

The Poisson Channel with Varying Dark Current Known to the Transmitter

Ligong Wang, *Member, IEEE*

Abstract—This paper studies the continuous-time Poisson channel whose dark current is random and may change for every τ -second time interval, where the actual values of the dark current are known to the transmitter as channel-state information (CSI). In the limit where τ tends to zero, the capacity gain provided by both causal and noncausal CSI is shown to vanish linearly with τ , so CSI at the transmitter provides almost no capacity improvement. The paper also considers a related problem of the state-dependent very noisy channel. In the “very noisy” limit, the capacity gain provided by noncausal CSI is shown to be same as that provided by causal CSI.

Index Terms—Channel-state information, Gel’fand-Pinsker problem, Poisson channel, Shannon strategy, very noisy channel.

I. INTRODUCTION

THE Poisson channel is a model for optical communication links where the transmitter modulates the intensity of the optical signal for transmission, and where the receiver performs direct detection using photodetectors. 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 was computed using different methods in [1]–[3]. The cut-off rate was computed in [2] (see also [4], [5]), and the error exponent in [3]. The reliability function of the ideal Poisson channel with noiseless feedback was determined in [6]. The Poisson channel with random or deterministic time-varying “dark current” was considered in [7] and later in [8]. A fading model where the dark current is fixed but the channel gain is time-varying was studied in [9]. Several multiple-user models were studied in [10]–[15]. Among them, [13] considers the peak-limited continuous-time Poisson channel with spurious photon arrivals at the receiver, where the exact positions of these spurious arrivals 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 arrivals, i.e., that the capacity is the same as when there are no spurious arrivals or, equivalently, as when these arrival positions are also known to the receiver.

Like [13], the current work also concerns the Poisson channel with CSI at the transmitter. However, while in [13] the CSI is the exact positions of the spurious arrivals, here

we consider CSI that is the time-varying *intensity* of the background light that generates these spurious arrivals. 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. As we shall see, the different types of CSI considered in [13] and in the current work lead to drastically different capacity results.

The main result of this paper is that, if the dark current varies very fast, then the benefit of both causal and noncausal CSI at the transmitter is small. Specifically, if the dark current can change after every τ seconds without memory, then the benefit of (causal or noncausal) CSI at the transmitter is at most $O(\tau)$ nats per second.

The current work is related to [7], which also studies Poisson channels with varying dark currents. In [7], 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 [8]. 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 do not provide an explicit capacity formula, but derive upper bounds on capacity that are sufficiently tight in the regime of interest.

Our work builds upon classic results on channels with CSI. Shannon [16] studied the discrete memoryless channel (DMC) affected by a random state that is independent and identically distributed (IID) across different channel uses, where the state is known to the transmitter causally. Gel’fand and Pinsker [17] studied the case where CSI is available to the transmitter noncausally. In both these settings [16], [17], the receiver does *not* have CSI. If CSI is available to both the transmitter and the receiver, then capacity is not affected by whether CSI at the transmitter is causal or noncausal, and it can be achieved with a simple rate-splitting scheme; see [18] and references therein. Some of these results have been generalized to broadcast channels [19]–[21].

Some examples of channels with CSI at the transmitter have been solved. One of them is computer memory (modeled by a perfect binary channel or a binary symmetric channel) with stuck-at faults [22], [23]. In this example, causal CSI does not increase capacity at all, whereas noncausal CSI increases it to the same capacity as when CSI is available on both (transmitter and receiver) sides. Another famous example is the additive white Gaussian noise channel with an additive state, known as the “dirty paper problem” [24], [25]. There, noncausal CSI

This work was presented in part at 2016 IEEE International Symposium of Information Theory (ISIT) in Barcelona, Spain.

The author is with ETIS—Université Paris Seine / Université de Cergy-Pontoise / ENSEA / CNRS, 95000 Cergy-Pontoise, France (e-mail: ligong.wang@ensea.fr).

again increases the capacity to that with CSI on both sides, hence the additive state can be completely canceled using only noncausal CSI at the transmitter. The capacity of this channel with causal CSI is unknown, and an upper bound was given in [26]. A third example is the above-mentioned Poisson channel with spurious arrivals [13]. On this channel, causal CSI does not increase capacity, whereas noncausal CSI, once again, can completely cancel the interference of the states. For a general DMC, an algorithm to numerically compute channel capacity with noncausal CSI was proposed in [27].

We note that, in most of these examples (with the exception of [26]), *the capacity upper bounds are trivial*. In particular, in all the above examples of channels with noncausal CSI, capacity upper bounds are obtained by revealing the CSI also to the receiver. In the current work, we demonstrate methods that can be used to prove nontrivial upper bounds on the capacity with CSI.

As a side result, we apply our proof techniques to another, simpler example, the *very noisy channel (VNC)* [12], [28], which is sometimes used as an abstract model for wideband channels. For the VNC, we show that, like for our Poisson model, noncausal CSI has little advantage in terms of capacity over causal CSI. Unlike the Poisson channel, however, causal CSI on the VNC may improve capacity over the no-CSI case.

The rest of this paper is arranged as follows. Section II introduces the Poisson channel model and provides some background. Section III presents the results on the Poisson channel, and provides some high-level intuition to our proof methods. Section IV contains the proof of the main result. Section V states and proves the result on the VNC. Section VI concludes the paper with some discussions. The Appendix proves a simple result that provides intuition to the Poisson channel with causal CSI.

II. CHANNEL MODEL AND PRELIMINARIES

A. Channel Model and Some Notation

We consider a continuous-time Poisson channel whose input signal must be a Lebesgue-measurable waveform on \mathbb{R}_0^+ subject to peak-power constraint \mathcal{A} : for every $t \in \mathbb{R}_0^+$,

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

We do not impose an average-power or a bandwidth constraint on the input. We assume that the channel is affected by background noise which has random, time-varying intensity $\Lambda(t)$, $t \in \mathbb{R}_0^+$; i.e., $\Lambda(\cdot)$ is the dark current. The output 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 $x(t) + \lambda(t)$. Let $Z(t)$, $t \in \mathbb{R}_0^+$, be the random counting function describing the output process, then, for any $t_0, \tau > 0$,

$$\begin{aligned} \Pr[Z(t_0 + \tau) - Z(t_0) = k | \Lambda(\cdot) = \lambda(\cdot), X(\cdot) = x(\cdot)] \\ = e^{-\sigma} \frac{\sigma^k}{k!}, \quad k \in \mathbb{Z}_0^+, \end{aligned} \quad (2)$$

where

$$\sigma = \int_{t_0}^{t_0 + \tau} (x(t) + \lambda(t)) dt. \quad (3)$$

Conditional on $X(\cdot) = x(\cdot)$ and $\Lambda(\cdot) = \lambda(\cdot)$, the increments of $Z(\cdot)$ over disjoint intervals are independent.

The Poisson channel is typically used to model optical communication with a direct-detection receiver. The input $X(t)$ describes the intensity emitted from the transmitter's laser or light-emitting diode at time t ; each unit jump in $Z(\cdot)$ —called a “point” or an “arrival”—describes the detection of one photon at the receiver's photodetector; and the dark current $\Lambda(\cdot)$ describes noise coming from the detector's thermal noise, background light, or light signal from interfering devices, etc.

To make the problem more tractable, we make some simplifying assumptions on the random time-varying dark current $\Lambda(\cdot)$. Divide \mathbb{R}_0^+ into slots of τ seconds long: $[0, \tau), [\tau, 2\tau), \dots$. We assume that $\Lambda(t)$ remains constant within each of these slots, while its values in different slots are IID. Further, 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 $q \in (0, 1)$.

The above model for $\Lambda(\cdot)$ is partially motivated by considering a situation where $\Lambda(\cdot)$ is an interfering signal generated by on-off modulation over τ -second rectangular waveforms. Such an interfering signal may be sent by the same transmitter but intended for another receiver (in a broadcast scenario), or sent by a nearby transmitter.

Since $\Lambda(\cdot)$ changes possibly every τ seconds without memory, we can think of this Poisson channel as a discrete-time memoryless channel, where each channel-use is itself a τ -second slot. To introduce some notation, consider, for example, the slot $[0, \tau)$. 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 within $[0, \tau)$: if $S = 0$ then $\Lambda(t) = 0$ for all $t \in [0, \tau)$; otherwise $\Lambda(t) = n$, $t \in [0, \tau)$. Let \mathbf{x} denote an input waveform $x(t)$, $t \in [0, \tau)$, and let \mathcal{X} denote the set of all admissible input waveforms on $[0, \tau)$ satisfying (1). For notational convenience, we introduce the following description of the output on $[0, \tau)$, which we shall use instead of the counting function $z(\cdot)$ in the rest of this paper. Let the (possibly empty) set

$$\mathbf{y} = \{t_1, \dots, t_k\} \subset [0, \tau) \quad (4)$$

be such that each $t_i \in \mathbf{y}$ corresponds to the position of an arrival, i.e., an increase by 1 in $z(t)$, $t \in [0, \tau)$. Note that, since with probability one $Z(\cdot)$ contains only unit jumps, the two descriptions of the output process in terms of \mathbf{y} and of $z(\cdot)$ are equivalent. Thus, each τ -second slot can be seen as one use of a discrete-time channel, whose input \mathbf{x} and output \mathbf{y} are both continuous-time objects, while the state s is binary. We further let \mathbf{u} denote any mapping

$$\mathbf{u}: \{0, 1\} \rightarrow \mathcal{X}, \quad s \mapsto \mathbf{u}_s, \quad (5)$$

i.e., each \mathbf{u} is determined by two input waveforms $\mathbf{u}_0, \mathbf{u}_1 \in \mathcal{X}$. Let \mathcal{U} denote the set of all such mappings (i.e., waveform pairs).

We use a letter like \mathbb{P} to denote a probability distribution,¹

¹For simplicity of notation, we shall not define σ -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.

and sometimes add indicative subscripts to it. In particular, \mathbb{P}_S denotes the state distribution on $\{0, 1\}$:

$$1 - \mathbb{P}_S(0) = \mathbb{P}_S(1) = q, \quad (6)$$

$\mathbb{P}_{\mathbf{X}}$ denotes a distribution on \mathcal{X} , and $\mathbb{P}_{\mathbf{U}}$ a distribution on \mathcal{U} . We use \times 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 $\mathbb{P}_S \times \mathbb{P}_{\mathbf{X}}$. Further, we use $\mathbb{P}_{\mathbf{U}|S}$ 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 σ -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 $S = s$.

B. Preliminaries

The capacity of the peak-limited Poisson channel with constant dark current was computed using different methods in [1]–[3]. For peak-power constraint \mathcal{A} and constant dark current ν (i.e., $\Lambda(t) = \nu$ with probability one for all t), it is given by [3, Theorem 1]:

$$C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, \nu) = \max_p \left(p(\mathcal{A} + \nu) \log \frac{\mathcal{A} + \nu}{p\mathcal{A} + \nu} + (1 - p)\nu \log \frac{\nu}{p\mathcal{A} + \nu} \right) \quad (7)$$

where the maximum is achieved by

$$p^* = \frac{(1 + \frac{\nu}{\mathcal{A}})^{1 + \frac{\nu}{\mathcal{A}}}}{(\frac{\nu}{\mathcal{A}})^{\frac{\nu}{\mathcal{A}}} \cdot e} - \frac{\nu}{\mathcal{A}}. \quad (8)$$

It is useful to note that p^* as a function of $\frac{\nu}{\mathcal{A}}$ is monotonically increasing, and satisfies [2]

$$e^{-1} \leq p^* \leq 0.5. \quad (9)$$

For the Poisson channel with random, time-varying dark current described in Section II-A, let us consider four cases: without CSI, with causal and noncausal CSI, respectively, at the transmitter, and with CSI at both the transmitter and the receiver. The capacity formulas for these four cases can be readily obtained by generalizing classic formulas to allow general input and output alphabets.

Case 1: No CSI.

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, denoted by $C_{\text{NoCSI}}^{\text{Pois}}(\mathcal{A}, n, q, \tau)$, 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.² Viewing the channel as a discrete-time memoryless channel as in Section II-A, one can generalize the standard capacity formula [29] to general input and output alphabets to obtain

$$C_{\text{NoCSI}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) = \frac{1}{\tau} \sup I(\mathbf{X}; \mathbf{Y}) \quad (10)$$

²For simplicity, we only consider average error probabilities.

with supremum over distributions of the form

$$\mathbb{P}_S(s)\mathbb{P}_{\mathbf{X}}(\mathbf{x})\mathbb{W}(\mathbf{y}|\mathbf{x}, s), \quad (11)$$

with \mathbb{P}_S given by (6).

Case 2: Causal CSI at the 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, $\lambda(s)$, $s \in [0, t]$. The decoder is of the same structure as in Case 1. Denote the capacity by $C_{\text{Cau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau)$, which is defined similarly as in Case 1. Generalizing [16] we have

$$C_{\text{Cau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) = \frac{1}{\tau} \sup I(\mathbf{U}; \mathbf{Y}) \quad (12)$$

with supremum over distributions of the form

$$\mathbb{P}_S(s)\mathbb{P}_{\mathbf{U}}(\mathbf{u})\mathbb{W}(\mathbf{y}|\mathbf{u}_s, s). \quad (13)$$

Case 3: Noncausal CSI at the transmitter only.

In this case, the encoder maps the message and the entire dark current $\lambda(t)$, $t \in [0, T]$, to the input waveform $x(t)$, $t \in [0, T]$. Denote the capacity by $C_{\text{NonCau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau)$. Generalizing [17] we have

$$C_{\text{NonCau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) = \frac{1}{\tau} \sup (I(\mathbf{U}; \mathbf{Y}) - I(\mathbf{U}; S)) \quad (14)$$

with supremum over distributions of the form

$$\mathbb{P}_S(s)\mathbb{P}_{\mathbf{U}|S}(\mathbf{u}|s)\mathbb{W}(\mathbf{y}|\mathbf{u}_s, s). \quad (15)$$

Remark 1: In some literature, the mapping \mathbf{U} in (12) and (14) is taken to be any auxiliary random variable, and the channel input \mathbf{X} is generated according to a conditional distribution $\mathbb{P}_{\mathbf{X}|\mathbf{U}S}$. It is however well understood that (in both cases of causal and noncausal CSI) it is optimal to take $\mathbb{P}_{\mathbf{X}|\mathbf{U}S}$ to be deterministic, thus reducing the auxiliary random variable \mathbf{U} to a random mapping that maps S to \mathbf{X} ; see, e.g., [18].

Case 4: CSI at both the transmitter and the receiver.

In this case, capacity is not affected by whether CSI is known to the transmitter causally or noncausally. Denote the capacity by $C_{\text{Both}}^{\text{Pois}}(\mathcal{A}, n, q, \tau)$. A simple time-sharing argument shows that (see [7], [8])

$$C_{\text{Both}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) = (1 - q)C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, 0) + qC_{\text{Const}}^{\text{Pois}}(\mathcal{A}, n), \quad (16)$$

where $C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, \nu)$ is the capacity of the Poisson channel with peak-power constraint \mathcal{A} and constant dark current ν .

The various capacities introduced above can be ordered as follows.

Proposition 2: For all \mathcal{A} , n , q , and τ ,

$$C_{\text{Both}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) \geq C_{\text{NonCau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) \geq C_{\text{Cau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) \geq C_{\text{NoCSI}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) \geq C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn). \quad (17)$$

Proof: The first three inequalities in (17) follow immediately from the fact that more knowledge cannot decrease capacity. It remains to prove the last inequality. To this end, consider the following scheme for the no-CSI case. The transmitter and the receiver first discretize the channel as in [3]: they divide the interval $[0, T]$ into small slots of Δ seconds (we restrict Δ to be such that τ/Δ is an integer),

and approximate each slot by a binary-input binary-output channel. Specifically, the binary input being 0 corresponds to transmitting 0, and it being 1 corresponds to transmitting \mathcal{A} during the slot; the output being 0 corresponds to having no arrival in the output in the slot, and the output being 1 corresponds to having at least one arrival. The transmitter and the receiver then use some common randomness (which cannot increase capacity in point-to-point settings) to permute these slots. After permutation, they ignore the common randomness they have used, and use the new sequence of channel-uses to communicate. The resulting channel is a memoryless one. Denote its transition law by W_b , then

$$W_b(0|0) = 1 - W_b(1|0) = (1 - q) + qe^{-n\Delta} \quad (18a)$$

$$W_b(0|1) = 1 - W_b(1|1) = e^{-A\Delta}((1 - q) + qe^{-n\Delta}). \quad (18b)$$

As Δ tends to zero, one can easily check that the capacity of the binary channel (18) divided by Δ tends to $C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn)$. (In fact, as Δ tends to zero, the above binary channel approaches the binary channel resulting from Wyner's original discretization scheme [3].) Hence using the above scheme we can achieve $C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn)$ on our model of Poisson channel with a random dark current without CSI. ■

III. RESULTS AND METHODS

A. Results

In this paper, we are interested in the regime where τ is small. We show that, in this regime, CSI at the transmitter, be it causal or noncausal, provides almost no benefit in terms of capacity. More precisely, the capacity increase due to CSI vanishes as τ tends to zero, and is no more than some constant times τ . Our main result is the following theorem.

Theorem 3 (Noncausal CSI): For all \mathcal{A} , n , q , and τ ,

$$\begin{aligned} & C_{\text{NonCau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) - C_{\text{Cons}}^{\text{Pois}}(\mathcal{A}, qn) \\ & \leq \left[5 + \frac{n}{\mathcal{A}} + \frac{(\mathcal{A} + n)^2}{8(\min\{q, 1 - q\})^2 n^2} \right] (\mathcal{A} + n)^2 \tau. \end{aligned} \quad (19)$$

Proof: See Section IV. ■

Note that Proposition 2 and Theorem 3 together imply that

$$\begin{aligned} & C_{\text{Cau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) - C_{\text{Cons}}^{\text{Pois}}(\mathcal{A}, qn) \\ & \leq \left[5 + \frac{n}{\mathcal{A}} + \frac{(\mathcal{A} + n)^2}{8(\min\{q, 1 - q\})^2 n^2} \right] (\mathcal{A} + n)^2 \tau. \end{aligned} \quad (20)$$

In particular, as claimed earlier, the capacity benefit of (causal or noncausal) CSI on our channel model vanishes in the limit where τ approaches zero:

$$\begin{aligned} C_{\text{Cons}}^{\text{Pois}}(\mathcal{A}, qn) &= \lim_{\tau \downarrow 0} C_{\text{NoCSI}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) \\ &= \lim_{\tau \downarrow 0} C_{\text{Cau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) \\ &= \lim_{\tau \downarrow 0} C_{\text{NonCau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau). \end{aligned} \quad (21)$$

In contrast, comparing (7) and (16) shows that

$$\begin{aligned} & C_{\text{Both}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) - C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn) \\ &= (1 - q)C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, 0) + qC_{\text{Const}}^{\text{Pois}}(\mathcal{A}, n) - C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, nq), \end{aligned} \quad (22)$$

which is positive for all \mathcal{A} , $n > 0$ and $q \in (0, 1)$, and does not depend on τ . Note that (21) and (22) together imply that, as τ approaches zero, the gap between $C_{\text{Both}}^{\text{Pois}}(\mathcal{A}, n, q, \tau)$ and $C_{\text{NonCau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau)$ approaches the right-hand side of (22). Hence, when CSI is available to both the transmitter and the receiver, it increases capacity compared to the no-CSI, causal-CSI, and noncausal-CSI cases.³

In fact, the bound for causal CSI (20) can be tightened to the following, which we state without a proof. A proof outline was presented in [30].⁴

Proposition 4 ([30, Proposition 2]): For all \mathcal{A} , n , q , and τ ,

$$C_{\text{Cau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) - C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn) \leq 4(\mathcal{A} + n)^2 \tau. \quad (23)$$

Remark 5: Recall that a different model of Poisson channel with CSI at the transmitter was studied in [13], where CSI consists of the exact positions of the arrivals caused by interference. It is shown there that such CSI, when available to the transmitter noncausally, can help to completely cancel the effect of interference: the capacity with noncausal CSI equals the capacity of the Poisson channel without this interference. In contrast, the current work shows that, if the transmitter only knows the intensity of the interference, and if this intensity changes fast, then noncausal CSI becomes nearly useless.

For causal CSI, [13] shows (somewhat heuristically) that causal knowledge of arrival times caused by interference does not increase capacity at all, whereas the current work shows that causal knowledge of interference intensity provides little capacity benefit in the regime where τ is small. These two observations are related in spirit. We note, however, that in our model the dark current has memory, so causal knowledge of the dark current value actually allows a short look-ahead, hence causal CSI in our model cannot be seen as strictly weaker than causal CSI considered in [13].

B. Some Intuition

First let us consider noncausal versus causal CSI. From (14) we observe the following. For a τ -second slot on a Poisson channel, when τ is small, the mutual information $I(\mathbf{U}; \mathbf{Y})$ must also be small. Thus, to make the right-hand side of (14) positive, one must choose $I(\mathbf{U}; S)$ to be small. This means that \mathbf{U} must be “nearly independent” of S . Such a choice will result in $I(\mathbf{U}; \mathbf{Y})$ being “nearly equal” to its value in the causal case as in (12), where \mathbf{U} is independent of S . Hence, one sees that noncausal CSI cannot be “much more useful” than causal CSI.

Next we discuss why capacity with causal CSI is not much larger than without CSI. For a τ -second slot on the Poisson channel as described in Section II-A, consider a binary approximation to the channel output: the binary output symbol y_b equals 1 if there is at least one point in the actual output y

³One might be tempted to think that a dark current that changes very fast would be nearly equivalent to a constant dark current, because, whatever the value of $\lambda(\cdot)$, the probability that it produces an arrival within a certain τ -second interval is very small. One can see from (22) that such a view is not valid.

⁴We note that the proof outlined for the *noncausal* result in [30] is incorrect. (It is unclear whether the bound [30, Proposition 3] itself is valid or not.) The causal result [30, Proposition 2] and its outlined proof are correct.

within this slot, and equals 0 otherwise. When τ is small, y_b captures most of the information in \mathbf{y} , because the probability that \mathbf{y} contains two or more points is very small (on the order of τ^2), and because the exact positions of the points, if any, cannot carry much information, as the interval is short. For Y_b we have

$$P_{Y_b|\mathbf{x}S}(1|\mathbf{x}, s) = 1 - e^{-\|\mathbf{x}\| + s\tau} \approx \|\mathbf{x}\| + sn\tau \quad (24a)$$

$$P_{Y_b|\mathbf{x}S}(0|\mathbf{x}, s) \approx 1 - \|\mathbf{x}\| - sn\tau, \quad (24b)$$

where $\|\cdot\|$ denotes the L_1 norm. Thus we have a channel where the input and the state act ‘‘almost independently,’’ where by ‘‘independently’’ we mean that the conditional output distribution can be written as the sum of a part that depends only on the input and another part that depends only on the state. But for DMCs where the input and the state act independently in this sense, one can show that causal CSI does not increase capacity at all: it is optimal to choose \mathbf{x} independently of s . We formally prove this as Proposition 8 in the Appendix.

C. Proof Method

Guided by the above observation, our method to prove an upper bound on $C_{\text{NonCau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau)$ is roughly as follows. We first fix a value $I(\mathbf{U}; S) = \alpha$, and then upper-bound the difference between $I(\mathbf{U}; \mathbf{Y})$ with noncausal CSI and $C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn)$ in terms of the parameter α . To this end, we apply an upper bound similar to the duality-based upper bound [31], [32] on $I(\mathbf{U}; \mathbf{Y})$, where we choose the reference output distribution as the output distribution that results when the transmitter ignores both the CSI and the time-varying nature of the dark current to choose the optimal input distribution in the constant-dark-current case, i.e., it chooses the input distribution that achieves $C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn)$. Let us denote the resulting bound on $I(\mathbf{U}; \mathbf{Y}) - C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn)$ by $f(\alpha)$, then

$$C_{\text{NonCau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) \leq C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn) + f(\alpha) - \alpha. \quad (25)$$

Our desired result is obtained by maximizing $f(\alpha) - \alpha$ over α .

Note that in the actual proof we must consider all possible continuous-time input and output signals; we *cannot* base our proof on a binary approximation similar to (24). This is because there exists no guarantee that such an approximation is optimal in the limit where $\tau \downarrow 0$. In fact, for any fixed $\tau > 0$, one can use Wyner’s method [3] to show that, if one further divides each τ -second slot into finer slots of Δ seconds long and applies binary approximation to each Δ -second slot, then, as Δ approaches zero, such a scheme will become optimal. But this does not imply that such a binary approximation is also optimal if one chooses $\Delta = \tau$, even if τ itself is small.

The proof method, where we fix $I(\mathbf{U}; S) = \alpha$ and then bound $I(\mathbf{U}; \mathbf{Y})$ in terms of α , can be applied to a broader class of channels where the input and the state have ‘‘little influence’’ on the output distribution. As an example, we shall prove in Section V that, for the very noisy channel, noncausal CSI provides little capacity improvement over causal CSI.

IV. PROOF OF THEOREM 3

Consider the formula (14). Fix $I(\mathbf{U}; S)$ to be

$$I(\mathbf{U}; S) = \alpha, \quad (26)$$

whose value can depend on τ . By Pinsker’s inequality [31],

$$\delta(\mathbb{P}_S \mathbb{P}_{\mathbf{U}|S}, \mathbb{P}_S \times \mathbb{P}_{\mathbf{U}}) \leq \sqrt{\frac{\alpha}{2}}, \quad (27)$$

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

We next look at the $I(\mathbf{U}; \mathbf{Y})$ term in (14). Let $\mathbb{P}_{S|\mathbf{U}}$ be the conditional distribution of S given \mathbf{U} according to the joint distribution $\mathbb{P}_S \mathbb{P}_{\mathbf{U}|S}$. For convenience, denote

$$\pi_{\mathbf{u}} \triangleq \mathbb{P}_{S|\mathbf{U}}(1|\mathbf{u}). \quad (28)$$

Let $\widehat{\mathbb{W}}$ denote the conditional distribution of \mathbf{Y} given \mathbf{U} , then

$$\widehat{\mathbb{W}}(\cdot|\mathbf{u}) = (1 - \pi_{\mathbf{u}})\mathbb{W}(\cdot|\mathbf{u}_0, 0) + \pi_{\mathbf{u}}\mathbb{W}(\cdot|\mathbf{u}_1, 1) \quad (29)$$

for every $\mathbf{u} \in \mathcal{U}$. For any distribution \mathbb{Q} on \mathcal{Y} , we have the following bound, which is related to the duality-based upper bound [31], [32]:

$$\begin{aligned} I(\mathbf{U}; \mathbf{Y}) &= \mathbb{E} \left[D \left(\widehat{\mathbb{W}}(\cdot|\mathbf{U}) \middle\| P_{\mathbf{U}} \widehat{\mathbb{W}} \right) \right] \\ &= \mathbb{E} \left[D \left(\widehat{\mathbb{W}}(\cdot|\mathbf{U}) \middle\| \mathbb{Q} \right) \right] - D \left(P_{\mathbf{U}} \widehat{\mathbb{W}} \middle\| \mathbb{Q} \right) \\ &\leq \mathbb{E} \left[D \left(\widehat{\mathbb{W}}(\cdot|\mathbf{U}) \middle\| \mathbb{Q} \right) \right]. \end{aligned} \quad (30)$$

We choose \mathbb{Q} to be doubly-stochastic Poisson under which, with probability $(1 - q)$, \mathbf{Y} is a homogeneous Poisson process of intensity $p^* \mathcal{A}$, and with probability q , it is a homogeneous Poisson process of intensity $p^* \mathcal{A} + n$, on $[0, \tau)$, where p^* is given in (8). Note that this \mathbb{Q} would be the output distribution if the transmitter ignored the CSI and chose the input distribution to be the one that achieves $C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn)$.

Conditional on $\mathbf{U} = \mathbf{u}$, \mathbf{Y} is doubly stochastic Poisson; its intensity is $u_0(t)$, $t \in [0, \tau)$, with probability $(1 - \pi_{\mathbf{u}})$, and is $u_1(t) + n$, $t \in [0, \tau)$, with probability $\pi_{\mathbf{u}}$. Thus we can obtain the following Radon-Nikodym derivative using [33, (19.125)]:

$$\begin{aligned} \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{y}) &= \frac{(1 - \pi_{\mathbf{u}})e^{-\|\mathbf{u}_0\|} \prod_{t \in \mathbf{y}} u_0(t) + \pi_{\mathbf{u}}e^{-\|\mathbf{u}_1\| - n\tau} \prod_{t \in \mathbf{y}} (u_1(t) + n)}{(1 - q)e^{-p^* \mathcal{A}\tau} (p^* \mathcal{A})^{|\mathbf{y}|} + qe^{-(p^* \mathcal{A} + n)\tau} (p^* \mathcal{A} + n)^{|\mathbf{y}|}} \end{aligned} \quad (31)$$

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

$$\|\mathbf{x}\| = \int_0^\tau x(t) dt. \quad (32)$$

We write the relative entropy in (30) as the following summation:

$$\begin{aligned} &D \left(\widehat{\mathbb{W}}(\cdot|\mathbf{u}) \middle\| \mathbb{Q} \right) \\ &= \sum_{k=0}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u}) \mathbb{E} \left[\log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{Y}) \middle\| |\mathbf{Y}| = k \right]. \end{aligned} \quad (33)$$

First consider the summand in (33) with $k = 0$, which means $\mathbf{Y} = \emptyset$. If $\widehat{\mathbb{W}}(\emptyset|\mathbf{u}) \geq \mathbb{Q}(\emptyset)$, then we use $\widehat{\mathbb{W}}(\emptyset|\mathbf{u}) \leq 1$ to derive the following bound:

$$\begin{aligned}
& \widehat{\mathbb{W}}(\emptyset|\mathbf{u}) \log \frac{\widehat{\mathbb{W}}(\emptyset|\mathbf{u})}{\mathbb{Q}(\emptyset)} \\
& \leq \log \frac{\widehat{\mathbb{W}}(\emptyset|\mathbf{u})}{\mathbb{Q}(\emptyset)} \\
& = \log \frac{(1 - \pi_{\mathbf{u}})e^{-\|\mathbf{u}_0\|} + \pi_{\mathbf{u}}e^{-\|\mathbf{u}_1\| - n\tau}}{(1 - q)e^{-p^* \mathcal{A}\tau} + qe^{-(p^* \mathcal{A} + n)\tau}} \\
& = \log \underbrace{\left((1 - \pi_{\mathbf{u}})e^{-\|\mathbf{u}_0\|} + \pi_{\mathbf{u}}e^{-\|\mathbf{u}_1\| - n\tau} \right)}_{\leq (1 - \pi_{\mathbf{u}})e^{-\|\mathbf{u}_0\|} + \pi_{\mathbf{u}}e^{-\|\mathbf{u}_1\| - n\tau} - 1} \\
& \quad - \log \underbrace{\left((1 - q)e^{-p^* \mathcal{A}\tau} + qe^{-(p^* \mathcal{A} + n)\tau} \right)}_{\geq (1 - q) \log(e^{-p^* \mathcal{A}\tau}) + q \log(e^{-(p^* \mathcal{A} + n)\tau})} \\
& \leq (1 - \pi_{\mathbf{u}})e^{-\|\mathbf{u}_0\|} + \pi_{\mathbf{u}}e^{-\|\mathbf{u}_1\| - n\tau} - 1 \\
& \quad + (1 - q)p^* \mathcal{A}\tau + q(p^* \mathcal{A} + n)\tau \\
& \leq (1 - \pi_{\mathbf{u}}) \left(1 - \|\mathbf{u}_0\| + \frac{\|\mathbf{u}_0\|^2}{2} \right) \\
& \quad + \pi_{\mathbf{u}} \left(1 - (\|\mathbf{u}_1\| + n\tau) + \frac{(\|\mathbf{u}_1\| + n\tau)^2}{2} \right) \\
& \quad - 1 + p^* \mathcal{A}\tau + qn\tau \\
& = p^* \mathcal{A}\tau - \|\hat{\mathbf{u}}\| + (q - \pi_{\mathbf{u}})n\tau \\
& \quad + \frac{(1 - \pi_{\mathbf{u}})\|\mathbf{u}_0\|^2}{2} + \frac{\pi_{\mathbf{u}}(\|\mathbf{u}_1\| + n\tau)^2}{2} \\
& \leq p^* \mathcal{A}\tau - \|\hat{\mathbf{u}}\| + (q - \pi_{\mathbf{u}})n\tau + (\mathcal{A} + n)^2 \tau^2, \\
& \quad \text{if } \widehat{\mathbb{W}}(\emptyset|\mathbf{u}) \geq \mathbb{Q}(\emptyset), \quad (34)
\end{aligned}$$

where

$$\hat{\mathbf{u}} \triangleq (1 - \pi_{\mathbf{u}})\mathbf{u}_0 + \pi_{\mathbf{u}}\mathbf{u}_1. \quad (35)$$

If $\widehat{\mathbb{W}}(\emptyset|\mathbf{u}) < \mathbb{Q}(\emptyset)$, we use $\widehat{\mathbb{W}}(\emptyset|\mathbf{u}) \geq 1 - (\mathcal{A} + n)\tau$ to write

$$\begin{aligned}
& \widehat{\mathbb{W}}(\emptyset|\mathbf{u}) \log \frac{\widehat{\mathbb{W}}(\emptyset|\mathbf{u})}{\mathbb{Q}(\emptyset)} \leq (1 - (\mathcal{A} + n)\tau) \log \frac{\widehat{\mathbb{W}}(\emptyset|\mathbf{u})}{\mathbb{Q}(\emptyset)} \\
& \leq \log \frac{\widehat{\mathbb{W}}(\emptyset|\mathbf{u})}{\mathbb{Q}(\emptyset)} - (\mathcal{A} + n)\tau \log \widehat{\mathbb{W}}(\emptyset|\mathbf{u}). \quad (36)
\end{aligned}$$

For the second term on the right-hand side of (36), we have

$$\begin{aligned}
& \log \widehat{\mathbb{W}}(\emptyset|\mathbf{u}) = \log \left((1 - \pi_{\mathbf{u}})e^{-\|\mathbf{u}_0\|} + \pi_{\mathbf{u}}e^{-\|\mathbf{u}_1\| - n\tau} \right) \\
& \geq (1 - \pi_{\mathbf{u}}) \log \left(e^{-\|\mathbf{u}_0\|} \right) + \pi_{\mathbf{u}} \log \left(e^{-\|\mathbf{u}_1\| - n\tau} \right) \\
& = -\|\hat{\mathbf{u}}\| - \pi_{\mathbf{u}}n\tau \\
& \geq -(\mathcal{A} + n)\tau. \quad (37)
\end{aligned}$$

Plugging the above into the second term on the right-hand side of (36), and upper-bounding the first term there by the right-hand side of (34), we obtain

$$\begin{aligned}
& \widehat{\mathbb{W}}(\emptyset|\mathbf{u}) \cdot \log \frac{\widehat{\mathbb{W}}(\emptyset|\mathbf{u})}{\mathbb{Q}(\emptyset)} \\
& \leq p^* \mathcal{A}\tau - \|\hat{\mathbf{u}}\| + (q - \pi_{\mathbf{u}})n\tau + 2(\mathcal{A} + n)^2 \tau^2, \\
& \quad \text{if } \widehat{\mathbb{W}}(\emptyset|\mathbf{u}) < \mathbb{Q}(\emptyset). \quad (38)
\end{aligned}$$

Combining (34) and (38) we conclude that (in both cases)

$$\begin{aligned}
& \widehat{\mathbb{W}}(\emptyset|\mathbf{u}) \cdot \log \frac{\widehat{\mathbb{W}}(\emptyset|\mathbf{u})}{\mathbb{Q}(\emptyset)} \\
& \leq p^* \mathcal{A}\tau - \|\hat{\mathbf{u}}\| + (q - \pi_{\mathbf{u}})n\tau + 2(\mathcal{A} + n)^2 \tau^2. \quad (39)
\end{aligned}$$

For the summand in (33) with $k = 1$, we have

$$\begin{aligned}
& \widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u}) \\
& = (1 - \pi_{\mathbf{u}})\|\mathbf{u}_0\|e^{-\|\mathbf{u}_0\|} + \pi_{\mathbf{u}}(\|\mathbf{u}_1\| + n\tau)e^{-\|\mathbf{u}_1\| - n\tau}. \quad (40)
\end{aligned}$$

Conditional on $\mathbf{U} = \mathbf{u}$, and further conditional on the event $|\mathbf{Y}| = 1$, the probability density function of the position of the point in \mathbf{Y} is

$$f(t) \triangleq \frac{(1 - \pi_{\mathbf{u}})e^{-\|\mathbf{u}_0\|}u_0(t) + \pi_{\mathbf{u}}e^{-\|\mathbf{u}_1\| - n\tau}(u_1(t) + n)}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \quad (41)$$

for $t \in [0, \tau]$, which can be bounded from above and below as

$$\begin{aligned}
& \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \cdot (1 - (\mathcal{A} + n)\tau) \\
& \leq f(t) \leq \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})}. \quad (42)
\end{aligned}$$

Hence, using (31),

$$\begin{aligned}
& \mathbb{E} \left[\log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{Y}) \middle| |\mathbf{Y}| = 1 \right] \\
& = \int_0^\tau f(t) \\
& \quad \log \frac{(1 - \pi_{\mathbf{u}})u_0(t)e^{-\|\mathbf{u}_0\|} + \pi_{\mathbf{u}}(u_1(t) + n)e^{-\|\mathbf{u}_1\| - n\tau}}{(1 - q)p^* \mathcal{A}e^{-p^* \mathcal{A}\tau} + q(p^* \mathcal{A} + n)e^{-(p^* \mathcal{A} + n)\tau}} dt \\
& \leq \int_0^\tau f(t) \log \frac{(1 - \pi_{\mathbf{u}})u_0(t) + \pi_{\mathbf{u}}(u_1(t) + n)}{((1 - q)p^* \mathcal{A} + q(p^* \mathcal{A} + n))e^{-(p^* \mathcal{A} + n)\tau}} dt \\
& = \int_0^\tau f(t) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{p^* \mathcal{A} + qn} dt + p^* \mathcal{A}\tau + n\tau \\
& = \int_0^\tau \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \log \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{p^* \mathcal{A} + qn} dt \\
& \quad + \int_0^\tau \left(f(t) - \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \right) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{p^* \mathcal{A} + qn} dt \\
& \quad + p^* \mathcal{A}\tau + n\tau. \quad (43)
\end{aligned}$$

Consider the second integral on the right-hand side of (43). Using (42) its integrand can be bounded as

$$\begin{aligned}
& \left(f(t) - \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \right) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{p^* \mathcal{A} + qn} \\
& = \underbrace{\left(\frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} - f(t) \right)}_{\geq 0} \log \frac{p^* \mathcal{A} + qn}{\hat{u}(t) + \pi_{\mathbf{u}}n} \\
& \leq \left(\frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} - f(t) \right) \cdot \max \left\{ 0, \log \frac{p^* \mathcal{A} + qn}{\hat{u}(t) + \pi_{\mathbf{u}}n} \right\} \\
& \leq \frac{(\hat{u}(t) + \pi_{\mathbf{u}}n)(\mathcal{A} + n)\tau}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \cdot \max \left\{ 0, \log \frac{p^* \mathcal{A} + qn}{\hat{u}(t) + \pi_{\mathbf{u}}n} \right\}
\end{aligned}$$

$$= \max \left\{ 0, \frac{(\hat{u}(t) + \pi_{\mathbf{u}}n)(\mathcal{A} + n)\tau}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \log \frac{p^*\mathcal{A} + qn}{\hat{u}(t) + \pi_{\mathbf{u}}n} \right\}. \quad (44)$$

Using the following inequality (which can be verified, e.g., by computing the derivative of its left-hand side with respect to a)

$$a \log \frac{b}{a} \leq be^{-1}, \quad a, b > 0,$$

we can continue (44) to

$$\begin{aligned} & \left(f(t) - \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \right) \log \frac{p^*\mathcal{A} + qn}{\hat{u}(t) + \pi_{\mathbf{u}}n} \\ & \leq \max \left\{ 0, \frac{(\mathcal{A} + n)\tau}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \cdot (p^*\mathcal{A} + qn)e^{-1} \right\} \\ & \leq \frac{(\mathcal{A} + n)^2\tau}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})}, \end{aligned} \quad (45)$$

hence

$$\begin{aligned} & \int_0^\tau \left(f(t) - \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \right) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{p^*\mathcal{A} + qn} dt \\ & \leq \frac{(\mathcal{A} + n)^2\tau^2}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})}. \end{aligned} \quad (46)$$

Plugging this into (43) we have

$$\begin{aligned} & \mathbb{E} \left[\log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{Y}) \middle| |\mathbf{Y}| = 1 \right] \\ & \leq \int_0^\tau \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} \log \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{p^*\mathcal{A} + qn} dt \\ & \quad + \frac{(\mathcal{A} + n)^2\tau^2}{\widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})} + p^*\mathcal{A}\tau + n\tau. \end{aligned} \quad (47)$$

Combining (40) and (47) we can upper-bound the summand in (33) with $k = 1$ by

$$\begin{aligned} & \widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u}) \mathbb{E} \left[\log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{Y}) \middle| |\mathbf{Y}| = 1 \right] \\ & \leq \int_0^\tau (\hat{u}(t) + \pi_{\mathbf{u}}n) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{p^*\mathcal{A} + qn} dt + (\mathcal{A} + n)^2\tau^2 \\ & \quad + \widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u})(p^*\mathcal{A}\tau + n\tau). \end{aligned} \quad (48)$$

For the remaining summands in (33) with $k \geq 2$, we have

$$\begin{aligned} & \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{y}) \\ & \leq \frac{(1 - \pi_{\mathbf{u}}) \prod_{t \in \mathbf{y}} u_0(t) + \pi_{\mathbf{u}} \prod_{t \in \mathbf{y}} (u_1(t) + n)}{\left((1 - q)(p^*\mathcal{A})^{|\mathbf{y}|} + q(p^*\mathcal{A} + n)^{|\mathbf{y}|} \right) e^{-(p^*\mathcal{A} + n)\tau}} \\ & \leq \frac{(1 - \pi_{\mathbf{u}})\mathcal{A}^{|\mathbf{y}|} + \pi_{\mathbf{u}}(\mathcal{A} + n)^{|\mathbf{y}|}}{\left((1 - q)(p^*\mathcal{A})^{|\mathbf{y}|} + q(p^*\mathcal{A} + n)^{|\mathbf{y}|} \right)} \cdot e^{(p^*\mathcal{A} + n)\tau} \\ & \leq \frac{(\mathcal{A} + n)^{|\mathbf{y}|}}{(p^*\mathcal{A})^{|\mathbf{y}|}} \cdot e^{(p^*\mathcal{A} + n)\tau} \\ & = \left(\frac{\mathcal{A} + n}{p^*\mathcal{A}} \right)^{|\mathbf{y}|} \cdot e^{(p^*\mathcal{A} + n)\tau}, \end{aligned} \quad (49)$$

therefore

$$\begin{aligned} & \sum_{k=2}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u}) \mathbb{E} \left[\log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{Y}) \middle| |\mathbf{Y}| = k \right] \\ & \leq \log \frac{\mathcal{A} + n}{p^*\mathcal{A}} \cdot \sum_{k=2}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u}) \cdot k \\ & \quad + (p^*\mathcal{A} + n)\tau \cdot \sum_{k=2}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u}). \end{aligned} \quad (50)$$

To bound the first term on the right-hand side, we use the following

$$\begin{aligned} & \sum_{k=2}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u}) \cdot k \\ & = \mathbb{E}[|\mathbf{Y}||\mathbf{u}] - \widehat{\mathbb{W}}(|\mathbf{Y}| = 1|\mathbf{u}) \cdot 1 \\ & = \|\bar{\mathbf{u}}\| + qn\tau - (1 - q)\|\mathbf{u}_0\|e^{-\|\mathbf{u}_0\|} \\ & \quad - q(\|\mathbf{u}_1\| + n\tau)e^{-\|\mathbf{u}_1\| - n\tau} \\ & \leq \|\bar{\mathbf{u}}\| + qn\tau - (1 - q)\|\mathbf{u}_0\|(1 - \|\mathbf{u}_0\|) \\ & \quad - q(\|\mathbf{u}_1\| + n\tau)(1 - \|\mathbf{u}_1\| - n\tau) \\ & = (1 - q)\|\mathbf{u}_0\|^2 + q(\|\mathbf{u}_1\| + n\tau)^2 \\ & \leq (\mathcal{A} + n)^2\tau^2. \end{aligned} \quad (51)$$

Hence we continue (50) as

$$\begin{aligned} & \sum_{k=2}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u}) \mathbb{E} \left[\log \frac{d\widehat{\mathbb{W}}(\cdot|\mathbf{u})}{d\mathbb{Q}}(\mathbf{Y}) \middle| |\mathbf{Y}| = k \right] \\ & \leq \log \frac{\mathcal{A} + n}{p^*\mathcal{A}} \cdot (\mathcal{A} + n)^2\tau^2 \\ & \quad + (p^*\mathcal{A} + n)\tau \cdot \sum_{k=2}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u}) \\ & \leq \left(\underbrace{\log \frac{1}{p^*}}_{\leq 1} + \underbrace{\log \frac{\mathcal{A} + n}{\mathcal{A}}}_{\leq \frac{n}{\mathcal{A}}} \right) (\mathcal{A} + n)^2\tau^2 \\ & \quad + (p^*\mathcal{A} + n)\tau \cdot \sum_{k=2}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u}) \\ & \leq \left(1 + \frac{n}{\mathcal{A}} \right) (\mathcal{A} + n)^2\tau^2 \\ & \quad + (p^*\mathcal{A} + n)\tau \cdot \sum_{k=2}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u}), \end{aligned} \quad (52)$$

where, to obtain $\log \frac{1}{p^*} \leq 1$ for the last step, we recall (9).

Plugging (39), (48), and (52) into (33), we obtain

$$\begin{aligned} & D\left(\widehat{\mathbb{W}}(\cdot|\mathbf{u}) \middle| \mathbb{Q}\right) \leq p^*\mathcal{A}\tau - \|\hat{\mathbf{u}}\| + (q - \pi_{\mathbf{u}})n\tau \\ & \quad + \int_0^\tau (\hat{u}(t) + \pi_{\mathbf{u}}n) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}}n}{p^*\mathcal{A} + qn} dt \\ & \quad + \left(4 + \frac{n}{\mathcal{A}} \right) (\mathcal{A} + n)^2\tau^2 \\ & \quad + (p^*\mathcal{A} + n)\tau \cdot \underbrace{\sum_{k=1}^{\infty} \widehat{\mathbb{W}}(|\mathbf{Y}| = k|\mathbf{u})}_{\leq (\mathcal{A} + n)\tau} \end{aligned}$$

$$\begin{aligned}
&\leq p^* \mathcal{A} \tau - \|\hat{\mathbf{u}}\| + (q - \pi_{\mathbf{u}}) n \tau \\
&\quad + \int_0^\tau (\hat{u}(t) + \pi_{\mathbf{u}} n) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}} n}{p^* \mathcal{A} + qn} dt \\
&\quad + \left(5 + \frac{n}{\mathcal{A}}\right) (\mathcal{A} + n)^2 \tau^2. \tag{53}
\end{aligned}$$

Consider the following expression:

$$\begin{aligned}
&p^* \mathcal{A} \tau - \|\bar{\mathbf{u}}\| + \int_0^\tau (\bar{u}(t) + qn) \log \frac{\bar{u}(t) + qn}{p^* \mathcal{A} + qn} dt \\
&= \int_0^\tau \left(p^* \mathcal{A} - \bar{u}(t) + (\bar{u}(t) + qn) \log \frac{\bar{u}(t) + qn}{p^* \mathcal{A} + qn} \right) dt, \tag{54}
\end{aligned}$$

where we define

$$\bar{\mathbf{u}} \triangleq (1 - q) \mathbf{u}_0 + q \mathbf{u}_1. \tag{55}$$

We shall continue (53) by first bounding (54), and then bounding the difference between the right-hand side of (53) and (54). To bound (54), note that, for every t , the second derivative of the integrand in (54) with respect to $\bar{u}(t)$ equals $(\bar{u}(t) + qn)^{-1}$, which is positive for all $\bar{u}(t) \in [0, \mathcal{A}]$, implying that the integrand is convex in $\bar{u}(t)$ on $[0, \mathcal{A}]$. Furthermore, we can check that, for both $\bar{u}(t) = 0$ and at $\bar{u}(t) = \mathcal{A}$, this integrand equals $C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn)$. Hence the integrand is always upper-bounded by $C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn)$. We thus conclude

$$\begin{aligned}
&\int_0^\tau \left(p^* \mathcal{A} - \bar{u}(t) + (\bar{u}(t) + qn) \log \frac{\bar{u}(t) + qn}{p^* \mathcal{A} + qn} \right) dt \\
&\leq \tau \cdot C_{\text{Const}}^{\text{Pois}}(\mathcal{A}, qn). \tag{56}
\end{aligned}$$

Comparing (53) with (54) and using (56) we obtain

$$\begin{aligned}
&D\left(\widehat{\mathbb{W}}(\cdot|\mathbf{u})\middle|\middle|\mathbb{Q}\right) - \tau \cdot C_{\text{Cons}}^{\text{Pois}}(\mathcal{A}, qn) - \left(5 + \frac{n}{\mathcal{A}}\right) (\mathcal{A} + n)^2 \tau^2 \\
&\leq \|\bar{\mathbf{u}}\| - \|\hat{\mathbf{u}}\| + qn\tau - \pi_{\mathbf{u}} n \tau \\
&\quad + \int_0^\tau (\hat{u}(t) + \pi_{\mathbf{u}} n) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}} n}{p^* \mathcal{A} + qn} dt \\
&\quad - \int_0^\tau (\bar{u}(t) + qn) \log \frac{\bar{u}(t) + qn}{p^* \mathcal{A} + qn} dt \\
&= \|\bar{\mathbf{u}}\| - \|\hat{\mathbf{u}}\| + qn\tau - \pi_{\mathbf{u}} n \tau \\
&\quad + \int_0^\tau (\hat{u}(t) + \pi_{\mathbf{u}} n) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}} n}{\bar{u}(t) + qn} dt \\
&\quad + \int_0^\tau (\hat{u}(t) + \pi_{\mathbf{u}} n - \bar{u}(t) - qn) \log \frac{\bar{u}(t) + qn}{p^* \mathcal{A} + qn} dt. \tag{57}
\end{aligned}$$

Using the log-sum inequality, we can bound the first integral on the right-hand side of (57) as

$$\begin{aligned}
&\int_0^\tau (\hat{u}(t) + \pi_{\mathbf{u}} n) \log \frac{\hat{u}(t) + \pi_{\mathbf{u}} n}{\bar{u}(t) + qn} dt \\
&\leq \int_0^\tau \left((1 - \pi_{\mathbf{u}}) u_0(t) \log \frac{(1 - \pi_{\mathbf{u}}) u_0(t)}{(1 - q) u_0(t)} \right. \\
&\quad \left. + \pi_{\mathbf{u}} (u_1(t) + n) \log \frac{\pi_{\mathbf{u}} (u_1(t) + n)}{q (u_1(t) + n)} \right) dt \\
&= \|\mathbf{u}_0\| (1 - \pi_{\mathbf{u}}) \log \frac{1 - \pi_{\mathbf{u}}}{1 - q} + (\|\mathbf{u}_1\| + n\tau) \pi_{\mathbf{u}} \log \frac{\pi_{\mathbf{u}}}{q} \\
&\leq \|\mathbf{u}_0\| (1 - \pi_{\mathbf{u}}) \cdot \frac{q - \pi_{\mathbf{u}}}{1 - q} + (\|\mathbf{u}_1\| + n\tau) \pi_{\mathbf{u}} \cdot \frac{\pi_{\mathbf{u}} - q}{q} \\
&\leq (\|\mathbf{u}_0\| (1 - \pi_{\mathbf{u}}) + (\|\mathbf{u}_1\| + n\tau) \pi_{\mathbf{u}}) \cdot \frac{|\pi_{\mathbf{u}} - q|}{\min\{q, 1 - q\}}
\end{aligned}$$

$$\leq \frac{(\mathcal{A} + n)\tau}{\min\{q, 1 - q\}} \cdot \delta(\mathbb{P}_{S|\mathbf{U}}, \mathbb{P}_S). \tag{58}$$

We bound the second integral on the right-hand side of (57) as

$$\begin{aligned}
&\int_0^\tau (\hat{u}(t) + \pi_{\mathbf{u}} n - \bar{u}(t) - qn) \log \frac{\bar{u}(t) + qn}{p^* \mathcal{A} + qn} dt \\
&\leq \int_0^\tau |\hat{u}(t) + \pi_{\mathbf{u}} n - \bar{u}(t) - qn| \cdot \left| \log \frac{\bar{u}(t) + qn}{p^* \mathcal{A} + qn} \right| dt \\
&\leq \int_0^\tau |\hat{u}(t) + \pi_{\mathbf{u}} n - \bar{u}(t) - qn| \cdot \log \frac{\mathcal{A} + qn}{qn} dt \\
&\leq \int_0^\tau |\hat{u}(t) + \pi_{\mathbf{u}} n - \bar{u}(t) - qn| \cdot \frac{\mathcal{A}}{qn} dt \\
&= \int_0^\tau |(\pi_{\mathbf{u}} - q)(u_1(t) + n - u_0(t))| \cdot \frac{\mathcal{A}}{qn} dt \\
&\leq \int_0^\tau |\pi_{\mathbf{u}} - q| \cdot (\mathcal{A} + n) \cdot \frac{\mathcal{A}}{qn} dt \\
&= \frac{(\mathcal{A} + n)\mathcal{A}\tau}{qn} \cdot \delta(\mathbb{P}_{S|\mathbf{U}}, \mathbb{P}_S). \tag{59}
\end{aligned}$$

Plugging (58) and (59) into (57) we obtain

$$\begin{aligned}
&D\left(\widehat{\mathbb{W}}(\cdot|\mathbf{u})\middle|\middle|\mathbb{Q}\right) - \tau \cdot C_{\text{Cons}}^{\text{Pois}}(\mathcal{A}, qn) - \left(5 + \frac{n}{\mathcal{A}}\right) (\mathcal{A} + n)^2 \tau^2 \\
&\leq \|\bar{\mathbf{u}}\| - \|\hat{\mathbf{u}}\| + qn\tau - \pi_{\mathbf{u}} n \tau \\
&\quad + \left(\frac{(\mathcal{A} + n)\tau}{\min\{q, 1 - q\}} + \frac{(\mathcal{A} + n)\mathcal{A}\tau}{qn} \right) \cdot \delta(\mathbb{P}_{S|\mathbf{U}}, \mathbb{P}_S) \\
&\leq \|\bar{\mathbf{u}}\| - \|\hat{\mathbf{u}}\| + qn\tau - \pi_{\mathbf{u}} n \tau \\
&\quad + \frac{(\mathcal{A} + n)^2 \tau}{\min\{q, 1 - q\} \cdot n} \cdot \delta(\mathbb{P}_{S|\mathbf{U}}, \mathbb{P}_S). \tag{60}
\end{aligned}$$

Taking expectation over \mathbf{U} and recalling (30), we have

$$\begin{aligned}
&I(\mathbf{U}; \mathbf{Y}) - \tau \cdot C_{\text{Cons}}^{\text{Pois}}(\mathcal{A}, qn) - \left(5 + \frac{n}{\mathcal{A}}\right) (\mathcal{A} + n)^2 \tau^2 \\
&\leq \mathbf{E} \left[\underbrace{\|\bar{\mathbf{U}}\| - \|\hat{\mathbf{U}}\| + qn\tau - \mathbb{P}_{S|\mathbf{U}}(1|\mathbf{U})n\tau}_{=0} \right] \\
&\quad + \frac{(\mathcal{A} + n)^2 \tau}{\min\{q, 1 - q\} \cdot n} \cdot \delta(\mathbb{P}_S \mathbb{P}_{\mathbf{U}|S}, \mathbb{P}_S \times \mathbb{P}_{\mathbf{U}}) \\
&= \frac{(\mathcal{A} + n)^2 \tau}{\min\{q, 1 - q\} \cdot n} \cdot \delta(\mathbb{P}_S \mathbb{P}_{\mathbf{U}|S}, \mathbb{P}_S \times \mathbb{P}_{\mathbf{U}}). \tag{61}
\end{aligned}$$

Recalling (14) and (27), we now have

$$\begin{aligned}
&C_{\text{NonCau}}^{\text{Pois}}(\mathcal{A}, n, q, \tau) - C_{\text{Cons}}^{\text{Pois}}(\mathcal{A}, qn) \\
&\leq \frac{(\mathcal{A} + n)^2}{\min\{q, 1 - q\} \cdot n} \cdot \sqrt{\frac{\alpha}{2}} - \frac{\alpha}{\tau} + \left(5 + \frac{n}{\mathcal{A}}\right) (\mathcal{A} + n)^2 \tau \\
&= \frac{(\mathcal{A} + n)^4}{8(\min\{q, 1 - q\})^2 n^2} \cdot \tau + \left(5 + \frac{n}{\mathcal{A}}\right) (\mathcal{A} + n)^2 \tau \\
&\quad - \frac{1}{\tau} \left(\sqrt{\alpha} - \sqrt{\frac{1}{8}} \cdot \frac{(\mathcal{A} + n)^2 \tau}{\min\{q, 1 - q\} \cdot n} \right)^2. \tag{62}
\end{aligned}$$

The right-hand side of the above is maximized at

$$\alpha = \frac{(\mathcal{A} + n)^4 \tau^2}{8(\min\{q, 1 - q\})^2 n^2}, \tag{63}$$

and the maximum equals the right-hand side of (19). This completes the proof.

V. A RELATED PROBLEM: THE VERY NOISY CHANNEL

In this section, we study the very noisy channel (VNC) [12], [28] with CSI. On a communication channel, if the number of degrees of freedom (e.g., bandwidth) available for transmission increases, while the transmitter resource (e.g., power) is held fixed, then the resource available to be used on each degree of freedom becomes small [12]. The channel thus becomes “very noisy” in the sense that, within each channel use, the transmitter has limited ability to influence the output. Consequently, the capacity per channel use is small. To model such a scenario, we generalize the model adopted in [12], [28] to include a channel state. The mathematical formulation is given below.⁵

Unlike the Poisson model we consider, the VNC capacity may benefit from causal CSI. However, we shall show that, similar to the Poisson case, noncausal CSI at the transmitter is “not much more useful” than causal CSI; see Theorem 6. Coming back to the picture where VNC is seen as an abstraction for a wideband channel, this result means that noncausal CSI does not increase the per-second capacity of such a channel over causal CSI.

A. Channel Model and Preliminaries

Consider a state-dependent channel with input $x \in \mathcal{X}$, output $y \in \mathcal{Y}$, and state $s \in \mathcal{S}$, where $\mathcal{X}, \mathcal{Y}, \mathcal{S}$ are all finite. The channel law is

$$W(y|x, s) = \beta(y)[1 + \epsilon\psi(y|x, s)], \quad x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}, \quad (64)$$

where $\beta(\cdot)$ is a distribution on \mathcal{Y} ; ψ satisfies

$$\sum_{y \in \mathcal{Y}} \beta(y)\psi(y|x, s) = 0, \quad x \in \mathcal{X}, s \in \mathcal{S}; \quad (65)$$

the state sequence \mathbf{S} is IID according to P_S and is independent of the message; and ϵ is some positive parameter, which is typically close to zero. This channel is a generalization of the VNC without states [12], [28]. Its capacity in various CSI settings is defined in the same way as for general state-dependent DMCs; see [18].

First consider the capacity of this channel with *causal* CSI at the transmitter and no CSI at the receiver, which we henceforth denote by $C_{\text{Cau}}^{\text{VNC}}(\epsilon)$. By Shannon’s classic result [16],

$$C_{\text{Cau}}^{\text{VNC}}(\epsilon) = \max I(U; Y), \quad (66)$$

where U takes value in the set \mathcal{U} containing all mappings \mathcal{S} to \mathcal{X} , and where the maximum is taken over joint distributions of the form $P_S(s)P_U(u)W(y|u(s), s)$, with $u(s)$ denoting result of applying the mapping u on the state s . For the VNC (64), the derived channel law from u to y is given by

$$\bar{W}(y|u) = \beta(y)[1 + \epsilon\bar{\psi}(y|u)], \quad u \in \mathcal{U}, y \in \mathcal{Y}, \quad (67)$$

with

$$\bar{\psi}(y|u) \triangleq \sum_s P_S(s)\psi(y|u(s), s). \quad (68)$$

⁵Note that the Poisson model considered in this paper *cannot* be seen as a special case of this VNC model.

The channel (67) itself can be seen as a VNC without states, hence we can apply the results in [12], [28] to obtain

$$\lim_{\epsilon \downarrow 0} \frac{C_{\text{Cau}}^{\text{VNC}}(\epsilon)}{\epsilon^2} = \frac{1}{2} \max_{P_U} \sum_{y, u} \beta(y)P_U(u)\bar{\eta}(y|u)^2 \quad (69)$$

where $\bar{\eta}$ depends on P_U and is given by

$$\bar{\eta}(y|u) = \bar{\psi}(y|u) - \sum_{u'} P_U(u')\bar{\psi}(y|u'), \quad u \in \mathcal{U}, y \in \mathcal{Y}. \quad (70)$$

Denote the capacity of (64) with *noncausal* CSI at the transmitter (and no CSI at the receiver) by $C_{\text{NonCau}}^{\text{VNC}}(\epsilon)$. We know that [17]

$$C_{\text{NonCau}}^{\text{VNC}}(\epsilon) = \max I(U; Y) - I(U; S), \quad (71)$$

where the maximization is over joint distributions of the form $P_S(s)P_{U|S}(u|s)W(y|u(s), s)$.

B. Result and Proof

Clearly, $C_{\text{NonCau}}^{\text{VNC}}(\epsilon) \geq C_{\text{Cau}}^{\text{VNC}}(\epsilon)$. Our main result concerning the VNC states that the difference between $C_{\text{NonCau}}^{\text{VNC}}(\epsilon)$ and $C_{\text{Cau}}^{\text{VNC}}(\epsilon)$ is at most on the order of ϵ^4 , and is hence negligible compared to $C_{\text{Cau}}^{\text{VNC}}(\epsilon)$ (or $C_{\text{NonCau}}^{\text{VNC}}(\epsilon)$).

Theorem 6: For the state-dependent VNC (64),

$$\limsup_{\epsilon \downarrow 0} \frac{C_{\text{NonCau}}^{\text{VNC}}(\epsilon) - C_{\text{Cau}}^{\text{VNC}}(\epsilon)}{\epsilon^4} \leq \frac{9}{2}\psi_{\max}^4, \quad (72)$$

where

$$\psi_{\max} \triangleq \max_{x, y, s} |\psi(y|x, s)|. \quad (73)$$

Proof: Under a joint distribution on $\mathcal{S} \times \mathcal{U} \times \mathcal{Y}$ of the form

$$P_{SUY}(s, u, y) = P_S(s)P_{U|S}(u|s)W(y|u(s), s), \quad (74)$$

and given channel law (64), the conditional distribution of $Y = y$ given $U = u$ is

$$\widehat{W}(y|u) = \beta(y) \left[1 + \epsilon\hat{\psi}(y|u) \right] \quad (75)$$

where

$$\hat{\psi}(y|u) \triangleq \sum_s P_{S|U}(s|u)\psi(y|u(s), s) \quad (76)$$

for all u, y .

Consider the capacity formula (71). We fix the second term on the right-and side of (71) as a function of ϵ

$$I(U; S) = \alpha(\epsilon) \quad (77)$$

and upper-bound the first term. Let R denote the capacity-achieving output distribution in the *causal* case.⁶ We have

$$\begin{aligned} I(U; Y) &= \mathbb{E} \left[D \left(\widehat{W}(\cdot|U) \middle\| P_U \widehat{W} \right) \right] \\ &= \mathbb{E} \left[D \left(\widehat{W}(\cdot|U) \middle\| R \right) \right] - D \left(P_U \widehat{W} \middle\| R \right) \\ &\leq \mathbb{E} \left[D \left(\widehat{W}(\cdot|U) \middle\| R \right) \right] \\ &= \sum_{u, y} P_U(u) \widehat{W}(y|u) \log \frac{\widehat{W}(y|u)}{R(y)} \end{aligned}$$

⁶The exact form of R is unknown, and is not needed for our proof.

$$\begin{aligned}
&= \sum_{u,y} P_U(u) \widehat{W}(y|u) \left(\log \frac{\widehat{W}(y|u)}{\overline{W}(y|u)} + \log \frac{\overline{W}(y|u)}{R(y)} \right) \\
&= \sum_{u,y} P_U(u) \widehat{W}(y|u) \log \frac{\widehat{W}(y|u)}{\overline{W}(y|u)} \\
&\quad + \sum_{u,y} P_U(u) \left(\widehat{W}(y|u) - \overline{W}(y|u) \right) \log \frac{\overline{W}(y|u)}{R(y)} \\
&\quad + \sum_{u,y} P_U(u) \overline{W}(y|u) \log \frac{\overline{W}(y|u)}{R(y)} \\
&= \sum_u P_U(u) D\left(\widehat{W}(\cdot|u) \parallel \overline{W}(\cdot|u)\right) \\
&\quad + \sum_{u,y} P_U(u) \left(\widehat{W}(y|u) - \overline{W}(y|u) \right) \log \frac{\overline{W}(y|u)}{R(y)} \\
&\quad + \sum_u P_U(u) D(\overline{W}(\cdot|u) \parallel R) \\
&\leq \sum_u P_U(u) D\left(\widehat{W}(\cdot|u) \parallel \overline{W}(\cdot|u)\right) \\
&\quad + \sum_{u,y} P_U(u) \left(\widehat{W}(y|u) - \overline{W}(y|u) \right) \log \frac{\overline{W}(y|u)}{R(y)} \\
&\quad + C_{\text{Cau}}^{\text{VNC}}(\epsilon). \tag{78}
\end{aligned}$$

Here, the last step follows because R is chosen to be the capacity-achieving output distribution in the causal case, hence the Kuhn-Tucker conditions imply

$$D(\overline{W}(\cdot|u) \parallel R) \leq C_{\text{Cau}}^{\text{VNC}}(\epsilon), \quad u \in \mathcal{U}. \tag{79}$$

Consider the first term on the right-hand side of (78). Each relative entropy in it satisfies (see, e.g., [34, Lemma 17.3.3])

$$\lim_{\epsilon \downarrow 0} \frac{D\left(\widehat{W}(\cdot|u) \parallel \overline{W}(\cdot|u)\right)}{\sum_y \frac{\left(\widehat{W}(y|u) - \overline{W}(y|u)\right)^2}{2\widehat{W}(y|u)}} = 1, \tag{80}$$

where

$$\begin{aligned}
&\sum_y \frac{\left(\widehat{W}(y|u) - \overline{W}(y|u)\right)^2}{2\widehat{W}(y|u)} \\
&= \sum_y \frac{\beta(y)^2 \epsilon^2 \left(\sum_s (P_{S|U}(s|u) - P_S(s)) \psi(y|u(s), s) \right)^2}{2\widehat{W}(y|u)} \\
&\leq \sum_y \frac{\beta(y)^2 \epsilon^2 (2\delta(P_{S|U}(\cdot|u), P_S) \psi_{\max})^2}{2\widehat{W}(y|u)} \\
&\leq \sum_y \frac{2\beta(y)^2 \epsilon^2 \delta(P_{S|U}(\cdot|u), P_S) \psi_{\max}^2}{\widehat{W}(y|u)} \tag{81}
\end{aligned}$$

where ψ_{\max} is given in (73), and where $\delta(\cdot, \cdot)$ denotes the total variation distance between two distributions. The last step (81) follows because the total variation distance is upper-bounded by 1. Next note that $\widehat{W}(y|u)$ tends to $\beta(y)$ as $\epsilon \downarrow 0$, and that $\sum_y \beta(y) = 1$. These imply that the right-hand side of (81)

asymptotically equals $2\epsilon^2 \delta(P_{S|U}(\cdot|u), P_S) \psi_{\max}^2$ in the limit $\epsilon \downarrow 0$. We can thus combine (80) and (81) to

$$\limsup_{\epsilon \downarrow 0} \frac{D\left(\widehat{W}(\cdot|u) \parallel \overline{W}(\cdot|u)\right)}{\epsilon^4} \leq \limsup_{\epsilon \downarrow 0} \frac{2\delta(P_{S|U}(\cdot|u), P_S) \psi_{\max}^2}{\epsilon^2}. \tag{82}$$

Because \mathcal{U} is finite, we can simultaneously take expectations on both sides of (82) over u :

$$\limsup_{\epsilon \downarrow 0} \frac{\sum_u P_U(u) D\left(\widehat{W}(\cdot|u) \parallel \overline{W}(\cdot|u)\right)}{\epsilon^4} \leq \limsup_{\epsilon \downarrow 0} \frac{2\delta(P_{SU}, P_S \times P_U) \psi_{\max}^2}{\epsilon^2}. \tag{83}$$

Next note that, by (77) and Pinsker's inequality [31],

$$\delta(P_{SU}, P_S \times P_U) \leq \sqrt{\frac{\alpha(\epsilon)}{2}}, \tag{84}$$

so (83) further implies

$$\limsup_{\epsilon \downarrow 0} \frac{\sum_u P_U(u) D\left(\widehat{W}(\cdot|u) \parallel \overline{W}(\cdot|u)\right)}{\epsilon^4} \leq \limsup_{\epsilon \downarrow 0} \frac{\sqrt{2\alpha(\epsilon)}}{\epsilon^2} \cdot \psi_{\max}^2. \tag{85}$$

Next consider the second term on the right-hand side of (78). For all u and y , we have (assuming ϵ is small enough so that $\epsilon\psi_{\max} < 1$)

$$\begin{aligned} \log \frac{\overline{W}(y|u)}{R(y)} &\leq \log \frac{\beta(y)(1 + \epsilon\psi_{\max})}{\beta(y)(1 - \epsilon\psi_{\max})} \\ &\leq \frac{2\epsilon\psi_{\max}}{1 - \epsilon\psi_{\max}}. \end{aligned}$$

Similarly,

$$\log \frac{\overline{W}(y|u)}{R(y)} \geq -\frac{2\epsilon\psi_{\max}}{1 - \epsilon\psi_{\max}}. \tag{86}$$

Therefore we have the following upper bound:

$$\begin{aligned}
&\sum_{u,y} P_U(u) \left(\widehat{W}(y|u) - \overline{W}(y|u) \right) \log \frac{\overline{W}(y|u)}{R(y)} \\
&\leq \sum_{u,y} P_U(u) \left| \widehat{W}(y|u) - \overline{W}(y|u) \right| \cdot \frac{2\epsilon\psi_{\max}}{1 - \epsilon\psi_{\max}} \\
&= \sum_{u,y} P_U(u) \beta(y) \epsilon \left| \sum_s (P_{S|U}(s|u) - P_S(s)) \psi(y|u, s) \right| \\
&\quad \cdot \frac{2\epsilon\psi_{\max}}{1 - \epsilon\psi_{\max}} \\
&\leq \sum_{u,y} P_U(u) \beta(y) \epsilon (2\delta(P_{S|U}(\cdot|u), P_S) \psi_{\max}) \cdot \frac{2\epsilon\psi_{\max}}{1 - \epsilon\psi_{\max}} \\
&= \frac{4\epsilon^2 \psi_{\max}^2}{1 - \epsilon\psi_{\max}} \cdot \delta(P_{SU}, P_S \times P_U) \\
&\leq \frac{4\epsilon^2 \psi_{\max}^2}{1 - \epsilon\psi_{\max}} \cdot \sqrt{\frac{\alpha(\epsilon)}{2}}, \tag{87}
\end{aligned}$$

where the last step again follows from (84). Hence we obtain

$$\limsup_{\epsilon \downarrow 0} \frac{\sum_{u,y} P_U(u) \left(\widehat{W}(y|u) - \overline{W}(y|u) \right) \log \frac{\widehat{W}(y|u)}{R(y)}}{\epsilon^4} \leq \limsup_{\epsilon \downarrow 0} \frac{2\sqrt{2\alpha(\epsilon)}}{\epsilon^2} \cdot \psi_{\max}^2. \quad (88)$$

Combining (71), (77), (78), (85), and (88) we now have

$$\limsup_{\epsilon \downarrow 0} \frac{C_{\text{NonCau}}^{\text{VNC}}(\epsilon) - C_{\text{Cau}}^{\text{VNC}}(\epsilon)}{\epsilon^4} \leq \limsup_{\epsilon \downarrow 0} \frac{\max_{\alpha(\epsilon)} \left\{ 3\sqrt{2}\epsilon^2 \psi_{\max}^2 \sqrt{\alpha(\epsilon)} - \alpha(\epsilon) \right\}}{\epsilon^4}. \quad (89)$$

The maximization in (89) is achieved by

$$\alpha(\epsilon) = \frac{9\epsilon^4 \psi_{\max}^4}{2}, \quad (90)$$

which yields (72). \blacksquare

C. Comparison to CSI at Both Transmitter and Receiver

Consider the capacity of the VNC (64) with CSI on both sides, which we denote by $C_{\text{Both}}^{\text{VNC}}(\epsilon)$.

Proposition 7: For the channel (64),

$$\lim_{\epsilon \downarrow 0} \frac{C_{\text{Both}}^{\text{VNC}}(\epsilon) - C_{\text{Cau}}^{\text{VNC}}(\epsilon)}{\epsilon^2} \geq \frac{1}{2} \sum_{y,u} \beta(y) P_U^*(u) \left[\sum_s P_S(s) \eta^*(y|u(s), s)^2 - \left(\sum_s P_S(s) \eta^*(y|u(s), s) \right)^2 \right], \quad (91)$$

where P_U^* is the distribution that achieves the maximum in (69), and where

$$\eta^*(y|u(s), s) = \psi(y|u(s), s) - \sum_{u'} P_U^*(u') \psi(y|u'(s), s) \quad (92)$$

for all y, u, s .

Proof: When \mathbf{S} is known to both the transmitter and the receiver, a rate-splitting strategy is optimal, and the capacity equals the weighted sum over s of the capacities given each realization $S = s$:

$$\lim_{\epsilon \downarrow 0} \frac{C_{\text{Both}}^{\text{VNC}}(\epsilon)}{\epsilon^2} = \frac{1}{2} \max_{P_{X|S}} \sum_{s,y,x} P_S(s) \beta(y) P_{X|S}(x|s) \eta(y|x, s)^2 \quad (93)$$

where η is determined by $P_{X|S}$ as

$$\eta(y|x, s) = \psi(y|x, s) - \sum_{x'} P_{X|S}(x'|s) \psi(y|x', s) \quad (94)$$

for all x, y, s . To derive a lower bound on (93), we make a possibly suboptimal choice for $P_{X|S}$:

$$P_{X|S}(x|s) = \sum_{u: u(s)=x} P_U^*(u) \quad (95)$$

for all x, s . Then (93) implies

$$\lim_{\epsilon \downarrow 0} \frac{C_{\text{Both}}^{\text{VNC}}(\epsilon)}{\epsilon^2} \geq \frac{1}{2} \sum_{s,y,u} P_S(s) \beta(y) P_U^*(u) \eta(y|u(s), s)^2. \quad (96)$$

On the other hand, we can rewrite (69) as

$$\lim_{\epsilon \downarrow 0} \frac{C_{\text{Cau}}^{\text{VNC}}(\epsilon)}{\epsilon^2} = \frac{1}{2} \sum_{y,u} \beta(y) P_U^*(u) \left(\sum_s P_S(s) \eta(y|u(s), s) \right)^2. \quad (97)$$

Comparing (96) and (97) yields the desired result. \blacksquare

Due to the convexity of the function $a \mapsto a^2$, the right-hand side of (91) is in general positive. Hence, comparing Theorem 6 and Proposition 7 we see that, in terms of the capacity of the VNC, noncausal CSI at the transmitter is almost only as beneficial as causal CSI, whereas additional CSI at the receiver is significantly more beneficial.

VI. CONCLUDING REMARKS

In this paper we have derived upper bounds on the capacities of two channels with CSI. Our main interest is in the continuous-time Poisson channel with a random time-varying dark current, where the values of the dark current are known to the transmitter as CSI, either causally or noncausally. Our capacity bound for the Poisson channel shows that the improvement in capacity from both causal and noncausal CSI vanishes like $O(\tau)$ when the coherence time τ of the dark current approaches zero. For a related side result, we consider the state-dependent VNC with noncausal CSI at the transmitter. For this channel our upper bound shows that the improvement in capacity due to noncausal CSI over causal CSI is negligible.

Our intuition regarding noncausal CSI is similar for both the Poisson channel and the VNC: the mutual information attainable over these channels is small, hence the dependence between the state and the auxiliary random variable in the Gel'fand-Pinsker formula must also be weak.

It is well-known that channels with noncausal CSI are closely related to Marton's inner bound on the capacity region of a broadcast channel [35], which employs binning. Our results in Theorems 3 and 6 imply that certain types of binning-based strategies are not useful on related broadcast channel models. This fact is related to, but not a consequence of previous capacity results on the Poisson and very noisy broadcast channels [12], [15].

APPENDIX

In this appendix we show a simple result: causal CSI cannot increase the capacity of a type of DMC on which the channel input and the state act "independently," in a sense we specify below.⁷

Consider a DMC with input $x \in \mathcal{X}$, state $s \in \mathcal{S}$, and output $y \in \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y}, \mathcal{S}$ are all finite. The state sequence is IID according to P_S and is independent of the message. Given input x and state s , the output Y is distributed according to $W(\cdot|x, s)$, which has the following form:

$$W(y|x, s) = V(y|x) + \phi(y|s), \quad x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}, \quad (98)$$

where V and ϕ are such that $W(\cdot|x, s)$ is a valid distribution on \mathcal{Y} for all $x \in \mathcal{X}$ and $s \in \mathcal{S}$.

⁷We are not aware whether this result has appeared in previous works or not.

Proposition 8: The capacity of the channel (98) with causal CSI at the transmitter equals its capacity without CSI.

Proof: The capacity of the channel (98) with causal CSI is given by Shannon [16] as $\max_{P_U} I(U; Y)$, where U takes value in the set of mappings from \mathcal{S} to \mathcal{X} and must be chosen independently of S . In this case,

$$P_{Y|U}(y|u) = \sum_s P_S(s)W(y|u(s), s), \quad u \in \mathcal{U}, y \in \mathcal{Y}. \quad (99)$$

We shall show that each mapping u induces the same output distribution as a random input that does not depend on s . Indeed, it follows immediately from (98) and (99) that

$$P_{Y|U}(y|u) = \sum_s P_S(s)[V(y|u(s)) + \phi(y|s)] \quad (100)$$

for all u, y . If we choose X randomly and independently of s according to distribution π_u :

$$\pi_u(x) = \sum_{s: u(s)=x} P_S(s), \quad x \in \mathcal{X}, \quad (101)$$

then the probability that $Y = y$ is given by

$$\begin{aligned} & \sum_{x,s} P_S(s)\pi_u(x)[V(y|x) + \phi(y|s)] \\ &= \sum_x \pi_u(x)V(y|x) + \sum_s P_S(s)\phi(y|s) \\ &= \sum_s P_S(s)V(y|u(s)) + \sum_s P_S(s)\phi(y|s) \\ &= \sum_s P_S(s)[V(y|u(s)) + \phi(y|s)], \end{aligned} \quad (102)$$

which is the same as $P_{Y|U}(y|u)$. Hence the transmitter choosing mapping u is equivalent to it choosing a random input according to π_u , but the latter can be done without CSI. This implies that there is no advantage in using Shannon strategies [16] over directly using input symbols. In particular, $\max_{P_U} I(U; Y)$, the capacity of (98) with causal CSI, cannot exceed $\max_{P_X} I(X; Y)$, the capacity without CSI. ■

ACKNOWLEDGEMENT

The author would like to thank Prakash Narayan, Aslan Tchamkerten, and the expert referee for their useful comments.

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] D. L. Snyder and I. B. Rhodes, "Some implications of the cutoff-rate criterion for coded direct-detection optical communication systems," *IEEE Trans. Inform. Theory*, vol. 26, pp. 327–338, May 1980.
- [5] J. L. Massey, "Capacity, cutoff rate, and coding for direct-detection optical channel," *IEEE Trans. Commun.*, vol. 29, pp. 1615–1621, Nov. 1981.
- [6] 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.
- [7] M. R. Frey, "Information capacity of the Poisson channel," *IEEE Trans. Inform. Theory*, vol. 37, pp. 244–256, Mar. 1991.
- [8] 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.
- [9] K. Chakraborty and P. Narayan, "The Poisson fading channel," *IEEE Trans. Inform. Theory*, vol. 53, pp. 2349–2364, July 2007.
- [10] P. Narayan and D. L. Snyder, "Signal set design for band-limited memoryless multiple-access channels with soft decision demodulation," *IEEE Trans. Inform. Theory*, vol. 33, pp. 539–556, July 1987.
- [11] A. Lapidoth and S. Shamai (Shitz), "The Poisson multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 488–501, Mar. 1998.
- [12] A. Lapidoth, İ. E. Telatar, and R. Urbanke, "On wide-band broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 3250–3258, Dec. 2003.
- [13] 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.
- [14] L. Lai, Y. Liang, and S. Shamai, "On the capacity bounds for Poisson interference channels," *IEEE Trans. Inform. Theory*, vol. 61, pp. 223–238, Jan. 2015.
- [15] H. Kim, B. Nachman, and A. El Gamal, "Superposition coding is almost always optimal for the Poisson broadcast channel," *IEEE Trans. Inform. Theory*, vol. 62, pp. 1782–1794, Apr. 2016.
- [16] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Research and Development*, vol. 2, pp. 289–293, 1958.
- [17] 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.
- [18] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [19] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2867–2877, Aug. 2005.
- [20] Y. Steinberg and S. Shamai (Shitz), "Achievable rates for the broadcast channel with states known at the transmitter," in *Proc. IEEE Int. Symp. Inform. Theory*, (Adelaide, Australia), pp. 2184–2188, Sept. 4–9, 2005.
- [21] A. Lapidoth and L. Wang, "The state-dependent semideterministic broadcast channel," *IEEE Trans. Inform. Theory*, vol. 59, pp. 2242–2251, Apr. 2013.
- [22] A. V. Kuznetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Peredachi Inf.*, vol. 10, Apr.–June 1974.
- [23] C. D. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inform. Theory*, vol. 29, pp. 731–739, Sept. 1983.
- [24] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.
- [25] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1639–1667, June 2002.
- [26] M. Kesal and U. Erez, "An upper bound on the capacity of the causal dirty-paper channel," *IEEE Trans. Inform. Theory*, vol. 57, pp. 5650–5659, Sept. 2011.
- [27] F. Dupuis, W. Yu, and F. M. J. Willems, "Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information," in *Proc. IEEE Int. Symp. Inform. Theory*, (Chicago, Illinois, USA), June 27 – July 2, 2004.
- [28] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, pp. 3–19, Jan. 1965.
- [29] C. E. Shannon, "A mathematical theory of communication," *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.
- [30] L. Wang, "On the continuous-time Poisson channel with varying dark current known to the transmitter," in *Proc. IEEE Int. Symp. Inform. Theory*, (Barcelona, Spain), July 10–15 2016.
- [31] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [32] 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.
- [33] R. S. Lipster and A. N. Shiryaev, *Statistics of Random Processes II: Applications*. Springer Verlag, 2nd ed., 2001.
- [34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, second ed., 2006.
- [35] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 306–311, May 1979.

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.