



# On the continuous-time Poisson channel with varying dark current known to the transmitter

Ligong Wang

## ► To cite this version:

Ligong Wang. On the continuous-time Poisson channel with varying dark current known to the transmitter. International Symposium on Information Theory, Jul 2016, Barcelona, Spain. pp.355 - 359, 10.1109/ISIT.2016.7541320 . hal-01390571

**HAL Id: hal-01390571**

**<https://hal.science/hal-01390571>**

Submitted on 13 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the Continuous-Time Poisson Channel with Varying Dark Current Known to the Transmitter

Ligong Wang

ETIS - ENSEA, Université de Cergy-Pontoise, CNRS UMR 8051

Cergy-Pontoise, France

ligong.wang@ensea.fr

**Abstract**—This paper considers a continuous-time Poisson channel whose dark current varies with time. The actual values of the dark current are revealed to the transmitter as channel-state information (CSI), either causally or noncausally. It is shown that, in the limit where the coherence time of the dark current tends to zero, the improvement on capacity provided by both causal and noncausal CSI vanishes linearly with the coherence time.

## I. INTRODUCTION

The Poisson channel is a model for optical communication links that use photodetectors at the receiving end. The continuous-time Poisson channel with infinite bandwidth subject to a peak-power constraint and possibly also an average-power constraint on its input signal has been extensively studied in the literature. The capacity of this channel is computed using different methods in [1]–[3]. Among them, [2] also determines the cut-off rate and [3] the error exponent of this channel. The Poisson channel with random or time-varying “dark current” was considered in [4]. The reliability function of the ideal Poisson channel with noiseless feedback is determined in [5]. Extension of the continuous-time Poisson channel model to multiple-user settings has also been studied [6]–[8]. In particular, [8] considers the peak-limited continuous-time Poisson channel with spurious photon counts at the receiver, where the exact positions of these spurious counts are known to the transmitter as *channel-state information (CSI)* before transmission starts. It shows that, in terms of capacity, the transmitter is able to completely cancel the influence of these spurious counts, i.e., that the capacity is the same as when there are no spurious counts.

Like [8], the current work also concerns the Poisson channel with CSI at the transmitter. However, while in [8] the CSI is the exact positions of the spurious counts, here we consider CSI that is the time-varying *intensity* of the background light that generates these spurious counts. Hence, effectively, we are looking at a Poisson channel whose *dark current* varies with time, where the values of the dark current are known to the transmitter as CSI. Compared to the one of [8], ours is perhaps a more realistic model for free-space or fiber optical communications. In such applications, the intensity of background radiation may be detected or predicted, but the exact positions of photon counts due to background radiation cannot be known in advance, due to their intrinsic quantum uncertainty.

As we shall see, the different types of CSI considered in [8] and in the current work lead to drastically different capacity results. Indeed, the main result of this paper is that, if the dark current varies very fast, i.e., if it has a very short *coherence time*, then the benefit of both causal and noncausal CSI at the transmitter is small. Specifically, if the dark current can change after every  $\tau$  seconds, then the benefit of (causal or noncausal) CSI at the transmitter is at most  $O(\tau)$  nats per second. In contrast, if the values of the dark current are known to the receiver, then this knowledge always helps to increase the capacity, no matter how short the coherence time is.

The current work is closely related to [4], which also studies Poisson channels with varying dark currents. In [4], the dark current is assumed to be either deterministic or random and unknown to either the transmitter or the receiver, and in the former case a closed-form capacity formula is derived. When an average-power constraint is present, computing this capacity becomes a problem of optimal power allocation, which is further addressed in [9]. In our setting, where the dark current is random and known only to the transmitter, the capacity seems more difficult to compute. In fact, within this work we are unable to provide an explicit capacity formula, but we derive upper bounds on capacity that are sufficiently tight in the regime of interest.

Single-letter capacity formulas for channels with causal and with noncausal CSI at the transmitter (but not at the receiver) are determined by Shannon [10] and by Gel’fand and Pinsker [11], respectively; see also [12]. Several examples of channels with noncausal CSI have been solved, before and after the discovery of Gel’fand and Pinsker’s formula. In most of these solved examples [13]–[15], including [8], capacity with noncausal CSI at the transmitter equals the capacity with CSI at both transmitter and receiver, and also equals the capacity of the same channel without states. The current work provides an example where this is not the case. Indeed, although we do not provide a capacity formula, our upper bound on the capacity with noncausal CSI at the transmitter shows that, when the coherence time of the dark current is sufficiently short, this capacity is strictly less than the capacity where receiver also has CSI.

The rest of this paper is arranged as follows. Section II formulates the problem and discusses some preliminary results. Section III studies the case where the transmitter has causal CSI, and shows that such CSI improves capacity by no more

than  $O(\tau)$ , where  $\tau$  is the coherence time according to which the dark current varies. Section IV shows that the capacity difference between noncausal and causal CSI is also like  $O(\tau)$ . Section V concludes the paper with some discussions. We provide the main ideas and key steps of the proofs. A full-length paper with complete proofs is under preparation.

## II. PROBLEM FORMULATION AND PRELIMINARIES

We consider a continuous-time Poisson channel whose input signal is a waveform on  $\mathbb{R}_0^+$  subject to peak-power constraint  $A$ . Thus, for every  $t \in \mathbb{R}_0^+$ ,

$$0 \leq X(t) \leq A \quad \text{with probability 1.} \quad (1)$$

(We do not impose an average-power constraint on the input.) The channel is affected by background noise which is a Poisson process of random, time-varying intensity  $\Lambda(t)$ ,  $t \in \mathbb{R}_0^+$ ; i.e.,  $\Lambda(\cdot)$  is the time-varying dark current. The output  $Y(t)$ ,  $t \in \mathbb{R}_0^+$ , is a doubly-stochastic Poisson process (Cox process): conditional on the input signal being  $x(\cdot)$  and the dark current being  $\lambda(\cdot)$ , the output is a Poisson process whose time- $t$  intensity equals  $x(t) + \lambda(t)$ .

In this work we are interested in the regime where  $\Lambda(\cdot)$  varies very fast. We assume that it varies in a block-fading manner as follows. Divide  $\mathbb{R}_0^+$  into slots of  $\tau$  seconds long:  $[0, \tau)$ ,  $[\tau, 2\tau)$ ,  $\dots$ , then  $\Lambda(t)$  remains constant within each one of these slots, while its values in different slots are independently distributed. We are interested in the regime where  $\tau$  is small. For simplicity, but without losing too much engineering insight, we assume that every  $\Lambda(t)$ ,  $t \in \mathbb{R}_0^+$ , takes only two values: 0 with probability  $(1 - q)$  and  $n$  with probability  $q$ , where  $n > 0$  and  $q \in (0, 1)$ .

### A. Some Notation

Let  $S$  be a binary random variable taking the values 0 and 1 with probabilities  $1 - q$  and  $q$ , respectively. We use  $S$  to characterize the dark current in the time slot  $[0, \tau)$ : if  $S = 0$  then  $\Lambda(t) = 0$  for all  $t \in [0, \tau)$ ; otherwise  $\Lambda(t) = n$  for all  $t \in [0, \tau)$ . Let  $\mathbf{x}$  denote an input waveform on  $[0, \tau)$ , and let  $\mathcal{X}$  denote the set of all admissible input waveforms satisfying (1). We describe the output signal on  $[0, \tau)$ , which is a point pattern, by a (possibly empty) set  $\mathbf{y} = \{t_1, \dots, t_k\} \subset [0, \tau)$  where each  $t_i \in \mathbf{y}$  corresponds to the position of a point, i.e., a photon count.<sup>1</sup> Let  $\mathcal{Y}$  denote the set of all finite subsets of  $[0, \tau)$ . Further, let  $\mathbf{u}$  denote a mapping from  $\{0, 1\}$  to  $\mathcal{X}$ , which maps  $s$  to  $\mathbf{u}_s(t)$ ,  $t \in [0, \tau)$ , the latter also written as  $\mathbf{u}_s$ ; note that  $\mathbf{u}_s \in \mathcal{X}$ . Let  $\mathcal{U}$  denote the set of all such mappings.

We use a letter like  $\mathbb{P}$  to denote a probability distribution,<sup>2</sup> and sometimes add subscripts to it. In particular,  $\mathbb{P}_S$  denotes a distribution on  $\{0, 1\}$ ,  $\mathbb{P}_{\mathbf{X}}$  one on  $\mathcal{X}$ , and  $\mathbb{P}_{\mathbf{U}}$  one on  $\mathcal{U}$ . We

<sup>1</sup>In the language of Stochastic Processes, a point process is usually characterized using its *counting function*, which is equivalent to our description as long as the probability that two counts occur simultaneously is zero.

<sup>2</sup>For simplicity of notation, we shall not define  $\sigma$ -algebras for probability measures. This should not cause any confusion within the scope of this work. However, it should be kept in mind that, in general, our “distributions” are indeed probability measures and not necessarily probability mass functions or probability density functions.

use  $\otimes$  to denote products of distributions. For example, when  $S$  and  $\mathbf{X}$  are generated independently according to  $\mathbb{P}_S$  and  $\mathbb{P}_{\mathbf{X}}$ , respectively, their joint distribution is  $\mathbb{P}_S \otimes \mathbb{P}_{\mathbf{X}}$ . Further, we use  $\mathbb{V}_{\mathbf{U}|S}$  to denote a “conditional distribution” of  $\mathbf{U}$  given  $S$ . (More precisely, it is a stochastic kernel with source  $\{0, 1\}$  and target  $\mathcal{U}$ , together with the  $\sigma$ -algebras defined on these sets.) We use  $\mathbb{W}$  to denote the transition law (again a stochastic kernel) of the Poisson channel on  $[0, \tau)$ :  $\mathbb{W}(\cdot|\mathbf{x}, s)$  is the conditional distribution for  $\mathbf{Y}$  given that  $\mathbf{X} = \mathbf{x}$  and  $S = s$ . Note that, because every  $\mathbf{u}$  is a (deterministic) mapping from  $\{0, 1\}$  to  $\mathcal{X}$ , the channel law  $\mathbb{W}$  also defines the law of the “induced” channel with input  $\mathbf{u}$ , state  $s$ , and output  $\mathbf{y}$ ; denote the law of the induced channel by  $\widetilde{\mathbb{W}}$ .

### B. CSI Settings and Capacity Formulas

Consider three settings for this channel: without CSI, with causal CSI, and with noncausal CSI.

*Case 1: CSI is available to neither transmitter nor receiver.*

In this case, an encoder working at blocklength  $T$  seconds maps a message to an input signal  $x(t)$ ,  $t \in [0, T]$ , and a corresponding decoder maps the output signal on  $[0, T]$  back to a message. The capacity in this case, denoted by  $C_{\text{NoCSI}}(A, n, q, \tau)$ , is defined as the maximum rate (in nats per second) at which information can be transmitted such that the probability of a decoding error, computed for a uniformly chosen message, can be made arbitrarily small as  $T$  tends to infinity.<sup>3</sup>

This channel can be thought of as a memoryless discrete-time channel with uncountable input and output alphabets, which are respectively  $\mathcal{X}$  and  $\mathcal{Y}$  defined in Section II-A. Thus one can generalize the standard capacity formula [16] to obtain

$$C_{\text{NoCSI}}(A, n, q, \tau) = \frac{1}{\tau} \sup I(\mathbf{X}; \mathbf{Y}) \quad (2)$$

with supremum over distributions of the form  $(\mathbb{P}_S \otimes \mathbb{P}_{\mathbf{X}}) \mathbb{W}$ .

*Case 2: Causal CSI is available to transmitter only.*

In this case, to choose the input value  $x(t)$  at time  $t$ , the encoder looks at both the message and the past and current values of the dark current,  $\lambda(s)$ ,  $s \in [0, t]$ .<sup>4</sup> The decoder is of the same structure as in Case 1. Denote the capacity by  $C_{\text{Cau}}(A, n, q, \tau)$ , which is defined similarly as in Case 1. Again, one can view this channel as a discrete-time channel with input and output alphabets  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively, and with state  $S$ . One can generalize the result of [10] to show

$$C_{\text{Cau}}(A, n, q, \tau) = \frac{1}{\tau} \sup I(\mathbf{U}; \mathbf{Y}) \quad (3)$$

with supremum over distributions of the form  $(\mathbb{P}_S \otimes \mathbb{P}_{\mathbf{U}}) \widetilde{\mathbb{W}}$ .

*Case 3: Noncausal CSI is available to transmitter only.*

In this case, the encoder maps the message and the dark current  $\lambda(t)$ ,  $t \in [0, T]$ , to the input waveform  $x(t)$ ,  $t \in [0, T]$ .

<sup>3</sup>We consider average error probabilities throughout this paper. However, in some cases, including the no-CSI case, capacity does not change when one considers maximum error probability instead.

<sup>4</sup>See [5] for a formal treatment of causality in continuous time.

Denote the capacity by  $C_{\text{NonCau}}(A, n, q, \tau)$ . Generalizing the result of [11] we have

$$C_{\text{NonCau}}(A, n, q, \tau) = \frac{1}{\tau} (\sup I(\mathbf{U}; \mathbf{Y}) - I(\mathbf{U}; S)) \quad (4)$$

with supremum over distributions of the form  $\mathbb{P}_S \mathbb{V}_{\mathbf{U}|S} \widetilde{\mathbb{W}}$ .

### C. Some Preliminary Results

We introduce two more quantities. Let  $C_{\text{Both}}(A, n, q, \tau)$  denote the capacity when CSI is available to both transmitter and receiver,<sup>5</sup> and let  $C_{\text{Uni}}(A, \nu)$  denote the capacity of the Poisson channel with peak-power constraint  $A$  and constant dark current  $\nu$ . Both quantities are readily known. Indeed,  $C_{\text{Uni}}(A, \nu)$  is given by [3, Theorem 1]:

$$C_{\text{Uni}}(A, \nu) = \max_p \left( p(A + \nu) \log \frac{A + \nu}{pA + \nu} + (1 - p)\nu \log \frac{\nu}{pA + \nu} \right) \quad (5)$$

where the maximum is achieved by

$$p^* = \frac{(1 + \frac{\nu}{A})^{1 + \frac{\nu}{A}}}{(\frac{\nu}{A})^{\frac{\nu}{A}} \cdot e} - \frac{\nu}{A}. \quad (6)$$

As for  $C_{\text{Both}}(A, n, q, \tau)$ , a simple time-sharing argument shows that (see also [4])

$$C_{\text{Both}}(A, n, q, \tau) = (1 - q)C_{\text{Uni}}(A, 0) + qC_{\text{Uni}}(A, n). \quad (7)$$

The various capacities introduced so far can be ordered as follows.

*Proposition 1:* For all  $A, n, q$ , and  $\tau$ ,

$$C_{\text{Both}}(A, n, q, \tau) \geq C_{\text{NonCau}}(A, n, q, \tau) \geq C_{\text{Cau}}(A, n, q, \tau) \geq C_{\text{NoCSI}}(A, n, q, \tau) \geq C_{\text{Uni}}(A, qn). \quad (8)$$

*Proof Sketch:* The first three inequalities in (8) follow immediately from the fact that more knowledge never harms capacity. For the last inequality, consider the following scheme. Transmitter and receiver first discretize the channel as in [3]: they divide the interval  $[0, T]$  into small slots of  $\Delta$  seconds, and approximate each slot by a binary-input binary-output channel. They then use some common randomness (which does not increase capacity in point-to-point settings) to permute these slots. After permutation, they forget how the slots have been permuted and use the new sequence of binary channels to communicate. One can verify that, as  $\Delta$  tends to zero, this new channel's law approaches the law of the binary channel resulting from applying the same discretization (but without permutation) on the Poisson channel with constant dark current  $qn$ . Finally, on the Poisson channel with constant dark current, this discretization approach achieves capacity in the limit where  $\Delta$  tends to zero [3]. ■

One can check from (5) and (7) that  $C_{\text{Both}}(A, n, q, \tau)$  is strictly larger than  $C_{\text{Uni}}(A, qn)$ , and that their difference does not depend on  $\tau$ . Thus, CSI that is available to both transmitter and receiver always helps to improve capacity, and the amount

<sup>5</sup>When CSI is available to both transmitter and receiver, capacity is not affected by whether it is known causally or noncausally.

of improvement is independent of  $\tau$ . In fact, one can show that even CSI that is available only to the receiver helps to increase capacity by an amount that is independent of  $\tau$ . In contrast, as we show in the rest of this paper, all the other three capacities in the chain of inequalities (8) tend to  $C_{\text{Uni}}(A, qn)$  as  $\tau$  tends to zero. In fact, we show that

$$C_{\text{NonCau}}(A, n, q, \tau) - C_{\text{Uni}}(A, qn) = O(\tau), \quad (9)$$

where the usual notation  $O(\tau)$  denotes a function of  $A, n, q$ , and  $\tau$  satisfying

$$\lim_{\tau \downarrow 0} \frac{O(\tau)}{\tau} < \infty, \quad \text{for all } A, q, n. \quad (10)$$

Note that (8) implies that (9) still holds when we replace  $C_{\text{NonCau}}$  by either  $C_{\text{Cau}}$  or  $C_{\text{NoCSI}}$ .

We prove (9) in two steps. In Section III we show that  $C_{\text{Cau}}(A, n, q, \tau) - C_{\text{Uni}}(A, qn) = O(\tau)$ , and then in Section IV we show that  $C_{\text{NonCau}}(A, n, q, \tau) - C_{\text{Cau}}(A, n, q, \tau) = O(\tau)$ . Explicit bounds are given in the corresponding sections.

### III. THE CAUSAL CASE

In this section we show that

$$C_{\text{Cau}}(A, n, q, \tau) - C_{\text{Uni}}(A, qn) = O(\tau). \quad (11)$$

The fact that the benefit of causal CSI vanishes for small  $\tau$  is not surprising: it can be shown as a consequence to a result in [8] which says that, when the exact positions of the photon counts produced by a constant interference are known to the transmitter causally, such information does not help to increase capacity. However, to show that the benefit of causal CSI behaves like  $O(\tau)$ , we need an explicit upper bound.

*Proposition 2:* For all  $A, n, q$ , and  $\tau$ ,

$$C_{\text{Cau}}(A, n, q, \tau) \leq C_{\text{Uni}}(A, qn) + 4(A + n)^2 \tau. \quad (12)$$

We try to provide some intuition to our proof of Proposition 2. To use the formula (3), consider a mapping  $\mathbf{u}$  that maps 0 to  $\mathbf{u}_0$  and maps 1 to  $\mathbf{u}_1$ . Also consider a random input waveform that does not depend on the state: it is  $\mathbf{u}_0$  with probability  $1 - q$  and is  $\mathbf{u}_1$  with probability  $q$ . Observe that the output distributions induced by these two input strategies are very close to each other. Indeed, simple calculation shows that the probabilities of the event of no photon count at the output when  $\mathbf{u}$  is used and when the random input waveform is used are approximately the same; their difference is on the order of  $\tau^2$ . The same is true for the probabilities of having exactly one photon count. The probabilities produced by the two input strategies differ in the dominant term only for the events of two or more photon counts at the output, while these probabilities are themselves on the order of  $\tau^2$  or higher. Hence the transmitter cannot do much better with causal CSI than what it can do without CSI. Our proof, albeit somewhat technical, is largely guided by this observation.

*Proof Sketch for Proposition 2:* We first generalize the duality bound [17], [18] to channels with causal CSI. For our channel, this bound becomes

$$C_{\text{Cau}}(A, n, q, \tau) \leq \frac{1}{\tau} \sup_{\mathbf{u}} D(\overline{\mathbb{W}}(\cdot | \mathbf{u}) \| \mathbb{Q}) \quad (13)$$

where  $\bar{\mathbb{W}}$  is the conditional distribution of  $\mathbf{Y}$  given  $\mathbf{U}$ :

$$\bar{\mathbb{W}}(\cdot|\mathbf{u}) = (1-q)\mathbb{W}(\cdot|\mathbf{u}_0, 0) + q\mathbb{W}(\cdot|\mathbf{u}_1, 1), \quad \mathbf{u} \in \mathcal{U}, \quad (14)$$

and  $\mathbb{Q}$  is any distribution on  $\mathcal{Y}$ . We choose  $\mathbb{Q}$  to be doubly-stochastic Poisson under which, with probability  $(1-q)$ ,  $\mathbf{Y}$  is a homogeneous Poisson process of intensity  $p^*A$ , and with probability  $q$ , it is a homogeneous Poisson process of intensity  $p^*A + n$ , on  $[0, \tau)$ , where  $p^*$  is given in (6). Note that this  $\mathbb{Q}$  would be the output distribution if we chose the input distribution to be the one that achieves  $C_{\text{Uni}}(A, qn)$ . Using [19, (19.125)] we can compute the Radon-Nikodym derivative

$$\frac{d\bar{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{y}) = \frac{(1-q)e^{-\|\mathbf{u}_0\|} \prod_{s \in \mathbf{y}} u_0(s) + qe^{-\|\mathbf{u}_1\| - n\tau} \prod_{s \in \mathbf{y}} (u_1(s) + n)}{(1-q)e^{-p^*A\tau} (p^*A)^{|\mathbf{y}|} + qe^{-(p^*A+n)\tau} (p^*A+n)^{|\mathbf{y}|}} \quad (15)$$

where  $|\cdot|$  denotes the cardinality of a set, and where  $\|\cdot\|$  denotes the  $L_1$  norm which, for a nonnegative waveform  $\mathbf{x}$ , equals

$$\|\mathbf{x}\| = \int_0^\tau x(t) dt. \quad (16)$$

We write the relative entropy in (13) as

$$D(\bar{\mathbb{W}}(\cdot|\mathbf{u})\|\mathbb{Q}) = \mathbb{E} \left[ \log \frac{d\bar{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{Y}) \right] \quad (17)$$

$$= \sum_{k=0}^{\infty} \bar{\mathbb{W}}(\{|\mathbf{Y}| = k\}|\mathbf{u}) \mathbb{E} \left[ \log \frac{d\bar{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{Y}) \middle| |\mathbf{Y}| = k \right]. \quad (18)$$

Using (15) we can express and upper-bound each summand in (18) separately (details omitted). We then arrive at the upper bound

$$D(\bar{\mathbb{W}}(\cdot|\mathbf{u})\|\mathbb{Q}) \leq \int_0^\tau \left( \bar{u}(t) - p^*A + (\bar{u}(t) + qn) \log \frac{\bar{u}(t) + qn}{p^*A + qn} \right) dt + 4(A+n)^2\tau^2. \quad (19)$$

One can check that the integrand in (19), for every  $t$ , is maximized at  $u(t) = 0$  and at  $u(t) = A$ , and equals  $C_{\text{Uni}}(A, qn)$ . Thus we have

$$D(\bar{\mathbb{W}}(\cdot|\mathbf{u})\|\mathbb{Q}) \leq \tau \cdot C_{\text{Uni}}(A, qn) + 4(A+n)^2\tau^2 \quad (20)$$

which, combined with (13), yields (12).  $\blacksquare$

#### IV. THE NONCAUSAL CASE

In this section we show that

$$C_{\text{NonCau}}(A, n, q, \tau) - C_{\text{Cau}}(A, n, q, \tau) = O(\tau). \quad (21)$$

Our bound is the following.

*Proposition 3:* For all  $A, n, q$ , and  $\tau$ ,

$$C_{\text{NonCau}}(A, n, q, \tau) \leq C_{\text{Cau}}(A, n, q, \tau) + \frac{(2A+n)^2}{8}\tau. \quad (22)$$

We again try to provide some intuition to our proof. Consider the formula (4). Note that, when  $\tau$  is small, the channel is “weak” in the sense that  $I(\mathbf{U}; \mathbf{Y})$  is small for all admissible joint distributions. Thus, in order for the right-hand side of (4) to be positive,  $I(\mathbf{U}; S)$  must also be small. In other words,  $\mathbf{U}$  can only depend on  $S$  “very weakly.” This means that  $\mathbb{P}_S \mathbb{V}_{\mathbf{U}|S}$  is close to the product distribution  $\mathbb{P}_S \otimes \mathbb{P}_{\mathbf{U}}$ , where  $\mathbb{P}_{\mathbf{U}}$  is the marginal of  $\mathbb{P}_S \mathbb{V}_{\mathbf{U}|S}$  on  $\mathbf{U}$ , but the latter distribution is admissible in the causal case (3). This is, roughly speaking, why the transmitter cannot do much better with noncausal CSI than with causal CSI.

*Proof Sketch for Proposition 3:* We use the formula (4). Fix

$$I(\mathbf{U}; S) = \alpha, \quad (23)$$

then by Pinsker’s inequality [17],

$$\delta(\mathbb{P}_S \mathbb{V}_{\mathbf{U}|S}, \mathbb{P}_S \otimes \mathbb{P}_{\mathbf{U}}) \leq \sqrt{\frac{\alpha}{2}}, \quad (24)$$

where  $\mathbb{P}_{\mathbf{U}}$  is the marginal distribution of  $\mathbb{P}_S \mathbb{V}_{\mathbf{U}|S}$  on  $\mathbf{U}$ , and where  $\delta(\cdot, \cdot)$  denotes the total variation distance.

We next consider the  $I(\mathbf{U}; \mathbf{Y})$  term in (4). Let  $\mathbb{R}$  be the distribution on  $\mathcal{Y}$  that achieves the capacity in the causal-CSI case. Again using (a generalized version of) the duality bound we have

$$I(\mathbf{U}; \mathbf{Y}) \leq \mathbb{E} \left[ \log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{U})}{d\mathbb{R}}(\mathbf{Y}) \right], \quad (25)$$

where  $\widehat{\mathbb{W}}$  denotes the conditional distribution of  $\mathbf{Y}$  given  $\mathbf{U}$  and is given by

$$\widehat{\mathbb{W}}(\cdot|\mathbf{u}) = \mathbb{P}_{S|\mathbf{U}}(0|\mathbf{u})\mathbb{W}(\cdot|\mathbf{u}_0, 0) + \mathbb{P}_{S|\mathbf{U}}(1|\mathbf{u})\mathbb{W}(\cdot|\mathbf{u}_0, 1) \quad (26)$$

for every  $\mathbf{u} \in \mathcal{U}$ , where  $\mathbb{P}_{S|\mathbf{U}}$  is the conditional distribution of  $S$  given  $\mathbf{U}$  according to the joint distribution  $\mathbb{P}_S \mathbb{V}_{\mathbf{U}|S}$ . Note that  $\widehat{\mathbb{W}}$  is different from  $\bar{\mathbb{W}}$  defined in (14). We can continue (25) as

$$I(\mathbf{U}; \mathbf{Y}) \leq \mathbb{E} \left[ \log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{U})}{d\bar{\mathbb{W}}(\cdot|\mathbf{U})}(\mathbf{Y}) \right] + \mathbb{E} \left[ \log \frac{d\bar{\mathbb{W}}(\cdot|\mathbf{U})}{d\mathbb{R}}(\mathbf{Y}) \right] \quad (27)$$

$$\leq \mathbb{E} \left[ \log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{U})}{d\bar{\mathbb{W}}(\cdot|\mathbf{U})}(\mathbf{Y}) \right] + \tau \cdot C_{\text{Cau}}(A, n, q, \tau), \quad (28)$$

where the last inequality follows because  $\mathbb{R}$  is the capacity-achieving output distribution in the causal case.

We fix any  $\mathbf{u} \in \mathcal{U}$  and look at

$$\mathbb{E} \left[ \log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{u})}{d\bar{\mathbb{W}}(\cdot|\mathbf{u})}(\mathbf{Y}) \right] = D(\widehat{\mathbb{W}}(\cdot|\mathbf{u})\|\bar{\mathbb{W}}(\cdot|\mathbf{u})). \quad (29)$$

First consider the case where

$$\mathbb{P}_{S|\mathbf{U}}(1|\mathbf{u}) \leq \mathbb{P}_S(1) = q. \quad (30)$$

In this case we can write

$$\widehat{\mathbb{W}}(\cdot|\mathbf{u}) = \frac{\mathbb{P}_{S|\mathbf{U}}(1|\mathbf{u})}{q} \cdot \bar{\mathbb{W}}(\cdot|\mathbf{u}) + \left( 1 - \frac{\mathbb{P}_{S|\mathbf{U}}(1|\mathbf{u})}{q} \right) \mathbb{W}(\cdot|\mathbf{u}_0, 0). \quad (31)$$

Then, by the convexity of  $D(\cdot\|\cdot)$ ,

$$D\left(\widehat{\mathbb{W}}(\cdot|\mathbf{u})\|\overline{\mathbb{W}}(\cdot|\mathbf{u})\right) \leq \left(1 - \frac{\mathbb{P}_{S|\mathbf{U}}(1|\mathbf{u})}{q}\right) D\left(\mathbb{W}(\cdot|\mathbf{u}_0, 0)\|\overline{\mathbb{W}}(\cdot|\mathbf{u})\right). \quad (32)$$

We can write down the Radon-Nikodym derivative between  $\mathbb{W}(\cdot|\mathbf{u}_0, 0)$  and  $\overline{\mathbb{W}}(\cdot|\mathbf{u})$ , again using [19, (19.125)]. After some calculations, we arrive at

$$D\left(\mathbb{W}(\cdot|\mathbf{u}, 0)\|\overline{\mathbb{W}}(\cdot|\mathbf{u})\right) \leq q(2A + n)\tau. \quad (33)$$

Thus, when (30) holds,

$$D\left(\widehat{\mathbb{W}}(\cdot|\mathbf{u})\|\overline{\mathbb{W}}(\cdot|\mathbf{u})\right) \leq (q - \mathbb{P}_{S|\mathbf{U}}(1|\mathbf{u}))(2A + n)\tau. \quad (34)$$

Similarly, one can show that, when (30) does not hold,

$$D\left(\widehat{\mathbb{W}}(\cdot|\mathbf{u})\|\overline{\mathbb{W}}(\cdot|\mathbf{u})\right) \leq (\mathbb{P}_{S|\mathbf{U}}(1|\mathbf{u}) - q)(2A + n)\tau. \quad (35)$$

We can summarize (34) and (35) into

$$D\left(\widehat{\mathbb{W}}(\cdot|\mathbf{u})\|\overline{\mathbb{W}}(\cdot|\mathbf{u})\right) \leq \delta(\mathbb{P}_{S|\mathbf{U}}(\cdot|\mathbf{u}), \mathbb{P}_S)(2A + n)\tau. \quad (36)$$

Now we have

$$\begin{aligned} & \mathbb{E}\left[\log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{U})}{d\overline{\mathbb{W}}(\cdot|\mathbf{U})}(\mathbf{Y})\right] \\ &= \mathbb{E}\left[D\left(\widehat{\mathbb{W}}(\cdot|\mathbf{U})\|\overline{\mathbb{W}}(\cdot|\mathbf{U})\right)\right] \end{aligned} \quad (37)$$

$$\leq \mathbb{E}\left[\delta(\mathbb{P}_{S|\mathbf{U}}(\cdot|\mathbf{U}), \mathbb{P}_S)\right](2A + n)\tau \quad (38)$$

$$= \delta(\mathbb{P}_S \mathbb{V}_{\mathbf{U}|\mathbf{S}}, \mathbb{P}_S \otimes \mathbb{P}_{\mathbf{U}})(2A + n)\tau \quad (39)$$

$$\leq \sqrt{\frac{\alpha}{2}}(2A + n)\tau, \quad (40)$$

where the last inequality follows from (24). Combining (4), (23), (28), and (40) we have

$$\begin{aligned} & C_{\text{NonCau}}(A, n, q, \tau) \\ &\leq C_{\text{Cau}}(A, n, q, \tau) + \frac{1}{\tau} \left( \sqrt{\frac{\alpha}{2}}(2A + n)\tau - \alpha \right) \end{aligned} \quad (41)$$

$$\begin{aligned} &= C_{\text{Cau}}(A, n, q, \tau) \\ &\quad - \frac{1}{\tau} \left( \sqrt{\alpha} - \sqrt{\frac{1}{8}}(2A + n)\tau \right)^2 + \frac{(2A + n)^2}{8}\tau \end{aligned} \quad (42)$$

$$\leq C_{\text{Cau}}(A, n, q, \tau) + \frac{(2A + n)^2}{8}\tau, \quad (43)$$

completing the proof.  $\blacksquare$

Finally, we state an upper bound on the overall benefit of noncausal CSI in terms of channel capacity. It is obtained immediately by combining Propositions 2 and 3 and the last inequality in (8), and by slightly loosening the bounds.

*Proposition 4:* For any  $A$ ,  $n$ ,  $q$ , and  $\tau$ ,

$$C_{\text{NonCau}}(A, n, q, \tau) - C_{\text{NoCSI}}(A, n, q, \tau) \leq 5(A + n)^2\tau. \quad (44)$$

## V. CONCLUDING REMARKS

In this paper we have derived upper bounds on the capacity of the continuous-time Poisson channel with time-varying dark current whose values are known to the transmitter, either causally or noncausally, as CSI. The bounds show that the improvement in capacity from such CSI vanishes like  $O(\tau)$  as  $\tau$  tends to zero, where  $\tau$  is the coherence time according to which the dark current varies.

We have adopted a simple model for the dark current: it varies in a block-fading manner, and takes only two possible values. It is possible to extend our proof techniques to Poisson channels where the dark current can take multiple values and varies, e.g., in a stationary way. Furthermore, the intuition to our proofs suggests that one might be able to show similar results for other wide-band channels, such as wide-band Gaussian channels and “very noisy” channels [7].

## REFERENCES

- [1] Y. Kabanov, “The capacity of a channel of the Poisson type,” *Theory of Probability and Its Appl.*, vol. 23, pp. 143–147, 1978.
- [2] M. H. A. Davis, “Capacity and cutoff rate for Poisson-type channels,” *IEEE Trans. Inform. Theory*, vol. 26, pp. 710–715, Nov. 1980.
- [3] A. D. Wyner, “Capacity and error exponent for the direct detection photon channel — part I,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 1449–1461, Nov. 1988.
- [4] M. R. Frey, “Information capacity of the Poisson channel,” *IEEE Trans. Inform. Theory*, vol. 37, pp. 244–256, Mar. 1991.
- [5] A. Lapidoth, “On the reliability function of the ideal Poisson channel with noiseless feedback,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 491–503, Mar. 1993.
- [6] A. Lapidoth and S. Shamai (Shitz), “The Poisson multiple-access channel,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 488–501, Mar. 1998.
- [7] A. Lapidoth, I. E. Telatar, and R. Urbanke, “On wide-band broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 3250–3258, Dec. 2003.
- [8] S. I. Bross, A. Lapidoth, and L. Wang, “The Poisson channel with side information,” in *Proc. 47th Allerton Conf. Comm., Contr. and Comp.*, (Monticello, IL), September 30–October 2, 2009.
- [9] Ain-ul-Aisha, L. Lai, and Y. Liang, “Optimal power allocation for Poisson channels with time-varying background light,” *IEEE Trans. Commun.*, vol. 63, pp. 4327–4338, Nov. 2015.
- [10] C. E. Shannon, “Channels with side information at the transmitter,” *IBM J. Research and Development*, vol. 2, pp. 289–293, 1958.
- [11] S. I. Gel’fand and M. S. Pinsker, “Coding for channels with random parameters,” *Prob. Contr. and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [12] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [13] A. V. Kusnetsov and B. S. Tsybakov, “Coding in a memory with defective cells,” *Probl. Peredachi Inf.*, vol. 10, Apr.–June 1974.
- [14] C. D. Heegard and A. A. El Gamal, “On the capacity of computer memory with defects,” *IEEE Trans. Inform. Theory*, vol. 29, pp. 731–739, Sept. 1983.
- [15] M. H. M. Costa, “Writing on dirty paper,” *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.
- [16] C. E. Shannon, “A mathematical theory of communication,” *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.
- [17] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [18] 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, pp. 2426–2467, Oct. 2003.
- [19] R. S. Lipster and A. N. Shiryaev, *Statistics of Random Processes II: Applications*. Springer Verlag, 2nd ed., 2001.