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

Stefan M. Moser*

Corrected Version
October 8, 2010

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 the dual expression for channel capacity and the notion of *capacity-achieving input distributions that escape to infinity*. The lower bound is based on a 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.

Index Terms — Channel capacity, direct detection, Gaussian noise, high signal-to-noise ratio (SNR), low signal-to-noise ratio (SNR), optical communication.

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.

*S. M. Moser is with the Department of Electrical Engineering at National Chiao Tung University (NCTU) in Hsinchu, Taiwan. This work has been partially sponsored by the MediaTek Research Center at National Chiao Tung University, Taiwan, and by the Industrial Technology Research Institute, Zhudong, Taiwan. Parts of this work have been published in S. M. Moser's Ph.D. thesis.

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 with the available power [3]–[8], while the capacity of the Poisson channel only grows logarithmically with the square root of the power [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 like the Poisson channel.

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

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

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, $Z_0 \sim \mathcal{N}_{\mathbb{R}}\big(0, \sigma^2\big)$ is a zero-mean, variance-σ^2 Gaussian random variable describing the independent noise, and $Z_1 \sim \mathcal{N}_{\mathbb{R}}(0,1)$ is a zero-mean, unit-variance Gaussian random variable describing the dependent noise, $Z_0 \perp\!\!\!\perp Z_1$. Note that without loss of generality we assume the input to be scaled such that Z_1 can be normalized to be of unit-variance.

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

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

[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].

and an average-power constraint by

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

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*

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

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].

Proposition 1. *Consider a channel*[2] $W(\cdot|\cdot)$ *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 briefly reviewed in Section 5.2.

The low-power results are 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

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

contains the bounds on capacity that are valid at high SNR, and Section 3 describes the low-SNR results. The derivations are then given in Section 4 (high-power lower bounds), Section 5 (asymptotic high-power upper bounds), and Section 6 (low-power results). 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*. We will conclude in Section 7.

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, $\mathcal{E}$ 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 $x \in \mathcal{X}$, $W(\cdot|x)$ 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 $\mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$ we denote a real Gaussian distribution with mean μ and variance σ^2. We write $Z_1 \perp\!\!\!\perp Z_2$ to express that the random variables Z_1 and Z_2 are statistically independent. All rates specified in this paper are in nats per channel use, and all logarithms are natural logarithms.

Finally, we give the following definitions.

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[3]

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

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

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

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 $\operatorname{erf}(\cdot)$*:*

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

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

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

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,[4] 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 $\frac{1}{3} \leq \alpha \leq 1$, which includes the situation with only a peak-power constraint $\alpha = 1$. And finally, in the third case we look at the situation with only an average-power constraint.

We begin with the first case.

Theorem 4. *The channel capacity $\mathsf{C}(\mathsf{A}, \mathcal{E})$ of a channel with conditional PDF* (1) *and under the input constraints* (3) *and* (4)*, where the ratio $\alpha = \frac{\mathcal{E}}{\mathsf{A}}$ lies in $\left(0, \frac{1}{3}\right)$, is bounded as follows:*

$$\begin{aligned} \mathsf{C}(\mathsf{A}, \mathcal{E}) \geq{}& \frac{1}{2}\log \mathsf{A} - \frac{1}{2}\log 2\pi e - (1-\alpha)\mu - \log\left(\frac{1}{2} - \alpha\mu\right) \\ & - e^{\mu}\left(\frac{1}{2} - \alpha\mu\right)\left(2\sqrt{\frac{\sigma^2}{\mathsf{A}}}\arctan\left(\sqrt{\frac{\mathsf{A}}{\sigma^2}}\right) + \log\left(1 + \frac{\sigma^2}{\mathsf{A}}\right)\right) \\ & + \frac{1}{2}\log\left(1 + \frac{2}{\mathcal{E}}\right) + \sqrt{\mathcal{E}(2+\mathcal{E})} - \mathcal{E} - 1; \end{aligned} \tag{13}$$

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

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

$$\frac{1}{2\mu} - \frac{e^{-\mu}}{\sqrt{\mu}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)} = \alpha. \tag{15}$$

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.

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

$$\lim_{\mathsf{A}\uparrow\infty}\left\{\mathsf{C}(\mathsf{A}, \alpha\mathsf{A}) - \frac{1}{2}\log \mathsf{A}\right\} = -\frac{1}{2}\log 2\pi e - (1-\alpha)\mu - \log\left(\frac{1}{2} - \alpha\mu\right), \quad 0 < \alpha < \frac{1}{3} \tag{16}$$

where μ is defined as above to be the solution to (15)*.*

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

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

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

$$\mathsf{C}(\mathsf{A}, \mathcal{E}) \geq \frac{1}{2} \log \mathsf{A} - \frac{1}{2} \log \frac{\pi e}{2} - \sqrt{\frac{\sigma^2}{\mathsf{A}}} \arctan\left(\sqrt{\frac{\mathsf{A}}{\sigma^2}}\right) - \frac{1}{2} \log\left(1 + \frac{\sigma^2}{\mathsf{A}}\right) + \frac{1}{2} \log\left(1 + \frac{6}{\mathsf{A}}\right) + \sqrt{\frac{\mathsf{A}}{3}\left(2 + \frac{\mathsf{A}}{3}\right)} - \frac{\mathsf{A}}{3} - 1; \tag{17}$$

$$\mathsf{C}(\mathsf{A}, \alpha\mathsf{A}) \leq \frac{1}{2} \log \mathsf{A} - \frac{1}{2} \log \frac{\pi e}{2} + o_{\mathsf{A}}(1). \tag{18}$$

Here 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

$$\lim_{\mathsf{A} \uparrow \infty} \left\{\mathsf{C}(\mathsf{A}, \alpha\mathsf{A}) - \frac{1}{2} \log \mathsf{A}\right\} = -\frac{1}{2} \log \frac{\pi e}{2}, \qquad \frac{1}{3} \leq \alpha \leq 1. \tag{19}$$

The bounds of Theorems 4 and 5 are depicted in Figure 1 for different values of α.

Remark 6. For $\alpha \uparrow \frac{1}{3}$ the solution μ to (15) tends to zero. If in (14) and (13) μ is chosen to be zero and α to be $\frac{1}{3}$, then (14) and (13) coincide with (18) and (17), 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 $\alpha \in \left[\frac{1}{3}, 1\right]$.

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

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

$$\mathsf{C}(\mathcal{E}) \geq \frac{1}{2} \log \mathcal{E} - \sqrt{\frac{\pi\sigma^2}{2\mathcal{E}}} + \frac{1}{2} \log\left(1 + \frac{2}{\mathcal{E}}\right) + \sqrt{\mathcal{E}(2 + \mathcal{E})} - \mathcal{E} - 1; \tag{20}$$

$$\mathsf{C}(\mathcal{E}) \leq \frac{1}{2} \log \mathcal{E} + o_{\mathcal{E}}(1). \tag{21}$$

Here the term $o_{\mathcal{E}}(1)$ tends to zero as $\mathcal{E} \uparrow \infty$.

The asymptotic expansion for the channel capacity is

$$\lim_{\mathcal{E} \uparrow \infty} \left\{\mathsf{C}(\mathcal{E}) - \frac{1}{2} \log \mathcal{E}\right\} = 0. \tag{22}$$

The bounds of Theorem 8 are shown in Figure 2 for various values of σ^2.

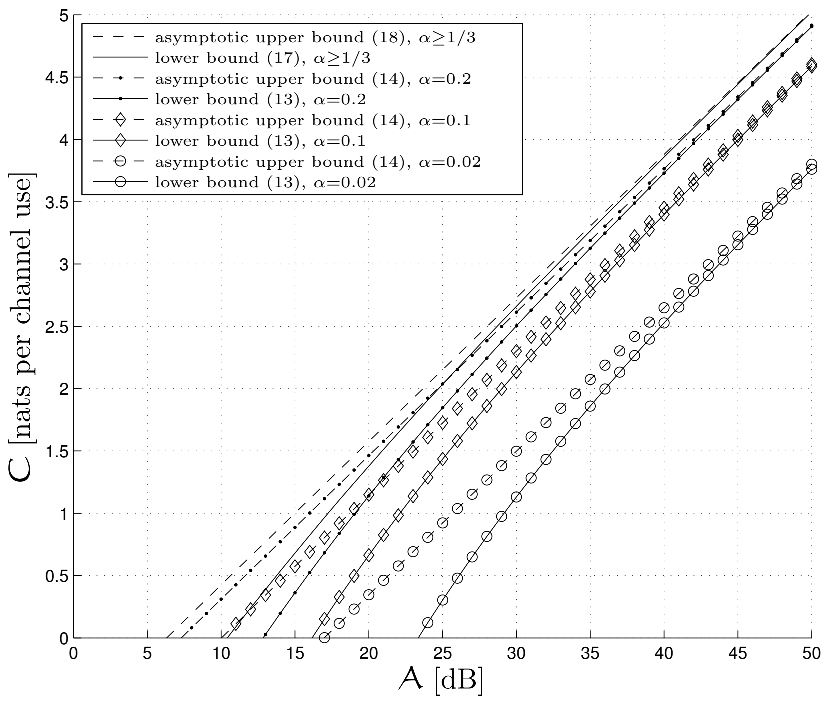

Figure 1: This plot depicts the firm lower bounds (13) and (17) (valid for all values of A) and the asymptotic upper bounds (14) and (18) (valid only in the limit when $\mathsf{A} \uparrow \infty$) on the capacity of the channel model (1) under an average- and a peak-power constraint with average-to-peak-power ratio α. For $\alpha \geq \frac{1}{3}$ (including the case of only a peak-power constraint $\alpha = 1$) the bounds do not depend on α. The upper bounds do not depend on the noise variance σ^2; for the lower bounds it is assumed $\sigma^2 = 2$. The horizontal axis is measured in dB where $\mathsf{A}\,[\text{dB}] = 10 \log_{10} \mathsf{A}$.

Remark 9. If we keep $\mathcal{E}$ fixed and let $\mathsf{A} \uparrow \infty$, we get $\alpha \downarrow 0$. For $\alpha \ll 1$ the solution μ to (15) tends to $\frac{1}{2\alpha} \gg 1$ which makes sure that (14) tends to (21). To see this note that for $\mu \gg 1$ we can approximate $\operatorname{erf}\left(\sqrt{\mu}\right) \approx 1$. Then we get from (15) that

$$\frac{1}{2} - \alpha\mu \approx \sqrt{\frac{\mu}{\pi}} e^{-\mu}. \tag{23}$$

Using this together with

$$\frac{1}{2} \log \mathsf{A} = \frac{1}{2} \log \mathcal{E} - \frac{1}{2} \log \alpha \tag{24}$$

$$\approx \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log 2\mu \tag{25}$$

we get from (14)

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

$$\approx \frac{1}{2} \log \mathcal{E} + \frac{1}{2} \log 2\mu - \frac{1}{2} \log 2\pi e - \mu + \frac{1}{2} - \log \sqrt{\frac{\mu}{\pi}} e^{-\mu} \tag{26}$$

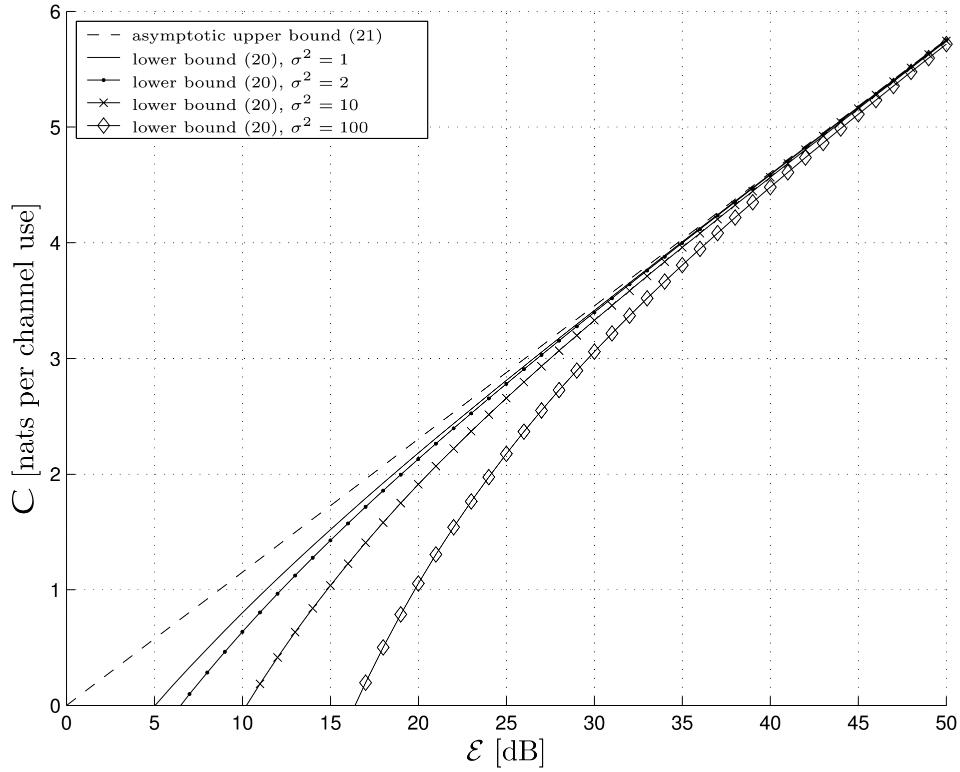

Figure 2: This plot depicts the firm lower bound (20) (valid for all values of $\mathcal{E}$) and the asymptotic upper bound (21) (valid only in the limit when $\mathcal{E} \uparrow \infty$) on the capacity of the channel model (1) with average-power constraint (4). The lower bound is shown for various different choices of the noise variance σ^2. The asymptotic upper bound does not depend on σ^2. The horizontal axis is measured in dB where $\mathcal{E}\,[\text{dB}] = 10\log_{10}\mathcal{E}$.

$$= \frac{1}{2}\log \mathcal{E}. \tag{27}$$

Similarly, (13) converges to (20) which can be seen by additionally noting that

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

$$\approx \sqrt{\frac{\mu}{\pi}}\left(2\sqrt{\frac{\alpha\sigma^2}{\mathcal{E}}}\cdot\frac{\pi}{2}+\underbrace{\frac{\alpha\sigma^2}{\mathcal{E}}}_{\ll\sqrt{\frac{\alpha\sigma^2}{\mathcal{E}}}}\right) \tag{28}$$

$$\approx \sqrt{\frac{\mu\alpha\pi\sigma^2}{\mathcal{E}}} \approx \sqrt{\frac{\pi\sigma^2}{2\mathcal{E}}}. \tag{29}$$

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 and the case where the average-power constraint is inactive.

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

$$\lim_{\mathsf{A}\downarrow 0} \frac{\mathsf{C}(\mathsf{A}, \alpha\mathsf{A})}{\mathsf{A}^2} = \alpha(1-\alpha)\frac{1+2\sigma^2}{4\sigma^4}. \tag{30}$$

In the case where the ratio $\alpha = \frac{\mathcal{E}}{\mathsf{A}}$ 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

$$\lim_{\mathsf{A}\downarrow 0} \frac{\mathsf{C}(\mathsf{A}, \alpha\mathsf{A})}{\mathsf{A}^2} = \frac{1+2\sigma^2}{16\sigma^4}. \tag{31}$$

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 $\alpha = \frac{1}{3}$.

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 $Q(\cdot)$:

$$\mathsf{C} = \sup_{Q(\cdot)} I(Q, W) \geq I(Q, W)\big|_{\text{for a specific } Q(\cdot)}. \tag{32}$$

This leads to a natural lower bound on capacity.

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

To avoid this problem we lower-bound $h(Y)$ in terms of $h(X)$, 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 $\mathsf{E}_Q[X] = \mathcal{E}$. Then*

$$h(Y) \geq h(X) + f_{\text{low}}(\mathcal{E}) > h(X) \tag{33}$$

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

$$\lim_{\mathcal{E}\uparrow\infty} f_{\text{low}}(\mathcal{E}) = 0 \tag{34}$$

given by

$$f_{\text{low}}(\mathcal{E}) \triangleq \frac{1}{2}\log\left(1+\frac{2}{\mathcal{E}}\right) - \mathcal{E} - 1 + \sqrt{\mathcal{E}(2+\mathcal{E})}, \qquad \mathcal{E} \geq 0. \tag{35}$$

Proof. See Appendix A. □

4.3 Proof of the Lower Bound (13)

Using (32) and Proposition 11 we get

$$\begin{aligned}
\mathsf{C} &\geq I(Q,W)\big|_{\text{any specified } Q(\cdot)} && (36)\\
&= h(Y) - h(Y|X) && (37)\\
&\geq h(X) + f_{\text{low}}(\mathcal{E}) - h(Y|X) && (38)\\
&= h(X) + f_{\text{low}}(\mathcal{E}) - \frac{1}{2}\mathsf{E}\left[\log 2\pi e(\sigma^2 + X)\right] && (39)\\
&= h(X) + f_{\text{low}}(\mathcal{E}) - \frac{1}{2}\log 2\pi e - \frac{1}{2}\mathsf{E}[\log X] - \frac{1}{2}\mathsf{E}\left[\log\left(1+\frac{\sigma^2}{X}\right)\right]. && (40)
\end{aligned}$$

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]:

$$Q(x) \triangleq \begin{cases} \frac{\sqrt{\mu}}{\sqrt{\mathsf{A}\pi x}\,\text{erf}(\sqrt{\mu})} \cdot e^{-\frac{\mu}{\mathsf{A}}x} & 0 \leq x \leq \mathsf{A} \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

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

$$\mathsf{E}[X] = \frac{\mathsf{A}}{2\mu} - \frac{\mathsf{A}e^{-\mu}}{\sqrt{\mu}\sqrt{\pi}\text{erf}\left(\sqrt{\mu}\right)} \stackrel{!}{=} \alpha\mathsf{A} \tag{42}$$

i.e., μ is the solution to (15).

Then we have

$$h(X) = \frac{1}{2}\log\frac{\mathsf{A}}{\mu} + \log\sqrt{\pi}\text{erf}\left(\sqrt{\mu}\right) + \alpha\mu + \frac{1}{2}\mathsf{E}[\log X] \tag{43}$$

and

$$\mathsf{E}\left[\log\left(1+\frac{\sigma^2}{X}\right)\right] = \int_0^{\mathsf{A}} \log\left(1+\frac{\sigma^2}{x}\right)\cdot\frac{\sqrt{\mu}}{\sqrt{\mathsf{A}\pi x}\cdot\operatorname{erf}\left(\sqrt{\mu}\right)}\underbrace{e^{-\frac{\mu}{\mathsf{A}}x}}_{\leq 1}\,\mathrm{d}x \tag{44}$$

$$\leq \int_0^{\mathsf{A}} \log\left(1+\frac{\sigma^2}{x}\right)\cdot\frac{\sqrt{\mu}}{\sqrt{\mathsf{A}\pi x}\cdot\operatorname{erf}\left(\sqrt{\mu}\right)}\,\mathrm{d}x \tag{45}$$

$$= \frac{2\pi\sqrt{\frac{\sigma^2}{\mathsf{A}}}\sqrt{\mu}+2\sqrt{\mu}\log\left(1+\frac{\sigma^2}{\mathsf{A}}\right)-4\sqrt{\frac{\sigma^2}{\mathsf{A}}}\sqrt{\mu}\arctan\left(\sqrt{\frac{\sigma^2}{\mathsf{A}}}\right)}{\sqrt{\pi}\cdot\operatorname{erf}\left(\sqrt{\mu}\right)} \tag{46}$$

$$= \frac{4\sqrt{\mu}\sqrt{\frac{\sigma^2}{\mathsf{A}}}\arctan\left(\sqrt{\frac{\mathsf{A}}{\sigma^2}}\right)+2\sqrt{\mu}\log\left(1+\frac{\sigma^2}{\mathsf{A}}\right)}{\sqrt{\pi}\cdot\operatorname{erf}\left(\sqrt{\mu}\right)} \tag{47}$$

where we have used that

$$\arctan\left(\frac{1}{\xi}\right) = \frac{\pi}{2} - \arctan(\xi), \qquad \xi \in \mathbb{R}. \tag{48}$$

Using (47) and (43) in (40) completes our proof.

4.4 Proof of the Lower Bounds (17) and (20)

As noted in Remarks 6 and 9, (17) and (20) 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 (41):

$$Q(x) \triangleq \begin{cases} \frac{1}{\sqrt{4\mathsf{A}x}} & 0 \leq x \leq \mathsf{A} \\ 0 & \text{otherwise} \end{cases} \tag{49}$$

and

$$Q(x) \triangleq \begin{cases} \frac{1}{\sqrt{2\pi\mathcal{E}x}}e^{-\frac{x}{2\mathcal{E}}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{50}$$

respectively.

For (49) we get

$$\mathsf{E}[X] = \frac{\mathsf{A}}{3} \tag{51}$$

$$h(X) = \log\mathsf{A} - 1 + \log 2 \tag{52}$$

$$\mathsf{E}\left[\log(\sigma^2+X)\right] = \log(\mathsf{A}+\sigma^2) - 2 + 2\sqrt{\frac{\sigma^2}{\mathsf{A}}}\arctan\left(\sqrt{\frac{\mathsf{A}}{\sigma^2}}\right) \tag{53}$$

which, when plugged into (39), yields (17).

For (50) we get

$$h(X) = \log \mathcal{E} - \frac{\gamma}{2} + \frac{1}{2}\log \pi e \tag{54}$$

$$\mathsf{E}[\log X] = \log \mathcal{E} - \gamma - \log 2 \tag{55}$$

$$\mathsf{E}\left[\log\left(1+\frac{\sigma^2}{X}\right)\right] = \int_0^\infty \frac{1}{\sqrt{2\pi\mathcal{E}x}} \underbrace{e^{-\frac{x}{2\mathcal{E}}}}_{\leq 1} \log\left(1+\frac{\sigma^2}{x}\right) \mathrm{d}x \tag{56}$$

$$\leq \int_0^\infty \frac{1}{\sqrt{2\pi\mathcal{E}x}} \log\left(1+\frac{\sigma^2}{x}\right) \mathrm{d}x \tag{57}$$

$$= \sqrt{\frac{2\pi\sigma^2}{\mathcal{E}}} \tag{58}$$

which, when plugged into (40), yields (20).

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.,

$$\mathsf{C} \leq \mathsf{E}_{Q^*}\left[D\big(W(\cdot|X)\big\|R(\cdot)\big)\right]. \tag{59}$$

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 (59). 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

$$\mathsf{E}_{Q^*}[o_X(1)] = o_{\mathsf{A}}(1). \tag{60}$$

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*

$$\mathsf{C}(\mathsf{A}, \mathcal{E}) \triangleq \sup_{Q(\cdot)} I(Q, W), \qquad \mathsf{A}, \mathcal{E} \geq \inf_{x\in\mathcal{X}} g(x) \tag{61}$$

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

$$Q\big(\{x \in \mathcal{X} \colon g(x) > \mathsf{A}\}\big) = 0 \tag{62}$$

and

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

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 a unique input distribution* $Q^*_{\mathsf{A},\mathcal{E}}(\cdot)$ *that achieves the supremum in the definition of the capacity-cost function as given in* (61)*. Similarly, for the situation with only an average-power constraint, a unique capacity-achieving input distribution* $Q^*_{\mathcal{E}}(\cdot)$ *exists.*

Proof. See [5]. □

We now will briefly review the notion of *input distributions that escape to infinity.* This notion is important because we can show that for most channels of interest, the capacity-achieving input distribution must escape to infinity. In fact, not only the capacity-achieving input distributions escape to infinity: every input distribution that achieves a mutual information with the same asymptotic growth in the cost as the capacity must 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.[5]

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

$$\alpha \triangleq \frac{\mathcal{E}}{\mathsf{A}} \tag{64}$$

we say that a family of input distributions

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

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

$$\lim_{\mathsf{A}\uparrow\infty} Q_{\mathsf{A},\alpha\mathsf{A}}\bigl(\{x \in \mathcal{X}\colon g(x) \leq \mathsf{A}_0\}\bigr) = 0. \tag{66}$$

Based on this definition, 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.

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

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

[5] 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 $g(x) = x$, $\forall\, x \geq 0$.

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

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

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

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

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

Proof. See [22, Sec. VII.C.3]. □

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), $\{Q^*_{\mathcal{E}}(\cdot)\}_{\mathcal{E}\geq 0}$ *escapes to infinity.*

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

$$\mathsf{C}_{\text{asy}}(\mathsf{A})=\frac{1}{2}\log\mathsf{A} \tag{70}$$

satisfies both conditions (67) and (68) 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.,*

$$f(x)=o_x(1). \tag{71}$$

Assume that f *is bounded. Then*

$$\lim_{\mathsf{A}\uparrow\infty}\mathsf{E}_{Q_{\mathsf{A},\alpha\mathsf{A}}}[f(X)]=0. \tag{72}$$

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

$$|f(\mathsf{A})|<\frac{\epsilon}{2}. \tag{73}$$

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

$$\lim_{\mathsf{A}\uparrow\infty} \int_0^{\mathsf{A}_1} |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x = 0. \tag{74}$$

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

$$\int_0^{\mathsf{A}_1} |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x < \frac{\epsilon}{2}. \tag{75}$$

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

$$\begin{aligned}
\Bigl|\mathsf{E}_{Q_{\mathsf{A},\alpha\mathsf{A}}}[f(X)]\Bigr| &\leq \mathsf{E}_{Q_{\mathsf{A},\alpha\mathsf{A}}}[|f(X)|] && (76)\\
&= \int_0^{\infty} |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x && (77)\\
&= \int_0^{\mathsf{A}_1} |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x + \int_{\mathsf{A}_1}^{\infty} |f(x)|\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x && (78)\\
&< \frac{\epsilon}{2} + \int_{\mathsf{A}_1}^{\infty} \frac{\epsilon}{2}\, Q_{\mathsf{A},\alpha\mathsf{A}}(x)\,\mathrm{d}x && (79)\\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. && (80)
\end{aligned}$$

Here the first inequality follows from Jensen's inequality and the convexity of $|\cdot|$; (79) follows from (73) and (75); 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 $R(\cdot)$:

$$R(y) = \begin{cases} \frac{2p}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} & y < 0\\ \frac{(1-2p)\sqrt{\mu}}{\sqrt{\mathsf{A}(1+\delta)y}\sqrt{\pi}\,\mathrm{erf}(\sqrt{\mu})} e^{-\frac{y\mu}{\mathsf{A}(1+\delta)}} & 0 \leq y \leq \mathsf{A}(1+\delta)\\ \frac{p}{\sqrt{2\pi}\mathcal{Q}(\mathsf{A}(1+\delta))} e^{-\frac{y^2}{2}} & y > \mathsf{A}(1+\delta) \end{cases} \tag{81}$$

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

$$\frac{\sqrt{\mu}}{\sqrt{\mathsf{A}(1+\delta)y}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)} e^{-\frac{y\mu}{\mathsf{A}(1+\delta)}} \tag{82}$$

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.[6] 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.

[6] Compare with (41).

With this choice we get

$$
\begin{aligned}
D\bigl(W(\cdot|x)\big\|R(\cdot)\bigr) = {}& -\frac{1}{2}\log 2\pi e(\sigma^2+x) \underbrace{- \int_{-\infty}^{0} \log\left(\frac{2p}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\right) W(y|x)\,\mathrm{d}y}_{c_1(x)} \\
& \underbrace{- \int_{0}^{\mathsf{A}(1+\delta)} \log\left(\frac{(1-2p)\sqrt{\mu}}{\sqrt{\mathsf{A}(1+\delta)y}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)} e^{-\frac{y\mu}{\mathsf{A}(1+\delta)}}\right) W(y|x)\,\mathrm{d}y}_{c_2(x)} \\
& \underbrace{- \int_{\mathsf{A}(1+\delta)}^{\infty} \log\left(\frac{p}{\sqrt{2\pi}\,\mathcal{Q}\left(\mathsf{A}(1+\delta)\right)} e^{-\frac{y^2}{2}}\right) W(y|x)\,\mathrm{d}y}_{c_3(x)}. \qquad (83)
\end{aligned}
$$

We evaluate each term separately:

$$
\begin{aligned}
c_1(x) &= \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\log\frac{\sqrt{2\pi}}{2p} + \frac{\sigma^2+x+x^2}{2}\mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right) \\
&\quad - \frac{x}{2}\sqrt{\frac{\sigma^2+x}{2\pi}}\,e^{-\frac{x^2}{2(\sigma^2+x)}} && (84) \\
&= o_x(1). && (85)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
c_3(x) &= \mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sqrt{\sigma^2+x}}\right)\log\frac{\mathcal{Q}\left(\mathsf{A}(1+\delta)\right)\sqrt{2\pi}}{p} \\
&\quad + \frac{\bigl(\mathsf{A}(1+\delta)+x\bigr)\sqrt{\sigma^2+x}}{2\sqrt{2\pi}}\,e^{-\frac{(\mathsf{A}(1+\delta)-x)^2}{2(\sigma^2+x)}} \\
&\quad + \frac{\sigma^2+x+x^2}{2}\mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sqrt{\sigma^2+x}}\right) && (86) \\
&\le \sqrt{\frac{\sigma^2+\mathsf{A}}{2\pi\delta^2\mathsf{A}^2}}\,e^{-\frac{\delta^2\mathsf{A}^2}{2(\sigma^2+\mathsf{A})}}\left|\log\frac{\mathcal{Q}\left(\mathsf{A}(1+\delta)\right)\sqrt{2\pi}}{p}\right| \\
&\quad + \frac{(2\mathsf{A}+\mathsf{A}\delta)\sqrt{\sigma^2+\mathsf{A}}}{2\sqrt{2\pi}}\,e^{-\frac{\mathsf{A}^2\delta^2}{2(\sigma^2+\mathsf{A})}} \\
&\quad + \frac{\sigma^2+\mathsf{A}+\mathsf{A}^2}{2}\sqrt{\frac{\sigma^2+\mathsf{A}}{2\pi\delta^2\mathsf{A}^2}}\,e^{-\frac{\delta^2\mathsf{A}^2}{2(\sigma^2+\mathsf{A})}} && (87) \\
&= o_{\mathsf{A}}(1) && (88)
\end{aligned}
$$

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

$$
\frac{1}{\sqrt{2\pi z^2}}e^{-\frac{z^2}{2}}\left(1-\frac{1}{z^2}\right) < \mathcal{Q}(z) < \frac{1}{\sqrt{2\pi z^2}}e^{-\frac{z^2}{2}} \qquad (89)
$$

such that

$$
\begin{aligned}
\mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sqrt{\sigma^2+x}}\right) &< \sqrt{\frac{\sigma^2+x}{2\pi(\mathsf{A}+\delta\mathsf{A}-x)^2}}\,e^{-\frac{(\mathsf{A}+\delta\mathsf{A}-x)^2}{2(\sigma^2+x)}} && (90) \\
&\le \sqrt{\frac{\sigma^2+\mathsf{A}}{2\pi\delta^2\mathsf{A}^2}}\,e^{-\frac{\delta^2\mathsf{A}^2}{2(\sigma^2+\mathsf{A})}}. && (91)
\end{aligned}
$$

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

$$
\begin{aligned}
c_2(x) &= \underbrace{\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right) - \mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sqrt{\sigma^2+x}}\right)\right)}_{\leq 1} \\
&\qquad \cdot \underbrace{\left(\frac{1}{2}\log \mathsf{A} + \log \frac{\sqrt{1+\delta}\sqrt{\pi}\mathrm{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}}\right)}_{\geq 0 \text{ for } \mathsf{A} \geq \frac{\mu}{\pi\mathrm{erf}^2(\sqrt{\mu})}} \\
&\quad + \frac{1}{2}\int_0^{\mathsf{A}(1+\delta)} \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}} \,\mathrm{d}y \\
&\quad + \frac{\mu\sqrt{\sigma^2+x}}{\mathsf{A}(1+\delta)\sqrt{2\pi}} \left(e^{-\frac{x^2}{2(\sigma^2+x)}} \underbrace{- e^{-\frac{(\mathsf{A}(1+\delta)-x)^2}{2(\sigma^2+x)}}}_{\leq 0}\right) \\
&\quad + \frac{x\mu}{\mathsf{A}(1+\delta)} \underbrace{\left(1 - \mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sqrt{\sigma^2+x}}\right) - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right)}_{\leq 1} && (92) \\
&\leq \frac{1}{2}\log \mathsf{A} + \log \frac{\sqrt{1+\delta}\sqrt{\pi}\mathrm{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}} + \frac{1}{2}\int_0^{\mathsf{A}(1+\delta)} \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}} \,\mathrm{d}y \\
&\quad + \frac{1}{\mathsf{A}} \cdot \frac{\mu\sqrt{\sigma^2+x}}{(1+\delta)\sqrt{2\pi}} e^{-\frac{x^2}{2(\sigma^2+x)}} + \frac{x\mu}{\mathsf{A}(1+\delta)}. && (93)
\end{aligned}
$$

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

$$
\begin{aligned}
&\left.\int_0^{\mathsf{A}(1+\delta)} \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}} \,\mathrm{d}y \right|_{x\geq 1} \\
&\quad = \underbrace{\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right) - \mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sqrt{\sigma^2+x}}\right)\right)}_{\leq 1} \underbrace{\log x}_{\geq 0} \\
&\qquad + \frac{x}{\sqrt{2\pi(\sigma^2+x)}} \int_{-1}^{\frac{\mathsf{A}(1+\delta)-x}{x}} \underbrace{\log(1+\tilde{y})}_{\leq \tilde{y}} e^{-\frac{x^2\tilde{y}^2}{2(\sigma^2+x)}} \,\mathrm{d}\tilde{y} && (94) \\
&\quad \leq \log x + \frac{x}{\sqrt{2\pi(\sigma^2+x)}} \int_{-1}^{\frac{\mathsf{A}(1+\delta)-x}{x}} \tilde{y} e^{-\frac{x^2\tilde{y}^2}{2(\sigma^2+x)}} \,\mathrm{d}\tilde{y} && (95) \\
&\quad = \log x + \sqrt{\frac{\sigma^2+x}{2\pi x^2}} \left(e^{-\frac{x^2}{2(\sigma^2+x)}} - \underbrace{e^{-\frac{(\mathsf{A}(1+\delta)-x)^2}{2(\sigma^2+x)}}}_{\geq 0}\right) && (96) \\
&\quad \leq \log x + \sqrt{\frac{\sigma^2+x}{2\pi x^2}} \cdot e^{-\frac{x^2}{2(\sigma^2+x)}}. && (97)
\end{aligned}
$$

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

$$
\begin{aligned}
&\left.\int_0^{\mathsf{A}(1+\delta)} \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}} \,\mathrm{d}y \right|_{x< 1} \\
&\quad \leq \int_0^{\mathsf{A}(1+\delta)} \frac{y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}} \,\mathrm{d}y && (98)
\end{aligned}
$$

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

$$+ x\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right) - \underbrace{\mathcal{Q}\left(\frac{\mathsf{A}(1+\delta)-x}{\sqrt{\sigma^2+x}}\right)}_{\geq 0}\right) \tag{99}$$

$$\leq \sqrt{\frac{\sigma^2+x}{2\pi}} e^{-\frac{x^2}{2(\sigma^2+x)}} + x\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right). \tag{100}$$

Hence, because (100) is bounded and from (97) we have

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

Plugging this into (93) yields

$$c_2(x) \leq \frac{1}{2}\log \mathsf{A} + \log\frac{\sqrt{1+\delta}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}} + \frac{1}{2}\log x + \frac{x\mu}{\mathsf{A}(1+\delta)} + o_x(1) + \frac{1}{\mathsf{A}}\cdot o_x(1). \tag{102}$$

Using all these results together with (83) and (7), we get

$$\mathsf{C} \leq \mathsf{E}_{Q^*}\left[D\big(W(\cdot|x)\big\|R(\cdot)\big)\right] \tag{103}$$

$$\leq \mathsf{E}_{Q^*}\Bigg[-\frac{1}{2}\log 2\pi e(\sigma^2+X) + o_X(1) + \frac{1}{2}\log\mathsf{A} + \log\frac{\sqrt{1+\delta}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}}$$

$$+ \frac{1}{2}\log X + \frac{X\mu}{\mathsf{A}(1+\delta)} + o_X(1) + \frac{1}{\mathsf{A}}\cdot o_X(1) + o_{\mathsf{A}}(1)\Bigg] \tag{104}$$

$$= -\frac{1}{2}\log 2\pi e + \mathsf{E}_{Q^*}\Bigg[\underbrace{-\frac{1}{2}\log\left(1+\frac{\sigma^2}{X}\right)}_{o_X(1)}\Bigg] + \frac{1}{2}\log\mathsf{A} + \frac{1}{2}\log(1+\delta) - \frac{1}{2}\log\mu$$

$$+ \log\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right) - \log(1-2p) + \frac{\mathsf{E}_{Q^*}[X]\,\mu}{\mathsf{A}(1+\delta)} + \mathsf{E}_{Q^*}[o_X(1)] + \frac{1}{\mathsf{A}}\cdot\mathsf{E}_{Q^*}[o_X(1)]$$

$$+ o_{\mathsf{A}}(1) \tag{105}$$

$$\leq \frac{1}{2}\log\mathsf{A} - \frac{1}{2}\log 2\pi e + \frac{1}{2}\log(1+\delta) - \frac{1}{2}\log\mu + \log\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\mu}\right)$$

$$- \log(1-2p) + \frac{\alpha\mu}{1+\delta} + \mathsf{E}_{Q^*}[o_X(1)] + \frac{1}{\mathsf{A}}\cdot\mathsf{E}_{Q^*}[o_X(1)] + o_{\mathsf{A}}(1). \tag{106}$$

Finally, we use[7] Claim 17 and choose μ to be the solution to (15). The result now follows since p and δ are arbitrary.

5.4 Proof of the Upper Bounds (18) and (21)

As noted in Remarks 6 and 9, (18) and (21) 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 (18) we make the same choice (81), but with

$$\mu \triangleq \frac{1}{\mathsf{A}}. \tag{107}$$

[7]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 (107) we have

$$\log \frac{\sqrt{\pi}\operatorname{erf}\left(\sqrt{\mu}\right)}{\sqrt{\mu}} = \log 2 + o_{\mathsf{A}}(1) \tag{108}$$

and hence, continuing from (104), we get

$$\begin{aligned}
\mathsf{C} &\le \mathsf{E}_{Q^*}\Bigg[-\frac{1}{2}\log 2\pi e(\sigma^2+X) + \frac{1}{2}\log\mathsf{A} + \log\frac{\sqrt{1+\delta}\sqrt{\pi}\operatorname{erf}\left(\sqrt{\mu}\right)}{(1-2p)\sqrt{\mu}} \\
&\qquad + \frac{1}{2}\log X + \frac{X\mu}{\mathsf{A}(1+\delta)} + o_X(1) + o_{\mathsf{A}}(1)\Bigg] &(109)\\
&= \frac{1}{2}\log\mathsf{A} - \frac{1}{2}\log 2\pi e + \frac{1}{2}\log(1+\delta) + \log 2 - \log(1-2p) \\
&\qquad + \mathsf{E}_{Q^*}\Bigg[\underbrace{-\frac{1}{2}\log\left(1+\frac{\sigma^2}{X}\right)}_{o_X(1)}\Bigg] + \frac{\mathsf{E}_{Q^*}[X]}{\mathsf{A}^2(1+\delta)} + \mathsf{E}_{Q^*}[o_X(1)] + o_{\mathsf{A}}(1) &(110)\\
&\le \frac{1}{2}\log\mathsf{A} - \frac{1}{2}\log\frac{\pi e}{2} + \frac{1}{2}\log(1+\delta) - \log(1-2p) + \underbrace{\frac{1}{\mathsf{A}(1+\delta)}}_{o_{\mathsf{A}}(1)} \\
&\qquad + \mathsf{E}_{Q^*}[o_X(1)] + o_{\mathsf{A}}(1) &(111)
\end{aligned}$$

where in the last step we bounded $\mathsf{E}_{Q^*}[X] \le \mathsf{A}$. The result (18) now follows from (60) and because p and δ are arbitrary.

For (21) we choose

$$R(y) = \begin{cases} \frac{2p}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} & y < 0 \\ \frac{1-p}{\sqrt{2\pi\mathcal{E}y}} e^{-\frac{y}{2\mathcal{E}}} & y \ge 0 \end{cases} \tag{112}$$

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

$$\begin{aligned}
D\big(W(\cdot|x)\big\|R(\cdot)\big) &= -\frac{1}{2}\log 2\pi e(\sigma^2+x) \underbrace{- \int_{-\infty}^{0} \log\left(\frac{2p}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\right) W(y|x)\,\mathrm{d}y}_{c_1(x)} \\
&\qquad \underbrace{- \int_{0}^{\infty} \log\left(\frac{1-p}{\sqrt{2\pi\mathcal{E}y}}e^{-\frac{y}{2\mathcal{E}}}\right) W(y|x)\,\mathrm{d}y}_{c_2'(x)}. &(113)
\end{aligned}$$

From (85) we know that $c_1(x) = o_x(1)$. For $c_2'(x)$ we get

$$\begin{aligned}
c_2'(x) &= \left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right)\log\frac{\sqrt{2\pi\mathcal{E}}}{1-p} + \frac{1}{2\mathcal{E}}\sqrt{\frac{\sigma^2+x}{2\pi}}e^{-\frac{x^2}{2(\sigma^2+x)}} \\
&\qquad + \frac{x}{2\mathcal{E}}\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right) + \frac{1}{2}\int_0^\infty \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}}e^{-\frac{(y-x)^2}{2(\sigma^2+x)}}\,\mathrm{d}y &(114)\\
&\le \left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right)\log\frac{\sqrt{2\pi\mathcal{E}}}{1-p} + \frac{1}{2\mathcal{E}}\sqrt{\frac{\sigma^2+x}{2\pi}}e^{-\frac{x^2}{2(\sigma^2+x)}} \\
&\qquad + \frac{x}{2\mathcal{E}}\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right) + \frac{1}{2}\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right)\log x
\end{aligned}$$

$$+\frac{1}{2}\sqrt{\frac{\sigma^2+x}{2\pi x^2}}e^{-\frac{x^2}{2(\sigma^2+x)}} \tag{115}$$

$$\leq \log\frac{\sqrt{2\pi\mathcal{E}}}{1-p}+\frac{1}{2}\sqrt{\frac{\sigma^2+x}{2\pi}}e^{-\frac{x^2}{2(\sigma^2+x)}}+\frac{x}{2\mathcal{E}}+\frac{1}{2}\log x$$

$$-\frac{1}{2}\mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\log x+\frac{1}{2}\sqrt{\frac{\sigma^2+x}{2\pi x^2}}e^{-\frac{x^2}{2(\sigma^2+x)}} \tag{116}$$

$$=\log\frac{\sqrt{2\pi\mathcal{E}}}{1-p}+\frac{x}{2\mathcal{E}}+\frac{1}{2}\log x+o_x(1) \tag{117}$$

where the first inequality follows in an equivalent way as shown in (94)–(96), and the second inequality holds for large enough $\mathcal{E}$ such that $\log\frac{\sqrt{2\pi\mathcal{E}}}{1-p}>0$ and $\mathcal{E}>1$.

We continue with (113) and take the expectation over Q^*:

$$\mathsf{C}\leq \mathsf{E}_{Q^*}\left[D\big(W(\cdot|X)\big\|R(\cdot)\big)\right] \tag{118}$$

$$\leq \mathsf{E}_{Q^*}\left[-\frac{1}{2}\log 2\pi e(\sigma^2+X)+\log\frac{\sqrt{2\pi\mathcal{E}}}{1-p}+\frac{X}{2\mathcal{E}}+\frac{1}{2}\log X+o_X(1)\right] \tag{119}$$

$$=\mathsf{E}_{Q^*}\left[\frac{1}{2}\log\mathcal{E}-\frac{1}{2}-\log(1-p)+\frac{X}{2\mathcal{E}}\underbrace{-\frac{1}{2}\log\left(1+\frac{\sigma^2}{X}\right)}_{o_X(1)}+o_X(1)\right] \tag{120}$$

$$=\frac{1}{2}\log\mathcal{E}-\frac{1}{2}-\log(1-p)+\frac{\mathcal{E}}{2\mathcal{E}}+\mathsf{E}_{Q^*}[o_X(1)]. \tag{121}$$

Analogously to (60) we have $\mathsf{E}_{Q^*}[o_X(1)]=o_{\mathcal{E}}(1)$. 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*

$$I(X;Y)=\frac{1}{2}J(0)\mathsf{Var}(X)+o(\mathsf{A}^2), \tag{122}$$

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:

$$J(x)\triangleq\int_{\mathcal{Y}}\frac{\left(\frac{\mathrm{d}}{\mathrm{d}x}W(y|x)\right)^2}{W(y|x)}\,\mathrm{d}y. \tag{123}$$

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 (30) and under a peak-power constraint only (31).

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

$$J(x) = \frac{1 + 2\sigma^2 + 2x}{2(\sigma^2 + x)^2} \tag{124}$$

such that

$$J(0) = \frac{1 + 2\sigma^2}{2\sigma^4}. \tag{125}$$

Moreover, it is not difficult to see that

$$\max_{\substack{Q \text{ s.t.} \\ (3) \text{ and } (4) \\ \text{are satisfied}}} \mathsf{Var}(X) = \begin{cases} \mathcal{E}(\mathsf{A} - \mathcal{E}) = \alpha(1-\alpha)\mathsf{A}^2 & \text{if } \mathcal{E} < \frac{\mathsf{A}}{2}, \\ \frac{\mathsf{A}^2}{4} & \text{if } \mathcal{E} \geq \frac{\mathsf{A}}{2}. \end{cases} \tag{126}$$

The theorem is now established by combining (125) and (126) with (122) and the definition of channel capacity.

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

For the asymptotic upper bounds we relied on two concepts introduced in [20] and [22]. Firstly, a technique of using duality-based upper bounds on mutual information in order to upper-bound capacity (see Proposition 1), and secondly the notion of *input distributions that escape to infinity* (see Section 5.2) that allows us to compute asymptotic expectations over the unknown capacity-achieving input distribution.

The capacity of the optical intensity channel with input-dependent Gaussian noise has also be established for the asymptotic low-SNR situation where the peak- and average-power tend to zero with their ratio held constant.

It is interesting to compare these 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 very 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: 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 (30) and (31) are very similar to the low-power asymptotic capacities of the free-space optical intensity channel [8], especially for large values of σ^2. 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

$$\frac{1}{2} \leq \alpha \leq 1 \tag{127}$$

while here at high power we have

$$\frac{1}{3} \leq \alpha \leq 1. \tag{128}$$

However, note that while the former result holds true for all values of A and $\mathcal{E}$, 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 expected to be somewhere in between, varying with A and σ^2.

On the other hand, it is very interesting to observe that the asymptotic high-power results given in this paper turn out to be *identical* to the asymptotic capacity of the Poisson channel given in [11], i.e., not only the prelog factor $\frac{1}{2}$ is the same, but also the second term in the high-SNR expansion of capacity!

Intuitively, this correspondence can be understood by realizing that for large values of λ, the cumulative distribution function of a Poisson random variable with mean λ approximates the cumulative distribution function of a Gaussian random variable with mean λ and variance λ. We prove this statement in Appendix C. If we in addition recall that the capacity-achieving input distribution escapes to infinity (see Corollary 16), we hence see that the channel model (1) asymptotically for large SNR converges with the Poisson channel.

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

A A Proof of Proposition 11

Conditional on $X = x$, Y can be written as $Y = Y_0 + Y_1$, where $Y_0 \sim \mathcal{N}_{\mathbb{R}}\big(0, \sigma^2\big)$ and $Y_1 \sim \mathcal{N}_{\mathbb{R}}(x, x)$, $Y_0 \perp\!\!\!\perp Y_1$. By the fact that conditioning reduces entropy we therefore have

$$h(Y) = h(Y_0 + Y_1) \geq h(Y_0 + Y_1 \,|\, Y_0) = h(Y_1 \,|\, Y_0) = h(Y_1). \tag{129}$$

Hence, we can restrict ourselves to the case where $\sigma^2 = 0$.

The proof of (33) 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 $\mathsf{E}_Q[X] = \mathcal{E}$. Let $Q_{\mathrm{E}}(\cdot)$ be an exponential probability distribution of mean $\mathcal{E}$

$$Q_{\mathrm{E}}(x) \triangleq \frac{1}{\mathcal{E}} e^{-\frac{x}{\mathcal{E}}}, \qquad x \geq 0. \tag{130}$$

If $Q_{\mathrm{E}}(\cdot)$ is used as input distribution to our channel, then the according output distribution is

$$(Q_{\mathrm{E}}W)(y) = \frac{1}{\sqrt{\mathcal{E}(\mathcal{E}+2)}} \exp\left(\frac{\sqrt{\mathcal{E}}\, y - \sqrt{\mathcal{E}+2}\, |y|}{\sqrt{\mathcal{E}}}\right), \quad y \in \mathbb{R}. \tag{131}$$

By the data processing theorem we now obtain:

$$D\bigl(Q \| Q_{\mathrm{E}}\bigr) \geq D\bigl((QW) \| (Q_{\mathrm{E}}W)\bigr) \tag{132}$$

where $(QW)(\cdot)$ denotes the corresponding output distribution of our channel when an input of law $Q(\cdot)$ is used. The first inequality in (33) in the proposition's statement now follows by evaluating the left-hand side of (132):

$$D\bigl(Q \| Q_{\mathrm{E}}\bigr) = -h_Q(X) - \mathsf{E}_Q\left[\log \frac{1}{\mathcal{E}} e^{-\frac{X}{\mathcal{E}}}\right] \tag{133}$$

$$= -h_Q(X) + \log \mathcal{E} + 1 \tag{134}$$

(where $h_Q(X)$ is computed based on the law $Q(\cdot)$) and by evaluating the right-hand side of (132):

$$D\bigl((QW) \| (Q_{\mathrm{E}}W)\bigr)$$

$$= -h_{(QW)}(Y) - \mathsf{E}_{(QW)}\left[\log\left(\frac{1}{\sqrt{\mathcal{E}(\mathcal{E}+2)}} \cdot \exp\left(\frac{\sqrt{\mathcal{E}}Y - \sqrt{\mathcal{E}+2}|Y|}{\sqrt{\mathcal{E}}}\right)\right)\right] \tag{135}$$

$$= -h_{(QW)}(Y) + \frac{1}{2}\log \mathcal{E} + \frac{1}{2}\log(\mathcal{E}+2) - \mathcal{E} + \sqrt{\frac{\mathcal{E}+2}{\mathcal{E}}}\mathsf{E}_{(QW)}[|Y|] \tag{136}$$

$$\geq -h_{(QW)}(Y) + \frac{1}{2}\log \mathcal{E} + \frac{1}{2}\log(\mathcal{E}+2) - \mathcal{E} + \sqrt{\mathcal{E}(2+\mathcal{E})}. \tag{137}$$

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

$$\mathsf{E}_{(QW)}[|Y|] \geq \bigl|\mathsf{E}_{(QW)}[Y]\bigr| = \mathcal{E}. \tag{138}$$

The proof of the monotonicity and positivity of $f_{\mathrm{low}}(\cdot)$ is straightforward and therefore omitted.

B A Proof of Corollary 16

To prove the claim of this lemma we rely on Proposition 15, i.e., we need to derive a function $\mathsf{C}_{\mathrm{asy}}(\cdot)$ that satisfies (67) and (68).

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

$$\varliminf_{\mathsf{A}\uparrow\infty} \frac{\mathsf{C}(\mathsf{A}, \alpha\mathsf{A})}{\frac{1}{2}\log \mathsf{A}} \geq 1, \tag{139}$$

and

$$\varliminf_{\mathcal{E}\uparrow\infty} \frac{\mathsf{C}(\mathcal{E})}{\frac{1}{2}\log \mathcal{E}} \geq 1, \tag{140}$$

respectively.

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

$$\mathsf{C}(\mathsf{A},\alpha\mathsf{A}) \le \mathsf{C}_{\text{peak}}(\mathsf{A}) \le \mathsf{C}_{\text{avg}}(\mathsf{A}) \tag{141}$$

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:

$$R(y) \triangleq \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} & y < 0 \\ \frac{1}{\sqrt{8\pi\mathcal{E}y}} e^{-\frac{y}{2\mathcal{E}}} & y \ge 0. \end{cases} \tag{142}$$

We get

$$\begin{aligned} D\bigl(W(\cdot|x)\big\|R(\cdot)\bigr) &= -\frac{1}{2}\log 2\pi e(\sigma^2+x) \underbrace{- \int_{-\infty}^{0} \log \frac{e^{-y^2/2}}{\sqrt{2\pi}} W(y|x)\,\mathrm{d}y}_{c_1'(x)} \\ &\quad \underbrace{- \int_{0}^{\infty} \log\left(\frac{1}{\sqrt{8\pi\mathcal{E}y}} e^{-\frac{y}{2\mathcal{E}}}\right) W(y|x)\,\mathrm{d}y}_{c_2'(x)}. \end{aligned} \tag{143}$$

From (83) and (84) with $p = \frac{1}{2}$ we know that

$$\begin{aligned} c_1'(x) &= \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\log\sqrt{2\pi} + \frac{\sigma^2+x+x^2}{2}\mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right) \\ &\quad - \frac{x}{2}\sqrt{\frac{\sigma^2+x}{2\pi}} e^{-\frac{x^2}{2(\sigma^2+x)}} \qquad (144) \\ &\le \frac{1}{4} e^{-\frac{x^2}{2(\sigma^2+x)}} \left(2\log\sqrt{2\pi} + \sigma^2 + x + x^2 - 2x\sqrt{\frac{\sigma^2+x}{2\pi}}\right) \qquad (145) \end{aligned}$$

where in the last step we used the bound

$$\mathcal{Q}\left(\xi\right) \le \frac{1}{2} e^{-\frac{\xi^2}{2}}, \qquad \xi \ge 0. \tag{146}$$

Note that (145) is bounded, i.e., there exists some finite constant $k_1 \in \mathbb{R}$ (independent of x and $\mathcal{E}$) such that

$$c_1'(x) \le k_1, \qquad \forall\, x \ge 0. \tag{147}$$

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

$$\begin{aligned} c_2'(x) &= \underbrace{\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right)}_{\le 1} \log\sqrt{8\pi\mathcal{E}} + \frac{1}{2\mathcal{E}} \underbrace{\sqrt{\frac{\sigma^2+x}{2\pi}} e^{-\frac{x^2}{2(\sigma^2+x)}}}_{\le k_2} \\ &\quad + \frac{x}{2\mathcal{E}} \underbrace{\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right)}_{\le 1} + \frac{1}{2}\int_0^{\infty} \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}}\,\mathrm{d}y \end{aligned} \tag{148}$$

$$\leq \max\left\{\frac{1}{2}\log 8\pi\mathcal{E}, 0\right\} + \frac{k_2}{2\mathcal{E}} + \frac{x}{2\mathcal{E}}$$
$$+ I\{x \geq 1\} \cdot \frac{1}{2}\int_0^\infty \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}}\,\mathrm{d}y\bigg|_{x\geq 1}$$
$$+ I\{x < 1\} \cdot \frac{1}{2}\int_0^\infty \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}}\,\mathrm{d}y\bigg|_{x<1}. \tag{149}$$

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

$$I\{\text{statement}\} = \begin{cases} 1 & \text{if statement is true} \\ 0 & \text{otherwise.} \end{cases} \tag{150}$$

Analogously to (97) we next have

$$\int_0^\infty \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}}\,\mathrm{d}y\bigg|_{x\geq 1} \leq \log x + \underbrace{\frac{1}{x}}_{\leq 1} \cdot \underbrace{\sqrt{\frac{\sigma^2+x}{2\pi}} e^{-\frac{x^2}{2(\sigma^2+x)}}}_{\leq k_2} \tag{151}$$
$$\leq \log x + k_2 \tag{152}$$

and analogously to (100) we get

$$\int_0^\infty \frac{\log y}{\sqrt{2\pi(\sigma^2+x)}} e^{-\frac{(y-x)^2}{2(\sigma^2+x)}}\,\mathrm{d}y\bigg|_{x<1}$$
$$\leq \underbrace{\sqrt{\frac{\sigma^2+x}{2\pi}} e^{-\frac{x^2}{2(\sigma^2+x)}}}_{\leq k_2} + \underbrace{x}_{\leq 1} \underbrace{\left(1 - \mathcal{Q}\left(\frac{x}{\sqrt{\sigma^2+x}}\right)\right)}_{\leq 1} \tag{153}$$
$$\leq k_2 + 1. \tag{154}$$

Plugging all this into (143) finally yields

$$D\big(W(\cdot|x)\big\|R(\cdot)\big)$$
$$\leq -\frac{1}{2}\log 2\pi e - \frac{1}{2}\log(\sigma^2+x) + k_1 + \max\left\{\frac{1}{2}\log 8\pi\mathcal{E}, 0\right\} + \frac{k_2}{2\mathcal{E}} + \frac{x}{2\mathcal{E}}$$
$$+ I\{x \geq 1\} \cdot \frac{1}{2}\big(\log x + k_2\big) + I\{x < 1\} \cdot \frac{1}{2}(k_2+1) \tag{155}$$
$$= \max\left\{\frac{1}{2}\log 4\mathcal{E}, -\frac{1}{2}\log 2\pi\right\} - \frac{1}{2} - I\{x \geq 1\} \cdot \frac{1}{2}\underbrace{\log(\sigma^2+x)}_{\geq \log x}$$
$$- I\{x < 1\} \cdot \frac{1}{2}\underbrace{\log(\sigma^2+x)}_{\geq \log \sigma^2} + k_1 + \frac{k_2}{2\mathcal{E}} + \frac{x}{2\mathcal{E}}$$
$$+ I\{x \geq 1\} \cdot \frac{1}{2}\log x + \underbrace{I\{x \geq 1\}}_{\leq 1} \cdot \frac{k_2}{2} + \underbrace{I\{x < 1\}}_{\leq 1} \cdot \frac{k_2+1}{2} \tag{156}$$
$$\leq \max\left\{\frac{1}{2}\log 4\mathcal{E}, -\frac{1}{2}\log 2\pi\right\} - \frac{1}{2} - I\{x < 1\} \cdot \frac{1}{2}\log \sigma^2 + k_1 + \frac{k_2}{2\mathcal{E}}$$
$$+ \frac{x}{2\mathcal{E}} + \frac{k_2}{2} + \frac{k_2+1}{2} \tag{157}$$
$$\leq \max\left\{\frac{1}{2}\log 4\mathcal{E}, -\frac{1}{2}\log 2\pi\right\} - \min\left\{0, \frac{1}{2}\log \sigma^2\right\} + k_1 + \frac{k_2}{2\mathcal{E}} + \frac{x}{2\mathcal{E}} + k_2. \tag{158}$$

Hence, we get

$$\mathsf{C}(\mathcal{E}) \leq \mathsf{E}_{Q^*}\Big[D\big(\tilde{W}(\cdot|X) \,\big\|\, R(\cdot)\big)\Big] \tag{159}$$

$$\leq \max\left\{\frac{1}{2}\log 4\mathcal{E}, -\frac{1}{2}\log 2\pi\right\} - \min\left\{0, \frac{1}{2}\log\sigma^2\right\} + k_1 + \frac{k_2}{2\mathcal{E}} + \frac{\mathcal{E}}{2\mathcal{E}} + k_2 \tag{160}$$

and therefore

$$\varlimsup_{\mathcal{E}\uparrow\infty} \frac{\mathsf{C}(\mathcal{E})}{\frac{1}{2}\log\mathcal{E}} \leq 1. \tag{161}$$

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 Poisson Distribution Approximates the Gaussian Distribution

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

$$T_{\text{c}} \triangleq T + U. \tag{162}$$

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

$$f_{T_{\text{c}}}(t) = \begin{cases} e^{-\lambda}\frac{\lambda^{\lfloor t\rfloor}}{\lfloor t\rfloor!} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{163}$$

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:

$$T = \lfloor T_{\text{c}} \rfloor. \tag{164}$$

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

$$S \triangleq \frac{T_{\text{c}} - \lambda}{\sqrt{\lambda}} \tag{165}$$

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, $f_S(\cdot)$, and the PDF of a zero-mean, unit-variance Gaussian random variable G, $f_G(\cdot)$, tends to zero as $\lambda \uparrow \infty$:

$$D(f_S \| f_G) = \mathsf{E}\left[\log \frac{f_S(S)}{f_G(S)}\right] \tag{166}$$

$$= \mathsf{E}\left[\log \frac{f_S\left(\frac{T_c-\lambda}{\sqrt{\lambda}}\right)}{f_G\left(\frac{T_c-\lambda}{\sqrt{\lambda}}\right)}\right] \tag{167}$$

$$= \frac{1}{2}\log\lambda - \lambda + \mathsf{E}[\lfloor T_c \rfloor]\log\lambda + \frac{1}{2}\log 2\pi - \mathsf{E}[\log(\lfloor T_c \rfloor !)] + \frac{1}{2\lambda}\mathsf{E}\left[(T_c-\lambda)^2\right] \tag{168}$$

$$= \frac{1}{2}\log 2\pi\lambda \underbrace{-\lambda + \lambda\log\lambda - \mathsf{E}[\log(T!)]}_{=-H(T)} + \frac{1}{2} + \frac{1}{6\lambda} \tag{169}$$

$$= \frac{1}{2}\log 2\pi e\lambda - H(T) + \frac{1}{6\lambda}, \tag{170}$$

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

$$\varliminf_{\lambda\uparrow\infty}\left\{H(T) - \frac{1}{2}\log 2\pi e\lambda\right\} \geq 0. \tag{171}$$

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

$$0 \leq \varlimsup_{\lambda\uparrow\infty} D(f_S\|f_G) \leq \varlimsup_{\lambda\uparrow\infty} \frac{1}{6\lambda} = 0. \tag{172}$$

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 help with the bounds at low SNR.

References

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

[2] S. Karp, R. M. Gagliardi, S. Moran, and L. Stotts, *Optical Channels.* Plenum Press, 1988.

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

[4] S. 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] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.

[6] A. A. Farid and S. Hranilovic, "Upper and lower bounds on the capacity of wireless optical intensity channels," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 24–30, 2007, pp. 2416–2420.

[7] ——, "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] A. Lapidoth, S. M. Moser, and M. 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] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," in *Proceedings of the IEE*, vol. 137, part I (Communications, Speech and Vision), no. 6, December 1990, pp. 424–430.

[10] D. Brady and S. Verdú, "The asymptotic capacity of the direct detection photon channel with a bandwidth constraint," in *Proceedings Twenty-Eighth Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, IL, USA, October 3–5, 1990, pp. 691–700.

[11] A. Lapidoth and S. 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] A. Lapidoth, J. H. Shapiro, V. Venkatesan, and L. Wang, "The Poisson channel at low input powers," in *Proceedings Twenty-Fifth IEEE Convention of Electrical & Electronics Engineers in Israel (IEEEI)*, Eilat, Israel, December 3–5, 2008.

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

[14] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 710–715, November 1980.

[15] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel — part I and II," *IEEE Transactions on Information Theory*, vol. 34, no. 6, pp. 1462–1471, November 1988.

[16] M. R. Frey, "Capacity of the L_p norm-constrained Poisson channel," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 445–450, March 1992.

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

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

[19] I. Bar-David and G. Kaplan, "Information rates of photon-limited overlapping pulse position modulation channels," *IEEE Transactions on Information Theory*, vol. 30, no. 3, pp. 455–464, May 1984.

[20] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, October 2003.

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

[22] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 437–453, February 2006.

[23] V. V. Prelov and E. 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] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.