

Information-theoretic aspects of optical communications

Doctoral Thesis**Author(s):**

Wang, Ligong

Publication date:

2011

Permanent link:

<https://doi.org/10.3929/ethz-a-006543410>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Originally published in:

ETH Series in Information Theory and its Applications 006

Diss. ETH No. 19708

Information-Theoretic Aspects of Optical Communications

A dissertation submitted to the
ETH Zurich
for the degree of
Doctor of Sciences ETH Zurich

presented by

Ligong Wang

M.Sc. ETH
born on 17 April 1982
citizen of P. R. China

accepted on the recommendation of
Prof. Dr. Amos Lapidot, examiner
Prof. Dr. Renato Renner, co-examiner
Prof. Dr. Yossef Steinberg, co-examiner

Hartung-Gorre Verlag, Konstanz, 2011

Reprint of Diss. ETH No. 19708

ETH Series in Information Theory and its Applications Vol. 6

edited by Amos Lapidot

Bibliographic Information published by Die Deutsche Nationalbibliothek

Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at <http://dnb.d-nb.de>.

Copyright © 2011 by Ligong Wang

First Edition 2011

Hartung-Gorre Verlag Konstanz

ISSN 1860 - 1081

ISBN-10: 3-86628-392-X

ISBN-13: 978-3-86628-392-3

Acknowledgments

During my first semester at ETH as a master student, I took a course on information theory, and found it extremely interesting. Later, I decided to do my PhD in this field, and the lecturer of that course, Prof. Amos Lapidoth, became my PhD advisor. In my research, Amos's deep understanding of information theory, vast knowledge in mathematics, and acute intuition have been very helpful and inspiring. What I value the most, however, is his rare crystal-clear logic, which strongly influenced my own research style. I am also grateful to Amos for encouraging me to finish my thesis at the right time, so that I would not miss opportunities for further postdoctoral research. When heard I was finishing, many people asked me: "How can your supervisor let you go when you are his only PhD student?"

My research on quantum information theory was mostly under the help of my co-examiner Prof. Renato Renner. Renato is a most kind, patient, yet persistent mentor. He has a physicist's view of large pictures, as well as specific knowledge to help me solve small technicalities. Without his guidance, I would probably have lost myself in the exciting but somewhat chaotic world of quantum information theory.

Many thanks also go to my other co-examiner Prof. Yossef Steinberg. His comments and questions were a big help on the final version of this thesis and on my future research.

During my PhD I had chances to collaborate with Dr. Shraga Bross and Prof. Jeffrey Shapiro, from whom I learned many things. I also wish to thank Prof. Hans-Andrea Loeliger and Prof. Stefan Wolf for their comments on my talks.

I always felt fortunate to have such a group of brilliant, modest, and friendly PhD-mates: Natalia, Stephan, Michèle, Tobi, and Christoph. I am deeply grateful to their help on my research and on my life as a foreigner in Switzerland. I enjoyed our time together discussing research, assisting lectures, and chatting about everything. I also very much enjoyed the friendships of the other members of ISI, despite the fact that I usually preferred a nap, which is of utmost importance to me, to a lunch hour with them.

I wish to thank all the students that worked on semester- or master-projects with me. Through working with them, I not only learned how to supervise projects, but also shared many of their interesting ideas. Special thanks go to Vinodh and Andreas who contributed to some of the results in this thesis. I would like to say “thank you” also to the students in the classes that I have assisted, as their feedback helped me improve my teaching. I will not forget the first time I assisted Information Transfer when a student said I should speak louder!

I find it very difficult, and somewhat embarrassing, to express how grateful I am to the people whom I owe the biggest “thank you”—my parents and my wife. Fortunately, I do not know how to type Chinese in Latex, which is a good excuse for not going into details.

Abstract

This dissertation studies several information-theoretic aspects of optical communications.

The first aspect it studies is the one-shot classical capacity of a quantum communication channel. It proves new upper and lower bounds on the amount of classical information that can be transmitted through a single use of a quantum channel, under a constraint on the average error probability. The bounds are expressed using a quantity defined via quantum hypothesis testing. Combined with the Quantum Stein's Lemma, these bounds provide a conceptually simple proof for the Holevo-Schumacher-Westmoreland Theorem for the classical capacity of a memoryless quantum channel. Further, they also give a general capacity formula that is valid for any, not necessarily memoryless, quantum channel.

The second topic studied in this dissertation is the capacity of a continuous-time peak-limited Poisson channel with spurious counts in the output. It is shown that, if the positions of the spurious counts are known noncausally to the encoder but not to the decoder, then the capacity of this channel equals the capacity of the same channel but with no spurious counts, regardless of whether the spurious counts are random or are chosen by a malicious adversary. On the other hand, if the positions of the spurious counts are known only causally to the encoder but not to the decoder, then such information does not help to increase the capacity of this channel.

Next, a rate-distortion problem for point processes is considered. In this problem, an encoder sees a point pattern on the interval $[0, T]$ and

describes it to a reconstructor using bits. Based on this description, the reconstructor produces a subset of $[0, T]$ of Lebesgue measure not exceeding DT for some $D > 0$ to cover all the points in the pattern. It is shown that, if the point pattern is the outcome of a homogeneous Poisson process of intensity λ , then, as T tends to infinity, the minimum number of bits per second needed for the encoder to describe the pattern is $-\lambda \log D$. Further, any point pattern containing no more than λ points per second can be successfully described in this sense using $-\lambda \log D$ bits per second. A Wyner-Ziv version of this problem is also studied where some points in the pattern are known to the reconstructor.

The last problem considered in this dissertation is the asymptotic capacity at low input powers of a discrete-time Poisson channel under average-power or average- and peak-power constraints. For a Poisson channel whose dark current is zero or decays to zero linearly with the allowed average input power E , capacity is shown to scale like $E \log \frac{1}{E}$ as E tends to zero. For a Poisson channel whose dark current is a nonzero constant, capacity is shown to scale, to within a constant, like $E \log \log \frac{1}{E}$.

Keywords: Arbitrarily varying channel, arbitrarily varying source, channel capacity, finite blocklength, hypothesis testing, low signal-to-noise ratio, optical communication, Poisson channel, Poisson process, quantum channel, rate-distortion theory, side-information.

Kurzfassung

Diese Dissertation befasst sich mit verschiedenen informationstheoretischen Aspekten der optischen Kommunikation.

Der erste Aspekt, der betrachtet wird, ist die klassische Kapazität von Quantenkommunikationskanälen bei einmaliger Benutzung des Kanals. Neue obere und untere Schranken für die Menge an klassischer Information, die bei einer einmaligen Nutzung eines Quantenkanals bei beschränkter durchschnittlicher Fehlerwahrscheinlichkeit übertragen werden kann, werden hergeleitet. Diese Schranken werden mit Hilfe von Größen ausgedrückt, die durch Hypothesenprüfungen definiert sind. In Kombination mit der Quanten-Version von Steins Lemma liefern die Schranken einen konzeptionell einfachen Beweis für den Satz von Holevo-Schumacher-Westmoreland über die klassische Kapazität von gedächtnislosen Quantenkanälen. Ausserdem bieten die Schranken eine allgemeine Formel für die Kapazität, die für alle Quantenkanäle, nicht nur für die gedächtnislosen, gilt.

Das zweite Thema, das in dieser Dissertation behandelt wird, ist die Kapazität von zeitkontinuierlichen, spitzenwertbegrenzten Poisson-Kanälen mit parasitären Zählereignissen am Ausgang. Es wird gezeigt, dass, falls die Zeitpunkte der parasitären Zählereignisse dem Codierer nicht-kausal bekannt sind, dem Decodierer jedoch unbekannt sind, die Kapazität des Kanals gleich der Kapazität des Kanals ohne parasitäre Zählereignisse ist. Dies gilt unabhängig davon, ob die parasitären Zählereignisse zufällig entstehen, oder von einem Widersacher böswillig gewählt werden. Falls die parasitären Zählereignisse dem Codierer nur kausal zur Verfügung stehen, dem Decodierer jedoch unbekannt sind, so erhöht sich durch diese Information die Kapazität des Kanals nicht.

Als nächstes wird ein Rate-Distortion-Problem für Punktprozesse betrachtet. In diesem Problem sieht ein Codierer ein Punktmuster auf dem Intervall $[0, T]$ und beschreibt es einem Empfänger mithilfe von Bits. Ausgehend von dieser Beschreibung produziert der Empfänger eine Teilmenge von $[0, T]$ mit Lebesgue-Maß nicht größer als DT , das alle Punkte im Punktmuster überdeckt, wobei $D > 0$. Es wird gezeigt, dass, falls das Punktmuster das Resultat eines homogenen Poisson-Prozesses mit Intensität λ ist, im Limes mit $T \rightarrow \infty$ die minimale Anzahl Bits pro Sekunde, die der Codierer braucht, um das Punktmuster zu beschreiben, gleich $-\lambda \log D$ ist. Darüber hinaus kann jedes Punktmuster mit nicht mehr als λ Punkten pro Sekunde erfolgreich mit $-\lambda \log D$ Bits pro Sekunde in diesem Sinne beschrieben werden. Es wird zudem eine Wyner-Ziv-Variante dieses Problems behandelt, in der einige Punkte des Punktmusters dem Empfänger bekannt sind.

Das letzte Problem, das in dieser Dissertation betrachtet wird, ist die asymptotische Kapazität des diskreten Poisson-Kanals bei niedrigen Eingangsleistungen und Durchschnittsleistungsbeschränkung oder Durchschnitts- und Spitzenleistungsbeschränkung. Für einen Poisson-Kanal ohne Dunkelstrom, oder mit Dunkelstrom, der linear mit der zulässigen Durchschnittsleistung abklingt, verhält sich die Kapazität wie $E \log \frac{1}{E}$, wenn die Durchschnittsleistung E nach Null strebt. Für Poisson-Kanäle mit konstantem Dunkelstrom wird gezeigt, dass die Kapazität sich bis auf einen Skalierungsfaktor wie $E \log \log \frac{1}{E}$ verhält.

Stichworte: Beliebig variierender Kanal, beliebig variierende Quelle, Kanalkapazität, endliche Blocklänge, Hypothesenprüfung, niedriger Signal-Rausch-Abstand, optische Kommunikation, Poisson-Kanal, Poisson-Prozess, Quantenkanal, Rate-Distortion-Theorie, Zusatzinformation.

Contents

Abstract	v
Kurzfassung	vii
1 Introduction	1
1.1 Motivation	1
1.2 Outline and Contributions	3
1.3 Notation	5
2 One-Shot Capacity Bounds for Quantum Channels	7
2.1 Introduction	7
2.2 Preliminary on Quantum Channels	8
2.3 Hypothesis Testing and $D_{\text{H}}^{\epsilon}(\rho\ \sigma)$	10
2.4 The Bounds	13
2.5 Asymptotic Analysis of the Bounds	20
3 The Poisson Channel with Side-Information	23
3.1 Introduction	23
3.2 Preliminary on the Poisson Channel	25
3.3 Noncausal SI	27
3.3.1 Random Codes against Arbitrary States	27
3.3.2 Deterministic Codes against Random States	31
3.4 Causal SI	33
4 Covering Point Patterns	35
4.1 Introduction	35
4.2 Covering a Poisson Process	36
4.3 Covering General Point Processes and Arbitrary Point Patterns	41

4.4	Some Points are Known to the Reconstructor	44
5	The Discrete-Time Poisson Channel at Low Input Powers	49
5.1	Introduction	49
5.2	Results	51
5.3	Lower Bounds	55
5.3.1	Dark Current Proportional to E	56
5.3.2	Constant Nonzero Dark Current	59
5.4	Upper Bounds	62
5.4.1	Zero Dark Current	62
5.4.2	Constant Nonzero Dark Current	69
A	One-Shot Capacity Bounds in Classical Settings	81
A.1	The Single-User Channel with SI	81
A.2	The Multiple-Access Channel	83
A.3	The MAC with SI	88
	Acronyms	91
	List of Symbols	93
	Bibliography	95
	About the Author	101

Chapter 1

Introduction

1.1 Motivation

This dissertation studies several information-theoretic aspects of optical communications.

Depending on the level of restrictions imposed on the transmitter and the receiver, optical communications can be modeled with different quantum and classical channels. In this dissertation we study quantum channels in general, and two specific classical channel models for optical communications: the continuous-time Poisson channel and the discrete-time Poisson channel.

To fully characterize optical communications, quantum effects need to be considered. Thus, the channel inputs and outputs are modeled by quantum states on Hilbert spaces, and the channel is modeled by a quantum operation. For such a quantum channel, Shannon's capacity formula [39], which has been proved for memoryless classical channels, no longer applies. Holevo [20] and Schumacher and Westmoreland [37] found a formula for the capacity of a memoryless quantum channel, known as the HSW Theorem, which is analogous to Shannon's formula. However, unlike in the classical case, this formula is not *single-letter* but needs *regularization* [16], which means it only gives the capacity of a quantum channel when computed for n channel uses where n tends to infinity.

In this dissertation we take a different approach to quantum channel coding. We derive bounds on the amount of information that can be transmitted through a *single use* of a quantum channel, under a constraint on the average error probability. We then combine these bounds with Quantum Stein's Lemma to obtain the same formula as in the HSW Theorem, but with much simpler proofs. Further, these bounds also give a general capacity formula, which is still not single-letter, but which is valid for any, not necessarily memoryless, quantum channels.

If the receiver is restricted to using *direct detection*, i.e., photon counting, then we can model optical communications purely classically as (continuous-time) *Poisson channels*. Here, the channel input describes the transmitted light intensity and is a continuous-time nonnegative-valued signal, and the channel output describes the positions of the detected photons and is a Poisson process whose intensity at a certain time-instant t depends on the channel input at t and the background noise. The Poisson channel with peak-limited inputs is one of the best understood continuous-time channels in information theory. The capacity of this channel has a simple closed-form expression which was computed in [21], [10] and [47].

We consider a variation of this channel model where, in addition to the photons produced by the input signal and by background noise, the receiver sees other spurious photon counts, and where the positions of these spurious counts are known to the encoder as *side-information (SI)*. The capacity formula for a discrete memoryless channel (DMC) with independent and identically distributed (IID) states and with *noncausal* SI, where the encoder has SI about the whole transmission block before starting to encode, was found by Gel'fand and Pinsker [13]. This formula has been used to obtain closed-form capacity expressions for several channels with noncausal SI, e.g., the Gaussian channel [5]. However, it cannot be directly used in our problem which is in continuous time. In this dissertation we derive the capacity of the Poisson channel with noncausal SI by constructing a coding scheme.

A *dual problem*, in the sense of [6], to the one of peak-limited Poisson channel is a rate-distortion problem for a homogeneous Poisson process. In this rate-distortion problem, the reconstructor produces a region in time to cover all the points (photon counts) in the Poisson process that is observed by the encoder. Using a technique similar to [47] we find the rate-distortion function in this problem. We also derive a bound on

the rate-distortion function for a point process that is not necessarily Poisson. Further, we consider a Wyner-Ziv [49] version of this rate-distortion problem where the reconstructor has SI about the source. This Wyner-Ziv problem can be considered as a dual problem to the one of peak-limited Poisson channel with noncausal SI.

If, in addition to the receiver being restricted to using direct detection, the transmitter is restricted to using *pulse-amplitude modulation*, then an optical communication channel further reduces to a *discrete-time Poisson channel*. Here, each channel use corresponds to a pulse, the input is proportional to the product of the transmitted light intensity by the pulse-duration, and the output is the number of photons detected during this pulse-duration. A closed-form expression of the capacity of this channel under average- and peak-power constraints is not yet known.

In this dissertation we study the *asymptotic* capacity of the discrete-time Poisson channel when the average input power tends to zero. We are interested in two regimes. The first is the wide-band regime where the pulse-duration tends to zero, so background noise tends to zero linearly with the average input power. We compare the asymptotic capacity in this regime to the capacity of the pure-loss bosonic channel [14] to evaluate the loss in capacity by using direct detection instead of a quantum receiver. The second is the weak-transmitter regime where background noise does not decrease with the average input power.

1.2 Outline and Contributions

The rest four chapters of this dissertation study the topics motivated in Section 1.1 separately: Chapter 2 proves one-shot capacity bounds for quantum channels; Chapter 3 studies peak-limited Poisson channels with SI; Chapter 4 studies rate-distortion problems of covering Poisson processes and other point patterns; Chapter 5 computes asymptotic capacities of discrete-time Poisson channels at low average input powers. The contributions in these chapters are summarized below:

Chapter 2: We derive bounds on the maximum amount of information that can be transmitted through a single use of a quantum communication channel subject to a condition on the average probability of a

decoding error. The bounds are expressed using a quantity called “hypothesis testing relative entropy” which is defined directly via hypothesis testing. The proofs for the bounds are also based on ideas of hypothesis testing. Combined with Quantum Stein’s Lemma, these bounds provide a conceptually simple proof for the HSW Theorem. They also give a general nonsingle-letter formula for the capacity of any quantum channel. In **Appendix A** we apply the techniques used in Chapter 2 to derive capacity bounds and general capacity formulas for several settings in classical information theory.

Chapter 3: We derive the capacity of a peak-limited Poisson channel with spurious counts at the decoder, where the positions of these spurious counts are known to the encoder but not to the decoder. We show that, if the spurious counts are known noncausally to the encoder, then the capacity of this channel equals the capacity of the same channel without spurious counts, irrespective of whether the spurious counts are random or are chosen by a malicious adversary. On the other hand, knowing the spurious counts only causally at the encoder does not increase capacity even if the spurious counts obey a homogeneous Poisson law.

Chapter 4: We consider a rate-distortion problem in which an encoder observes a point pattern—a finite number of points in the interval $[0, T]$ —which is to be described to a reconstructor using bits. Based on these bits, the reconstructor wishes to select a subset of $[0, T]$ of average Lebesgue measure not exceeding DT , where $D > 0$, that covers all the points in the pattern. We show that, if the point pattern is produced by a homogeneous Poisson process of intensity λ , then, as T tends to infinity, the minimum number of bits per second needed by the encoder is $-\lambda \log D$. We also show that, as T tends to infinity, any point pattern on $[0, T]$ containing no more than λT points can be successfully described using $-\lambda \log D$ bits per second in this sense. Finally, we consider a Wyner-Ziv version of this problem where some of the points in the pattern are known to the reconstructor, but where the encoder does not know which points they are. We show that, in this case, only $-\nu \log D$ bits are needed, where νT is the number of points that are unknown to the reconstructor.

Chapter 5: We study the asymptotic capacity at low input powers of an average-power limited or an average- and peak-power limited discrete-time Poisson channel. For a Poisson channel whose dark current is zero or decays to zero linearly with its average input power E , we show that its capacity scales like $E \log \frac{1}{E}$ for small E . For a Poisson channel whose dark current is a nonzero constant, we show that its capacity scales, to within a constant, like $E \log \frac{1}{E}$ for small E .

1.3 Notation

We use a lower-case letter like x to denote a constant, and an upper-case letter like X to denote a random variable. There are exceptions where we use upper-case letters to denote constants, for example: C denotes the capacity, R denotes the rate, and D denotes the expected distortion.

We use a lower-case boldface letter like \mathbf{x} to denote 1) a vector, 2) a signal, i.e., a function on the real line, or 3) a point pattern, i.e., a set of points on the real line. It will be clear from the context which one we mean. Accordingly, we use an upper-case boldface letter like \mathbf{X} to denote 1) a random vector, 2) a random signal, or 3) a random point pattern, i.e., a point process. If \mathbf{x} (\mathbf{X}) is a (random) signal, then we use $x(t)$ ($X(t)$) to denote its value at t . If \mathbf{x} (\mathbf{X}) is a (random) point pattern, we use \mathbf{n}_x (\mathbf{N}_X) to denote its counting function, so $n_x(t_2) - n_x(t_1)$ is the number of points in \mathbf{x} that lie in the interval $(t_1, t_2]$.

A special class of vectors, namely, norm-one vectors on Hilbert spaces, which are used to describe pure quantum states, are denoted like $|\psi\rangle$, and the conjugate of the vector $|\psi\rangle$ is denoted by $\langle\psi|$.

We use a font like \mathbf{A} or $\mathbf{\Pi}$ to denote a matrix or an operator on a Hilbert space. However, density operators, i.e., trace-one positive-semidefinite Hermitian operators are usually denoted by lower-case Greek letters like ρ .

We use a font like \mathbb{H} to denote a Hilbert space, and a font like \mathcal{S} to denote a set. Exceptions are: \mathbb{R} denotes the set of reals and \mathbb{Z} denotes the set of integers.

We use a font like \mathfrak{C} to denote a quantum operation, i.e., a completely positive map between density operators.

We use \otimes to denote the tensor product of vectors, matrices, operators, or Hilbert spaces. For consistency, we also use \otimes to denote the Cartesian product of sets, and the product of distributions. For example, the distribution on (X, Y) where X and Y are independent, X is distributed according to P_X , and Y is distributed according to P_Y is denoted by $P_X \otimes P_Y$. We use $\rho^{\otimes n}$ to denote the tensor product of n copies of the operator ρ , and similarly for vectors, Hilbert spaces, etc.

Chapter 2

One-Shot Capacity Bounds for Quantum Channels

2.1 Introduction

The coding theorem for transmitting classical information through a memoryless quantum channel has been established by Holevo [20] and by Schumacher and Westmoreland [37]. A formula for general quantum channels has been obtained by Hayashi and Nagaoka [18]. These results concern the asymptotic regime where the number of channel uses tends to infinity and the probability of error is required to tend to zero.

In this chapter we look at a different scenario where the channel is used only once, and a finite error probability is allowed. We provide upper and lower bounds on the amount of (classical) information that can be transmitted through a single use of the quantum channel such that the average probability of error is below a certain value. Combined with the Quantum Stein's Lemma [31], they give a conceptually simple proof for the Holevo-Schumacher-Westmoreland (HSW) Theorem. The bounds can also be directly applied to “many” uses of an arbitrary channel,

where no assumption is made on the channel or the input states. In the asymptotic limit as the number of channel-uses tends to infinity and the probability of error is required to tend to zero, the upper bound and the lower bound coincide and lead to a capacity expression which is equivalent to that in [18]. These results require remarkably simple proofs, despite their strength and generality.

Channel coding is closely related to the problem of hypothesis testing, and this connection has been used in several works (see, e.g., [17, 18, 31, 32]). Here, we use hypothesis testing directly to define a relative-entropy-type quantity (Section 2.3). Our bounds on the one-shot channel capacity will then be expressed in terms of this quantity (Section 2.4). This quantity is similar to the “smooth min-relative entropy” introduced in [9], but its position in the smooth entropy framework (see, e.g., [34]) is still to be clarified.

This work is closely related to recent work of Mosonyi and Datta [29] who also studied the one-shot classical capacities of quantum channels. However, the bounds on the capacity they derive are different from ours (and the quantitative relation between them is unknown). In particular, the upper and lower bounds in our work coincide asymptotically for arbitrary channels, which is not shown to be true for the bounds in [29].

The rest of this chapter is arranged as follows: in Section 2.2 we review some mathematical formulations for quantum channels; in Section 2.3 we introduce the hypothesis testing relative entropy; in Section 2.4 we prove the one-shot capacity bounds; and in Section 2.5 we apply the one-shot bounds to infinitely many channel uses to obtain general capacity formulas.

2.2 Preliminary on Quantum Channels

In this section we give a brief review of the mathematical formulations of quantum channels, and explain their connections to their classical counterparts. For more details see, e.g., [30].

A *quantum state* is described by a trace-one, positive-semidefinite, Hermitian operator acting on a Hilbert space \mathbb{H} . We call such an operator a *density operator*. If a density operator ρ is of rank one, then we call it

a *pure state* and also write it as $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a normalized vector on \mathbb{H} , and where $\langle\psi|$ is the conjugate of $|\psi\rangle$.

We say a group of quantum states $\{\rho_i\}$ are *classical*¹ if they are simultaneously diagonalizable. In this case, there exists an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of \mathbb{H} such that every ρ_i can be written in the form

$$\rho_i = \sum_{x\in\mathcal{X}} P_i(x)|x\rangle\langle x|, \quad (2.1)$$

where P_i is a probability mass function (PMF) on the set \mathcal{X} . Note that if \mathbb{H} has a continuous spectrum, then (2.1) may be replaced by an integral, and P_i may be replaced by a density or a general probability measure. Hence, classical states are fully characterized by their corresponding probability distributions.

A state with two independent parts, $\rho^{\mathbb{A}}$ acting on Hilbert space \mathbb{A} and $\sigma^{\mathbb{B}}$ acting on Hilbert space \mathbb{B} , is described by the tensor product $\rho^{\mathbb{A}} \otimes \sigma^{\mathbb{B}}$.

A *measurement* on a Hilbert space \mathbb{H} is called a *POVM (Positive Operator-Valued Measure)* and is a set of positive-semidefinite Hermitian operators $\{E_j\}$ acting on \mathbb{H} which sum to \mathbb{I} . When applied to state ρ , the probability that the measurement $\{E_j\}$ yields outcome j is $\text{tr}(E_j\rho)$.

If the measurement operators of a POVM are simultaneously diagonalizable in the basis $\{|x\rangle\}_{x\in\mathcal{X}}$, then this POVM is a (classical) guessing rule, namely, a (possibly random) mapping from \mathcal{X} to the set of measurement outcomes.

A quantum operation is a completely-positive, trace-preserving² linear map between density operators. We call such a map a *Completely Positive Map (CPM)*.

If a CPM maps every diagonal operator in the basis $\{|x\rangle\}_{x\in\mathcal{X}}$ to a diagonal operator in the basis $\{|y\rangle\}_{y\in\mathcal{Y}}$, then it can be considered as a stochastic kernel from \mathcal{X} to \mathcal{Y} .

¹In the literature of quantum information theory, the word “classical” has several different meanings.

²It is often assumed that a quantum operation is *trace-nonincreasing*. In this thesis, however, we shall insist that it is trace-preserving.

2.3 Hypothesis Testing and $D_{\text{H}}^{\epsilon}(\rho\|\sigma)$

Consider a hypothesis testing problem between two quantum states ρ and σ . We wish to minimize the probability of guessing ρ when the real state is σ , subject to the condition that the probability of guessing σ when the real state is ρ is at most ϵ . Denote this minimum probability by $p_{\epsilon}^*(\rho|\sigma)$. Since any binary hypothesis test is equivalent to a POVM with two elements, it is easy to verify the following:

$$p_{\epsilon}^*(\rho|\sigma) = \inf_{\substack{Q: 0 \leq Q \leq I, \\ \text{tr}(Q\rho) \geq 1-\epsilon}} \text{tr}(Q\sigma). \quad (2.2)$$

Motivated by this observation we define a new type of relative entropy:

Definition 2.1. *The hypothesis testing relative entropy with parameter ϵ between two quantum states ρ and σ is*

$$D_{\text{H}}^{\epsilon}(\rho\|\sigma) \triangleq \sup_{\substack{Q: 0 \leq Q \leq I, \\ \text{tr}(Q\rho) \geq 1-\epsilon}} -\log \text{tr}(Q\sigma). \quad (2.3)$$

Remark 2.2. *For two classical states, i.e., probability distributions P and Q on a probability space (Ω, \mathcal{F}) , (2.3) becomes*

$$D_{\text{H}}^{\epsilon}(P\|Q) = \sup_{\substack{\Phi: \Omega \rightarrow [0,1] \\ \int_{\Omega} \Phi dP \geq 1-\epsilon}} -\log \int_{\Omega} \Phi dQ. \quad (2.4)$$

The next lemma shows some properties of the hypothesis testing relative entropy:

Lemma 2.3. *1. Relation to hypothesis testing:*

$$D_{\text{H}}^{\epsilon}(\rho\|\sigma) = -\log p_{\epsilon}^*(\rho|\sigma). \quad (2.5)$$

2. Positivity: for any ρ , σ , and $\epsilon \in [0, 1]$,

$$D_{\text{H}}^{\epsilon}(\rho\|\sigma) \geq 0, \quad (2.6)$$

with equality if and only if ρ and σ have the same support space and $\epsilon = 0$.

3. *Relation to Rényi's relative entropy:*

$$D_{\mathbb{H}}^0(\rho\|\sigma) = -\log \text{tr}(\Pi\sigma) = D_0(\rho\|\sigma), \quad (2.7)$$

where Π denotes the projector onto the support of ρ , and where $D_0(\rho\|\sigma)$ denotes Rényi's relative entropy of order 0.

4. *Data Processing Inequality (DPI): for any states ρ and σ , any CPM \mathfrak{E} acting on them, and any $\epsilon \in [0, 1]$,*

$$D_{\mathbb{H}}^{\epsilon}(\rho\|\sigma) \geq D_{\mathbb{H}}^{\epsilon}(\mathfrak{E}(\rho)\|\mathfrak{E}(\sigma)). \quad (2.8)$$

Proof. The first three properties are immediate from the definition of $D_{\mathbb{H}}^{\epsilon}(\rho\|\sigma)$. We next prove the DPI. Consider any POVM to distinguish $\mathfrak{E}(\rho)$ from $\mathfrak{E}(\sigma)$. We construct a new POVM to distinguish ρ from σ by preceding the given POVM with the CPM \mathfrak{E} . Clearly, this new POVM gives the same error probabilities (in distinguishing ρ and σ) as the given POVM (in distinguishing $\mathfrak{E}(\rho)$ and $\mathfrak{E}(\sigma)$). Thus

$$p_{\epsilon}^*(\rho|\sigma) \leq p_{\epsilon}^*(\mathfrak{E}(\rho)|\mathfrak{E}(\sigma)), \quad (2.9)$$

which, combined with Property 1, yields the DPI. \square

The Quantum Stein's Lemma [31], which is a generalization of the Chernoff-Stein Lemma in classical information theory (see, e.g., [7]), shows how $D_{\mathbb{H}}^{\epsilon}(\rho\|\sigma)$ is related to the normal quantum relative entropy $D(\rho\|\sigma)$. We restate this lemma in the following way to highlight this relation:

Lemma 2.4. *For any two states ρ and σ on a Hilbert space, and any $\epsilon \in (0, 1)$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} D_{\mathbb{H}}^{\epsilon}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma). \quad (2.10)$$

For a general sequence of state-pairs $\{\rho_n, \sigma_n\}_{n \in \mathbb{N}}$, the limit as n tends to infinity of $\frac{1}{n} D_{\mathbb{H}}^{\epsilon}(\rho_n\|\sigma_n)$ need not exist; when it does exist, it is in general not equal to a relative entropy. However, such limits (or limit inferiors/superiors) have equivalent expressions in the language of *quantum information spectrum*. We show one such equivalence in the following lemma:

Lemma 2.5. For any sequence of state-pairs $\{\rho_n, \sigma_n\}_{n \in \mathbb{N}}$,

$$\lim_{\epsilon \downarrow 0} \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} D_{\text{H}}^{\epsilon}(\rho_n \| \sigma_n) = \sup \left\{ a \mid \lim_{n \rightarrow \infty} \text{tr}(\Pi_n(a) \rho_n) = 1 \right\}, \quad (2.11)$$

where $\Pi_n(a)$ is the projector onto the support space of $(\rho_n - e^{na} \sigma_n)^+$.

Proof. Let

$$a^* \triangleq \sup \left\{ a \mid \lim_{n \rightarrow \infty} \text{tr}(\Pi_n(a) \rho_n) = 1 \right\}. \quad (2.12)$$

We first show

$$\lim_{\epsilon \downarrow 0} \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} D_{\text{H}}^{\epsilon}(\rho_n \| \sigma_n) \geq a^*. \quad (2.13)$$

To this end, let

$$\epsilon_n \triangleq 1 - \text{tr}(\Pi_n(a^*) \rho_n), \quad n \in \mathbb{N}, \quad (2.14)$$

then

$$\lim_{n \rightarrow \infty} \epsilon_n = 0. \quad (2.15)$$

We have

$$0 = \text{tr} \left(\Pi_n(a^*) (\rho_n - e^{na^*} \sigma_n)^- \right) \quad (2.16)$$

$$\leq \text{tr} \left(\Pi_n(a^*) (\rho_n - e^{na^*} \sigma_n) \right) \quad (2.17)$$

$$\leq 1 - e^{na^*} \text{tr}(\Pi_n(a^*) \sigma_n), \quad (2.18)$$

so

$$-\log \text{tr}(\Pi_n(a^*) \sigma_n) \geq na^*, \quad n \in \mathbb{N}, \quad (2.19)$$

and further

$$D_{\text{H}}^{\epsilon_n}(\rho_n \| \sigma_n) \geq a^*, \quad n \in \mathbb{N}. \quad (2.20)$$

Now (2.13) follows by (2.15) and (2.20).

We next show that, for any $b > a^*$,

$$\lim_{\epsilon \downarrow 0} \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} D_{\text{H}}^{\epsilon}(\rho_n \| \sigma_n) \leq b. \quad (2.21)$$

To this end, for a fixed $b > a^*$, let

$$c \triangleq 1 - \underline{\lim}_{n \rightarrow \infty} \text{tr}(\Pi_n(b) \rho_n), \quad (2.22)$$

then $c > 0$. Fix any $\epsilon \in (0, c)$ and any sequence of operators $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$ satisfying $0 \leq \mathbb{Q}_n \leq \mathbb{I}$ and

$$\mathrm{tr}(\mathbb{Q}_n \rho_n) \geq 1 - \epsilon, \quad n \in \mathbb{N}. \quad (2.23)$$

We have the following chain of (in)equalities:

$$\begin{aligned} & \mathrm{tr}((\mathbb{Q}_n - \Pi_n(b))(\rho_n - e^{nb}\sigma_n)) \\ &= \mathrm{tr}((\mathbb{Q}_n - \Pi_n(b))(\rho_n - e^{nb}\sigma_n)^+) \\ & \quad + \mathrm{tr}((\mathbb{Q}_n - \Pi_n(b))(\rho_n - e^{nb}\sigma_n)^-) \end{aligned} \quad (2.24)$$

$$\begin{aligned} & \leq \mathrm{tr}((\mathbb{I} - \Pi_n(b))(\rho_n - e^{nb}\sigma_n)^+) \\ & \quad - \mathrm{tr}(\Pi_n(b)(\rho_n - e^{nb}\sigma_n)^-) \end{aligned} \quad (2.25)$$

$$= 0 + 0 \quad (2.26)$$

$$= 0. \quad (2.27)$$

We can now bound $\mathrm{tr}(\mathbb{Q}_n \sigma_n)$ as

$$\mathrm{tr}(\mathbb{Q}_n \sigma_n) \geq \mathrm{tr}((\mathbb{Q}_n - \Pi_n(b))\sigma_n) \quad (2.28)$$

$$\geq e^{-nb} \mathrm{tr}((\mathbb{Q}_n - \Pi_n(b))\rho_n) \quad (2.29)$$

$$\geq e^{-nb} ((1 - \epsilon) - \mathrm{tr}(\Pi_n(b)\rho_n)), \quad (2.30)$$

where (2.29) follows from (2.27) and (2.30) follows from (2.23). Combining (2.22) and (2.30) implies

$$\underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \mathrm{tr}(\mathbb{Q}_n \sigma_n) \leq \lim_{n \rightarrow \infty} -\frac{1}{n} \log(e^{-nb}(c - \epsilon)) = b. \quad (2.31)$$

Since (2.31) holds for every $\epsilon \in (0, c)$ and every sequence $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$ satisfying (2.23), we obtain (2.21).

Combining (2.13) and (2.21) proves the lemma. \square

2.4 The Bounds

A quantum channel has a set of quantum states available as input, where each state is labeled by a different $x \in \mathcal{X}$. We define a family of normalized and mutually orthogonal vectors $\{|x\rangle\}_{x \in \mathcal{X}}$, and let \mathbb{A} be the Hilbert space spanned by $\{|x\rangle\}_{x \in \mathcal{X}}$. When the state labeled x is fed into the

channel, the output state is denoted as ρ_x which is a density operator acting on some Hilbert space \mathbb{B} . The channel can be described by a CPM \mathfrak{W} from $\mathcal{D}(\mathbb{A})$ to $\mathcal{D}(\mathbb{B})$, where $\mathcal{D}(\cdot)$ denotes the set of density operators on a Hilbert space, such that

$$\rho_x = \mathfrak{W}(|x\rangle\langle x|), \quad x \in \mathcal{X}. \quad (2.32)$$

For any PMF P_X on \mathcal{X} we denote

$$\pi^{\mathbb{A}\mathbb{B}} \triangleq \sum_{x \in \mathcal{X}} P_X(x) |x\rangle\langle x|^{\mathbb{A}} \otimes \rho_x^{\mathbb{B}}, \quad (2.33)$$

and denote the partial traces of $\pi^{\mathbb{A}\mathbb{B}}$ by $\pi^{\mathbb{A}}$ and $\pi^{\mathbb{B}}$, respectively, i.e.,

$$\pi^{\mathbb{A}} = \sum_{x \in \mathcal{X}} P_X(x) |x\rangle\langle x|^{\mathbb{A}}, \quad (2.34)$$

$$\pi^{\mathbb{B}} = \sum_{x \in \mathcal{X}} P_X(x) \rho_x^{\mathbb{B}}. \quad (2.35)$$

We consider the *one-shot* scenario where the channel is used only once. In this scenario, a *codebook* \mathcal{C} of size m is a list of input labels $x_i \in \mathcal{X}$, $i \in \{1, \dots, m\}$, and a corresponding *decoding POVM* acts on \mathbb{B} and has m measurement operators. We say a decoding error occurs if a state x_i is fed into the channel but the output of the decoding POVM is not i . An (R, ϵ) -*code* consists of a codebook of size 2^R and a corresponding decoding POVM such that, when the message is chosen uniformly over $\{1, \dots, 2^R\}$, the probability of a decoding error is at most ϵ .³

Note that a one-shot classical channel is a special case of a one-shot quantum channel where the CPM is restricted to being a stochastic kernel, and where the decoding POVM is restricted to being a guessing rule.

We are now ready to prove one-shot converse and achievability bounds expressed using $D_{\text{H}}^{\epsilon}(\rho||\sigma)$ on the amount of information that can be transmitted through a quantum channel. We begin with the converse bound.

³It is well known that, for single-user scenarios, the capacity does not depend on whether the average or the maximum probability of error is considered. For finite blocklength analysis, one can construct a code that has maximum probability of error not larger than 2ϵ from a code of average probability of error ϵ by sacrificing one bit.

Theorem 2.6. *If an (R, ϵ) -code exists for a one-shot quantum channel, then*

$$R \leq \sup_{P_X} D_{\text{H}}^{\epsilon}(\pi^{\text{A}\text{B}} \parallel \pi^{\text{A}} \otimes \pi^{\text{B}}). \quad (2.36)$$

Remark 2.7. *For a classical channel, (2.36) becomes*

$$R \leq \sup_{P_X} D_{\text{H}}^{\epsilon}(P_{XY} \parallel P_X \otimes P_Y), \quad (2.37)$$

where P_{XY} is the joint distribution on the channel input X and the channel output Y induced by P_X and the channel law.

We give two proofs for Theorem 2.6.

Proof 1. We choose a uniform distribution on the input labels used in the codebook. This yields the state

$$\pi^{\text{A}\text{B}} = 2^{-R} \sum_{i=1}^{2^R} |x_i\rangle\langle x_i| \otimes \rho_{x_i}. \quad (2.38)$$

To prove (2.36), it suffices to show

$$R \leq D_{\text{H}}^{\epsilon}(\pi^{\text{A}\text{B}} \parallel \pi^{\text{A}} \otimes \pi^{\text{B}}) \quad (2.39)$$

for the above $\pi^{\text{A}\text{B}}$. To this end, denote the decoding POVM operators by \mathbf{E}_i , $i \in \{1, \dots, 2^R\}$. Let

$$\mathbf{Q}^{\text{A}\text{B}} \triangleq \sum_{i=1}^{2^R} |x_i\rangle\langle x_i|^{\text{A}} \otimes \mathbf{E}_i^{\text{B}}, \quad (2.40)$$

then

$$0 \leq \mathbf{Q} \leq \mathbf{I}. \quad (2.41)$$

Because the average probability of error is not larger than ϵ ,

$$\text{tr}(\mathbf{Q}\pi^{\text{A}\text{B}}) = 2^{-R} \sum_{i=1}^{2^R} \text{tr}(\mathbf{E}_i \rho_{x_i}) \geq 1 - \epsilon. \quad (2.42)$$

Thus, to prove the theorem, it suffices to show

$$R \leq -\log \text{tr}(\mathbf{Q}(\pi^{\text{A}} \otimes \pi^{\text{B}})). \quad (2.43)$$

In fact, (2.43) holds with equality, as we justify as follows:

$$\mathrm{tr}(\mathbb{Q}(\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}})) = \mathrm{tr} \left(2^{-R} \sum_{j=1}^{2^R} \mathbb{E}_j \left(2^{-R} \sum_{i=1}^{2^R} \rho_{x_i} \right) \right) \quad (2.44)$$

$$= 2^{-R} \mathrm{tr} \left(\mathbb{1} \left(2^{-R} \sum_{i=1}^{2^R} \rho_{x_i} \right) \right) \quad (2.45)$$

$$= 2^{-R}. \quad (2.46)$$

□

Proof 2. As in Proof 1, we choose a uniform distribution on the input labels used in the codebook and show that (2.39) holds for the state $\pi^{\mathbb{A}\mathbb{B}}$ as in (2.38). To this end, it is enough to show that

$$R \leq D_{\mathrm{H}}^{\epsilon}(P_{MM'} \| P_M \otimes P_{M'}), \quad (2.47)$$

where $P_{MM'}$ is the joint distribution of the transmitted message M and the decoder's guess M' . Indeed, the decoding POVM combined with the inverse of the encoding map can be viewed as a CPM which maps $\pi^{\mathbb{A}\mathbb{B}}$ to $P_{MM'}$ and which maps $\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}$ to $P_M \otimes P_{M'}$. Thus it follows from the DPI for $D_{\mathrm{H}}^{\epsilon}(\rho \| \sigma)$ that

$$D_{\mathrm{H}}^{\epsilon}(P_{MM'} \| P_M \otimes P_{M'}) \leq D_{\mathrm{H}}^{\epsilon}(\pi^{\mathbb{A}\mathbb{B}} \| \pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}), \quad (2.48)$$

and hence (2.47) implies (2.39).

To prove (2.47), we suggest a (possibly suboptimal) scheme to distinguish between $P_{MM'}$ and $P_M \otimes P_{M'}$. The scheme guesses $P_{MM'}$ if $M = M'$, and guesses $P_M \otimes P_{M'}$ otherwise. In this scheme, the probability of guessing $P_M \otimes P_{M'}$ when $P_{MM'}$ is true is exactly the probability that $M \neq M'$ computed from $P_{MM'}$, namely, the average probability of a decoding error, and is thus not larger than ϵ by assumption. On the other hand, the probability of guessing $P_{MM'}$ when $P_M \otimes P_{M'}$ is true is

$$\sum_{i=1}^{2^R} P_M(i) \cdot P_{M'}(i) = 2^{-R} \sum_{i=1}^{2^R} P_{M'}(i) \quad (2.49)$$

$$= 2^{-R}. \quad (2.50)$$

Thus we obtain (2.47) and hence (2.39). □

We next show the achievability bound.

Theorem 2.8. *For any $\epsilon > \epsilon' > 0$ and $c > 0$ there exists an (R, ϵ) -code for a one-shot quantum channel with*

$$R \geq \sup_{P_X} D_H^{\epsilon'}(\pi^{\mathbb{A}\mathbb{B}} \parallel \pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}}) - \log \frac{2 + c + c^{-1}}{\epsilon - (1 + c)\epsilon'}. \quad (2.51)$$

The main technique we need for proving Theorem 2.8 is the following lemma by Hayashi and Nagaoka [18, Lemma 2]:

Lemma 2.9. *For any positive real c and any operators $0 \leq S \leq I$ and $T \geq 0$,*

$$I - (S + T)^{-1/2} S (S + T)^{-1/2} \leq (1 + c)(I - S) + (2 + c + c^{-1})T. \quad (2.52)$$

Proof of Theorem 2.8. Fix $\epsilon' \in [0, \epsilon]$ and $c > 0$. For any P_X , we shall first show that, for any Q acting on $\mathbb{A}\mathbb{B}$ such that $0 \leq Q \leq I$ and $\text{tr}(Q\pi^{\mathbb{A}\mathbb{B}}) \geq 1 - \epsilon'$, there exist a codebook of size 2^R and a decoding POVM which satisfy

$$\Pr(\text{error}) \leq (1 + c)\epsilon' + (2 + c + c^{-1})(2^R - 1) \text{tr}(Q(\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}})). \quad (2.53)$$

To this end, for any such Q , we define

$$A_x^{\mathbb{B}} \triangleq \text{tr}_{\mathbb{A}}(|x\rangle\langle x|^{\mathbb{A}} \otimes I^{\mathbb{B}}) Q. \quad (2.54)$$

We randomly generate a codebook by choosing the codewords IID P_X . Let the corresponding decoding POVM have elements

$$E_i = \left(\sum_{j=1}^{2^R} A_{x_j} \right)^{-\frac{1}{2}} A_{x_i} \left(\sum_{j=1}^{2^R} A_{x_j} \right)^{-\frac{1}{2}}. \quad (2.55)$$

For a specific codebook \mathcal{C} consisting of $\{x_j\}_{j \in \{1, \dots, 2^R\}}$, and the transmitted codeword x_i , the probability of error is given by

$$\Pr(\text{error} | x_i, \mathcal{C}) = \text{tr}((I - E_i)\rho_{x_i}). \quad (2.56)$$

Using Lemma 2.9 we obtain

$$\begin{aligned} \Pr(\text{error} | x_i, \mathcal{C}) &\leq (1 + c)(1 - \text{tr}(A_{x_i}\rho_{x_i})) \\ &\quad + (2 + c + c^{-1}) \sum_{j \neq i} \text{tr}(A_{x_j}\rho_{x_i}). \end{aligned} \quad (2.57)$$

Averaging over all codebooks we have

$$\begin{aligned} \Pr(\text{error}|x_i) &\leq (1+c)(1 - \text{tr}(\mathbf{A}_{x_i}\rho_{x_i})) \\ &\quad + (2+c+c^{-1})(2^R-1) \text{tr} \left(\left(\sum_{x \in \mathcal{X}} P_X(x)\mathbf{A}_x \right) \rho_{x_i} \right). \end{aligned} \quad (2.58)$$

Further averaging the above inequality over the transmitted codeword x_i we obtain

$$\begin{aligned} \Pr(\text{error}) &\leq (1+c) \left(1 - \sum_{x \in \mathcal{X}} P_X(x) \text{tr}(\mathbf{A}_x \rho_x) \right) \\ &\quad + (2+c+c^{-1})(2^R-1) \\ &\quad \cdot \text{tr} \left(\left(\sum_{x \in \mathcal{X}} P_X(x)\mathbf{A}_x \right) \left(\sum_{x \in \mathcal{X}} P_X(x)\rho_x \right) \right). \end{aligned} \quad (2.59)$$

To see that this is the desired inequality, we first check

$$1 - \epsilon' \leq \text{tr}(\mathbf{Q}\pi^{\mathbb{A}\mathbb{B}}) \quad (2.60)$$

$$= \sum_{x \in \mathcal{X}} P_X(x) \text{tr}(\mathbf{Q}|x\rangle\langle x|^{\mathbb{A}} \otimes \rho_x^{\mathbb{B}}) \quad (2.61)$$

$$= \sum_{x \in \mathcal{X}} P_X(x) \text{tr}(\mathbf{A}_x \rho_x); \quad (2.62)$$

and then check

$$\text{tr}(\mathbf{Q}(\pi^{\mathbb{A}} \otimes \pi^{\mathbb{B}})) = \sum_{x' \in \mathcal{X}} P_X(x') \text{tr} \left(\mathbf{Q}|x'\rangle\langle x'| \otimes \left(\sum_{x \in \mathcal{X}} P_X(x)\rho_x \right) \right) \quad (2.63)$$

$$= \sum_{x' \in \mathcal{X}} P_X(x') \text{tr} \left(\mathbf{A}_{x'} \left(\sum_{x \in \mathcal{X}} P_X(x)\rho_x \right) \right) \quad (2.64)$$

$$= \text{tr} \left(\left(\sum_{x \in \mathcal{X}} P_X(x)\mathbf{A}_x \right) \left(\sum_{x \in \mathcal{X}} P_X(x)\rho_x \right) \right). \quad (2.65)$$

Using (2.59), (2.62) and (2.65) we see that (2.53) holds for the average probability of error averaged over the randomly chosen codebooks. Thus there must exist at least one codebook that satisfies (2.53). Furthermore, since a codebook that satisfies (2.53) can be found for any \mathbf{Q} satisfying $0 \leq \mathbf{Q} \leq \mathbf{I}$ and $\text{tr}(\mathbf{Q}\pi^{\mathbb{A}\mathbb{B}}) \geq 1 - \epsilon'$, we conclude that there must exist a codebook that satisfies

$$\Pr(\text{error}) \leq (1+c)\epsilon' + (2+c+c^{-1})2^{R-D_{\text{H}}^{\epsilon'}(\pi^{\mathbb{A}\mathbb{B}}\|\pi^{\mathbb{A}}\otimes\pi^{\mathbb{B}})}. \quad (2.66)$$

By rearranging terms in the above inequality we obtain (2.51). \square

For classical channels the achievability bound of Theorem 2.8 can be tightened, and its proof can be simplified. This classical bound is very similar to [33, Theorem 17].

Theorem 2.10. *For any $\epsilon > \epsilon' > 0$, there exists an (R, ϵ) -code for a one-shot classical channel with*

$$R \geq \sup_{P_X} D_0^{\epsilon'}(P_{XY} \| P_X \otimes P_Y) - \log \frac{1}{\epsilon - \epsilon'}, \quad (2.67)$$

where P_{XY} is the joint distribution on the channel input X and the channel output Y induced by P_X and the channel law.

Proof. For any distribution P_X on \mathcal{X} , we randomly generate a codebook of size 2^R such that the 2^R codewords are chosen IID P_X . For any $\Phi : \mathcal{X} \otimes \mathcal{Y} \rightarrow [0, 1]$ satisfying

$$\int_{\mathcal{X} \otimes \mathcal{Y}} \Phi \, dP_{XY} \geq 1 - \epsilon', \quad (2.68)$$

we use the following random decoding rule: when y is received, select some or none of the messages such that message j is selected with probability $\Phi(x_j, y)$ independently of the other messages. If only one message is selected, output this message; otherwise declare an error.

To analyze the error probability, suppose i was the transmitted message. The error event is the union of \mathcal{E}_1 and \mathcal{E}_2 , where \mathcal{E}_1 denotes the event that some message other than i is selected; \mathcal{E}_2 denotes the event that message i is not selected.

We first bound $\Pr(\mathcal{E}_1)$ averaged over all codebooks. Fix x_i and y . The probability averaged over all codebooks of selecting a particular message other than i is given by

$$\int_{\mathcal{X}} \Phi(x, y) \, dP_X. \quad (2.69)$$

Since there are $(2^R - 1)$ such messages, we can use the union bound to obtain

$$\mathbb{E}[\Pr(\mathcal{E}_1 | x_i, y)] \leq (2^R - 1) \cdot \int_{\mathcal{X}} \Phi(x, y) \, dP_X. \quad (2.70)$$

Since the RHS of (2.70) does not depend on x_i , we further have

$$\mathbb{E}[\Pr(\mathcal{E}_1|y)] \leq (2^R - 1) \cdot \int_{\mathcal{X}} \Phi(x, y) dP_X. \quad (2.71)$$

Averaging this inequality over y gives

$$\mathbb{E}[\Pr(\mathcal{E}_1)] \leq (2^R - 1) \int_{\mathcal{Y}} \left(\int_{\mathcal{X}} \Phi(x, y) dP_X \right) dP_Y \quad (2.72)$$

$$= (2^R - 1) \int_{\mathcal{X} \otimes \mathcal{Y}} \Phi d(P_X \otimes P_Y). \quad (2.73)$$

On the other hand, the probability of \mathcal{E}_2 averaged over all generated codebooks can be bounded as

$$\begin{aligned} \mathbb{E}[\Pr(\mathcal{E}_2)] &= \int_{\mathcal{X} \otimes \mathcal{Y}} (1 - \Phi) dP_{XY} \\ &\leq \epsilon'. \end{aligned} \quad (2.74)$$

Combining (2.73) and (2.74) yields

$$\Pr(\text{error}) \leq (2^R - 1) \int_{\mathcal{X} \otimes \mathcal{Y}} \Phi d(P_X \otimes P_Y) + \epsilon'. \quad (2.75)$$

Since (2.75) holds for every Φ satisfying (2.68), we obtain

$$\Pr(\text{error}) \leq (2^R - 1) \cdot 2^{-D_0^{\epsilon'}(P_{XY} \| P_X \otimes P_Y)} + \epsilon'. \quad (2.76)$$

Hence, there exist at least one codebook whose error probability satisfies (2.76). By rearranging terms we obtain the desired bound. \square

2.5 Asymptotic Analysis of the Bounds

The results of Section 2.4 apply to the transmission of a message in a one-shot channel. Obviously, n uses of a channel can be viewed as a big one-shot channel. We can thus employ Theorems 2.6 and 2.8 to derive the known expressions for channel capacities. The simplest such case is the capacity of a memoryless channel. Here we can directly apply Stein's lemma to recover the HSW Theorem [20,37] which says that the capacity of a memoryless channel is given by

$$C = \sup_n \sup_{P_{X_n}} D(\pi^{\mathbb{A}_n \mathbb{B}^{\otimes n}} \| \pi^{\mathbb{A}_n} \otimes \pi^{\mathbb{B}^{\otimes n}}) \quad (2.77)$$

$$= \sup_n \sup_{P_{X_n}} I(\mathbb{A}_n; \mathbb{B}^{\otimes n}), \quad (2.78)$$

where $I(\cdot; \cdot)$ denotes the mutual information between two Hilbert spaces, and where \mathbb{A}_n denotes the space spanned by the orthonormal states $\{|x_n\rangle\}_{x_n \in \mathcal{X}_n}$ which label input states over n channel uses.⁴

In the following, we show how analogous asymptotic formulas can be obtained for a channel whose structure is arbitrary. Such a channel is described by CPMs from $\mathcal{D}(\mathbb{A}_n)$ to $\mathcal{D}(\mathbb{B}^{\otimes n})$ for all $n \in \mathbb{Z}^+$. An (n, R, ϵ) -code on a channel consists of a codebook with entries $(x_{i,1}, \dots, x_{i,n}) \in \mathcal{X}_n$, $i \in \{1, \dots, 2^R\}$, and a decoding POVM acting on $\mathbb{B}^{\otimes n}$ such that average probability of error is no larger than ϵ . We define capacity and optimistic capacity in the same way as in classical information theory [18, 42]:

Definition 2.11. *The capacity C of a channel is the supremum over all R for which there exists a sequence of (n, R, ϵ_n) -codes such that*

$$\lim_{n \rightarrow \infty} \epsilon_n = 0. \quad (2.79)$$

The optimistic capacity \overline{C} of a channel is the supremum over all R for which there exists a sequence of (n, R, ϵ_n) -codes such that

$$\underline{\lim}_{n \rightarrow \infty} \epsilon_n = 0. \quad (2.80)$$

Given Definition 2.11, the next theorem is an immediate consequence to Theorems 2.6 and 2.8.

Theorem 2.12. *For any quantum channel, capacity is given by*

$$C = \lim_{\epsilon \downarrow 0} \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sup_{P_{X_n}} D_{\text{H}}^{\epsilon}(\pi^{\mathbb{A}_n \mathbb{B}^{\otimes n}} \| \pi^{\mathbb{A}_n} \otimes \pi^{\mathbb{B}^{\otimes n}}), \quad (2.81)$$

and optimistic capacity is given by

$$\overline{C} = \lim_{\epsilon \downarrow 0} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sup_{P_{X_n}} D_{\text{H}}^{\epsilon}(\pi^{\mathbb{A}_n \mathbb{B}^{\otimes n}} \| \pi^{\mathbb{A}_n} \otimes \pi^{\mathbb{B}^{\otimes n}}). \quad (2.82)$$

Theorem 2.12 is equivalent to [18, Theorem 1]. Here the equivalence of (2.81) to [18, (7)] follows from Lemma 2.5, and the rest follows in

⁴Note that \mathbb{A}_n cannot be replaced by $\mathbb{A}^{\otimes n}$, which is spanned by labels of *product* input states over n channel uses.

similar ways. The general capacity formula for a classical channel [46, (1.4)] is a special case of (2.81).

We can also use Theorems 2.6 and 2.8 to study the ϵ -capacities which are usually defined as follows in classical information theory (see, for example, [46]):

Definition 2.13. *The ϵ -capacity C_ϵ of a channel is the supremum over all R such that, for every large enough n , there exists an (n, R, ϵ) -code. The optimistic ϵ -capacity \overline{C}_ϵ of a channel is the supremum over all R for which there exist (n, R, ϵ) -codes for infinitely many n 's.*

From Theorems 2.6 and 2.8 we immediately have the following:

Theorem 2.14. *For any quantum communication channel and any $\epsilon \in (0, 1)$, the ϵ -capacity of the channel satisfies*

$$C_\epsilon \leq \varliminf_{n \rightarrow \infty} \frac{1}{n} \sup_{P_{X_n}} D_{\text{H}}^\epsilon(\pi^{\mathbb{A}_n \mathbb{B}^{\otimes n}} \parallel \pi^{\mathbb{A}_n} \otimes \pi^{\mathbb{B}^{\otimes n}}), \quad (2.83)$$

$$C_\epsilon \geq \lim_{\epsilon' \uparrow \epsilon} \varliminf_{n \rightarrow \infty} \frac{1}{n} \sup_{P_{X_n}} D_{\text{H}}^{\epsilon'}(\pi^{\mathbb{A}_n \mathbb{B}^{\otimes n}} \parallel \pi^{\mathbb{A}_n} \otimes \pi^{\mathbb{B}^{\otimes n}}); \quad (2.84)$$

and the optimistic ϵ -capacity of the channel satisfies

$$\overline{C}_\epsilon \leq \overline{\varliminf}_{n \rightarrow \infty} \frac{1}{n} \sup_{P_{X_n}} D_{\text{H}}^\epsilon(\pi^{\mathbb{A}_n \mathbb{B}^{\otimes n}} \parallel \pi^{\mathbb{A}_n} \otimes \pi^{\mathbb{B}^{\otimes n}}), \quad (2.85)$$

$$\overline{C}_\epsilon \geq \lim_{\epsilon' \uparrow \epsilon} \overline{\varliminf}_{n \rightarrow \infty} \frac{1}{n} \sup_{P_{X_n}} D_{\text{H}}^{\epsilon'}(\pi^{\mathbb{A}_n \mathbb{B}^{\otimes n}} \parallel \pi^{\mathbb{A}_n} \otimes \pi^{\mathbb{B}^{\otimes n}}). \quad (2.86)$$

As special cases of Theorem 2.14, we obtain the corresponding results for classical channels as in [46, Theorem 6] and [41, Theorem 7].

Chapter 3

The Poisson Channel with Side-Information

3.1 Introduction

The Poisson channel is often used to model optical communication with a direct-detection receiver. Its capacity has been computed using different methods in [21], [10], and [47]. In this chapter we consider a variation of the Poisson channel where there are spurious counts in the output, and where the positions of the spurious counts are known to the encoder as side-information (SI).

Communication aided by SI at the encoder provides a rich source for interesting problems in information theory. A DMC with IID states is described by a transition law

$$W(y|x, s) \tag{3.1a}$$

and a state law

$$P_S(s), \tag{3.1b}$$

where x denotes the channel input, y denotes the channel output, and s denotes the state of the channel.

Shannon [40] studied the case where the SI is unknown to the decoder and known *causally* to the encoder. In this scenario, before sending x_k , the encoder knows $\{s_1, \dots, s_k\}$. He showed that the capacity of the channel (3.1) is given by

$$\sup_{P_U} I(U; Y), \quad (3.2)$$

where U is a random variable taking value in the set of mappings from channel states to channel inputs, and where $P_{Y|U}$ is given by

$$P_{Y|U}(y|u) = \sum_s P_S(s)W(y|u(s), s). \quad (3.3)$$

A different scenario is where the SI is still unknown to the decoder but known *noncausally* to the encoder. In this case, the encoder knows the whole state sequence before starting to encode. The capacity of (3.1) in this case was found by Gel'fand and Pinsker [13] to be

$$\sup_{\substack{P_{U|S} \\ g: (U,S) \rightarrow X}} I(U; Y) - I(U; S), \quad (3.4)$$

where the supremum is over all joint distributions on (S, U, Y) of the form

$$P_S(s)P_{U|S}(u|s)W(y|g(u, s), s). \quad (3.5)$$

The capacity with noncausal SI was computed in various cases. The case of writing on binary memory with defects was solved by Kusnetsov and Tsybakov [23] before the discovery of the formula (3.4); the case of writing on binary memory with defects *and noise* was solved by Heegard and El Gamal [19]; and the case of additive Gaussian noise channel with additive states (“writing on dirty paper”) was solved by Costa [5].

We consider the continuous-time Poisson channel where the channel states correspond to spurious counts in the output, and where the decoder does not know which counts are spurious. Such a scenario might occur, e.g., when the encoder communicates by exposing a piece of film with light, but when some pixels on the film have already been exposed. We study first the case where the SI is known noncausally to the encoder, and then the case where the SI is known causally to the encoder.

In the noncausal SI case, we distinguish between two settings. In the first setting we assume that the states are chosen by an adversary

subject to a constraint on the average number of spurious counts per second, and we allow the encoder and the decoder to use random codes. Since the state sequence can be arbitrary, we cannot use Gel'fand and Pinsker's formula (3.4). Neither can we use Ahlswede's result for an arbitrarily varying channel (AVC) with states known to the encoder [1], which does not hold when the adversary is subject to a constraint. Instead, as in [23] and [19], we show by construction that the capacity with no spurious counts can be achieved on this channel with random codes. In the second setting we assume that the spurious counts are random (but not necessarily a homogeneous Poisson process). Using the result from the first setting, we show that the capacity with no spurious counts is achievable on this channel with deterministic codes.

For the causal SI setting, we show that, even if the spurious counts obey a homogeneous Poisson law, causal SI does not increase the capacity of this channel. Thus, as in [23] and [19], for the peak-limited Poisson channel causal SI does not increase capacity at all, while noncausal SI increases it to that of the same channel model but without states.

The rest of this chapter is arranged as follows: in Section 3.2 we give a brief review of the peak-limited Poisson channel; in Section 3.3 we state and prove the capacity results for the noncausal SI case; and in Section 3.4 we discuss the causal SI case.

3.2 Preliminary on the Poisson Channel

In a Poisson channel, the channel input \mathbf{x} is a nonnegative, real-valued, continuous-time signal, and the channel output \mathbf{y} is a point pattern, i.e., a set of points on the real line. For a given input signal \mathbf{x} , the random output \mathbf{Y} is a Poisson process whose time- t intensity is $x(t) + \lambda$, where λ is a nonnegative constant called the *dark current*. Hence, for a given \mathbf{x} , the number of points in \mathbf{Y} that fall in the interval $(t_1, t_2]$ has a Poisson distribution of mean

$$\xi = \int_{t_1}^{t_2} (x(t) + \lambda) dt. \quad (3.6)$$

We denote such a distribution by $\text{Pois}_\xi(\cdot)$ so

$$\text{Pois}_\xi(\vartheta) = e^{-\xi} \frac{\xi^\vartheta}{\vartheta!}, \quad \vartheta \in \mathbb{Z}_0^+. \quad (3.7)$$

We impose a peak-power constraint on the input so

$$x(t) \leq \mathbf{A}, \quad t \in \mathbb{R}, \quad (3.8)$$

where $\mathbf{A} > 0$ is the maximal allowed input power. We denote the capacity of this channel (in bits per second) by $C_{\text{Pois}}(\mathbf{A}, \lambda)$.

The value of $C_{\text{Pois}}(\mathbf{A}, \lambda)$ was first found by Kabanov [21] and by Davis [10] using martingale techniques. Later, Wyner [47] showed that $C_{\text{Pois}}(\mathbf{A}, \lambda)$ can be achieved by dividing the channel into small time-slots and by then looking at the resulting DMC. We describe this approach in the following scheme:

Scheme 3.1. *We divide the time-interval $[0, T]$ into slots of Δ seconds long. The encoder first maps the message m to a $\{0, 1\}$ -valued vector \mathbf{x}^Δ of length $\frac{T}{\Delta}$.¹ It then maps \mathbf{x}^Δ to the channel input \mathbf{x} as*

$$x(t) = \mathbf{A}x_{\lceil \frac{t}{\Delta} \rceil}^\Delta, \quad t \in [0, T]. \quad (3.9)$$

The decoder maps the channel output \mathbf{y} , which is a point pattern on $[0, T]$, to a vector \mathbf{y}^Δ of length $\frac{T}{\Delta}$ as follows: if \mathbf{y} has at least one point in the time-slot $((i-1)\Delta, i\Delta]$, choose $y_i^\Delta = 1$; otherwise choose $y_i^\Delta = 0$. The decoder then maps \mathbf{y}^Δ to m' .

Scheme 3.1 reduces the continuous-time Poisson channel to a DMC with input x^Δ and output y^Δ whose transition law, when $\Delta \ll 1$, is approximately

$$W(1|x^\Delta) = \begin{cases} \lambda\Delta, & x^\Delta = 0 \\ (\mathbf{A} + \lambda)\Delta, & x^\Delta = 1. \end{cases} \quad (3.10)$$

We denote the capacity of the channel (3.10) (in bits per channel use) by $C^\Delta(\mathbf{A}, \lambda)$. Wyner [47] showed that

$$C_{\text{Pois}}(\mathbf{A}, \lambda) = \lim_{\Delta \downarrow 0} \frac{C^\Delta(\mathbf{A}, \lambda)}{\Delta} \quad (3.11)$$

$$= \max_{p \in (0,1)} \left\{ p(\mathbf{A} + \lambda) \log(\mathbf{A} + \lambda) + (1-p)\lambda \log \lambda - (p\mathbf{A} + \lambda) \log(p\mathbf{A} + \lambda) \right\}. \quad (3.12)$$

¹When T is not divisible by Δ , we consider \mathbf{x} as a signal on $[0, T']$ where $T' = \lceil \frac{T}{\Delta} \rceil \Delta$. When we let Δ tend to zero, the difference between T and T' also tends to zero. Henceforth we ignore this technicality and assume T is divisible by Δ .

Note that $C(\mathbf{A}, \lambda)$ can also be written as

$$C_{\text{Pois}}(\mathbf{A}, \lambda) = \max_{p \in (0,1)} (p\mathbf{A} + \lambda) D \left(\text{Ber} \left(\frac{p(\mathbf{A} + \lambda)}{p\mathbf{A} + \lambda} \right) \parallel \text{Ber}(p) \right), \quad (3.13)$$

where $\text{Ber}(\pi)$ denotes the Bernoulli distribution of parameter π .

3.3 Noncausal SI

We now consider the continuous-time Poisson channel as described in Section 3.2, but with spurious counts in its output. We assume that the positions of these spurious counts are known noncausally to the encoder but not to the decoder. We consider two settings: for Section 3.3.1 we assume that the spurious counts are generated by a malicious adversary and that the encoder and the decoder are allowed to use random codes; for Section 3.3.2 we assume that the spurious counts occur randomly according to a law known to both encoder and decoder and that only deterministic codes are allowed.

3.3.1 Random Codes against Arbitrary States

Consider the Poisson channel as described in Section 3.2 but with an adversary who generates spurious counts at the decoder and reveals the times at which they occur to the encoder before communication begins. These spurious counts form a point pattern \mathbf{s} on the interval $[0, T]$. Thus, conditional on the input being \mathbf{x} , output process \mathbf{Y} is the union of \mathbf{s} and \mathbf{Z} , where \mathbf{Z} is a Poisson process whose time- t intensity is $x(t) + \lambda$. In terms of counting functions, this can be written as

$$N_Y(t) = N_Z(t) + n_s(t), \quad t \in \mathbb{R}. \quad (3.14)$$

We assume that the adversary is subject to the restriction that, within each time-interval $[0, T]$, the total number of spurious counts cannot exceed νT for some constant ν which is known to both encoder and decoder:

$$n_s(T) - n_s(0) \leq \nu T. \quad (3.15)$$

In a (deterministic) (T, R) -code, the encoder maps the message $m \in \{1, \dots, 2^{RT}\}$ and the channel state \mathbf{s} to an input signal \mathbf{x} , and the decoder guesses the message m from the channel output \mathbf{y} . A *random* (T, R) -code is a probability distribution on all deterministic (T, R) -codes.²

A rate R (in bits per second) is said to be *achievable with random codes* on the channel (3.14) if, for every $T > 0$, there exists a random (T, R) -code such that, as T tends to infinity, the average probability of a guessing error tends to zero for *all possible* \mathbf{s} . The *random coding capacity* of this channel is defined as the supremum over all rates achievable with random codes.

Since the adversary can choose not to introduce any spurious counts, the random coding capacity of the channel (3.14) is upper-bounded by $C_{\text{Pois}}(\mathbf{A}, \lambda)$, which is given in (3.13). Our first result in this chapter is that this bound is tight:

Theorem 3.2. *For any positive A , λ and ν , the random coding capacity of the channel (3.14), where \mathbf{s} is known noncausally to the encoder but unknown to the decoder, is equal to $C_{\text{Pois}}(\mathbf{A}, \lambda)$.*

Before proving Theorem 3.2, we describe the following scheme which is based on Scheme 3.1 to reduce the continuous-time channel to a discrete-time channel.

Scheme 3.3. *As in Scheme 3.1, we divide the time-interval $[0, T]$ into slots of Δ seconds long. From the channel state \mathbf{s} , we produce a length- $\frac{T}{\Delta}$ vector \mathbf{s}^Δ in the following way: if \mathbf{s} has at least one point in the interval $((i-1)\Delta, i\Delta]$, choose $s_i^\Delta = 1$; otherwise choose $s_i^\Delta = 0$. The encoder then maps the message m and the discrete-time state sequence \mathbf{s}^Δ to a $\{0, 1\}$ -valued, length- $\frac{T}{\Delta}$ vector \mathbf{x}^Δ . The rest is the same as in Scheme 3.1.*

Proof of Theorem 3.2. We only need to prove the lower bound. Namely, we need to show that any rate below the RHS of (3.13) is achievable with random codes on the channel (3.14). To this end, for fixed positive constants T , R and α , we shall construct a block-coding scheme to transmit RT bits of information using $(1 + \alpha)T$ seconds. (Later we shall choose α arbitrarily small.) We divide the block into two phases, first the *training phase* and then the *information phase*, where the training

²For more explicit formulations of random and deterministic codes, see [27] and references therein.

phase is αT seconds long and the information phase is T seconds long. Within each phase we apply Scheme 3.3 to reduce the channel to a DMC with an arbitrary state sequence. When $\Delta \ll 1$, the transition law of this DMC is approximated by

$$W(1|x^\Delta, s^\Delta) = \begin{cases} \lambda\Delta, & x^\Delta = 0, s^\Delta = 0 \\ (A + \lambda)\Delta, & x^\Delta = 1, s^\Delta = 0 \\ 1, & s^\Delta = 1. \end{cases} \quad (3.16)$$

From now on we shall refer to time-slots where $s^\Delta = 1$ as “stuck slots.”

Denote the total number of stuck slots in the information phase by μT . Then, by the state constraint,

$$\mu \leq (1 + \alpha)\nu. \quad (3.17)$$

In the training phase the encoder tells the decoder the value of μT . To do this, the encoder and the decoder use the channel as an AVC. Namely, in this phase the encoder ignores its knowledge about the times of the spurious counts. Since, by the state constraint, the total number of stuck slots in the whole transmission block cannot exceed $(1 + \alpha)\nu T$, we know that the total number of stuck slots in the training phase also cannot exceed $(1 + \alpha)\nu T$. It can be easily verified using the formula for random coding capacity of the AVC with state constraints [8] that the random coding capacity of the AVC (3.16) under this constraint is proportional to Δ for small Δ . Thus, the amount of information that can be reliably transmitted in the training phase is proportional to αT for large T and small Δ . On the other hand, according to (3.17), we only need $\log((1 + \alpha)\nu T)$ bits to describe μT . Thus we conclude that, for any $\alpha > 0$, for large enough T and small enough Δ , the training phase can be accomplished successfully with high probability.

We next describe the information phase (which is T seconds long). If the training phase is successful, then in the information phase, both the encoder and the decoder know the total number of stuck slots μT , but only the encoder knows their positions. With such knowledge, they can use the following random coding scheme to transmit RT bits of information in this phase:

- **Codebook:** Generate $2^{(R+R')T}$ codewords independently with the symbols within every codeword chosen IID $\text{Ber}(p)$. Label the codewords as $\mathbf{x}^\Delta(m, k)$ where $m \in \{1, \dots, 2^{RT}\}$ and $k \in \{1, \dots, 2^{R'T}\}$.

- **Encoder:** For a given message $m \in \{1, \dots, 2^{RT}\}$ and a state sequence \mathbf{s} , find a k such that

$$\sum_{j=1}^{T/\Delta} \mathbb{I}\left\{ (x_j^\Delta(m, k), s_j) = (1, 1) \right\} \geq (1 - \epsilon) \frac{p(\mathcal{A} + \lambda)}{p\mathcal{A} + \lambda} \mu T, \quad (3.18)$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function so the left-hand side (LHS) is the number of slots where $\mathbf{x}^\Delta(m, k)$ and \mathbf{s}^Δ are both one. Send $\mathbf{x}^\Delta(m, k)$. If no such k can be found, send an arbitrary sequence.

- **Decoder:** Find a codeword $\mathbf{x}^\Delta(m', k')$ in the codebook such that, for the observed \mathbf{y}^Δ ,

$$\begin{aligned} \sum_{j=1}^{T/\Delta} \mathbb{I}\left\{ (x_j^\Delta(m', k'), y_j^\Delta) = (1, 1) \right\} \\ \geq (1 - \epsilon) \frac{p(\mathcal{A} + \lambda)}{p\mathcal{A} + \lambda} (p\mathcal{A} + \lambda + \mu) T. \end{aligned} \quad (3.19)$$

Output m' . If no such codeword can be found, declare an error.

We next analyze the error probability of this random codebook. There are three types of errors which we analyze separately:

- The encoder cannot find a k such that $\mathbf{x}^\Delta(m, k)$ satisfies (3.18). We know that the total number of stuck slots in this phase is μT . Since the codebook is generated independently of the stuck slots, we know that the symbols of a particular codeword at these slots are drawn IID $\text{Ber}(p)$. By Sanov's theorem [7], for large T , the probability that a particular codeword satisfies the requirement, i.e., has at least $(1 - \epsilon) \frac{p(\mathcal{A} + \lambda)}{p\mathcal{A} + \lambda} \mu T$ ones in these μT slots, is approximately $2^{\mu T D\left(\text{Ber}\left((1 - \epsilon) \frac{p(\mathcal{A} + \lambda)}{p\mathcal{A} + \lambda}\right) \parallel \text{Ber}(p)\right)}$. Therefore, when T tends to infinity, the probability of this error tends to zero if we choose

$$R' > \mu D\left(\text{Ber}\left(\left(1 - \epsilon\right) \frac{p(\mathcal{A} + \lambda)}{p\mathcal{A} + \lambda}\right) \parallel \text{Ber}(p)\right). \quad (3.20)$$

- There are less than $(1 - \epsilon) \frac{p(\mathcal{A} + \lambda)}{p\mathcal{A} + \lambda} (p\mathcal{A} + \lambda + \mu) T$ slots where the transmitted codeword and the output are both equal to one. The

probability of this error tends to zero as T tends to infinity by the Law of Large Numbers.

- There is some $\tilde{\mathbf{x}}^\Delta$ which is not the transmitted codeword but which satisfies (3.19). To analyze the probability of this error, we first note that, by the Law of Large Numbers, when T is large, the number of slots where $y_j^\Delta = 1$ is close to $(p\mathbf{A} + \lambda + \mu)T$. We also note that a particular codeword that is not transmitted is drawn IID $\text{Ber}(p)$ independently of \mathbf{y}^Δ . Therefore, again by Sanov's theorem, the probability that this codeword has at least $(1 - \epsilon)\frac{p(\mathbf{A} + \lambda)}{p\mathbf{A} + \lambda}(p\mathbf{A} + \lambda + \mu)T$ ones at the approximately $(p\mathbf{A} + \lambda + \mu)T$ slots where $y_j^\Delta = 1$ is approximately $2^{(p\mathbf{A} + \lambda + \mu)TD} \left(\text{Ber}\left(\left(1 - \epsilon\right)\frac{p(\mathbf{A} + \lambda)}{p\mathbf{A} + \lambda}\right) \parallel \text{Ber}(p) \right)$. Thus, when T tends to infinity, this probability tends to zero if we choose

$$R + R' < (p\mathbf{A} + \lambda + \mu)D \left(\text{Ber} \left(\left(1 - \epsilon\right)\frac{p(\mathbf{A} + \lambda)}{p\mathbf{A} + \lambda} \right) \parallel \text{Ber}(p) \right). \quad (3.21)$$

By combining (3.20) and (3.21) and noting that ϵ can be chosen to be arbitrarily small, we conclude that, for every $p \in (0, 1)$, when T is large and Δ is small, successful transmission in the information phase can be achieved with high probability as long as

$$R < (p\mathbf{A} + \lambda)D \left(\text{Ber} \left(\frac{p(\mathbf{A} + \lambda)}{p\mathbf{A} + \lambda} \right) \parallel \text{Ber}(p) \right). \quad (3.22)$$

Furthermore, since we have shown that the training phase can be accomplished with any positive α , the overall transmission rate, which is equal to $\frac{R}{1 + \alpha}$, can also be made arbitrarily close to the RHS of (3.22). Optimization over p implies that we can achieve all rates up to the RHS of (3.13) with random coding. \square

3.3.2 Deterministic Codes against Random States

We next consider the case where, rather than being an arbitrary point pattern chosen by an adversary, the spurious counts are random. Such random counts can be modeled by a random point process \mathbf{S} which is independent of the message, so the channel output \mathbf{Y} is the union of \mathbf{S} and the Poisson process \mathbf{Z} whose time- t intensity is $x(t) + \lambda$ which is

independent of \mathbf{S} conditional on \mathbf{x} . In terms of counting functions

$$N_Y(t) = N_Z(t) + N_S(t), \quad t \in \mathbb{R}. \quad (3.23)$$

We assume that \mathbf{S} has a finite expected number of points per second:

$$\overline{\lim}_{t \rightarrow \infty} \frac{\mathbb{E}[N_S(t)]}{t} < \infty. \quad (3.24)$$

Note that (3.24) is satisfied, for example, when \mathbf{S} is a homogeneous Poisson process. We also assume that the law of \mathbf{S} is known to both encoder and decoder (and, in particular, the code may depend on the law of \mathbf{S}), while the realization of \mathbf{S} is known noncausally to the encoder but unknown to the decoder. A rate is said to be achievable on this channel if, for every $T > 0$, there exists a deterministic (T, R) -code such that the average probability of error averaged over \mathbf{S} tends to zero as T tends to infinity. The capacity of this channel is defined as the supremum over all achievable rates.

Theorem 3.4. *The capacity of the channel (3.23), where the realization of \mathbf{S} is known noncausally to the encoder but unknown to the decoder, is equal to $C_{\text{Pois}}(\mathbf{A}, \lambda)$, irrespective of the law of \mathbf{S} .*

Proof. We first observe that the capacity of the channel (3.23) cannot exceed $C_{\text{Pois}}(\mathbf{A}, \lambda)$. This is because we can mimic the channel (3.23) over a channel without spurious counts as follows: The encoder and the decoder use common randomness (which does not increase the capacity of a single-user channel without states) to generate \mathbf{S} and then the decoder ignores its realization.

We shall next show that any rate below $C_{\text{Pois}}(\mathbf{A}, \lambda)$ is achievable on (3.23). Fix any $\epsilon > 0$ and $R < C_{\text{Pois}}(\mathbf{A}, \lambda)$. Let

$$\zeta \triangleq \overline{\lim}_{t \rightarrow \infty} \frac{\mathbb{E}[N_S(t)]}{t}. \quad (3.25)$$

Since $\zeta < \infty$, there exists a t_0 such that

$$\mathbb{E}[N_S(t)] \leq 2\zeta, \quad t > t_0. \quad (3.26)$$

Using this and Markov's inequality we have

$$\Pr \left[N_S(t) \geq \frac{2\zeta}{\epsilon} \right] \leq \epsilon, \quad t > t_0. \quad (3.27)$$

Thus the error probability of a random (T, R) -code where $T > t_0$ can be bounded as

$$\Pr[\text{error}] \leq \Pr \left[\text{error}, N_S(T) \geq \frac{2\zeta}{\epsilon} \right] + \Pr \left[\text{error}, N_S(T) < \frac{2\zeta}{\epsilon} \right] \quad (3.28)$$

$$\begin{aligned} &\leq \Pr \left[N_S(T) \geq \frac{2\zeta}{\epsilon} \right] \\ &\quad + \Pr \left[\text{error} \mid N_S(T) < \frac{2\zeta}{\epsilon} \right] \cdot \Pr \left[N_S(T) < \frac{2\zeta}{\epsilon} \right] \end{aligned} \quad (3.29)$$

$$\leq \epsilon + \Pr \left[\text{error} \mid N_S(T) < \frac{2\zeta}{\epsilon} \right], \quad T > t_0. \quad (3.30)$$

To bound the second term on the RHS of (3.30) we use Theorem 3.2 which says that there exists t_1 such that, for any $T > t_1$, there exists a random (T, R) -code whose average error probability conditional on any realization of \mathbf{S} satisfying $n_s(T) < \frac{2\zeta}{\epsilon}$ is not larger than ϵ . Therefore, for such codes,

$$\Pr \left[\text{error} \mid N_S(T) < \frac{2\zeta}{\epsilon} \right] \leq \epsilon, \quad T > t_1. \quad (3.31)$$

Combining (3.30) and (3.31) yields that for any $T > \max\{t_0, t_1\}$ there exists a random (T, R) -code for which

$$\Pr[\text{error}] \leq 2\epsilon. \quad (3.32)$$

Since this is true for all $\epsilon > 0$ and $R < C_{\text{Pois}}(\mathbf{A}, \lambda)$, we conclude that all rates below $C_{\text{Pois}}(\mathbf{A}, \lambda)$ are achievable on (3.23) with random codes.

We next observe that we do not need to use random codes. Indeed, picking for every T and R the best deterministic (T, R) -code yields at worst the same average error probability as that of any random (T, R) -code. Thus we conclude that any rate below $C_{\text{Pois}}(\mathbf{A}, \lambda)$ is achievable on (3.23) with deterministic codes, and hence the capacity of (3.23) is equal to $C_{\text{Pois}}(\mathbf{A}, \lambda)$. \square

3.4 Causal SI

In this section we shall argue that causal SI does not increase the capacity of a peak-limited Poisson channel. We look at the simplest case where the spurious counts occur as a homogeneous Poisson process of intensity μ .

We shall be imprecise regarding the definition of causality in continuous time by directly looking at the DMC (3.16).³ Since the continuous-time state \mathbf{S} is a Poisson process, for $\Delta \ll 1$, the discrete-time state sequence \mathbf{S}^Δ is IID $\text{Ber}(\mu\Delta)$.

To argue that having causal SI does not increase capacity, we shall show that every mapping u from channel states to input symbols (as in (3.2)) is equivalent to a *deterministic input symbol* in the sense that it induces the same output distribution as the latter. Indeed, since when $s^\Delta = 1$ the input symbol x^Δ has no influence on the output Y^Δ , we know that the value of $u(1)$ does not influence the output distribution. Therefore $u: s^\Delta \mapsto u(s^\Delta)$ is equivalent to the mapping that maps both 0 and 1 to $u(0)$, and is thus also equivalent to the deterministic input symbol $x^\Delta = u(0)$.

In a more explicit argument, we use the capacity expression (3.2). For any distribution P_U on \mathcal{U} , we let

$$P_{X^\Delta}(x^\Delta) = \sum_{u: u(0)=x^\Delta} P_U(u). \quad (3.33)$$

Then, for the above P_U and P_X ,

$$I(U; Y^\Delta) = I(X^\Delta; Y^\Delta). \quad (3.34)$$

Therefore

$$\sup_{P_U} I(U; Y^\Delta) \leq \sup_{P_{X^\Delta}} I(X^\Delta; Y^\Delta), \quad (3.35)$$

where the LHS is the capacity of the channel with causal SI, and where the RHS is the capacity of the channel with no SI. Thus we conclude that the capacity of our channel model with causal SI is not larger than that of the channel with no SI.

³For formulations of causality in continuous time see [24] or [22].

Chapter 4

Covering Point Patterns

4.1 Introduction

Chapter 3 concerns the Poisson channel where, in order to convey a number of bits, the encoder tries to *produce* a point pattern at the decoder's end. In this chapter, we consider a related problem where an encoder *observes* a point pattern which is to be described to a reconstructor using bits. Based on these bits, the reconstructor wishes to produce a covering-set—a subset of $[0, T]$ containing all the points—of least Lebesgue measure. There is a trade-off between the number of bits used and the Lebesgue measure of the covering-set. This trade-off can be formulated as a continuous-time rate-distortion problem. In this chapter we investigate this trade-off in the limit where $T \rightarrow \infty$.

When the point pattern is produced by a homogeneous Poisson process, this problem is the “dual” problem to that of the peak-limited Poisson channel with zero dark current in the sense of [6]. However, the duality results of [6] only apply to discrete memoryless channels and sources, so they cannot be directly used to solve our problem. Instead, we shall use a technique that is similar to Wyner's [47, 48] to find the desired rate-distortion function. We shall show that, if the point pattern is the outcome of a homogeneous Poisson process of intensity λ , and if the reconstructor is restricted to select covering-sets of average Lebesgue

measure not exceeding DT for some $D > 0$, then the minimum number of bits per second needed by the encoder to describe the pattern is $-\lambda \log D$.

Previous works [3, 4, 12, 36, 45] have studied rate-distortion functions of the Poisson process with different distortion measures. It is interesting to notice that our rate-distortion function, $-\lambda \log D$, is equal to the ones in [4] and in [45], where a queueing distortion measure is considered. This is no coincidence, since the Poisson channel is closely related to the queueing channel introduced in [2].

We also show that the Poisson process is the most difficult to cover, in the sense that any point process that, with high probability, has no more than λT points in $[0, T]$ can be described with $-\lambda \log D$ bits per second. This is even true if an adversary selects the point pattern, provided that the pattern contains no more than λ points per second and that the encoder and the reconstructor are allowed to use random codes.

Finally, we consider a Wyner-Ziv setting [49] of the problem where some points in the pattern are known to the reconstructor but the encoder does not know which ones they are. This can be viewed as a dual problem to the Poisson channel with noncausal SI studied in Chapter 3. We show that in this setting one can achieve the same minimum rate as when the transmitter *does* know the reconstructor's SI.

The rest of this chapter is arranged as follows: in Section 4.2 we present the result for the Poisson process; in Section 4.3 we present the results for general point processes and arbitrary point patterns; and in Section 4.4 we present the results for the Wyner-Ziv setting.

4.2 Covering a Poisson Process

Let \mathbf{X} be a homogeneous Poisson process of intensity λ on the interval $[0, T]$. Then its counting function $N_X(\cdot)$ satisfies

$$\Pr[N_X(t + \tau) - N_X(t) = k] = \text{Pois}_{\lambda\tau}(k) \quad (4.1)$$

for all $\tau \in [0, T]$, $t \in [0, T - \tau]$ and $k \in \{0, 1, \dots\}$, where $\text{Pois}_\xi(\cdot)$ denotes the Poisson distribution of mean ξ as in (3.7).

The encoder maps the realization of the Poisson process to a message in $\{1, \dots, 2^{TR}\}$. The reconstructor then maps this message to a $\{0, 1\}$ -valued, Lebesgue-measurable signal $\hat{\mathbf{x}}$ on $[0, T]$. We wish to minimize the total length of the region where $\hat{x}(t) = 1$ while guaranteeing that all points in the Poisson process lie in this region. See Figure 4.1 for an illustration.

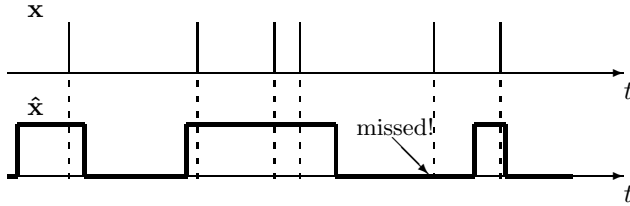


Figure 4.1: Covering a point pattern.

More formally, we formulate this problem as a continuous-time rate-distortion problem, where the distortion between the point pattern \mathbf{x} and the reproduction signal $\hat{\mathbf{x}}$ is

$$d(\mathbf{x}, \hat{\mathbf{x}}) \triangleq \begin{cases} \frac{\mu(\hat{x}^{-1}(1))}{T}, & \text{if all points in } \mathbf{x} \text{ are in } \hat{x}^{-1}(1) \\ \infty, & \text{otherwise} \end{cases} \quad (4.2)$$

where $\mu(\cdot)$ denotes the Lebesgue measure.

We say that (R, D) is an achievable rate-distortion pair for the homogeneous Poisson process of intensity λ if, for every $\epsilon > 0$, there exists some $T_0 > 0$ such that, for every $T > T_0$, there exists an encoder $f_T(\cdot)$ and a reconstructor $\phi_T(\cdot)$ of rate $R + \epsilon$ bits per second which, when applied to the Poisson process \mathbf{X} on $[0, T]$, gives

$$\mathbb{E}[d(\mathbf{X}, \phi_T(f_T(\mathbf{X})))] \leq D + \epsilon. \quad (4.3)$$

Denote by $R_{\text{Pois}}(D, \lambda)$ the minimum rate R such that (R, D) is achievable for the homogeneous Poisson process of intensity λ .

Theorem 4.1. *For all $D, \lambda > 0$,*

$$R_{\text{Pois}}(D, \lambda) = \begin{cases} -\lambda \log D \text{ bits per second,} & D \in (0, 1) \\ 0, & D \geq 1. \end{cases} \quad (4.4)$$

To prove Theorem 4.1, we propose a scheme to reduce the original problem to one for a discrete memoryless source. This is reminiscent of Wyner's scheme for reducing the peak-limited Poisson channel to a DMC [47], which we restated as Scheme 3.1. We shall show the optimality of this scheme in Lemma 4.3, and we shall then prove Theorem 4.1 by computing the best rate that is achievable using this scheme.

Scheme 4.2. *We divide the time-interval $[0, T]$ into slots of Δ seconds long. The encoder first maps the original point pattern \mathbf{x} to a $\{0, 1\}$ -valued vector \mathbf{x}^Δ of length $\frac{T}{\Delta}$ in the following way: if \mathbf{x} has at least one point in the time-slot $((i-1)\Delta, i\Delta]$, choose $x_i^\Delta = 1$; otherwise choose $x_i^\Delta = 0$. The encoder then maps \mathbf{x}^Δ to a message in $\{1, \dots, 2^{TR}\}$.*

Based on the encoder's message, the reconstructor produces a $\{0, 1\}$ -valued length- $\frac{T}{\Delta}$ vector $\hat{\mathbf{x}}^\Delta$ to meet the distortion criterion

$$\mathbb{E} \left[d^\Delta(\mathbf{X}^\Delta, \hat{\mathbf{X}}^\Delta) \right] \leq D, \quad (4.5)$$

where the distortion measure $d^\Delta(\cdot, \cdot)$ is given by

$$d^\Delta(0, 0) = 0 \quad (4.6a)$$

$$d^\Delta(0, 1) = 1 \quad (4.6b)$$

$$d^\Delta(1, 0) = \infty \quad (4.6c)$$

$$d^\Delta(1, 1) = 1. \quad (4.6d)$$

It then maps $\hat{\mathbf{x}}^\Delta$ to a continuous-time reconstruction signal $\hat{\mathbf{x}}$ through

$$\hat{x}(t) = \hat{x}_{\lceil \frac{t}{\Delta} \rceil}^\Delta, \quad t \in [0, T]. \quad (4.7)$$

Scheme 4.2 reduces the task of designing a code for the point process \mathbf{X} subject to distortion $d(\cdot, \cdot)$ to the task of designing a code for the random vector \mathbf{X}^Δ subject to the distortion $d^\Delta(\cdot, \cdot)$. The way we define $d^\Delta(\cdot, \cdot)$ yields the simple relation

$$d(\mathbf{x}, \hat{\mathbf{x}}) = d^\Delta(\mathbf{x}^\Delta, \hat{\mathbf{x}}^\Delta). \quad (4.8)$$

When \mathbf{X} is the homogeneous Poisson process of intensity λ , the components of \mathbf{X}^Δ are IID $\text{Ber}(1 - e^{-\lambda\Delta})$. Let $R_\Delta(D, \lambda)$ denote the rate-distortion function for \mathbf{X}^Δ and $d^\Delta(\cdot, \cdot)$. If we combine Scheme 4.2 with

¹As in Scheme 3.1, we ignore the issue whether T is divisible by Δ or not.

an optimal code for \mathbf{X}^Δ subject to $\mathbb{E} \left[d^\Delta(\mathbf{X}^\Delta, \hat{\mathbf{X}}^\Delta) \right] \leq D$, we can achieve any rate that is larger than

$$\frac{R_\Delta(D, \lambda) \text{ bits}}{\Delta \text{ seconds}}. \quad (4.9)$$

The next lemma, which is reminiscent of [48, Theorem 2.1], shows that when we let Δ tend to zero, there is no loss in optimality in using Scheme 4.2.

Lemma 4.3. *For all $D, \lambda > 0$,*

$$R_{\text{Pois}}(D, \lambda) = \lim_{\Delta \downarrow 0} \frac{R_\Delta(D, \lambda)}{\Delta}. \quad (4.10)$$

Proof. Given any rate-distortion code with 2^{TR} codewords $\hat{\mathbf{x}}_m$, $m \in \{1, \dots, 2^{TR}\}$ that achieves expected distortion D , we shall construct a new code that can be constructed through Scheme 4.2, that contains $(2^{TR} + 1)$ codewords, and that achieves an expected distortion that is arbitrarily close to D .

Denote the codewords of our new code by $\hat{\mathbf{w}}_m$, $m \in \{1, \dots, 2^{TR} + 1\}$. We choose the last codeword to be the constant 1. We next describe our choices for the other codewords. For every $\epsilon > 0$ and every $\hat{\mathbf{x}}_m$, we can approximate the set $\{t: \hat{x}_m(t) = 1\}$ by a set \mathcal{A}_m that is equal to a finite, say N_m , union of open intervals. More specifically,

$$\mu(\hat{x}_m^{-1}(1) \triangle \mathcal{A}_m) \leq 2^{-TR} \epsilon, \quad (4.11)$$

where \triangle denotes the symmetric difference between two sets (see, e.g., [35, Chapter 3, Proposition 15]). Define

$$\mathcal{B} \triangleq \bigcup_{m=1}^{2^{TR}} (\hat{x}_m^{-1}(1) \setminus \mathcal{A}_m), \quad (4.12)$$

and note that by (4.11)

$$\mu(\mathcal{B}) \leq \epsilon. \quad (4.13)$$

For each \mathcal{A}_m , $m \in \{1, \dots, 2^{TR}\}$, let

$$\hat{\mathbf{w}}_m = \mathbb{I} \left\{ \left((\lceil t/\Delta \rceil - 1) \Delta, \lceil t/\Delta \rceil \Delta \right) \cap \mathcal{A}_m \neq \emptyset \right\}, \quad (4.14)$$

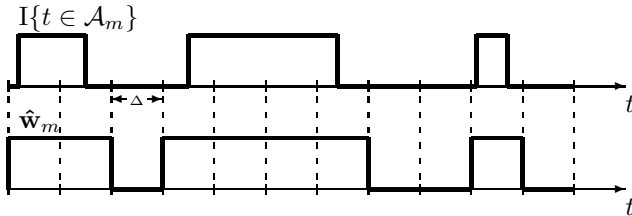


Figure 4.2: Constructing \hat{w}_m from \mathcal{A}_m .

where $I\{\cdot\}$ denotes the indicator function. Note that $\mathcal{A}_m \subseteq \hat{w}_m^{-1}(1)$. See Figure 4.2 for an illustration of this construction. Let

$$N \triangleq \max_{m \in \{1, \dots, 2^{TR}\}} N_m. \quad (4.15)$$

It can be seen that

$$\mu(\hat{w}_m^{-1}(1)) - \mu(\mathcal{A}_m) \leq 2N\Delta, \quad m \in \{1, \dots, 2^{TR}\}. \quad (4.16)$$

Our encoder works as follows: if \mathbf{x} contains no point in \mathcal{B} , it maps \mathbf{x} to the same message as the given encoder; otherwise it maps \mathbf{x} to the index $(2^{TR} + 1)$ of the all-one codeword. To analyze the distortion, first consider the case where \mathbf{x} contains no point in \mathcal{B} . In this case, all points in \mathbf{x} must be covered by the selected codeword \hat{w}_m . By (4.11) and (4.16), the difference $d(\mathbf{x}, \hat{w}_m) - d(\mathbf{x}, \hat{x}_m)$, if positive, can be made arbitrarily small by choosing small ϵ and Δ . Next consider the case where \mathbf{x} does contain points in \mathcal{B} . By (4.13), the probability that this happens can be made arbitrarily small by choosing ϵ small, therefore its contribution to the expected distortion can also be made arbitrarily small. We conclude that our code $\{\hat{w}_m\}$ can achieve a distortion that is arbitrarily close to the distortion achieved by the original code $\{\hat{x}_m\}$. \square

Proof of Theorem 4.1. We derive $R_{\text{Pois}}(D, \lambda)$ by computing the RHS of (4.10). To compute $R_\Delta(D, \lambda)$ we apply Shannon's formula of the rate-distortion function for a discrete memoryless source [39]:

$$R_\Delta(D, \lambda) = \min_{P_{\hat{X}^\Delta | X^\Delta} : \mathbb{E}[d_\Delta(X^\Delta, \hat{X}^\Delta)] \leq D} I(X^\Delta; \hat{X}^\Delta).^2 \quad (4.17)$$

²Strictly speaking, since our distortion measure is unbounded, we need to modify

When $D \in (0, 1)$, the conditional distribution $P_{\hat{X}^\Delta | X^\Delta}$ which achieves the minimum on the RHS of (4.17) is

$$P_{\hat{X}^\Delta | X^\Delta}^*(1|0) = De^{\lambda\Delta} - e^{\lambda\Delta} + 1, \quad (4.18a)$$

$$P_{\hat{X}^\Delta | X^\Delta}^*(1|1) = 1. \quad (4.18b)$$

Computing the mutual information $I(X^\Delta; \hat{X}^\Delta)$ under this $P_{\hat{X}^\Delta | X^\Delta}^*$ yields

$$R_\Delta(D, \lambda) = H_b(D) - e^{-\lambda\Delta} H_b(De^{\lambda\Delta} - e^{\lambda\Delta} + 1), \quad D \in (0, 1), \quad (4.19)$$

where $H_b(\cdot)$ denotes the binary entropy function.

When $D \geq 1$, it is optimal to choose $\hat{X}^\Delta = 1$ (deterministically), yielding

$$R_\Delta(D, \lambda) = 0, \quad D \geq 1. \quad (4.20)$$

Combining (4.10), (4.19) and (4.20) and computing the limit as Δ tends to zero yields (4.4). \square

4.3 Covering General Point Processes and Arbitrary Point Patterns

We next consider a general point process \mathbf{Y} . We assume that there exists some λ such that

$$\lim_{t \rightarrow \infty} \Pr \left[\frac{N_{\mathbf{Y}}(t)}{t} > \lambda + \delta \right] = 0 \quad \text{for all } \delta > 0. \quad (4.21)$$

Condition (4.21) is satisfied, for example, when \mathbf{Y} is an ergodic process whose expected number of points per second is less than or equal to λ .

Since the Poisson process is memoryless, one naturally expects it to be the most difficult to describe. This is indeed the case, as the next theorem shows.

Shannon's proof of this formula in order to use it for our problem. This can be done by letting the reconstructor produce the all-one sequence, which yields bounded distortion for any source sequence, whenever no codeword can be found that is jointly typical with the source sequence.

Theorem 4.4. *The pair $(R_{\text{Pois}}(D, \lambda), D)$ is achievable on any point process satisfying (4.21).*

Before proving Theorem 4.4, we state a stronger result. Consider a point pattern \mathbf{z} chosen by an adversary on the interval $[0, T]$ which contains no more than λT points. The corresponding counting function $n_z(\cdot)$ must then satisfy

$$n_z(T) \leq \lambda T. \quad (4.22)$$

The encoder and the reconstructor are allowed to use random codes. Namely, they fix a distribution on all (deterministic) codes of a certain rate on $[0, T]$. According to this distribution, they randomly pick a code (which is not revealed to the adversary). They then apply it to the point pattern \mathbf{z} chosen by the adversary. We say that (R, D) is achievable with random coding against an adversary subject to (4.22) if, for every $\epsilon > 0$, there exists some T_0 such that, for every $T > T_0$, there exists a random code on $[0, T]$ of rate $R + \epsilon$ such that the expected distortion between any \mathbf{z} satisfying (4.22) and its reconstruction is smaller than $D + \epsilon$.

Theorem 4.5. *The pair $(R_{\text{Pois}}(D, \lambda), D)$ is achievable with random coding against an adversary subject to (4.22).*

Proof. First note that when $D \geq 1$, the encoder does not need to describe the pattern: the reconstructor simply produces the all-one function, yielding distortion 1 for any \mathbf{z} . Hence the pair $(0, D)$ is achievable with random coding.

Next consider $D \in (0, 1)$. We use Scheme 4.2 to reduce the original problem to one of random coding for an arbitrary discrete-time sequence \mathbf{z}^Δ . Here \mathbf{z}^Δ is $\{0, 1\}$ -valued, has length $\frac{T}{\Delta}$, and satisfies

$$\sum_{i=1}^{T/\Delta} z_i^\Delta \leq \lambda T. \quad (4.23)$$

We shall construct a random code of rate $\frac{R}{\Delta}$ which, when applied to any \mathbf{z}^Δ satisfying (4.23), yields

$$\mathbb{E} \left[d^\Delta(\mathbf{z}^\Delta, \hat{\mathbf{Z}}^\Delta) \right] < D + \epsilon, \quad (4.24)$$

where the random vector $\hat{\mathbf{Z}}^\Delta$ is the result of applying the random encoder and reconstructor to \mathbf{z}^Δ . Combined with Scheme 4.2 this random code

will yield a random code on the continuous-time point pattern \mathbf{z} that achieves the rate-distortion pair (R, D) .

Our discrete-time random code consists of 2^{TR} $\{0, 1\}$ -valued, length- $\frac{T}{\Delta}$ random sequences $\hat{\mathbf{Z}}^\Delta(m)$, $m \in \{1, \dots, 2^{TR}\}$. The first sequence $\hat{\mathbf{Z}}^\Delta(1)$ is chosen deterministically to be the all-one sequence. The other $2^{TR} - 1$ sequences are drawn independently, with each sequence drawn IID $\text{Ber}(D)$.

To describe source sequence \mathbf{z}^Δ , the encoder looks for a codeword $\hat{\mathbf{z}}^\Delta(m)$, $m \in \{2, \dots, 2^{TR}\}$ such that

$$\hat{z}_i^\Delta(m) = 1 \text{ whenever } z_i^\Delta = 1. \quad (4.25)$$

If it finds one or more such codewords, it sends the index of the first one; otherwise it sends 1. The reconstructor outputs the sequence $\hat{\mathbf{z}}^\Delta(m)$ where m is the message it receives from the encoder.

We next analyze the expected distortion of this random code for a fixed \mathbf{z}^Δ satisfying (4.23). Define

$$\mu \triangleq \frac{\sum_{i=1}^{T/\Delta} z_i^\Delta}{T}, \quad (4.26)$$

and note that by (4.23) $\mu \leq \lambda$. Denote by \mathcal{E} the event that the encoder cannot find $\hat{\mathbf{z}}^\Delta(m)$, $m \in \{2, \dots, 2^{TR}\}$ satisfying (4.25). If \mathcal{E} occurs, the encoder sends 1 and the resulting distortion is equal to 1.

The probability that a randomly drawn codeword $\hat{\mathbf{Z}}^\Delta(m)$ satisfies (4.25) is

$$D^{\mu T} \geq D^{\lambda T} = 2^{(\lambda \log D)T}. \quad (4.27)$$

Because the codewords $\hat{\mathbf{Z}}^\Delta(m)$, $m \in \{2, \dots, 2^{TR}\}$ are chosen independently, if we choose $R > -\lambda \log D$, then $\Pr[\mathcal{E}] \rightarrow 0$ as $T \rightarrow \infty$. Hence, for large enough T , the contribution to the expected distortion from the event \mathcal{E} can be ignored.

We next analyze the expected distortion conditional on \mathcal{E}^c . In this case, the reproduction $\hat{\mathbf{Z}}^\Delta$ has the following distribution: at positions where \mathbf{z}^Δ takes the value 1, $\hat{\mathbf{Z}}^\Delta$ must also be 1; at other positions the elements of $\hat{\mathbf{Z}}^\Delta$ have the IID $\text{Ber}(D)$ distribution. Thus the expected value of $\sum_{i=1}^{T/\Delta} \hat{z}_i^\Delta$ is $\mu T + D(\frac{T}{\Delta} - \mu T)$, and

$$\mathbb{E} \left[d^\Delta(\mathbf{z}^\Delta, \hat{\mathbf{Z}}^\Delta) \middle| \mathcal{E}^c \right] = D + (1 - D)\mu\Delta. \quad (4.28)$$

When we let Δ tend to zero, this value tends to D . We have thus shown that, for small enough Δ , we can achieve the pair $(\frac{R}{\Delta}, D)$ on \mathbf{z}^Δ with random coding whenever $R > -\lambda \log D$, and therefore we can also achieve (R, D) on the continuous-time point pattern \mathbf{z} with random coding if $R > -\lambda \log D$. \square

We next use Theorem 4.5 to prove Theorem 4.4.

Proof of Theorem 4.4. It follows from Theorem 4.5 that, on any point process satisfying (4.21), the pair $(R_{\text{Pois}}(D, \lambda + \delta), D)$ is achievable with *random coding*. Further, since there is no adversary, the existence of a good random code guarantees the existence of a good deterministic code. Hence $(R_{\text{Pois}}(D, \lambda + \delta), D)$ is also achievable on this process with deterministic coding. Theorem 4.4 now follows when we let δ tend to zero, since $R_{\text{Pois}}(D, \cdot)$ is a continuous function. \square

4.4 Some Points are Known to the Reconstructor

In this section we consider a Wyner-Ziv setting for our problem. We first consider the case where \mathbf{X} is a homogeneous Poisson process of intensity λ . (Later we consider an arbitrary point pattern.) Assume that each point in \mathbf{X} is known to the reconstructor independently with probability p . Also assume that the encoder does not know which points are known to the reconstructor. The encoder maps \mathbf{X} to a message in $\{1, \dots, 2^{TR}\}$, and the reconstructor produces a Lebesgue-measurable, $\{0, 1\}$ -valued signal $\hat{\mathbf{X}}$ on $[0, T]$ based on this message and the positions of the points that he knows. The achievability of a rate-distortion pair is defined in the same way as in Section 4.2. Denote the smallest rate R for which (R, D) is achievable by $R_{\text{WZ}}(D, \lambda, p)$.

Obviously, $R_{\text{WZ}}(D, \lambda, p)$ is lower-bounded by the smallest achievable rate when the transmitter *does* know which points are known to the reconstructor. The latter rate is equal to $R_{\text{Pois}}(D, (1-p)\lambda)$. Indeed, when the encoder knows which points are known to the reconstructor, it is optimal for it to describe only the remaining points, which themselves form a homogeneous Poisson process of intensity $(1-p)\lambda$. The reconstructor

then selects a set based on this description to cover the points unknown to it and adds to this set the points it knows. Thus,

$$R_{\text{WZ}}(D, \lambda, p) \geq R_{\text{Pois}}(D, (1-p)\lambda). \quad (4.29)$$

The next theorem shows that (4.29) holds with equality.

Theorem 4.6. *Knowing the points at the reconstructor only is as good as knowing them also at the encoder:*

$$R_{\text{WZ}}(D, \lambda, p) = R_{\text{Pois}}(D, (1-p)\lambda). \quad (4.30)$$

To prove Theorem 4.6, it remains to show that the pair $(R_{\text{Pois}}(D, (1-p)\lambda), D)$ is achievable. We shall show this as a consequence of a stronger result concerning arbitrary sources.

Consider an arbitrary point pattern \mathbf{z} on $[0, T]$ chosen by an adversary. The adversary is allowed to put at most λT points in \mathbf{z} . Also, it must reveal all but at most νT points to the reconstructor, without telling the encoder which points it has revealed. The encoder and the reconstructor are allowed to use random codes, where the encoder is a random mapping from \mathbf{z} to a message in $\{1, \dots, 2^{TR}\}$, and where the reconstructor is a random mapping from this message, together with the point pattern that it knows, to a $\{0, 1\}$ -valued, Lebesgue-measurable signal $\hat{\mathbf{z}}$. The distortion $d(\mathbf{z}, \hat{\mathbf{z}})$ is defined as in (4.2).

Theorem 4.7. *Against an adversary who puts at most λT points on $[0, T]$ and reveals all but at most νT points to the reconstructor, the rate-distortion pair $(R_{\text{Pois}}(D, \nu), D)$ is achievable with random coding.*

In order to prove Theorem 4.7, we describe a variation of Scheme 4.2.

Scheme 4.8. *We apply Scheme 4.2, but, additionally, the reconstructor maps the point pattern \mathbf{s} that it knows to a vector \mathbf{s}^Δ : if \mathbf{s} has at least one point in the interval $((i-1)\Delta, i\Delta]$, it chooses $s_i^\Delta = 1$; otherwise it chooses $s_i^\Delta = 0$. The discrete-time reconstruction sequence $\hat{\mathbf{x}}^\Delta$ is determined by the encoder's message and \mathbf{s}^Δ .*

Proof of Theorem 4.7. The case $D \geq 1$ is trivial, so we shall only consider the case where $D \in (0, 1)$. The encoder and the reconstructor first use Scheme 4.8 to reduce the problem to discrete-time. Define

$$\mu \triangleq \frac{\sum_{i=1}^{T/\Delta} z_i \Delta}{T}, \quad (4.31)$$

and note that, by assumption, $\mu \leq \lambda$. If $\mu \leq \nu$, then we can ignore the reconstructor's SI and use the random code of Theorem 4.5. Henceforth we assume $\mu > \nu$.

Since there are at most νT points that are unknown to the reconstructor,

$$\sum_{i=1}^{T/\Delta} s_i^\Delta \geq (\mu - \nu)T. \quad (4.32)$$

The encoder conveys the value of μT to the receiver using bits. Since μT is an integer between 0 and λT , the number of bits per second needed to describe it tends to zero as T tends to infinity.

Next, the encoder and the reconstructor randomly generate $2^{T(R+\tilde{R})}$ independent codewords

$$\hat{\mathbf{z}}^\Delta(m, k), \quad m \in \{1, \dots, 2^{TR}\}, \quad k \in \{1, \dots, 2^{T\tilde{R}}\}, \quad (4.33)$$

where each codeword is generated IID $\text{Ber}(D)$.

To describe \mathbf{z}^Δ , the encoder looks for a codeword $\hat{\mathbf{z}}^\Delta(m, k)$ such that

$$\hat{z}_i^\Delta(m, k) = 1 \text{ whenever } z_i^\Delta = 1. \quad (4.34)$$

If it finds one or more such codewords, it sends the index m of the first one; otherwise it tells the reconstructor to produce the all-one sequence.

When the reconstructor receives the index m , it looks for an index $\tilde{k} \in \{1, \dots, 2^{T\tilde{R}}\}$ such that

$$\hat{z}_i^\Delta(m, \tilde{k}) = 1 \text{ whenever } s_i^\Delta = 1. \quad (4.35)$$

If there is only one such codeword, it outputs it as the reconstruction; if there are more than one such codewords, it outputs the all-one sequence.

To analyze the expected distortion for \mathbf{z}^Δ over this random code, first consider the event that the encoder cannot find a codeword satisfying (4.34). Note that the probability that a randomly generated codeword satisfies (4.34) is $D^{\mu T}$, so the probability of this event tends to zero as T tends to infinity provided that

$$R + \tilde{R} > -\mu \log D. \quad (4.36)$$

Next consider the event that the reconstructor finds more than one \tilde{k} satisfying (4.35). The probability that a randomly generated codeword satisfies (4.35) is $D^{\sum_{i=1}^{T/\Delta} s_i^\Delta}$. Consequently, by (4.32) the probability of this event tends to zero as T tends to infinity provided

$$\tilde{R} < -(\mu - \nu) \log D. \quad (4.37)$$

Finally, if the encoder finds a codeword satisfying (4.34) and the reconstructor finds only one codeword satisfying (4.35), then the two codewords must be the same. Following the same calculations as in the proof of Theorem 4.5, the expected distortion in this case tends to D as Δ tends to zero.

Combining (4.36) and (4.37), we can make the expected distortion arbitrarily close to D as $T \rightarrow \infty$ if

$$R > -\nu \log D. \quad (4.38)$$

□

Proof of Theorem 4.6. The claim follows from (4.29), Theorem 4.7, and the Law of Large Numbers. □

Chapter 5

The Discrete-Time Poisson Channel at Low Input Powers

5.1 Introduction

We consider the discrete-time memoryless Poisson channel whose input x is in the set \mathbb{R}_0^+ of nonnegative reals and whose output y is in the set \mathbb{Z}_0^+ of nonnegative integers. Conditional on the input $X = x$, the output Y has a Poisson distribution of mean $(\lambda + x)$, where $\lambda \geq 0$ is called the *dark current*. Using the notation (3.7) the channel law $W(\cdot|\cdot)$ can be written as

$$W(y|x) = \text{Pois}_{\lambda+x}(y), \quad x \in \mathbb{R}_0^+, y \in \mathbb{Z}_0^+. \quad (5.1)$$

As discussed in Chapter 1, this channel is often used to model pulse-amplitude modulated optical communication with a direct-detection receiver. Here the input x is proportional to the product of the transmitted light intensity by the pulse duration; the output y models the number of photons arriving at the receiver during the pulse duration; and λ models the average number of extraneous counts that appear in y in addition to those associated with the illumination x .

The *average-power constraint*¹ is

$$\mathbb{E}[X] \leq E \quad (5.2)$$

where $E > 0$ is the maximum allowed average power.

The *peak-power constraint* is that with probability one

$$X \leq A. \quad (5.3)$$

We assume throughout this chapter that $A > 0$. In the absence of a peak-power constraint we write $A = \infty$.

No analytic expression for the capacity of this channel is known. In [38] Shamai showed that the capacity-achieving input distribution is discrete with the number of mass points depending on E and A . In [26] Lapidoth and Moser derived the asymptotic capacity of this channel in the regime where both the allowed average power and allowed peak power tend to infinity with their ratio held fixed.

In this chapter we seek the asymptotic capacity of the Poisson channel when the allowed average input power tends to zero with the allowed peak-power—if finite—held fixed. We consider two different cases for the dark current λ . The first is when the dark current tends to zero proportionally to the average power. This corresponds to the wide-band regime where the pulse duration tends to zero.² The second case is when the dark current is constant. This corresponds to the regime where the transmitter is weak.

Our lower bounds on channel capacity in the various cases are all based on binary inputs. Our upper bounds are derived using duality (see [25] and references therein). In some cases our lower and upper bounds asymptotically coincide (Theorem 5.1). An efficient way to compute asymptotic capacities at low average input powers is to compute the

¹The word “power” here has the meaning “average number of photons transmitted per channel use.” If we denote by P the standard “power” in physics, namely, energy per unit time (in watts), then the notion of “power” in this chapter is really $\eta P \Delta / \hbar \omega$, where η is the detector’s quantum efficiency, Δ is the pulse duration (in sec), and $\hbar \omega$ is the photon energy (in joules) at the operating frequency ω (in rad/sec).

²Note that by “wide-band” we mean that the communication bandwidth, i.e., the reciprocal of the pulse duration Δ , is large enough so that $\eta P \Delta / \hbar \omega \ll 1$, but this bandwidth is still much smaller than the optical center frequency ω . Once the bandwidth becomes comparable to the optical center frequency, photon-flux is no longer proportional to input power, and therefore our channel model becomes inadequate.

capacity per unit cost [44]. However, we shall see that, apart from one case (equation (5.10)), the capacity per unit cost is infinite, i.e., the capacity tends to zero more slowly than linearly with the average power.

Among the results in this chapter, the special case of zero dark current was derived independently in [28, 43].

The rest of this chapter is arranged as follows: in Section 5.2 we provide the results with some discussions; in Section 5.3 we prove the lower bounds; and in Section 5.4 we prove the upper bounds.

5.2 Results

Let $C(\lambda, E, A)$ denote the capacity of the Poisson channel with dark current λ under constraints (5.2) and (5.3):

$$C(\lambda, E, A) = \sup I(X; Y) \quad (5.4)$$

where the supremum is over input distributions satisfying (5.2) and (5.3).

When λ is proportional to E , the low-average-power asymptotic capacity of the Poisson channel is given in the following theorem. Note that this also includes the case where the dark current is zero.

Theorem 5.1. *For any $c \geq 0$ and $A \in (0, \infty]$,*

$$\lim_{E \downarrow 0} \frac{C(cE, E, A)}{E \log \frac{1}{E}} = 1. \quad (5.5)$$

Recall that, for any $\alpha, \beta > 0$, the sum of two independent random variables with the Poisson distributions $\text{Pois}_\alpha(\cdot)$ and $\text{Pois}_\beta(\cdot)$ has the Poisson distribution $\text{Pois}_{\alpha+\beta}(\cdot)$. Thus, we can produce any Poisson channel with nonzero dark current from a Poisson channel with zero dark current by having the receiver add an independent Poisson random variable to the channel output. Since this cannot increase capacity,

$$C(0, E, A) \geq C(cE, E, A), \quad c, E, A > 0. \quad (5.6)$$

Consequently, to prove Theorem 5.1, we only need to show

$$\lim_{\mathbb{E} \downarrow 0} \frac{C(c\mathbb{E}, \mathbb{E}, \mathcal{A})}{\mathbb{E} \log \frac{1}{\mathbb{E}}} \geq 1, \quad c > 0, \mathcal{A} \in (0, \infty], \quad (5.7)$$

$$\lim_{\mathbb{E} \downarrow 0} \frac{C(0, \mathbb{E}, \infty)}{\mathbb{E} \log \frac{1}{\mathbb{E}}} \leq 1. \quad (5.8)$$

We shall prove (5.7) in Section 5.3.1 and (5.8) in Section 5.4.1.

Remarks about Theorem 5.1:

- If we set $\mathcal{A} = \infty$, then the model considered in Theorem 5.1 can be used to describe pulse-amplitude modulation on a *continuous-time* Poisson channel with constant dark current under an average input-power constraint and in the absence of a peak-power constraint. Theorem 5.1 shows that, as we let the pulse duration Δ tend to zero, capacity grows like $\frac{\mathbb{E}}{\Delta} \log \frac{1}{\Delta}$, where $\frac{\mathbb{E}}{\Delta}$ is the continuous-time average power³ which remains constant as \mathbb{E} tends to zero proportionally with Δ .
- Note that (5.5) does not depend on the peak input power \mathcal{A} . In fact, as the proof shows, (5.5) can be achieved using on-off signaling, where the “on” signal is chosen small but constant. In the continuous-time picture, this choice corresponds to the peak power growing like the constant divided by Δ . As Δ tends to zero, the maximum continuous-time input power thus tends to infinity. Note that, to achieve unbounded capacity—in our case $\frac{\mathbb{E}}{\Delta} \log \frac{1}{\Delta}$ —on the continuous-time Poisson channel, it is necessary to use inputs that tend to infinity since peak-limited continuous-time Poisson channels have bounded capacities [47] (see also Chapter 3).
- It is somewhat surprising that the RHS of (5.5) does not depend on the dark current. In particular, it does not depend on whether the dark current is zero or not. Intuitively this is because, when \mathbb{E} is small, our “on” signal, which we hold constant, dominates the dark-current floor.
- The bound (5.8) can also be derived by noting that the capacity of the Poisson channel with zero dark current under only an average-power constraint is upper-bounded by the capacity of the pure-loss

³To be precise, $\frac{\mathbb{E}}{\Delta}$ is $\eta P / \hbar \omega$ where P is the “power” in physics. See Footnote 1.

bosonic channel, and by using the explicit formula [14]

$$C_{\text{bosonic}}(\mathbb{E}) = (1 + \mathbb{E}) \log(1 + \mathbb{E}) - \mathbb{E} \log \mathbb{E} \quad (5.9)$$

of the latter.

- Because the pure-loss bosonic channel with coherent input states and direct detection reduces to a Poisson channel, the formula (5.5) and the achievability of its LHS using binary signaling combine with (5.9) to show that the asymptotic (quantum-receiver) capacity of the pure-loss bosonic channel is achievable with binary modulation (on-off signaling) and direct detection.
- To see how well capacity is approximated by its asymptotic expression, we compare this expression with nonasymptotic upper and lower bounds in Figure 5.1. The upper bound is computed using (5.52) and (5.85); the lower bounds are computed using (5.16), (5.20) and (5.27). It can be seen that this approximation is useful for $c = 0.1$, and for $c = 1$ and $\mathbb{E} < 10^{-3}$. For $c = 1$ and $\mathbb{E} > 10^{-3}$ our choice of the input distribution becomes highly suboptimal, which is why the lower bound deviates significantly from the capacity asymptote.

For our second case where the dark current is constant and does not scale with \mathbb{E} , the asymptotics depend critically on whether a peak-power constraint is present ($\mathbb{A} < \infty$) or not ($\mathbb{A} = \infty$):

Theorem 5.2. *For any $\lambda > 0$,*

$$\lim_{\mathbb{E} \downarrow 0} \frac{C(\lambda, \mathbb{E}, \mathbb{A})}{\mathbb{E}} = \left(1 + \frac{\lambda}{\mathbb{A}}\right) \log \left(1 + \frac{\mathbb{A}}{\lambda}\right) - 1, \quad \mathbb{A} < \infty, \quad (5.10)$$

and

$$\frac{1}{2} \leq \lim_{\mathbb{E} \downarrow 0} \frac{C(\lambda, \mathbb{E}, \infty)}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} \leq \overline{\lim}_{\mathbb{E} \downarrow 0} \frac{C(\lambda, \mathbb{E}, \infty)}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} \leq 2. \quad (5.11)$$

The proof of (5.10) is a simple application of the formula for the capacity per unit cost [44, Theorem 2]. We shall prove the lower bound in (5.11) in Section 5.3.2 and the upper bound in Section 5.4.2.

Remarks about Theorem 5.2:

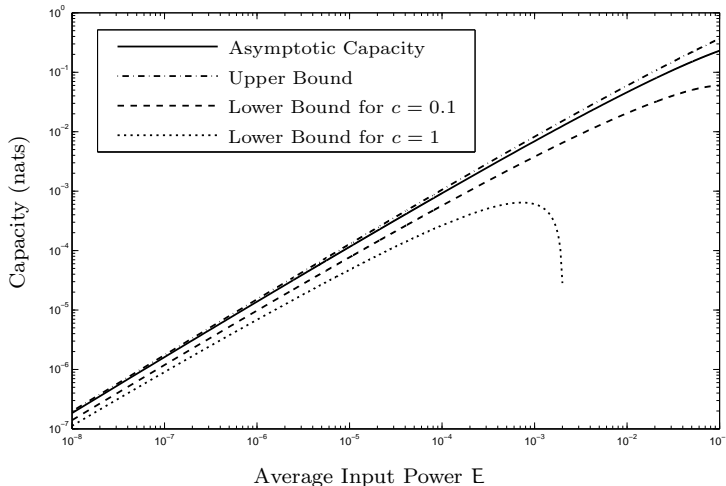


Figure 5.1: Comparison of capacity asymptote $E \log \frac{1}{E}$ with nonasymptotic upper and lower bounds for dark current proportional to E .

- In contrast to Theorem 5.1, here the capacity asymptote depends heavily on the peak input power A . In particular, it is linear in E if A is finite, and it is proportional to $E \log \log \frac{1}{E}$ if A is infinite.
- As the proof shows, both (5.10) and (5.11) can be achieved with on-off signaling. In the case of (5.10), the “on” signal is equal to A ; while in the case of (5.11), the “on” signal tends to infinity as E tends to zero. These signaling schemes are in the same spirit as the one that achieves (5.5) in the sense that the “on” signal should be large compared to the dark-current floor.
- We compare the asymptotic and nonasymptotic lower bounds in Figure 5.2. The nonasymptotic lower bounds are computed using (5.43). Interestingly, for most realistic values of E , this nonasymptotic lower bound for $\lambda = 1$ is better than that for $\lambda = 0.1$. This is because, when $\lambda = 0.1$, our choice of the input distribution is good only for extremely small input powers. To get a sense of how good the asymptotic approximation is, we can always lower-bound the capacity when $\lambda = 0.1$ by the lower bound for $\lambda = 1$, which is rather close to the asymptotic lower bound whenever $E < 10^{-3}$.

Our nonasymptotic upper bounds are difficult to compute and are therefore not included in this figure.

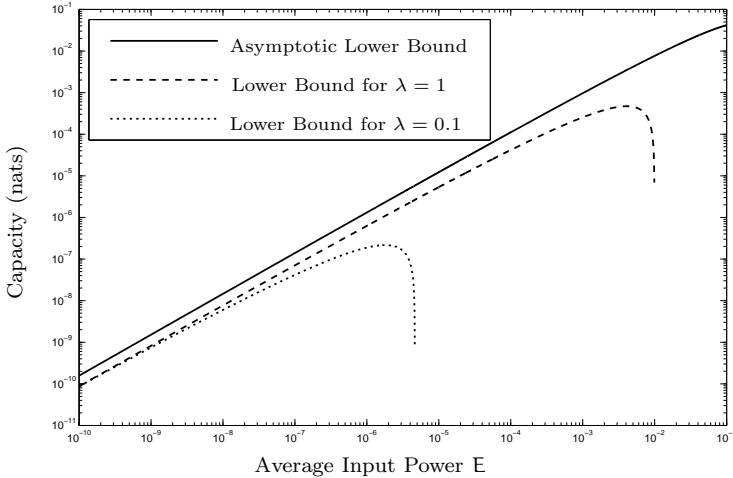


Figure 5.2: Comparison of asymptotic lower bound $\frac{1}{2}E \log \log \frac{1}{E}$ with nonasymptotic lower bound for constant nonzero dark current.

5.3 Lower Bounds

The lower bounds in this section are obtained by choosing binary input distributions and then studying the corresponding mutual informations. We denote by Q^b the binary distribution

$$X = \begin{cases} 0, & \text{w.p. } (1-p) \\ \zeta, & \text{w.p. } p \end{cases} \quad (5.12)$$

where $\zeta > 0$ and $p \in (0, 1)$. If we choose the parameters ζ and p so that constraints (5.2) and (5.3) are satisfied, then

$$C(\lambda, E, A) \geq I(Q^b, W). \quad (5.13)$$

5.3.1 Dark Current Proportional to E

We next derive (5.7). To this end, we write out the mutual information $I(Q^b, W)$ for the input distribution Q^b of (5.12) as

$$I(Q^b, W) = H(Y) - H(Y|X) \quad (5.14)$$

$$\begin{aligned} &= - \sum_{y=0}^{\infty} ((1-p)\text{Pois}_{\lambda}(y) + p\text{Pois}_{\lambda+\zeta}(y)) \\ &\quad \cdot \log((1-p)\text{Pois}_{\lambda}(y) + p\text{Pois}_{\lambda+\zeta}(y)) \\ &\quad + (1-p) \sum_{y=0}^{\infty} \text{Pois}_{\lambda}(y) \log \text{Pois}_{\lambda}(y) \\ &\quad + p \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \log \text{Pois}_{\lambda+\zeta}(y) \end{aligned} \quad (5.15)$$

$$= I_0(\lambda, \zeta, p) + I_1(\lambda, \zeta, p) \quad (5.16)$$

where in the last equality we defined

$$\begin{aligned} I_0(\lambda, \zeta, p) &\triangleq -((1-p)e^{-\lambda} + pe^{-(\lambda+\zeta)}) \\ &\quad \cdot \log((1-p)e^{-\lambda} + pe^{-(\lambda+\zeta)}) \\ &\quad - (1-p)\lambda e^{-\lambda} - p(\lambda + \zeta)e^{-(\lambda+\zeta)} \end{aligned} \quad (5.17)$$

$$\begin{aligned} I_1(\lambda, \zeta, p) &\triangleq - \sum_{y=1}^{\infty} ((1-p)\text{Pois}_{\lambda}(y) + p\text{Pois}_{\lambda+\zeta}(y)) \\ &\quad \cdot \log((1-p)\text{Pois}_{\lambda}(y) + p\text{Pois}_{\lambda+\zeta}(y)) \\ &\quad + (1-p) \sum_{y=1}^{\infty} \text{Pois}_{\lambda}(y) \log \text{Pois}_{\lambda}(y) \\ &\quad + p \sum_{y=1}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \log \text{Pois}_{\lambda+\zeta}(y). \end{aligned} \quad (5.18)$$

Note that in the above decomposition we took out the terms corresponding to $y = 0$ in all three summations to form $I_0(\lambda, \zeta, p)$ and collected the remaining terms in $I_1(\lambda, \zeta, p)$.

We lower-bound $I_0(\lambda, \zeta, p)$ as

$$I_0(\lambda, \zeta, p) \geq 0 - (1-p)\lambda e^{-\lambda} - p(\lambda + \zeta)e^{-(\lambda+\zeta)} \quad (5.19)$$

$$\geq -\lambda - p(\lambda + \zeta). \quad (5.20)$$

We lower-bound $I_1(\lambda, \zeta, p)$ as

$$\begin{aligned}
 & I_1(\lambda, \zeta, p) \\
 &= - \sum_{y=1}^{\infty} ((1-p)\text{Pois}_{\lambda}(y) + p\text{Pois}_{\lambda+\zeta}(y)) \\
 &\quad \cdot \log \left((1-p) \frac{\text{Pois}_{\lambda}(y)}{\text{Pois}_{\lambda+\zeta}(y)} + p \right) \\
 &\quad + (1-p) \sum_{y=1}^{\infty} \text{Pois}_{\lambda}(y) \log \frac{\text{Pois}_{\lambda}(y)}{\text{Pois}_{\lambda+\zeta}(y)} \tag{5.21}
 \end{aligned}$$

$$\begin{aligned}
 &= - \sum_{y=1}^{\infty} ((1-p)\text{Pois}_{\lambda}(y) + p\text{Pois}_{\lambda+\zeta}(y)) \\
 &\quad \cdot \left(\log p + \log \left(1 + \frac{1-p}{p} \frac{\text{Pois}_{\lambda}(y)}{\text{Pois}_{\lambda+\zeta}(y)} \right) \right) \\
 &\quad + (1-p) \sum_{y=1}^{\infty} \text{Pois}_{\lambda}(y) \log \frac{e^{-\lambda} \frac{\lambda^y}{y!}}{e^{-(\lambda+\zeta)} \frac{(\lambda+\zeta)^y}{y!}} \tag{5.22}
 \end{aligned}$$

$$\begin{aligned}
 &= - \sum_{y=1}^{\infty} ((1-p)\text{Pois}_{\lambda}(y) + p\text{Pois}_{\lambda+\zeta}(y)) \\
 &\quad \cdot \left(\underbrace{\log p + \log \left(1 + \frac{1-p}{p} \frac{\text{Pois}_{\lambda}(y)}{\text{Pois}_{\lambda+\zeta}(y)} \right)}_{\leq \frac{1-p}{p} \frac{\text{Pois}_{\lambda}(y)}{\text{Pois}_{\lambda+\zeta}(y)}} \right) \\
 &\quad + (1-p)\zeta \underbrace{\sum_{y=1}^{\infty} \text{Pois}_{\lambda}(y)}_{=1-e^{-\lambda}} \\
 &\quad + (1-p) \log \frac{\lambda}{\lambda+\zeta} \underbrace{\sum_{y=1}^{\infty} \text{Pois}_{\lambda}(y)y}_{=\lambda} \tag{5.23}
 \end{aligned}$$

$$\begin{aligned}
 &\geq - \underbrace{\sum_{y=1}^{\infty} ((1-p)\text{Pois}_{\lambda}(y) + p\text{Pois}_{\lambda+\zeta}(y)) \log p}_{=(1-p)(1-e^{-\lambda})+p(1-e^{-(\lambda+\zeta)})} \\
 &\quad - \sum_{y=1}^{\infty} ((1-p)\text{Pois}_{\lambda}(y) + p\text{Pois}_{\lambda+\zeta}(y)) \frac{1-p}{p} \frac{\text{Pois}_{\lambda}(y)}{\text{Pois}_{\lambda+\zeta}(y)}
 \end{aligned}$$

$$\begin{aligned}
& + (1-p)(1-e^{-\lambda})\zeta - (1-p)\lambda \log \left(1 + \frac{\zeta}{\lambda}\right) \tag{5.24} \\
& = \left((1-p)(1-e^{-\lambda}) + p(1-e^{-(\lambda+\zeta)}) \right) \log \frac{1}{p} \\
& \quad - \frac{(1-p)^2}{p} \sum_{y=1}^{\infty} \underbrace{\frac{(\text{Pois}_{\lambda}(y))^2}{\text{Pois}_{\lambda+\zeta}(y)}}_{=e^{\frac{\zeta^2}{\lambda+\zeta}} \text{Pois}_{\frac{\lambda^2}{\lambda+\zeta}}(y)} - (1-p) \underbrace{\sum_{y=1}^{\infty} \text{Pois}_{\lambda}(y)}_{=1-e^{-\lambda}}
\end{aligned}$$

$$\begin{aligned}
& + (1-p)(1-e^{-\lambda})\zeta - (1-p)\lambda \log \left(1 + \frac{\zeta}{\lambda}\right) \tag{5.25} \\
& = \left((1-p)(1-e^{-\lambda}) + p(1-\underbrace{e^{-(\lambda+\zeta)}}_{\leq e^{-\zeta}}) \right) \log \frac{1}{p} \\
& \quad - \underbrace{\frac{(1-p)^2}{p}}_{\leq \frac{1}{p}} \underbrace{e^{\frac{\zeta^2}{\lambda+\zeta}}}_{\leq e^{\zeta}} \underbrace{\left(1 - e^{-\frac{\lambda^2}{\lambda+\zeta}}\right)}_{\leq \frac{\lambda^2}{\lambda+\zeta} \leq \frac{\lambda^2}{\zeta}} - \underbrace{(1-p)}_{\leq 1} \underbrace{(1-e^{-\lambda})}_{\leq \lambda} \\
& \quad - \underbrace{(1-p)}_{\leq 1} \underbrace{(1-e^{-\lambda})\zeta}_{\leq \lambda} - (1-p)\lambda \log \left(1 + \frac{\zeta}{\lambda}\right) \tag{5.26}
\end{aligned}$$

$$\begin{aligned}
& \geq (1-p)(1-e^{-\lambda}) \log \frac{1}{p} + p(1-e^{-\zeta}) \log \frac{1}{p} \\
& \quad - \frac{1}{p} \frac{\lambda^2}{\zeta} e^{\zeta} - \lambda - \lambda\zeta - (1-p)\lambda \log \left(1 + \frac{\zeta}{\lambda}\right). \tag{5.27}
\end{aligned}$$

Choose any $\zeta \in (0, \mathbf{A}]$ and, for small enough \mathbf{E} , let $p = \mathbf{E}/\zeta$. Then the distribution (5.12) satisfies both constraints (5.2) and (5.3). Let $\lambda = c\mathbf{E}$. Using (5.20) we can bound the asymptote of $I_0(\lambda, \zeta, p)$ as

$$\lim_{\mathbf{E} \downarrow 0} \frac{I_0\left(c\mathbf{E}, \zeta, \frac{\mathbf{E}}{\zeta}\right)}{\mathbf{E} \log \frac{1}{\mathbf{E}}} \geq -\lim_{\mathbf{E} \downarrow 0} \frac{c\mathbf{E}}{\mathbf{E} \log \frac{1}{\mathbf{E}}} - \lim_{\mathbf{E} \downarrow 0} \frac{\frac{\mathbf{E}}{\zeta}(c\mathbf{E} + \zeta)}{\mathbf{E} \log \frac{1}{\mathbf{E}}} = 0. \tag{5.28}$$

Similarly, using (5.27) we can bound the asymptote of $I_1(\lambda, \zeta, p)$ as

$$\begin{aligned}
& \lim_{\mathbf{E} \downarrow 0} \frac{I_1\left(c\mathbf{E}, \zeta, \frac{\mathbf{E}}{\zeta}\right)}{\mathbf{E} \log \frac{1}{\mathbf{E}}} \\
& \geq \lim_{\mathbf{E} \downarrow 0} \frac{\left(1 - \frac{\mathbf{E}}{\zeta}\right) (1 - e^{c\mathbf{E}}) \log \frac{\zeta}{\mathbf{E}}}{\mathbf{E} \log \frac{1}{\mathbf{E}}} + \lim_{\mathbf{E} \downarrow 0} \frac{\frac{\mathbf{E}}{\zeta} (1 - e^{-\zeta}) \log \frac{\zeta}{\mathbf{E}}}{\mathbf{E} \log \frac{1}{\mathbf{E}}}
\end{aligned}$$

$$\begin{aligned}
& - \lim_{\mathbb{E} \downarrow 0} \frac{\frac{\zeta}{\mathbb{E}} \frac{c^2 \mathbb{E}^2}{\zeta} e^\zeta}{\mathbb{E} \log \frac{1}{\mathbb{E}}} - \lim_{\mathbb{E} \downarrow 0} \frac{c \mathbb{E}}{\mathbb{E} \log \frac{1}{\mathbb{E}}} - \lim_{\mathbb{E} \downarrow 0} \frac{c \mathbb{E} \zeta}{\mathbb{E} \log \frac{1}{\mathbb{E}}} \\
& - \lim_{\mathbb{E} \downarrow 0} \frac{\left(1 - \frac{\mathbb{E}}{\zeta}\right) c \mathbb{E} \log \left(1 + \frac{\zeta}{c \mathbb{E}}\right)}{\mathbb{E} \log \frac{1}{\mathbb{E}}} \tag{5.29}
\end{aligned}$$

$$= c + \frac{1 - e^{-\zeta}}{\zeta} - 0 - 0 - 0 - c \tag{5.30}$$

$$= \frac{1 - e^{-\zeta}}{\zeta}. \tag{5.31}$$

Combining (5.13), (5.16), (5.28), and (5.31) we obtain

$$\lim_{\mathbb{E} \downarrow 0} \frac{C(c\mathbb{E}, \mathbb{E}, \mathbf{A})}{\mathbb{E} \log \frac{1}{\mathbb{E}}} \geq \frac{1 - e^{-\zeta}}{\zeta}, \quad \text{for all } \zeta \in (0, \mathbf{A}]. \tag{5.32}$$

We can make the RHS of (5.32) arbitrarily close to 1 by choosing arbitrarily small ζ . Thus we obtain (5.7).

5.3.2 Constant Nonzero Dark Current

We next prove the lower bound in (5.11). To this end, we lower-bound the mutual information $I(Q^b, W)$ for the input distribution (5.12) as follows:

$$\begin{aligned}
& I(Q^b, W) \\
& = H(Y) - H(Y|X) \tag{5.33}
\end{aligned}$$

$$\begin{aligned}
& = - \sum_{y=0}^{\infty} \left((1-p) \text{Pois}_\lambda(y) + p \text{Pois}_{\lambda+\zeta}(y) \right) \\
& \quad \cdot \log \left((1-p) \text{Pois}_\lambda(y) + p \text{Pois}_{\lambda+\zeta}(y) \right) \\
& \quad + (1-p) \sum_{y=0}^{\infty} \text{Pois}_\lambda(y) \log \text{Pois}_\lambda(y) \\
& \quad + p \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \log \text{Pois}_{\lambda+\zeta}(y) \tag{5.34} \\
& = - p \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \log \left((1-p) \frac{\text{Pois}_\lambda(y)}{\text{Pois}_{\lambda+\zeta}(y)} + p \right)
\end{aligned}$$

$$-(1-p) \sum_{y=0}^{\infty} \text{Pois}_{\lambda}(y) \log \left((1-p) + p \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} \right) \quad (5.35)$$

$$\begin{aligned} &= -p \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \\ &\quad \cdot \left(\log \frac{\text{Pois}_{\lambda}(y)}{\text{Pois}_{\lambda+\zeta}(y)} + \log \left((1-p) + p \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} \right) \right) \\ &\quad - (1-p) \sum_{y=0}^{\infty} \text{Pois}_{\lambda}(y) \log \left((1-p) + p \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} \right) \quad (5.36) \end{aligned}$$

$$\begin{aligned} &= p \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \log \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} - \sum_{y=0}^{\infty} \underbrace{\left((1-p) \text{Pois}_{\lambda}(y) \right)}_{\geq 0} \\ &\quad + \underbrace{p \text{Pois}_{\lambda+\zeta}(y)}_{\geq 0} \underbrace{\log \left((1-p) + p \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} \right)}_{\leq \log \left(1 + p \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} \right) \leq p \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)}} \quad (5.37) \end{aligned}$$

$$\begin{aligned} &\geq p \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \log \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} \\ &\quad - \sum_{y=0}^{\infty} \left((1-p) \text{Pois}_{\lambda}(y) + p \text{Pois}_{\lambda+\zeta}(y) \right) p \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} \quad (5.38) \end{aligned}$$

$$\begin{aligned} &= p \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \log \left(\frac{e^{-(\zeta+\lambda)} \frac{(\zeta+\lambda)^y}{y!}}{e^{-\lambda} \frac{\lambda^y}{y!}} \right) \\ &\quad - (1-p)p \sum_{y=0}^{\infty} \text{Pois}_{\lambda}(y) \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} \\ &\quad - p^2 \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \frac{\text{Pois}_{\lambda+\zeta}(y)}{\text{Pois}_{\lambda}(y)} \quad (5.39) \end{aligned}$$

$$\begin{aligned} &= p \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \log \left(e^{-\zeta} \left(1 + \frac{\zeta}{\lambda} \right)^y \right) \\ &\quad - (1-p)p \underbrace{\sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y)}_{=1} \end{aligned}$$

$$-p^2 \sum_{y=0}^{\infty} \frac{\left(e^{-(\zeta+\lambda)} \frac{(\zeta+\lambda)^y}{y!} \right)^2}{e^{-\lambda} \frac{\lambda^y}{y!}} \quad (5.40)$$

$$= p \sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y) \left(-\zeta + y \log \left(1 + \frac{\zeta}{\lambda} \right) \right) - (1-p)p$$

$$- p^2 \underbrace{\left(\sum_{y=0}^{\infty} e^{-(\lambda+2\zeta)} \frac{\left(\lambda + 2\zeta + \frac{\zeta^2}{\lambda} \right)^y}{y!} e^{-\frac{\zeta^2}{\lambda}} \right)}_{=\sum_{y=0}^{\infty} \text{Pois}_{\lambda+2\zeta+\frac{\zeta^2}{\lambda}}(y)=1} e^{\frac{\zeta^2}{\lambda}} \quad (5.41)$$

$$= -p\zeta \underbrace{\sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y)}_{=1} + p \underbrace{\sum_{y=0}^{\infty} \text{Pois}_{\lambda+\zeta}(y)y}_{=(\zeta+\lambda)} \log \left(1 + \frac{\zeta}{\lambda} \right)$$

$$- p + p^2 - p^2 e^{\frac{\zeta^2}{\lambda}} \quad (5.42)$$

$$= p(\zeta + \lambda) \log \left(1 + \frac{\zeta}{\lambda} \right) - p\zeta - p - p^2 \left(e^{\frac{\zeta^2}{\lambda}} - 1 \right). \quad (5.43)$$

For small enough \mathbb{E} , we choose

$$\zeta = \sqrt{\lambda \log \frac{1}{\mathbb{E}}} \quad (5.44)$$

and

$$p = \frac{\mathbb{E}}{\zeta} = \frac{\mathbb{E}}{\sqrt{\lambda \log \frac{1}{\mathbb{E}}}}. \quad (5.45)$$

From (5.13) and (5.43) we then obtain

$$\lim_{\mathbb{E} \downarrow 0} \frac{C(\lambda, \mathbb{E}, \infty)}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}}$$

$$\geq \lim_{\mathbb{E} \downarrow 0} \frac{\frac{\mathbb{E}}{\sqrt{\lambda \log \frac{1}{\mathbb{E}}}} \left(\sqrt{\lambda \log \frac{1}{\mathbb{E}}} + \lambda \right) \log \left(1 + \frac{\sqrt{\lambda \log \frac{1}{\mathbb{E}}}}{\lambda} \right)}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}}$$

$$+ \lim_{\mathbb{E} \downarrow 0} \frac{-\mathbb{E} - \frac{\mathbb{E}}{\sqrt{\lambda \log \frac{1}{\mathbb{E}}}} - \left(\frac{\mathbb{E}}{\sqrt{\lambda \log \frac{1}{\mathbb{E}}}} \right)^2 \left(e^{\frac{\lambda \log \frac{1}{\mathbb{E}}}{\lambda}} - 1 \right)}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} \quad (5.46)$$

$$= \frac{1}{2} + 0 \quad (5.47)$$

$$= \frac{1}{2}. \quad (5.48)$$

This establishes the lower bound in (5.11).

5.4 Upper Bounds

In this section we prove the upper bounds on the asymptotic capacities of the Poisson channel. We shall use the duality bound [25] which states that, for any distribution $R(\cdot)$ on the output, the channel capacity satisfies

$$C \leq \sup \mathbb{E} [D(W(\cdot|X) \| R(\cdot))], \quad (5.49)$$

where the supremum is taken over all allowed input distributions.

5.4.1 Zero Dark Current

We next prove (5.8). To this end, as in [26], we shall introduce in Section 5.4.1.1 the *Poisson channel with continuous output*. This channel is equivalent to our channel but its output alphabet is not the nonnegative integers but the nonnegative reals. We shall then prove a lemma in Section 5.4.1.2 before finally proving (5.8) in Section 5.4.1.3.

5.4.1.1 Poisson Channel with Continuous Output

We introduce the Poisson channel with continuous output whose dark current is equal to zero. Its input x is the same as that of the original Poisson channel, and its output $\tilde{Y} \in \mathbb{R}_0^+$ is

$$\tilde{Y} \triangleq Y + U \quad (5.50)$$

where Y is the output of the original Poisson channel with zero dark current, and the random variable U is independent of (X, Y) and uniformly distributed on the interval $[0, 1)$. Then \tilde{Y} is a continuous random variable whose conditional density $\tilde{w}(\tilde{y}|x)$ given $X = x$ is

$$\tilde{w}(\tilde{y}|x) = W(\lfloor \tilde{y} \rfloor | x) = \text{Pois}_x(\lfloor \tilde{y} \rfloor) \quad (5.51)$$

where $\lfloor a \rfloor$ denotes the largest integer not exceeding a , and the second equality follows because $\lambda = 0$. Note that $W(\cdot|x)$ is a probability mass function on \mathbb{Z}^+ whereas $\tilde{w}(\cdot|x)$ is a density on \mathbb{R}^+ .

Denoting the capacity of the channel $\tilde{w}(\cdot|\cdot)$ under constraints (5.2) and (5.3) by $\tilde{C}(0, E, A)$,

$$C(0, E, A) = \tilde{C}(0, E, A), \quad E > 0, A \in (0, \infty] \quad (5.52)$$

because Y can be computed from \tilde{Y} [26, Lemma 17].

5.4.1.2 A Lemma

The following lemma lower-bounds the differential entropy $h(\tilde{Y}|X = x)$ of \tilde{Y} conditional on $X = x$.

Lemma 5.3. *Let \tilde{Y} be defined as in (5.50) and \tilde{w} be given by (5.51), then*

$$h(\tilde{Y}|X = x) \geq \frac{1 - e^{-x}}{2} \log 2\pi + \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \log y. \quad (5.53)$$

Proof. By (5.50) we have as in [26, Lemma 17]

$$h(\tilde{Y}|X = x) = H(Y|X = x). \quad (5.54)$$

The RHS of (5.54) can be bounded as

$$\begin{aligned} H(Y|X = x) &= - \sum_{y=0}^{\infty} W(y|x) \log W(y|x) \end{aligned} \quad (5.55)$$

$$= - \sum_{y=0}^{\infty} \text{Pois}_x(y) \log \left(e^{-x} \frac{x^y}{y!} \right) \quad (5.56)$$

$$= x \underbrace{\sum_{y=0}^{\infty} \text{Pois}_x(y)}_{=1} - \log x \underbrace{\sum_{y=0}^{\infty} \text{Pois}_x(y) y}_{=x} + \sum_{y=0}^{\infty} \text{Pois}_x(y) \log(y!) \quad (5.57)$$

$$= x - x \log x + \sum_{y=1}^{\infty} \text{Pois}_x(y) \log(y!) \quad (5.58)$$

$$\geq x - x \log x + \sum_{y=1}^{\infty} \text{Pois}_x(y) \log \left(\sqrt{2\pi y} \left(\frac{y}{e} \right)^y \right) \quad (5.59)$$

$$\begin{aligned} &= x - x \log x + \frac{1}{2} \log 2\pi \underbrace{\sum_{y=1}^{\infty} \text{Pois}_x(y)}_{=1 - \text{Pois}_x(0) = 1 - e^{-x}} \\ &\quad + \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \log y + \sum_{y=0}^{\infty} \text{Pois}_x(y) y \log y - \underbrace{\sum_{y=0}^{\infty} \text{Pois}_x(y) y}_{=x} \end{aligned} \quad (5.60)$$

$$\begin{aligned} &= -x \log x + \frac{1 - e^{-x}}{2} \log 2\pi \\ &\quad + \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \log y + \sum_{y=0}^{\infty} \text{Pois}_x(y) \underbrace{y \log y}_{\geq x \log x + (1 + \log x)(y-x) - x} \end{aligned} \quad (5.61)$$

$$\begin{aligned} &\geq -x \log x + \frac{1 - e^{-x}}{2} \log 2\pi + \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \log y \\ &\quad + \sum_{y=0}^{\infty} \text{Pois}_x(y) (x \log x + (1 + \log x)(y - x)) \end{aligned} \quad (5.62)$$

$$\begin{aligned} &= -x \log x + \frac{1 - e^{-x}}{2} \log 2\pi + \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \log y \\ &\quad + x \log x \underbrace{\sum_{y=0}^{\infty} \text{Pois}_x(y)}_{=1} + (1 + \log x) \underbrace{\sum_{y=0}^{\infty} \text{Pois}_x(y)(y - x)}_{=0} \end{aligned} \quad (5.63)$$

$$= \frac{1 - e^{-x}}{2} \log 2\pi + \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \log y. \quad (5.64)$$

Here, (5.59) follows by Stirling's Bound [11]

$$n! \geq \sqrt{2\pi n} \left(\frac{n}{e} \right)^n e^{\frac{1}{12n+1}} \geq \sqrt{2\pi n} \left(\frac{n}{e} \right)^n, \quad n \in \mathbb{Z}_0^+ \quad (5.65)$$

and (5.62) follows by bounding $y \log y$ by its Taylor expansion at $y = x$. Combining (5.54) and (5.64) proves the lemma. \square

5.4.1.3 Proof of (5.8)

According to (5.52), to prove (5.8), we need to prove

$$\overline{\lim}_{\mathbb{E} \downarrow 0} \frac{\tilde{C}(0, \mathbb{E}, \infty)}{\mathbb{E} \log \frac{1}{\mathbb{E}}} \leq 1. \quad (5.66)$$

To prove (5.66) using the duality bound (5.49), we choose the distribution $\tilde{R}(\cdot)$ on \tilde{Y} to be of density

$$f_{\tilde{R}}(\tilde{y}) = \begin{cases} 1 - p, & 0 \leq \tilde{y} < 1 \\ p \cdot \frac{\tilde{y}^{\nu-1} e^{-\frac{\tilde{y}}{\beta}}}{\beta^\nu \Gamma(\nu, \frac{1}{\beta})}, & \tilde{y} \geq 1 \end{cases} \quad (5.67)$$

where $\beta > 0$ is arbitrary (e.g. $\beta = 1$), whereas $\nu \in (0, 1]$ and $p \in (0, 1)$ will be specified later, and $\Gamma(\cdot, \cdot)$ denotes the Incomplete Gamma Function

$$\Gamma(a, \xi) = \int_{\xi}^{\infty} t^{a-1} e^{-t} dt, \quad a, \xi \geq 0. \quad (5.68)$$

We next apply the duality bound to upper-bound the capacity of the channel $\tilde{w}(\cdot|x)$ using the above output distribution:

$$\tilde{C}(0, \lambda, \infty) \leq \sup_{\mathbb{E}[X] \leq \mathbb{E}} D(\tilde{w}(\cdot|x) || f_{\tilde{R}}). \quad (5.69)$$

We bound $D(\tilde{w}(\cdot|x) || f_{\tilde{R}}(\cdot))$ as follows:

$$\begin{aligned} & D(\tilde{w}(\cdot|x) || f_{\tilde{R}}(\cdot)) \\ &= \int_0^{\infty} \tilde{w}(\tilde{y}|x) \log \frac{\tilde{w}(\tilde{y}|x)}{f_{\tilde{R}}(\tilde{y})} d\tilde{y} \end{aligned} \quad (5.70)$$

$$\begin{aligned} &= -h(\tilde{Y}|X = x) + \int_0^1 \tilde{w}(\tilde{y}|x) \log \frac{1}{f_{\tilde{R}}(\tilde{y})} d\tilde{y} \\ &+ \int_1^{\infty} \tilde{w}(\tilde{y}|x) \log \frac{1}{f_{\tilde{R}}(\tilde{y})} d\tilde{y} \end{aligned} \quad (5.71)$$

$$\begin{aligned} &= -h(\tilde{Y}|X = x) + \int_0^1 \tilde{w}(\tilde{y}|x) \log \frac{1}{1-p} d\tilde{y} \\ &+ \int_1^{\infty} \tilde{w}(\tilde{y}|x) \log \frac{\beta^\nu \Gamma\left(\nu, \frac{1}{\beta}\right)}{p \tilde{y}^{\nu-1} e^{-\frac{\tilde{y}}{\beta}}} d\tilde{y} \end{aligned} \quad (5.72)$$

$$\begin{aligned}
&= -h(\tilde{Y}|X = x) + \log \frac{1}{1-p} \underbrace{\int_0^1 \tilde{w}(\tilde{y}|x) d\tilde{y}}_{=e^{-x}} \\
&\quad + \log \frac{\beta^\nu \Gamma\left(\nu, \frac{1}{\beta}\right)}{p} \underbrace{\int_1^\infty \tilde{w}(\tilde{y}|x) d\tilde{y}}_{=1-e^{-x}} \\
&\quad + (1-\nu) \int_1^\infty \tilde{w}(\tilde{y}|x) \log \tilde{y} d\tilde{y} \\
&\quad + \frac{1}{\beta} \underbrace{\int_1^\infty \tilde{w}(\tilde{y}|x) \tilde{y} d\tilde{y}}_{=\mathbb{E}[\tilde{Y}|X=x] - \int_0^1 \tilde{w}(\tilde{y}|x) \tilde{y} d\tilde{y}} \tag{5.73}
\end{aligned}$$

$$\begin{aligned}
&= -h(\tilde{Y}|X = x) + e^{-x} \log \frac{1}{1-p} \\
&\quad + (1-e^{-x}) \log \frac{\beta^\nu \Gamma\left(\nu, \frac{1}{\beta}\right)}{p} \\
&\quad + (1-\nu) \int_1^\infty \tilde{w}(\tilde{y}|x) \log \tilde{y} d\tilde{y} \\
&\quad + \frac{1}{\beta} \left(\underbrace{\mathbb{E}[\tilde{Y}|X = x]}_{=x+\frac{1}{2}} - \underbrace{\int_0^1 \tilde{w}(\tilde{y}|x) \tilde{y} d\tilde{y}}_{=e^{-x}} \right) \tag{5.74}
\end{aligned}$$

$$\begin{aligned}
&= -h(\tilde{Y}|X = x) + e^{-x} \log \frac{1}{1-p} \\
&\quad + (1-e^{-x}) \log \frac{\beta^\nu \Gamma\left(\nu, \frac{1}{\beta}\right)}{p} \\
&\quad + \underbrace{(1-\nu)}_{\geq 0} \underbrace{\int_1^\infty \tilde{w}(\tilde{y}|x) \log \tilde{y} d\tilde{y}}_{\leq \sum_{y=1}^\infty W(y|x) \log(y+1)} \\
&\quad + \frac{1}{\beta} \left(x + \frac{1}{2} - \frac{1}{2} e^{-x} \right) \tag{5.75} \\
&\leq -h(\tilde{Y}|X = x) + e^{-x} \log \frac{1}{1-p} \\
&\quad + (1-e^{-x}) \log \frac{\beta^\nu \Gamma\left(\nu, \frac{1}{\beta}\right)}{p}
\end{aligned}$$

$$\begin{aligned}
& + (1 - \nu) \sum_{y=1}^{\infty} \underbrace{W(y|x)}_{=\text{Pois}_x(y)} \log(y+1) \\
& + \frac{x}{\beta} + \frac{(1 - e^{-x})}{2\beta} \tag{5.76} \\
\leq & - \frac{(1 - e^{-x})}{2} \log 2\pi - \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \log y \\
& + e^{-x} \log \frac{1}{1-p} + (1 - e^{-x}) \log \frac{\beta^\nu \Gamma\left(\nu, \frac{1}{\beta}\right)}{p} \\
& + (1 - \nu) \sum_{y=1}^{\infty} \text{Pois}_x(y) \log(y+1) \\
& + \frac{x}{\beta} + \frac{(1 - e^{-x})}{2\beta} \tag{5.77}
\end{aligned}$$

where the last inequality follows from Lemma 5.3.

Substituting $\nu = \frac{1}{2}$ in (5.77) yields

$$\begin{aligned}
& D(\tilde{w}(\cdot|x) || f_{\tilde{R}}(\cdot)) \\
& \leq - \frac{(1 - e^{-x})}{2} \log 2\pi - \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \log y \\
& + e^{-x} \log \frac{1}{1-p} + (1 - e^{-x}) \log \frac{\beta^{\frac{1}{2}} \Gamma\left(\frac{1}{2}, \frac{1}{\beta}\right)}{p} \\
& + \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \log(y+1) \\
& + \frac{x}{\beta} + \frac{(1 - e^{-x})}{2\beta} \tag{5.78}
\end{aligned}$$

$$\begin{aligned}
& = - \frac{(1 - e^{-x})}{2} \log 2\pi + \frac{1}{2} \sum_{y=1}^{\infty} \text{Pois}_x(y) \underbrace{\log\left(1 + \frac{1}{y}\right)}_{\leq \log 2} \\
& + e^{-x} \log \frac{1}{1-p} + (1 - e^{-x}) \log \frac{\beta^{\frac{1}{2}} \Gamma\left(\frac{1}{2}, \frac{1}{\beta}\right)}{p} \\
& + \frac{x}{\beta} + \frac{(1 - e^{-x})}{2\beta} \tag{5.79}
\end{aligned}$$

$$\begin{aligned}
&\leq -\frac{(1-e^{-x})}{2} \log 2\pi + \frac{1}{2} \underbrace{\sum_{y=1}^{\infty} \text{Pois}_x(y)}_{=(1-e^{-x})} \log 2 \\
&\quad + e^{-x} \log \frac{1}{1-p} + (1-e^{-x}) \log \frac{\beta^{\frac{1}{2}} \Gamma\left(\frac{1}{2}, \frac{1}{\beta}\right)}{p} \\
&\quad + \frac{x}{\beta} + \frac{(1-e^{-x})}{2\beta} \tag{5.80}
\end{aligned}$$

$$\begin{aligned}
&= -\frac{(1-e^{-x})}{2} \log \pi + e^{-x} \log \frac{1}{1-p} \\
&\quad + (1-e^{-x}) \log \left(\beta^{\frac{1}{2}} \Gamma\left(\frac{1}{2}, \frac{1}{\beta}\right) \right) + (1-e^{-x}) \log \frac{1}{p} \\
&\quad + \frac{x}{\beta} + \frac{(1-e^{-x})}{2\beta} \tag{5.81}
\end{aligned}$$

$$\begin{aligned}
&= \underbrace{(1-e^{-x}) \log \frac{1}{p}}_{\leq x} + \underbrace{e^{-x} \log \frac{1}{1-p}}_{\leq 1} + \frac{x}{\beta} \\
&\quad + \underbrace{(1-e^{-x})}_{\leq x} \left(\frac{1}{2} \log \beta + \log \frac{\Gamma\left(\frac{1}{2}, \frac{1}{\beta}\right)}{\sqrt{\pi}} + \frac{1}{2\beta} \right) \tag{5.82}
\end{aligned}$$

$$\begin{aligned}
&\leq x \log \frac{1}{p} + \log \frac{1}{1-p} + \frac{x}{\beta} \\
&\quad + x \max \left\{ 0, \left(\frac{1}{2} \log \beta + \log \frac{\Gamma\left(\frac{1}{2}, \frac{1}{\beta}\right)}{\sqrt{\pi}} + \frac{1}{2\beta} \right) \right\}. \tag{5.83}
\end{aligned}$$

From (5.83) and (5.69), we get

$$\begin{aligned}
&\tilde{C}(0, \mathbb{E}, \infty) \\
&\leq \sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E} \left[X \log \frac{1}{p} + \log \frac{1}{1-p} + \frac{X}{\beta} \right. \\
&\quad \left. + X \max \left\{ 0, \left(\frac{1}{2} \log \beta + \log \frac{\Gamma\left(\frac{1}{2}, \frac{1}{\beta}\right)}{\sqrt{\pi}} + \frac{1}{2\beta} \right) \right\} \right] \tag{5.84} \\
&\leq \mathbb{E} \log \frac{1}{p} + \log \frac{1}{1-p} + \frac{\mathbb{E}}{\beta}
\end{aligned}$$

$$+ \mathbb{E} \max \left\{ 0, \left(\frac{1}{2} \log \beta + \log \frac{\Gamma\left(\frac{1}{2}, \frac{1}{\beta}\right)}{\sqrt{\pi}} + \frac{1}{2\beta} \right) \right\}. \quad (5.85)$$

Note that (5.85) holds for all $p \in (0, 1)$ and $\beta > 0$. Choosing $p = \frac{\mathbb{E}}{1+\mathbb{E}}$ in (5.85) and letting \mathbb{E} tend to zero yields (5.66) and hence concludes the proof of (5.8).

5.4.2 Constant Nonzero Dark Current

In this section we shall prove the upper bound in (5.11), namely

$$\overline{\lim}_{\mathbb{E} \downarrow 0} \frac{C(\lambda, \mathbb{E}, \infty)}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} \leq 2. \quad (5.86)$$

To this end, we shall prove two lemmas in Section 5.4.2.1, then derive a general upper bound on $C(\lambda, \mathbb{E}, \infty)$ in Section 5.4.2.2, and finally prove (5.86) in Section 5.4.2.3.

5.4.2.1 Lemmas

We next present two lemmas. The first lemma shows that the tail of a Poisson distribution of mean ξ behaves like $\left(\frac{\xi}{n}\right)^n$ for large n .

Lemma 5.4. *If Y is a mean- ξ Poisson random variable, then, for any $n \in \mathbb{Z}^+$,*

$$L_\xi(n) \leq \Pr[Y \geq n] \leq U_\xi(n), \quad (5.87)$$

where

$$L_\xi(n) \triangleq \exp \left(n - \xi + n \log \xi - n \log n - \frac{1}{12n} - \frac{1}{2} \log(2\pi n) \right) \quad (5.88)$$

$$U_\xi(n) \triangleq \begin{cases} \exp(n - \xi + n \log \xi - n \log n), & n > \xi \\ 1, & n \leq \xi. \end{cases} \quad (5.89)$$

Proof. To prove the lower bound, we observe that for every $n \in \mathbb{Z}^+$

$$\Pr[Y \geq n] \geq \Pr[Y = n] = e^{-\xi} \frac{\xi^n}{n!}. \quad (5.90)$$

Using Stirling's Bound

$$n! \leq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}} \quad (5.91)$$

we obtain from (5.90) that

$$\Pr[Y \geq n] \geq e^{-\xi} \xi^n \left(\sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}\right)^{-1} = L_\xi(n). \quad (5.92)$$

This establishes the lower bound in (5.87).

To prove the upper bound, we recall that the moment generating function of the Poisson distribution of mean ξ is

$$\mathbb{E}[e^{\theta Y}] = e^{\xi(e^\theta - 1)}, \quad \theta \in \mathbb{R}. \quad (5.93)$$

Consequently, by the Chernoff Bound,

$$\Pr[Y \geq n] \leq \frac{\mathbb{E}[e^{\theta Y}]}{e^{\theta n}} = \frac{e^{\xi(e^\theta - 1)}}{e^{\theta n}}, \quad \theta > 0. \quad (5.94)$$

When $n > \xi$, letting $\theta = \log\left(\frac{n}{\xi}\right)$ in (5.94) yields the upper bound in (5.87)

$$\Pr[Y \geq n] \leq \frac{e^{n-\xi}}{e^{n \log \frac{n}{\xi}}} = U_\xi(n), \quad n > \xi. \quad (5.95)$$

When $n \leq \xi$, the upper bound in (5.87) is trivial. \square

The second lemma is a simple property of convex functions.

Lemma 5.5. *For any convex function $f: [a, b] \rightarrow \mathbb{R}$*

$$\frac{f(b) - f(a)}{b - a} \geq \frac{f(c) - f(a)}{c - a}, \quad c \in (a, b). \quad (5.96)$$

Proof. Since $c \in (a, b)$, we can write $c = \alpha a + (1 - \alpha)b$ for some $\alpha \in (0, 1)$. The convexity of $f(\cdot)$ implies

$$f(c) \leq \alpha f(a) + (1 - \alpha)f(b). \quad (5.97)$$

Thus,

$$\frac{f(c) - f(a)}{c - a} \leq \frac{\alpha f(a) + (1 - \alpha)f(b) - f(a)}{(\alpha a + (1 - \alpha)b) - a} \quad (5.98)$$

$$= \frac{(1 - \alpha)(f(b) - f(a))}{(1 - \alpha)(b - a)} \quad (5.99)$$

$$= \frac{f(b) - f(a)}{b - a}. \quad (5.100)$$

□

5.4.2.2 An Upper Bound on $C(\lambda, E, \infty)$

We shall next apply (5.49) to upper-bound $C(\lambda, E, \infty)$. To this end, we choose $R(\cdot)$ to be

$$R(y) = \begin{cases} \text{Pois}_\lambda(y), & y \in \{0, 1, \dots, \eta - 1\} \\ \delta(1 - p)p^{y-\eta}, & y \in \{\eta, \eta + 1, \dots\} \end{cases} \quad (5.101)$$

where $\eta \in \mathbb{Z}^+$ and $p \in (0, 1)$ are constants that will be specified in Section 5.4.2.3, and δ is the normalizing factor

$$\delta \triangleq \Pr[Y \geq \eta | X = 0] = \sum_{y=\eta}^{\infty} \text{Pois}_\lambda(y). \quad (5.102)$$

In the following calculations we shall assume that η is large compared to λ .

To upper-bound the capacity using (5.49), we write $D(W(\cdot|x)||R(\cdot))$ as

$$\begin{aligned} D(W(\cdot|x)||R(\cdot)) &= \underbrace{\sum_{y=0}^{\eta-1} W(y|x) \log \frac{W(y|x)}{R(y)}}_{\triangleq D_1(x)} \\ &\quad + \underbrace{\sum_{y=\eta}^{\infty} W(y|x) \log \frac{W(y|x)}{R(y)}}_{\triangleq D_2(x)} \end{aligned} \quad (5.103)$$

and study $D_1(x)$ and $D_2(x)$ separately. Substituting (5.101) and the channel law (5.1) in $D_1(x)$ yields

$$\begin{aligned} D_1(x) &= \sum_{y=0}^{\eta-1} \text{Pois}_{\lambda+x}(y) \log \frac{e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!}}{e^{-\lambda} \frac{\lambda^y}{y!}} \\ &= -x \sum_{y=0}^{\eta-1} \text{Pois}_{\lambda+x}(y) + \log \left(1 + \frac{x}{\lambda} \right) \sum_{y=0}^{\eta-1} \text{Pois}_{\lambda+x}(y). \end{aligned} \quad (5.104)$$

The second term $D_2(x)$ can be upper-bounded as

$$D_2(x) = \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) \log \frac{e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!}}{\delta(1-p)p^{y-\eta}} \quad (5.105)$$

$$\begin{aligned} &= \left(-(\lambda+x) + \log \frac{1}{\delta} + \log \frac{1}{1-p} - \eta \log \frac{1}{p} \right) \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) \\ &\quad + \log \frac{\lambda+x}{p} \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)y - \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) \log(y!) \end{aligned} \quad (5.106)$$

$$\begin{aligned} &\leq \left(-(\lambda+x) + \log \frac{1}{\delta} + \log \frac{1}{1-p} - \eta \log \frac{1}{p} \right) \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) \\ &\quad + \log \frac{\lambda+x}{p} \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)y \\ &\quad - \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)(y \log y - y) \end{aligned} \quad (5.107)$$

where the inequality follows from Stirling's Bound

$$n! \geq \sqrt{2\pi n} \left(\frac{n}{e} \right)^n e^{\frac{1}{12n+1}} \geq \left(\frac{n}{e} \right)^n, \quad n \in \mathbb{Z}^+. \quad (5.108)$$

Substituting (5.104) and (5.107) in (5.103) we obtain

$$\begin{aligned} &D(W(\cdot|x)||R(\cdot)) \\ &\leq -x \sum_{y=0}^{\eta-1} \text{Pois}_{\lambda+x}(y) + \log \left(1 + \frac{x}{\lambda} \right) \sum_{y=0}^{\eta-1} \text{Pois}_{\lambda+x}(y)y \\ &\quad + \left(-(\lambda+x) + \log \frac{1}{\delta} + \log \frac{1}{1-p} - \eta \log \frac{1}{p} \right) \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) \end{aligned}$$

$$\begin{aligned}
& + \log \frac{\lambda + x}{p} \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)y \\
& - \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)(y \log y - y) \tag{5.109} \\
= & -x \underbrace{\left(\sum_{y=0}^{\eta-1} \text{Pois}_{\lambda+x}(y) + \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) \right)}_{\leq 0}
\end{aligned}$$

$$\begin{aligned}
& + \left(\underbrace{-\lambda}_{\leq 0} + \log \frac{1}{\delta} + \log \frac{1}{1-p} - \underbrace{\eta \log \frac{1}{p}}_{\leq 0} \right) \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) \\
& + \log(\lambda + x) \underbrace{\sum_{y=0}^{\infty} \text{Pois}_{\lambda+x}(y)y}_{=\lambda+x} - \log \lambda \underbrace{\sum_{y=0}^{\eta-1} \text{Pois}_{\lambda+x}(y)y}_{=(\lambda+x) - \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)y} \\
& + \left(1 + \log \frac{1}{p} \right) \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)y - \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)y \log y \tag{5.110}
\end{aligned}$$

$$\begin{aligned}
\leq & \underbrace{\left(\log \frac{1}{\delta} + \log \frac{1}{1-p} \right) \Pr[Y \geq \eta | X = x]}_{\triangleq D_3(x)} \\
& + \underbrace{\left(1 + \log \frac{1}{p} + \log \lambda \right) \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)y}_{\triangleq D_4(x)} \\
& + \underbrace{(\lambda + x) \log \left(1 + \frac{x}{\lambda} \right) - \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y)y \log y}_{\triangleq D_5(x)} \tag{5.111}
\end{aligned}$$

It follows from (5.49) and (5.111) that $C(\lambda, E, \infty)$ is upper-bounded by

$$\begin{aligned}
C(\lambda, E, \infty) & \leq \sup_{\mathbb{E}[X] \leq E} \mathbb{E}[D_3(X) + D_4(X) + D_5(X)] \tag{5.112} \\
& \leq \sup_{\mathbb{E}[X] \leq E} \mathbb{E}[D_3(X)] + \sup_{\mathbb{E}[X] \leq E} \mathbb{E}[D_4(X)]
\end{aligned}$$

$$+ \sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_5(X)]. \quad (5.113)$$

To find an upper bound on the capacity, we shall next upper-bound the three terms on the RHS of (5.113) separately.

We first consider $\sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_3(X)]$. By Lemma 5.4,

$$\mathbb{E}[D_3(X)] \leq \left(\log \frac{1}{\delta} + \log \frac{1}{1-p} \right) \mathbb{E}[U_{\lambda+X}(\eta)]. \quad (5.114)$$

Further, when $\mathbb{E}[X] \leq \mathbb{E}$, we can bound $\mathbb{E}[U_{\lambda+X}(\eta)]$ as

$$\mathbb{E}[U_{\lambda+X}(\eta)] = U_\lambda(\eta) + \mathbb{E} \left[X \cdot \frac{U_{\lambda+X}(\eta) - U_\lambda(\eta)}{X} \right] \quad (5.115)$$

$$\leq U_\lambda(\eta) + \mathbb{E}[X] \cdot \sup_{x \in \mathbb{R}^+} \left(\frac{U_{\lambda+x}(\eta) - U_\lambda(\eta)}{x} \right) \quad (5.116)$$

$$\leq U_\lambda(\eta) + \mathbb{E} \cdot \sup_{x \in \mathbb{R}^+} \left(\frac{U_{\lambda+x}(\eta) - U_\lambda(\eta)}{x} \right) \quad (5.117)$$

where in the first step we adopt the convention $\frac{0}{0} = 0$. To upper-bound the supremum in (5.117), we first observe that $U_{\lambda+x}(\eta)$ is convex in x for $x \in [0, \eta - \sqrt{\eta} - \lambda]$. Then, by Lemma 5.5, for $x < \eta - \sqrt{\eta} - \lambda$,

$$\begin{aligned} & \frac{U_{\lambda+x}(\eta) - U_\lambda(\eta)}{x} \\ & \leq \frac{U_{\eta - \sqrt{\eta}}(\eta) - U_\lambda(\eta)}{\eta - \sqrt{\eta} - \lambda} \end{aligned} \quad (5.118)$$

$$\leq \frac{1}{\eta - \sqrt{\eta} - \lambda}, \quad x < \eta - \sqrt{\eta} - \lambda. \quad (5.119)$$

On the other hand, when $x \geq \eta - \sqrt{\eta} - \lambda$,

$$\frac{U_{\lambda+x}(\eta) - U_\lambda(\eta)}{x} \leq \frac{1}{x} \leq \frac{1}{\eta - \sqrt{\eta} - \lambda}, \quad x \geq \eta - \sqrt{\eta} - \lambda. \quad (5.120)$$

Thus we obtain

$$\sup_{x \in \mathbb{R}^+} \left(\frac{U_{\lambda+x}(\eta) - U_\lambda(\eta)}{x} \right) \leq \frac{1}{\eta - \sqrt{\eta} - \lambda}. \quad (5.121)$$

Combining (5.117) and (5.121) yields that for η larger than some $\eta(\lambda)$,

$$\begin{aligned} & \sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[U_{\lambda+X}(\eta)] \\ & \leq U_{\lambda}(\eta) + \frac{\mathbb{E}}{\eta - \sqrt{\eta} - \lambda} \end{aligned} \quad (5.122)$$

$$= \exp(\eta - \lambda + \eta \log \lambda - \eta \log \eta) + \frac{\mathbb{E}}{\eta - \sqrt{\eta} - \lambda} \quad (5.123)$$

$$\leq \exp(\eta + \eta \log \lambda - \eta \log \eta) + \frac{\mathbb{E}}{\eta - \sqrt{\eta} - \lambda}. \quad (5.124)$$

We next use Lemma 5.4 to bound δ as

$$\delta = \Pr[Y \geq \eta | X = 0] \geq L_{\lambda}(\eta). \quad (5.125)$$

Therefore,

$$\log \frac{1}{\delta} \leq \log \frac{1}{L_{\lambda}(\eta)} \leq \eta \log \eta + \frac{1}{12\eta} + \frac{1}{2} \log(2\pi\eta) + \lambda - \eta \log \lambda \quad (5.126)$$

where the second inequality is obtained by omitting nonpositive terms. Using (5.114), (5.124) and (5.126) we obtain that for η larger than some $\eta(\lambda)$,

$$\begin{aligned} & \sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_3(X)] \\ & \leq \left(\eta \log \eta + \frac{1}{12\eta} + \frac{1}{2} \log(2\pi\eta) + \lambda - \eta \log \lambda + \log \frac{1}{1-p} \right) \\ & \quad \cdot \left(\exp \left(\eta + \eta \log \lambda - \eta \log \eta + \frac{\mathbb{E}}{\eta - \sqrt{\eta} - \lambda} \right) \right). \end{aligned} \quad (5.127)$$

We next derive an upper bound on $\sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_4(X)]$. To this end, we observe

$$\sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) y = \sum_{y=\eta}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \cdot y \quad (5.128)$$

$$= (\lambda+x) \sum_{y=\eta}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^{y-1}}{(y-1)!} \quad (5.129)$$

$$= (\lambda+x) \sum_{y=\eta-1}^{\infty} \text{Pois}_{\lambda+x}(y) \quad (5.130)$$

$$= (\lambda+x) \Pr[Y \geq \eta-1 | X = x] \quad (5.131)$$

$$\leq (\lambda+x) U_{\lambda+x}(\eta-1) \quad (5.132)$$

where the inequality follows by Lemma 5.4. Using the above inequality we obtain

$$\begin{aligned} & \mathbb{E}[D_4(X)] \\ & \leq \max \left\{ 0, \left(1 + \log \frac{1}{p} + \log \lambda \right) \mathbb{E}[(\lambda + X)U_{\lambda+X}(\eta - 1)] \right\}. \end{aligned} \quad (5.133)$$

Similarly to (5.117), when $\mathbb{E}[X] \leq \mathbb{E}$ we can bound the expectation on the RHS of (5.133) as

$$\begin{aligned} & \mathbb{E}[(\lambda + X)U_{\lambda+X}(\eta - 1)] \\ & \leq \lambda U_\lambda(\eta - 1) \\ & \quad + \mathbb{E} \cdot \sup_{x \in \mathbb{R}^+} \left(\frac{(\lambda + x)U_{\lambda+x}(\eta - 1) - \lambda U_\lambda(\eta - 1)}{x} \right). \end{aligned} \quad (5.134)$$

To bound the supremum on the RHS of (5.134), we observe that $(\lambda + x)U_{\lambda+x}(\eta - 1)$ is convex in x on $[0, \eta - \sqrt{\eta} - \lambda]$. Thus, by Lemma 5.5, for $x < \eta - \sqrt{\eta} - \lambda$,

$$\begin{aligned} & \frac{(\lambda + x)U_{\lambda+x}(\eta - 1) - \lambda U_\lambda(\eta - 1)}{x} \\ & \leq \frac{(\eta - \sqrt{\eta})U_{\eta - \sqrt{\eta}}(\eta - 1) - \lambda U_\lambda(\eta - 1)}{\eta - \sqrt{\eta} - \lambda} \end{aligned} \quad (5.135)$$

$$\leq \frac{\lambda}{\eta - \sqrt{\eta} - \lambda} + 1, \quad x < \eta - \sqrt{\eta} - \lambda. \quad (5.136)$$

When $x \geq \eta - \sqrt{\eta} - \lambda$,

$$\begin{aligned} & \frac{(\lambda + x)U_{\lambda+x}(\eta - 1) - \lambda U_\lambda(\eta - 1)}{x} \\ & \leq \frac{\lambda + x}{x} \end{aligned} \quad (5.137)$$

$$\leq \frac{\lambda}{\eta - \sqrt{\eta} - \lambda} + 1. \quad (5.138)$$

Thus the supremum on the RHS of (5.134) can be bounded as

$$\begin{aligned} & \sup_{x \in \mathbb{R}^+} \left(\frac{(\lambda + x)U_{\lambda+x}(\eta - 1) - \lambda U_\lambda(\eta - 1)}{x} \right) \\ & \leq \frac{\lambda}{\eta - \sqrt{\eta} - \lambda} + 1. \end{aligned} \quad (5.139)$$

Combining (5.133), (5.134), (5.139) and the definition of $U_\lambda(\cdot)$ we have that for η larger than some $\eta(\lambda)$,

$$\begin{aligned} & \sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_4(X)] \\ & \leq \max \left\{ 0, \left(1 + \log \frac{1}{p} + \log \lambda \right) \cdot \left(\mathbb{E} + \frac{\lambda \mathbb{E}}{\eta - \sqrt{\eta} - \lambda} \right. \right. \\ & \quad \left. \left. + \lambda \cdot e^{\eta-1-\lambda+(\eta-1) \log \lambda - (\eta-1) \log(\eta-1)} \right) \right\}. \end{aligned} \quad (5.140)$$

We now consider $\sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_5(X)]$. For $x \leq \eta - \lambda$, we simply bound $D_5(x)$ by

$$D_5(x) \leq (\lambda + x) \log \left(1 + \frac{x}{\lambda} \right), \quad x \leq \eta - \lambda. \quad (5.141)$$

When $x > \eta - \lambda$, we use the inequality

$$y \log y \geq (\lambda + x) \log(\lambda + x) + (1 + \log(\lambda + x))(y - (\lambda + x)) \quad (5.142)$$

to obtain

$$\begin{aligned} & \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) y \log y \\ & \geq (\lambda + x) \log(\lambda + x) \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) \\ & \quad + (1 + \log(\lambda + x)) \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) (y - (\lambda + x)) \end{aligned} \quad (5.143)$$

$$\begin{aligned} & \geq (\lambda + x) \log(\lambda + x) \sum_{y=\eta}^{\infty} \text{Pois}_{\lambda+x}(y) \\ & \quad + (1 + \log(\lambda + x)) \underbrace{\sum_{y=0}^{\infty} \text{Pois}_{\lambda+x}(y) (y - (\lambda + x))}_{=0} \end{aligned} \quad (5.144)$$

$$= (\lambda + x) \log(\lambda + x) \Pr[Y \geq \eta | X = x] \quad (5.145)$$

$$= (\lambda + x) \log(\lambda + x) (1 - \Pr[Y \leq \eta - 1 | X = x]), \quad (5.146)$$

where the second inequality is obtained by adding nonpositive terms. Thus we may bound $D_5(x)$, when $x > \eta - \lambda$, by

$$D_5(x) \leq (\lambda + x) \log \left(1 + \frac{x}{\lambda} \right) \Pr[Y \leq \eta - 1 | X = x] \\ + (\lambda + x) \max \left\{ \log \frac{1}{\lambda}, 0 \right\}. \quad (5.147)$$

By combining (5.141) and (5.147) and by adding nonnegative terms, we can upper-bound $\mathbb{E}[D_5(X)]$, for all input distributions satisfying $\mathbb{E}[X] \leq \mathbb{E}$, by

$$\mathbb{E}[D_5(X)] \leq \mathbb{E}[X] \max \left\{ \log \frac{1}{\lambda}, 0 \right\} \\ + \lambda \max \left\{ \log \frac{1}{\lambda}, 0 \right\} \cdot \Pr[X \geq \eta - \lambda] + \mathbb{E}[D_6(X)] \quad (5.148) \\ \leq \mathbb{E} \max \left\{ \log \frac{1}{\lambda}, 0 \right\} + \frac{\lambda \mathbb{E}}{\eta - \lambda} \max \left\{ \log \frac{1}{\lambda}, 0 \right\} \\ + \mathbb{E}[D_6(X)] \quad (5.149)$$

where in the second step we applied Markov's inequality to $\Pr[X \geq \eta - \lambda]$, and where

$$D_6(x) = (\lambda + x) \log \left(1 + \frac{x}{\lambda} \right) \cdot \\ \begin{cases} 1, & x \leq \eta - \lambda \\ \Pr[Y \leq \eta - 1 | X = x], & x > \eta - \lambda. \end{cases} \quad (5.150)$$

To upper-bound $\Pr[Y \leq \eta - 1 | X = x]$ when $x > \eta - \lambda$, we use the Chernoff Bound and (5.93) to write

$$\Pr[Y \leq \eta - 1 | X = x] \leq \frac{\mathbb{E}[e^{\theta Y}]}{e^{\theta(\eta-1)}} = \frac{e^{(\lambda+x)(e^\theta-1)}}{e^{\theta(\eta-1)}}, \quad \theta < 0. \quad (5.151)$$

Letting $\theta = \log \left(\frac{\eta-1}{\lambda+x} \right)$ in the above inequality yields

$$\Pr[Y \leq \eta - 1 | X = x] \\ \leq e^{\eta-1-(\lambda+x)+(\eta-1) \log(\lambda+x)-(\eta-1) \log(\eta-1)}, \\ x > \eta - \lambda. \quad (5.152)$$

Substituting (5.152) into the definition of $D_6(x)$ we obtain

$$D_6(x) \leq D_7(x), \quad x \in \mathbb{R}_0^+ \quad (5.153)$$

where

$$D_7(x) \triangleq (\lambda + x) \log \left(1 + \frac{x}{\lambda} \right) \cdot \begin{cases} 1, & x \leq \eta - \lambda \\ e^{\eta-1-(\lambda+x)+(\eta-1) \log(\lambda+x)-(\eta-1) \log(\eta-1)}, & x > \eta - \lambda. \end{cases} \quad (5.154)$$

Thus, when $\mathbb{E}[X] \leq \mathbb{E}$,

$$\mathbb{E}[D_6(X)] \leq \mathbb{E}[D_7(X)] \quad (5.155)$$

$$= \mathbb{E} \left[X \cdot \frac{D_7(X)}{X} \right] \quad (5.156)$$

$$\leq \mathbb{E} \cdot \sup_{x \in \mathbb{R}^+} \left(\frac{D_7(x)}{x} \right) \quad (5.157)$$

where we adopt the convention $\frac{0}{0} = 0$. It can be checked by computing the derivative of $\frac{D_7(x)}{x}$ with respect to x that, for η larger than some $\eta(\lambda)$, $\frac{D_7(x)}{x}$ is monotonically decreasing in x for $x > \eta - \lambda$. Using this observation, Lemma 5.5, and the fact that $(\lambda + x) \log \left(1 + \frac{x}{\lambda} \right)$ is convex in x on \mathbb{R}^+ , we have that the supremum on the RHS of (5.157) is achieved when $x = \eta - \lambda$. Therefore, when $\mathbb{E}[X] \leq \mathbb{E}$,

$$\mathbb{E}[D_6(X)] \leq \mathbb{E} \cdot \frac{\eta \log \frac{\eta}{\lambda}}{\eta - \lambda}. \quad (5.158)$$

Combining (5.149) and (5.158) we obtain

$$\begin{aligned} \sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_5(X)] &\leq \mathbb{E} \cdot \left(1 + \frac{\lambda}{\eta - \lambda} \right) \cdot \max \left\{ \log \frac{1}{\lambda}, 0 \right\} \\ &+ \mathbb{E} \cdot \frac{\eta \log \left(\frac{\eta}{\lambda} \right)}{\eta - \lambda}. \end{aligned} \quad (5.159)$$

Thus, for any $\lambda > 0$, $\mathbb{E} > 0$, $p \in (0, 1)$, and η larger than some $\eta(\lambda)$, we can combine (5.113), (5.127), (5.140) and (5.159) to obtain an upper bound on $C(\lambda, \mathbb{E}, \infty)$.

5.4.2.3 Proof of (5.86)

For small enough \mathbb{E} , we choose

$$\eta = \left\lfloor \log \frac{1}{\mathbb{E}} \right\rfloor, \quad (5.160)$$

and let $p \in (0, 1)$ have any fixed value that does not depend on \mathbb{E} . In this case, it follows by (5.113) that

$$\begin{aligned} \overline{\lim}_{\mathbb{E} \downarrow 0} \frac{C(\lambda, \mathbb{E}, \infty)}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} &\leq \overline{\lim}_{\mathbb{E} \downarrow 0} \frac{\sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_3(X)]}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} \\ &\quad + \overline{\lim}_{\mathbb{E} \downarrow 0} \frac{\sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_4(X)]}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} \\ &\quad + \overline{\lim}_{\mathbb{E} \downarrow 0} \frac{\sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_5(X)]}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}}. \end{aligned} \quad (5.161)$$

Substituting (5.160) into (5.127) yields

$$\overline{\lim}_{\mathbb{E} \downarrow 0} \frac{\sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_3(X)]}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} \leq 1. \quad (5.162)$$

Substituting (5.160) into (5.140) yields

$$\overline{\lim}_{\mathbb{E} \downarrow 0} \frac{\sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_4(X)]}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} \leq 0. \quad (5.163)$$

Finally, substituting (5.160) into (5.159) yields

$$\overline{\lim}_{\mathbb{E} \downarrow 0} \frac{\sup_{\mathbb{E}[X] \leq \mathbb{E}} \mathbb{E}[D_5(X)]}{\mathbb{E} \log \log \frac{1}{\mathbb{E}}} \leq 1. \quad (5.164)$$

Combining (5.161), (5.162), (5.163) and (5.164) proves (5.86).

Appendix A

One-Shot Capacity Bounds in Classical Settings

The simplicity of the bounds in Chapter 2 makes it possible to extend them to more complicated scenarios. In this appendix we apply the techniques of Chapter 2 to derive capacity bounds in three settings in classical information theory: the single-user channel with SI, the multiple-access channel (MAC), and the MAC with SI. The bounds are applied to infinitely many channel uses to obtain general formulas for capacities or capacity regions. In the case of the MAC, these general formulas are equivalent to the known ones [15], while in the other two cases, these general formulas are new, to the best of our knowledge.

A.1 The Single-User Channel with SI

Consider a one-shot channel with a state S taking value in \mathcal{S} according to distribution P_S where, conditional on the channel input being x and the state being s , the output distribution is $W(\cdot|x, s)$. Suppose that the realization of the state is known to the encoder but not to the decoder, so

the encoder is a mapping from (m, s) to x , where m denotes the message, and the decoder is a mapping from y to its guess m' of the transmitted message. An (R, ϵ) -code consists of an encoder-decoder pair of size 2^R such that, when a message is chosen uniformly over $\{1, \dots, 2^R\}$, the probability of a decoding error is at most ϵ .

Proposition A.1. *If an (R, ϵ) -code exists for a one-shot channel with SI, then*

$$R \leq \sup_{\substack{U: U \perp S \\ g: (U, S) \mapsto X}} D_{\text{H}}^{\epsilon}(P_{UY} \| P_U \otimes P_Y). \quad (\text{A.1})$$

Proof. Note that (2.47) holds for the one-shot channel with SI. Also note that $M \rightarrow Y \rightarrow M'$ forms a Markov chain. We hence have

$$R \leq D_{\text{H}}^{\epsilon}(P_{MY} \| P_M \otimes P_Y). \quad (\text{A.2})$$

The proposition now follows from the facts that $M \perp S$ and that the encoder is a mapping from (M, S) to X . \square

Proposition A.2. *For any $\epsilon > \epsilon' > 0$, there exists an (R, ϵ) -code for a one-shot channel with SI satisfying*

$$R \geq \sup_{\substack{U: U \perp S \\ g: (U, S) \mapsto X}} D_{\text{H}}^{\epsilon'}(P_{UY} \| P_U \otimes P_Y) - \log \frac{1}{\epsilon - \epsilon'}. \quad (\text{A.3})$$

Proof. For any $U \perp S$ and $g: (U, S) \rightarrow X$, consider the derived channel whose input is U and whose output is Y . Any code for the derived channel yields a code for the original channel with the same error probability. Hence applying Theorem 2.8 to the derived channel gives the desired bound. \square

We next apply Propositions A.1 and A.2 to many uses of a general channel with *noncausal* SI. Such a channel is described by a sequence of transition laws $W_n(y^n | x^n, s^n)$, $n \in \mathbb{N}$, where the state sequence s^n is drawn according to P_{S^n} and is known noncausally to the encoder but unknown to the decoder. Thus, the encoder is a mapping from (m, s^n) to x^n , and the decoder is a mapping from y^n to m' . The definition for the capacity of such a channel is the same as in Section 2.5.

From Propositions A.1 and A.2 we immediately obtain the following general capacity formula:

Proposition A.3. *The capacity of any channel with noncausal SI is*

$$C = \lim_{\epsilon \downarrow 0} \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sup_{\substack{U_n: U_n \perp S^n \\ g_n: (U_n, S^n) \mapsto X^n}} D_{\text{H}}^{\epsilon}(P_{U_n Y^n} \| P_{U_n} \otimes P_{Y^n}). \quad (\text{A.4})$$

Remark A.4. *According to Lemma 2.5, (A.4) is equivalent to*

$$C = \sup_{\substack{\{U_n: U_n \perp S^n\} \\ \{g_n: (U_n, S^n) \mapsto X^n\}}} \underline{I}(\mathbf{U}; \mathbf{Y}), \quad (\text{A.5})$$

where $\underline{I}(\mathbf{U}; \mathbf{Y})$ is the inf-information rate [46] between $\{U_n\}$ and $\{Y^n\}$:

$$\underline{I}(\mathbf{U}; \mathbf{Y}) \triangleq \sup \left\{ a \left| \lim_{n \rightarrow \infty} \Pr \left[\frac{1}{n} \log \frac{P_{U_n Y^n}}{P_{U_n} \otimes P_{Y^n}} \geq a \right] = 1 \right. \right\}, \quad (\text{A.6})$$

where the probability is computed according to $P_{U_n Y^n}$.

Remark A.5. *The single-letter formula for a DMC with IID states is [13]*

$$C_{\text{memoryless}} = \sup_{\substack{P_{U|S} \\ g: (U, S) \mapsto X}} I(U; Y) - I(U; S). \quad (\text{A.7})$$

An analogous expression in terms of information spectrums is

$$\tilde{C} \triangleq \sup \left\{ a \left| \lim_{n \rightarrow \infty} \Pr \left[\frac{1}{n} \log \frac{P_{S^n} \otimes P_{U_n Y^n}}{P_{S^n U_n} \otimes P_{Y^n}} \geq 1 \right] = 1 \right. \right\}, \quad (\text{A.8})$$

where the probability is computed according to $P_{S^n U_n Y^n}$. It can be seen that

$$C \leq \tilde{C}. \quad (\text{A.9})$$

We do not know yet whether (A.9) holds with equality or not.

A.2 The Multiple-Access Channel

We now consider a one-shot MAC with two encoders and one decoder, described by the transition law $W(y|x_1, x_2)$. Encoder 1 is a mapping from message m_1 to input x_1 , Encoder 2 is a mapping from message m_2 to input x_2 , and decoder is a mapping from the channel output y to guesses of both messages (m'_1, m'_2) . An (R_1, R_2, ϵ) -code consists of a

codebook for Encoder 1 of size 2^{R_1} , a codebook for Encoder 2 of size 2^{R_2} , and a decoder such that, when both messages are chosen uniformly, the probability of a decoding error is at most ϵ :

$$\Pr [(M_1, M_2) \neq (M'_1, M'_2)] \leq \epsilon. \quad (\text{A.10})$$

Proposition A.6. *If an (R_1, R_2, ϵ) -code exists for the one-shot MAC, then there exists a joint distribution of the form*

$$P_{X_1 X_2 Y}(x_1, x_2, y) = P_{X_1}(x_1)P_{X_2}(x_2)W(y|x_1, x_2) \quad (\text{A.11})$$

and a mapping $\Phi: \mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \mathcal{Y} \rightarrow [0, 1]$ such that

$$\int \Phi \, dP_{X_1 X_2 Y} \geq 1 - \epsilon \quad (\text{A.12})$$

and

$$R_1 \leq -\log \int \Phi \, d(P_{X_1} \otimes P_{X_2 Y}) \quad (\text{A.13a})$$

$$R_2 \leq -\log \int \Phi \, d(P_{X_2} \otimes P_{X_1 Y}) \quad (\text{A.13b})$$

$$R_1 + R_2 \leq -\log \int \Phi \, d(P_{X_1} \otimes P_{X_2} \otimes P_Y). \quad (\text{A.13c})$$

Proof. We first look at the joint distribution $P_{M_1 M_2 M'_1 M'_2}$ which is generated by uniform and independent messages (M_1, M_2) and the encoders, the channel, and the decoder. Let $\Psi(m_1, m_2, m'_1, m'_2)$ equal 1 when $(m_1, m_2) = (m'_1, m'_2)$ and equal 0 otherwise. Then, since the average error probability is no larger than ϵ ,

$$\int \Psi \, dP_{M_1 M_2 M'_1 M'_2} \geq 1 - \epsilon. \quad (\text{A.14})$$

On the other hand, it is easy to see

$$R_1 = -\log \int \Psi \, d(P_{M_1} \otimes P_{M_2 M'_1 M'_2}) \quad (\text{A.15a})$$

$$R_2 = -\log \int \Psi \, d(P_{M_2} \otimes P_{M_1 M'_1 M'_2}) \quad (\text{A.15b})$$

$$R_1 + R_2 = -\log \int \Psi \, d(P_{M_1} \otimes P_{M_2} \otimes P_{M'_1 M'_2}). \quad (\text{A.15c})$$

Clearly, the distribution $P_{X_1 X_2 Y}$ on (X_1, X_2, Y) generated by the uniform and independent messages (M_1, M_2) is of the form (A.11). Note that the product of the inverse mapping of the encoders combined with the decoder is a stochastic kernel that maps $P_{X_1 X_2 Y}$ to $P_{M_1 M_2 M'_1 M'_2}$, $P_{X_1} \otimes P_{X_2 Y}$ to $P_{M_1} \otimes P_{M_2 M'_1 M'_2}$, $P_{X_2} \otimes P_{X_1 Y}$ to $P_{M_2} \otimes P_{M_1 M'_1 M'_2}$, and $P_{X_1} \otimes P_{X_2} \otimes P_Y$ to $P_{M_1} \otimes P_{M_2} \otimes P_{M'_1 M'_2}$. Hence from this stochastic kernel and the above chosen Ψ one can construct a $\Phi: \mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \mathcal{Y} \rightarrow [0, 1]$ that satisfies (A.12) and (A.13). \square

Proposition A.7. *For any joint distribution of the form*

$$P_{X_1 X_2 Y}(x_1, x_2, y) = P_{X_1}(x_1)P_{X_2}(x_2)W(y|x_1, x_2), \quad (\text{A.16})$$

any $\epsilon' > 0$, and any mapping $\Phi: \mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \mathcal{Y} \rightarrow [0, 1]$ such that

$$\int \Phi dP_{X_1 X_2 Y} \geq 1 - \epsilon', \quad (\text{A.17})$$

there exists an (R_1, R_2, ϵ) -code for the one-shot MAC with

$$\begin{aligned} \epsilon \leq & \epsilon' + 2^{R_1} \int \Phi d(P_{X_1} \otimes P_{X_2 Y}) + 2^{R_2} \int \Phi d(P_{X_2} \otimes P_{X_1 Y}) \\ & + 2^{R_1 + R_2} \int \Phi d(P_{X_1} \otimes P_{X_2} \otimes P_Y). \end{aligned} \quad (\text{A.18})$$

Proof. We generate two codebooks $\{x_i\}_{i \in \{1, \dots, 2^{R_1}\}}$ and $\{x_j\}_{j \in \{1, \dots, 2^{R_2}\}}$ for Encoders 1 and 2, respectively, independently according to IID P_{X_1} and IID P_{X_2} . When y is received, the decoder selects some or none of the message-pairs in $\{1, \dots, 2^{R_1}\} \otimes \{1, \dots, 2^{R_2}\}$ such that (k, l) is selected with probability $\Phi(x_{1,k}, x_{2,l}, y)$. If only one message-pair is selected, the decoder outputs it; otherwise it declares an error.

The error probability is upper-bounded by the sum of the probabilities of the following four events, which we analyze in the same way as in the proof of Theorem 2.10 (the probability of each event is averaged over all codebooks):

- The correct message-pair is not selected. This probability is upper-bounded by ϵ' according to (A.17).
- A message-pair with an incorrect message from Encoder 1 but a correct message from Encoder 2 is selected. This probability is

upper-bounded by

$$2^{R_1} \int \Phi d(P_{X_1} \otimes P_{X_2 Y}). \quad (\text{A.19})$$

- A message-pair with a correct message from Encoder 1 but an incorrect message from Encoder 2 is selected. This probability is upper-bounded by

$$2^{R_2} \int \Phi d(P_{X_2} \otimes P_{X_1 Y}). \quad (\text{A.20})$$

- A message-pair with two incorrect messages is selected. This probability is upper-bounded by

$$2^{R_1+R_2} \int \Phi d(P_{X_1} \otimes P_{X_2} \otimes P_Y). \quad (\text{A.21})$$

Combining these four bounds yields (A.18). \square

We next consider many uses of a general MAC. Such a channel is described by a sequence of transition laws $W_n(y^n|x_1^n, x_2^n)$. An (n, R_1, R_2, ϵ) -code consists of Encoder 1 which maps an element from $\{1, \dots, 2^{R_1}\}$ to x_1^n , Encoder 2 which maps an element from $\{1, \dots, 2^{R_2}\}$ to x_2^n , and a decoder which maps y^n to an element in $\{1, \dots, 2^{R_1}\} \otimes \{1, \dots, 2^{R_2}\}$, such that the average probability of decoding error is at most ϵ . A rate-pair (R_1, R_2) is said to be achievable if there exists a sequence of $(n, R_1, R_2, \epsilon_n)$ -codes with ϵ_n tending to zero as n tends to infinity. The *capacity region* of the MAC is the set of all achievable rate-pairs.

Using Propositions A.6 and A.7 we can obtain capacity region and ϵ -capacity region formulas for the general MAC, which are equivalent to those in [15]. We show the capacity region formula in the following.

Let $\mathcal{R}(\mathbf{W})$ be the set of all rate-pairs for which there exists a sequence of distributions of the form

$$P_{X_1^n X_2^n Y^n}(x_1^n, x_2^n, y^n) = P_{X_1^n}(x_1^n) P_{X_2^n}(x_2^n) W(y^n|x_1^n, x_2^n) \quad (\text{A.22})$$

such that

$$R_1 \leq \lim_{\epsilon \downarrow 0} \varliminf_{n \rightarrow \infty} D_{\text{H}}^{\epsilon}(P_{X_1^n X_2^n Y^n} \| P_{X_1^n} \otimes P_{X_2^n Y^n}) \quad (\text{A.23a})$$

$$R_2 \leq \lim_{\epsilon \downarrow 0} \varliminf_{n \rightarrow \infty} D_{\text{H}}^{\epsilon}(P_{X_1^n X_2^n Y^n} \| P_{X_2^n} \otimes P_{X_1^n Y^n}) \quad (\text{A.23b})$$

$$R_1 + R_2 \leq \lim_{\epsilon \downarrow 0} \varliminf_{n \rightarrow \infty} D_{\text{H}}^{\epsilon}(P_{X_1^n X_2^n Y^n} \| P_{X_1^n} \otimes P_{X_2^n} \otimes P_{Y^n}). \quad (\text{A.23c})$$

Proposition A.8. *The capacity region of a general MAC is $\mathcal{R}(\mathbf{W})$.*

Proof. Let $\tilde{\mathcal{R}}(\mathbf{W})$ be the set of all rate-pairs (R_1, R_2) for which there exists a sequence of distributions of the form (A.22) and a sequence of Φ_n with

$$\lim_{n \rightarrow \infty} \int \Phi_n \, dP_{X_1^n X_2^n Y^n} = 1, \quad (\text{A.24})$$

such that

$$R_1 \leq \varliminf_{n \rightarrow \infty} -\log \int \Phi_n \, d(P_{X_1^n} \otimes P_{X_2^n Y^n}) \quad (\text{A.25a})$$

$$R_2 \leq \varliminf_{n \rightarrow \infty} -\log \int \Phi_n \, d(P_{X_2^n} \otimes P_{X_1^n Y^n}) \quad (\text{A.25b})$$

$$R_1 + R_2 \leq \varliminf_{n \rightarrow \infty} -\log \int \Phi_n \, d(P_{X_1^n} \otimes P_{X_2^n} \otimes P_{Y^n}). \quad (\text{A.25c})$$

From Propositions A.6 and A.7 it follows immediately that the capacity region is $\mathcal{R}(\mathbf{W})$. Thus it remains to show

$$\mathcal{R}(\mathbf{W}) = \tilde{\mathcal{R}}(\mathbf{W}). \quad (\text{A.26})$$

To this end, note that the bounds (A.23) are equivalent to: there exists sequences $\Phi_{1,n}$, $\Phi_{2,n}$ and $\Phi_{3,n}$ with

$$\lim_{n \rightarrow \infty} \int \Phi_{i,n} \, dP_{X_1^n X_2^n Y^n} = 1, \quad i = 1, 2, 3 \quad (\text{A.27})$$

such that

$$R_1 \leq \varliminf_{n \rightarrow \infty} -\log \int \Phi_{1,n} \, d(P_{X_1^n} \otimes P_{X_2^n Y^n}) \quad (\text{A.28a})$$

$$R_2 \leq \varliminf_{n \rightarrow \infty} -\log \int \Phi_{2,n} \, d(P_{X_2^n} \otimes P_{X_1^n Y^n}) \quad (\text{A.28b})$$

$$R_1 + R_2 \leq \varliminf_{n \rightarrow \infty} -\log \int \Phi_{3,n} \, d(P_{X_1^n} \otimes P_{X_2^n} \otimes P_{Y^n}). \quad (\text{A.28c})$$

Restricting $\Phi_{1,n} = \Phi_{2,n} = \Phi_{3,n} \triangleq \Phi_n$ in the above yields the region $\tilde{\mathcal{R}}(\mathbf{W})$. Hence

$$\tilde{\mathcal{R}}(\mathbf{W}) \subseteq \mathcal{R}(\mathbf{W}). \quad (\text{A.29})$$

On the other hand, for given $\Phi_{i,n}$, $i = 1, 2, 3$, if we choose

$$\tilde{\Phi}_n \triangleq \Phi_{1,n}\Phi_{2,n}\Phi_{3,n}, \quad (\text{A.30})$$

then

$$\tilde{\Phi}_n \geq 1 - \sum_{i=1,2,3} (1 - \Phi_{i,n}), \quad (\text{A.31})$$

yielding

$$\begin{aligned} \lim_{n \rightarrow \infty} \int \tilde{\Phi}_n dP_{X_1^n X_2^n Y^n} \\ \geq 1 - \sum_{i=1,2,3} \left\{ 1 - \lim_{n \rightarrow \infty} \int \Phi_{i,n} dP_{X_1^n X_2^n Y^n} \right\} \end{aligned} \quad (\text{A.32})$$

$$= 1. \quad (\text{A.33})$$

Because

$$\tilde{\Phi}_n \leq \Phi_{i,n}, \quad i = 1, 2, 3, \quad (\text{A.34})$$

the bounds (A.25) computed with $\tilde{\Phi}_n$ are not tighter than the bounds (A.28). Hence

$$\mathcal{R}(\mathbf{W}) \subseteq \tilde{\mathcal{R}}(\mathbf{W}). \quad (\text{A.35})$$

Combining (A.29) and (A.35) yields (A.26). \square

A.3 The MAC with SI

We shall next combine the techniques used in Sections A.1 and A.2 to study the MAC with SI. Though a single-letter expression for the capacity region of a memoryless MAC with noncausal SI is still unknown, as we shall see, it is rather easy to obtain (nonsingle-letter) expressions for the capacity region of a general MAC with noncausal SI.

We start with one-shot bounds. Consider a one-shot MAC with states (S_1, S_2) of joint distribution $P_{S_1 S_2}$ and channel law $W(y|x_1, x_2, s_1, s_2)$. Assume that the state s_1 is known to Encoder 1 only, and that the state s_2 is known to Encoder 2 only.

Proposition A.9. *If an (R_1, R_2, ϵ) -code exists for a one-shot MAC with SI, then there exist a joint distribution*

$$P_{U_1 U_2 Y}(u_1, u_2, y) = \sum_{s_1, s_2} P_{S_1 S_2}(s_1, s_2) P_{U_1}(u_1) P_{U_2}(u_2) W(y | g_1(u_1, s_1), g_2(u_2, s_2), s_1, s_2) \quad (\text{A.36})$$

where $g_1: (u_1, s_1) \mapsto x_1$ and $g_2: (u_2, s_2) \mapsto x_2$ are deterministic mappings, and a $\Phi: \mathcal{U}_1 \otimes \mathcal{U}_2 \otimes \mathcal{Y} \rightarrow [0, 1]$ such that

$$\int \Phi dP_{U_1 U_2 Y} \geq 1 - \epsilon \quad (\text{A.37})$$

and

$$R_1 \leq -\log \int \Phi d(P_{U_1} \otimes P_{U_2 Y}) \quad (\text{A.38a})$$

$$R_2 \leq -\log \int \Phi d(P_{U_2} \otimes P_{U_1 Y}) \quad (\text{A.38b})$$

$$R_1 + R_2 \leq -\log \int \Phi d(P_{U_1} \otimes P_{U_2} \otimes P_Y). \quad (\text{A.38c})$$

Proof. For the given code, let U_1 be the message M_1 and U_2 be the message M_2 , then the proof follows in the same way as that of Proposition A.6. \square

Proposition A.10. *For any joint distribution of the form (A.36) and any $\Phi: \mathcal{U}_1 \otimes \mathcal{U}_2 \otimes \mathcal{Y} \rightarrow [0, 1]$ such that*

$$\int \Phi dP_{U_1 U_2 Y} \geq 1 - \epsilon', \quad (\text{A.39})$$

there exists an (R_1, R_2, ϵ) -code for the one-shot MAC with

$$\begin{aligned} \epsilon &\leq \epsilon' + 2^{R_1} \int \Phi d(P_{U_1} \otimes P_{X_2 Y}) + 2^{R_2} \int \Phi d(P_{U_2} \otimes P_{U_1 Y}) \\ &\quad + 2^{R_1 + R_2} \int \Phi d(P_{U_1} \otimes P_{U_2} \otimes P_Y). \end{aligned} \quad (\text{A.40})$$

Proof. For given $g_1: (u_1, s_1) \mapsto x_1$ and $g_2: (u_2, s_2) \mapsto x_2$ we have a derived MAC with inputs U_1 and U_2 and output Y . Applying Proposition A.7 to this derived channel yields the result. \square

We next consider many uses of a general MAC with noncausal SI. Such a channel has state sequences (S_1^n, S_2^n) of joint distributions $P_{S_1^n S_2^n}$ and channel laws $W_n(y^n | x_1^n, x_2^n, s_1^n, s_2^n)$, $n \in \mathbb{N}$. The state sequence s_1^n is known noncausally to Encoder 1, but not to Encoder 2 or the decoder; and the state sequence s_2^n is known noncausally to Encoder 2, but not to Encoder 1 or the decoder. The capacity region of this channel is defined in the same way as in Section A.2.

Using Propositions A.9 and A.10 and the proof techniques used for Proposition A.8, we obtain the expression for the capacity region of the general Gel'fand-Pinsker MAC. Let $\mathcal{R}(\mathbf{W}, \mathbf{S}_1, \mathbf{S}_2)$ be the set of rate-pairs (R_1, R_2) for which there exists a sequence of distributions

$$\begin{aligned} & P_{U_n U_n Y^n}(u_{1,n}, u_{2,n}, y^n) \\ &= \sum_{s_1, s_2} P_{S_1^n S_2^n}(s_1^n, s_2^n) P_{U_{1,n}}(u_{1,n}) P_{U_{2,n}}(u_{2,n}) \\ & \quad W(y^n | g_{1,n}(u_{1,n}, s_1^n), g_{2,n}(u_{2,n}, s_2^n), s_1^n, s_2^n), \end{aligned} \quad (\text{A.41})$$

where $g_{1,n}: (u_{1,n}, s_1^n) \mapsto x_1^n$ and $g_{2,n}: (u_{2,n}, s_2^n) \mapsto x_2^n$, $n \in \mathbb{N}$, such that

$$R_1 \leq \lim_{\epsilon \downarrow 0} \underline{\lim}_{n \rightarrow \infty} D_{\text{H}}^\epsilon(P_{U_{1,n} U_{2,n} Y^n} \| P_{U_{1,n}} \otimes P_{U_{2,n} Y^n}) \quad (\text{A.42a})$$

$$R_2 \leq \lim_{\epsilon \downarrow 0} \underline{\lim}_{n \rightarrow \infty} D_{\text{H}}^\epsilon(P_{U_{1,n} U_{2,n} Y^n} \| P_{U_{2,n}} \otimes P_{U_{1,n} Y^n}) \quad (\text{A.42b})$$

$$R_1 + R_2 \leq \lim_{\epsilon \downarrow 0} \underline{\lim}_{n \rightarrow \infty} D_{\text{H}}^\epsilon(P_{U_{1,n} U_{2,n} Y^n} \| P_{U_{1,n}} \otimes P_{U_{2,n}} \otimes P_{Y^n}) \quad (\text{A.42c})$$

Proposition A.11. *The capacity region of a general Gel'fand-Pinsker MAC is $\mathcal{R}(\mathbf{W}, \mathbf{S}_1, \mathbf{S}_2)$.*

Acronyms

AVC	arbitrarily varying channel
CPM	completely positive map
DMC	discrete memoryless channel
DPI	data processing inequality
HSW	Holevo-Schumacher-Westmoreland (Theorem)
IID	independent and identically distributed
LHS	left-hand side
MAC	multiple-access channel
PMF	probability mass function
POVM	positive operator-valued measure
RHS	right-hand side
SI	side-information

List of Symbols

Sets

Δ	symmetric difference
\mathcal{A}^c	complement set of set \mathcal{A}
\mathbb{R}	set of reals
\mathbb{R}^+	set of positive reals
\mathbb{R}_0^+	set of nonnegative reals
\mathbb{Z}^+	set of positive integers
\mathbb{Z}_0^+	set of nonnegative integers

Calculus and Functions

$\lfloor x \rfloor$	largest integer not greater than x
$\lceil x \rceil$	smallest integer not less than x
sup	supremum
inf	infimum
max	maximum
min	minimum
\lim	limit
$\overline{\lim}$	limit superior
$\underline{\lim}$	limit inferior
log	base-2 logarithm function
$\Gamma(\cdot, \cdot)$	Incomplete Gamma Function
$I\{\text{statement}\}$	indicator function, equals 1 if statement is true and equals 0 otherwise
$n_x(\cdot)$	counting function for the point pattern \mathbf{x}

Algebra

\otimes	tensor product
$\rho^{\otimes n}$	tensor product of n copies of ρ
$ \psi\rangle$	normalized vector on a Hilbert space
$\langle\psi $	conjugate of $ \psi\rangle$
$\text{tr}(\cdot)$	trace
$\text{tr}_{\mathbb{A}}(\cdot)$	partial trace over space \mathbb{A}
$\mathbf{A} \leq \mathbf{B}$	operators \mathbf{A} and \mathbf{B} are such that $\mathbf{B} - \mathbf{A}$ is positive semidefinite
\mathbf{A}^+	the positive part of operator \mathbf{A} , i.e., the operator with the same positive eigenvalues as \mathbf{A} , but with the negative eigenvalues of \mathbf{A} changed to zeros
\mathbf{I}	identity operator
$\mathcal{D}(\mathbb{H})$	set of density operators on Hilbert space \mathbb{H}

Probability

$\text{Pr}[\cdot]$	probability
$\text{E}[\cdot]$	expectation
$X \text{---} \circ \text{---} Y \text{---} \circ \text{---} Z$	random variables X , Y and Z form a Markov chain
$\text{Ber}(p)$	Bernoulli Distribution of parameter p
$\text{Pois}_{\xi}(\cdot)$	Poisson Distribution of mean ξ

Information Theory

$H(\cdot)$	entropy
$h(\cdot)$	differential entropy
$D(\cdot\ \cdot)$	relative entropy
$D_0(\cdot\ \cdot)$	Rényi's relative entropy of order 0
$D_{\mathbb{H}}^{\epsilon}(\cdot\ \cdot)$	Hypothesis Testing Relative Entropy of parameter ϵ
$I(X; Y)$	mutual information between X and Y
$I(Q, W)$	mutual information over channel W with input distribution Q
$\underline{I}(\mathbf{X}; \mathbf{Y})$	inf-information rate between \mathbf{X} and \mathbf{Y}

Bibliography

- [1] R. Ahlswede, “Arbitrarily varying channels with states sequence known to the sender,” *IEEE Trans. Inform. Theory*, vol. 32, no. 5, pp. 621–629, Sept. 1986.
- [2] V. Anantharam and S. Verdú, “Bits through queues,” *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.
- [3] A. S. Bedekar, “On the information about message arrival times required for in-order decoding,” in *Proc. IEEE Int. Symposium on Inf. Theory*, Washington DC, June 24–29, 2001.
- [4] T. P. Coleman, N. Kiyavash, and V. G. Subramanian, “The rate-distortion function of a Poisson process with a queueing distortion measure,” in *Proceedings Data Compression Conference 2008*, Cliff Lodge, Snowbird, Utah, USA, Mar. 2008.
- [5] M. H. M. Costa, “Writing on dirty paper,” *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [6] T. M. Cover and M. Chiang, “Duality between channel capacity and rate distortion with two-sided state information,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1629–1638, June 2002.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [8] I. Csiszár and P. Narayan, “Arbitrarily varying channels with constrained inputs and states,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 27–34, Jan. 1988.

- [9] N. Datta, “Min- and max-relative entropies and a new entanglement monotone,” 2008. [Online]. Available: <http://arxiv.org/abs/0803.2770>
- [10] M. H. A. Davis, “Capacity and cutoff rate for Poisson-type channels,” *IEEE Trans. Inform. Theory*, vol. 26, pp. 710–715, Nov. 1980.
- [11] W. Feller, *An Introduction to Probability Theory and its Applications*, 3rd ed. New York: John Wiley & Sons, 1957/1968, vol. 1.
- [12] R. G. Gallager, “Basic limits on protocol information in data communication networks,” *IEEE Trans. Inform. Theory*, vol. 22, no. 4, pp. 385–398, July 1976.
- [13] S. I. Gel’fand and M. S. Pinsker, “Coding for channel with random parameters,” *Prob. Contr. and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [14] V. Giovanetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, “Classical capacity of the lossy bosonic channel: the exact solution,” *Phys. Rev. Lett.*, vol. 92, no. 2, p. 027902(4), 2004.
- [15] T. S. Han, “An information-spectrum approach to capacity theorems for the general multiple-access channel,” *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2773–2795, 1998.
- [16] M. B. Hastings, “Superadditivity of communication capacity using entangled inputs,” *Nature Physics*, vol. 5, pp. 255–257, 2009.
- [17] M. Hayashi, “Error exponent in asymmetric hypothesis testing and its application to classical-quantum channel coding,” *Phys. Rev. A*, vol. 76, no. 062301, 2007.
- [18] M. Hayashi and H. Nagaoka, “General formulas for capacity of classical-quantum channels,” *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1753–1768, July 2003.
- [19] C. D. Heegard and A. A. El Gamal, “On the capacity of computer memory with defects,” *IEEE Trans. Inform. Theory*, vol. 29, no. 5, pp. 731–739, Sept. 1983.
- [20] A. S. Holevo, “The capacity of the quantum channel with general signal states,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 269–273, Jan. 1998.

- [21] Y. Kabanov, “The capacity of a channel of the Poisson type,” *Theory of Probability and Its Appl.*, vol. 23, pp. 143–147, 1978.
- [22] T. T. Kadota, M. Zakai, and J. Ziv, “Mutual information of the white Gaussian channel with and without feedback,” *IEEE Trans. Inform. Theory*, vol. 17, no. 4, pp. 368–371, July 1971.
- [23] A. V. Kusnetsov and B. S. Tsybakov, “Coding in a memory with defective cells,” *Probl. Peredachi Inf.*, vol. 10, no. 2, Apr.–June 1974.
- [24] A. Lapidoth, “On the reliability function of the ideal Poisson channel with noiseless feedback,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 491–503, Mar. 1993.
- [25] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
- [26] —, “On the capacity of the discrete-time Poisson channel,” *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 303–322, Jan. 2009.
- [27] A. Lapidoth and P. Narayan, “Reliable communication under channel uncertainty,” *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
- [28] A. Martinez, “Low-signal-energy asymptotics of capacity and mutual information for the discrete-time poisson channel,” Aug. 2008, subm. to *IEEE Trans. Inform. Theory*.
- [29] M. Mosonyi and N. Datta, “Generalized relative entropies and the capacity of classical-quantum channels,” 2008. [Online]. Available: <http://arxiv.org/abs/0810.3478>
- [30] M. A. Nielsen and I. L. Chuang, *Quantum Computation and quantum information*. Cambridge University Press, 2000.
- [31] T. Ogawa and H. Nagaoka, “Strong converse and Stein’s lemma in quantum hypothesis testing,” *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2428–2433, Nov. 2000.
- [32] —, “A new proof of the channel coding theorem via hypothesis testing in quantum information theory,” 2002. [Online]. Available: <http://arxiv.org/abs/quant-ph/0208139>

- [33] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [34] R. Renner, “Security of QKD,” Ph.D. thesis, Swiss Federal Institute of Technology, Zurich, 2005. [Online]. Available: <http://arxiv.org/abs/quant-ph/0512258>
- [35] H. L. Royden, *Real Analysis*, 3rd ed. Macmillan Publishing Company, 1988.
- [36] I. Rubin, “Information rates and data-compression schemes for Poisson processes,” *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 200–210, Mar. 1974.
- [37] B. Schumacher and M. D. Westmoreland, “Sending classical information via noisy quantum channels,” *Phys. Rev. A*, vol. 56, pp. 131–138, 1997.
- [38] S. Shamai (Shitz), “Capacity of a pulse amplitude modulated direct detection photon channel,” in *Proc. IEE*, vol. 137, pt. I (Communications, Speech and Vision), no. 6, Dec. 1990, pp. 424–430.
- [39] C. E. Shannon, “A mathematical theory of communication,” *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.
- [40] —, “Channels with side information at the transmitter,” *IBM J. Research and Development*, vol. 2, pp. 289–293, 1958.
- [41] Y. Steinberg, “New converses in the theory of identification via channels,” *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 984–998, May 1998.
- [42] S. Vembu, S. Verdú, and Y. Steinberg, “The source-channel separation theorem revisited,” *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995.
- [43] V. Venkatesan, “On low power capacity of the Poisson channel,” Master’s thesis, Signal and Inform. Proc. Lab., ETH Zurich, Switzerland, Apr. 2008, supervised by Prof. Dr. Amos Lapidoth, Ligong Wang.
- [44] S. Verdú, “On channel capacity per unit cost,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.

-
- [45] —, “The exponential distribution in information theory,” *Problems of Information Transmission*, vol. 31, no. 1, pp. 86–95, 1996.
- [46] S. Verdú and T. S. Han, “A general formula for channel capacity,” *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.
- [47] A. D. Wyner, “Capacity and error exponent for the direct detection photon channel — part I,” *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.
- [48] —, “Capacity and error exponent for the direct detection photon channel — part II,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 1462–1471, Nov. 1988.
- [49] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

About the Author

Ligong Wang was born in Qinhuangdao, Hebei Province, China, on 17 April 1982.

He attended primary school and high school in Beijing. During his school years he developed great interests in mathematics and physics, and took part in many competitions on these subjects.

He entered Tsinghua University, Beijing, in 2000 to study Electronic Engineering. There he obtained his bachelor's degree in 2004.

In October 2004, he entered ETH Zurich as a master student. At ETH he became interested in information theory and wrote his master thesis in this field. He obtained his master's degree in 2006.

In July 2006 he joined the Signal and Information Processing Laboratory at ETH as a PhD student to further work on information theory under supervision of Prof. Amos Lapidoth. Fascinated by quantum physics, he also worked on projects in quantum information theory with Prof. Renato Renner (Institute of Theoretical Physics, ETH Zurich).

He loves reading and classical music. He has sung in several choirs and practices solo-singing regularly. He was a regular visitor to the Tonhalle and the Opernhaus in Zurich, and the KKL in Lucerne, often purchasing last-minute tickets at reduced student-prices.

He married Tianyun Gao in 2006. Their daughter, Xinyuan, was born in 2010.

ETH Series in Information Theory and its Applications

edited by Amos Lapidoth

- Vol. 1: Stefan M. Moser, **Duality-Based Bounds on Channel Capacity**. ISBN 3-89649-956-4
- Vol. 2: Daniel Hösl, **On the Role of the Line-of-Sight Component in Coherent MIMO Ricean Channels**. ISBN 3-86628-044-0
- Vol. 3: Michèle Angela Wigger, **Cooperation on the Multiple-Access Channel**. ISBN 3-86628-231-1
- Vol. 4: Stéphane Tinguely, **Transmitting Correlated Sources over Wireless Networks**. ISBN 3-86628-248-6
- Vol. 5: Tobias Koch, **On Heating Up and Fading in Communication Channels**. ISBN 3-86628-261-3