Capacity Results of an Optical Intensity Channel with Input-Dependent Gaussian Noise

Intermediate Report of NSC Project

"Capacity Analysis of Multiple-Users Communication Systems using Joint Estimation and Detection"

Date:	31 May 2011
Project-Number:	NSC 99-2628-E-009-002
Project Duration:	1 August 2010 – 31 July 2013
Funded by:	National Science Council, Taiwan
Author:	Stefan M. Moser
Organization:	Information Theory Laboratory Department of Electrical Engineering National Chiao Tung University
Address:	Engineering Building IV, Office 727 1001 Ta Hsueh Rd. Hsinchu 30010, Taiwan
E-mail:	stefan.moser@ieee.org

Abstract

This paper investigates a channel model describing optical communication based on intensity modulation. It is assumed that the main distortion is caused by additive Gaussian noise, however, with a noise variance depending on the current signal strength. Both the high-power and low-power asymptotic capacities under simultaneously both a peak-power and an average-power constraint are derived.

The high-power results are based on a new firm (nonasymptotic) lower bound and a new asymptotic upper bound. The upper bound relies on a dual expression for channel capacity and the notion of *capacity-achieving input distributions that escape to infinity*. The lower bound is based on a new lower bound on the differential entropy of the channel output in terms of the differential entropy of the channel input.

The low-power results make use of a theorem by Prelov and van der Meulen.

Keywords: Channel capacity, direct detection, escaping to infinity, Gaussian noise, high signal-to-noise ratio (SNR), low signal-to-noise ratio (SNR), optical communication.

Contents

1 Introduction

In optical communication, systems often implement some form of *intensity modulation*, where the input signal modulates the optical intensity of the emitted light, i.e., it is proportional to the light intensity and is therefore nonnegative. The receiver usually consists of a photo-detector that measures the optical intensity of the incoming light and produces an output signal which is proportional to the detected intensity, corrupted by noise.

In the *free-space optical intensity channel* [1] [2] it is assumed that the corrupting noise is additive white Gaussian distributed and independent of the signal. This assumption is reasonable if the ambient light is strong or if the receiver suffers from intensive thermal noise. However, particularly at high power, this model neglects a fundamental issue of optical communication: the noise depends on the signal itself due to the random nature of photon emission in the laser diode.

A more accurate (but for analysis also more difficult) model is the *Poisson channel* [1] [2]. There the channel output is modeled as a discrete Poisson random variable with a rate that depends on the current input. This model reflects the physical nature of the transmitted signal consisting of many photons. The noisiness of the received signal is caused by two main effects. Firstly, the exact number of arriving photons at the receiver during a given time interval is implicitly random and is modeled by the mentioned Poisson distribution with a rate proportional to the input signal. Secondly, the signal is impaired by background radiation (called *dark current*) that is modeled by an additional constant rate added to the rate of the Poisson distribution.

Not surprisingly, the behavior of channel capacity of these two channels differ significantly: at high signal-to-noise ratios (SNR), the free-space optical intensity channel has a capacity that grows logarithmically in the power with the multiplicative factor in front of the logarithm—the so-called *pre-log*—being 1 [3]–[8]. The capacity growth of the Poisson channel, on the other hand, is logarithmic with a pre-log of only $\frac{1}{2}$ [9]–[11] [4]. At low SNR, the capacity of the free-space optical intensity channel grows quadratically in the peak-power [8], while the Poisson channel exhibits a linear or stronger growth in the average-power, depending on the exact assumptions about peak power and dark current [12]. Note that for both models the exact capacity is in general not known.[1]

In this paper we will consider a channel model that is in-between the free-space optical intensity channel and the Poisson channel: we keep the less involved assumption of additive white Gaussian noise, but we make the variance of the noise dependent on the current input signal to better reflect the physical properties of optical communication. So basically, we consider an "improved" free-space optical intensity channel. We will analyze the capacity of this improved model and ask the question whether it behaves more like its sibling model (the free-space optical intensity channel) or more like the Poisson channel.

The conditional probability density function (PDF) of this input-dependent Gaussian noise channel is given by

$$W(y|x) = \frac{1}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}}\, e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}}, \quad y \in \mathbb{R},\ x \geq 0. \tag{1}$$

[1]Interestingly, for the more general form of the Poisson channel that uses continuous-time signals and that is not restricted to a fixed pulse-amplitude modulation, the capacity is known exactly [13]–[19].

Alternatively, we can describe the channel model by writing the channel output Y as

$$Y = x + \sqrt{x} Z_1 + Z_0 \tag{2}$$

where $x \geq 0$ denotes the channel input, $Z_0 \sim \mathcal{N}_{\mathbb{R}}\left(0, \sigma^2\right)$ is a zero-mean, variance-σ^2 Gaussian random variable describing the input-independent noise, and $Z_1 \sim \mathcal{N}_{\mathbb{R}}\left(0, \varsigma^2\sigma^2\right)$ is a zero-mean, variance-($\varsigma^2\sigma^2$) Gaussian random variable describing the input-dependent noise. Here Z_0 and Z_1 are assumed to be independent. The parameter $\sigma^2 > 0$ describes the strength of the input-independent noise, while $\varsigma^2 > 0$ is the ratio of the input-dependent noise variance to input-independent noise variance, i.e., it describes the strength of input-dependent noise with respect to the input-independent noise.[2]

We simultaneously consider two types of input constraints: a peak-power constraint is accounted for by the peak-input constraint

$$\Pr[X > \mathsf{A}] = 0 \tag{3}$$

and an average-power constraint by

$$\mathsf{E}[X] \leq \mathsf{E}. \tag{4}$$

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). Moreover, we once more emphasize that for the same reason the input must be nonnegative:

$$x \geq 0. \tag{5}$$

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

$$\alpha \triangleq \frac{\mathsf{E}}{\mathsf{A}}. \tag{6}$$

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.

In this paper we investigate the channel capacity of this channel model. We will present lower bounds on capacity that are based on a new result that proves that the differential entropy of the output of our channel model is always larger than the differential entropy of the channel's input (see Section 4.2 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. 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 [20] using a dual expression of 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 [20, Sec. V] [4, Ch. 2].

[2]Note that for $\varsigma^2 = 0$ the model (1) reduces to the free-space optical intensity channel [1] [2]. However, as we will see, the capacity behavior of the free-space optical intensity channel is fundamentally different from (1), particularly at high SNR. Therefore, to prevent to have to make case distinctions, we restrict ς^2 to be strictly positive.

Proposition 1. *Consider a channel*[3] $W(\cdot|\cdot)$ *with input alphabet* $\mathcal{X} = \mathbb{R}_0^+$ *and output alphabet* $\mathcal{Y} = \mathbb{R}$. *Then for an* arbitrary *distribution* $R(\cdot)$ *over* $\mathcal{Y}$, *the channel capacity is upper-bounded by*

$$\mathsf{C} \leq \mathsf{E}_{Q^*}\left[D\big(W(\cdot|X)\big\|R(\cdot)\big)\right]. \tag{7}$$

Here, $D(\cdot\|\cdot)$ *stands for the relative entropy [21, Ch. 2], and* $Q^*(\cdot)$ *denotes the capacity-achieving input distribution.*

The challenge of using (7) lies in a clever choice of the arbitrary law $R(\cdot)$ that will lead to a good upper bound. Moreover, note that the bound (7) 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 [20] [22]. This concept will be reviewed in Section 5.2.

Finally, we present the asymptotic low-power capacity of the optical intensity channel with input-dependent noise. This result is based on a theorem by Prelov and van der Meulen [23] (see Section 6).

The remainder of this paper is structured as follows. After some brief remarks about our notation, we summarize our main results in Sections 2 and 3: Section 2 contains the bounds on capacity that are tight at high SNR, and Section 3 describes the low-SNR results. The derivations are then given in Section 4 (lower bounds), Section 5 (asymptotic high-power upper bounds), and Section 6 (asymptotic low-power capacity). The first two derivation sections both contain a subsection with mathematical preliminaries. In particular, in Section 4.2 we prove that the differential entropy of the channel output $h(Y)$ is lower-bounded by the differential entropy of its input $h(X)$, and in Section 5.2 we review the concept of *input distributions that escape to infinity.* Finally, in Section 7, we will discuss the results and summarize the main points of the techniques used to derived them.

For random quantities we use uppercase letters and for their realizations lowercase letters. Scalars are typically denoted using Greek letters or lowercase Roman letters. A few exceptions are the following symbols: C stands for capacity, E and A are the average and peak power, respectively, $D(\cdot\|\cdot)$ denotes the relative entropy between two probability measures, and $I(\cdot;\cdot)$ stands for the mutual information. Moreover, the capitals Q, W, and R denote PDFs:

- $Q(\cdot)$ denotes a generic PDF on the channel input;
- for any input letter $x \in \mathcal{X}$, $W(\cdot|x)$ represents a PDF on the channel output when the channel input is x;
- $R(\cdot)$ denotes a generic PDF on the channel output.

The expression $I(Q, W)$ stands for the mutual information between input X and output Y of a channel with transition probability measure $W(\cdot|\cdot)$ when the input has distribution $Q(\cdot)$, i.e., $I(Q, W) \triangleq I(X; Y)$. The starred version $Q^*(\cdot)$ is used to represent a capacity-achieving input distribution.

By $\mathcal{N}_{\mathbb{R}}\big(\mu, \sigma^2\big)$ we denote a real Gaussian distribution with mean μ and variance σ^2. All rates specified in this paper are in nats per channel use, and all logarithms are natural logarithms.

Finally, we give the following definitions.

[3] There are certain measurability assumptions on the channel that we omit for simplicity. See [20, Sec. V] [4, Ch. 2].

Definition 2. *Let $f\colon \mathbb{R}_0^+ \to \mathbb{R}$ 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 $z > z_0$*

$$|f(z)| < \epsilon. \tag{8}$$

Then we write[4]

$$f(z) = o_z(1). \tag{9}$$

Definition 3. *The* $\mathcal{Q}$-function *is defined as*

$$\mathcal{Q}(\xi) \triangleq \int_{\xi}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,\mathrm{d}t, \quad \forall\, \xi \in \mathbb{R}. \tag{10}$$

It describes the partial integration of the zero-mean, unit-variance Gaussian PDF. Note that the $\mathcal{Q}$-function is closely related to the error function $\operatorname{erf}(\cdot)$*:*

$$\operatorname{erf}(\xi) \triangleq \frac{2}{\sqrt{\pi}} \int_0^{\xi} e^{-t^2}\,\mathrm{d}t, \qquad \forall\, \xi \in \mathbb{R} \tag{11}$$

$$= 1 - 2\mathcal{Q}\left(\sqrt{2}\xi\right). \tag{12}$$

2 High-Power 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,[5] 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 we restrict the average-to-peak-power ratio (6) to be $0 < \alpha < \frac{1}{3}$. In the second case we have $\frac{1}{3} \le \alpha \le 1$, 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 4. *The channel capacity $\mathsf{C}(\mathsf{A}, \mathsf{E})$ of a channel with channel law* (1) *and under the input constraints* (3) *and* (4)*, where the ratio $\alpha = \frac{\mathsf{E}}{\mathsf{A}}$ lies in $\left(0, \frac{1}{3}\right)$, is bounded as follows:*

$$\mathsf{C}(\mathsf{A}, \alpha\mathsf{A}) \ge \frac{1}{2}\log\frac{\mathsf{A}}{\sigma^2} - \frac{1}{2}\log 2\pi e\varsigma^2 - (1-\alpha)\mu - \log\left(\frac{1}{2} - \alpha\mu\right) + \beta(\mathsf{A}, \alpha, \mu, \sigma, \varsigma) \tag{13}$$

$$\mathsf{C}(\mathsf{A}, \alpha\mathsf{A}) \le \frac{1}{2}\log\frac{\mathsf{A}}{\sigma^2} - \frac{1}{2}\log 2\pi e\varsigma^2 - (1-\alpha)\mu - \log\left(\frac{1}{2} - \alpha\mu\right) + o_{\mathsf{A}}(1) \tag{14}$$

[4]Note that by the subscript z we want to imply that $o_z(1)$ does not depend on any other nonconstant variable apart from z.

[5]Note, however, that while these bounds are valid for any value of the SNR, they are only useful at medium to high SNR.

where

$$\beta(\mathsf{A},\alpha,\mu,\sigma,\varsigma) \triangleq -e^{\mu}\left(\frac{1}{2}-\alpha\mu\right)\left(\frac{2}{\sqrt{\varsigma^2\mathsf{A}}}\arctan\left(\sqrt{\varsigma^2\mathsf{A}}\right)+\log\left(1+\frac{1}{\varsigma^2\mathsf{A}}\right)\right) + \frac{1}{2}\log\left(1+\frac{2\varsigma^2\sigma^2}{\alpha\mathsf{A}}\right)-\frac{\alpha\mathsf{A}}{\varsigma^2\sigma^2}-1+\frac{\sqrt{\alpha\mathsf{A}(\alpha\mathsf{A}+2\varsigma^2\sigma^2)}}{\varsigma^2\sigma^2} \tag{15}$$

and where $\mu \in \left(0,\frac{1}{2\alpha}\right)$ *is the solution to*

$$\frac{1}{2\mu}-\frac{e^{-\mu}}{\sqrt{\mu}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)}=\alpha. \tag{16}$$

Note that the function $\mu \mapsto \frac{1}{2\mu}-\frac{e^{-\mu}}{\sqrt{\mu}\sqrt{\pi}\mathrm{erf}(\sqrt{\mu})}$ *is monotonically decreasing in* $[0,\infty)$ *and tends to* $\frac{1}{3}$ *for* $\mu \downarrow 0$ *and to* 0 *for* $\mu \uparrow \infty$. *Hence, a solution always exists and is unique. Moreover, from* (16) *it also follows that*

$$\alpha\mu = \frac{1}{2}-\frac{\sqrt{\mu}e^{-\mu}}{\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)} \tag{17}$$

i.e., $\alpha\mu \in \left(0,\frac{1}{2}\right)$ *for all* $\alpha \in \left(0,\frac{1}{3}\right)$. *Therefore* $\log\left(\frac{1}{2}-\alpha\mu\right)$ *is always defined and finite.*

The term $o_{\mathsf{A}}(1)$ *tends to zero as the average power and the peak power tend to infinity with their ratio held fixed at* α, $0<\alpha<\frac{1}{3}$.

The asymptotic expansion of the channel capacity is

$$\lim_{\mathsf{A}\uparrow\infty}\left\{\mathsf{C}(\mathsf{A},\alpha\mathsf{A})-\frac{1}{2}\log\frac{\mathsf{A}}{\sigma^2}\right\} = -\frac{1}{2}\log 2\pi e\varsigma^2-(1-\alpha)\mu-\log\left(\frac{1}{2}-\alpha\mu\right), \quad 0<\alpha<\frac{1}{3} \tag{18}$$

where μ *is defined as above to be the solution to* (16).

The bounds of Theorem 4 are depicted in Figs. 1 and 2 for different values of α and ς^2. The asymptotic expansion (18) is shown in Fig. 3.

In the second case $\alpha \geq \frac{1}{3}$ (that includes $\alpha = 1$ corresponding to the case when we only have a peak-power constraint), we have the following bounds.

Theorem 5. *The channel capacity* $\mathsf{C}(\mathsf{A},\mathsf{E})$ *of a channel with channel law* (1) *and under the input constraints* (3) *and* (4), *where the ratio* $\alpha = \frac{\mathsf{E}}{\mathsf{A}}$ *lies in* $\left[\frac{1}{3},1\right]$, *is bounded as follows:*

$$\mathsf{C}(\mathsf{A},\alpha\mathsf{A}) \geq \frac{1}{2}\log\frac{\mathsf{A}}{\sigma^2}-\frac{1}{2}\log\frac{\pi e\varsigma^2}{2}+\beta\left(\mathsf{A},\frac{1}{3},0,\sigma,\varsigma\right) \tag{19}$$

$$\mathsf{C}(\mathsf{A},\alpha\mathsf{A}) \leq \frac{1}{2}\log\frac{\mathsf{A}}{\sigma^2}-\frac{1}{2}\log\frac{\pi e\varsigma^2}{2}+o_{\mathsf{A}}(1) \tag{20}$$

where β *is defined in* (15), *i.e.,*

$$\beta\left(\mathsf{A},\frac{1}{3},0,\sigma,\varsigma\right) = -\frac{1}{\sqrt{\varsigma^2\mathsf{A}}}\arctan\left(\sqrt{\varsigma^2\mathsf{A}}\right)-\frac{1}{2}\log\left(1+\frac{1}{\varsigma^2\mathsf{A}}\right) + \frac{1}{2}\log\left(1+\frac{6\varsigma^2\sigma^2}{\mathsf{A}}\right)-\frac{\mathsf{A}}{3\varsigma^2\sigma^2}-1+\frac{\sqrt{\mathsf{A}(\mathsf{A}+6\varsigma^2\sigma^2)}}{3\varsigma^2\sigma^2} \tag{21}$$

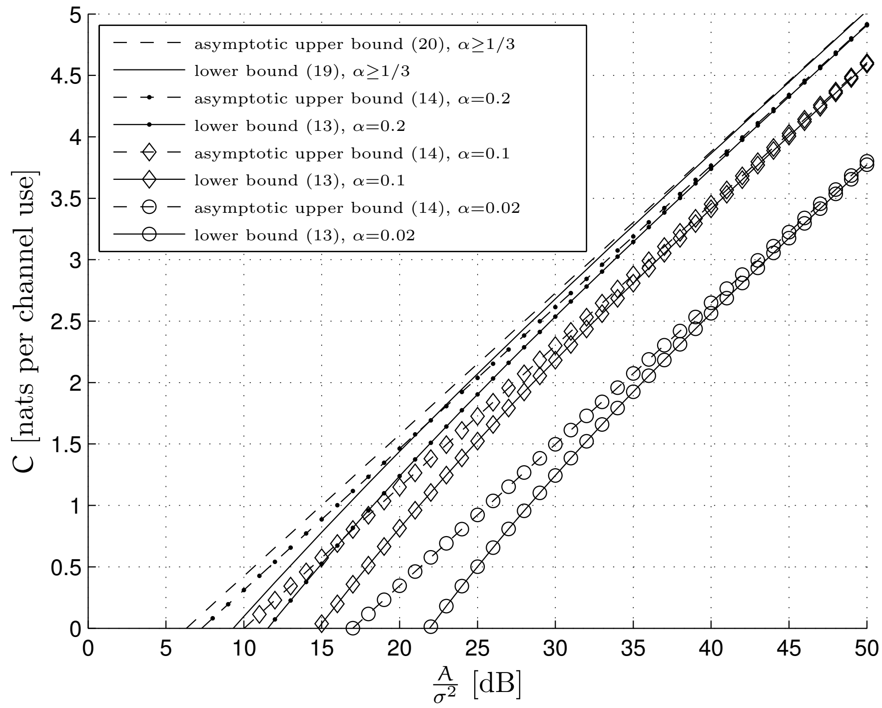

Figure 1: This plot depicts the firm lower bounds (13) and (19) (valid for all values of A) and the asymptotic upper bounds (14) and (20) (valid only in the limit when $\mathsf{A} \uparrow \infty$) on the capacity of the channel model (1) under an average- and a peak-power constraint with average-to-peak-power ratio α. The bounds are depicted for various values of α. The noise variance is set to $\sigma^2 = 1$, and the noise variance ratio is $\varsigma^2 = 1$. 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 horizontal axis is measured in dB where $\frac{\mathsf{A}}{\sigma^2}$ [dB] $= 10 \log_{10} \frac{\mathsf{A}}{\sigma^2}$.

and where the term $o_{\mathsf{A}}(1)$ tends to zero as the average-power and the peak-power tend to infinity with their ratio held fixed at α, $\frac{1}{3} \leq \alpha \leq 1$.

The asymptotic expansion of the channel capacity is

$$\lim_{\mathsf{A}\uparrow\infty} \left\{ \mathsf{C}(\mathsf{A}, \alpha\mathsf{A}) - \frac{1}{2} \log \frac{\mathsf{A}}{\sigma^2} \right\} = -\frac{1}{2} \log \frac{\pi e \varsigma^2}{2}, \qquad \frac{1}{3} \leq \alpha \leq 1. \tag{22}$$

The bounds of Theorem 5 are depicted in Figs. 1 and 2.

Remark 6. For $\alpha \uparrow \frac{1}{3}$ the solution μ to (16) tends to zero. If in (13) and (14) μ is chosen to be zero and α to be $\frac{1}{3}$, then (13) and (14) coincide with (19) and (20), respectively.

Remark 7. Note that in Theorem 5 both the lower and the upper bound do not depend on α, i.e., they are invariant to changes of the average-power constraint. This means that at least asymptotically the average-power constraint becomes inactive for $\alpha \in \left[\frac{1}{3}, 1\right]$. This will be discussed further in Section 7.

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

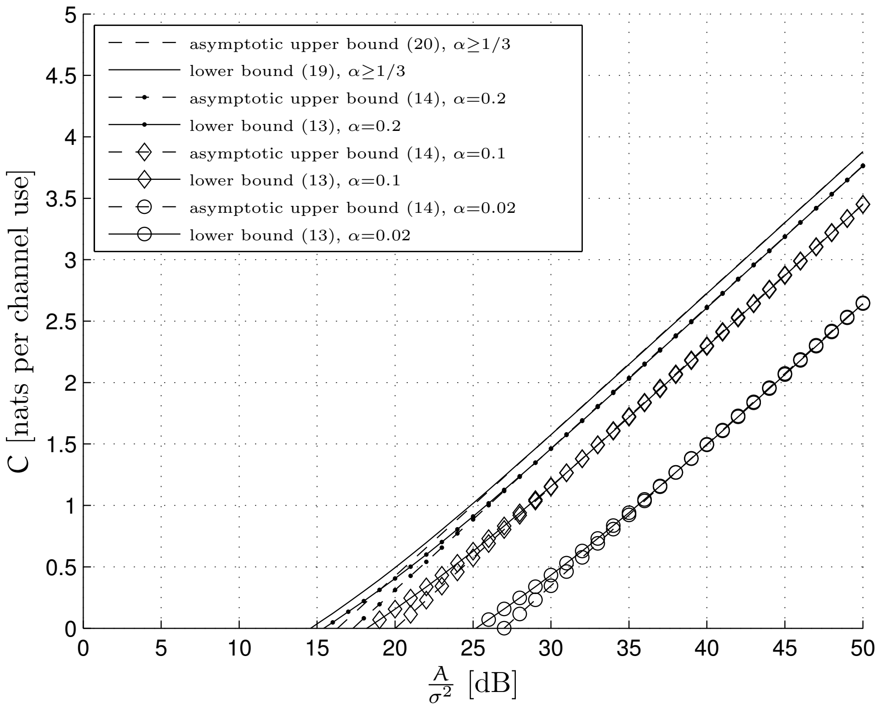

Figure 2: This plot depicts the same bounds as shown in Fig. 1, but with a noise variance ratio $\varsigma^2 = 10$.

Theorem 8. *The channel capacity* $\mathsf{C}(\mathsf{E})$ *of a channel with channel law* (1) *and under the average-power constraint* (4) *is bounded as follows:*

$$\mathsf{C}(\mathsf{E}) \geq \frac{1}{2}\log\frac{\mathsf{E}}{\sigma^2} - \frac{1}{2}\log\varsigma^2 + \frac{1}{2}\log\left(1 + \frac{2\varsigma^2\sigma^2}{\mathsf{E}}\right) - \frac{\mathsf{E}}{\varsigma^2\sigma^2} - 1 + \frac{\sqrt{\mathsf{E}(\mathsf{E}+2\varsigma^2\sigma^2)}}{\varsigma^2\sigma^2} - \sqrt{\frac{\pi}{2\varsigma^2\mathsf{E}}} \tag{23}$$

$$\mathsf{C}(\mathsf{E}) \leq \frac{1}{2}\log\frac{\mathsf{E}}{\sigma^2} - \frac{1}{2}\log\varsigma^2 + o_{\mathsf{E}}(1) \tag{24}$$

where the term $o_{\mathsf{E}}(1)$ *tends to zero as* $\mathsf{E}\uparrow\infty$.

The asymptotic expansion for the channel capacity is

$$\lim_{\mathsf{E}\uparrow\infty}\left\{\mathsf{C}(\mathsf{E}) - \frac{1}{2}\log\frac{\mathsf{E}}{\sigma^2}\right\} = -\frac{1}{2}\log\varsigma^2. \tag{25}$$

The bounds of Theorem 8 are shown in Fig. 4.

Remark 9. If we keep E fixed and let $\mathsf{A}\uparrow\infty$, we get $\alpha\downarrow 0$. For $\alpha \ll 1$ the solution μ to (16) tends to $\frac{1}{2\alpha} \gg 1$ which makes sure that (14) tends to (24). To see this note that for $\mu \gg 1$ we can approximate $\operatorname{erf}\left(\sqrt{\mu}\right) \approx 1$. Then we get from (16) that

$$\frac{1}{2} - \alpha\mu \approx \sqrt{\frac{\mu}{\pi}}e^{-\mu}. \tag{26}$$

Using this together with

$$\frac{1}{2}\log\mathsf{A} = \frac{1}{2}\log\mathsf{E} - \frac{1}{2}\log\alpha \tag{27}$$

$$\approx \frac{1}{2}\log\mathsf{E} + \frac{1}{2}\log 2\mu \tag{28}$$

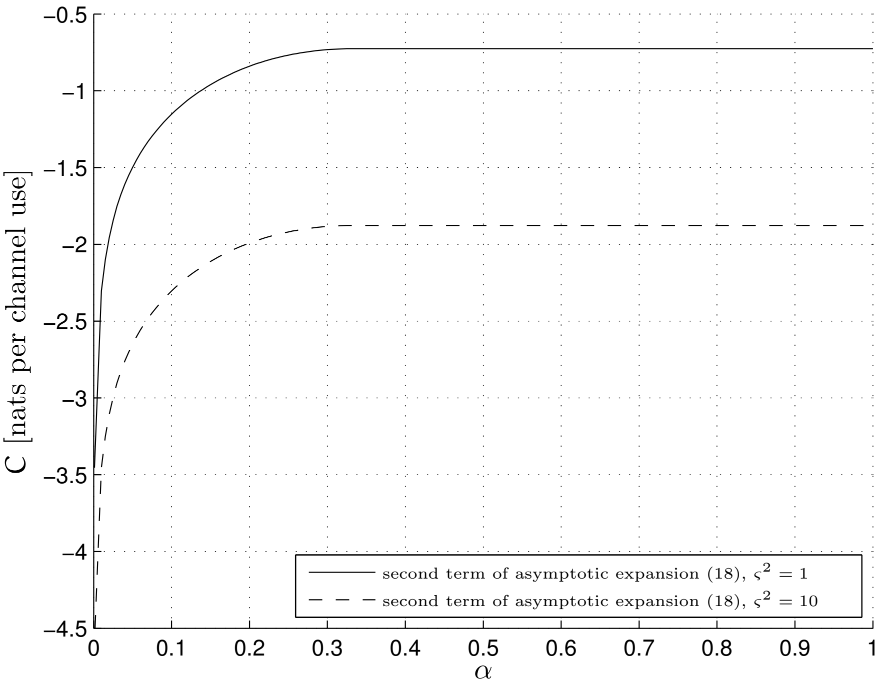

Figure 3: This plot shows the second term of the asymptotic expansion of capacity as given in (18) and (22) as a function of α. For $\alpha \geq \frac{1}{3}$ this expansion does not depend on α anymore. The noise variance ratio is assumed to be $\varsigma^2 = 1$ and $\varsigma^2 = 10$, respectively.

we get from (14)

$$\frac{1}{2}\log\frac{\mathsf{A}}{\sigma^2} - \frac{1}{2}\log 2\pi e\varsigma^2 - \mu + \underbrace{\alpha\mu}_{\approx\frac{1}{2}} - \log\left(\underbrace{\frac{1}{2} - \alpha\mu}_{\approx\sqrt{\frac{\mu}{\pi}}e^{-\mu}}\right)$$

$$\approx \frac{1}{2}\log\frac{\mathsf{E}}{\sigma^2} + \frac{1}{2}\log 2\mu - \frac{1}{2}\log 2\pi e\varsigma^2 - \mu + \frac{1}{2} - \log\left(\sqrt{\frac{\mu}{\pi}}e^{-\mu}\right) \tag{29}$$

$$= \frac{1}{2}\log\frac{\mathsf{E}}{\sigma^2} - \frac{1}{2}\log\varsigma^2. \tag{30}$$

Similarly, (13) converges to (23), which can be seen by additionally noting that

$$e^{\mu}\underbrace{\left(\frac{1}{2} - \alpha\mu\right)}_{\approx\sqrt{\frac{\mu}{\pi}}e^{-\mu}}\left(\frac{2}{\sqrt{\varsigma^2\mathsf{A}}}\underbrace{\arctan\left(\sqrt{\varsigma^2\mathsf{A}}\right)}_{\approx\frac{\pi}{2}} + \underbrace{\log\left(1 + \frac{1}{\varsigma^2\mathsf{A}}\right)}_{\approx\frac{1}{\varsigma^2\mathsf{A}}}\right)$$

$$\approx \sqrt{\frac{\mu}{\pi}}\left(2\sqrt{\frac{\alpha}{\varsigma^2\mathsf{E}}}\cdot\frac{\pi}{2} + \underbrace{\frac{\alpha}{\varsigma^2\mathsf{E}}}_{\ll\sqrt{\frac{\alpha}{\varsigma^2\mathsf{E}}}}\right) \tag{31}$$

$$\approx \sqrt{\frac{\mu\alpha\pi}{\varsigma^2\mathsf{E}}} \approx \sqrt{\frac{\pi}{2\varsigma^2\mathsf{E}}}. \tag{32}$$

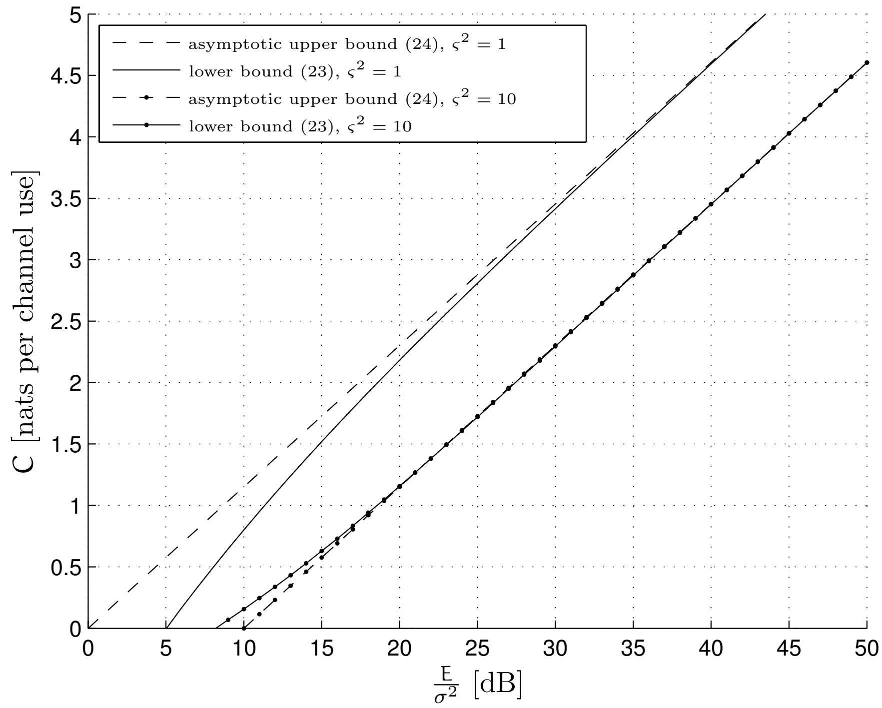

Figure 4: This plot depicts the firm lower bound (23) (valid for all values of E) and the asymptotic upper bound (24) (valid only in the limit when $\mathsf{E} \uparrow \infty$) on the capacity of the channel model (1) under an average-power constraint. The noise variance is fixed to $\sigma^2 = 1$, while the noise variance ratio is $\varsigma^2 = 1$ and $\varsigma^2 = 10$, respectively. The horizontal axis is measured in dB where $\frac{\mathsf{E}}{\sigma^2}$ [dB] $= 10 \log_{10} \frac{\mathsf{E}}{\sigma^2}$.

3 Low-Power Results

For low SNR, we only give the asymptotic behavior of capacity in the limit of a vanishing peak power. We distinguish two cases: the case where we have both a peak- and average-power constraint $0 < \alpha < \frac{1}{2}$, and the case where the average-power constraint is inactive $\frac{1}{2} \leq \alpha \leq 1$.

Theorem 10. *For $\mathsf{A} \downarrow 0$, the asymptotic low-power channel capacity $\mathsf{C}(\mathsf{A}, \alpha\mathsf{A})$ of a channel with channel law* (1) *and under the input constraints* (3) *and* (4)*, where the ratio $\alpha = \frac{\mathsf{E}}{\mathsf{A}}$ lies in $\left(0, \frac{1}{2}\right)$, satisfies*

$$\lim_{\mathsf{A} \downarrow 0} \frac{\mathsf{C}(\mathsf{A}, \alpha\mathsf{A})}{\mathsf{A}^2/\sigma^2} = \frac{\alpha(1-\alpha)}{2} \left(1 + \frac{1}{2}\varsigma^4\sigma^2\right). \tag{33}$$

In the case where the ratio $\alpha = \frac{\mathsf{E}}{\mathsf{A}}$ lies in $\left[\frac{1}{2}, 1\right]$, or if only a peak-power constraint A is imposed (which corresponds to $\alpha = 1$), the asymptotic low-power channel capacity satisfies

$$\lim_{\mathsf{A} \downarrow 0} \frac{\mathsf{C}(\mathsf{A}, \alpha\mathsf{A})}{\mathsf{A}^2/\sigma^2} = \frac{1}{8} \left(1 + \frac{1}{2}\varsigma^4\sigma^2\right). \tag{34}$$

We notice that the threshold between the case with both a peak- and an average-power constraint and the case where the average-power constraint is inactive is at $\alpha = \frac{1}{2}$, and—contrary to the high-power regime—not at $\alpha = \frac{1}{3}$. This will be discussed further in Section 7.

4 Derivation of the High-Power Lower Bounds

4.1 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 $Q(\cdot)$:

$$\mathsf{C} = \sup_{Q(\cdot)} I(Q,W) \geq I(Q,W)\big|_{\text{for a specific } Q(\cdot)}. \tag{35}$$

This leads to a natural lower bound on capacity.

We would like to choose a distribution $Q(\cdot)$ that is reasonably close to a 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 $I(Q,W)$ is intractable. Note that even for relatively "simple" input distributions, 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 $h(X)$, i.e., we "transfer" the problem of computing (or bounding) $h(Y)$ to the input-side of the channel, where it is much easier to choose an appropriate distribution that leads to a tight lower bound.

4.2 Mathematical Preliminaries

The channel model (1) has a useful property relating the differential entropy of the input with the differential entropy of the output: $h(Y)$ can be lower-bounded in terms of $h(X)$. This is shown in the following proposition.

Proposition 11. *Let Y be the output of a channel defined by* (1) *with an input $x \geq 0$. Assume some distribution $Q(\cdot)$ on X having a finite positive mean $\mathsf{E}_Q[X] = \mathsf{E}$. Then*

$$h(Y) \geq h(X) + f_{\text{low}}(\mathsf{E}) > h(X) \tag{36}$$

where $f_{\text{low}}(\cdot)$ is a monotonically decreasing positive function with

$$\lim_{\mathsf{E}\uparrow\infty} f_{\text{low}}(\mathsf{E}) = 0 \tag{37}$$

given by

$$f_{\text{low}}(\mathsf{E}) \triangleq \frac{1}{2}\log\left(1+\frac{2\varsigma^2\sigma^2}{\mathsf{E}}\right) - \frac{\mathsf{E}+\varsigma^2\sigma^2}{\varsigma^2\sigma^2} + \frac{\sqrt{\mathsf{E}(\mathsf{E}+2\varsigma^2\sigma^2)}}{\varsigma^2\sigma^2}, \qquad \mathsf{E} > 0. \tag{38}$$

Proof. See Appendix A. □

4.3 Proof of the Lower Bound (13)

Using (35) and Proposition 11 we get

$$\begin{aligned}
\mathsf{C} &\geq I(Q,W)\big|_{\text{any specified } Q(\cdot)} && (39)\\
&= h(Y) - h(Y|X) && (40)\\
&\geq h(X) + f_{\text{low}}(\mathsf{E}) - h(Y|X) && (41)\\
&= h(X) + f_{\text{low}}(\mathsf{E}) - \frac{1}{2}\mathsf{E}\left[\log 2\pi e\sigma^2(1+\varsigma^2 X)\right] && (42)\\
&= h(X) + f_{\text{low}}(\mathsf{E}) - \frac{1}{2}\log 2\pi e\varsigma^2\sigma^2 - \frac{1}{2}\mathsf{E}[\log X] - \frac{1}{2}\mathsf{E}\left[\log\left(1+\frac{1}{\varsigma^2 X}\right)\right] && (43)
\end{aligned}$$

where in the last equality we have restricted the choice of the input to have zero mass at 0. We choose an input distribution $Q(\cdot)$ that maximizes the entropy $h(X)$ under the given power constraints (3) and (4) and under the additional constraint that $\mathsf{E}[\log X]$ is constant [21, Ch. 12]:

$$Q(x) \triangleq \begin{cases} \frac{\sqrt{\mu}}{\sqrt{\mathsf{A}\pi x}\,\mathrm{erf}(\sqrt{\mu})} \cdot e^{-\frac{\mu}{\mathsf{A}}x} & 0 < x \leq \mathsf{A} \\ 0 & \text{otherwise.} \end{cases} \tag{44}$$

The parameter μ is chosen to satisfy the average-power constraint with equality:

$$\mathsf{E}[X] = \frac{\mathsf{A}}{2\mu} - \frac{\mathsf{A}e^{-\mu}}{\sqrt{\mu}\sqrt{\pi}\mathrm{erf}\left(\sqrt{\mu}\right)} \stackrel{!}{=} \alpha\mathsf{A} \tag{45}$$

i.e., μ is the solution to (16).

Then we have

$$h(X) = \frac{1}{2}\log\frac{\mathsf{A}}{\mu} + \log\sqrt{\pi}\mathrm{erf}\left(\sqrt{\mu}\right) + \alpha\mu + \frac{1}{2}\mathsf{E}[\log X] \tag{46}$$

and

$$\begin{aligned}
\mathsf{E}\left[\log\left(1+\frac{1}{\varsigma^2 X}\right)\right] &= \int_0^{\mathsf{A}} \log\left(1+\frac{1}{\varsigma^2 x}\right)\cdot\frac{\sqrt{\mu}}{\sqrt{\mathsf{A}\pi x}\cdot\mathrm{erf}\left(\sqrt{\mu}\right)}\underbrace{e^{-\frac{\mu}{\mathsf{A}}x}}_{\leq 1}\,\mathrm{d}x && (47)\\
&\leq \int_0^{\mathsf{A}} \log\left(1+\frac{1}{\varsigma^2 x}\right)\cdot\frac{\sqrt{\mu}}{\sqrt{\mathsf{A}\pi x}\cdot\mathrm{erf}\left(\sqrt{\mu}\right)}\,\mathrm{d}x && (48)\\
&= \frac{1}{\sqrt{\pi}\cdot\mathrm{erf}\left(\sqrt{\mu}\right)}\cdot\left(2\pi\frac{\sqrt{\mu}}{\sqrt{\varsigma^2\mathsf{A}}} + 2\sqrt{\mu}\log\left(1+\frac{1}{\varsigma^2\mathsf{A}}\right)\right.\\
&\qquad\left. - 4\frac{\sqrt{\mu}}{\sqrt{\varsigma^2\mathsf{A}}}\arctan\left(\frac{1}{\sqrt{\varsigma^2\mathsf{A}}}\right)\right) && (49)\\
&= \frac{4\frac{\sqrt{\mu}}{\sqrt{\varsigma^2\mathsf{A}}}\arctan\left(\sqrt{\varsigma^2\mathsf{A}}\right) + 2\sqrt{\mu}\log\left(1+\frac{1}{\varsigma^2\mathsf{A}}\right)}{\sqrt{\pi}\cdot\mathrm{erf}\left(\sqrt{\mu}\right)} && (50)
\end{aligned}$$

where we have used that

$$\arctan\left(\frac{1}{\xi}\right) = \frac{\pi}{2} - \arctan(\xi), \quad \xi \in \mathbb{R}. \tag{51}$$

Using (50) and (46) in (43) completes our proof.

4.4 Proof of the Lower Bounds (19) and (23)

As noted in Remarks 6 and 9, (19) and (23) turn out to be the limiting cases of (13) for $\alpha \uparrow \frac{1}{3}$ and $\alpha \downarrow 0$, respectively. This is because we choose the input distributions as the corresponding limiting distributions of (44):

$$Q(x) \triangleq \begin{cases} \frac{1}{\sqrt{4\mathsf{A}x}} & 0 < x \leq \mathsf{A} \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

and

$$Q(x) \triangleq \begin{cases} \frac{1}{\sqrt{2\pi\mathsf{E}x}}e^{-\frac{x}{2\mathsf{E}}} & x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{53}$$

respectively.

For (52) we get

$$\mathsf{E}[X] = \frac{\mathsf{A}}{3} \tag{54}$$
$$h(X) = \log \mathsf{A} - 1 + \log 2 \tag{55}$$
$$\mathsf{E}\left[\log(1+\varsigma^2 X)\right] = \log(1+\varsigma^2\mathsf{A}) - 2 + 2\sqrt{\frac{1}{\varsigma^2\mathsf{A}}}\arctan\left(\sqrt{\varsigma^2\mathsf{A}}\right) \tag{56}$$

which, when plugged into (42), yields (19).

For (53) we get

$$h(X) = \log \mathsf{E} - \frac{\gamma}{2} + \frac{1}{2}\log \pi e \tag{57}$$
$$\mathsf{E}[\log X] = \log \mathsf{E} - \gamma - \log 2 \tag{58}$$
$$\mathsf{E}\left[\log\left(1+\frac{1}{\varsigma^2 X}\right)\right] = \int_0^\infty \frac{1}{\sqrt{2\pi \mathsf{E} x}} \underbrace{e^{-\frac{x}{2\mathsf{E}}}}_{\leq 1} \log\left(1+\frac{1}{\varsigma^2 x}\right) \mathrm{d}x \tag{59}$$
$$\leq \int_0^\infty \frac{1}{\sqrt{2\pi \mathsf{E} x}} \log\left(1+\frac{1}{\varsigma^2 x}\right) \mathrm{d}x \tag{60}$$
$$= \sqrt{\frac{2\pi}{\varsigma^2\mathsf{E}}} \tag{61}$$

which, when plugged into (43), yields (23).

5 Derivation of the High-Power Upper Bounds

5.1 Overview

We rely on Proposition 1 to derive the upper bounds on capacity, i.e.,

$$\mathsf{C} \leq \mathsf{E}_{Q^*}\left[D\big(W(\cdot|X)\big\|R(\cdot)\big)\right]. \tag{62}$$

Hence, there are two main parts in the derivation: firstly, we need to specify a certain distribution $R(\cdot)$ and try to evaluate the relative entropy in (62). Secondly, we have the difficulty to compute an expectation over the capacity-achieving input distribution $Q^*(\cdot)$, which of course is unknown. To solve this problem we resort to the concept of *input distributions that escape to infinity* as introduced in [20] and further refined in [22]. This concept tells that under $Q^*(\cdot)$ the probability of any set of finite-power input symbols tends to zero as the power is loosened to infinity. This will allow us to prove that

$$\mathsf{E}_{Q^*}[o_X(1)] = o_{\mathsf{A}}(1). \tag{63}$$

for integrable $o_x(1)$. The price we pay for using this concept is that our results are only valid asymptotically as A tends to infinity.

5.2 Mathematical Preliminaries

Recall the following definition of a capacity-cost function with an average- and a peak-power 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 $g : \mathcal{X} \to \mathbb{R}_0^+$, we define the capacity-cost function $\mathsf{C} : \big([\inf_{x\in\mathcal{X}} g(x), \infty)\big)^2 \to \mathbb{R}_0^+$ by*

$$\mathsf{C}(\mathsf{A},\mathsf{E}) \triangleq \sup_{Q(\cdot)} I(Q,W), \qquad \mathsf{A},\mathsf{E} \geq \inf_{x\in\mathcal{X}} g(x) \tag{64}$$

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

$$Q\big(\{x \in \mathcal{X} \colon g(x) > \mathsf{A}\}\big) = 0 \tag{65}$$

and

$$\mathsf{E}_Q[g(X)] \leq \mathsf{E}. \tag{66}$$

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

The following lemma shows that capacity-achieving input distributions do exist for the channel under consideration.

Lemma 13. *Consider the channel* (1) *with the cost function $g(x) = x$, i.e., the constraints* (3) *and* (4)*. Then there exists*[6] *an input distribution $Q^*_{\mathsf{A},\mathsf{E}}(\cdot)$ that achieves the supremum in the definition of the capacity-cost function as given in* (64)*. Similarly, for the situation with only an average-power constraint, a capacity-achieving input distribution $Q^*_{\mathsf{E}}(\cdot)$ exists.*

Proof. See [5]. □

We will now review the notion of *input distributions that escape to infinity.* The statements in this section are valid in general, i.e., they are not restricted to the channel model under study. 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. We then consider a general cost function $g \colon \mathcal{X} \to [0,\infty)$ which is assumed measurable.[7]

Definition 14. *Fixing $\alpha \in (0,1]$ as ratio of available average to peak cost*

$$\alpha \triangleq \frac{\mathsf{E}}{\mathsf{A}} \tag{67}$$

we say that a family of input distributions $\{Q_{\mathsf{A},\mathsf{E}}(\cdot)\}$ on $\mathcal{X}$ parametrized[8] *by A and E* escapes to infinity *if for any fixed $\mathsf{A}_0 > 0$*

$$\lim_{\mathsf{A}\uparrow\infty} Q_{\mathsf{A},\alpha\mathsf{A}}\big(\{x \in \mathcal{X} \colon g(x) \leq \mathsf{A}_0\}\big) = 0. \tag{68}$$

To put it into simple words and taking the example of a finite input alphabet $\mathcal{X}$, an input distribution $\{Q_{\mathsf{A},\mathsf{E}}(\cdot)\}$ that escapes to infinity will assign zero probability to any finite-cost input letter once the cost constraints are relaxed completely, i.e., it will only use letters of infinite cost. So, for example, the binary on/off-distribution that with a fixed probability p generates A and with the remaining probability $1-p$ the zero symbol 0 is not escaping to infinity, because the probability of the zero symbol remains finite even if $\mathsf{A} \to \infty$. On the other hand, the binary distribution that with equal probability chooses between $\frac{\mathsf{A}}{2}$ and A does escape to infinity as both symbols with positive probability tend to infinity for $\mathsf{A} \to \infty$.

Definition 14 is of interest because in [22] a general theorem was presented demonstrating that if the ratio of mutual information to channel capacity is to approach one, then the input distribution must escape to infinity.

---

[6] Note that while the capacity-achieving input distribution might not be unique, it is shown in [5] that the capacity-achieving output distribution is.

[7] For an intuitive understanding of the following definition and some of its consequences, it is best to focus on the example of the channel model (1) where the channel inputs are nonnegative real numbers and where the cost function $g(\cdot)$ is $g(x) = x$, $\forall\, x \geq 0$.

[8] Note that due to the given cost function $g(\cdot)$ and the given ratio α, the parameters A and E must be chosen such that $\mathsf{A} \geq \inf_{x\in\mathcal{X}} \frac{g(x)}{\alpha}$ and $\mathsf{E} = \alpha\mathsf{A}$.

Proposition 15. *Let the capacity-cost function* $\mathsf{C}(\mathsf{A},\mathsf{E})$ *of a channel* $W(\cdot|\cdot)$ *be finite but unbounded. Suppose there exists a function* $\mathsf{C}_{\text{asy}}(\cdot)$ *that captures the asymptotic behavior of the capacity-cost function* $\mathsf{C}(\mathsf{A},\alpha\mathsf{A})$ *in the sense that*

$$\lim_{\mathsf{A}\uparrow\infty}\frac{\mathsf{C}(\mathsf{A},\alpha\mathsf{A})}{\mathsf{C}_{\text{asy}}(\mathsf{A})}=1. \tag{69}$$

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

$$\varliminf_{\mathsf{A}\uparrow\infty}\left\{\sup_{\mu\in(0,\mu_0]}\frac{\mu\mathsf{C}_{\text{asy}}\left(\frac{\mathsf{A}}{\mu}\right)}{\mathsf{C}_{\text{asy}}(\mathsf{A})}\right\}<1,\qquad\forall\,0<\mu_0<1. \tag{70}$$

Let $\{Q_{\mathsf{A},\alpha\mathsf{A}}(\cdot)\}_{\mathsf{A}\geq 0}$ *be a family of input distributions satisfying the cost constraints* (65) *and* (66) *such that*

$$\lim_{\mathsf{A}\uparrow\infty}\frac{I(Q_{\mathsf{A},\alpha\mathsf{A}},W)}{\mathsf{C}_{\text{asy}}(\mathsf{A})}=1. \tag{71}$$

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

Proof. See [22, Sec. VII.C.3]. □

To put it again into simple words, this theorem states that the optimal (i.e., capacity-achieving) input distribution escapes to infinity. Actually, the statement is even stronger: any input distribution that induces a mutual information growing with the same speed in the cost constraints as the capacity must escape to infinity. So, for example, if we are not necessarily interested in achieving the exact asymptotic capacity, but are content to have an input that will achieve the correct pre-log,[9] this input still must escape to infinity.

To better understand why Proposition 15 holds, consider for the moment the example of a channel with only an average-power constraint E, and separate the input alphabet into the two subsets $\mathcal{X}_1\triangleq\{x\colon 0\leq g(x)\leq\mathsf{E}_0\}$ and $\mathcal{X}_2\triangleq\{x\colon g(x)>\mathsf{E}_0\}$ for some fixed finite E_0. Any input distribution Q with average power $\mathsf{E}_Q[g(X)]=\mathsf{E}$ can now be regarded as a two-stage process: in a first stage $\mathcal{X}_1$ or $\mathcal{X}_2$ is chosen with probability $p\triangleq Q(\mathcal{X}_1)$ and $1-p=Q(\mathcal{X}_2)$, respectively. In a second stage a value is picked from the chosen subset $\mathcal{X}_i$ (with probability $\frac{Q(\cdot)}{p}$ if $\mathcal{X}_1$ has been chosen in the first stage, or probability $\frac{Q(\cdot)}{1-p}$ if $\mathcal{X}_2$ has been chosen). Now note that the first stage (being a binary process) can at most contribute 1 bit to the achieved rate. If in the second stage $\mathcal{X}_1$ is chosen, the maximum contribution is $\mathsf{C}(\mathsf{E}_0)<\infty$ due to the fixed limitation of the inputs in $\mathcal{X}_1$. If $\mathcal{X}_2$ is chosen, then we can achieve a rate of $\mathsf{C}\big(\frac{\mathsf{E}}{1-p}\big)$ for a proper choice of Q because the probability distribution in this case is $\frac{Q(\cdot)}{1-p}$. Since the first two contributions are finite, for large E they become negligible and we end up with a rate of approximately $(1-p)\mathsf{C}\left(\frac{\mathsf{E}}{1-p}\right)$. If now $Q(\cdot)$ does *not* escape to infinity, this means that p remains positive for all E. But condition (70) says that for large E,

$$\frac{(1-p)\mathsf{C}\left(\frac{\mathsf{E}}{1-p}\right)}{\mathsf{C}(\mathsf{E})}<1,\quad\forall\,p>0 \tag{72}$$

i.e., we are strictly suboptimal. Hence, a good input distribution will make sure that p tends to zero once E gets very large.

[9] Recall that by the *pre-log* we refer to the limiting ratio of channel capacity to the logarithm of the available cost. It is sometimes also known as "multiplexing gain".

Note that Proposition 15 holds in vast generality. For example, condition (70) is satisfied by any monotonically increasing, concave function with a slope that grows not faster than $\frac{\eta}{\mathsf{A}}$ (for some fixed $\eta > 0$). This includes, e.g.,

$$\log\log \mathsf{A}, \quad \log \mathsf{A}, \quad (\log \mathsf{A})^{\kappa} \text{ for } \kappa > 0$$

and any positive multiple thereof. Hence, most channels of interest fall under the assumptions of Proposition 15.

Before we show some consequences of Proposition 15, we next prove that the channel model (1) under investigation indeed also falls under the assumptions of Proposition 15.

Corollary 16. *Fix the average-to-peak-power ratio* α*. Then the capacity-achieving input distribution* $\{Q^*_{\mathsf{A},\alpha\mathsf{A}}(\cdot)\}_{\mathsf{A}\geq 0}$ *of the channel model* (1) *with peak- and average-power constraints* (3) *and* (4) *escapes to infinity. Similarly, for the situation with only an average-power constraint* (4)*,* $\{Q^*_{\mathsf{E}}(\cdot)\}_{\mathsf{E}\geq 0}$ *escapes to infinity.*

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

$$\mathsf{C}_{\text{asy}}(\mathsf{A}) = \frac{1}{2}\log \mathsf{A} \tag{73}$$

satisfies both conditions (69) and (70) of Proposition 15. The latter already has been shown in [22, 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.

Our proof relies heavily on the lower bounds derived in Section 4 and on Proposition 1. The details are deferred to Appendix B. □

The fact that $Q^*(\cdot)$ escapes to infinity will be used in this paper mainly in the following way.

Claim 17. *Let* $\{Q_{\mathsf{A},\alpha\mathsf{A}}(\cdot)\}_{\mathsf{A}\geq 0}$ *be a family of input distributions that escapes to infinity, and let* $f\colon \mathbb{R}_0^+ \to \mathbb{R}$ *be as in Definition 2, i.e.,*

$$f(x) = o_x(1). \tag{74}$$

Assume that f *is bounded. Then*

$$\lim_{\mathsf{A}\uparrow\infty} \mathsf{E}_{Q_{\mathsf{A},\alpha\mathsf{A}}}[f(X)] = 0. \tag{75}$$

Proof. Let $\epsilon > 0$ be arbitrary. Choose A_1 such that for all $\mathsf{A} > \mathsf{A}_1$

$$|f(\mathsf{A})| < \frac{\epsilon}{2}. \tag{76}$$

Recall that because $\{Q_{\mathsf{A},\alpha\mathsf{A}}(\cdot)\}_{\mathsf{A}\geq 0}$ escapes to infinity and because f is bounded, we have

$$\lim_{\mathsf{A}\uparrow\infty} \int_0^{\mathsf{A}_1} |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x = 0. \tag{77}$$

Hence, there exists an A_2 such that for $\mathsf{A} > \mathsf{A}_2$ we have

$$\int_0^{\mathsf{A}_1} |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x < \frac{\epsilon}{2}. \tag{78}$$

Therefore, for $\mathsf{A} > \mathsf{A}_0 \triangleq \max\{\mathsf{A}_1, \mathsf{A}_2\}$ we have

$$\begin{aligned}
\left|\mathsf{E}_{Q_{\mathsf{A},\alpha\mathsf{A}}}[f(X)]\right| &\leq \mathsf{E}_{Q_{\mathsf{A},\alpha\mathsf{A}}}[|f(X)|] && (79)\\
&= \int_0^\infty |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x && (80)\\
&= \int_0^{\mathsf{A}_1} |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x + \int_{\mathsf{A}_1}^\infty |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x && (81)\\
&< \frac{\epsilon}{2} + \int_{\mathsf{A}_1}^\infty \frac{\epsilon}{2}\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x && (82)\\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. && (83)
\end{aligned}$$

Here the first inequality follows from Jensen's inequality and the convexity of $|\cdot|$; (82) follows from (76) and (78); and in the last inequality we take $\epsilon/2$ out of the integration and upper-bound the integral by 1.

Hence, $\mathsf{E}_{Q_{\mathsf{A},\alpha\mathsf{A}}}[o_X(1)] = o_{\mathsf{A}}(1)$. □

5.3 Proof of the Upper Bound (14)

The derivation of (14) is based on (7) with the following choice of an output distribution $R(\cdot)$:

$$R(y) = \begin{cases} \frac{2p}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} & y < 0 \\ \frac{(1-2p)\sqrt{\mu}}{\sqrt{\mathsf{A}(1+\delta)y}\sqrt{\pi}\,\mathrm{erf}(\sqrt{\mu})} e^{-\frac{y\mu}{\mathsf{A}(1+\delta)}} & 0 \leq y \leq \mathsf{A}(1+\delta) \\ \frac{p}{\sqrt{2\pi}\mathcal{Q}(\mathsf{A}(1+\delta))} e^{-\frac{y^2}{2}} & y > \mathsf{A}(1+\delta) \end{cases} \tag{84}$$

where $\mu, \delta > 0$ and $0 < p < 1$ are arbitrary. Note that

$$\frac{\sqrt{\mu}}{\sqrt{\mathsf{A}(1+\delta)y}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)} e^{-\frac{y\mu}{\mathsf{A}(1+\delta)}} \tag{85}$$

is a PDF on $[0, \mathsf{A}(1+\delta)]$ that maximizes differential entropy under an average-power constraint and under the constraint that $\mathsf{E}[\log Y]$ is constant.[10] The choice of Gaussian "tails" for $y < 0$ and $y > \mathsf{A}(1+\delta)$ is motivated by simplicity. It will turn out that asymptotically they have no influence on the result.

With this choice we get

$$\begin{aligned}
D\big(W(\cdot|x)\big\|R(\cdot)\big) = &-\frac{1}{2}\log 2\pi e\sigma^2(1+\varsigma^2 x)\\
&\underbrace{-\int_{-\infty}^0 \log\left(\frac{2pe^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\right) W(y|x)\,\mathrm{d}y}_{c_1(x)}\\
&\underbrace{-\int_0^{\mathsf{A}(1+\delta)} \log\left(\frac{(1-2p)\sqrt{\mu}e^{-\frac{y\mu}{\mathsf{A}(1+\delta)}}}{\sqrt{\mathsf{A}(1+\delta)y}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)}\right) W(y|x)\,\mathrm{d}y}_{c_2(x)}\\
&\underbrace{-\int_{\mathsf{A}(1+\delta)}^\infty \log\left(\frac{pe^{-\frac{y^2}{2}}}{\sqrt{2\pi}\mathcal{Q}\left(\mathsf{A}(1+\delta)\right)}\right) W(y|x)\,\mathrm{d}y}_{c_3(x)}.
\end{aligned} \tag{86}$$

[10] Compare with (44).

We evaluate each term separately:

$$c_1(x) = \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\log\frac{\sqrt{2\pi}}{2p} + \frac{\sigma^2+\varsigma^2\sigma^2 x + x^2}{2}\mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right) - \frac{x}{2}\sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi}}e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}} \tag{87}$$

$$= o_x(1). \tag{88}$$

Similarly,

$$c_3(x) = \mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\log\frac{\mathcal{Q}\left(\mathsf{A}(1+\delta)\right)\sqrt{2\pi}}{p} + \frac{\sigma^2+\sigma^2\varsigma^2 x + x^2}{2}\mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sigma\sqrt{1+\varsigma^2 x}}\right) + \frac{\left(\mathsf{A}(1+\delta)+x\right)\sqrt{\sigma^2(1+\varsigma^2 x)}}{2\sqrt{2\pi}}e^{-\frac{(\mathsf{A}(1+\delta)-x)^2}{2\sigma^2(1+\varsigma^2 x)}} \tag{89}$$

$$\leq \sqrt{\frac{\sigma^2(1+\varsigma^2\mathsf{A})}{2\pi\delta^2\mathsf{A}^2}}e^{-\frac{\delta^2\mathsf{A}^2}{2\sigma^2(1+\varsigma^2\mathsf{A})}}\left|\log\frac{\mathcal{Q}\left(\mathsf{A}(1+\delta)\right)\sqrt{2\pi}}{p}\right| + \frac{\sigma^2+\sigma^2\varsigma^2\mathsf{A}+\mathsf{A}^2}{2}\sqrt{\frac{\sigma^2(1+\varsigma^2\mathsf{A})}{2\pi\delta^2\mathsf{A}^2}}e^{-\frac{\delta^2\mathsf{A}^2}{2\sigma^2(1+\varsigma^2\mathsf{A})}} + \frac{(2\mathsf{A}+\mathsf{A}\delta)\sqrt{\sigma^2(1+\varsigma^2\mathsf{A})}}{2\sqrt{2\pi}}e^{-\frac{\mathsf{A}^2\delta^2}{2\sigma^2(1+\varsigma^2\mathsf{A})}} \tag{90}$$

$$= o_{\mathsf{A}}(1) \tag{91}$$

where the inequality follows because $0 \leq x \leq \mathsf{A}$ and because the $\mathcal{Q}$-function as defined in (10) satisfies

$$\frac{1}{\sqrt{2\pi z^2}}e^{-\frac{z^2}{2}}\left(1-\frac{1}{z^2}\right) < \mathcal{Q}(z) < \frac{1}{\sqrt{2\pi z^2}}e^{-\frac{z^2}{2}} \tag{92}$$

such that

$$\mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sigma\sqrt{1+\varsigma^2 x}}\right) < \sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi(\mathsf{A}+\delta\mathsf{A}-x)^2}}e^{-\frac{(\mathsf{A}+\delta\mathsf{A}-x)^2}{2\sigma^2(1+\varsigma^2 x)}} \tag{93}$$

$$\leq \sqrt{\frac{\sigma^2(1+\varsigma^2\mathsf{A})}{2\pi\delta^2\mathsf{A}^2}}e^{-\frac{\delta^2\mathsf{A}^2}{2\sigma^2(1+\varsigma^2\mathsf{A})}}. \tag{94}$$

Finally, for $\mathsf{A} \geq \frac{\mu}{\pi\mathrm{erf}^2(\sqrt{\mu})}$,

$$c_2(x) = \underbrace{\left(1-\mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)-\mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right)}_{\leq 1} \cdot \underbrace{\left(\frac{1}{2}\log\mathsf{A}+\log\frac{\sqrt{1+\delta}\sqrt{\pi}\mathrm{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}}\right)}_{\geq 0 \text{ for } \mathsf{A}\geq\frac{\mu}{\pi\mathrm{erf}^2(\sqrt{\mu})}} + \frac{1}{2}\int_0^{\mathsf{A}(1+\delta)}\frac{\log y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}}e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}}\,\mathrm{d}y + \frac{\mu\sqrt{\sigma^2(1+\varsigma^2 x)}}{\mathsf{A}(1+\delta)\sqrt{2\pi}}\left(e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}}\underbrace{-e^{-\frac{(\mathsf{A}(1+\delta)-x)^2}{2\sigma^2(1+\varsigma^2 x)}}}_{\leq 0}\right)$$

$$+\frac{x\mu}{\mathsf{A}(1+\delta)}\cdot\underbrace{\left(1-\mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sigma\sqrt{1+\varsigma^2 x}}\right)-\mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right)}_{\leq 1} \tag{95}$$

$$\leq \frac{1}{2}\log\mathsf{A}+\log\frac{\sqrt{1+\delta}\sqrt{\pi}\operatorname{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}} + \frac{1}{2}\int_0^{\mathsf{A}(1+\delta)}\frac{\log y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}}e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}}\,\mathrm{d}y + \frac{1}{\mathsf{A}}\cdot\frac{\mu\sqrt{\sigma^2(1+\varsigma^2 x)}}{(1+\delta)\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma(1+\varsigma^2 x)}}+\frac{x\mu}{\mathsf{A}(1+\delta)}. \tag{96}$$

Next, we assume $x \geq 1$ and derive (using the substitution $\tilde{y} \triangleq \frac{y-x}{x}$):

$$\left.\int_0^{\mathsf{A}(1+\delta)}\frac{\log y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}}e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}}\,\mathrm{d}y\right|_{x\geq 1}$$

$$= \underbrace{\left(1-\mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)-\mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right)}_{\leq 1}\underbrace{\log x}_{\geq 0} + \frac{x}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}}\int_{-1}^{\frac{\mathsf{A}(1+\delta)-x}{x}}\underbrace{\log(1+\tilde{y})}_{\leq\tilde{y}}e^{-\frac{x^2\tilde{y}^2}{2\sigma^2(1+\varsigma^2 x)}}\,\mathrm{d}\tilde{y} \tag{97}$$

$$\leq \log x+\frac{x}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}}\int_{-1}^{\frac{\mathsf{A}(1+\delta)-x}{x}}\tilde{y}e^{-\frac{x^2\tilde{y}^2}{2\sigma^2(1+\varsigma^2 x)}}\,\mathrm{d}\tilde{y} \tag{98}$$

$$= \log x+\sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi x^2}}\left(e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}}-\underbrace{e^{-\frac{(\mathsf{A}(1+\delta)-x)^2}{2\sigma^2(1+\varsigma^2 x)}}}_{\geq 0}\right) \tag{99}$$

$$\leq \log x+\sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi x^2}}\cdot e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}}. \tag{100}$$

For $x < 1$ we bound $\log y \leq y$ and get

$$\left.\int_0^{\mathsf{A}(1+\delta)}\frac{\log y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}}e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}}\,\mathrm{d}y\right|_{x<1}$$

$$\leq \int_0^{\mathsf{A}(1+\delta)}\frac{y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}}e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}}\,\mathrm{d}y \tag{101}$$

$$= \sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi}}\left(e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}}-\underbrace{e^{-\frac{(\mathsf{A}(1+\delta)-x)^2}{2\sigma^2(1+\varsigma^2 x)}}}_{\geq 0}\right) + x\left(1-\mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)-\underbrace{\mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sigma\sqrt{1+\varsigma^2 x}}\right)}_{\geq 0}\right) \tag{102}$$

$$\leq \sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi}}e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}}+x\left(1-\mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right). \tag{103}$$

Hence, because (103) is bounded and from (100) we have

$$\int_0^{\mathsf{A}(1+\delta)} \frac{\log y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}} e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}} \,\mathrm{d}y = \log x + o_x(1). \tag{104}$$

Plugging this into (96) yields

$$\begin{aligned} c_2(x) \le \frac{1}{2}\log \mathsf{A} + \log \frac{\sqrt{1+\delta}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}} + \frac{1}{2}\log x \\ + \frac{x\mu}{\mathsf{A}(1+\delta)} + o_x(1) + \frac{1}{\mathsf{A}}\cdot o_x(1). \end{aligned} \tag{105}$$

Using all these results together with (86) and (7), we get

$$\begin{aligned} \mathsf{C} &\le \mathsf{E}_{Q^*}\left[D\big(W(\cdot|X)\big\|R(\cdot)\big)\right] &(106)\\ &\le \mathsf{E}_{Q^*}\Bigg[-\frac{1}{2}\log 2\pi e\sigma^2(1+\varsigma^2 X) + o_X(1) + \frac{1}{2}\log \mathsf{A} \\ &\qquad + \log \frac{\sqrt{1+\delta}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}} + \frac{1}{2}\log X + \frac{X\mu}{\mathsf{A}(1+\delta)} \\ &\qquad + o_X(1) + \frac{1}{\mathsf{A}}\cdot o_X(1) + o_{\mathsf{A}}(1)\Bigg] &(107)\\ &= -\frac{1}{2}\log 2\pi e\varsigma^2\sigma^2 - \mathsf{E}_{Q^*}\Bigg[\underbrace{\frac{1}{2}\log\left(1+\frac{1}{\varsigma^2 X}\right)}_{o_X(1)}\Bigg] \\ &\quad + \frac{1}{2}\log\mathsf{A} + \frac{1}{2}\log(1+\delta) - \frac{1}{2}\log\mu + \log\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right) \\ &\quad - \log(1-2p) + \frac{\mathsf{E}_{Q^*}[X]\,\mu}{\mathsf{A}(1+\delta)} + \mathsf{E}_{Q^*}[o_X(1)] \\ &\quad + \frac{1}{\mathsf{A}}\cdot\mathsf{E}_{Q^*}[o_X(1)] + o_{\mathsf{A}}(1) &(108)\\ &\le \frac{1}{2}\log\frac{\mathsf{A}}{\sigma^2} - \frac{1}{2}\log 2\pi e\varsigma^2 + \frac{1}{2}\log(1+\delta) - \frac{1}{2}\log\mu \\ &\quad + \log\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right) - \log(1-2p) + \frac{\alpha\mu}{1+\delta} \\ &\quad + \mathsf{E}_{Q^*}[o_X(1)] + \frac{1}{\mathsf{A}}\cdot\mathsf{E}_{Q^*}[o_X(1)] + o_{\mathsf{A}}(1). &(109) \end{aligned}$$

Finally, we use[11] Claim 17 and choose μ to be the solution to (16). The result now follows since p and δ are arbitrary.

5.4 Proof of the Upper Bounds (20) and (24)

As noted in Remarks 6 and 9, (20) and (24) can be seen as limiting cases of (14) for $\alpha \uparrow \frac{1}{3}$ and $\alpha \downarrow 0$, respectively. They are derived analogously to (14).

For (20) we make the same choice (84), but with

$$\mu \triangleq \frac{1}{\mathsf{A}}. \tag{110}$$

[11] Note that all $o_x(1)$ functions are integrable and bounded.

Note that in order to avoid any dependence on α, we also upper-bound any occurrence of $\mathsf{E}_{Q^*}[X]$ by A instead of $\alpha\mathsf{A}$. Note that from (110) we have

$$\log \frac{\sqrt{\pi}\mathrm{erf}\left(\sqrt{\mu}\right)}{\sqrt{\mu}} = \log 2 + o_{\mathsf{A}}(1) \tag{111}$$

and hence, continuing from (107), we get

$$\begin{aligned}
\mathsf{C} \leq \mathsf{E}_{Q^*}\Bigg[& -\frac{1}{2}\log 2\pi e\sigma^2(1+\varsigma^2 X) + \frac{1}{2}\log X + \frac{1}{2}\log \mathsf{A} \\
& + \log \frac{\sqrt{1+\delta}\sqrt{\pi}\mathrm{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}} + \frac{X\mu}{\mathsf{A}(1+\delta)} \\
& + o_X(1) + \frac{1}{\mathsf{A}}\cdot o_X(1) + o_{\mathsf{A}}(1)\Bigg] && (112)\\
= {} & \frac{1}{2}\log \mathsf{A} - \frac{1}{2}\log 2\pi e\varsigma^2\sigma^2 + \frac{1}{2}\log(1+\delta) + \log 2 \\
& - \log(1-2p) - \mathsf{E}_{Q^*}\Bigg[\underbrace{\frac{1}{2}\log\left(1+\frac{1}{\varsigma^2 X}\right)}_{o_X(1)}\Bigg] \\
& + \frac{\mathsf{E}_{Q^*}[X]}{\mathsf{A}^2(1+\delta)} + \mathsf{E}_{Q^*}[o_X(1)] + \frac{1}{\mathsf{A}}\cdot \mathsf{E}_{Q^*}[o_X(1)] + o_{\mathsf{A}}(1) && (113)\\
\leq {} & \frac{1}{2}\log\frac{\mathsf{A}}{\sigma^2} - \frac{1}{2}\log\frac{\pi e\varsigma^2}{2} + \frac{1}{2}\log(1+\delta) - \log(1-2p) \\
& + \underbrace{\frac{1}{\mathsf{A}(1+\delta)}}_{o_{\mathsf{A}}(1)} + \mathsf{E}_{Q^*}[o_X(1)] + \frac{1}{\mathsf{A}}\cdot\mathsf{E}_{Q^*}[o_X(1)] + o_{\mathsf{A}}(1) && (114)
\end{aligned}$$

where in the last step we bounded $\mathsf{E}_{Q^*}[X] \leq \mathsf{A}$. The result (20) now follows from (63) and because p and δ are arbitrary.

For (24) we choose

$$R(y) = \begin{cases} \frac{2p}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} & y < 0 \\ \frac{1-p}{\sqrt{2\pi\mathsf{E}y}} e^{-\frac{y}{2\mathsf{E}}} & y \geq 0 \end{cases} \tag{115}$$

where $0 < p < 1$ is a free parameter. Then we get

$$\begin{aligned}
D\big(W(\cdot|x)\big\|R(\cdot)\big) = {} & -\frac{1}{2}\log 2\pi e\sigma^2(1+\varsigma^2 x) \\
& \underbrace{-\int_{-\infty}^{0}\log\left(\frac{2p}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\right)W(y|x)\,\mathrm{d}y}_{c_1(x)} \\
& \underbrace{-\int_{0}^{\infty}\log\left(\frac{1-p}{\sqrt{2\pi\mathsf{E}y}}e^{-\frac{y}{2\mathsf{E}}}\right)W(y|x)\,\mathrm{d}y}_{c_2'(x)}. && (116)
\end{aligned}$$

From (88) we know that $c_1(x) = o_x(1)$. For $c_2'(x)$ we get

$$\begin{aligned}
c_2'(x) &= \left(1 - \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right)\log\frac{\sqrt{2\pi\mathsf{E}}}{1-p} + \frac{\sigma}{2\mathsf{E}}\sqrt{\frac{1+\varsigma^2 x}{2\pi}}e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}} \\
&\quad + \frac{x}{2\mathsf{E}}\left(1 - \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right) + \frac{1}{2}\int_0^\infty \frac{\log y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}}e^{-\frac{(y-x)^2}{2(\sigma^2+x)}}\,\mathrm{d}y \qquad (117)\\
&\le \left(1 - \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right)\log\frac{\sqrt{2\pi\mathsf{E}}}{1-p} + \underbrace{\frac{\sigma}{2\mathsf{E}}}_{\le\frac{\sigma}{2}\text{ for }\mathsf{E}>1}\sqrt{\frac{1+\varsigma^2 x}{2\pi}}e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}} \\
&\quad + \frac{x}{2\mathsf{E}}\left(1 - \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right) + \frac{1}{2}\left(1 - \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right)\log x \\
&\quad + \frac{1}{2}\sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi x^2}}e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}} \qquad (118)\\
&\le \log\frac{\sqrt{2\pi\mathsf{E}}}{1-p} + \frac{\sigma}{2}\sqrt{\frac{1+\varsigma^2 x}{2\pi}}e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}} + \frac{x}{2\mathsf{E}} + \frac{1}{2}\log x \\
&\quad - \frac{1}{2}\mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\log x + \frac{\sigma}{2}\sqrt{\frac{1+\varsigma^2 x}{2\pi x^2}}e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}} \qquad (119)\\
&= \log\frac{\sqrt{2\pi\mathsf{E}}}{1-p} + \frac{x}{2\mathsf{E}} + \frac{1}{2}\log x + o_x(1) \qquad (120)
\end{aligned}$$

where the first inequality follows in an equivalent way as shown in (97)–(99), and the second inequality holds for large enough E such that $\log\frac{\sqrt{2\pi\mathsf{E}}}{1-p} > 0$ and $\mathsf{E} > 1$.

We continue with (116) and take the expectation over Q^*:

$$\begin{aligned}
\mathsf{C} &\le \mathsf{E}_{Q^*}\left[D\big(W(\cdot|X)\big\|R(\cdot)\big)\right] \qquad (121)\\
&\le \mathsf{E}_{Q^*}\left[-\frac{1}{2}\log 2\pi e\sigma^2(1+\varsigma^2 X) + \log\frac{\sqrt{2\pi\mathsf{E}}}{1-p} + \frac{X}{2\mathsf{E}} + \frac{1}{2}\log X + o_X(1)\right] \qquad (122)\\
&= \mathsf{E}_{Q^*}\Bigg[\frac{1}{2}\log\frac{\mathsf{E}}{\sigma^2} - \frac{1}{2}\log\varsigma^2 - \frac{1}{2} - \log(1-p) + \frac{X}{2\mathsf{E}} \underbrace{- \frac{1}{2}\log\left(1 + \frac{1}{\varsigma^2 X}\right)}_{o_X(1)} \\
&\qquad + o_X(1)\Bigg] \qquad (123)\\
&\le \frac{1}{2}\log\frac{\mathsf{E}}{\sigma^2} - \frac{1}{2}\log\varsigma^2 - \frac{1}{2} - \log(1-p) + \frac{\mathsf{E}}{2\mathsf{E}} + \mathsf{E}_{Q^*}[o_X(1)]. \qquad (124)
\end{aligned}$$

Analogously to (63) we have $\mathsf{E}_{Q^*}[o_X(1)] = o_{\mathsf{E}}(1)$. The result now follows since p is arbitrary.

6 Derivation of the Low-Power Behavior

For scenarios where the peak-power constraint tends to 0, a result by Prelov and van der Meulen [23] can be used to obtain the exact asymptotic low-power capacity. The following theorem is included as a special case in [23, Theorem 2].

Theorem 18 ([23]). *Consider a channel that for all sufficiently small inputs x produces an output that is Gaussian distributed with mean m_x and variance σ_x^2 that can* *depend on x. Then, for sufficiently small* A *and* $|X| \leq \mathsf{A}$*, the mutual information between the channel's input X and output Y satisfies*

$$I(X;Y) = \frac{1}{2} J(0) \mathsf{Var}(X) + o(\mathsf{A}^2) \tag{125}$$

where $o(\mathsf{A}^2)$ *denotes a term that tends to 0 faster than* $\mathsf{A}^2 \downarrow 0$*, and where* $J(0)$ *denotes the Fisher information of the channel at* 0*:*

$$J(x) \triangleq \int_{\mathcal{Y}} \frac{\left(\frac{\mathrm{d}}{\mathrm{d}x} W(y|x)\right)^2}{W(y|x)} \, \mathrm{d}y. \tag{126}$$

It is quite obvious that the optical intensity channel with input-dependent Gaussian noise satisfies the assumption in the theorem. Thus, we can use it to derive the asymptotic low-power capacity under both peak- and average-power constraints (33) and under a peak-power constraint only (34).

We briefly sketch the derivation of Theorem 10. For the channel law (1) we have

$$J(x) = \frac{2 + \varsigma^4 \sigma^2 + 2\varsigma^2 x}{2\sigma^2 (1 + \varsigma^2 x)^2} \tag{127}$$

such that

$$J(0) = \frac{2 + \varsigma^4 \sigma^2}{2\sigma^2}. \tag{128}$$

Moreover, it is not difficult to see that

$$\max_{\substack{Q \text{ s.t.} \\ (3) \text{ and } (4) \\ \text{are satisfied}}} \mathsf{Var}(X) = \begin{cases} \mathsf{E}(\mathsf{A} - \mathsf{E}) = \alpha(1-\alpha)\mathsf{A}^2 & \text{if } \mathsf{E} < \frac{\mathsf{A}}{2} \\ \frac{\mathsf{A}^2}{4} & \text{if } \mathsf{E} \geq \frac{\mathsf{A}}{2}. \end{cases} \tag{129}$$

The theorem is now established by combining (128) and (129) with (125) and the definition of channel capacity.

7 Discussion and Conclusions

New (firm) lower bounds and new (asymptotic) upper bounds on the capacity of the optical intensity channel with input-dependent Gaussian noise 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 capacity of the optical intensity channel with input-dependent Gaussian noise has also be established exactly for the asymptotic low-SNR situation where the peak- and average-power tend to zero with their ratio held constant.

7.1 Comparison to other Channel Models

It is interesting to compare the presented results with the results of the free-space optical intensity channel [8] and the Poisson channel [11] [12]. As mentioned above, the former model is similar to the given channel (1) because in both channels the noise is modeled to be additive and Gaussian distributed. The free-space optical intensity channel, however, neglects a fundamental property of optical intensity communication: it ignores that the noise is implicitly dependent on the current input signal.

At low power this disregard does not have a large impact on the behavior of capacity. The asymptotic capacities (33) and (34) are similar to the low-power asymptotic capacities of the free-space optical intensity channel [8], especially for small values of ς^2. In this regime, the two models also share the same threshold $\alpha = \frac{1}{2}$ between the case with both peak- and average-power constraints being active and the case where the average-power constraint is inactive.

At high power, however, the input-dependent part of the noise becomes dominant. This can be seen very clearly by the fact that in [8] the capacity grows like $\log \mathsf{A}$ for large A, whereas here we have an asymptotic growth of only $\frac{1}{2}\log \mathsf{A}$. Moreover, for the free-space optical intensity channel the range of the average-to-peak power ratio α with no impact on the asymptotic high-SNR capacity is

$$\frac{1}{2} \leq \alpha \leq 1 \tag{130}$$

while here at high power we have

$$\frac{1}{3} \leq \alpha \leq 1. \tag{131}$$

However, note that while the former result holds true for all values of A and E, in the current paper we have only been able to prove that for $\mathsf{A} \uparrow \infty$ the threshold is $\frac{1}{3}$, and for $\mathsf{A} \downarrow 0$ it is $\frac{1}{2}$. For any finite value of A the threshold is likely to be somewhere in between, varying with A, σ^2, and ς^2.

On the other hand, it is very interesting to observe that the asymptotic high-power results (18), (22), and (25) turn out to be *identical*[12] to the asymptotic capacity of the Poisson channel [11] for the case $\varsigma^2\sigma^2 = 1$.

Intuitively, this correspondence can be understood by realizing that, for large x, the cumulative distribution function of a Gaussian random variable with mean x and variance x approximates the cumulative distribution function of a Poisson random variable with mean x. We prove this statement in Appendix C. Then recall from Corollary 16 about the capacity-achieving input distribution escaping to infinity that asymptotically as the available peak and average power A and E tend to infinity, the optimal input distribution does not put any finite mass on any finite input. Hence, for an optimal distribution and for $\mathsf{A}, \mathsf{E} \to \infty$, we really do have $x \to \infty$. So we see that asymptotically for large SNR and if $\varsigma^2\sigma^2 = 1$, the channel model (1) converges to the Poisson channel.

Thus, we conclude that whereas the capacity of the optical intensity channel with input-dependent Gaussian noise at low power behaves similar to the capacity of the free-space optical intensity channel, at high power it behaves similar to the capacity of the discrete-time Poisson channel.

7.2 Lower Bounds

In principle, the derivation of lower bounds on capacity is straightforward: since capacity is defined as a maximization of mutual information over a set of possible input distributions, a lower bound can be found by simply dropping the maximization and picking any input distribution from the candidate set. The problems lie in

[12] I.e., not only the pre-log factor $\frac{1}{2}$ is the same, but also the second term in the high-SNR expansion of capacity!

the details: firstly, it is not clear which distribution to pick that would yield a good lower bound. And secondly, it usually becomes very difficult to evaluate the mutual information analytically for a chosen input distribution. In particular the latter is a big hurdle if one is not content with numerical evaluations, but would like to derive analytical bounds.

In this work, we have solved this issue using Proposition 11. There we prove that it is possible to replace (i.e., lower-bound) the differential entropy of the channel's output with the differential entropy of its input. This change simplifies the evaluation of mutual information drastically. Note that the presented lower bounds perform poorly at low power. This is to be expected because at low power, the entropy of the channel output will be dominated by the uncertainty of the noise and not the input. Hence, the lower bound in Proposition 11 is poor in this regime. On the other hand, at high power the input's uncertainty will dominate the output entropy. Indeed, as we have shown, asymptotically the lower bound is tight!

Beside the simplification in the evaluation of mutual information, Proposition 11 fulfills another important task: it also provides us with clues on how to pick a good candidate input distribution. Once $h(Y)$ is lower-bounded according to Proposition 11, we face an expression that depends on the input distribution Q only. A good choice is then to choose Q such as to maximize this expression under the given power constraints. For example, in (42) we choose Q such as to maximize $h(X) - \frac{1}{2}\mathsf{E}[\log X]$.

We would like to point out that the behavior of the lower bound is relatively robust. If we choose instead of the optimized Q given in (44), a much simpler uniform distribution[13]

$$Q(x) = \begin{cases} \frac{1}{2\alpha\mathsf{A}} & \text{if } x \in [0, 2\alpha\mathsf{A}] \\ 0 & \text{otherwise} \end{cases} \tag{132}$$

(where $0 < \alpha \leq \frac{1}{2}$), then we get a lower bound that has the following asymptotic behavior:

$$\mathsf{C}(\mathsf{A}, \alpha\mathsf{A}) \geq \frac{1}{2}\log\frac{\mathsf{A}}{\sigma^2} - \frac{1}{2}\log\frac{\pi\varsigma^2}{\alpha} + o_{\mathsf{A}}(1). \tag{133}$$

We see that the pre-log $\frac{1}{2}$ is preserved, but only the second constant term is slightly too small. For example, for $\alpha = \frac{1}{3}$ we get a gap between the asymptotic capacity and this lower bound of

$$\Delta_{\{\alpha=\frac{1}{3}\}} = -\frac{1}{2}\log\frac{e}{2} + \frac{1}{2}\log\frac{1}{\alpha}\bigg|_{\alpha=\frac{1}{3}} \approx 0.40 \text{ nats} \tag{134}$$

for $\alpha = \frac{1}{2}$ we get a gap $\Delta_{\{\alpha=\frac{1}{2}\}} \approx 0.19$ nats, or for $\alpha = 0.02$ we get a gap $\Delta_{\{\alpha=0.02\}} \approx 0.57$ nats.

Another interesting observation we can gain from this discussion concerns the threshold of α where the average-power constraint becomes inactive. We have already seen that for the optical intensity channel with input-dependent Gaussian noise (1) this threshold is at $\alpha = \frac{1}{2}$ asymptotically as $\mathsf{A} \downarrow 0$, and at $\alpha = \frac{1}{3}$ asymptotically as $\mathsf{A} \uparrow \infty$. For finite values of A we expect it to be in between these two extreme values. Since the structure of our lower bounds is optimized for very large

[13]Note that the choice of the right boundary of this uniform distribution is designed such that the average-power constraint is satisfied. Also note that this distribution does escape to infinity: for any fixed $\mathsf{A}_0 > 0$ we have $\Pr[X \leq \mathsf{A}_0] = \frac{\mathsf{A}_0}{2\alpha\mathsf{A}} \to 0$ as $\mathsf{A} \to \infty$.

values of A, they do not reflect this change of threshold, but only follow the asymptotic behavior with a threshold of $\frac{1}{3}$. A different design of the lower bound might lead to a different behavior here, as can be seen from the example (132)–(133) that exemplifies a threshold $\alpha = \frac{1}{2}$.

7.3 Upper Bounds

For the asymptotic upper bounds at high power we relied on two concepts introduced in [20] and [22]. Firstly, we use a technique of a duality-based upper bound on mutual information (see Proposition 1). One needs to pick a distribution on the channel output alphabet and use it to evaluate an expression containing the relative entropy. While we are completely free in this choice of the output distribution, the problems are similar to the situation of the lower bounds: we need to find an output distribution that simultaneously is simple enough for evaluation, but complex enough to lead to an acceptable upper bound. The choices used in the presented proofs here have been inspired by the input distributions that we have chosen in the derivations of the corresponding lower bounds.

In addition, there is a further problem: the expression of the upper bound given in Proposition 1 also depends on the capacity-achieving input distribution. While obviously it is not known, we do know some properties of a capacity-achieving input distribution: firstly, it must satisfy the given power constraints, and secondly, it must escape to infinity.

The notion of *input distributions that escape to infinity* is the second main concept used in the derivation of the upper bounds. As introduced in [20] and [22] and reviewed in Section 5.2, most channels of interest have the basic property that any finite-power input symbol will become less and less desirable the larger the allowed input power becomes. In the asymptotic limit, an optimal input distribution will assign zero probability to any finite-power input. This property has been used extensively in the derivation of the upper bounds. Its main application is that any expression of the form

$$\lim_{\mathsf{A}\uparrow\infty} \mathsf{E}_{Q^*}[\ldots] \tag{135}$$

can be replaced by

$$\lim_{x\uparrow\infty} \ldots \tag{136}$$

The price we pay is that in contrast to the lower bounds, the upper bounds are only valid asymptotically.

A A Proof of Proposition 11

We start by reducing the problem to the situation without input-independent noise. To that goal note that Y can be written as

$$Y = \underbrace{x + \sqrt{x}Z_1}_{Y_1} + \underbrace{Z_0}_{Y_0} = Y_1 + Y_0, \tag{137}$$

where Y_0 and Y_1 are independent, $Y_0 \sim \mathcal{N}_{\mathbb{R}}\left(0, \sigma^2\right)$, and conditional on $X = x$, $Y_1 \sim \mathcal{N}_{\mathbb{R}}\left(x, x\varsigma^2\sigma^2\right)$. By the fact that conditioning reduces entropy we have

$$h(Y) = h(Y_0 + Y_1) \geq h(Y_0 + Y_1 \,|\, Y_0) = h(Y_1 \,|\, Y_0) = h(Y_1). \tag{138}$$

Hence, we can reduce the problem to finding a lower bound to $h(Y_1)$ only.

The proof of (36) is based on the data processing inequality for relative entropies [24, Ch. 1, Lemma 3.11(ii)].

According to the assumptions of this proposition we have a distribution $Q(\cdot)$ on the input X with $\mathsf{E}_Q[X] = \mathsf{E}$. Let $Q_{\text{Exp}}(\cdot)$ be an exponential probability distribution of mean $\mathsf{E} > 0$

$$Q_{\text{Exp}}(x) \triangleq \frac{1}{\mathsf{E}} e^{-\frac{x}{\mathsf{E}}}, \quad x \geq 0. \tag{139}$$

If $Q_{\text{Exp}}(\cdot)$ is used as input distribution to our reduced channel Y_1 as given in (137), then the corresponding output distribution is

$$(Q_{\text{Exp}} W_1)(y) = \frac{1}{\sqrt{\mathsf{E}(\mathsf{E} + 2\varsigma^2\sigma^2)}} \cdot \exp\left(\frac{\sqrt{\mathsf{E}}\, y - \sqrt{\mathsf{E} + 2\varsigma^2\sigma^2}\, |y|}{\varsigma^2\sigma^2\sqrt{\mathsf{E}}}\right), \quad y \in \mathbb{R}. \tag{140}$$

By the data processing theorem for relative entropy [24, Ch. 1, Lemma 3.11(ii)] we now obtain:

$$D\big(Q \big\| Q_{\text{Exp}}\big) \geq D\big((QW_1) \big\| (Q_{\text{Exp}} W_1)\big) \tag{141}$$

where $(QW_1)(\cdot)$ denotes the corresponding output distribution of the channel Y_1 when an input of law $Q(\cdot)$ is used. The first inequality in (36) in the proposition's statement now follows by evaluating the left-hand side of (141):

$$\begin{aligned} D\big(Q \big\| Q_{\text{Exp}}\big) &= -h_Q(X) - \mathsf{E}_Q\left[\log\left(\frac{1}{\mathsf{E}} e^{-\frac{X}{\mathsf{E}}}\right)\right] && (142)\\ &= -h(X) + \log \mathsf{E} + 1 && (143) \end{aligned}$$

(where $h_Q(X)$ is computed based on the law $Q(\cdot)$), and by evaluating the right-hand side of (141):

$$\begin{aligned} &D\big((QW_1) \big\| (Q_{\text{Exp}} W_1)\big) \\ &\quad = -h_{(QW_1)}(Y) \\ &\qquad - \mathsf{E}_{(QW_1)}\left[\log\left(\frac{1}{\sqrt{\mathsf{E}(\mathsf{E} + 2\varsigma^2\sigma^2)}} \cdot \exp\left(\frac{\sqrt{\mathsf{E}}\, Y - \sqrt{\mathsf{E} + 2\varsigma^2\sigma^2}\, |Y|}{\varsigma^2\sigma^2\sqrt{\mathsf{E}}}\right)\right)\right] && (144)\\ &\quad = -h(Y_1) + \frac{1}{2}\log \mathsf{E} + \frac{1}{2}\log(\mathsf{E} + 2\varsigma^2\sigma^2) - \frac{\mathsf{E}}{\varsigma^2\sigma^2} \\ &\qquad + \frac{1}{\varsigma^2\sigma^2}\sqrt{\frac{\mathsf{E} + 2\varsigma^2\sigma^2}{\mathsf{E}}}\, \mathsf{E}_{(QW_1)}[|Y|] && (145)\\ &\quad \geq -h(Y_1) + \frac{1}{2}\log \mathsf{E} + \frac{1}{2}\log(\mathsf{E} + 2\varsigma^2\sigma^2) - \frac{\mathsf{E}}{\varsigma^2\sigma^2} + \frac{\sqrt{\mathsf{E}(\mathsf{E} + 2\varsigma^2\sigma^2)}}{\varsigma^2\sigma^2}. && (146) \end{aligned}$$

Here we have used Jensen's inequality with the convex function $|\cdot|$ to get

$$\mathsf{E}_{(QW_1)}[|Y|] \geq \big|\mathsf{E}_{(QW_1)}[Y]\big| = \mathsf{E}. \tag{147}$$

The proof of the monotonicity and positivity of $f_{\text{low}}(\cdot)$ is straightforward and therefore omitted. To see that $f_{\text{low}}(\mathsf{E})$ tends to 0 as $\mathsf{E} \uparrow \infty$, note that

$$\frac{\sqrt{\mathsf{E}(\mathsf{E} + 2\varsigma^2\sigma^2)}}{\varsigma^2\sigma^2} = \frac{\sqrt{(\mathsf{E} + \varsigma^2\sigma^2)^2 - \varsigma^4\sigma^4}}{\varsigma^2\sigma^2}. \tag{148}$$

B A Proof of Corollary 16

To prove the claim of this lemma we rely on Proposition 15, i.e., we need to find a function $\mathsf{C}_{\text{asy}}(\cdot)$ that satisfies (69) and (70).

From the lower bounds in Theorems 4, 5 and 8 (which are proven in Section 4) we know that

$$\varliminf_{\mathsf{A}\uparrow\infty} \frac{\mathsf{C}(\mathsf{A},\alpha\mathsf{A})}{\frac{1}{2}\log\mathsf{A}} \geq 1 \tag{149}$$

and

$$\varliminf_{\mathsf{E}\uparrow\infty} \frac{\mathsf{C}(\mathsf{E})}{\frac{1}{2}\log\mathsf{E}} \geq 1 \tag{150}$$

respectively.

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

$$\mathsf{C}(\mathsf{A},\alpha\mathsf{A}) \leq \mathsf{C}_{\text{peak}}(\mathsf{A}) \leq \mathsf{C}_{\text{avg}}(\mathsf{A}) \tag{151}$$

where $\mathsf{C}_{\text{peak}}(\cdot)$ and $\mathsf{C}_{\text{avg}}(\cdot)$ 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 only case.

Our derivation is based on Proposition 1 with the following choice of an output distribution:

$$R(y) \triangleq \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} & y < 0 \\ \frac{1}{\sqrt{8\pi\mathsf{E}y}} e^{-\frac{y}{2\mathsf{E}}} & y \geq 0. \end{cases} \tag{152}$$

We get

$$D\big(W(\cdot|x)\big\|R(\cdot)\big) = -\frac{1}{2}\log 2\pi e\sigma^2(1+\varsigma^2 x) \underbrace{-\int_{-\infty}^{0} \log\frac{e^{-y^2/2}}{\sqrt{2\pi}} W(y|x)\,\mathrm{d}y}_{c_1'(x)} \underbrace{-\int_0^\infty \log\left(\frac{1}{\sqrt{8\pi\mathsf{E}y}} e^{-\frac{y}{2\mathsf{E}}}\right) W(y|x)\,\mathrm{d}y}_{c_2'(x)}. \tag{153}$$

From (86) and (87) with $p = \frac{1}{2}$ we know that

$$\begin{aligned} c_1'(x) &= \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\log\sqrt{2\pi} + \frac{\sigma^2+\varsigma^2\sigma^2 x + x^2}{2}\mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right) \\ &\quad - \frac{x}{2}\sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi}} e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}} & (154) \\ &\leq \frac{1}{4} e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}} \left(2\log\sqrt{2\pi} + \sigma^2 + \varsigma^2\sigma^2 x + x^2 - 2x\sqrt{\frac{\sigma^2(1+\varsigma^2 x}{2\pi}}\right) & (155) \end{aligned}$$

where in the last step we used the bound

$$\mathcal{Q}(\xi) \leq \frac{1}{2} e^{-\frac{\xi^2}{2}}, \quad \xi \geq 0. \tag{156}$$

Note that (155) is bounded, i.e., there exists some finite constant $k_1 \in \mathbb{R}$ (independent of x and E) such that

$$c_1'(x) \leq k_1, \quad \forall\, x \geq 0. \tag{157}$$

For $c_2'(x)$ we bound as follows (compare with (117)):

$$c_2'(x) = \underbrace{\left(1 - \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right)}_{\leq 1} \log\sqrt{8\pi\mathsf{E}} + \frac{1}{2\mathsf{E}} \underbrace{\sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi}} e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}}}_{\leq k_2}$$
$$+ \frac{x}{2\mathsf{E}} \underbrace{\left(1 - \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right)}_{\leq 1} + \frac{1}{2}\int_0^\infty \frac{\log y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}} e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}} \, \mathrm{d}y \tag{158}$$

$$\leq \max\left\{\frac{1}{2}\log 8\pi\mathsf{E}, 0\right\} + \frac{k_2}{2\mathsf{E}} + \frac{x}{2\mathsf{E}} + \frac{I\{x \geq 1\}}{2} \int_0^\infty \frac{\log y \cdot e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}}}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}} \, \mathrm{d}y \bigg|_{x\geq 1}$$
$$+ \frac{I\{x < 1\}}{2} \int_0^\infty \frac{\log y \cdot e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}}}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}} \, \mathrm{d}y \bigg|_{x<1} \tag{159}$$

where $k_2 \in \mathbb{R}$ is another finite constant independent of x and E, and $I\{\cdot\}$ denotes the indicator function

$$I\{\text{statement}\} = \begin{cases} 1 & \text{if statement is true} \\ 0 & \text{otherwise.} \end{cases} \tag{160}$$

Analogously to (100) we next have

$$\int_0^\infty \frac{\log y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}} e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}} \, \mathrm{d}y \bigg|_{x\geq 1} \leq \log x + \underbrace{\frac{1}{x}}_{\leq 1} \cdot \underbrace{\sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi}} e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}}}_{\leq k_2} \tag{161}$$
$$\leq \log x + k_2 \tag{162}$$

and analogously to (103) we get

$$\int_0^\infty \frac{\log y}{\sqrt{2\pi\sigma^2(1+\varsigma^2 x)}} e^{-\frac{(y-x)^2}{2\sigma^2(1+\varsigma^2 x)}} \, \mathrm{d}y \bigg|_{x<1}$$
$$\leq \underbrace{\sqrt{\frac{\sigma^2(1+\varsigma^2 x)}{2\pi}} e^{-\frac{x^2}{2\sigma^2(1+\varsigma^2 x)}}}_{\leq k_2} + \underbrace{x}_{\leq 1} \underbrace{\left(1 - \mathcal{Q}\left(\frac{x}{\sigma\sqrt{1+\varsigma^2 x}}\right)\right)}_{\leq 1} \tag{163}$$
$$\leq k_2 + 1. \tag{164}$$

Plugging all this into (153) finally yields

$$D\big(W(\cdot|x)\big\|R(\cdot)\big) \leq -\frac{1}{2}\log 2\pi e\sigma^2 - \frac{1}{2}\log(1+\varsigma^2 x) + k_1 + \max\left\{\frac{1}{2}\log 8\pi\mathsf{E}, 0\right\}$$
$$+ \frac{k_2}{2\mathsf{E}} + \frac{x}{2\mathsf{E}} + \frac{I\{x \geq 1\}}{2}\big(\log x + k_2\big) + \frac{I\{x < 1\}}{2}(k_2 + 1) \tag{165}$$
$$= \max\left\{\frac{1}{2}\log\frac{4\mathsf{E}}{\sigma^2}, -\frac{1}{2}\log 2\pi\sigma^2\right\} - \frac{1}{2} - \frac{I\{x \geq 1\}}{2} \underbrace{\log(1+\varsigma^2 x)}_{\geq \log x + \log\varsigma^2}$$

$$
\begin{aligned}
&- \frac{I\{x<1\}}{2} \underbrace{\log(1+\varsigma^2 x)}_{\geq 0} + k_1 + \frac{k_2}{2\mathsf{E}} + \frac{x}{2\mathsf{E}} + I\{x \geq 1\} \cdot \frac{1}{2} \log x \\
&+ \underbrace{I\{x \geq 1\}}_{\leq 1} \cdot \frac{k_2}{2} + \underbrace{I\{x<1\}}_{\leq 1} \cdot \frac{k_2+1}{2} && (166)\\
&\leq \max\left\{\frac{1}{2}\log\frac{4\mathsf{E}}{\sigma^2}, -\frac{1}{2}\log 2\pi\sigma^2\right\} - \frac{1}{2} - I\{x \geq 1\}\cdot\frac{1}{2}\log\varsigma^2 \\
&+ k_1 + \frac{k_2}{2\mathsf{E}} + \frac{x}{2\mathsf{E}} + \frac{k_2}{2} + \frac{k_2+1}{2} && (167)\\
&\leq \max\left\{\frac{1}{2}\log\frac{4\mathsf{E}}{\sigma^2}, -\frac{1}{2}\log 2\pi\sigma^2\right\} - \min\left\{0, \frac{1}{2}\log\varsigma^2\right\} \\
&+ k_1 + \frac{k_2}{2\mathsf{E}} + \frac{x}{2\mathsf{E}} + k_2. && (168)
\end{aligned}
$$

Hence, we get

$$
\begin{aligned}
\mathsf{C}(\mathsf{E}) &\leq \mathsf{E}_{Q^*}\big[D\big(W(\cdot|X) \,\|\, R(\cdot)\big)\big] && (169)\\
&\leq \max\left\{\frac{1}{2}\log\frac{4\mathsf{E}}{\sigma^2}, -\frac{1}{2}\log 2\pi\sigma^2\right\} - \min\left\{0, \frac{1}{2}\log\varsigma^2\right\} \\
&\quad + k_1 + \frac{k_2}{2\mathsf{E}} + \frac{\mathsf{E}}{2\mathsf{E}} + k_2 && (170)
\end{aligned}
$$

and therefore

$$
\varlimsup_{\mathsf{E}\uparrow\infty} \frac{\mathsf{C}(\mathsf{E})}{\frac{1}{2}\log\mathsf{E}} \leq 1. \tag{171}
$$

Hence, we have shown that $\mathsf{C}_{\text{asy}}(\zeta) \triangleq \frac{1}{2}\log\zeta$ satisfies the conditions of Proposition 15. This proves our claim.

C The Gaussian Distribution Approximates the Poisson Distribution

In this appendix we will show that for large values of λ, a Gaussian distribution of mean λ and variance λ will approximate a Poisson distribution of mean λ.

Note that strictly speaking we have to compare the cumulative distribution functions (CDF) as a Poisson random variable is discrete, while a Gaussian random variable is continuous. To simplify the proof, however, we will use a trick to create a "continuous Poisson random variable". Let $T \sim \mathcal{P}o(\lambda)$ be a Poisson random variable with mean λ, and let $U \sim \mathcal{U}([0,1))$ be a random variable that is uniformly distributed on the interval $[0,1)$ and that is independent of T. We now define the "continuous Poisson random variable" T_{c} as

$$
T_{\text{c}} \triangleq T + U. \tag{172}
$$

Obviously, T_{c} is a continuous random variable with PDF

$$
f_{T_{\text{c}}}(t) = \begin{cases} e^{-\lambda}\frac{\lambda^{\lfloor t \rfloor}}{\lfloor t \rfloor !} & t \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{173}
$$

But also note that from T_{c} one can always retrieve the value of the Poisson random variable T by simply applying the flooring operation:

$$
T = \lfloor T_{\text{c}} \rfloor. \tag{174}
$$

To prove our claim of T approximating a Gaussian random variable for large λ, we will now show that

$$S \triangleq \frac{T_{\text{c}} - \lambda}{\sqrt{\lambda}} \tag{175}$$

will converge to a zero-mean, unit-variance Gaussian random variable if λ tends to infinity. Note that once λ gets very large, the influence of U will vanish, i.e., S will tend to $\frac{T-\lambda}{\sqrt{\lambda}}$.

Concretely, we will now show that the relative entropy between the PDF of S,

$$f_S(s) = \begin{cases} \sqrt{\lambda} e^{-\lambda} \frac{\lambda^{\lfloor \sqrt{\lambda} s + \lambda \rfloor}}{\lfloor \sqrt{\lambda} s + \lambda \rfloor !} & s \geq -\sqrt{\lambda} \\ 0 & \text{otherwise} \end{cases} \tag{176}$$

and the PDF of a zero-mean, unit-variance Gaussian random variable G,

$$f_G(s) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}, \quad s \in \mathbb{R} \tag{177}$$

tends to zero as $\lambda \uparrow \infty$:

$$\begin{aligned}
D(f_S \| f_G) &= \mathsf{E}\left[\log \frac{f_S(S)}{f_G(S)}\right] && (178)\\
&= \mathsf{E}\left[\log \frac{f_S\left(\frac{T_{\text{c}}-\lambda}{\sqrt{\lambda}}\right)}{f_G\left(\frac{T_{\text{c}}-\lambda}{\sqrt{\lambda}}\right)}\right] && (179)\\
&= \frac{1}{2}\log\lambda - \lambda + \mathsf{E}[\lfloor T_{\text{c}} \rfloor]\log\lambda - \mathsf{E}[\log(\lfloor T_{\text{c}} \rfloor !)] + \frac{1}{2}\log 2\pi + \frac{1}{2\lambda}\mathsf{E}\left[(T_{\text{c}} - \lambda)^2\right] && (180)\\
&= \frac{1}{2}\log 2\pi\lambda \underbrace{- \lambda + \mathsf{E}[T]\log\lambda - \mathsf{E}[\log(T!)]}_{=-H(T)} + \frac{1}{2\lambda}\mathsf{E}\left[(T + U - \lambda)^2\right] && (181)\\
&= \frac{1}{2}\log 2\pi\lambda - H(T) + \frac{1}{2} + \frac{1}{6\lambda} && (182)
\end{aligned}$$

where $H(T)$ denotes the entropy of T. From [11, Lemma 19] we know that

$$\varliminf_{\lambda \uparrow \infty} \left\{ H(T) - \frac{1}{2}\log 2\pi e\lambda \right\} \geq 0. \tag{183}$$

Hence, noting that relative entropy is nonnegative, we see that

$$0 \leq \varlimsup_{\lambda \uparrow \infty} D(f_S \| f_G) \leq \varlimsup_{\lambda \uparrow \infty} \frac{1}{6\lambda} = 0. \tag{184}$$

The claim now follows because the relative entropy is equal to zero if, and only if, its two arguments are identical.

Acknowledgment

The author is indebted to Frank Kschischang who pointed out the input-dependent Gaussian noise channel as an interesting model for analysis; to Amos Lapidoth for his superb coaching and many valuable inputs and comments; to Nick Letzepis who contributed the insight of Appendix C; and to Michèle Wigger for her critical reading and her help with the bounds at low SNR. Moreover, many thanks go also to the two anonymous reviewers for their very constructive comments.

References

[1] Joseph M. Kahn and John R. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265–298, February 1997.

[2] Sherman Karp, Robert M. Gagliardi, Steven E. Moran, and Lawrence B. Stotts, *Optical Channels*. Plenum Press, 1988.

[3] Steve Hranilovic and Frank 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.

[4] Stefan M. Moser, "Duality-based bounds on channel capacity," Ph.D. dissertation, ETH Zurich, October 2004, Diss. ETH No. 15769. [Online]. Available: http://moser.cm.nctu.edu.tw/

[5] Terence H. Chan, Steve Hranilovic, and Frank 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.

[6] Ahmed A. Farid and Steve 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, pp. 2416–2420.

[7] Ahmed A. Farid and Steve Hranilovic, "Channel capacity and non-uniform signalling for free-space optical intensity channels," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 9, pp. 1553–1563, December 2009.

[8] Amos Lapidoth, Stefan M. Moser, and Michèle A. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4449–4461, October 2009.

[9] Shlomo 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.

[10] David Brady and Sergio 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.

[11] Amos Lapidoth and Stefan M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 303–322, January 2009.

[12] Amos Lapidoth, Jeffrey H. Shapiro, Vinodh Venkatesan, and Ligong 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.

[13] Yurii M. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and Its Applications*, vol. 23, pp. 143–147, 1978.

[14] Mark 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.

[15] Aaron 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.

[16] Michael 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.

[17] Michael R. Frey, "Information capacity of the Poisson channel," *IEEE Transactions on Information Theory*, vol. 37, no. 2, pp. 244–256, March 1991.

[18] Shlomo Shamai (Shitz) and Amos 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.

[19] Israel Bar-David and Gideon 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.

[20] Amos Lapidoth and Stefan 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.

[21] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

[22] Amos Lapidoth and Stefan 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.

[23] Vyacheslav V. Prelov and Edward C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals," *IEEE Transactions on Information Theory*, vol. 39, no. 5, pp. 1728–1735, September 1993.

[24] Imre Csiszár and János Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, 1981.