

On the Capacity of the Discrete-Time Poisson Channel

Amos Lapidoth, *Fellow, IEEE*, and Stefan M. Moser, *Member, IEEE*

Abstract—The large-inputs asymptotic capacity of a peak-power and average-power limited discrete-time Poisson channel is derived using a new firm (nonasymptotic) lower bound and an asymptotic upper bound. The upper bound is based on the dual expression for channel capacity and the notion of *capacity-achieving input distributions that escape to infinity*. The lower bound is based on a lower bound on the entropy of a conditionally Poisson random variable in terms of the differential entropy of its conditional mean.

Index Terms—Channel capacity, direct detection, high signal-to-noise ratio (SNR), optical communication, PAM, photon, Poisson channel.

I. INTRODUCTION

WE consider a memoryless discrete-time channel whose output Y takes value in the set of nonnegative integers \mathbb{Z}_0^+ and whose input takes value in the set of nonnegative real numbers \mathbb{R}_0^+ . Conditional on the input $x \geq 0$, the output is Poisson distributed with mean $x + \lambda_0$, where λ_0 is some nonnegative constant, called *dark current*. Thus, the conditional channel law is given by

$$W(y|x) = e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^y}{y!}, \quad y \in \mathbb{Z}_0^+, x \in \mathbb{R}_0^+. \quad (1)$$

This channel is often used to model pulse-amplitude modulated (PAM) optical communication with a direct-detection receiver [1]. Here the input x is proportional to the product of the transmitted light intensity by the pulse duration; the dark current λ_0 similarly models the time-by-intensity product of the background radiation; and the output Y models the number of photons arriving at the receiver during the pulse duration. A peak-power constraint on the transmitter is accounted for by the peak-input constraint

$$\Pr[X > A] = 0 \quad (2)$$

Manuscript received October 12, 2007; revised March 01, 2008. Current version published December 24, 2008. The work of S. M. Moser was supported in part by the ETH under Grant TH-23 02-2. The material in this paper was presented in part at the 2003 Winter School on Coding and Information Theory, Monte Verità, Ascona, Switzerland, February 2003, and at the 41st Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, IL, October 2003, and have been published as part of S. M. Moser's Ph.D. dissertation.

A. Lapidoth is with the Department of Information Technology and Electrical Engineering, Swiss Federal Institute of Technology (ETH), 8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch).

S. M. Moser is with the Department of Communication Engineering, National Chiao Tung University (NCTU), Hsinchu 30010, Taiwan (stefan.moser@ieec.org).

Communicated by Y. Steinberg, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2008.2008121

and an average-power constraint by

$$E[X] \leq \mathcal{E}. \quad (3)$$

Note that since the input is proportional to the light intensity, the power constraints apply to the input directly and not to the square of its magnitude (as is usually the case for electrical transmission models).

We use $0 < \alpha \leq 1$ to denote the average-to-peak-power ratio

$$\alpha \triangleq \frac{\mathcal{E}}{A}. \quad (4)$$

The case $\alpha = 1$ corresponds to the absence of an average-power constraint, whereas $\alpha \ll 1$ corresponds to a very weak peak-power constraint.

Although we also provide firm lower bounds on channel capacity that are valid for all values of the peak and average power, our main interest in this paper is mostly in the case where both the allowed average power and the allowed peak power are large. In fact, we shall compute the asymptotic behavior of channel capacity as both A and \mathcal{E} tend to infinity with the ratio α held fixed. The low-input regime where the input power is small was studied in [2] and [3].

No analytic expression for the capacity of the Poisson channel is known. In [1] Shamai showed that capacity-achieving input distributions are discrete with a finite number of mass points, where the number of mass points increases to infinity as the constraints are relaxed.

In [4] Brady & Verdú considered the case of the Poisson channel with only an average-power constraint. The following bounds were derived. Let \mathcal{E} and λ_0 tend to infinity with their ratio $\text{SNR} \triangleq \frac{\mathcal{E}}{\lambda_0}$ held fixed. Given $\epsilon > 0$ there exists an \mathcal{E}_ϵ such that for all $\mathcal{E} > \mathcal{E}_\epsilon$ the capacity is bounded by

$$\mathcal{C}(\mathcal{E}) \geq \frac{1}{2} \log \frac{\mathcal{E}}{2\pi} - \frac{1}{2} \log \left(1 + \frac{1}{\text{SNR}} \right) - \epsilon \quad (5)$$

$$\begin{aligned} \mathcal{C}(\mathcal{E}) \leq & \frac{1}{2} \log \frac{\mathcal{E}}{2\pi} + \log \left(\sqrt{\text{SNR}} \left(1 + \frac{1}{\mathcal{E}_\epsilon} \right) + \frac{1}{\sqrt{\text{SNR}}} \right) \\ & + 1 + \log \frac{3}{2} + \epsilon. \end{aligned} \quad (6)$$

Note that the difference between the upper and lower bound is unbounded if the dark current is held constant while \mathcal{E} tends to infinity.

While the capacity of the discrete-time Poisson channel is unknown, the capacity of the general continuous-time Poisson channel where the input signal is not restricted to be pulse-amplitude modulated (PAM) has been derived exactly: the case with a peak-power constraint only was solved by Kabanov [5]; the more general situation of peak- and average-power

constraints was treated by Davis [6]; Wyner [7] found the reliability function of the channel; and Frey [8] [9] studied the capacity of the Poisson channel under an L -norm constraint.

The capacity of the continuous-time Poisson channel can only be achieved by input processes that have unbounded bandwidth. Since this is not realistic Shamai and Lapidot [10] investigated the channel capacity of a Poisson channel with some spectral constraints, but without restricting the input to use PAM. Note that even though the penalty incurred by the PAM scheme tends to zero once the pulse duration is shortened to zero, a PAM scheme is not optimal if we only limit the minimal pulse width, but not the pulse shape [11].

Besides the Poisson channel model there are a few related channel models used to describe optical communication. The free-space optical intensity channel has been investigated in [12, Ch. 3], [13], [14], [15], [16], [17], [18]. A variation of this model where the noise depends on the input has been studied in [12, Ch. 4], [19].

One of the obstacles to an exact expression for the capacity of the Poisson channel is that the Poisson distribution does not seem to admit a simple analytic expression. Recently, however, Martinez [20] derived a new expression for the entropy of a Poisson random variable based on an integral representation that can be easily computed numerically. Using this expression he derived firm lower and upper bounds on the capacity for the discrete-time Poisson channel with only an average-power constraint and no dark current:

$$\mathcal{C}(\mathcal{E}) \geq \frac{1}{2} \log(1 + \mathcal{E}) \quad (7)$$

$$\mathcal{C}(\mathcal{E}) \leq \left(\mathcal{E} + \frac{1}{2}\right) \log\left(\mathcal{E} + \frac{1}{2}\right) - \mathcal{E} \log \mathcal{E} - \frac{1}{2} + \log\left(1 + \frac{\sqrt{2e} - 1}{\sqrt{1 + 2\mathcal{E}}}\right). \quad (8)$$

Similarly to the bounds presented here, the derivation of (8) is based on a duality approach. We would like to emphasize that (8) is a firm bound valid for all values of \mathcal{E} whereas we will present upper bounds that are only valid asymptotically as the available power tends to infinity. However, in the derivation in [20] there is a tiny gap in the proof that is shown only numerically. Nevertheless Martinez' bounds are very close and actually tighter than the bounds presented here (see Figure 2 in Section II below).

Here we present results for the more general case where we enforce both peak- and average-power constraints and assume a general (nonnegative) dark current λ_0 . We will derive new lower bounds on channel capacity that are tighter than previous bounds. These bounds are based on a new result that proves that the entropy of the output of a Poisson channel is always larger than the differential entropy of the channel's input (see Section III-B for more details).

We will also introduce an asymptotic upper bound on channel capacity, where "asymptotic" means that the bound is valid when the available peak and average power tend to infinity with their ratio held fixed.¹ The upper and lower bounds

asymptotically coincide, thus yielding the exact asymptotic behavior of channel capacity.

The derivation of the upper bounds is based on a technique introduced in [21] using a dual expression for mutual information. We will not state it in its full generality but adapted to the form needed in this paper. For more details and for a proof we refer to [21, Sec. V], [12, Ch. 2].

Proposition 1. Assume a channel² $\tilde{W}(\cdot|\cdot)$ with input alphabet $\mathcal{X} = \mathbb{R}_0^+$ and output alphabet $\mathcal{Y} = \mathbb{R}_0^+$. Then for an arbitrary distribution $R(\cdot)$ over the channel output alphabet, the channel capacity is upper-bounded by

$$\mathcal{C} \leq \mathbb{E}_{Q^*} \left[D(\tilde{W}(\cdot|X) \| R(\cdot)) \right]. \quad (9)$$

Here, $D(\cdot|\cdot)$ stands for the relative entropy [22, Ch. 2], and $Q^*(\cdot)$ denotes the capacity-achieving input distribution.

Proof: See [21, Sec. V]. ■

The challenge of using (9) lies in a clever choice of the arbitrary law $R(\cdot)$ that will lead to a good upper bound. Moreover, note that the bound (9) still contains an expectation over the (unknown) capacity-achieving input distribution $Q^*(\cdot)$. To handle this expectation we will need to resort to the concept of *input distributions that escape to infinity* as introduced in [21], [23]. This concept will be briefly reviewed in Section IV-B1.

The results of this paper are partially based on [24] and have appeared in the Ph.D. thesis [12, Ch. 5].

The remainder of this paper is structured as follows. After some brief remarks about our notation, we summarize our main results in the subsequent section. The derivations are then given in Section III (lower bounds) and Section IV (upper bounds). These two derivation sections both contain a subsection with mathematical preliminaries. In particular, in Section III-B we prove that the entropy of the output of a Poisson channel is lower-bounded by the differential entropy of its input, in Section IV-B1 we review the concept of *input distributions that escape to infinity*, and in Section IV-B2 we show an adapted version of the channel model with continuous channel output. We will conclude in Section V.

We try to distinguish between those quantities that are random and those that are constant: for random quantities we use uppercase letters and for their realizations lowercase letters. Scalars are typically denoted using Greek letters or lower-case Roman letters. However, there will be a few exceptions to these rules. Since they are widely used in the literature, we will stick with the common customary shape of the following symbols: \mathcal{C} stands for capacity, $H(\cdot)$ denotes the entropy of a discrete random variable, $D(\cdot|\cdot)$ denotes the relative entropy between two probability measures, and $I(\cdot;\cdot)$ stands for the mutual information functional. Moreover, we have decided to use the capitals Q , W , and R to denote probability mass functions (PMF) in case of discrete random variables or cumulative distribution functions (CDF) in case of continuous random variables, respectively:

- $Q(\cdot)$ denotes a distribution on an input of a channel;

¹In contrast to [4] we regard the dark current as a parameter of the channel that remains unchanged, *i.e.*, we will always keep λ_0 constant.

²There are certain measurability assumptions on the channel that we omit for simplicity. See [21, Sec. V], [12, Ch. 2].

- $W(\cdot|\cdot)$ denotes a channel law, *i.e.*, the distribution of the channel output conditioned on the channel input; and
- $R(\cdot)$ denotes a distribution on the channel output.

In the case when $Q(\cdot)$ or $R(\cdot)$ represents a CDF, the corresponding probability density function (PDF) is denoted by $Q'(\cdot)$ and $R'(\cdot)$, respectively.

The symbol \mathcal{E} denotes average power and A stands for peak power. We shall denote the mean- η Poisson distribution by $\mathcal{P}o(\eta)$ and the uniform distribution on the interval $[a, b)$ by $\mathcal{U}([a, b))$. All rates specified in this paper are in nats per channel use, and all logarithms are natural logarithms.

Finally, we give the following definition.

Definition 2. Let $f : \mathbb{R}_0^+ \rightarrow \mathbb{R}$ be a function that tends to zero as its argument tends to infinity, *i.e.*, for any $\epsilon > 0$ there exists a constant z_0 such that for all $z > z_0$

$$|f(z)| < \epsilon. \quad (10)$$

Then we write³

$$f(z) = o_z(1). \quad (11)$$

II. MAIN RESULTS

We present upper and lower bounds on the capacity of channel (1). While the lower bounds are valid for all values of the power, the upper bounds are valid asymptotically only, *i.e.*, only in the limit when the average power and the peak power tend to infinity with their ratio kept fixed. It will turn out that in this limit the lower and upper bounds coincide, *i.e.*, asymptotically we can specify the capacity precisely.

We distinguish between three cases: in the first case we have both an average- and a peak-power constraint where the average-to-peak-power ratio (4) is in the range $0 < \alpha < \frac{1}{3}$. In the second case $\frac{1}{3} \leq \alpha \leq 1$, which includes the situation with only a peak-power constraint $\alpha = 1$. And finally, in the third case we look at the situation with only an average-power constraint.

We begin with the first case.

Theorem 3. The channel capacity $\mathcal{C}(A, \mathcal{E})$ of a Poisson channel with dark current λ_0 under a peak-power constraint (2) and an average-power constraint (3), where the ratio $\alpha = \frac{\mathcal{E}}{A}$ lies in $(0, \frac{1}{3})$, is bounded as follows.

$$\begin{aligned} \mathcal{C}(A, \mathcal{E}) \geq & \frac{1}{2} \log A - (1 - \alpha)\mu - \log \left(\frac{1}{2} - \alpha\mu \right) \\ & - e^\mu \left(\frac{1}{2} - \alpha\mu \right) \left[\log \left(1 + \frac{\lambda_0 + \frac{1}{12}}{A} \right) \right. \\ & \left. + 2\sqrt{\frac{\lambda_0 + \frac{1}{12}}{A}} \arctan \left(\sqrt{\frac{A}{\lambda_0 + \frac{1}{12}}} \right) \right] \\ & + (\mathcal{E} + 1) \log \left(1 + \frac{1}{\mathcal{E}} \right) - 1 - \frac{1}{2} \log 2\pi e \end{aligned} \quad (12)$$

³Note that by this notation we want to imply that $o_z(1)$ does not depend on any other nonconstant variable apart from z .

$$\mathcal{C}(A, \alpha A) \leq \frac{1}{2} \log A - (1 - \alpha)\mu - \log \left(\frac{1}{2} - \alpha\mu \right) - \frac{1}{2} \log 2\pi e + o_A(1). \quad (13)$$

Here μ is the solution to

$$\alpha = \frac{1}{2\mu} - \frac{e^{-\mu}}{\sqrt{\mu}\sqrt{\pi}\operatorname{erf}(\sqrt{\mu})} \quad (14)$$

where the error function $\operatorname{erf}(\cdot)$ is defined as

$$\operatorname{erf}(\xi) \triangleq \frac{2}{\sqrt{\pi}} \int_0^\xi e^{-t^2} dt = 1 - 2\mathcal{Q}(\sqrt{2}\xi), \quad \forall \xi \in \mathbb{R} \quad (15)$$

with the Gaussian \mathcal{Q} -function

$$\mathcal{Q}(\xi) \triangleq \int_\xi^\infty \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt, \quad \forall \xi \in \mathbb{R}. \quad (16)$$

Note that the function $\mu \mapsto \frac{1}{2\mu} - \frac{e^{-\mu}}{\sqrt{\mu}\sqrt{\pi}\operatorname{erf}(\sqrt{\mu})}$ is monotonically decreasing in $[0, \infty)$ and tends to $\frac{1}{3}$ for $\mu \downarrow 0$ and to 0 for $\mu \uparrow \infty$.

The error term $o_A(1)$ tends to zero as the average power and the peak power tend to infinity with their ratio held fixed at α , $0 < \alpha < \frac{1}{3}$. Hence, the asymptotic expansion of channel capacity is

$$\begin{aligned} \lim_{A \uparrow \infty} \left\{ \mathcal{C}(A, \alpha A) - \frac{1}{2} \log A \right\} \\ = -\frac{1}{2} \log 2\pi e - (1 - \alpha)\mu - \log \left(\frac{1}{2} - \alpha\mu \right), \\ 0 < \alpha < \frac{1}{3} \quad (17) \end{aligned}$$

where μ is defined as above to be the solution to (14).

In the second case $\alpha \geq \frac{1}{3}$, we have the following bounds.

Theorem 4. The channel capacity $\mathcal{C}(A, \mathcal{E})$ of a Poisson channel with dark current λ_0 under a peak-power constraint (2) and an average-power constraint (3), where the ratio $\alpha = \frac{\mathcal{E}}{A}$ lies in $[\frac{1}{3}, 1]$, is bounded as follows:

$$\begin{aligned} \mathcal{C}(A, \mathcal{E}) \geq & \frac{1}{2} \log A + \left(\frac{A}{3} + 1 \right) \log \left(1 + \frac{3}{A} \right) - 1 \\ & - \sqrt{\frac{\lambda_0 + \frac{1}{12}}{A}} \left(\frac{\pi}{4} + \frac{1}{2} \log 2 \right) - \frac{1}{2} \log \frac{\pi e}{2} \end{aligned} \quad (18)$$

$$\mathcal{C}(A, \alpha A) \leq \frac{1}{2} \log A - \frac{1}{2} \log \frac{\pi e}{2} + o_A(1). \quad (19)$$

Here the error term $o_A(1)$ tends to zero as the average power and the peak power tend to infinity with their ratio held fixed at α , $\frac{1}{3} \leq \alpha \leq 1$. Hence, the asymptotic expansion for the channel capacity is

$$\begin{aligned} \lim_{A \uparrow \infty} \left\{ \mathcal{C}(A, \alpha A) - \frac{1}{2} \log A \right\} = -\frac{1}{2} \log \frac{\pi e}{2}, \\ \frac{1}{3} \leq \alpha \leq 1. \quad (20) \end{aligned}$$

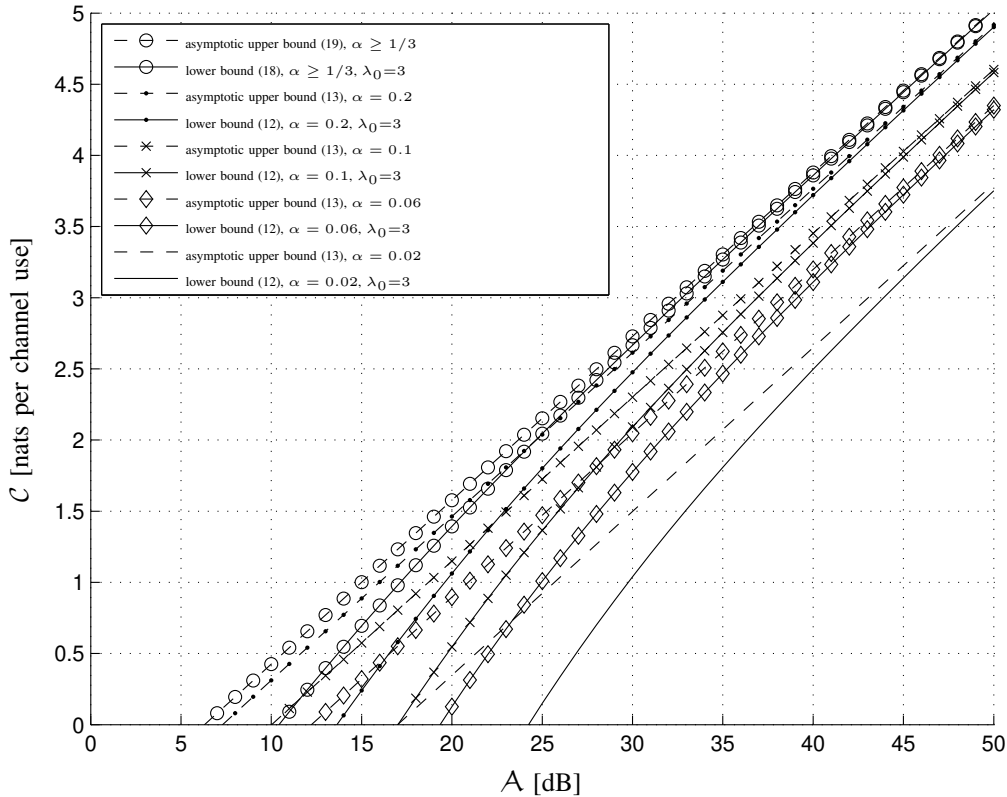


Fig. 1. This plot depicts the firm lower bounds (12) and (18) (valid for all values of A) and the asymptotic upper bounds (13) and (19) (valid only in the limit when $A \uparrow \infty$) on the capacity of a Poisson channel under an average- and a peak-power constraint with average-to-peak-power ratio α . For $\alpha \geq \frac{1}{3}$ (including the case of only a peak-power constraint $\alpha = 1$) the bounds do not depend on α . The upper bounds do not depend on the dark current. For the lower bounds the dark current is assumed to be $\lambda_0 = 3$. The horizontal axis is measured in dB where $A [\text{dB}] = 10 \log_{10} A$.

The bounds of Theorem 3 and 4 are depicted in Figure 1 *is bounded as follows:*
for different values of α .

Remark 5. For $\alpha \uparrow \frac{1}{3}$ the solution μ to (14) tends to zero. If in (13) μ is chosen to be zero, then (13) coincides with (19). On the other hand the lower bound (12) does not converge to (18) for $\mu \downarrow 0$. The reason for this lies in a detail of the derivations shown in Section III-D: in the case of only a peak-power constraint we are able to derive the value of $E \left[\log \left(1 + \frac{a}{X} \right) \right]$ exactly (see (49)), whereas in the case of a peak- and average-power constraint we need to bound this value (see (45)).

Remark 6. Note that in Theorem 4 both the lower and the upper bound do not depend on α . Asymptotically, the average-power constraint becomes inactive for $\alpha \in \left[\frac{1}{3}, 1 \right]$ so the transmitter uses less than the available average power.

Finally, for the case with only an average-power constraint the results are as follows.

Theorem 7. The channel capacity $\mathcal{C}(\mathcal{E})$ of a Poisson channel with dark current λ_0 under an average-power constraint (3)

$$\mathcal{C}(\mathcal{E}) \geq \frac{1}{2} \log \mathcal{E} - \sqrt{\frac{\pi \left(\lambda_0 + \frac{1}{12} \right)}{2\mathcal{E}}} + (\mathcal{E} + 1) \log \left(1 + \frac{1}{\mathcal{E}} \right) - 1 \quad (21)$$

$$\mathcal{C}(\mathcal{E}) \leq \frac{1}{2} \log \mathcal{E} + o_{\mathcal{E}}(1). \quad (22)$$

Here the error term $o_{\mathcal{E}}(1)$ tends to zero as $\mathcal{E} \uparrow \infty$. Hence, the asymptotic expansion for the channel capacity is

$$\lim_{\mathcal{E} \uparrow \infty} \left\{ \mathcal{C}(\mathcal{E}) - \frac{1}{2} \log \mathcal{E} \right\} = 0. \quad (23)$$

The bounds of Theorem 7 are shown in Figure 2, together with the lower and upper bound (5) and (6) from [4] and (7) and (8) from [20].

Remark 8. If we keep \mathcal{E} fixed and let $A \uparrow \infty$, we get $\alpha \downarrow 0$. For $\alpha \ll 1$ the solution μ to (14) tends to $\frac{1}{2\alpha} \gg 1$ which makes sure that (13) tends to (22). To see this note that for

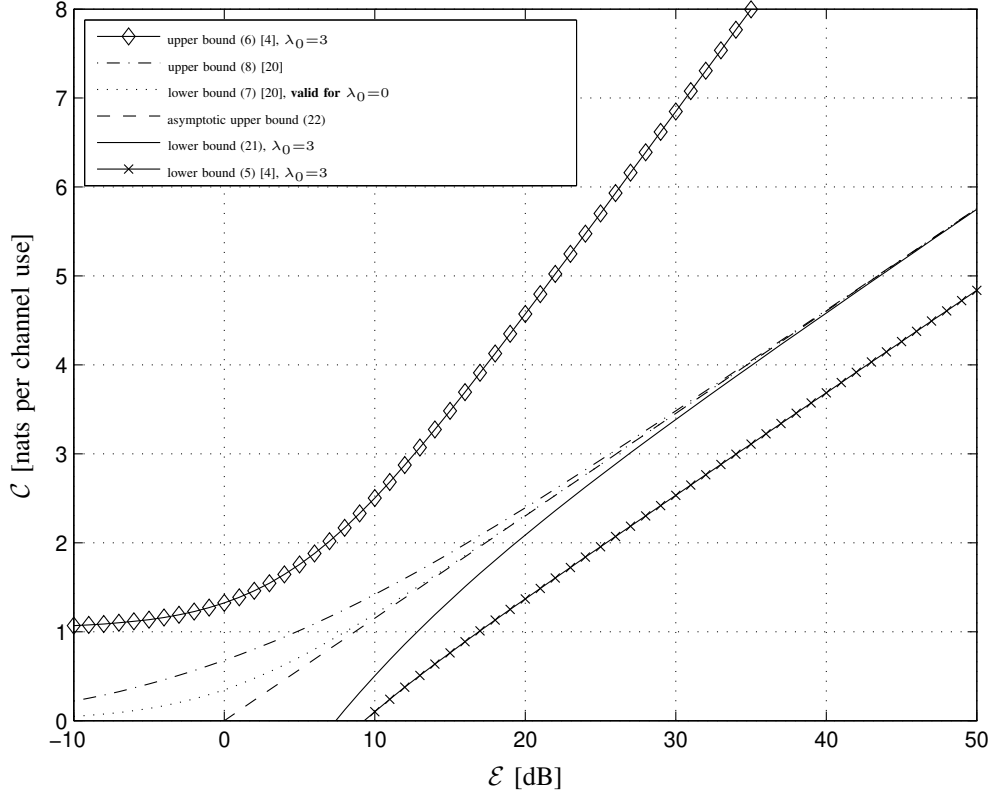


Fig. 2. This plot depicts the firm lower bound (21) (valid for all values of \mathcal{E}) and the asymptotic upper bound (22) (valid only in the limit when $\mathcal{E} \uparrow \infty$) on the capacity of a Poisson channel with average-power constraint $\mathbb{E}[X] \leq \mathcal{E}$. The lower bound assumes a dark current $\lambda_0 = 3$. Additionally asymptotic versions of the lower and upper bound (5) and (6) by Brady & Verdú [4] are plotted where we have assumed $\mathcal{E} \uparrow \infty$, $\epsilon = 0$, and $\lambda_0 = 3$, and the firm lower and upper bounds (7) and (8) by Martinez [20] are shown. Note that the lower bound (7) assumes $\lambda_0 = 0$ and is therefore not directly comparable with (21). The horizontal axis is measured in dB where $\mathcal{E} [\text{dB}] = 10 \log_{10} \mathcal{E}$.

$\mu \gg 1$ we can approximate $\text{erf}(\sqrt{\mu}) \approx 1$. Then we get from (14) that

$$\frac{1}{2} - \alpha\mu \approx \sqrt{\frac{\mu}{\pi}} e^{-\mu}. \quad (24)$$

Using this together with

$$\frac{1}{2} \log \mathcal{A} = \frac{1}{2} \log \mathcal{E} - \frac{1}{2} \log \alpha \quad (25)$$

$$\approx \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log 2\mu \quad (26)$$

we get from (13)

$$\begin{aligned} & \frac{1}{2} \log \mathcal{A} - \mu + \underbrace{\alpha\mu}_{\approx \frac{1}{2}} - \log \left(\frac{1}{2} - \alpha\mu \right) - \frac{1}{2} \log 2\pi e \\ & \approx \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log 2\mu - \mu + \frac{1}{2} - \log \sqrt{\frac{\mu}{\pi}} e^{-\mu} \\ & \quad - \frac{1}{2} \log 2\pi e \\ & = \frac{1}{2} \log \mathcal{E}. \end{aligned} \quad (27)$$

Similarly, (12) converges to (21) which can be seen by noting that for \mathcal{E} fixed and $\mathcal{A} \uparrow \infty$ such that $\alpha \downarrow 0$ we get

$$\begin{aligned} & e^\mu \left(\frac{1}{2} - \alpha\mu \right) \left[\log \left(1 + \frac{\lambda_0 + \frac{1}{12}}{\mathcal{A}} \right) \right. \\ & \quad \left. + 2\sqrt{\frac{\lambda_0 + \frac{1}{12}}{\mathcal{A}}} \arctan \left(\sqrt{\frac{\mathcal{A}}{\lambda_0 + \frac{1}{12}}} \right) \right] \\ & \rightarrow \sqrt{\frac{\pi(\lambda_0 + \frac{1}{12})}{2\mathcal{E}}}, \quad \alpha \downarrow 0. \end{aligned} \quad (29)$$

Hence Theorem 7 can be seen as corollary to Theorem 3.

III. DERIVATION OF THE LOWER BOUNDS

A. Overview

The key ideas of the derivation of the lower bounds are as follows. We drop the optimization in the definition of capacity and simply choose one particular $Q(\cdot)$:

$$\mathcal{C} = \sup_{Q(\cdot)} I(X; Y) \geq I(X; Y) \Big|_{\text{for a specific } Q(\cdot)}. \quad (30)$$

(28) This leads to a natural lower bound on capacity.

We would like to choose a distribution $Q(\cdot)$ that is reasonably close to the capacity-achieving input distribution in order to get a tight lower bound. However, we might have the difficulty that for such a $Q(\cdot)$ the evaluation of $I(X; Y)$ is intractable. Note that even for relatively “simple” distributions $Q(\cdot)$ the distribution of the corresponding channel output Y may be difficult to compute, let alone $H(Y)$.

To avoid this problem we lower-bound $H(Y)$ in terms of $h(X)$ and upper-bound $H(Y|X)$ in terms of $\mathbb{E}[\log X]$. This will lead to a lower bound on \mathcal{C} that only depends on $Q(\cdot)$ through the expression

$$h(X) - \frac{1}{2}\mathbb{E}[\log X]. \quad (31)$$

We then choose the CDF $Q(\cdot)$ to maximize this expression under the given power constraints.

B. Mathematical Preliminaries

The following lemma summarizes some basic properties of a Poisson distribution.

Lemma 9. *Let K be Poisson distributed with mean η , $K \sim \mathcal{Po}(\eta)$, i.e.,*

$$\Pr[K = k] = e^{-\eta} \frac{\eta^k}{k!}, \quad k \in \mathbb{Z}_0^+. \quad (32)$$

Then the following holds:

$$\mathbb{E}[K] = \eta \quad (33)$$

$$\text{Var}(K) = \eta \quad (34)$$

$$\mathbb{E}[(K - \mathbb{E}[K])^3] = \eta \quad (35)$$

and $\Pr[K = k]$ is monotonically nondecreasing for $k \leq \mathbb{E}[K]$ and monotonically nonincreasing for $k \geq \mathbb{E}[K]$.

Proof: See, e.g., [25]. ■

Since no simple analytic expression for the entropy of a Poisson random variable is known, we shall resort to simple bounds. We begin with an upper bound.

Lemma 10. *If $K \sim \mathcal{Po}(\eta)$ is a mean- η Poisson random variable, then its entropy $H(K)$ is upper-bounded by*

$$H(K) \leq \frac{1}{2} \log 2\pi e \left(\eta + \frac{1}{12} \right). \quad (36)$$

Proof: See [22, Theorem 16.3.3]. ■

In Section IV-B2 we will present a lower bound on $H(K)$ that is valid asymptotically when the mean η tends to infinity.

The following proposition is the key in the derivation of the lower bounds on channel capacity. It demonstrates that if Y is conditionally Poisson given a mean $X + \lambda_0 \geq 0$, then the entropy $H(Y)$ can be lower-bounded in terms of the differential entropy $h(X)$.

Proposition 11. *Let Y be the output of a Poisson channel with input $X \geq 0$ and dark current λ_0 according to (1). Assume that X has a finite positive expectation $\mathbb{E}[X] > 0$. Then*

$$H(Y) \geq h(X) + (1 + \mathbb{E}[X]) \log \left(1 + \frac{1}{\mathbb{E}[X]} \right) - 1 \quad (37)$$

$$> h(X). \quad (38)$$

Proof: A proof is given in Appendix A. ■

C. Proof of the Lower Bound (12)

Using Lemma 10 and Proposition 11 we get

$$\begin{aligned} I(X; Y) &\geq h(X) + (\mathbb{E}[X] + 1) \log \left(1 + \frac{1}{\mathbb{E}[X]} \right) - 1 \\ &\quad - \frac{1}{2} \log 2\pi e - \frac{1}{2} \mathbb{E}[\log X] \\ &\quad - \frac{1}{2} \mathbb{E} \left[\log \left(1 + \frac{\lambda_0 + \frac{1}{12}}{X} \right) \right]. \end{aligned} \quad (39)$$

We choose an input distribution $Q(\cdot)$ with the following density:

$$Q'(x) = \frac{\sqrt{\mu}}{\sqrt{A\pi x} \cdot \text{erf}(\sqrt{\mu})} \cdot e^{-\frac{\mu}{x}}, \quad 0 \leq x \leq A \quad (40)$$

where $\text{erf}(\cdot)$ is defined in (15) and where μ is chosen to achieve the average-power constraint:

$$\mathbb{E}[X] = \frac{A}{2\mu} - \frac{Ae^{-\mu}}{\sqrt{\pi\mu} \cdot \text{erf}(\sqrt{\mu})} \stackrel{!}{=} \alpha A \quad (41)$$

i.e., μ is the solution to (14). Note that the choice (40) corresponds to the distribution that maximizes (31) under the constraints (2) and (3) [22, Ch. 12]. We then have

$$h(X) = \frac{1}{2} \log \frac{A}{\mu} + \log \sqrt{\pi} \text{erf}(\sqrt{\mu}) + \alpha\mu + \frac{1}{2} \mathbb{E}[\log X] \quad (42)$$

and for $a \geq 0$

$$\begin{aligned} &\mathbb{E} \left[\log \left(1 + \frac{a}{X} \right) \right] \\ &= \int_0^A \log \left(1 + \frac{a}{x} \right) \cdot \frac{\sqrt{\mu}}{\sqrt{A\pi x} \cdot \text{erf}(\sqrt{\mu})} \cdot \underbrace{e^{-\frac{\mu}{x}}}_{\leq 1} dx \end{aligned} \quad (43)$$

$$\leq \int_0^A \log \left(1 + \frac{a}{x} \right) \cdot \frac{\sqrt{\mu}}{\sqrt{A\pi x} \cdot \text{erf}(\sqrt{\mu})} dx \quad (44)$$

$$= \frac{4\sqrt{\frac{a}{A}}\sqrt{\mu} \arctan \left(\sqrt{\frac{A}{a}} \right) + 2\sqrt{\mu} \log \left(1 + \frac{a}{A} \right)}{\sqrt{\pi} \cdot \text{erf}(\sqrt{\mu})}. \quad (45)$$

The result (12) now follows from (39) with (14), (42), and (45) where a is replaced by $\lambda_0 + \frac{1}{12}$.

D. Proof of the Lower Bounds (18) and (21)

The lower bound (18) follows from (39) with the following choice of an input distribution $Q(\cdot)$:

$$Q'(x) \triangleq \frac{1}{\sqrt{4Ax}}, \quad 0 \leq x \leq A. \quad (46)$$

Note that this choice corresponds to (40) with $\mu \downarrow 0$. It is the distribution that maximizes (31) under the peak-power constraint (2) [22, Ch. 12].

We then get

$$\mathbb{E}[X] = \frac{A}{3} \quad (47)$$

$$h(X) = \frac{1}{2} \log A + \frac{1}{2} \log 4 + \frac{1}{2} \mathbb{E}[\log X] \quad (48)$$

$$\mathbb{E} \left[\log \left(1 + \frac{a}{X} \right) \right] = \sqrt{\frac{a}{4A}} (\pi + 2 \log 2). \quad (49)$$

Plugging this into (39) with a substituted by $\lambda_0 + \frac{1}{12}$ yields the desired result.

As noted in Remark 8, (21) can be seen as limiting case of (12) for $\alpha \downarrow 0$. It could also be derived analogously to (12) with the choice

$$Q'(x) \triangleq \frac{1}{\sqrt{2\pi\mathcal{E}x}} \cdot e^{-\frac{x}{2\mathcal{E}}}, \quad x \geq 0 \quad (50)$$

(which is the limiting PDF of (40) for $\alpha \downarrow 0$).

IV. DERIVATION OF THE UPPER BOUNDS

A. Overview

The derivation of the upper bounds is based on the following key ideas:

- We will assume that the dark current is zero, *i.e.*, $\lambda_0 = 0$. This is no loss in generality because any upper bound to the capacity of a Poisson channel without dark current is also an upper bound to the case with nonzero dark current. This can be seen as follows: conditional on $X = x$ let $Y \sim \mathcal{Po}(x + \lambda_0)$. Then Y can be written as

$$Y = Y_1 + Y_2 \quad (51)$$

where $Y_1 \sim \mathcal{Po}(x)$, $Y_2 \sim \mathcal{Po}(\lambda_0)$, and where $Y_1 \perp\!\!\!\perp Y_2$. Expanding mutual information twice using the chain rule we get

$$I(X; Y_1, Y) = I(X; Y_1) + I(X; Y|Y_1) \quad (52)$$

$$= I(X; Y_1) + I(X; Y_1 + Y_2|Y_1) \quad (53)$$

$$= I(X; Y_1) + I(X; Y_2|Y_1) \quad (54)$$

$$= I(X; Y_1) \quad (55)$$

and

$$I(X; Y_1, Y) = I(X; Y) + I(X; Y_1|Y) \quad (56)$$

$$\geq I(X; Y) \quad (57)$$

where the inequality follows from the nonnegativity of mutual information. Hence

$$I(X; Y_1) \geq I(X; Y_1 + Y_2) \quad (58)$$

which proves our claim.

Actually, we will show that asymptotically the dark current has no impact on the capacity.

- One difficulty of the Poisson channel model (1) is that while we have a continuous input, the output is discrete. This complicates the application of the technique explained in Proposition 1 considerably. To circumvent this problem we slightly change the channel model without changing its capacity value. The idea is to add some independent continuous noise U to the channel output Y that is uniformly distributed between 0 and 1, *i.e.*,

$$\tilde{Y} \triangleq Y + U \quad (59)$$

where $U \sim \mathcal{U}([0, 1])$, independent of X and Y . There is no loss in information because, given \tilde{Y} , we can always recover Y by applying the “floor”-operation

$$Y = \lfloor \tilde{Y} \rfloor \quad (60)$$

where for any $a \in \mathbb{R}$, $\lfloor a \rfloor$ denotes the largest integer smaller than or equal to a .

- We will rely on Proposition 1 to derive an upper bound on the capacity of this new channel model with input X and output \tilde{Y} , *i.e.*, we will choose an output distribution $R(\cdot)$ and evaluate (9). In various places we will need to resort to further upper-bounding.
- To evaluate the expectation in (9) over the unknown capacity-achieving input distribution $Q^*(\cdot)$ we will resort to the concept of *input distributions that escape to infinity* as introduced in [21] and further refined in [23]. In short, even if $Q^*(\cdot)$ is unknown, this concept allows us to compute $\mathbb{E}_{Q^*}[f(X)]$ for arbitrary bounded functions $f(\cdot)$ in the asymptotic limit when the available power \mathcal{E} tends to infinity. The price we pay is that our upper bounds are only valid asymptotically for infinite power. For more details, see Section IV-B1.
- As mentioned before no strictly analytic expression for the entropy of a Poisson distributed random variable is known. We will resort to an asymptotic lower bound on $H(Y|X = x)$ that is valid as x tends to infinity. We then again use the concept of *input distributions that escape to infinity* to show that if the available power tends to infinity also x tends to infinity.

B. Mathematical Preliminaries

In Section IV-B1 we will review the concept of *input distributions that escape to infinity* and some of its implications. Note that the stated results are general and not restricted to the case of a Poisson channel. Section IV-B2 shows how the Poisson channel model can be modified to have a continuous output.

1) *Input Distributions that Escape To Infinity*: In this section we will briefly review the notion of *input distributions that escape to infinity* as introduced in [21] and further refined in [23]. Loosely speaking, a sequence of input distributions parametrized by the allowed cost is said to escape to infinity if it assigns to any fixed compact set a probability that tends to zero as the allowed cost tends to infinity.

This notion is important because we can show that for most channels of interest, the capacity-achieving input distribution must escape to infinity. In fact, not only the capacity-achieving input distributions escape to infinity: every sequence of input distributions that achieves a mutual information having the same asymptotic growth rate as capacity must escape to infinity.

The statements in this section are valid in general, *i.e.*, they are not restricted to the Poisson channel. We will only assume that the input and output alphabets \mathcal{X} and \mathcal{Y} of some channel are separable metric spaces, and that for any set $\mathcal{B} \subset \mathcal{Y}$ the mapping $x \mapsto W(\mathcal{B}|x)$ from \mathcal{X} to $[0, 1]$ is Borel measurable.⁴ We then consider a general cost function $g : \mathcal{X} \rightarrow [0, \infty)$ which is assumed measurable.

⁴In the case of the Poisson channel, the channel output alphabet is discrete. However, it will be shown in Section IV-B2 that this channel can be easily modified to have a continuous output without changing its basic properties.

Recall the following definition of a capacity-cost function with an average and a peak constraint.

Definition 12. Given a channel $W(\cdot|\cdot)$ over the input alphabet \mathcal{X} and the output alphabet \mathcal{Y} and given some nonnegative cost function $g : \mathcal{X} \rightarrow \mathbb{R}_0^+$, we define the capacity-cost function $\mathcal{C} : (\inf_{x \in \mathcal{X}} g(x), \infty)^2 \rightarrow \mathbb{R}_0^+$ by

$$\mathcal{C}(\mathcal{A}, \mathcal{E}) \triangleq \sup_{Q(\cdot)} I(X; Y), \quad \mathcal{A}, \mathcal{E} \geq \inf_{x \in \mathcal{X}} g(x) \quad (61)$$

where the supremum is over all input distributions $Q(\cdot)$ that satisfy

$$Q(\{x \in \mathcal{X} : g(x) > \mathcal{A}\}) = 0 \quad (62)$$

and

$$\mathbb{E}_Q[g(X)] \leq \mathcal{E}. \quad (63)$$

Note that all following results also hold in the case of only an average constraint, without limitation on the peak-power. However, for brevity we will omit the explicit statements for this case.

We will now define the notion of input distributions that escape to infinity. For an intuitive understanding of the following definition and some of its consequences, it is best to focus on the example of the Poisson channel where the channel inputs are nonnegative real numbers and where the cost function $g(\cdot)$ is $g(x) = x, \forall x \geq 0$.

Definition 13. Fixing $\alpha \in (0, 1]$ as ratio of available average to peak cost

$$\alpha \triangleq \frac{\mathcal{E}}{\mathcal{A}} \quad (64)$$

we say that a family of input distributions

$$\{Q_{\mathcal{A}, \mathcal{E}}(\cdot)\}_{\mathcal{A} \geq \inf_x \frac{g(x)}{\alpha}, \mathcal{E} = \alpha \mathcal{A}} \quad (65)$$

on \mathcal{X} parametrized by \mathcal{A} and \mathcal{E} escapes to infinity if for any $\mathcal{A}_0 > 0$

$$\lim_{\mathcal{A} \uparrow \infty} Q_{\mathcal{A}, \alpha \mathcal{A}}(\{x \in \mathcal{X} : g(x) \leq \mathcal{A}_0\}) = 0. \quad (66)$$

Based on this definition, in [23] a general theorem was presented demonstrating that if the ratio of mutual information to channel capacity is to approach one, then the input distributions must escape to infinity.

Proposition 14. Let the capacity-cost function $\mathcal{C}(\cdot, \cdot)$ be finite but unbounded. Let $\mathcal{C}_{\text{asy}}(\cdot)$ be a function that captures the asymptotic behavior of the capacity-cost function $\mathcal{C}(\mathcal{A}, \alpha \mathcal{A})$ in the sense that

$$\lim_{\mathcal{A} \uparrow \infty} \frac{\mathcal{C}(\mathcal{A}, \alpha \mathcal{A})}{\mathcal{C}_{\text{asy}}(\mathcal{A})} = 1. \quad (67)$$

Assume that $\mathcal{C}_{\text{asy}}(\cdot)$ satisfies the growth condition

$$\lim_{\mathcal{A} \uparrow \infty} \left\{ \sup_{\mu \in (0, \mu_0]} \frac{\mu \mathcal{C}_{\text{asy}}\left(\frac{\mathcal{A}}{\mu}\right)}{\mathcal{C}_{\text{asy}}(\mathcal{A})} \right\} < 1, \quad \forall 0 < \mu_0 < 1. \quad (68)$$

Let $\{Q_{\mathcal{A}, \alpha \mathcal{A}}(\cdot)\}_{\mathcal{A} \geq 0}$ be a family of input distributions satisfying the cost constraints (62) and (63) such that if $X_{\mathcal{A}} \sim Q_{\mathcal{A}, \alpha \mathcal{A}}(\cdot)$

$$\lim_{\mathcal{A} \uparrow \infty} \frac{I(X_{\mathcal{A}}; Y)}{\mathcal{C}_{\text{asy}}(\mathcal{A})} = 1. \quad (69)$$

Then $\{Q_{\mathcal{A}, \alpha \mathcal{A}}(\cdot)\}_{\mathcal{A} \geq 0}$ escapes to infinity.

Proof: See [23, Sec. VII.C.3]. ■

Note that in [1] it has been shown that the Poisson channel has a unique capacity-achieving input distribution. We will now show that this distribution falls into the setting of Proposition 14, i.e., that it escapes to infinity.

Corollary 15. Fix the average-to-peak-power ratio

$$\alpha \triangleq \frac{\mathcal{E}}{\mathcal{A}}. \quad (70)$$

Then the capacity-achieving input distribution $\{Q_{\mathcal{A}, \alpha \mathcal{A}}^*(\cdot)\}_{\mathcal{A} \geq 0}$ of a Poisson channel (1) with peak- and average-power constraints (2) and (3) escapes to infinity. Similarly, for the situation with only an average-power constraint (3), $\{Q_{\mathcal{E}}^*(\cdot)\}_{\mathcal{E} \geq 0}$ escapes to infinity.

Proof: To prove this statement, we will show that the function

$$\mathcal{C}_{\text{asy}}(\mathcal{A}) = \frac{1}{2} \log \mathcal{A} \quad (71)$$

satisfies both conditions (67) and (68) of Proposition 14. The latter has already been shown in [23, Remark 9] and is therefore omitted. The former condition is more tricky. The difficulty lies in the fact that we need to derive the asymptotic behavior of the capacity at this early stage of the proof, even though precisely this asymptotic behavior is our main result of this paper. Note, however, that for the proof of this corollary it is sufficient to find the first term in the asymptotic expansion of capacity.

Nevertheless, our proof relies heavily on the lower bounds derived in Section III, on Proposition 1, and also on Lemmas 17 to 19 of Section IV-B2. Of course, we made sure that none of the used results relies in turn on this corollary!

The details are deferred to the very end of this paper in Appendix F. ■

Remark 16. If a family of input distributions $\{Q_{\mathcal{A}, \alpha \mathcal{A}}(\cdot)\}_{\mathcal{A} \geq 0}$ escapes to infinity, then for every bounded function $f(\cdot)$ that decays to zero, i.e., that satisfies

$$f(x) = o_x(1) \quad (72)$$

we have

$$\lim_{\mathcal{A} \uparrow \infty} \mathbb{E}_{Q_{\mathcal{A}, \alpha \mathcal{A}}} [f(X)] = 0. \quad (73)$$

2) *A Poisson Channel with Continuous Output:* In the following we define an adapted Poisson channel model which has a continuous output. To this end let Y be the output of a Poisson channel with input x as given in (1). We define a new random variable

$$\tilde{Y} \triangleq Y + U \quad (74)$$

where U is independent of Y and uniformly distributed between 0 and 1, $U \sim \mathcal{U}([0, 1])$. Then $\tilde{Y} \geq 0$ is continuous with the probability density function⁵

$$\tilde{W}(\tilde{y}|x) = W(\lfloor \tilde{y} \rfloor | x) = e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^{\lfloor \tilde{y} \rfloor}}{\lfloor \tilde{y} \rfloor!},$$

$$\tilde{y} \geq 0, x \geq 0, \lambda_0 \geq 0. \quad (75)$$

The Poisson channel with continuous output is equivalent to the Poisson channel as defined in Section I. This is shown in the following lemma.

Lemma 17. *Let the random variables Y , \tilde{Y} , and X be defined as above. Then*

$$a) \quad I(X; \tilde{Y}) = I(X; Y) \quad (76)$$

$$b) \quad h(\tilde{Y}|X = x) = H(Y|X = x) \quad (77)$$

$$c) \quad h(\tilde{Y}) = H(Y). \quad (78)$$

Proof: Define $Y' \triangleq \lfloor \tilde{Y} \rfloor$. The random variables

$$X \circlearrowleft Y \circlearrowleft \tilde{Y} \circlearrowleft Y' \quad (79)$$

form a Markov chain. Hence, from the data processing inequality it follows

$$I(X; Y) \geq I(X; \tilde{Y}) \geq I(X; Y'). \quad (80)$$

However, since $Y' = Y$, Part a) is proven.

Part b) follows from the definition of $h(\cdot)$ and $H(\cdot)$, respectively, and the fact that, for any $\tilde{y} \geq 0$, $\tilde{W}(\tilde{y}|x) = W(\lfloor \tilde{y} \rfloor | x)$:

$$h(\tilde{Y}|X = x) = - \int_0^\infty \tilde{W}(\tilde{y}|x) \log \tilde{W}(\tilde{y}|x) d\tilde{y} \quad (81)$$

$$= - \sum_{y=0}^\infty \int_y^{y+1} W(\lfloor \tilde{y} \rfloor | x) \log W(\lfloor \tilde{y} \rfloor | x) d\tilde{y} \quad (82)$$

$$= - \sum_{y=0}^\infty W(y|x) \log W(y|x) \int_y^{y+1} d\tilde{y} \quad (83)$$

$$= - \sum_{y=0}^\infty W(y|x) \log W(y|x) \quad (84)$$

$$= H(Y|X = x). \quad (85)$$

Part c) now follows from a) and b). \blacksquare

We will next derive some more properties of the ‘‘continuous Poisson’’ distribution (75). Without loss of generality, in the rest of this section we will restrict ourselves to the case of $\lambda_0 = 0$.

The expected logarithm of a Poisson distributed random variable is unbounded since the random variable takes on the value zero with a nonzero probability. However, $\mathbb{E}[\log \tilde{Y}|X = x]$ is well defined. It can be bounded as follows.

⁵Slightly misusing our notation we will write $\tilde{W}(\cdot|\cdot)$ to denote a probability density function rather than a CDF. We believe that it simplifies the reading.

Lemma 18. *Let \tilde{Y} be defined as above with PDF $\tilde{W}(\cdot|\cdot)$ given in (75) and assume that $\lambda_0 = 0$. Fix an arbitrary $0 < \delta < 1$. Then*

$$\mathbb{E}[\log \tilde{Y}|X = x] \leq \log(1+x) - \log\left(1 + \frac{1}{x}\right) + 2\log(1+\delta) + \left(\frac{1}{x} + \frac{1}{12x^2}\right) \frac{1}{\left(1 + \frac{\delta}{2}\right)^2} \left(1 + \frac{2}{\delta} + \log \frac{\delta/2}{1+\delta}\right),$$

$$x \geq \frac{1}{\delta} \quad (86)$$

$$\mathbb{E}[\log \tilde{Y}|X = x] \leq \log(1+x) + \log(1+\delta) + \left(4x + \frac{1}{3}\right) \frac{1}{2x+1},$$

$$x < \frac{1}{\delta} \quad (87)$$

$$\mathbb{E}[\log \tilde{Y}|X = x] \geq \log(1+x) - \log\left(1 + \frac{1}{x}\right) + 2\log(1-\delta) - \frac{1}{\left(\delta + \frac{1}{2x}\right)^2} \left(\frac{1}{x} + \frac{1}{12x^2}\right),$$

$$x \geq \frac{1}{\delta} \quad (88)$$

$$\mathbb{E}[\log \tilde{Y}|X = x] \geq \log(1+x) - 1 - \log(1+x),$$

$$x < \frac{1}{\delta}. \quad (89)$$

From this it follows that

$$\mathbb{E}[\log \tilde{Y}|X = x] = \log(1+x) + o_x(1) \quad (90)$$

where the $o_x(1)$ term is bounded and tends to zero as x tends to infinity.

Proof: A proof is given in Appendix B. \blacksquare

We next derive a lower bound on the entropy of a Poisson random variable of sufficiently large mean.

Lemma 19. *Let \tilde{Y} be defined as above with PDF $\tilde{W}(\cdot|\cdot)$ given in (75) and assume that $\lambda_0 = 0$. Fix an arbitrary $0 < \delta < \frac{1}{2}$. Then*

$$h(\tilde{Y}|X = x) \geq \frac{1}{2} \log 2\pi e(1+x) + \log(1-\delta) + \frac{1}{2} \log \frac{x}{1+x} - \frac{1}{6x} - e^{-x} \frac{x^2}{2} \log x - \frac{1}{2} e^{-x} \log 2\pi - \frac{1}{\left(\delta - \frac{1}{2x}\right)^2} \left(\frac{1}{x} + \frac{1}{12x^2}\right),$$

$$x \geq \frac{1}{\delta} \quad (91)$$

$$h(\tilde{Y}|X = x) \geq 0, \quad x \geq 0. \quad (92)$$

Consequently,

$$\lim_{x \rightarrow \infty} \left\{ h(\tilde{Y}|X = x) - \frac{1}{2} \log x \right\} \geq \frac{1}{2} \log 2\pi e \quad (93)$$

which together with Lemma 10 and Lemma 17b) implies

$$h(\tilde{Y}|X = x) = \frac{1}{2} \log 2\pi e(1+x) + o_x(1) \quad (94)$$

where the $o_x(1)$ term is bounded and tends to zero as x tends to infinity.

Proof: A proof is given in Appendix C.

Finally, we state some other properties of \tilde{Y} .

Lemma 20. *Let \tilde{Y} be defined as above with PDF $\tilde{W}(\cdot|\cdot)$ given in (75) and assume that $\lambda_0 = 0$. Let $x = A$ and let $\delta > 0$ be fixed (in particular, δ is not allowed to depend on A). Then we have the following:*

$$\lim_{A \uparrow \infty} \int_{A(1+\delta)}^{\infty} \tilde{W}(\tilde{y}|A) d\tilde{y} = 0 \quad (95)$$

$$\lim_{A \uparrow \infty} \int_{A(1+\delta)}^{\infty} \tilde{W}(\tilde{y}|A) \log A d\tilde{y} = 0 \quad (96)$$

$$\lim_{A \uparrow \infty} \int_{A(1+\delta)}^{\infty} \tilde{W}(\tilde{y}|A) \log \frac{1}{\tilde{W}(\tilde{y}|A)} d\tilde{y} = 0. \quad (97)$$

Proof: A proof is given in Appendix D. ■

C. Proof of the Upper Bound (13)

The derivation of (13) is based on (9) with the following choice of an output distribution $R(\cdot)$:

$$R'(\tilde{y}) \triangleq \begin{cases} p \cdot \frac{\tilde{y}^{\nu-1} e^{-\frac{\tilde{y}}{\beta}}}{\beta^\nu \gamma\left(\nu, \frac{A(1+\delta)}{\beta}\right)} & \forall 0 \leq \tilde{y} \leq A(1+\delta) \\ (1-p) \cdot e^{-(\tilde{y}-A(1+\delta))} & \forall \tilde{y} > A(1+\delta) \end{cases} \quad (98)$$

where $\nu, \beta, \delta > 0$ are free parameters that will be specified later, where

$$p = p(A) \triangleq \Pr\left[\tilde{Y} \leq A(1+\delta) \mid X = A\right] \quad (99)$$

and where $\gamma(\cdot, \cdot)$ denotes the incomplete gamma function

$$\gamma(\nu, \xi) \triangleq \int_0^\xi e^{-t} t^{\nu-1} dt, \quad \nu > 0. \quad (100)$$

Note that

$$\frac{\tilde{y}^{\nu-1} e^{-\frac{\tilde{y}}{\beta}}}{\beta^\nu \gamma\left(\nu, \frac{A(1+\delta)}{\beta}\right)} \quad (101)$$

is the PDF on $[0, A(1+\delta)]$ that maximizes differential entropy under the constraints that $\mathbf{E}[\tilde{Y}]$ and $\mathbf{E}[\log \tilde{Y}]$ are constant. The choice of an exponential distribution on $(A(1+\delta), \infty)$ is motivated by simplicity. It will turn out that asymptotically this ‘‘tail’’ of our output distribution has no influence on the result.

With this choice we get

$$\begin{aligned} & - \int_0^\infty \tilde{W}(\tilde{y}|x) \log R'(\tilde{y}) d\tilde{y} \\ &= \underbrace{-\log p \cdot \Pr\left[\tilde{Y} \leq A(1+\delta) \mid X = x\right]}_{c_a} \\ & \quad - \underbrace{\int_0^{A(1+\delta)} \tilde{W}(\tilde{y}|x) \log \frac{\tilde{y}^{\nu-1} e^{-\frac{\tilde{y}}{\beta}}}{\beta^\nu \gamma\left(\nu, \frac{A(1+\delta)}{\beta}\right)} d\tilde{y}}_{c_b} \\ & \quad - \underbrace{\log(1-p) \cdot \Pr\left[\tilde{Y} > A(1+\delta) \mid X = x\right]}_{c_c} \\ & \quad + \underbrace{\int_{A(1+\delta)}^\infty \tilde{W}(\tilde{y}|x) (\tilde{y} - A(1+\delta)) d\tilde{y}}_{c_d}. \end{aligned} \quad (102)$$

■ We will now consider each term individually. We start with a simple bound on c_a :

$$c_a = \underbrace{\Pr\left[\tilde{Y} \leq A(1+\delta) \mid X = x\right]}_{\leq 1} \underbrace{\log \frac{1}{p}}_{\geq 0} \leq \log \frac{1}{p}. \quad (103)$$

Next, we bound c_b as follows:

$$\begin{aligned} c_b &= \int_0^{A(1+\delta)} \tilde{W}(\tilde{y}|x) \log \frac{\beta^\nu \gamma\left(\nu, \frac{A(1+\delta)}{\beta}\right)}{\tilde{y}^{\nu-1} e^{-\frac{\tilde{y}}{\beta}}} d\tilde{y} \\ &= \left(\nu \log \beta + \log \gamma\left(\nu, \frac{A(1+\delta)}{\beta}\right) \right) \\ & \quad \cdot \Pr\left[\tilde{Y} \leq A(1+\delta) \mid X = x\right] \end{aligned} \quad (104)$$

$$\begin{aligned} & + (1-\nu) \left(\mathbf{E}[\log \tilde{Y} \mid X = x] \right. \\ & \quad \left. - \underbrace{\int_{A(1+\delta)}^\infty \tilde{W}(\tilde{y}|x) \log \tilde{y} d\tilde{y}}_{\geq 0 \text{ for } A > 1} \right) \\ & + \frac{1}{\beta} \left(\mathbf{E}[\tilde{Y} \mid X = x] - \underbrace{\int_{A(1+\delta)}^\infty \tilde{y} \tilde{W}(\tilde{y}|x) d\tilde{y}}_{\geq 0} \right) \end{aligned} \quad (105)$$

$$\begin{aligned} & \leq \left(\nu \log \beta + \log \gamma\left(\nu, \frac{A(1+\delta)}{\beta}\right) \right) \\ & \quad \cdot \Pr\left[\tilde{Y} \leq A(1+\delta) \mid X = x\right] \\ & \quad + (1-\nu) \mathbf{E}[\log \tilde{Y} \mid X = x] + \frac{1}{\beta} \left(x + \frac{1}{2} \right) \end{aligned} \quad (106)$$

where for the inequality (106) we have assumed that $\nu < 1$ and $A > 1$, and where we have used that $\mathbf{E}[\tilde{Y} \mid X = x] = x + \frac{1}{2}$.

For c_c we use the monotonicity of the Poisson distribution (Lemma 9) and the peak-power constraint $\Pr[X > A] = 0$ to get

$$c_c = \Pr\left[\tilde{Y} > A(1+\delta) \mid X = x\right] \log \frac{1}{1-p} \quad (107)$$

$$\leq \Pr\left[\tilde{Y} > A(1+\delta) \mid X = A\right] \log \frac{1}{1-p} \quad (108)$$

$$= (1-p) \log \frac{1}{1-p} \quad (109)$$

where the last equality follows from (99).

Finally, we bound c_d as follows:

$$c_d = \int_{A(1+\delta)}^\infty \tilde{y} \tilde{W}(\tilde{y}|x) d\tilde{y} - A(1+\delta) \int_{A(1+\delta)}^\infty \tilde{W}(\tilde{y}|x) d\tilde{y} \quad (110)$$

$$\leq \int_{\lfloor A(1+\delta) \rfloor}^\infty \tilde{y} \tilde{W}(\tilde{y}|x) d\tilde{y} - A(1+\delta) \int_{A(1+\delta)}^\infty \tilde{W}(\tilde{y}|x) d\tilde{y} \quad (111)$$

$$\begin{aligned} &= \sum_{y=\lfloor A(1+\delta) \rfloor}^\infty \left(y + \frac{1}{2} \right) e^{-x} \frac{1}{y!} x^y \\ & \quad - \underbrace{A(1+\delta) \Pr\left[\tilde{Y} > A(1+\delta) \mid X = x\right]}_{\geq 0} \end{aligned} \quad (112)$$

$$\leq \sum_{y=\lfloor \mathbf{A}(1+\delta) \rfloor}^{\infty} y e^{-x} \frac{1}{y!} x^y + \frac{1}{2} \Pr[Y > \lfloor \mathbf{A}(1+\delta) \rfloor \mid X = x] \quad (113)$$

$$= x \sum_{y=\lfloor \mathbf{A}(1+\delta) \rfloor}^{\infty} e^{-x} \frac{1}{(y-1)!} x^{y-1} + \frac{1}{2} \Pr[Y > \lfloor \mathbf{A}(1+\delta) \rfloor \mid X = x] \quad (114)$$

$$= x \sum_{y=\lfloor \mathbf{A}(1+\delta) \rfloor - 1}^{\infty} e^{-x} \frac{1}{y!} x^y + \frac{1}{2} \Pr[Y > \lfloor \mathbf{A}(1+\delta) \rfloor \mid X = x] \quad (115)$$

$$= x \int_{\lfloor \mathbf{A}(1+\delta) \rfloor - 1}^{\infty} \tilde{W}(\tilde{y}|x) d\tilde{y} + \frac{1}{2} \Pr[\tilde{Y} > \lfloor \mathbf{A}(1+\delta) \rfloor \mid X = x] \quad (116)$$

$$= x \Pr[\tilde{Y} > \lfloor \mathbf{A}(1+\delta) \rfloor - 1 \mid X = x] + \frac{1}{2} \Pr[\tilde{Y} > \lfloor \mathbf{A}(1+\delta) \rfloor \mid X = x] \quad (117)$$

$$\leq \mathbf{A} \Pr[\tilde{Y} > \mathbf{A}(1+\delta_2) \mid X = x] + \frac{1}{2} \Pr[\tilde{Y} > \mathbf{A}(1+\delta_2) \mid X = x] \quad (118)$$

$$\leq \left(\mathbf{A} + \frac{1}{2}\right) \Pr[\tilde{Y} > \mathbf{A}(1+\delta_2) \mid X = \mathbf{A}] \quad (119)$$

$$\leq \left(\mathbf{A} + \frac{1}{2}\right) \mathbf{E}\left[e^{r\tilde{Y}} \mid X = \mathbf{A}\right] e^{-r\mathbf{A}(1+\delta_2)}. \quad (120)$$

Here in (118) we have chosen an arbitrary δ_2 , $0 < \delta_2 < \delta$, assuming that \mathbf{A} is large enough such that

$$\mathbf{A}(1+\delta_2) < \lfloor \mathbf{A}(1+\delta) \rfloor - 1 < \lfloor \mathbf{A}(1+\delta) \rfloor. \quad (121)$$

Eq. (119) follows again from monotonicity of the Poisson distribution; and the final inequality (120) follows from Chernov's bound [26]

$$\Pr[V \geq a] \leq \mathbf{E}\left[e^{rV}\right] e^{-ra}, \quad \forall r > 0, \forall a \geq \mathbf{E}[V]. \quad (122)$$

Next, we upper-bound the moment-generating function of \tilde{Y}

$$\mathbf{E}\left[e^{r\tilde{Y}} \mid X = \mathbf{A}\right] = \int_0^{\infty} e^{r\tilde{y}} \tilde{W}(\tilde{y}|\mathbf{A}) d\tilde{y} \quad (123)$$

$$\leq \sum_{y=0}^{\infty} e^{r(y+1)} e^{-\mathbf{A}} \frac{1}{y!} \mathbf{A}^y \quad (124)$$

$$= e^{r-\mathbf{A}+\mathbf{A}e^r} \quad (125)$$

and choose $r = \log(1+\delta_2)$. This yields:

$$c_d \leq \left(\mathbf{A} + \frac{1}{2}\right) (1+\delta_2) e^{-\mathbf{A}((1+\delta_2)\log(1+\delta_2)-\delta_2)}. \quad (126)$$

Note that $(1+\delta_2)\log(1+\delta_2) - \delta_2 > 0$ for $\delta_2 > 0$, i.e.,

$$\left(\mathbf{A} + \frac{1}{2}\right) (1+\delta_2) e^{-\mathbf{A}((1+\delta_2)\log(1+\delta_2)-\delta_2)} = o_{\mathbf{A}}(1). \quad (127)$$

Plugging all these bounds together with (102) into (9) yields:

$$\begin{aligned} \mathcal{C} \leq \mathbf{E}_{Q^*} \left[-h(\tilde{Y}|X=x) + \log \frac{1}{p} \right. \\ \left. + \left(\nu \log \beta + \log \gamma \left(\nu, \frac{\mathbf{A}(1+\delta)}{\beta} \right) \right) \right. \\ \left. \cdot \Pr[\tilde{Y} \leq \mathbf{A}(1+\delta) \mid X=x] \right. \\ \left. + (1-\nu) \mathbf{E}[\log \tilde{Y} \mid X=x] + \frac{1}{\beta} \left(X + \frac{1}{2} \right) \right. \\ \left. + (1-p) \log \frac{1}{1-p} \right. \\ \left. + \left(\mathbf{A} + \frac{1}{2} \right) (1+\delta_2) e^{-\mathbf{A}((1+\delta_2)\log(1+\delta_2)-\delta_2)} \right]. \quad (128) \end{aligned}$$

Next, we introduce

$$\tilde{p}(\mathbf{A}) \triangleq \mathbf{E}_{Q^*} \left[\Pr[\tilde{Y} \leq \mathbf{A}(1+\delta) \mid X=x] \right] \quad (129)$$

and we choose

$$\beta \triangleq \frac{\mathbf{A}(1+\delta)}{\mu} \quad (130)$$

where μ is the solution to (14). Note that such a solution always exists, is unique, and is nonnegative as long as $0 < \alpha \leq \frac{1}{3}$.

Then, using (90) from Lemma 18 and (94) from Lemma 19 we get

$$\begin{aligned} \mathcal{C} \leq -\frac{1}{2} \log 2\pi e - \frac{1}{2} \mathbf{E}_{Q^*}[\log(1+X)] + \log \frac{1}{p(\mathbf{A})} \\ + \underbrace{\left(\nu \log \frac{\mathbf{A}(1+\delta)}{\mu} + \log \gamma(\nu, \mu) \right)}_{\geq 0 \text{ if } \mathbf{A} \text{ is large enough}} \underbrace{\tilde{p}(\mathbf{A})}_{\leq 1} \\ + (1-\nu) \mathbf{E}_{Q^*}[\log(1+X)] + \frac{\mu}{\mathbf{A}(1+\delta)} \left(\mathcal{E} + \frac{1}{2} \right) \\ + (1-p(\mathbf{A})) \log \frac{1}{1-p(\mathbf{A})} \\ + \underbrace{\left(\mathbf{A} + \frac{1}{2} \right) (1+\delta_2) e^{-\mathbf{A}((1+\delta_2)\log(1+\delta_2)-\delta_2)}}_{=o_{\mathbf{A}}(1)} \\ + \mathbf{E}_{Q^*}[o_X(1)] \quad (131) \\ \leq -\frac{1}{2} \log 2\pi e + \log \frac{1}{p(\mathbf{A})} + \nu \log \mathbf{A} + \nu \log \frac{1+\delta}{\mu} \\ + \log \gamma(\nu, \mu) + \left(\frac{1}{2} - \nu \right) \mathbf{E}_{Q^*}[\log(1+X)] \\ + \frac{\mu \mathcal{E}}{\mathbf{A}(1+\delta)} + \frac{\mu}{2\mathbf{A}(1+\delta)} + (1-p(\mathbf{A})) \log \frac{1}{1-p(\mathbf{A})} \\ + o_{\mathbf{A}}(1) + \mathbf{E}_{Q^*}[o_X(1)] \quad (132) \\ = \frac{1}{2} \log \mathbf{A} - \frac{1}{2} \log 2\pi e + \frac{1}{2} \log \frac{1+\delta}{\mu} + \log \sqrt{\pi} \operatorname{erf}(\sqrt{\mu}) \\ + \frac{\alpha \mu}{1+\delta} + \log \frac{1}{p(\mathbf{A})} + \frac{\mu}{2\mathbf{A}(1+\delta)} \\ + (1-p(\mathbf{A})) \log \frac{1}{1-p(\mathbf{A})} + o_{\mathbf{A}}(1) + \mathbf{E}_{Q^*}[o_X(1)]. \quad (133) \end{aligned}$$

Here, in (132) we upper-bound $\tilde{p}(\mathbf{A}) \leq 1$ assuming that \mathbf{A} is large enough so that the terms in the brackets are larger than zero. In (133) we choose $\nu \triangleq \frac{1}{2}$, use the relation

$$\gamma\left(\frac{1}{2}, \mu\right) = \sqrt{\pi} \operatorname{erf}(\sqrt{\mu}) \quad (134)$$

and recall that $\frac{\varepsilon}{\Lambda} = \alpha$.

Finally, we recall from Lemma 20 that

$$\lim_{\Lambda \uparrow \infty} p(\mathbf{A}) = 1 \quad (135)$$

and therefore

$$\lim_{\Lambda \uparrow \infty} (1 - p(\mathbf{A})) \log \frac{1}{1 - p(\mathbf{A})} = 0. \quad (136)$$

Together with (73) and (14) this yields

$$\begin{aligned} \mathcal{C} \leq & \frac{1}{2} \log \Lambda - \frac{1}{2} \log 2\pi e + \frac{1}{2} \log(1 + \delta) - \mu \\ & - \log\left(\frac{1}{2} - \alpha\mu\right) + \frac{\alpha\mu}{1 + \delta} + o_{\Lambda}(1). \end{aligned} \quad (137)$$

Since δ is arbitrary, this concludes our proof.

D. Proof of the Upper Bounds (19) and (22)

The derivation of the asymptotic upper bound (22) could be done according to the scheme described in Section IV-C with a different choice of an output distribution⁶ $R(\cdot)$:

$$R'(\tilde{y}) \triangleq \frac{\tilde{y}^{\nu-1} e^{-\frac{\tilde{y}}{\beta}}}{\beta^{\nu} \Gamma(\nu)}, \quad \tilde{y} \geq 0 \quad (138)$$

where $\nu, \beta > 0$. However, because (22) can be seen as limiting case of (13) for $\alpha \downarrow 0$ as explained in Remark 8 we omit the details of the proof.

The bound (19) could also be derived very similarly. However, there is an alternative derivation that is less general than the derivation shown in Section IV-C and implicitly demonstrates the power of the duality approach (9). We will derive (19) using this alternative approach. Details can be found in Appendix E.

V. CONCLUSIONS

New (firm) lower bounds and new (asymptotic) upper bounds on the capacity of the discrete-time Poisson channel subject to a peak-power constraint and an average-power constraint were derived. The gap between the lower bounds and the upper bounds tends to zero asymptotically as the peak-power and average-power tend to infinity with their ratio held fixed. The bounds thus yield the asymptotic expansion of channel capacity in this regime.

The derivation of the lower bounds relies on a new result that relates the differential entropy of a Poisson channel's input to the entropy of its output (see Proposition 11).

The asymptotic upper bounds were derived in two ways: in a less elegant version we lower-bound the conditional entropy $h(Y|X)$ in such a way that we get an expression that depends

solely on the distribution of the channel output. Then we upper-bound this expression by choosing the maximizing distribution. In a more powerful approach we rely on a technique that has been introduced in [21]: we upper-bound capacity using duality-based upper bounds on mutual information (see Proposition 1).

In both version we additionally need to rely on another concept introduced in [21] and [23]: the notion of *input distributions that escape to infinity* (see Section IV-B1) that allows us to compute asymptotic expectations over the unknown capacity-achieving input distribution.

APPENDIX A

A PROOF OF PROPOSITION 11

Given $X = x$, Y can be written as $Y = Y_1 + Y_2$, where $Y_1 \sim \mathcal{Po}(x)$ and $Y_2 \sim \mathcal{Po}(\lambda_0)$, $Y_1 \perp\!\!\!\perp Y_2$. But

$$H(Y) = H(Y_1 + Y_2) \quad (139)$$

$$\geq H(Y_1 + Y_2 | Y_2) \quad (140)$$

$$= H(Y_1 | Y_2) \quad (141)$$

$$= H(Y_1) \quad (142)$$

and we can restrict ourselves to the case where $\lambda_0 = 0$.

The proof is based on the data processing inequality of the relative entropy [27, Ch. 1, Lemma 3.11(ii)].

Let $Q(\cdot)$ denote an arbitrary CDF on \mathbb{R}_0^+ with a certain finite mean $\mathbb{E}_Q[X] = \eta > 0$. Let $Q_E(\cdot)$ denote the mean- η exponential CDF on \mathbb{R}_0^+ . Let $R(\cdot)$ be the PMF of Y when Y is conditionally Poisson given X and $X \sim Q(\cdot)$, and let $R_G(\cdot)$ be the PMF of Y when Y is conditionally Poisson given X and $X \sim Q_E(\cdot)$. It is straightforward to show that $R_G(\cdot)$ is a mean- η geometric PMF on \mathbb{Z}_0^+ .

By the data processing theorem we obtain:

$$D(Q(\cdot) \parallel Q_E(\cdot)) \geq D(R(\cdot) \parallel R_G(\cdot)) \quad (143)$$

where $D(\cdot \parallel \cdot)$ denotes relative entropy. The first inequality in the proposition's statement now follows by evaluating the left-hand side of (143)

$$D(Q(\cdot) \parallel Q_E(\cdot)) = \int_0^{\infty} Q'(x) \log \frac{Q'(x)}{\frac{1}{\eta} e^{-\frac{x}{\eta}}} dx \quad (144)$$

$$= -h(X) + \log \eta + 1 \quad (145)$$

and evaluating the right-hand side of (143)

$$\begin{aligned} D(R(\cdot) \parallel R_G(\cdot)) &= \sum_{y=0}^{\infty} R(y) \log \frac{R(y)}{\frac{1}{1+\eta} \left(\frac{\eta}{1+\eta}\right)^y} \\ &= -H(Y) + (1 + \eta) \log(1 + \eta) - \eta \log \eta. \end{aligned} \quad (146)$$

The second inequality in the proposition's statement follows by noting that $(1 + \eta) \log(1 + \eta^{-1}) - 1$ is monotonically decreasing in η and approaches zero, as $\eta \uparrow \infty$.

⁶The PDF (138) maximizes the entropy $h(\tilde{Y})$ under the constraints that $\mathbb{E}[\tilde{Y}]$ and $\mathbb{E}[\log \tilde{Y}]$ are constant.

APPENDIX B
A PROOF OF LEMMA 18

Everything in the following derivation is conditional on $X = x$. Recall that here we assume $\lambda_0 = 0$. We start with the proof of (89):

$$\begin{aligned} \mathbf{E}[\log \tilde{Y} | X = x] &= \int_0^\infty \tilde{W}(\tilde{y}|x) \log \tilde{y} \, d\tilde{y} \quad (148) \end{aligned}$$

$$= \int_0^1 \underbrace{\tilde{W}(\tilde{y}|x)}_{\leq 1} \underbrace{\log \tilde{y}}_{\leq 0} \, d\tilde{y} + \underbrace{\int_1^\infty \tilde{W}(\tilde{y}|x) \log \tilde{y} \, d\tilde{y}}_{\geq 0} \quad (149)$$

$$\geq \int_0^1 \log \tilde{y} \, d\tilde{y} = -1. \quad (150)$$

For the derivation of (86)–(88) we let

$$Z_x \triangleq \frac{\tilde{Y}}{x} \quad (151)$$

be a nonnegative continuous random variable with density

$$f_x(z) \triangleq \frac{d}{dz} \Pr[Z_x \leq z] = -\frac{d}{dz} \Pr[Z_x \geq z] = x \tilde{W}(zx|x). \quad (152)$$

Then from the definition of \tilde{Y} in (74) and from Lemma 9 we have

$$\mathbf{E}[Z_x] = \frac{\mathbf{E}[\tilde{Y}]}{x} = \frac{x + \frac{1}{2}}{x} = 1 + \frac{1}{2x} \quad (153)$$

$$\text{Var}(Z_x) = \frac{\text{Var}(\tilde{Y})}{x^2} = \frac{1}{x^2} \left(x + \frac{1}{12} \right) = \frac{1}{x} + \frac{1}{12x^2}. \quad (154)$$

Moreover, using the fact that the probability distribution of a Poisson random variable is monotonically increasing for all values below its mean (see Lemma 9), we get for $z \leq 1$

$$\Pr[Z_x \leq z] = \Pr[\tilde{Y} \leq zx] \quad (155)$$

$$= \int_0^{zx} \tilde{W}(\tilde{y}|x) \, d\tilde{y} \quad (156)$$

$$= \int_0^{zx} W(\lfloor \tilde{y} \rfloor | x) \, d\tilde{y} \quad (157)$$

$$\leq \int_0^{zx} W(\lfloor zx \rfloor | x) \, d\tilde{y} \quad (158)$$

$$= zx W(\lfloor zx \rfloor | x) \quad (159)$$

$$= zx \tilde{W}(zx|x) \quad (160)$$

$$= z f_x(z), \quad z \leq 1. \quad (161)$$

Here, from Lemma 9 the inequality (158) holds as long as $zx \leq x$, i.e., $z \leq 1$. The last equality follows from (152).

We now have

$$\begin{aligned} \mathbf{E}[\log \tilde{Y} | X = x] &= \log x + \mathbf{E}[\log Z_x] \quad (162) \end{aligned}$$

$$\begin{aligned} &= \log x + \int_0^{1-\delta} f_x(z) \log z \, dz + \int_{1-\delta}^{1+\delta} f_x(z) \log z \, dz \\ &\quad + \int_{1+\delta}^\infty f_x(z) \log z \, dz \quad (163) \end{aligned}$$

where $0 < \delta < 1$ is arbitrary. We will now find upper and lower bounds to each of the three integrals separately.

$$0 \leq - \int_0^{1-\delta} f_x(z) \log z \, dz \quad (164)$$

$$= - \lim_{a \downarrow 0} \int_a^{1-\delta} f_x(z) \log z \, dz \quad (165)$$

$$\begin{aligned} &= \lim_{a \downarrow 0} \left\{ \underbrace{-\log(1-\delta) \Pr[Z_x \leq 1-\delta]}_{\geq 0} + \underbrace{\Pr[Z_x \leq a] \log a}_{\leq 0} \right. \\ &\quad \left. + \int_a^{1-\delta} \Pr[Z_x \leq z] \frac{1}{z} \, dz \right\} \quad (166) \end{aligned}$$

$$\leq -\log(1-\delta) + \int_0^{1-\delta} \Pr[Z_x \leq z] \frac{1}{z} \, dz \quad (167)$$

$$\leq -\log(1-\delta) + \int_0^{1-\delta} f_x(z) \, dz \quad (168)$$

$$= -\log(1-\delta) + \Pr[Z_x \leq 1-\delta] \quad (169)$$

$$= -\log(1-\delta) \quad (170)$$

$$+ \Pr \left[Z_x - \mathbf{E}[Z_x] \leq 1-\delta - 1 - \frac{1}{2x} \right]$$

$$= -\log(1-\delta) \quad (171)$$

$$+ \Pr \left[(Z_x - \mathbf{E}[Z_x])^2 \geq \left(-\delta - \frac{1}{2x} \right)^2 \right]$$

$$\leq -\log(1-\delta) + \frac{1}{\left(\delta + \frac{1}{2x} \right)^2} \left(\frac{1}{x} + \frac{1}{12x^2} \right). \quad (172)$$

Here (166) follows from integration by parts; (168) from (161); and (172) follows from Chebyshev's inequality [26]

$$\Pr \left[|V - \mathbf{E}[V]|^2 > \epsilon^2 \right] \leq \frac{\text{Var}(V)}{\epsilon^2}. \quad (173)$$

For the second integral we only use the monotonicity of $\log z$:

$$\int_{1-\delta}^{1+\delta} f_x(z) \log z \, dz \leq \log(1+\delta) \underbrace{\int_{1-\delta}^{1+\delta} f_x(z) \, dz}_{\leq 1} \quad (174)$$

$$\leq \log(1+\delta) \quad (175)$$

$$\int_{1-\delta}^{1+\delta} f_x(z) \log z \, dz \geq \underbrace{\log(1-\delta)}_{\leq 0} \underbrace{\int_{1-\delta}^{1+\delta} f_x(z) \, dz}_{\leq 1} \quad (176)$$

$$\geq \log(1-\delta). \quad (177)$$

For the last integral term we use integration by parts, similarly to the first integral:

$$0 \leq \int_{1+\delta}^\infty f_x(z) \log z \, dz \quad (178)$$

$$= \lim_{a \uparrow \infty} \int_{1+\delta}^a f_x(z) \log z \, dz \quad (179)$$

$$\begin{aligned} &= \lim_{a \uparrow \infty} \left\{ \underbrace{-\Pr[Z_x \geq a] \log a}_{\leq 0} + \underbrace{\Pr[Z_x \geq 1+\delta] \log(1+\delta)}_{\leq 1} \right. \\ &\quad \left. + \int_{1+\delta}^a \Pr[Z_x \geq z] \frac{1}{z} \, dz \right\} \quad (180) \end{aligned}$$

$$\leq \log(1 + \delta) + \int_{1+\delta}^{\infty} \Pr[Z_x \geq z] \frac{1}{z} dz. \quad (181)$$

Now we distinguish between two cases. In the first case we assume that $x \geq \frac{1}{\delta}$. Then

$$1 + \delta \geq 1 + \frac{1}{2x} \quad (182)$$

and we can use Chebyshev's inequality (173):

$$\begin{aligned} & \int_{1+\delta}^{\infty} \Pr[Z_x \geq z] \frac{1}{z} dz \\ &= \int_{1+\delta}^{\infty} \Pr\left[(Z_x - \mathbf{E}[Z_x])^2 \geq \left(z - 1 - \frac{1}{2x}\right)^2\right] \frac{1}{z} dz \quad (183) \end{aligned}$$

$$\leq \int_{1+\delta}^{\infty} \frac{1}{\left(z - 1 - \frac{1}{2x}\right)^2} \left(\frac{1}{x} + \frac{1}{12x^2}\right) dz \quad (184)$$

$$\leq \left(\frac{1}{x} + \frac{1}{12x^2}\right) \int_{1+\delta}^{\infty} \frac{1}{\left(z - 1 - \frac{\delta}{2}\right)^2} dz \quad (185)$$

$$= \left(\frac{1}{x} + \frac{1}{12x^2}\right) \frac{1}{\left(1 + \frac{\delta}{2}\right)^2} \left(1 + \frac{2}{\delta} + \log \frac{\delta/2}{1 + \delta}\right) \quad (186)$$

where in (185) we use once more that $x \geq \frac{1}{\delta}$.

For $x < \frac{1}{\delta}$ we need to make one additional step:

$$\begin{aligned} & \int_{1+\delta}^{\infty} \Pr[Z_x \geq z] \frac{1}{z} dz \\ &= \int_{1+\delta}^{1+\frac{1}{x}} \Pr[Z_x \geq z] \frac{1}{z} dz + \int_{1+\frac{1}{x}}^{\infty} \Pr[Z_x \geq z] \frac{1}{z} dz \quad (187) \end{aligned}$$

$$\begin{aligned} &= \int_{1+\delta}^{1+\frac{1}{x}} \Pr[Z_x \geq z] \frac{1}{z} dz \\ &+ \int_{1+\frac{1}{x}}^{\infty} \Pr\left[(Z_x - \mathbf{E}[Z_x])^2 \geq \left(z - 1 - \frac{1}{2x}\right)^2\right] \frac{1}{z} dz \quad (188) \end{aligned}$$

$$\begin{aligned} &\leq \int_{1+\delta}^{1+\frac{1}{x}} \frac{1}{z} dz + \left(\frac{1}{x} + \frac{1}{12x^2}\right) \frac{1}{\left(1 + \frac{1}{2x}\right)^2} \\ &\quad \cdot \left(1 + 2x + \log \frac{\frac{1}{2x}}{1 + \frac{1}{x}}\right) \quad (189) \end{aligned}$$

$$\begin{aligned} &= \log\left(1 + \frac{1}{x}\right) - \log(1 + \delta) \\ &+ \left(\frac{1}{x} + \frac{1}{12x^2}\right) \frac{4x^2}{(2x+1)^2} \underbrace{\left(1 + 2x + \log \frac{1}{2x+2}\right)}_{\leq 0} \quad (190) \end{aligned}$$

$$\leq \log\left(1 + \frac{1}{x}\right) - \log(1 + \delta) + \left(4x + \frac{1}{3}\right) \frac{1}{2x+1} \quad (191)$$

where (189) follows again from Chebyshev's inequality.

The claimed results now follow by combining the corresponding terms.

APPENDIX C A PROOF OF LEMMA 19

Everything in the following derivation is conditional on $X = x$. Recall that here we assume $\lambda_0 = 0$. The bound

(92) follows from Lemma 17b) and the fact that entropy is nonnegative.

To derive (91) we write

$$\begin{aligned} & h(\tilde{Y}|X = x) \\ &= - \int_0^{\infty} \tilde{W}(\tilde{y}|x) \log \tilde{W}(\tilde{y}|x) d\tilde{y} \quad (192) \end{aligned}$$

$$= - \int_0^{\infty} \tilde{W}(\tilde{y}|x) \log \left(e^{-x} \frac{1}{[\tilde{y}]!} x^{[\tilde{y}]} \right) d\tilde{y} \quad (193)$$

$$\begin{aligned} &= x - \mathbf{E}[\lfloor \tilde{Y} \rfloor | X = x] \log x \\ &+ \int_0^{\infty} \tilde{W}(\tilde{y}|x) \log([\tilde{y}]!) d\tilde{y} \quad (194) \end{aligned}$$

$$\begin{aligned} &= x - \mathbf{E}[Y|X = x] \log x \\ &+ \int_1^{\infty} \tilde{W}(\tilde{y}|x) \log([\tilde{y}]!) d\tilde{y} \quad (195) \end{aligned}$$

$$= x - x \log x + \sum_{y=1}^{\infty} W(y|x) \log(y!) \quad (196)$$

where in (195) we use that $\lfloor \tilde{Y} \rfloor = Y$.

Using Stirling's bound [28], [29]

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} \leq n! \leq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}} \quad (197)$$

and the Taylor expansion of $y \log y$ around x

$$\begin{aligned} & y \log y = x \log x + (1 + \log x)(y - x) + \frac{1}{2x}(y - x)^2 \\ & - \frac{1}{6x^2}(y - x)^3 + \underbrace{\frac{1}{12(x + \theta(y - x))^3}(y - x)^4}_{\geq 0} \quad (198) \end{aligned}$$

$$\begin{aligned} &\geq x \log x + (1 + \log x)(y - x) + \frac{1}{2x}(y - x)^2 \\ & - \frac{1}{6x^2}(y - x)^3 \quad (199) \end{aligned}$$

(where $0 < \theta < 1$), we get

$$\begin{aligned} & h(\tilde{Y}|X = x) \\ &\geq x - x \log x \\ &+ \sum_{y=1}^{\infty} W(y|x) \left(\frac{1}{2} \log 2\pi y + y \log y - y + \underbrace{\frac{1}{12y+1}}_{\geq 0} \right) \quad (200) \end{aligned}$$

$$\begin{aligned} &\geq -x \log x + \frac{1}{2}(1 - e^{-x}) \log 2\pi + \frac{1}{2} \sum_{y=1}^{\infty} W(y|x) \log y \\ &+ \sum_{y=0}^{\infty} W(y|x) y \log y \quad (201) \end{aligned}$$

$$\begin{aligned} &\geq -x \log x + \frac{1}{2}(1 - e^{-x}) \log 2\pi + \frac{1}{2} \sum_{y=2}^{\infty} W(y|x) \log y \\ &+ \sum_{y=0}^{\infty} W(y|x) \left(x \log x + (1 + \log x)(y - x) \right. \\ &\quad \left. + \frac{1}{2x}(y - x)^2 - \frac{1}{6x^2}(y - x)^3 \right) \quad (202) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2} \log 2\pi + \frac{1}{2} - \frac{1}{6x} + \frac{1}{2} \sum_{y=2}^{\infty} W(y|x) \log y \\
&\quad - \frac{1}{2} e^{-x} \log 2\pi. \tag{203}
\end{aligned}$$

Here, (200) follows from the lower bound in (197); in (201) we use the mean of a Poisson distribution, however, noting that the summation starts at $y = 1$ instead of $y = 0$; then in (202) we insert (199); and in the final step (203) we again use Lemma 9.

In order to evaluate the remaining sum in (203) we introduce Z_x as shown in (151)–(161) in Appendix B.

$$\begin{aligned}
&\sum_{y=2}^{\infty} W(y|x) \log y \\
&\geq \int_2^{\infty} \tilde{W}(\tilde{y}|x) \log(\tilde{y} - 1) d\tilde{y} \tag{204}
\end{aligned}$$

$$= \int_1^{\infty} \tilde{W}(\tilde{y} + 1|x) \log \tilde{y} d\tilde{y} \tag{205}$$

$$= \int_{\frac{1}{x}}^{\infty} x \tilde{W}(xz + 1|x) \log(xz) dz \tag{206}$$

$$= \int_{\frac{1}{x}}^{\infty} f_x\left(z + \frac{1}{x}\right) (\log x + \log z) dz \tag{207}$$

$$= \Pr\left[Z_x \geq \frac{2}{x}\right] \log x + \int_{\frac{1}{x}}^{\infty} f_x\left(z + \frac{1}{x}\right) \log z dz \tag{208}$$

$$\begin{aligned}
&= \Pr\left[Z_x \geq \frac{2}{x}\right] \log x + \int_{\frac{1}{x}}^{1-\delta} f_x\left(z + \frac{1}{x}\right) \log z dz \\
&\quad + \int_{1-\delta}^{1+\delta} f_x\left(z + \frac{1}{x}\right) \log z dz \\
&\quad + \int_{1+\delta}^{\infty} f_x\left(z + \frac{1}{x}\right) \log z dz \tag{209}
\end{aligned}$$

where we introduce an arbitrary $0 < \delta < \frac{1}{2}$ and assume that $x \geq \frac{1}{\delta}$.

We will now find bounds for each integral separately, similarly to the derivation in Appendix B. We again start with integration by parts:

$$\begin{aligned}
&-\int_{\frac{1}{x}}^{1-\delta} f_x\left(z + \frac{1}{x}\right) \log z dz \\
&= \underbrace{-\log(1-\delta)}_{\geq 0} \underbrace{\Pr\left[Z_x \leq 1 - \delta + \frac{1}{x}\right]}_{\leq 1} \\
&\quad + \underbrace{\Pr\left[Z_x \leq \frac{2}{x}\right] \log \frac{1}{x}}_{\leq 0 \text{ because } x \geq \frac{1}{\delta} > 2} + \int_{\frac{1}{x}}^{1-\delta} \Pr\left[Z_x \leq z + \frac{1}{x}\right] \frac{1}{z} dz \tag{210}
\end{aligned}$$

$$\leq -\log(1-\delta) + \int_{\frac{1}{x}}^{1-\delta} f_x\left(z + \frac{1}{x}\right) \frac{z + \frac{1}{x}}{z} dz \tag{211}$$

$$\leq -\log(1-\delta) + \frac{\frac{1}{x} + \frac{1}{x}}{\frac{1}{x}} \int_{\frac{1}{x}}^{1-\delta} f_x\left(z + \frac{1}{x}\right) dz \tag{212}$$

$$= -\log(1-\delta) + 2 \Pr\left[Z_x \leq 1 - \delta + \frac{1}{x}\right]$$

$$- 2 \Pr\left[Z_x \leq \frac{2}{x}\right] \tag{213}$$

$$\begin{aligned}
&\leq -\log(1-\delta) \\
&\quad + 2 \Pr\left[Z_x - \mathbf{E}[Z_x] \leq 1 - \delta + \frac{1}{x} - 1 - \frac{1}{2x}\right] \tag{214}
\end{aligned}$$

$$\begin{aligned}
&= -\log(1-\delta) \\
&\quad + 2 \Pr\left[(Z_x - \mathbf{E}[Z_x])^2 \geq \left(\delta - \frac{1}{2x}\right)^2\right] \tag{215}
\end{aligned}$$

$$\leq -\log(1-\delta) + \frac{2}{\left(\delta - \frac{1}{2x}\right)^2} \left(\frac{1}{x} + \frac{1}{12x^2}\right). \tag{216}$$

Here, (211) follows from (161) using our assumption that $x \geq \frac{1}{\delta}$; (212) is due to the monotonicity in z of $\frac{z+1/x}{z}$; in (215) we use that $x \geq \frac{1}{\delta} > \frac{1}{2\delta}$; and the last inequality (216) follows from Chebyshev's inequality (173).

For the second integral we use the monotonicity of $\log z$:

$$\begin{aligned}
&\int_{1-\delta}^{1+\delta} f_x\left(z + \frac{1}{x}\right) \log z dz \\
&\geq \underbrace{\log(1-\delta)}_{\leq 0} \underbrace{\int_{1-\delta}^{1+\delta} f_x\left(z + \frac{1}{x}\right) dz}_{\leq 1} \\
&\geq \log(1-\delta). \tag{218}
\end{aligned}$$

The third integral we simply lower-bound by zero:

$$\int_{1+\delta}^{\infty} f_x\left(z + \frac{1}{x}\right) \log z dz \geq 0. \tag{219}$$

Combined this yields for $x \geq \frac{1}{\delta}$,

$$\begin{aligned}
&h(\tilde{Y}|X = x) \\
&\geq \frac{1}{2} \log 2\pi + \frac{1}{2} - \frac{1}{6x} + \frac{1}{2} \Pr\left[Z_x \geq \frac{2}{x}\right] \log x \\
&\quad + \log(1-\delta) - \frac{1}{\left(\delta - \frac{1}{2x}\right)^2} \left(\frac{1}{x} + \frac{1}{12x^2}\right) \\
&\quad - \frac{1}{2} e^{-x} \log 2\pi \tag{220} \\
&= \frac{1}{2} \log 2\pi e + \log(1-\delta) + \frac{1}{2} \log x - \frac{1}{6x} \\
&\quad - \frac{1}{2} \Pr\left[Z_x < \frac{2}{x}\right] \log x - \frac{1}{2} e^{-x} \log 2\pi \\
&\quad - \frac{1}{\left(\delta - \frac{1}{2x}\right)^2} \left(\frac{1}{x} + \frac{1}{12x^2}\right). \tag{221}
\end{aligned}$$

We then again use (161) under the condition $x \geq \frac{1}{\delta} > 2$ to show that

$$\frac{1}{2} \Pr\left[Z_x < \frac{2}{x}\right] \log x \leq \frac{1}{2} \frac{2}{x} f_x\left(\frac{2}{x}\right) \log x \tag{222}$$

$$= \frac{1}{x} \tilde{W}(2|x) \log x \tag{223}$$

$$= e^{-x} \frac{x^2}{2} \log x. \tag{224}$$

Hence,

$$\begin{aligned} h(\tilde{Y}|X=x) &\geq \frac{1}{2} \log 2\pi e + \log(1-\delta) + \frac{1}{2} \log x - \frac{1}{6x} \\ &\quad - e^{-x} \frac{x^2}{2} \log x - \frac{1}{2} e^{-x} \log 2\pi \\ &\quad - \frac{1}{(\delta - \frac{1}{2x})^2} \left(\frac{1}{x} + \frac{1}{12x^2} \right). \end{aligned} \quad (225)$$

APPENDIX D
A PROOF OF LEMMA 20

Let $\delta_1 \triangleq \frac{\delta}{2} > 0$. From the definition of $W(\cdot|\cdot)$ we have for any integer $y \geq \lfloor A(1+\delta_1) \rfloor$:

$$\begin{aligned} W(y+1|A) &= e^{-A} \cdot \frac{A^{y+1}}{(y+1)!} \end{aligned} \quad (226)$$

$$= e^{-A} \cdot \frac{A^y}{y!} \cdot \frac{A}{y+1} \quad (227)$$

$$\leq e^{-A} \cdot \frac{A^y}{y!} \cdot \frac{A}{\lfloor A(1+\delta_1) \rfloor + 1} \quad (228)$$

$$\leq e^{-A} \cdot \frac{A^y}{y!} \cdot \frac{1}{1+\delta_1} \quad (229)$$

$$= W(y|A) \cdot \frac{1}{1+\delta_1}, \quad y \geq \lfloor A(1+\delta_1) \rfloor \quad (230)$$

where for (229) we use

$$\lfloor A(1+\delta_1) \rfloor + 1 \geq A(1+\delta_1). \quad (231)$$

Therefore, knowing that $W(\cdot|\cdot) \leq 1$, we get

$$W(y|A) \leq W(\lfloor A(1+\delta_1) \rfloor | A) \cdot \left(\frac{1}{1+\delta_1} \right)^{y-\lfloor A(1+\delta_1) \rfloor} \quad (232)$$

$$\leq \left(\frac{1}{1+\delta_1} \right)^{y-\lfloor A(1+\delta_1) \rfloor}, \quad y \geq \lfloor A(1+\delta_1) \rfloor. \quad (233)$$

Hence,

$$\begin{aligned} &\int_{A(1+\delta)}^{\infty} \tilde{W}(\tilde{y}|A) d\tilde{y} \\ &\leq \int_{\lfloor A(1+\delta) \rfloor}^{\infty} \tilde{W}(\tilde{y}|A) d\tilde{y} \end{aligned} \quad (234)$$

$$= \sum_{y=\lfloor A(1+\delta) \rfloor}^{\infty} W(y|A) \quad (235)$$

$$\leq \sum_{y=\lfloor A(1+\delta) \rfloor}^{\infty} \left(\frac{1}{1+\delta_1} \right)^{y-\lfloor A(1+\delta) \rfloor} \quad (236)$$

$$= \left(\frac{1}{1+\delta_1} \right)^{\lfloor A(1+\delta) \rfloor - \lfloor A(1+\delta_1) \rfloor} \sum_{y=0}^{\infty} \left(\frac{1}{1+\delta_1} \right)^y \quad (237)$$

$$= \left(\frac{1}{1+\delta_1} \right)^{\lfloor A(1+\delta) \rfloor - \lfloor A(1+\delta_1) \rfloor} \cdot \frac{1+\delta_1}{\delta_1} \quad (238)$$

$$\leq \left(\frac{2}{2+\delta} \right)^{A\frac{\delta}{2}-1} \cdot \frac{2+\delta}{\delta} \quad (239)$$

where for (239) we use

$$\begin{aligned} &\lfloor A(1+\delta) \rfloor - \lfloor A(1+\delta_1) \rfloor \\ &\geq A(1+\delta) - 1 - A(1+\delta_1) \end{aligned} \quad (240)$$

$$= A\frac{\delta}{2} - 1. \quad (241)$$

Note that the left-hand side of (95) and (96) are trivially lower-bounded by zero. Hence, (95) and (96) follow from (239).

To prove (97) we again assume $y \geq \lfloor A(1+\delta_1) \rfloor$:

$$\begin{aligned} &W(y+1|A) \log \frac{1}{W(y+1|A)} \\ &\leq W(y|A) \frac{1}{1+\delta_1} \log \frac{1}{W(y|A) \frac{1}{1+\delta_1}} \end{aligned} \quad (242)$$

$$= \frac{1}{1+\delta_1} W(y|A) \log \frac{1}{W(y|A)} + \frac{\log(1+\delta_1)}{1+\delta_1} \cdot W(y|A) \quad (243)$$

$$\begin{aligned} &\leq \frac{1}{1+\delta_1} W(y|A) \log \frac{1}{W(y|A)} \\ &\quad + \frac{\log(1+\delta_1)}{1+\delta_1} \cdot W(y|A) \log \frac{1}{W(y|A)} \end{aligned} \quad (244)$$

$$= \frac{1+\log(1+\delta_1)}{1+\delta_1} \cdot W(y|A) \log \frac{1}{W(y|A)}, \quad y \geq \lfloor A(1+\delta_1) \rfloor. \quad (245)$$

Here, (242) can be argued as follows: for $\xi \ll 1$ the function $\xi \mapsto \xi \log \frac{1}{\xi}$ is monotonically increasing. For large A (as a matter of fact $A \geq 2$ is already sufficiently large) $W(\cdot|A)$ is small enough such that $W(\cdot|A) \log \frac{1}{W(\cdot|A)}$ is monotonically increasing. Therefore, we can use (230). Inequality (244) follows because for large A (again $A \geq 2$ is sufficiently large) we have $\log \frac{1}{W(y|A)} > 1$. Hence,

$$\begin{aligned} &W(y|A) \log \frac{1}{W(y|A)} \\ &\leq \underbrace{W(\lfloor A(1+\delta_1) \rfloor | A) \log \frac{1}{W(\lfloor A(1+\delta_1) \rfloor | A)}}_{\leq 1} \\ &\quad \cdot \left(\frac{1+\log(1+\delta_1)}{1+\delta_1} \right)^{y-\lfloor A(1+\delta_1) \rfloor} \end{aligned} \quad (246)$$

$$\leq \left(\frac{1+\log(1+\delta_1)}{1+\delta_1} \right)^{y-\lfloor A(1+\delta_1) \rfloor}, \quad y \geq \lfloor A(1+\delta_1) \rfloor. \quad (247)$$

Therefore (97) can be derived as follows:

$$\begin{aligned} &\int_{A(1+\delta)}^{\infty} \tilde{W}(\tilde{y}|A) \log \frac{1}{\tilde{W}(\tilde{y}|A)} d\tilde{y} \\ &\leq \int_{\lfloor A(1+\delta) \rfloor}^{\infty} \tilde{W}(\tilde{y}|A) \log \frac{1}{\tilde{W}(\tilde{y}|A)} d\tilde{y} \end{aligned} \quad (248)$$

$$= \sum_{y=\lfloor A(1+\delta) \rfloor}^{\infty} W(y|A) \log \frac{1}{W(y|A)} \quad (249)$$

$$\leq \sum_{y=\lfloor A(1+\delta) \rfloor}^{\infty} \left(\frac{1+\log(1+\delta_1)}{1+\delta_1} \right)^{y-\lfloor A(1+\delta_1) \rfloor} \quad (250)$$

$$= \sum_{y=0}^{\infty} \left(\frac{1 + \log(1 + \delta_1)}{1 + \delta_1} \right)^{y + \lfloor A(1+\delta) \rfloor - \lfloor A(1+\delta_1) \rfloor} \quad (251)$$

$$= \frac{1 + \delta_1}{\delta_1 - \log(1 + \delta_1)} \left(\frac{1 + \log(1 + \delta_1)}{1 + \delta_1} \right)^{\lfloor A(1+\delta) \rfloor - \lfloor A(1+\delta_1) \rfloor} \quad (252)$$

$$\leq \frac{1 + \delta_1}{\delta_1 - \log(1 + \delta_1)} \left(\frac{1 + \log(1 + \delta_1)}{1 + \delta_1} \right)^{A \frac{\delta}{2} - 1}. \quad (253)$$

Here for (253) we have again used (241). Note that the left-hand side of (97) is trivially lower-bounded by zero since $\tilde{W}(\cdot|\cdot) \leq 1$. Hence, (97) follows from (253).

APPENDIX E

PROOF OF THE UPPER BOUND (19)

To derive (19) we firstly note that the capacity of a channel with an imposed peak- and average-power constraint is upper-bounded by the capacity of the same channel with a peak-power constraint only. Hence, any upper bound on the capacity for the case $\alpha = 1$ is implicitly an upper bound on the capacity for all $0 < \alpha \leq 1$, *i.e.*, we will derive an upper bound for the case $\alpha = 1$ only.

The derivation of this upper bound could be done according to the scheme in Section IV-C with a choice of an output distribution $R(\cdot)$ with PDF

$$R'(\tilde{y}) \triangleq \begin{cases} p \cdot \frac{\beta \cdot \tilde{y}^{\beta-1}}{(\mathbf{A}(1+\delta))^\beta} & \forall 0 \leq \tilde{y} \leq \mathbf{A}(1+\delta) \\ (1-p) \cdot e^{-(\tilde{y}-\mathbf{A}(1+\delta))} & \forall \tilde{y} > \mathbf{A}(1+\delta) \end{cases} \quad (254)$$

where $\beta > -1$, $\delta > 0$, and

$$p = p(\mathbf{A}) \triangleq \Pr[\tilde{Y} \leq \mathbf{A}(1+\delta) | X = \mathbf{A}]. \quad (255)$$

However, we will show a different approach here that does not rely on the duality-based technique of Proposition 1. This approach is more cumbersome and less general, but clearly illustrates the elegance and power of the duality approach.

The new approach uses the trick to “transfer” the problem of computing the mutual information between input and output of the channel to a problem that depends only on the distribution of the channel output. More specifically, we will lower-bound $h(\tilde{Y}|X = x)$ by an expression containing $\frac{1}{2}\mathbb{E}[\log \tilde{Y}]$ such that the mutual information is upper-bounded by an expression that contains

$$h(\tilde{Y}) - \frac{1}{2}\mathbb{E}[\log \tilde{Y}] \quad (256)$$

and does not directly depend on X . We can then find a valid upper bound by maximizing this expression over all allowed output distributions. Unfortunately, this maximum is unbounded as we do not have a peak-power constraint on the *output*. Hence, we additionally need to “transfer” the peak-power constraint to the output side, *i.e.*, we need to show that the contribution of the terms for $\tilde{y} > \mathbf{A}$ are asymptotically negligible. As a matter of fact, we will only be able to show that the terms for $\tilde{y} > \mathbf{A}(1+\delta)$ are negligible for an arbitrary $\delta > 0$.

Interestingly, the PDF that will achieve the maximum in (256) is (almost) our choice $R'(\cdot)$ of (254). Therefore the derivations of both approaches are very similar in many

aspects. The main difference—and also the reason why this alternative derivation is much less powerful—is that in this alternative derivation we have to *transfer* the problem to the output side (including the peak-power constraint!) and then *prove* that our choice of $R(\cdot)$ is entropy-maximizing. This is in stark contrast to the approach of Section IV-C where we may simply specify $R(\cdot)$ without any justification. In the case of only a peak-power constraint such a justification is possible; in the more complicated scenario of both a peak- and an average-power constraint such a proof may be very difficult.

We will now show the details. Again assume $\lambda_0 = 0$. Using Lemma 17a), Lemma 18, and Lemma 19 we get

$$I(X; Y)$$

$$= h(\tilde{Y}) - \mathbb{E}_Q \left[\frac{1}{2} \log(1 + X) + \frac{1}{2} \log 2\pi e + o_X(1) \right] \quad (257)$$

$$= h(\tilde{Y}) - \mathbb{E}_Q \left[\frac{1}{2} \mathbb{E}[\log \tilde{Y} | X = x] - o_X(1) \right] - \frac{1}{2} \log 2\pi e + \mathbb{E}_Q[o_X(1)] \quad (258)$$

$$= h(\tilde{Y}) - \frac{1}{2} \mathbb{E}[\log \tilde{Y}] - \frac{1}{2} \log 2\pi e + \mathbb{E}_Q[o_X(1)] \quad (259)$$

$$\leq \int_0^{\mathbf{A}(1+\delta)} p_{\tilde{Y}}(\tilde{y}) \log \frac{1}{p_{\tilde{Y}}(\tilde{y})} d\tilde{y} + \int_{\mathbf{A}(1+\delta)}^{\infty} p_{\tilde{Y}}(\tilde{y}) \log \frac{1}{p_{\tilde{Y}}(\tilde{y})} d\tilde{y} - \frac{1}{2} \int_0^{\mathbf{A}(1+\delta)} p_{\tilde{Y}}(\tilde{y}) \log \tilde{y} d\tilde{y} - \frac{1}{2} \log 2\pi e + \mathbb{E}_Q[o_X(1)] \quad (260)$$

where (260) follows because for large \mathbf{A} ,

$$\int_{\mathbf{A}(1+\delta)}^{\infty} p_{\tilde{Y}}(\tilde{y}) \log \tilde{y} d\tilde{y} \geq 0. \quad (261)$$

In order to prove that $\int_{\mathbf{A}(1+\delta)}^{\infty} p_{\tilde{Y}}(\tilde{y}) \log \frac{1}{p_{\tilde{Y}}(\tilde{y})} d\tilde{y}$ is small we note that for $\tilde{y} > \mathbf{A}(1+\delta)$ the output distribution can be bounded as follows:

$$p_{\tilde{Y}}(\tilde{y}) = \int_0^{\mathbf{A}} Q(x) \tilde{W}(\tilde{y}|x) dx \quad (262)$$

$$\leq \max_{0 \leq x \leq \mathbf{A}} \tilde{W}(\tilde{y}|x) \quad (263)$$

$$= \tilde{W}(\tilde{y}|\mathbf{A}), \quad \tilde{y} > \mathbf{A}(1+\delta). \quad (264)$$

Since $\tilde{W}(\tilde{y}|\mathbf{A})$ is small for large \mathbf{A} , we can use the monotonicity of $\xi \log \frac{1}{\xi}$ for small ξ to bound

$$\int_{\mathbf{A}(1+\delta)}^{\infty} p_{\tilde{Y}}(\tilde{y}) \log \frac{1}{p_{\tilde{Y}}(\tilde{y})} d\tilde{y} \leq \int_{\mathbf{A}(1+\delta)}^{\infty} \tilde{W}(\tilde{y}|\mathbf{A}) \log \frac{1}{\tilde{W}(\tilde{y}|\mathbf{A})} d\tilde{y} \quad (265)$$

$$= o_{\mathbf{A}}(1) \quad (266)$$

where (266) follows from (97).

Let

$$\tilde{p} = \tilde{p}(\mathbf{A}) \triangleq \Pr[\tilde{Y} \leq \mathbf{A}(1+\delta)] \quad (267)$$

where $\tilde{p} \uparrow 1$ as $A \uparrow \infty$. Let further

$$\hat{p}_{\tilde{Y}}(\tilde{y}) \triangleq \begin{cases} \frac{1}{\tilde{p}} p_{\tilde{Y}}(\tilde{y}) & \text{for } 0 \leq \tilde{y} \leq A(1+\delta) \\ 0 & \text{otherwise.} \end{cases} \quad (268)$$

Note that $\hat{p}_{\tilde{Y}}(\cdot)$ is a probability density, *i.e.*, it is nonnegative and integrates to one. Hence,

$$\begin{aligned} I(X; Y) &\leq \int_0^{A(1+\delta)} \tilde{p} \cdot \hat{p}_{\tilde{Y}}(\tilde{y}) \log \frac{1}{\tilde{p} \cdot \hat{p}_{\tilde{Y}}(\tilde{y})} d\tilde{y} - \frac{1}{2} \log 2\pi e \\ &\quad - \frac{1}{2} \int_0^{A(1+\delta)} \tilde{p} \cdot \hat{p}_{\tilde{Y}}(\tilde{y}) \log \tilde{y} d\tilde{y} + \mathbf{E}_Q[o_X(1)] \\ &\quad + o_A(1) \end{aligned} \quad (269)$$

$$\begin{aligned} &= \tilde{p} \left(\int_0^{A(1+\delta)} \hat{p}_{\tilde{Y}}(\tilde{y}) \log \frac{1}{\hat{p}_{\tilde{Y}}(\tilde{y})} d\tilde{y} \right. \\ &\quad \left. - \frac{1}{2} \int_0^{A(1+\delta)} \hat{p}_{\tilde{Y}}(\tilde{y}) \log \tilde{y} d\tilde{y} \right) \\ &\quad + \tilde{p} \log \frac{1}{\tilde{p}} - \frac{1}{2} \log 2\pi e + \mathbf{E}_Q[o_X(1)] + o_A(1) \end{aligned} \quad (270)$$

$$\begin{aligned} &\leq \tilde{p} \cdot \sup_{\hat{p}_{\tilde{Y}} \in \tilde{\mathcal{Y}}} \left\{ \int_0^{A(1+\delta)} \hat{p}_{\tilde{Y}}(\tilde{y}) \log \frac{1}{\hat{p}_{\tilde{Y}}(\tilde{y})} d\tilde{y} \right. \\ &\quad \left. - \frac{1}{2} \int_0^{A(1+\delta)} \hat{p}_{\tilde{Y}}(\tilde{y}) \log \tilde{y} d\tilde{y} \right\} \\ &\quad + \tilde{p} \log \frac{1}{\tilde{p}} - \frac{1}{2} \log 2\pi e + \mathbf{E}_Q[o_X(1)] + o_A(1) \end{aligned} \quad (271)$$

$$\begin{aligned} &= \tilde{p} \cdot \sup_{\eta} \left\{ \sup_{\hat{p}_{\tilde{Y}} \in \tilde{\mathcal{Y}}_{\eta}} h(\tilde{Y}) - \frac{1}{2}\eta \right\} + \tilde{p} \log \frac{1}{\tilde{p}} - \frac{1}{2} \log 2\pi e \\ &\quad + \mathbf{E}_Q[o_X(1)] + o_A(1) \end{aligned} \quad (272)$$

where $\tilde{\mathcal{Y}}$ denotes the set of all distributions over $[0, A(1+\delta)]$, and $\tilde{\mathcal{Y}}_{\eta}$ the set of all distributions over $[0, A(1+\delta)]$ that satisfy the constraint $\mathbf{E}[\log \tilde{Y}] = \eta$.

The supremum over $\tilde{\mathcal{Y}}_{\eta}$ is achieved by the distributions [22, Ch. 11]

$$\hat{p}_{\tilde{Y}}^*(\tilde{y}) = \frac{\beta \cdot \tilde{y}^{\beta-1}}{(A(1+\delta))^{\beta}}, \quad \beta > -1 \quad (273)$$

with

$$\mathbf{E}_{\hat{p}_{\tilde{Y}}^*}[\log \tilde{Y}] = \log(A(1+\delta)) - \frac{1}{\beta} \quad (274)$$

$$h_{\hat{p}_{\tilde{Y}}^*}(\tilde{Y}) = \log(A(1+\delta)) - \log \beta + \frac{\beta-1}{\beta}. \quad (275)$$

Hence,

$$\begin{aligned} I(X; Y) &\leq \tilde{p} \cdot \sup_{\beta} \left\{ \frac{1}{2} \log A + \frac{1}{2} \log(1+\delta) - \log \beta + 1 - \frac{1}{2\beta} \right\} \\ &\quad + \tilde{p} \log \frac{1}{\tilde{p}} - \frac{1}{2} \log 2\pi e + \mathbf{E}_Q[o_X(1)] + o_A(1) \end{aligned} \quad (276)$$

$$\begin{aligned} &= \tilde{p} \left(\frac{1}{2} \log A + \frac{1}{2} \log(1+\delta) - \log \frac{1}{2} + 1 - \frac{1}{2 \cdot \frac{1}{2}} \right) \\ &\quad + \tilde{p} \log \frac{1}{\tilde{p}} - \frac{1}{2} \log 2\pi e + \mathbf{E}_Q[o_X(1)] + o_A(1) \end{aligned} \quad (277)$$

$$\begin{aligned} &\leq \frac{1}{2} \log A + \frac{1}{2} \log(1+\delta) + \log 2 + \tilde{p}(A) \log \frac{1}{\tilde{p}(A)} \\ &\quad - \frac{1}{2} \log 2\pi e + \mathbf{E}_Q[o_X(1)] + o_A(1) \end{aligned} \quad (278)$$

$$\begin{aligned} &= \frac{1}{2} \log A - \frac{1}{2} \log \frac{\pi e}{2} + \frac{1}{2} \log(1+\delta) + \mathbf{E}_Q[o_X(1)] \\ &\quad + o_A(1) \end{aligned} \quad (279)$$

where the supremum in (276) is achieved for $\beta = \frac{1}{2}$, and where in (278) we have bounded $\tilde{p} \leq 1$.

Finally, we use (73) and Corollary 15. The result now follows since δ is arbitrary.

APPENDIX F A PROOF OF COROLLARY 15

To prove the claim of this corollary we rely on Proposition 14, *i.e.*, we need to derive a function $\mathcal{C}_{\text{asy}}(\cdot)$ that satisfies (67) and (68).

From the lower bounds in Theorem 3, 4, and 7 (which are proven in Section III) we know that

$$\lim_{A \uparrow \infty} \frac{\mathcal{C}(A, \alpha A)}{\frac{1}{2} \log A} \geq 1 \quad (280)$$

and

$$\lim_{\mathcal{E} \uparrow \infty} \frac{\mathcal{C}(\mathcal{E})}{\frac{1}{2} \log \mathcal{E}} \geq 1 \quad (281)$$

respectively.

We next derive upper bounds on the channel capacity. Note that

$$\mathcal{C}(A, \alpha A) \leq \mathcal{C}_{\text{PP}}(A) \leq \mathcal{C}_{\text{avg}}(A) \quad (282)$$

where $\mathcal{C}_{\text{PP}}(\cdot)$ and $\mathcal{C}_{\text{avg}}(\cdot)$ denote the capacity under an peak-power and average-power constraint, respectively. Hence, it will be sufficient to show an upper bound for the average-power constraint case only. Moreover, as shown in (58), we can further upper-bound capacity by assuming $\lambda_0 = 0$.

Our derivation is based on Lemma 17a) and on (9) with the choice of an output distribution $R(\cdot)$ on \mathbb{R}_0^+ having the following density:

$$R'(\tilde{y}) \triangleq \frac{e^{-\frac{\tilde{y}}{\mathcal{E}}}}{\sqrt{\pi \tilde{y} \mathcal{E}}}, \quad \tilde{y} \geq 0. \quad (283)$$

For $X = x$ we get

$$\begin{aligned} D(\tilde{W}(\cdot|x) \| R(\cdot)) &= -h(\tilde{Y}|X=x) + \frac{1}{2} \mathbf{E}[\log \tilde{Y}|X=x] \\ &\quad + \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log \pi + \frac{1}{\mathcal{E}} \mathbf{E}[\tilde{Y}|X=x] \end{aligned} \quad (284)$$

$$\begin{aligned} &= -h(\tilde{Y}|X=x) + \frac{1}{2} \mathbf{E}[\log \tilde{Y}|X=x] \\ &\quad + \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log \pi + \frac{x}{\mathcal{E}} + \frac{1}{2\mathcal{E}} \end{aligned} \quad (285)$$

where we use that $\mathbf{E}[\tilde{Y}|X=x] = x + \frac{1}{2}$ (the additional term $\frac{1}{2}$ follows from (74)). We fix an arbitrary $0 < \delta < \frac{1}{2}$ and

continue by a case distinction. For $x \geq \frac{1}{\delta}$ we use the bounds (86) and (91) to get

$$\begin{aligned}
D(\tilde{W}(\cdot|x) \parallel R(\cdot)) &\leq -\frac{1}{2} \log 2\pi e - \log(1-\delta) - \frac{1}{2} \log x + \frac{1}{6x} \\
&\quad + e^{-x} \frac{x^2}{2} \log x + \frac{1}{2} e^{-x} \log 2\pi \\
&\quad + \frac{1}{(\delta - \frac{1}{2x})^2} \left(\frac{1}{x} + \frac{1}{12x^2} \right) + \frac{1}{2} \log x + \log(1+\delta) \\
&\quad + \frac{1}{2} \left(\frac{1}{x} + \frac{1}{12x^2} \right) \frac{1}{(1+\frac{\delta}{2})^2} \left(1 + \frac{2}{\delta} + \log \frac{\delta/2}{1+\delta} \right) \\
&\quad + \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log \pi + \frac{x}{\mathcal{E}} + \frac{1}{2\mathcal{E}} \tag{286} \\
&\leq -\frac{1}{2} - \log(1-\delta) + \frac{\delta}{6} + 1 + \frac{4}{\delta^2} \left(\delta + \frac{\delta^2}{12} \right) \\
&\quad + \frac{1}{2} \left(\delta + \frac{\delta^2}{12} \right) \frac{1}{(1+\frac{\delta}{2})^2} \left(1 + \frac{2}{\delta} + \log \frac{\delta/2}{1+\delta} \right) \\
&\quad + \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log \pi + \frac{x}{\mathcal{E}} + \frac{1}{2\mathcal{E}} + \log(1+\delta) \tag{287} \\
&= \frac{1}{2} \log \mathcal{E} + \frac{x}{\mathcal{E}} + \frac{1}{2\mathcal{E}} + O(1) \tag{288}
\end{aligned}$$

where in (287) we use $e^{-x} \frac{x^2}{2} \log x \leq 1$ and $e^{-x} \leq 1$, and where we use in various places that $x \geq \frac{1}{\delta}$. Here $O(1)$ denotes some finite terms that only depend on δ , but not on \mathcal{E} or x .

For $x < \frac{1}{\delta}$ we use (87) and (92) to get

$$\begin{aligned}
D(\tilde{W}(\cdot|x) \parallel R(\cdot)) &\leq \frac{1}{2} \log(1+x) + \frac{1}{2} \log(1+\delta) + \frac{1}{2} \left(4x + \frac{1}{3} \right) \frac{1}{2x+1} \\
&\quad + \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log \pi + \frac{x}{\mathcal{E}} + \frac{1}{2\mathcal{E}} \tag{289} \\
&\leq \frac{1}{2} \log \left(1 + \frac{1}{\delta} \right) + \frac{1}{2} \log(1+\delta) + \frac{1}{2} \left(\frac{4}{\delta} + \frac{1}{3} \right) \\
&\quad + \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log \pi + \frac{x}{\mathcal{E}} + \frac{1}{2\mathcal{E}} \tag{290} \\
&= \frac{1}{2} \log \mathcal{E} + \frac{x}{\mathcal{E}} + \frac{1}{2\mathcal{E}} + O(1) \tag{291}
\end{aligned}$$

where we upper-bound $\frac{1}{2x+1} \leq 1$ and where we use in various places that $x < \frac{1}{\delta}$. Again $O(1)$ does not depend on \mathcal{E} or x .

Hence, we get

$$\mathcal{C}(\mathcal{E}) \leq \mathbf{E}_{Q^*} \left[D(\tilde{W}(\cdot|X) \parallel R(\cdot)) \right] \tag{292}$$

$$\leq \frac{1}{2} \log \mathcal{E} + \frac{\mathcal{E}}{\mathcal{E}} + \frac{1}{2\mathcal{E}} + O(1) \tag{293}$$

$$= \frac{1}{2} \log \mathcal{E} + o_{\mathcal{E}}(1) + O(1) \tag{294}$$

and therefore

$$\overline{\lim}_{\mathcal{E} \uparrow \infty} \frac{\mathcal{C}(\mathcal{E})}{\frac{1}{2} \log \mathcal{E}} \leq 1. \tag{295}$$

Hence, we have shown that $\mathcal{C}_{\text{asy}}(\zeta) \triangleq \frac{1}{2} \log \zeta$ satisfies the conditions of Proposition 14. This proves our claim.

REFERENCES

- [1] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," in *Proceedings of the IEE*, vol. 137, part I (Communications, Speech and Vision), no. 6, December 1990, pp. 424–430.
- [2] A. Lapidoth, J. H. Shapiro, V. Venkatesan, and L. Wang, "The Poisson channel at low input powers," in *Proceedings Twenty-Fifth IEEE Convention of Electrical & Electronics Engineers in Israel (IEEEI)*, Eilat, Israel, December 3–5, 2008.
- [3] V. Venkatesan, "On low power capacity of the Poisson channel," Master's thesis, Signal and Information Processing Laboratory, ETH Zurich, Switzerland, April 2008, supervised by Prof. Dr. Amos Lapidoth.
- [4] D. Brady and S. Verdú, "The asymptotic capacity of the direct detection photon channel with a bandwidth constraint," in *Proceedings Twenty-Eighth Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, IL, USA, October 3–5, 1990, pp. 691–700.
- [5] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and Its Applications*, vol. 23, pp. 143–147, 1978.
- [6] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 710–715, November 1980.
- [7] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel — part I and II," *IEEE Transactions on Information Theory*, vol. 34, no. 6, pp. 1462–1471, November 1988.
- [8] M. R. Frey, "Capacity of the L_p norm-constrained Poisson channel," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 445–450, March 1992.
- [9] —, "Information capacity of the Poisson channel," *IEEE Transactions on Information Theory*, vol. 37, no. 2, pp. 244–256, March 1991.
- [10] S. Shamai (Shitz) and A. Lapidoth, "Bounds on the capacity of a spectrally constrained Poisson channel," *IEEE Transactions on Information Theory*, vol. 39, no. 1, pp. 19–29, January 1993.
- [11] I. Bar-David and G. Kaplan, "Information rates of photon-limited overlapping pulse position modulation channels," *IEEE Transactions on Information Theory*, vol. 30, no. 3, pp. 455–464, May 1984.
- [12] S. M. Moser, "Duality-based bounds on channel capacity," Ph.D. dissertation, Swiss Federal Institute of Technology (ETH), Zurich, October 2004, Diss. ETH No. 15769. [Online]. Available: <http://moser.cm.nctu.edu.tw/>
- [13] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.
- [14] S. Hranilovic and F. R. Kschischang, "Capacity bounds for power- and band-limited optical intensity channels corrupted by Gaussian noise," *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 784–795, May 2004.
- [15] A. A. Farid and S. Hranilovic, "Upper and lower bounds on the capacity of wireless optical intensity channels," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 24–30, 2007.
- [16] A. Lapidoth, S. M. Moser, and M. A. Wigger, "On the capacity of free-space optical intensity channels," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, July 6–11, 2008, pp. 2419–2423.
- [17] —, "On the capacity of free-space optical intensity channels," June 2008, submitted.
- [18] A. A. Farid and S. Hranilovic, "Design of non-uniform capacity-approaching signaling for optical wireless intensity channels," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, July 6–11, 2008, pp. 2327 – 2331.
- [19] A. Lapidoth and S. M. Moser, "On the capacity of an optical intensity channel with input-dependent noise," 2008, in preparation.
- [20] A. Martinez, "Spectral efficiency of optical direct detection," *Journal of the Optical Society of America B*, vol. 24, no. 4, pp. 739–749, April 2007.
- [21] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, October 2003.
- [22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [23] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 437–453, February 2006.

- [24] B. Rankov and D. Lenz, "Bounds on the capacity of Poisson channels," Master's thesis, Signal and Information Processing Laboratory, ETH Zurich, Switzerland, March 2002, supervised by Prof. Dr. Amos Lapidoth.
- [25] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. John Wiley & Sons, 1994, vol. 1.
- [26] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.
- [27] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [28] W. Feller, *An Introduction to Probability Theory and its Applications*, 3rd ed. New York: John Wiley & Sons, 1957/1968, vol. 1.
- [29] H. Robbins, "A remark on Stirling's formula," *The American Mathematical Monthly*, vol. 62, pp. 26–29, 1955.

Amos Lapidoth (S'89–M'95–SM'00–F'04) received the B.A. degree in mathematics (*summa cum laude*, 1986), the B.Sc. degree in electrical engineering (*summa cum laude*, 1986), and the M.Sc. degree in electrical engineering (1990), all from the Technion–Israel Institute of Technology, Haifa. He received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1995.

During 1995–1999, he was an Assistant and Associate Professor in the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge, and was the KDD Career Development Associate Professor in Communications and Technology. He is now Professor of Information Theory at the Signal and Information Processing Laboratory, ETH Zurich, Switzerland. His research interests are in digital communications and information theory.

Dr. Lapidoth served in the years 2003–2004 as Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.

Stefan M. Moser (S'01–M'05) was born in Switzerland. He received the M.Sc. degree in electrical engineering (with distinction) in 1999, the M.Sc. degree in industrial management (MBA) in 2003, and the Ph.D. degree in the field of information theory in 2004, all from the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland.

During 1999 to 2003, he was a Research and Teaching Assistant, and from 2004 to 2005, he was a Senior Research Assistant with the Signal and Information Processing Laboratory, ETH Zurich. Since August 2005, he has been an Assistant Professor with the Department of Communication Engineering, National Chiao Tung University (NCTU), Hsinchu, Taiwan. His research interests are in information theory and digital communications.

Dr. Moser received the National Chiao Tung University Outstanding Researchers Award in 2007 and 2008, the National Chiao Tung University Excellent Teaching Award and the National Chiao Tung University Outstanding Mentoring Award both in 2007, the Willi Studer Award of ETH in 1999, and the ETH Silver Medal for an excellent Master's thesis in 1999.