. 8620, 2015.

[26] L. Wang, "On Gaussian covert communication in continuous time," *EURASIP J. App. Sig. Proc.*, Dec. 2019.

[27] Q. Zhang, M. R. Bloch, M. Bakshi, and S. Jaggi, "Undetectable radios: Covert communication under spectral mask constraints," in *Proc. IEEE Int. Symp. Inform. Theory*, (Paris, France), July 7–12 2019.

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. In the years 2011–2014 he was a Postdoctoral Associate at the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge, MA, USA. He is now a researcher (chargé de recherche) with CNRS, France, and is affiliated with ETIS laboratory in Cergy-Pontoise. His research interests include classical and quantum information theory, physical-layer security, and digital, in particular optical communications. Since 2019 he has been serving as an Associate Editor for Shannon Theory for the IEEE Transactions on Information Theory.