The Poisson Fading Channel

Kaushik Chakraborty, *Member, IEEE*, and Prakash Narayan, *Fellow, IEEE*

Abstract—In this first paper of a two-part series, a single-user single-input single-output (SISO) shot-noise-limited Poisson channel is considered over which an information signal is transmitted by modulating the intensity of an optical beam, and individual photon arrivals are counted at the photodetector receiver. The transmitted signal, which is chosen to satisfy peak and average constraints, undergoes multiplicative fading, which occurs over coherence time intervals of fixed duration. The fade coefficient (channel state) remains constant in each coherence interval, and varies across successive such intervals in an independent and identically distributed (i.i.d.) fashion. A single-letter characterization of the capacity of this channel is obtained when the receiver is provided with perfect channel state information (CSI) while the transmitter CSI can be imperfect. The asymptotic behavior of channel capacity in the low and high peak-signal-to-shot-noise ratio (SNR) regimes is studied.

Index Terms—Channel capacity, channel state information, coherence time, free-space optical communication, lognormal fading, Poisson fading channel, single input single output.

I. INTRODUCTION

THE development of the technology of free-space optical communication has long been motivated by deep space applications, which, in fact, have served as an important impetus for much of the literature on communication in the Poisson regime ([3], [16], [19], [23], [25], [28], [29], among others). Additionally, free space optics has emerged recently as an attractive technology for several applications, e.g., metro network extensions, last mile connectivity, fiber backup, radio-frequency (RF)-wireless backhaul, and enterprise connectivity [38]. The many benefits of wireless optical systems include rapid deployment time, high security, inexpensive components, seamless wireless extension of the optical fiber backbone, immunity from RF interference, and lack of licensing regulations. Consequently, free-space optical communication has received much attention in recent years [14], [15], [17], [22], [39], [41], [42].

In free-space optical communication links, atmospheric turbulence can cause random variations in the refractive index of air at optical wavelengths which, in turn, result in random fluctuations in both the intensity and phase of a propagating optical signal [18], [36]. Such fluctuations, which in practice can routinely exceed 10 dB, lead to an increase in link error probability thereby degrading communication performance [42]. The fluctuations in the intensity of the transmitted optical signal, termed "fading," can be modeled in terms of an ergodic log-normal process with a correlation time or "coherence time" interval of the order of 1–10 ms [34]. For the systems under consideration, data rates typically can be of the order of gigabits per second. Therefore, the free-space optical channel is a slowly varying fading channel with occasional deep fades that can affect millions of consecutive bits [15].

A general approach that is often followed to achieve higher rates of reliable communication over fading channels is to use estimates of the channel fade (also referred to as path gain or channel state) at the transmitter and the receiver. For RF communication, a comprehensive review can be found in [31]. In optical fading channels, instantaneous realizations of the channel state can be estimated at the receiver; at typical data rates, more than 10^6 bits are transmitted during each coherence time, a small fraction of which can be used by the receiver to form good estimates of the channel fade.[1] Then, depending on the availability of a feedback link and the amount of acceptable delay, the transmitter can be provided with complete or partial knowledge of the channel state, which can be used for adaptive power control, thereby achieving higher throughputs (cf., e.g., [5], [31] in the context of RF communication).

Another approach for combating the detrimental effects of fading entails the use of spatial diversity in the form of multiple transmit and receive elements. In RF communication, the use of multiple transmit and receive antennas has been shown to significantly improve communication throughput in the presence of channel fading; for a survey of recent results, see [11]. In the context of free-space optical communication, in several experimental studies [2], [19], [21], multiple laser beams were shown to improve communication performance. Attempts have since been made to characterize analytically the benefits of multiple-input multiple-output (MIMO) communication over optical fading channels [14], [15], [22], [39].

The study of the Shannon capacity of the direct detection optical channel *without fading*, known popularly as the *Poisson channel*, has a rich history, beginning with [9]; see [37] for

K. Chakraborty was with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA. He is now with the California Institute for Telecommunications and Information Technology, University of California at San Diego, La Jolla, CA 92037 USA (e-mail: kchakrab@ucsd.edu).

P. Narayan is with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: prakash@eng.umd.edu).

[1]Clearly, the degree of accuracy of the receiver's estimated knowledge of the channel state (and the fraction of transmitted bits used for this purpose) will affect receiver performance. This issue is beyond the scope of the present work. For a related study in the context of RF communication, see, for instance, [27].

a literature review. Bar-David [1] examined the capacity of a Poisson channel with pulse position modulation (PPM), a model further advanced by Pierce [28]. Kabanov [20] applied martingale techniques to obtain the capacity of this channel subject to a peak constraint on the transmitted optical power, sans the PPM restriction. Davis [8] extended Kabanov's result by considering both peak and average constraints on the transmitted optical power. A new derivation of the Poisson channel capacity by Wyner [40] also felicitously yielded the entire reliability function of this channel for all rates below capacity. This capacity is achieved by information-bearing "rate functions" of unbounded bandwidth. Models that take into account bandwidth limitations have been considered by several authors (cf., e.g., [3], [23], [24], [32], [33]); the corresponding capacities and associated optimum coding and modulation techniques for the bandwidth-constrained Poisson channel are significantly different from those for the unbounded bandwidth Poisson channel. Of direct relevance to our work is [15], where ergodic and outage capacity issues for the single-user MIMO Poisson channel with random path gains are addressed; this channel is not subject to bandwidth constraints.

In a two-part series of papers, we consider several issues concerning reliable communication over a single-user optical channel with fading. These include the formulation of a block-fading channel model that takes into account the slowly varying nature of optical fade; the determination of channel capacity when the receiver has accurate estimates of channel state while the transmitter is provided with varying degrees of channel state information (CSI); a characterization of good transmitter power control strategies that achieve capacity; and the limiting behavior of channel capacity in low and high signal-to-noise power ratio (SNR) regimes.

In this first paper, we consider a single-user single-input single-output (SISO) Poisson channel with fading and limited by shot noise, thereafter referred to as the *Poisson fading channel*. An information signal is transmitted by modulating the intensity of an optical beam, and the receiver performs direct detection by counting individual photon arrivals at the photodetector. The (nonnegative) transmitted signal is constrained in its peak and average power. A block-fading channel model is introduced that accounts for the slowly varying nature of optical fade; the channel fade remains constant for a coherence time of a fixed duration, and varies across successive such intervals in an independent and identically distributed (i.i.d.) fashion. We provide a characterization of channel capacity for different levels of CSI at the transmitter and with perfect CSI at the receiver. While our model resembles that studied in [15], we restrict ourselves in this paper to the case of a SISO channel. In this setting, we consider a more general class of problems than those considered in [15], and from our capacity results, conclusions are drawn that differ considerably from those reported therein. A shortcoming of our model is that it ignores bandwidth limitations associated with the transmitter and receiver devices currently used in practice. We also ignore the effects of infrared and visible background light, and assume that the dark current at the photodetector is the dominant source of noise. However, these assumptions lead to a simple channel model which is amenable to an exact analysis. Other
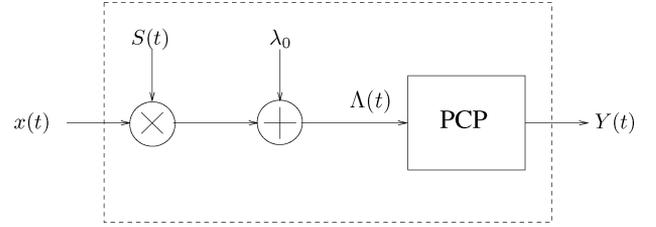

Fig. 1.   Poisson fading channel.

models have been proposed in the literature that describe the background noise as additive white Gaussian (cf., e.g., [17], [41]), and bounds on channel capacity have been computed.

In the sequel paper [6], we study a single-user MIMO Poisson fading channel with multiple transmit and receive apertures. Building on the results of the present paper, we provide in [6] a single-letter characterization of the MIMO channel capacity. We also identify pertinent properties of the optimal transmission strategy of the transmit apertures for a symmetric channel with isotropically distributed fading when the transmitter and the receiver have perfect CSI.

The remainder of this paper is organized as follows. Section II deals with the problem formulation. Our results are stated in Section III and proved in Section IV. An illustrative example of a channel with a lognormal fade is considered in Section V. Finally, Section VI contains our concluding remarks.

II. PROBLEM FORMULATION

The following notation will be used throughout this paper. Random variables (rvs) are denoted individually by upper case letters and collectively by bold upper case letters. We use the notation $\boldsymbol{X}_i^j$ to denote the collection of rvs $\{X_i, X_{i+1}, \ldots, X_j\}$; when $i = 1$, we use $\boldsymbol{X}^j = \{X_1, \ldots, X_j\}$. A continuous-time random process $\{X(t), a \leq t \leq b\}$ is denoted in shorthand notation by X_a^b; when $a = 0$, we use $X^b = \{X(t), 0 \leq t \leq b\}$. Realizations of rvs and random processes, which are denoted in lower case letters, follow the same convention.

A block schematic diagram of the channel model is given in Fig. 1. For a given $\mathbb{R}_0^+$-valued[2] transmitted signal $\{x(t), t \geq 0\}$, the corresponding received signal at the channel output $Y^\infty = \{Y(t), t \geq 0\}$ is a $\mathbb{Z}_0^+$-valued nondecreasing (left-continuous) Poisson counting process (PCP) with rate (or intensity) equal to

$$\Lambda(t) = S(t)x(t) + \lambda_0, \qquad t \geq 0 \qquad (1)$$

where $\{S(t), t \geq 0\}$ is a $\mathbb{R}_0^+$-valued random fade, and $\lambda_0 \geq 0$ is the background noise (dark current) rate which is assumed to be constant. Note that Y^∞ is an independent increments process with $Y(0) = 0$, and for $0 \leq \tau, t < \infty$

$$\Pr\left\{Y(t+\tau) - Y(t) = j | \Lambda_t^{t+\tau} = \lambda_t^{t+\tau}\right\}$$
$$= \frac{1}{j!}e^{-\Gamma(\lambda_t^{t+\tau})}\Gamma^j(\lambda_t^{t+\tau}), \qquad j = 0, 1, \ldots$$

---

[2]The set of nonnegative real numbers is denoted by $\mathbb{R}_0^+$, the set of positive integers by $\mathbb{Z}^+$, and the set of nonnegative integers by $\mathbb{Z}_0^+$.
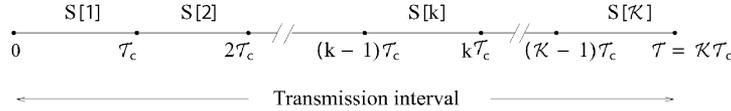
Fig. 2. Block-fading channel.

where $\Gamma(\lambda_t^{t+\tau}) = \int_t^{t+\tau} \lambda(u)du$. Physically, the jumps in Y^∞ correspond to the arrival of photons in the receiver. Let $[0, \mathcal{T}]$ denote the time interval of signal transmission and reception. Let $\Sigma(\mathcal{T})$ denote the space of nondecreasing, left-continuous, piecewise-constant, $\mathbb{Z}_0^+$-valued functions $\{g(t), 0 \leq t \leq \mathcal{T}\}$ with $g(0) = 0$. The output process $Y^\mathcal{T} = \{Y(t), 0 \leq t \leq \mathcal{T}\}$ takes values in $\Sigma(\mathcal{T})$.

The input to the channel is an $\mathbb{R}_0^+$-valued signal $x^\mathcal{T} = \{x(t), 0 \leq t \leq \mathcal{T}\}$ which is proportional to the transmitted optical power, and which satisfies peak and average "power" constraints of the form

$$0 \leq x(t) \leq \alpha, \qquad 0 \leq t \leq \mathcal{T}$$
$$\frac{1}{\mathcal{T}} \int_0^\mathcal{T} x(t)dt \leq \sigma\alpha \qquad (2)$$

where the peak power $\alpha > 0$ and the ratio of average-to-peak power σ, $0 \leq \sigma \leq 1$, are fixed.

The channel fade, i.e., path gain, is modeled by a $\mathbb{R}_0^+$-valued random process $\{S(t), t \geq 0\}$ of known distribution. The channel coherence time $\mathcal{T}_c$ is a measure of the intermittent coherence of the time-varying channel fade. We assume that the channel fade remains fixed over time intervals of width $\mathcal{T}_c$, and varies in an i.i.d. manner across successive such intervals. For $k \in \mathbb{Z}^+$, let the channel fade during $((k-1)\mathcal{T}_c, k\mathcal{T}_c]$ be denoted by the rv $S[k]$, i.e.,

$$S(t) = S[k], \qquad t \in ((k-1)\mathcal{T}_c, k\mathcal{T}_c], \; k \in \mathbb{Z}^+$$

and $S(0) = S[1]$. See Fig. 2. The channel fade is then described by the random sequence $\boldsymbol{S}^\infty = \{S[1], S[2], \ldots\}$ of i.i.d. repetitions of a rv S with known distribution. Note that the rate of the received signal Y^∞ in (1) is now given by[3]

$$\Lambda(t) = S[\lceil t/\mathcal{T}_c \rceil]x(t) + \lambda_0, \qquad t \geq 0. \qquad (3)$$

In an illustrative example will be discussed in Section V, we shall consider a channel fade with a normalized lognormal distribution, as advocated in [15].

We shall assume throughout that the receiver has perfect CSI. This assumption, justified in practice,[4] is pivotal for the mathematical tractability of our analysis. Various degrees of CSI can be made available to the transmitter. In general, we can model the CSI available at the transmitter in terms of a mapping $h : \mathbb{R}_0^+ \to \mathbb{U}$, where $\mathbb{U}$ is an arbitrary subset of $\mathbb{R}_0^+$, not necessarily finite. For $\boldsymbol{S}^\infty = \boldsymbol{s}^\infty$, the transmitter (respectively, receiver) is

provided with CSI $u[k] = h(s[k])$ (respectively, $s[k]$) during $((k-1)\mathcal{T}_c, k\mathcal{T}_c]$, $k \in \mathbb{Z}^+$. Let $\{U[k] = h(S[k]), k \in \mathbb{Z}^+\}$ denote the CSI at the transmitter, hereafter referred to as the transmitter CSI h. We shall be particularly interested in two special cases of this general framework. In the first case, the transmitter has perfect CSI, i.e., h is the identity mapping. In the second case, the transmitter is provided no CSI, i.e., h is the trivial (constant) mapping.

We assume, without loss of generality, that the signal transmission duration $\mathcal{T}$ is an integer multiple of the channel coherence time $\mathcal{T}_c$, i.e., $\mathcal{T} = \mathcal{K}\mathcal{T}_c$, $\mathcal{K} \in \mathbb{Z}^+$. For the channel under consideration, a $(\mathcal{W}, \mathcal{T})$–code (f, ϕ) is defined as follows.

1) For each $\boldsymbol{u}^\mathcal{K} \in \mathbb{U}^\mathcal{K}$, the codebook comprises a set of $\mathcal{W}$ waveforms

$$f(w, \boldsymbol{u}^\mathcal{K}) = \{x_w(t, \boldsymbol{u}^{\lceil t/\mathcal{T}_c \rceil}), 0 \leq t \leq \mathcal{T}\}$$

$w \in \mathbb{W} = \{1, \ldots, \mathcal{W}\}$, satisfying peak and average power constraints which follow from (2)

$$0 \leq x_w(t, \boldsymbol{u}^{\lceil t/\mathcal{T}_c \rceil}) \leq \alpha, \qquad 0 \leq t \leq \mathcal{T}$$
$$\frac{1}{\mathcal{T}} \int_0^\mathcal{T} x_w\left(t, \boldsymbol{u}^{\lceil t/\mathcal{T}_c \rceil}\right) dt \leq \sigma\alpha \qquad (4)$$

for all $w \in \mathbb{W}$. Note that the transmitter output $x_w(t, \cdot)$ at time t is allowed to be a function of the (causal) transmitter CSI $\boldsymbol{u}^{\lceil t/\mathcal{T}_c \rceil}$.

2) The decoder is a mapping $\phi : \Sigma(\mathcal{T}) \times (\mathbb{R}_0^+)^\mathcal{K} \to \mathbb{W}$.

For each message $w \in \mathbb{W}$ and transmitter CSI $\boldsymbol{u}^\mathcal{K} \in \mathbb{U}^\mathcal{K}$ corresponding to the fade $\boldsymbol{s}^\mathcal{K} \in (\mathbb{R}_0^+)^\mathcal{K}$, the transmitter sends a waveform $x_w(t, \boldsymbol{u}^{\lceil t/\mathcal{T}_c \rceil})$, $0 \leq t \leq \mathcal{T}$, over the channel. The receiver, upon observing $y^\mathcal{T}$ and being provided with $\boldsymbol{s}^\mathcal{K}$, produces an output $\hat{w} = \phi(y^\mathcal{T}, \boldsymbol{s}^\mathcal{K})$. The rate of this $(\mathcal{W}, \mathcal{T})$–code is $\mathcal{R} = \frac{1}{\mathcal{T}} \ln \mathcal{W}$ nats/s, and its average probability of decoding error is

$$P_e(f, \phi)$$
$$= \frac{1}{\mathcal{W}} \sum_{w=1}^\mathcal{W} \mathbb{E}\left[\Pr\left\{ \phi(Y^\mathcal{T}, \boldsymbol{S}^\mathcal{K}) \neq w \, \middle| \, x_w^\mathcal{T}(\boldsymbol{U}^\mathcal{K}), \boldsymbol{S}^\mathcal{K} \right\} \right] \quad (5)$$

where we have used the shorthand notation

$$x_w^\mathcal{T}\left(\boldsymbol{U}^\mathcal{K}\right) = \left\{ x_w\left(t, \boldsymbol{U}^{\lceil t/\mathcal{T}_c \rceil}\right), \; 0 \leq t \leq \mathcal{T} \right\}, \qquad w \in \mathbb{W}. \tag{6}$$

Definition 1: Let α, λ_0, σ, $\mathcal{T}_c$ be fixed. Given $0 < \epsilon < 1$, a number $\mathcal{R} > 0$ is an ϵ-achievable rate if for every $\delta > 0$ and for all $\mathcal{T}$ sufficiently large, there exist $(\mathcal{W}, \mathcal{T})$–codes (f, ϕ) such that $\frac{1}{\mathcal{T}} \ln \mathcal{W} > \mathcal{R} - \delta$ and $P_e(f, \phi) < \epsilon$; $\mathcal{R}$ is an achievable rate if it is ϵ-achievable for all $0 < \epsilon < 1$. The supremum of all achievable rates is the capacity of the channel, and will be denoted by $\mathcal{C}$.

---

[3]We denote by $\lceil x \rceil$ the smallest integer greater than or equal to x.

[4]The channel coherence time $\mathcal{T}_c$ is typically of the order of 1–10 ms, during which 1–10 Mb can be transmitted at a rate of 1 Gb/s. A negligible fraction of the bits transmitted and received during a coherence time can be used by the receiver to estimate the prevailing fade value.

III. STATEMENT OF RESULTS

Our first and main result provides a single-letter characterization of the capacity of the Poisson fading channel model described above. The two special cases of perfect and no CSI at the transmitter are considered next. Finally, we analyze the limiting behavior of capacity in the high and low peak-signal-to-shot-noise ratio (SNR) regimes, viz. in the limits as $\lambda_0 \to 0$ and $\lambda_0 \to \infty$, respectively. Recall our standing assumption throughout that the receiver has perfect CSI.

For stating our results, it is convenient to set

$$\zeta(x, y) \triangleq (x+y)\ln(x+y) - y\ln y, \qquad x, y \geq 0 \quad (7)$$

(with $0\ln 0 \triangleq 0$), and

$$\xi(x) \triangleq \frac{1}{x}\left(e^{-1}(1+x)^{(1+1/x)} - 1\right), \qquad x \geq 0 \quad (8)$$

whence it can be verified that for $x, y \geq 0$

$$\zeta(x, y) - x(1+\ln y) = x\ln(1 + \xi(x/y)x/y). \quad (9)$$

Note that $\zeta(\cdot, y)$ is strictly convex on $[0, \infty)$ for every $y \geq 0$. *We shall assume the following technical conditions throughout, which are satisfied by a broad class of distributions[5] for S*:

$$\Pr\{S > 0\} = 1$$
$$\mathbb{E}[S] < \infty$$
$$\mathbb{E}[|\zeta(S\alpha, \lambda_0)|] < \infty, \qquad \forall\, \lambda_0 \geq 0. \quad (10)$$

Before stating our main capacity result for the Poisson fading channel, we provide a heuristic argument that progressively anticipates its formula. First, let $S = s$ with probability 1; then, the corresponding capacity is clearly (cf., e.g., [40, Theorem 1] or [8, Theorem 1])

$$\max_{0 \leq \mu \leq \sigma} \mu\zeta(s\alpha, \lambda_0) - \zeta(\mu s\alpha, \lambda_0) \quad (11)$$

and is achieved by an i.i.d. $\{0, \alpha\}$-valued channel input. The capacity value in (11) is to be interpreted as a maximum with respect to the parameter μ of the conditional mutual information between the channel input and output conditioned on $S = s$, where μ is the probability of the channel input taking the value α. Second, with the rv S as in the previous section, let the receiver possess perfect CSI while the transmitter has no CSI. The corresponding capacity can be expected to be a similar maximum of the averaged conditional mutual information, where the average is taken over the distribution of the rv S; by analogy with (11), its value would then be

$$\max_{0 \leq \mu \leq \sigma} \mathbb{E}\left[\mu\zeta(S\alpha, \lambda_0) - \zeta(\mu S\alpha, \lambda_0)\right]. \quad (12)$$

Finally, we turn to the case at hand, namely, when the receiver has perfect CSI while the transmitter CSI is given in terms of the rv $U = h(S)$, a function of the receiver CSI. The capacity value should once again be a maximum of the averaged conditional mutual information as above, where the channel input now has

a conditional distribution conditioned on the rv U; thus, this last conditional mutual information, in effect, is a convexification with respect to U of the counterparts of the sort which led to (12); also, the maximization with respect to μ (which now is a function of transmitter CSI) should be subject to an overall average power constraint. Hence, the anticipated capacity value is

$$\max_{\substack{\mu:\mathbb{U}\to[0,1] \\ \mathbb{E}[\mu(U)]\leq\sigma}} \mathbb{E}\left[\mu(U)\zeta(S\alpha, \lambda_0) - \zeta(\mu(U)S\alpha, \lambda_0)\right]. \quad (13)$$

A formal proof of (13) as the capacity expression for the Poisson fading channel with transmitter CSI h could be constructed along the lines above. However, this would entail establishing rigorously the appropriate maxima of conditional mutual information quantities involving mixed rvs as the capacities in the second and last steps above, combined with showing that their values are then (12) and (13), respectively. Instead, we prefer to give in Section IV an elemental proof which is directly crafted for the problem at hand, thereby avoiding a derivation of (more general) capacity results for mixed alphabet channels with CSI[6] from which the sought-after expressions then must be extracted.

Theorem 1: Let α, λ_0, σ, $\mathcal{T}_c$ be fixed. The capacity for transmitter CSI h is

$$\mathcal{C} = \max_{\substack{\mu:\mathbb{U}\to[0,1] \\ \mathbb{E}[\mu(U)]\leq\sigma}} \mathbb{E}\left[\mu(U)\zeta(S\alpha, \lambda_0) - \zeta(\mu(U)S\alpha, \lambda_0)\right] \quad (14)$$

where $U = h(S)$.

The capacities for the special cases of perfect and no CSI at the transmitter follow directly from Theorem 1.

Corollary 1: The capacity for perfect CSI at the transmitter is

$$\mathcal{C}_{\mathrm{P}} = \max_{\substack{\mu:\mathbb{R}_0^+\to[0,1] \\ \mathbb{E}[\mu(S)]\leq\sigma}} \mathbb{E}\left[\mu(S)\zeta(S\alpha, \lambda_0) - \zeta(\mu(S)S\alpha, \lambda_0)\right].$$

$$(15)$$

The capacity for no CSI at the transmitter is

$$\mathcal{C}_{\mathrm{N}} = \max_{0 \leq \mu \leq \sigma} \mathbb{E}\left[\mu\zeta(S\alpha, \lambda_0) - \zeta(\mu S\alpha, \lambda_0)\right]. \quad (16)$$

Remarks:

(i) The optimization in (14) (as also in (15), (16)) is that of a concave functional over a convex compact set, so that the maximum clearly exists.

(ii) The assumption of perfect receiver CSI is consequential for the capacity result of Theorem 1. From this result, we see that the value of the capacity does not depend on the coherence time $\mathcal{T}_c$. Conditioned on the transmitted signal $\{x(t), 0 \leq t \leq \mathcal{T}\}$, and perfect receiver CSI $\boldsymbol{s}^{\mathcal{K}}$, the segments of the received signal in different coherence intervals, viz. $Y_{(k-1)\mathcal{T}_c}^{k\mathcal{T}_c}$, $k = 1, \ldots, \mathcal{K}$, are (conditionally) independent across the coherence intervals; hence, it suffices to look at a single coherence interval in the mutual information computations. Furthermore, in a single coherence interval, conditioned on perfect receiver CSI, the optimality of i.i.d. transmitted signals leads to a lack of dependence of capacity on $\mathcal{T}_c$. The fact that the capacity of a block-fading channel with perfect receiver CSI does not

---

[5]In particular, the lognormal distribution, which is widely used to describe the channel fade in free-space optical channels [15], [22], [34], [39], [42], satisfies these conditions.

[6]See [4], [7] for channels with finite alphabets.

depend on the block size has been reported in the literature in various settings in other contexts (cf., e.g., [26] for such a result on block interference channels).

(iii) Our proof of the achievability part of Theorem 1, based on [40], shows that $\{0, \alpha\}$-valued transmitted signals which are i.i.d. (conditioned on the current transmitter CSI) with arbitrarily fast intertransition times, can achieve capacity. This is in concordance with previous results (e.g., [40]), where the optimality of binary signaling for Poisson channels has been established.

(iv) The optimizing "power control law" μ in (14), (15), and (16) depicts the probability with which the transmitter picks the signal level α depending on the available transmitter CSI. Thus, it can be interpreted as the optimal average conditional "duty cycle" of the transmitted signal as a function of transmitter CSI.

(v) Some consequences of imperfect receiver CSI are discussed in Section VI.

Theorem 2: The optimal power control law $\mu^* : \mathbb{U} \to [0, 1]$ that achieves the maximum in (14) is given as follows. For $\rho \geq 0$, $u \in \mathbb{U}$, let $\mu = \mu_\rho(u)$ be the solution of the equation

$$\mathbb{E}\left[S\alpha \ln\left(\frac{1 + \xi(S\alpha/\lambda_0)S\alpha/\lambda_0}{1 + \mu S\alpha/\lambda_0} \right) \bigg| U = u \right] = \rho, \quad (17)$$

and define

$$\sigma_0 \triangleq \mathbb{E}[\mu_0(U)]. \quad (18)$$

If $\sigma_0 > \sigma$, let $\rho = \rho^* > 0$ be the solution of the equation[7]

$$\mathbb{E}\left[[\mu_\rho(U)]^+ \right] = \sigma. \quad (19)$$

Then the optimal power control law μ^* in (14) is given by

$$\mu^*(u) = \begin{cases} \mu_0(u), & \sigma_0 \leq \sigma \\ [\mu_{\rho^*}(u)]^+, & \sigma_0 > \sigma, \end{cases} \quad u \in \mathbb{U}. \quad (20)$$

The following corollary particularizes the previous optimal power control law in the special cases of perfect and no CSI at the transmitter.

Corollary 2: For perfect CSI at the transmitter, the optimal power control law $\mu^* : \mathbb{R}_0^+ \to [0, 1]$ that achieves the maximum in (15) is given as follows. For $\rho \geq 0$, $s \in \mathbb{R}_0^+$, let

$$\mu_\rho(s) \triangleq \frac{\lambda_0}{\alpha s} \left(e^{-\left(1 + \frac{\rho}{\alpha s}\right)} \left(1 + \frac{\alpha s}{\lambda_0} \right)^{\left(1 + \frac{\lambda_0}{\alpha s}\right)} - 1 \right) \quad (21)$$

and let

$$\sigma_0 \triangleq \mathbb{E}[\mu_0(S)]. \quad (22)$$

If $\sigma_0 > \sigma$, let $\rho = \rho^* > 0$ be the solution of the equation $\mathbb{E}\left[[\mu_\rho(S)]^+ \right] = \sigma$. Then the optimal power control law μ^* in (15) is given by

$$\mu^*(s) = \begin{cases} \mu_0(s), & \sigma_0 \leq \sigma, \\ [\mu_{\rho^*}(s)]^+, & \sigma_0 > \sigma, \end{cases} \quad s \in \mathbb{R}_0^+. \quad (23)$$

[7]We denote $\max\{x, 0\}$ by $[x]^+$.

For no CSI at the transmitter, the maximum in (16) is achieved by

$$\mu^* = \min\{\sigma, \mu_0\} \quad (24)$$

with μ_0 being the solution of the equation

$$\mathbb{E}\left[S\alpha \ln\left(\frac{1 + \xi(S\alpha/\lambda_0)S\alpha/\lambda_0}{1 + \mu_0 S\alpha/\lambda_0} \right) \right] = 0. \quad (25)$$

Remarks:

(i) It can be verified that the left-hand side of (19) decreases monotonically from σ_0 to 0 as ρ increases from 0 to ∞, so that for each $0 \leq \sigma < \sigma_0$, there exists a unique $\rho = \rho^* > 0$ that solves (19). Furthermore, it can be shown that for each $u \in \mathbb{U}$, $0 \leq \mu^*(u) \leq 1/2$, so that the power control law given by (20) is well defined.

(ii) For the case of perfect transmitter CSI, the optimal power control law in (23) differs from, and yields a higher capacity value than, the claimed optimal power control law in [15, eq. (4)]. An example in Section V shows the difference in the values of capacity when computed using the two power control laws for a range of values of α/λ_0 and σ.

The peak signal-to-noise ratio, denoted SNR, is defined as $\text{SNR} = \alpha/\lambda_0$. We characterize next the capacity in the high- and low-SNR regimes (equivalently, the low- and high-shot-noise regimes) when the peak signal power α is fixed.

Theorem 3: In the high-SNR regime in the limit as $\lambda_0 \to 0$, the capacity for transmitter CSI h is

$$\mathcal{C}^{\text{H}} = \mathbb{E}\left[\mu^{\text{H}}(U)\zeta(S\alpha, 0) - \zeta(\mu^{\text{H}}(U)S\alpha, 0) \right] \quad (26)$$

with $U = h(S)$, where the optimal power control law $\mu^{\text{H}} : \mathbb{U} \to [0, 1]$ is given as follows: if $\sigma < 1/e$, let $\rho = \rho^* > 0$ be the solution of the equation $\mathbb{E}\left[e^{-\rho/\alpha\mathbb{E}[S|U]} \right] = e\sigma$; then μ^{H} is given by

$$\mu^{\text{H}}(u) = \begin{cases} e^{-1-\rho^*/\alpha\mathbb{E}[S|U=u]}, & \sigma < 1/e \\ e^{-1}, & \sigma \geq 1/e, \end{cases} \quad u \in \mathbb{U}. \quad (27)$$

In the low-SNR regime for $\lambda_0 \gg 1$, the capacity is[8]

$$\mathcal{C}^{\text{L}} = \mu^{\text{L}}(1 - \mu^{\text{L}})\mathbb{E}[S^2]\alpha^2/2\lambda_0 + O\left(\lambda_0^{-2}\right) \quad (28)$$

where $\mu^{\text{L}} = \min\{\sigma, 1/2\}$.

In the next two immediate corollaries of Theorem 3, we characterize the capacity in the high- and low-SNR regimes, in the special cases when the transmitter is provided with perfect and no CSI, respectively.

Corollary 3: For perfect CSI at the transmitter, in the high-SNR regime in the limit as $\lambda_0 \to 0$, the capacity is

$$\mathcal{C}_{\text{P}}^{\text{H}} = \mathbb{E}\left[\mu^{\text{H}}(S)\zeta(S\alpha, 0) - \zeta(\mu^{\text{H}}(S)S\alpha, 0) \right] \quad (29)$$

[8]By the standard notation $f(x) = O(g(x))$, we mean that there exists a number $0 \leq A < \infty$, not depending on x, such that $f(x) \leq Ag(x) \, \forall \, x \geq x_0$, where $x_0 = x_0(A) < \infty$.

where the optimal power control law $\mu^{\mathrm{H}} : \mathbb{R}_0^+ \to [0, 1]$ is given as follows: if $\sigma < 1/e$, let $\rho = \rho^* > 0$ be the solution of $\mathbb{E}\left[e^{-\rho/S\alpha}\right] = e\sigma$; then μ^{H} is given by

$$\mu^{\mathrm{H}}(s) = \begin{cases} e^{-1-\rho^*/\alpha s}, & \sigma < 1/e \\ e^{-1}, & \sigma \geq 1/e, \ s \in \mathbb{R}_0^+. \end{cases} \qquad (30)$$

In the low-SNR regime for $\lambda_0 \gg 1$, the capacity is the same as in (28), i.e.,

$$\mathcal{C}_{\mathrm{P}}^{\mathrm{L}} = \mu^{\mathrm{L}}(1 - \mu^{\mathrm{L}})\mathbb{E}[S^2]\alpha^2/2\lambda_0 + O\left(\lambda_0^{-2}\right) \qquad (31)$$

where $\mu^{\mathrm{L}} = \min\{\sigma, 1/2\}$.

Corollary 4: For no CSI at the transmitter, in the high-SNR regime in the limit as $\lambda_0 \to 0$, the capacity is

$$\mathcal{C}_{\mathrm{N}}^{\mathrm{H}} = \mathbb{E}\left[\mu^{\mathrm{H}}\zeta(S\alpha, 0) - \zeta(\mu^{\mathrm{H}}S\alpha, 0)\right] \qquad (32)$$

where $\mu^{\mathrm{H}} = \min\{\sigma, 1/e\}$.

In the low-SNR regime for $\lambda_0 \gg 1$, the capacity is the same as in (28), i.e.,

$$\mathcal{C}_{\mathrm{N}}^{\mathrm{L}} = \mu^{\mathrm{L}}(1 - \mu^{\mathrm{L}})\mathbb{E}[S^2]\alpha^2/2\lambda_0 + O\left(\lambda_0^{-2}\right) \qquad (33)$$

where $\mu^{\mathrm{L}} = \min\{\sigma, 1/2\}$.

Remark: In the low-SNR regime for $\lambda_0 \gg 1$, we see from (28) that transmitter CSI does not improve capacity. The capacity increases with SNR (approximately) linearly with a slope proportional to $\mu^{\mathrm{L}}(1 - \mu^{\mathrm{L}})$, where the optimal duty cycle $\mu^{\mathrm{L}} = \min\{1/2, \sigma\}$ does not depend on the transmitter CSI. However, in the high-SNR regime as $\lambda_0 \to 0$, if $\sigma < 1/e$, the optimal duty cycle μ^{H} (and hence the capacity in (26)) depends on transmitter CSI.[9]

IV. PROOFS

We begin this section with some additional definitions and observations that will be needed in our proofs. First, observe that for any $t > 0$, the number of photon arrivals N_t in $[0, t]$ together with the corresponding (ordered) arrival times $\boldsymbol{T}^{N_t} = (T_1, \ldots, T_{N_t})$ suffice to describe the statistics of Y^t, so that the rv $(N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}})$ is a complete description of $Y^{\mathcal{T}}$.

The channel is characterized as follows. For an input signal $x^{\mathcal{T}}$ satisfying (2) and a fade $\boldsymbol{S}^{\mathcal{K}} = \boldsymbol{s}^{\mathcal{K}}$, the channel output rv $(N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}})$ has the "conditional sample function density" (cf., e.g., [35])

$$f_{N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}}|X^{\mathcal{T}}, \boldsymbol{S}^{\mathcal{K}}}\left(n_{\mathcal{T}}, \boldsymbol{t}^{n_{\mathcal{T}}} \,\middle|\, x^{\mathcal{T}}, \boldsymbol{s}^{\mathcal{K}}\right)$$
$$= \begin{cases} \exp\left(-\int_0^{\mathcal{T}} \lambda(t)dt\right), & n_{\mathcal{T}} = 0 \\ \exp\left(-\int_0^{\mathcal{T}} \lambda(t)dt\right) \cdot \prod_{i=1}^{n_{\mathcal{T}}} \lambda(t_i), & n_{\mathcal{T}} \geq 1 \end{cases} \qquad (34)$$

where

$$\lambda(t) = s[\lceil t/\mathcal{T}_{\mathrm{c}}\rceil]x(t) + \lambda_0, \qquad 0 \leq t \leq \mathcal{T}.$$

---

[9]This is in contrast with the results in [15], where the authors have argued that at high SNR, a knowledge of the fade at the transmitter does not improve capacity.

In order to write the channel output sample function density conditioned only on the fade for a given joint distribution of $(X^{\mathcal{T}}, \boldsymbol{S}^{\mathcal{K}})$, consider the conditional mean of $X(t)$ (conditioned causally on the channel output and the fade) defined by

$$\hat{X}(t) \triangleq \mathbb{E}\left[X(t) \,\middle|\, N_t, \boldsymbol{T}^{N_t}, \boldsymbol{S}^{\lceil t/\mathcal{T}_{\mathrm{c}}\rceil}\right], \qquad 0 \leq t \leq \mathcal{T} \quad (35)$$

where we have suppressed for notational convenience the dependence of $\hat{X}(t)$ on $(N_t, \boldsymbol{T}^{N_t}, \boldsymbol{S}^{\lceil t/\mathcal{T}_{\mathrm{c}}\rceil})$. Define

$$\Lambda(t) \triangleq S[\lceil t/\mathcal{T}_{\mathrm{c}}\rceil]X(t) + \lambda_0, \qquad 0 \leq t \leq \mathcal{T} \qquad (36)$$

and

$$\hat{\Lambda}(t) \triangleq \mathbb{E}\left[\Lambda(t) \,\middle|\, N_t, \boldsymbol{T}^{N_t}, \boldsymbol{S}^{\lceil t/\mathcal{T}_{\mathrm{c}}\rceil}\right]$$
$$= S[\lceil t/\mathcal{T}_{\mathrm{c}}\rceil]\hat{X}(t) + \lambda_0, \qquad 0 \leq t \leq \mathcal{T}. \qquad (37)$$

From [35, Theorem 7.2.1], it follows that conditioned on $\boldsymbol{S}^{\mathcal{K}}$, the rv $(N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}})$ corresponds to a "self-exciting" PCP with rate process $\hat{\Lambda}^{\mathcal{T}}$, and its (conditional) sample function density is given by

$$f_{N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}}|\boldsymbol{S}^{\mathcal{K}}}\left(n_{\mathcal{T}}, \boldsymbol{t}^{n_{\mathcal{T}}} \,\middle|\, \boldsymbol{s}^{\mathcal{K}}\right)$$
$$= \begin{cases} \exp\left(-\int_0^{\mathcal{T}} \hat{\lambda}(t)dt\right), & n_{\mathcal{T}} = 0 \\ \exp\left(-\int_0^{\mathcal{T}} \hat{\lambda}(t)dt\right) \cdot \prod_{i=1}^{n_{\mathcal{T}}} \hat{\lambda}(t_i), & n_{\mathcal{T}} \geq 1 \end{cases} \qquad (38)$$

where

$$\hat{\lambda}(t) = s[\lceil t/\mathcal{T}_{\mathrm{c}}\rceil]\hat{x}(t) + \lambda_0$$

with

$$\hat{x}(t) = \mathbb{E}\left[X(t) \,\middle|\, N_t = n_t, \boldsymbol{T}^{N_t} = \boldsymbol{t}^{n_t}, \boldsymbol{S}^{\lceil t/\mathcal{T}_{\mathrm{c}}\rceil} = \boldsymbol{s}^{\lceil t/\mathcal{T}_{\mathrm{c}}\rceil}\right]$$

for $0 \leq t \leq \mathcal{T}$.

Proof of Theorem 1:

Converse Part: With $\mathcal{T} = \mathcal{K}\mathcal{T}_{\mathrm{c}}$, $\mathcal{K} \in \mathbb{Z}^+$, consider $(\mathcal{W}, \mathcal{T})$–codes (f, ϕ) of rate $\mathcal{R} = \frac{1}{\mathcal{T}}\ln \mathcal{W}$ and with $P_e(f, \phi) \leq \epsilon$ (cf. (5)), where $0 < \epsilon < 1$ is arbitrary but fixed. Let W be a rv uniformly distributed on (the message set) $\mathbb{W}$, with W being independent of $\boldsymbol{S}^{\mathcal{K}}$. Denote $X(t) \triangleq x_W(t, \boldsymbol{U}^{\lceil t/\mathcal{T}_{\mathrm{c}}\rceil})$, $0 \leq t \leq \mathcal{T}$ (cf. (6)). Note that (4) then implies that

$$0 \leq X(t) \leq \alpha, \qquad 0 \leq t \leq \mathcal{T}$$
$$\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}} \mathbb{E}[X(t)]dt \leq \sigma\alpha. \qquad (39)$$

Let $(N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}})$ be the channel output corresponding to an input $X^{\mathcal{T}}$ when the channel fade is $\boldsymbol{S}^{\mathcal{K}}$, i.e., $(N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}})$ is a PCP with rate process $\{\Lambda(t), 0 \leq t \leq \mathcal{T}\}$ as in (36) with $X(t) = x_W(t, \boldsymbol{U}^{\lceil t/\mathcal{T}_{\mathrm{c}}\rceil})$, $0 \leq t \leq \mathcal{T}$. Clearly, the following Markov property holds:

$$W \multimap X^t, \boldsymbol{S}^{\lceil t/\mathcal{T}_{\mathrm{c}}\rceil} \multimap N_t, \boldsymbol{T}^{N_t}, \qquad 0 \leq t \leq \mathcal{T}. \quad (40)$$

By a standard argument

$$\mathcal{R} = \frac{1}{\mathcal{T}}H(W)$$
$$= \frac{1}{\mathcal{T}}I\left(W \wedge \phi\left(N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}}, \boldsymbol{S}^{\mathcal{K}}\right)\right)$$
$$\quad + \frac{1}{\mathcal{T}}H\left(W \,\middle|\, \phi\left(N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}}, \boldsymbol{S}^{\mathcal{K}}\right)\right)$$

which, upon using Fano's inequality

$$H\left(W\,|\,\phi(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}},\,\boldsymbol{S}^{\mathcal{K}})\right)\le\epsilon\ln\mathcal{W}+h_b(\epsilon)$$
$$=\epsilon\,\mathcal{T}\mathcal{R}+h_b(\epsilon)$$

leads to

$$\mathcal{R}\le\frac{1}{(1-\epsilon)}\left[\frac{1}{\mathcal{T}}I\left(W\wedge\phi(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}},\,\boldsymbol{S}^{\mathcal{K}})\right)+\frac{1}{\mathcal{T}}h_b(\epsilon)\right]$$

where h_b denotes binary entropy. Since $0<\epsilon<1$ was arbitrary, we get the standard outcome that the rate $\mathcal{R}$ of the $(\mathcal{W},\,\mathcal{T})$–code $(f,\,\phi)$ with $P_e(f,\,\phi)\le\epsilon$ must satisfy

$$\mathcal{R}\lesssim\frac{1}{\mathcal{T}}I\left(W\wedge\phi(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}},\,\boldsymbol{S}^{\mathcal{K}})\right). \tag{41}$$

Proceeding further with the right-hand side of (41)

$$I\left(W\wedge\phi(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}},\,\boldsymbol{S}^{\mathcal{K}})\right)$$
$$\le I\left(W\wedge N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}},\,\boldsymbol{S}^{\mathcal{K}}\right) \tag{42}$$
$$= I\left(W\wedge N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,\boldsymbol{S}^{\mathcal{K}}\right) \tag{43}$$
$$= h(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,\boldsymbol{S}^{\mathcal{K}})-h(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,W,\,\boldsymbol{S}^{\mathcal{K}})$$
$$\le h(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,\boldsymbol{S}^{\mathcal{K}})-h(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,X^{\mathcal{T}},\,\boldsymbol{S}^{\mathcal{K}}) \tag{44}$$

where (42)–(44) follow, respectively, by the data processing result for mixed rvs,[10] the independence of W from $\boldsymbol{S}^{\mathcal{K}}$, and (40). The difference between the conditional entropies of the mixed rvs[11] on the right-hand side of (44) is

$$h(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,\boldsymbol{S}^{\mathcal{K}})-h(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,X^{\mathcal{T}},\,\boldsymbol{S}^{\mathcal{K}})$$
$$=\mathbb{E}\left[-\ln f_{N_{\mathcal{T}},\boldsymbol{T}^{N_{\mathcal{T}}}|\boldsymbol{S}^{\mathcal{K}}}(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,\boldsymbol{S}^{\mathcal{K}})\right]$$
$$-\mathbb{E}\left[-\ln f_{N_{\mathcal{T}},\boldsymbol{T}^{N_{\mathcal{T}}}|X^{\mathcal{T}},\boldsymbol{S}^{\mathcal{K}}}(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,X^{\mathcal{T}},\,\boldsymbol{S}^{\mathcal{K}})\right]$$
$$=\mathbb{E}\left[\int_0^{\mathcal{T}}\left(\hat{\Lambda}(t)-\Lambda(t)\right)dt\right]$$
$$+\mathbb{E}\left[\ln\left\{\mathbb{1}(N_{\mathcal{T}}=0)+\mathbb{1}(N_{\mathcal{T}}\ge1)\prod_{i=1}^{N_{\mathcal{T}}}\Lambda(T_i)\right\}\right.$$
$$\left.-\ln\left\{\mathbb{1}(N_{\mathcal{T}}=0)+\mathbb{1}(N_{\mathcal{T}}\ge1)\prod_{i=1}^{N_{\mathcal{T}}}\hat{\Lambda}(T_i)\right\}\right] \tag{45}$$

where (45) is by (34), (38), with $\mathbb{1}(\cdot)$ denoting the indicator function. The first expectation term in (45), by an interchange of operations,[12] is

$$\mathbb{E}\left[\int_0^{\mathcal{T}}\left(\hat{\Lambda}(t)-\Lambda(t)\right)dt\right]$$
$$=\int_0^{\mathcal{T}}\mathbb{E}\left[\hat{\Lambda}(t)-\Lambda(t)\right]dt$$
$$=0 \tag{46}$$

---

[10]This result can be deduced, for instance, from [30, the last paragraph of Section 3.4, p. 36, and Kolmogorov's formula (3.6.3) on p. 37].

[11]Our definition of the conditional entropy of mixed rvs is consistent with the general formulation developed, for instance, in [30, Ch. 3].

[12]The interchange is permissible as the assumed condition $\mathbb{E}[S]<\infty$ implies the integrability of $\{\hat{\Lambda}(t)-\Lambda(t),0\le t\le\mathcal{T}\}$.

since $\mathbb{E}[\hat{\Lambda}(t)]=\mathbb{E}[\Lambda(t)]$, $0\le t\le\mathcal{T}$, by (37). The second expectation term in (45) equals

$$\mathbb{E}\left[\mathbb{E}\left[\ln\left\{\mathbb{1}(N_{\mathcal{T}}=0)+\mathbb{1}(N_{\mathcal{T}}\ge1)\prod_{i=1}^{N_{\mathcal{T}}}\Lambda(T_i)\right\}\right.\right.$$
$$\left.\left.-\ln\left\{\mathbb{1}(N_{\mathcal{T}}=0)+\mathbb{1}(N_{\mathcal{T}}\ge1)\prod_{i=1}^{N_{\mathcal{T}}}\hat{\Lambda}(T_i)\right\}\,\right|\,N_{\mathcal{T}}\right]\right]. \tag{47}$$

The (inner) conditional expectation term above, conditioned on $N_{\mathcal{T}}=0$, equals 0; and, conditioned on $N_{\mathcal{T}}=n_{\mathcal{T}}\ge1$, it equals

$$\mathbb{E}\left[\ln\prod_{i=1}^{n_{\mathcal{T}}}\Lambda(T_i)-\ln\prod_{i=1}^{n_{\mathcal{T}}}\hat{\Lambda}(T_i)\,\right|\,N_{\mathcal{T}}=n_{\mathcal{T}}\right]$$
$$=\mathbb{E}\left[\sum_{i=1}^{n_{\mathcal{T}}}\left(\ln\Lambda(T_i)-\ln\hat{\Lambda}(T_i)\right)\,\right|\,N_{\mathcal{T}}=n_{\mathcal{T}}\right];$$

so that the conditional expectation in (47) is

$$\mathbb{1}(N_{\mathcal{T}}\ge1)\mathbb{E}\left[\sum_{i=1}^{N_{\mathcal{T}}}\left(\ln\Lambda(T_i)-\ln\hat{\Lambda}(T_i)\right)\,\right|\,N_{\mathcal{T}}\right]$$
$$=\mathbb{E}\left[\mathbb{1}(N_{\mathcal{T}}\ge1)\sum_{i=1}^{N_{\mathcal{T}}}\left(\ln\Lambda(T_i)-\ln\hat{\Lambda}(T_i)\right)\,\right|\,N_{\mathcal{T}}\right]. \tag{48}$$

Then, by (46)–(48), we get in (45) that

$$h(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,\boldsymbol{S}^{\mathcal{K}})-h(N_{\mathcal{T}},\,\boldsymbol{T}^{N_{\mathcal{T}}}\,|\,X^{\mathcal{T}},\,\boldsymbol{S}^{\mathcal{K}})$$
$$=\mathbb{E}\left[\mathbb{1}(N_{\mathcal{T}}\ge1)\sum_{i=1}^{N_{\mathcal{T}}}\left(\ln\Lambda(T_i)-\ln\hat{\Lambda}(T_i)\right)\right] \tag{49}$$
$$=\int_0^{\mathcal{T}}\left\{\mathbb{E}\left[\zeta\left(S[\lceil t/\mathcal{T}_c\rceil]X(t),\,\lambda_0\right)\right]\right.$$
$$\left.-\mathbb{E}\left[\zeta\left(S[\lceil t/\mathcal{T}_c\rceil]\hat{X}(t),\,\lambda_0\right)\right]\right\}dt \tag{50}$$

where (50) is shown to result from (49) in Appendix A.

Next, in the right-hand side of (50)

$$\mathbb{E}\left[\zeta\left(S[\lceil t/\mathcal{T}_c\rceil]\hat{X}(t),\,\lambda_0\right)\right]$$
$$=\mathbb{E}\left[\mathbb{E}\left[\zeta\left(S[\lceil t/\mathcal{T}_c\rceil]\hat{X}(t),\,\lambda_0\right)\,\middle|\,S[\lceil t/\mathcal{T}_c\rceil]\right]\right]$$
$$\ge\mathbb{E}\left[\zeta\left(\mathbb{E}\left[S[\lceil t/\mathcal{T}_c\rceil]\hat{X}(t)\,\middle|\,S[\lceil t/\mathcal{T}_c\rceil]\right],\,\lambda_0\right)\right] \tag{51}$$
$$=\mathbb{E}\left[\zeta\left(S[\lceil t/\mathcal{T}_c\rceil]\mathbb{E}\left[\hat{X}(t)\,\middle|\,S[\lceil t/\mathcal{T}_c\rceil]\right],\,\lambda_0\right)\right]$$
$$=\mathbb{E}\left[\zeta\left(S[\lceil t/\mathcal{T}_c\rceil]\mathbb{E}\left[X(t)\,\middle|\,S[\lceil t/\mathcal{T}_c\rceil]\right],\,\lambda_0\right)\right] \tag{52}$$
$$=\mathbb{E}\left[\zeta\left(S[\lceil t/\mathcal{T}_c\rceil]\mathbb{E}\left[X(t)\,\middle|\,U[\lceil t/\mathcal{T}_c\rceil]\right],\,\lambda_0\right)\right] \tag{53}$$

where (51) is by Jensen's inequality applied to the convex function $\zeta(\cdot,\,\lambda_0)$; (52) is from

$$\mathbb{E}\left[\hat{X}(t)\,\middle|\,S[\lceil t/\mathcal{T}_c\rceil]\right]$$
$$=\mathbb{E}\left[\mathbb{E}\left[X(t)\,\middle|\,N_t,\,\boldsymbol{T}^{N_t},\,\boldsymbol{S}^{\lceil t/\mathcal{T}_c\rceil}\right]\,\middle|\,S[\lceil t/\mathcal{T}_c\rceil]\right]$$
$$=\mathbb{E}\left[X(t)\,\middle|\,S[\lceil t/\mathcal{T}_c\rceil]\right];$$

and (53) holds as

$$\mathbb{E}\Big[X(t)\Big|S[\lceil t/\mathcal{T}_c\rceil]\Big]$$
$$= \mathbb{E}\Big[X(t)\Big|S[\lceil t/\mathcal{T}_c\rceil], U[\lceil t/\mathcal{T}_c\rceil]\Big]$$
$$= \mathbb{E}\Big[X(t)\Big|U[\lceil t/\mathcal{T}_c\rceil]\Big] \tag{54}$$

by virtue of the Markov condition

$$X(t) \multimap U[\lceil t/\mathcal{T}_c\rceil] \multimap S[\lceil t/\mathcal{T}_c\rceil], \qquad 0 \le t \le \mathcal{T} \tag{55}$$

which itself is a simple consequence of the independence of $(W, \boldsymbol{U}^{\lceil t/\mathcal{T}_c\rceil - 1})$ and $S_{\lceil t/\mathcal{T}_c\rceil}$, and the fact that $X(t)$ is a function of $(W, \boldsymbol{U}^{\lceil t/\mathcal{T}_c\rceil - 1})$ and $U[\lceil t/\mathcal{T}_c\rceil]$ with the latter being determined by $S[\lceil t/\mathcal{T}_c\rceil]$. Summarizing collectively (44), (50), (53), we get that

$$I(W \wedge \phi(N_\mathcal{T}, \boldsymbol{T}^{N_\mathcal{T}}, \boldsymbol{S}^\mathcal{K}))$$
$$\le \int_0^\mathcal{T} \mathbb{E}\Big[\zeta\Big(S[\lceil t/\mathcal{T}_c\rceil]X(t), \lambda_0\Big)$$
$$- \zeta\Big(S[\lceil t/\mathcal{T}_c\rceil]\mathbb{E}\Big[X(t)\Big|U[\lceil t/\mathcal{T}_c\rceil]\Big], \lambda_0\Big)\Big]\,dt. \tag{56}$$

The right-hand side of (56) is further bounded above by a suitable modification of an argument of Davis [8]. Considering the integrand in (56), fix $0 \le t \le \mathcal{T}$ and condition on $S[\lceil t/\mathcal{T}_c\rceil] = s$, $s \in \mathbb{R}_0^+$. Then

$$\mathbb{E}\Big[\zeta\Big(S[\lceil t/\mathcal{T}_c\rceil]X(t), \lambda_0\Big)$$
$$- \zeta\Big(S[\lceil t/\mathcal{T}_c\rceil]\mathbb{E}\Big[X(t)\Big|U[\lceil t/\mathcal{T}_c\rceil]\Big], \lambda_0\Big)\Big|S[\lceil t/\mathcal{T}_c\rceil] = s\Big]$$
$$= \mathbb{E}\Big[\zeta(sX(t), \lambda_0)\Big|S[\lceil t/\mathcal{T}_c\rceil] = s\Big]$$
$$- \zeta\Big(s\mathbb{E}\Big[X(t)\Big|U[\lceil t/\mathcal{T}_c\rceil] = h(s)\Big], \lambda_0\Big)$$
$$= \mathbb{E}\Big[\zeta(sX(t), \lambda_0)\Big|U[\lceil t/\mathcal{T}_c\rceil] = h(s)\Big]$$
$$- \zeta\Big(s\mathbb{E}\Big[X(t)\Big|U[\lceil t/\mathcal{T}_c\rceil] = h(s)\Big], \lambda_0\Big) \tag{57}$$

by (55). Consider maximizing the right-hand side of (57) over all conditional distributions of $X(t)$ conditioned on $U[\lceil t/\mathcal{T}_c\rceil] = h(s)$ with a fixed conditional mean

$$\mathbb{E}[X(t)|U[\lceil t/\mathcal{T}_c\rceil] = h(s)] = \pi_t(h(s))$$

say, and subject to the first constraint (alone) in (39). Then, the right-hand side of (57) equals

$$\mathbb{E}\Big[\zeta(sX(t), \lambda_0)\Big|U[\lceil t/\mathcal{T}_c\rceil] = h(s)\Big] - \zeta(s\pi_t(h(s)), \lambda_0) \tag{58}$$

and is maximized by considering the first term above. Using the strict convexity of $\zeta(\cdot, \lambda_0)$, this term is largest (see [8, proof of Theorem 1], or [33, Lemma 1])[13] iff $X(t)$ is a $\{0, \alpha\}$-valued rv with

$$\Pr\Big(X(t) = \alpha\Big|U[\lceil t/\mathcal{T}_c\rceil] = h(s)\Big)$$
$$= 1 - \Pr\Big(X(t) = 0\Big|U[\lceil t/\mathcal{T}_c\rceil] = h(s)\Big)$$
$$= \frac{\pi_t(h(s))}{\alpha} \tag{59}$$

and the corresponding largest value of (58) is

$$\frac{\pi_t(h(s))\zeta(\alpha s, \lambda_0)}{\alpha} - \zeta(s\pi_t(h(s)), \lambda_0). \tag{60}$$

Noting that (39) implies that

$$0 \le \pi_t(h(s)) \le \alpha, \qquad 0 \le t \le \mathcal{T}, \qquad s \in \mathbb{R}_0^+ \tag{61}$$

and

$$\frac{1}{\mathcal{T}}\int_0^\mathcal{T} \mathbb{E}\Big[\pi_t\big(h(S[\lceil t/\mathcal{T}_c\rceil])\big)\Big]\,dt \le \sigma\alpha; \tag{62}$$

we thus obtain from (56)–(58), (60)–(62) the expressions that appear in (63) at the bottom of the page.

[13]An alternative proof can be gleaned from the following simple observation. Let X be a $[0, \alpha]$-valued rv, $\alpha > 0$, of arbitrary distribution but with fixed mean $\mu = \mathbb{E}[X]$. Let $g : [0, \alpha] \to \mathbb{R}_0^+$ be a strictly convex mapping, with $g(0) = 0$. Then, $g(X) \le \frac{g(\alpha)}{\alpha}X$, whence $\mathbb{E}[g(X)] \le \frac{g(\alpha)\mu}{\alpha}$. It is readily seen that this upper bound on $\mathbb{E}[g(X)]$ is achieved if $X \in \{0, \alpha\}$ with $\Pr\{X = 0\} = 1 - \Pr\{X = \alpha\} = 1 - \frac{\mu}{\alpha}$. The necessity of this choice of (optimal) X follows from the strict convexity of $g(\cdot)$ on $[0, \alpha]$.

$$\frac{1}{\mathcal{T}}I\Big(W \wedge \phi(N_\mathcal{T}, \boldsymbol{T}^{N_\mathcal{T}}, \boldsymbol{S}^\mathcal{K})\Big) \le \max_{\substack{\pi_t:\mathcal{U}\to[0,\alpha],\, 0\le t\le\mathcal{T} \\ \frac{1}{\mathcal{T}}\int_0^\mathcal{T}\mathbb{E}\left[\pi_t\left(h(S[\lceil t/\mathcal{T}_c\rceil])\right)\right]dt\le\sigma\alpha}} \frac{1}{\mathcal{T}}\int_0^\mathcal{T}\Bigg\{\mathbb{E}\Big[\pi_t\big(h(S[\lceil t/\mathcal{T}_c\rceil])\big)\zeta(S[\lceil t/\mathcal{T}_c\rceil]\alpha, \lambda_0)/\alpha\Big]$$
$$- \mathbb{E}\Big[\zeta\Big(S[\lceil t/\mathcal{T}_c\rceil]\pi_t\big(h(S[\lceil t/\mathcal{T}_c\rceil])\big), \lambda_0\Big)\Big]\Bigg\}\,dt$$
$$= \max_{\substack{\pi_t:\mathcal{U}\to[0,\alpha],\, 0\le t\le\mathcal{T} \\ \frac{1}{\mathcal{T}}\int_0^\mathcal{T}\mathbb{E}\left[\pi_t\left(U[\lceil t/\mathcal{T}_c\rceil]\right)\right]dt\le\sigma\alpha}} \frac{1}{\mathcal{T}}\int_0^\mathcal{T}\Bigg\{\mathbb{E}\Big[\pi_t(U[\lceil t/\mathcal{T}_c\rceil])\zeta(S[\lceil t/\mathcal{T}_c\rceil]\alpha, \lambda_0)/\alpha\Big]$$
$$- \mathbb{E}\Big[\zeta\Big(S[\lceil t/\mathcal{T}_c\rceil]\pi_t(U[\lceil t/\mathcal{T}_c\rceil]), \lambda_0\Big)\Big]\Bigg\}\,dt. \tag{63}$$

In order to simplify the right-hand side of (63), for every $u \in \mathbb{U}$, define

$$\nu_k(u) \triangleq \frac{1}{\mathcal{T}_c} \int_{(k-1)\mathcal{T}_c}^{k\mathcal{T}_c} \pi_t(u)dt, \qquad k = 1, \ldots, \mathcal{K} \quad (64)$$

$$\mu(u) \triangleq \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \frac{\nu_k(u)}{\alpha}. \quad (65)$$

From (62), we get

$$0 \le \mu(u) \le 1, \qquad u \in \mathbb{U}, \quad (66)$$

and

$$\sigma \ge \frac{1}{\alpha\mathcal{T}} \int_0^{\mathcal{T}} \mathbb{E}\left[\pi_t(h(S[\lceil t/\mathcal{T}_c\rceil]))\right] dt$$

$$= \frac{1}{\alpha\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \frac{1}{\mathcal{T}_c} \int_{(k-1)\mathcal{T}_c}^{k\mathcal{T}_c} \mathbb{E}\left[\pi_t(U[k])\right] dt$$

$$= \frac{1}{\alpha\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \mathbb{E}\left[\frac{1}{\mathcal{T}_c} \int_{(k-1)\mathcal{T}_c}^{k\mathcal{T}_c} \pi_t(U)dt\right] \quad (67)$$

$$= \frac{1}{\alpha\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \mathbb{E}[\nu_k(U)] \quad (68)$$

$$= \mathbb{E}[\mu(U)] \quad (69)$$

where (67) is by an interchange of operations,[14] and by the i.i.d. nature of the channel fade sequence $\boldsymbol{S}^\infty$; (68) is by (64); and (69) is by (65).

The time-averaged integral on the right-hand side of (63) can be written as

$$\frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} \left\{ \mathbb{E}\left[\pi_t(U[\lceil t/\mathcal{T}_c\rceil])\zeta(S[\lceil t/\mathcal{T}_c\rceil]\alpha, \lambda_0)/\alpha\right] \right.$$
$$\left. - \mathbb{E}\left[\zeta(S[\lceil t/\mathcal{T}_c\rceil]\pi_t(U[\lceil t/\mathcal{T}_c\rceil]), \lambda_0)\right] \right\} dt$$

$$= \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \frac{1}{\mathcal{T}_c} \int_{(k-1)\mathcal{T}_c}^{k\mathcal{T}_c} \left\{ \mathbb{E}\left[\pi_t(U[k])\zeta(S[k]\alpha, \lambda_0)/\alpha\right.\right.$$
$$\left.\left. - \zeta(S[k]\pi_t(U[k]), \lambda_0)\right] \right\} dt$$

$$\le \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \mathbb{E}\left[\nu_k(U[k])\zeta(S[k]\alpha, \lambda_0)/\alpha\right.$$
$$\left. - \zeta(S[k]\nu_k(U[k]), \lambda_0)\right] \quad (70)$$

$$= \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \mathbb{E}\left[\nu_k(U)\zeta(S\alpha, \lambda_0)/\alpha - \zeta(S\nu_k(U), \lambda_0)\right] \quad (71)$$

$$\le \mathbb{E}[\mu(U)\zeta(S\alpha, \lambda_0) - \zeta(S\mu(U)\alpha, \lambda_0)] \quad (72)$$

where (70) is by Jensen's inequality applied to the convex function $\zeta(\cdot, \lambda_0)$ and (64); (71) holds by the i.i.d. nature of $\boldsymbol{S}^\infty$; and (72) is by Jensen's inequality and (65). Summarizing collectively (41), (63), (66), (69), (72), we get that

$$\mathcal{R} \lesssim \max_{\substack{\mu: \mathbb{U} \to [0, 1] \\ \mathbb{E}[\mu(U)] \le \sigma}} \mathbb{E}[\mu(U)\zeta(S\alpha, \lambda_0) - \zeta(\mu(U)S\alpha, \lambda_0)] \quad (73)$$

thereby completing the proof of the converse part.

[14]The interchange is permissible since by definition, $0 \le \pi_t(u) \le 1, u \in \mathbb{U}$, $0 \le t \le \mathcal{T}$.
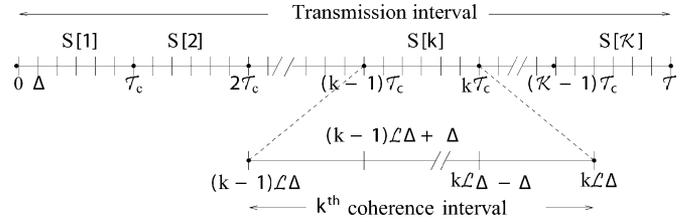

Fig. 3. Subintervals of $[0, \mathcal{T}]$.

Achievability Part: We closely follow Wyner's approach [40].

Fix $\mathcal{L} \in \mathbb{Z}^+$ and set $\Delta = \mathcal{T}_c/\mathcal{L}$. Divide the time interval $[0, \mathcal{T}]$, where $\mathcal{T} = \mathcal{K}\mathcal{T}_c$ with $\mathcal{K} \in \mathbb{Z}^+$, into $\mathcal{K}\mathcal{L}$ equal subintervals, each of duration Δ. See Fig. 3. Then, in the channel fade sequence $\boldsymbol{S}^\infty = \{S[k]\}_{k=1}^\infty$, each $S[k]$ remains unvarying for a block of $\mathcal{L}$ consecutive Δ-duration subintervals within $((k-1)\mathcal{T}_c, k\mathcal{T}_c]$, and $\{S[k]\}_{k=1}^\infty$ varies across such blocks in an i.i.d. manner.

Now, consider the following class of restricted $(\mathcal{W}, \mathcal{T})$-codes with message set $\mathbb{W}$. The codebook consists of waveforms $f(w, \boldsymbol{u}^\mathcal{K}) = \{x_w(t, \boldsymbol{u}^{\lceil t/\mathcal{T}_c\rceil}), 0 \le t \le \mathcal{T}\}, w \in \mathbb{W}, \boldsymbol{u}^\mathcal{K} \in \mathbb{U}^\mathcal{K}$, which are restricted to be $\{0, \alpha\}$-valued and piecewise constant within each of the $\mathcal{K}\mathcal{L}$ time slots of duration Δ. Define

$$\tilde{x}_n(w, \boldsymbol{u}^{\lceil n/L\rceil}) \triangleq \begin{cases} 0, & \text{if } x_w(t, \boldsymbol{u}^{\lceil t/\mathcal{T}_c\rceil}) = 0 \\ 1, & \text{if } x_w(t, \boldsymbol{u}^{\lceil t/\mathcal{T}_c\rceil}) = \alpha \end{cases}$$
$$t \in [(n-1)\Delta, n\Delta), \quad n = 1, \ldots, \mathcal{K}\mathcal{L}. \quad (74)$$

Note that the condition (4) requires that

$$\frac{1}{\mathcal{K}\mathcal{L}} \sum_{n=1}^{\mathcal{K}\mathcal{L}} \tilde{x}_n(w, \boldsymbol{u}^{\lceil n/\mathcal{L}\rceil}) \le \sigma, \qquad w \in \mathbb{W}, \boldsymbol{u}^\mathcal{K} \in \mathbb{U}^\mathcal{K}. \quad (75)$$

Next, the (restricted) decoder $\phi : \{0, 1\}^{\mathcal{K}\mathcal{L}} \times (\mathbb{R}_0^+)^\mathcal{K} \to \mathbb{W}$ is based on limited observations over the $\mathcal{K}\mathcal{L}$ time slots, comprising

$$\tilde{Y}_n = \mathbb{1}(Y(n\Delta) - Y((n-1)\Delta) = 1), \quad n = 1, \ldots, \mathcal{K}\mathcal{L} \quad (76)$$

(with $Y(0) = 0$), and $\boldsymbol{S}^\mathcal{K}$. This decoder considers the arrival of two or more photons within an interval of width Δ, as an "error event" of negligible probability (for small Δ); this event is combined with the positive probability event of no photon arrival within the same interval.

The largest achievable rate of the restricted $(\mathcal{W}, \mathcal{T})$-codes as above—and, hence, the capacity $\mathcal{C}$ for transmitter CSI h—is clearly no smaller than $\frac{\mathcal{C}(\mathcal{L})}{\mathcal{T}_c}$, where $\mathcal{C}(\mathcal{L})$ is the capacity of a $(\mathcal{L}\text{-})$block discrete memoryless channel (in nats per block channel use) with input alphabet $\tilde{\mathbb{X}}^\mathcal{L} = \{0, 1\}^\mathcal{L}$; output alphabet $\tilde{\mathbb{Y}}^\mathcal{L} = \{0, 1\}^\mathcal{L}$; state alphabet $\mathbb{R}_0^+$; transition probability mass function (pmf)

$$W^{(\mathcal{L})}(\tilde{\boldsymbol{y}}^\mathcal{L}|\tilde{\boldsymbol{x}}^\mathcal{L}, s) = \prod_{l=1}^\mathcal{L} W_{\tilde{Y}|\tilde{X}, S}(\tilde{y}_l|\tilde{x}_l, s),$$

$$\tilde{\boldsymbol{x}}^\mathcal{L}, \tilde{\boldsymbol{y}}^\mathcal{L} \in \{0, 1\}^\mathcal{L}, s \in \mathbb{R}_0^+ \quad (77)$$

where $\tilde{X}, S, \tilde{Y}$, respectively, are $\tilde{\mathbb{X}}$-, $\mathbb{R}_0^+$-, and $\tilde{\mathbb{Y}}$-valued rvs, and $W_{\tilde{Y}|\tilde{X}, S}(\cdot|\cdot, s), s \in \mathbb{R}_0^+$, is given by (see Fig. 4)

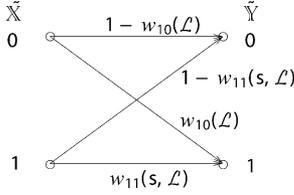$$W_{\tilde{Y}|\tilde{X}, S}(1|0, s) = 1 - W_{\tilde{Y}|\tilde{X}, S}(0|0, s)$$

Fig. 4. Discrete channel approximation.

$$= \frac{\lambda_0 \mathcal{T}_c}{\mathcal{L}} \exp\left(-\frac{\lambda_0 \mathcal{T}_c}{\mathcal{L}}\right)$$

$$\triangleq w_{10}(\mathcal{L}),$$

$$W_{\tilde{Y}|\tilde{X}, S}(1|1, s) = 1 - W_{\tilde{Y}|\tilde{X}, S}(0|1, s)$$

$$= \frac{(\lambda_0 + s\alpha)\mathcal{T}_c}{\mathcal{L}} \exp\left(-\frac{(\lambda_0 + s\alpha)\mathcal{T}_c}{\mathcal{L}}\right)$$

$$\triangleq w_{11}(s, \mathcal{L}) \tag{78}$$

with transmitter CSI h (and perfect receiver CSI); and under constraint (75).

By [7, Remark A2 following Proposition 1], it is readily obtained[15] that

$$\mathcal{C}(\mathcal{L}) = \max_{P_{\tilde{\boldsymbol{X}}^{\mathcal{L}}|U}: \sum_{l=1}^{\mathcal{L}} \mathbb{E}[\tilde{X}_l] \leq \mathcal{L}\sigma} I(\tilde{\boldsymbol{X}}^{\mathcal{L}} \wedge \tilde{\boldsymbol{Y}}^{\mathcal{L}}|S) \tag{79}$$

with $U = h(S)$, where the joint conditional pmf $P_{\tilde{\boldsymbol{X}}^{\mathcal{L}}, \tilde{\boldsymbol{Y}}^{\mathcal{L}}|S}$ is given by

$$P_{\tilde{\boldsymbol{X}}^{\mathcal{L}}, \tilde{\boldsymbol{Y}}^{\mathcal{L}}|S}(\tilde{\boldsymbol{x}}^{\mathcal{L}}, \tilde{\boldsymbol{y}}^{\mathcal{L}}|s) = P_{\tilde{\boldsymbol{X}}^{\mathcal{L}}|U}(\tilde{\boldsymbol{x}}^{\mathcal{L}}|h(s))W^{(\mathcal{L})}(\tilde{\boldsymbol{y}}^{\mathcal{L}}|\tilde{\boldsymbol{x}}^{\mathcal{L}}, s),$$

$$\tilde{\boldsymbol{x}}^{\mathcal{L}}, \tilde{\boldsymbol{y}}^{\mathcal{L}} \in \{0, 1\}^{\mathcal{L}}, \; s \in \mathbb{R}_0^+. \tag{80}$$

By a standard argument which uses (77), the maximum in (79) is achieved by

$$P_{\tilde{\boldsymbol{X}}^{\mathcal{L}}|U}(\tilde{\boldsymbol{x}}^{\mathcal{L}}|h(s)) = \prod_{l=1}^{\mathcal{L}} P_{\tilde{X}_l|U}(\tilde{x}_l|h(s)),$$

$$\tilde{\boldsymbol{x}}^{\mathcal{L}} \in \{0, 1\}^{\mathcal{L}}, \; s \in \mathbb{R}_0^+ \tag{81}$$

so that from (79)

$$\mathcal{C}(\mathcal{L}) = \mathcal{L} \max_{P_{\tilde{X}|U}: \mathbb{E}[\tilde{X}] \leq \sigma} I(\tilde{X} \wedge \tilde{Y}|S) \tag{82}$$

where $P_{\tilde{X}, \tilde{Y}|S}$ is given by

$$P_{\tilde{X}, \tilde{Y}|S}(\tilde{x}, \tilde{y}|s) = P_{\tilde{X}|U}(\tilde{x}|h(s))W_{\tilde{Y}|\tilde{X}, S}(\tilde{y}|\tilde{x}, s),$$

$$\tilde{x}, \tilde{y} \in \{0, 1\}, \; s \in \mathbb{R}_0^+. \tag{83}$$

Setting

$$\mu(u) \triangleq \Pr\{\tilde{X} = 1|U = u\} = \mathbb{E}[\tilde{X}|U = u], \qquad u \in \mathbb{U}, \tag{84}$$

and

$$\beta_{\mathcal{L}}(s)$$
$$\triangleq I(\tilde{X} \wedge \tilde{Y}|S = s)$$
$$= H(\tilde{Y}|S = s) - H(\tilde{Y}|\tilde{X}, S = s)$$
$$= h_b\left(\mu(h(s))w_{11}(s, \mathcal{L}) + (1 - \mu(h(s)))w_{10}(\mathcal{L})\right)$$
$$\quad - \mu(h(s))h_b(w_{11}(s, \mathcal{L}))$$
$$\quad - (1 - \mu(h(s)))h_b(w_{10}(\mathcal{L})), \; s \in \mathbb{R}_0^+ \tag{85}$$

[15] In [7], S is taken to be finite-valued; however, the result (79) can be seen to hold also when S is $\mathbb{R}_0^+$-valued.

we can express (82) as

$$\mathcal{C}(\mathcal{L}) = \max_{\substack{\mu:\mathbb{U}\to[0, 1] \\ \mathbb{E}[\mu(U)] \leq \sigma}} \mathcal{L}\mathbb{E}[\beta_{\mathcal{L}}(S)]. \tag{86}$$

Since $\mathcal{L} \in \mathbb{Z}^+$ was arbitrary, we have

$$\mathcal{C} \geq \lim_{\mathcal{L}\to\infty} \frac{\mathcal{C}(\mathcal{L})}{\mathcal{T}_c}$$

$$\geq \max_{\substack{\mu:\mathbb{U}\to[0, 1] \\ \mathbb{E}[\mu(U)] \leq \sigma}} \lim_{\mathcal{L}\to\infty} \frac{\mathbb{E}[\beta_{\mathcal{L}}(S)]}{\mathcal{T}_c/\mathcal{L}} \tag{87}$$

by (86). Finally, it is shown in Appendix B that

$$\lim_{\mathcal{L}\to\infty} \frac{\mathbb{E}[\beta_{\mathcal{L}}(S)]}{\mathcal{T}_c/\mathcal{L}} = \mathbb{E}[\mu(U)\zeta(S\alpha, \lambda_0) - \zeta(S\mu(U)\alpha, \lambda_0)] \tag{88}$$

whence

$$\mathcal{C} \geq \max_{\substack{\mu:\mathbb{U}\to[0, 1] \\ \mathbb{E}[\mu(U)] \leq \sigma}} \mathbb{E}[\mu(U)\zeta(S\alpha, \lambda_0) - \zeta(S\mu(U)\alpha, \lambda_0)] \tag{89}$$

thereby completing the proof of the achievability part of Theorem 1. □

Remarks:

(i) In Section III, Remark (iii) following Theorem 1 constitutes an interpretation of (81) when $\mathcal{L} \to \infty$.

(ii) In the proof above of the achievability part of Theorem 1, we could also have considered a restricted decoder ϕ with $\tilde{Y}_n$ in (76) replaced by

$$\tilde{Y}_n = 1 - \mathbb{1}(Y(n\Delta) - Y((n-1)\Delta) = 0),$$

$$n = 1, \ldots, \mathcal{KL}$$

which combines the event corresponding to the arrival of two or more photons in an interval of width Δ, of negligible probability, with the positive probability event of a single photon arrival.

Proof of Theorem 2: From (14), we can write

$$\mathcal{C} = \max_{\substack{\mu:\mathbb{U}\to[0, 1] \\ \mathbb{E}[\mu(U)] \leq \sigma}} \mathbb{E}[\psi(U, \mu(U))] \tag{90}$$

where for $u \in \mathbb{U}$

$$\psi(u, \mu(u)) \triangleq \mathbb{E}[\mu(U)\zeta(S\alpha, \lambda_0) - \zeta(S\mu(U)\alpha, \lambda_0)|U = u]$$
$$= \mathbb{E}[\mu(u)\zeta(S\alpha, \lambda_0) - \zeta(S\mu(u)\alpha, \lambda_0)|U = u]. \tag{91}$$

It is useful to note that

$$\frac{\partial\psi}{\partial\mu} = \mathbb{E}\left[\zeta(S\alpha, \lambda_0) - S\alpha(1 + \ln(S\mu(u)\alpha + \lambda_0))\Big|U = u\right]$$

$$= \mathbb{E}\left[S\alpha \ln\left(\frac{1 + \xi(S\alpha/\lambda_0)S\alpha/\lambda_0}{1 + \mu(u)S\alpha/\lambda_0}\right)\Big|U = u\right]$$

$$\frac{\partial^2\psi}{\partial\mu^2} = -\mathbb{E}\left[\frac{S^2\alpha^2}{(S\mu(u)\alpha + \lambda_0)}\Big|U = u\right], \qquad u \in \mathbb{U} \tag{92}$$

where $\xi(\cdot)$ is as defined in (8). Note that by (10), the last expectation is that of a nonnegative rv, so that $\frac{\partial^2\psi}{\partial\mu^2} < 0$, i.e., $\psi(\cdot)$ is a strictly concave function of μ in the range of our interest.

In order to determine the optimal power control law $\mu^*: \mathbb{U} \to [0, 1]$, we proceed as follows. First consider the "unconstrained" optimization problem, i.e., without the constraints $\mu: \mathbb{U} \to [0, 1]$ and $\mathbb{E}[\mu(U)] \leq \sigma$

$$\max_{\mu: \mathbb{U} \to \mathbb{R}} \mathbb{E}[\psi(U, \mu(U))] \qquad (93)$$

and let $\mu_0: \mathbb{U} \to \mathbb{R}$ denote a maximizer in (93). Then μ_0 must satisfy the (necessary) Euler–Lagrange condition (cf., e.g., [10])

$$\left. \frac{\partial \psi}{\partial \mu} \right|_{\mu_0} = 0 \qquad (94)$$

which, by (92), implies that μ_0 satisfies

$$\mathbb{E}\left[S\alpha \ln\left(\frac{1 + \xi(S\alpha/\lambda_0)S\alpha/\lambda_0}{1 + \mu_0(u)S\alpha/\lambda_0} \right) \Big| U = u \right] = 0 \qquad (95)$$

for every $u \in \mathbb{U}$.

Furthermore, since $\psi(\cdot)$ is a strictly concave function of μ in the region of interest, (94) also constitutes a sufficient condition for μ_0 to be the (unique) maximizer in (93). It can be verified (cf., e.g., [8, Fig. 2]) that $\xi(\cdot)$ is monotone decreasing on $[0, \infty)$ with $\xi(0) = \frac{1}{2}$ and $\lim_{x \to \infty} \xi(x) = \frac{1}{e}$. Therefore, from (95), it follows that $\frac{1}{e} \leq \mu_0(u) \leq \frac{1}{2}$, $u \in \mathbb{U}$. With $\sigma_0 \triangleq \mathbb{E}[\mu_0(U)]$, we see that $\frac{1}{e} \leq \sigma_0 \leq \frac{1}{2}$.

Consider next the constrained optimization problem on the right-hand side of (90), i.e., now with the inclusion of the constraints $\mu: \mathbb{U} \to [0, 1]$ and $\mathbb{E}[\mu(U)] \leq \sigma$. If $\sigma \geq \sigma_0$, then the "unconstrained" maximizer μ_0 in (93) satisfies the previous two constraints, and hence is the solution of the constrained problem in (90) as well. Suppose next that $\sigma < \sigma_0$. First ignore the (local) constraint $0 \leq \mu(u) \leq 1$, $u \in \mathbb{U}$, and define the Lagrangian functional

$$L(\mu) = \mathbb{E}[\chi(U, \mu(U))] \qquad (96)$$

where

$$\chi(u, \mu(u)) \triangleq \psi(u, \mu(u)) - \rho\mu(u), \qquad u \in \mathbb{U} \qquad (97)$$

and $\rho \geq 0$ is a Lagrange multiplier. Using the strict concavity of $L(\cdot)$, we conclude that a necessary and sufficient condition for $\mu_\rho: \mathbb{U} \to \mathbb{R}$ to be the maximizer in (90) is given by the Euler–Lagrange equation

$$\left. \frac{\partial \chi}{\partial \mu} \right|_{\mu_\rho} = 0. \qquad (98)$$

By (92), (97), and (98), we then see that μ_ρ satisfies

$$\mathbb{E}\left[S\alpha \ln\left(\frac{1 + \xi(S\alpha/\lambda_0)S\alpha/\lambda_0}{1 + \mu_\rho(u)S\alpha/\lambda_0} \right) \Big| U = u \right] = \rho \qquad (99)$$

for every $u \in \mathbb{U}$.

We now impose the constraint $0 \leq \mu(u) \leq 1$, $u \in \mathbb{U}$. Note that $\mu_\rho(u) \leq \mu_0(u) \leq \frac{1}{2}$, $u \in \mathbb{U}$ for all $\rho \geq 0$. However, given $\rho > 0$, $\mu_\rho(u)$ can be < 0 for some $u \in \mathbb{U}$. By the strict concavity of $\chi(u, \cdot)$, it follows that if for some $u \in \mathbb{U}$, $\mu_\rho(u) < 0$, then for all $\omega \geq 0$, $\chi(u, \omega)$ is a strictly decreasing function of ω. Therefore, if $\mu_\rho(u) < 0$ for some $u \in \mathbb{U}$, the constraint $\mu(u) \geq 0$ dictates the maximizing solution to be $\mu^*(u) = 0$;

otherwise, the maximizing solution is given by $\mu^*(u) = \mu_\rho(u)$. Summarizing the previous observations, we get that $\mu^*(u) = [\mu_\rho(u)]^+$, $u \in \mathbb{U}$. Finally, the optimal Lagrange multiplier ρ^* is chosen to satisfy the power constraint $\mathbb{E}[[\mu_{\rho^*}(U)]^+] = \sigma$. This concludes the proof of Theorem 2. $\square$

Proof of Theorem 3: First consider the limit as $\lambda_0 \to 0$. Let $\mathcal{C}^{\mathrm{H}} = \lim_{\lambda_0 \to 0} \mathcal{C}$, $\mu_\rho^{\mathrm{H}}(u) = \lim_{\lambda_0 \to 0} \mu_\rho(u)$, and $\mu^{\mathrm{H}}(u) = \lim_{\lambda_0 \to 0} \mu^*(u)$, $u \in \mathbb{U}$, $\rho \geq 0$, where $\mathcal{C}$, $\mu_\rho(\cdot)$ and $\mu^*(\cdot)$ are as defined in (14), (17), and (20), respectively. Since

$$\lim_{\lambda_0 \to 0} \mathbb{E}\left[S\alpha \ln\left(\frac{1 + \xi(S\alpha/\lambda_0)S\alpha/\lambda_0}{1 + \mu S\alpha/\lambda_0} \right) \Big| U = u \right]$$
$$= \mathbb{E}[S\alpha \ln(1/e\mu)|U = u], \qquad 0 \leq \mu \leq 1$$

by (17), it follows that $\mu_\rho^{\mathrm{H}}(u) = e^{-1-\rho/\alpha\mathbb{E}[S|U=u]}$. Clearly, $0 \leq \mu_\rho^{\mathrm{H}}(u) \leq e^{-1}$, $\rho \geq 0$, $u \in \mathbb{U}$. Furthermore, $\mu_0^{\mathrm{H}}(u) = e^{-1}$, $u \in \mathbb{U}$, so that $\sigma_0 = e^{-1}$. By (20), it follows that

$$\mu^{\mathrm{H}}(u) = \begin{cases} e^{-1}, & \sigma \geq e^{-1} \\ e^{-1-\rho^*/\alpha\mathbb{E}[S|U=u]}, & \sigma < e^{-1} \end{cases} \qquad (100)$$

where for $\sigma < e^{-1}$, $\rho^* \geq 0$ satisfies $\mathbb{E}[e^{-1-\rho^*/\alpha\mathbb{E}[S|U]}] = \sigma$. Finally, by (14), in the limit as $\lambda_0 \to 0$, we get

$$\mathcal{C}^{\mathrm{H}} = \mathbb{E}\left[\mu^{\mathrm{H}}(U)\zeta(S\alpha, 0) - \zeta(\mu^{\mathrm{H}}(U)S\alpha, 0) \right]. \qquad (101)$$

This establishes the first part of Theorem 3.

Next consider the case $\lambda_0 \gg 1$. For $u \in \mathbb{U}$, $\rho \geq 0$, let $\mu_\rho^{\mathrm{L}}(u) = \lim_{\lambda_0 \to \infty} \mu_\rho(u)$ and $\mu^{\mathrm{L}}(u) = \lim_{\lambda_0 \to \infty} \mu^*(u)$. Given $s \geq 0$, $q \geq 0$

$$\zeta(s\alpha, \lambda_0) - s\alpha(1 + \ln(\lambda_0 + qs\alpha))$$
$$= s\alpha \ln \lambda_0 + (s\alpha + \lambda_0) \ln (1 + s\alpha/\lambda_0)$$
$$- s\alpha (1 + \ln \lambda_0 + \ln (1 + qs\alpha/\lambda_0))$$

which, for $\lambda_0 \gg 1$, with the approximation $\ln x = x - x^2/2 + O(x^3)$ for $x \ll 1$, equals

$$(s\alpha + \lambda_0) \left(s\alpha/\lambda_0 - (s\alpha/\lambda_0)^2/2 + O(\lambda_0^{-3}) \right)$$
$$- s\alpha \left(1 + qs\alpha/\lambda_0 - (qs\alpha/\lambda_0)^2/2 + O(\lambda_0^{-3}) \right)$$
$$= (1 - 2q)s^2\alpha^2/2\lambda_0 + O(\lambda_0^{-2})$$
$$\to 0, \qquad \text{as } \lambda_0 \to \infty. \qquad (102)$$

Hence, for $\lambda_0 \gg 1$, by (17), (92), and (102), it follows that

$$\rho = \mathbb{E}[\zeta(S\alpha, \lambda_0) - S\alpha(1 + \ln(\lambda_0 + \mu_\rho^{\mathrm{L}}(u)S\alpha))|U = u]$$
$$= (1 - 2\mu_\rho^{\mathrm{L}}(u))\mathbb{E}[S^2|U = u]\alpha^2/2\lambda_0 + O(\lambda_0^{-2})$$

so that as $\lambda_0 \to \infty$, we get that $\rho \to 0$ and $\mu_\rho^{\mathrm{L}}(u) = 1/2 + O(\lambda_0^{-1}) \to 1/2$, $u \in \mathbb{U}$. In this case, $\sigma_0 = 1/2$, and therefore $\mu^{\mathrm{L}}(u) = \min\{\sigma, 1/2\}$, $u \in \mathbb{U}$, a constant which we denote by μ^{L}.

Finally, we compute $\mathcal{C}^{\mathrm{L}}$. Given $s \geq 0$, for $\lambda_0 \gg 1$, we have

$$\mu^{\mathrm{L}}\zeta(s\alpha, \lambda_0) - \zeta(\mu^{\mathrm{L}}s\alpha, \lambda_0)$$
$$= \mu^{\mathrm{L}}(s\alpha + \lambda_0)\ln(1 + s\alpha/\lambda_0)$$
$$- (\mu^{\mathrm{L}}s\alpha + \lambda_0)\ln(1 + \mu^{\mathrm{L}}s\alpha/\lambda_0)$$
$$= \mu^{\mathrm{L}}(s\alpha + \lambda_0) \left(s\alpha/\lambda_0 - (s\alpha/\lambda_0)^2/2 + O(\lambda_0^{-3}) \right)$$
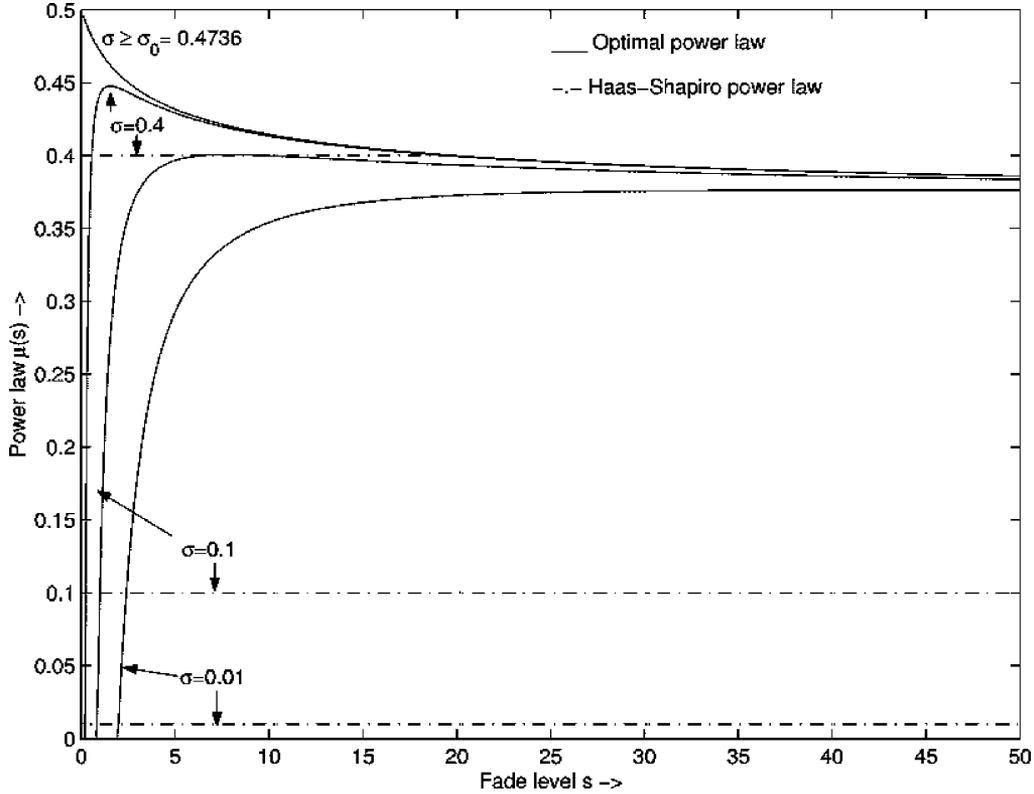
Fig. 5. Optimal power control law with perfect CSI at transmitter and receiver ($\mathrm{SNR} = 0$ dB).

$$- (\mu^{\mathrm{L}} s\alpha + \lambda_0)\mu^{\mathrm{L}} s\alpha/\lambda_0$$
$$+ (\mu^{\mathrm{L}} s\alpha + \lambda_0)\left((\mu^{\mathrm{L}} s\alpha/\lambda_0)^2/2 + O(\lambda_0^{-3})\right)$$
$$= \mu^{\mathrm{L}}(1 - \mu^{\mathrm{L}})s^2\alpha^2/2\lambda_0 + O\left(\lambda_0^{-2}\right) \qquad (103)$$

where we have used the approximation $\ln x = x - x^2/2 + O(x^3)$ for $x \ll 1$ in (103). By (14), we then obtain

$$\mathcal{C}^{\mathrm{L}} = \mu^{\mathrm{L}}(1 - \mu^{\mathrm{L}})\mathbb{E}[S^2]\alpha^2/2\lambda_0 + O(\lambda_0^{-2}). \qquad (104)$$

This establishes the second part of Theorem 3.   □

V. NUMERICAL EXAMPLE

In this example, we consider a channel fade with a lognormal distribution. We compute the capacity values for the two special cases—of perfect and no CSI at the transmitter—discussed in Section III, and compare these values with those determined by the results of [15] as well as those for the channel without fading.

The channel fade is assumed to be an i.i.d. lognormal process, i.e., $S_k \sim S = \exp(2G)$, $k \in \mathbb{Z}^+$, where G is Gaussian with mean μ_{G} and variance σ_{G}^2; take $\mu_{\mathrm{G}} = -\sigma_{\mathrm{G}}^2$ so that the fade is normalized, i.e., $\mathbb{E}[S] = 1$. This implies that the optical medium, on average, does not attenuate or amplify the transmitted signal. The log amplitude variance σ_{G}^2 can vary from 0 (negligible fading) to 0.5 (severe turbulence) [15]; we pick $\sigma_{\mathrm{G}}^2 = 0.1$, which corresponds to a moderately turbulent fade. Denote the (peak) SNR by $\mathrm{SNR} = 10\log_{10}(\alpha/\lambda_0)$ (in decibels). We fix $\alpha = 1$ and vary the parameters σ and λ_0 to study the effect of the average power constraint and SNR on

the channel capacity. Fig. 5 shows the behavior of the optimal power control law $\mu_{\mathrm{P}}^*(\cdot)$ for perfect transmitter CSI (cf. (23)) as a function of the channel fade for different values of σ for $\mathrm{SNR} = 0$ dB. The power control law in [15, eq. (4)] is also plotted for comparison. If $\sigma \leq \sigma_0$ (cf. (22)), the average power constraint is satisfied with equality at optimality, i.e., $\mathbb{E}[\mu_{\mathrm{P}}^*(S)] = \sigma$. In this case, the optimal power control law dictates that the transmitter should not transmit when the channel fade is debilitating (i.e., for small values of s). For $\sigma \ll \sigma_0$, this behavior is similar to the *water-filling* power control law in RF Gaussian fading channels (cf., e.g., [31]). On the other hand, if $\sigma > \sigma_0$, the average power constraint is not satisfied with equality at optimality. In this case, the optimal power control law is given by

$$\mu_{\mathrm{P}}^*(s) = \mu_0(s) = \frac{\lambda_0}{\alpha s}\left(e^{-1}\left(1 + \frac{\alpha s}{\lambda_0}\right)^{\left(1 + \frac{\lambda_0}{\alpha s}\right)} - 1\right) \quad (105)$$

which is equal to the parameter $q_0\left(\frac{s\alpha}{\lambda_0}\right)$ defined in [40, eq. (1.5d)]. This corroborates the fact that when the transmitter and receiver have perfect CSI and the average power constraint σ is ineffective, the instantaneous optical fade can be treated as a multiplicative constant for the peak SNR, and the results from the constant fade case can be directly applied. It is interesting to note that there is a transition region, which corresponds to the case when $\sigma \lesssim \sigma_0$, where the optimal power control law is not monotone. We contrast our results with the previously reported power control law in [15, eq. (4)], which is held constant over a wide range of values of σ, and does not properly exploit the transmitter's knowledge of the channel fade.

In Figs. 6 and 7, we compare the capacity values obtained in different situations. Specifically, we plot the capacity for perfect
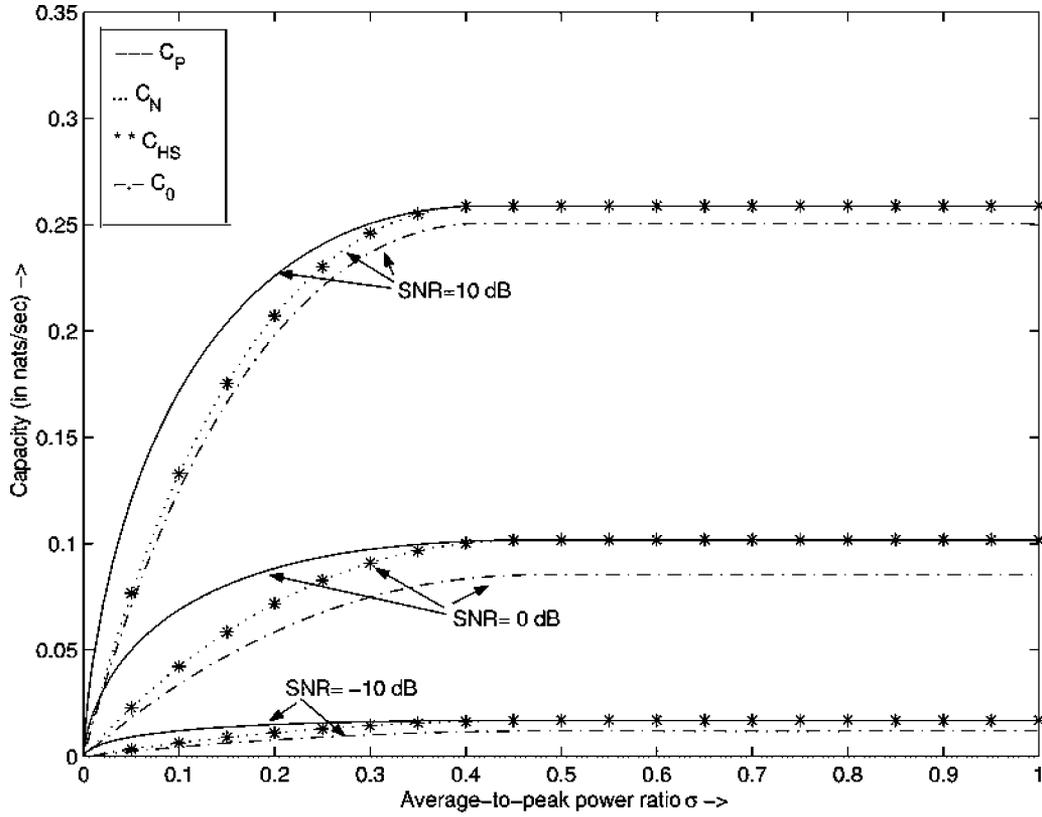
Fig. 6. Comparison of capacity versus σ for various assumptions on transmitter CSI.
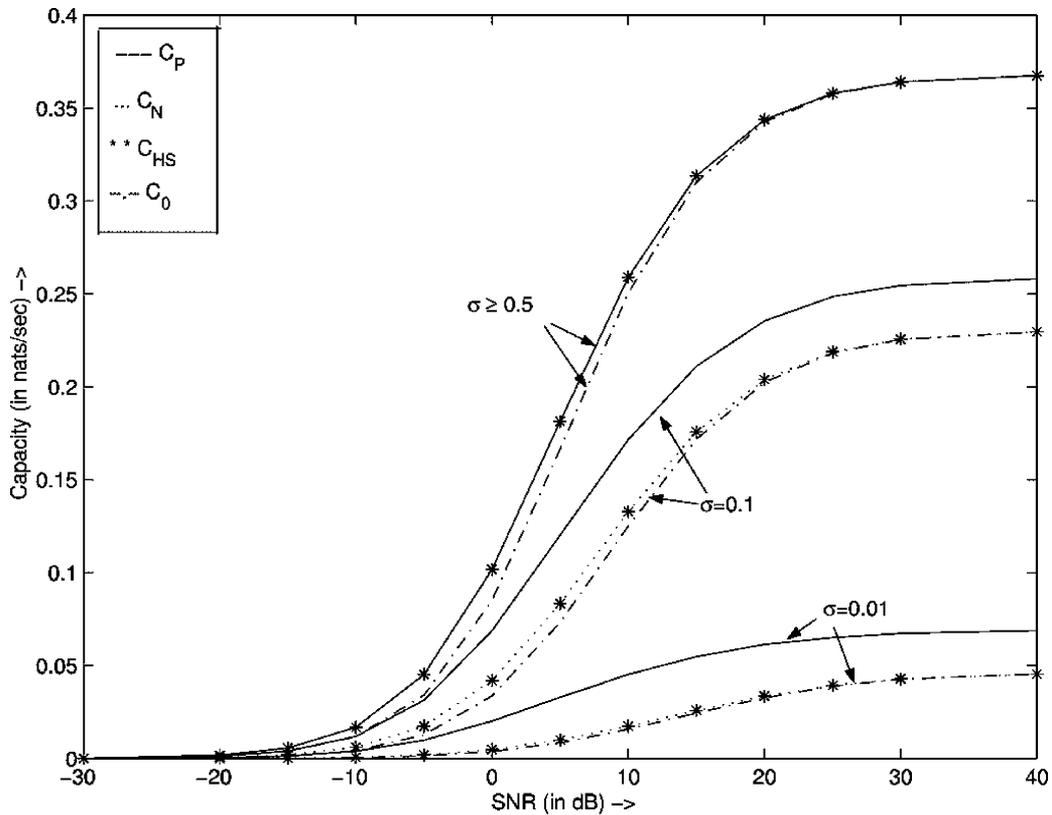

Fig. 7. Comparison of capacity versus SNR for various assumptions on transmitter CSI.

transmitter CSI ($\mathcal{C}_P$), the capacity for no transmitter CSI ($\mathcal{C}_N$), the capacity obtained from [15] ($\mathcal{C}_{HS}$), and the capacity of the Poisson channel without fading ($\mathcal{C}_0$) (cf., e.g., [40, eq. (1.5)]) for various values of σ and SNR. From the plots, we observe that the knowledge of CSI at the transmitter can increase the channel capacity. The improvement is greater for small σ

values, i.e., when the average power constraint is severe. Furthermore, knowledge of CSI at the transmitter can significantly improve capacity at high SNR when σ is small. It is clear that $\mathcal{C}_{\mathrm{HS}}$, the capacity computed from [15], is very close to $\mathcal{C}_{\mathrm{N}}$, the capacity when no CSI is available at the transmitter. This can be understood from the power control law associated with $\mathcal{C}_{\mathrm{HS}}$ in Fig. 5, which does not properly exploit the transmitter's knowledge of the channel condition for a large range of values of σ.

VI. DISCUSSION

We have studied the capacity problem for a single-user SISO shot-noise-limited direct detection block-fading Poisson channel. A channel model for the free-space Poisson fading channel has been proposed in which the channel fade remains unvarying in coherence time intervals of duration $\mathcal{T}_{\mathrm{c}}$, and varies across successive such intervals in an i.i.d. fashion. Under the assumptions of perfect receiver CSI and different extents of transmitter CSI, a (single-letter) characterization of channel capacity is obtained when the transmitted signal is constrained in its peak and average power levels.

Binary signaling with arbitrarily fast intertransition times is shown to be optimal for this channel. The two signaling levels correspond to no transmission ("OFF" state) and transmission at the peak power level ("ON" state). Furthermore, with perfect CSI at the receiver, the channel capacity does not depend on the channel coherence time $\mathcal{T}_{\mathrm{c}}$. An exact characterization of the optimal power control law, which represents the conditional probability of the transmit aperture being in the ON state as a function of current transmitter CSI and which can be viewed as the optimal average duty cycle of the transmitter, has been obtained.

We have also studied the effects of different degrees of CSI at the transmitter on channel capacity as a function of the peak SNR. In the high-SNR regime, an availability of different amounts of CSI at the transmitter can lead to a significant gain in capacity when the average power constraint is stringent. On the other hand, in the low-SNR regime, a knowledge of CSI at the transmitter does not provide any additional advantage. This is in contrast to the results cited in [15], where the authors have claimed that in the high-SNR regime, a knowledge of CSI at the transmitter does not improve capacity. In [15], the capacity formula for perfect transmitter CSI is reported as

$$\mathcal{C}_{\mathrm{P}} = \mathbb{E}\left[\max_{0 \leq \mu \leq \sigma} \mu\zeta(S\alpha, \lambda_0) - \zeta(\mu S\alpha, \lambda_0)\right]$$

and leads to lower values than the exact capacity formula (cf. (15)) in which the transmitter power constraint is less restrictive.

It is natural to ask what happens when perfect CSI is not available at the receiver. An answer, along the lines of Theorems 1–3 above, is not yet known. In the context of RF Gaussian fading channels, it is well known that i.i.d. Gaussian channel inputs no longer achieve channel capacity (cf., e.g., [31]). In the present context of the Poisson fading channel, suppose that the receiver CSI is described by $\{D[k] = g(S[k]), k = 1, 2, \ldots\}$ for a given mapping $g : \mathbb{R}_0^+ \to \mathbb{D}, \mathbb{D} \subseteq \mathbb{R}_0^+$, with $D[k]$ being the receiver CSI during $((k-1)\mathcal{T}_{\mathrm{c}}, k\mathcal{T}_{\mathrm{c}}], k = 1, 2, \ldots$. Then, with $\mathcal{T} = \mathcal{K}\mathcal{T}_{\mathrm{c}}, \mathcal{K} \in \mathbb{Z}^+$, conditioned on a channel input signal $x^{\mathcal{T}}$

and receiver CSI $\boldsymbol{D}^{\mathcal{K}} = \boldsymbol{d}^{\mathcal{K}}$, the channel output rv $(N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}})$ has a conditional sample density function

$$\begin{aligned} &f_{N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}} | X^{\mathcal{T}}, \boldsymbol{D}^{\mathcal{K}}}\left(n_{\mathcal{T}}, \boldsymbol{t}^{n_{\mathcal{T}}} \middle| x^{\mathcal{T}}, \boldsymbol{d}^{\mathcal{K}}\right) \\ &= \int_{\boldsymbol{s}^{\mathcal{K}} \in (\mathbb{R}_0^+)^{\mathcal{K}}} f_{N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}} | X^{\mathcal{T}}, \boldsymbol{S}^{\mathcal{K}}}\left(n_{\mathcal{T}}, \boldsymbol{t}^{n_{\mathcal{T}}} \middle| x^{\mathcal{T}}, \boldsymbol{s}^{\mathcal{K}}\right) \\ &\quad \times \prod_{k=1}^{\mathcal{K}} \mathbb{1}\left(d[k] = g(s[k])\right) d\boldsymbol{s}^{\mathcal{K}} \end{aligned} \tag{106}$$

where $f_{N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}} | X^{\mathcal{T}}, \boldsymbol{S}^{\mathcal{K}}}(\cdot, \cdot | \cdot, \cdot)$ is given by (34). Now, with regard to Remark (ii) following Theorem 1 in Section II, it is clear from (106) that conditioned on a transmitted signal and the receiver CSI, the segments of the received signal in different coherence intervals need no longer be independent across these intervals; and channel capacity, in general, can depend on the coherence time $\mathcal{T}_{\mathrm{c}}$.

APPENDIX A
PROOF OF (50)

We begin with the following proposition which paraphrases a result from [12], [13], and provide a different and simpler proof.

Proposition 1: If $(\bar{N}_{\mathcal{T}}, \bar{\boldsymbol{T}}^{\bar{N}_{\mathcal{T}}}) = (\bar{N}_{\mathcal{T}}, (\bar{T}_1, \ldots, \bar{T}_{\bar{N}_{\mathcal{T}}}))$ is a PCP with (deterministic) rate function $\{\lambda(t), 0 \leq t \leq \mathcal{T}\}$, then for any function $g(\cdot)$ integrable on $[0, \mathcal{T}]$

$$\mathbb{E}\left[\mathbb{1}(\bar{N}_{\mathcal{T}} \geq 1) \sum_{i=1}^{\bar{N}_{\mathcal{T}}} g(\bar{T}_i)\right] = \int_0^{\mathcal{T}} \lambda(t)g(t)dt. \tag{107}$$

Proof: Write

$$\begin{aligned} &\mathbb{E}\left[\mathbb{1}(\bar{N}_{\mathcal{T}} \geq 1) \sum_{i=1}^{\bar{N}_{\mathcal{T}}} g(\bar{T}_i)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}(\bar{N}_{\mathcal{T}} \geq 1) \sum_{i=1}^{\bar{N}_{\mathcal{T}}} g(\bar{T}_i) \middle| \bar{N}_{\mathcal{T}}\right]\right]. \end{aligned} \tag{108}$$

Note that conditioned on $\bar{N}_{\mathcal{T}} = 0$, the value of the conditional expectation in (108) is 0. Next, use the fact (cf., e.g., [35, pp. 62–63]) that conditioned on $\bar{N}_{\mathcal{T}} = n \geq 1$, the conditional joint distribution of the rvs $\bar{T}_1, \ldots, \bar{T}_n$ is the same as the joint distribution of the i.i.d. rvs $U_1, \ldots, U_n$ with (common) probability density function

$$f_U(u) = \begin{cases} \frac{\lambda(u)}{\int_0^{\mathcal{T}} \lambda(t)dt}, & \text{if } 0 \leq u \leq \mathcal{T} \\ 0, & \text{otherwise} \end{cases} \tag{109}$$

to obtain that, conditioned on $\bar{N}_{\mathcal{T}} = n \geq 1$, the value of the conditional expectation in (108) is

$$\begin{aligned} &\mathbb{E}\left[\mathbb{1}(\bar{N}_{\mathcal{T}} \geq 1) \sum_{i=1}^{\bar{N}_{\mathcal{T}}} g(\bar{T}_i) \middle| \bar{N}_{\mathcal{T}} = n\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n} g(\bar{T}_i) \middle| N_{\mathcal{T}} = n\right] \end{aligned}$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} g(U_i)\right]$$

$$= n\mathbb{E}[g(U_1)]$$

$$= n\frac{\int_0^{\mathcal{T}} g(t)\lambda(t)dt}{\int_0^{\mathcal{T}} \lambda(t)dt} \tag{110}$$

where the last equality is by (109). Hence, the conditional expectation in (108) equals

$$\frac{\int_0^{\mathcal{T}} g(t)\lambda(t)dt}{\int_0^{\mathcal{T}} \lambda(t)dt}\bar{N}_{\mathcal{T}}$$

whence the left-hand side of (107) is

$$\mathbb{E}\left[\mathbb{1}(\bar{N}_{\mathcal{T}} \geq 1)\sum_{i=1}^{\bar{N}_{\mathcal{T}}} g(\bar{T}_i)\right]$$

$$= \frac{\int_0^{\mathcal{T}} g(t)\lambda(t)dt}{\int_0^{\mathcal{T}} \lambda(t)dt}\mathbb{E}[\bar{N}_{\mathcal{T}}]$$

$$= \frac{\int_0^{\mathcal{T}} g(t)\lambda(t)dt}{\int_0^{\mathcal{T}} \lambda(t)dt} \cdot \int_0^{\mathcal{T}} \lambda(t)dt \tag{111}$$

$$= \int_0^{\mathcal{T}} g(t)\lambda(t)dt \tag{112}$$

where (111) follows by the fact that $\bar{N}_{\mathcal{T}}$ is a Poisson rv with mean $\int_0^{\mathcal{T}} \lambda(t)dt$. $\qquad\square$

Next, recall that $(N_{\mathcal{T}}, \boldsymbol{T}^{N_{\mathcal{T}}})$ is a PCP with rate process $\{\Lambda(t), 0 \leq t \leq \mathcal{T}\}$. Therefore, by (107)

$$\mathbb{E}\left[\mathbb{1}(N_{\mathcal{T}} \geq 1)\sum_{i=1}^{N_{\mathcal{T}}} \ln \Lambda(T_i)\Bigg| \Lambda^{\mathcal{T}} = \lambda^{\mathcal{T}}\right] = \int_0^{\mathcal{T}} \lambda(t)\ln \lambda(t)dt$$

so that

$$\mathbb{E}\left[\mathbb{1}(N_{\mathcal{T}} \geq 1)\sum_{i=1}^{N_{\mathcal{T}}} \ln \Lambda(T_i)\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}(N_{\mathcal{T}} \geq 1)\sum_{i=1}^{N_{\mathcal{T}}} \ln \Lambda(T_i)\Bigg| \Lambda^{\mathcal{T}}\right]\right]$$

$$= \mathbb{E}\left[\int_0^{\mathcal{T}} \Lambda(t)\ln \Lambda(t)dt\right]. \tag{113}$$

Similarly, we can write akin to (113) that

$$\mathbb{E}\left[\mathbb{1}(N_{\mathcal{T}} \geq 1)\sum_{i=1}^{N_{\mathcal{T}}} \ln \hat{\Lambda}(T_i)\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}(N_{\mathcal{T}} \geq 1)\sum_{i=1}^{N_{\mathcal{T}}} \ln \hat{\Lambda}(T_i)\Bigg| \hat{\Lambda}^{\mathcal{T}}\right]\right]$$

$$= \mathbb{E}\left[\int_0^{\mathcal{T}} \Lambda(t)\ln \hat{\Lambda}(t)dt\right]. \tag{114}$$

Combining (113), (114), we get

$$\mathbb{E}\left[\mathbb{1}(N_{\mathcal{T}} \geq 1)\sum_{i=1}^{N_{\mathcal{T}}} \left(\ln \Lambda(T_i) - \ln \hat{\Lambda}(T_i)\right)\right]$$

$$= \mathbb{E}\left[\int_0^{\mathcal{T}} \left\{\Lambda(t)\ln \Lambda(t) - \Lambda(t)\ln \hat{\Lambda}(t)\right\}dt\right]$$

$$= \int_0^{\mathcal{T}} \left\{\mathbb{E}\left[\Lambda(t)\ln \Lambda(t) - \Lambda(t)\ln \hat{\Lambda}(t)\right]\right\}dt \tag{115}$$

where (115) is by an interchange of operations.[16] Noting that

$$\mathbb{E}\left[\Lambda(t)\ln \hat{\Lambda}(t)\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\Lambda(t)\ln \hat{\Lambda}(t)\Big| N_t, \boldsymbol{T}^{N_t}, \boldsymbol{S}^{\lceil t/\mathcal{T}_c\rceil}\right]\right]$$

$$= \mathbb{E}\left[\ln \hat{\Lambda}(t)\mathbb{E}\left[\Lambda(t)\Big| N_t, \boldsymbol{T}^{N_t}, \boldsymbol{S}^{\lceil t/\mathcal{T}_c\rceil}\right]\right]$$

$$= \mathbb{E}\left[\hat{\Lambda}(t)\ln \hat{\Lambda}(t)\right]$$

we get from (115) that

$$\mathbb{E}\left[\mathbb{1}(N_{\mathcal{T}} \geq 1)\sum_{i=1}^{N_{\mathcal{T}}} \left(\ln \Lambda(T_i) - \ln \hat{\Lambda}(T_i)\right)\right]$$

$$= \int_0^{\mathcal{T}} \left\{\mathbb{E}\left[\Lambda(t)\ln \Lambda(t) - \hat{\Lambda}(t)\ln \hat{\Lambda}(t)\right]\right\}dt$$

$$= \int_0^{\mathcal{T}} \left\{\mathbb{E}\left[\zeta\left(S[\lceil t/\mathcal{T}_c\rceil]X(t), \lambda_0\right)\right]\right.$$

$$\left. - \mathbb{E}\left[\zeta\left(S[\lceil t/\mathcal{T}_c\rceil]\hat{X}(t), \lambda_0\right)\right]\right\}dt \tag{116}$$

where (116) is by (7), (36), (37).

APPENDIX B
PROOF OF (88)

In (78), for $\mathcal{L} \gg 1$, we use the approximation

$$\exp(x/\mathcal{L}) = 1 + x/\mathcal{L} + O(\mathcal{L}^{-2}), \qquad x \in \mathbb{R}$$

to get

$$w_{10}(\mathcal{L}) = \lambda_0 \mathcal{T}_c/\mathcal{L} + O\left(\mathcal{L}^{-2}\right)$$

$$w_{11}(s, \mathcal{L}) = (\lambda_0 + s\alpha)\mathcal{T}_c/\mathcal{L} + O\left(\mathcal{L}^{-2}\right) \tag{117}$$

and in (85), for $x \ll 1$, we use the approximations [40]

$$h_b(x) = -x\ln x + x + O(x^2)$$

$$h_b(x + O(x^2)) = h_b(x) + O(x^2 \ln x). \tag{118}$$

Then, by (85) and (117), we have that for $\mathcal{L} \gg 1$, $s \in \mathbb{R}_0^+$

$$\beta_{\mathcal{L}}(s) = h_b\left((\lambda_0 + s\mu(h(s))\alpha)\mathcal{T}_c/\mathcal{L} + O\left(\mathcal{L}^{-2}\right)\right)$$

$$- \mu(h(s))h_b\left((\lambda_0 + s\alpha)\mathcal{T}_c/\mathcal{L} + O\left(\mathcal{L}^{-2}\right)\right)$$

$$- (1 - \mu(h(s)))h_b\left(\lambda_0 \mathcal{T}_c/\mathcal{L} + O\left(\mathcal{L}^{-2}\right)\right)$$

$$= \frac{\mathcal{T}_c}{\mathcal{L}}\left[\mu(h(s))(\lambda_0 + s\alpha)\ln(\lambda_0 + s\alpha)\right.$$

$$+ (1 - \mu(h(s)))\lambda_0 \ln \lambda_0$$

$$\left. - (\lambda_0 + s\mu(h(s))\alpha)\ln\left(\lambda_0 + s\mu(h(s))\alpha\right)\right]$$

$$+ O\left(\mathcal{L}^{-2}\ln \mathcal{L}\right) \tag{119}$$

$$= \frac{\mathcal{T}_c}{\mathcal{L}}\left[\mu(h(s))\zeta(s\alpha, \lambda_0) - \zeta(s\mu(h(s))\alpha, \lambda_0)\right]$$

$$+ O\left(\mathcal{L}^{-2}\ln \mathcal{L}\right) \tag{120}$$

[16]The assumed condition $\mathbb{E}[|\zeta(S\alpha, \lambda_0)|] < \infty$ ensures the integrability of $\{\Lambda(t)\ln \Lambda(t) - \Lambda(t)\ln \hat{\Lambda}(t), 0 \leq t \leq \mathcal{T}\}$, thereby permitting the interchange.

where (119) is by (118), and (120) is by (7). Hence, for each $s \in \mathbb{R}_0^+$

$$\lim_{\mathcal{L} \to \infty} \frac{\beta_{\mathcal{L}}(s)}{\mathcal{T}_c/\mathcal{L}} = \mu(h(s))\zeta(s\alpha, \lambda_0) - \zeta(s\mu(h(s))\alpha, \lambda_0)$$

whereby (88) follows upon interchanging the limit and the expectation which is permissible by virtue of the boundedness of $\beta_{\mathcal{L}}(\cdot)$.

ACKNOWLEDGMENT

The authors express their gratitude to the referees and the Associate Editor for their thorough reviews and comprehensive comments. In particular, they thank the referees for suggestions which led to the inclusion of a heuristic argument that anticipates the result of Theorem 1, and for providing several helpful pointers to work on free-space optical communication.

REFERENCES

[1] I. Bar-David, "Information in the time of arrival of a photon packet: Capacity of PPM channels," *J. Opt. Soc. Amer.*, vol. 63, no. 2, pp. 166–170, Feb. 1973.

[2] A. Biswas and M. W. Wright, "Mountain-Top-to-Mountain-Top Optical Link Demonstration: Part I: The Interplanetary Network," Jet Propulsion Lab., Pasadena, CA, 2002, Progr. Rep. 42–149.

[3] S. A. Butman, J. Katz, and J. R. Lesh, "Bandwidth limitations on noiseless optical channel capacity," *IEEE Trans. Commun.*, vol. COMM-30, no. 5, pp. 1262–1264, May 1982.

[4] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

[5] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, Jul. 1999.

[6] K. Chakraborty and P. Narayan, "The MIMO Poisson fading channel," *IEEE Trans. Inf. Theory*, unpublished.

[7] A. Das and P. Narayan, "Capacities of time-varying multiple-access channels with side information," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 4–25, Jan. 2002.

[8] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 710–715, Nov. 1980.

[9] G. Fillmore and G. Lachs, "Information rates for photocount detection systems," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 4, pp. 465–468, Jul. 1969.

[10] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Englewood Cliffs, NJ: Prentice-Hall, 1963.

[11] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[12] S. M. Haas, "Capacity and Coding for Multiple-Aperture, Wireless, Optical Communications," Ph.D. dissertation, MIT, Cambridge, MA, 2003.

[13] S. M. Haas and J. H. Shapiro, "Capacity of the multiple-input, multiple-output Poisson channel," in *Proc. Kansas Workshop on Stochastic Theory and Control*, B. Pasik-Duncan, Ed., Lawrence, KS, Oct. 2001, Lecture Notes in Control and Information Sciences.

[14] S. M. Haas, J. H. Shapiro, and V. Tarokh, "Space-time codes for wireless optical communications," *Eurasip J. Appl. Signal Process.*, vol. 2002, no. 3, pp. 211–220, Mar. 2002.

[15] S. M. Haas and J. H. Shapiro, "Capacity of wireless optical communications," *IEEE J. Select. Areas Commun.*, vol. 21, no. 8, pp. 1346–1357, Oct. 2003.

[16] J. Hamkins and B. Moision, "Multipulse Pulse-Position Modulation on Discrete Memoryless Channels," Jet Propulsion Lab., Pasadena, CA, 2005, The Interplanetary Network Progr. Rep. 42–161.

[17] S. Hranilovic and F. R. Kschischang, "Capacity bounds for power- and band-limited optical intensity channels corrupted by Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 784–795, May 2004.

[18] A. Ishimaru, *Wave Propagation and Scattering in Random Media*. New York: IEEE Press and Oxford Univ. Press, 1997.

[19] M. Jeganathan, K. E. Wilson, and J. R. Lesh, "Preliminary Analysis of Fluctuations in the Received Uplink-Beacon-Power Data Obtained From the GOLD Experiments," Jet Propulsion Lab., Pasadena, California, 1996, The Telecommunications and Data Acquisition Project Rep. 42–124.

[20] Y. M. Kabanov, "Capacity of a channel of the Poisson type," *Theory Probab. its Applic.*, vol. 23, no. 1, pp. 143–147, 1978.

[21] I. I. Kim, H. Hakakha, P. Adhikari, E. Korevaar, and A. K. Mazumdar, "Scintillation reduction using multiple transmitters," *Free Space Laser Communication Technologies IX, Proc. SPIE*, vol. 2990, pp. 102–113, 1997.

[22] E. J. Lee and V. W. S. Chan, "Part 1: Optical communication over the clear turbulent atmospheric channel using diversity," *IEEE J. Select. Areas Commun.*, vol. 22, no. 9, pp. 1896–1906, Nov. 2004.

[23] J. R. Lesh, "Capacity limit of the noiseless, energy-efficient optical PPM channel," *IEEE Trans. Commun.*, vol. COMM-31, no. 4, pp. 546–548, Apr. 1983.

[24] R. G. Lipes, "Pulse-Position-Modulation Coding as Near-Optimum Utilization of Photon Counting Channel With Bandwidth and Power Constraints," Jet Propulsion Lab., Pasadena, CA, 1980, The Deep Space Network Progr. Rep. 42–56.

[25] R. J. McEliece, "Practical codes for photon communication," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 4, pp. 393–397, Jul. 1981.

[26] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 1, pp. 44–53, Jan. 1984.

[27] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[28] J. R. Pierce, "Optical channels—Practical limits with photon counting," *IEEE Trans. Commun.*, vol. COMM-26, no. 12, pp. 1819–1821, Dec. 1978.

[29] J. R. Pierce, E. C. Posner, and E. R. Rodemich, "The capacity of the photon counting channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 61–76, Jan. 1981.

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

[31] J. Proakis, E. Biglieri, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[32] S. Shamai (Shitz), "On the capacity of a direct-detection photon channel with intertransition-constrained binary input," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1540–1550, Nov. 1991.

[33] S. Shamai (Shitz) and A. Lapidoth, "Bounds on the capacity of a spectrally constrained Poisson channel," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 19–29, Jan. 1993.

[34] J. H. Shapiro, "Imaging and optical communication through atmospheric turbulence," in *Laser Beam Propagation in the Atmosphere*, J. W. Strohbehn, Ed. Berlin, Germany: Springer, 1978, ch. 6.

[35] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, 2nd ed. New York: Springer-Verlag, 1991.

[36] V. I. Tatarskii, *Wave Propagation in a Turbulent Medium*. New York: Dover, 1961.

[37] S. Verdú, "Poisson communication theory," presented at the International Technion Communication Day in Honor of Israel Bar-David, Haifa, Israel, Mar. 25, 1999, Invited Talk.

[38] H. A. Willebrand and B. S. Ghuman, "Fiber optics without fiber," *IEEE Spectrum*, vol. 38, no. 8, pp. 41–45, Aug. 2001.

[39] S. G. Wilson, M. Brandt-Pierce, Q. Cao, and J. Levesque, "Free-space optical MIMO communication with Q-ary PPM," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1402–1412, Aug. 2005.

[40] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel, parts I and II," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1449–1471, Nov. 1988.

[41] R. You and J. M. Kahn, "Upper-bounding the capacity of optical IM/DD channels with multiple subcarrier modulation and fixed bias using trigonometric moment space method," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 514–523, Feb. 2002.

[42] X. Zhu and J. M. Kahn, "Free-space optical communication through atmospheric turbulent channels," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1293–1300, Aug. 2002.