# 國 立 交 通 大 學

# 電信工程研究所

# 碩 士 論 文

## 離散時間中卜瓦松通道容量分析

## On the Capacity of the Discrete-Time Poisson Channel

研究生： 陳偉翔

指導教授： 莫詩台方

中 華 民 國 九十九 年 八 月

# 離散時間中卜瓦松通道容量分析

# On the Capacity of Discrete-Time Poisson Channel


研　究　生：陳偉翔　　　　　　Student：Wei-Hsiang Chen

指導教授：莫詩台方　副教授　　Advisor：Stefan M. Moser

國　立　交　通　大　學

電信工程研究所

碩　士　論　文


A Thesis

Submitted to Institute of Communication Engineering

College of Electrical Engineering and Computer Engineering

National Chiao Tung University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

In

Communication Engineering

August 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年八月

離散時間中卜瓦松通道容量分析

On the Capacity of Discrete-Time Poisson Channel

研　究　生：陳偉翔　　　　　　指導教授：莫詩台方　博士

國立交通大學電信工程研究所碩士班

## 中文摘要

本論文中將研究離散時間中卜瓦松通道之通道容量。此通道模型可描述一個包含直接感測接收端的光通道，其接收端可視為一個計算光子的計數器，而到達接收端之光子數量由傳送端信號以及環境中游離的光子決定。基於能量消耗和安全的考量，我們將同時對傳送信號的平均能量和瞬時能量作限制。

我們嘗試找出此通道容量的上下界。藉著某些近日研究卜瓦松分布的新成果以及應用通道容量的二元性，可以推導出包含部份數值估計的數學式，其中數值估計的複雜度十分的低。我們上下界所夾的區域並不大，換句話說，我們成功的縮小了估計通道容量時可能的區間。

# On the Capacity of the Discrete-Time Poisson Channel

Student: Chen Wei-Hsiang Advisor: Prof. Stefan M. Moser

Institute of Communication Engineering

National Chiao Tung University

# Abstract

The capacity of the discrete time Poisson channel is studied. This channel law describes optical communication with a direct-detection receiver, where the output models the receiving photons due to the transmitted laser signal and due to surrounding light. For battery issues and safety reasons the inputs are simultaneously constrained on both their average and peak power.

We try to find an upper and a lower bound on the capacity. Applying some recent results on the Poisson distribution and a dual expression of channel capacity, the bounds are partially analytic and partially numerical, where the computation complexity is very low. The gap between our bounds is not large, *i.e.* we ave succeeded in narrowing the range of the capacity.

# Acknowledgments

I would like to thank my advisor, Prof. Stefan M. Moser. He gives me the ideas of this thesis, and keeps guiding me through the difficulties. I've learned a lot from him how to analyze a problem rationally and develop a reasonable solution. Beside, I've improved my skills of presenting my opinions from discussions with my advisor.

I'm thankful to join the family of information-theory lab. Our lab is a comfortable working space with some powerful iMacs, a good place for inspiration with a sofa and a white board, and a relaxing corner with a fantastic coffee machine! I really enjoy the atmosphere here and appreciate my colleagues. They are always willing to help me while I'm typing this thesis and preparing the oral presentation.

Finally, I would like to thank my family. They always give me a lot of support and love.

Hsinchu, 20 August 2010

Chen Wei-Hsiang

# Contents

# List of Figures

# Chapter 1

# Introduction

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

$$W(y|x) = e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^y}{y!}, \qquad y \in \mathbb{Z}_0^+, x \in \mathbb{R}_0^+ \tag{1.1}$$

The channel law is often used to model pulse-amplitude modulated optical communication with a direct-detection receiver. The input $x$ is proportional to the product of light intensity by the pulse duration, and the dark current $\lambda_0$ is proportional to the product of background radiation and the pulse duration. The output $Y$ models the number of photons arriving at the receiver during the pulse.

Due to power and safety issues, we have peak-power and average-power constraints on the transmitter. 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 usual for radio communication:

$$\Pr[X > \mathsf{A}] = 0 \tag{1.2}$$

$$\mathsf{E}[X] \leq \mathcal{E} \tag{1.3}$$

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

$$\alpha \triangleq \frac{\mathcal{E}}{\mathsf{A}} \tag{1.4}$$

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

Our work builds on the results in [1] with the same power constraints. There proposed asymptotic upper bounds valid only for infinite power and lower bounds valid for all value of power condition but only tight for high power. These upper and lower bounds will asymptotically coincide, thus yielding the exact asymptotic behavior of channel capacity.

On the other hand, we propose upper bounds which are valid for any specified power, and we derive lower bounds that are better in the low-SNR region. Our lower bounds contain an integral that needs numerical computation, however the complexity for this integral is very low.

The derivation of the upper bounds is based on a technique introduced in [2], which uses a dual expression for mutual information. The main idea is to use dual expression to change the maximization problem of capacity into a minimization. Thereby, it becomes possible to find an upper bound to capacity by simply dropping the minimization. Briefly speaking, although capacity is usually not easy to compute because it's a supremum of mutual information, we can find an expression of capacity where we can upper-bound it by choosing specific output distribution.

The idea of the lower bound is easier. It follows the definition of capacity and mutual information. Since capacity is the supremum of mutual information, an intuitive lower bound is the mutual information with a specific input distribution.

This thesis is structured as follows. After some remarks about our notations, we summarize our main results in the subsequent chapter. We also show some figures with brief explanations in Chapter 2. The derivation are then given in Chapter 3 (upper bounds) and Chapter 4 (lower bounds). These two derivation chapters both contain a section with mathematical preliminaries.

# Chapter 2

# Main Results

## 2.1 Notation and Definition

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

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

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

- $R(\cdot)$ denotes a distribution on the channel output.

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

The symbol $\mathcal{E}$ denotes average power and $\mathsf{A}$ stands for peak power. We shall denote the mean-$\lambda$ Poisson distribution by $\mathcal{P}_0(\lambda)$ and the uniform distribution on the interval $[a,b)$ by $\mathcal{U}([a,b))$. All rates specified in this thesis are in nats per channel use, and all logarithms are natural logarithms.

Next, we provide some definitions required in our bounds. According to our introduction, different bounds can be found with different choices of input and output distributions. Here we focus on notations related to the distributions we choose in this thesis, where the main ideas of derivation will be shown in later chapters.

First we define the probability mass function of a Poisson random variable $N_\lambda$ with parameter $\lambda$, *i.e.*

$$\Pr[N_\lambda = n] = e^{-\lambda}\frac{\lambda^n}{n!}, \quad n \in \mathbb{N}_0 \tag{2.1}$$

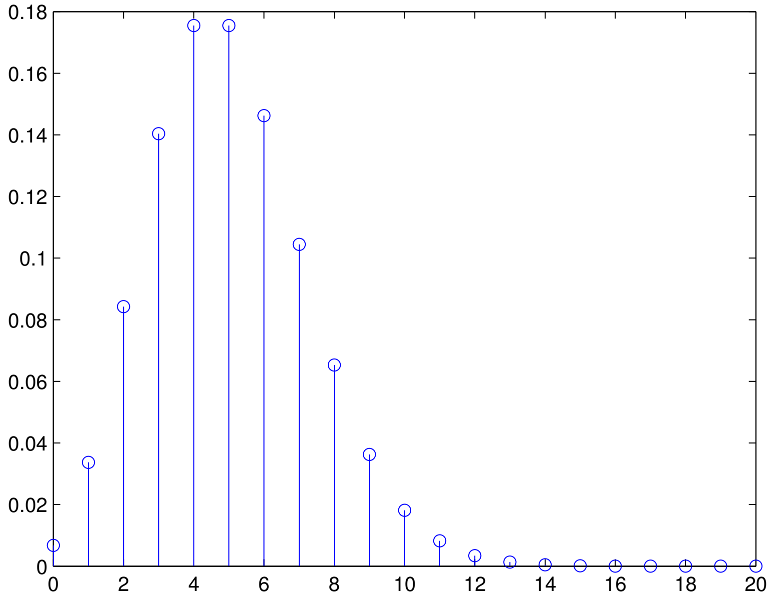


Figure 2.1: Poisson law with mean equal to 5

We observe that the Poisson channel (1.1) with input $x$ and discrete output $y$ follows (2.1). Note that $Y$ is discrete, *i.e.*, $Y \in \mathbb{N}_0$. In certain situations, it will be easier to have a channel model with a continuous output. To that goal, we define an adapted Poisson random variable as follows: $\tilde{N}_\lambda$ with mean $\lambda$

$$\tilde{N}_\lambda = N_\lambda + U \tag{2.2}$$

where

$$U \sim \mathcal{U}([0,1)) \tag{2.3}$$

and its probability density function yields

$$f_{\tilde{N}_\lambda}(r) = e^{-\lambda}\frac{\lambda^{\lfloor r \rfloor}}{\lfloor r \rfloor !}, \quad r \in \mathbb{R}_0^+ \tag{2.4}$$

According to (2.4), we introduce a new channel output random variable $\tilde{Y}$

$$\tilde{Y} = Y + U \tag{2.5}$$

where $U$ satisfies (2.3). The channel model with dark current $\lambda_0$ can be adapted to

$$\tilde{W}'(\tilde{y}|x) = e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^{\lfloor \tilde{y} \rfloor}}{\lfloor \tilde{y} \rfloor !}, \qquad \tilde{y} \geq 0, x \geq 0, \lambda_0 \geq 0 \tag{2.6}$$

We call it an adapted Poisson channel model with continuous output. Properties of this channel will be given in Chapter 3. Here we just define some notations related to this model.

$$p_x \triangleq \Pr[\tilde{Y} \leq (\mathsf{A} + \lambda_0)(1+\delta)|X = x] \tag{2.7}$$
$$p_{\mathrm{T},x} \triangleq \Pr[\tilde{Y} > (\mathsf{A} + \lambda_0)(1+\delta)|X = x] = 1 - p_x \tag{2.8}$$

where $\mathsf{A}$ denotes peak-power constraint, and $\delta$ is a constant.

As an example, Figure 2.2 shows $p_3$ and $p_{\mathrm{T},3}$ in adapted Poisson channel model. We assume that

$$x = 3$$
$$\mathsf{A} = 10.5$$
$$\lambda_0 = 2$$
$$\delta = 0$$

which leads to

$$x + \lambda_0 = 5$$
$$(\mathsf{A} + \lambda_0)(1+\delta) = 12.5$$

and we can derive

$$p_k = \sum_{n=0}^{\lfloor (\mathsf{A}+\lambda_0)(1+\delta) \rfloor} \Pr[N_\lambda = n] - \Pr[N_\lambda = \lfloor (\mathsf{A}+\lambda_0)(1+\delta) \rfloor] \cdot (1 - \lfloor (\mathsf{A}+\lambda_0)(1+\delta) \rfloor + (\mathsf{A}+\lambda_0)(1+\delta)) \tag{2.9}$$
$$= \sum_{n=0}^{12} \Pr[N_\lambda = n] - 0.5 \cdot \Pr[N_\lambda = 12] \tag{2.10}$$

Finally we present some functions we use in this thesis.

- The Gamma function and the incomplete Gamma function:

$$\Gamma(\nu) \triangleq \int_0^\infty e^{-t} t^{\nu-1} \, \mathrm{d}t, \tag{2.11}$$
$$\gamma(\nu, \xi) \triangleq \int_0^\xi e^{-t} t^{\nu-1} \, \mathrm{d}t, \qquad \nu > 0 \tag{2.12}$$

Note that $\Gamma(n) = (n-1)!$ for all positive integer $n$.

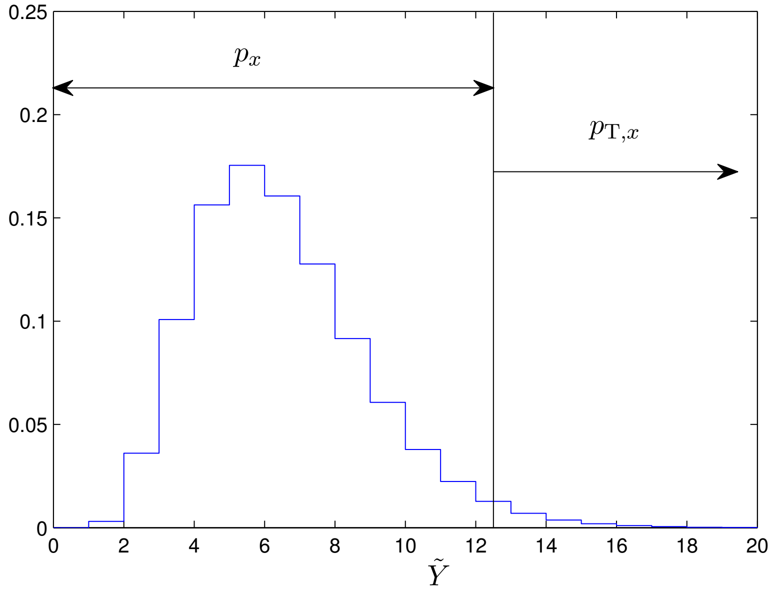


Figure 2.2: Adapted Poisson channel law with continuous output

- The Gaussian $\mathcal{Q}$-function and the error function:

$$\mathcal{Q}(\xi) \triangleq \int_{\xi}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} \mathrm{d}t, \qquad \forall \xi \in \mathbb{R} \tag{2.13}$$

$$\mathrm{erf}\,(\xi) \triangleq \frac{2}{\sqrt{\pi}} \int_{0}^{\xi} e^{-t^2}\,\mathrm{d}t, \qquad \forall \xi \in \mathbb{R} \tag{2.14}$$

Note that $Q(\xi) = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{\xi}{\sqrt{2}}\right)$.

- The hypergeometric functions:

$$\begin{aligned} &{}_p\mathcal{F}_q\left(a_1, a_2, \cdots, a_p; b_1, b_2, \cdots, b_q; x\right) \\ &\triangleq \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k \cdots (a_p)_k}{(b_1)_k (b_2)_k \cdots (b_q)_k} \cdot \frac{x^k}{k!} \end{aligned} \tag{2.15}$$

where

$$(a)_k \triangleq \frac{\Gamma(a+k)}{\Gamma(a)} \tag{2.16}$$

## 2.2 Results

We present upper and lower bounds on the capacity of channel (1.1). The upper bounds are found analytically, and the lower bounds are expressed with some integral terms. These numerical terms are easy to approximate theoretically, but some computing problems occur in the simulation of Matlab. We will show the results first and defer the simulation discussion to Chapter 5.

We distinguish between three cases: in the first case, we have both an average- and peak-power constraint where the average-to-peak ratio (1.4) is in the range $0 < \alpha < \frac{1}{3}$. In the second case, $\frac{1}{3} \leq \alpha \leq 1$, which includes the situation with only peak-power constraint $\alpha = 1$. And finally, in the third case, we look at the situation with only an average-power constraint. Each case contains two parts with different assumptions. First part contains the upper and lower bounds on a Poisson channel with dark current. However, the lower bound here is not good for low SNR, so we assume a channel without dark current and derive another lower bound. Note that we don't propose upper bounds on a channel without dark current due to some limitation of our derivation in Chapter 3.

We begin with the first case.

**Theorem 1.** *The channel capacity* $\mathsf{C}(\mathsf{A}, \mathcal{E})$ *of a Poisson channel with dark current* $\lambda_0$ *under a peak-power constraint* (1.2) *and an average-power constraint* (1.3)*, where the ratio* $\alpha = \frac{\mathcal{E}}{\mathsf{A}}$ *lies in* $(0, \frac{1}{3})$*, is bounded as follows*

$$
\begin{aligned}
\mathsf{C} \leq\ & -\frac{1}{2}\log(2\pi e) + \left(\frac{1}{2} - \nu\right)\log(\alpha\mathsf{A} + \lambda_0) + \left(\frac{5}{6}(\alpha\mathsf{A} + \lambda_0)^{-1} + \frac{1}{6}(\alpha\mathsf{A} + \lambda_0)^{-2}\right) \\
& + (1 - \nu)\left(\frac{1}{2}(\alpha\mathsf{A} + \lambda_0)^{-1} + \frac{5}{6}(\alpha\mathsf{A} + \lambda_0)^{-2} + \frac{1}{3}(\alpha\mathsf{A} + \lambda_0)^{-3}\right. \\
& \quad + (\alpha\mathsf{A} + \lambda_0)\log\left(1 + \frac{1}{\alpha\mathsf{A} + \lambda_0}\right) - 1\Bigg) + \frac{\mu\left(\alpha\mathsf{A} + \lambda_0 + \frac{1}{2}\right)}{(\mathsf{A} + \lambda_0)(1 + \delta)} \\
& + \Bigg(1 - \frac{\gamma(\lfloor(\mathsf{A} + \lambda_0)(1 + \delta)\rfloor, \alpha\mathsf{A} + \lambda_0)}{\Gamma(\lfloor(\mathsf{A} + \lambda_0)(1 + \delta)\rfloor)} \\
& \quad + \Pr\left[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A} + \lambda_0)(1 + \delta)\rfloor\right] \cdot \Big((\mathsf{A} + \lambda_0)(1 + \delta) - \lfloor(\mathsf{A} + \lambda_0)(1 + \delta)\rfloor\Big)\Bigg) \\
& \quad \cdot \left(\nu\log\left(\frac{(\mathsf{A} + \lambda_0)(1 + \delta)}{\mu}\right) + \log\gamma(\nu, \mu) + \log\frac{1}{1 - p_{\mathrm{T,A}}}\right) \\
& + p_{\mathrm{T,A}}\log\frac{1}{p_{\mathrm{T,A}}} + \left(\frac{1}{2} - (\mathsf{A} + \lambda_0)\delta\right) \cdot \frac{\gamma(\lfloor(\mathsf{A} + \lambda_0)(1 + \delta)\rfloor + 1, \mathsf{A} + \lambda_0)}{\Gamma(\lfloor(\mathsf{A} + \lambda_0)(1 + \delta)\rfloor)} \\
& + \Pr\left[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A} + \lambda_0)(1 + \delta)\rfloor\right] \cdot \Bigg((\mathsf{A} + \lambda_0) \\
& \quad + \frac{1}{2}\Big((\mathsf{A} + \lambda_0)(1 + \delta) - \lfloor(\mathsf{A} + \lambda_0)(1 + \delta)\rfloor - 1\Big)^2\Bigg)
\end{aligned}
\tag{2.17}
$$

$$
\begin{aligned}
\mathsf{C} \geq\ & r\log\mathsf{A} + (\alpha\mathsf{A} + r)\log\left(1 + \frac{r}{\alpha\mathsf{A}}\right) - \alpha\mathsf{A} \\
& + (2r - 1)e^{\mu}(2\alpha\mu - 1) \cdot {}_2\mathcal{F}_2\left(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}, \frac{3}{2}; -\mu\right) - \mu - \log\left(\frac{1}{2} - \alpha\mu\right) + \alpha\mu - r - \lambda \\
& + \int_0^{A} \frac{\sqrt{\mu} \cdot (x + \lambda_0)\log(x + \lambda_0)}{\sqrt{\mathsf{A}\pi x} \cdot \mathrm{erf}\left(\sqrt{\mu}\right)} e^{-\frac{\mu x}{\mathsf{A}}}\, \mathrm{d}x
\end{aligned}
$$

$$
\begin{aligned}
&+\int_0^1 \left( \frac{1}{1-t} + \frac{(t^{r-1} - 1 - e^{-\lambda_0(1-t)})\sqrt{\mu}\cdot \operatorname{erf}\left(\sqrt{\mu + (1-t)\mathsf{A}}\right)}{(1-t)\sqrt{\mu+(1-t)\mathsf{A}}\cdot \operatorname{erf}\left(\sqrt{\mu}\right)} \right. \\
&\qquad\qquad \left. + \alpha\mathsf{A} + \lambda_0 + r - 1 \right) \frac{1}{\log t}\,\mathrm{d}t
\end{aligned} \tag{2.18}
$$

Recall that

$$
p_{\mathrm{T},\mathsf{A}} \triangleq \Pr[\tilde{Y} > (\mathsf{A}+\lambda_0)(1+\delta) | X = \mathsf{A}] = 1 - p_{\mathsf{A}} \tag{2.19}
$$

and $\Gamma(\cdot)$, $\gamma(\cdot)$ are defined by (2.11), (2.12). In the bounds, $\mu$ is the solution to

$$
\alpha = \frac{1}{2\mu} - \frac{e^{-\mu}}{\sqrt{\mu\pi}\cdot \operatorname{erf}\left(\sqrt{\mu}\right)} \tag{2.20}
$$

where the error function $\operatorname{erf}(\cdot)$ is defined by (2.14). This solution is well-defined because $\mu \mapsto \frac{1}{2\mu} - \frac{e^{-\mu}}{\sqrt{\mu\pi}\cdot\operatorname{erf}(\sqrt{\mu})}$ is monotonically decreasing in $[0,\infty)$ and tends to $\frac{1}{3}$ for $\mu \downarrow 0$ and to 0 for $\mu \uparrow \infty$. Besides, we choose $\nu$ a specific value for our derivation of upper bounds. $\delta$ and $r$ are free parameters we can choose and we will show some examples in the plots.

**Theorem 2.** *The channel capacity $\mathsf{C}(\mathsf{A},\mathcal{E})$ of a Poisson channel without dark current $\lambda_0$ under a peak-power constraint* (1.2) *and an average-power constraint* (1.3)*, where the ratio $\alpha = \frac{\mathcal{E}}{\mathsf{A}}$ lies in $(0, \frac{1}{3})$, is lower-bounded as follows*

$$
\begin{aligned}
\mathsf{C} \geq\; & r\log\mathsf{A} + (\alpha\mathsf{A} + r)\log\left(1 + \frac{r}{\alpha\mathsf{A}}\right) + \alpha\mathsf{A}\log\mathsf{A} \\
& + \frac{2}{9}e^{\mu}(2\alpha\mu - 1)\cdot {}_2\mathcal{F}_2\left(\frac{3}{2},\frac{3}{2};\frac{5}{2},\frac{5}{2};-\mu\right)\mathsf{A} - \alpha\mathsf{A} \\
& + (2r-1)e^{\mu}(2\alpha\mu - 1)\cdot {}_2\mathcal{F}_2\left(\frac{1}{2},\frac{1}{2};\frac{3}{2},\frac{3}{2};-\mu\right) - \mu - \log\left(\frac{1}{2} - \alpha\mu\right) + \alpha\mu - r \\
& + \int_0^1 \left[ \frac{1}{1-t}\left( 1 - t^{r-1}\frac{\sqrt{\mu}\cdot\operatorname{erf}\left(\sqrt{\mu + \mathsf{A}(1-t)}\right)}{\operatorname{erf}\left(\sqrt{\mu}\right)\sqrt{\mu + \mathsf{A}(1-t)}} \right) - (\alpha\mathsf{A} + r - 1) \right] \frac{1}{\log t}\,\mathrm{d}t
\end{aligned} \tag{2.21}
$$

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

**Theorem 3.** *The channel capacity $\mathsf{C}(\mathsf{A},\mathcal{E})$ of a Poisson channel with dark current $\lambda_0$ under a peak-power constraint* (1.2) *and an average-power constraint* (1.3)*, where the ratio $\alpha = \frac{\mathcal{E}}{\mathsf{A}}$ lies in $[\frac{1}{3}, 1]$, is bounded as follows*

$$
\begin{aligned}
\mathsf{C} \leq\; & -\frac{1}{2}\log(2\pi e) + \left(\frac{1}{2} - \beta\right)\log(\alpha\mathsf{A} + \lambda_0) + \left(\frac{5}{6}(\alpha\mathsf{A}+\lambda_0)^{-1} + \frac{1}{6}(\alpha\mathsf{A}+\lambda_0)^{-2}\right) \\
& + (1-\beta)\left(\frac{1}{2}(\alpha\mathsf{A}+\lambda_0)^{-1} + \frac{5}{6}(\alpha\mathsf{A}+\lambda_0)^{-2} + \frac{1}{3}(\alpha\mathsf{A}+\lambda_0)^{-3}\right.
\end{aligned}
$$

$$
\begin{aligned}
&\quad + (\alpha\mathsf{A} + \lambda_0)\log\left(1 + \frac{1}{\alpha\mathsf{A} + \lambda_0}\right) - 1\Bigg) \\
&+ \Bigg(1 - \frac{\gamma(\lfloor(\mathsf{A} + \lambda_0)(1+\delta)\rfloor, \alpha\mathsf{A} + \lambda_0)}{\Gamma(\lfloor(\mathsf{A} + \lambda_0)(1+\delta)\rfloor)} \\
&\quad + \Pr\left[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A} + \lambda_0)(1+\delta)\rfloor\right] \cdot \left((\mathsf{A} + \lambda_0)(1+\delta) - \lfloor(\mathsf{A} + \lambda_0)(1+\delta)\rfloor\right)\Bigg) \\
&\quad \cdot \left(\beta \log\left((\mathsf{A} + \lambda_0)(1+\delta)\right) + \log\frac{1}{\beta \cdot (1 - p_{\mathrm{T},\mathsf{A}})}\right) + p_{\mathrm{T},\mathsf{A}} \log\frac{1}{p_{\mathrm{T},\mathsf{A}}} \\
&+ \left(\frac{1}{2} - (\mathsf{A} + \lambda_0)\delta\right) \cdot \frac{\gamma(\lfloor(\mathsf{A} + \lambda_0)(1+\delta)\rfloor + 1, \mathsf{A} + \lambda_0)}{\Gamma(\lfloor(\mathsf{A} + \lambda_0)(1+\delta)\rfloor + 1)} \\
&+ \Pr\left[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A} + \lambda_0)(1+\delta)\rfloor\right] \cdot \Bigg((\mathsf{A} + \lambda_0) \\
&\quad + \frac{1}{2}\Big((\mathsf{A} + \lambda_0)(1+\delta) - \lfloor(\mathsf{A} + \lambda_0)(1+\delta)\rfloor - 1\Big)^2\Bigg)
\end{aligned}
\tag{2.22}
$$

$$
\begin{aligned}
\mathsf{C} \geq\ & r\log\mathsf{A} + \left(\frac{1}{3}\mathsf{A} + r\right)\log\left(1 + \frac{3r}{\mathsf{A}}\right) - \frac{1}{3}\mathsf{A} + \log 2 - 3r + 1 - \lambda_0 \\
&+ \int_0^{\mathsf{A}} \frac{(x + \lambda_0)\log(x + \lambda_0)}{\sqrt{4\mathsf{A}x}}\,\mathrm{d}x \\
&+ \int_0^1 \Bigg(\frac{1}{1-t} + \frac{(t^{r-1} - 1 - e^{-\lambda_0(1-t)})\sqrt{\pi}\cdot\operatorname{erf}\left(\sqrt{(1-t)\mathsf{A}}\right)}{(1-t)\sqrt{4\mathsf{A}(1-t)}} \\
&\qquad + \frac{1}{3}\mathsf{A} + \lambda_0 + r - 1\Bigg)\frac{1}{\log t}\,\mathrm{d}t
\end{aligned}
\tag{2.23}
$$

**Remark 4.** In [1], it is discovered that asymptotically $\alpha = \frac{1}{3}$ is the threshold of activating the average-power constraint. That is, whenever $\alpha \geq \frac{1}{3}$, the average-power constraint is inactive and the upper- and lower-bound can be expressed without $\alpha$. However, here we reach a result that the upper bound is always related to $\alpha$. Note that for $\mathsf{A} \to \infty$, almost all terms of (2.22) containing $\alpha$ tend to some constant such that the bounds become again independent of $\alpha$. Hence we make a conjecture: for $\alpha \geq \frac{1}{3}$, average- and peak-power constraints are both active for low SNR while only the peak-power constraint is active for high SNR.

**Theorem 5.** *The channel capacity $\mathsf{C}(\mathsf{A}, \mathcal{E})$ of a Poisson channel without dark current $\lambda_0$ under a peak-power constraint* (1.2) *and an average-power constraint* (1.3), *where the ratio $\alpha = \frac{\mathcal{E}}{\mathsf{A}}$ lies in $[\frac{1}{3}, 1]$, is bounded as follows*

$$
\begin{aligned}
\mathsf{C} \geq\ & r\log\mathsf{A} + \left(\frac{1}{3}\mathsf{A} + r\right)\log\left(1 + \frac{3r}{\mathsf{A}}\right) + \frac{1}{3}\mathsf{A}\log\mathsf{A} - \frac{5}{9}\mathsf{A} + \log 2 - 3r + 1 \\
&+ \int_0^1 \left[\frac{1}{1-t}\left(1 - t^{r-1}\frac{\sqrt{\pi}\operatorname{erf}\left(\sqrt{(1-t)\mathsf{A}}\right)}{\sqrt{4\mathsf{A}(1-t)}}\right) - \frac{1}{3}\mathsf{A} - r + 1\right]\frac{1}{\log t}\,\mathrm{d}t
\end{aligned}
\tag{2.24}
$$

Finally, for the case with only a average-power constraint the result are as follows.

**Theorem 6.** *The channel capacity* $\mathsf{C}(\mathcal{E})$ *of a Poisson channel with dark current* $\lambda_0$ *under an average-power constraint* (1.3) *is bounded as follows:*

$$\begin{aligned}\mathsf{C} \leq &-\frac{1}{2}\log(2\pi e) + \left(\frac{1}{2}-\nu\right)\log(\mathcal{E}+\lambda_0) + \left(\frac{5}{6}(\mathcal{E}+\lambda_0)^{-1} + \frac{1}{6}(\mathcal{E}+\lambda_0)^{-2}\right)\\ &+(1-\nu)\left(\frac{1}{2}(\mathcal{E}+\lambda_0)^{-1} + \frac{5}{6}(\mathcal{E}+\lambda_0)^{-2} + \frac{1}{3}(\mathcal{E}+\lambda_0)^{-3}\right.\\ &\quad\left. + (\mathcal{E}+\lambda_0)\log\left(1+\frac{1}{\mathcal{E}+\lambda_0}\right) - 1\right) + \nu\log\frac{\mathcal{E}+\lambda_0+\frac{1}{2}}{\nu} + \log\Gamma(\nu) + \nu \end{aligned} \tag{2.25}$$

$$\begin{aligned}\mathsf{C} \geq &\, r\log\mathcal{E} + (\mathcal{E}+r)\log\left(1+\frac{r}{\mathcal{E}}\right) - \mathcal{E} + \frac{1}{2}\log\pi - (r-1)(\log 2+\gamma) + \frac{1}{2}(1-\gamma)\\ &- r - \lambda_0 + \int_0^\infty \frac{(x+\lambda_0)\log(x+\lambda_0)}{\sqrt{2\pi\mathcal{E}x}}e^{-\frac{x}{2\mathcal{E}}}\,\mathrm{d}x\\ &+\int_0^1\left(\frac{1}{1-t} + \frac{t^{r-1}-1-e^{-\lambda_0(1-t)}}{1-t}\cdot\sqrt{\frac{2}{2+4\mathcal{E}-4\mathcal{E}t}}\right.\\ &\qquad\left. + \mathcal{E} + \lambda_0 + r - 1\right)\frac{1}{\log t}\,\mathrm{d}t \end{aligned} \tag{2.26}$$

*where* $\gamma$ *is the Euler's constant defined as*

$$\gamma \triangleq \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \log n\right) \simeq 0.57721\cdots \tag{2.27}$$

**Theorem 7.** *The channel capacity* $\mathsf{C}(\mathcal{E})$ *of a Poisson channel without dark current* $\lambda_0$ *under an average-power constraint* (1.3) *is bounded as follows:*

$$\begin{aligned}\mathsf{C} \geq &\, r\log\mathcal{E} + (\mathcal{E}+r)\log\left(1+\frac{r}{\mathcal{E}}\right) + \mathcal{E}\log\mathcal{E} + (1-\log 2-\gamma)\mathcal{E} + \frac{1}{2}\log\pi\\ &+\frac{1}{2}(1-\gamma) - (r-1)(\log 2+\gamma) - r\\ &+\int_0^1\left[\frac{1}{1-t}\left(1-t^{r-1}\sqrt{\frac{2}{2+4\mathcal{E}+4\mathcal{E}t}}\right) - \mathcal{E} - r + 1\right]\frac{1}{\log t}\,\mathrm{d}t \end{aligned} \tag{2.28}$$

## 2.3 Plots

Plots of our upper bounds and lower bounds will be shown here with some discussions. We first observe some behaviors of upper bounds and lower bounds individually, and then put them together at last.

One important parameter in the computation of upper bounds is $\delta$. As Figure 2.3 shows, choosing small $\delta$ causes zigzag lines for low SNR and choosing $\delta$ large leads to failures of numerical computation for high SNR. This problem basically comes from the

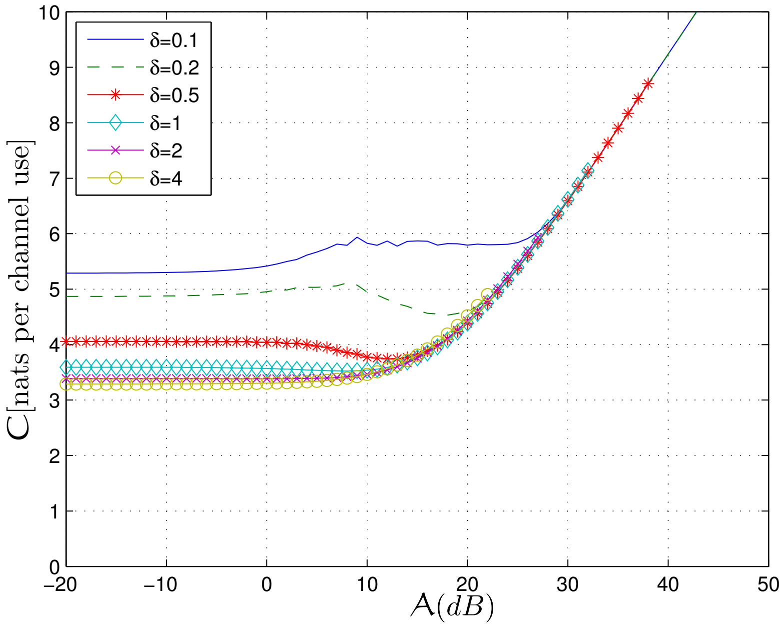


Figure 2.3: Upper bounds with $\alpha = 0.2$, $\lambda_0 = 12$

evaluation of $\mathsf{E}[p_X]$, where $p_x$ is defined in (2.7). This behavior always appears whatever power constraints we have.

We also provide upper bounds with different $\lambda_0$ and $\alpha$ in Figure 2.4 and Figure 2.5. It can be observed in Figure 2.4 that the upper bounds on capacity at high SNR become small for large dark current $\lambda_0$, which is compatible to the idea that noise makes channel worse. In Figure 2.5 we show upper bounds of (2.17) and (2.22) based on different power constraints. In comparison with [1], our upper bounds here are valid for finite $\mathsf{A}$. However, in [1] it is proved that the upper bounds become the same asymptotically with $\alpha \geq \frac{1}{3}$. Our bounds could be less good for large $\mathsf{A}$ since they don't reach that statement.

For the lower bounds, $r$ is an important parameter, we can reach a better bound for low SNR by choosing $r$ properly. As shown in Figure 2.6, $r = \frac{1}{2}$ is the optimal solution among our options. So we will use this choice of $r$ for the following plots. The optimal choice of $r$ is not found in this thesis, but we will provide some principles later.

A comparison of lower bounds with and without dark current is shown in Figure 2.7. It's obvious that the bounds without dark current is tighter than ones with dark current. However, it's shown in Figure 2.8 that the lower bounds are good for high SNR since they coincide with the asymptotic upper bounds from [1]. Besides, Figure 2.9 depicts both upper and lower bounds on capacity. We can observe that our upper bound is loose for high SNR. Nevertheless, we provide bounds for finite power.

Figure 2.10 shows bounds with only average power constraint.

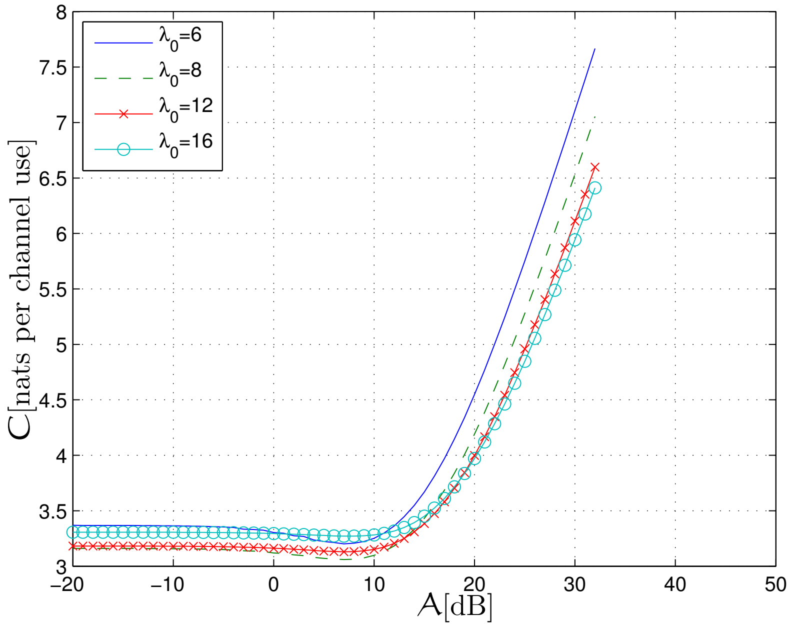


Figure 2.4: Upper bounds with $\alpha = 0.2$, $\delta = 1$

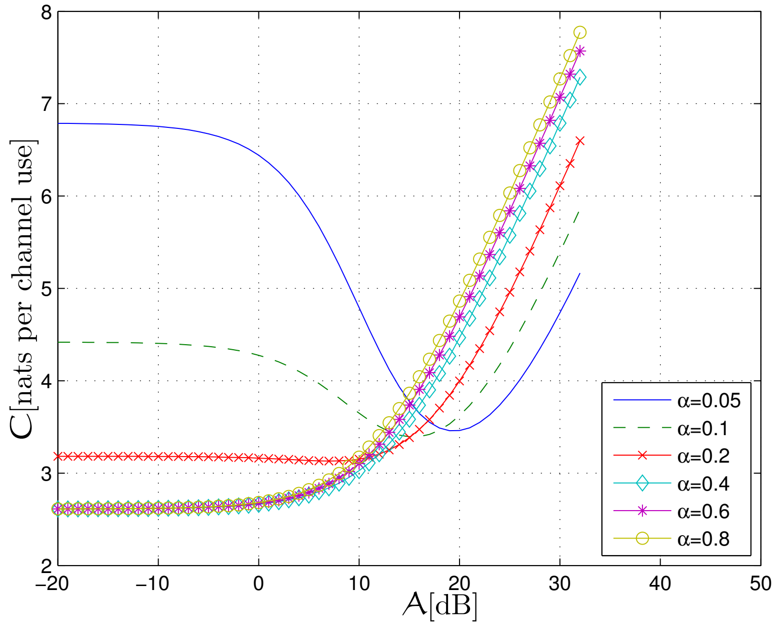


Figure 2.5: Upper bounds with $\lambda_0 = 12$, $\delta = 1$

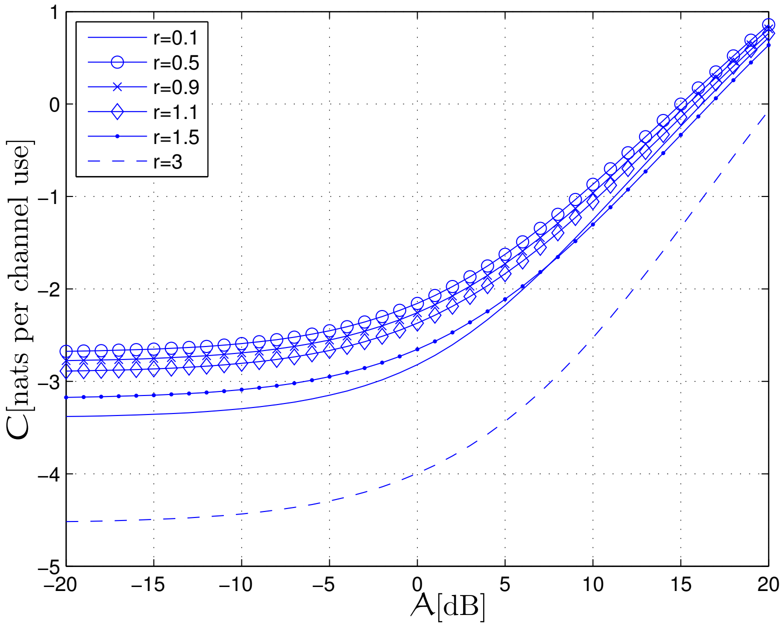


Figure 2.6: Lower bounds with $\alpha = 0.2$, $\lambda_0 = 12$

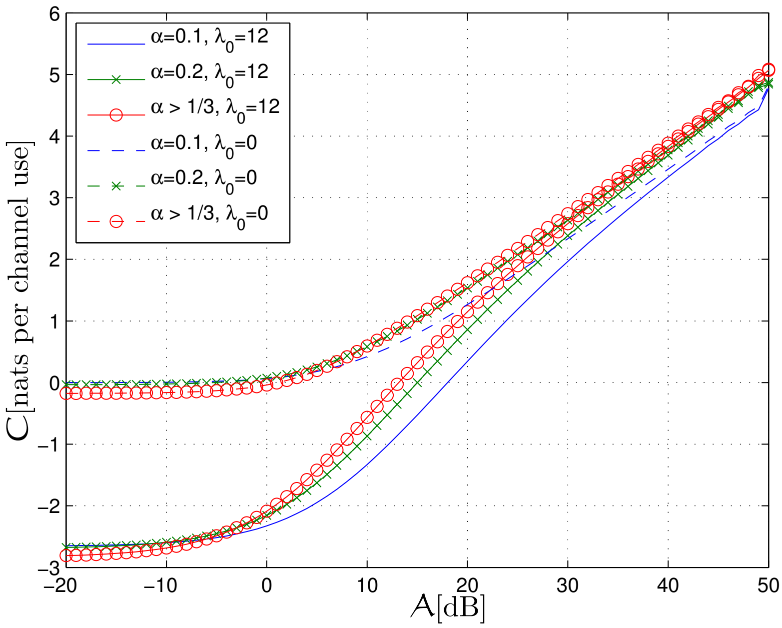


Figure 2.7: Lower bounds with different $\lambda_0$, $r = 0.5$

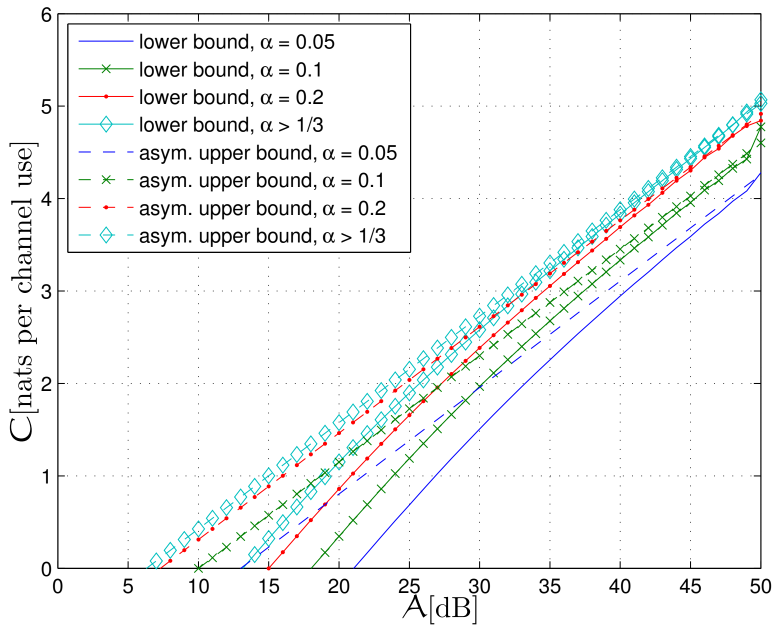


Figure 2.8: Lower bounds with dark current and asymptotic upper bounds in [1], $\lambda_0 = 12$

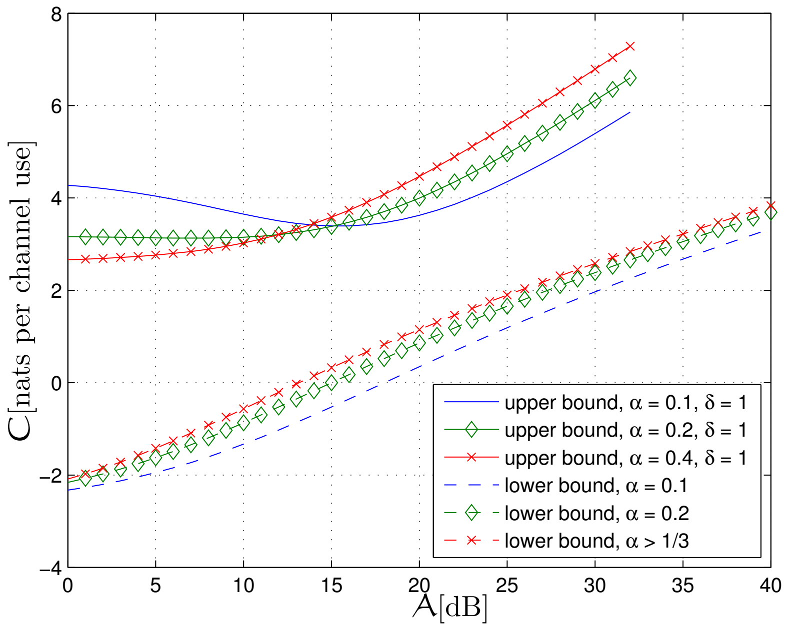


Figure 2.9: Lower bounds and upper bounds with dark current $\lambda_0 = 12$

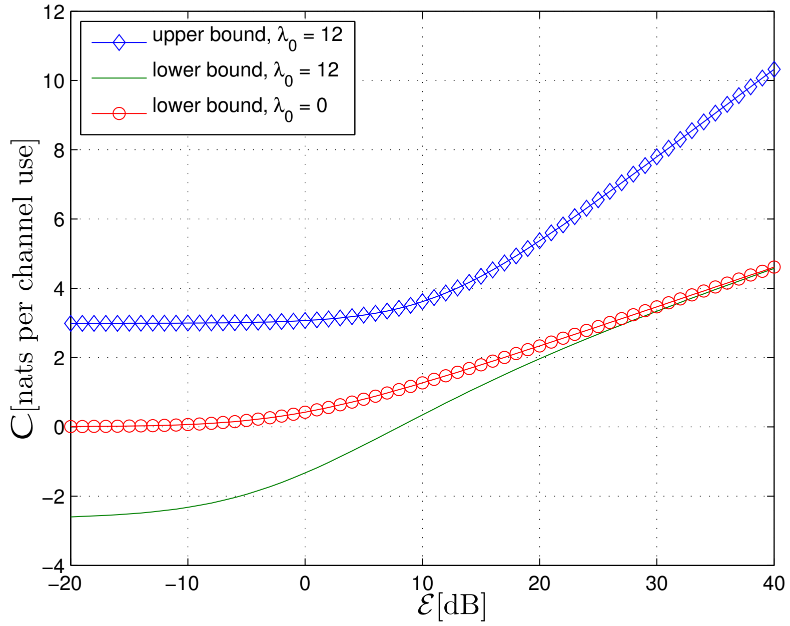


Figure 2.10: Lower bounds and upper bounds with only average power constraint

# Chapter 3

# Derivation of the Upper Bound

## 3.1 Overview

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

- To evaluate an upper bound of capacity, we refer to the duality of mutual information and propose the following proposition:

  **Proposition 8.** *Assume a channel* $\tilde{W}(\cdot|\cdot)$ *with input alphabet* $\mathcal{X} = \mathbb{R}_0^+$ *and output alphabet* $\mathcal{Y} = \mathbb{R}_0^+$*. Then for an arbitrary distribution* $R(\cdot)$ *over the channel output alphabet, the channel capacity is upper-bounded by*

  $$\mathsf{C} \leq \mathsf{E}_{Q^*}\left[D(\tilde{W}'(\cdot|X)\|R'(\cdot))\right] \tag{3.1}$$

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

  The challenge of using (3.1) lies in a clever choice of the arbitrary law $R'(\cdot)$ that will lead to a good upper bound. Moreover, note that the bound (3.1) still contains an expectation over the (unknown) capacity-achieving input distribution $Q^*(\cdot)$. To handle this expectation we will use Jensen's inequality with some assumptions based on channel noise, *i.e.*, dark current.

- We will refer to [3] and adapt their bounds on the entropy $H(N_\lambda)$ and expected logarithm $\mathsf{E}[\log N_\lambda]$ to our channel, where the Poisson law is given in (2.1).

- One difficulty of the Poisson channel model (1.1) is that while we have a continuous input, the output is discrete. This complicates the application of the technique explained in Proposition 8 considerably. To circumvent this problem we slightly change the channel model without changing its capacity value. The idea is to add some independent continuous noise $U$ to the channel output $Y$ that is uniformly distributed between 0 and 1, *i.e.*,

  $$\tilde{Y} \triangleq Y + U \tag{3.2}$$

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

$$Y = \lfloor \tilde{Y} \rfloor \tag{3.3}$$

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

- To evaluate the expectation in (3.1) over the unknown capacity-achieving input distribution $Q^*(\cdot)$ we need the following trick. We will further bound the expectation with the aid of Jensen's inequality, which states that

$$\mathsf{E}[f(X)] \leq f\left(\mathsf{E}[X]\right) \tag{3.4}$$

for a concave function $f(\cdot)$, *i.e.*, $\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} \leq 0$.

In the following part of this chapter we will focus on the case that the average-to-peak power ratio is between zero and one third, *i.e.*,

$$\alpha = \frac{\mathcal{E}}{\mathsf{A}} \in \left(0, \frac{1}{3}\right) \tag{3.5}$$

where the other cases are derived with similar steps in Appendix A.

## 3.2 Mathematical Preliminaries

We start with some preliminaries to our channel model (1.1). We will first state some theorems we need in this thesis and then make some adaption.

### 3.2.1 Bounds on the Entropy of the Poisson Law

Recall that $N_\lambda$ denotes a Poisson random variable with mean $\lambda$ where the mass function is defined in (2.1). We then have the following lemmas.

**Lemma 9.** *For $\lambda \geq 0$, the entropy of a Poisson random variable $H(\lambda)$ can be bounded as*

$$-\gamma_m(\lambda) \leq H(\lambda) - \frac{1}{2}\log(2\pi\lambda) - \frac{1}{2} - \beta_m(\lambda) \leq 0 \tag{3.6}$$

*where*

$$\gamma_m(\lambda) = \sum_{k=1}^{2m-1} \frac{b(m,k)}{\lambda^k} \tag{3.7}$$

$$\beta_m(\lambda) = \sum_{k=m}^{2m} \frac{a(m,k)}{\lambda^k} \tag{3.8}$$

*and $b(m,k)$, $a(m,k)$ are some constants.*

*Proof.* See [3]. □

**Lemma 10.** *Assume the kth central moment of the Poisson distribution is*

$$\mu_k(\lambda) = \mathsf{E}\left[(Y-\lambda)^k\right] \tag{3.9}$$

*Then we can state*

$$0 \le \mathsf{E}\left[\log\frac{Y+1}{\lambda}\right] - \sum_{k=2}^{2m+1}\frac{(-1)^k\mu_k(\lambda)}{k(k-1)\lambda^k} \le \frac{\mu_{2m+2}(\lambda)}{(2m+1)\lambda^{2m+2}} \tag{3.10}$$

*Proof.* See [3]. □

The above lemmas bound the entropy and log-expectation of Poisson distribution with polynomial. According to the integer $m$ we choose, we can gain different coefficients for the bounds. The bounds become tighter with larger $m$ and longer formula.

We choose $m = 1$ for our following derivation and adapt its form to our model.

$$H(Y|X=x) \le \frac{1}{2}\log(2\pi(x+\lambda_0)) + \frac{1}{2} + \frac{1}{6}(x+\lambda_0)^{-1} \tag{3.11}$$

$$H(Y|X=x) \ge \frac{1}{2}\log(2\pi(x+\lambda_0)) + \frac{1}{2} - \frac{5}{6}(x+\lambda_0)^{-1} - \frac{1}{6}(x+\lambda_0)^{-2} \tag{3.12}$$

$$\mathsf{E}[\log Y + 1 \,|\, X=x] \le \log(x+\lambda_0) + \frac{1}{2}(x+\lambda_0)^{-1} + \frac{5}{6}(x+\lambda_0)^{-2} + \frac{1}{3}(x+\lambda_0)^{-3} \tag{3.13}$$

### 3.2.2 A Poisson Channel with Continuous Output

In the following we define an adapted Poisson channel model which has a continuous output. To this end, let $Y$ be the output of a Poisson channel with input $x$ as given in (1.1). We define a new random variable with PDF

$$\tilde{W}'(\tilde{y}|x) = W(\lfloor\tilde{y}\rfloor|x) = e^{-(x+\lambda_0)}\frac{(x+\lambda_0)^{\lfloor\tilde{y}\rfloor}}{\lfloor\tilde{y}\rfloor!}, \qquad \tilde{y}\ge 0, x\ge 0, \lambda_0 \ge 0 \tag{3.14}$$

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

**Lemma 11.** *Let the random variables $Y$, $\tilde{Y}$, and $X$ be defined as above. Then*

*a)*

$$I(X;\tilde{Y}) = I(X;Y) \tag{3.15}$$

*b)*

$$h(\tilde{Y}|X=x) = H(Y|X=x) \tag{3.16}$$

*c)*

$$h(\tilde{Y}) = H(Y) \tag{3.17}$$

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

$$X \multimap\!\!- Y \multimap\!\!- \tilde{Y} \multimap\!\!- Y' \tag{3.18}$$

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

$$I(X;Y) \geq I(X;\tilde{Y}) \geq I(X;Y') \tag{3.19}$$

On the other hand, $Y' = Y$, thus Part a) is proven.

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

$$h(\tilde{Y}|X=x) = -\int_0^\infty \tilde{W}'(\tilde{y}|x) \log \tilde{W}'(\tilde{y}|x)\, \mathrm{d}\tilde{y} \tag{3.20}$$

$$= -\sum_{y=0}^{\infty} \int_y^{y+1} W(\lfloor \tilde{y} \rfloor|x) \log W(\lfloor \tilde{y} \rfloor|x)\, \mathrm{d}\tilde{y} \tag{3.21}$$

$$= -\sum_{y=0}^{\infty} W(y|x) \log W(y|x) \int_y^{y+1} \mathrm{d}\tilde{y} \tag{3.22}$$

$$= -\sum_{y=0}^{\infty} W(y|x) \log W(y|x) \tag{3.23}$$

$$= H(Y|X=x) \tag{3.24}$$

Part c) now follows from a) and b). □

Note that the expected logarithm of the continuous Poisson distribution can be bounded as follows:

$$\mathsf{E}\left[\log \tilde{Y} \,\middle|\, X=x\right] \leq \mathsf{E}[\log(Y+1) \,|\, X=x] \tag{3.25}$$

## 3.3 Proof of the Upper Bound (2.17)

The derivation of (2.17) is based on Proposition 8 and the following choice of an output distribution $R'(\cdot)$:

$$R'(\tilde{y}) \triangleq \begin{cases} p_{\mathsf{A}} \cdot \dfrac{\tilde{y}^{\nu-1} e^{-\frac{\tilde{y}}{\beta}}}{\beta^\nu \gamma\left(\nu, \frac{(\mathsf{A}+\lambda_0)(1+\delta)}{\beta}\right)}, & \forall 0 \leq \tilde{y} \leq (\mathsf{A}+\lambda_0)(1+\delta) \\ (1-p_{\mathsf{A}}) \cdot e^{-(\tilde{y}-(\mathsf{A}+\lambda_0)(1+\delta))}, & \forall \tilde{y} > (\mathsf{A}+\lambda_0)(1+\delta) \end{cases} \tag{3.26}$$

where $p_{\mathsf{A}}$ is a variable of $\mathsf{A}$ defined in (2.7) and $\nu$, $\beta$, $\delta$ are free parameters that will be specified later.

With this choice we get

$$\mathsf{C} \leq \mathsf{E}_{Q*}\Big[D\big(\tilde{W}'(\tilde{y}|x)\,\big\|R'(\tilde{y})\big)\Big] \tag{3.27}$$

$$\begin{aligned}
&\leq \mathsf{E}_{Q*}\Bigg[-h(\tilde{Y}|X=x)+\log\frac{1}{p_{\mathsf{A}}}\cdot\Pr\big[\tilde{Y}\leq(\mathsf{A}+\lambda_0)(1+\delta)\big|X=x\big]\\
&\qquad+(1-\nu)\mathsf{E}\Big[\log\tilde{Y}\;\Big|\;X=x\Big]+\left(\nu\log\beta+\log\gamma\left(\nu,\frac{(\mathsf{A}+\lambda_0)(1+\delta)}{\beta}\right)\right)\\
&\qquad\cdot\Pr\big[\tilde{Y}\leq(\mathsf{A}+\lambda_0)(1+\delta)\big|X=x\big]+\frac{1}{\beta}\Bigg(x+\lambda_0+\underbrace{\frac{1}{2}}_{\mathsf{E}[U]}\Bigg)\\
&\qquad+\log\frac{1}{1-p_{\mathsf{A}}}\cdot\Pr\big[\tilde{Y}>(\mathsf{A}+\lambda_0)(1+\delta)\big|X=x\big]\\
&\qquad+\int_{(\mathsf{A}+\lambda_0)(1+\delta)}^{\infty}\tilde{W}'(\tilde{y}|x)\,(\tilde{y}-(\mathsf{A}+\lambda_0)(1+\delta))\,\mathrm{d}\tilde{y}\Bigg]
\end{aligned} \tag{3.28}$$

The above equation can be separated into four parts

$$\mathsf{C}_{\mathrm{a}} \triangleq -h(\tilde{Y}|X=x)+(1-\nu)\mathsf{E}\Big[\log\tilde{Y}\;\Big|\;X=x\Big] \tag{3.29}$$

$$\begin{aligned}
\mathsf{C}_{\mathrm{b}} &\triangleq \left(\nu\log\beta+\log\gamma\left(\nu,\frac{(\mathsf{A}+\lambda_0)(1+\delta)}{\beta}\right)\right)\Pr\big[\tilde{Y}\leq(\mathsf{A}+\lambda_0)(1+\delta)\big|X=x\big]\\
&\quad+\frac{1}{\beta}\left(x+\lambda_0+\frac{1}{2}\right)
\end{aligned} \tag{3.30}$$

$$\begin{aligned}
\mathsf{C}_{\mathrm{c}} &\triangleq \log\frac{1}{p_{\mathsf{A}}}\cdot\Pr\big[\tilde{Y}\leq(\mathsf{A}+\lambda_0)(1+\delta)\big|X=x\big]\\
&\quad+\log\frac{1}{1-p_{\mathsf{A}}}\cdot\Pr\big[\tilde{Y}>(\mathsf{A}+\lambda_0)(1+\delta)\big|X=x\big]
\end{aligned} \tag{3.31}$$

$$\mathsf{C}_{\mathrm{d}} \triangleq \int_{(\mathsf{A}+\lambda_0)(1+\delta)}^{\infty}\tilde{W}'(\tilde{y}|x)\,(\tilde{y}-(\mathsf{A}+\lambda_0)(1+\delta))\,\mathrm{d}\tilde{y} \tag{3.32}$$

and we will consider each term individually. We first start with $\mathsf{C}_{\mathrm{a}}$.

$$\mathsf{E}\Big[\log\tilde{Y}\;\Big|\;X=x\Big]$$

$$= \int_0^{\infty}\log\tilde{y}\cdot e^{-(x+\lambda_0)}\frac{(x+\lambda_0)^{\lfloor\tilde{y}\rfloor}}{\lfloor\tilde{y}\rfloor!}\,\mathrm{d}\tilde{y} \tag{3.33}$$

$$= \sum_{k=0}^{\infty}\int_k^{k+1}\log\tilde{y}\cdot e^{-(x+\lambda_0)}\frac{(x+\lambda_0)^k}{k!}\,\mathrm{d}\tilde{y} \tag{3.34}$$

$$= \sum_{k=0}^{\infty}e^{-(x+\lambda_0)}\frac{(x+\lambda_0)^k}{k!}\int_k^{k+1}\log\tilde{y}\,\mathrm{d}\tilde{y} \tag{3.35}$$

$$= \sum_{k=0}^{\infty}e^{-(x+\lambda_0)}\frac{(x+\lambda_0)^k}{k!}\left(\log(k+1)+k\log\left(1+\frac{1}{k}\right)-1\right) \tag{3.36}$$

$$= \mathsf{E}[\log(Y+1)\,|\,X=x]+\mathsf{E}\left[Y\log\left(1+\frac{1}{Y}\right)\;\middle|\;X=x\right]-1 \tag{3.37}$$

$$\leq \mathsf{E}[\log(Y+1) \mid X=x] + \mathsf{E}[Y \mid X=x] \log\left(1+\frac{1}{\mathsf{E}[Y \mid X=x]}\right) - 1 \tag{3.38}$$

$$= \mathsf{E}[\log(Y+1) \mid X=x] + (x+\lambda_0)\log\left(1+\frac{1}{x+\lambda_0}\right) - 1 \tag{3.39}$$

Hence, applying Lemma 9 and Lemma 10, we get

$$\mathsf{C}_{\mathrm{a}} = -h(\tilde{Y}|X=x) + (1-\nu)\mathsf{E}\left[\log \tilde{Y} \,\middle|\, X=x\right] \tag{3.40}$$

$$\leq -h(\tilde{Y}|X=x) + (1-\nu)\Bigg(\mathsf{E}[\log(Y+1)|X=x] + (x+\lambda_0)\log\left(1+\frac{1}{x+\lambda_0}\right) - 1\Bigg) \tag{3.41}$$

$$\leq -\frac{1}{2}\log(2\pi e) + \left(\frac{1}{2}-\nu\right)\log(x+\lambda_0) + \left(\frac{5}{6}(x+\lambda_0)^{-1} + \frac{1}{6}(x+\lambda_0)^{-2}\right) + (1-\nu)\Bigg(\frac{1}{2}(x+\lambda_0)^{-1} + \frac{5}{6}(x+\lambda_0)^{-2} + \frac{1}{3}(x+\lambda_0)^{-3} + (x+\lambda_0)\log\left(1+\frac{1}{x+\lambda_0}\right) - 1\Bigg) \tag{3.42}$$

We observe that $\mathsf{C}_{\mathrm{a}}$ is a function of $x$ given specific $\nu$ and $\lambda_0$. So we try to explore some properties of it with respect to $\nu$ and $\lambda_0$. First, we split $\mathsf{C}_{\mathrm{a}}$ into two parts and define a new variable $\tilde{x}$ as follows

$$\tilde{x} \triangleq x+\lambda_0 \tag{3.43}$$

$$\mathsf{C}_{\mathrm{a},1} \triangleq (1-\nu)\left((x+\lambda_0)\log\left(1+\frac{1}{x+\lambda_0}\right) - 1\right) \tag{3.44}$$

$$= (1-\nu)\left(\tilde{x}\log\left(1+\frac{1}{\tilde{x}}\right) - 1\right) \tag{3.45}$$

$$\mathsf{C}_{\mathrm{a},2} \triangleq -\frac{1}{2}\log(2\pi e) + \left(\frac{1}{2}-\nu\right)\log(x+\lambda_0) + \left(\frac{5}{6}(x+\lambda_0)^{-1} + \frac{1}{6}(x+\lambda_0)^{-2}\right) + (1-\nu)\left(\frac{1}{2}(x+\lambda_0)^{-1} + \frac{5}{6}(x+\lambda_0)^{-2} + \frac{1}{3}(x+\lambda_0)^{-3}\right) \tag{3.46}$$

$$= -\frac{1}{2}\log(2\pi e) + \left(\frac{1}{2}-\nu\right)\log\tilde{x} + \left(\frac{5}{6}\tilde{x}^{-1} + \frac{1}{6}\tilde{x}^{-2}\right) + (1-\nu)\left(\frac{1}{2}\tilde{x}^{-1} + \frac{5}{6}\tilde{x}^{-2} + \frac{1}{3}\tilde{x}^{-3}\right) \tag{3.47}$$

Now we consider the first and second derivative of $\mathsf{C}_{\mathrm{a},1}$:

$$\frac{\partial \mathsf{C}_{\mathrm{a},1}}{\partial \tilde{x}} = (1-\nu)\left(\log\frac{1+\tilde{x}}{\tilde{x}} - \frac{1}{1+\tilde{x}}\right) \tag{3.48}$$

$$\frac{\partial^2 \mathsf{C}_{\mathrm{a},1}}{\partial \tilde{x}^2} = (1-\nu)\frac{-1}{\tilde{x}(1+\tilde{x})^2} \tag{3.49}$$

We can observe that $\mathsf{C}_{\mathrm{a},1}$ is monotonically increasing and concave as long as

$$\nu \leq 1 \tag{3.50}$$

For $\mathsf{C}_{\mathrm{a},2}$ we derive

$$\frac{\partial \mathsf{C}_{\mathrm{a},2}}{\partial \tilde{x}} = \left(\frac{1}{2}-\nu\right)\tilde{x}^{-1} - \left(\frac{5}{6}\tilde{x}^{-2}+\frac{1}{3}\tilde{x}^{-3}\right) - (1-\nu)\left(\frac{1}{2}\tilde{x}^{-2}+\frac{5}{3}\tilde{x}^{-3}+\tilde{x}^{-4}\right) \quad (3.51)$$

$$= \left(\frac{1}{2}-\nu\right)\tilde{x}^{-1} - \left(\frac{8}{6}-\frac{\nu}{2}\right)\tilde{x}^{-2} - \left(2-\frac{5}{3}\nu\right)\tilde{x}^{-3} - (1-\nu)\,\tilde{x}^{-4} \quad (3.52)$$

$$\frac{\partial^2 \mathsf{C}_{\mathrm{a},2}}{\partial \tilde{x}^2} = -\left(\frac{1}{2}-\nu\right)\tilde{x}^{-2} + \left(\frac{5}{3}\tilde{x}^{-3}+\tilde{x}^{-4}\right) + (1-\nu)\left(\tilde{x}^{-3}+5\tilde{x}^{-4}+4\tilde{x}^{-5}\right) \quad (3.53)$$

$$= \left(\nu-\frac{1}{2}\right)\tilde{x}^{-2} + \left(\frac{8}{3}-\nu\right)\tilde{x}^{-3} + (6-5\nu)\,\tilde{x}^{-4} + (4-4\nu)\,\tilde{x}^{-5} \quad (3.54)$$

We wish to choose $\nu$ such that $\mathsf{C}_{\mathrm{a},2}$ becomes monotonically increasing and concave, so we assume that

$$\frac{\partial \mathsf{C}_{\mathrm{a},2}}{\partial \tilde{x}} \geq 0 \quad (3.55)$$

$$\frac{\partial^2 \mathsf{C}_{\mathrm{a},2}}{\partial \tilde{x}^2} \leq 0 \quad (3.56)$$

where we can find that for $\tilde{x}$ large enough($e.g,\ \tilde{x}=5$), a possible value of $\nu$ can be specified satisfying (3.55) and (3.56). Meanwhile, such $\nu$ also satisfies (3.50), *i.e.*

$$\nu \leq \frac{3\tilde{x}^{-1}-8\tilde{x}^{-2}-12\tilde{x}^{-3}-6\tilde{x}^{-4}}{6\tilde{x}^{-1}-3\tilde{x}^{-2}-10\tilde{x}^{-3}-6\tilde{x}^{-4}} \quad (3.57)$$

$$\nu \leq \frac{3\tilde{x}^{-2}+16\tilde{x}^{-3}+36\tilde{x}^{-4}+24\tilde{x}^{-5}}{6\tilde{x}^{-2}+6\tilde{x}^{-3}+30\tilde{x}^{-4}+24\tilde{x}^{-5}} \quad (3.58)$$

We conclude that for a Poisson channel with large enough dark current $\lambda_0$, the function $\mathsf{C}_{\mathrm{a}}$ can always be set as monotonically increasing and concave for a specified value of $\nu$. And we can find such $\nu$ easily with the help of (3.50), (3.57) and (3.58).

Now we apply these two properties to our derivation.

$$\begin{aligned}\mathsf{E}_{Q^*}[\mathsf{C}_{\mathrm{a}}] \triangleq \mathsf{E}_{Q*}\Bigg[&-\frac{1}{2}\log(2\pi e) + \left(\frac{1}{2}-\nu\right)\log(X+\lambda_0) + \left(\frac{5}{6}(X+\lambda_0)^{-1}+\frac{1}{6}(X+\lambda_0)^{-2}\right)\\ &+ (1-\nu)\left(\frac{1}{2}(X+\lambda_0)^{-1}+\frac{5}{6}(X+\lambda_0)^{-2}+\frac{1}{3}(X+\lambda_0)^{-3}\right.\\ &\left.+ (X+\lambda_0)\log\left(1+\frac{1}{X+\lambda_0}\right)-1\right)\Bigg]\end{aligned} \quad (3.59)$$

$$\begin{aligned}&\leq -\frac{1}{2}\log(2\pi e) + \left(\frac{1}{2}-\nu\right)\log(\mathsf{E}[X]+\lambda_0) + \left(\frac{5}{6}(\mathsf{E}[X]+\lambda_0)^{-1}+\frac{1}{6}(\mathsf{E}[X]+\lambda_0)^{-2}\right)\\ &+ (1-\nu)\left(\frac{1}{2}(\mathsf{E}[X]+\lambda_0)^{-1}+\frac{5}{6}(\mathsf{E}[X]+\lambda_0)^{-2}+\frac{1}{3}(\mathsf{E}[X]+\lambda_0)^{-3}\right.\\ &\left.+ (\mathsf{E}[X]+\lambda_0)\log\left(1+\frac{1}{\mathsf{E}[X]+\lambda_0}\right)-1\right)\end{aligned} \quad (3.60)$$

$$\begin{aligned}&\leq -\frac{1}{2}\log(2\pi e) + \left(\frac{1}{2}-\nu\right)\log(\alpha\mathsf{A}+\lambda_0) + \left(\frac{5}{6}(\alpha\mathsf{A}+\lambda_0)^{-1}+\frac{1}{6}(\alpha\mathsf{A}+\lambda_0)^{-2}\right)\\ &+ (1-\nu)\left(\frac{1}{2}(\alpha\mathsf{A}+\lambda_0)^{-1}+\frac{5}{6}(\alpha\mathsf{A}+\lambda_0)^{-2}+\frac{1}{3}(\alpha\mathsf{A}+\lambda_0)^{-3}\right.\end{aligned}$$

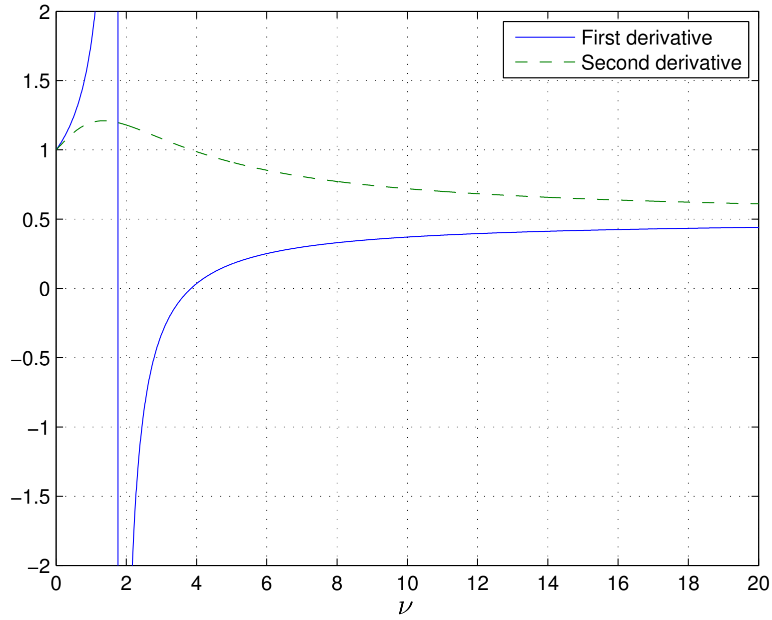


Figure 3.11: Derivatives of $\mathsf{C}_{a,2}$ for different $\nu$



$$+ (\alpha\mathsf{A} + \lambda_0)\log\left(1 + \frac{1}{\alpha\mathsf{A} + \lambda_0}\right) - 1\Bigg) \tag{3.61}$$

Note that for (3.60) we apply Jensen's inequality (3.4) and for (3.61) we use the fact that $\mathsf{C}_a$ is monotonically increasing.

Next, we bound $\mathsf{C}_b$ as follows:

$$\mathsf{C}_b = \left(\nu\log\beta + \log\gamma\left(\nu, \frac{(\mathsf{A}+\lambda_0)(1+\delta)}{\beta}\right)\right)\underbrace{\Pr\left[\tilde{Y} \leq (\mathsf{A}+\lambda_0)(1+\delta)|X = x\right]}_{=1-p_{\mathrm{T},x}}$$

$$+ \frac{1}{\beta}\left(x + \lambda_0 + \frac{1}{2}\right) \tag{3.62}$$

$$= (1 - p_{\mathrm{T},x})\left(\nu\log\beta + \log\gamma\left(\nu, \frac{(\mathsf{A}+\lambda_0)(1+\delta)}{\beta}\right)\right) + \frac{1}{\beta}\left(x + \lambda_0 + \frac{1}{2}\right) \tag{3.63}$$

We choose the free parameter $\beta$ as

$$\beta \triangleq \frac{(\mathsf{A}+\lambda_0)(1+\delta)}{\mu} \tag{3.64}$$

where $\mu$ is the solution of

$$\alpha = \frac{1}{2\mu} - \frac{e^{-\mu}}{\sqrt{\mu\pi}\cdot\operatorname{erf}\left(\sqrt{\mu}\right)} \tag{3.65}$$

Note that this is a suboptimal choice.

Before going further, we first look at some properties related to $p_{\mathrm{T},x}$.

$$p_{\mathrm{T},x} \triangleq \Pr\left[\tilde{Y} > (\mathsf{A}+\lambda_0)(1+\delta) \middle| X = x\right] \tag{3.66}$$

$$= \int_{(\mathsf{A}+\lambda_0)(1+\delta)}^{\infty} e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^{\lfloor \tilde{y} \rfloor}}{\lfloor \tilde{y} \rfloor !} \,\mathrm{d}\tilde{y} \tag{3.67}$$

$$= \sum_{y=\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor}^{\infty} e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^{y}}{y!} - e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^{\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor}}{\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor !} \left((\mathsf{A}+\lambda_0)(1+\delta) - \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right) \tag{3.68}$$

$$= \frac{\gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor, x+\lambda_0\right)}{\Gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right)} - \Pr\left[N_{x+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right] \cdot \left((\mathsf{A}+\lambda_0)(1+\delta) - \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right) \tag{3.69}$$

where

$$\sum_{y=\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor}^{\infty} e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^{y}}{y!} = \frac{\gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor, x+\lambda_0\right)}{\Gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right)} \tag{3.70}$$

Note that

$$\frac{\partial^2}{\partial x^2}\gamma(x,n) = x^{n-2}e^{-x}(n-1-x) \tag{3.71}$$

Hence, $\gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor, x+\lambda_0\right)$ is convex as long as

$$\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor - 1 - (\mathsf{A}+\lambda_0) \geq 0 \tag{3.72}$$

*i.e.*, $\delta$ is big enough. Moreover, the second term of $p_{\mathrm{T},x}$ is decreasing in $x$. According to the above properties and Jensen's inequality,

$$\frac{\gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor, x+\lambda_0\right)}{\Gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right)} \geq \frac{\gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor, \alpha\mathsf{A}+\lambda_0\right)}{\Gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right)} \tag{3.73}$$

$$\Pr\left[N_{x+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right] \leq \Pr\left[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right] \tag{3.74}$$

so we can derive

$$\mathsf{E}_{Q^*}[p_{\mathrm{T},X}] \geq \frac{\gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor, \alpha\mathsf{A}+\lambda_0\right)}{\Gamma\left(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right)} - \Pr\left[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right] \cdot \left((\mathsf{A}+\lambda_0)(1+\delta) - \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right) \tag{3.75}$$

which leads to

$$\mathsf{E}_{Q^*}[\mathsf{C}_{\mathrm{b}}] = \mathsf{E}_{Q^*}[1-p_{\mathrm{T},X}]\left(\nu \log\frac{(\mathsf{A}+\lambda_0)(1+\delta)}{\mu} + \log\gamma(\nu,\mu)\right) + \frac{\mu\left(\alpha\mathsf{A}+\lambda_0+\frac{1}{2}\right)}{(\mathsf{A}+\lambda_0)(1+\delta)} \tag{3.76}$$

$$
\begin{aligned}
&\leq \Bigg(1 - \frac{\gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor, \alpha\mathsf{A}+\lambda_0)}{\Gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor)} + \Pr\left[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right] \\
&\qquad \cdot \left((\mathsf{A}+\lambda_0)(1+\delta) - \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right)\Bigg) \left(\nu \log \frac{(\mathsf{A}+\lambda_0)(1+\delta)}{\mu} + \log\gamma(\nu,\mu)\right) \\
&\quad + \frac{\mu\left(\alpha\mathsf{A}+\lambda_0+\frac{1}{2}\right)}{(\mathsf{A}+\lambda_0)(1+\delta)}
\end{aligned} \tag{3.77}
$$

Now we consider $\mathsf{C}_\mathrm{c}$ and bound it as follows:

$$
\mathsf{E}_{Q^*}[\mathsf{C}_\mathrm{c}] = \mathsf{E}_{Q^*}[1-p_{\mathrm{T},X}] \log\frac{1}{1-p_{\mathrm{T},\mathsf{A}}} + \mathsf{E}_{Q^*}[p_{\mathrm{T},X}] \log\frac{1}{p_{\mathrm{T},\mathsf{A}}} \tag{3.78}
$$

$$
\begin{aligned}
&\leq \Bigg(1 - \frac{\gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor, \alpha\mathsf{A}+\lambda_0)}{\Gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor)} + \Pr\left(N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right) \\
&\qquad \cdot \left((\mathsf{A}+\lambda_0)(1+\delta) - \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right)\Bigg) \log\frac{1}{1-p_{\mathrm{T},\mathsf{A}}} + p_{\mathrm{T},\mathsf{A}} \log\frac{1}{p_{\mathrm{T},\mathsf{A}}}
\end{aligned} \tag{3.79}
$$

where for the first part we use (3.75) and for the second part we use the fact that $p_{\mathrm{T},x}$ is decreasing in $x$.

Finally we bound $\mathsf{C}_\mathrm{d}$ as follows:

$$
\mathsf{E}_{Q^*}[\mathsf{C}_\mathrm{d}] = \mathsf{E}_{Q^*}\left[\int_{(\mathsf{A}+\lambda_0)(1+\delta)}^{\infty} \tilde{W}'(\tilde{y}|x) \underbrace{(\tilde{y} - (\mathsf{A}+\lambda_0)(1+\delta))}_{\geq 0} \,\mathrm{d}\tilde{y}\right] \tag{3.80}
$$

$$
\leq \int_{(\mathsf{A}+\lambda_0)(1+\delta)}^{\infty} (\tilde{y} - (\mathsf{A}+\lambda_0)(1+\delta))\, \tilde{W}'(\tilde{y}|\mathsf{A}) \,\mathrm{d}\tilde{y} \tag{3.81}
$$

$$
= \int_{(\mathsf{A}+\lambda_0)(1+\delta)}^{\infty} (\tilde{y} - (\mathsf{A}+\lambda_0)(1+\delta))\, e^{-(\mathsf{A}+\lambda_0)} \frac{(\mathsf{A}+\lambda_0)^{\lfloor\tilde{y}\rfloor}}{\lfloor\tilde{y}\rfloor!} \,\mathrm{d}\tilde{y} \tag{3.82}
$$

$$
\begin{aligned}
&= \sum_{k=\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor+1}^{\infty} \int_{k}^{k+1} (\tilde{y} - (\mathsf{A}+\lambda_0)(1+\delta))\, e^{-(\mathsf{A}+\lambda_0)} \frac{(\mathsf{A}+\lambda_0)^{k}}{k!} \,\mathrm{d}\tilde{y} \\
&\quad + e^{-(\mathsf{A}+\lambda_0)} \frac{(\mathsf{A}+\lambda_0)^{\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor}}{\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor} \\
&\qquad \cdot \int_{(\mathsf{A}+\lambda_0)(1+\delta)}^{\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor+1} (\tilde{y} - (\mathsf{A}+\lambda_0)(1+\delta)) \,\mathrm{d}\tilde{y}
\end{aligned} \tag{3.83}
$$

$$
\begin{aligned}
&= \sum_{k=\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor+1}^{\infty} \left(k + \frac{1}{2} - (\mathsf{A}+\lambda_0)(1+\delta)\right) e^{-(\mathsf{A}+\lambda_0)} \frac{(\mathsf{A}+\lambda_0)^{k}}{k!} \\
&\quad + e^{-(\mathsf{A}+\lambda_0)} \frac{(\mathsf{A}+\lambda_0)^{\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor}}{\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor} \\
&\qquad \cdot \frac{1}{2}\Big((\mathsf{A}+\lambda_0)(1+\delta) - \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor - 1\Big)^2
\end{aligned} \tag{3.84}
$$

$$
\begin{aligned}
&= \sum_{k=\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor+1}^{\infty} \left(\frac{1}{2} - (\mathsf{A}+\lambda_0)(1+\delta)\right) e^{-(\mathsf{A}+\lambda_0)} \frac{(\mathsf{A}+\lambda_0)^k}{k!} \\
&\quad + \sum_{k=\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor+1}^{\infty} (\mathsf{A}+\lambda_0) e^{-(\mathsf{A}+\lambda_0)} \frac{(\mathsf{A}+\lambda_0)^{k-1}}{(k-1)!} \\
&\quad + \Pr[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor] \\
&\qquad \cdot \frac{1}{2}\Big((\mathsf{A}+\lambda_0)(1+\delta) - \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor - 1\Big)^2 && (3.85) \\
&= \sum_{k=\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor+1}^{\infty} \left(\frac{1}{2} - (\mathsf{A}+\lambda_0)\delta\right) e^{-(\mathsf{A}+\lambda_0)} \frac{(\mathsf{A}+\lambda_0)^k}{k!} \\
&\quad + (\mathsf{A}+\lambda_0) \cdot \Pr[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor] \\
&\quad + \Pr[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor] \\
&\qquad \cdot \frac{1}{2}\Big((\mathsf{A}+\lambda_0)(1+\delta) - \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor - 1\Big)^2 && (3.86) \\
&= \left(\frac{1}{2} - (\mathsf{A}+\lambda_0)\delta\right) \frac{\gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor + 1, \mathsf{A}+\lambda_0)}{\Gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor + 1)} \\
&\quad + \Pr[N_{\mathsf{A}+\lambda_0} = \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor] \cdot \Bigg((\mathsf{A}+\lambda_0) \\
&\qquad + \frac{1}{2}\Big((\mathsf{A}+\lambda_0)(1+\delta) - \lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor - 1\Big)^2\Bigg) && (3.87)
\end{aligned}
$$

where

$$\tilde{W}'(\tilde{y}|x) \le \tilde{W}'(\tilde{y}|\mathsf{A}) \quad \forall (x \le \mathsf{A}) \cap (y \ge (\mathsf{A}+\lambda_0)(1+\delta)) \tag{3.88}$$

Since we have bounded all parts of (3.28) individually, we can derive an upper bound of capacity (2.17). The derivation for the cases that with different power constraints follows along the same lines. We defer the details to Appendix A.

# Chapter 4

# Derivation of the Lower Bound

## 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 a specific input distribution

$$\mathsf{C} = \sup_{Q'(\cdot)} I(X;Y) \geq I(X;Y)|_{\text{for a specific } Q'(\cdot)} \tag{4.1}$$

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, it might be difficult to evaluate $I(X;Y)$ for such a $Q'(\cdot)$. Note that even for a relative "simple" input distribution the corresponding channel output $Y$ may be hard to compute, let alone $H(Y)$.

To avoid this problem we first lower-bound $H(Y)$ in terms of $h(X)$. Besides, we simplify the channel by reducing the dark current $\lambda_0$ to zero in some parts of the proof, which will help our derivation but also loose the results. And we use some numerical evaluations in our bounds.

## 4.2 Mathematical Preliminaries

Given $X = x$, we assume $Y \sim \mathcal{P}o(x + \lambda_0)$ is conditionally Poisson with rate $x + \lambda_0$,

$$W(y|x) = e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^y}{y!}, \quad y \in \mathbb{N}_0,\ x \geq 0,\ \lambda_0 \geq 0 \tag{4.2}$$

Then $Y$ can be written as $Y = Y_1 + Y_2$, , where $Y_1 \sim \mathcal{P}o(x)$ and $Y_2 \sim \mathcal{P}o(\lambda_0)$, $Y_1 \perp\!\!\!\perp Y_2$. Note that

$$\begin{aligned} H(Y) &= H(Y_1 + Y_2) && (4.3)\\ &\geq H(Y_1 + Y_2 | Y_2) && (4.4)\\ &= H(Y_1 | Y_2) && (4.5)\\ &= H(Y_1) && (4.6) \end{aligned}$$

Hence, we will assume that the dark current is zero in some parts of our derivations.

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

**Proposition 12.** *Let $Y$ be the output of a Poisson channel with input $X \geq 0$ and dark current $\lambda_0$ according to* (1.1)*. Assume that $X$ has a finite positive expectation $\mathsf{E}[X] > 0$. Then*

$$\begin{aligned} H(Y) &\geq h(X) - (\mathsf{E}[X] + r) \log \frac{\mathsf{E}[X]}{\mathsf{E}[X] + r} - r + (r-1)\mathsf{E}[\log X] \\ &\quad + \mathsf{E}[\log \Gamma(Y_1 + 1) - \log \Gamma(Y_1 + r)] \qquad (4.7) \\ &= h(X) - (\mathsf{E}[X] + r) \log \frac{\mathsf{E}[X]}{\mathsf{E}[X] + r} - r + (r-1)\mathsf{E}[\log X] \\ &\quad + I\{r \neq 1\} \cdot \int_0^1 \left( \frac{t^{r-1} - 1}{1-t} \mathsf{E}\left[e^{-X(1-t)}\right] + r - 1 \right) \frac{1}{\log t} \, \mathrm{d}t. \qquad (4.8) \end{aligned}$$

*where $r$ is a parameter we can choose and conditioned on $X = x$, $Y_1 \sim \mathcal{P}o(x)$.*

*Proof.* A proof is given in Appendix C. □

## 4.3 Proof of the Lower Bound

### 4.3.1 General Form

The first step of the derivation follows (4.1) using the definition of mutual information.

$$\mathsf{C} = \sup_{Q(\cdot)} I(X;Y) \geq I(X;Y)|_{\text{for a specific } Q'(\cdot)} \qquad (4.9)$$

$$= H(Y) - H(Y|X) \qquad (4.10)$$

for a good choice of $Q'(\cdot)$.

Note that the conditional entropy of the channel output given the channel input is given by

$$H(Y|X) = \mathsf{E}[X] + \lambda_0 - \mathsf{E}[(X + \lambda_0)\log(X + \lambda_0)] + \mathsf{E}[\log \Gamma(Y+1)] \qquad (4.11)$$

Since $X$ is nonnegative, we can derive

$$H(Y|X) = \mathsf{E}[X] + \lambda_0 - \mathsf{E}[(X + \lambda_0)\log(X + \lambda_0)] + \mathsf{E}[\log \Gamma(Y+1)] \qquad (4.12)$$

$$= \mathsf{E}[X] + \lambda_0 - \mathsf{E}[X \log(X + \lambda_0)] - \lambda_0 \mathsf{E}[\log(X + \lambda_0)] + \mathsf{E}[\log \Gamma(Y+1)] \qquad (4.13)$$

Moreover, we can derive a lower bound of $H(Y)$ according to Proposition 12.

According to the definition of mutual information (4.10), we can derive a lower bound

$$\begin{aligned} I(X;Y) &\geq h(X) - (\mathsf{E}[X] + r) \log \frac{\mathsf{E}[X]}{\mathsf{E}[X] + r} + (r-1)\mathsf{E}[\log X] + \mathsf{E}[(X + \lambda_0)\log(X + \lambda_0)] \\ &\quad - \mathsf{E}[X] - r - \lambda_0 + \mathsf{E}[\log \Gamma(Y_1 + 1) - \log \Gamma(Y_1 + r) - \log \Gamma(Y+1)] \qquad (4.14) \end{aligned}$$

where we can express

$$\begin{aligned}&\mathsf{E}[\log\Gamma(Y_1+1)-\log\Gamma(Y_1+r)-\log\Gamma(Y+1)]\\&\quad=\int_0^1\left(\frac{\left(t^{r-1}-1+e^{-\lambda_0(1-t)}\right)\mathsf{E}\left[e^{-X(1-t)}\right]-1}{1-t}+\mathsf{E}[X]+\lambda_0+r-1\right)\frac{1}{\log t}\,\mathrm{d}t \quad (4.15)\end{aligned}$$

A lower bound can be derived if the expected values in (4.14) can be solved. However, it becomes difficult if we try to make a tighter bound and choose input distributions more complicated. Hence, we loose the channel condition and investigate a channel without dark current. The lower bound will become

$$\begin{aligned}I(X;Y)\geq h(X)-(\mathsf{E}[X]+r)\log\frac{\mathsf{E}[X]}{\mathsf{E}[X]+r}+(r-1)\mathsf{E}[\log X]+\mathsf{E}[X\log X]\\-\mathsf{E}[X]-r-\mathsf{E}[\Gamma(Y_1+r)] \quad (4.16)\end{aligned}$$

where

$$\mathsf{E}[\log\Gamma(Y_1+r)]=\int_0^1\left(\frac{1-\mathsf{E}\left[e^{-X(1-t)}\right]}{1-t}-\mathsf{E}[X]-r+1\right)\frac{1}{\log t}\,\mathrm{d}t \quad (4.17)$$

In the remainder of this chapter we will show the proof of (2.18). Derivation with different power constraints are similar and deferred to Appendix B.

### 4.3.2 Proof of (2.21)

We choose the input distribution as follows:

$$Q'(x)=\frac{\sqrt{\mu}}{\sqrt{\mathsf{A}\pi x}\cdot\operatorname{erf}\left(\sqrt{\mu}\right)}e^{-\frac{\mu x}{\mathsf{A}}},\quad 0\leq x\leq\mathsf{A}, \quad (4.18)$$

where $\mu$ is a solution to (2.20), or

$$\frac{\sqrt{\mu}}{\sqrt{\pi}\operatorname{erf}\left(\sqrt{\mu}\right)}=\left(\frac{1}{2}-\alpha\mu\right)e^{\mu}. \quad (4.19)$$

Then we have:

$$\mathsf{E}[X]=\alpha\mathsf{A} \quad (4.20)$$

$$\mathsf{E}[\log X]=\log\mathsf{A}+2e^{\mu}(2\alpha\mu-1)\cdot{}_2\mathcal{F}_2\left(\frac{1}{2},\frac{1}{2};\frac{3}{2},\frac{3}{2};-\mu\right) \quad (4.21)$$

$$\begin{aligned}h(X)=\log\mathsf{A}+e^{\mu}(2\alpha\mu-1)\cdot{}_2\mathcal{F}_2\left(\frac{1}{2},\frac{1}{2};\frac{3}{2},\frac{3}{2};-\mu\right)-\mu-\log\left(\frac{1}{2}-\alpha\mu\right)\\+\alpha\mu \quad (4.22)\end{aligned}$$

$$\mathsf{E}\left[e^{-X(1-t)}\right]=\frac{\sqrt{\mu}}{\operatorname{erf}\left(\sqrt{\mu}\right)}\cdot\frac{\operatorname{erf}\left(\sqrt{\mu+(1-t)\mathsf{A}}\right)}{\sqrt{\mu+(1-t)\mathsf{A}}} \quad (4.23)$$

$$\mathsf{E}[X\log X]=\alpha\mathsf{A}\log\mathsf{A}+\frac{2}{9}e^{\mu}(2\alpha\mu-1)\cdot\mathsf{A}\cdot{}_2\mathcal{F}_2\left(\frac{3}{2},\frac{3}{2};\frac{5}{2},\frac{5}{2};-\mu\right) \quad (4.24)$$

Here, ${}_2\mathcal{F}_2$ denotes the hypergeometric function defined in (2.15).

According to (4.14) and (4.16), we can derive the lower bound (2.18) and (2.21) by plugging in above expectations.

# Chapter 5

# Conclusion and Discussion

Some lower bounds and upper bounds on capacity of discrete-time Poisson channel subject to a peak-power constraint and average-power constraint are derived in this thesis. The bounds help us to evaluate a region the capacity supposed to be with given power. We propose upper bounds for a channel with dark current. Besides, lower bounds are derived for a channel with or without dark current separately.

Since the entropy of Poisson is not known yet, we use different methods to evaluate it in the derivation of upper bounds and lower bounds. For upper bounds, we use Lemma 9 and Lemma 10, which can be accurate if we use more terms in evaluation. On the other hand, we use Proposition 12 in the derivation of lower bounds, where some integral expression are used.

In derivation of upper bounds, we choose some parameters of output distribution $R'(\tilde{y})$ in order to satisfy Jensen's inequality. It's shown that these choices lead to some suboptimal bounds, especially for high SNR. It might be improved if we can choose the output distribution in a more clever way.

Some general form of lower bounds are derived with data processing lemma. However, we make sacrifice in (4.4) and result in poor lower bounds at low SNR. We need to evaluate the entropy of output $H(Y)$ more precisely to solve this problem. Besides, there are some integral expressions in the lower bounds as (4.15) and (4.17). They are theoretically computable but cause some problems in Matlab simulations. Hence, we evaluate them with the help of Taylor expansions.

# Appendix A

# A Proof of Upper Bound

The derivations shown in this appendix follow the same line as presented in Chapter 3. We first choose an output distribution and use Proposition 8, and then separate the formula into different parts. Every part can be bounded individually to yield an upper bound on capacity.

## A.1 Average-to-Peak Ratio $\alpha \in [\frac{1}{3}, 1]$

We choose the following output distribution:

$$R'(\tilde{y}) \triangleq \begin{cases} p_{\mathsf{A}} \cdot \frac{\beta \cdot \tilde{y}^{\beta-1}}{((\mathsf{A}+\lambda)(1+\delta))^{\beta}}, & \forall 0 \leq \tilde{y} \leq (\mathsf{A}+\lambda_0)(1+\delta) \\ (1-p_{\mathsf{A}}) \cdot e^{-(\tilde{y}-(\mathsf{A}+\lambda_0)(1+\delta))}, & \forall \tilde{y} > (\mathsf{A}+\lambda_0)(1+\delta) \end{cases} \tag{A.1}$$

The upper bound can be derived proceeding similar to Chapter 3:

$$\mathsf{C} \leq \mathsf{E}_{Q*}\Big[D\big(\tilde{W}'(\tilde{y}|x)\,\big\|\,R'(\tilde{y})\big)\Big] \tag{A.2}$$

$$\begin{aligned} &\leq \mathsf{E}_{Q*}\bigg[-h(\tilde{Y}|X=x) + \log\frac{1}{\beta \cdot p_{\mathsf{A}}} \cdot \Pr\Big[\tilde{Y} \leq (\mathsf{A}+\lambda_0)(1+\delta)\Big| X=x\Big] \\ &\quad + (1-\beta)\mathsf{E}\Big[\log \tilde{Y}\ \Big|\ X=x\Big] \\ &\quad + \beta\log\left((\mathsf{A}+\lambda_0)(1+\delta)\right) \cdot \Pr\Big[\tilde{Y} \leq (\mathsf{A}+\lambda_0)(1+\delta)\Big|X=x\Big] \\ &\quad + \log\frac{1}{1-p_{\mathsf{A}}} \cdot \Pr\Big[\tilde{Y} > (\mathsf{A}+\lambda_0)(1+\delta)\Big|X=x\Big] \\ &\quad + \int_{(\mathsf{A}+\lambda_0)(1+\delta)}^{\infty} \tilde{W}'(\tilde{y}|x)\left(\tilde{y}-(\mathsf{A}+\lambda_0)(1+\delta)\right)\mathrm{d}\tilde{y}\bigg] \end{aligned} \tag{A.3}$$

We define

$$\mathsf{C}_{\mathrm{a}} \triangleq -h(\tilde{Y}|X=x) + (1-\beta)\mathsf{E}\Big[\log\tilde{Y}\ \Big|\ X=x\Big] \tag{A.4}$$

$$\mathsf{C}_{\mathrm{b}} \triangleq \beta\log\left((\mathsf{A}+\lambda_0)(1+\delta)\right) \cdot \Pr\Big[\tilde{Y} \leq (\mathsf{A}+\lambda_0)(1+\delta)\Big|X=x\Big] \tag{A.5}$$

$$\mathsf{C}_{\mathrm{c}} \triangleq \log\frac{1}{\beta \cdot p_{\mathsf{A}}} \cdot \Pr\Big[\tilde{Y} \leq (\mathsf{A}+\lambda_0)(1+\delta)\Big|X=x\Big]$$

$$+\log\frac{1}{1-p_{\mathsf{A}}}\cdot\Pr\left[\tilde{Y}>(\mathsf{A}+\lambda_0)(1+\delta)\middle|X=x\right] \tag{A.6}$$

$$\mathsf{C}_{\mathrm{d}}\triangleq\int_{(\mathsf{A}+\lambda_0)(1+\delta)}^{\infty}\tilde{W}'(\tilde{y}|x)\left(\tilde{y}-(\mathsf{A}+\lambda_0)(1+\delta)\right)\mathrm{d}\tilde{y} \tag{A.7}$$

We start with $C_{\mathrm{a}}$ and follow the same steps we had before. From (3.59) and (3.60) we can choose a specific value of $\beta$ and the upper bound will be similar to (3.61) with $\beta$ substituting for $\nu$.

The derivation of $C_{\mathrm{b}}$ and $C_{\mathrm{c}}$ is as follows:

$$\mathsf{E}_{Q^*}[\mathsf{C}_{\mathrm{b}}]=\mathsf{E}_{Q^*}[1-p_{\mathrm{T},X}]\cdot\beta\log\left((\mathsf{A}+\lambda_0)(1+\delta)\right) \tag{A.8}$$

$$\leq\Bigg(1-\frac{\gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor,\alpha\mathsf{A}+\lambda_0)}{\Gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor)}+\Pr\left[N_{\mathsf{A}+\lambda_0}=\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right]$$
$$\cdot\left((\mathsf{A}+\lambda_0)(1+\delta)-\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right)\Bigg)\cdot\beta\log\left((\mathsf{A}+\lambda_0)(1+\delta)\right) \tag{A.9}$$

and

$$\mathsf{C}_{\mathrm{c}}=(1-p_{\mathrm{T},x})\log\frac{1}{\beta\cdot(1-p_{\mathrm{T},\mathsf{A}})}+p_{\mathrm{T},x}\log\frac{1}{p_{\mathrm{T},\mathsf{A}}} \tag{A.10}$$

thus we can have the following bound:

$$\mathsf{E}_{Q^*}[\mathsf{C}_{\mathrm{c}}]=\mathsf{E}_{Q^*}[1-p_{\mathrm{T},X}]\log\frac{1}{1-p_{\mathrm{T},\mathsf{A}}}+\mathsf{E}_{Q^*}[p_{\mathrm{T},X}]\log\frac{1}{p_{\mathrm{T},\mathsf{A}}} \tag{A.11}$$

$$\leq\Bigg(1-\frac{\gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor,\alpha\mathsf{A}+\lambda_0)}{\Gamma(\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor)}+\Pr\left[N_{\mathsf{A}+\lambda_0}=\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right]$$
$$\cdot\left((\mathsf{A}+\lambda_0)(1+\delta)-\lfloor(\mathsf{A}+\lambda_0)(1+\delta)\rfloor\right)\Bigg)\log\frac{1}{\beta\cdot(1-p_{\mathrm{T},\mathsf{A}})}$$
$$+p_{\mathrm{T},\mathsf{A}}\log\frac{1}{p_{\mathrm{T},\mathsf{A}}} \tag{A.12}$$

The term $C_{\mathrm{d}}$ is the same as before so we ignore the derivation.

## A.2 Only Average-Power Constraint

We choose the output distribution as follows:

$$R'(\tilde{y})\triangleq\frac{\tilde{y}^{\nu-1}\cdot e^{-\frac{\tilde{y}}{\beta}}}{\beta^{\nu}\cdot\Gamma(\nu)},\quad\tilde{y}\geq 0 \tag{A.13}$$

and we derive

$$\mathsf{C}\leq\mathsf{E}_{Q*}\left[D\left(\tilde{W}'(\tilde{y}|x)\middle\|R'(\tilde{y})\right)\right] \tag{A.14}$$

$$\leq\mathsf{E}_{Q*}\Bigg[-h(\tilde{Y}|X=x)+(1-\nu)\mathsf{E}\left[\log\tilde{Y}\,\middle|\,X=x\right]$$
$$+\frac{1}{\beta}\left(X+\lambda_0+\frac{1}{2}\right)+\nu\log\beta+\log\Gamma(\nu)\Bigg] \tag{A.15}$$

We set

$$\mathsf{C}_{\mathrm{a}} \triangleq -h(\tilde{Y}|X=x) + (1+\nu)\mathsf{E}\left[\log \tilde{Y} \,\middle|\, X=x\right] \tag{A.16}$$

$$\mathsf{C}_{\mathrm{b}} \triangleq \nu \log \beta + \log \Gamma(\nu) + \frac{1}{\beta}\left(x + \lambda_0 + \frac{1}{2}\right) \tag{A.17}$$

where $\mathsf{C}_{\mathrm{a}}$ here can be bounded as in (3.61). And $\mathsf{C}_{\mathrm{b}}$ is bounded as follows:

$$\mathsf{E}_{Q^*}[\mathsf{C}_{\mathrm{b}}] = \nu \log \beta + \log \Gamma(\nu) + \frac{1}{\beta}\left(\mathsf{E}_{Q^*}[X] + \lambda_0 + \frac{1}{2}\right) \tag{A.18}$$

$$\leq \nu \log \beta + \log \Gamma(\nu) + \frac{1}{\beta}\left(\mathcal{E} + \lambda_0 + \frac{1}{2}\right) \tag{A.19}$$

where we choose

$$\beta \triangleq \frac{\mathcal{E} + \lambda_0 + \frac{1}{2}}{\nu} \tag{A.20}$$

# Appendix B

# A Proof of Lower Bound

This chapter contains parts of the derivations of the lower bound in Chapter 4. According to (4.14), a lower bound can be derive by solving some expectations of our chosen distribution. In the following we show the input distribution we choose and the corresponding parts needed for a lower bound.

## B.1 Average-to-Peak Ratio $\alpha \in [\frac{1}{3}, 1]$

Here we choose the input distribution as

$$Q'(x) = \frac{1}{\sqrt{4\mathsf{A}x}}, \quad \forall 0 \leq x \leq \mathsf{A} \tag{B.1}$$

Then we have:

$$\mathsf{E}[X] = \frac{1}{3}\mathsf{A} \tag{B.2}$$

$$\mathsf{E}[\log X] = \log \mathsf{A} - 2 \tag{B.3}$$

$$h(X) = \log(2\mathsf{A}) - 1 \tag{B.4}$$

$$\mathsf{E}\left[e^{-X(1-t)}\right] = \frac{\sqrt{\pi}\mathrm{erf}\left(\sqrt{(1-t)\mathsf{A}}\right)}{\sqrt{4\mathsf{A}(1-t)}} \tag{B.5}$$

$$\mathsf{E}[X \log X] = \frac{1}{3}\mathsf{A}\log \mathsf{A} - \frac{2}{9}\mathsf{A} \tag{B.6}$$

Plugging above results in (4.14) we get (2.24).

## B.2 Only Average-Power Constraint

Here we choose the input distribution as

$$Q'(x) = \frac{1}{\sqrt{2\pi\mathcal{E}x}} e^{-\frac{x}{2\mathcal{E}}}, \quad x \geq 0 \tag{B.7}$$

and derive

$$\mathsf{E}[X] = \mathcal{E} \tag{B.8}$$

$$\mathsf{E}[\log X] = \log \mathcal{E} - \log 2 - \gamma \tag{B.9}$$

$$h(X) = \frac{1}{2}\log \pi + \log \mathcal{E} + \frac{1}{2}(1-\gamma) \tag{B.10}$$

$$\mathsf{E}\left[e^{-X(1-t)}\right] = \sqrt{\frac{2}{2+4\mathcal{E}-4\mathcal{E}t}} \tag{B.11}$$

$$\mathsf{E}[X\log X] = \mathcal{E}\left(\log \mathcal{E} - \log 2 + 2 - \gamma\right) \tag{B.12}$$

where $\gamma$ is defined in (2.27).

# Appendix C

# A Proof of Proposition 12

The proof is based on the data processing inequality of the relative entropy [4]. We focus our discussion on the Poisson channel described in (1.1). Besides, we use an numerical evaluation for the logarithm Euler's gamma function [5]:

$$\log \Gamma(x) = \int_0^1 \left( \frac{1 - u^{x-1}}{1 - u} - (x - 1) \right) \frac{\mathrm{d}u}{\log u} \tag{C.1}$$

where Euler's gamma function is defined as

$$\Gamma(x) = a^x \int_0^\infty u^{x-1} e^{-au} \, \mathrm{d}u \tag{C.2}$$

where $a$ is a positive real number. We often take $a = 1$, the common definition of the gamma function.

Let $Q'(\cdot)$ denote an arbitrary PDF on $\mathbb{R}^+$ with a certain finite mean $\mathsf{E}_{Q'}[X] = \eta > 0$, and let $R(\cdot)$ be the PMF of $Y$ when $Y$ is conditionally Poisson given $X \sim Q'(\cdot)$. On the other hand, let another input distribution $Q'_{\Gamma,r,\beta}$ denotes the Gamma distribution:

$$Q'_{\Gamma,r,\beta}(x) = \frac{x^{r-1} e^{-\frac{x}{\beta}}}{\beta^r \Gamma(r)}, \quad x \geq 0, \ \beta > 0, \ r > 0 \tag{C.3}$$

where its corresponding output $Y$ will be negative binomial distributed, *i.e*,

$$R_{\mathrm{NB},r,p}(y) = \frac{\Gamma(r + y)}{y! \Gamma(r)} (1 - p)^r p^y, \quad y \in \mathbb{N}_0, \ 0 < p < 1, \ r > 0 \tag{C.4}$$

where $R_{\mathrm{NB},r,p}(\cdot)$ is the output distribution with the following choice:

$$\beta = \frac{p}{1 - p}, \, 0 < p < 1 \tag{C.5}$$

By the data processing theorem we obtain

$$D(Q' \| Q'_{\Gamma,r,\beta}) \geq D(R \| R_{\mathrm{NB},r,p}) \tag{C.6}$$

We evaluate both sides of the inequality as follows:

$$D(Q'\|Q'_{\Gamma,r,\beta}) = -h(X) + \frac{1-p}{p}\mathsf{E}[X] + (1-r)\mathsf{E}[\log X] + r\log\frac{p}{1-p} + \log\Gamma(r) \tag{C.7}$$

$$\begin{aligned} D(R\|R_{\mathrm{NB},r,p}) = {} & -H(Y_1) + \mathsf{E}[\log\Gamma(Y_1+1)] + \log\Gamma(r) - \mathsf{E}[\log\Gamma(Y_1+r)] \\ & - r\log(1-p) - \mathsf{E}[Y_1]\log p \end{aligned} \tag{C.8}$$

Next we use that $\mathsf{E}[Y] = \mathsf{E}[X]$, apply (C.1)

$$\begin{aligned} & \mathsf{E}[\log\Gamma(Y_1+1) - \log\Gamma(Y_1+r)] \\ & = I\{r \neq 1\} \cdot \int_0^1 \left(\frac{t^{r-1}-1}{1-t}\mathsf{E}\left[e^{-X(1-t)}\right] + r - 1\right)\frac{1}{\log t}\,\mathrm{d}t \end{aligned} \tag{C.9}$$

and choose the optimal value for $p$ to derive (4.8)

$$p = \frac{\mathsf{E}[X]}{\mathsf{E}[X]+r} \tag{C.10}$$

# Bibliography


[1] 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.

[2] ——, "On the fading number of multi-antenna systems over flat fading channels with memory and incomplete side information," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Lausanne, Switzerland, June 30 – July 5, 2002, p. 478.

[3] S. M. Moser, "Sharp bounds on the Poisson law and related quantities," January 2010, arXiv:cs.IT/1001.2897v1. [Online]. Available: http://arxiv.org/abs/1001.2897v1

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

[5] G. E. Andrews, R. Askey, and R. Roy, *Special Functions.* Cambridge University Press, 1999.