Multiple-User Quantum Information Theory for Optical Communication Channels

by

Saikat Guha

B. Tech., Electrical Engineering
Indian Institute of Technology Kanpur, 2002

S. M., Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2004

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2008

Author .....
Department of Electrical Engineering and Computer Science
May 23, 2008

Certified by .....
Jeffrey H. Shapiro
Julius A. Stratton Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Terry P. Orlando
Chair, Department Committee on Graduate Students

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2008		**00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Multiple-User Quantum Information Theory for Optical Communication Channels	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Massachusetts Institute of Technology,Department of Electrical Engineering and Computer Science,Cambridge,MA,02139**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
U.S. Government or Federal Rights License

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**239**	

Multiple-User Quantum Information Theory for Optical Communication Channels

by

Saikat Guha

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Research in the past decade has established capacity theorems for point-to-point bosonic channels with additive thermal noise, under the presumption of a conjecture on the minimum output von Neumann entropy. In the first part of this thesis, we evaluate the optimum capacity for free-space line-of-sight optical communication using Gaussian-attenuation apertures. Optimal power allocation across all the spatio-temporal modes is studied, in both the far-field and near-field propagation regimes. We establish the gap between ultimate capacity and data rates achievable using classical encoding states and structured receivers. The remainder of the thesis addresses the ultimate capacity of bosonic broadcast channels, i.e., when one transmitter is used to send information to more than one receiver. We show that when coherent-state encoding is employed in conjunction with coherent detection, the bosonic broadcast channel is equivalent to the classical degraded Gaussian broadcast channel whose capacity region is known. We draw upon recent work on the capacity region of the two-user degraded quantum broadcast channel to establish the ultimate capacity region for the bosonic broadcast channel, under the presumption of another conjecture on the minimum output entropy. We also generalize the degraded broadcast channel capacity theorem to more than two receivers, and prove that if the above conjecture is true, then the rate region achievable using a coherent-state encoding with optimal joint-detection measurement at the receivers would be the ultimate capacity region of the bosonic broadcast channel with loss and additive thermal noise. We show that the minimum output entropy conjectures restated for Wehrl entropy, are immediate consequences of the entropy power inequality (EPI). We then show that an EPI-like inequality for von Neumann entropy would imply all the minimum output entropy conjectures needed for our channel capacity results. We call this new conjectured result the Entropy Photon-Number Inequality (EPnI).

Thesis Supervisor: Jeffrey H. Shapiro
Title: Julius A. Stratton Professor of Electrical Engineering

Acknowledgments

This work would not have been possible without the able guidance of my supervisor Prof. Jeffrey H. Shapiro. I have yet to meet someone as meticulous, detail-oriented, rigorous and organized as Prof. Shapiro. His mentoring style has always been to urge students to find for themselves the interesting questions to answer, and to help them by steering their thought processes in the right direction, rather than predisposing them to tackle well-defined problems — a philosophy that has been key to my growth as a researcher, and will be a guiding light for me in the years to come.

I am immensely grateful to my thesis committee members Prof. Vincent Chan, Prof. Seth Lloyd and Prof. Lizhong Zheng for taking the time to read this thesis, and for providing valuable and constructive feedback on my work.

I would like to thank my present and former colleagues Dr. Baris I. Erkmen, Dr. Brent J. Yen and Dr. Mohsen Razavi for the numerous interesting dialogues we have had on a wide variety of topics, the amount I have learned from which is invaluable. I would especially like to thank Brent and Baris for patiently answering all my stupid technical questions for all these years. I thank Dr. Vittorio Giovannetti and Dr. Lorenzo Maccone, former post-doctoral scholars in our group, for all that I have learned from them. I am grateful to Dr. Dongning Guo, Assistant Professor of Electrical Engineering at Northwestern University, for the discussions on the Entropy Power Inequality. I thank Dr. Franco Wong for answering all my questions about the experiments, from which I learned a lot. I thank Prof. Seth Lloyd for many enriching discussions on a variety of topics. I really admire his zeal for research, his ever-cheerful demeanor and his superb whiteboard presentations. I thank Prof. G. David Forney for mentoring me patiently over many months while we worked on quantum convolutional codes. I owe my understanding of error correction completely to Prof. Forney. I thank Prof. Sanjoy Mitter for many interesting discussions that provided me a great deal of useful insight into the relationship between the entropy power inequality and the monotonicity of entropy.

I really enjoyed my one term as a teaching assistant for the course 6.003 (Signals

and Systems). I thank Prof. Joel Voldvan and Prof. Qing Hu for having given me the opportunity to teach tutorials and mentor students in 6.003. I also thank profusely all my erstwhile students in the class for asking me numerous questions that I would never have thought of myself. Answering their questions enriched my own understanding of the subject tremendously, and I thank them also for the brilliant feedback they gave me at the end of the term.

I am what I am because of my parents Mrs. Shikha and Dr. Shambhu Nath Guha, and no words are enough to thank them. Throughout my childhood, my father, being a physicist himself, would always give answers patiently, though very accurately, to all my naive and silly questions. I still remember the day I learned about *inertia*, when I asked him why the ceiling fan, unlike the light bulb, would not shut off immediately when I turned the switch off! It is because of my father's encouragement and support that I prepared for the Mathematics Olympiad. Even though I did not secure a place in the Indian IMO team, the preparation itself was crucial in sharpening my mathematical abilities that is an asset to me, even to this day. He later encouraged (and trained) me to participate in the Physics Olympiad, which led me to make it through all the levels of selection to the Indian IPhO team, and to secure an honorable mention at the IPhO 1998 held at Reykjavik. Apart from all the values I have learned from my mother, which still form an indelible part of my life today, I learnt from her *Sanskrit*, the beautiful ancient language of India, and one of the most scientifically structured languages in my opinion, that has ever been spoken across the world. I thank my sister Somrita, for all the fun times, laughs and fights we have shared while growing up. I am really grateful to my best friend Arindam for having been there for me all these years. Amongst many friends that I made at MIT, Debajyoti Bera and Siddharth Ray, particularly, have rendered my stay here profoundly memorable. I thank my wife's parents Mrs. Nivedita and Mr. Ashok Ghosh, and her sisters Ronita and Sorita for all their love and support. I thank Josephina Lee for many wonderful discussions we have had, and for helping me get through many things while I was at MIT.

The last one and a half years of my Ph.D., during the time that I have known

and spent with my wife Sujata, have certainly been the most extraordinary chapter of my life so far. From the fits of laughter at the most inconsequential of events, the fervent narrations of her day-to-day anecdotes, to the patient listener she has been to the countless discourses on my research, and the long and passionate discussions on an array of topics that we have had on our endless drives all over New England and elsewhere, she has unveiled a world to me that I never knew existed.

Finally, I would like to thank all the agencies that have funded my doctoral work. This research was supported at various stages by the Army Research Office, DARPA and the W. M. Keck Foundation Center for Extreme Quantum Information Theory (xQIT) at MIT.

To my wonderful wife Sujata, to whom I am indebted for all the love and support that she has given me, for every moment of my life that I have spent with her, and for every moment of our lives together that I eagerly look forward to ...

Contents

List of Figures

3-14 In order to evaluate the capacity region of the M-receiver noiseless bosonic degraded broadcast channel depicted in Fig. 3-13 using a coherent-state input alphabet $\{|\alpha\rangle\}$, $\alpha \in \mathbb{C}$ and $\langle \hat{a}^\dagger \hat{a} \rangle = \langle |\alpha|^2 \rangle \leq \bar{N}$, we choose the $M-1$ auxiliary classical Markov random variables (in Eqs. (3.35)) as complex-valued random variables T_k, $k \in \{1, \ldots, M-1\}$, taking values $\tau_k \in \mathbb{C}$. In order to visualize the postulated optimal Gaussian distributions for the random variables T_k, let us associate with T_k, a quantum system, i.e., a coherent-set alphabet $\{|\tau_k\rangle\}$ and modal annihilation operator $\hat{t}_k$, $\forall k$. In accordance with the Markov property of the random variables T_k, let $\hat{t}_{M-1}$ be in an isotropic zero-mean Gaussian mixture of coherent-states with a variance $\bar{N}$ (see Eq. (3.104)), and for $k \in \{1, \ldots, M-2\}$, let $\hat{t}_k$ be obtained from $\hat{t}_{k+1}$ by mixing it with another mode $\hat{u}_{k+1}$ excited in a zero-mean thermal state with mean photon number $\bar{N}$, through a beam splitter with transmissivity $1-\gamma_{k+1}$, as shown in the figure above, for some $\gamma_{k+1} \in (0,1)$. We complete the Markov chain $T_{M-1} \rightarrow \ldots \rightarrow T_1 \rightarrow A$, by obtaining the transmitter mode $\hat{a}$ by mixing $\hat{t}_1$ with a mode $\hat{u}_1$ excited in a zero-mean thermal state with mean photon number $\bar{N}$, through a beam splitter with transmissivity $1-\gamma_1$, for $\gamma_1 \in (0,1)$. The above setup of the auxiliary modes gives rise to the distributions given in Eqs. (3.104), which we use to evaluate the achievable rate region of the M-receiver bosonic broadcast channel using coherent-state encoding. 101

Chapter 1

Introduction

The objective of any communication system is to transfer information from one point to another efficiently, given the constraints on the available physical resources. In most communication systems, the transfer of information is done by superimposing the information onto an electromagnetic (EM) wave. The EM wave is known as the *carrier* and the process of superimposing information onto the carrier wave is known as *modulation*. The modulated carrier is then transmitted to the destination through a noisy medium, called the *communication channel*. At the receiver, the noisy wave is received and *demodulated* to retrieve the information as accurately as possible. Such systems are often characterized by the location of the carrier wave's frequency within the electromagnetic spectrum. In radio systems for example, the carrier wave is selected from the radio frequency (RF) portion of the spectrum.

In an optical communication system, the carrier wave is selected from the optical range of frequencies, which includes the infrared, visible light, and ultraviolet frequencies. The main advantage of communicating with optical frequencies is the potential increase in information that can be transmitted because of the possibility of harnessing an immense amount of bandwidth. The amount of information transmitted in any communication system depends directly on the bandwidth of the modulated carrier, which is usually a fraction of the carrier wave's frequency. Thus increasing the carrier frequency increases the available transmission bandwidth. For example, the frequencies in the optical range would typically have a usable transmission band-

width about three to four orders of magnitude greater than that of a carrier wave in the RF region. Another important advantage of optical communications relative to RF systems comes from their narrower transmitted beams — μRad beam divergences are possible with optical systems. These narrower beamwidths deliver power more efficiently to the receiver aperture. Narrow beams also enhance communication security by making it hard for an eavesdropper to intercept an appreciable amount of the transmitted power. Communicating with optical frequencies has some challenges associated with it as well. As optical frequencies are accompanied by extremely small wavelengths, the design of optical components require completely different techniques than conventional microwave or RF communication systems. Also, the advantage that optical communication derives from its comparatively narrow beam introduces the need for high-accuracy beam pointing. RF beams require much less pointing accuracy. Progress in the theoretical study of optical communication, the advent of *laser* - a high-power optical carrier source, the developments in the field of optical fiber-based communication, and the development of novel wideband optical modulators and efficient detectors, have made optical communication emerge as a field of immense technological importance [1].

The field of information theory, which was born from Claude Shannon's revolutionary 1948 paper [2], addresses ultimate limits on data compression and communication rates over noisy communication channels. It tells us how to compute the maximum rate at which reliable data communication can be achieved over a noisy communication channel by appropriately encoding and decoding the data. This ultimate data rate is known as the *channel capacity* [2, 3, 4]. Information theory also tells us how to compute the maximum extent a given set of data can be compressed so that the original data can be recovered within a specified amount of tolerable distortion level. Unfortunately, information theory does not give us the exact algorithm (or the optimal code) that would achieve capacity on a given channel, nor does it tell us how to optimally compress a given set of data. Nevertheless, it sets ultimate limits on communication and data compression that are essential to meaningfully determine how well a real system is actually performing.

The performance of communication systems that rely on electromagnetic wave propagation are ultimately limited by noise of quantum-mechanical origin. Moreover, high-sensitivity photodetection systems have long been close to this noise limit. Hence determining the ultimate capacities of lasercom channels is of immediate relevance. Much work has already been done on quantum information theory [5, 6], which sets ultimate limits on the rates of reliable communication of classical information and quantum information over quantum communication channels. As in classical information theory, quantum information theory does not tell us the transmitter and receiver structures that would achieve the best communication rates for specific forms of quantum noise. Nevertheless, the limits set by quantum information theory are extremely useful in determining the degree to which available technology can approach the ultimate performance bounds.

The most famous classical channel capacity formula is Shannon's result for the classical additive white Gaussian noise channel. For a complex-valued channel model in which we transmit a and receive $c = \sqrt{\eta}\, a + \sqrt{1-\eta}\, b$, where $0 < \eta < 1$ is the channel's transmissivity and b is a zero-mean, isotropic, complex-valued Gaussian random variable that is independent of a, Shannon's capacity is

$$C_{\text{classical}} = \ln[1 + \eta\bar{N}/(1-\eta)N] \text{ nats/use}, \tag{1.1}$$

when $E(|a|^2) \leq \bar{N}$ and $E(|b|^2) = N$.

The lossy bosonic channel provides a quantum model for optical communication systems that rely on fiber or free-space propagation. In this quantum channel model, we control the state of an electromagnetic mode with photon annihilation operator $\hat{a}$ at the transmitter, and receive another mode with photon annihilation operator $\hat{c} = \sqrt{\eta}\, \hat{a} + \sqrt{1-\eta}\, \hat{b}$, where $\hat{b}$ is the annihilation operator of a noise mode that is in a zero-mean, isotropic, complex-valued Gaussian state. For lasercom, if quantum measurements corresponding to ideal optical homodyne or heterodyne detection are employed at the receiver, this quantum channel reduces to a real-valued (homodyne) or complex-valued (heterodyne) additive Gaussian noise channel, from which the

following capacity formulas (in nats/use) follow:

$$C_{\text{homodyne}} = \frac{1}{2}\ln[1 + 4\eta\bar{N}/(2(1-\eta)N + 1)] \tag{1.2}$$

$$C_{\text{heterodyne}} = \ln[1 + \eta\bar{N}/((1-\eta)N + 1)], \tag{1.3}$$

where $\langle\hat{a}^\dagger\hat{a}\rangle \leq \bar{N}$ and $\langle\hat{b}^\dagger\hat{b}\rangle = N$, with angle brackets used to denote quantum averaging. The +1 terms in the noise denominators are quantum contributions, so that even when the noise mode $\hat{b}$ is unexcited these capacities remain finite, unlike the situation in Eq. (1.1).

The classical capacity of the pure-loss bosonic channel—in which the $\hat{b}$ mode is unexcited ($N = 0$)—was shown in [7] to be $C_{\text{pure-loss}} = g(\eta\bar{N})$ nats/use, where $g(x) \equiv (x+1)\ln(x+1) - x\ln(x)$ is the Shannon entropy of the Bose-Einstein probability distribution with mean x. This capacity exceeds the $N = 0$ versions of Eqs. (1.2) and (1.3), as well as the best known bound on the capacity of ideal optical direct detection [8]. For this pure-loss case, capacity has been shown to be achievable using single-use coherent-state encoding with a Gaussian prior density [7]. The ultimate capacity of the thermal-noise ($N > 0$) version of this channel is bounded below by $C_{\text{thermal}} \geq g(\eta\bar{N} + (1-\eta)N) - g((1-\eta)N)$, and this bound was shown to be the capacity if the thermal channel obeyed a certain minimum output entropy conjecture [9]. This conjecture states that the von Neumann entropy at the output of the thermal channel is minimized when the $\hat{a}$ mode is in its vacuum state. Considerable evidence in support of this conjecture has been accumulated [10], but it has yet to be proven. Nevertheless, the preceding lower bound already exceeds Eqs. (1.2) and (1.3) as well as the best known bounds on the capacity of direct detection [8].

Less is known about the classical-information capacity of multi-user bosonic channels. For multiple-access bosonic communications—in which two or more senders communicate to a common receiver over a shared propagation medium—single-use coherent-state encoding with a Gaussian prior and optimum measurement achieves the sum-rate capacity, but it falls short of achieving the ultimate capacity in the "corner regions" [11]. Moreover, the capacity region that is lost when coherent de-

tection is employed instead of the optimum measurement has been quantified for this multiple-access channel. In this thesis we will report our capacity analysis for the bosonic broadcast channel. As we described in [12], this work led to an inner bound on the capacity region, which we showed to be the capacity region under the presumption of a second minimum output entropy conjecture. Both of these minimum output entropy conjectures have been proven if the input states are restricted to be Gaussian, and, as we will describe later in this thesis, we have shown them to be equivalent under this input-state restriction. We will also show that the second conjecture will establish the privacy capacity of the lossy bosonic channel, as well as its ultimate quantum information carrying capacity [13].

The Entropy Power Inequality (EPI) from classical information theory is widely used in coding theorem converse proofs for Gaussian channels. By analogy with the EPI, we conjecture its quantum version, viz., the Entropy Photon-number Inequality (EPnI). We will show that the two minimum output entropy conjectures cited above are simple corollaries of the EPnI. Hence, proving the EPnI would immediately establish some key capacity results for the capacities of bosonic communication channels [13].

We will assume that the reader has had some prior acquaintance with quantum mechanics, quantum optics and information theory. We will use standard notation widely in use in the quantum optics and information theory literature. For a quick summary of the background material and notation, see Appendix A. Chapter 2 of this thesis reviews some of our early work on the single-mode bosonic channel capacity, and describes capacity calculations for the free-space optical channel using Gaussian-attenuation transmitter and receiver apertures. Chapter 3 starts with a brief introduction to the capacity of classical discrete memoryless broadcast channels and then walks the reader through the classical-information capacity analysis for the bosonic broadcast channel in which a single sender communicates to two or more receivers through a lossless optical beam splitter with no extra noise or with additive thermal noise. We prove the ultimate classical information capacities of the bosonic broadcast channel subject to the minimum output entropy conjectures elucidated in

Chapter 4. In that chapter we describe three conjectures on the minimum output entropy of bosonic channels, none of which have yet been proven. Proving these conjectures would, respectively, complete the proofs of the ultimate channel capacity of the lossy bosonic channel with additive thermal noise, the ultimate capacity region of the the multiple-user bosonic broadcast channel with no extra noise, and that of the bosonic broadcast channel with additive thermal noise. Chapter 5 begins with motivating the thought process that led us to conjecture the quantum version of the Entropy Power Inequality (EPI), which we call the Entropy Photon-number Inequality (EPnI). There we show that the EPnI subsumes all the minimum output entropy conjectures described in Chapter 4. We also discuss some recent progress made towards a proof of the EPnI. The rest of Chapter 5 delves briefly into some interesting problems in the area of quantum optical information theory, including the additivity properties of quantum information theoretic quantities, a quantum version of the central limit theorem, and a conjecture on the monotonicity of quantum entropy. Chapter 6 concludes the thesis with remarks on the major open problems ahead of us in the theory of bosonic communications and comments on lines of future work in this area.

Chapter 2

Point-to-point Bosonic Communication Channel

2.1 Background

Reliable, high data rate communication—carried by electromagnetic waves at microwave to optical frequencies—is an essential ingredient of our technological age. Information theory seeks to delineate the ultimate limits on reliable communication that arise from the presence of noise and other disturbances, and to establish means by which these limits can be approached in practical systems. The mathematical foundation for this assessment of limits is Shannon's Noisy Channel Coding Theorem [2], which introduced the notion of channel capacity—the maximum mutual information between a channel's input and output—as the highest rate at which error-free communication could be maintained. Textbook treatments of channel capacity [4],[3] study channel models—ranging from the binary symmetric channel's digital abstraction to the additive white-Gaussian-noise channel's idealization of thermal-noise-limited waveform transmission—for which classical physics is the underlying paradigm. Fundamentally, however, electromagnetic waves are quantum mechanical, i.e., they are boson fields [14],[15]. Moreover, high-sensitivity photodetection systems have long been limited by noise of quantum mechanical origin [16]. Thus it would seem that determining the ultimate limits on optical communication would necessarily involve

an explicitly quantum analysis, but such has not been the case. Nearly all work on the communication theory of optical channels—viz., that done for systems with laser transmitters and either coherent-detection or direct-detection receivers—uses semiclassical (shot-noise) models (see, e.g., [1],[17]). Here, electromagnetic waves are taken to be classical entities, and the fundamental noise is due to the random release of discrete charge carriers in the process of photodetection. Inasmuch as the quantitative results obtained from shot-noise analyses of such systems are known to coincide with those derived in rigorous quantum-mechanical treatments [18], it might be hoped that the semiclassical approach would suffice. But, Helstrom's derivation [19] of the optimum quantum receiver for binary coherent-state (laser light) signaling demonstrated that the lowest error probability, at constant average photon number, required a receiver that was neither coherent detection nor direct detection. That Dolinar [20] was able to show how Helstrom's optimum receiver could be realized with a photodetection feedback system which admits to a semiclassical analysis did not alleviate the need for a fully quantum-mechanical theory of optical communication, as Shapiro et al. [21] soon proved that even better binary-communication performance could be obtained by use of two-photon coherent state (now known as squeezed state) light, for which semiclassical photodetection theory did not apply.

In quantum mechanics, the state of a physical system together with the measurement that is made on that system determine the statistics of the outcome of that measurement, see, e.g., [14]. Thus in seeking the classical information capacity of a bosonic channel, we must allow for optimization over *both* the transmitted quantum states *and* the receiver's quantum measurement. In particular, it is *not* appropriate to immediately restrict consideration to coherent-state transmitters and coherent-detection or direct-detection receivers. Imposing these structural constraints leads to Gaussian-noise (Shannon-type) capacity formulas for coherent (homodyne and heterodyne) detection [22] and a variety of Poisson-noise capacity results (depending on the power and/or bandwidth constraints that are enforced) for shot-noise-limited direct detection [8, 23, 24, 25, 26]. None of these results, however, can be regarded as specifying the ultimate limit on reliable communication at optical frequencies. What is

needed for deducing the fundamental limits on optical communication is the analog of Shannon's Noisy Channel Coding Theorem—free of unjustified structural constraints on the transmitter and receiver—that applies to transmission of classical information over a noisy quantum channel, viz., the Holevo-Schumacher-Westmoreland (HSW) Theorem [27, 28, 29].

Until recently, little had been done to address the classical information capacity of bosonic quantum channels. As will be seen below, the HSW Theorem renders quantum measurement optimization an implicit—rather than explicit—part of capacity determination, and confronts a superadditivity property that is absent from classical Shannon theory. Prior to this theorem—and well after its proof—about the only bosonic channel whose classical information capacity had been determined was the lossless channel [30, 31], in which the field modes (with annihilation operators $\{\hat{a}_j\}$) controlled by the transmitter are available for measurement (without loss, hence without additional quantum noise) at the receiver. This situation changed dramatically when we obtained the capacity of the pure-loss channel [7], i.e., one in which photons may be lost en route from the transmitter to the receiver while incurring the minimal additional quantum noise required to preserve the Heisenberg uncertainty relation. We then considered active channel models—in which noise photons are injected from an external environment or the signal is amplified with unavoidable quantum noise—obtaining upper and lower bounds on the resulting channel capacities, which are asymptotically tight at low and high noise levels [9]. [We conjectured that our lower bounds are in fact the capacities, but we have yet to prove that assertion.] Collectively, the preceding channel models can represent line-of-sight free-space optical communications (see [7],[9]) and loss-limited fiber-optic communications with or without pre-detection optical amplification. Furthermore, the classical-noise channel—in which optical amplification is used to balance the attenuation due to free-space diffraction or fiber propagation—is the quantum analog of Shannon's additive white-Gaussian-noise channel, thus its capacity is especially interesting in comparison to Shannon's well-known formula.

For the pure-loss case, it turns out that capacity is achievable with coherent-state

(laser light) encoding, but a multi-symbol quantum measurement (a joint measurement over entire codewords) is required. Heterodyne detection is asymptotically optimum in the limit of large average photon number for single-mode operation [7]. The same is true in the limit of high average power level for wideband operation over the far-field free space channel [7],[9]. However, all coherent reception techniques fall short of the HSW Theorem capacity for the pure-loss channel in photon/power starved scenarios such as deep space communication. We show later in this chapter that at very low photon numbers per mode, the direct detection receiver along with a coherent-state on-off-keying modulation can achieve data rates very close to the ultimate capacity. For these applications it becomes especially important to find practical ways to reap the capacity advantage that multi-symbol quantum measurement affords. In the remainder of this chapter we review the results we have obtained so far, towards developing these approaches, and applying them, to the thermal-noise and classical-noise channels, and as well as to broadcast channels.

Section 2.2 provides a quick summary of bosonic channel models and the HSW theorem. Section 2.3 presents our capacity results for the point-to-point single-mode channels. Section 2.4 then addresses multiple spatio-temporal modes of the free-space optical channel using Gaussian apertures, something that is easily analyzed by tensoring up a collection of single-mode models. Finally, section 2.5 presents our capacity results for modulation schemes using coherent-state codewords that are geared towards achieving high data rates at very low input power regimes.

2.2 Bosonic communication channels

We are interested in the classical communication capacities of point-to-point bosonic channels with additive quantum Gaussian noise and practical means for communicating at rates approaching these capacities. The three main categories of point-to-point bosonic channels that we describe below are, the lossy channel, the amplifying channel, and the classical-noise channel. For each single-mode channel, the transmitter Alice (A) sends out an electromagnetic-field mode with annihilation operator $\hat{a}$ and

the output is received by the receiver Bob (B), which is another field mode with annihilation operator $\hat{b}$. The channels of interest are *not* unitary evolutions, so they are all governed by TPCP maps that relate their output density operators, $\hat{\rho}^A$, to their input density operators, $\hat{\rho}^B$.

2.2.1 The lossy channel

The TPCP map $\mathcal{E}_\eta^N(\cdot)$ for the single-mode lossy channel can be derived from the commutator preserving beam splitter relation

$$\hat{b} = \sqrt{\eta}\,\hat{a} + \sqrt{1-\eta}\,\hat{e}, \tag{2.1}$$

in which the annihilation operator $\hat{e}$ is associated with an environmental (noise) quantum system E, and $0 \leq \eta \leq 1$ is the channel transmissivity. [See [32] for how this single-mode map leads to the quantum version of the Huygens-Fresnel diffraction integral, and for a quantum characteristic function specification of its associated TPCP map.] For the pure-loss channel, the $\hat{e}$ mode is in its vacuum state; for the thermal-noise channel this mode is in a thermal state, viz., an isotropic-Gaussian mixture of coherent states with average photon number $N > 0$,

$$\hat{\rho}^E = \int \frac{\exp(-|\mu|^2/N)}{\pi N} |\mu\rangle\langle\mu| \mathrm{d}^2\mu. \tag{2.2}$$

2.2.2 The amplifying channel

The TPCP map $\mathcal{A}_\kappa^M(\cdot)$ for the single-mode amplifying channel can be derived from the commutator-preserving phase-insensitive amplifier relation [33]

$$\hat{b} = \sqrt{\kappa}\,\hat{a} + \sqrt{\kappa-1}\,\hat{e}^\dagger, \tag{2.3}$$

where $\hat{e}$ is now the modal annihilation operator for the noise introduced by the amplifier and $\kappa \geq 1$ is the amplifier gain. This amplifier injects the minimum possible noise when the $\hat{e}$-mode is in its vacuum state; in the excess-noise case this mode's

density operator is the isotropic-Gaussian coherent-state mixture (2.2).

2.2.3 The classical-noise channel

The classical-noise channel can be viewed as the cascade of a pure-loss channel $\mathcal{E}_\eta^0$ followed by a minimum-noise amplifying channel $\mathcal{A}_\kappa^0$ whose gain exactly compensates for the loss, $\kappa = 1/\eta$. Then, with $\eta = 1/(M+1)$, we obtain the following TPCP map for the classical-noise channel,

$$\hat{\rho}^B = \mathcal{N}_M(\hat{\rho}^A) \equiv \int \frac{\exp(-|\mu|^2/M)}{\pi M} \hat{D}(\mu)\hat{\rho}^A\hat{D}^\dagger(\mu)\mathrm{d}^2\mu, \tag{2.4}$$

where $\hat{D}(\mu)$ is the displacement operator, i.e., $\hat{b} = \hat{a} + m$ where m is a zero-mean, isotropic Gaussian noise with variance given by $\langle|m|^2\rangle = M$, so that this channel is the quantum version of the additive white-Gaussian-noise channel.

2.3 Point-to-point, Single-Mode Channels

Let us begin with a brief survey of recent work on the capacity of the point-to-point single-mode bosonic communication channel, done by various members of our research group at MIT, led by Prof. J. H. Shapiro. The details appeared in several published articles (viz. [10], [7],[9], [11], and [34]). The capacity of the single-mode, pure-loss channel (2.1), whose transmitter is constrained to use no more than $\bar{N}$ photons on average in a single use of the channel, is given by

$$C = g(\eta\bar{N}) \,\text{nats/use}, \tag{2.5}$$

where

$$g(x) \equiv (x+1)\ln(x+1) - x\ln(x) \tag{2.6}$$

is the Shannon entropy of the Bose-Einstein probability distribution with mean x. This capacity is achieved by single-use random coding over coherent states using an isotropic Gaussian distribution which meets the bound on the average number of

transmitted photons per use of the channel. [Note that the optimality of single-use encoding means that the capacity of the single-mode pure-loss channel is *not* super-additive.] This capacity exceeds what is achievable with homodyne and heterodyne detection,

$$C_{\rm hom} = \frac{1}{2}\ln(1+4\eta\bar{N}) \quad \text{and} \quad C_{\rm het} = \ln(1+\eta\bar{N}), \tag{2.7}$$

although heterodyne detection is asymptotically optimal as $\bar{N} \to \infty$. The direct-detection capacity $C_{\rm dir}$ obtained by using a coherent-state encoding and photon-counting measurement is not known. $C_{\rm dir}$ has been shown to satisfy [35],

$$C_{\rm dir} \leq \frac{1}{2}\ln(\eta\bar{N}) + o(1) \quad \text{and} \quad \lim_{\bar{N}\to\infty}(C_{\rm dir}) = \frac{1}{2}\ln(\eta\bar{N}), \tag{2.8}$$

and so is dominated by (2.5) for $\ln(\eta\bar{N}) > 1$. The best known bounds to the direct-detection capacity have recently been evaluated by Martinez [8], who has shown that tight lower bounds (achievable rates) to the direct-detection capacity can be obtained by constraining the input distribution to be a gamma density with parameter ν. For instance, a lower bound that is obtained with a gamma density input distribution with $\nu = 1$ is given by

$$C_{\rm dir} \geq (1+\eta\bar{N})\ln(1+\eta\bar{N}) + \int_0^1 \frac{\eta^2\bar{N}^2(1-u)}{1+\eta\bar{N}(1-u)}\frac{\mathrm{u}}{\ln u} - \eta\bar{N}\gamma_e, \tag{2.9}$$

where $\gamma_e = 0.5772\ldots$ is the Euler's constant. The best known upper bound to the direct-detection capacity is given by [8]:

$$C_{\rm dir} \leq \left(\frac{1}{2}+\eta\bar{N}\right)\ln\left(\frac{1}{2}+\eta\bar{N}\right) - \eta\bar{N}\ln(\eta\bar{N}) - \frac{1}{2} + \ln\left(1+\frac{\sqrt{2e}-1}{\sqrt{1+2\eta\bar{N}}}\right). \tag{2.10}$$

Employing the pure-loss channel's optimal random code ensemble over the thermal-noise, amplifying, and classical-noise channels leads to the following lower bounds on

their channel capacities:

$$C \geq \begin{cases} g(\eta\bar{N} + (1-\eta)N) - g((1-\eta)N) & \text{thermal-noise channel} \\ g(\kappa\bar{N} + (\kappa-1)(N+1)) - g((\kappa-1)(N+1)) & \text{amplifying channel} \\ g(\bar{N} + M) - g(M) & \text{classical-noise channel} \end{cases} \tag{2.11}$$

which was conjectured to be their capacities [9]. The proof of that conjecture is intimately related to the problem of determining the minimum von Neumann entropies that can be realized at the output of these channels by choice of their input states. In particular, showing that coherent-state inputs are the entropy-minimizing input states would complete the proof of the capacity conjecture stated above, and lower bounds on the minimum output entropies immediately imply upper bounds on the corresponding channel capacities. So far, among many other things, it is known that coherent-state inputs lead to *local* minima in the output entropies, and we have a suite of output-entropy lower bounds for single-use encoding over the thermal-noise and classical-noise channels. We also know that coherent-state inputs minimize the integer-order Rényi output entropies [34],[36], from which a proof of our capacity conjecture would follow were a rigorous foundation available for the replica method of statistical mechanics, see, e.g., [37, 38] for recent classical-communication applications of the replica method. As additional evidence towards the conjecture, we collected numerical evidence supporting a stronger version of the conjecture, that the output-state of the bosonic channels for a vacuum-state input majorizes all other output states. Our further quest into the theory of bosonic multiple-user communication has led us to propose two new conjectures on the minimum von Neumann entropy at the output of bosonic channels. Our three minimum output-entropy conjectures are elaborated in Chapter 4. Proving **conjecture 1** would prove the capacity of the single-user bosonic channel with additive thermal noise. Proving **conjecture 2** would prove the ultimate capacity region of the M-user bosonic broadcast channel with vacuum-state noise. Proving **conjecture 3** would prove the ultimate capacity region of the M-user bosonic broadcast channel with additive thermal noise. As

evidence supporting our conjectures, we prove the Wehrl entropy versions of the conjectures. Also, in the thesis, we will prove that if we restrict our optimization only to Gaussian states, then the minimum output entropy conjectures 2 and 3 are both true. The proof of the Gaussian-state version of conjecture 1 appeared in [10]. In Chapter 5 we will report the quantum version of the Entropy Power Inequality, viz., the Entropy Photon-number Inequality (EPnI), and we will show that the minimum output entropy conjectures cited above can be derived as simple special cases of the EPnI. Hence, proving the EPnI would immediately establish some key capacity results for the capacities of bosonic communication channels [13].

2.4 Multiple-Spatial-Mode, Pure-Loss, Free-Space Channel

As an explicit example of the mean-energy constrained, pure-loss channel, we now treat the case of free-space optical communication. My SM thesis [39] treated the wideband pure-loss channel with frequency-independent loss. Despite its providing insight into multi-mode capacity, this analysis does not necessarily pertain to a realistic scenario. In [39] we also studied the far-field, scalar free-space channel in which line-of-sight propagation of a single polarization occurs over an L-m-long path from a circular transmitter pupil (area A_t) to a circular receiver pupil (area A_r) with the transmitter restricted to use frequencies $\{\,\omega : 0 \leq \omega \leq \omega_c \ll \omega_0 \equiv 2\pi cL/\sqrt{A_t A_r}\,\}$. This frequency range is the far-field power transfer regime, wherein there is only a single spatial mode that couples appreciable power from the transmitter pupil to the receiver pupil, and its transmissivity at frequency ω is $\eta(\omega) = (\omega/\omega_0)^2 \ll 1$. Figure 2-1 shows the geometry, the power allocations versus frequency for heterodyne, homodyne, and optimal reception, and their corresponding capacities versus transmitted power normalized by $P_0 \equiv 2\pi\hbar c^2 L^2/A_t A_r$, when only this dominant spatial mode is employed [7]. Far-field, free-space transmissivity increases as ω^2, thus high frequencies are used preferentially for this channel because the transmissivity

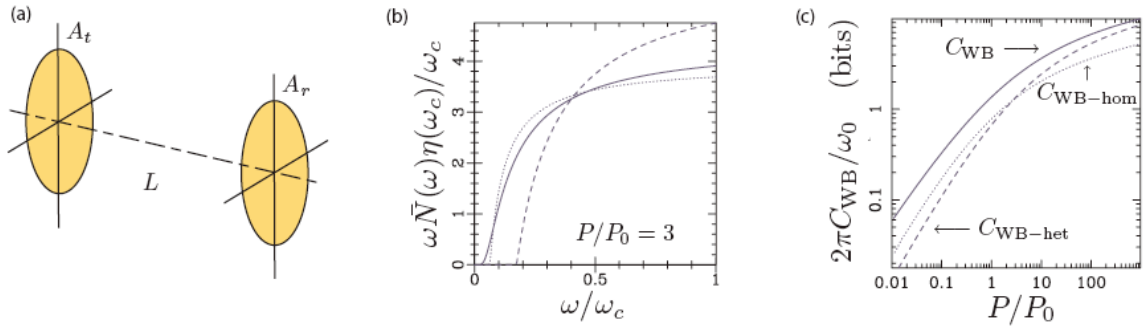


Figure 2-1: Capacity results for the far-field, free-space, pure-loss channel: (a) propagation geometry; (b) capacity-achieving power allocations $\hbar\omega\bar{N}(\omega)$ versus frequency ω for heterodyne (dashed curve), homodyne (dotted curve), and optimal reception (solid curve), with ω_c and $\hbar\omega_c/\eta(\omega_c)$ being used to normalize the frequency and the power-spectra axes, respectively; and (c) wideband capacities of optimal, homodyne, and heterodyne reception versus transmitter power P, with $P_0 \equiv 2\pi\hbar c^2 L^2/A_t A_r$ used for the reference power.

advantage of high-frequency photons more than compensates for their higher energy consumption.

We also explored the near-field behavior of the pure-loss free-space channel [40], by employing the full prolate-spheroidal wave function normal-mode decomposition associated with the propagation geometry shown in Fig. 2-1(a) [41, 42]. Near-field propagation at frequency $\omega = 2\pi c/\lambda$ prevails when $D_f = A_t A_r/(\lambda L)^2$, the product of the transmitter and receiver Fresnel numbers, is much greater than unity. In this case there are approximately D_f spatial modes with near-unity transmissivities, with all other modes affording insignificant power transfer from the transmitter pupil to the receiver pupil.

We also sketched out a general wideband capacity analysis for the free-space channel in [39], which applies when neither the far-field nor the near-field assumptions may be made for the entire channel spectrum. At very low frequencies the channel looks like the far-field channel we analyzed earlier, in which the channel transmissivity $\eta(\omega) \propto \omega^2$. So in that region, we expect that the optimal power allocation uses high frequency photons preferentially, and that the power goes to zero at low frequencies. At higher frequencies, the channel is closer to a lossless wideband channel we con-

sidered earlier, for which we know that the optimal power allocation goes to zero at very high frequencies [39]. So, in the ultra wideband case, we would expect the power allocation to vanish both for very low and very high frequencies. This intuition is validated later in this section.

The actual capacity calculation for the general wideband free-space channel for the hard circular-apertures case is difficult owing to the complicated nonlinear dependence of modal transmissivity on center frequency of transmission, for which closed-form expressions are not available. In [43], we took another approach to the wideband capacity of the pure-loss free-space channel, by employing either the Hermite-Gaussian (HG) or Laguerre-Gaussian (LG) mode sets that are associated with the soft-aperture (Gaussian-attenuation pupil) version of the Fig. 2-1(a) propagation geometry. Two benefits are derived from this approach. First, closed-form expressions become available for the modal transmissivities, as opposed to the hard-aperture case [Fig. 2-1(a)], for which numerical evaluations or analytical approximations must be employed. Second, the LG modes have been the subject of a great deal of interest, in the quantum optics and quantum information communities [44], owing to their carrying orbital angular momentum. Thus it was germane to explore whether they conferred any special advantage in regards to classical information transmission. As we shall describe, in the next subsection, the modal transmissivities of the LG modes are isomorphic to those of the HG modes. Inasmuch as the latter do not convey orbital angular momentum, it is clear that such conveyance is not essential to capacity-achieving classical communication over the pure-loss free-space channel. After this, we will compute the classical capacity of the general wideband free-space channel with soft apertures, and will describe the scheme for doing optimal power-allocation across spatio-temporal modes of the quantized optical field to achieve the ultimate rate limits afforded by coherent-state encoding with both conventional coherent detectors and that with the optimum joint-detection quantum measurement.

2.4.1 Propagation Model: Hermite-Gaussian and Laguerre-Gaussian Mode Sets

In lieu of the hard-aperture propagation geometry from Fig. 2-1(a), wherein the transmitter and receiver pupils are perfectly transmitting apertures within otherwise opaque planar screens, we now introduce the soft-aperture propagation geometry of Fig. 2-2. From the quantum version of scalar Fresnel diffraction theory [32], we know that it is sufficient, insofar as this propagation geometry is concerned, to identify a complete set of monochromatic spatial modes, for a single electromagnetic polarization of frequency $\omega = 2\pi c/\lambda = ck$, that maintain their orthogonality when transmitted through this channel. The resulting input and output mode sets constitute a singular-value decomposition (SVD) of the linear propagation kernel (spatial impulse response) associated with this geometry, which we will now develop.

Let $u_i(\vec{x}\,)$, for $\vec{x}$ a 2D vector in the transmitter's exit-pupil plane, denote a frequency-ω field entering the transmitter pupil that is normalized to satisfy

$$\int \mathrm{d}^2\vec{x}\ |u_i(\vec{x}\,)|^2 = 1. \tag{2.12}$$

After masking of the field by Gaussian intensity transmitter and receiver apertures, and undergoing free-space Fresnel diffraction over an L-m-long path, the field immediately after the receiver pupil is given by

$$u_o(\vec{x}\,') = \int \mathrm{d}^2\vec{x}\ u_i(\vec{x}\,)h(\vec{x}\,',\vec{x}\,), \tag{2.13}$$

where

$$h(\vec{x}\,',\vec{x}\,) \equiv \exp(-|\vec{x}\,'|^2/r_R^2)\frac{\exp(ikL + ik|\vec{x}-\vec{x}\,'|^2/2L)}{i\lambda L}\exp(-|\vec{x}\,|^2/r_T^2), \tag{2.14}$$

is the channel's spatial impulse response.

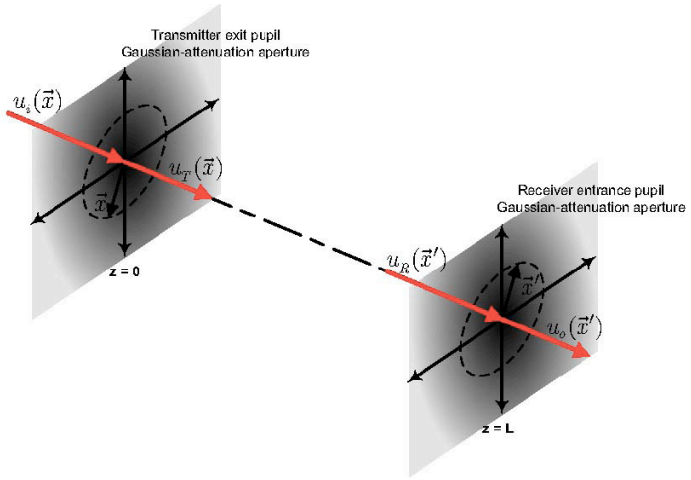


Figure 2-2: Propagation geometry with soft apertures.

The singular-value (normal-mode) decomposition of $h(\vec{x}\,', \vec{x}\,)$ is

$$h(\vec{x}\,', \vec{x}\,) = \sum_{m=1}^{\infty} \sqrt{\eta_m}\, \phi_m(\vec{x}\,') \Phi_m^*(\vec{x}\,), \tag{2.15}$$

where

$$1 \geq \eta_1 \geq \eta_2 \geq \eta_3 \geq \cdots \geq 0, \tag{2.16}$$

are the modal transmissivities, $\{\Phi_m(\vec{x}\,)\}$ is a complete orthonormal (CON) set of functions (input modes) on the transmitter's exit-pupil plane, and $\{\phi_m(\vec{x}\,')\}$ is a CON set of functions (output modes) on the receiver's entrance-pupil plane. Physically, this decomposition implies that $h(\vec{x}\,', \vec{x}\,)$ can be separated into a countably-infinite set of parallel channels in which transmission of $u_i(\vec{x}\,) = \Phi_m(\vec{x}\,)$ results in reception of $u_o(\vec{x}\,') = \sqrt{\eta_m}\, \phi_m(\vec{x}\,')$. Singular-value decompositions are unique if their $\{\eta_m\}$ are distinct. When degeneracies exist, the SVD is not unique. In particular, a linear combination of input modes with the same η_m value produces $\sqrt{\eta_m}$ times that same linear combination of the associated output modes after propagation through $h(\vec{x}\,', \vec{x}\,)$.

The spatial impulse response $h(\vec{x}\,', \vec{x}\,)$ has both rectangular and cylindrical symmetries. The Hermite-Gaussian (HG) modes $\Phi_{n,m}(x, y)$ provide an SVD of this channel that has rectangular symmetry, whereas Laguerre-Gaussian (LG) modes $\Phi_{p,l}(r, \theta)$ provide an alternative SVD for this channel with cylindrical symmetry. Even though

the spatial forms of the two sets of CON spatial modes are completely different, the associated modal transmissivities for the HG and the LG modes are respectively given by

$$\eta_q = \left(\frac{1 + 2D_f - \sqrt{1 + 4D_f}}{2D_f} \right)^q , \tag{2.17}$$

for $q = 1, 2, \ldots$. $D_f = (kr_T^2/4L)(kr_R^2/4L)$ is the product of the transmitter-pupil and receiver-pupil Fresnel numbers for this soft-aperture configuration. Also, there are q spatial modes with transmissivity η_q. The doubly-indexed HG modes $\Phi_{n,m}(x, y)$ with $n+m+1 = q$ span the same eigenspace as the doubly-indexed LG modes $\Phi_{p,l}(r, \theta)$ with $2p+|\ell|+1 = q$, and hence are related by a unitary transformation. Channel capacity, when either the HG or LG modes are employed for information transmission depends only on their modal transmissivities. Hence owing to singular-value degeneracies, the HG and LG modes of the soft-aperture free-space channel are equivalent mode sets as far as channel capacity is concerned. A single frequency-ω photon in the LG mode $\Phi_{p,l}(r, \theta)$ carries orbital angular momentum $\hbar\ell$ directed along the propagation (z) axis, whereas that same photon in the HG mode $\Phi_{n,m}(x, y)$ carries no z-directed orbital angular momentum. The equivalence of the $\{\eta_{p,l}\}$ and the $\{\eta_{n,m}\}$ then implies that angular momentum does not play a role in determining the channel capacity for classical information transmission over the free-space channel shown in Fig. 2-2.

2.4.2 Wideband Capacities with Multiple Spatial Modes

In this section, we shall address the wideband capacities that can be achieved over the pure-loss, scalar free-space channel shown in Fig. 2-2 using either heterodyne detection, homodyne detection, or the optimum joint-detection receiver. We will allow the transmitter to use multiple spatial modes, from either the HG or LG mode sets, and all frequencies $\omega \in [0, \infty)$ subject to a constraint, P, on the average power in the field entering the transmitter's exit pupil. It follows from our prior work [7, 40]

that the capacities we are seeking satisfy,

$$C(P) = \max_{\bar{N}_q(\omega)} \sum_{q=1}^{\infty} q \int_0^{\infty} \frac{\mathrm{d}\omega}{2\pi} C_{\mathrm{SM}}(\eta(\omega)^q, \bar{N}_q(\omega)), \tag{2.18}$$

where the maximization is subject to the average power constraint,

$$P = \sum_{q=1}^{\infty} q \int_0^{\infty} \frac{\mathrm{d}\omega}{2\pi} \hbar\omega \bar{N}_q(\omega), \tag{2.19}$$

and

$$\eta(\omega)^q \equiv \left(\frac{1 + 2(\omega/\omega_0)^2 - \sqrt{1 + 4(\omega/\omega_0)^2}}{2(\omega/\omega_0)^2} \right)^q \tag{2.20}$$

is the modal transmissivity at frequency ω with q-fold degeneracy, with $\omega_0 = 4cL/r_t r_R$ being the frequency at which $D_f = 1$. In (2.18),

$$C_{\mathrm{SM}}(\eta, \bar{N}) \equiv \begin{cases} g(\eta\bar{N}), & \text{for optimum reception} \\ \ln(1 + \eta\bar{N}), & \text{for heterodyne detection} \\ \frac{1}{2}\ln(1 + 4\eta\bar{N}), & \text{for homodyne detection} \end{cases} \tag{2.21}$$

are the relevant single-mode capacities as functions of the modal transmissivity, η, and the average photon number, $\bar{N}$, for that mode. Regardless of the frequency dependence of $\eta(\omega)$ the single-mode capacity formulas for heterodyne and homodyne detection imply that their wideband multiple-spatial-mode capacities bear the following relationship,

$$C_{\mathrm{hom}}(P) = \frac{1}{2} C_{\mathrm{het}}(4P). \tag{2.22}$$

Thus, only two maximizations need to be performed, both of which can be done via Lagrange multipliers, to obtain the wideband multiple-spatial-mode capacities for optimum reception, heterodyne detection, and homodyne detection.

The results we have obtained by performing the preceding maximizations are as follows. The optimum-reception capacity (in nats/sec) and its associated optimum

modal-power spectra are given by

$$C(P) = \frac{P}{\hbar\omega_0\sigma} - \sum_{q=1}^{\infty} q \int_0^{\infty} \frac{\mathrm{d}\omega}{2\pi} \ln[1 - \exp(-\omega/\omega_0\eta(\omega)^q\sigma)], \tag{2.23}$$

and

$$\hbar\omega\bar{N}_q(\omega) = \frac{\hbar\omega/\eta(\omega)^q}{\exp(\omega/\omega_0\eta(\omega)^q\sigma) - 1}, \tag{2.24}$$

respectively, where σ is a Lagrange multiplier chosen to enforce the average power constraint. The corresponding capacity and optimum modal-power spectra for heterodyne detection are

$$C_{\mathrm{het}}(P) = \sum_{q=1}^{\infty} q \int \frac{\mathrm{d}\omega}{2\pi} \ln\left(\frac{\beta\omega_0\eta(\omega)^q}{\omega}\right), \tag{2.25}$$

and

$$\hbar\omega\bar{N}_q(\omega) = \max\left[\hbar\omega_0\left(\beta - \frac{\omega}{\omega_0\eta(\omega)^q}\right), 0\right], \tag{2.26}$$

where β is another Lagrange multiplier, again chosen to enforce the average power constraint. Finally, the capacity and optimum power allocation for homodyne detection are given by

$$C_{\mathrm{hom}}(P) = \sum_{q=1}^{\infty} q \int \frac{\mathrm{d}\omega}{2\pi} \left[\frac{1}{2}\ln\left(\frac{2\beta\omega_0\eta(\omega)^q}{\omega}\right)\right], \tag{2.27}$$

and

$$\hbar\omega\bar{N}_q(\omega) = \max\left[\hbar\omega_0\left(\frac{\beta}{2} - \frac{\omega}{4\omega_0\eta(\omega)^q}\right), 0\right], \tag{2.28}$$

where β is a Lagrange multiplier, chosen to enforce the average power constraint.

2.4.3 Optimum power allocation: water-filling

The capacity-achieving power spectrum for optimal reception employs all spatial modes and all frequencies. On the other hand, the capacity-achieving power spectra for heterodyne and homodyne detection are "water-filling" allocations, i.e., they

fill spatial-mode/frequency volumes above their appropriate noise-to-transmissivity-ratio contours until the average power constraint is met (Fig. 2-3). That water-filling power allocation should be capacity achieving for these coherent detection cases is hardly a surprise, as water-filling power allocation has long been known to be optimal for additive Gaussian noise channels [4]. A consequence of water-filling power allocation is that heterodyne and homodyne detection only employ a finite number of spatial modes to achieve their respective capacities, whereas optimal-reception capacity needs all spatial modes. This behavior is illustrated in Fig. 2-4(a)-(c), where we have plotted the capacity-achieving power spectra for optimum reception, homodyne detection, and heterodyne detection when $P = 8.12\hbar\omega_0^2$. In this case, heterodyne detection uses $1 \leq q \leq 3$ (a total of 6 spatial modes) with non-zero power, and homodyne detection uses $1 \leq q \leq 4$ (a total of 10 spatial modes) with non-zero power. Optimum reception uses all spatial modes, but we have only plotted the spectra for $1 \leq q \leq 6$.

In Fig. 2-4(d) we have plotted the heterodyne detection, homodyne detection, and optimum reception capacities in bits/sec, normalized by ω_0, versus the normalized power, $P/\hbar\omega_0^2$. Unlike the case seen in Fig. 2-1(c) for the wideband capacities of the single-spatial-mode, far-field pure-loss channel, in which heterodyne detection outperforms homodyne detection at high power levels, Fig. 2-4(d) shows that homodyne detection is consistently better than heterodyne detection for the multiple-spatial-mode scenario. This behavior has a simple physical explanation. Consider first the single-spatial mode wideband capacities. At low power levels, when capacity is power limited, homodyne detection outperforms heterodyne detection because at every frequency it suffers less noise. On the other hand, at high enough power levels single-spatial mode communication becomes bandwidth limited. In this case heterodyne detection's factor-of-two bandwidth advantage over homodyne detection carries the day. Things are different when multiple spatial modes are available. In this case, increasing power never reaches bandwidth-limited operation; additional, lower transmissivity, spatial modes get employed as the power is increased so that the noise advantage of homodyne detection continues to give a higher channel capacity than

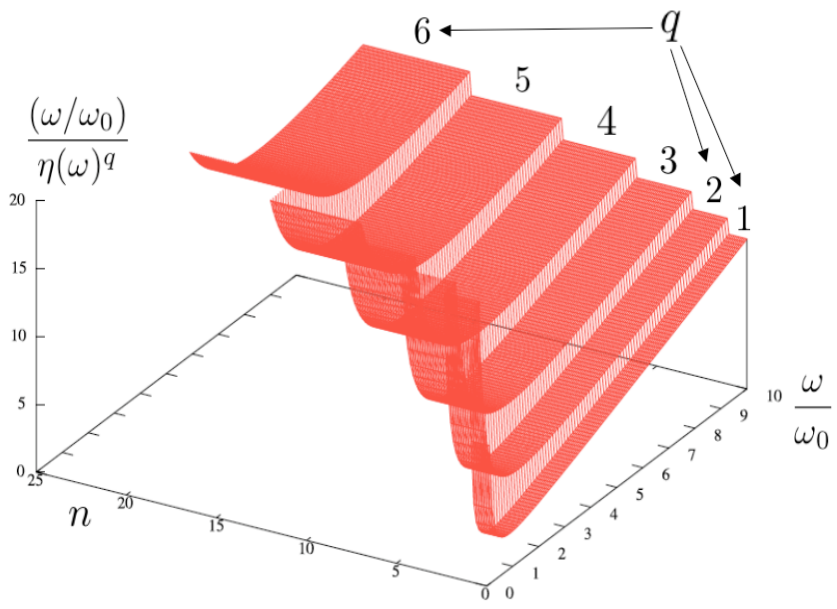


Figure 2-3: Visualization of the capacity-achieving power allocation for the wideband, multiple-spatial-mode, free-space channel, with coherent-state encoding and heterodyne detection as 'water-filling' into bowl-shaped steps of a terrace. The horizontal axis ω/ω_0, is a normalized frequency; n is the total number of spatial modes used. The vertical axis is $(\omega/\omega_0)/\eta(\omega)^q$. Power starts 'filling' into this terrace starting from the $q = 1$ step. It keeps spilling over to the higher steps as input power increases.

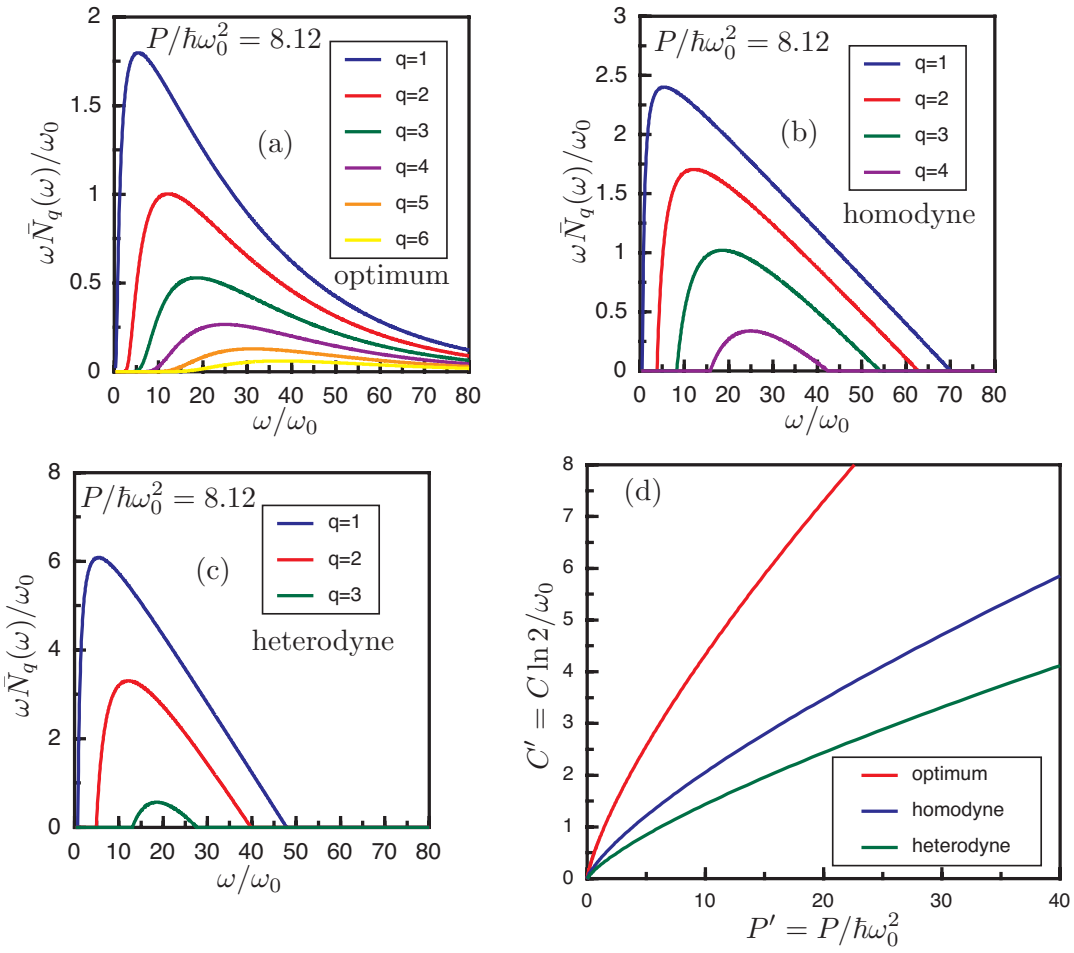


Figure 2-4: Capacity-achieving power spectra for wideband, multiple-spatial-mode communication over the scalar, pure-loss, free-space channel when $P = 8.12\hbar\omega_0^2$: (a) optimum reception uses all spatial modes although spectra are only shown (from top to bottom) for $1 \leq q \leq 6$; (b) homodyne detection uses 10 spatial modes with (from top to bottom) $1 \leq q \leq 4$; (c) heterodyne detection uses 6 spatial modes with (from top to bottom) $1 \leq q \leq 3$. (d) Wideband, multiple-spatial-mode capacities (in bits per second) for the scalar, pure-loss, free-space channel that are realized with optimum reception (top curve), homodyne detection (middle curve), and heterodyne detection (bottom curve). The capacities, in bits/sec, are normalized by $\omega_0 = 4cL/r_T r_R$, the frequency at which $D_f = 1$, and plotted versus the average transmitter power normalized by $\hbar\omega_0^2$.

does heterodyne detection.

Figure 2-4 shows that the wideband capacity realized with optimum reception, on the multiple-spatial-mode pure-loss channel, increasingly outstrips that of homodyne detection with increasing transmitter power. This advantage indicates that joint measurements over entire codewords afford performance that is unapproachable with homodyne detection, which is a single-use quantum measurement.

2.5 Low-power Coherent-State Modulation

We computed the classical information capacities of the single-mode and wideband lossy bosonic communication channels, using various structured transmitter encodings and receiver measurements, in [39]. Out of the various modulation states, of particular importance are the coherent-state encoding techniques, as coherent-states are classical states of light which can be generated readily using lasers. Moreover, we have shown [7] that coherent-state encoding with an isotropic complex-Gaussian prior density over all coherent states, along with an optimum receiver measurement, achieves capacity for the pure-loss bosonic channel. Coherent-state encodings would be provably optimum for encoding classical messages for thermal-noise bosonic channels and bosonic broadcast channels, if certain conjectures on the minimum output entropy of bosonic channel were proven to be true [9, 12]. When the transmitter is starved for photons, instead of using the full-blown Gaussian distribution over all coherent states, several simplified encoding techniques using a few coherent states do remarkably well. These low-power coherent-state based encoding schemes are the subject of study for this section.

2.5.1 On-Off Keying (OOK)

A common scheme for optical modulation, which has been in use for many years, is On-Off Keying (OOK) using coherent states with direct detection measurement. With direct detection (or photon counting) receivers, the bosonic channel, from the coherent-state transmitter to the measurement outcome, becomes a classical Pois-

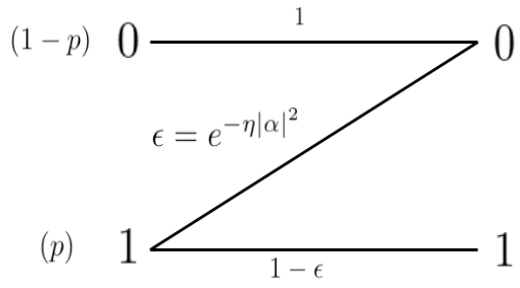


Figure 2-5: The "Z"-channel model. The single-mode bosonic channel, when used with OOK-modulated coherent-states and photon number measurement, reduces to a "Z"-channel when the mean photon number constraint at the input satisfies $\bar{N} \ll 1$. The transition probability from logical 1 (input coherent state $|\alpha\rangle$) to logical 0 (vacuum state) is given by $\epsilon = e^{-\eta|\alpha|^2}$.

son channel, because of the Poisson statistics of the photon-number measurement on coherent states. This encoding-decoding scheme is widely employed in real systems because of easy availability of coherent-state modulators, and direct-detection receivers[1].

OOK entails either sending a coherent-state $|\alpha\rangle$ or the vacuum state $|0\rangle$ in each use of the channel. Consider a single-mode lossy bosonic channel with transmissivity η and a mean photon number constraint $\bar{N}$ at the input of the channel. In the limit of $\bar{N} \ll 1$, the bosonic channel for these encoding states reduces to a "Z"-channel (Figure 2-5), wherein, the transition probability from logical 1 (input coherent state $|\alpha\rangle$) to logical 0 (vacuum state) is given by $\epsilon = e^{-\eta|\alpha|^2}$. The capacity of the channel in bits per use is given by

$$C_{\mathrm{OOK}}(\eta, \bar{N}) = \max_{p} \left[H\left(p(1 - e^{-\eta\bar{N}/p})\right) - pH\left(e^{-\eta\bar{N}/p}\right) \right], \tag{2.29}$$

where $H(p) = -p\log p - (1-p)\log 1 - p$ is the binary Shannon entropy. The channel capacity of OOK with direct-detection gets closer and closer to optimal capacity as $\bar{N} \to 0$, as we see in Figure 2-6. The approach of the OOK capacity to the optimal capacity is exponentially slow as $\bar{N} \to 0$. At $\bar{n} = 10^{-7}$, C_{OOK} is about 77.5% of the ultimate capacity $g(\eta\bar{N})$ and the ratio $C_{\mathrm{OOK}}/g(\eta\bar{N})$ increases at about 0.03 per

[1]Although, typical direct-detection receivers are not signal-shot-noise limited photon counters.

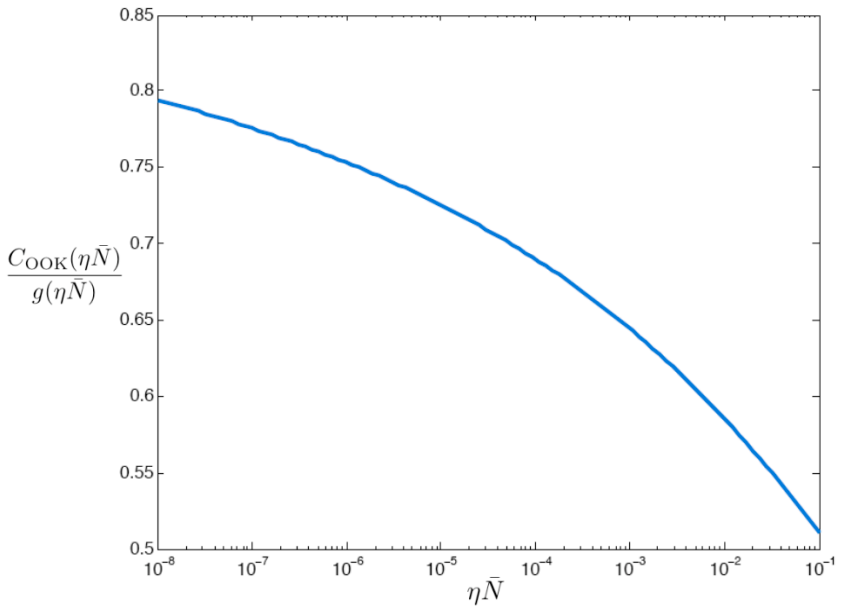


Figure 2-6: This figure shows that capacity achieved using OOK modulation and direct-detection gets closer and closer to optimal capacity as $\bar{N} \to 0$. The ordinate is the ratio of the OOK and the ultimate capacities in bits per channel use. The approach of the OOK capacity to the optimal capacity gets exponentially slow as $\bar{N} \to 0$, as is evident from the log-scale used for the $\eta\bar{N}$-axis of the graph. At $\bar{N} = 10^{-7}$, C_{OOK} is about 77.5% of the ultimate capacity $g(\eta\bar{N})$.

decade of decrease of $\bar{N}$, at very low values of $\bar{N}$.

2.5.2 Binary Phase-Shift Keying (BPSK)

Another common modulation scheme using coherent-state inputs is Binary Phase-Shift Keying (BPSK), in which the input alphabet comprises two coherent states of equal magnitude that are 180 degrees out of phase: $\{|\alpha\rangle, -|\alpha\rangle\}$. With a two-element quantum POVM measurement that result in symmetric outcomes for the two symbol states, the BPSK channel becomes a binary symmetric channel (BSC). With a mean photon number constraint of $\bar{N}$ at the input, it is easy to show that the achievable capacity using the best symbol-by-symbol measurement at the output (realized by a sequence of Dolinar receivers [20]) is given by the BSC capacity formula:

$$C_{\text{BPSK}}(\eta\bar{N}) = 1 - H\left(\frac{1-\sqrt{1-e^{-4\eta\bar{N}}}}{2}\right). \tag{2.30}$$

Comparing performance of BPSK to that of OOK

Figure 2-7 compares classical communication rates achievable by OOK (with direct detection) and BPSK (with Dolinar reception) modulation schemes, with the rates achieved by doing homodyne or heterodyne detection with an input alphabet over all coherent states, chosen from an isotropic Gaussian distribution of coherent states. The ultimate capacity is given by $g(\eta\bar{N})$ bits per channel use. Figure 2-7(a) is for low $\bar{N}$, whereas Figure 2-7(b) compares the achievable rates at higher $\bar{N}$. At very low mean photon number, OOK performs the best of the conventional schemes. In the low $\bar{N}$ regime, both the binary modulation schemes, viz., OOK and BPSK perform better than the unrestricted coherent-state modulation with coherent detection. In the high $\bar{N}$ regime, coherent-detection capacities outperform the binary schemes, because the maximum rate achievable using any binary modulation system is 1 bit per channel use.

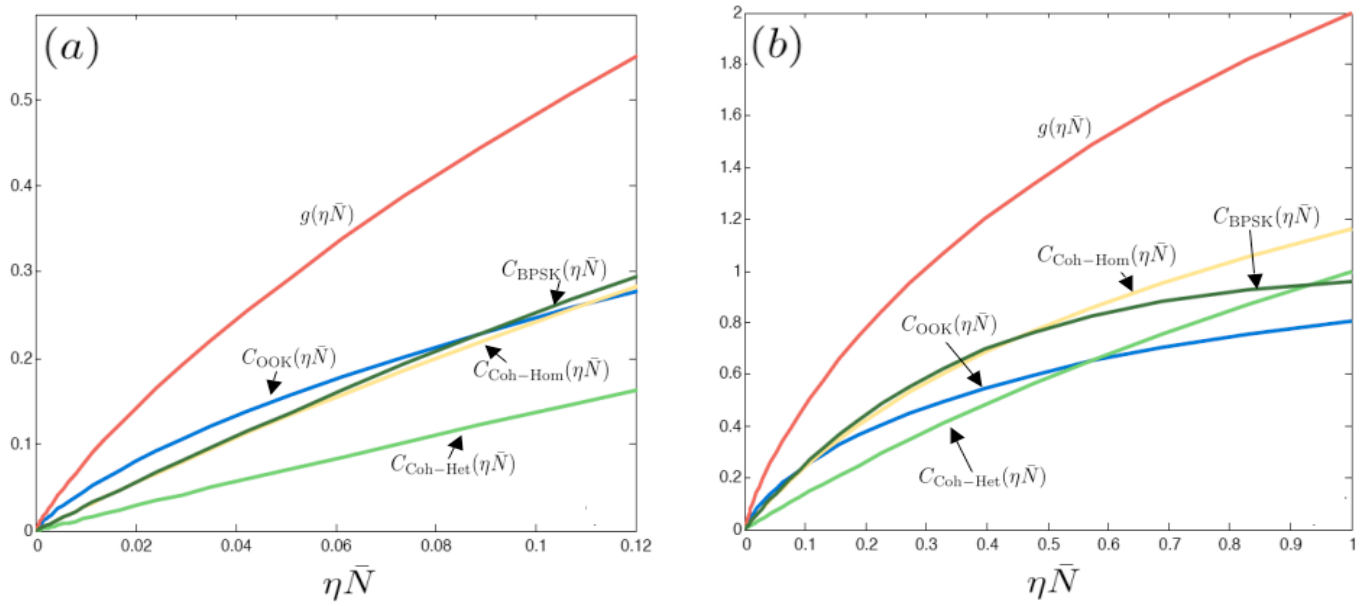


Figure 2-7: Comparison of capacities (in bits per channel use) of the single-mode lossy bosonic channel achieved by: OOK modulation with direct detection; $\{|\alpha\rangle, -|\alpha\rangle\}$-BPSK modulation using coherent-states; and homodyne and heterodyne detection with isotropic-Gaussian random coding over coherent states. For very low values of $\bar{N}$, the average transmitter photon number, shown in (a), OOK outperforms all but the ultimate capacity. At somewhat higher values of $\bar{N}$, both OOK and BPSK are better than isotropic-Gaussian random coding with coherent detection. In the high $\bar{N}$ regime, coherent-detection capacities outperform the binary schemes, because, the maximum rate achievable by the latter approaches cannot exceed 1 bit per channel use.

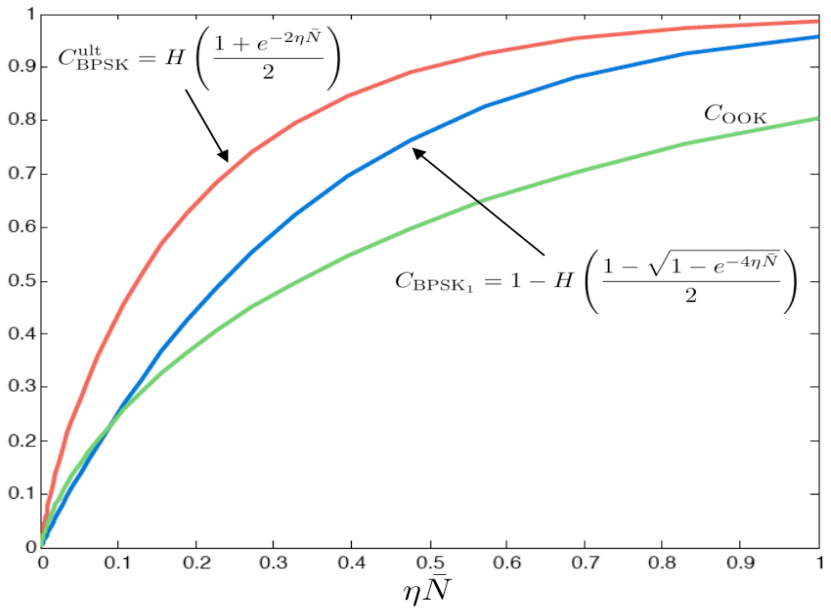


Figure 2-8: This figure illustrates the gap between the ultimate BPSK coherent-state capacity (Equation (2.31)) and the achievable rate using a BPSK coherent-state alphabet and symbol-by-symbol "Dolinar receiver" measurement (Equation (2.30)). In order to bridge the gap between these two capacities, optimal multi-symbol joint measurement schemes must be used at the receiver. All capacities are plotted in units of bits per channel use.

Ultimate capacity using the BPSK alphabet

The ultimate capacity that can be achieved using a binary coherent-state alphabet $\{|\alpha\rangle, |-\alpha\rangle\}$, with an average input-photon-number constraint $\bar{N}$ can be computed by maximizing the Holevo information for the binary alphabet over all binary prior probability densities $\{p, 1-p\}$. The ultimate capacity using the binary coherent-state alphabet is given by

$$C_{\text{BPSK}}^{\text{ult}} = H\left(\frac{1+e^{-2\eta\bar{N}}}{2}\right). \tag{2.31}$$

Figure 2-8 shows the gap between the ultimate BPSK capacity and the achievable rate using a BPSK coherent-state alphabet and symbol-by-symbol Dolinar-receiver measurement. In order to bridge the gap between these two capacities, optimal multi-symbol joint measurement schemes must be used at the receiver. Some examples of such improvement over single-symbol measurement schemes (and implementations thereof) were worked out by Sasaki et. al., in [45, 46]. Recently, Ishida et. al. worked out best achievable rate regions for the lossy bosonic channel using various coherent-state modulation schemes [47], such as Quadrature Phase Shift Keying (QPSK), and Quadrature Amplitude Modulation (QAM).

Chapter 3

Broadcast and Wiretap Channels

3.1 Background

A *broadcast channel* is the congregation of communication media connecting a single transmitter to two or more receivers. The transmitter encodes and sends out information to each receiver in a way that each receiver can reliably decode its respective information. The information sent out to the receivers may be independent or nested. The capacity region of a broadcast channel is the set of all rate M-tuples $\{R_0, \ldots, R_{M-1}\}$, at which independent information can be sent perfectly reliably to the respective M receivers by using suitable encoding and decoding schemes. The classical discrete-memoryless broadcast channel was first studied by Cover [48], whose capacity region still remains an open problem. The capacity region of a special case of the broadcast channel, known as the degraded broadcast channel – in which the channel symbols received by one of the receivers is a stochastically degraded version of the symbols received by the other receiver – was conjectured by Cover [48], and later proved to be achievable by Bergmans [49]. The converse to the degraded broadcast channel capacity theorem was established later by Bergmans [50] and Gallager [51].

A quantum broadcast channel is a quantum-mechanical communication link connecting one transmitter to two or more receivers. Quantum broadcast channels, like point-to-point quantum communication channels, may be used to send classical information, quantum information, or a combination thereof. We will restrict our attention

only to the case of classical information transmission over quantum broadcast channels. The transmitter encodes information intended to be sent to various receivers into quantum states of the transmission medium, and the receivers extract classical information from received quantum states by performing suitable quantum measurements. Even though the capacity region of the general quantum broadcast channel is still an open problem, like its classical counterpart, the capacity region of the two-user degraded quantum broadcast channel for finite-dimensional Hilbert spaces was found by Yard, et. al.[52]. bosonic broadcast channels constitute a special class of quantum broadcast channels in which the information is encoded into quantum states of an optical-frequency quantized electromagnetic field.

In this chapter, we will show that when coherent-state encoding is employed in conjunction with coherent detection, the bosonic broadcast channel is equivalent to a classical degraded Gaussian broadcast channel whose capacity region is known, and known to be dual to that of the classical Gaussian multiple-access channel [53]. Thus, under these coding and detection assumptions, the capacity region for the bosonic broadcast channel is dual to that for the bosonic multiple-access channel (MAC) with coherent-state encoding and coherent detection. To treat more general transmitter and receiver conditions, we use a limiting argument to apply the degraded quantum broadcast-channel coding theorem for finite-dimensional state spaces [52] to the infinite-dimensional bosonic channel with an average photon-number constraint. We first consider the lossless two-receiver case in which Alice (A) simultaneously transmits to Bob (B), via the transmissivity $\eta > 1/2$ port of a lossless beam splitter, and to Charlie (C), via that beam splitter's reflectivity $1 - \eta < 1/2$ port. Alice uses arbitrary encoding with an average photon number $\bar{N}$, while Bob and Charlie employ optimum measurements. Given a conjecture about the minimum output entropy of a lossy bosonic channel is true (see chapter 4), we show that the ultimate capacity region is achieved by a coherent-state encoding, and is given by

$$R_B \leq g(\eta\beta\bar{N}), \quad R_C \leq g((1-\eta)\bar{N}) - g((1-\eta)\beta\bar{N}), \tag{3.1}$$

where $g(x) \equiv (x+1)\log(x+1) - x\log(x)$ is the entropy of the Bose Einstein distribution with mean x, and $\beta \in [0,1]$. Interestingly, this capacity region is *not* dual to that of the bosonic multiple-access channel with coherent-state encoding and optimum measurement that was found in [11].

We begin this chapter by reviewing the capacity region of the degraded classical broadcast channel, and we evaluate the capacity region of the Gaussian broadcast channel as an example. We then present a brief review of Yard et. al.'s capacity theorem for the degraded quantum broadcast channel with two receivers, following which we present our generalization of their result for an arbitrary number of receivers. Thereafter we present our results on the classical information capacity of the bosonic broadcast channel. We first analyze the two-receiver lossless case with no additional noise and that with additive thermal noise. We then generalize our results to the lossy broadcast channel with multiple receivers. We compare the rate regions obtained by using coherent-state encoding for the bosonic broadcast channel with that of the bosonic multiple access channel and we find that a duality that is observed between capacity regions of the classical Gaussian-noise broadcast and multiple-access channels is not seen in the quantum case. The chapter concludes with a section on the privacy capacity of the bosonic wiretap channel, which is a special kind of a two-receiver broadcast channel in which one of the receivers is an eavesdropper, while the other is the intended receiver.

3.2 Classical Broadcast Channel

In classical information theory, a two-user discrete-memoryless broadcast channel is modeled by a classical probability transition matrix $p_{B,C|A}(\beta,\gamma|\alpha)$, where α, β, and γ belong to Alice's (input) alphabet $\mathcal{A}$, and Bob and Charlie's (output) alphabets, $\mathcal{B}$ and $\mathcal{C}$ respectively. A broadcast channel is said to be *memoryless* if successive uses of the channel are independent, i.e., $p_{B^n,C^n|A^n}(\beta^n,\gamma^n|\alpha^n) = \Pi_{i=1}^n p_{B,C|A}(\beta_i,\gamma_i|\alpha_i)$. M-user broadcast channels, for $M > 2$, are defined similarly. A $((2^{nR_B}, 2^{nR_C}), n)$ code for a two-receiver broadcast channel consists on an encoder

$$\alpha^n : 2^{nR_B} \times 2^{nR_C} \to \mathcal{A}^n, \tag{3.2}$$

and two decoders

$$\hat{W}_B : \mathcal{B}^n \to 2^{nR_B} \tag{3.3}$$

$$\hat{W}_C : \mathcal{C}^n \to 2^{nR_C}. \tag{3.4}$$

The probability of error $P_e^{(n)}$ is the probability that the overall decoded message doesn't match with the transmitted message, i.e.,

$$P_e^{(n)} = P(\hat{W}_B(B^n) \neq W_B \quad \text{OR} \quad \hat{W}_C(C^n) \neq W_C),$$

where the message (W_B, W_C) is assumed to be uniformly distributed over $2^{nR_B} \times 2^{nR_C}$. A rate pair (R_B, R_C) is said to be *achievable* for the broadcast channel if there exists a sequence of $((2^{nR_B}, 2^{nR_C}), n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$. The *capacity region* of the broadcast channel is the closure of the set of all achievable rates.

Although the capacity region for general broadcast channels is still an open problem, the capacity region is known for a special class of broadcast channels known as *degraded broadcast channels*. It is often the case that one receiver (say C) is further downstream from the first receiver (say B), so that C always receives a degraded version of B's message. When $A \to B \to C$ forms a Markov chain, i.e., when $p_{B,C|A}(\beta, \gamma|\alpha) = p_{B|A}(\beta|\alpha)p_{C|B}(\gamma|\beta)$ we say that the receiver C is a *physically degraded* version of B, and that $A \to B \to C$ is a *physically degraded* broadcast channel. The probabilities of error $P(\hat{W}_B(B^n) \neq W_B)$ and $P(\hat{W}_C(C^n) \neq W_C)$ depend only on the marginal distributions $p_{B|A}(\beta|\alpha)$ and $p_{C|B}(\gamma|\beta)$ and not on the joint distribution $p_{B,C|A}(\beta, \gamma|\alpha)$. Thus we define a weaker notion of degraded broadcast channel — a broadcast channel $p_{B,C|A}(\beta, \gamma|\alpha)$ is said to be *degraded* (also known as stochastically degraded to distinguish from the stronger notion of degraded in the Markov sense), if there exists a distribution $\tilde{p}(\gamma|\beta)$, such that

$$p_{C|A}(\gamma|\alpha) = \sum_{\beta} p_{B|A}(\beta|\alpha)\tilde{p}(\gamma|\beta). \tag{3.5}$$

Such channels were first studied by Cover [48], who conjectured that the capacity region for Alice to send independent information to Bob and Charlie at rates R_B and R_C respectively over a degraded broadcast channel[1] $A \to B \to C$ is the convex hull of the closure of all (R_B, R_C) satisfying

$$R_B \leq I(A; B|T) \tag{3.6}$$

$$R_C \leq I(T; C) \tag{3.7}$$

for some joint distribution $p_T(\tau)p_{A|T}(\alpha|\tau)p_{B,C|A}(\beta, \gamma|\alpha)$, where T is an auxiliary random variable with cardinality $|\mathcal{T}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}|\}$. The achievability of the above capacity result was proved by Bergmans [49], whereas Gallager came up with a particularly novel proof of the converse [51].

3.2.1 Degraded broadcast channel with M receivers

A formal proof of the capacity region for a degraded discrete memoryless broadcast channel with an arbitrary number of receivers, was done recently by Borade et. al. [54], in which they also proved bounds for capacity regions for general multiple-level broadcast networks. Consider a discrete memoryless broadcast channel with transmitter Alice (A) sending information to M receivers, Y_0, Y_1, ..., Y_{M-1}. Such a channel is completely specified by the transition probabilities $p_{Y_0,\ldots,Y_{M-1}|A}(y_0, \ldots, y_{M-1}|\alpha)$. Let us also assume that the channel map is stochastically degraded (in the same sense as described in Eq. (3.5)), as $A \to Y_0 \to Y_1 \to \ldots \to Y_{M-1}$; i.e., Y_0 being the least noisy receiver and Y_{M-1} the noisiest receiver. The optimal capacity region is given by the

[1] In all that follows, a degraded broadcast channel $A \to B \to C$ will be understood to mean a stochastically degraded channel (3.5) with transmitter A, and receivers B and C.

convex hull of all rate M-tuples $(R_0, R_1, \ldots, R_{M-1})$ satisfying

$$\begin{aligned} R_0 &\leq I(A; Y_0|T_1), \\ R_k &\leq I(T_k; Y_k|T_{k+1}), \quad \text{for } k \in \{1, \ldots, M-2\}, \\ R_{M-1} &\leq I(T_{M-1}; Y_{M-1}), \end{aligned} \tag{3.8}$$

where T_k, $k \in \{1, \ldots, M-1\}$ are auxiliary random variables such that $T_{M-1} \rightarrow T_{M-2} \rightarrow \ldots \rightarrow T_1 \rightarrow A$ forms a Markov chain, i.e.,

$$p_{T_{M-1},\ldots,T_1,A}(\tau_{M-1}, \ldots, \tau_1, \alpha) = p_{T_{M-1}}(\tau_{M-1}) \left(\prod_{k=M-1}^{2} p_{T_{k-1}|T_k}(\tau_{k-1}|\tau_k) \right) p_{A|T_1}(\alpha|\tau_1). \tag{3.9}$$

The above Markov chain structure of the auxiliary random variables T_k, $k \in \{1, \ldots, M-1\}$ has been shown to be optimal [54]. In a degraded broadcast channel, messages intended for noisier receivers can always be decoded by less noisy receivers[2]. Hence the k^{th} receiver actually receives $M-k$ messages at a rate $R_k + \ldots + R_{M-1}$.

3.2.2 The Gaussian broadcast channel

A Gaussian broadcast channel is one in which each receiver receives the transmitted symbols corrupted by zero-mean additive Gaussian noise of a fixed noise variance. The Gaussian broadcast channel is an example of a degraded broadcast channel because the channel can be recharacterized as a stochastically degraded channel in which the noisier receiver's received symbols can be thought of as being obtained from the less noisy receiver's received symbols by passing them through a hypothetical additive Gaussian noise channel with a noise variance equaling the difference of the Gaussian noise variances seen by the two receivers (see Fig. 3-1).

[2]For a more detailed description of how messages are encoded and decoded in a degraded broadcast channel using superposition coding, please see [3].

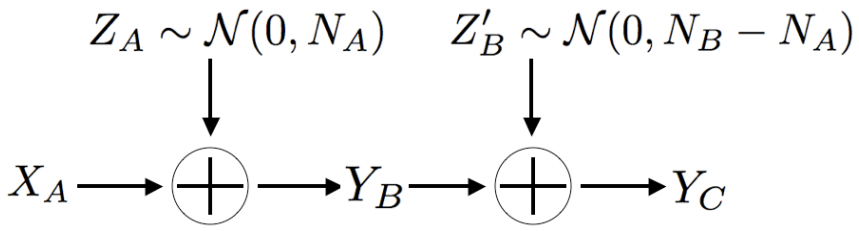


Figure 3-1: Classical additive Gaussian noise broadcast channel

The two-user Gaussian broadcast channel

The simplest case of the Gaussian broadcast channel is the scalar two-receiver case. There are two receivers, Bob and Charlie, whose received symbols Y_B and Y_C are given in terms of Alice's transmitted symbol X_A by

$$Y_B = X_A + Z_B \quad \text{and} \tag{3.10}$$

$$Y_C = X_A + Z_C, \tag{3.11}$$

where $Z_A \sim \mathcal{N}(0, N_B)$ and $Z_B \sim \mathcal{N}(0, N_C)$ are zero-mean Gaussian distributed random variables with variances N_B and N_C respectively. This channel can be characterized by an equivalent degraded channel as shown in Fig. 3-1.

Let us use $C_G(\gamma)$ to denote the capacity of a memoryless scalar additive white Gaussian channel (AWGN) with signal to noise ratio (SNR) γ. It is well known that,

$$C_G(\gamma) = \frac{1}{2}\ln(1+\gamma) \quad \text{nats per use.} \tag{3.12}$$

It is easily shown [3], that an achievable capacity region for the Gaussian broadcast channel, with signal power constraint $E[|X_A|^2] \leq \bar{N}$, can be obtained by choosing both $p_T(\tau)$ and $p_{A|T}(\alpha|\tau)$ to be Gaussian. The resulting achievable region is given by,

$$R_B \leq C_G\left(\frac{\beta\bar{N}}{N_B}\right), \tag{3.13}$$

$$R_C \leq C_G\left(\frac{(1-\beta)\bar{N}}{\beta\bar{N}+N_C}\right), \tag{3.14}$$

for $0 \leq \beta \leq 1$. Bergmans proved the converse statement for the Gaussian broadcast channel [50], thereby showing that the capacity region given above is the ultimate capacity region for the Gaussian broadcast channel. Using Bergmans's notation[3],

$$g_C(S) \equiv \frac{1}{2} \ln (2\pi e S) \tag{3.15}$$

to denote the Shannon entropy (in nats) of a Gaussian random variable with variance S, the above two-receiver Gaussian broadcast capacity region can alternatively be expressed as,

$$\begin{aligned} R_B &\leq g_C(\beta \bar{N} + N_B) - g_C(N_B), & (3.16) \\ R_C &\leq g_C(\bar{N} + N_C) - g_C(\beta \bar{N} + N_C) & (3.17) \end{aligned}$$

for $0 \leq \beta \leq 1$. An example plot of the capacity region of a two-user Gaussian broadcast channel is given in Fig. 3-2.

An example from optical communications

Let us consider a special case of the two-user Gaussian broadcast channel, in which Bob and Charlie receive attenuated versions of Alice's message corrupted by Gaussian noise, i.e.,

$$\begin{aligned} Y_B &= \sqrt{\eta} X_A + \sqrt{1-\eta} Z_B \quad \text{and} \\ Y_C &= \sqrt{1-\eta} X_A + \sqrt{\eta} Z_C, & (3.18) \end{aligned}$$

[3]We use a subscript (C) for Bergman's $g(\cdot)$ function to distinguish it from the function $g(x) = (1+x)\ln(1+x) - x \ln x$ — which is the Shannon entropy of the Bose-Einstein probability mass function with mean x (and also the von Neumann entropy of the bosonic thermal state with mean photon-number x) — that will be used throughout this thesis. We will see later in this chapter, that the functions $g_C(\cdot)$ and $g(\cdot)$ play analogous roles in defining classical capacity regions for the classical Gaussian broadcast channel and that of the quantum (bosonic) broadcast channel, respectively. As we will see in Chapter 5, the functions $g_C(\cdot)$ and $g(\cdot)$ also play analogous roles in defining the (classical) Entropy Power Inequality (EPI) and the (quantum) Entropy Photon-Number Inequality (EPnI).

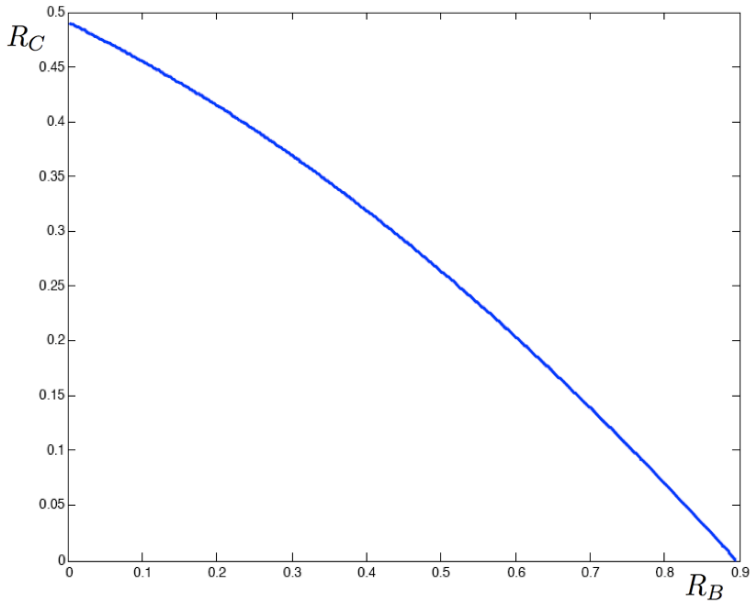


Figure 3-2: Capacity region of the classical additive Gaussian noise broadcast channel, with an input power constraint $E[|X_A|^2] \leq 10$, and noise powers given by, $N_B = 2$ and $N_C = 6$. The rates R_B and R_C are in nats per channel use.

where $1/2 < \eta < 1$, and Z_B and Z_C are independent, identically distributed (i.i.d.) $\mathcal{N}(0, N)$ random variables. Such a channel model arises when the transmitter Alice encodes classical information into the magnitude of the complex electromagnetic field of a classical laser beam and the beam splits into two through a lossless beam splitter of transmissivity η, in presence of an ambient thermal environment that is sufficiently strong that its noise contribution dominates over the quantum noise. Bob and Charlie, the two receivers receive their respective classical signals at the two output ports of the beam splitter by performing optical homodyne detection (see Fig. 3-3). Using Bergman's results, it is not hard to see that the capacity region of this channel will be given by,

$$R_B \leq g_C(\eta\beta\bar{N} + (1-\eta)N) - g_C((1-\eta)N), \tag{3.19}$$

$$R_C \leq g_C((1-\eta)\bar{N} + \eta N) - g_C((1-\eta)\beta\bar{N} + \eta N), \tag{3.20}$$

where $0 \leq \beta \leq 1$.

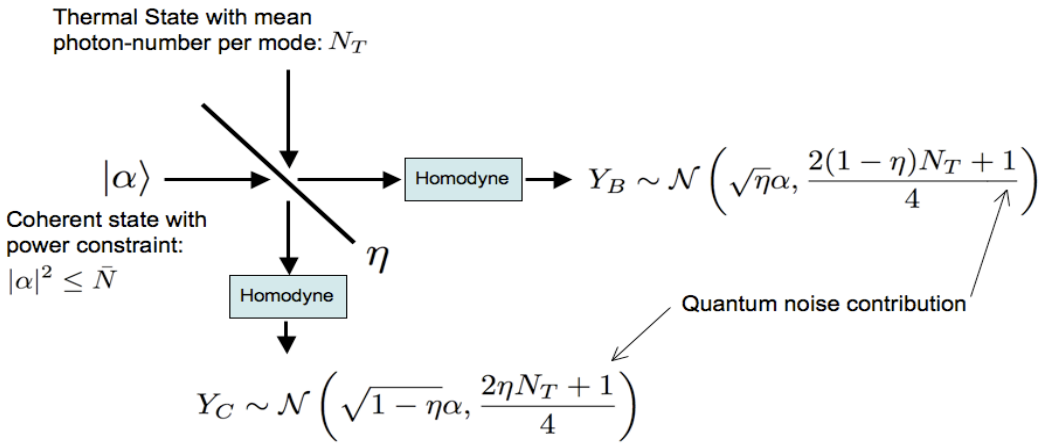


Figure 3-3: A broadcast channel in which the transmitter Alice encodes information into a real-valued α for a classical electromagnetic field (coherent state $|\alpha\rangle$) and the beam splits into two, through a lossless beam splitter with transmissivity η, in presence of an ambient thermal environment with an average of N_T photons per mode. Bob and Charlie, the two receivers, receive their respective classical signals Y_B and Y_C at the two output ports of the beam splitter by performing optical homodyne detection. In the limit of high noise ($N_T \gg 1$), and with the substitutions $X_A = \alpha; \alpha \in \mathbb{R}$, and $N_T = 2N$, this channel reduces to the broadcast channel model described by (3.18).

The M-receiver Gaussian broadcast channel

As an example of the capacity region of a degraded broadcast channel with M receivers, let us consider an M-receiver version of the lossy thermal noise optical channel model from Eq. (3.18). Each of the M receivers receive an attenuated version of Alice's transmitted message with an additive zero-mean Gaussian noise, given by

$$Y_k = \sqrt{\eta_k} A + \sqrt{1-\eta_k} Z_k, \quad k \in \{0, \ldots, M-1\}, \tag{3.21}$$

where the transmitter has a mean power constraint given by $E[|A|^2] \leq \bar{N}$, and Z_k are i.i.d. Gaussian $\mathcal{N}(0, N)$ random variables. The optimal capacity region of the Gaussian broadcast channel for M receivers was first found by Bergmans [50], and is given by

$$R_k \leq g_C(\eta_k \beta_{k+1} \bar{N} + (1-\eta_k)N) - g_C(\eta_k \beta_k \bar{N} + (1-\eta_k)N), \quad k \in \{0, \ldots, M-1\}, \tag{3.22}$$

where,

$$0 = \beta_0 < \beta_1 < \ldots < \beta_{M-1} < \beta_M = 1. \tag{3.23}$$

3.3 Quantum Broadcast Channel

In this section, we study the classical information capacity of quantum broadcast channels, which are quantum channels from one transmitter to two or more receivers. The transmitter encodes information intended to be sent to various receivers into the quantum states of the transmission medium, and the receivers extract classical information from received quantum states by performing suitable quantum measurements. Even though the capacity region of the general quantum broadcast channel is still an open problem, like its classical counterpart, the capacity region of the two-user degraded quantum broadcast channel for finite-dimensional Hilbert spaces was found by Yard, et. al.[52]. We begin this section by stating Yard et. al.'s capacity theorem, and then we prove its straightforward extension to the case of an arbitrary number

of receivers.

3.3.1 Quantum degraded broadcast channel with two receivers

A quantum channel $\mathcal{N}_{A-B}$ from Alice to Bob is a trace-preserving completely positive map that maps Alice's single-use density operators $\hat{\rho}^A$ to Bob's, $\hat{\rho}^B = \mathcal{N}_{A-B}(\hat{\rho}^A)$. The two-user quantum broadcast channel $\mathcal{N}_{A-BC}$ is a quantum channel from sender Alice (A) to two independent receivers Bob (B) and Charlie (C). The quantum channel from Alice to Bob is obtained by tracing out C from the channel map, i.e., $\mathcal{N}_{A-B} \equiv \mathrm{Tr}_C\left(\mathcal{N}_{A-BC}\right)$, with a similar definition for $\mathcal{N}_{A-C}$. We say that a broadcast channel $\mathcal{N}_{A-BC}$ is *degraded* if there exists a *degrading channel* $\mathcal{N}^{\mathrm{deg}}_{B-C}$ from B to C satisfying $\mathcal{N}_{A-C} = \mathcal{N}^{\mathrm{deg}}_{B-C} \circ \mathcal{N}_{A-B}$. The degraded broadcast channel describes a physical scenario in which for each successive n uses of $\mathcal{N}_{A-BC}$ Alice communicates a randomly generated classical message $(m,k) \in (W_B, W_C)$ to Bob and Charlie, where the message-sets W_B and W_C are sets of classical indices of sizes 2^{nR_B} and 2^{nR_C} respectively. The messages (m,k) are assumed to be uniformly distributed over (W_B, W_C). Because of the degraded nature of the channel, Bob receives the entire message (m,k) whereas Charlie only receives the index k. To convey these messages (m,k), Alice prepares n-channel use states that, after transmission through the channel, result in bipartite conditional density matrices $\left\{\hat{\rho}^{B^nC^n}_{m,k}\right\}$, $\forall (m,k) \in (W_B, W_C)$. The quantum states received by Bob and Charlie, $\left\{\hat{\rho}^{B^n}_{m,k}\right\}$ and $\left\{\hat{\rho}^{C^n}_{m,k}\right\}$ respectively, can be found by tracing out the other receiver, viz., $\hat{\rho}^{B^n}_{m,k} \equiv \mathrm{Tr}_{C^n}\left(\hat{\rho}^{B^nC^n}_{m,k}\right)$, etc. A $(2^{nR_B}, 2^{nR_C}, n, \epsilon)$ code for this channel consists of an encoder

$$x^n : (W_B, W_C) \to \mathcal{A}^n, \tag{3.24}$$

a positive operator-valued measure (POVM) $\{\Lambda_{mk}\}$ on $\mathcal{B}^n$ and a POVM $\{\Lambda'_k\}$ on $\mathcal{C}^n$ which satisfy[4]

$$\mathrm{Tr}\left(\hat{\rho}^{B^nC^n}_{m,k}(\Lambda_{mk} \otimes \Lambda'_k)\right) \geq 1 - \epsilon \tag{3.25}$$

[4] $\mathcal{A}^n$, $\mathcal{B}^n$, and $\mathcal{C}^n$ are the n channel use alphabets of Alice, Bob, and Charlie, with respective sizes $|\mathcal{A}^n|$, $|\mathcal{B}^n|$, and $|\mathcal{C}^n|$.

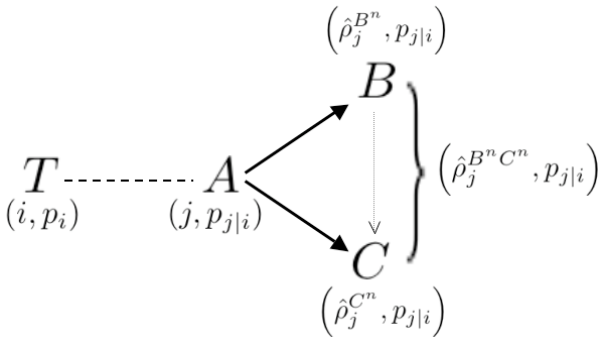


Figure 3-4: Schematic diagram of the degraded single-mode bosonic broadcast channel. The transmitter Alice (A) encodes her messages to Bob (B) and Charlie (C) in a classical index j, and, over n successive uses of the channel, creates a bipartite state $\hat{\rho}_j^{B^n C^n}$ at the receivers.

for every $(m, k) \in (W_B, W_C)$. A rate-pair (R_B, R_C) is *achievable* if there exists a sequence of $(2^{nR_B}, 2^{nR_C}, n, \epsilon_n)$ codes with $\epsilon_n \to 0$. The classical *capacity region* of the broadcast channel is defined as the convex hull of the closure of all achievable rate pairs (R_B, R_C). The classical capacity region of the two-user degraded quantum broadcast channel $\mathcal{N}_{A-BC}$ was recently derived by Yard et. al. [52], and can be expressed in terms of the Holevo information [27, 28, 29],

$$\chi(p_j, \hat{\sigma}_j) \equiv S\left(\sum_j p_j \hat{\sigma}_j\right) - \sum_j p_j S(\hat{\sigma}_j), \tag{3.26}$$

where $\{p_j\}$ is a probability distribution associated with the density operators $\hat{\sigma}_j$, and $S(\hat{\rho}) \equiv -\mathrm{Tr}(\hat{\rho} \log \hat{\rho})$ is the von Neumann entropy of the quantum state $\hat{\rho}$. Because χ may not be additive, the rate region (R_B, R_C) of the degraded broadcast channel

must be computed by maximizing over successive uses of the channel, i.e., for n uses

$$\begin{aligned} R_B &\leq \sum_i p_i \chi\big(p_{j|i}, \mathcal{N}_{A-B}^{\otimes n}(\hat{\rho}_j^{A^n})\big)/n \\ &= \frac{1}{n}\sum_i p_i \left[S\left(\sum_j p_{j|i}\hat{\rho}_j^{B^n}\right) - \sum_{i,j} p_{j|i} S\big(\hat{\rho}_j^{B^n}\big)\right], \quad \text{and} \qquad (3.27) \\ R_C &\leq \chi\left(p_i, \sum_j p_{j|i}\mathcal{N}_{A-C}^{\otimes n}(\hat{\rho}_j^{A^n})\right)/n \\ &= \frac{1}{n}\left[S\left(\sum_{i,j} p_i p_{j|i}\hat{\rho}_j^{C^n}\right) - \sum_i p_i S\left(\sum_j p_{j|i}\hat{\rho}_j^{C^n}\right)\right], \qquad (3.28) \end{aligned}$$

where $j \equiv (m,k)$ is a collective index and the states $\left\{\hat{\rho}_j^{A^n}\right\}$ live in the Hilbert space $\mathcal{H}^{\otimes n}$ of n successive uses of the broadcast channel[5]. The probabilities $\{p_i\}$ form a distribution over an auxiliary classical alphabet $\mathcal{T}$, of size $|\mathcal{T}|$, satisfying $|\mathcal{T}| \leq \min\{|\mathcal{A}|^n, |\mathcal{B}|^{2n} + |\mathcal{C}|^{2n} - 1\}$. The ultimate rate-region is computed by maximizing the region specified by Eqs. (3.27) and (3.28)[6], over $\{p_i\}$, $\left\{p_{j|i}\right\}$, $\left\{\hat{\rho}_j^{A^n}\right\}$, and n, subject to the cardinality constraint on $|\mathcal{T}|$. Fig. 3-4 illustrates the setup of the two-user degraded quantum channel.

[5]Note that, as the actual n-channel-use quantum states sent out by Alice $\hat{\rho}_j^{A^n}$ do not appear in the expressions for R_B or R_C in Eqs. (3.27) and (3.28), the quantum broadcast channel (set up to transmit classical information to multiple receivers) may be seen without any ambiguity, as a cq-broadcast channel, in which Alice's n-use alphabet A^n is a classical random variable, that takes values on a classical index set $\{j\}$ over n successive uses of the channel.

[6] **An alternative notation used in the literature** — An alternative notation, widely used in published literature on quantum information theory, employs $I(A;B)_\rho \equiv H(A)_\rho - H(A|B)_\rho$ to denote the Holevo information between (classical or quantum) systems A and B in a joint state ρ. The classical capacity region of the quantum degraded broadcast channel expressed in this notation closely resembles that of the classical degraded broadcast channel. Consider a degraded broadcast channel $\mathcal{N}^{\mathcal{A}\to BC}$ with n-use conditional density matrices $\left\{\rho_j^{B^nC^n}\right\}$. The capacity region for Alice (A) to send information to Bob (B) and Charlie (C) at rates R_B and R_C respectively is the convex hull of the closure of all (R_B, R_C) satisfying

$$R_B \leq I(A^n; B^n|T)_\sigma/n \qquad (3.29)$$
$$R_C \leq I(T; C^n)_\sigma/n \qquad (3.30)$$

for some $n \geq 1$ and some $p_{T,A^n}(i,j)$ giving rise to the state $\sigma^{TA^nB^nC^n} = \bigoplus_{i,j} p_T(i) p_{A^n|T}(j|i) \rho_j^{B^nC^n}$.

3.3.2 Quantum degraded broadcast channel with M receivers

In this section, we generalize the capacity region of the two-receiver quantum degraded broadcast channel in the previous section, to an arbitrary number of receivers. Using this result, later in this chapter, we evaluate the capacity region of the bosonic broadcast channel with an arbitrary number of receivers. The M-receiver quantum broadcast channel $\mathcal{N}_{A-Y_0\ldots Y_{M-1}}$ is a quantum channel from a sender Alice (A) to M independent receivers $Y_0, \ldots, Y_{M-1}$. The quantum channel from A to Y_0 is obtained by tracing out all the other receivers from the channel map, i.e., $\mathcal{N}_{A-Y_0} \equiv \mathrm{Tr}_{Y_1,\ldots,Y_{M-1}}\left(\mathcal{N}_{A-Y_0\ldots Y_{M-1}}\right)$, with a similar definition for $\mathcal{N}_{A-Y_k}$ for $k \in \{1, \ldots, M-1\}$. We say that a broadcast channel $\mathcal{N}_{A-Y_0\ldots Y_{M-1}}$ is *degraded* if there exists a series of *degrading channels* $\mathcal{N}^{\mathrm{deg}}_{Y_k-Y_{k+1}}$ from Y_k to Y_{k+1}, for $k \in \{0, \ldots, M-2\}$, satisfying

$$\mathcal{N}_{A-Y_{M-1}} = \mathcal{N}^{\mathrm{deg}}_{Y_{M-2}-Y_{M-1}} \circ \mathcal{N}^{\mathrm{deg}}_{Y_{M-3}-Y_{M-2}} \circ \ldots \circ \mathcal{N}^{\mathrm{deg}}_{Y_0-Y_1} \circ \mathcal{N}_{A-Y_0}. \tag{3.31}$$

The M-receiver degraded broadcast channel (see Fig. 3-5) describes a physical scenario in which for each successive n uses of the channel $\mathcal{N}_{A-Y_0\ldots Y_{M-1}}$ Alice communicates a randomly generated classical message $(m_0, \ldots, m_{M-1}) \in (W_0, \ldots, W_{M-1})$ to the receivers $Y_0, \ldots, Y_{M-1}$, where the message-sets W_k are sets of classical indices of sizes 2^{nR_k}, for $k \in \{0, \ldots, M-1\}$. The messages $(m_0, \ldots, m_{M-1})$ are assumed to be independent and uniformly distributed over $(W_0, \ldots, W_{M-1})$, i.e.,

$$p_{W_0,\ldots,W_{M-1}}(m_0, \ldots, m_{M-1}) = \prod_{k=0}^{M-1} p_{W_k}(m_k) = \prod_{k=0}^{M-1} \frac{1}{2^{nR_k}} \tag{3.32}$$

Because of the degraded nature of the channel, given that the transmission rates are within the capacity region and proper encoding and decoding is employed at the transmitter and at the receivers, Y_0 can decode the entire message M-tuple $(m_0, \ldots, m_{M-1})$, Y_1 can decode the reduced message $(M-1)$-tuple $(m_1, \ldots, m_{M-1})$, and so on, until the noisiest receiver Y_{M-1} can only decode the single message-

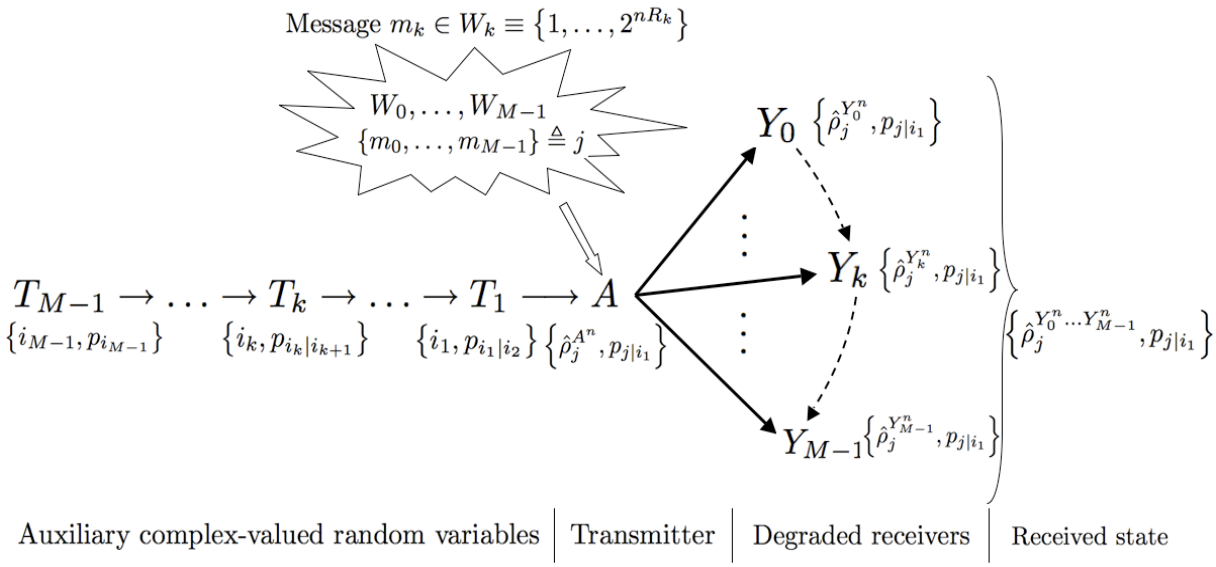


Figure 3-5: This figure summarizes the setup of the transmitter and the channel model for the M-receiver quantum degraded broadcast channel. In each successive n uses of the channel, the transmitter A sends a randomly generated classical message $(m_0, \ldots, m_{M-1}) \in (W_0, \ldots, W_{M-1})$ to the M receivers $Y_0, \ldots, Y_{M-1}$, where the message-sets W_k are sets of classical indices of sizes 2^{nR_k}, for $k \in \{0, \ldots, M-1\}$. The dashed arrows indicate the direction of degradation, i.e., Y_0 is the least noisy receiver, and Y_{M-1} is the noisiest receiver. In this degraded channel model, the quantum state received at the receiver Y_k, $\hat{\rho}^{Y_k}$ can always be reconstructed from the quantum state received at the receiver $Y_{k'}$, $\hat{\rho}^{Y_{k'}}$, for $k' < k$, by passing $\hat{\rho}^{Y_{k'}}$ through a trace-preserving completely positive map (a quantum channel). For sending the classical message $(m_0, \ldots, m_{M-1}) \triangleq j$, Alice chooses a n-use state (codeword) $\hat{\rho}_j^{A^n}$ using a prior distribution $p_{j|i_1}$, where i_k denotes the complex values taken by an auxiliary random variable T_k. It can be shown that, in order to compute the capacity region of the quantum degraded broadcast channel, we need to choose $M-1$ complex valued auxiliary random variables with a Markov structure as shown above, i.e., $T_{M-1} \to T_{M-2} \to \ldots \to T_k \to \ldots \to T_1 \to A^n$ is a Markov chain.

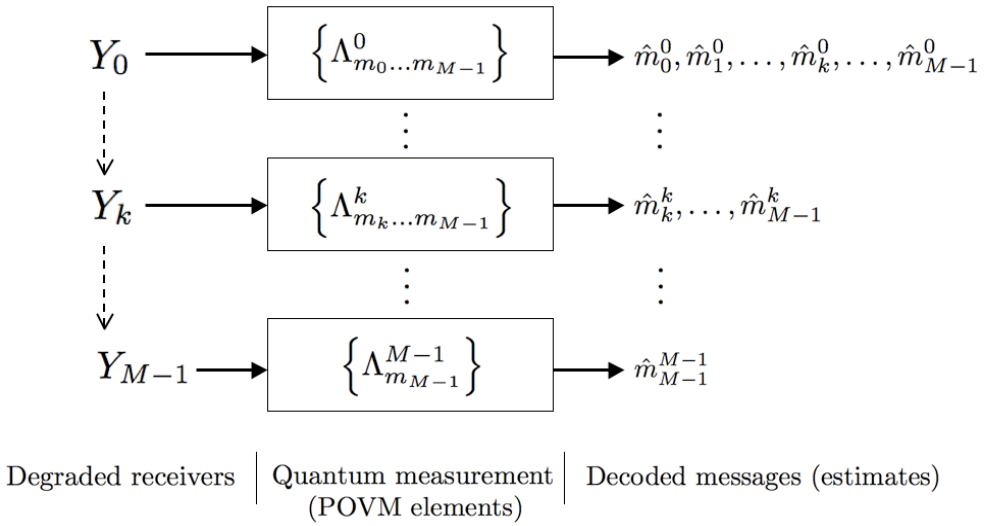


Figure 3-6: This figure illustrates the decoding end of the M-receiver quantum degraded broadcast channel. The decoder consists of a set of measurement operators, described by positive operator-valued measures (POVMs) for each receiver; $\left\{\Lambda^0_{m_0 \ldots m_{M-1}}\right\}$, $\left\{\Lambda^1_{m_1 \ldots m_{M-1}}\right\}$, ..., $\left\{\Lambda^{M-1}_{m_{M-1}}\right\}$ on ${\mathcal{Y}_0}^n$, ${\mathcal{Y}_1}^n$, ..., ${\mathcal{Y}_{M-1}}^n$ respectively. Because of the degraded nature of the channel, if the transmission rates are within the capacity region and proper encoding and decoding are employed at the transmitter and at the receivers respectively, Y_0 can decode the entire message M-tuple to obtain estimates $(\hat{m}^0_0, \ldots, \hat{m}^0_{M-1})$, Y_1 can decode the reduced message $(M-1)$-tuple to obtain its own estimates $(\hat{m}^1_1, \ldots, \hat{m}^1_{M-1})$, and so on, until the noisiest receiver Y_{M-1} can only decode the single message-index m_{M-1} to obtain an estimate $\hat{m}^{M-1}_{M-1}$. Even though the less noisy receivers can decode the messages of the noisier receivers, the message m_k is intended to be sent to receiver Y_k, $\forall k$. Hence, when we say that a broadcast channel is operating at a rate $(R_0, \ldots, R_{M-1})$, we mean that the message m_k is reliably decoded by receiver Y_k at the rate R_k bits per channel use.

index m_{M-1}. To convey the message-set[7] $\mathbf{m}_0^{M-1}$, Alice prepares n-channel use states that, after transmission through the channel, result in M-partite conditional density matrices $\left\{\hat{\rho}_{\mathbf{m}_0^{M-1}}^{Y_0^n \ldots Y_{M-1}^n}\right\}$, $\forall \mathbf{m}_0^{M-1} \in \mathbf{W}_0^{M-1}$. The quantum states received by a particular receiver, say Y_0, can be found by tracing out the other receivers, viz. $\hat{\rho}_{\mathbf{m}_0^{M-1}}^{Y_0^n} \equiv \mathrm{Tr}_{Y_1^n,\ldots,Y_{M-1}^n}\left(\hat{\rho}_{\mathbf{m}_0^{M-1}}^{Y_0^n \ldots Y_{M-1}^n}\right)$, etc. Fig. 3-6 illustrates this decoding process.

A $(2^{nR_0}, \ldots, 2^{nR_{M-1}}, n, \epsilon)$ code for this channel consists of an encoder

$$x^n : (\mathbf{W}_0^{M-1}) \to \mathcal{A}^n, \tag{3.33}$$

a set of positive operator-valued measures (POVMs) — $\left\{\Lambda_{m_0 \ldots m_{M-1}}^0\right\}$, $\left\{\Lambda_{m_1 \ldots m_{M-1}}^1\right\}$, …, $\left\{\Lambda_{m_{M-1}}^{M-1}\right\}$ on ${\mathcal{Y}_0}^n$, ${\mathcal{Y}_1}^n$, …, ${\mathcal{Y}_{M-1}}^n$ respectively, such that the mean probability of a collective correct decision satisfies[8]

$$\mathrm{Tr}\left(\hat{\rho}_{\mathbf{m}_0^{M-1}}^{Y_0^n \ldots Y_{M-1}^n}\left(\bigotimes_{k=0}^{M-1} \Lambda_{m_k \ldots m_{M-1}}^k\right)\right) \geq 1 - \epsilon, \tag{3.34}$$

for $\forall \mathbf{m}_0^{M-1} \in \mathbf{W}_0^{M-1}$. A rate M-tuple $(R_0, \ldots, R_{M-1})$ is *achievable* if there exists a sequence of $(2^{nR_0}, \ldots, 2^{nR_{M-1}}, n, \epsilon)$ codes with $\epsilon_n \to 0$. The classical *capacity region* of the broadcast channel is defined as the convex hull of the closure of all achievable rate M-tuples $(R_0, \ldots, R_{M-1})$. The classical capacity region of the two-user degraded quantum broadcast channel with discrete alphabet was derived by Yard et. al. [52], and we used the infinite-dimensional extension of Yard et. al.'s capacity theorem to prove the capacity region of the bosonic broadcast channel, subject to the minimum output entropy conjecture 2. The capacity region of the degraded quantum broadcast channel can easily be extended to the case of an arbitrary number M, of receivers. For notational similarity to the capacity region of the classical degraded broadcast channel, we state the capacity theorem first, using the shorthand notation for Holevo

[7] From here on, we use the shorthand notation $\mathbf{m}_0^{M-1}$ to denote the message M-tuple $(m_0, \ldots, m_{M-1})$. Similarly, the notation $\mathbf{W}_k^{M-1}$ will be used to denote the set $(W_k, \ldots, W_{M-1})$. We will also use the shorthand notation for probability distributions, such as $p_{\mathbf{W}_1^{M-1}}(\mathbf{m}_1^{M-1}) \triangleq p_{W_1,\ldots,W_{M-1}}(m_1, \ldots, m_{M-1})$.

[8] $\mathcal{A}^n$ and ${\mathcal{Y}_k}^n$ are the n channel use alphabets of Alice, and the k^{th} receiver Y_k respectively, with respective sizes $|\mathcal{A}^n|$ and $|{\mathcal{Y}_k}^n|$, for $k \in [0, \ldots, M-1]$.

information we introduced in footnote 6 earlier in this chapter.

Theorem 3.1 — The capacity region of the M-receiver degraded broadcast channel $\mathcal{N}_{A-Y_0\ldots Y_{M-1}}$, as defined in Eq. (3.31), is given by

$$\begin{aligned} R_0 &\leq \frac{1}{n} I\left(A^n; Y_0^n | T_1\right), \\ R_k &\leq \frac{1}{n} I\left(T_k; Y_k^n | T_{k+1}\right) \quad \forall k \in \{1, \ldots, M-2\}, \\ R_{M-1} &\leq \frac{1}{n} I\left(T_{M-1}; Y_{M-1}^n\right), \end{aligned} \tag{3.35}$$

where T_k, $k \in \{1, \ldots, M-1\}$ form a set of auxiliary complex valued random variables such that $T_{M-1} \to T_{M-2} \to \ldots \to T_k \to \ldots \to T_1 \to A^n$ is a Markov chain[9], i.e.,

$$p_{T_{M-1},\ldots,T_1,A^n}(i_{M-1}, \ldots, i_1, j) = p_{T_{M-1}}(i_{M-1}) \left(\prod_{k=M-1}^{2} p_{T_{k-1}|T_k}(i_{k-1}|i_k) \right) p_{A^n|T_1}(j|i_1). \tag{3.36}$$

In order to find the optimum capacity region, the above rate region must be optimized over the joint distribution $p_{T_{M-1},\ldots,T_1,A^n}(i_{M-1}, \ldots, i_1, j)$. As Holevo information is not necessarily additive (unlike Shannon mutual information), the rate region must also be optimized over the codeword block-length n. The above Markov chain structure of the auxiliary random variables T_k, $k \in \{1, \ldots, M-1\}$ is shown to be optimal in the converse proof which proves the optimality of the above capacity region without assuming any special structure of the auxiliary random variables. Also, note the striking similarity of the expressions for the capacity region given above, with the capacity region of the classical M-receiver degraded broadcast channel, given in Eqs. (3.8). Holevo information takes place of Shannon mutual information in the quantum case, and because of superadditivity of Holevo information, an additional regularization over number of channel uses n, is required.

Proof — The proof of the achievability and converse to the above capacity region is a straightforward extension of Yard et. al.'s two-receiver degraded broadcast channel capacity region. The proof, though simple, involves notational complexity. In order

[9]Here, we have used A^n to denote a classical random variable with a slight abuse of notation. See footnote 5.

to preserve the flow of this chapter, we have omitted the formal proof of the M-receiver quantum degraded broadcast capacity region from this section, but for the sake of completeness and for the more interested readers, we have included the proof (*achievability* for $M = 3$ with a brief sketch of the general case, and *converse* for the general M-receiver case) in Appendix B.

M-receiver degraded broadcast capacity region in the Holevo information ($\chi(p_i, \hat{\rho}_i)$) notation

The capacity region above can be re-cast in the Holevo-information notation that we used earlier in this chapter for the two-receiver quantum broadcast channel. For the channel model of the multiple-user quantum degraded broadcast channel we described in the section above (pictorially depicted in Fig. 3-5), our proposed capacity region

(in Eqs. (3.35)) can alternatively be expressed as[10]

$$\begin{aligned}
R_0 &\leq \frac{1}{n}\sum_{i_1} p_{T_1}(i_1)\chi\left(p_{A^n|T_1}(j|i_1), \hat{\rho}_j^{Y_0^n}\right) \\
&= \frac{1}{n}\sum_{i_1} p_{T_1}(i_1)\left[S\left(\sum_j p_{A^n|T_1}(j|i_1)\hat{\rho}_j^{Y_0^n}\right) - \sum_j p_{A^n|T_1}(j|i_1) S\left(\hat{\rho}_j^{Y_0^n}\right)\right], \\
R_k &\leq \frac{1}{n}\sum_{i_{k+1}} p_{T_{k+1}}(i_{k+1})\chi\left(p_{T_k|T_{k+1}}(i_k|i_{k+1}), \hat{\rho}_{i_k}^{Y_k^n}\right), \quad \forall k \in \{1,\ldots,M-2\}, \\
&= \frac{1}{n}\sum_{i_{k+1}} p_{T_{k+1}}(i_{k+1})\left[S\left(\sum_{i_k} p_{T_k|T_{k+1}}(i_k|i_{k+1})\hat{\rho}_{i_k}^{Y_k^n}\right) - \sum_{i_k} p_{T_k|T_{k+1}}(i_k|i_{k+1}) S\left(\hat{\rho}_{i_k}^{Y_k^n}\right)\right], \\
R_{M-1} &\leq \frac{1}{n}\chi\left(p_{T_{M-1}}(i_{M-1}), \hat{\rho}_{i_{M-1}}^{Y_{M-1}^n}\right) \\
&= \frac{1}{n} S\left(\sum_{i_{M-1}} p_{T_{M-1}}(i_{M-1})\hat{\rho}_{i_{M-1}}^{Y_{M-1}^n}\right) - \sum_{i_{M-1}} p_{T_{M-1}}(i_{M-1}) S\left(\hat{\rho}_{i_{M-1}}^{Y_{M-1}^n}\right). \qquad (3.38)
\end{aligned}$$

Even though the capacity-region expressions above have been written for a discrete alphabet, in Section 3.4.6, we will generalize it to a continuous alphabet of quantum states over an infinite-dimensional Hilbert space, in which case the summations in Eqs. (3.38) will be replaced by integrals. We will use the infinite-dimensional extension of this capacity theorem in the following section to evaluate the capacity region of the M-receiver bosonic broadcast channel.

[10] In Fig. 3-5, we define $j \triangleq \{m_0, \ldots, m_{M-1}\}$ to be a collective index for the M messages that Alice encodes into her n-use transmitted codeword state $\rho_j^{A^n}$, and $\rho_j^{Y_k^n}$ is defined to be the state received by Y_k over n successive channel uses. We introduce more notation here for conditional received states:

$$\begin{aligned}
\hat{\rho}_{i_1}^{Y_1^n} &\triangleq \sum_j p_{A^n|T_1}(j|i_1)\hat{\rho}_j^{Y_1^n}, \\
\hat{\rho}_{i_k}^{Y_l^n} &\triangleq \sum_{j,i_1,\ldots,i_{k-1}} p_{A^n|T_1}(j|i_1) p_{T_1|T_2}(i_1|i_2)\ldots p_{T_{k-1}|T_k}(i_{k-1}|i_k)\hat{\rho}_j^{Y_l^n} \qquad (3.37)
\end{aligned}$$

3.4 Bosonic Broadcast Channel

3.4.1 Channel model

The two-user noiseless bosonic broadcast channel $\mathcal{N}_{A-BC}$ consists of a collection of spatial and temporal bosonic modes at the transmitter (Alice), that interact with a minimal-quantum-noise environment and split into two sets of spatio-temporal modes en route to two independent receivers (Bob and Charlie). The multi-mode two-user bosonic broadcast channel $\mathcal{N}_{A-BC}$ is given by $\bigotimes_s \mathcal{N}_{A_s-B_sC_s}$, where $\mathcal{N}_{A_s-B_sC_s}$ is the broadcast-channel map for the sth mode, which can be obtained from the Heisenberg evolutions

$$\hat{b}_s = \sqrt{\eta_s}\,\hat{a}_s + \sqrt{1-\eta_s}\,\hat{e}_s, \quad \text{and} \tag{3.39}$$

$$\hat{c}_s = \sqrt{1-\eta_s}\,\hat{a}_s - \sqrt{\eta_s}\,\hat{e}_s, \tag{3.40}$$

where $\{\hat{a}_s\}$ are Alice's modal annihilation operators, and $\{\hat{b}_s\}$, $\{\hat{c}_s\}$ are the corresponding modal annihilation operators for Bob and Charlie, respectively. The modal transmissivities $\{\eta_s\}$ satisfy $0 \leq \eta_s \leq 1$, $\forall s$, and the environment modes $\{\hat{e}_s\}$ are in their vacuum states. We will limit our treatment here to the single-mode bosonic broadcast channel, as the capacity of the multi-mode channel can in principle be obtained by summing up capacities of all spatio-temporal modes and maximizing the sum capacity region subject to an overall input-power budget using Lagrange multipliers, cf. [55], where this was done for the capacity of the multi-mode single-user lossy bosonic channel.

We are interested in finding the capacity region (R_B, R_C) of achievable rate-pairs at which Alice can send information to Bob and Charlie, with vanishingly low probabilities of error. Alice is constrained by a mean photon-number (power) constraint $\langle \hat{a}^\dagger \hat{a} \rangle \leq \bar{N}$. The principal result we have for the single-mode bosonic broadcast channel stems from the fact that the bosonic broadcast channel is a degraded broadcast channel, and hence the capacity theorem we stated in the previous section can be adapted to this case by extending the result to infinite-dimensional Hilbert spaces.

Our capacity result depends on a minimum output entropy conjecture (dealt with in detail in chapter 4). Assuming this conjecture to be true, we prove in this section, that the ultimate capacity region of the single-mode noiseless bosonic broadcast channel (see Fig. 3-7) with a mean input photon-number constraint $\langle \hat{a}^\dagger \hat{a} \rangle \leq \bar{N}$ is given by

$$R_B \leq g(\eta\beta\bar{N}), \quad \text{and} \tag{3.41}$$

$$R_C \leq g((1-\eta)\bar{N}) - g((1-\eta)\beta\bar{N}), \tag{3.42}$$

for $0 \leq \beta \leq 1$, where $g(x) = (1+x)\ln(1+x) - x\ln(x)$. We further prove, assuming the validity of the minimum output entropy conjecture, that this rate region is additive and is achievable with single channel use coherent-state encoding with the following Gaussian prior and conditional distributions:

$$p_T(\tau) = \frac{1}{\pi\bar{N}} \exp\left(-\frac{|\tau|^2}{\bar{N}}\right), \quad \text{and} \tag{3.43}$$

$$p_{A|T}(\alpha|\tau) = \frac{1}{\pi\bar{N}\beta} \exp\left(-\frac{|\sqrt{1-\beta}\,\tau - \alpha|^2}{\bar{N}\beta}\right), \tag{3.44}$$

where T is a complex-valued auxiliary classical random variable taking values $\tau \in \mathbb{C}$, and A is a complex-valued classical random variable taking value $\alpha \in \mathbb{C}$ when Alice sends out the single-mode coherent state $|\alpha\rangle$.

3.4.2 Degraded broadcast condition

Lemma 3.2 — The pure-loss bosonic broadcast channel $\mathcal{N}_{A-BC}$, with transmissity $\eta > 1/2$, is stochastically equivalent to a degraded cq-broadcast channel $A \to B \to C$, in which the degrading channel from Bob to Charlie $\mathcal{N}^{\text{deg}}_{B-C}$ is another beam splitter with transmissivity $\eta' = (1-\eta)/\eta$ (Fig. 3-8).

Proof — Refer to Figure 3-8. The annihilation operator $\hat{g}$ corresponds to the output of the degrading channel, which is excited in a state $\hat{\rho}_g$. In order to prove that the bosonic broadcast channel $\mathcal{N}_{A-BC}$ is indeed equivalent to a degraded broadcast channel, we need to show that the states $\hat{\rho}_g$ and $\hat{\rho}_c$ are identical quantum states,

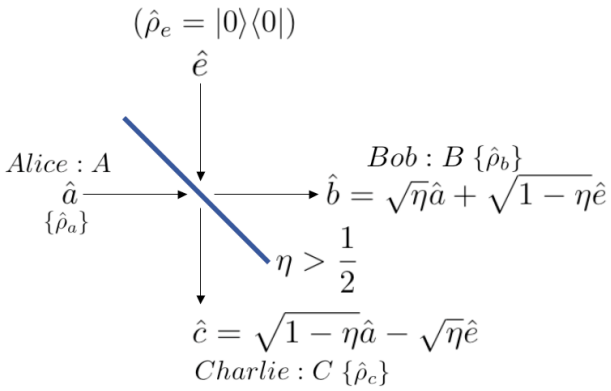


Figure 3-7: A single-mode noiseless bosonic broadcast channel with two receivers $\mathcal{N}_{A-BC}$, can be envisioned as a beam splitter with transmissivity η. With $\eta > 1/2$, the bosonic broadcast channel reduces to a degraded quantum broadcast channel, where Bob (B) is the less-noisy receiver and Charlie (C) is the more noisy (degraded) receiver.

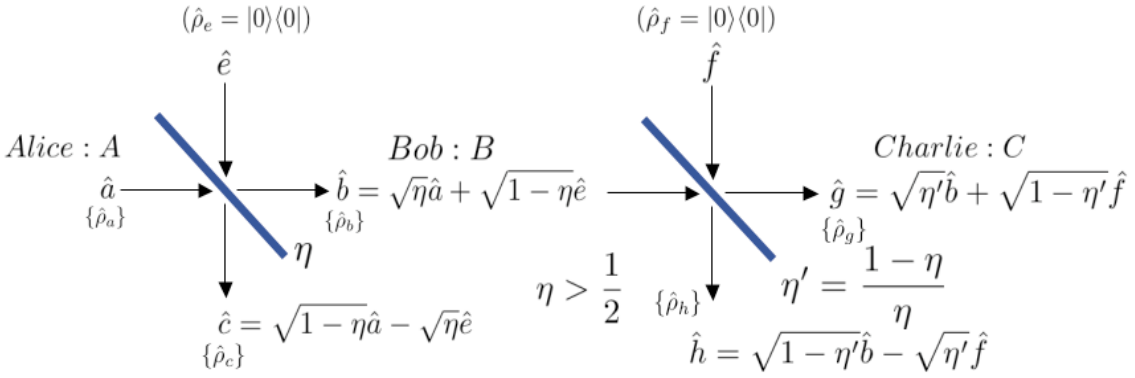


Figure 3-8: The stochastically degraded version of the single-mode bosonic broadcast channel

i.e., the classical statistics of the results of measuring the states $\hat{\rho}_g$ and $\hat{\rho}_c$ using any POVM, will be exactly the same, provided $\eta > 1/2$.

Let us compute the antinormally ordered characteristic functions of the states $\hat{\rho}_c$ and $\hat{\rho}_g$. We have

$$\begin{aligned}
\chi_A^{\hat{\rho}_c}(\zeta) &= \langle e^{-\zeta^* \hat{c}} e^{\zeta \hat{c}^\dagger} \rangle \\
&= \langle e^{-\zeta^* \sqrt{1-\eta}\hat{a}} e^{\zeta \sqrt{1-\eta}\hat{a}^\dagger} \rangle \langle e^{\zeta^* \sqrt{\eta}\hat{e}} e^{-\zeta \sqrt{\eta}\hat{e}^\dagger} \rangle \\
&= \chi_A^{\hat{\rho}_a}(\sqrt{1-\eta}\zeta) \chi_A^{\hat{\rho}_e}(-\sqrt{\eta}\zeta) \\
&= \chi_A^{\hat{\rho}_a}(\sqrt{1-\eta}\zeta) e^{-\eta|\zeta|^2}, \qquad (3.45)
\end{aligned}$$

and

$$\begin{aligned}
\chi_A^{\hat{\rho}_g}(\zeta) &= \chi_A^{\hat{\rho}_b}(\sqrt{\eta'}\zeta) \chi_A^{\hat{\rho}_f}(\sqrt{1-\eta'}\zeta) \\
&= \chi_A^{\hat{\rho}_a}(\sqrt{\eta\eta'}\zeta) \chi_A^{\hat{\rho}_e}(\sqrt{\eta'(1-\eta)}\zeta) \\
&\times \chi_A^{\hat{\rho}_f}(\sqrt{1-\eta'}\zeta) \\
&= \chi_A^{\hat{\rho}_a}(\sqrt{\eta\eta'}\zeta) e^{-\eta'(1-\eta)|\zeta|^2} e^{-(1-\eta')|\zeta|^2} \\
&= \chi_A^{\hat{\rho}_a}(\sqrt{1-\eta}\zeta) e^{-\eta|\zeta|^2}, \qquad (3.46)
\end{aligned}$$

so that $\chi_A^{\hat{\rho}_c}(\zeta) = \chi_A^{\hat{\rho}_g}(\zeta), \quad \forall \hat{\rho}_a$. Inverse Fourier transforming these characteristic functions thus yields the same expressions for $\hat{\rho}_c$ and $\hat{\rho}_g$. Hence $\hat{\rho}_g$ and $\hat{\rho}_c$ are identical states, and the pure-loss bosonic broadcast channel $\mathcal{N}_{A-BC}$ is a degraded broadcast channel for $\eta > 1/2$.

3.4.3 Noiseless bosonic broadcast channel with two receivers

It is known [10, 7, 39] that coherent-state modulation using isotropic Gaussian prior distribution achieves the ultimate classical capacity (maximizes the Holevo information) for a single-mode pure-loss bosonic channel. It is also known however, that for quantum multiple-access channels, coherent-state encodings are not optimal [11].

So it is not clear, at the outset, whether coherent-state encoding will be capacity achieving for the bosonic broadcast channel. Nevertheless, it is worth assessing the capacity region realized by coherent-state encoding.

Consider the two-user bosonic broadcast channel $\mathcal{N}_{A-BC}$ and assume that Alice has access to all coherent states $|\alpha\rangle$ to encode her information, with a mean photon-number constraint $\langle \hat{a}^\dagger \hat{a} \rangle \leq \bar{N}$. Bob and Charlie thus receive attenuated versions of the coherent states that Alice transmits at each channel use. Let us introduce an auxiliary classical complex-valued random variable T, and an associated coherent-state alphabet $|\tau\rangle$ and prior probability distribution $p_T(\tau)$. Alice transmits coherent states $|\alpha\rangle$ with conditional probability $p_{A|T}(\alpha|\tau)$. The first step towards proving that the ultimate capacity region of the two-user bosonic broadcast channel is given by Eqs. (3.41) and (3.42), is to show that the probability distributions $p_T(\tau)$ and $p_{A|T}(\alpha|\tau)$, as given by Eqs. (3.43) and (3.44), achieve these rates.

Yard et al.'s capacity region in Equations (3.27) and (3.28) require finite-dimensional Hilbert spaces. Nevertheless, we will use their result for the bosonic broadcast channel which has an infinite-dimensional state space, as their result can be extended to infinite-dimensional state spaces by means of a limiting argument.[11]

Theorem 3.3 — Assuming the truth of strong conjecture 2 (see Section 4.1), the ultimate capacity region of the single-mode noiseless bosonic broadcast channel (see Fig. 3-7) with a mean input photon-number constraint $\langle \hat{a}^\dagger \hat{a} \rangle \leq \bar{N}$ is given by

$$R_B \leq g(\eta\beta\bar{N}), \quad \text{and} \tag{3.47}$$

$$R_C \leq g((1-\eta)\bar{N}) - g((1-\eta)\beta\bar{N}), \tag{3.48}$$

[11] When $|T|$ and $|A|$ are finite, and we are using coherent states, we land up with a finite number of possible transmitted states, which leads to a finite number of possible states received by Bob and Charlie. To be more explicit, let us limit the auxiliary-input alphabet (T) – and hence the input (A) and the output alphabets (B, and C) – to coherent states in the finite-dimensional subspace spanned by the Fock states $\{|0\rangle, |1\rangle, \ldots, |K\rangle\}$, where $K \gg \bar{N}$. Applying Yard et al.'s thereom to the Hilbert space spanned by these states then gives us a broadcast channel capacity region that must be strictly an inner bound of the rate region given by Eqs. (3.49) and (3.50). In the limit that we choose K sufficiently large, (maintaining the cardinality condition $|T| \leq |A|$ that is required by the theorem), clearly the rate-region expressions given by Yard et. al.'s theorem can be brought to as close as we wish, to those given by Eqs. (3.49) and (3.50).

for $0 \leq \beta \leq 1$, where $g(x) = (1+x)\ln(1+x) - x\ln(x)$. This rate region is additive and is achievable with single channel use coherent-state encoding with the Gaussian prior and conditional distributions given in Eqs. (3.43) and (3.44).

Proof [Achievability] — Using the infinite-dimensional (continuous-variable) extension of Eqs. (3.27) and (3.28), the $n = 1$ rate-region for the bosonic broadcast channel using coherent-state encoding is given by:

$$\begin{aligned} R_B &\leq \int p_T(\tau) S\left(\int p_{A|T}(\alpha|\tau) |\sqrt{\eta}\,\alpha\rangle\langle\sqrt{\eta}\,\alpha|\,\mathrm{d}^2\alpha\right)\mathrm{d}^2\tau \qquad (3.49) \\ R_C &\leq S\left(\int p_T(\tau) p_{A|T}(\alpha|\tau) |\sqrt{1-\eta}\,\alpha\rangle\langle\sqrt{1-\eta}\,\alpha|\,\mathrm{d}^2\alpha\,\mathrm{d}^2\tau\right) \\ &\quad - \int p_T(\tau) S\Bigg(\int p_{A|T}(\alpha|\tau) \times \\ &\qquad |\sqrt{1-\eta}\,\alpha\rangle\langle\sqrt{1-\eta}\,\alpha|\,\mathrm{d}^2\alpha\Bigg)\mathrm{d}^2\tau, \qquad (3.50) \end{aligned}$$

where we need to maximize the bounds for R_B and R_C over all joint distributions $p_T(\tau)p_{A|T}(\alpha|\tau)$ subject to $\langle|\alpha|^2\rangle \leq \bar{N}$. Note that A and T are complex-valued random variables, and the second term in the R_B bound (3.27) vanishes, because the von Neumann entropy of a pure state is zero. Substituting Eqs. (3.43) and (3.44) into Eqs. (3.49) and (3.50), shows that the rate-region Eqs. (3.41) and (3.42) is achievable using single-use coherent state encoding.

Proof [Converse] — Assume that the rate pair (R_B, R_C) is achievable. Let $\{x^n(m,k)\}$, and POVMs $\{\Lambda_{mk}\}$ and $\{\Lambda'_k\}$ comprise any $(2^{nR_B}, 2^{nR_C}, n, \epsilon)$ code in the achieving sequence. Suppose that Bob and Charlie store their decoded messages in the classical registers $\hat{W}_B$ and $\hat{W}_C$ respectively. Let us use $p_{W_B,W_C}(m,k) = p_{W_B}(m)p_{W_C}(k)$ to denote the joint probability mass function of the independent message registers W_B and W_C. As (R_B, R_C) is an achievable rate-pair, there must exist $\epsilon'_n \to 0$, such that

$$\begin{aligned} nR_C &= H(W_C) \\ &\leq I(W_C; \hat{W}_C) + n\epsilon'_n \\ &\leq \chi(p_{W_C}(k), \hat{\rho}_k^{C^n}) + n\epsilon'_n, \qquad (3.51) \end{aligned}$$

where $I(W_C;\hat{W}_C) \equiv H(\hat{W}_C) - H(\hat{W}_C|W_C)$ is the Shannon mutual information, and $\hat{\rho}_k^{C^n} = \sum_m p_{W_B}(m)\hat{\rho}_{m,k}^{C^n}$. The second line follows from Fano's inequality and the third line follows from Holevo's bound[12]. Similarly, for an $\epsilon_n'' \to 0$, we can bound nR_B as

$$\begin{aligned} nR_B &= H(W_B) \\ &\leq I(W_B;\hat{W}_B) + n\epsilon_n'' \\ &\leq \chi(p_{W_B}(m), \hat{\rho}_m^{B^n}) + n\epsilon_n'' \\ &\leq \sum_k p_{W_C}(k)\chi(p_{W_B}(m), \hat{\rho}_{m,k}^{B^n}) + n\epsilon_n'', \end{aligned} \tag{3.52}$$

where the three lines above follow from Fano's inequality, Holevo's bound and the concavity of Holevo information. In order to prove the converse, we now need to show that there exists a number $\beta \in [0,1]$, such that

$$\begin{aligned} \sum_k p_{W_C}(k)\chi(p_{W_B}(m), \hat{\rho}_{m,k}^{B^n}) &\leq ng(\eta\beta\bar{N}), \\ \text{and} \quad \chi(p_{W_C}(k), \hat{\rho}_k^{C^n}) &\leq ng((1-\eta)\bar{N}) - ng((1-\eta)\beta\bar{N}). \end{aligned}$$

From the non-negativity of the von Neumann entropy $S\big(\hat{\rho}_{m,k}^{B^n}\big)$, it follows that

$$\sum_k p_{W_C}(k)\chi(p_{W_B}(m), \hat{\rho}_{m,k}^{B^n}) \leq \sum_k p_{W_C}(k) S\left(\sum_m p_{W_B}(m)\hat{\rho}_{m,k}^{B^n}\right),$$

as the second term of the Holevo information above is non-negative. Because the maximum von Neumann entropy of a single-mode bosonic state with $\langle \hat{a}^\dagger \hat{a} \rangle \leq \bar{N}$ is given by $g(\bar{N})$, we have that

$$0 \leq S\big(\hat{\rho}_k^{B^n}\big) \leq \sum_{j=1}^{n} g\big(\eta\bar{N}_{k_j}\big) \leq ng\big(\eta\bar{N}_k\big), \tag{3.53}$$

where $\bar{N}_k \equiv \sum_{j=1}^n \frac{1}{n}\bar{N}_{k_j}$, and $\bar{N}_{k_j}$ is the mean photon number of the j^{th} symbol $\hat{\rho}_k^{B_j^n}$

[12]Holevo's bound [27, 28, 29]: Let X be the input alphabet for a channel, $\{p_i, \hat{\rho}_i\}$ the priors and modulating states, $\{\Pi_j\}$ be a POVM, and Y the resulting output (classical) alphabet. The Shannon mutual information $I(X;Y)$ is upper bounded by the Holevo information $\chi(p_i, \hat{\rho}_i)$

of the n-symbol codeword $\hat{\rho}_k^{B^n}$, for $j \in \{1, \ldots, n\}$. The last inequality above follows because $g(x)$ is concave. Therefore, $\exists \beta_k \in [0,1]$, $\forall k \in W_C$, such that

$$S\left(\hat{\rho}_k^{B^n}\right) = ng\left(\eta\beta_k \bar{N}_k\right), \tag{3.54}$$

because $g(x)$ is a monotonically increasing function of $x \geq 0$. Because of the degraded nature of the channel, Charlie's state can be obtained as the output of a beam splitter whose input states are Bob's state (coupling coefficient $\eta' = (1-\eta)/\eta$ to Charlie) and a vacuum state (coupling coefficient $1-\eta'$ to Charlie). It follows, from assuming the truth of strong conjecture 2 (see chapter 4), that

$$S\big(\hat{\rho}_k^{C^n}\big) \geq ng\big((1-\eta)\beta_k \bar{N}_k\big). \tag{3.55}$$

$\bar{N}$ is the average number of photons per-use at the transmitter (Alice) averaged over the entire codebook. Thus, the mean photon-number of the n-use average codeword at Bob, $\hat{\rho}^{B^n} \equiv \sum_k p_{W_C}(k)\hat{\rho}_k^{B^n}$, is $\eta\bar{N}$. Hence,

$$0 \leq \sum_k p_{W_C}(k) S\big(\hat{\rho}_k^{B^n}\big) \leq S(\hat{\rho}^{B^n}) \leq ng\left(\eta\bar{N}\right), \tag{3.56}$$

where the second inequality follows from the concavity of von Neumann entropy, and the third inequality arises from maximizing the entropy subject to the average photon number constraint. The monotonicity of $g(x)$ then implies that there is a $\beta \in [0,1]$, such that $\sum_k p_{W_C}(k) S\big(\hat{\rho}_k^{B^n}\big) = ng(\eta\beta\bar{N})$. Hence we have,

$$\sum_k p_{W_C}(k)\chi(p_{W_B}(m), \hat{\rho}_{m,k}^{B^n}) \leq ng(\eta\beta\bar{N}). \tag{3.57}$$

for some $\beta \in [0,1]$. Equation (3.54), and the uniform distribution $p_{W_C}(k) = 1/2^{nR_C}$ imply that

$$\sum_k \frac{1}{2^{nR_C}} g\left(\eta\beta_k \bar{N}_k\right) = g\left(\eta\beta\bar{N}\right). \tag{3.58}$$

Using (3.58), the concavity of $g(x)$, and $\eta > 1/2$, we have shown (proof in Appendix C)

that

$$\sum_k \frac{1}{2^{nR_C}} g\left((1-\eta)\beta_k \bar{N}_k\right) \geq g\left((1-\eta)\beta\bar{N}\right). \tag{3.59}$$

From Eq. (3.59), and Eq. (3.55) summed over k, we then obtain

$$\sum_k p_{W_C}(k) S\left(\hat{\rho}_k^{C^n}\right) \geq ng((1-\eta)\beta\bar{N}). \tag{3.60}$$

Finally, writing Charlie's Holevo information as

$$\begin{aligned} \chi(p_{W_C}(k), \hat{\rho}_k^{C^n}) &= S\left(\sum_k p_{W_C}(k)\hat{\rho}_k^{C^n}\right) - \sum_k p_{W_C}(k) S\left(\hat{\rho}_k^{C^n}\right) \\ &\leq ng((1-\eta)\bar{N}) - \sum_k p_{W_C}(k) S\left(\hat{\rho}_k^{C^n}\right), \end{aligned} \tag{3.61}$$

we can use Eq. (3.60) to get

$$\chi(p_{W_C}(k), \hat{\rho}_k^{C^n}) \leq ng((1-\eta)\bar{N}) - ng((1-\eta)\beta\bar{N}), \tag{3.62}$$

which completes the proof. The capacity region is additive, because the achievability part of the proof above shows that a product distribution over single-use coherent-state alphabet achieves the rate region.

3.4.4 Achievable rate region using coherent detection receivers

Unless we have a proof of strong conjecture 2, we cannot assert that Eqs. (3.41) and (3.42) define the capacity region of the two-user bosonic broadcast channel. However, because the rate region specified by these equations is achievable with single-use coherent-state encoding, we know that they comprise an inner bound on the ultimate capacity region. In this regard, it is instructive to examine how the rate region defined by Eqs. (3.41) and (3.42) compares with what can be realized by conventional, coherent detection schemes used in optical communications.

Suppose Alice sends a coherent state $|\alpha\rangle$, into the channel in Fig. 3-7. Bob and Charlie will then receive coherent states $|\sqrt{\eta}\alpha\rangle$ and $|\sqrt{1-\eta}\alpha\rangle$, respectively. More-

over, if Bob and Charlie employ homodyne-detection receivers, with local oscillator phases set to observe the real quadrature, their results of measurement will be $\sqrt{\eta}\Re(\alpha) + \nu_B$ for Bob and $\sqrt{1-\eta}\Re(\alpha) + \nu_C$ for Charlie, where ν_B and ν_C are independent, identically distributed, real-valued Gaussian random variables with variance 1/4 [18]. Similarly, if Bob and Charlie employ heterodyne-detection receivers, their results of measurement will be $\sqrt{\eta}\alpha + z_B$ and $\sqrt{1-\eta}\alpha + z_C$, where z_B and z_C are independent, identically distributed complex-valued zero-mean Gaussian random variables with variance 1/2 [18]. These results imply that the $\eta > 1/2$ bosonic broadcast channel with coherent-state encoding and homodyne detection is a classical degraded scalar-Gaussian broadcast channel, whose capacity region is known to be [3]

$$R_B \leq \frac{1}{2}\ln\left(1 + 4\eta\beta\bar{N}\right) \tag{3.63}$$

$$R_C \leq \frac{1}{2}\ln\left(1 + \frac{4(1-\eta)(1-\beta)\bar{N}}{1 + 4(1-\eta)\beta\bar{N}}\right), \tag{3.64}$$

for $0 \leq \beta \leq 1$. Similarly, the $\eta > 1/2$ bosonic broadcast channel with coherent-state encoding and heterodyne detection is a classical degraded vector-Gaussian broadcast channel, whose capacity region is known to be

$$R_B \leq \ln\left(1 + \eta\beta\bar{N}\right) \tag{3.65}$$

$$R_C \leq \ln\left(1 + \frac{(1-\eta)(1-\beta)\bar{N}}{1 + (1-\eta)\beta\bar{N}}\right), \tag{3.66}$$

for $0 \leq \beta \leq 1$. In Fig. 3-9 we compare the capacity regions attained by a coherent-state input alphabet using homodyne, heterodyne, and optimum reception. As is known for single-user bosonic communication, homodyne detection performs better than heterodyne detection when the transmitters are starved for photons, because it has lower noise. Conversely, heterodyne detection outperforms homodyne detection when the transmitters are photon rich, because it has a factor-of-two bandwidth advantage over homodyne detection. In order to bridge the gap between the coherent-detection capacity regions and the ultimate capacity region, one must use joint detection over long codewords. Future investigation will be needed to develop receivers that

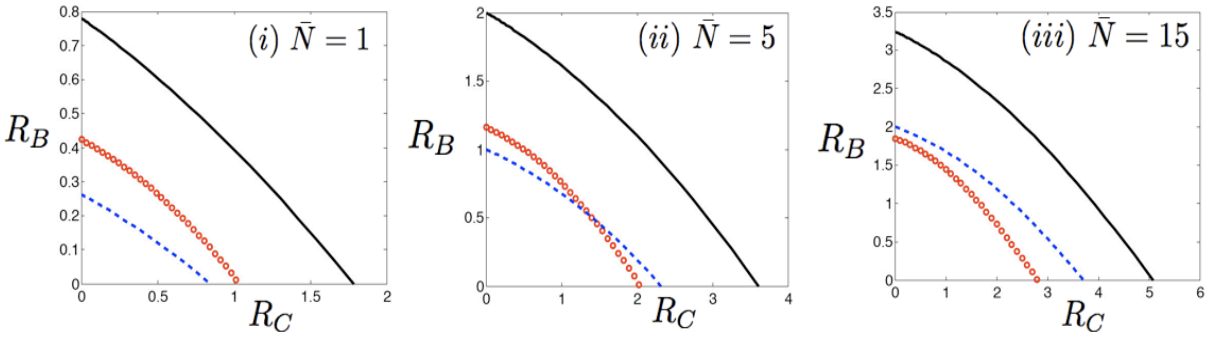


Figure 3-9: Comparison of bosonic broadcast channel capacity regions, in bits per channel use, achieved by coherent-state encoding using homodyne detection (the capacity region lies inside the boundary marked by circles), heterodyne detection (the capacity region lies inside the boundary marked by dashes), and optimum reception (the capacity region lies inside the boundary marked by the solid curve), for $\eta = 0.8$, and $\bar{N} = 1$, 5, and 15.

can approach the ultimate communication rates over the bosonic broadcast channel.

3.4.5 Thermal-noise bosonic broadcast channel with two receivers

Now assume that the environment mode $\hat{e}$ in the bosonic broadcast channel in Fig. 3-7) is in a zero-mean thermal state with mean photon number N (see Fig. 3-10), i.e.,

$$\hat{\rho}_e \equiv \hat{\rho}_{T,N} \triangleq \frac{1}{\pi N} \int e^{-|\mu|^2/N} |\mu\rangle\langle\mu| d\mu. \tag{3.67}$$

Theorem 3.4 — Provided the minimum output entropy conjectures strong conjecture 1 and strong conjecture 3 (see Section 4.1) are true, the capacity region for the bosonic broadcast channel with additive thermal noise, with mean photon number constraint $\bar{N}$ at the input and an additive zero-mean thermal noise with N photons per mode, on average, is given by,

$$R_B \leq g(\eta\beta\bar{N} + (1-\eta)N) - g((1-\eta)N) \tag{3.68}$$

$$R_C \leq g((1-\eta)\bar{N} + \eta N) - g((1-\eta)\beta\bar{N} + \eta N), \tag{3.69}$$

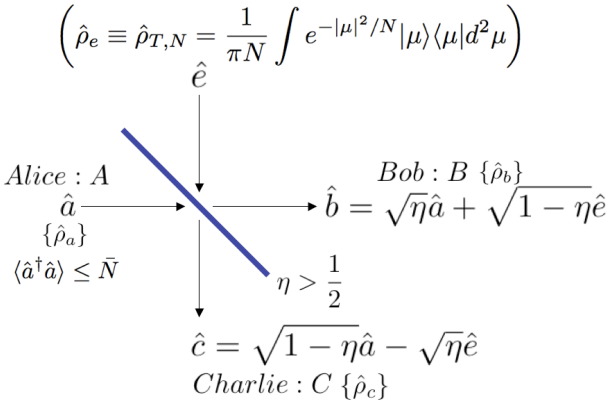


Figure 3-10: A single-mode noiseless bosonic broadcast channel with two receivers $\mathcal{N}_{A-BC}$, with additive thermal noise. The transmitter Alice (A) is constrained to use $\bar{N}$ photons per use of the channel, and the noise (environment) mode is in a zero-mean thermal state $\hat{\rho}_{T,N}$, with mean photon number N. With $\eta > 1/2$, the bosonic broadcast channel reduces to a degraded quantum broadcast channel, where Bob (B) is the less-noisy receiver and Charlie (C) is the more noisy (degraded) receiver. See the degraded version of the channel in Fig. 3-11.

and capacity is achieved using product-coherent-state encoding with a Gaussian prior density as in the case of the noiseless bosonic broadcast channel[13].

Proof [Achievability] — It can be readily verified that the degraded broadcast condition still holds for the case of the bosonic broadcast channel with additive thermal noise (See Fig. 3-11). We generalize Yard et. al.'s rate regions for degraded quantum broadcast channels, from Eqs. (3.27) and (3.28), to the case of the bosonic broadcast channel with coherent-state encoding and additive thermal noise in a similar way to

[13]Note the striking similarity between the expressions for the rate region for the classical Gaussian-noise broadcast channel as given in Eqs. (3.19) and (3.20) and that for the rate region of the bosonic thermal-noise broadcast channel as we propose above in Eqs. (3.68) and (3.69). The expressions for these two rate regions are exactly identical except for the fact that the logarithmic function $g_C(\cdot)$ is replaced by the bosonic thermal-state entropy function $g(\cdot)$ in the quantum case. We will repeatedly encounter in this thesis instances of this analogous role that $g(\cdot)$ plays in the bosonic case, which the logarithmic function $g_C(\cdot)$ does in the classical Gaussian case. The observation of this analogy was one of the key initial hints that led us to conjecture the Entropy Photon-number Inequality (EPnI) [13] in analogy with the Entropy Power Inequality (EPI) of classical information theory. The EPnI subsumes all the three minimum output entropy conjectures that we describe in chapter 4. We will talk about the EPnI in detail in Chapter 5 of this thesis, where we will see why the existence of a simple inverse of $g_C(\cdot)$ (i.e., the $\exp(\cdot)$-function) makes it a great deal easier to prove the EPI as opposed to the EPnI (whose general proof is still an open problem), because the inverse function of $g(\cdot)$ doesn't admit a nice analytic form.

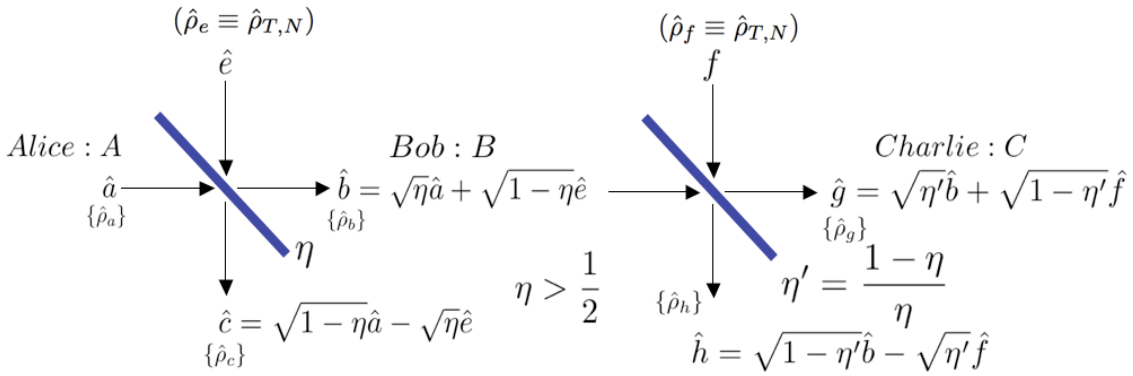


Figure 3-11: The stochastically degraded version of the single-mode bosonic broadcast channel with additive thermal noise.

what we did for the noiseless Broadcast channel[14]:

$$
\begin{aligned}
R_B &\leq \int p_T(\tau) S\left(\int p_{A|T}(\alpha|\tau)\left(\frac{1}{\pi(1-\eta)N}\int e^{-\frac{|\gamma-\sqrt{\eta}\alpha|^2}{(1-\eta)N}}|\gamma\rangle\langle\gamma|\mathrm{d}^2\gamma\right)\mathrm{d}^2\alpha\right)\mathrm{d}^2\tau \\
&- \int\int p_T(\tau)p_{A|T}(\alpha|\tau) S\left(\frac{1}{\pi(1-\eta)N}\int e^{-\frac{|\gamma-\sqrt{\eta}\alpha|^2}{(1-\eta)N}}|\gamma\rangle\langle\gamma|\mathrm{d}^2\gamma\right)\mathrm{d}^2\alpha\mathrm{d}^2\tau \quad (3.70) \\
R_C &\leq S\left(\int p_T(\tau)p_{A|T}(\alpha|\tau)\left(\frac{1}{\pi\eta N}\int e^{-\frac{|\gamma-\sqrt{1-\eta}\alpha|^2}{\eta N}}|\gamma\rangle\langle\gamma|\mathrm{d}^2\gamma\right)\mathrm{d}^2\alpha\mathrm{d}^2\tau\right) \\
&- \int p_T(\tau) S\left(\int p_{A|T}(\alpha|\tau)\left(\frac{1}{\pi\eta N}\int e^{-\frac{|\gamma-\sqrt{1-\eta}\alpha|^2}{\eta N}}|\gamma\rangle\langle\gamma|\mathrm{d}^2\gamma\right)\mathrm{d}^2\alpha\right)\mathrm{d}^2\tau \quad (3.71)
\end{aligned}
$$

where, in order to get the $n = 1$ capacity region, we need to maximize the bounds for R_B and R_C over all complex-valued joint distributions $p_T(\tau)p_{A|T}(\alpha|\tau)$ subject to $\langle|\alpha|^2\rangle \leq \bar{N}$. Note that A and T are two complex-valued random variables, and the second term in the bound for R_B (see Equation (3.27)) is non-zero, because the conditional output states at the two receivers are now mixed states in general. Substituting the distributions from Eqs. (3.43), and (3.44) into the expressions for

[14]Let us limit the auxiliary-input alphabet (T) to coherent states in the finite-dimensional subspace spanned by the Fock states $\{|0\rangle, |1\rangle, \ldots, |K_1\rangle\}$, and limit the thermal-noise state $\hat{\rho}_e$ to the span of $\{|0\rangle, |1\rangle, \ldots, |K_2\rangle\}$, such that $K_1 + K_2 \gg \bar{N} + N$. Applying Yard et al.'s thereom to the Hilbert space spanned by these states then gives us a broadcast channel capacity region that must be strictly an inner bound of the rate region given by Eqs. (3.70) and (3.71). In the limit in which we choose K_1 and K_2 sufficiently large, (maintaining the cardinality condition $|T| \leq |A|$ that is required by the theorem), the rate-region expressions given by Yard et. al.'s theorem can be brought to as close as we wish to that given by Eqs. (3.70) and (3.71).

the rate-bounds in Eqs. (3.70) and (3.71), and using the fact that the von Neumann entropy of a thermal state with mean photon-number N is equal to $g(N)$, we obtain the rate-bounds in the capacity theorem above. It follows that the rate region (3.68), (3.69) is achievable.

Proof [Converse] — Assume that the rate pair (R_B, R_C) is achievable. Let us begin with the same initial steps as in the proof of the converse of the capacity theorem for the noiseless bosonic broadcast channel. Equations (3.51) and (3.52) still hold. Thus, in order to prove the converse for the thermal noise broadcast channel, we now need to show that there exists a number $\beta \in [0, 1]$, such that

$$\sum_k p_{W_C}(k)\chi(p_{W_B}(m), \hat{\rho}^{B^n}_{m,k}) \leq ng(\eta\beta\bar{N} + (1-\eta)N) - ng((1-\eta)N), \tag{3.72}$$

$$\chi(p_{W_C}(k), \hat{\rho}^{C^n}_k) \leq ng((1-\eta)\bar{N} + \eta N) - ng((1-\eta)\beta\bar{N} + \eta N). \tag{3.73}$$

Assuming the truth of strong conjecture 1 (see chapter 4), the minimum entropy of Bob's n-mode state is achieved when Alice sends a product of vacuum states (or a product of arbitrary coherent states). Thus using strong conjecture 1 we have for all $(m, k) \in (W_B, W_C)$,

$$S(\hat{\rho}^{B^n}_{m,k}) \geq ng((1-\eta)N). \tag{3.74}$$

From the non-negativity of Holevo information $\chi(p_{W_B}(m), \hat{\rho}^{B^n}_{m,k})$, it follows that[15]

$$\begin{aligned} S(\hat{\rho}^{B^n}_k) &\geq \sum_m p_{W_B}(m) S(\hat{\rho}^{B^n}_{m,k}) && (3.75)\\ &\geq ng((1-\eta)N). && (3.76) \end{aligned}$$

Let $\bar{N}^A_k = \sum_{j=1}^n \frac{1}{n}\bar{N}^A_{k_j}$, where $\bar{N}^A_{k_j}$ is the mean photon number of the j^{th} symbol $\hat{\rho}^{A^n_j}_k$ of

[15] From the definition of Holevo information, we have

$$\begin{aligned} \chi(p_{W_B}(m), \hat{\rho}^{B^n}_{m,k}) &\equiv S(\sum_m p_{W_B}(m)\hat{\rho}^{B^n}_{m,k}) - \sum_m p_{W_B}(m) S(\hat{\rho}^{B^n}_{m,k}) \\ &= S(\hat{\rho}^{B^n}_k) - \sum_m p_{W_B}(m) S(\hat{\rho}^{B^n}_{m,k}) \\ &\geq 0. \end{aligned}$$

the n-symbol codeword $\hat{\rho}_k^{A^n}$, for $j \in \{1, \ldots, n\}$. Similarly, let $\bar{N}_k^B = \sum_{j=1}^n \frac{1}{n} \bar{N}_{k_j}^B$, where $\bar{N}_{k_j}^B$ is the mean photon number of the j^{th} symbol $\hat{\rho}_k^{B_j^n}$ of the n-symbol codeword $\hat{\rho}_k^{B^n}$, for $j \in \{1, \ldots, n\}$. The overall mean photon numbers per channel use for Alice and Bob are thus given by an average over the codebook W_C, i.e., $\bar{N} = 2^{-nR_C} \sum_{k=1}^{2^{nR_C}} \bar{N}_k^A$, and $\bar{N}^B = 2^{-nR_C} \sum_{k=1}^{2^{nR_C}} \bar{N}_k^B$. From the input-output relation of the channel, the following must hold:

$$\bar{N}_{k_j}^B = \eta \bar{N}_{k_j}^A + (1-\eta)N, \quad \forall k, j \tag{3.77}$$

$$\bar{N}_k^B = \eta \bar{N}_k^A + (1-\eta)N, \quad \forall k, \quad \text{and} \tag{3.78}$$

$$\bar{N}^B = \eta \bar{N} + (1-\eta)N. \tag{3.79}$$

Using Eq. (3.76), the fact that the maximum von Neumann entropy of a single-mode bosonic state with mean photon number $\bar{N}$ is given by $g(\bar{N})$, and the concavity of $g(x)$, we have

$$ng((1-\eta)N) \leq S\left(\hat{\rho}_k^{B^n}\right) \leq \sum_{j=1}^n g\left(\bar{N}_{k_j}^B\right) \leq ng(\bar{N}_k^B) = ng\left(\eta \bar{N}_k^A + (1-\eta)N\right). \tag{3.80}$$

Therefore given the monotonicity of the $g(x)$-function, $\exists \beta_k \in [0,1]$, $\forall k \in W_C$, such that

$$S\left(\hat{\rho}_k^{B^n}\right) = ng\left(\eta \beta_k \bar{N}_k^A + (1-\eta)N\right). \tag{3.81}$$

The average number of photons per use at the transmitter (Alice) averaged over the entire codebook (W_B, W_C), is $\bar{N}$. Thus, the mean photon-number of the n-use average codeword for Bob, $\hat{\rho}^{B^n} \equiv \sum_k p_{W_C}(k) \hat{\rho}_k^{B^n}$, is $\eta \bar{N} + (1-\eta)N$. Hence,

$$ng((1-\eta)N) \leq \sum_k p_{W_C}(k) S\left(\hat{\rho}_k^{B^n}\right) \leq S(\hat{\rho}^{B^n}) \leq ng\left(\eta \bar{N} + (1-\eta)N\right), \tag{3.82}$$

where the first inequality assumes strong conjecture 1 and the second inequality follows from the concavity of von Neumann entropy. The monotonicity of $g(x)$ then

implies that there is a $\beta \in [0,1]$, such that

$$\sum_k p_{W_C}(k) S\left(\hat{\rho}_k^{B^n}\right) = ng(\eta\beta\bar{N} + (1-\eta)N). \tag{3.83}$$

We thus have,

$$\begin{aligned}
& \sum_k p_{W_C}(k)\chi(p_{W_B}(m), \hat{\rho}_{m,k}^{B^n}) \\
= & \sum_k p_{W_C}(k) S\left(\sum_m p_{W_B}(m)\hat{\rho}_{m,k}^{B^n}\right) - \sum_k \sum_m p_{W_C}(k) p_{W_B}(m) S(\hat{\rho}_{m,k}^{B^n}) \quad (3.84) \\
= & \sum_k p_{W_C}(k) S\left(\hat{\rho}_k^{B^n}\right) - \sum_k \sum_m p_{W_C}(k) p_{W_B}(m) S(\hat{\rho}_{m,k}^{B^n}) \quad (3.85) \\
\leq & ng(\eta\beta\bar{N} + (1-\eta)N) - ng((1-\eta)N). \quad (3.86)
\end{aligned}$$

where the last inequality follows from Eqs. (3.83) and (3.74). This completes the first part of the converse proof, i.e., inequality (3.72).

Because of the degraded nature of the channel, Charlie's state can be obtained as the output of a beam splitter of transmissivity $\eta' = (1-\eta)/\eta$, whose input states are Bob's state and a thermal state of mean photon number N (See Fig. 3-11). It follows, from assuming the truth of strong conjecture 3 (see chapter 4), that

$$\begin{aligned}
S\left(\hat{\rho}_k^{C^n}\right) &\geq ng\left(\eta'(\eta\beta_k\bar{N}_k^A + (1-\eta)N) + (1-\eta')N\right) \quad (3.87) \\
&= ng((1-\eta)\beta_k\bar{N}_k^A + \eta N). \quad (3.88)
\end{aligned}$$

Equations (3.81), (3.83), and the uniform distribution $p_{W_C}(k) = 1/2^{nR_C}$ imply that

$$\sum_k \frac{1}{2^{nR_C}} g\left(\eta\beta_k\bar{N}_k^A + (1-\eta)N\right) = g\left(\eta\beta\bar{N} + (1-\eta)N\right). \tag{3.89}$$

Using (3.89), the concavity of $g(x)$-function, and $\eta > 1/2$, we have shown (proof in Appendix C) that

$$\sum_k \frac{1}{2^{nR_C}} g\left((1-\eta)\beta_k\bar{N}_k^A + \eta N\right) \geq g\left((1-\eta)\beta\bar{N} + \eta N\right). \tag{3.90}$$

From Eq. (3.90), and (3.88) summed over k, we then obtain

$$\sum_k p_{W_C}(k) S\left(\hat{\rho}_k^{C^n}\right) \geq ng((1-\eta)\beta\bar{N} + \eta N). \tag{3.91}$$

Finally, we bound Charlie's Holevo information using the standard maximum entropy bound with a mean photon number constraint and Eq. (3.91), which yields:

$$\begin{aligned} \chi(p_{W_C}(k), \hat{\rho}_k^{C^n}) &= S\left(\sum_k p_{W_C}(k)\hat{\rho}_k^{C^n}\right) - \sum_k p_{W_C}(k) S\left(\hat{\rho}_k^{C^n}\right) \\ &\leq ng((1-\eta)\bar{N} + \eta N) - ng((1-\eta)\beta\bar{N} + \eta N), \end{aligned} \tag{3.92}$$

completing the proof of the second piece of the converse, i.e., that of inequality (3.73). The capacity region is additive, because the achievability part of the proof above shows that a product distribution over single-use coherent-state alphabet achieves the rate region.

3.4.6 Noiseless bosonic broadcast channel with M receivers

Let us now consider a bosonic broadcast channel in which the transmitter Alice (A) sends independent messages to M receivers, $Y_0, \ldots, Y_{M-1}$. Let us label Alice's modal annihilation operator as $\hat{a}$, and the annihilation operators for the receivers Y_l as $\hat{y}_l$, $\forall l \in \{0, \ldots, M-1\}$. In order to characterize the bosonic broadcast channel as a quantum-mechanically correct representation of the evolution of a closed system, we must incorporate $M-1$ environment inputs $\{E_1, \ldots, E_{M-1}\}$ along with the transmitter A, such that the M output annihilation operators are related to the M input annihilation operators through a unitary matrix, i.e.,

$$\begin{pmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \vdots \\ \hat{y}_{M-1} \end{pmatrix} = U \begin{pmatrix} \hat{a} \\ \hat{e}_1 \\ \vdots \\ \hat{e}_{M-1} \end{pmatrix}, \tag{3.93}$$

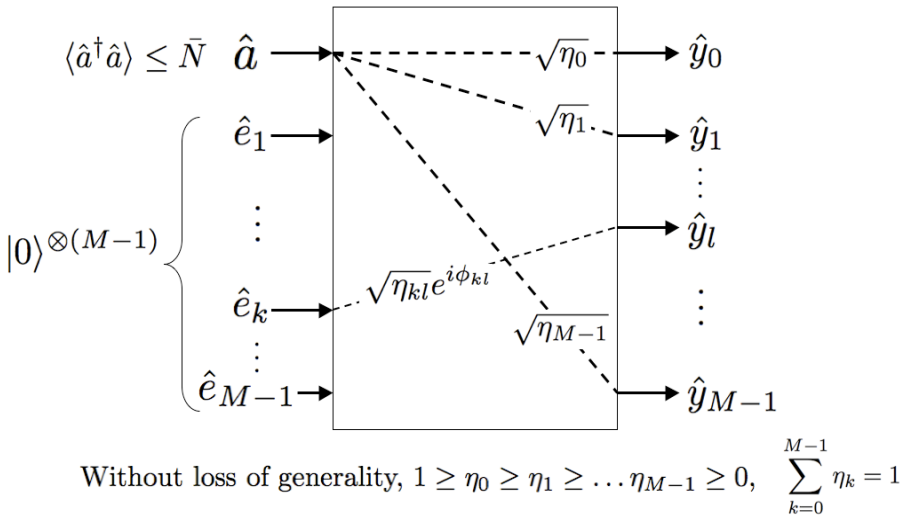


Figure 3-12: An M-receiver noiseless bosonic broadcast channel. Transmitter Alice (A) sends independent messages to M receivers, $Y_0, \ldots, Y_{M-1}$. We have labeled Alice's modal annihilation operator as $\hat{a}$, and those of the receivers Y_l as $\hat{y}_l$, $\forall l \in \{0, \ldots, M-1\}$. In order to characterize the bosonic broadcast channel as a quantum-mechanically correct representation of the evolution of a closed system, we must incorporate $M-1$ environment inputs $\{E_1, \ldots, E_{M-1}\}$ along with the transmitter A (whose modal annihilation operators have been labeled as $\{\hat{e}_1, \ldots, \hat{e}_{M-1}\}$), such that the M output annihilation operators are related to the M input annihilation operators through a unitary matrix, as given in Eq. (3.93). For the noiseless bosonic broadcast channel, all the $M-1$ environment modes $\hat{e}_k$ are in their vacuum states. The transmitter is constrained to at most $\bar{N}$ photons on an average per channel use, for encoding the data. The fractional power coupling from the transmitter to the receiver Y_k is taken to be η_k. We have labeled the receivers in such a way, that $1 \geq \eta_0 \geq \eta_1 \geq \ldots \geq \eta_{M-1} \geq 0$. This ordering of the transmissivities renders this channel a degraded quantum broadcast channel $A \rightarrow Y_0 \rightarrow \ldots \rightarrow Y_{M-1}$ (See Fig. 3-13). The fractional power coupling from E_k to Y_l has been taken to be η_{kl}. For $M = 2$, the above channel model reduces to the familiar two-receiver beam splitter channel model as given in Fig. 3-7.

where $\{\hat{e}_1, \ldots, \hat{e}_{M-1}\}$ are the modal annihilation operators of the $M-1$ environment modes (see Fig. 3-12). The unitary matrix describing the channel can be expressed in the most general form as:

$$U = \begin{pmatrix} \sqrt{\eta_0} & \sqrt{\eta_{10}}e^{i\phi_{10}} & \cdots & \sqrt{\eta_{M-1,0}}e^{i\phi_{M-1,0}} \\ \sqrt{\eta_1} & \sqrt{\eta_{11}}e^{i\phi_{11}} & \cdots & \sqrt{\eta_{M-1,1}}e^{i\phi_{M-1,1}} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\eta_{M-1}} & \sqrt{\eta_{1,M-1}}e^{i\phi_{1,M-1}} & \cdots & \sqrt{\eta_{M-1,M-1}}e^{i\phi_{M-1,M-1}} \end{pmatrix}, \tag{3.94}$$

where $\{\eta_0, \ldots, \eta_{M-1}\}$ are the transmissivities (fractional power couplings) from the transmitter A to the $M-1$ receivers $Y_0, \ldots, Y_{M-1}$. Without loss of generality, we have numbered the receivers, so that the transmissivities are in decreasing order, i.e.,

$$1 \geq \eta_0 \geq \eta_1 \geq \ldots \geq \eta_{M-1} \geq 0. \tag{3.95}$$

The power coupling from the environment mode $\hat{e}_k$ to the output mode $\hat{y}_l$ is η_{kl}. Without loss of generality, the phases for the entries of the first column of U have been taken to be 0, as an overall phase is inconsequential in each of the $M-1$ input-output relations,

$$\hat{y}_k = \sqrt{\eta_k}\hat{a} + \sum_{l=1}^{M-1} \sqrt{\eta_{lk}}e^{i\phi_{lk}}\hat{e}_l. \tag{3.96}$$

The fractional power-couplings must satisfy the following normalization constraints,

$$\sum_{k=0}^{M-1} \eta_k = 1, \tag{3.97}$$

$$\sum_{k=0}^{M-1} \eta_{lk} = 1, \quad \forall l \in \{1, \ldots, M-1\}, \tag{3.98}$$

$$\eta_k + \sum_{l=1}^{M-1} \eta_{lk} = 1, \quad \forall k \in \{0, \ldots, M-1\}. \tag{3.99}$$

Theorem 3.5 — For the noiseless bosonic broadcast channel, i.e., when the environment modes $\{\hat{e}_k : 1 \leq k \leq M-1\}$ are in a product of $M-1$ vacuum states, $|0\rangle^{\otimes(M-1)}$,

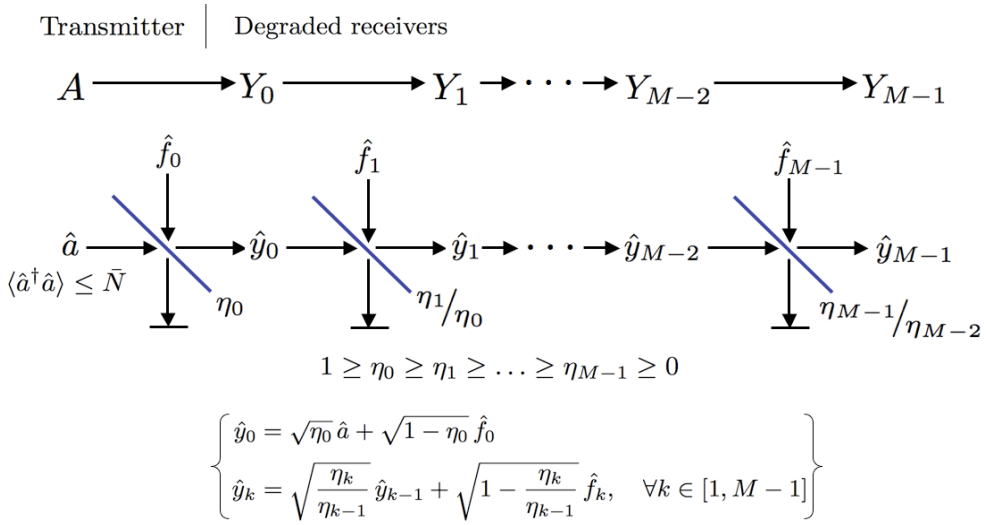


Figure 3-13: An equivalent stochastically degraded model for the M-receiver noiseless bosonic broadcast channel depicted in Fig. 3-12. If the receivers are ordered in a way such that the fractional power couplings η_k from the transmitter to the receiver Y_k are in decreasing order, the quantum states at each receiver Y_k, for $k \in \{1, \ldots, M-1\}$, can be obtained from the state received at receiver Y_{k-1} by mixing it with a vacuum state, through a beam splitter of transmissivity η_k/η_{k-1}. This equivalent representation of the M-receiver bosonic broadcast channel confirms that the bosonic broadcast channel is indeed a degraded broadcast channel, whose capacity region is given by the infinite-dimensional (continuous-variable) extension of Yard et. al.'s theorem in Eqs. (3.38).

and with an input mean photon-number constraint $\langle \hat{a}^\dagger \hat{a} \rangle \leq \bar{N}$, the ultimate capacity region[16] is given by

$$R_k \leq g(\eta_k \beta_{k+1} \bar{N}) - g(\eta_k \beta_k \bar{N}), \quad k \in \{0, \ldots, M-1\}, \tag{3.100}$$

where,

$$0 = \beta_0 < \beta_1 < \ldots < \beta_{M-1} < \beta_M = 1. \tag{3.101}$$

Proof [Achievability] — Using the infinite-dimensional (continuous-variable) extension of Eqs. (3.38), the $n = 1$ rate-region for the bosonic broadcast channel using

[16]Note the similarity with the capacity region for the classical Gaussian broadcast channel, as given in Eq. (3.22), with $N = 0$. Also note that Eq. (3.100) reduces to the two-user noiseless bosonic broadcast capacity region, as given in Eqs. (3.41) and (3.42), with the substitutions $\eta_0 = \eta$, and $\eta_1 = 1 - \eta$.

coherent-state encoding is given by[17] (see Fig. 3-13 and Fig. 3-14 for notation):

$$
\begin{aligned}
R_0 &\leq \int p_{T_1}(\tau_1) S\left(\int p_{A|T_1}(\alpha|\tau_1)|\sqrt{\eta}_0\alpha\rangle\langle\sqrt{\eta}_0\alpha|\mathrm{d}^2\alpha\right)\mathrm{d}^2\tau_1 \\
R_k &\leq \int p_{T_{k+1}}(\tau_{k+1})\chi\left(p_{T_k|T_{k+1}}(\tau_k|\tau_{k+1}),\hat{\rho}_{\tau_k}^{Y_k}\right)\mathrm{d}^2\tau_{k+1} \\
&= \int p_{T_{k+1}}(\tau_{k+1})\left(S\left(\int p_{T_k|T_{k+1}}(\tau_k|\tau_{k+1})\hat{\rho}_{\tau_k}^{Y_k}\mathrm{d}^2\tau_k\right)\right. \\
&\quad \left. - \int p_{T_k|T_{k+1}}(\tau_k|\tau_{k+1})S\left(\hat{\rho}_{\tau_k}^{Y_k}\right)\mathrm{d}^2\tau_k\right)\mathrm{d}^2\tau_{k+1}, \quad \text{for } k\in\{1,\ldots,M-2\}, \\
R_{M-1} &\leq \chi\left(p_{T_{M-1}}(\tau_{M-1}),\hat{\rho}_{\tau_{M-1}}^{Y_{M-1}}\right) \\
&= S\left(\int p_{T_{M-1}}(\tau_{M-1}),\hat{\rho}_{\tau_{M-1}}^{Y_{M-1}}\right) \\
&\quad -\int p_{T_{M-1}}(\tau_{M-1})S\left(\hat{\rho}_{\tau_{M-1}}^{Y_{M-1}}\right)\mathrm{d}^2\tau_{M-1}
\end{aligned}
\tag{3.103}
$$

where we need to maximize the above rate region $\{R_0,\ldots,R_{M-1}\}$ over all joint distributions $p_{T_{M-1}}(\tau_{M-1})p_{T_{M-2|M-1}}(\tau_{M-2}|\tau_{M-1})\ldots p_{T_1|T_2}(\tau_1|\tau_2)p_{A|T_1}(\alpha|\tau_1)$ subject to $\langle|\alpha|^2\rangle\leq\bar{N}$. Note that A, and the auxiliary random variables $T_1,\ldots,T_{M-1}$ are complex-valued, and the second term in the R_0 bound (see (3.38)) vanishes, because the von Neumann entropy of a pure state is zero.

Let us associate with each random variable T_k, a quantum system, i.e. a coherent-state alphabet $\{|\tau_k\rangle\}$ and a modal annihilation operator $\hat{t}_k$, $\forall k\in\{1,\ldots,M-1\}$. In

[17]Here, we use a continuous-variable version of the notation we used in Eqs. (3.38). When the cardinalities $|A|$ and $|T_k|$, $1\leq k\leq M-1$ are finite, and we are using coherent states, we end up with a finite number of possible transmitted states, which leads to a finite number of possible states received by Bob and Charlie. To be more explicit, let us limit the auxiliary-input alphabets (T_k, $1\leq k\leq M-1$) – and hence the input (A) and the output alphabets (Y_k, $0\leq k\leq M-1$) – to coherent states in the finite-dimensional subspace spanned by the Fock states $\{|0\rangle,|1\rangle,\ldots,|K\rangle\}$, where $K\gg\bar{N}$. Applying the extension of Yard et al.'s theorem to M receivers (3.38), the Hilbert space spanned by these states then gives us a broadcast channel capacity region that must be strictly an inner bound of the rate region given by Eqs. (3.103). In the limit that we choose K sufficiently large, clearly the rate-region expressions given by Eqs. (3.38) can be brought to as close as we wish, to those given by Eqs. (3.103). The summations in Eqs. (3.38) get replaced by integrals. The collective message index j is now replaced by the complex number α, the indices i_k are replaced by τ_k, and the density matrices of the conditional received states are given by: ,

$$
\hat{\rho}_{\tau_k}^{Y_k}=\int\ldots\int p_{A|T_1}(\alpha|\tau_1)p_{T_1|T_2}(\tau_1|\tau_2)\ldots p_{T_{k-1}|T_k}(\tau_{k-1}|\tau_k)\hat{\rho}_{\alpha}^{Y_k}\mathrm{d}^2\tau_{k-1}\ldots\mathrm{d}^2\tau_1\mathrm{d}^2\alpha, \tag{3.102}
$$

where, $\hat{\rho}_{\alpha}^{Y_k}=|\sqrt{\eta_k}\alpha\rangle\langle\sqrt{\eta_k}\alpha|$ is the state received by the receiver Y_k, when the transmitter sends a coherent state $\hat{\rho}_{\alpha}^{A}=|\alpha\rangle\langle\alpha|$.

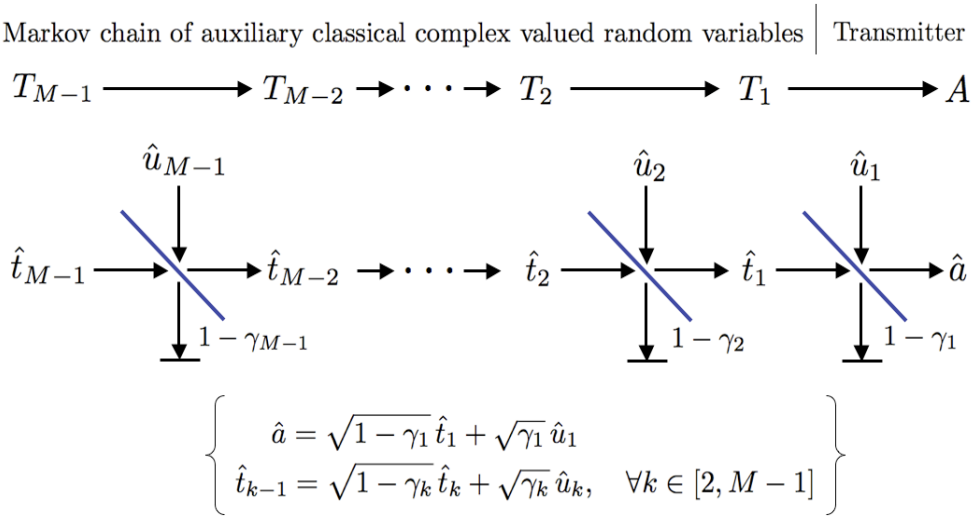


Figure 3-14: In order to evaluate the capacity region of the M-receiver noiseless bosonic degraded broadcast channel depicted in Fig. 3-13 using a coherent-state input alphabet $\{|\alpha\rangle\}$, $\alpha \in \mathbb{C}$ and $\langle \hat{a}^\dagger \hat{a} \rangle = \langle |\alpha|^2 \rangle \leq \bar{N}$, we choose the $M-1$ auxiliary classical Markov random variables (in Eqs. (3.35)) as complex-valued random variables T_k, $k \in \{1, \ldots, M-1\}$, taking values $\tau_k \in \mathbb{C}$. In order to visualize the postulated optimal Gaussian distributions for the random variables T_k, let us associate with T_k, a quantum system, i.e., a coherent-set alphabet $\{|\tau_k\rangle\}$ and modal annihilation operator $\hat{t}_k$, $\forall k$. In accordance with the Markov property of the random variables T_k, let $\hat{t}_{M-1}$ be in an isotropic zero-mean Gaussian mixture of coherent-states with a variance $\bar{N}$ (see Eq. (3.104)), and for $k \in \{1, \ldots, M-2\}$, let $\hat{t}_k$ be obtained from $\hat{t}_{k+1}$ by mixing it with another mode $\hat{u}_{k+1}$ excited in a zero-mean thermal state with mean photon number $\bar{N}$, through a beam splitter with transmissivity $1-\gamma_{k+1}$, as shown in the figure above, for some $\gamma_{k+1} \in (0,1)$. We complete the Markov chain $T_{M-1} \to \ldots \to T_1 \to A$, by obtaining the transmitter mode $\hat{a}$ by mixing $\hat{t}_1$ with a mode $\hat{u}_1$ excited in a zero-mean thermal state with mean photon number $\bar{N}$, through a beam splitter with transmissivity $1-\gamma_1$, for $\gamma_1 \in (0,1)$. The above setup of the auxiliary modes gives rise to the distributions given in Eqs. (3.104), which we use to evaluate the achievable rate region of the M-receiver bosonic broadcast channel using coherent-state encoding.

accordance with the Markov property of the random variables T_k, let $\hat{t}_{M-1}$ be in an isotropic zero-mean Gaussian mixture of coherent-states with a variance $\bar{N}$ (see Eq. (3.104)), and for $k \in \{1, \ldots, M-2\}$, let $\hat{t}_k$ be obtained from $\hat{t}_{k+1}$ by mixing it with another mode $\hat{u}_{k+1}$ excited in a zero-mean thermal state with mean photon number $\bar{N}$, through a beam splitter with transmissivity $1-\gamma_{k+1}$, as shown in Fig. 3-14, for real numbers $\gamma_{k+1} \in (0,1)$. We complete the Markov chain $T_{M-1} \rightarrow \ldots \rightarrow T_1 \rightarrow A$, by obtaining the transmitter mode $\hat{a}$ by mixing $\hat{t}_1$ with a mode $\hat{u}_1$ in a vacuum state, through a beam splitter with transmissivity $1-\gamma_1$, for $\gamma_1 \in (0,1)$. This setup of the auxiliary modes gives rise to the distributions given below, which we use to evaluate the achievable rate region using coherent-state encoding:

$$\begin{aligned}
p_{A|T_1}(\alpha|\tau_1) &= \frac{1}{\pi\gamma_1\bar{N}} \exp\left(-\frac{|\sqrt{1-\gamma_1}\tau_1 - \alpha|^2}{\gamma_1\bar{N}}\right) \\
p_{T_k|T_{k+1}}(\tau_k|\tau_{k+1}) &= \frac{1}{\pi\gamma_{k+1}\bar{N}} \exp\left(-\frac{|\sqrt{1-\gamma_{k+1}}\tau_{k+1} - \tau_k|^2}{\gamma_{k+1}\bar{N}}\right), \quad \text{for } k \in \{1, \ldots, M-2\}, \\
p_{T_{M-1}}(\tau_{M-1}) &= \frac{1}{\bar{N}} \exp\left(-\frac{|\tau_{M-1}|^2}{\bar{N}}\right).
\end{aligned} \tag{3.104}$$

Substituting Eqs. (3.104) into Eqs. (3.103), we get

$$\begin{aligned}
R_0 &\leq g(\eta_0\beta_1\bar{N}), \\
R_k &\leq g(\eta_k\beta_{k+1}\bar{N}) - g(\eta_k\beta_k\bar{N}), \quad \text{for } k \in \{1, \ldots, M-2\}, \\
R_{M-1} &\leq g(\eta_{M-1}\bar{N}) - g(\eta_{M-1}\beta_{M-1}\bar{N}),
\end{aligned} \tag{3.105}$$

where we define

$$\beta_k \triangleq 1 - \prod_{i=1}^{k}(1-\gamma_i), \quad \text{for } k \in \{1, \ldots, M-1\}. \tag{3.106}$$

By further defining $\beta_0 \triangleq 0$, and $\beta_M \triangleq 1$, we have by construction, $0 = \beta_0 < \beta_1 < \ldots < \beta_{M-1} < \beta_M = 1$. With these definitions, Eqs. (3.105) reduce to the rate-region expression given in Eq. (3.100). Hence the postulated rate region is achievable using

single-use coherent state encoding.

Proof [Converse] — Our goal in proving the converse is to show that any achievable rate M-tuple $(R_0, \ldots, R_{M-1})$ must be inside the ultimate rate-region proposed by Eqs. (3.105). Let us assume that $(R_0, \ldots, R_{M-1})$ is achievable. Using the notation in Eq. (3.33), let $\{x^n(m_0, \ldots, m_{M-1})\}$, and POVMs $\left\{\Lambda^0_{m_0 \ldots m_{M-1}}\right\}$, $\left\{\Lambda^1_{m_1 \ldots m_{M-1}}\right\}$, …, $\left\{\Lambda^{M-1}_{m_{M-1}}\right\}$ comprise a $(2^{nR_0}, \ldots, 2^{nR_{M-1}}, n, \epsilon)$ code in the achieving sequence. Let us suppose that the receivers $Y_0, \ldots, Y_{M-1}$ store their respective decoded messages in registers $\hat{W}_0, \ldots, \hat{W}_{M-1}$. By assuming a good source encoder prior to the broadcast channel-encoder, it is fair to assume a uniform distribution over the messages, i.e.,

$$\begin{aligned} p_{\mathbf{W}_0^{M-1}}(\mathbf{m}_0^{M-1}) &= \prod_{k=0}^{M-1} p_{W_k}(m_k) \\ &= \prod_{k=0}^{M-1} \frac{1}{2^{nR_k}} \\ &= \frac{1}{2^{n\sum_{k=0}^{M-1} R_k}}. \end{aligned} \tag{3.107}$$

Lemma 3.6 — For every $k \in \{1, \ldots, M-1\}$, $\exists \beta_k \in [0,1]$, s.t.[18]

$$\sum_{\mathbf{m}_k^{M-1}} p_{\mathbf{W}_k^{M-1}}(\mathbf{m}_k^{M-1}) S\left(\hat{\rho}^{Y_{k-1}^n}_{\mathbf{m}_k^{M-1}}\right) = ng\left(\eta_{k-1}\beta_k \bar{N}\right). \tag{3.111}$$

Proof — We have

$$0 \leq \sum_{\mathbf{m}_k^{M-1}} p_{\mathbf{W}_k^{M-1}}(\mathbf{m}_k^{M-1}) S\left(\hat{\rho}^{Y_{k-1}^n}_{\mathbf{m}_k^{M-1}}\right) \leq S\left(\hat{\rho}^{Y_{k-1}^n}\right) \leq ng(\eta_{k-1}\bar{N}), \tag{3.112}$$

where the first inequality follows from the non-negativity of von-Neumann entropy. The second inequality follows from concavity of von-Neumann entropy or equivalently from the non-negativity of Holevo information (see footnote 15), because

$$\hat{\rho}^{Y_{k-1}^n} = \sum_{\mathbf{m}_k^{M-1}} p_{\mathbf{W}_k^{M-1}}(\mathbf{m}_k^{M-1}) \hat{\rho}^{Y_{k-1}^n}_{\mathbf{m}_k^{M-1}}.$$

The third inequality above is due to the fact that the maximum entropy of a n-mode state with a mean photon number $\bar{n}$ per mode, is given by $ng(\bar{n})$. From the monotonicity of the function $g(\cdot)$, there must therefore exist a real number $\beta_k \in [0,1]$,

[18] We defined earlier in this chapter $\{m_0, \ldots, m_{M-1}\} \triangleq \mathbf{m}_0^{M-1}$ to be a collective index for the M messages that Alice encodes into her n-use transmitted codeword state $\rho^{A^n}_{\mathbf{m}_0^{M-1}}$, and $\rho^{Y_k^n}_{\mathbf{m}_0^{M-1}}$ was defined to be the state received by Y_k over n successive channel uses. We also used the compact notation $\mathbf{W}_k^{M-1}$ for the vectors of random variables $(W_k, \ldots, W_{M-1})$. Y_k^n represents the n-use quantum system of the k^{th} receiver. By averaging a conditional received state that is indexed by a set of messages $\mathbf{m}_k^{M-1}$, over the probability mass function of a subset of the message-sets $\mathbf{W}_k^{M-1}$, we get a new conditional received state now indexed only by the remaining (smaller set of) messages. The received state that has been averaged over all messages is not indexed by any message. Also, by taking the trace of a joint conditional received state over a set of receiver Hilbert spaces, we obtain the conditional received state for the remaining (smaller set of) receivers. Thus, the following (and other similar) identities hold:

$$\hat{\rho}^{Y_k^n}_{m_k} = \sum_{\mathbf{m}_{k+1}^{M-1}} p_{\mathbf{W}_{k+1}^{M-1}}(\mathbf{m}_{k+1}^{M-1}) \hat{\rho}^{Y_k^n}_{\mathbf{m}_k^{M-1}} \tag{3.108}$$

$$\hat{\rho}^{Y_{M-1}^n} = \sum_{m_{M-1}} p_{W_{M-1}}(m_{M-1}) \hat{\rho}^{Y_{M-1}^n}_{m_{M-1}} \tag{3.109}$$

$$\hat{\rho}^{Y_k^n}_{\mathbf{m}_k^{M-1}} = \text{Tr}_{Y_{k+1}^n, \ldots, Y_{M-1}^n}\left(\hat{\rho}^{Y_k^n \ldots Y_{M-1}^n}_{\mathbf{m}_k^{M-1}}\right) \tag{3.110}$$

such that

$$\sum_{\mathbf{m}_k^{M-1}} p_{\mathbf{W}_k^{M-1}}(\mathbf{m}_k^{M-1}) S\left(\hat{\rho}^{Y_{k-1}^n}_{\mathbf{m}_k^{M-1}}\right) = ng\left(\eta_{k-1}\beta_k \bar{N}\right), \tag{3.113}$$

which completes the proof of Lemma 3.6.

Now, as $(R_0, \ldots, R_{M-1})$ is an achievable rate M-tuple, there exist $\epsilon_{k,n} \to 0$ as $n \to \infty$, for $k \in \{0, \ldots, M-1\}$, such that,

$$\begin{aligned}
0 \leq nR_k &= H(W_k) \\
&\leq I(W_k; \hat{W}_k) + n\epsilon_{k,n} && (3.114)\\
&\leq \chi\left(p_{W_k}(m_k), \hat{\rho}^{Y_k^n}_{m_k}\right) + n\epsilon_{k,n} && (3.115)\\
&\leq \sum_{\mathbf{m}_{k+1}^{M-1}} p_{\mathbf{W}_{k+1}^{M-1}}(\mathbf{m}_{k+1}^{M-1}) \chi\left(p_{W_k}(m_k), \hat{\rho}^{Y_k^n}_{\mathbf{m}_k^{M-1}}\right) + n\epsilon_{k,n}, && (3.116)
\end{aligned}$$

where $I(W_k; \hat{W}_k) = H(W_k) - H(W_k|\hat{W}_k)$ is the Shannon mutual information. Inequality (3.114) follows from Fano's inequality, (3.115) follows from the Holevo's bound [27, 28, 29], and (3.116) follows from the concavity of Holevo information, as $\hat{\rho}^{Y_k^n}_{m_k} = \sum_{\mathbf{m}_{k+1}^{M-1}} p_{\mathbf{W}_{k+1}^{M-1}}(\mathbf{m}_{k+1}^{M-1}) \hat{\rho}^{Y_k^n}_{\mathbf{m}_k^{M-1}}$. Specializing inequality (3.116) to $k = 0$ we obtain,

$$\begin{aligned}
nR_0 &\leq \sum_{\mathbf{m}_1^{M-1}} p_{\mathbf{W}_1^{M-1}}(\mathbf{m}_1^{M-1}) \chi\left(p_{W_0}(m_0), \hat{\rho}^{Y_0^n}_{\mathbf{m}_0^{M-1}}\right) + n\epsilon_{0,n} && (3.117)\\
&\leq \sum_{\mathbf{m}_1^{M-1}} p_{\mathbf{W}_1^{M-1}}(\mathbf{m}_1^{M-1}) S\left(\sum_{m_0} p_{W_0}(m_0) \hat{\rho}^{Y_0^n}_{\mathbf{m}_0^{M-1}}\right) + n\epsilon_{0,n} && (3.118)\\
&= \sum_{\mathbf{m}_1^{M-1}} p_{\mathbf{W}_1^{M-1}}(\mathbf{m}_1^{M-1}) S\left(\hat{\rho}^{Y_0^n}_{\mathbf{m}_1^{M-1}}\right) + n\epsilon_{0,n} && (3.119)\\
&= ng(\eta_0 \beta_1 \bar{N}) + n\epsilon_{0,n}, && (3.120)
\end{aligned}$$

where inequality (3.118) follows from dropping out the second term of Holevo information in (3.117). Inequality (3.120) follows from *Lemma* 3.2, for $k = 1$. For $k \in \{1, \ldots, M-2\}$, continuing from (3.116) we have,

$$
\begin{aligned}
nR_k &\leq \sum_{\mathbf{m}_{k+1}^{M-1}} p_{\mathbf{W}_{k+1}^{M-1}}(\mathbf{m}_{k+1}^{M-1}) \left[S\left(\sum_{m_k} p_{W_k}(m_k) \hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_k^n} \right) - \sum_{m_k} p_{W_k}(m_k) S\left(\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_k^n} \right) \right] + n\epsilon_{k,n} \\
&= \sum_{\mathbf{m}_{k+1}^{M-1}} p_{\mathbf{W}_{k+1}^{M-1}}(\mathbf{m}_{k+1}^{M-1}) S\left(\hat{\rho}_{\mathbf{m}_{k+1}^{M-1}}^{Y_k^n} \right) - \sum_{\mathbf{m}_k^{M-1}} p_{\mathbf{W}_k^{M-1}}(\mathbf{m}_k^{M-1}) S\left(\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_k^n} \right) + n\epsilon_{k,n} \quad (3.121) \\
&= ng\left(\eta_k \beta_{k+1} \bar{N} \right) - \sum_{\mathbf{m}_k^{M-1}} p_{\mathbf{W}_k^{M-1}}(\mathbf{m}_k^{M-1}) S\left(\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_k^n} \right) + n\epsilon_{k,n}, \quad (3.122)
\end{aligned}
$$

where (3.121) and (3.122) follow from the definition of Holevo information and *Lemma* 3.2 respectively. Next, we shall bound the second term in (3.122). Let us define $\bar{N}^A_{\mathbf{m}_k^{M-1},j}$ to be the mean photon number of the j^{th} symbol $\hat{\rho}_{\mathbf{m}_k^{M-1}}^{A_j^n}$ of the n-symbol codeword $\hat{\rho}_{\mathbf{m}_k^{M-1}}^{A^n}$, whose mean photon number is given by $\bar{N}^A_{\mathbf{m}_k^{M-1}} = \frac{1}{n}\sum_{j=1}^n \bar{N}^A_{\mathbf{m}_k^{M-1},j}$. Hence, $\eta_{k-1}\bar{N}^A_{\mathbf{m}_k^{M-1},j}$ is the mean photon number of the j^{th} symbol $\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_{k-1,j}^n}$ of the n-symbol codeword $\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_{k-1}^n}$, whose mean photon number is given by $\eta_{k-1}\bar{N}^A_{\mathbf{m}_k^{M-1}}$. The overall mean photon number of the transmitter codeword per channel use $\bar{N}$, is thus given by averaging $\bar{N}^A_{\mathbf{m}_k^{M-1}}$ over the codebooks $\mathbf{W}_k^{M-1}$, i.e.,

$$
\bar{N} = 2^{-n\sum_{j=k}^{M-1} R_j} \sum_{\mathbf{m}_k^{M-1}} \bar{N}^A_{\mathbf{m}_k^{M-1}}.
$$

From the non-negativity of von-Neumann entropy, the fact that the maximum von Neumann entropy of a single-mode bosonic state with mean photon number $\bar{N}$ is given by $g(\bar{N})$, and the concavity of $g(x)$, we have the following inequalities:

$$
0 \leq S\left(\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_{k-1}^n} \right) \leq \sum_{j=1}^n g\left(\eta_{k-1}\bar{N}^A_{\mathbf{m}_k^{M-1},j} \right) \leq ng\left(\eta_{k-1}\bar{N}^A_{\mathbf{m}_k^{M-1}} \right). \quad (3.123)
$$

Therefore, there must exist real numbers $\beta_{\mathbf{m}_k^{M-1}} \in [0,1]$, $\forall \mathbf{m}_k^{M-1} \in \mathbf{W}_k^{M-1}$, such that

$$
S\left(\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_{k-1}^n} \right) = ng\left(\eta_{k-1}\beta_{\mathbf{m}_k^{M-1}}\bar{N}^A_{\mathbf{m}_k^{M-1}} \right). \quad (3.124)
$$

Because of the degraded nature of the channel, $\hat{y}_k = \sqrt{\eta_k/\eta_{k-1}}\hat{y}_{k-1} + \sqrt{1-(\eta_k/\eta_{k-1})}\hat{f}_k$,

with $\hat{f}_k$ in a vacuum state (see Fig. 3-12). Hence, using Eq. (3.124) and strong conjecture 2 (see chapter 4), we have

$$S\left(\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_k^n}\right) \geq ng\left(\eta_k \beta_{\mathbf{m}_k^{M-1}} \bar{N}_{\mathbf{m}_k^{M-1}}^{A}\right). \tag{3.125}$$

Taking an average of both sides of Eq. (3.124) over the codebooks $\mathbf{W}_k^{M-1}$, and using *Lemma* 3.2, we have

$$\begin{aligned} \sum_{\mathbf{m}_k^{M-1}} p_{\mathbf{W}_k^{M-1}}(\mathbf{m}_k^{M-1}) S\left(\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_{k-1}^n}\right) &= \frac{n}{2^{n\sum_{j=k}^{M-1} R_j}} \sum_{\mathbf{m}_k^{M-1}} g\left(\eta_{k-1} \beta_{\mathbf{m}_k^{M-1}} \bar{N}_{\mathbf{m}_k^{M-1}}^{A}\right) \\ &= ng\left(\eta_{k-1}\beta_k \bar{N}\right). \end{aligned} \tag{3.126}$$

Equation (3.126) and a theorem on a property of the $g(\cdot)$ function (see Appendix C), then gives us

$$\frac{n}{2^{n\sum_{j=k}^{M-1} R_j}} \sum_{\mathbf{m}_k^{M-1}} g\left(\eta_k \beta_{\mathbf{m}_k^{M-1}} \bar{N}_{\mathbf{m}_k^{M-1}}^{A}\right) \geq ng\left(\eta_k \beta_k \bar{N}\right). \tag{3.127}$$

Taking an average of both sides of Eq. (3.125) over the codebooks $\mathbf{W}_k^{M-1}$, and using Eq. (3.127), we get

$$\begin{aligned} \sum_{\mathbf{m}_k^{M-1}} p_{\mathbf{W}_k^{M-1}}(\mathbf{m}_k^{M-1}) S\left(\hat{\rho}_{\mathbf{m}_k^{M-1}}^{Y_k^n}\right) &\geq \frac{n}{2^{n\sum_{j=k}^{M-1} R_j}} \sum_{\mathbf{m}_k^{M-1}} g\left(\eta_k \beta_{\mathbf{m}_k^{M-1}} \bar{N}_{\mathbf{m}_k^{M-1}}^{A}\right) \\ &\geq ng\left(\eta_k \beta_k \bar{N}\right). \end{aligned} \tag{3.128}$$

Combining Eqs. (3.122) and (3.128), we finally get the desired bound for R_k, for $k \in \{1, \ldots, M-2\}$, i.e.,

$$nR_k \leq ng\left(\eta_k \beta_{k+1} \bar{N}\right) - ng\left(\eta_k \beta_k \bar{N}\right) + n\epsilon_{k,n}. \tag{3.129}$$

Since $nR_k \geq 0$, the monotonicity of $g(\cdot)$ implies that

$$\beta_{k+1} \geq \beta_k, \quad \forall k \in \{1, \ldots, M-2\}. \tag{3.130}$$

To prove the final piece of the converse proof, i.e., to prove that the proposed rate bound for R_{M-1} holds, we proceed as follows:

$$\begin{aligned}
nR_{M-1} &= H(W_{M-1}) \\
&\leq I(W_{M-1};\hat{W}_{M-1}) + n\epsilon_{M-1,n} && (3.131)\\
&\leq \chi\left(p_{W_{M-1}}(m_{M-1}), \hat{\rho}_{m_{M-1}}^{Y_{M-1}^n}\right) + n\epsilon_{M-1,n} && (3.132)\\
&= S\left(\sum_{m_{M-1}} p_{W_{M-1}}(m_{M-1})\hat{\rho}_{m_{M-1}}^{Y_{M-1}^n}\right) - \sum_{m_{M-1}} p_{W_{M-1}}(m_{M-1}) S\left(\hat{\rho}_{m_{M-1}}^{Y_{M-1}^n}\right) + n\epsilon_{M-1,n} \\
&= S\left(\hat{\rho}^{Y_{M-1}^n}\right) - \sum_{m_{M-1}} p_{W_{M-1}}(m_{M-1}) S\left(\hat{\rho}_{m_{M-1}}^{Y_{M-1}^n}\right) + n\epsilon_{M-1,n} && (3.133)\\
&\leq ng\left(\eta_{M-1}\bar{N}\right) - \sum_{m_{M-1}} p_{W_{M-1}}(m_{M-1}) S\left(\hat{\rho}_{m_{M-1}}^{Y_{M-1}^n}\right) + n\epsilon_{M-1,n} && (3.134)\\
&\leq ng\left(\eta_{M-1}\bar{N}\right) - ng\left(\eta_{M-1}\beta_{M-1}\bar{N}\right) + n\epsilon_{M-1,n}, && (3.135)
\end{aligned}$$

where inequality (3.131) follows from Fano's inequality, (3.132) results from the Holevo bound, (3.134) follows from the fact that the maximum von Neumann entropy of a single-mode bosonic state with mean photon number $\bar{N}$ is given by $g(\bar{N})$. The last inequality (3.135) follows from[19] Eq. (3.128) with $k = M-1$. As $\epsilon_{k,n} \to 0$ as $n \to \infty$, going to the limit of large block length codes, Eqs. (3.120), (3.129), (3.130) and (3.135), along with the definitions $\beta_0 = 0$, and $\beta_M = 1$, we have shown that if $(R_0, \ldots, R_{M-1})$ is an achievable rate M-tuple, then they must satisfy,

$$R_k \leq g(\eta_k\beta_{k+1}\bar{N}) - g(\eta_k\beta_k\bar{N}), \quad k \in \{0, \ldots, M-1\}, \tag{3.136}$$

for real numbers β_k satisfying

$$0 = \beta_0 < \beta_1 < \ldots < \beta_{M-1} < \beta_M = 1, \tag{3.137}$$

which is what we set out to prove.

[19]Note that the same method we used to bound the second term in Eq. (3.122) for $k \in \{1, \ldots, M-2\}$ can also be used for $k = M-1$. All the steps from Eq. (3.122) to Eq. (3.128) follow through exactly in the same way if we substitute $k = M-1$ everywhere.

3.4.7 Thermal-noise bosonic broadcast channel with M receivers

Consider an extension of the noiseless M-receiver bosonic broadcast channel as depicted in Fig. 3-12, in which each environment mode $\hat{e}_k$, for $k \in \{1, \ldots, M-1\}$, is in a zero-mean thermal state with mean photon number N (see Eq. (3.67)). This channel can also be equivalently represented by a degraded model as depicted in Fig. 3-13, in which each of the modes f_k, for $k \in \{1, \ldots, M-1\}$, is now in a zero-mean thermal state with mean photon number N.

Theorem 3.7 — With a mean photon number constraint of $\bar{N}$ photons per channel use at the transmitter, the ultimate capacity region of the thermal-noise bosonic broadcast channel, with uniform noise coupling of N photons on an average in each mode, can be achieved by coherent-state encoding with an isotropic Gaussian prior distribution. Given the truth of strong conjectures 1 and 3, the ultimate capacity region is given by[20]

$$R_k \leq g(\eta_k \beta_{k+1} \bar{N} + (1-\eta_k)N) - g(\eta_k \beta_k \bar{N} + (1-\eta_k)N), \quad k \in \{0, \ldots, M-1\}, \tag{3.138}$$

for real numbers β_k satisfying

$$0 = \beta_0 < \beta_1 < \ldots < \beta_{M-1} < \beta_M = 1. \tag{3.139}$$

Proof — The proof of this theorem follows exactly as in the proof of the ultimate capacity region of the noiseless bosonic broadcast channel with M receivers, using ideas from the capacity-region proof for the thermal-noise bosonic broadcast channel with two receivers. We omit the proof from the thesis due to its notational complexity.

[20] Note that the expression for this capacity region resembles the expression for the capacity region of the M-receiver classical Gaussian broadcast channel, as given in Eq. (3.22). The only difference between these two capacity-region expressions is that the Bergman's $g_C(\cdot)$ function in the classical Gaussian case is replaced by the $g(\cdot)$ function in the quantum bosonic case.

3.4.8 Comparison of bosonic broadcast and multiple-access channel capacity regions

In classical information theory, Vishwanath et. al. [53] established a duality between what is termed the dirty paper achievable region (but recently proved to be the ultimate capacity region [56]) for the classical Multiple-Input-Multiple-Output (MIMO) Gaussian broadcast channel (BC) and the capacity region of the MIMO Gaussian multiple-access channel (MAC), which is easy to compute. Using this duality, the computational complexity required for obtaining the capacity region for the MIMO broadcast channel was greatly reduced. The duality result states that if we were to trace out the capacity regions of the MIMO Gaussian MAC with a certain fixed value of the total received power P and channel-gain values, and for all the various possible power-allocations between the users, the corners of all those capacity regions would trace out the capacity region of the MIMO Gaussian broadcast channel with transmitter power P and the exact same channel-gain values. Unlike this classical result, it turns out that the capacity region of the bosonic broadcast channel using coherent-state inputs is not the exact dual of the envelope of the capacity regions of a multiple-access channel (MAC) using coherent-state inputs. In Figure 3-15, for $\eta = 0.8$, and $\bar{N} = 15$, we show that the capacity region of the bosonic broadcast channel lies below the envelope of the multiple-access capacity regions of the dual MAC. The capacity region of the bosonic MAC using coherent-state inputs was first computed by Yen [11]. So, assuming that the optimum modulation, coding, and receivers are available, on a fixed beam splitter with the same power budget, more collective classical information can be sent when this beam splitter is used as a multiple-access channel, as opposed to when it is used as a broadcast channel. We believe that the duality between the classical MIMO MAC and BC capacity regions arises solely due to the special structure of the $\log(\cdot)$-function in the capacity region expressions of the classical Gaussian-noise channels, rather than for any physical reason. The capacity expressions for the quantum bosonic channels have the $g(\cdot)$-function instead which does not exhibit the same duality properties.

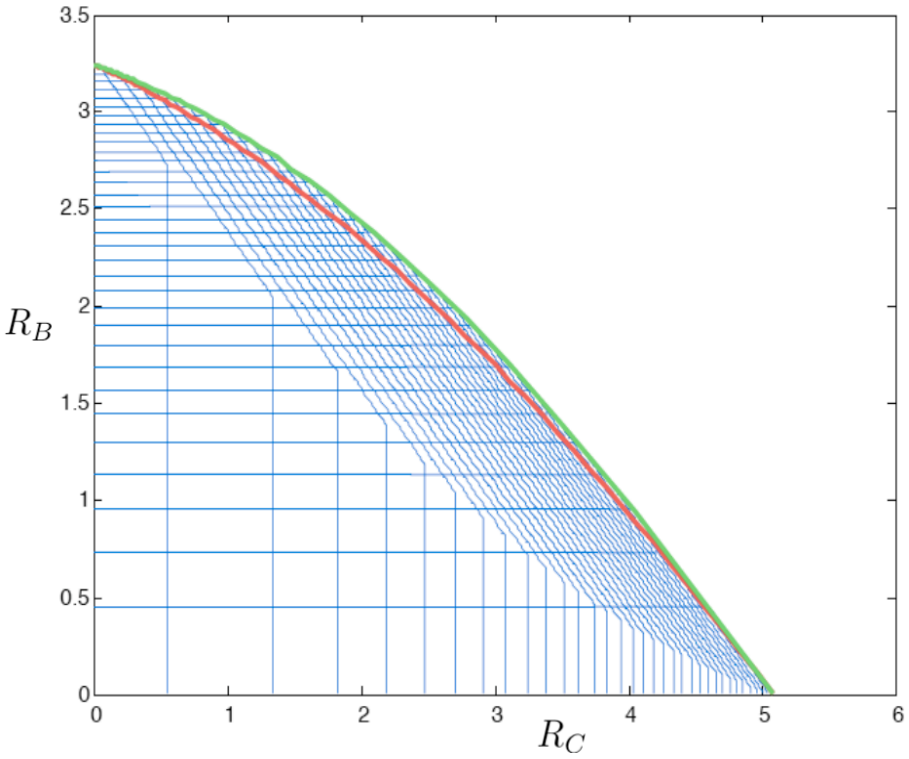


Figure 3-15: Comparison of bosonic broadcast and multiple-access channel capacity regions for $\eta = 0.8$, and $\bar{N} = 15$. The rates are in the units of bits per channel use. The red line is the conjectured ultimate broadcast capacity region, which lies below the green line - the envelope of the MAC capacity regions. Assuming that the optimum modulation, coding, and receivers are available, on a fixed beam splitter with the same power budget, more collective classical information can be sent when this beam splitter is used as a multiple-access channel, as opposed to when it is used as a broadcast channel. This is unlike the case of the classical MIMO Gaussian multiple-access and broadcast channels (BC), where a duality holds between the MAC and BC capacity regions.

3.5 The Wiretap Channel and Privacy Capacity

The term "wiretap channel" was coined by Wyner [57] to describe a communication system, in which Alice wishes to communicate classical information to Bob over a point-to-point discrete memoryless channel that is subjected to a wiretap by an eavesdropper Eve. Alice's goal is to reliably and securely communicate classical data to Bob, in such a way that Eve gets no information whatsoever about the message. Wyner used the conditional entropy rate of the signal received by Eve, given Alice's transmitted message, to measure the secrecy level guaranteed by the system. He gave a single-letter characterization of the rate-equivocation region under the limiting assumption that the signal received by Eve is a degraded version of the one received by Bob. Csiszár and Körner later generalized Wyner's results to the case in which the signal received by Eve is not a degraded version of the one received by Bob [58]. These classical-channel results were later extended by Devetak [59] to encompass classical transmission over a quantum wiretap channel.

3.5.1 Quantum wiretap channel

In earlier sections in this chapter, we have defined a quantum channel $\mathcal{N}_{A\text{-}B}$ from Alice to Bob to be a trace-preserving completely positive map that transforms Alice's single-use density operator $\hat{\rho}^A$ to Bob's, $\hat{\rho}^B = \mathcal{N}_{A\text{-}B}(\hat{\rho}^A)$. The quantum wiretap channel $\mathcal{N}_{A\text{-}BE}$ is a quantum channel from Alice to an intended receiver Bob and an eavesdropper Eve . The quantum channel from Alice to Bob is obtained by tracing out E from the channel map, i.e., $\mathcal{N}_{A\text{-}B} \equiv \mathrm{Tr}_E\left(\mathcal{N}_{A\text{-}BE}\right)$, and similarly for $\mathcal{N}_{A\text{-}E}$. A quantum wiretap channel is degraded if there exists a degrading channel $\mathcal{N}_{B\text{-}E}^{\mathrm{deg}}$ such that $\mathcal{N}_{A\text{-}E} = \mathcal{N}_{B\text{-}E}^{\mathrm{deg}} \circ \mathcal{N}_{A\text{-}B}$.

The wiretap channel describes a physical scenario in which for each successive n uses of $\mathcal{N}_{A\text{-}BE}$ Alice communicates a randomly generated classical message $m \in W$ to Bob, where m is a classical index that is uniformly distributed over the set, W, of 2^{nR} possibilities. To encode and transmit m, Alice generates an instantiation $k \in K$ of a discrete random variable, and then prepares n-channel-use states that after

transmission through the channel, result in bipartite conditional density operators $\{\hat{\rho}_{m,k}^{B^n E^n}\}$. A $(2^{nR}, n, \epsilon)$ code for this channel consists of an encoder, $x^n : (W, K) \to \mathcal{A}^n$, and a positive operator-valued measure (POVM) $\{\Lambda_m^{B^n}\}$ on $\mathcal{B}^n$ such that the following conditions are satisfied for every $m \in W$.[21]

1. Bob's probability of decoding error is at most ϵ, i.e.,

$$\mathrm{Tr}\big(\hat{\rho}_{m,k}^{B^n} \Lambda_m^{B^n}\big) > 1 - \epsilon, \quad \forall k, \quad \text{and} \tag{3.140}$$

2. For any POVM $\{\Lambda_m^{E^n}\}$ on $\mathcal{E}^n$, no more than ϵ bits of information is revealed about the secret message m. Using $j \equiv (m, k)$, this condition can be expressed, in terms of the Holevo information [27, 28, 29], as follows,

$$\chi\big(p_j, \mathcal{N}_{A-E}^{\otimes n}(\rho_j^{A^n})\big) \leq \epsilon. \tag{3.141}$$

Because Holevo information may not be additive, the classical privacy capacity C_p of the quantum wiretap channel must be computed by maximizing over successive uses of the channel, i.e., for n being the number of uses of the channel [59],

$$\begin{aligned} &C_p(\mathcal{N}_{A\text{-}BE}) \\ &= \sup_n \max_{p_T(i) p_{A|T}(j|i)} \frac{1}{n} \Bigg[\chi(p_T(i), \sum_j p_{A|T}(j|i) \hat{\rho}_j^{B^n}) \\ &\quad - \chi(p_T(i), \sum_j p_{A|T}(j|i) \hat{\rho}_j^{E^n}) \Bigg] \end{aligned} \tag{3.142}$$

where the $\{\hat{\rho}_j^{A^n}\}$ are density operators on the Hilbert space $\mathcal{H}^{\otimes n}$ of n successive channel uses. The probabilities $\{p_i\}$ form a distribution over an auxiliary classical alphabet $\mathcal{T}$, of size $|\mathcal{T}|$. The ultimate privacy capacity is computed by maximizing the expression specified in (3.142) over $\{p_T(i)\}$, $\{p_{A|T}(j|i)\}$, $\{\hat{\rho}_j^{A^n}\}$, and n. For a degraded wiretap channel, the auxiliary random variable is unnecessary, and Eq. (3.142) reduces

[21] $\mathcal{A}^n$, $\mathcal{B}^n$, and $\mathcal{E}^n$ are the n-channel-use alphabets of Alice, Bob, and Eve, with respective sizes $|\mathcal{A}^n|$, $|\mathcal{B}^n|$, and $|\mathcal{E}^n|$.

to

$$C_p(\mathcal{N}_{A\text{-}BE}) = \sup_n \max_{p_A(j)} \frac{1}{n}[\chi(p_A(j), \hat{\rho}_j^{B^n}) - \chi(p_A(j), \hat{\rho}_j^{E^n})]. \tag{3.143}$$

3.5.2 Noiseless bosonic wiretap channel

The noiseless bosonic wiretap channel consists of a collection of spatial and temporal bosonic modes at the transmitter that interact with a minimal-quantum-noise environment and split into two sets of spatio-temporal modes en route to two independent receivers, one being the intended receiver and the other being the eavesdropper. The multi-mode bosonic wiretap channel is given by $\bigotimes_s \mathcal{N}_{A_s\text{-}B_sE_s}$, where $\mathcal{N}_{A_s\text{-}B_sE_s}$ is the wiretap-channel map for the sth mode, which can be obtained from the Heisenberg evolutions

$$\hat{b}_s = \sqrt{\eta_s}\,\hat{a}_s + \sqrt{1-\eta_s}\,\hat{f}_s, \tag{3.144}$$

$$\hat{e}_s = \sqrt{1-\eta_s}\,\hat{a}_s - \sqrt{\eta_s}\,\hat{f}_s, \tag{3.145}$$

where the $\{\hat{a}_s\}$ are Alice's modal annihilation operators, and $\{\hat{b}_s\}$, $\{\hat{e}_s\}$ are the corresponding modal annihilation operators for Bob and Eve, respectively. The modal transmissivities $\{\eta_s\}$ satisfy $0 \leq \eta_s \leq 1$, and the environment modes $\{\hat{f}_s\}$ are in their vacuum states. We will limit our treatment here to the single-mode bosonic wiretap channel, as the privacy capacity of the multi-mode channel can in principle be obtained by summing up capacities of all spatio-temporal modes and maximizing the sum capacity subject to an overall input-power budget using Lagrange multipliers, cf. [9], where this was done for the multi-mode single-user lossy bosonic channel.

Theorem 3.8 — Assuming the truth of minimum output entropy conjecture 2 (see chapter 4), the ultimate privacy capacity of the single-mode noiseless bosonic wiretap channel (see Fig. 3-16) with mean input photon-number constraint $\langle \hat{a}^\dagger \hat{a} \rangle \leq \bar{N}$ is

$$C_p(\mathcal{N}_{A\text{-}BE}) = g(\eta\bar{N}) - g((1-\eta)\bar{N}) \text{ nats/use}, \tag{3.146}$$

for $\eta > 1/2$ and $C_p = 0$ for $\eta \leq 1/2$. This capacity is additive and achievable with

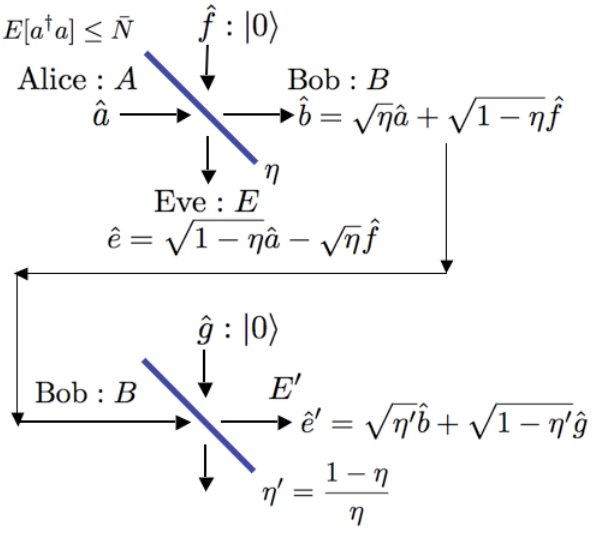


Figure 3-16: Schematic diagram of the single-mode bosonic wiretap channel. The transmitter Alice (A) encodes her messages to Bob (B) in a classical index j, and over n successive uses of the channel, thus preparing a bipartite state $\hat{\rho}_j^{B^n E^n}$ where E^n represents n channel uses of an eavesdropper Eve (E).

single-channel-use coherent-state encoding with a zero-mean isotropic Gaussian prior distribution $p_A(\alpha) = \exp(-|\alpha|^2/\bar{N})/\pi\bar{N}$.

Proof — Devetak's result for the privacy capacity of the degraded quantum wiretap channel in Eq. (3.143) requires finite-dimensional Hilbert spaces. Nevertheless, we will use this result for the bosonic wiretap channel, which has an infinite-dimensional state space, by extending it to infinite-dimensional state spaces through a limiting argument[22]. Furthermore, it was recently shown that the privacy capacity of a degraded wiretap channel is additive, and equal to the single-letter quantum capacity

[22]When $|\mathcal{T}|$ and $|\mathcal{A}|$ are finite and we are using coherent states in Eq. (3.143), there will be a finite number of possible transmitted states, leading to a finite number of possible states received by Bob and Eve. Suppose we limit the auxiliary-input alphabet (T)—and hence the input (A) and the output alphabets (B and E)—to truncated coherent states within the finite-dimensional Hilbert space spanned by the Fock states $\{\, |m\rangle : 0 \leq m \leq M \,\}$, where $M \gg \bar{N}$. Applying Devetak's theorem to the Hilbert space spanned by these truncated coherent states then gives us a lower bound on the privacy capacity of the bosonic wiretap channel when the entire, infinite-dimensional Hilbert space is employed. By taking M sufficiently large, while maintaining the cardinality condition for $\mathcal{T}$, the rate-region expressions given by Devetak's theorem will converge to Eq. (3.146).

of the channel from Alice to Bob [60], i.e.,

$$C_p(\mathcal{N}_{A\text{-}BE}) = C_p^{(1)}(\mathcal{N}_{A\text{-}BE}) = Q^{(1)}(\mathcal{N}_{A\text{-}B}), \tag{3.147}$$

where the superscript (1) denotes single-letter capacity. It is straightforward to show that if $\eta > 1/2$, the bosonic wiretap channel is a degraded channel, in which Bob's is the less-noisy receiver and Eve's is the more-noisy receiver. The degraded nature of the bosonic wiretap channel has been depicted in Fig. 3-16, where the quantum states $\hat{\rho}^{E'}$ of the constructed system E' are identical to the quantum states $\hat{\rho}^{E}$ for a given input quantum state $\hat{\rho}^{A}$. Using Eq. (3.147) for the bosonic wiretap channel, we have

$$\begin{aligned}
C_p(\mathcal{N}_{A\text{-}BE}) &= \max_{\langle \hat{a}^\dagger \hat{a} \rangle \leq \bar{N}} \left[S\left(\hat{\rho}^B\right) - S\left(\hat{\rho}^E\right) \right] \\
&= \max_{\langle \hat{b}^\dagger \hat{b} \rangle \leq \eta \bar{N}} [S(\hat{\rho}^B) - S(\hat{\rho}^{E'})] \\
&= \max_{0 \leq K \leq g(\eta \bar{N})} \{ \max\nolimits_{\langle \hat{b}^\dagger \hat{b} \rangle \leq \eta \bar{N}, S(\hat{\rho}^B) = K} [S(\hat{\rho}^B) - S(\hat{\rho}^{E'})] \} \\
&= \max_{0 \leq K \leq g(\eta \bar{N})} \{ K - \min\nolimits_{\langle \hat{b}^\dagger \hat{b} \rangle \leq \eta \bar{N}, S(\hat{\rho}^B) = K} [S(\hat{\rho}^{E'})] \} \\
&= \max_{0 \leq K \leq g(\eta \bar{N})} \{ K - g[(1-\eta) g^{-1}(K)/\eta] \} \\
&= g(\eta \bar{N}) - g((1-\eta)\bar{N}) \text{ nats/use} \\
&= Q^{(1)}(\mathcal{N}_{A\text{-}B}).
\end{aligned} \tag{3.148}$$

The first equality above follows from Lemma 3 of [60]. The second equality follows from $\mathcal{N}_{A\text{-}BE}$ being a degraded channel. The restriction to $0 \leq K \leq g(\eta \bar{N})$ in the third equality is permissible because $\max_{\langle \hat{b}^\dagger \hat{b} \rangle \leq \eta \bar{N}} S(\hat{\rho}^B) = g(\eta \bar{N})$. The fifth equality follows[23] from minimum output entropy conjecture 2 (see chapter 4), which also implies that the optimum $\hat{\rho}^B$ is a thermal state with $\langle \hat{b}^\dagger \hat{b} \rangle = \eta \bar{N}$. Hence, capacity is attained when Alice encodes using coherent-state inputs $|\alpha\rangle$ with a zero-mean isotropic

[23]Here, $g^{-1}(S)$ is the inverse of the function $g(N)$. Because $g(N)$ for $N \geq 0$ is a non-negative, monotonically increasing, concave function of N, it has an inverse, $g^{-1}(S)$ for $S \geq 0$, that is non-negative, monotonically increasing, and convex.

Gaussian prior distribution $p_A(\alpha) = (1/\pi\bar{N}) \exp\left(-|\alpha|^2/\bar{N}\right)$. The sixth equality follows from the monotonicity of the function $g(x) - g(\eta x)$ for $0 \leq \eta \leq 1$, and equality to the single-letter quantum capacity follows from Eq. (3.147). Note that the privacy capacity of this channel is zero when $\eta \leq 1/2$. It is straightforward to show that in the limit of high input photon number $\bar{N}$,

$$C_p(\mathcal{N}_{A\text{-}BE}) = Q^{(1)}(\mathcal{N}_{A\text{-}B}) = \max\left\{0, \ln(\eta) - \ln(1-\eta)\right\},$$

a result that Wolf et. al. [61] independently derived by a different approach without use of an unproven output entropy conjecture.

Chapter 4

Minimum Output Entropy Conjectures for Bosonic Channels

In general, the evolution of a quantum state resulting from the state's propagation through a quantum communication channel is not unitary, so that a pure state loses some coherence in its transit through that channel. Various measures of a channel's ability to preserve the coherence of its input state have been introduced. One of the most useful of these is the channel's capacity. In this chapter, we will focus on a different, but somewhat related measure, namely the minimum von Neumann entropy $\mathbb{S}(\mathcal{E}(\hat{\rho}))$ at the output of a quantum channel $\mathcal{E}$ optimized over the input state $\hat{\rho}$. This quantity is related to the minimum amount of noise implicit in the channel. The output entropy associated with a pure-state input measures the entanglement that such a state establishes with the environment during the communication process. Because the state of the environment is not accessible, this entanglement is responsible for the loss of quantum coherence, and hence for the injection of noise into the channel output. Low values of entanglement established with the environment correspond to low-noise communication channels. Furthermore, the study of $\mathbb{S}$ yields important information about channel capacities. In particular, we have shown that an upper bound on the classical capacity derives from a lower bound on the output entropy of multiple channel uses, see, e.g., [55]. Finally, the additivity of the minimum entropy has been shown to imply the additivity of the classical capacity and of the entan-

glement of formation [62, 63], which is a problem of huge interest to the quantum information research community.

Our study of minimum output entropy will be restricted to bosonic channels in which the optical-frequency electromagnetic field, used as the information carrier, interacts with a source of additive thermal noise. For these channels, we proposed a conjecture for the minimum output entropy [10] that, if shown to be true, would prove the ultimate rate limits to point-to-point bosonic communications, as we mentioned in Chapter 2. Even though a rigorous proof of the conjecture is yet to be seen, several attempts have been made in order to prove the conjecture, and partial results, bounds, and other supporting evidence have been found, see, e.g., [10, 55, 9, 39]. We call this conjecture, the conjecture 1. As we described in the previous chapter, a capacity analysis of the bosonic broadcast channel with two receivers and no additional noise led us to an inner bound on the capacity region, which we showed to be the ultimate capacity region under the presumption of a second minimum output entropy conjecture [12], the conjecture 2. We further saw in Chapter 3 that capacity analysis of the two-receiver and the general M-receiver bosonic broadcast channel with additive thermal noise leads to an inner bound on the capacity region achievable using coherent-state encoding. We proved that this inner bound is the ultimate capacity region under the presumption of a slightly generalized version of conjecture 2, which we call conjecture 3. We also showed in Chapter 3 that proving the single-mode version of conjecture 2 will establish the privacy capacity of the lossy bosonic channel [13]. In what follows, all these conjectures will be termed 'weak' when they are applied to single-mode states, and they will be termed 'strong' when they are applied to general n-mode bosonic states. The strong version of each conjecture subsumes the respective weak version as a special case. Neither the weak nor the strong version of these conjectures have been proven yet, but a variety of supporting evidence has been obtained, especially for conjecture 1 [10].

We will spend the next two sections of this chapter describing each minimum output entropy conjecture and its significance, along with the work that has been done so far in attempting to prove these conjectures and to obtain evidence in support of

their validity. The final section of this chapter discusses proofs of the strong versions of each minimum output conjecture for Wehrl entropy, which is an alternative measure of entropy that provides a measurement of a quantum state in phase space. The Wehrl-entropy proofs elucidate the thought process that led us recently to conjecture the Entropy Photon-Number Inequality (EPnI) [13], in analogy with the Entropy Power Inequality (EPI) from classical information theory. The EPnI subsumes all the minimum output entropy conjectures presented in this chapter, and will be the subject matter of the next chapter.

4.1 Minimum Output Entropy Conjectures

4.1.1 Conjecture 1

Weak Conjecture 1 — *Let a lossless beam splitter have input $\hat{a}$ in state $\hat{\rho}^A$, input $\hat{b}$ in a zero-mean thermal state with mean photon number N, and output $\hat{c}$ from its transmissivity-η port, i.e., $\hat{c} = \sqrt{\eta}\hat{a} + \sqrt{1-\eta}\hat{b}$. Then $S(\hat{\rho}^C)$, the von Neumann entropy of output $\hat{c}$, is minimized when the input state $\hat{\rho}^A$ is in the vacuum state (or any non-zero-mean coherent-state), and the minimum output entropy is given by $\mathbb{S}(\hat{\rho}^C) = g((1-\eta)N)$.*

Strong Conjecture 1 — *Consider n uses of a lossless beam splitter in which the output modes of the n uses, $\hat{c}_i : 1 \leq i \leq n$, are related to the input modes by*

$$\hat{c}_i = \sqrt{\eta}\hat{a}_i + \sqrt{1-\eta}\hat{b}_i, \quad \forall 1 \leq i \leq n. \tag{4.1}$$

Let the input modes $\hat{b}_i : 1 \leq i \leq n$ be in a product state of mean-photon-number N thermal states. Then putting all the $\hat{a}_i : 1 \leq i \leq n$ in their vacuum states (or equivalently in coherent states of arbitrary mean values) minimizes the output von Neumann entropy of the joint state of the $\hat{c}_i : 1 \leq i \leq n$. The resulting minimum output entropy is $\mathbb{S}(\hat{\rho}^{C^n}) = ng((1-\eta)N)$.

In [55], we showed that proving strong conjecture 1 would complete the classical-capacity proof of the point-to-point bosonic channel with additive thermal noise, and

will also prove that the capacity is achieved using a coherent-state encoding and an optimum detection scheme that employs joint measurements over long codeword blocks.

4.1.2 Conjecture 2

Weak Conjecture 2 — *Let a lossless beam splitter have input $\hat{a}$ in its vacuum state, input $\hat{b}$ in a zero-mean state with von Neumann entropy $S(\hat{\rho}^B) = g(K)$, and output $\hat{c}$ from its transmissivity-η port. Then the von Neumann entropy of output $\hat{c}$ is minimized when input $\hat{b}$ is in a thermal state with average photon number K, and the minimum output entropy is given by $\mathbb{S}(\hat{\rho}^C) = g((1-\eta)K)$.*

Strong Conjecture 2 — *Consider n uses of the beam splitter in which the output modes of the n uses, $\hat{c}_i : 1 \leq i \leq n$, are related to the input modes by Eq. 4.1. Let the input modes $\hat{a}_i : 1 \leq i \leq n$ be in a product state of n vacuum states. Also, the von Neumann entropy of the joint state of the inputs $\hat{b}_i : 1 \leq i \leq n$ is constrained to be $ng(K)$. Then, putting all the $\hat{b}_i : 1 \leq i \leq n$ in a product state of mean-photon-number K thermal states minimizes the output von Neumann entropy of the joint state of the $\hat{c}_i : 1 \leq i \leq n$. The resulting minimum output entropy is $\mathbb{S}(\hat{\rho}^{C^n}) = ng((1-\eta)K)$.*

In Chapter 3, we showed that proving strong conjecture 2 would complete the converse proof to the capacity region theorem for the general M-receiver noiseless bosonic broadcast channel. Proving the conjecture would also establish the fact that a product coherent-state encoder and optimum joint measurement detectors at each receiver achieves the ultimate capacity region for the noiseless bosonic broadcast channel.

4.1.3 Conjecture 3: An extension of Conjecture 2

Weak Conjecture 3 — *Let a lossless beam splitter have input $\hat{a}$ in a zero-mean thermal state with mean photon number N, input $\hat{b}$ in a zero-mean state with von Neumann entropy $S(\hat{\rho}^B) = g(K)$, and output $\hat{c}$ from its transmissivity-η port. Then the von Neumann entropy of output $\hat{c}$ is minimized when input $\hat{b}$ is in a thermal*

state with average photon number K, and the minimum output entropy is given by $\mathbb{S}(\hat{\rho}^C) = g(\eta N + (1-\eta)K)$.

Strong Conjecture 3 — *Consider n uses of the beam splitter in which the output modes of the n uses, $\hat{c}_i : 1 \leq i \leq n$ are related to the input modes by Equation 4.1. Let the input modes $\hat{a}_i : 1 \leq i \leq n$ be in a product state of n mean-photon-number N thermal states. Also, the von Neumann entropy of the joint state of the inputs $\hat{b}_i : 1 \leq i \leq n$ is constrained to be $ng(K)$. Then, putting all the $\hat{b}_i : 1 \leq i \leq n$ in a product state of mean-photon-number K thermal states minimizes the output von Neumann entropy of the joint state of the $\hat{c}_i : 1 \leq i \leq n$. The resulting minimum output entropy is* $\mathbb{S}(\hat{\rho}^{C^n}) = ng(\eta N + (1-\eta)K)$.

In Chapter 3, we showed that proving strong conjecture 3 would complete the converse proof to the capacity region theorem for the general M-receiver bosonic broadcast channel with additive thermal noise. Proving the conjecture would also establish the fact that a product coherent-state encoder and optimum joint measurement detectors at each receiver achieves the ultimate capacity region for the thermal-noise bosonic broadcast channel.

4.2 Evidence in Support of the Conjectures

In this section, we list all the supporting evidence that has been collected, so far, in favor of the above minimum output entropy conjectures. Most of the supporting evidence we have, is for conjecture 1, although there is some for the others.

1. **Proofs for entropy measures other than von Neumann entropy** — It turns out to be easier to work analytically with certain entropy measures that are alternatives to the von Neumann entropy, e.g., the quantum-state Wehrl entropy, Rényi entropy, and the Rényi-Wehrl entropy. Proofs for identical statements in conjectures 1, 2 and 3 have been attempted for the above alternative measures of entropy. Following are the results that were obtained.

(i) **Wehrl entropy** is the Shannon differential entropy (with an offset of $\ln \pi$) of the Husimi probability function $Q_{\hat{\rho}}(\mu)$ for the state $\hat{\rho}$ [64],

$$W(\hat{\rho}) \equiv -\int Q_{\hat{\rho}}(\mu) \ln [\pi Q_{\hat{\rho}}(\mu)] \mathrm{d}^2\mu, \tag{4.2}$$

$$= h(Q_{\hat{\rho}}(\mu)) - \ln \pi, \tag{4.3}$$

where $Q_{\hat{\rho}}(\mu) \equiv \langle \mu|\hat{\rho}|\mu\rangle/\pi$ with $|\mu\rangle$ a coherent state. The Wehrl entropy provides a measurement of the state $\hat{\rho}$ in phase space and its minimum value is achieved for coherent states [64]. Conjecture 1 (both the strong and weak forms) was proved for the Wehrl entropy measure by Giovannetti, et. al. [34]. We have proven weak conjectures 2 and 3 for Wehrl entropy using a technique similar to that was used in the Wehrl-entropy proof of conjecture 1 (see Appendix D). Later, we proved both the strong and the weak conjectures 1, 2 and 3 by using the Entropy Power Inequality (EPI) of classical information theory.

(ii) **Rényi entropy** of order z, $S_z(\hat{\rho})$, of a quantum state $\hat{\rho}$ is defined in an analogous way to the definition of Rényi entropy of order z for a classical random variable X with probability mass function $\{p_i\}$, i.e., $H_z(X) = (-1/(z-1)) \ln(\sum_i p_i^z)$:

$$S_z(\hat{\rho}) = -\frac{1}{z-1} \ln \mathrm{Tr}(\hat{\rho}^z), \quad \text{for } 0 < z < \infty, z \neq 1. \tag{4.4}$$

It is a monotonic function of the z-purity of a density operator, and reduces to the definition of the von Neumann entropy in the limit $z \to 1$. Weak and strong versions of conjecture 1 have been proven for integer-ordered Rényi entropies for $z \in \{2, 3, \ldots\}$ [34].

(iii) **Rényi-Wehrl entropy** of order z is defined by

$$W_z(\hat{\rho}) = -\frac{1}{z-1} \ln \left(\frac{1}{\pi} \int (\pi Q_{\hat{\rho}}(\mu))^z \mathrm{d}^2\mu \right), \quad \text{for } z \geq 1. \tag{4.5}$$

Thus the Wehrl entropy is the limit of $W_z(\hat{\rho})$ as $z \to 1$. Weak conjecture 1 has been proved for the Rényi-Wehrl entropy measure [34].

2. **Proof for Gaussian states** — Strong conjectures 1 and 2 have been proven for the special case in which the input states are restricted to be Gaussian, and we have shown them to be equivalent to each other under the Gaussian-input-state restriction [12]. The proofs result from the fact that Gaussian states are completely characterized by their means and covariance matrices, and if the two inputs to a beam splitter are independent Gaussian states, then the outputs of the beam splitter are a jointly-Gaussian state whose means and covariance matrix are linear functions of the means and covariance matrices of the input Gaussian states. The Gaussian-state proof for conjecture 1 appeared in [10]. Weak conjecture 3 can be proved for Gaussian-state inputs, but the strong form of conjecture 3 hasn't been proved yet under the Gaussian input-state restriction.

3. **Majorization conjecture and simulated annealing** — In [10], we proposed the majorization conjecture (which is stronger than weak conjecture 1), whose truth would imply the truth of weak conjecture 1: *The output states produced by coherent state inputs majorize all other output states.* By definition, a state $\hat{\rho}$ majorizes a state $\hat{\sigma}$ (which we denote by $\hat{\rho} \succ \hat{\sigma}$), if all ordered partial sums of the eigenvalues of $\hat{\rho}$ equal or exceed the corresponding sums for $\hat{\sigma}$, i.e.,

$$\hat{\rho} \succ \hat{\sigma} \Rightarrow \sum_{i=0}^{k} \lambda_i \geq \sum_{i=0}^{k} \mu_i, \quad \forall k \geq 0, \tag{4.6}$$

where λ_i and μ_i are the eigenvalues of $\hat{\rho}$ and $\hat{\sigma}$, respectively, arranged in decreasing order (i.e. $\lambda_0 \geq \lambda_1 \geq \ldots$). If $\hat{\rho} \succ \hat{\sigma}$, then $S(\hat{\rho}) \leq S(\hat{\sigma})$. Thus, if the majorization conjecture holds, it would imply weak conjecture 1. As a test of this conjecture, we used simulated annealing – a well-known algorithm to search for the global minimum of multivariate functions – to minimize the output entropy of the lossy thermal-noise channel. We used a variety of randomly-

generated input states to initiate the minimization, and for each case the final input state after a few hundred iterations of the algorithm was extremely close to a coherent-state, as proposed by conjecture 1. In fact, we found for all the cases we studied, that not only did the output-state at every successive iteration of the algorithm have a lower entropy than the output-state of the previous iteration, the eigenvalues of the output-state at every iteration majorized those for the preceding iteration.

4. **Lower and upper bounds** — A suite of lower and upper bounds were found for the output entropy of the lossy thermal-noise channel that support the weak conjecture 1. The details and plots appeared in [10].

5. **Local minimum condition** — In support of the strong conjecture 1, it was also shown in [10], that the product n-mode vacuum state is a local minimum of output entropy for n uses of the lossy thermal noise channel.

6. **Thermal state best of all Fock-state diagonal states** — A weaker version of conjecture 2 would be to propose that the thermal state input yields the lowest output entropy among all other input states (with the same entropy as required by conjecture 2) that are diagonal in the number-state (Fock-state) basis. We verified that this is indeed the case for several input states diagonal in the number-state basis (see Fig. 4-1).

4.3 Proof of all Strong Conjectures for Wehrl Entropy

Inasmuch as we were unable to prove the strong conjectures for von Neumann entropy, once we had the Wehrl-entropy proofs of weak conjectures 2 and 3 (see Appendix D) and the Wehrl-entropy proof of the strong conjecture 1 [65], we wanted to generalize the Wehrl-entropy proofs of conjectures 2 and 3 to their respective strong forms as well. We found that the proofs of all the strong Wehrl-entropy conjectures followed

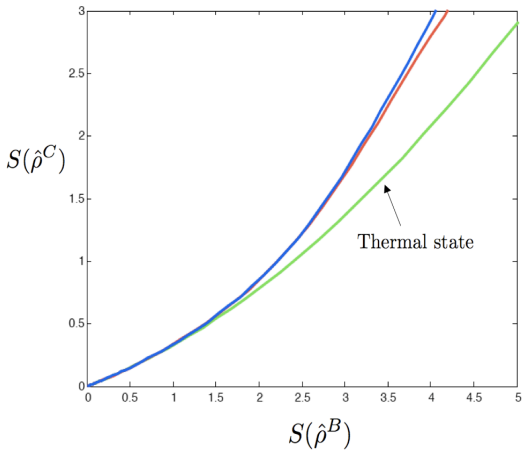


Figure 4-1: This figure presents empirical evidence in support of weak conjecture 2. The input $\hat{\rho}^A = |0\rangle\langle 0|$ is in its vacuum state. For a fixed value of $S(\hat{\rho}^B)$, we choose three different inputs $\hat{\rho}^B$, each one diagonal in the Fock-state basis, i.e. $\hat{\rho}^B = \sum_{n=0}^{\infty} p_n |n\rangle\langle n|$ with $\sum_{n=0}^{\infty} p_n = 1$. The three different inputs $\hat{\rho}^B$ correspond to choosing the distribution $\{p_n\}$ to be a Binomial distribution (blue curve), a Poisson distribution (red curve) and a Bose-Einstein distribution (green curve). As expected, we see that the output state $\hat{\rho}^C$ has the lowest entropy when $\hat{\rho}^B$ is a thermal state, i.e. when $\{p_n\}$ is a Bose-Einstein distribution.

from a simple observation that Wehrl entropy is the Shannon entropy of the Husimi function (with a fixed offset term), and that the Entropy Power Inequality (EPI) [66] for Shannon entropy encompasses the Wehrl entropy conjectures as special cases.

The Wehrl entropy is defined for an n-mode density operator $\hat{\rho}$ in a way analogous to that for a single-mode state (4.2),

$$W(\hat{\rho}) \triangleq -\int Q_{\hat{\rho}}(\boldsymbol{\mu}) \ln \left(\pi^n Q_{\hat{\rho}}(\boldsymbol{\mu})\right) \mathrm{d}^{2n}\boldsymbol{\mu} \tag{4.7}$$

$$= h(Q_{\hat{\rho}}(\boldsymbol{\mu})) - n \ln \pi, \tag{4.8}$$

where the Husimi function $Q_{\hat{\rho}}(\boldsymbol{\mu}) \equiv \langle \boldsymbol{\mu}|\hat{\rho}|\boldsymbol{\mu}\rangle/\pi^n$ is a $2n$-dimensional probability density function, with $|\boldsymbol{\mu}\rangle \triangleq |\mu_1\rangle \otimes |\mu_2\rangle \otimes \ldots \otimes |\mu_n\rangle$ being an n-mode coherent state, $\boldsymbol{\mu} \in \mathbb{C}^n$. Before we embark on the proofs, let us first state the strong versions of the minimum output entropy conjectures for Wehrl entropy.

Strong Conjecture 1 (Wehrl) — *Consider n uses of the beam splitter in which the output modes of the n uses, $\hat{c}_i : 1 \leq i \leq n$, are related to the input modes by*

Eq. 4.1. Let the input modes $\hat{b}_i : 1 \leq i \leq n$ be in a product state of n mean-photon-number K thermal states. Then, putting all the modes $\hat{a}_i : 1 \leq i \leq n$ in a product of n vacuum states minimizes the output Wehrl entropy of the joint state of the modes $\hat{c}_i : 1 \leq i \leq n$, and the minimum output entropy is given by $\mathbb{W}(\hat{\rho}^{C^n}) = n(1 + \ln(1 + (1-\eta)K))$.

Strong Conjecture 2 (Wehrl) — *Consider n uses of the beam splitter in which the output modes of the n uses, $\hat{c}_i : 1 \leq i \leq n$, are related to the input modes by Eq. 4.1. Let the input modes $\hat{a}_i : 1 \leq i \leq n$ be in a product state of n vacuum states. Also, the Wehrl entropy of the joint state of the inputs $\hat{b}_i : 1 \leq i \leq n$ is constrained to be $\hat{\rho}^{B^n} = n(1 + \ln(1 + K))$. Then, putting all the modes $\hat{b}_i : 1 \leq i \leq n$ in a product state of mean-photon-number K thermal states minimizes the output Wehrl entropy of the joint state of the modes $\hat{c}_i : 1 \leq i \leq n$, and the minimum output entropy is given by $\mathbb{W}(\hat{\rho}^{C^n}) = n(1 + \ln(1 + (1-\eta)K))$.*

Strong Conjecture 3 (Wehrl) — *Consider n uses of the beam splitter in which the output modes of the n uses, $\hat{c}_i : 1 \leq i \leq n$, are related to the input modes by Eq. 4.1. Let the input modes $\hat{a}_i : 1 \leq i \leq n$ be in a product state of n mean-photon-number N thermal states. Also, the Wehrl entropy of the joint state of the inputs $\hat{b}_i : 1 \leq i \leq n$ is constrained to be $\hat{\rho}^{B^n} = n(1 + \ln(1 + K))$. Then, putting all the modes $\hat{b}_i : 1 \leq i \leq n$ in a product state of mean-photon-number K thermal states minimizes the output Wehrl entropy of the joint state of the modes $\hat{c}_i : 1 \leq i \leq n$, and the minimum output entropy is given by $\mathbb{W}(\hat{\rho}^{C^n}) = n(1 + \ln(1 + \eta N + (1-\eta)K))$.*

Theorem 4.1 (Entropy Power Inequality (EPI)) [66] — Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be independent random m-vectors taking values in $\mathbb{R}^m$, and let $\boldsymbol{Z} = \sqrt{\eta}\boldsymbol{X} + \sqrt{1-\eta}\boldsymbol{Y}$. Then,

$$e^{2h(\boldsymbol{Z})/m} \geq \eta e^{2h(\boldsymbol{X})/m} + (1-\eta)e^{2h(\boldsymbol{Y})/m}, \tag{4.9}$$

where $h(\boldsymbol{X}) = -\int p_{\boldsymbol{X}}(\boldsymbol{x}) \ln[p_{\boldsymbol{X}}(\boldsymbol{x})] \, \mathrm{d}^m\boldsymbol{x}$ is the Shannon differential entropy of $\mathbf{X}$. Equality in (4.9) holds if and only if $\mathbf{X}$ and $\mathbf{Y}$ are both Gaussian random vectors with proportional covariance matrices.

Corollary 4.2 [Shapiro, 2007] — Consider n uses of the beam splitter in which the

output modes of the n uses, $\hat{\boldsymbol{c}} \equiv \{\hat{c}_i : 1 \leq i \leq n\}$, are related to the input modes $\hat{\boldsymbol{a}} \equiv \{\hat{a}_i : 1 \leq i \leq n\}$ and $\hat{\boldsymbol{b}} \equiv \left\{\hat{b}_i : 1 \leq i \leq n\right\}$ by Eq. 4.1. Let $\hat{\rho}^{A^n}$, $\hat{\rho}^{B^n}$ and $\hat{\rho}^{C^n}$ be the joint density operators of the n uses of the inputs and the output respectively. Then,

$$e^{W(\hat{\rho}^{C^n})/n} \geq \eta e^{W(\hat{\rho}^{A^n})/n} + (1-\eta) e^{W(\hat{\rho}^{B^n})/n}, \tag{4.10}$$

where $W(\hat{\rho})$ is the Wehrl entropy of the n-mode state $\hat{\rho}$.

Proof — Let us first recall a few definitions. The antinormally ordered characteristic function $\chi_A^{\hat{\rho}}(\boldsymbol{\zeta})$ of an n-mode density operator $\hat{\rho}$ is given by:

$$\chi_A^{\hat{\rho}}(\boldsymbol{\zeta}) = \mathrm{tr}\left(\hat{\rho} e^{-\boldsymbol{\zeta}^\dagger \hat{\boldsymbol{a}}} e^{\boldsymbol{\zeta} \hat{\boldsymbol{a}}^\dagger}\right), \tag{4.11}$$

where $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_n)^T$ is a column vector of n complex numbers. Also, the antinormally ordered characteristic function $\chi_A^{\hat{\rho}}(\boldsymbol{\zeta})$ and the Husimi function $Q_{\hat{\rho}}(\boldsymbol{\mu}) \equiv \langle \boldsymbol{\mu}|\hat{\rho}|\boldsymbol{\mu}\rangle/\pi^n$ of a state $\hat{\rho}$ form a 2-D Fourier-Transform Inverse-Transform pair:

$$\chi_A^{\hat{\rho}}(\boldsymbol{\zeta}) = \int Q_{\hat{\rho}}(\boldsymbol{\mu}) e^{\boldsymbol{\mu}^\dagger \boldsymbol{\zeta} - \boldsymbol{\zeta}^\dagger \boldsymbol{\mu}} \mathrm{d}^{2n}\boldsymbol{\mu}, \tag{4.12}$$

$$Q_{\hat{\rho}}(\boldsymbol{\mu}) = \frac{1}{\pi^{2n}} \int \chi_A^{\hat{\rho}}(\boldsymbol{\zeta}) e^{-\boldsymbol{\mu}^\dagger \boldsymbol{\zeta} + \boldsymbol{\zeta}^\dagger \boldsymbol{\mu}} \mathrm{d}^{2n}\boldsymbol{\zeta}, \tag{4.13}$$

with $\boldsymbol{\mu}, \boldsymbol{\zeta} \in \mathbb{C}^n$. As the two n-use input states $\hat{\rho}^{A^n}$ and $\hat{\rho}^{B^n}$ are statistically independent, Eq. 4.11 implies that the output state characteristic function is a product of the input state characteristic functions with scaled arguments:

$$\chi_A^{\hat{\rho}^{C^n}}(\boldsymbol{\zeta}) = \chi_A^{\hat{\rho}^{A^n}}(\sqrt{\eta}\boldsymbol{\zeta}) \chi_A^{\hat{\rho}^{B^n}}(\sqrt{1-\eta}\boldsymbol{\zeta}) \tag{4.14}$$

From Eq. 4.14, using the multiplication-convolution property of Fourier transforms (FT), we get

$$Q_{\hat{\rho}^{C^n}}(\boldsymbol{\mu}) = \frac{1}{\eta^n} Q_{\hat{\rho}^{A^n}}\left(\frac{\boldsymbol{\mu}}{\sqrt{\eta}}\right) \star \frac{1}{(1-\eta)^n} Q_{\hat{\rho}^{B^n}}\left(\frac{\boldsymbol{\mu}}{\sqrt{1-\eta}}\right) \tag{4.15}$$

where, we used the scaling-property of FT: $\chi_A^{\hat{\rho}}(\sqrt{\eta}\boldsymbol{\zeta}) \longleftrightarrow (1/\eta^n) Q_{\hat{\rho}}(\boldsymbol{\mu}/\sqrt{\eta})$.

Now, as the Husimi function $Q_{\hat{\rho}}(\cdot)$ is a proper probability density function, we can define two $2n$-dimensional statistically-independent real random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, with distributions $p_{\boldsymbol{X}}(\boldsymbol{\mu}) \triangleq Q_{\hat{\rho}^{A^n}}(\boldsymbol{\mu})$, and $p_{\boldsymbol{Y}}(\boldsymbol{\mu}) \triangleq Q_{\hat{\rho}^{B^n}}(\boldsymbol{\mu})$, and define the linear combination $\boldsymbol{Z} = \sqrt{\eta}\boldsymbol{X} + \sqrt{1-\eta}\boldsymbol{Y}$. Thus, the p.d.f. of $\boldsymbol{Z}$ is given by $p_{\boldsymbol{Z}}(\boldsymbol{\mu}) = Q_{\hat{\rho}^{C^n}}(\boldsymbol{\mu})$ as found from Eq. (4.15). Using Eq. (4.8), we have that the differential entropies of $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ can be expressed in terms of the Wehrl entropies of the n-mode quantum systems A^n, B^n and C^n respectively, by $h(\boldsymbol{X}) = W(\hat{\rho}^{A^n}) + n\ln\pi$, $h(\boldsymbol{Y}) = W(\hat{\rho}^{B^n}) + n\ln\pi$, and $h(\boldsymbol{Z}) = W(\hat{\rho}^{C^n}) + n\ln\pi$. Using these relations, Corollary 4.2 is immediately equivalent to the Entropy Power Inequality (Theorem 4.1) with $m \equiv 2n$.

Proof: Strong Conjecture 1 (Wehrl) — The input $\hat{\boldsymbol{a}}$ is given to be in a pure state. Thus the Wehrl entropy of the input $\hat{\boldsymbol{a}}$ is given by [67]

$$W(\hat{\rho}^{A^n}) = n. \tag{4.16}$$

The state of the input $\hat{\boldsymbol{b}}$ is in a product of K-photon thermal states. Therefore,

$$\begin{aligned} \hat{\rho}^{B^n} &= \left(\frac{1}{\pi K}\int e^{-|\alpha|^2/K}|\alpha\rangle\langle\alpha|\mathrm{d}^2\alpha\right)^{\otimes n}, & (4.17)\\ Q_{\hat{\rho}^{B^n}}(\boldsymbol{\mu}) &= \frac{1}{(\pi(1+K))^n}e^{-|\boldsymbol{\mu}|^2/(1+K)}, \text{and} \\ W(\hat{\rho}^{B^n}) &= n(1+\ln(1+K)), & (4.18) \end{aligned}$$

Therefore, Corollary 4.2 implies the following bound:

$$e^{W(\hat{\rho}^{C^n})/n} \geq \eta e + (1-\eta)e^{1+\ln(1+K)}, \tag{4.19}$$

which on taking the natural logarithm of both sides translates into a lower bound for the Wehrl entropy of the output $\hat{\boldsymbol{c}}$,

$$\begin{aligned} W(\hat{\rho}^{C^n}) &\geq n\ln\left(e(\eta + (1-\eta)e^{\ln(1+K)})\right) & (4.20)\\ &= n(1+\ln(1+(1-\eta)K)). & (4.21) \end{aligned}$$

It is readily verified that a product of n vacuum states at the input $\hat{\boldsymbol{a}}$, i.e. $\hat{\rho}^{A^n} = (|0\rangle\langle 0|)^{\otimes n}$ achieves the lower bound (4.21), for in this case $Q_{\hat{\rho}^{A^n}}(\boldsymbol{\mu}) = (1/\pi^n)e^{-|\boldsymbol{\mu}|^2}$, and the convolution (4.15) yields $Q_{\hat{\rho}^{C^n}}(\boldsymbol{\mu}) = 1/(\pi(1+(1-\eta)K))^n e^{-|\boldsymbol{\mu}|^2/(1+(1-\eta)K)}$, which gives $W(\hat{\rho}^{C^n}) = n(1+\ln(1+(1-\eta)K))$. Hence, a product vacuum state for the input $\hat{\boldsymbol{a}}$ achieves minimum output entropy $\mathbb{W}(\hat{\rho}^{C^n})$, and the minimum output entropy is given by

$$\mathbb{W}(\hat{\rho}^{C^n}) = n(1+\ln(1+(1-\eta)K)). \tag{4.22}$$

Proof: Strong Conjecture 2 (Wehrl) — The input $\hat{\boldsymbol{a}}$ is given to be in a an n-mode vacuum state. Thus the Husimi function and the Wehrl entropy of the input $\hat{\boldsymbol{a}}$ are given by

$$\begin{aligned} Q_{\hat{\rho}^{A^n}}(\boldsymbol{\mu}) &= \frac{1}{\pi^n} e^{-|\boldsymbol{\mu}|^2}, & (4.23)\\ W(\hat{\rho}^{A^n}) &= n. & (4.24) \end{aligned}$$

The state of the input $\hat{\boldsymbol{b}}$ is mixed with fixed Wehrl entropy $W(\hat{\rho}^{B^n}) = n(1+\ln(1+K))$. Therefore, Corollary 4.2 implies the following bound:

$$e^{W(\hat{\rho}^{C^n})/n} \geq \eta e + (1-\eta)e^{1+\ln(1+K)}, \tag{4.25}$$

which on taking the natural logarithm of both sides translates into a lower bound for the Wehrl entropy of the output $\hat{\boldsymbol{c}}$,

$$\begin{aligned} W(\hat{\rho}^{C^n}) &\geq n\ln\left(e(\eta+(1-\eta)e^{\ln(1+K)})\right) & (4.26)\\ &= n(1+\ln(1+(1-\eta)K)). & (4.27) \end{aligned}$$

It is readily verified that a product of n K-photon thermal states at the input $\hat{\boldsymbol{b}}$, i.e. $\hat{\rho}^{B^n} = \left((1/\pi K)\int e^{-|\alpha|^2/K}|\alpha\rangle\langle\alpha|\mathrm{d}^2\alpha\right)^{\otimes n}$ achieves the lower bound (4.27), for in this case $Q_{\hat{\rho}_{B^n}}(\boldsymbol{\mu}) = (1/(\pi(1+K))^n)e^{-|\boldsymbol{\mu}|^2/(1+K)}$, and the convolution (4.15) yields $Q_{\hat{\rho}^{C^n}}(\boldsymbol{\mu}) = (1/(\pi(1+(1-\eta)K))^n)e^{-|\boldsymbol{\mu}|^2/(1+(1-\eta)K)}$, which gives $W(\hat{\rho}^{C^n}) = n(1+\ln(1+(1-\eta)K))$. Hence, a product vacuum state for the input $\hat{\boldsymbol{a}}$ achieves

minimum output entropy $\mathbb{W}(\hat{\rho}^{C^n})$, and the minimum output entropy is given by

$$\mathbb{W}(\hat{\rho}^{C^n}) = n(1 + \ln(1 + (1-\eta)K)). \tag{4.28}$$

Proof: Strong Conjecture 3 (Wehrl) — The input $\hat{\boldsymbol{a}}$ is given to be in a an n-mode product thermal state with N photons on an average in each mode. Thus the Husimi function and the Wehrl entropy of the input $\hat{\boldsymbol{a}}$ are given by

$$\begin{aligned} Q_{\hat{\rho}^{A^n}}(\boldsymbol{\mu}) &= \frac{1}{(\pi(1+N))^n} e^{-|\boldsymbol{\mu}|^2/(1+N)}, \quad \text{and} & (4.29)\\ W(\hat{\rho}^{A^n}) &= n(1 + \ln(1+N)). & (4.30) \end{aligned}$$

The state of the input $\hat{\boldsymbol{b}}$ is mixed with fixed Wehrl entropy $W(\hat{\rho}^{B^n}) = n(1+\ln(1+K))$. Therefore, Corollary 4.2 implies the following bound:

$$e^{W(\hat{\rho}^{C^n})/n} \geq \eta e^{1+\ln(1+N)} + (1-\eta)e^{1+\ln(1+K)}, \tag{4.31}$$

which on taking the natural logarithm of both sides translates into a lower bound for the Wehrl entropy of the output $\hat{\boldsymbol{c}}$,

$$\begin{aligned} W(\hat{\rho}^{C^n}) &\geq n \ln\left(e(\eta(1+N) + (1-\eta)(1+K))\right) & (4.32)\\ &= n(1 + \ln(1 + \eta N + (1-\eta)K)). & (4.33) \end{aligned}$$

It is readily verified that a product of n K-photon thermal states at the input $\hat{\boldsymbol{b}}$, i.e. $\hat{\rho}^{B^n} = \left((1/\pi K)\int e^{-|\alpha|^2/K}|\alpha\rangle\langle\alpha|\mathrm{d}^2\alpha\right)^{\otimes n}$ achieves the lower bound (4.33), for in this case $Q_{\hat{\rho}_{B^n}}(\boldsymbol{\mu}) = (1/(\pi(1+K))^n)e^{-|\boldsymbol{\mu}|^2/(1+K)}$, and the convolution (4.15) yields $Q_{\hat{\rho}^{C^n}}(\boldsymbol{\mu}) = (1/(\pi(1+\eta N+(1-\eta)K))^n)e^{-|\boldsymbol{\mu}|^2/(1+\eta N+(1-\eta)K)}$, which gives $W(\hat{\rho}^{C^n}) = n(1+\ln(1+\eta N+(1-\eta)K))$. Hence, a product vacuum state for the input $\hat{\boldsymbol{a}}$ achieves minimum output entropy $\mathbb{W}(\hat{\rho}^{C^n})$, and the minimum output entropy is given by

$$\mathbb{W}(\hat{\rho}^{C^n}) = n(1 + \ln(1 + \eta N + (1-\eta)K)). \tag{4.34}$$

Chapter 5

The Entropy Photon-Number Inequality and its Consequences

In the previous chapter we saw that the Entropy Power Inequality (EPI) can be used to prove all the Wehrl-entropy versions of the minimum output entropy conjectures as special cases. The reason Wehrl entropies of the input and output states of a beam splitter admit an EPI-like inequality (corollary 4.2), is that Wehrl entropy is essentially the Shannon entropy of the Husimi function, and the Husimi function of the output state of a beam splitter is the convolution (with properly scaled arguments) of the Husimi functions of the two input states — much like how the probability distribution function (p.d.f.) of the weighted sum of two random variables is the convolution (with properly scaled arguments) of the p.d.f.'s of the two individual random variables. In order to prove the minimum output entropy conjectures for the von Neumann entropy measure, therefore, it is natural to conjecture an EPI-like inequality similar to that in corollary 4.2, that would supersede all the minimum output entropy conjectures.

In section 5.1 below, we restate the EPI in three equivalent forms, in terms of the "entropy powers" of the random variables. In section 5.2 we first restate corollary 4.2 in terms of what we define as "Wehrl-entropy photon-numbers" of the quantum states, in analogy to the notion of entropy power of a random variable introduced in section 5.1. After that we state two equivalent forms of our conjectured Entropy

Photon-number Inequality (EPnI). Section 5.3 describes how the EPnI, if true, would immediately imply all the minimum output entropy conjectures from Chapter 4. In section 5.4, we describe some recent progress that we have made towards a proof of the EPnI.

5.1 The Entropy Power Inequality (EPI)

Because a real-valued, zero-mean Gaussian random variable U has differential (Shannon) entropy given by $h(U) = \frac{1}{2}\ln(2\pi e\langle U^2\rangle)$, where the mean-squared value $\langle U^2\rangle$ is considered to be the *power* of U, we can define the **entropy power** of a random variable X, $P(X)$ to be the mean-squared value $\langle \tilde{X}^2\rangle$ of the zero-mean Gaussian random variable $\tilde{X}$ having an entropy equal to the entropy of X, i.e. $h(\tilde{X}) = h(X)$ and $P(X) = (1/2\pi e)e^{2h(X)}$. Further, let $\mathbf{X}$ and $\mathbf{Y}$ be statistically independent, n-dimensional, real-valued random vectors that possess differential entropies $h(\mathbf{X})$ and $h(\mathbf{Y})$ respectively. The entropy powers of $\mathbf{X}$ and $\mathbf{Y}$ are defined analogously:

$$P(\mathbf{X}) \equiv \frac{e^{2h(\mathbf{X})/n}}{2\pi e} \quad \text{and} \quad P(\mathbf{Y}) \equiv \frac{e^{2h(\mathbf{Y})/n}}{2\pi e}. \tag{5.1}$$

In this way, an n-dimensional, real-valued, random vector $\tilde{\mathbf{X}}$ comprised of independent, identically distributed (i.i.d.), real-valued, zero-mean, variance-$P(\mathbf{X})$, Gaussian random variables has differential entropy $h(\tilde{\mathbf{X}}) = h(\mathbf{X})$. We can similarly define an i.i.d. Gaussian random vector $\tilde{\mathbf{Y}}$ with differential entropy $h(\tilde{\mathbf{Y}}) = h(\mathbf{Y})$. We define a new random vector by the convex combination

$$\mathbf{Z} \equiv \sqrt{\eta}\,\mathbf{X} + \sqrt{1-\eta}\,\mathbf{Y}, \tag{5.2}$$

where $0 \leq \eta \leq 1$. This random vector has differential entropy $h(\mathbf{Z})$ and entropy power $P(\mathbf{Z})$. Furthermore, let $\tilde{\mathbf{Z}} \equiv \sqrt{\eta}\,\tilde{\mathbf{X}} + \sqrt{1-\eta}\,\tilde{\mathbf{Y}}$. Three equivalent forms of the

Entropy Power Inequality (EPI), see, e.g., [68], are given by

$$P(\mathbf{Z}) \geq \eta P(\mathbf{X}) + (1-\eta)P(\mathbf{Y}), \tag{5.3}$$

$$h(\mathbf{Z}) \geq h(\tilde{\mathbf{Z}}), \tag{5.4}$$

$$h(\mathbf{Z}) \geq \eta h(\mathbf{X}) + (1-\eta)h(\mathbf{Y}). \tag{5.5}$$

5.2 The Entropy Photon-Number Inequality (EPnI)

Let $\hat{\boldsymbol{a}} = [\ \hat{a}_1 \quad \hat{a}_2 \quad \cdots \quad \hat{a}_n\ ]$ and $\hat{\boldsymbol{b}} = [\ \hat{b}_1 \quad \hat{b}_2 \quad \cdots \quad \hat{b}_n\ ]$ be vectors of photon annihilation operators for a collection of $2n$ different electromagnetic field modes of frequency ω [15]. Let the joint states of the modes associated with $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ be statistically independent of each other, and thus be given by the product-state density operator $\hat{\rho}_{\boldsymbol{ab}} = \hat{\rho}_{\boldsymbol{a}} \otimes \hat{\rho}_{\boldsymbol{b}}$, where $\hat{\rho}_{\boldsymbol{a}}$ and $\hat{\rho}_{\boldsymbol{b}}$ are the density operators associated with the $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ modes, respectively. The von Neumann entropies of the $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ modes are $S(\hat{\rho}_{\boldsymbol{a}}) = -\mathrm{tr}[\hat{\rho}_{\boldsymbol{a}} \ln(\hat{\rho}_{\boldsymbol{a}})]$ and $S(\hat{\rho}_{\boldsymbol{b}}) = -\mathrm{tr}[\hat{\rho}_{\boldsymbol{b}} \ln(\hat{\rho}_{\boldsymbol{b}})]$. We define a new vector of photon annihilation operators, $\hat{\boldsymbol{c}} = [\ \hat{c}_1 \quad \hat{c}_2 \quad \cdots \quad \hat{c}_n\ ]$, by the convex combination

$$\hat{\boldsymbol{c}} \equiv \sqrt{\eta}\,\hat{\boldsymbol{a}} + \sqrt{1-\eta}\,\hat{\boldsymbol{b}}, \quad \text{for } 0 \leq \eta \leq 1, \tag{5.6}$$

and use $\hat{\rho}_{\boldsymbol{c}}$ to denote its density operator. This is equivalent to saying that $\hat{c}_i$ is the output of a lossless beam splitter whose inputs, $\hat{a}_i$ and $\hat{b}_i$, couple to that output with transmissivity η and reflectivity $1-\eta$, respectively.

5.2.1 EPnI for Wehrl entropy: Corollary 4.2

In analogy to the notion of entropy power of a random variable, let us define the **Wehrl-entropy photon numbers** of the n-mode density operators $\hat{\rho}_{\boldsymbol{a}}$ and $\hat{\rho}_{\boldsymbol{b}}$ as

follows:

$$N_W(\hat{\rho}_{\boldsymbol{a}}) \equiv g_W^{-1}\left(\frac{S(\hat{\rho}_{\boldsymbol{a}})}{n}\right), \tag{5.7}$$

$$N_W(\hat{\rho}_{\boldsymbol{b}}) \equiv g_W^{-1}\left(\frac{S(\hat{\rho}_{\boldsymbol{b}})}{n}\right), \tag{5.8}$$

where $g_W(N) \triangleq 1+\ln(1+N)$ is the Wehrl entropy of the thermal state $\hat{\rho}_T$ with mean photon number N and $g_W^{-1}(x) = e^{x-1}-1$ is the well-defined inverse function of $g_W(\cdot)$ for $x \geq 0$. Thus, if $\hat{\rho}_{\tilde{\boldsymbol{a}}} \equiv \bigotimes_{i=1}^n \hat{\rho}_{T_{a_i}}$ and $\hat{\rho}_{\tilde{\boldsymbol{b}}} \equiv \bigotimes_{i=1}^n \hat{\rho}_{T_{b_i}}$, where $\hat{\rho}_{T_{a_i}}$ is the thermal state of average photon number $N_W(\hat{\rho}_{\boldsymbol{a}})$ for the $\hat{a}_i$ mode and $\hat{\rho}_{T_{b_i}}$ is the thermal state of average photon number $N_W(\hat{\rho}_{\boldsymbol{b}})$ for the $\hat{b}_i$ mode, we have that $W(\hat{\rho}_{\tilde{\boldsymbol{a}}}) = W(\hat{\rho}_{\boldsymbol{a}})$ and $W(\hat{\rho}_{\tilde{\boldsymbol{b}}}) = W(\hat{\rho}_{\boldsymbol{b}})$.

For the vector of photon annihilation operators $\hat{\boldsymbol{c}} = [\ \hat{c}_1 \quad \hat{c}_2 \quad \cdots \quad \hat{c}_n\ ]$ that is given by the convex combination (5.6) it is straightforward to see that Eqs. (5.3)-(5.5) can be recast into the following three equivalent forms, that we call the *Wehrl-Entropy Photon-number Inequality* (WEPnI):

$$N_W(\hat{\rho}_{\boldsymbol{c}}) \geq \eta N_W(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta)N_W(\hat{\rho}_{\boldsymbol{b}}), \tag{5.9}$$

$$W(\hat{\rho}_{\boldsymbol{c}}) \geq W(\hat{\rho}_{\tilde{\boldsymbol{c}}}), \quad \text{and} \tag{5.10}$$

$$W(\hat{\rho}_{\boldsymbol{c}}) \geq \eta W(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta)W(\hat{\rho}_{\boldsymbol{b}}), \tag{5.11}$$

where $\hat{\rho}_{\tilde{\boldsymbol{c}}} \equiv \bigotimes_{i=1}^n \hat{\rho}_{T_{c_i}}$ with $\hat{\rho}_{T_{c_i}}$ being the thermal state of average photon number $\eta N_W(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta)N_W(\hat{\rho}_{\boldsymbol{b}})$ for $\hat{c}_i$. Equation (5.9) is the same as Corollary 4.2.

5.2.2 EPnI for von Neumann entropy: Conjectured

Let us define the **entropy photon numbers** of the n-mode density operators $\hat{\rho}_{\boldsymbol{a}}$ and $\hat{\rho}_{\boldsymbol{b}}$ as follows:

$$N(\hat{\rho}_{\boldsymbol{a}}) \equiv g^{-1}\left(\frac{S(\hat{\rho}_{\boldsymbol{a}})}{n}\right) \quad \text{and} \tag{5.12}$$

$$N(\hat{\rho}_{\boldsymbol{b}}) \equiv g^{-1}\left(\frac{S(\hat{\rho}_{\boldsymbol{b}})}{n}\right), \tag{5.13}$$

where $g^{-1}(y)$ is the well-defined inverse function of $y = g(x) = (1+x)\ln(1+x) - x\ln(x)$, for $x \geq 0$. Thus, if $\hat{\rho}_{\tilde{\boldsymbol{a}}} \equiv \bigotimes_{i=1}^{n} \hat{\rho}_{T_{a_i}}$ and $\hat{\rho}_{\tilde{\boldsymbol{b}}} \equiv \bigotimes_{i=1}^{n} \hat{\rho}_{T_{b_i}}$, where $\hat{\rho}_{T_{a_i}}$ is the thermal state of average photon number $N(\hat{\rho}_{\boldsymbol{a}})$ for the $\hat{a}_i$ mode and $\hat{\rho}_{T_{b_i}}$ is the thermal state of average photon number $N(\hat{\rho}_{\boldsymbol{b}})$ for the $\hat{b}_i$ mode, we have $S(\hat{\rho}_{\tilde{\boldsymbol{a}}}) = S(\hat{\rho}_{\boldsymbol{a}})$ and $S(\hat{\rho}_{\tilde{\boldsymbol{b}}}) = S(\hat{\rho}_{\boldsymbol{b}})$.

For the vector of photon annihilation operators $\hat{\boldsymbol{c}} = [\ \hat{c}_1 \quad \hat{c}_2 \quad \cdots \quad \hat{c}_n\ ]$ that is given by the convex combination (5.6), we conjecture the following two equivalent forms of the *Entropy Photon-number Inequality* (EPnI):

$$N(\hat{\rho}_{\boldsymbol{c}}) \geq \eta N(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta) N(\hat{\rho}_{\boldsymbol{b}}) \tag{5.14}$$

$$S(\hat{\rho}_{\boldsymbol{c}}) \geq S(\hat{\rho}_{\tilde{\boldsymbol{c}}}), \tag{5.15}$$

where $\hat{\rho}_{\tilde{\boldsymbol{c}}} \equiv \bigotimes_{i=1}^{n} \hat{\rho}_{T_{c_i}}$ with $\hat{\rho}_{T_{c_i}}$ being the thermal state of average photon number $\eta N(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta) N(\hat{\rho}_{\boldsymbol{b}})$ for $\hat{c}_i$. By analogy with the classical EPI and the quantum WEPnI, we might expect there to be a third equivalent form of the quantum EPnI, viz.,

$$S(\hat{\rho}_{\boldsymbol{c}}) \geq \eta S(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta) S(\hat{\rho}_{\boldsymbol{b}}). \tag{5.16}$$

It is easily shown (see below) that (5.14) implies (5.16), but we have not been able to prove the converse. Indeed, we suspect that the converse might be false.

Proof of equivalence between different forms of the EPnI

Below, we prove the equivalence of the two forms of the EPnI in Eqs. (5.14) and (5.15), and we also prove that (5.14) implies (5.16). If we can also prove that (5.16) implies (5.14), all the three forms of the conjectured EPnI would be equivalent.

1. To show that (5.14) implies (5.15), assume (5.14) is true:

$$N(\hat{\rho}_{\boldsymbol{c}}) \geq \eta N(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta) N(\hat{\rho}_{\boldsymbol{b}}) \tag{5.17}$$

$$= \eta N(\hat{\rho}_{\tilde{\boldsymbol{a}}}) + (1-\eta) N(\hat{\rho}_{\tilde{\boldsymbol{b}}}) \tag{5.18}$$

Now, if $\hat{\rho}_{\tilde{\boldsymbol{a}}\tilde{\boldsymbol{b}}} = \hat{\rho}_{\tilde{\boldsymbol{a}}} \otimes \hat{\rho}_{\tilde{\boldsymbol{b}}}$ is the joint density operator of the $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ modes, we find that the state of the $\hat{\boldsymbol{c}}$ modes is $\hat{\rho}_{\tilde{\boldsymbol{c}}} \equiv \bigotimes_{i=1}^{n} \hat{\rho}_{T_{c_i}}$, where $\hat{\rho}_{T_{c_i}}$ is a thermal state with average photon number given by $N(\hat{\rho}_{\tilde{\boldsymbol{c}}}) = \eta N(\hat{\rho}_{\tilde{\boldsymbol{a}}}) + (1-\eta)N(\hat{\rho}_{\tilde{\boldsymbol{b}}})$, so that $S(\hat{\rho}_{\tilde{\boldsymbol{c}}}) = ng[N(\hat{\rho}_{\tilde{\boldsymbol{c}}})]$. Thus, from (5.18) we get $N(\hat{\rho}_{\boldsymbol{c}}) \geq N(\hat{\rho}_{\tilde{\boldsymbol{c}}}) = g^{-1}(S(\hat{\rho}_{\tilde{\boldsymbol{c}}})/n)$. Taking $g(\cdot)$ of both sides of this inequality completes the proof.

2. To show that (5.15) implies (5.14), assume (5.15) is true:

$$\begin{aligned} N(\hat{\rho}_{\boldsymbol{c}}) &= g^{-1}(S(\hat{\rho}_{\boldsymbol{c}})/n) \\ &\geq g^{-1}(S(\hat{\rho}_{\tilde{\boldsymbol{c}}})/n) = g^{-1}[g(\eta N(\hat{\rho}_{\tilde{\boldsymbol{a}}}) + (1-\eta)N(\hat{\rho}_{\tilde{\boldsymbol{b}}}))] \\ &= \eta N(\hat{\rho}_{\tilde{\boldsymbol{a}}}) + (1-\eta)N(\hat{\rho}_{\tilde{\boldsymbol{b}}}) \\ &= \eta N(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta)N(\hat{\rho}_{\boldsymbol{b}}), \end{aligned} \tag{5.19}$$

where the inequality is due to $g^{-1}(S)$ being a monotonically increasing function of S, and the proof is complete.

3. To show that (5.14) implies (5.16), assume that (5.14) is true. We then have that $N(\hat{\rho}_{\boldsymbol{c}}) \geq \eta N(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta)N(\hat{\rho}_{\boldsymbol{b}})$, so that

$$S(\hat{\rho}_{\boldsymbol{c}}) = ng[N(\hat{\rho}_{\boldsymbol{c}})] \geq ng[\eta N(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta)N(\hat{\rho}_{\boldsymbol{b}})] \tag{5.20}$$

$$\geq \eta ng[N(\hat{\rho}_{\boldsymbol{a}})] + (1-\eta)ng[N(\hat{\rho}_{\boldsymbol{b}})] \tag{5.21}$$

$$= \eta S(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta)S(\hat{\rho}_{\boldsymbol{b}}), \tag{5.22}$$

where the second inequality follows from $g(N)$ being concave, and the proof is complete.

5.3 Relationship of the EPnI with the Minimum Output Entropy Conjectures

More important than whether or not (5.16) is equivalent to (5.14) and (5.15) is the role of the EPnI in proving classical information capacity results for Bosonic channels. In particular, the EPnI (5.14) provides simple proofs of the strong versions of the three minimum output entropy conjectures we stated in Section 4.1. These conjectures are important because proving minimum output entropy conjecture 1 also proves the conjectured capacity of the thermal-noise channel [9], proving minimum output entropy conjecture 2 also proves the conjectured capacity region of the Bosonic broadcast channel [12], and proving minimum output entropy conjecture 3 also proves the conjectured capacity region of the Bosonic broadcast channel with additive thermal noise (see Chapter 3). Furthermore, as we have shown in Chapter 3, proving minimum output entropy conjecture 2 also establishes the privacy capacity of the Bosonic wiretap channel and the single-letter quantum capacity of the lossy Bosonic channel. Before we prove that the EPnI subsumes all the minimum output entropy conjectures, we restate the conjectures below for ease of reference.

Minimum Output Entropy Conjecture 1 — Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be n-dimensional vectors of annihilation operators, with joint density operator $\hat{\rho}_{\boldsymbol{ab}} = (|\psi\rangle_{\boldsymbol{aa}}\langle\psi|) \otimes \hat{\rho}_{\boldsymbol{b}}$, where $|\psi\rangle_{\boldsymbol{a}}$ is an arbitrary zero-mean-field pure state of the $\boldsymbol{a}$ modes and $\hat{\rho}_{\boldsymbol{b}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{b_i}}$ with $\hat{\rho}_{T_{b_i}}$ being the $\hat{b}_i$ mode's thermal state of average photon number N. Define a new vector of photon annihilation operators, $\hat{\boldsymbol{c}} = [\ \hat{c}_1 \ \ \hat{c}_2 \ \ \cdots \ \ \hat{c}_n\ ]$, by the convex combination (5.6) and use $\hat{\rho}_{\boldsymbol{c}}$ to denote its density operator and $S(\hat{\rho}_{\boldsymbol{c}})$ to denote its von Neumann entropy. Then choosing $|\psi\rangle_{\boldsymbol{a}}$ to be the n-mode vacuum state minimizes $S(\hat{\rho}_{\boldsymbol{c}})$. The resulting minimum output entropy is $S(\hat{\rho}_{\boldsymbol{c}}) = ng((1-\eta)N)$.

Minimum Output Entropy Conjecture 2 — Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be n-dimensional vectors of annihilation operators with joint density operator $\hat{\rho}_{\boldsymbol{ab}} = (|\psi\rangle_{\boldsymbol{aa}}\langle\psi|) \otimes \hat{\rho}_{\boldsymbol{b}}$, where $|\psi\rangle_{\boldsymbol{a}} = \bigotimes_{i=1}^{n} |0\rangle_{a_i}$ is the n-mode vacuum state and $\hat{\rho}_{\boldsymbol{b}}$ has von Neumann entropy $S(\hat{\rho}_{\boldsymbol{b}}) = ng(K)$ for some $K \geq 0$. Define a new vector of photon annihilation operators, $\hat{\boldsymbol{c}} = [\ \hat{c}_1 \ \ \hat{c}_2 \ \ \cdots \ \ \hat{c}_n\ ]$, by the convex combination (5.6) and use $\hat{\rho}_{\boldsymbol{c}}$ to denote its

density operator and $S(\hat{\rho}_{\boldsymbol{c}})$ to denote its von Neumann entropy. Then choosing $\hat{\rho}_{\boldsymbol{b}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{b_i}}$ with $\hat{\rho}_{T_{b_i}}$ being the $\hat{b}_i$ mode's thermal state of average photon number K minimizes $S(\hat{\rho}_{\boldsymbol{c}})$. The resulting minimum output entropy is $S(\hat{\rho}_{\boldsymbol{c}}) = ng((1-\eta)K)$.

Minimum Output Entropy Conjecture 3 — Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be n-dimensional vectors of annihilation operators with joint density operator $\hat{\rho}_{\boldsymbol{ab}} = \hat{\rho}_{\boldsymbol{a}} \otimes \hat{\rho}_{\boldsymbol{b}}$, where $\hat{\rho}_{\boldsymbol{a}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{a_i}}$ with $\hat{\rho}_{T_{a_i}}$ being the $\hat{a}_i$ mode's thermal state of average photon number N, and $\hat{\rho}_{\boldsymbol{b}}$ has von Neumann entropy $S(\hat{\rho}_{\boldsymbol{b}}) = ng(K)$ for some $K \geq 0$. Define a new vector of photon annihilation operators, $\hat{\boldsymbol{c}} = [\ \hat{c}_1 \quad \hat{c}_2 \quad \cdots \quad \hat{c}_n\ ]$, by the convex combination (5.6) and use $\hat{\rho}_{\boldsymbol{c}}$ to denote its density operator and $S(\hat{\rho}_{\boldsymbol{c}})$ to denote its von Neumann entropy. Then choosing $\hat{\rho}_{\boldsymbol{b}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{b_i}}$ with $\hat{\rho}_{T_{b_i}}$ being the $\hat{b}_i$ mode's thermal state of average photon number K minimizes $S(\hat{\rho}_{\boldsymbol{c}})$. The resulting minimum output entropy is $S(\hat{\rho}_{\boldsymbol{c}}) = ng(\eta N + (1-\eta)K)$.

To see that the EPnI encompasses all three of the preceding minimum output entropy conjectures, we begin by using the premise of conjecture 1 in (5.14). Because the $\hat{\boldsymbol{a}}$ modes are in a pure state, we get $S(\hat{\rho}_{\boldsymbol{a}}) = 0$ and hence the EPnI tells us that

$$N(\hat{\rho}_{\boldsymbol{c}}) \geq (1-\eta) N(\hat{\rho}_{\boldsymbol{b}}) = (1-\eta) N. \tag{5.23}$$

Taking $g(\cdot)$ on both sides of this inequality, we get $S(\hat{\rho}_{\boldsymbol{c}})/n \geq g[(1-\eta)N]$. But, if $|\psi\rangle_{\boldsymbol{a}}$ is the n-mode vacuum state, we can easily show that $\hat{\rho}_{\boldsymbol{c}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{c_i}}$, with $\hat{\rho}_{T_{c_i}}$ being the $\hat{c}_i$ mode's thermal state of average photon number $(1-\eta)N$. Thus, when $|\psi\rangle_{\boldsymbol{a}}$ is the n-mode vacuum state we get $S(\hat{\rho}_{\boldsymbol{c}}) = ng[(1-\eta)N]$, which completes the proof.

Next, we apply the premise of conjecture 2 in (5.14). Once again, the $\hat{\boldsymbol{a}}$ modes are in a pure state, so we get

$$N(\hat{\rho}_{\boldsymbol{c}}) \geq (1-\eta) N(\hat{\rho}_{\boldsymbol{b}}) = (1-\eta) K, \tag{5.24}$$

and hence $S(\hat{\rho}_{\boldsymbol{c}})/n \geq g[(1-\eta)K]$. But, taking $\hat{\rho}_{\boldsymbol{b}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{b_i}}$, with $\hat{\rho}_{T_{b_i}}$ being the $\hat{b}_i$ mode's thermal state of average photon number K, satisfies the premise of minimum output entropy conjecture 2 and implies that $\hat{\rho}_{\boldsymbol{c}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{c_i}}$, with $\hat{\rho}_{T_{c_i}}$ being the

$\hat{c}_i$ mode's thermal state of average photon number $(1-\eta)K$. In this case we have $S(\hat{\rho}_{\boldsymbol{c}}) = ng[(1-\eta)K]$, which completes the proof.

Finally, we apply the premise of conjecture 3 in (5.14). The input state $\hat{\rho}_{\boldsymbol{a}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{a_i}}$ with $\hat{\rho}_{T_{a_i}}$ being the $\hat{a}_i$ mode's thermal state of average photon number N. So we get

$$N(\hat{\rho}_{\boldsymbol{c}}) \geq \eta N(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta)N(\hat{\rho}_{\boldsymbol{b}}) = \eta N + (1-\eta)K, \tag{5.25}$$

and hence $S(\hat{\rho}_{\boldsymbol{c}})/n \geq g[\eta N + (1-\eta)K]$. But, taking $\hat{\rho}_{\boldsymbol{b}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{b_i}}$, with $\hat{\rho}_{T_{b_i}}$ being the $\hat{b}_i$ mode's thermal state of average photon number K, satisfies the premise of minimum output entropy conjecture 3 and implies that $\hat{\rho}_{\boldsymbol{c}} = \bigotimes_{i=1}^{n} \hat{\rho}_{T_{c_i}}$, with $\hat{\rho}_{T_{c_i}}$ being the $\hat{c}_i$ mode's thermal state of average photon number $\eta N + (1-\eta)K$. In this case we have $S(\hat{\rho}_{\boldsymbol{c}}) = ng[\eta N + (1-\eta)K]$, which completes the proof.

5.4 Evidence in Support of the EPnI

As opposed to the extensive body of evidence we have that supports the validity of conjectures 1 and 2, we do not yet have nearly as much evidence for the conjectured EPnI. The EPnI might turn out to be harder to prove than our earlier conjectures, because it is a more powerful result. However, there is a huge existing literature on various ways to prove the classical EPI [68]. By drawing upon those approaches we may be able to prove the quantum EPnI. Below, we summarize the evidence we have collected so far supporting the validity of the EPnI.

5.4.1 Proof of EPnI for product Gaussian state inputs

A natural starting point in trying to prove the EPnI in its most general form would be to prove it when the input states $\hat{\rho}_{\boldsymbol{a}}$ and $\hat{\rho}_{\boldsymbol{b}}$ (and thus the output state $\hat{\rho}_{\boldsymbol{c}}$) are restricted to be Gaussian states[1]. Even though we can prove strong conjectures 1 and 2 when restricted to Gaussian input states [12], we haven't been able to prove the EPnI with this input restriction. Nevertheless, we have been able to prove the EPnI

[1]Gaussian states are states that are completely described by all the first and the second order moments of their field operators. For a quick overview of Gaussian states, see [69].

for single-mode states ($n = 1$) with the Gaussian-input restriction. In other words, we have proved the EPnI, when both the inputs $\hat{\rho}_{\boldsymbol{a}}$ and $\hat{\rho}_{\boldsymbol{b}}$ are tensor products of single-mode Gaussian states.

Theorem 5.1: [EPnI for product Gaussian state inputs: Guha, Erkmen, 2008] — Single-mode fields $\hat{a}$ and $\hat{b}$ excited in statistically independent Gaussian states $\hat{\rho}_a$ and $\hat{\rho}_b$ are inputs to a beam splitter of transmissivity η, resulting in the output mode, $\hat{c} = \sqrt{\eta}\hat{a} + \sqrt{1-\eta}\hat{b}$, in a Gaussian state $\hat{\rho}_c$. Then the following inequality holds:

$$g^{-1}\left(S(\hat{\rho}_c)\right) \geq \eta g^{-1}\left(S(\hat{\rho}_a)\right) + (1-\eta) g^{-1}\left(S(\hat{\rho}_b)\right), \tag{5.26}$$

with equality when $\hat{a}$ and $\hat{b}$ are in thermal states.

Proof — The von Neumann entropy $S(\hat{\rho}_a)$ is independent of the mean-field $\langle\hat{a}\rangle$. Hence without loss of generality, let us suppress the mean-field values of all the states and assume that $\langle\hat{a}\rangle = \langle\hat{b}\rangle = \langle\hat{c}\rangle = 0$. For a single mode Gaussian state $\hat{\rho}_a$, with mean-field $\langle\hat{a}\rangle = 0$, and covariance matrix[2],

$$K_a \triangleq \begin{pmatrix} \langle\Delta\hat{a}\Delta\hat{a}^\dagger\rangle & \langle\Delta\hat{a}^2\rangle \\ \langle\Delta\hat{a}^{\dagger 2}\rangle & \langle\Delta\hat{a}^\dagger\Delta\hat{a}\rangle \end{pmatrix} = \begin{pmatrix} \langle\hat{a}\hat{a}^\dagger\rangle & \langle\hat{a}^2\rangle \\ \langle\hat{a}^{\dagger 2}\rangle & \langle\hat{a}^\dagger\hat{a}\rangle \end{pmatrix} = \begin{pmatrix} 1+\bar{N}_a & P_a \\ P_a^* & \bar{N}_a \end{pmatrix}, \tag{5.27}$$

where $\Delta\hat{a} \equiv \hat{a} - \langle\hat{a}\rangle$, the Wigner characteristic function $\chi_W^{\hat{\rho}_a}(\zeta) \equiv \mathrm{Tr}\left(\hat{\rho}_a e^{-\zeta^*\hat{a}+\zeta\hat{a}^\dagger}\right)$ can be shown to be given by (see Appendix A)

$$\chi_W^{\hat{\rho}_a}(\zeta) = \exp\left((\alpha^*\zeta - \alpha\zeta^*) + \Re(P_a^*\zeta^2) - (\bar{N}_a + \frac{1}{2})|\zeta|^2\right). \tag{5.28}$$

Let the input state $\hat{\rho}_b$ be a Gaussian state with mean-field $\langle\hat{b}\rangle = 0$, and covariance matrix,

$$K_b \triangleq \begin{pmatrix} \langle\Delta\hat{b}\Delta\hat{b}^\dagger\rangle & \langle\Delta\hat{b}^2\rangle \\ \langle\Delta\hat{b}^{\dagger 2}\rangle & \langle\Delta\hat{b}^\dagger\Delta\hat{b}\rangle \end{pmatrix} = \begin{pmatrix} \langle\hat{b}\hat{b}^\dagger\rangle & \langle\hat{b}^2\rangle \\ \langle\hat{b}^{\dagger 2}\rangle & \langle\hat{b}^\dagger\hat{b}\rangle \end{pmatrix} = \begin{pmatrix} 1+\bar{N}_b & P_b \\ P_b^* & \bar{N}_b \end{pmatrix}. \tag{5.29}$$

[2]The commutation relation $[\hat{a}, \hat{a}^\dagger] = 1$ implies that $\langle\Delta\hat{a}\Delta\hat{a}^\dagger\rangle = 1 + \langle\Delta\hat{a}^\dagger\Delta\hat{a}\rangle$. Also, for a zero mean field ($\langle\hat{a}\rangle = 0$) state, $\langle\Delta\hat{a}^\dagger\Delta\hat{a}\rangle = \langle\hat{a}^\dagger\hat{a}\rangle$ is the mean photon number in the state, hence justifying the notation $\bar{N}_a$, as we can always choose $\langle\hat{a}\rangle = 0$ because von Neumann entropy is invariant to shifts in the mean field.

Using the beam splitter transformation $\hat{c} = \sqrt{\eta}\hat{a} + \sqrt{1-\eta}\hat{b}$, and the fact that $\hat{a}$ and $\hat{b}$ are independent modes, we can compute the Wigner characteristic function of $\hat{\rho}_c$ via $\chi_W^{\hat{\rho}_c}(\zeta) = \chi_W^{\hat{\rho}_a}(\sqrt{\eta}\zeta)\chi_W^{\hat{\rho}_b}(\sqrt{1-\eta}\zeta)$. Thus it is easy to see that $\hat{\rho}_c$ is a Gaussian state with mean field $\langle\hat{c}\rangle = \sqrt{\eta}\alpha + \sqrt{1-\eta}\beta$, and covariance matrix $K_c = \eta K_a + (1-\eta)K_b$, i.e.,

$$K_c = \begin{pmatrix} 1+\bar{N}_c & P_c \\ P_c^* & \bar{N}_c \end{pmatrix}, \tag{5.30}$$

with $\bar{N}_c = \eta\bar{N}_a + (1-\eta)\bar{N}_b$, and $P_c = \eta P_c + (1-\eta)P_b$.

When the phase-sensitive (off-diagonal) term in the covariance matrix K_a, $P_a = 0$, the Gaussian state $\hat{\rho}_a$ is a thermal state, whose Wigner characteristic function is circularly symmetric Gaussian about its mean. Using the symplectic diagonalization[3] $\hat{\rho}_a = U\hat{\rho}_{T,N_a}U^\dagger$ where $\hat{\rho}_{T,N_a}$ is a zero-mean thermal state with mean photon number $N_a = \sqrt{(\bar{N}_a + 1/2)^2 - |P_a|^2} - 1/2$, we have $S(\hat{\rho}_a) = g(N_a)$. Using symplectic diagonalizations of $\hat{\rho}_b$ and $\hat{\rho}_c$, we similarly have $S(\hat{\rho}_b) = g(N_b) = g(\sqrt{(\bar{N}_b + 1/2)^2 - |P_b|^2} - 1/2)$ and $S(\hat{\rho}_c) = g(N_c) = g(\sqrt{(\bar{N}_c + 1/2)^2 - |P_c|^2} - 1/2)$. Hence, the statement of theorem 5.1 is equivalent to the following:

For complex numbers $P_a, P_b \in \mathbb{C}$, and non-negative real numbers $N_a, N_b \in \mathbb{R}^+$, it follows that

$$\begin{aligned}\sqrt{(\bar{N}_c + 1/2)^2 - |P_c|^2} - \frac{1}{2} \;\geq\; & \eta\left(\sqrt{(\bar{N}_a + 1/2)^2 - |P_a|^2} - \frac{1}{2}\right) \\ & + (1-\eta)\left(\sqrt{(\bar{N}_b + 1/2)^2 - |P_b|^2} - \frac{1}{2}\right),\end{aligned} \tag{5.32}$$

where $P_c = \eta P_a + (1-\eta)P_b$ and $\bar{N}_c = \eta\bar{N}_a + (1-\eta)\bar{N}_b$.

[3] Any n-mode Gaussian state $\hat{\rho}_{\boldsymbol{a}}$ can be shown to be unitarily equivalent to a tensor-product of n independent thermal states with mean photon numbers λ_i, for $1 \leq i \leq n$, i.e.

$$\hat{\rho}_{\boldsymbol{a}} = U\left(\bigotimes_{i=1}^{n}\hat{\rho}_{T_i}\right)U^\dagger, \tag{5.31}$$

with $\hat{\rho}_{T_i}$ being a thermal state of average photon number λ_i. The λ_i are known as the symplectic eigenvalues of the Gaussian state $\hat{\rho}_{\boldsymbol{a}}$. Because a unitary operation leaves the von Neumann entropy of a state unchanged, $S(\hat{\rho}_{\boldsymbol{a}}) = \sum_{i=1}^{n} g(\lambda_i)$. See [70] for details of a systematic algorithm to compute the symplectic eigenvalues λ_i for an arbitrary n-mode Gaussian state, given its covariance matrix $K_{\boldsymbol{a}}$.

Lemma 5.2 — For non-negative real numbers m_1, m_2, r_1, r_2 and $\alpha \in \mathbb{R}$, satisfying $m_i \geq r_i$ for $i = 1, 2$,

$$m_1 m_2 + r_1 r_2 \cos\alpha \geq \sqrt{(m_1^2 - r_1^2)(m_2^2 - r_2^2)}. \tag{5.33}$$

Proof — Since $-1 \leq \cos\alpha \leq 1$, $m_1 m_2 + r_1 r_2 \cos\alpha \geq m_1 m_2 - r_1 r_2$. Now,

$$\begin{aligned}
(m_1 r_2 - m_2 r_1)^2 &\geq 0, \quad \text{or} && (5.34)\\
m_1^2 r_2^2 + m_2^2 r_1^2 &\geq 2 m_1 m_2 r_1 r_2, \quad \text{or} && (5.35)\\
m_1^2 m_2^2 + r_1^2 r_2^2 - 2 m_1 m_2 r_1 r_2 &\geq m_1^2 m_2^2 + r_1^2 r_2^2 - m_1^2 r_2^2 - m_2^2 r_1^2, \quad \text{or} && (5.36)\\
m_1 m_1 - r_1 r_2 &\geq \sqrt{(m_1^2 - r_1^2)(m_2^2 - r_2^2)}. && (5.37)\\
\therefore m_1 m_2 + r_1 r_2 \cos\alpha &\geq \sqrt{(m_1^2 - r_1^2)(m_2^2 - r_2^2)}. && (5.38)
\end{aligned}$$

Using *Lemma* 5.2 with the substitutions $m_1 = \bar{N}_a + 1/2$, $m_2 = \bar{N}_b + 1/2$, $P_a = r_1 e^{i\theta_1}$, $P_b = r_2 e^{i\theta_2}$ and $\alpha = \theta_1 - \theta_2$, we get[4],

$$(\bar{N}_a + \frac{1}{2})(\bar{N}_b + \frac{1}{2}) + \Re(P_a P_b^*) \geq \sqrt{\left((\bar{N}_a + \frac{1}{2})^2 - |P_a|^2\right)\left((\bar{N}_b + \frac{1}{2})^2 - |P_b|^2\right)}, \tag{5.39}$$

which can be seen to be equivalent to Eq. (5.32) with a few steps of simplification. It is readily verified from Eq. (5.32), that the inequality (5.26) is met with equality when $P_a = P_b = P_c = 0$, i.e. all the input and output states are thermal states.

5.4.2 Proof of the third form of EPnI for $\eta = 1/2$

We showed in section 5.2.2 that the conjectured EPnI (5.14) is equivalent to a second form (5.15), both of which imply a third form (5.16). We have not been able to show whether or not the third form of the EPnI is equivalent to the first two forms. In this section, we will prove (5.16) for $\eta = 1/2$.

Theorem 5.3 [Giovannetti, 2008] — Suppose that n-mode fields, $\hat{\boldsymbol{a}} = [\ \hat{a}_1 \quad \hat{a}_2 \quad \cdots \quad \hat{a}_n\ ]$

[4]Note that with these substitutions, the condition $m_i \geq r_i$ in *Lemma* 5.2 is automatically satisfied, because the symplectic eigenvalue of a Gaussian state must be non-negative. Hence, $\sqrt{(\bar{N}_a + 1/2)^2 - |P_a|^2} - \frac{1}{2} \geq 0 \Rightarrow \sqrt{(\bar{N}_a + 1/2)^2 - |P_a|^2} \geq \frac{1}{2} > 0$.

and $\hat{\boldsymbol{b}} = [\ \hat{b}_1 \quad \hat{b}_2 \quad \cdots \quad \hat{b}_n\ ]$ in statistically independent states $\hat{\rho}_{\boldsymbol{a}}$ and $\hat{\rho}_{\boldsymbol{b}}$, are the inputs to a beam splitter of transmissivity $\eta = 1/2$, resulting in the n-mode output $\hat{\boldsymbol{c}} = [\ \hat{c}_1 \quad \hat{c}_2 \quad \cdots \quad \hat{c}_n\ ]$ such that $\hat{\boldsymbol{c}} = \sqrt{\eta}\hat{\boldsymbol{a}} + \sqrt{1-\eta}\hat{\boldsymbol{b}}$. Then,

$$S(\hat{\rho}_{\boldsymbol{c}}) \geq \frac{1}{2}S(\hat{\rho}_{\boldsymbol{a}}) + \frac{1}{2}S(\hat{\rho}_{\boldsymbol{b}}). \tag{5.40}$$

Proof — Consider a beam splitter of transmissivity η with two sets of statistically independent n-mode fields $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ as inputs, producing outputs $\hat{\boldsymbol{c}} = \sqrt{\eta}\hat{\boldsymbol{a}} + \sqrt{1-\eta}\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{d}} = \sqrt{1-\eta}\hat{\boldsymbol{a}} - \sqrt{\eta}\hat{\boldsymbol{b}}$. As the evolution from the joint input state $\hat{\rho}_{\boldsymbol{ab}}$ to the joint output state $\hat{\rho}_{\boldsymbol{cd}}$ is unitary, the total entropy remains unchanged, i.e.

$$\begin{aligned} S(\hat{\rho}_{\boldsymbol{cd}}) &= S(\hat{\rho}_{\boldsymbol{ab}}) && (5.41)\\ &= S(\hat{\rho}_{\boldsymbol{a}} \otimes \hat{\rho}_{\boldsymbol{b}}) = S(\hat{\rho}_{\boldsymbol{a}}) + S(\hat{\rho}_{\boldsymbol{b}}), && (5.42) \end{aligned}$$

where the second equality follows from the independence of $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$.

Lemma 5.4 — Either one of the following must be true:

$$\begin{aligned} S(\hat{\rho}_{\boldsymbol{c}}) &\geq \eta S(\hat{\rho}_{\boldsymbol{a}}) + (1-\eta)S(\hat{\rho}_{\boldsymbol{b}}), \quad \text{OR} && (5.43)\\ S(\hat{\rho}_{\boldsymbol{d}}) &\geq (1-\eta)S(\hat{\rho}_{\boldsymbol{a}}) + \eta S(\hat{\rho}_{\boldsymbol{b}}). && (5.44) \end{aligned}$$

Proof — Assume that both (5.43) and (5.44) are false. From subadditivity of von Neumann entropy (see [6]),

$$\begin{aligned} S(\hat{\rho}_{\boldsymbol{cd}}) &\leq S(\hat{\rho}_{\boldsymbol{c}}) + S(\hat{\rho}_{\boldsymbol{d}}) && (5.45)\\ &< S(\hat{\rho}_{\boldsymbol{a}}) + S(\hat{\rho}_{\boldsymbol{b}}), && (5.46) \end{aligned}$$

where the second inequality follows from our assumption that both (5.43) and (5.44) are false. Equations (5.42) and (5.46) then imply $S(\hat{\rho}_{\boldsymbol{cd}}) < S(\hat{\rho}_{\boldsymbol{ab}})$, which is a contradiction.

Now, let $\eta = 1/2$. Using *Lemma* 5.4, either one of the following must be true:

$$S(\hat{\rho}_{\boldsymbol{c}}) \geq \frac{1}{2}S(\hat{\rho}_{\boldsymbol{a}}) + \frac{1}{2}S(\hat{\rho}_{\boldsymbol{b}}), \quad \text{OR} \tag{5.47}$$

$$S(\hat{\rho}_{\boldsymbol{d}}) \geq \frac{1}{2}S(\hat{\rho}_{\boldsymbol{a}}) + \frac{1}{2}S(\hat{\rho}_{\boldsymbol{b}}). \tag{5.48}$$

But, for $\eta = 1/2$, the Wigner characteristic functions of the two output states $\hat{\rho}_{\boldsymbol{c}}$ and $\hat{\rho}_{\boldsymbol{d}}$ are identical, i.e., $\chi_W^{\hat{\rho}_{\boldsymbol{c}}}(\boldsymbol{\zeta}) = \chi_W^{\hat{\rho}_{\boldsymbol{d}}}(\boldsymbol{\zeta}) = \chi_W^{\hat{\rho}_{\boldsymbol{a}}}(\boldsymbol{\zeta}/\sqrt{2})\chi_W^{\hat{\rho}_{\boldsymbol{b}}}(\boldsymbol{\zeta}/\sqrt{2})$, and hence the states $\hat{\rho}_{\boldsymbol{c}}$ and $\hat{\rho}_{\boldsymbol{d}}$ are identical. Therefore, $S(\hat{\rho}_{\boldsymbol{c}}) = S(\hat{\rho}_{\boldsymbol{d}})$. It follows that, Eqs. (5.47) and (5.48) imply,

$$S(\hat{\rho}_{\boldsymbol{c}}) \geq \frac{1}{2}S(\hat{\rho}_{\boldsymbol{a}}) + \frac{1}{2}S(\hat{\rho}_{\boldsymbol{b}}). \tag{5.49}$$

5.5 Monotonicity of Quantum Information

The following result is a straightforward corollary of Theorem 5.3:

Corollary 5.5 — Let $\hat{a}_1$ and $\hat{a}_2$ be single-mode inputs to a 50-50 beam splitter, producing output mode $\hat{b}_2 = (\hat{a}_1 + \hat{a}_2)/\sqrt{2}$ in state $\hat{\rho}_{b_2}$. If $\hat{a}_1$ and $\hat{a}_2$ are in identical states $\hat{\rho}_a$, then $S(\hat{\rho}_{b_2}) \geq S(\hat{\rho}_a)$.

The classical version of corollary 5.5 was proved by Shannon [2], who showed that if $Y_2 = (X_1 + X_2)/\sqrt{2}$ is a linear combination of two i.i.d. random variables with the same distribution as a random variable X, then $H(Y_2) \geq H(X)$. Shannon also proposed a general conjecture on the monotonicity of entropy, which was first proved only very recently [71].

Corollary 5.5 led us to propose a yet another conjecture, on the monotonicity of von Neumann entropy, in analogy with Shannon's conjecture on the monotonicity of classical entropy. The proof of our monotonicity conjecture is yet to be seen for the general case, even though we have been able to prove it for some special cases. In addition to the ABBN proof from [71], Shannon's monotonicity conjecture has also been proven by Tulino and Verdú [72] and by Madiman and Barron [72], each one using a different technique. In proving Shannon's monotonicity conjecture, Tulino and Verdú used the same result on the relationship between minimum mean-squared

error (MMSE) and mutual information that Verdú and Guo used to proved the EPI [66]. Hence, this suggests there might be complementary proofs for the EPnI and the quantum version of Shannon's monotonicity conjecture (see Section 5.5.2 below).

5.5.1 Shannon's conjecture on the monotonicity of entropy

The following theorem is the original form of Shannon's monotonicity conjecture:

Theorem 5.6 [Entropy increases at every step: [71, 72, 72]] — Let $\{X_1, X_2, \ldots\}$ be i.i.d. random variables, and let Y_n be the normalized running-sum defined by

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}}. \tag{5.50}$$

Then, $H(Y_{n+1}) \geq H(Y_n)$, $\forall n \in \{1, 2, \ldots\}$.

Theorem 5.6 was proved first by Artstein, Ball, Barthe, and Naor in 2004 [71] using relationships between Shannon entropy and Fisher information. Two other proofs ([72, 73]) followed a few years later.

5.5.2 A conjecture on the monotonicity of quantum entropy

In analogy to theorem 5.6, it is natural to conjecture the following generalization of corollary 5.5:

Conjecture 5.7 [von Neumann entropy increases at every step: Guha, 2008] — Let $\{\hat{a}_1, \hat{a}_2, \ldots\}$ be independent modes in identical states $\hat{\rho}_{a_i} \equiv \hat{\rho}_a$. Let us define

$$\hat{b}_n = \frac{\hat{a}_1 + \hat{a}_2 + \ldots + \hat{a}_n}{\sqrt{n}}. \tag{5.51}$$

Then, $S(\hat{\rho}_{b_{n+1}}) \geq (\hat{\rho}_{b_n})$, $\forall n \in \{1, 2, \ldots\}$.

Even though we don't have a proof of the above conjecture, we have the following two pieces of evidence that support its validity.

Proof of the monotonicity conjecture for steps of powers of 2

The following theorem proves a slightly less general version of the conjecture. We will show that $S(\hat{\rho}_{b_{2^{k+1}}}) \geq S(\hat{\rho}_{b_{2^k}})$. Thus, von Neumann entropy does increase monotonically (at steps $n = 2^k, \forall k$) as we mix in more and more modes in identical independent states, but whether or not the entropy increases at every step n is not yet known.

Theorem 5.8 [von Neumann entropy increases at powers-of-2 steps: Guha, 2008] — Let $\{\hat{a}_1, \hat{a}_2, \ldots\}$ be independent modes in identical states $\hat{\rho}_{a_i} \equiv \hat{\rho}_a$. Let us define

$$\hat{b}_n = \frac{\hat{a}_1 + \hat{a}_2 + \ldots + \hat{a}_n}{\sqrt{n}}. \tag{5.52}$$

Then, $S(\hat{\rho}_{b_{2^{k+1}}}) \geq S(\hat{\rho}_{b_{2^k}}), \forall k \in \{0, 1, \ldots\}$.

Proof — Consider

$$\begin{aligned}
\hat{b}_{2^{k+1}} &= \frac{\hat{a}_1 + \ldots + \hat{a}_{2^{k+1}}}{\sqrt{2^{k+1}}} && (5.53)\\
&= \frac{1}{\sqrt{2}}\left(\frac{\hat{a}_1 + \ldots + \hat{a}_{2^k}}{\sqrt{2^k}} + \frac{\hat{a}_{2^k+1} + \ldots + \hat{a}_{2^{k+1}}}{\sqrt{2^k}}\right) && (5.54)\\
&= \frac{1}{\sqrt{2}}\left(\hat{b}_{2^k} + \hat{b}'_{2^k}\right), \forall k \in \{0, 1, \ldots\}, && (5.55)
\end{aligned}$$

where we define $\hat{b}'_{2^k} \triangleq \frac{\hat{a}_{2^k+1} + \ldots + \hat{a}_{2^{k+1}}}{\sqrt{2^k}}$. As the $\hat{a}_i$'s are mutually independent and are in identical states $\hat{\rho}_a$, therefore $\hat{b}_{2^k}$ and $\hat{b}'_{2^k}$ must be in independent identical states, $\hat{\rho}_{b_{2^k}}$. The proof now follows from applying corollary 5.5 to the modes $\hat{b}_{2^k}$ and $\hat{b}'_{2^k}$ mixing on a 50-50 beam splitter to produce $\hat{b}_{2^{k+1}}$.

The quantum central limit theorem

An important conequence of Shannon's monotonicity result (Theorem 5.6 above) is that the convergence in the central limit theorem is monotonic. The Central Limit Theorem (CLT) states that:

Theorem 5.9 [Central Limit Theorem (CLT)] — Let $\{X_1, X_2, \ldots\}$ be independent identically distributed copies of a zero-mean random variable X with variance σ_X^2,

and let Y_n be the normalized running-sum defined by

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}}. \tag{5.56}$$

Then, Y_n converges in distribution to a zero-mean Gaussian random variable X_G with variance $\mathrm{Var}(X_G) \triangleq \sigma_X^2$, as $n \to \infty$. Hence, $\lim_{n\to\infty} H(Y_n) = H(X_G) = \frac{1}{2}\ln(2\pi e\sigma_X^2)$. The monotonicity result (Theorem 5.6) proves that $H(Y_n)$ increases monotonically as n increases, but the CLT (Theorem 5.7) says that $H(Y_n)$ converges as n increases without bound, and converges to the Gaussian random variable with the same variance as X.

In the quantum case, we have yet to prove our conjectured monotonicity result (Conjecture 5.7). However we can prove that von Neumann entropy is monotonic in n, for $n \in \{1, 2, 4, \ldots, 2^k, \ldots\}$ (Theorem 5.8). We will show below that the von Neumann entropy $S(\hat{\rho}_{b_{2^k}})$ in Theorem 5.8 also converges as $n = 2^k$ increases without bound – like the Shannon entropy in the classical case – and converges to the von Neumann entropy of a single-mode zero-mean Gaussian state with the same second order moments as the zero-mean single-mode state $\hat{\rho}_a$. To state it more precisely:

Theorem 5.10 [Quantum Central Limit Theorem (QCLT): Shapiro, 2008] — Let $\{\hat{a}_1, \hat{a}_2, \ldots\}$ be independent modes in identical zero-mean states $\hat{\rho}_{a_i} \equiv \hat{\rho}_a$. Let us define

$$\hat{b}_n = \frac{\hat{a}_1 + \hat{a}_2 + \ldots + \hat{a}_n}{\sqrt{n}}. \tag{5.57}$$

Then, the state $\hat{\rho}_{b_n}$ converges to the single-mode zero-mean Gaussian state $\hat{\rho}_G$ with covariance matrix $K_{\hat{\rho}_G} = K_a$ as $n \to \infty$. Hence, $\lim_{n\to\infty} S(\hat{\rho}_{b_n}) = S(\hat{\rho}_G) = g(\sqrt{|K_{\hat{\rho}_G}|} - 1/2)$.

Proof — From the independence of the modes $\hat{a}_i$, $1 \leq i \leq n$, we have

$$\chi_W^{\hat{\rho}_{b_n}}(\zeta) = \left[\chi_W^{\hat{\rho}_a}\left(\frac{\zeta}{\sqrt{n}}\right)\right]^n. \tag{5.58}$$

Expressing the Wigner characteristic functions in terms of the real and imaginary

parts of $\zeta = \zeta_1 + j\zeta_2$, we have

$$\ln\left[\chi_W^{\hat{\rho}_{b_n}}(\zeta)\right] = n\ln\left[\chi_W^{\hat{\rho}_a}\left(\frac{\zeta}{\sqrt{n}}\right)\right] \tag{5.59}$$

$$= n\ln\left[\langle\exp\left(\frac{-2j\zeta_1\hat{a}_2}{\sqrt{n}} + \frac{2j\zeta_2\hat{a}_1}{\sqrt{n}}\right)\rangle_{\hat{\rho}_a}\right]. \tag{5.60}$$

Note that $\chi_W^{\hat{\rho}_a}(0,0) = 1$ and that we are given $\langle\hat{a}\rangle = 0$. For a function $f(x,y)$, such that $f(0,0) = 1$, we have the following Taylor series expansion for $\ln(f(x,y))$ around $(x,y) \equiv (0,0)$:

$$\begin{aligned}\ln(f(x,y)) &= xf_x(0,0) + yf_y(0,0) + \frac{1}{2!}\left[x^2(f_{xx}(0,0) - f_x(0,0)^2)\right. \\ &\quad \left. + xy(f_{xy}(0,0) - f_x(0,0)f_y(0,0) + f_{yx}(0,0)) + y^2(f_{yy}(0,0) - f_y(0,0)^2)\right] \\ &\quad + \text{h.o.t.},\end{aligned} \tag{5.61}$$

where using which we expand $\ln\left[\chi_W^{\hat{\rho}_{b_n}}(\zeta)\right] = n\ln\left[\chi_W^{\hat{\rho}_a}\left(\frac{\zeta}{\sqrt{n}}\right)\right]$ around $(\zeta_1,\zeta_2) = (0,0)$ by evaluating all the first and second order partial derivatives of $\chi_W^{\hat{\rho}_a}(\zeta_1,\zeta_2)$. We obtain the following:

$$\ln\left[\chi_W^{\hat{\rho}_{b_n}}(\zeta)\right] = n\left[-2\left(\frac{\zeta_1^2V_2 + \zeta_2^2V_1 - 2\zeta_1\zeta_2V_{12}}{n}\right) + o\left(\frac{1}{n^{3/2}}\right)\right], \tag{5.62}$$

which implies that

$$\chi_W^{\hat{\rho}_{b_n}}(\zeta) = \exp\left[-2\left(\zeta_1^2V_2 + \zeta_2^2V_1 - 2\zeta_1\zeta_2V_{12}\right) + o\left(\frac{1}{n^{1/2}}\right)\right]. \tag{5.63}$$

Hence in the limit $n \to \infty$, $\chi_W^{\hat{\rho}_{b_n}}(\zeta)$ is identical to the Wigner characteristic function of a Gaussian state whose covariance matrix equals that of the state $\hat{\rho}_a$ (see Appendix A).

It can be shown that for a state $\hat{\rho}_a$ with covariance matrix K_a, the von Neumann entropy $S(\hat{\rho}_a)$ is maximum when $\hat{\rho}_a$ is Gaussian. Thus, the proof of the Monotonicity Conjecture for $n = 2^k$ (Theorem 5.8) along with the Quantum Central Limit Theorem (Theorem 5.10) suggest that the entropy $S\left(\hat{\rho}_{b_n}\right)$ increases monotonically as n increases, and converges to the entropy of the Gaussian state $\hat{\rho}_G$ with covariance

matrix that is the same as that of $\hat{\rho}_a$, i.e. $\lim_{n\to\infty} S\left(\hat{\rho}_{b_n}\right) = g\left(\sqrt{|K_a|} - \frac{1}{2}\right)$.

Chapter 6

Conclusions and Future Work

In this chapter, we summarize the accomplishments of the thesis, and make suggestions for future work.

6.1 Summary

Classical information theory was born with Claude Shannon's seminal 1948 paper [2], in which he derived the ultimate limits to data rates at which reliable communications can be achieved over a channel. It took almost half a century of painstaking research to come up with error-correcting codes that actually approach operating near the Shannon bound [74]. The past 40 years have also witnessed tremendous growth in the complexity and power of digital computing, and with the advent of nanoscale technologies modern-day digital computing chips are coming close to reaching their physical limits imposed by quantum mechanics. The advent of Shor's factoring algorithm [75] and some other quantum algorithms that were discovered in the past decade, has shown us that the interesting though somewhat counter-intuitive implications of the quantum nature of matter can be potentially used to our advantage in performing computing and communications tasks, and can solve some problems efficiently that have no known efficient classical solutions.

The primary motivation behind this thesis derives from the overwhelming interest in today's communications and information theory communities in pursuing the quan-

tum parallel of the half a century of work on information theory, error-control coding and the theory of digital communications that began with Shannon's work. Quantum information science has seen several advances in the past decade, and we already understand fairly well the information theory behind sending classical data reliably over point-to-point quantum communication channels, i.e., encoding classical data by modulating the quantum states of carrier particles of the medium. What is less well understood is the information theory behind sending classical data in multiple-user settings, over point-to-point quantum channels with feedback, over fading channels, over channels in which the transmitter and receiver have multiple antennas, sending quantum data reliably over quantum channels, etc. Peter Shor and Seth Lloyd have shown that the maximum of a quantity called coherent information of a channel is the maximum achievable data rate, in qubits per channel use, at which quantum information can be transmitted reliably over a quantum channel by appropriately encoding and decoding the quantum information [76, 77].

The performance of communication systems that use electromagnetic waves to carry the information are ultimately limited by noise of quantum-mechanical origin. At optical frequencies the quantum-mechanical effects are fairly pronounced and perceivable, and shot-noise-limited semiclassical photo-detection theory falls short of explaining the measurement statistics obtained by standard optical receivers detecting non-classical states of light. Thus, determining the ultimate classical information carrying capacity of optical communication channels requires quantum-mechanical analysis to properly account for the bosonic nature of optical waves. Recent research by several theorists in our group and by several others, has established capacity theorems for point-to-point bosonic channels with additive thermal noise, under the presumption of a minimum output entropy conjecture for such channels [55]. Towards the beginning of this thesis, we drew upon our work on the capacity of the point-to-point lossy bosonic channel to evaluate the optimum capacity of the free-space line-of-sight optical communication channel with Gaussian-attenuation transmit and receive apertures. Optimal power allocation across all the spatio-temporal modes was studied, in the far and near-field propagation regimes. We also compared and estab-

lished the gap between the ultimate capacity and date rates that can be achieved by using classical encoding states and structured receiver measurements.

The latter part of this the was an attempt to further the pursuit of the ultimate classical information capacity of bosonic channels, albeit in the multiple-user setting; particularly for the case in which one transmitter sends independent streams of bits to more than one receiver, viz., the broadcast channel. We drew upon recent work on the capacity region of two-user degraded quantum broadcast channels to establish ultimate capacity-region theorems for the bosonic broadcast channel, under the presumption of another conjecture on the minimum output entropy of bosonic channels. We also generalized the degraded broadcast channel capacity theorem to the case of more than two receivers, and we proved that if the above conjecture is true, the rate region achievable using a coherent-state encoding with optimal joint-detection measurement at the receivers would in fact be the ultimate capacity region of the bosonic broadcast channel with additive thermal noise and loss, and with an arbitrary number of receivers. In an attempt to the prove the minimum output entropy conjectures, we realized that these conjectures, restated for the Wehrl-entropy measure instead of von Neumann entropy, could all be shown to be immediate consequences of the entropy power inequality (EPI) – a very well known inequality in classical information theory, primarily used in proving coding-theorem converses for Gaussian channels. The upshot of the equivalence established between the EPI and the Wehrl-entropy conjectures, was our realization that an EPI-like inequality, restated in terms of the von Neumann entropy measure, would imply all the minimum output entropy conjectures that lie at the heart of several capacity results for bosonic communication channels. We therefore conjectured the entropy photon-number inequality (EPnI) in analogy with the EPI, that connects von Neumann entropies and mean photon-numbers of states of bosonic modes that linearly interact with one another. We showed that the minimum output entropy conjectures can be derived as special cases of the EPnI. We conjectured two forms of the EPnI that we proved to be equivalent to each other. We also conjectured a third form of the EPnI in analogy with the EPI, which the former two forms can be readily shown to imply, but we have not been able to show

the converse. We proved the EPnI under a product-Gaussian-state restriction, and proved the third form of the EPnI for the special case in which the input states mix in equal proportions (i.e. $\eta = 1/2$). This proof of the third form of EPnI for $\eta = 1/2$ instigated investigation into the monotonicity properties of information, which is – in its classical form – very closely tied with the EPI. In analogy with an old conjecture by Shannon, on the monotonicity of Shannon entropy of the sum of i.i.d. random variables, we proposed a quantum version of the monotonicity conjecture. We proved the conjecture but only for the special case in which the number of independent modes in the mixture increment as powers of 2, i.e. $n = 2^k$. We also proved a quantum version of the central limit theorem which along with the proof of the monotonicity conjecture for $n = 2^k$ provides strong evidence in favor of the quantum version of the monotonicity conjecture.

6.2 Future work

In what follows, we describe some of the primary open problems in line with the research done in this thesis.

6.2.1 Bosonic fading channels

In realistic unguided-propagation scenarios, transmission loss in the propagation medium is frequency-dependent, time-varying and is of probabilistic nature. Our work on the capacity of wideband free-space optical channels in Chapter 2 takes into consideration only diffraction-limited propagation and additive ambient noise from a thermal environment. Atmospheric optical transmission suffers from a variety of other propagation problems, many of which are time-varying and random, e.g., the fading that arises from the refractive-index fluctuations known as atmospheric turbulence. Drawing on our work on the lossy bosonic channel with fixed transmission loss, an outage-capacity model can be set up for the slow-fading bosonic channel, i.e., in the case in which the transmissivity changes slowly over time in comparison to the data rate. Contrary to the case of fixed transmission loss, there is no transmission

rate R, for the fading channel for which the probability of error can be driven down arbitrarily close to zero. So, in the strict sense, the capacity of the slow-fading channel is zero. An ϵ-outage capacity is the maximum rate at which one can transmit data reliably over the channel successfully, on at least a $1-\epsilon$ fraction of the total number of large blocks of channel uses in which transmission is attempted. For the fast-fading case, similar to the classical scenario, it is not unreasonable to suspect that it will be meaningful to assign a positive capacity to the channel in the usual sense, in the limit that codewords have a block-length that is much longer than the coherence time of the fade. The way one would find the fast-fading capacity, say, for the lossy bosonic channel using coherent-state inputs under a mean photon number constraint of $\bar{N}$ photons per mode at the input, would be by maximizing the Holevo quantity

$$C_{\text{fast}-\text{fade}-\text{coh}} = \max_{p(\alpha):\langle|\alpha|^2\rangle\leq\bar{N}} \chi\left(p(\alpha), \int_{\mathbb{C}}\int_0^1 p_\eta(x)|\sqrt{x}\alpha\rangle\langle\sqrt{x}\alpha|dxd^2\alpha\right), \tag{6.1}$$

where $\chi(p(\alpha),\hat{\rho}_\alpha) = S(\sum_\alpha p(\alpha)\hat{\rho}_\alpha) - \sum_\alpha p(\alpha)S(\hat{\rho}_\alpha)$ is the Holevo information for the ensemble $\{p(\alpha),\hat{\rho}_\alpha\}$, $S(\hat{\rho}) = -\text{Tr}(\hat{\rho}\log\hat{\rho})$ is the von Neumann entropy of the quantum state $\hat{\rho}$, and $p_\eta(x)$ is the probability distribution of the fast-fading transmissivity parameter η of the channel. Even though the above is an achievable rate using coherent (classical) states, for a realistic fading model such as Rayleigh or Rician fading, whether or not there would be any capacity advantage by using non-classical states for encoding, is yet to be answered.

6.2.2 The bosonic multiple-acess channel (MAC)

It was shown by Yen and Shapiro in [11] that coherent states achieve the sum-rate capacity for the bosonic MAC with two transmitters and one receiver. It was also shown that at the two corners of the capacity region of the two-user MAC (i.e., when the transmission rate for one of the two transmitters is zero), using non-classical (squeezed) states yields substantial rate-benefit over using classical (coherent) states for encoding. Finding the best achievable rate region for the bosonic MAC for two or

more users, and the best encoding states and measurement that would achieve that capacity, is still an open problem.

6.2.3 Multiple-input multiple-output (MIMO) or multiple-antenna channels

Under the presumption of a minimum output entropy conjecture, we found in this thesis the ultimate capacity region for the bosonic broadcast channel with additive thermal noise, and an arbitrary number of receivers. The degraded nature of the bosonic broadcast channel is instrumental in finding the capacity region, using extensions of known results on degraded quantum broadcast channels [52] to infinite dimensional Hilbert spaces. Multiple Input Multiple Output (MIMO) channels are those in which each transmitter and receiver may have more than one antenna. A MIMO channel can be a point-to-point, multiple-access, or a broadcast channel based on how many physical transmitters and receivers it has. The famous classical example of a degraded broadcast channel is the Gaussian-noise broadcast channel, whose capacity region was found by Bergmans [49]. The capacity region of the MIMO Gaussian broadcast channel, however,, was a long-standing open problem because of the non-degraded nature of the MIMO Gaussian channel. Very recently, the capacity of the MIMO additive-Gaussian-noise broadcast channel was found by Weingarten et. al. [78]. Finding the classical capacity region for the general bosonic MIMO broadcast channel remains an open problem.

6.2.4 The Entropy photon-number inequality (EPnI) and its consequences

The Entropy Power Inequality (EPI) from classical information theory is widely used in coding theorem converse proofs for Gaussian channels. By analogy with the EPI, we conjectured in this thesis a quantum version of the EPI, which we call the Entropy Photon-number Inequality (EPnI). We showed that the three minimum output entropy conjectures cited in Chapter 4 are simple corollaries of the EPnI. Hence, prov-

ing the EPnI would immediately establish key results for the capacities of bosonic communication channels, including (i) the classical capacity of the single-user lossy bosonic channel with additive thermal noise, (ii) the classical capacity region of the general multiple-receiver bosonic broadcast channel, – and thanks to recent work by Graeme Smith on privacy capacity of degradable channels [60] – (iii) the privacy capacity of the bosonic wiretap channel, and (iv) the ultimate quantum capacity of the lossy bosonic channel[1].

Even though the EPnI's being a stronger conjecture might make it harder to prove than the less powerful minimum output entropy conjectures, the huge literature on various wave to prove the EPI may potentially help in trying to prove the EPnI. For example, proving the EPnI for integer-ordered Renyi entropy might be a good first step as the Renyi entropy is simpler to deal with analytically than the von Neumann entropy.

6.3 Outlook for the Future

The ultimate aim of research on information theory for bosonic channels is to characterize completely the ultimate rate-limits of communications over the most general quantum network. In particular, this goal entails developing a complete theory of continuous-variable communications, error-correction and cryptography (for instance, CV quantum key distribution) for transmission of information over quantum optical channels, at rates approaching the ultimate information theoretic limits. Toward that end we need to develop a theoretical framework with which we might be able to port known robust block and convolutional qubit error-correcting codes (and design new codes) for bosonic channels where the quantum state of every field mode lives in an infinite dimensional Hilbert space, as opposed to qubit spaces for which the theory of quantum error-correcting codes (QECC) has been built. In classical communications, by sampling and quantizing band-limited signals, it is possible to use bit-error

[1]The ultimate quantum capacity of the lossy bosonic channel has been found by Wolf. et. al. by a technique that doesn't make use of any unproven conjecture. Wolf's capacity result agrees with ours and hence lends more evidence to the truth of the second minimum output entropy conjecture.

correcting block and convolutional codes on analog continuous-time channels, such as the band-limited additive white Gaussian noise (AWGN) channel. Plots of symbol-error probability versus channel signal-to-noise ratio (SNR) quantify the performance of specific codes over a given channel, in terms of the distance from the theoretical bound imposed by Shannon. For instance, state-of-the-art turbo codes [74] with soft-input soft-output (SISO) iterative decoding are known to perform within 0.1 dB of the Shannon bound at a probability of symbol error of 10^{-5}. It would be nice to be able to make a similar statement about the performance of, say, a quantum convolutional code (QCC) over a lossy bosonic channel with additive thermal noise for transmission of quantum information, e.g.,"The fidelity of decoding a certain QCC over a lossy thermal noise channel increases as a function of the channel SNR, and is within 0.1 dB of the theoretical bound set by the quantum coherent information". Continuous-variable quantum key distribution is a topic on which a great deal of work has been done recently [79], but more work is still needed to find the best secret key rates, and the optimal protocols to achieve those rates over bosonic channels. Some work has been done by Gottesman, Kitaev, and Preskill [80] on encoding qubit states into continuous variable field modes.

Quantum information processing has seen a huge surge of interest in the past decade, largely in academia but increasingly in industry. Whereas making a quantum computer crack a 128-bit RSA encryption code using Shor's algorithm is still a distant dream, obtaining better data rates over lasercom channels for terrestrial and deep-space applications using quantum modulation and detection schemes, or obtaining progressively more secure communications using reliable quantum key distribution (QKD) systems over existing optical channels with novel encoding schemes and quantum measurement, seem a lot more realizable in a relatively short time frame.

Appendix A

Preliminaries

This appendix will provide a brief background on quantum mechanics, quantum optics, and quantum information theory that will be useful in reading this thesis.

A.1 Quantum mechanics: states, evolution, and measurement

It was found in the early 1900s by Max Planck that the energy of electromagnetic waves must be described as consisting of small packets of energy or 'quanta' in order to explain the spectrum of black-body radiation. He postulated that a radiating body consisted of an enormous number of elementary electronic oscillators, some vibrating at one frequency and some at another, with all frequencies from zero to infinity being represented. The energy E of any one oscillator was not permitted to take on any arbitrary value, but was proportional to some integral multiple of the frequency f of the oscillator, i.e., $E = hf$, where $h = 6.626 \times 10^{-34}$ Joule seconds is the Planck's constant. In 1905, Albert Einstein used Planck's constant to explain the photoelectric effect by postulating that the energy in a beam of light occurs in concentrations that he called light quanta, that later on came to be known as photons. This led to a theory that established a duality between subatomic particles and electromagnetic waves in which particles and waves were neither one nor the other, but had certain

properties of both.

The foundations of quantum mechanics date from the early 1800s, but the real beginnings of modern quantum mechanics date from the work of Max Planck in the 1900s. The term "quantum mechanics" was first coined by Max Born in 1924. The acceptance of quantum mechanics by the general physics community is due to its accurate prediction of the physical behavior of systems, particularly of systems showing previously unexplained phenomena in which Newtonian mechanics fails, such as the black body radiation, photoelectric effect, and stable electron orbits. Most of classical physics is now recognized to be composed of special cases of quantum mechanics and/or relativity theory. Paul Dirac brought relativity theory to bear on quantum physics, so that it could properly deal with events that occur at a substantial fraction of the speed of light. Classical physics, however, also deals with gravitational forces, and no one has yet been able to bring gravity into a unified theory with the relativized quantum theory.

We will provide below a very brief account on the mathematical formulation of quantum mechanics, that will be a useful foundation for the material covered in this thesis. For detailed study of quantum mechanics, the reader is referred to one of the many popular texts on the subject, such as [81] and [82].

A.1.1 Pure and mixed states

A *pure state* in quantum mechanics is the entirety of information that may be known about a physical system. Mathematically, a pure state is a unit length vector, $|\psi\rangle$ (known as a 'ket' in Dirac notation) that lives in a complex Hilbert space $\mathcal{H}$ of possible states for that system. Expressed in terms of a set of complete basis vectors $\{|\phi_n\rangle\} \in \mathcal{H}$, $|\psi\rangle = \sum_n c_n|\phi_n\rangle$ becomes a column vector of (a possibly infinite) set of complex numbers c_n, where $\sum_n |c_n|^2 = 1$. With each pure state $|\psi\rangle$ we associate its Hermitian conjugate vector (known as a 'bra') $\langle\psi|$, which is a row vector when expressed in a basis of $\mathcal{H}$. The simplest example of a pure state is the state of a two-level system also known as a 'qubit', which is the fundamental unit of quantum information, in analogy with a 'bit' of classical information. A qubit lives in the two-

dimensional complex vector space $\mathbb{C}^2$ spanned by two orthonormal vectors $|0\rangle$ and $|1\rangle$, and can be expressed as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where $\alpha, \beta \in \mathbb{C}$, and $|\alpha|^2 + |\beta|^2 = 1$.

A *mixed state* in quantum mechanics represents classical (statistical) uncertainty about a physical system. Mathematically, a mixed state is represented by a 'density matrix' (or a density operator) $\hat{\rho}$, which is a positive definite, unit-trace operator in $\mathcal{H}$. The canonical form of a density matrix is

$$\hat{\rho} = \sum_k p_k |\psi_k\rangle\langle\psi_k|, \tag{A.1}$$

for any collection of pure states $\{|\psi_k\rangle\}$, and $\sum_k p_k = 1$. The mixed state $\hat{\rho}$ can be thought of as a statistical mixture of pure states $|\psi_k\rangle$, where the projection $|\psi_k\rangle\langle\psi_k|$ is the density operator for the pure state $|\psi_k\rangle$, though it is worth pointing out that the decomposition of a mixed state $\hat{\rho}$ as a mixture of pure states (A.1) is by no means unique. As we know, a positive definite operator $\hat{\rho}$ must have a spectral decomposition $\hat{\rho} = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$, in terms of the eigenkets $|\lambda_i\rangle$, with the unit-trace condition on $\hat{\rho}$ requiring that the eigenvalues λ_i must form a probability distribution.

A.1.2 Composite quantum systems

We shall henceforth use symbols such as A, B, C to refer to quantum systems, with $\mathcal{H}_A$ referring to the Hilbert space whose unit vectors are the pure states of the quantum system A. Given two systems A and B, the pure states of the composite system AB correspond to unit vectors in $\mathcal{H}_{AB} \equiv \mathcal{H}_A \otimes \mathcal{H}_B$. We use superscripts on pure state vectors and density matrices to identify the quantum system with which they are associated. For a multipartite density matrix $\hat{\rho}^{ABC}$, we use the notation $\hat{\rho}^{AB} = \mathrm{Tr}_C \hat{\rho}^{ABC}$ to denote the partial trace over one of the constituent quantum systems.

Let $\left\{|\phi_m\rangle^A\right\}$ and $\left\{|\phi_n\rangle^B\right\}$ represent sets of basis vectors for the state spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ of quantum systems A and B respectively. Pure states $|\psi\rangle^{AB}$ and mixed states $\hat{\rho}^{AB}$ of the composite system AB are defined similarly as above with an underlying

set of basis vectors $|\phi_{mn}\rangle^{AB} \triangleq |\phi_m\rangle^A \otimes |\phi_n\rangle^B \in \mathcal{H}_{AB}$, viz.,

$$|\psi\rangle^{AB} = \sum_{mn} c_{mn}|\phi_{mn}\rangle^{AB}, \quad \text{with } \sum_{mn}|c_{mn}|^2 = 1, \quad \text{and} \tag{A.2}$$

$$\hat{\rho}^{AB} = \sum_k p_k|\psi_k\rangle^{AB\,AB}\langle\psi_k|, \quad \text{with } p_k \geq 0, \quad \sum_k p_k = 1, \tag{A.3}$$

for pure states $|\psi_k\rangle^{AB} \in \mathcal{H}_{AB}$.

A Pure state $|\psi\rangle^{AB} \in \mathcal{H}_{AB}$ of a composite system AB can be classified into:

1. A product state — when $|\psi\rangle^{AB}$ can be decomposed into a tensor product of two pure states in A and B, i.e. $|\psi\rangle^{AB} = |\psi\rangle^A \otimes |\psi\rangle^B$.

2. An entangled state — when $|\psi\rangle^{AB}$ cannot be expressed as a tensor product of two pure states in A and B (for instance, the state $(|0\rangle|0\rangle + |1\rangle|1\rangle)/\sqrt{2}$ is a pure entangled state of a two-qubit system).[1]

A mixed state $\hat{\rho}^{AB} \in B(\mathcal{H}_{AB})$ of a composite system[2] AB can be classified into:

1. A product state — when $\hat{\rho}^{AB}$ can be decomposed into a tensor product of two states in A and B, i.e. $\hat{\rho}^{AB} = \hat{\rho}^A \otimes \hat{\rho}^B$, with at least one of $\hat{\rho}^A$ or $\hat{\rho}^B$ being a mixed state.

2. A classically-correlated state — when $\hat{\rho}^{AB}$ is not a product state, but can be expressed nevertheless as a statistical mixture of product pure states of the systems A and B, i.e. $\hat{\rho}^{AB} = \sum_k p_k(|\alpha_k\rangle^A \otimes |\beta_k\rangle^B)({}^A\langle\alpha_k| \otimes {}^B\langle\beta_k|)$, for any set of pure states $|\alpha_k\rangle \in \mathcal{H}_\mathcal{A}$ and $|\beta_k\rangle \in \mathcal{H}_B$, with $p_k \geq 0$ and $\sum_k p_k = 1$.

3. An entangled state — when $\hat{\rho}^{AB}$ is a mixed state of the composite system AB which is neither a product state nor a classically-correlated state, i.e. the joint state of the composite system has a correlation between the systems A and B

[1] Entanglement is inherently a quantum-mechanical property of composite physical systems and is stronger than any probabilistic correlation between the constituent systems that classical physics might permit. The individual states of the systems A and B, when their joint state is pure and entangled, are mixed states, which are obtained by taking a partial trace over the other system, i.e. $\hat{\rho}^A = \mathrm{Tr}_B(\hat{\rho}^{AB}) = \mathrm{Tr}_B(|\psi\rangle^{AB\,AB}\langle\psi|) \equiv \sum_n {}^B\langle\phi_n|\hat{\rho}^{AB}|\phi_n\rangle^B$, and vice versa.

[2] $B(\mathcal{H})$ is the set of all bounded operators in $\mathcal{H}$.

which is stronger than any (classical) probabilistic correlation. For instance, consider equal mixtures of the Bell states $|\alpha\rangle = (|0\rangle|0\rangle + |1\rangle|1\rangle)/\sqrt{2}$ and $|\beta\rangle = (|1\rangle|0\rangle + |0\rangle|1\rangle)/\sqrt{2}$. This is a mixed entangled state, $(|\alpha\rangle\langle\alpha| + |\beta\rangle\langle\beta|)/2$, of a two-qubit system.[3]

A.1.3 Evolution

The time evolution of a **closed system** is defined in terms of the unitary time-evolution operator $\hat{U}(t, t_0) = \exp(-i\hat{H}(t - t_0)/\hbar)$, where $\hat{H}$ is the time-independent Hamiltonian of the closed system. The evolution of the system when it is in a pure state $|\psi(t_0)\rangle$ at time t_0, and when it is in a mixed state $\hat{\rho}(t_0)$ at time t_0 are respectively given by:

$$|\psi(t)\rangle = \hat{U}(t, t_0)|\psi(t_0)\rangle, \quad \text{and} \tag{A.4}$$

$$\hat{\rho}(t) = \hat{U}(t, t_0)\hat{\rho}(t_0)\hat{U}^\dagger(t, t_0). \tag{A.5}$$

The time evolution of a general **open system**, i.e. a system that interacts with an environment is not a unitary evolution in general. The joint state of the system and the environment is a closed system and hence must follow a unitary evolution as stated above. But when we look at the evolution of the state of the system alone, it is non-unitary and is represented by what we call a *trace-preserving, completely-positive* (TPCP) map. All **quantum channels** that we study in this thesis are TPCP maps in general. A TPCP map $\mathcal{E}$ takes density operator $\hat{\rho}_{\text{in}} \in B(\mathcal{H}_{\text{in}})$ to density operator $\hat{\rho}_{\text{out}} \in B(\mathcal{H}_{\text{out}})$, and must satisfy the following properties:

(i) $\mathcal{E}$ preserves the trace, i.e., $\text{Tr}(\mathcal{E}(\hat{\rho})) = 1$ for any $\hat{\rho}_{\text{in}} \in B(\mathcal{H}_{\text{in}})$.

[3] We reiterate that if a mixed state $\hat{\rho}^{AB}$ is not decomposable into a tensor product of mixed states, i.e. $\hat{\rho}^{AB} \neq \hat{\rho}^A \otimes \hat{\rho}^B$, the joint state $\hat{\rho}^{AB}$ is NOT necessarily entangled, and it could just have classical correlations between the two constituent systems. There has been a long ongoing debate about whether the experimentally demonstrated enhancement in imaging characteristics of optical coherence tomography (OCT) systems using the entangled bi-photon state generated by spontaneous parametric downconversion (SPDC), should really be attributed to the *entanglement* property of the photon pairs. It has been shown that almost all performance enhancements obtained by using Gaussian entangled bi-photon imagers over thermal-light sources are also obtainable by using *classically-correlated* Gaussian states with phase-sensitive correlations. See [69] for details.

(ii) $\mathcal{E}$ is a convex linear map on the set of density operators $\hat{\rho}_{\text{in}} \in B(\mathcal{H}_{\text{in}})$, i.e. $\mathcal{E}(\sum_k p_k \hat{\rho}_k) = \sum_k p_k \mathcal{E}(\hat{\rho}_k)$, for any probability distribution $\{p_k\}$.

(iii) $\mathcal{E}$ is a completely positive map. This means that $\mathcal{E}$ maps positive operators in $B(\mathcal{H}_{\text{in}})$ to positive operators on $B(\mathcal{H}_{\text{out}})$, and, for any reference system R and for any positive operator $\hat{\rho} \in B(\mathcal{H}_{\text{in}} \otimes R)$, we have that $(\mathcal{E} \otimes I_R)\hat{\rho} \geq 0$ where $\hat{I}_R$ is the identity operator on R.

It can be shown that any TPCP map can be expressed in an *operator sum representation* [6], $\mathcal{E}(\hat{\rho}) = \sum_k \hat{A}_k \hat{\rho} \hat{A}_k^\dagger$, where the Kraus operators A_k must satisfy $\sum_k \hat{A}_k^\dagger \hat{A}_k = \hat{I}$ in order to preserve the trace of $\mathcal{E}(\hat{\rho})$.

A.1.4 Observables and measurement

In quantum mechanics, each dynamical observable (for instance position, momentum, energy, angular momentum, etc.) is represented by a Hermitian operator $\hat{M}$. Being a Hermitian operator, M must have a complete orthonormal set of eigenvectors $\{|\phi_m\rangle\}$ with associated real eigenvalues ϕ_m that satisfy $\hat{M}|\phi_m\rangle = \phi_m|\phi_m\rangle$. The outcome of a measurement of $\hat{M}$ on a quantum state $\hat{\rho}$ always leads to an eigenvalue ϕ_n with probability, $p(n) = \langle\phi_n|\hat{\rho}|\phi_n\rangle$. Given that the measurement result obtained is ϕ_n, the post-measurement state of the system is the eigenstate $|\phi_n\rangle$ corresponding to the eigenvalue ϕ_n. This phenomenon is known as the "collapse" of the wave function. Thus, if the system is in an eigenstate of a measurement operator $\hat{M}$ to begin with, the measurement result is known with certainty and the measurement of $\hat{M}$ doesn't alter the state of the system. The Hermitian operator $\hat{H}$ corresponding to measuring the total energy of a closed quantum system is known as the Hamiltonian for the system. The measurement of an observable as described above is also known as a *projective measurement*, as the measurement projects the state onto an eigenspace of the measurement operator.

In analogy to the evolution of an open system described above, a more general measurement on a system entails a projective measurement performed on the joint state of the system in question along with an auxiliary environment prepared in some

initial state. This general measurement scheme can be described by a set of positive semi-definite operators $\left\{\hat{\Pi}_m\right\}$ that satisfy $\sum_m \hat{\Pi}_m = \hat{I}$. If a measurement is performed on a quantum state $\hat{\rho}$, the outcome of the measurement is n with probability $p(n) = \mathrm{Tr}(\hat{\rho}\hat{\Pi}_n)$. The above description of a quantum measurement is known as the *positive operator-valued measure* (POVM) formalism and the operators $\left\{\hat{\Pi}_m\right\}$ are known as POVM operators. The POVM operators by themselves do not determine a post-measurement state. We use the POVM formalism throughout the thesis.

A.2 Quantum entropy and information measures

Amongst various measures of how *mixed* a quantum state $\hat{\rho}$ is, the information-theoretically most relevant one is the *von Neumann entropy* $S(\hat{\rho})$, which is defined as

$$S(\hat{\rho}) = -\mathrm{Tr}(\hat{\rho}\ln\hat{\rho}) \tag{A.6}$$

$$= H(\{\lambda_n\}), \tag{A.7}$$

where $H(\{\lambda_n\}) \equiv -\sum_n \lambda_n \ln\lambda_n$ is the Shannon entropy of the eigenvalues λ_n of $\hat{\rho}$. Hence, it is obvious that the von Neumann entropy of a pure state is zero, i.e. $S(|\psi\rangle\langle\psi|) = 0$. Most of quantum information theory is built around the von Neumann entropy measure of a quantum state. Below, we list a few important properties of von Neumann entropy:

A.2.1 Data Compression

In analogy with the role that Shannon entropy plays in classical information theory, it can be shown that $S(\hat{\rho}^A)$ is the optimal compression rate on the quantum system A in the state $\hat{\rho}^A \in B(\mathcal{H}_A)$. In other words, for large n, the density matrix $\hat{\rho}^{A\otimes n}$ has nearly all of its support on a subspace of $\mathcal{H}_A^{\otimes n}$ (called the *typical subspace*) of dimension $2^{nS(\hat{\rho}^A)}$. We will henceforth use the notation $S(A)$ interchangeably with $S(\hat{\rho}^A)$ to mean von Neumann entropy of the system A (or the von Neumann entropy

of the state $\hat{\rho}^A$). If A is a classical random variable, we use the function $H(A)$ to denote the Shannon entropy of A.

A.2.2 Subadditivity

The joint entropy $S(A, B)$ of a bipartite system AB is always upper bounded by the sum of the entropies of the individual systems A and B, i.e.

$$S(A, B) \leq S(A) + S(B), \tag{A.8}$$

with equality when the joint state of AB is a product state, i.e. $\hat{\rho}^{AB} = \hat{\rho}^A \otimes \hat{\rho}^B$. Another well-known inequality, known as the strong subadditivity of von Neumann entropy is given by

$$S(A, B, C) + S(B) \leq S(A, B) + S(B, C), \tag{A.9}$$

with equality when the tripartite system ABC is in a product state, i.e. $\hat{\rho}^{ABC} = \hat{\rho}^A \otimes \hat{\rho}^B \otimes \hat{\rho}^C$.

A.2.3 Joint and conditional entropy

The entropy of a bipartite system AB in a joint state $\hat{\rho}^{AB}$ is defined as $S(A, B) = -\mathrm{Tr}(\hat{\rho}^{AB} \ln \hat{\rho}^{AB})$. Even though there is no direct definition of quantum conditional entropy as in classical information theory, one may define a conditional entropy (in analogy to its classical counterpart) as $S(A|B) = S(A, B) - S(B)$. The quantum conditional entropy can be negative, contrary to its classical counterpart[4]. Furthermore, conditioning can only reduce entropy, i.e., $S(A|B, C) \leq S(A|B)$, and discarding a quantum system can never increase quantum mutual information (see Section A.2.5), i.e. $I(A; B) \leq I(A; B, C)$.

[4]For the bipartite two-qubit Bell state $|\psi\rangle^{AB} = (|00\rangle + |11\rangle)/\sqrt{2}$, $S(A|B) = S(A, B) - S(B) = 0 - 1 = -1$. The joint state of the system AB is a pure state, hence $S(A, B) = 0$, whereas the state of system B, $\hat{\rho}^B = \mathrm{Tr}_A(\hat{\rho}^{AB}) = (|0\rangle\langle 0| + |1\rangle\langle 1|)/2$ is a mixed state with entropy $S(B) = 1$.

A.2.4 Classical-quantum states

We define here the notion of classical-quantum states and classical-quantum channels. To any classical set $\mathcal{X}$, we associate a Hilbert space $\mathcal{H}_X$ with orthonormal basis $\left\{|x\rangle^X\right\}_{x\in\mathcal{X}}$, so that for any classical random variable X which takes the values $x \in \mathcal{X}$ with probability $p(x)$, we may write a density matrix

$$\hat{\rho}^X = \sum_x p(x)|x\rangle\langle x|^X \equiv \bigoplus_x p(x)$$

which is diagonal in that basis. An ensemble of quantum states $\left\{\hat{\rho}_x^B, p(x)\right\}$ can be associated, in a similar way, to a block diagonal *classical-quantum (cq)* state for the system XB:

$$\hat{\rho}^{XB} = \sum_x p(x)|x\rangle\langle x|^X \otimes \hat{\rho}_x^B \equiv \bigoplus_x p(x)\hat{\rho}_x^B, \tag{A.10}$$

where X is a classical random variable and B is a quantum system, with conditional density matrices $\hat{\rho}_x^B$. Then the conditional entropy $S(B|X)$ is then,

$$S(B|X) = \sum_x p(x)S(\hat{\rho}_x^B). \tag{A.11}$$

A.2.5 Quantum mutual information

The quantum mutual information $I(A;B)$ of a bipartite system AB is defined in analogy to Shannon mutual information as:

$$\begin{aligned} I(A;B) &= S(A) + S(B) - S(A,B) && \text{(A.12)}\\ &= S(A) - S(A|B) && \text{(A.13)}\\ &= S(B) - S(B|A). && \text{(A.14)} \end{aligned}$$

A bipartite product mixed state $\hat{\rho}^A \otimes \hat{\rho}^B$ has zero quantum mutual information. The quantum mutual information of a cq-state (A.10) is given by

$$\begin{aligned} I(X;B) &= S(B) - S(B|X) && \text{(A.15)} \\ &= S\left(\sum_x p(x)\hat{\rho}_x^B\right) - \sum_x p(x) S(\hat{\rho}_x^B) && \text{(A.16)} \\ &\triangleq \chi\left(p(x), \hat{\rho}_x^B\right), && \text{(A.17)} \end{aligned}$$

where $\chi\left(p(x), \hat{\rho}_x^B\right)$ is defined as the *Holevo information* of the ensemble of states $\left\{p(x), \hat{\rho}_x^B\right\}$. This equivalence between the input-output quantum mutual information $I(X;B)$ of a cq-system and the Holevo information $\chi\left(p(x), \hat{\rho}_x^B\right)$ will be used extensively in the thesis.

A.2.6 The Holevo bound

Suppose Alice chooses a classical message index $x \in \mathcal{X}$ with probability $p(x)$ and encodes x by preparing a quantum state $\hat{\rho}_x^A$. She sends her state to Bob through a channel $\mathcal{E}$ which then produces a state $\hat{\rho}_x^B = \mathcal{E}(\hat{\rho}_x^A)$ at Bob's end, conditioned on the classical index x. In order to obtain information about x, Bob measures his state $\hat{\rho}_x^B$ using a POVM $\left\{\hat{\Pi}_y\right\}$. The probability that the outcome of his POVM measurement is y given Alice sent x is given by $p(y|x) = \mathrm{Tr}(\hat{\rho}_x^B \hat{\Pi}_y)$. Using X and Y to denote the random variables of which x and y are instances, we know from Shannon information theory that, when Bob uses the POVM $\left\{\hat{\Pi}_y\right\}$, the maximum rate at which Alice can transmit information to Bob by a suitable encoding and decoding scheme is given by the maximum of the mutual information $I(X;Y)$ over all input distributions $p(x)$. Holevo, Schumacher and Westmoreland showed [27, 28, 29] that for a given prior $p(x)$ and POVM $\left\{\hat{\Pi}_y\right\}$, the single-use Holevo information is an upper bound on Shannon mutual information,

$$I(X;Y) \leq \chi\left(p(x), \hat{\rho}_x^B\right), \qquad \text{(A.18)}$$

which is known as the *Holevo bound*. Maximizing over $p(x)$ on both sides, one gets

$$\max_{p(x)} I(X;Y) \leq \max_{p(x)} \chi\left(p(x), \mathcal{E}(\hat{\rho}_x^A)\right). \tag{A.19}$$

As the right-hand side does not depend on the choice of the POVM elements $\left\{\hat{\Pi}_y\right\}$, the inequality is preserved by a further maximization of the left hand side over the measurements,

$$\begin{aligned} \max_{p(x),\left\{\hat{\Pi}_y\right\}} I(X;Y) &\leq \max_{p(x)} \chi\left(p(x), \mathcal{E}(\hat{\rho}_x^A)\right), \quad \text{or} & \text{(A.20)} \\ C_{1,1}(\mathcal{E}) &\leq C_{1,\infty}(\mathcal{E}), & \text{(A.21)} \end{aligned}$$

where $C_{1,1}(\mathcal{E})$ is the maximum value of the Shannon Information $I(X;Y)$ optimized over all possible symbol-by-symbol POVM measurements $\left\{\hat{\Pi}_y\right\}$. $C_{1,\infty}(\mathcal{E})$ on the other hand, is the maximum value of the Shannon Information $I(X;Y)$ optimized not only over all possible symbol-by-symbol POVM measurements, but also over arbitrary multiple-channel-use POVM measurements. As we will see below, $C_{1,\infty}(\mathcal{E})$ is the capacity of the channel $\mathcal{E}$ for transmission of classical information if Alice is limited to send single-channel-use symbols $\hat{\rho}_x^A$ and Bob may choose any joint measurement at the receiver.

A.2.7 Ultimate classical communication capacity: The HSW theorem

The classical capacity of a quantum channel is established by random coding arguments akin to those employed in classical information theory. A set of symbols $\{j\}$ is represented by a collection of input states $\{\hat{\rho}_j\}$ that are selected according to some prior distribution $\{p_j\}$. The output states $\{\hat{\rho}'_j\}$ are obtained by applying the channel's TPCP map $\mathcal{E}(\cdot)$ to these input symbols. According to the HSW Theorem, the

capacity of this channel, in nats per use, is

$$C = \sup_n (C_{n,\infty}/n) = \sup_n \{ \max_{\{p_j, \hat{\rho}_j\}} [\chi(p_j, \mathcal{E}^{\otimes n}(\hat{\rho}_j))/n] \}, \tag{A.22}$$

where $C_{n,\infty}$ is the capacity achieved when coding is performed over n-channel-use symbols and arbitrary joint-detection measurement is used at the receiver. The supremum over n is necessitated by the fact that channel capacity may be superadditive, viz., $C_{n,\infty} > nC_{1,\infty}$ is possible for quantum channels, whereas such is not the case for classical channels. The HSW Theorem tells us that Holevo information plays the role for classical information transmission over a quantum channel that Shannon's mutual information does for a classical channel.

Neither Eq. (A.17) nor Eq. (A.22) have any explicit dependence on the quantum measurement used at the receiver, so that measurement optimization is implicit within the HSW Theorem. To obtain the same capacity C by maximizing a Shannon mutual information we can introduce a positive-operator-valued measure (POVM) [6], representing the multi-symbol quantum measurement (a joint measurement over an entire codeword) performed at the receiver. For example, if single-use encoding is performed with priors $\{p_j\}$, the probability of receiving a particular m-symbol codeword, $\mathbf{k} \equiv (k_1, k_1, \ldots, k_m)$, given that $\mathbf{j} \equiv (j_1, j_2, \ldots, j_m)$ was sent is

$$\Pr(\,\mathbf{k} \mid \mathbf{j}\,) \equiv \mathrm{Tr}\left\{ \hat{\Pi}_{\mathbf{k}} \left[\bigotimes_{l=1}^{m} \mathcal{E}(\hat{\rho}_{j_l}) \right] \right\}, \tag{A.23}$$

where the POVM, $\{\hat{\Pi}_{\mathbf{k}}\}$, is a set of Hermitian operators on the Hilbert space of output states for m channel uses that resolve the identity. From $\{\, p_j, \Pr(\,\mathbf{k} \mid \mathbf{j}\,)\}$ we can then write down a Shannon mutual information for single-use encoding and m-symbol codewords that must be maximized. Ultimately, by allowing for n-channel-use symbols and optimizing over the priors, the signal states, *and* the POVM, we would arrive at the capacity predicted by the HSW Theorem. Evidently, determining capacity is easier via the HSW Theorem than it is via Shannon mutual information, because one less optimization is required. However, finding a practical system that

can approach capacity will require that we pay attention to the receiver measurement.

A.3 Quantum optics

Classical electromagnetic (EM) waves in free space in the absence of free electrostatic charge and current densities are governed by the following Maxwell's equations[5]:

$$\begin{aligned}
\boldsymbol{\nabla} \times \boldsymbol{E}(\boldsymbol{r},t) &= -\mu_0 \frac{\partial \boldsymbol{H}(\boldsymbol{r},t)}{\partial t} && \text{(A.24)}\\
\boldsymbol{\nabla} \cdot \epsilon_0 \boldsymbol{E}(\boldsymbol{r},t) &= 0 && \text{(A.25)}\\
\boldsymbol{\nabla} \times \boldsymbol{H}(\boldsymbol{r},t) &= \epsilon_0 \frac{\partial \boldsymbol{E}(\boldsymbol{r},t)}{\partial t} && \text{(A.26)}\\
\boldsymbol{\nabla} \cdot \mu_0 \boldsymbol{H}(\boldsymbol{r},t) &= 0, && \text{(A.27)}
\end{aligned}$$

where $\boldsymbol{E}(\boldsymbol{r},t)$ and $\boldsymbol{H}(\boldsymbol{r},t)$ are the electric and magnetic field intensity vectors in free space as a function of the 3D spatial coordinates $\boldsymbol{r}$ and time t. The permittivity (ϵ_0) and permeability (μ_0) of free space are constants satisfying $\mu_0\epsilon_0 = c^{-2}$, where c is the speed of light in vacuum. General solutions to these equations can be obtained by introducing a vector potential $\boldsymbol{A}(\boldsymbol{r},t)$ defined by $\boldsymbol{E} = -\partial\boldsymbol{A}/\partial t$ and $\boldsymbol{H} = (\boldsymbol{\nabla}\times\boldsymbol{A})/\mu_0$. By working in the Coulomb gauge ($\boldsymbol{\nabla}\cdot\boldsymbol{A} = 0$), it is straightforward to show that $\boldsymbol{A}(\boldsymbol{r},t)$ must satisfy the vector wave equation

$$\boldsymbol{\nabla}^2 \boldsymbol{A}(\boldsymbol{r},t) - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{A}(\boldsymbol{r},t)}{\partial t^2} = \boldsymbol{0}. \qquad \text{(A.28)}$$

By using the method of separation of variables to solve for the complex vector potential, we may express $\boldsymbol{A}(\boldsymbol{r},t) = q_{\boldsymbol{l},\sigma}(t)\boldsymbol{u}_{\boldsymbol{l},\sigma}(\boldsymbol{r})$ so that Eq. (A.28) is now expressed as the decoupled mode equations

$$\begin{aligned}
\boldsymbol{\nabla}^2 \boldsymbol{u}_{\boldsymbol{l},\sigma}(\boldsymbol{r}) + \frac{\omega_{\boldsymbol{l}}^2}{c^2}\boldsymbol{u}_{\boldsymbol{l},\sigma}(\boldsymbol{r}) &= \boldsymbol{0}, \quad \text{and} && \text{(A.29)}\\
\frac{\mathrm{d}^2 q_{\boldsymbol{l},\sigma}(t)}{\mathrm{d}t^2} + \omega_{\boldsymbol{l}}^2 q_{\boldsymbol{l},\sigma}(t) &= 0, && \text{(A.30)}
\end{aligned}$$

[5]The development of field quantization in this section has been taken from the lecture notes of MIT class 6.972, Fall 2002, taught by Prof. Jeffrey H. Shapiro.

where Eq. (A.29) is the vector Helmholtz equation, Eq. (A.30) represents the dynamics of a simple harmonic oscillator (SHO), and $-\omega_{\boldsymbol{l}}^2/c^2$ is the separation constant for doing the separation of variables. The spatial mode index $\boldsymbol{l} \equiv (l_x, l_y, l_z)$ is a triplet of non-negative integers (not all zero) and $\sigma \in (0, 1)$ is a polarization mode index. Upon solving with the simplest boundary conditions in 3D cartesian coordinates, i.e., the $V \equiv L \times L \times L$ cubical cavity, we obtain the following solutions,

$$\boldsymbol{u}_{\boldsymbol{l},\sigma}(\boldsymbol{r}) = \frac{1}{L^{3/2}} e^{j(\boldsymbol{k}_{\boldsymbol{l}} \cdot \boldsymbol{r})} \boldsymbol{e}_{\boldsymbol{l},\sigma} \quad \text{and} \tag{A.31}$$

$$q_{\boldsymbol{l},\sigma}(t) = q_{\boldsymbol{l},\sigma} e^{-j\omega_{\boldsymbol{l}} t}, \quad \text{for } t \geq 0, \tag{A.32}$$

where $\boldsymbol{k}_{\boldsymbol{l}} = (2\pi l_x/L, 2\pi l_y/L, 2\pi l_z/L)$ is the wave vector for the spatial mode $\boldsymbol{l}$, satisfying $\boldsymbol{k}_{\boldsymbol{l}} \cdot \boldsymbol{k}_{\boldsymbol{l}} = (2\pi/L)^2 \boldsymbol{l} \cdot \boldsymbol{l} = \omega_{\boldsymbol{l}}^2/c^2$. Let us renormalize the harmonic oscillator temporal mode function $q_{\boldsymbol{l},\sigma}(t)$ as follows,

$$a_{\boldsymbol{l},\sigma}(t) = \sqrt{\frac{\omega_{\boldsymbol{l}}}{2\hbar}} q_{\boldsymbol{l},\sigma}(t) \tag{A.33}$$

$$= a_{\boldsymbol{l},\sigma} e^{-j\omega_{\boldsymbol{l}} t}, \tag{A.34}$$

where $a_{\boldsymbol{l},\sigma}(t)$ is a dimensionless complex-valued mode function. By taking the appropriate derivatives of the vector potential, we can compute the complex electric and magnetic fields:

$$\boldsymbol{E}(\boldsymbol{r}, t) = \sum_{\boldsymbol{l},\sigma} j \sqrt{\frac{\hbar \omega_{\boldsymbol{l}}}{2\epsilon_0 L^3}} \left(a_{\boldsymbol{l},\sigma} e^{-j(\omega_{\boldsymbol{l}} t - \boldsymbol{k}_{\boldsymbol{l}} \cdot \boldsymbol{r})} - a_{\boldsymbol{l},\sigma}^* e^{j(\omega_{\boldsymbol{l}} t - \boldsymbol{k}_{\boldsymbol{l}} \cdot \boldsymbol{r})} \right) \boldsymbol{e}_{\boldsymbol{l},\sigma} \tag{A.35}$$

$$\boldsymbol{H}(\boldsymbol{r}, t) = \sum_{\boldsymbol{l},\sigma} jc \sqrt{\frac{\hbar}{2\omega_{\boldsymbol{l}} \mu_0 L^3}} \left(a_{\boldsymbol{l},\sigma} e^{-j(\omega_{\boldsymbol{l}} t - \boldsymbol{k}_{\boldsymbol{l}} \cdot \boldsymbol{r})} - a_{\boldsymbol{l},\sigma}^* e^{j(\omega_{\boldsymbol{l}} t - \boldsymbol{k}_{\boldsymbol{l}} \cdot \boldsymbol{r})} \right) \boldsymbol{k}_{\boldsymbol{l}} \times \boldsymbol{e}_{\boldsymbol{l},\sigma}. \tag{A.36}$$

The stored energy in the EM field in the cavity is given by

$$H = \int_V \left(\frac{1}{2}\epsilon_0 \boldsymbol{E}\cdot\boldsymbol{E} + \frac{1}{2}\mu_0 \boldsymbol{H}\cdot\boldsymbol{H}\right) dv, \quad \text{which simplifies to} \tag{A.37}$$

$$= \sum_{\boldsymbol{l},\sigma} \hbar\omega_{\boldsymbol{l}} (a^*_{\boldsymbol{l},\sigma} a_{\boldsymbol{l},\sigma}). \tag{A.38}$$

Note that the total energy is time independent as $a^*_{\boldsymbol{l},\sigma}(t)a_{\boldsymbol{l},\sigma}(t)$ is phase-insensitive. The radiation field in Eqs. (A.35) and (A.36) is quantized by associating operators $\hat{a}_{\boldsymbol{l},\sigma}(t)$ with normalized SHO mode function $a_{\boldsymbol{l},\sigma}(t)$, whose real and imaginary parts are the normalized canonical position and momentum operators, i.e.,

$$\hat{a}_{\boldsymbol{l},\sigma}(t) = \hat{a}_{1\boldsymbol{l},\sigma}(t) + j\hat{a}_{2\boldsymbol{l},\sigma}(t), \tag{A.39}$$

where the quadrature operators of the same spatial mode must satisfy the canonical commutation relation $[\hat{a}_{1\boldsymbol{l},\sigma}, \hat{a}_{2\boldsymbol{l},\sigma}] = j/2$. The field operator and its complex conjugate for a pair of spatial modes must thus satisfy the commutation relation

$$\left[\hat{a}_{\boldsymbol{l},\sigma}(t), \hat{a}^\dagger_{\boldsymbol{l}',\sigma'}(t)\right] = \delta_{\boldsymbol{l},\boldsymbol{l}'}\delta_{\sigma,\sigma'}. \tag{A.40}$$

The quantized field operators and the Hamiltonian (the total energy operator) are thus given by

$$\hat{\boldsymbol{E}}(\boldsymbol{r},t) = \sum_{\boldsymbol{l},\sigma} j\sqrt{\frac{\hbar\omega_{\boldsymbol{l}}}{2\epsilon_0 L^3}} \left(\hat{a}_{\boldsymbol{l},\sigma} e^{-j(\omega_l t - \boldsymbol{k}_l\cdot\boldsymbol{r})} - \hat{a}^\dagger_{\boldsymbol{l},\sigma} e^{j(\omega_l t - \boldsymbol{k}_l\cdot\boldsymbol{r})}\right) \boldsymbol{e}_{\boldsymbol{l},\sigma} \tag{A.41}$$

$$\hat{\boldsymbol{H}}(\boldsymbol{r},t) = \sum_{\boldsymbol{l},\sigma} jc\sqrt{\frac{\hbar}{2\omega_{\boldsymbol{l}}\mu_0 L^3}} \left(\hat{a}_{\boldsymbol{l},\sigma} e^{-j(\omega_l t - \boldsymbol{k}_l\cdot\boldsymbol{r})} - \hat{a}^\dagger_{\boldsymbol{l},\sigma} e^{j(\omega_l t - \boldsymbol{k}_l\cdot\boldsymbol{r})}\right) \boldsymbol{k}_{\boldsymbol{l}} \times \boldsymbol{e}_{\boldsymbol{l},\sigma}, \tag{A.42}$$

$$\hat{H} = \sum_{\boldsymbol{l},\sigma} \frac{\hbar\omega_{\boldsymbol{l}}}{2} \left[\hat{a}_{\boldsymbol{l},\sigma}\hat{a}^\dagger_{\boldsymbol{l},\sigma} + \hat{a}^\dagger_{\boldsymbol{l},\sigma}\hat{a}_{\boldsymbol{l},\sigma}\right] \tag{A.43}$$

$$= \sum_{\boldsymbol{l},\sigma} \hbar\omega_{\boldsymbol{l}} \left[\hat{a}^\dagger_{\boldsymbol{l},\sigma}\hat{a}_{\boldsymbol{l},\sigma} + \frac{1}{2}\right] \tag{A.44}$$

$$= \sum_{\boldsymbol{l},\sigma} \hbar\omega_{\boldsymbol{l}} \left[\hat{N}_{\boldsymbol{l},\sigma} + \frac{1}{2}\right], \tag{A.45}$$

where $\hat{N}_{\boldsymbol{l},\sigma} \triangleq \hat{a}^{\dagger}_{\boldsymbol{l},\sigma}\hat{a}_{\boldsymbol{l},\sigma}$ is the photon number operator for the mode indexed by $(\boldsymbol{l},\sigma)$. It is evident that from Eqs. (A.41) and (A.42) that the electric and magnetic field operators can be written as the sum of a positive-frequency component and a complex-conjugate negative-frequency component, i.e.,

$$\hat{\boldsymbol{E}}(\boldsymbol{r},t) = \hat{\boldsymbol{E}}^{(+)}(\boldsymbol{r},t) + \hat{\boldsymbol{E}}^{(-)}(\boldsymbol{r},t), \tag{A.46}$$

$$\hat{\boldsymbol{H}}(\boldsymbol{r},t) = \hat{\boldsymbol{H}}^{(+)}(\boldsymbol{r},t) + \hat{\boldsymbol{H}}^{(-)}(\boldsymbol{r},t), \tag{A.47}$$

where $\hat{\boldsymbol{E}}^{(-)}(\boldsymbol{r},t) = \hat{\boldsymbol{E}}^{(+)\dagger}(\boldsymbol{r},t)$ and $\hat{\boldsymbol{H}}^{(-)}(\boldsymbol{r},t) = \hat{\boldsymbol{H}}^{(+)\dagger}(\boldsymbol{r},t)$.

A.3.1 Semiclassical vs. quantum theory of photodetection: coherent states

Let us assume that only one polarization is excited, the only excited modes are $+z$ going plane waves with wave-number $\omega_l/c = k_l = (2\pi l)/L$; $l \in \{1,2,\ldots\}$, i.e. $l_x = l_y = 0, l_z = l$, impinging on an ideal photodetector. Also assume that the only modes excited lie within a frequency band $\omega_0 \pm \Delta\omega$, with $\Delta\omega \ll \omega$. Further assuming that we only look at the electric field in the time window $t_0 \leq t \leq t_0 + T$ where $T = L/c$, and normalizing the field operator to $\sqrt{\text{photons/sec}}$ units by integrating the field over the photosensitive surface of the photodetector, we have for the positive-frequency field operator

$$\hat{E}^{(+)}(t) = \frac{1}{\sqrt{T}} \sum_{l=-\infty}^{\infty} \hat{a}_l e^{-j2\pi lt/T}, \quad \text{for } t_0 \leq t \leq t_0 + T, \tag{A.48}$$

where $[\hat{a}_n, \hat{a}^{\dagger}_m] = \delta_{nm}$. Semiclassical theory predicts the photocurrent $i(t)$ to be an inhomogeneous Poisson impulse train with rate function $q|E(t)|^2$, given that the detector is illuminated by a deterministic classical field $E(t)$. The noise inherent to this Poisson process is what defines the shot-noise limit of semiclassical photodetection. Quantum theory of photodetection, on the other hand, predicts the photocurrent produced by the ideal photodetector to be a stochastic process whose statistics are those of the Hermitian photocurrent operator $\hat{i}(t) = q\hat{E}^{(+)\dagger}(t)\hat{E}^{(+)}(t)$. Just like the

measurement of any other dynamical observable in the framework of quantum mechanics, the photocurrent statistics are governed by the *quantum state* of the field. Non-classical states of the field such as photon number states, quadrature squeezed states, etc., do not obey the photocurrent statistics predicted by the semiclassical theory. We define *classical states* of the field to be those whose photocurrent measurement statistics predicted by the quantum theory comply with what is predicted by the semiclassical theory. Such states are known to be *coherent states*, and are eigenstates of the positive-field operator $\hat{E}^{(+)}(t)$ indexed by the complex amplitude of the field $E^{(+)}(t)$. The general multi-mode coherent state of the field $\hat{\boldsymbol{E}}^{(+)}(\boldsymbol{r},t)$ is given by

$$|\boldsymbol{\alpha}\rangle = \bigotimes_{\boldsymbol{l},\sigma} |\alpha_{\boldsymbol{l},\sigma}\rangle_{\boldsymbol{l},\sigma}, \tag{A.49}$$

$$\triangleq |\boldsymbol{E}^{(+)}(\boldsymbol{r},t)\rangle. \tag{A.50}$$

where $\hat{a}_{\boldsymbol{l},\sigma}|\alpha_{\boldsymbol{l},\sigma}\rangle_{\boldsymbol{l},\sigma} = \alpha_{\boldsymbol{l},\sigma}|\alpha_{\boldsymbol{l},\sigma}\rangle_{\boldsymbol{l},\sigma}$ is satisfied for each mode $(\boldsymbol{l},\sigma)$. It is easily verified that the multi-mode coherent state is an eigenstate of

$$\hat{\boldsymbol{E}}^{(+)}(\boldsymbol{r},t) = \sum_{\boldsymbol{l},\sigma} j\sqrt{\frac{\hbar\omega_{\boldsymbol{l}}}{2\epsilon_0 L^3}}\left(\hat{a}_{\boldsymbol{l},\sigma}e^{-j(\omega_l t - \boldsymbol{k}_{\boldsymbol{l}}\cdot\boldsymbol{r})}\right)\boldsymbol{e}_{\boldsymbol{l},\sigma},$$

i.e.,

$$\hat{\boldsymbol{E}}^{(+)}(\boldsymbol{r},t)|\boldsymbol{E}^{(+)}(\boldsymbol{r},t)\rangle = \boldsymbol{E}^{(+)}(\boldsymbol{r},t)|\boldsymbol{E}^{(+)}(\boldsymbol{r},t)\rangle, \tag{A.51}$$

with eigenfunction $\boldsymbol{E}^{(+)}(\boldsymbol{r},t) = \sum_{\boldsymbol{l},\sigma} j\sqrt{\frac{\hbar\omega_{\boldsymbol{l}}}{2\epsilon_0 L^3}}\left(\alpha_{\boldsymbol{l},\sigma}e^{-j(\omega_l t - \boldsymbol{k}_{\boldsymbol{l}}\cdot\boldsymbol{r})}\right)\boldsymbol{e}_{\boldsymbol{l},\sigma}$.

A.3.2 Photon-number (Fock) states

Photon-number states (or Fock states) are states of the quantized field that have a fixed number of photons in each mode, i.e. the measurement statistics of an ideal photodetector on a Fock state is deterministic. A multi-mode Fock state is given by the tensor product

$$|\boldsymbol{n}\rangle = \bigotimes_{\boldsymbol{l},\sigma} |n_{\boldsymbol{l},\sigma}\rangle_{\boldsymbol{l},\sigma}, \tag{A.52}$$

in which each single-mode Fock state $|n_{\boldsymbol{l},\sigma}\rangle_{\boldsymbol{l},\sigma}$ is the eigenstate of the corresponding mode's photon number operator $\hat{N}_{\boldsymbol{l},\sigma} = \hat{a}^\dagger_{\boldsymbol{l},\sigma}\hat{a}_{\boldsymbol{l},\sigma}$, i.e.,

$$\hat{N}_{\boldsymbol{l},\sigma}|n_{\boldsymbol{l},\sigma}\rangle_{\boldsymbol{l},\sigma} = n_{\boldsymbol{l},\sigma}|n_{\boldsymbol{l},\sigma}\rangle_{\boldsymbol{l},\sigma}, \tag{A.53}$$

for $n_{\boldsymbol{l},\sigma} \in \{0, 1, 2, \ldots\}$.

A.3.3 Single-mode states and characteristic functions

In all that follows, we shall drop the mode-index subscripts $(\boldsymbol{l}, \sigma)$ and will refer only to a single mode of the bosonic field, unless noted otherwise. A single mode, as we have seen, is characterized by the non-Hermitian operator $\hat{a}$, whose eigenstates $|\alpha\rangle$, $\alpha \in \mathbb{C}$ are classical states, i.e., they yield Poisson statistics for an ideal photon-counting measurement. The photon number operator $\hat{N} = \hat{a}^\dagger\hat{a}$ is a Hermitian operator whose measurement counts the number of photons in the mode. Its eigenstates $|n\rangle$, $n \in \{0, 1, \ldots\}$ are called Fock states or photon-number states, and they are non-classical states. It can be easily verified that the field operator $\hat{a}$ takes a Fock state $|n\rangle$ to a Fock state with one less number of photons, $|n-1\rangle$, and the conjugate operator $\hat{a}^\dagger$ takes a Fock state $|n\rangle$ to another Fock state with one additional number of photons $|n+1\rangle$, i.e.

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle \tag{A.54}$$

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle. \tag{A.55}$$

Because of the above property, we shall call the operator $\hat{a}$ the *annihilation operator* and $\hat{a}^\dagger$ the *creation operator* of the mode. They are sometimes also known as *ladder operators*. The Fock states form a complete orthonormal (CON) basis for all states of a single-mode bosonic field, viz., $\langle m|n\rangle = \delta_{mn}$ and $\hat{I} = \sum_n |n\rangle\langle n|$, for $\hat{I}$ the identity operator. Therefore, coherent states can be expanded in the Fock basis. Not surprisingly, we obtain

$$|\alpha\rangle = \sum_{n=0}^{\infty} \frac{e^{-|\alpha|^2/2}\alpha^n}{\sqrt{n!}}|n\rangle, \tag{A.56}$$

confirming the fact that the probability of counting m photons when a single-mode coherent state is subject to ideal photon counting measurement is given by the Poisson formula $p(m) = e^{-|\alpha|^2}|\alpha|^{2m}/m!$. The *displacement operator* is defined as

$$\hat{D}(\alpha) \equiv \exp(\alpha\hat{a}^\dagger - \alpha^*\hat{a}). \tag{A.57}$$

It displaces the vacuum state to a coherent state, $\hat{D}(\alpha)|0\rangle = |\alpha\rangle$. Coherent states do not form an orthonormal set, unlike number states. The inner product of two coherent states is given by

$$\langle\alpha|\beta\rangle = \exp\left[\alpha^*\beta - \frac{1}{2}(|\alpha|^2 + |\beta|^2)\right], \tag{A.58}$$

and the squared magnitude of the inner product is given by $|\langle\alpha|\beta\rangle|^2 = e^{-|\alpha-\beta|^2}$, so that $|\alpha\rangle$ and $|\beta\rangle$ are nearly orthogonal when $|\alpha - \beta| \gg 1$. The coherent states form an overcomplete basis of the single-mode state space, i.e., they resolve the identity via

$$\hat{I} = \int |\alpha\rangle\langle\alpha|\frac{\mathrm{d}^2\alpha}{\pi} = \sum_{n=0}^{\infty} |n\rangle\langle n|. \tag{A.59}$$

The thermal state of a mode with annihilation operator $\hat{a}$ is an isotropic Gaussian mixture of coherent states, i.e.,

$$\hat{\rho}_T = \int \frac{e^{-|\alpha|^2/N}}{\pi N} |\alpha\rangle\langle\alpha|\mathrm{d}^2\alpha, \tag{A.60}$$

where $N = \langle\hat{N}\rangle$ is the average photon number in the state $\hat{\rho}_T$. The thermal state can also be equivalently expressed as a statistical mixture of Fock states with a Bose-Einstein distribution, i.e.,

$$\hat{\rho}_T = \sum_{n=0}^{\infty} \frac{N^n}{(N+1)^{n+1}} |n\rangle\langle n|. \tag{A.61}$$

From Eq. (A.61) we immediately have that the von Neumann entropy of the thermal state $S(\hat{\rho}_T) = g(N) \triangleq (1+N)\ln(1+N) - N\ln N$, because the photon-number states

are orthonormal.

We define three kinds of characteristic functions for a single-mode state $\hat{\rho}$:

1. Normally ordered: $\chi_N^{\hat{\rho}}(\zeta) = \mathrm{Tr}(\hat{\rho}e^{\zeta\hat{a}^\dagger}e^{-\zeta^*\hat{a}}) = e^{|\zeta|^2/2}\langle\hat{D}(\zeta)\rangle$,

2. Anti-normally ordered: $\chi_A^{\hat{\rho}}(\zeta) = \mathrm{Tr}(\hat{\rho}e^{-\zeta^*\hat{a}}e^{\zeta\hat{a}^\dagger}) = e^{-|\zeta|^2/2}\langle\hat{D}(\zeta)\rangle$,

3. Wigner: $\chi_W^{\hat{\rho}}(\zeta) = \mathrm{Tr}(\hat{\rho}e^{-\zeta^*\hat{a}+\zeta\hat{a}^\dagger}) = \langle\hat{D}(\zeta)\rangle$.

As is evident from the definitions above, if one of the characteristic functions is known, the others can be computed easily. As examples, the antinormally-ordered characteristic function for a coherent state $|\alpha\rangle$ is $e^{\zeta\alpha^*-\zeta^*\alpha-|\zeta|^2}$, for the thermal state with mean photon number N it is, $e^{-(1+N)|\zeta|^2}$ and for the vacuum state it is $e^{-|\zeta|^2}$. The Husimi function $Q_{\hat{\rho}}(\alpha) = \langle\alpha|\hat{\rho}|\alpha\rangle/\pi$ is a proper probability distribution over the complex plane $\alpha \in \mathbb{C}$ and is the 2D Fourier transform of the antinormally ordered characteristic function $\chi_A^{\hat{\rho}}(\zeta)$, i.e.,

$$\chi_A^{\hat{\rho}}(\zeta) = \int Q_{\hat{\rho}}(\alpha)e^{\zeta\alpha^*-\zeta^*\alpha}\mathrm{d}^2\alpha \tag{A.62}$$

$$Q_{\hat{\rho}}(\alpha) = \frac{1}{\pi^2}\int \chi_A^{\hat{\rho}}(\zeta)e^{-\zeta\alpha^*+\zeta^*\alpha}\mathrm{d}^2\zeta. \tag{A.63}$$

The state $\hat{\rho}$ can be retrieved from $\chi_A^{\hat{\rho}}(\zeta)$ as follows

$$\hat{\rho} = \int \chi_A^{\hat{\rho}}(\zeta)e^{-\zeta\hat{a}^\dagger}e^{\zeta^*\hat{a}}\frac{\mathrm{d}^2\zeta}{\pi}. \tag{A.64}$$

A.3.4 Coherent detection

Besides the photon counting measurement of an optical field that we described above, the most commonly used optical detection schemes are the coherent-detection techniques, known as homodyne and heterodyne detection.

1. Homodyne detection — Homodyne detection is used to measure a single quadrature of the field. The measurement corresponds to measuring the Hermitian quadrature operator $\Re(\hat{a}e^{-j\theta})$. The actual realization of a homodyne detector

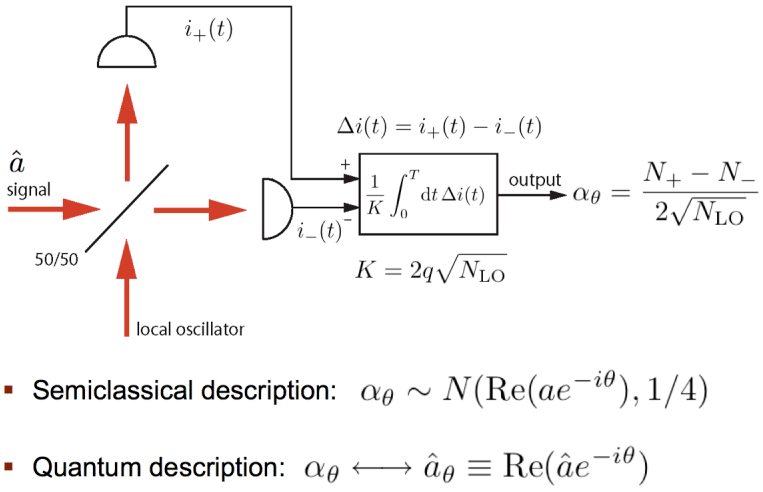
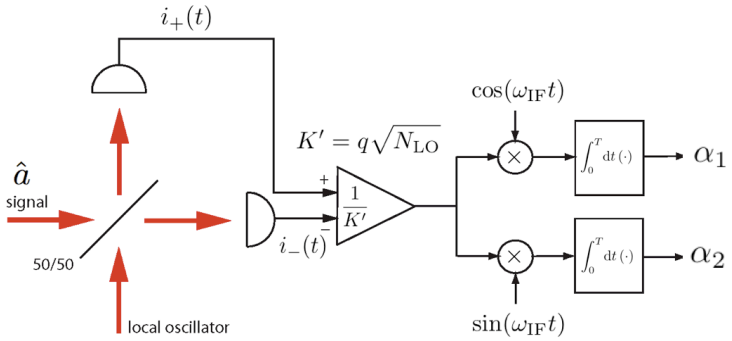


Figure A-1: Balanced homodyne detection. Homodyne detection is used to measure one quadrature of the field. The signal field $\hat{a}$ is mixed on a 50-50 beam splitter with a local oscillator excited in a strong coherent state with phase θ, that has the same frequency as the signal. The outputs beams are incident on a pair of photodiodes whose photocurrent outputs are passed through a differential amplifier and a matched filter to produce the classical output α_θ. If the input $\hat{a}$ is in a coherent state $|\alpha\rangle$, then the output of homodyne detection is predicted correctly by both the semiclassical and the quantum theories, i.e., a Gaussian-distributed real number α_θ with mean $\alpha\cos\theta$ and variance 1/4. If the input state is not a classical (coherent) state, then the quantum theory must be used to correctly account for the statistics of the outcome, which is given by the measurement of the quadrature operator $\Re(\hat{a}e^{-j\theta})$.

is depicted in Fig. A-1. If the input $\hat{a}$ is in a coherent state $|\alpha\rangle$, then the output of homodyne detection is a Gaussian distributed real number α_θ with mean $\alpha\cos\theta$ and variance 1/4. If the local oscillator phase $\theta = 0$, homodyne detection measures $\hat{a}_1$, the real quadrature of the field. If the detected state is a Gaussian state (see next section), then the outcome of homodyne measurement is a real Gaussian random variable with mean $\langle\hat{a}_1\rangle$ and variance $\langle\Delta\hat{a}_1^2\rangle = \langle(\hat{a}_1 - \langle\hat{a}_1\rangle)^2\rangle$.

2. Heterodyne detection — Heterodyne detection is used to measure both quadratures of the bosonic field simultaneously. For a general input state $\hat{\rho}$, the outcome of heterodyne measurement (α_1, α_2) has a probability distribution given by the Husimi function of $\hat{\rho}$ given by $Q_{\hat{\rho}}(\alpha) = \langle\alpha|\hat{\rho}|\alpha\rangle/\pi$. If the input is a coherent state $|\alpha\rangle$, then the outcome of measurement is a pair of real variance-1/2 Gaussian random variables with means $(\Re(\alpha), \Im(\alpha))$.



- Semiclassical description: $\{\alpha_1, \alpha_2\}$ SI, $\alpha_i \sim N(a_i, 1/2)$
- Quantum description: $\alpha \longleftrightarrow \hat{a}$

Figure A-2: Balanced heterodyne detection. Heterodyne detection is used to measure both quadratures of the field simultaneously. The signal field $\hat{a}$ is mixed on a 50-50 beam splitter with a local oscillator excited in a strong coherent state with phase $\theta = 0$, whose frequency is offset by an intermediate (radio) frequency, $\omega_{\rm IF}$, from that of the signal. The outputs beams are incident on a pair of photodiodes whose photocurrent outputs are passed through a differential amplifier. The output current of the differential amplifier is split into two paths and the two are multiplied by a pair of strong orthogonal intermediate-frequency oscillators followed by detection by a pair of matched filters, to yield two classical outcomes α_1 and α_2. If the input is a coherent state $|\alpha\rangle$, then both semiclassical and quantum theories predict the outputs (α_1, α_2) to be a pair of real variance-1/2 Gaussian random variables with means $(\Re(\alpha), \Im(\alpha))$. For a general input state $\hat{\rho}$, the outcome of heterodyne measurement (α_1, α_2) has a distribution given by the Husimi function of $\hat{\rho}$ given by $Q_{\hat{\rho}}(\alpha) = \langle\alpha|\hat{\rho}|\alpha\rangle/\pi$.

A.3.5 Gaussian states

For a single-mode state $\hat{\rho}$, let us define the mean field $\langle \hat{a} \rangle = \mathrm{Tr}(\hat{\rho}\hat{a})$ and the covariance matrix,

$$K \triangleq \begin{pmatrix} \langle \Delta\hat{a}\Delta\hat{a}^\dagger \rangle & \langle \Delta\hat{a}^2 \rangle \\ \langle \Delta\hat{a}^{\dagger 2} \rangle & \langle \Delta\hat{a}^\dagger \Delta\hat{a} \rangle \end{pmatrix} \tag{A.65}$$

where $\Delta\hat{a} \equiv \hat{a} - \langle \hat{a} \rangle$. The commutation relation $[\hat{a}, \hat{a}^\dagger] = 1$ implies that $\langle \Delta\hat{a}\Delta\hat{a}^\dagger \rangle = 1 + \langle \Delta\hat{a}^\dagger\Delta\hat{a} \rangle$. Also, the off-diagonal terms are complex conjugates of each other, i.e., $\langle \Delta\hat{a}^{\dagger 2} \rangle = \langle \Delta\hat{a}^2 \rangle^*$. Thus, the covariance matrix takes a form,

$$K = \begin{pmatrix} 1+N & P \\ P^* & N \end{pmatrix}. \tag{A.66}$$

For a zero mean field ($\langle \hat{a} \rangle = 0$) state, $\langle \Delta\hat{a}^\dagger\Delta\hat{a} \rangle = \langle \hat{a}^\dagger\hat{a} \rangle$ is the mean photon number in the state. Also, for states with $\langle \hat{a} \rangle = 0$, the correlation matrix

$$R \triangleq \begin{pmatrix} \langle \hat{a}\hat{a}^\dagger \rangle & \langle \hat{a}^2 \rangle \\ \langle \hat{a}^{\dagger 2} \rangle & \langle \hat{a}^\dagger\hat{a} \rangle \end{pmatrix} \tag{A.67}$$

is identical to the covariance matrix K defined in Eq. (A.65). The symmetrized covariance matrix is defined as $K_S = K - Q/2$, where

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A.68}$$

The Wigner covariance matrix (or the quadrature covariance matrix) is another equivalent form of the covariance matrix of $\hat{\rho}$ and is given by

$$K_Q \triangleq \begin{pmatrix} \langle \Delta\hat{a}_1^2 \rangle & \frac{1}{2}\langle \Delta\hat{a}_1\Delta\hat{a}_2 + \Delta\hat{a}_2\Delta\hat{a}_1 \rangle \\ \frac{1}{2}\langle \Delta\hat{a}_1\Delta\hat{a}_2 + \Delta\hat{a}_2\Delta\hat{a}_1 \rangle & \langle \Delta\hat{a}_2^2 \rangle \end{pmatrix} = \begin{pmatrix} V_1 & V_{12} \\ V_{12} & V_2 \end{pmatrix}, \tag{A.69}$$

where $\hat{a} = \hat{a}_1 + j\hat{a}_2$, $\Delta\hat{a}_1 \equiv \hat{a}_1 - \langle\hat{a}_1\rangle$ and $\Delta\hat{a}_2 \equiv \hat{a}_2 - \langle\hat{a}_2\rangle$. The relationship between these different forms of the covariance matrix is given by

$$UK_QU^\dagger = K_S, \tag{A.70}$$

where

$$U = \begin{pmatrix} 1 & j \\ 1 & -j \end{pmatrix}, \tag{A.71}$$

satisfies $U^\dagger U = 2I$, so that it is a scaled unitary matrix. The relationship between the elements of K_Q and K work out to be $N + 1/2 = V_1 + V_2$ and $P = (V_1 - V_2) + 2jV_1V_2$.

One definition of a bosonic Gaussian state is a state $\hat{\rho}$ whose Wigner characteristic function $\chi_W^{\hat{\rho}}(\zeta) \equiv \mathrm{Tr}\left(\hat{\rho}e^{-\zeta^*\hat{a}+\zeta\hat{a}^\dagger}\right)$ is quadratic in (ζ, ζ^*). An equivalent definition of a Gaussian state is a state that is completely described by only the first and second moments of the field.

Theorem 1.1 — The Wigner characteristic function $\chi_W^{\hat{\rho}}(\zeta)$ of a single-mode Gaussian state $\hat{\rho}$ with complex mean $\langle\hat{a}\rangle = \alpha$ and covariance matrix (A.66), is given by

$$\chi_W^{\hat{\rho}}(\zeta) = \exp\left[(\alpha^*\zeta - \alpha\zeta^*) + \Re(P^*\zeta^2) - (N + \frac{1}{2})|\zeta|^2\right]. \tag{A.72}$$

Proof — Expressing the Wigner characteristic function $\chi_W^{\hat{\rho}}(\zeta) \equiv \mathrm{Tr}\left(\hat{\rho}e^{-\zeta^*\hat{a}+\zeta\hat{a}^\dagger}\right)$ in terms of the real and imaginary parts of $\zeta = \zeta_1 + j\zeta_2$, we have

$$\ln\left[\chi_W^{\hat{\rho}}(\zeta_1, \zeta_2)\right] = \ln\left[\langle\exp\left(-2j\zeta_1\hat{a}_2 + 2j\zeta_2\hat{a}_1\right)\rangle_{\hat{\rho}}\right]. \tag{A.73}$$

Note that $\chi_W^{\hat{\rho}_a}(0,0) = 1$. For a function $f(\zeta_1, \zeta_2)$, such that $f(0,0) = 1$, we have the

following Taylor series expansion for $\ln(f(\zeta_1,\zeta_2))$ around $(\zeta_1,\zeta_2)\equiv(0,0)$:

$$\begin{aligned}\ln(f(\zeta_1,\zeta_2)) &= \zeta_1 f_{\zeta_1}(0,0)+\zeta_2 f_{\zeta_2}(0,0)+\frac{1}{2!}\left[{\zeta_1}^2(f_{\zeta_1\zeta_1}(0,0)-f_{\zeta_1}(0,0)^2)\right.\\ &\quad +\zeta_1\zeta_2(f_{\zeta_1\zeta_2}(0,0)-2f_{\zeta_1}(0,0)f_{\zeta_2}(0,0)+f_{\zeta_2\zeta_1}(0,0))\\ &\quad \left.+{\zeta_2}^2(f_{\zeta_2\zeta_2}(0,0)-f_{\zeta_2}(0,0)^2)\right]\\ &\quad +\text{h.o.t.} \qquad (A.74)\end{aligned}$$

Let us assign $f(\zeta_1,\zeta_2)=\chi_W^{\hat{\rho}}(\zeta_1,\zeta_2)=\langle\exp\left(-2j\zeta_1\hat{a}_2+2j\zeta_2\hat{a}_1\right)\rangle$, where the expectation is taken in the state $\hat{\rho}$. As $\hat{\rho}$ is a Gaussian state, the Wigner characteristic function must be a quadratic in (ζ_1,ζ_2) by definition. Hence, the expansion in Eq. (A.74) is *exact* without the h.o.t. (higher order terms). The partial derivatives of $f(\zeta_1,\zeta_2)$ are given by:

$$f_{\zeta_1}(\zeta_1,\zeta_2) = \langle -2j\hat{a}_2 e^{-2j\zeta_1\hat{a}_2+2j\zeta_2\hat{a}_1}\rangle \qquad (A.75)$$

$$f_{\zeta_2}(\zeta_1,\zeta_2) = \langle 2j\hat{a}_1 e^{-2j\zeta_1\hat{a}_2+2j\zeta_2\hat{a}_1}\rangle \qquad (A.76)$$

$$f_{\zeta_1\zeta_1}(\zeta_1,\zeta_2) = \langle -4\hat{a}_2^2 e^{-2j\zeta_1\hat{a}_2+2j\zeta_2\hat{a}_1}\rangle \qquad (A.77)$$

$$f_{\zeta_2\zeta_2}(\zeta_1,\zeta_2) = \langle -4\hat{a}_1^2 e^{-2j\zeta_1\hat{a}_2+2j\zeta_2\hat{a}_1}\rangle \qquad (A.78)$$

$$f_{\zeta_1\zeta_2}(\zeta_1,\zeta_2) = \langle (-2j\hat{a}_2)(2j\hat{a}_1) e^{-2j\zeta_1\hat{a}_2+2j\zeta_2\hat{a}_1}\rangle \qquad (A.79)$$

$$f_{\zeta_2\zeta_1}(\zeta_1,\zeta_2) = \langle (2j\hat{a}_1)(-2j\hat{a}_2) e^{-2j\zeta_1\hat{a}_2+2j\zeta_2\hat{a}_1}\rangle \qquad (A.80)$$

Evaluating each partial derivative at $(0,0)$ and substituting in Eq. (A.74) we get

$$\begin{aligned}\ln(f(\zeta_1,\zeta_2)) &= -2j\zeta_1\langle\hat{a}_2\rangle+2j\zeta_2\langle\hat{a}_1\rangle+\frac{1}{2}\left[\zeta_1^2\left(-4\langle\hat{a}_2^2\rangle+4\langle\hat{a}_2\rangle^2\right)\right.\\ &\quad +\zeta_1\zeta_2\left(4\langle\hat{a}_2\hat{a}_1\rangle-8\langle\hat{a}_2\hat{a}_1\rangle+4\langle\hat{a}_1\hat{a}_2\rangle\right)\\ &\quad \left.+\zeta_2^2\left(-4\langle\hat{a}_1^2\rangle+4\langle\hat{a}_1\rangle^2\right)\right] \qquad (A.81)\\ &= (-2j\zeta_1\alpha_2+2j\zeta_2\alpha_1)+2\left(-\zeta_1^2\langle\Delta\hat{a}_2^2\rangle-\zeta_2^2\langle\Delta\hat{a}_1^2\rangle\right.\\ &\quad \left.+\zeta_1\zeta_2(\langle\hat{a}_2\hat{a}_1\rangle-2\langle\hat{a}_2\rangle\langle\hat{a}_1\rangle+\langle\hat{a}_1\hat{a}_2\rangle)\right), \qquad (A.82)\end{aligned}$$

where we used (α_1,α_2) to denote the real and the imaginary parts of α. We can express

$\chi^{\hat{\rho}}_W(\zeta_1,\zeta_2)$ in terms of the entries of the Wigner covariance matrix K_Q, by observing that $V_{12} = \frac{1}{2}\langle \Delta\hat{a}_1\Delta\hat{a}_2 + \Delta\hat{a}_2\Delta\hat{a}_1\rangle = (\langle \hat{a}_2\hat{a}_1\rangle - 2\langle\hat{a}_2\rangle\langle\hat{a}_1\rangle + \langle\hat{a}_1\hat{a}_2\rangle)/2$. Therefore,

$$\ln\left[\chi^{\hat{\rho}}_W(\zeta_1,\zeta_2)\right] = \left[(-2j\zeta_1\alpha_2 + 2j\zeta_2\alpha_1) - 2\left(\zeta_1^2V_2 + \zeta_2^2V_1 - 2\zeta_1\zeta_2V_{12}\right)\right], \tag{A.83}$$

which implies,

$$\chi^{\hat{\rho}}_W(\zeta_1,\zeta_2) = \exp\left[(-2j\zeta_1\alpha_2 + 2j\zeta_2\alpha_1) - 2\left(\zeta_1^2V_2 + \zeta_2^2V_1 - 2\zeta_1\zeta_2V_{12}\right)\right], \tag{A.84}$$

Substituting $\zeta_1 = (\zeta+\zeta^*)/2$, $\zeta_2 = (\zeta-\zeta^*)/2j$, $N + 1/2 = V_1 + V_2$ and $P = (V_1 - V_2) + 2jV_1V_2$, we can express $\chi^{\hat{\rho}}_W(\zeta)$ in terms of entries of the covariance matrix K as follows,

$$\chi^{\hat{\rho}}_W(\zeta) = \exp\left[(\alpha^*\zeta - \alpha\zeta^*) + \Re(P^*\zeta^2) - (N+\frac{1}{2})|\zeta|^2\right]. \tag{A.85}$$

Multi-mode Gaussian states and the symplectic diagonalization[6] — Let us introduce vector-valued annihilation operators by stacking the annihilation operators of N independent modes as follows,

$$\hat{\boldsymbol{a}} = [\hat{a}_1 \ldots \hat{a}_N]^T \tag{A.86}$$

is an $N \times 1$ column vector of annihilation operators. Similarly, the column vector of creation operators is denoted

$$\hat{\boldsymbol{a}}^\dagger = [\hat{a}_1^\dagger \ldots \hat{a}_N^\dagger]^T. \tag{A.87}$$

With no loss of generality let us initially restrict our attention to zero-mean Gaussian states of N modes, such that the state is completely characterized by the $2N \times 2N$ correlation matrix

$$\mathrm{R} = \left\langle \begin{bmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{a}}^\dagger \end{bmatrix} \begin{bmatrix} (\hat{\boldsymbol{a}}^\dagger)^T & \hat{\boldsymbol{a}}^T \end{bmatrix} \right\rangle = \begin{bmatrix} \langle \hat{\boldsymbol{a}}^\dagger\hat{\boldsymbol{a}}^T\rangle + \mathrm{I}_N & \langle \hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^T\rangle \\ \langle \hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^T\rangle^* & \langle \hat{\boldsymbol{a}}^\dagger\hat{\boldsymbol{a}}^T\rangle \end{bmatrix}, \tag{A.88}$$

[6] The author thanks his colleague Baris I. Erkmen for this section, which has been partly adapted from [12]

where I_N is an $N \times N$ identity matrix and $*$ refers to element-wise complex conjugation.

Theorem 1.2 — Let $\hat{\boldsymbol{a}} = [\hat{a}_1 \dots \hat{a}_N]^T$ be N modes of a field that are in a zero-mean Gaussian state with $2N \times 2N$ correlation matrix R, as given in (A.88). Then, there exists $\mathrm{S} \in \mathbb{C}^{2N \times 2N}$ and $\Lambda \in \mathbb{C}^{2N \times 2N}$, such that

$$\mathrm{R} = \mathrm{S}\Lambda\mathrm{S}^\dagger \,, \tag{A.89}$$

where $\mathrm{S}^\dagger\mathrm{Q}\mathrm{S} = \mathrm{S}\mathrm{Q}\mathrm{S}^\dagger = \mathrm{Q}$ and $\Lambda = \mathrm{diag}\{\lambda_1 + 1, \dots, \lambda_N + 1, \lambda_1, \dots, \lambda_N\}$, with

$$\mathrm{Q} = \begin{bmatrix} \mathrm{I}_N & 0 \\ 0 & -\mathrm{I}_N \end{bmatrix} \tag{A.90}$$

and $\lambda_1, \dots, \lambda_N \geq 0$.

Proof — We use Williamson's symplectic decomposition theorem on the symmetrized (real-valued) correlation matrix for the quadratures, $\hat{\boldsymbol{a}}_{\boldsymbol{1}} \equiv [\hat{\boldsymbol{a}} + \hat{\boldsymbol{a}}^\dagger]/2$ and $\hat{\boldsymbol{a}}_{\boldsymbol{2}} \equiv [\hat{\boldsymbol{a}} - \hat{\boldsymbol{a}}^\dagger]/2i$, of the annihilation operators [83]. Then the expressions in the theorem are obtained by transforming this quadrature correlation matrix decomposition into the annihilation operator correlation matrix via the transformation

$$U = \begin{bmatrix} \mathrm{I}_N & i\mathrm{I}_N \\ \mathrm{I}_N & -i\mathrm{I}_N \end{bmatrix} . \tag{A.91}$$

The strength of a symplectic decomposition is the expansion of $\hat{\boldsymbol{a}}$ into a new set of unsqueezed modes with average photon number $\lambda_n, n = 1, \dots, N$ per mode.

Corollary 1.3 — Let $\hat{\boldsymbol{a}} = [\hat{a}_1 \dots \hat{a}_N]^T$ be in an arbitrary N-mode Gaussian state with mean $\langle \boldsymbol{a} \rangle$ and covariance matrix R. Then $\hat{\boldsymbol{a}}$ can be obtained via a symplectic transformation on an N-mode field $\hat{\boldsymbol{d}}$ that is in a tensor product of N uncorrelated thermal (Gaussian) states.

Proof — Consider the following linear transformation on $\hat{\boldsymbol{a}}$:

$$\begin{bmatrix} \hat{\boldsymbol{d}} \\ \hat{\boldsymbol{d}}^\dagger \end{bmatrix} = \mathrm{S}^{-1} \begin{bmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{a}}^\dagger \end{bmatrix}, \tag{A.92}$$

where $S^{-1} = \mathrm{QS}^\dagger\mathrm{Q}$ is the inverse of the symplectic matrix that diagonalizes R. Utilizing the symplectic diagonalization of R, we find that

$$\mathrm{R_d} = \Lambda\,. \tag{A.93}$$

Consequently, $\hat{d}_n$ has average photon number $\langle \hat{d}_n^\dagger \hat{d}_n \rangle = \lambda_n$, for $n = 1, \ldots, N$, where $\lambda_n \geq 0$ are the symplectic eigenvalues of R found in Theorem 1.2. Furthermore, all modes $\{d_n\}$ are uncorrelated. Therefore, each mode can be represented as an isotropic mixture of coherent states displaced by the corresponding mean, and the joint state is the tensor product of N such states.

Corollary 1.4 — Let $\hat{\boldsymbol{d}} = [\hat{d}_1 \ldots \hat{d}_N]^T$ be N modes in an arbitrary state. A symplectic transformation on the N-modes, mapping $\hat{\boldsymbol{d}}$ into $\hat{\boldsymbol{a}}$ as

$$\begin{bmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{a}}^\dagger \end{bmatrix} = \mathrm{S} \begin{bmatrix} \hat{\boldsymbol{d}} \\ \hat{\boldsymbol{d}}^\dagger \end{bmatrix}, \tag{A.94}$$

does not alter the von-Neumann entropy of the state; i.e. if $\hat{\rho}_d$ and $\hat{\rho}_a$ denote input and output the density operators respectively, then $S(\hat{\rho}_d) = S(\hat{\rho}_a)$.

Proof — The symplectic transformation given in (A.94) is a canonical transformation, i.e., it preserves the commutation relations. Thus it can be implemented with a unitary operator $\hat{U}$, satisfying $\hat{U}\hat{U}^\dagger = \hat{U}^\dagger\hat{U} = \hat{I}$ [84]. The theorem and corollaries collectively show that an arbitrary N-mode Gaussian state can always be linearly transformed into a tensor product of N thermal states with no change in the entropy of the joint state.

As a simple example, using the symplectic diagonalization of a single-mode zero-mean Gaussian state $\hat{\rho}$ whose covariance matrix is given by Eq. (A.66), a unitary squeezing transformation exists that transforms $\hat{\rho}$ to a zero-mean thermal state $\hat{\rho}_{T,N}$,

i.e., $\hat{\rho} = U\hat{\rho}_{T,N}U^{\dagger}$ where $\hat{\rho}_{T,N}$ is a zero-mean thermal state with mean photon number $\bar{N} = \sqrt{(N+1/2)^2 - |P|^2} - 1/2$. Thus the von Neumann entropy of a Gaussian state whose covariance matrix is given by Eq. A.66, is given by $S(\hat{\rho}) = g(\bar{N})$.

Appendix B

Capacity region of a degraded quantum broadcast channel with M receivers

In this appendix, we generalize the capacity region of the two-receiver quantum degraded broadcast channel proved by Yard et. al.[52], to an arbitrary number of receivers. In chapter 3, we postponed the general proof of the capacity region to this appendix, but we used this result to evaluate the capacity region of the Bosonic broadcast channel with an arbitrary number of receivers. For the sake of completeness, and ease of reading, we restate the set-up of the problem and go through the notation before we do the proof.

B.1 The Channel Model

The M-receiver quantum broadcast channel $\mathcal{N}_{A-Y_0\ldots Y_{M-1}}$ is a quantum channel from a sender Alice (A) to M independent receivers $Y_0, \ldots, Y_{M-1}$. The quantum channel from A to Y_0 is obtained by tracing out all the other receivers from the channel map, i.e., $\mathcal{N}_{A-Y_0} \equiv \mathrm{Tr}_{Y_1,\ldots,Y_{M-1}}\left(\mathcal{N}_{A-Y_0\ldots Y_{M-1}}\right)$, with a similar definition for $\mathcal{N}_{A-Y_k}$ for $k \in \{1, \ldots, M-1\}$. We say that a broadcast channel $\mathcal{N}_{A-Y_0\ldots Y_{M-1}}$ is *degraded* if there exists a series of *degrading channels* $\mathcal{N}^{\mathrm{deg}}_{Y_k-Y_{k+1}}$ from Y_k to Y_{k+1}, for $k \in \{0, \ldots, M-2\}$,

satisfying

$$\mathcal{N}_{A-Y_{M-1}} = \mathcal{N}^{\text{deg}}_{Y_{M-2}-Y_{M-1}} \circ \mathcal{N}^{\text{deg}}_{Y_{M-3}-Y_{M-2}} \circ \ldots \circ \mathcal{N}^{\text{deg}}_{Y_0-Y_1} \circ \mathcal{N}_{A-Y_0}. \tag{B.1}$$

The M-receiver degraded broadcast channel (see Fig. B-1) describes a physical scenario in which for each successive n uses of the channel $\mathcal{N}_{A-Y_0\ldots Y_{M-1}}$ Alice communicates a randomly generated classical message $(m_0, \ldots, m_{M-1}) \in (W_0, \ldots, W_{M-1})$ to the receivers $Y_0, \ldots, Y_{M-1}$, where the message-sets W_k are sets of classical indices of sizes 2^{nR_k}, for $k \in \{0, \ldots, M-1\}$. The messages $(m_0, \ldots, m_{M-1})$ are assumed to be independent and uniformly distributed over $(W_0, \ldots, W_{M-1})$, i.e.

$$p_{W_0,\ldots,W_{M-1}}(m_0, \ldots, m_{M-1}) = \prod_{k=0}^{M-1} p_{W_k}(m_k) = \prod_{k=0}^{M-1} \frac{1}{2^{nR_k}} \tag{B.2}$$

Because of the degraded nature of the channel, given that the transmission rates are within the capacity region and proper encoding and decoding is employed at the transmitter and at the receivers, Y_0 can decode the entire message M-tuple $(m_0, \ldots, m_{M-1})$, Y_1 can decode the reduced message $(M-1)$-tuple $(m_1, \ldots, m_{M-1})$, and so on, until the noisiest receiver Y_{M-1} can only decode the single message-index m_{M-1}. To convey the message-set $\mathbf{m}_0^{M-1}$, Alice prepares n-channel use states that, after transmission through the channel, result in M-partite conditional density matrices $\left\{\hat{\rho}^{Y_0^n \ldots Y_{M-1}^n}_{\mathbf{m}_0^{M-1}}\right\}, \forall \mathbf{m}_0^{M-1} \in \mathbf{W}_0^{M-1}$. The quantum states received by a receiver, say Y_0 can be found by tracing out the other receivers, viz. $\hat{\rho}^{Y_0^n}_{\mathbf{m}_0^{M-1}} \equiv \text{Tr}_{Y_1^n,\ldots,Y_{M-1}^n}\left(\hat{\rho}^{Y_0^n \ldots Y_{M-1}^n}_{\mathbf{m}_0^{M-1}}\right)$, etc. Fig. B-2 illustrates this decoding process.

B.2 Capacity Region: Theorem

A $(2^{nR_0}, \ldots, 2^{nR_{M-1}}, n, \epsilon)$ code for this channel consists of an encoder

$$x^n : (\mathbf{W}_0^{M-1}) \to \mathcal{A}^n, \tag{B.3}$$

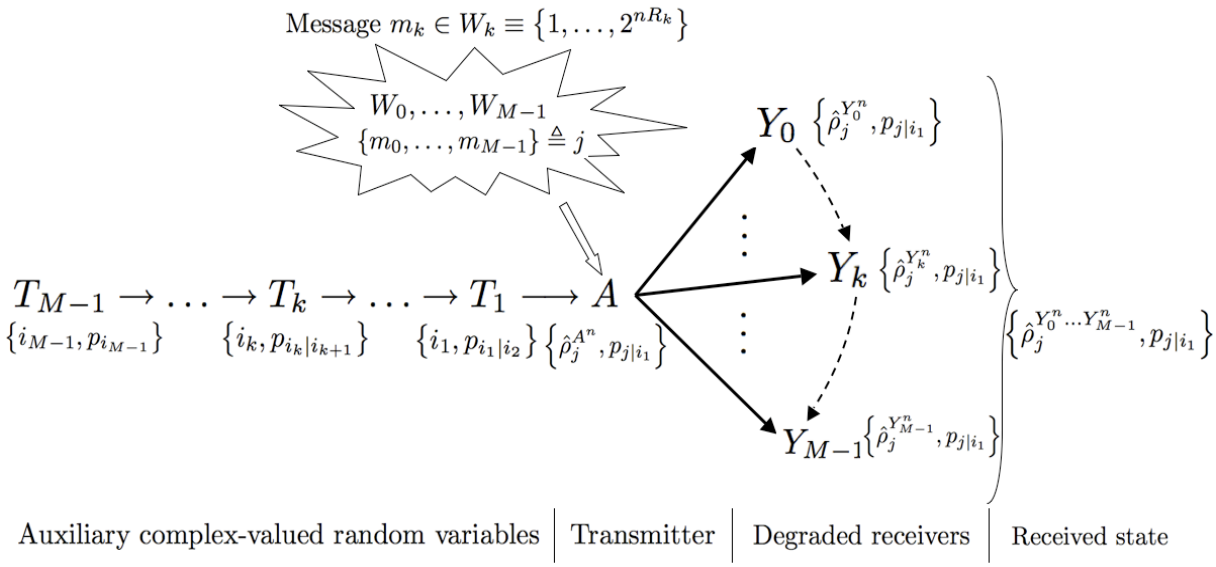


Figure B-1: This figure summarizes the setup of the transmitter and the channel model for the M-receiver quantum degraded broadcast channel. In each successive n uses of the channel, the transmitter A sends a randomly generated classical message $(m_0, \ldots, m_{M-1}) \in (W_0, \ldots, W_{M-1})$ to the M receivers $Y_0, \ldots, Y_{M-1}$, where the message-sets W_k are sets of classical indices of sizes 2^{nR_k}, for $k \in \{0, \ldots, M-1\}$. The dashed arrows indicate the direction of degradation, i.e. Y_0 is the least noisy receiver, and Y_{M-1} is the noisiest receiver. In this degraded channel model, the quantum state received at the receiver Y_k, $\hat{\rho}^{Y_k}$ can always be reconstructed from the quantum state received at the receiver $Y_{k'}$, $\hat{\rho}^{Y_{k'}}$, for $k' < k$, by passing $\hat{\rho}^{Y_{k'}}$ through a trace-preserving completely positive map (a quantum channel). For sending the classical message $(m_0, \ldots, m_{M-1}) \triangleq j$, Alice chooses a n-use state (codeword) $\hat{\rho}_j^{A^n}$ using a prior distribution $p_{j|i_1}$, where i_k denotes the complex values taken by an auxiliary random variable T_k. It can be shown that, in order to compute the capacity region of the quantum degraded broadcast channel, we need to choose $M-1$ complex valued auxiliary random variables with a Markov structure as shown above, i.e. $T_{M-1} \to T_{M-2} \to \ldots \to T_k \to \ldots \to T_1 \to A^n$ is a Markov chain.

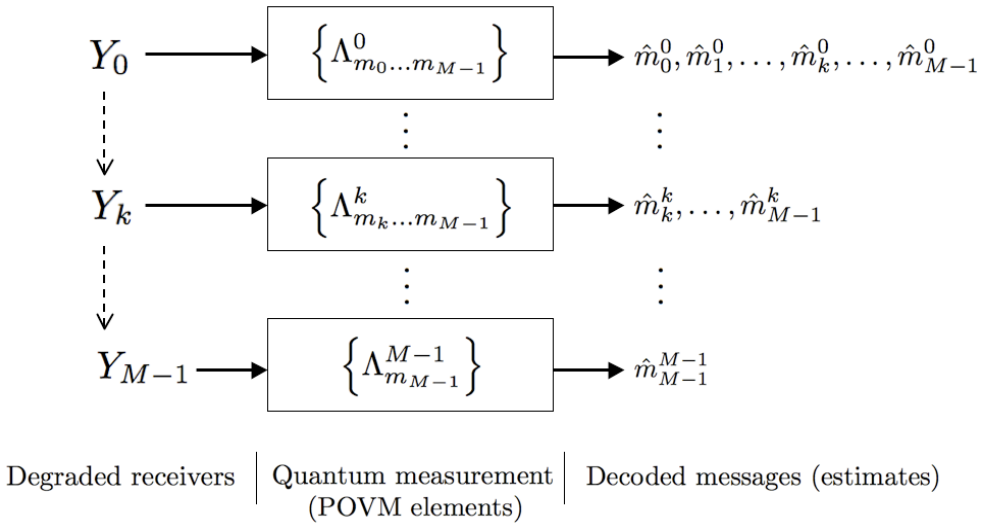


Figure B-2: This figure illustrates the decoding end of the M-receiver quantum degraded broadcast channel. The decoder consists of a set of measurement operators, described by positive operator-valued measures (POVMs) for each receiver; $\left\{\Lambda^0_{m_0 \ldots m_{M-1}}\right\}$, $\left\{\Lambda^1_{m_1 \ldots m_{M-1}}\right\}$, …, $\left\{\Lambda^{M-1}_{m_{M-1}}\right\}$ on ${\mathcal{Y}_0}^n$, ${\mathcal{Y}_1}^n$, …, ${\mathcal{Y}_{M-1}}^n$ respectively. Because of the degraded nature of the channel, if the transmission rates are within the capacity region and proper encoding and decoding are employed at the transmitter and at the receivers respectively, Y_0 can decode the entire message M-tuple to obtain estimates $(\hat{m}^0_0, \ldots, \hat{m}^0_{M-1})$, Y_1 can decode the reduced message $(M-1)$-tuple to obtain its own estimates $(\hat{m}^1_1, \ldots, \hat{m}^1_{M-1})$, and so on, until the noisiest receiver Y_{M-1} can only decode the single message-index m_{M-1} to obtain an estimate $\hat{m}^{M-1}_{M-1}$. Even though the less noisy receivers can decode the messages of the noisier receivers, the message m_k is intended to be sent to receiver Y_k, $\forall k$. Hence, when we say that a broadcast channel is operating at a rate $(R_0, \ldots, R_{M-1})$, we mean that the message m_k is reliably decoded by receiver Y_k at the rate R_k bits per channel use.

a set of positive operator-valued measures (POVMs) — $\left\{\Lambda^0_{m_0\ldots m_{M-1}}\right\}$, $\left\{\Lambda^1_{m_1\ldots m_{M-1}}\right\}$, …, $\left\{\Lambda^{M-1}_{m_{M-1}}\right\}$ on ${\mathcal{Y}_0}^n$, ${\mathcal{Y}_1}^n$, …, ${\mathcal{Y}_{M-1}}^n$ respectively, such that the mean probability of a collective correct decision satisfies

$$\mathrm{Tr}\left(\hat{\rho}^{Y_0^n\ldots Y_{M-1}^n}_{\mathbf{m}_0^{M-1}}\left(\bigotimes_{k=0}^{M-1}\Lambda^k_{m_k\ldots m_{M-1}}\right)\right)\geq 1-\epsilon, \tag{B.4}$$

for $\forall\mathbf{m}_0^{M-1}\in\mathbf{W}_0^{M-1}$. A rate M-tuple $(R_0,\ldots,R_{M-1})$ is *achievable* if there exists a sequence of $(2^{nR_0},\ldots,2^{nR_{M-1}},n,\epsilon)$ codes with $\epsilon_n\to 0$. The classical *capacity region* of the broadcast channel is defined as the convex hull of the closure of all achievable rate M-tuples $(R_0,\ldots,R_{M-1})$. The classical capacity region of the two-user degraded quantum broadcast channel with discrete alphabet was derived by Yard et. al. [52], and we used the infinite-dimensional extension of Yard et. al.'s capacity theorem to prove the capacity region of the Bosonic broadcast channel, subject to the minimum output entropy conjecture 2. The capacity region of the degraded quantum broadcast channel can easily be extended to the case of an arbitrary number M, of receivers. For notational similarity to the capacity region of the classical degraded broadcast channel, we state the capacity theorem first, using the shorthand notation for Holevo information we introduced in footnote 6 in chapter 3.

Theorem B.1 — The capacity region of the M-receiver degraded broadcast channel $\mathcal{N}_{A-Y_0\ldots Y_{M-1}}$ as defined in Eq. (B.1), is given by

$$\begin{aligned}
R_0 &\leq \frac{1}{n}I\left(A^n;Y_0^n|T_1\right),\\
R_k &\leq \frac{1}{n}I\left(T_k;Y_k^n|T_{k+1}\right)\quad\forall k\in\{1,\ldots,M-2\},\\
R_{M-1} &\leq \frac{1}{n}I\left(T_{M-1};Y_{M-1}^n\right),
\end{aligned} \tag{B.5}$$

where T_k, $k\in\{1,\ldots,M-1\}$ form a set of auxiliary complex valued random variables

such that $T_{M-1} \rightarrow T_{M-2} \rightarrow \ldots \rightarrow T_k \rightarrow \ldots \rightarrow T_1 \rightarrow A^n$ forms a Markov chain, i.e.

$$p_{T_{M-1},\ldots,T_1,A^n}(i_{M-1},\ldots,i_1,j) = p_{T_{M-1}}(i_{M-1}) \left(\prod_{k=M-1}^{2} p_{T_{k-1}|T_k}(i_{k-1}|i_k) \right) p_{A^n|T_1}(j|i_1), \tag{B.6}$$

where with a slight abuse of notation, we have used the symbols $T_1, \ldots, T_{M-1}$ to denote complex-valued classical random variables taking values $i_k \in \mathcal{T}_k$ where $\mathcal{T}_k$ denotes a complex alphabet, as well as to denote quantum systems, by associating a complete orthonormal set of pure quantum states with the complex probability densities $p_{T_k}(i_k)$ of these auxiliary random variables. With further abuse of notation, we have used A^n to denote a classical random variable. See footnote 5 in chapter 3.

In order to find the optimum capacity region, the above rate region must be optimized over the joint distribution $p_{T_{M-1},\ldots,T_1,A^n}(i_{M-1},\ldots,i_1,j)$. As Holevo information is not necessarily additive (unlike Shannon mutual information), the rate region must also be optimized over the codeword block-length n. The above Markov chain structure of the auxiliary random variables T_k, $k \in \{1,\ldots,M-1\}$ is shown to be optimal in the converse proof which proves the optimality of the above capacity region without assuming any special structure of the auxiliary random variables. Also, note the striking similarity of the expressions for the capacity region given above, with the capacity region of the classical M-receiver degraded broadcast channel, given in Eqs. (3.8). Holevo information takes place of Shannon mutual information in the quantum case, and because of superadditivity of Holevo information, an additional regularization over number of channel uses n, is required.

The capacity region above can be re-cast in the Holevo-information notation that we used earlier in this chapter for the two-receiver quantum broadcast channel. For the channel model of the multiple-user quantum degraded broadcast channel we described in the section above (pictorially depicted in Fig. B-1), our proposed capacity

region (in Eqs. (B.5)) can alternatively be expressed as

$$
\begin{aligned}
R_0 &\leq \frac{1}{n}\sum_{i_1} p_{T_1}(i_1)\chi\left(p_{A^n|T_1}(j|i_1), \hat{\rho}_j^{Y_0^n}\right) \\
&= \frac{1}{n}\sum_{i_1} p_{T_1}(i_1)\left[S\left(\sum_j p_{A^n|T_1}(j|i_1)\hat{\rho}_j^{Y_0^n}\right) - \sum_j p_{A^n|T_1}(j|i_1) S\left(\hat{\rho}_j^{Y_0^n}\right)\right], \\
R_k &\leq \frac{1}{n}\sum_{i_{k+1}} p_{T_{k+1}}(i_{k+1})\chi\left(p_{T_k|T_{k+1}}(i_k|i_{k+1}), \hat{\rho}_{i_k}^{Y_k^n}\right), \quad \forall k \in \{1, \ldots, M-2\}, \\
&= \frac{1}{n}\sum_{i_{k+1}} p_{T_{k+1}}(i_{k+1})\left[S\left(\sum_{i_k} p_{T_k|T_{k+1}}(i_k|i_{k+1})\hat{\rho}_{i_k}^{Y_k^n}\right) - \sum_{i_k} p_{T_k|T_{k+1}}(i_k|i_{k+1}) S\left(\hat{\rho}_{i_k}^{Y_k^n}\right)\right], \\
R_{M-1} &\leq \frac{1}{n}\chi\left(p_{T_{M-1}}(i_{M-1}), \hat{\rho}_{i_{M-1}}^{Y_{M-1}^n}\right) \\
&= \frac{1}{n} S\left(\sum_{i_{M-1}} p_{T_{M-1}}(i_{M-1})\hat{\rho}_{i_{M-1}}^{Y_{M-1}^n}\right) - \sum_{i_{M-1}} p_{T_{M-1}}(i_{M-1}) S\left(\hat{\rho}_{i_{M-1}}^{Y_{M-1}^n}\right). \qquad \text{(B.7)}
\end{aligned}
$$

Even though the capacity-region expressions above have been written for a discrete alphabet, it can be generalized to a continuous alphabet of quantum states over an infinite-dimensional Hilbert space, in which case the summations in Eqs. (B.7) are replaced by integrals (see footnote 17 in Chapter 3).

B.3 Capacity Region: Proof (Achievability)

Proof [Achievability ($M = 3$, single channel use)] — It is more instructive to do the "achievability" part of the proof first, for $M = 3$ receivers. The general proof for the M-receiver case is a logical extension of this proof. We need to prove achievability only for the single-channel-use rate region (i.e., for $n = 1$ in Eqs. (B.5)), because the same proof can be applied to multiple-use (larger) quantum systems of the transmitter and the receiver alphabets to obtain the general capacity region. For any $\epsilon, \delta > 0$, we

will show that for rate 3-tuples (R_0, R_1, R_2) satisfying[1]

$$\begin{aligned} I(A;Y_0|T_1) - \delta(1+2d_0) &\leq R_0 \leq I(A;Y_0|T_1) + 2\delta I_0, && \text{(B.8)} \\ I(T_1;Y_1|T_2) - \delta(1+d_1) &\leq R_1 \leq I(T_1;Y_1|T_2) + \delta I_1, \text{and} && \text{(B.9)} \\ 0 &\leq R_2 = I(T_2;Y_2) - \delta, && \text{(B.10)} \end{aligned}$$

for finite positive real numbers d_0, d_1, I_0, I_1, there exists an $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, O(\epsilon))$ code for the degraded broadcast channel $\mathcal{N}_{A-Y_0Y_1Y_2}$. Below is a brief heuristic of the proof, followed by the actual proof.

We will construct the required triply-indexed set of codewords $\left\{\hat{\rho}^{A^n}_{m_0,m_1,m_2}\right\}_{m_k \in 2^{nR_k}}$ as follows. First, we will select a rate R_2 code for the channel $\mathcal{N}_{T_2-Y_0Y_1Y_2}$ with codewords selected in an independent, identically distributed (i.i.d.) manner from the distribution $p_{T_2}(i_2)$, which conveys the message index $m_2 \in 2^{nR_2}$ to all the three receivers Y_0, Y_1 and Y_2. We call these codewords the "primary cloud-centers"[2]. Thereafter for each $i_2 \in \mathcal{T}_2$, we pick a code of rate R_{1,i_2} and blocklength approximately $p_{T_2}(i_2)n$ for the conditional channel $\mathcal{N}^{i_2}_{T_1-Y_0Y_1}$ with codewords selected i.i.d. according to $p_{T_1|T_2}(i_1|i_2)$. These codewords are called the "secondary cloud-centers". If the receiver Y_1 knows i_2, it can decode at rates approaching $R_{1,i_2} \approx I(T_1;Y_1|T_2 = i_2)$, such that the average rate $R_1 \approx \sum_{i_2} p_{T_2}(i_2) R_{1,i_2}$ is close to the desired rate at which Y_1 can decode the message index m_1. Y_0 can similarly learn m_1 reliably at rates approaching R_1. Then finally, for each $i_2 \in \mathcal{T}_2$, and $i_1 \in \mathcal{T}_1$, we pick a random HSW code of blocklength approximately $np_{T_2}(i_2)p_{T_1|T_2}(i_1|i_2)$ for the conditional channel $\mathcal{N}^{i_2,i_1}_{A-Y_0}$, with codewords selected i.i.d. according to $p_{A|T_1,T_2}(j|i_1,i_2)$. If the receiver Y_0 knows both i_2 and i_1 (for our case as $T_2 \to T_1 \to A$ is a Markov chain, Y_0 just needs to know i_1), it can decode at rates approaching $R_{0,i_2,i_1} \approx I(A;Y_0|T_1 = i_1)$, such that the average rate $R_0 \approx \sum_{i_2,i_1} p_{T_2}(i_2)p_{T_1|T_2}(i_1|i_2)R_{0,i_2,i_1}$ is close to the desired rate at which Y_0 can decode the private message index m_0.

[1] From now on, we will freely use both the $I(X;Y)$, and the more explicit $\chi(p_X(x), \hat{\rho}^Y_x)$ notations interchangeably, for Holevo quantities.

[2] To read more about layered-encoding techniques for the classical degraded broadcast channel, using "cloud-centers" and "clouds", see [3].

B.3.1 Constructing codebooks with the desired rate-bounds

Let us choose arbitrary $\delta, \epsilon > 0$. Pick

$$\begin{aligned} R_2 &= I(T_2; Y_2) - \delta && \text{(B.11)} \\ &= \chi\left(p_{T_2}(i_2), \hat{\rho}_{i_2}^{Y_2}\right) - \delta && \text{(B.12)} \\ &\leq \chi\left(p_{T_2}(i_2), \hat{\rho}_{i_2}^{Y_0 Y_1 Y_2}\right) - \delta. && \text{(B.13)} \end{aligned}$$

Because of the degraded nature of the channel,

$$I(T_2; Y_2) \leq I(T_2; Y_1) \leq I(T_2; Y_0) \leq I(T_2; Y_0, Y_1, Y_2), \tag{B.14}$$

where the last inequality follows from the fact that the point-to-point channel $\mathcal{N}_{T_2 - Y_0 Y_1 Y_2}$ between T_2 and the joint receiver (Y_0, Y_1, Y_2) can transmit information reliably at a rate as least as high as the capacity of the channel $\mathcal{N}_{T_2 - Y_k}$ between T_2 and one of the receivers Y_k alone. Hence by using HSW theorem for the channel $\mathcal{N}_{T_2 - Y_0 Y_1 Y_2}$, we obtain an (R_2, n, ϵ) code $\left\{\hat{\rho}_{m_2}^{T_2^n}, \Lambda_{m_2}^0, \Lambda_{m_2}^1, \Lambda_{m_2}^2\right\}$ with all codewords chosen i.i.d. from $p_{T_2}(i_2)$ and of type $P_2(i_2)$, satisfying $|P_2(\cdot) - p_{T_2}(\cdot)|_1 \leq \delta$, and for all $m_2 \in W_2$,

$$\mathrm{Tr}\left(\Lambda_{m_2}^0 \otimes \Lambda_{m_2}^1 \otimes \Lambda_{m_2}^2\right) \hat{\rho}_{m_2}^{Y_0^n Y_1^n Y_2^n} \geq 1 - \epsilon, \tag{B.15}$$

where

$$\hat{\rho}_{m_2}^{Y_0^n Y_1^n Y_2^n} = \bigotimes_{l=1}^{n} \hat{\rho}_{m_2}^{Y_{0,l}^n Y_{1,l}^n Y_{2,l}^n}$$

are product-state codewords, with $\hat{\rho}_{m_2}^{Y_{0,l}^n Y_{1,l}^n Y_{2,l}^n} = \left(\mathcal{N}_{T_2 - Y_0 Y_1 Y_2}\right)^{\otimes n}\left(\hat{\rho}_{m_2}^{T_{2,l}^n}\right)$, for $l \in \{1, \ldots, n\}$, being the l^{th} symbol of the received n-symbol codeword[3].

Let us define the cardinalities of the alphabets of T_2, T_1, and the transmitter A,

[3]Note that throughout this discussion, each codeword symbol is transmitted in a *single use* of the channel.

as $|\mathcal{T}_2| = d_2$, $|\mathcal{T}_1| = d_1$ and $|\mathcal{A}| = d_0$. For each i_2, define

$$\begin{aligned} R_{1,i_2} &\triangleq I(T_1; Y_1|T_2 = i_2) - \delta_{i_2} \leq d_1 && \text{(B.16)} \\ &\leq I(T_1; Y_0, Y_1|T_2 = i_2) - \delta_{i_2} \leq d_1 && \text{(B.17)} \\ &= \chi\left(p_{T_1|T_2}(i_1|i_2), \hat{\rho}_{i_1}^{Y_0Y_1}\right) - \delta_{i_2}, && \text{(B.18)} \end{aligned}$$

where

$$\delta_{i_2} \triangleq \delta P_2(i_2), \tag{B.19}$$

for $i_2 \in \{1, \ldots, d_2\}$. Define

$$\begin{aligned} \epsilon_{i_2} &\triangleq \epsilon P_2(i_2), \text{and} && \text{(B.20)} \\ n_{i_2} &\triangleq nP_2(i_2), && \text{(B.21)} \end{aligned}$$

for $i_2 \in \{1, \ldots, d_2\}$. For each $i_2 \in \mathcal{T}_2$, there exists an $(R_{1,i_2}, n_{i_2}, \epsilon_{i_2})$ random HSW code $\left\{\hat{\rho}_{m_{1,i_2}}^{T_1^{n_{i_2}}}, \Lambda_{m_{1,i_2}}^{0(i_2)}, \Lambda_{m_{1,i_2}}^{1(i_2)}\right\}$, $\forall m_{1,i_2} \in \left\{1, \ldots, 2^{nR_{1,i_2}}\right\}$, for the conditional channel $\mathcal{N}_{T_1-Y_0Y_1}^{i_2}$, which satisfies

$$E\left[2^{-n_{i_2}R_{1,i_2}} \sum_{m_{1,i_2}=1}^{2^{nR_{1,i_2}}} \mathrm{Tr}\left((\Lambda_{m_{1,i_2}}^{0(i_2)} \otimes \Lambda_{m_{1,i_2}}^{1(i_2)})\hat{\rho}_{m_{1,i_2}}^{Y_0^{n_{i_2}}Y_1^{n_{i_2}}}\right)\right] \geq 1 - \epsilon_{i_2}, \tag{B.22}$$

where each codeword $\hat{\rho}_{m_{1,i_2}}^{T_1^{n_{i_2}}}$ is chosen i.i.d. from $p_{T_1|T_2}(i_1|i_2)$ and each codeword is of the type $P_{1|2}(i_1|i_2)$, such that $|P_{1|2}(\cdot|i_2) - p_{T_1|T_2}(\cdot|i_2)|_1 \leq \delta_{i_2}$, and the expectation is over the randomness in the HSW codes. Note that owing to the symmetry of the random code construction, (B.22) may be equivalently expressed as

$$\mathrm{E}\left[\mathrm{Tr}\left((\Lambda_1^{0(i_2)} \otimes \Lambda_1^{1(i_2)})\hat{\rho}_1^{Y_0^{n_{i_2}}Y_1^{n_{i_2}}}\right)\right] \geq 1 - \epsilon_{i_2}. \tag{B.23}$$

Also note that the personal rate to Y_1 (to decode message m_1), is given by

$$R_1 = \sum_{i_2} P_2(i_2) R_{1,i_2}. \tag{B.24}$$

We also have,

$$|P_2(\cdot) - p_{T_2}(\cdot)|_1 = \sum_{i_2=1}^{d_2} |P_2(i_2) - p_{T_2}(i_2)| \leq \delta. \tag{B.25}$$

Using Eqs. (B.24) and (B.25), we now derive lower and upper bounds for R_1 as follows.

$$\begin{aligned}
R_1 = \sum_{i_2} P_2(i_2) R_{1,i_2} &= \sum_{i_2} p_{T_2}(i_2) R_{1,i_2} - \sum_{i_2} \left(p_{T_2}(i_2) - P_2(i_2)\right) R_{1,i_2} \\
&\geq \sum_{i_2} p_{T_2}(i_2) R_{1,i_2} - |P_2(\cdot) - p_{T_2}(\cdot)|_1 d_1 \qquad \text{(B.26)} \\
&= \sum_{i_2} p_{T_2} \left[I(T_1; Y_1|T_2 = i_2) - \delta_{i_2}\right] - |P_2(\cdot) - p_{T_2}(\cdot)|_1 d_1 \\
&= I(T_1; Y_1|T_2) - \sum_{i_2} p_{T_2}(i_2)\delta_{i_2} - |P_2(\cdot) - p_{T_2}(\cdot)|_1 d_1 \\
&= I(T_1; Y_1|T_2) - \delta \sum_{i_2} p_{T_2}(i_2) P_2(i_2) - |P_2(\cdot) - p_{T_2}(\cdot)|_1 d_1 \\
&\geq I(T_1; Y_1|T_2) - \delta - \delta d_1 \\
&\geq I(T_1; Y_1|T_2) - \delta(1 + d_1), \qquad \text{(B.27)}
\end{aligned}$$

where inequality (B.26) follows from (B.16). The upper bound is derived as follows:

$$\begin{aligned}
R_1 = \sum_{i_2} P_2(i_2) R_{1,i_2} &\leq \sum_{i_2} P_2(i_2) I(T_1; Y_1|T_2 = i_2) \\
&= \sum_{i_2} p_{T_2}(i_2) I(T_1; Y_1|T_2 = i_2) + \sum_{i_2} \left(P_2(i_2) - p_{T_2}(i_2)\right) I(T_1; Y_1|T_2 = i_2) \\
&\leq I(T_1; Y_1|T_2) + |P_2(\cdot) - p_{T_2}(\cdot)|_1 \max_{i_2} I(T_1; Y_1|T_2 = i_2) \\
&\leq I(T_1; Y_1|T_2) + \delta I_1, \qquad \text{(B.28)}
\end{aligned}$$

where $I_1 \triangleq \max_{i_2} I(T_1; Y_1|T_2 = i_2)$ is a finite non-negative real number. Combining Eqs. (B.27) and (B.28), we have

$$I(T_1; Y_1|T_2) - \delta(1 + d_1) \leq R_1 \leq I(T_1; Y_1|T_2) + \delta I_1. \tag{B.29}$$

Now, given each $i_2 \in \mathcal{T}_2$, we define for each $i_1 \in \mathcal{T}_1$,

$$R_{0,i_2,i_1} \triangleq I(A;Y_0|T_1=i_1) - \delta_{i_2,i_1} \leq d_0 \tag{B.30}$$

$$= \chi\left(p_{A|T_1}(j|i_1), \hat{\rho}_j^{Y_0}\right) - \delta_{i_2,i_1}, \tag{B.31}$$

where

$$\delta_{i_2,i_1} \triangleq \delta_{i_2} P_{1|2}(i_1|i_2) = \delta P_2(i_2) P_{1|2}(i_1|i_2). \tag{B.32}$$

Let us also define

$$\epsilon_{i_2,i_1} \triangleq \epsilon_{i_2} P_{1|2}(i_1|i_2) = \epsilon P_2(i_2) P_{1|2}(i_1|i_2), \text{and} \tag{B.33}$$

$$n_{i_2,i_1} \triangleq n_{i_2} P_{1|2}(i_1|i_2) = n P_2(i_2) P_{1|2}(i_1|i_2). \tag{B.34}$$

Given a fixed i_2, for each i_1, there exists an $(R_{0,i_2,i_1}, n_{i_2,i_1}, \epsilon_{i_2,i_1})$ random HSW code $\left\{\hat{\rho}^{A^{n_{i_2,i_1}}}_{m_{0,i_2,i_1}}, \Lambda^{0(i_2,i_1)}_{m_{0,i_2,i_1}}\right\}$; $m_{0,i_2,i_1} \in \left\{1,\ldots,2^{nR_{0,i_2,i_1}}\right\}$, for the conditional channel $\mathcal{N}^{i_2,i_1}_{A-Y_0}$, with each codeword chosen i.i.d. from $p_{A|T_1,T_2}(j|i_1,i_2) \equiv p_{A|T_1}(j|i_1)$, and each codeword satisfying

$$\mathrm{E}\left[2^{-n_{i_2,i_1}R_{0,i_2,i_1}} \sum_{m_{0,i_2,i_1}}^{2^{nR_{0,i_2,i_1}}} \mathrm{Tr}\left(\Lambda^{0(i_2,i_1)}_{m_{0,i_2,i_1}} \hat{\rho}^{Y_0^{n_{i_2,i_1}}}_{m_{0,i_2,i_1}}\right)\right] \geq 1 - \epsilon_{i_2,i_1}. \tag{B.35}$$

Note that owing to the symmetry of random code construction, (B.35) can alternatively be expressed as

$$\mathrm{E}\left[\mathrm{Tr}\left(\Lambda_1^{0(i_2,i_1)} \hat{\rho}_1^{Y_0^{n_{i_2,i_1}}}\right)\right] \geq 1 - \epsilon_{i_2,i_1}. \tag{B.36}$$

The personal rate to Y_0 (to decode its personal message m_0), is given by

$$R_0 = \sum_{i_2,i_1} P_2(i_2) P_{1|2}(i_1|i_2) R_{0,i_2,i_1}. \tag{B.37}$$

Lemma B.2 — Given two probability density functions $p(x)$ and $q(x)$ defined on the

same alphabet $\mathcal{X}$ that satisfy

$$\sum_{x\in\mathcal{X}} |p(x) - q(x)| \leq \delta, \tag{B.38}$$

and given that the conditional distributions $p(y|x)$ and $q(y|x)$ defined on the alphabets $\mathcal{X}$ and $\mathcal{Y}$, ($x \in \mathcal{X}$, $y \in \mathcal{Y}$) satisfy

$$\sum_{y\in\mathcal{Y}} |p(y|x) - q(y|x)| \leq \delta_x, \quad \forall x, \tag{B.39}$$

Then the joint distributions $p(y,x) = p(y|x)p(x)$ and $q(y,x) = q(y|x)q(x)$ must satisfy

$$\sum_{(x,y)\in(\mathcal{X},\mathcal{Y})} |p(y,x) - q(y,x)| \leq \delta + \sum_{x\in\mathcal{X}} \delta_x q(x). \tag{B.40}$$

Proof —

$$\begin{aligned}
& \sum_{(x,y)\in(\mathcal{X},\mathcal{Y})} |p(y,x) - q(y,x)| && \text{(B.41)}\\
= & \sum_{(x,y)\in(\mathcal{X},\mathcal{Y})} |\left(p(x) - q(x)\right) p(y|x) + (p(y|x) - q(y|x))q(x)| && \text{(B.42)}\\
\leq & \sum_{(x,y)\in(\mathcal{X},\mathcal{Y})} |p(x) - q(x)|\, p(y|x) + \sum_{(x,y)\in(\mathcal{X},\mathcal{Y})} |p(y|x) - q(y|x)|\, q(x) && \text{(B.43)}\\
= & \sum_{x\in\mathcal{X}} |p(x) - q(x)| + \sum_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{Y}} |p(y|x) - q(y|x)| \right) q(x) && \text{(B.44)}\\
\leq & \delta + \sum_{x\in\mathcal{X}} \delta_x q(x). && \text{(B.45)}
\end{aligned}$$

Now, we use Eq. (B.37) and *Lemma* B.2 to derive lower and upper bounds on R_0.

The derivation of the lower bound proceeds as follows.

$$
\begin{aligned}
R_0 &= \sum_{i_2,i_1} P_2(i_2)P_{1|2}(i_1|i_2)R_{0,i_2,i_1} && \text{(B.46)}\\
&= \sum_{i_2,i_1} p_{T_1|T_2}(i_1|i_2)p_{T_2}(i_2)R_{0,i_2,i_1} - \sum_{i_2,i_1}\left(p_{T_1|T_2}(i_1|i_2)p_{T_2}(i_2) - P_2(i_2)P_{1|2}(i_1|i_2)\right)R_{0,i_2,i_1}\\
&\geq \sum_{i_2,i_1} p_{T_1,T_2}(i_1,i_2)R_{0,i_2,i_1} - d_0\sum_{i_2,i_1}\left|p_{T_1|T_2}(i_1|i_2)p_{T_2}(i_2) - P_2(i_2)P_{1|2}(i_1|i_2)\right| && \text{(B.47)}\\
&\geq \sum_{i_2,i_1} p_{T_1,T_2}(i_1,i_2)\left(I(A;Y_0|T_1=i_1) - \delta_{i_2,i_1}\right) - d_0\left(\delta + \sum_{i_2} P_2(i_2)\delta_{i_2}\right) && \text{(B.48)}\\
&= I(A;Y_0|T_1) - \delta\sum_{i_2,i_1} p_{T_1,T_2}(i_1,i_2)P_2(i_2)P_{1|2}(i_1|i_2) - d_0\left(\delta + \delta\sum_{i_2} P_2(i_2)^2\right) && \text{(B.49)}\\
&\geq I(A;Y_0|T_1) - \delta - d_0(\delta+\delta) && \text{(B.50)}\\
&= I(A;Y_0|T_1) - \delta(1+2d_0), && \text{(B.51)}
\end{aligned}
$$

where (B.47) follows from Eq. (B.30), (B.48) follows from Eq. (B.30) and *Lemma* B.2, and (B.50) follows from the fact that $\sum_x p_1(x)p_2(x) \leq 1$ for two probability distribution functions $p_1(x)$ and $p_2(x)$ defined on a common alphabet. The derivation of the upper bound proceeds as follows.

$$
\begin{aligned}
R_0 &= \sum_{i_2,i_1} P_2(i_2)P_{1|2}(i_1|i_2)R_{0,i_2,i_1} && \text{(B.52)}\\
&\leq \sum_{i_2,i_1} P_2(i_2)P_{1|2}(i_1|i_2)I(A;Y_0|T_1=i_1) && \text{(B.53)}\\
&= \sum_{i_2,i_1} p_{T_2}(i_2)P_{T_1|T_2}(i_1|i_2)I(A;Y_0|T_1=i_1)\\
&\quad + \sum_{i_2,i_1}\left(P_2(i_2)P_{1|2}(i_1|i_2) - p_{T_2}(i_2)P_{T_1|T_2}(i_1|i_2)\right)I(A;Y_0|T_1=i_1) && \text{(B.54)}\\
&\leq I(A;Y_0|T_1)\\
&\quad + \max_{i_1} I(A;Y_0|T_1=i_1)\sum_{i_2,i_1}\left|P_2(i_2)P_{1|2}(i_1|i_2) - p_{T_2}(i_2)P_{T_1|T_2}(i_1|i_2)\right| && \text{(B.55)}\\
&\leq I(A;Y_0|T_1) + 2\delta I_0, && \text{(B.56)}
\end{aligned}
$$

where $I_0 \triangleq \max_{i_1} I(A;Y_0|T_1=i_1)$ is a finite non-negative real number. Combining

Eqs. (B.51) and (B.56), we have

$$I(A;Y_0|T_1) - \delta(1+2d_0) \leq R_0 \leq I(A;Y_0|T_1) + 2\delta I_0. \tag{B.57}$$

Combining inequalities (B.11), (B.29) and (B.57), we have constructed codebooks for the degraded broadcast channel $\mathcal{N}_{A-Y_0Y_1Y_2}$ transmitting the messages (m_0, m_1, m_2) at a rate 3-tuple (R_0, R_1, R_2), that can be brought arbitrarily close to the postulated ultimate capacity region (B.5) with $M = 3$ and $n = 1$, by choosing δ small enough. What remains to be shown, in order to complete the proof of achievability of the postulated capacity-region, is to

(i) instantiate the codewords of the codes we constructed above, and

(ii) to construct measurement operators for the receivers to decode the messages, and show that those measurement operators lead to an average overall error-probability that goes as $O(\epsilon)$ for sufficiently large blocklength n.

The above tasks are dealt with in the following two sections.

B.3.2 Instantiating the codewords

Let us denote the quantum states associated with the auxiliary random variables T_1 and T_2 as follows — $\mathcal{T}_k \equiv \{\hat{\sigma}_{k,1}, \hat{\sigma}_{k,2}, \ldots, \hat{\sigma}_{k,d_k}\}$, for $k \in \{1,2\}$. Recall that all the codewords $\hat{\rho}_{m_2}^{T_2^n}$ are of the same type $P_2(\cdot)$, for $m_2 \in W_2$. Without loss of generality, let us assume that the primary-cloud-center codewords are[4]

$$\hat{\rho}_1^{T_2^n} \triangleq \hat{\sigma}_{2,1}^{n_1} \otimes \hat{\sigma}_{2,2}^{n_2} \otimes \ldots \otimes \hat{\sigma}_{2,d_2}^{n_{d_2}} \tag{B.58}$$

$$= \bigotimes_{i_2=1}^{d_2} \hat{\sigma}_{2,i_2}^{n_{i_2}}, \tag{B.59}$$

[4]Note that $\hat{\sigma}_{2,k}^{n_k} \triangleq \hat{\sigma}_{2,k}^{\otimes n_k} = \hat{\sigma}_{2,k} \otimes \ldots \otimes \hat{\sigma}_{2,k}$ (n_k-fold tensor product). Also, recall from Eq. (B.21), that $n = n_1 + n_2 + \ldots + n_{d_2}$.

and $\pi_2(m_2)$ is a collection of permutations on n elements, such that

$$\hat{\rho}_{m_2}^{T_2^n} = \pi_2(m_2)\left(\hat{\rho}_1^{T_2^n}\right), \quad \forall m_2 \in W_2. \tag{B.60}$$

For each primary-cloud-center codeword $\hat{\rho}_{m_2}^{T_2^n}$, 2^{nR_1} secondary-cloud-center codewords $\hat{\rho}_{m_1,m_2}^{T_1^n}$ are chosen for every $m_1 \in W_1$. Each symbol of the secondary-cloud-center codewords $\hat{\rho}_{m_1,m_2}^{T_1^n}$ is chosen from i.i.d. from $\mathcal{T}_1$ according to the distribution $p_{T_1|T_2}(i_1|i_2)$. As $2^{nR_1} = \prod_{i_2=1}^{d_2} 2^{n_{i_2}R_{1,i_2}}$ (using Eqs. (B.24) and (B.21)), we may uniquely identify each message $m_1 \in W_1$ with a collection of messages m_{1,i_2} for $i_2 \in \{1,\ldots,d_2\}$, and $m_{1,i_2} \in W_{1,i_2} \triangleq \left\{1,\ldots,2^{nR_{1,i_2}}\right\}$. Hence, we have

$$\begin{aligned}
\hat{\rho}_{m_1,m_2}^{T_1^n} &= \pi_2(m_2)\left(\hat{\rho}_{m_1,1}^{T_1^n}\right) && \text{(B.61)}\\
&= \pi_2(m_2)\left(\hat{\rho}_{m_{1,1}}^{T_1^{n_1}} \otimes \hat{\rho}_{m_{1,2}}^{T_1^{n_2}} \otimes \ldots \otimes \hat{\rho}_{m_{1,d_2}}^{T_1^{n_{d_2}}}\right) && \text{(B.62)}\\
&= \pi_2(m_2)\left(\bigotimes_{i_2=1}^{d_2} \hat{\rho}_{m_{1,i_2}}^{T_1^{n_{i_2}}}\right). && \text{(B.63)}
\end{aligned}$$

Now, each one of the codewords $\hat{\rho}_{m_{1,i_2}}^{T_1^{n_{i_2}}}$ is of the same type $P_{1|2}(\cdot|i_2)$. Hence, without loss of generality, we can assume that[5]

$$\begin{aligned}
\hat{\rho}_1^{T_1^{n_{i_2}}} &\triangleq \hat{\sigma}_{1,1}^{n_{i_2,1}} \otimes \hat{\sigma}_{1,2}^{n_{i_2,2}} \otimes \ldots \otimes \hat{\sigma}_{1,d_1}^{n_{i_2,d_1}} && \text{(B.64)}\\
&= \bigotimes_{i_1=1}^{d_1} \hat{\sigma}_{1,i_1}^{n_{i_2,i_1}}, && \text{(B.65)}
\end{aligned}$$

and $\pi_{1,i_2}(m_{1,i_2})$ is a collection of permutations on n_{i_2} elements, such that for each $i_2 \in \mathcal{T}_2$,

$$\hat{\rho}_{m_{1,i_2}}^{T_1^{n_{i_2}}} = \pi_{1,i_2}(m_{1,i_2})\left(\hat{\rho}_1^{T_1^{n_{i_2}}}\right), \quad \forall m_{1,i_2} \in W_{1,i_2}. \tag{B.66}$$

Without loss of generality, $m_1 = 1$ can be mapped to $(m_{1,1}, m_{1,2}, \ldots, m_{1,d_2}) \equiv (1,1,\ldots,1)$, i.e.

$$\hat{\rho}_{1,1}^{T_1^n} = \hat{\rho}_1^{T_1^{n_1}} \otimes \hat{\rho}_1^{T_1^{n_2}} \otimes \ldots \otimes \hat{\rho}_1^{T_1^{n_{d_2}}}. \tag{B.67}$$

[5]Note that $n_{i_2,i_1} = n_{i_2}P_{1|2}(i_1|i_2)$, and thus, $n_{i_2} = n_{i_2,1} + n_{i_2,2} + \ldots + n_{i_2,d_1}$.

Now we can define a permutation by cascading the permutations $\pi_{1,i_2}(m_{1,i_2})$,

$$\pi_1(m_1) \triangleq \bigotimes_{i_2=1}^{d_2} \pi_{1,i_2}(m_{1,i_2}), \tag{B.68}$$

such that $\hat{\rho}_{m_1,1}^{T_1^n} = \pi_1(m_1)\left(\hat{\rho}_{1,1}^{T_1^n}\right)$. Combining this with Eq. (B.61) we have,

$$\begin{aligned}
\hat{\rho}_{m_1,m_2}^{T_1^n} &= \pi_2(m_2) \circ \pi_1(m_1)\left(\hat{\rho}_{1,1}^{T_1^n}\right) && \text{(B.69)}\\
&= \pi_2(m_2) \circ \pi_1(m_1)\left(\bigotimes_{i_2=1}^{d_2}\bigotimes_{i_1=1}^{d_1} \hat{\sigma}_{1,i_1}^{n_{i_2,i_1}}\right). && \text{(B.70)}
\end{aligned}$$

However, neither the primary nor the secondary cloud-center codewords are the actual codewords sent out by the transmitter, as they are after all drawn from hypothetical auxiliary alphabets. With $\hat{\sigma}_{1,i_1}^{n_{i_2,i_1}} \in \mathcal{T}_1$ given, the final transmitted codewords are drawn from Alice's alphabet around each secondary-cloud-center codeword, and are chosen i.i.d. symbol-by-symbol from the conditional distribution $p_{A|T_1,T_2}(j|i_1,i_2) \equiv p_{A|T_1}(j|i_1), \forall i_2$ (because $T_2 \to T_1 \to A$ is a Markov chain). As $2^{nR_0} = \prod_{i_2,i_1} 2^{n_{i_2,i_1}R_{0,i_2,i_1}}$ (using Eqs. (B.37) and (B.34)), we may uniquely identify each message $m_0 \in W_0$ with a collection of messages m_{0,i_2,i_1} for $(i_1,i_2) \in (\mathcal{T}_1,\mathcal{T}_2)$, and $m_{0,i_2,i_1} \in W_{0,i_2,i_1} \triangleq \left\{1,\ldots,2^{nR_{0,i_2,i_1}}\right\}$, $\forall i_1,i_2$. Hence, the transmitted codewords are given by

$$\begin{aligned}
\hat{\rho}_{m_0,m_1,m_2}^{A^n} &= \pi_2(m_2) \circ \pi_1(m_1)\left(\hat{\rho}_{m_0,1,1}^{A^n}\right) && \text{(B.71)}\\
&= \pi_2(m_2) \circ \pi_1(m_1)\left(\left(\hat{\rho}_{m_{0,1,1}}^{A^{n_{1,1}}} \otimes \hat{\rho}_{m_{0,1,2}}^{A^{n_{1,2}}} \otimes \ldots \hat{\rho}_{m_{0,1,d_1}}^{A^{n_{1,d_1}}}\right) \otimes \right.\\
&\qquad \left.\left(\hat{\rho}_{m_{0,2,1}}^{A^{n_{2,1}}} \otimes \hat{\rho}_{m_{0,2,2}}^{A^{n_{2,2}}} \otimes \ldots \hat{\rho}_{m_{0,2,d_1}}^{A^{n_{2,d_1}}}\right) \otimes \ldots \otimes \left(\hat{\rho}_{m_{0,d_2,1}}^{A^{n_{d_2,1}}} \otimes \hat{\rho}_{m_{0,d_2,2}}^{A^{n_{d_2,2}}} \otimes \ldots \hat{\rho}_{m_{0,d_2,d_1}}^{A^{n_{d_2,d_1}}}\right)\right)\\
&= \pi_2(m_2) \circ \pi_1(m_1)\left(\bigotimes_{i_2=1}^{d_2}\bigotimes_{i_1=1}^{d_1} \hat{\rho}_{m_{0,i_2,i_1}}^{A^{n_{i_2,i_1}}}\right). && \text{(B.72)}
\end{aligned}$$

In summary, given a message triplet (m_0,m_1,m_2), Alice first represents the message m_0 as a collection of messages from smaller index-sets $m_{0,i_2,i_1} \in W_{0,i_2,i_1}$, and generates the codeword $\hat{\rho}_{m_0,1,1}^{A^n}$ for $(m_0,m_1=1,m_2=1)$ as shown above. Thereafter, she applies the permutations $\pi_1(m_1)$ and $\pi_2(m_2)$ respectively in that order, to obtain the final

codeword $\hat{\rho}^{A^n}_{m_0,m_1,m_2}$ to be broadcast on the channel[6].

B.3.3 Receiver measurement and decoding error probability

The decoding process proceeds in three stages (M stages in general), which unravel the information from the layered cloud-center and cloud encoding technique we employed earlier. We start this section with a brief description of the decoding process and how it works. We then follow it up with constructing the actual measurement operators for the three receivers, and provide a rigorous error analysis in order to bound the overall average probability of decoding error.

Steps of the decoding process

The following are the steps of the decoding process:

1. Y_0, Y_1, and Y_2 measure $\left\{\Lambda^0_{m_2}\right\}$, $\left\{\Lambda^1_{m_2}\right\}$, and $\left\{\Lambda^2_{m_2}\right\}$ respectively on their respective received states $\hat{\rho}^{Y^n_k}_{m_0,m_1,m_2}$, and they declare their respective results of measurement, $\left\{\hat{m}_2^{(0)}, \hat{m}_2^{(1)}, \hat{m}_2^{(2)}\right\}$ to be the common message W_2.

2. Y_0 and Y_1 permute their respective codewords according to $\pi_2^{-1}(\hat{m}_2^{(k)})$, for $k \in \{0,1\}$ respectively. If Y_0 and Y_1 correctly decoded m_2 in step 1 above, after applying the permutations, they should jointly see a state that is close to $\hat{\rho}^{Y^n_0 Y^n_1}_{m_0,m_1,1}$. They measure each block of n_{i_2} symbols, $i_2 \in \{1,\ldots,d_2\}$, using

---

[6]

(i) The joint received codewords are given by

$$\hat{\rho}^{Y^n_0 Y^n_1 Y^n_2}_{m_0,m_1,m_2} = \mathcal{N}^{\otimes n}_{A-Y_0Y_1Y_2}\left(\hat{\rho}^{A^n}_{m_0,m_1,m_2}\right). \tag{B.73}$$

(ii) On averaging out the received codeword $\hat{\rho}^{Y^n_0 Y^n_1 Y^n_2}_{m_0,m_1,m_2}$ over messages m_0 and m_1, we obtain

$$\begin{aligned} \mathrm{E}_{m_0,m_1}\left[\hat{\rho}^{Y^n_0 Y^n_1 Y^n_2}_{m_0,m_1,m_2}\right] &= \sum_{(m_0,m_1)\in(W_0,W_1)} p_{W_0,W_1}(m_0,m_1)\hat{\rho}^{Y^n_0 Y^n_1 Y^n_2}_{m_0,m_1,m_2} \\ &= \hat{\rho}^{Y^n_0 Y^n_1 Y^n_2}_{m_2} = \mathcal{N}^{\otimes n}_{T_2-Y_0Y_1Y_2}\left(\hat{\rho}^{T^n_2}_{m_2}\right). \end{aligned} \tag{B.74}$$

(iii) To find the state received by Y_0, we must trace out the other receivers:

$$\hat{\rho}^{Y^n_0}_{m_0,m_1,m_2} = \mathrm{Tr}_{Y^n_1 Y^n_2}\left(\hat{\rho}^{Y^n_0 Y^n_1 Y^n_2}_{m_0,m_1,m_2}\right). \tag{B.75}$$

$\left\{\Lambda^{0(i_2)}_{m_{1,i_2}}\right\}$ and $\left\{\Lambda^{1(i_2)}_{m_{1,i_2}}\right\}$ respectively, and concatenate their measurement results $\left\{\hat{m}^{(k)}_{1,1}, \hat{m}^{(k)}_{1,2}, \ldots, \hat{m}^{(k)}_{1,d_2}\right\} \triangleq \hat{m}^{(k)}_1$, $k \in \{0,1\}$, which they declare to be their decoded message W_1.

3. Finally Y_0 applies the permutation $\pi_1^{-1}(\hat{m}^{(0)}_1)$ and obtains a state close to $\hat{\rho}^{Y_0^n}_{m_{0,1,1}}$. It measures using the measurement operators $\bigotimes_{i_2=1}^{d_2}\left(\bigotimes_{i_1=1}^{d_1}\Lambda^{0(i_2,i_1)}_{m_{0,i_2,i_1}}\right)$ and concatenates its results $\{m_{0,i_2,i_1}\}_{i_2=1,i_1=1}^{d_2,d_1}$ to obtain the estimate $\hat{m}^{(0)}_0$ which it declares as its decoded message W_0.

Construction of the measurement operators

The above procedure can be summarized by the action of the following POVM elements (measurement operators) for the three receivers, which (adhering to the notation set forth in the beginning of section B.2 above) are given by:

1. Y_2 — $\left\{\Lambda^2_{m_2}\right\}$.

2. Y_1 — $\left\{\Lambda^1_{m_1 m_2}\right\}$, where

$$\Lambda^1_{m_1 m_2} \triangleq \sqrt{\Lambda^1_{m_2}}\Lambda^1_{m_1|m_2}\sqrt{\Lambda^1_{m_2}}, \quad \text{and} \tag{B.76}$$

$$\Lambda^1_{m_1|m_2} \triangleq \pi_2(m_2)\left(\bigotimes_{i_2=1}^{d_2}\Lambda^{1(i_2)}_{m_{1,i_2}}\right). \tag{B.77}$$

3. Y_0 — $\left\{\Lambda^0_{m_0 m_1 m_2}\right\}$, where

$$\left\{\Lambda^0_{m_0 m_1 m_2}\right\} \triangleq \sqrt{\Lambda^0_{m_2}}\sqrt{\Lambda^0_{m_1|m_2}}\Lambda^0_{m_0|m_1 m_2}\sqrt{\Lambda^0_{m_1|m_2}}\sqrt{\Lambda^0_{m_2}}, \tag{B.78}$$

$$\Lambda^0_{m_1|m_2} \triangleq \pi_2(m_2)\left(\bigotimes_{i_2=1}^{d_2}\Lambda^{0(i_2)}_{m_{1,i_2}}\right), \quad \text{and} \tag{B.79}$$

$$\begin{aligned}\Lambda^0_{m_0|m_1 m_2} &\triangleq \bigotimes_{i_2=1}^{d_2}\left(\pi_{1,i_2}(m_{1,i_2})\left(\bigotimes_{i_1=1}^{d_1}\Lambda^{0(i_2,i_1)}_{m_{0,i_2,i_1}}\right)\right)\\ &= \pi_1(m_1)\left(\bigotimes_{i_2=1}^{d_2}\bigotimes_{i_1=1}^{d_1}\Lambda^{0(i_2,i_1)}_{m_{0,i_2,i_1}}\right).\end{aligned} \tag{B.80}$$

Error analysis

Our goal is to prove that with the codewords and the measurement operators we have constructed above, the overall average probability of correct decision $P_{m_0m_1m_2} = 1 - O(\epsilon)$, where

$$P_{m_0m_1m_2} = \mathrm{Tr}\left(\Lambda^0_{m_0m_1m_2} \otimes \Lambda^1_{m_1m_2} \otimes \Lambda^2_{m_2}\right) \hat{\rho}^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2}. \tag{B.81}$$

We will use the following two lemmas, whose proofs can be found in [52]:

Lemma B.3 — If $0 \leq \Lambda \leq 1$, then

$$\mathrm{Tr}(\Lambda\sigma) \geq \mathrm{Tr}(\Lambda\rho) - |\rho - \sigma|_1. \tag{B.82}$$

Lemma B.4 — If $0 \leq \Lambda \leq 1$, and $\mathrm{E}[\mathrm{Tr}(\Lambda\rho)] \geq 1 - \epsilon$ then

$$\mathrm{E}\left[|\sqrt{\Lambda}\rho\sqrt{\Lambda} - \rho|_1\right] \leq \sqrt{8\epsilon}. \tag{B.83}$$

Let us begin by defining two intermediate states in the decoding process:

$$\begin{aligned}
\hat{\rho}'^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2} &\triangleq \left(\sqrt{\Lambda^0_{m_2}} \otimes \sqrt{\Lambda^1_{m_2}} \otimes \sqrt{\Lambda^2_{m_2}}\right) \hat{\rho}^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2} \left(\sqrt{\Lambda^0_{m_2}} \otimes \sqrt{\Lambda^1_{m_2}} \otimes \sqrt{\Lambda^2_{m_2}}\right), \\
\hat{\rho}''^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2} &\triangleq \left(\sqrt{\Lambda^0_{m_1|m_2}} \otimes \sqrt{\Lambda^1_{m_1|m_2}}\right) \hat{\rho}'^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2} \left(\sqrt{\Lambda^0_{m_1|m_2}} \otimes \sqrt{\Lambda^1_{m_1|m_2}}\right).
\end{aligned}$$

The average probability of correct decision $P_{m_0m_1m_2}$ can be expressed as

$$\begin{aligned}
\mathrm{E}[P_{m_0m_1m_2}] &= \mathrm{E}\left[\mathrm{Tr}\left(\Lambda^0_{m_0|m_1m_2} \hat{\rho}''^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2}\right)\right] && \text{(B.84)} \\
&\geq \mathrm{E}\left[\mathrm{Tr}\left(\Lambda^0_{m_0|m_1m_2} \hat{\rho}'^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2}\right)\right] \\
&\quad -\mathrm{E}\left[|\hat{\rho}'^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2} - \hat{\rho}''^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2}|_1\right] && \text{(B.85)} \\
&\geq \mathrm{E}\left[\mathrm{Tr}\left(\Lambda^0_{m_0|m_1m_2} \hat{\rho}^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2}\right)\right] \\
&\quad -\mathrm{E}\left[|\hat{\rho}^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2} - \hat{\rho}'^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2}|_1\right] \\
&\quad -\mathrm{E}\left[|\hat{\rho}'^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2} - \hat{\rho}''^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2}|_1\right], && \text{(B.86)}
\end{aligned}$$

where (B.85) and (B.86) follow from *Lemma* B.3. In order to bound the last term in (B.86), let us consider the following:

$$
\begin{aligned}
&\mathrm{E}\left[\mathrm{Tr}\left(\Lambda^0_{m_1|m_2}\otimes\Lambda^1_{m_1|m_2}\right)\hat{\rho}'^{Y_0^nY_1^nY_2^n}_{m_0,m_1,m_2}\right] \\
&= \mathrm{E}\left[\mathrm{Tr}\left((\pi_2(m_2)\left(\bigotimes_{i_2=1}^{d_2}\Lambda^{0(i_2)}_{m_{1,i_2}}\otimes\bigotimes_{i_2=1}^{d_2}\Lambda^{1(i_2)}_{m_{1,i_2}}\right)\right)\hat{\rho}'^{Y_0^nY_1^nY_2^n}_{m_1,m_2}\right)\right] && \text{(B.87)}\\
&= \mathrm{E}\left[\mathrm{Tr}\left(\left(\bigotimes_{i_2=1}^{d_2}\Lambda^{0(i_2)}_{m_{1,i_2}}\otimes\bigotimes_{i_2=1}^{d_2}\Lambda^{1(i_2)}_{m_{1,i_2}}\right)\left(\pi_2^{-1}(m_2)\hat{\rho}'^{Y_0^nY_1^nY_2^n}_{m_1,m_2}\right)\right)\right] && \text{(B.88)}\\
&= \mathrm{E}\left[\mathrm{Tr}\left(\left(\bigotimes_{i_2=1}^{d_2}\Lambda^{0(i_2)}_{m_{1,i_2}}\otimes\bigotimes_{i_2=1}^{d_2}\Lambda^{1(i_2)}_{m_{1,i_2}}\right)\hat{\rho}'^{Y_0^nY_1^nY_2^n}_{m_1,1}\right)\right] && \text{(B.89)}\\
&\geq \mathrm{E}\left[\mathrm{Tr}\left(\left(\bigotimes_{i_2=1}^{d_2}\Lambda^{0(i_2)}_{m_{1,i_2}}\otimes\bigotimes_{i_2=1}^{d_2}\Lambda^{1(i_2)}_{m_{1,i_2}}\right)\hat{\rho}^{Y_0^nY_1^nY_2^n}_{m_1,1}\right)\right] \\
&\quad -\mathrm{E}\left[|\hat{\rho}^{Y_0^nY_1^nY_2^n}_{m_1,1}-\hat{\rho}'^{Y_0^nY_1^nY_2^n}_{m_1,1}|_1\right] && \text{(B.90)}\\
&\geq \mathrm{E}\left[\mathrm{Tr}\left(\left(\bigotimes_{i_2=1}^{d_2}\Lambda^{0(i_2)}_{m_{1,i_2}}\otimes\bigotimes_{i_2=1}^{d_2}\Lambda^{1(i_2)}_{m_{1,i_2}}\right)\left(\bigotimes_{i_2=1}^{d_2}\hat{\rho}^{Y_0^{n_{i_2}}Y_1^{n_{i_2}}}_{m_{1,i_2}}\right)\right)\right]-\sqrt{8\epsilon} && \text{(B.91)}\\
&= \prod_{i_2=1}^{d_2}\mathrm{E}\left[\mathrm{Tr}\left(\Lambda_1^{0(i_2)}\otimes\Lambda_1^{1(i_2)}\right)\hat{\rho}_1^{Y_0^{n_{i_2}}Y_1^{n_{i_2}}}\right]-\sqrt{8\epsilon} && \text{(B.92)}\\
&\geq 1-\sum_{i_2=1}^{d_2}\epsilon_{i_2}-\sqrt{8\epsilon} && \text{(B.93)}\\
&= 1-\sum_{i_2=1}^{d_2}\epsilon P_2(i_2)-\sqrt{8\epsilon} && \text{(B.94)}\\
&= 1-(\epsilon+\sqrt{8\epsilon}) && \text{(B.95)}\\
&\triangleq 1-\epsilon_1, && \text{(B.96)}
\end{aligned}
$$

where we define $\epsilon_1 \triangleq \epsilon+\sqrt{8\epsilon}$. Eq. (B.87) follows from Eqs. (B.77) and (B.79). Also note that we drop the message index m_0 in (B.87), because the expectation averages out m_0, as the measurement operator $(\Lambda^0_{m_1|m_2}\otimes\Lambda^1_{m_1|m_2})$ has no m_0 dependence. Equation (B.88) simply results from the fact that permuting the measurement operators is equivalent to inverse-permuting the codeword instead. Equation (B.89) follows from the definition of the permutation $\pi_2(m_2)$, and (B.90) follows from *Lemma* B.3. In

obtaining the first term in inequality (B.91), we drop the superscript Y_2^n from the received-state density operator (because it doesn't change the value of the expectation, as the measurement $(\Lambda^0_{m_1|m_2} \otimes \Lambda^1_{m_1|m_2})$ acts only on the joint Hilbert space of Y_0^n and Y_1^n), and we use Eqs. (B.61) and (B.63) to express $\hat{\rho}'^{Y_0^n Y_1^n}_{m_1,1} \equiv \bigotimes_{i_2=1}^{d_2} \hat{\rho}^{Y_0^{n_{i_2}} Y_1^{n_{i_2}}}_{m_{1,i_2}}$. To obtain the second term of the inequality (B.91), first note that (B.15) specialized to $m_2 = 1$, implies that $\mathrm{Tr}\left((\Lambda^0_1 \otimes \Lambda^1_1 \otimes \Lambda^2_1)\hat{\rho}^{Y_0^n Y_1^n Y_2^n}_{m_1,1}\right) \geq 1-\epsilon$, $\forall m_1$. Also note that by definition, $\hat{\rho}'^{Y_0^n Y_1^n Y_2^n}_{m_1,1} = \left(\sqrt{\Lambda^0_1} \otimes \sqrt{\Lambda^1_1} \otimes \sqrt{\Lambda^2_1}\right) \hat{\rho}^{Y_0^n Y_1^n Y_2^n}_{m_1,1} \left(\sqrt{\Lambda^0_1} \otimes \sqrt{\Lambda^1_1} \otimes \sqrt{\Lambda^2_1}\right)$. Hence, *Lemma* B.4 implies $\mathrm{E}\left[|\hat{\rho}^{Y_0^n Y_1^n Y_2^n}_{m_1,1} - \hat{\rho}'^{Y_0^n Y_1^n Y_2^n}_{m_1,1}|_1\right] \leq \sqrt{8\epsilon}$. Equation (B.92) follows from the symmetry of random code construction, that we earlier observed in going from (B.22) to (B.23). Inequality (B.93) follows from (B.23) and Eq. (B.94) follows from the definition (B.20).

Continuing from (B.86), we have

$$
\begin{aligned}
\mathrm{E}[P_{m_0 m_1 m_2}] &\geq \mathrm{E}\left[\mathrm{Tr}\left(\Lambda^0_{m_0|m_1 m_2} \hat{\rho}^{Y_0^n Y_1^n Y_2^n}_{m_0,m_1,m_2}\right)\right] - \sqrt{8\epsilon} - \sqrt{8\epsilon_1} && \text{(B.97)}\\
&= \mathrm{E}\left[\mathrm{Tr}\left(\pi_1(m_1) \bigotimes_{i_2=1}^{d_2} \bigotimes_{i_1=1}^{d_1} \Lambda^{0(i_2,i_1)}_{m_0,i_2,i_1}\right)\left(\pi_2(m_2) \circ \pi_1(m_1) \bigotimes_{i_2=1}^{d_2} \bigotimes_{i_1=1}^{d_1} \hat{\rho}^{Y_0^{n_{i_2,i_1}}}_{m_0,i_2,i_1}\right)\right] \\
&\quad -\sqrt{8\epsilon} - \sqrt{8\epsilon_1} && \text{(B.98)}\\
&= \mathrm{E}\left[\mathrm{Tr}\left(\pi_2(m_2) \circ \pi_1(m_1) \bigotimes_{i_2=1}^{d_2} \bigotimes_{i_1=1}^{d_1} \Lambda^{0(i_2,i_1)}_{m_0,i_2,i_1}\right)\left(\pi_2(m_2) \circ \pi_1(m_1) \bigotimes_{i_2=1}^{d_2} \bigotimes_{i_1=1}^{d_1} \hat{\rho}^{Y_0^{n_{i_2,i_1}}}_{m_0,i_2,i_1}\right)\right] \\
&\quad -\sqrt{8\epsilon} - \sqrt{8\epsilon_1} && \text{(B.99)}\\
&= \mathrm{E}\left[\mathrm{Tr}\left(\bigotimes_{i_2=1}^{d_2} \bigotimes_{i_1=1}^{d_1} \Lambda^{0(i_2,i_1)}_{m_0,i_2,i_1}\right)\left(\bigotimes_{i_2=1}^{d_2} \bigotimes_{i_1=1}^{d_1} \hat{\rho}^{Y_0^{n_{i_2,i_1}}}_{m_0,i_2,i_1}\right)\right] - \sqrt{8\epsilon} - \sqrt{8\epsilon_1} && \text{(B.100)}\\
&= \prod_{i_2=1}^{d_2} \prod_{i_1=1}^{d_1} \mathrm{E}\left[\mathrm{Tr}\left(\Lambda_1^{0(i_2,i_1)} \hat{\rho}_1^{Y_0^{n_{i_2,i_1}}}\right)\right] - \sqrt{8\epsilon} - \sqrt{8\epsilon_1} && \text{(B.101)}\\
&\geq 1 - \sum_{i_2=1}^{d_2} \sum_{i_1=1}^{d_1} \epsilon_{i_2,i_1} - \sqrt{8\epsilon} - \sqrt{8\epsilon_1} && \text{(B.102)}\\
&= 1 - \sum_{i_2=1}^{d_2} \sum_{i_1=1}^{d_1} \epsilon P_{1|2}(i_i|i_2) P_2(i_2) - \sqrt{8\epsilon} - \sqrt{8\epsilon_1} && \text{(B.103)}\\
&= 1 - \left(\epsilon + \sqrt{8\epsilon} + \sqrt{8(\epsilon + \sqrt{8\epsilon})}\right) && \text{(B.104)}\\
&= 1 - O(\epsilon), && \text{(B.105)}
\end{aligned}
$$

where (B.97) follows from (B.86), (B.96), and two applications of *Lemma* B.4. Equation (B.98) follows from (B.80) and (B.72). Note that dropping the superscripts Y_1^n and Y_2^n on the received joint quantum state in Eq. (B.97) doesn't make a difference, as the measurement operators $\left\{\Lambda^0_{m_0|m_1m_2}\right\}$ act only on the Hilbert space of Y_0^n. Equation (B.99) follows from the fact that the measurement operators $\left\{\Lambda^0_{m_0|m_1m_2}\right\}$ do not depend on m_2, and hence can be chosen arbitrarily up to a permutation $\pi_2(m_2)$. Next, we remove the permutations $\pi_2(m_2) \circ \pi_1(m_1)$ from both the parentheses in (B.99), so that the trace remains unchanged in Eq. (B.100). Equation (B.101) follows from the symmetry of the HSW code construction, (B.102) follows from (B.36), (B.103) follows from the definition (B.33), and (B.105) completes the proof.

B.3.4 Proof of achievability with M receivers

The proof of the achievability of the capacity region of the M-receiver degraded quantum broadcast channel (B.5) is a straightforward generalization of the $M = 3$ case we proved above. We will not go through every single detail of the M-receiver achievability proof here, but we will rather sketch the proof. Similar to the $M = 3$ case, we need to prove achievability only for $n = 1$, because the same proof can be applied to n-use (larger) quantum systems of the transmitter and the receivers to obtain the general $n > 1$ capacity region (B.5).

For any $\epsilon, \delta > 0$, we aim to show here that for rate M-tuples $(R_0, \ldots, R_{M-1})$ satisfying

$$\begin{aligned} I(A;Y_0|T_1) - \delta(1 + (M-1)d_0) &\leq R_0 \leq I(A;Y_0|T_1) + (M-1)\delta I_0, \\ I(T_k;Y_k|T_{k+1}) - \delta(1 + (M-k-1)d_k) &\leq R_k \leq I(T_k;Y_k|T_{k+1}) + (M-k-1)\delta I_k, \\ 0 &\leq R_{M-1} = I(T_{M-1};B_{M-1}) - \delta, \end{aligned} \tag{B.106}$$

there exists an $(2^{nR_0}, \ldots, 2^{nR_{M-1}}, n, O(\epsilon))$ code for the degraded broadcast channel $\mathcal{N}_{A-Y_0\ldots Y_{M-1}}$, where $d_k \triangleq |\mathcal{T}_k|$ is the cardinality of the alphabet associated with the auxiliary random variable T_k and the cardinality of the transmitter's alphabet, $|\mathcal{A}| \triangleq d_0$. $I_k \triangleq \max_{i_{k+1}} I(T_k;Y_k|T_{k+1} = i_{k+1})$ are finite non-negative real numbers.

Using HSW theorem [27, 28, 29] for the channel $\mathcal{N}_{T_{M-1}-Y_0\ldots Y_{M-1}}$, let us obtain a (R_{M-1}, n, ϵ) code $\left\{\hat{\rho}^{T^n_{M-1}}_{m_{M-1}}, \Lambda^0_{m_{M-1}}, \Lambda^1_{m_{M-1}}, \ldots, \Lambda^{M-1}_{m_{M-1}}\right\}$ with all codewords chosen i.i.d. from the distribution $p_{T_{M-1}}(i_{M-1})$ of type P_{M-1}, satisfying $|P_{M-1} - p_{T_{M-1}}(\cdot)|_1 \leq \delta$, and for all $m_{M-1} \in W_{M-1}$,

$$\mathrm{Tr}\left(\bigotimes_{k=0}^{M-1} \Lambda^k_{m_{M-1}}\right) \hat{\rho}^{Y^n_0\ldots Y^n_{M-1}}_{m_{M-1}} \geq 1 - \epsilon, \tag{B.107}$$

where $\hat{\rho}^{Y^n_0\ldots Y^n_{M-1}}_{m_{M-1}} = \bigotimes_{l=1}^n \hat{\rho}^{Y^n_{0,l}\ldots Y^n_{M-1,l}}_{m_{M-1}}$ are product-state codewords. Treating these codewords as the primary cloud-centers, for each $i_{M-1} \in \mathcal{T}_{M-1}$, we choose another layer of codewords $\hat{\rho}^{T^{n_{i_{M-1}}}_{M-2}}_{m_{M-2,i_{M-1}}}$ for the conditional channel $\mathcal{N}^{i_{M-1}}_{T_{M-2}-Y_0\ldots Y_{M-2}}$, picked i.i.d. from the distribution $p_{T_{M-2}|T_{M-1}}(i_{M-2}|i_{M-1})$, which form a random HSW code of rate $R_{M-2,i_{M-1}}$. Taking the average of these rates over the entire codebook, the desired rate bound $I(T_{M-2}; Y_{M-2}|T_{M-1}) - \delta(1 + d_{M-1}) \leq R_{M-2} \leq I(T_{M-2}; Y_{M-2}|T_{M-1}) + \delta I_{M-2}$ can be established for R_{M-2}. Continuing in this manner, we keep selecting HSW codewords from the alphabets of the auxiliary random variables with the appropriate conditional distributions, viz. by applying HSW theorem to the channel $\mathcal{N}^{i_{M-1},\ldots,i_l}_{T_{l-1}-Y_0\ldots Y_{l-1}}$, to select a code of overall rate R_{l-1} close to the desired bound (B.106). Proving the rate bounds involve applications of *Lemma* B.2 and simple manipulations similar to those leading to the rate bounds for R_1 and R_0 in the $M = 3$ proof we did earlier.

Codewords and measurement operators are selected in a layered way, exactly as we did earlier for the $M = 3$ case. For the chosen measurement and codewords, the bound for the average probability of correct decision works out to be

$$\mathrm{E}\left[P_{m_0\ldots m_{M-1}}\right] \geq 1 - \left(\epsilon + \sqrt{8\epsilon} + \sqrt{8\epsilon_1} + \sqrt{8\epsilon_2} + \ldots + \sqrt{8\epsilon_{M-2}}\right), \tag{B.108}$$

where $\epsilon_{i+1} = \epsilon_i + \sqrt{8\epsilon_i}$, for $i \in \{0, \ldots, M-3\}$, and $\epsilon_0 \triangleq \epsilon$. Hence, $\mathrm{E}\left[P_{m_0\ldots m_{M-1}}\right] \geq 1 - O(\epsilon)$, as desired. The proof parallels the layered codebook construction technique used for classical degraded broadcast channels, and works out pretty much in the same manner as the $M = 3$ proof.

B.4 Capacity Region: Proof (Converse)

Our goal in proving the converse to the capacity-region proof is to show that any achievable rate M-tuple $(R_0, \ldots, R_{M-1})$ must be inside the ultimate rate-region proposed by Eqs. (B.5). Let us assume that $(R_0, \ldots, R_{M-1})$ is achievable. Let $\{x^n(m_0, \ldots, m_{M-1})\}$, and POVMs $\left\{\Lambda^0_{m_0 \ldots m_{M-1}}\right\}$, $\left\{\Lambda^1_{m_1 \ldots m_{M-1}}\right\}$, $\ldots$, $\left\{\Lambda^{M-1}_{m_{M-1}}\right\}$ comprise a $(2^{nR_0}, \ldots, 2^{nR_{M-1}}, n, \epsilon)$ code in the achieving sequence. Let us suppose that the receivers $Y_0, \ldots, Y_{M-1}$ store their respective decoded messages in registers $\hat{W}_0, \ldots, \hat{W}_{M-1}$. Then, for real numbers $\epsilon_{n,k} \to 0$, we have for $k \in \{0, 1, \ldots, M-2\}$

$$\begin{aligned}
nR_k &= H(W_k) && \text{(B.109)}\\
&\leq I(W_k; \hat{W}_k) + n\epsilon_{n,k} && \text{(B.110)}\\
&\leq \chi\left(p_{W_k}(m_k), \hat{\rho}^{Y_k^n}_{m_k}\right) + n\epsilon_{n,k} && \text{(B.111)}\\
&< \sum_{m_{k+1}} p_{W_{k+1}}(m_{k+1}) \chi\left(p_{W_k}(m_k), \hat{\rho}^{Y_k^n}_{\mathbf{m}_k^{k+1}}\right) n\epsilon_{n,k} && \text{(B.112)}\\
&= I(W_k; Y_k^n | W_{k+1}) + n\epsilon_{n,k}, && \text{(B.113)}
\end{aligned}$$

where (B.110) and (B.111) follow from Fano's inequality and the Holevo bound respectively. Equation (B.112) follows from concavity of Holevo information (as $\hat{\rho}^{Y_k^n}_{m_k} = \sum_{m_{k+1}} p_{W_{k+1}}(m_{k+1}) \hat{\rho}^{Y_k^n}_{\mathbf{m}_k^{k+1}}$). For $k = 0$, we further have

$$\begin{aligned}
nR_0 &\leq I(W_0; Y_0^n | W_1) + \epsilon_{n,0} && \text{(B.114)}\\
&\leq I(A^n; Y_0^n | W_1) + \epsilon_{n,0}, && \text{(B.115)}
\end{aligned}$$

where (B.115) follows from the Markov nature of $(W_0, \ldots, W_{M-1}) \to A^n \to Y_0^n \to \ldots \to Y_{M-1}^n$. We also have similarly, for $\epsilon_{n,M-1} \to 0$,

$$\begin{aligned}
nR_{M-1} &= nH(W_{M-1}) && \text{(B.116)}\\
&\leq I(W_{M-1}; \hat{W}_{M-1}) + n\epsilon_{n,M-1} && \text{(B.117)}\\
&\leq \chi\left(p_{W_{M-1}}(m_{M-1}), \hat{\rho}^{Y_{M-1}^n}_{m_{M-1}}\right) + n\epsilon_{n,M-1} && \text{(B.118)}\\
&= I(W_{M-1}; Y_{M-1}^n) + n\epsilon_{n,M-1}. && \text{(B.119)}
\end{aligned}$$

Choosing $T_k = W_k$ for $k \in \{1, 2, \ldots, M-1\}$ completes the proof.

Appendix C

Theorem on property of $g(x)$

The converse proofs of the capacity region for the Bosonic broadcast channel with and without thermal noise, in chapter 3, use a theorem on a property of the Bose-Einstein entropy function, $g(x) = (1+x)\ln(1+x) - x\ln x$, in order to conclude Eqs. (3.59) and (3.90). In this appendix, we prove two lemmas which lead to the proof of a theorem. After that, we show how the theorem implies Eqs. (3.59) and (3.90), as two simple special cases.

Lemma A.1 — For all real numbers $x \geq 0$, $C \geq 0$, and $0 \leq \kappa \leq 1$, the following inequality holds:

$$\frac{\ln\left(1+\frac{1}{\kappa x+C}\right)}{\ln\left(1+\frac{1}{x}\right)} \geq \frac{\kappa x(1+x)}{(\kappa x+C)(1+\kappa x+C)}. \tag{C.1}$$

Proof — Define a function $f(x) \triangleq x(1+x)\ln(1+1/x)$. We claim that $f(x)$ has the following properties[1]:

[1]Proofs —

1. We can express $f(x)$ as, $f(x) = x(g(x) - \ln x)$. Therefore, $\lim_{x\to 0} f(x) = \lim_{x\to 0}(xg(x)) - \lim_{x\to 0}(x\ln x)$. It is readily verified by applying the L' Hopital's rule, that $\lim_{x\to 0}(xg(x)) = \lim_{x\to 0}(x\ln x) = 0$.

2. By straightforward differentiation, $f''(x) = 2\ln(1+1/x) - (2x+1)/(x(1+x))$. Claim: $\ln(1+y) \leq y(y+2)/2(y+1)$, $\forall y \geq 0$. Proof: It is easy to see the following:

   - Both the left and right hand sides of the proposed inequality go to zero at $y = 0$.
   - Both $\ln(1+y)$ and $y(y+2)/2(y+1)$ are positive for $y > 0$.

1. $\lim_{x\to 0} f(x) = 0$.

2. $f(x)$ is a concave function, i.e., the second derivative $f''(x) \leq 0$, for $x \geq 0$.

3. $f(x)$ is monotonically increasing for $x \geq 0$.

Given properties 1 and 2 above, we have $f(\kappa x) \geq \kappa f(x)$, for $x \geq 0$ and $0 \leq \kappa \leq 1$. We further have from property 3 above, that for any non-negative real number $C \geq 0$, $f(\kappa x + C) \geq f(\kappa x)$, for $x \geq 0$ and $0 \leq \kappa \leq 1$. Combining the two above, we obtain $f(\kappa x + C) \geq \kappa f(x)$. Substituting the explicit form of $f(x)$, we have Eq. (C.1), that we set out to prove.

Lemma A.2 — The following holds:

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2} g\left(\kappa g^{-1}(y) + C\right) \geq 0, \tag{C.2}$$

for $y \geq 0$, where C is a non-negative real number.

Proof — Let us define $p(y) \triangleq g\left(\kappa g^{-1}(y) + C\right)$. Differentiating twice with respect to y, we get

$$\begin{aligned} \frac{\mathrm{d}^2 p(y)}{\mathrm{d}y^2} &= \kappa \ln\left(1 + \frac{1}{\kappa g^{-1}(y) + C}\right)\left(\frac{\mathrm{d}^2}{\mathrm{d}y^2} g^{-1}(y)\right) \\ &\quad -\kappa^2 \frac{1}{(\kappa g^{-1}(y) + C)(1 + \kappa g^{-1}(y) + C)}\left(\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right)^2. \end{aligned} \tag{C.3}$$

Now consider the identity $g(g^{-1}(y)) = y$, and substitute $g^{-1}(y) = x$. Differentiating

---

- $\frac{\mathrm{d}}{\mathrm{d}y}\ln(1+y) \leq \frac{\mathrm{d}}{\mathrm{d}y}\left[\frac{y(y+2)}{2(y+1)}\right]$, for $y \geq 0$.

Hence, $\ln(1+y) \leq y(y+2)/2(y+1)$, $\forall y \geq 0$. Substituting $y = 1/x$, we get $f''(x) \leq 0$, for $x \geq 0$.

3. By straightforward differentiation, $f'(x) = (2x+1)\ln(1+1/x) - 1$. Claim: $\ln(1+y) \geq y/(y+2)$, $\forall y \geq 0$. Proof: It is easy to see the following:

   - Both the left and right hand sides of the proposed inequality go to zero at $y = 0$.
   - Both $\ln(1+y)$ and $y/(y+2)$ are positive for $y > 0$.
   - $\frac{\mathrm{d}}{\mathrm{d}y}\ln(1+y) \geq \frac{\mathrm{d}}{\mathrm{d}y}\left[\frac{y}{y+2}\right]$, for $y \geq 0$.

   Hence, $\ln(1+y) \geq y/(y+2)$, $\forall y \geq 0$. Substituting $y = 1/x$, we get $f'(x) \geq 0$, for $x \geq 0$. Since $\lim_{x\to 0} f(x) = 0$, $f(x)$ must be monotonically increasing for $x \geq 0$.

both sides of the identity with respect to y, we get $(\mathrm{d}g(x)/\mathrm{d}x)(\mathrm{d}x/\mathrm{d}y) = 1$, which implies $\mathrm{d}x/\mathrm{d}y = 1/(\mathrm{d}g(x)/\mathrm{d}x)$. Therefore, we get

$$\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y) = \frac{1}{\ln\left(1+\frac{1}{g^{-1}(y)}\right)}, \tag{C.4}$$

and thus,

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2}g^{-1}(y) = \frac{1}{g^{-1}(y)(1+g^{-1}(y))}\frac{1}{\left[\ln\left(1+\frac{1}{g^{-1}(y)}\right)\right]^3}. \tag{C.5}$$

Substituting Eqs. (C.4) and (C.5) into Eq. (C.3) we finally obtain,

$$\begin{aligned}\frac{\mathrm{d}^2p(y)}{\mathrm{d}y^2} &= \frac{\kappa}{g^{-1}(y)(1+g^{-1}(y))\left[\ln\left(1+\frac{1}{g^{-1}(y)}\right)\right]^2}\left[\frac{\ln\left(1+\frac{1}{\kappa g^{-1}(y)+C}\right)}{\ln\left(1+\frac{1}{\kappa g^{-1}(y)}\right)}\right. \\ &\quad \left. - \frac{\kappa g^{-1}(y)(1+g^{-1}(y))}{(\kappa g^{-1}(y)+C)(1+\kappa g^{-1}(y)+C)}\right] \qquad \text{(C.6)} \\ &\geq 0, \qquad \text{(C.7)}\end{aligned}$$

where the last inequality follows from using *Lemma* A.1, along with the fact that $g^{-1}(y) \geq 0$, $\forall y \geq 0$.

Theorem A.3 — Given non-negative real numbers $x_k \in \mathbb{R}^+$, for $k \in \{1,\ldots,n\}$, and $0 \leq \kappa \leq 1$, if x_0 is defined by

$$\sum_{k=1}^{n}\frac{1}{n}g(x_k) = g(x_0), \tag{C.8}$$

then the following inequality holds:

$$\sum_{k=1}^{n}\frac{1}{n}g(\kappa x_k + C) \geq g(\kappa x_0 + C), \tag{C.9}$$

where $g(x) \equiv (1+x)\log(1+x) - x\log(x)$, and $C \geq 0$.

Proof — Because $g(x)$ is a $1-1$ function, we can define unambiguously the inverse function $h(y) \equiv g^{-1}(y)$, such that $y = g(x) \equiv x = h(y)$ for $x, y \geq 0$. Define $y_k \triangleq g(x_k)$, $y'_k \triangleq g\left(\kappa g^{-1}(y_k) + C\right)$ and $l(y_k) \triangleq y_k - y'_k$, for $k \in \{0,1,\ldots,n\}$. Rephrasing

the theorem in terms of $h(y)$, we have the following theorem. Given

$$y_0 = \frac{1}{n}\sum_{k=1}^{n} y_k, \quad y_k \geq 0, \forall k, \tag{C.10}$$

the following is true:

$$\frac{1}{n}\sum_{k=1}^{n} y'_k \geq y'_0. \tag{C.11}$$

Using *Lemma* A.2, it follows that $l(y) = y - y' = y - g\left(\kappa g^{-1}(y) + C\right)$ is a convex function in y, i.e. $l''(y) \leq 0$. Thus, Eqn. (C.10) implies

$$l(y_0) \geq \frac{1}{n}\sum_{k=1}^{n} l(y_k), \tag{C.12}$$

which implies

$$\begin{aligned} y_0 - y'_0 &\geq \frac{1}{n}\sum_{k=1}^{n}(y_k - y'_k) && \text{(C.13)} \\ &\geq \frac{1}{n}\sum_{k=1}^{n} y_k - \frac{1}{n}\sum_{k=1}^{n} y'_k. && \text{(C.14)} \end{aligned}$$

Using Eq. (C.10), we thus have

$$\frac{1}{n}\sum_{k=1}^{n} y'_k \geq y'_0, \tag{C.15}$$

which completes the proof. Eqs. (3.59) and (3.90) follow as straightforward consequences of Theorem A.3, as shown below.

Corollary A.4 — Given

$$\sum_k \frac{1}{2^{nR_C}} g\left(\eta\beta_k \bar{N}_k\right) = g\left(\eta\beta\bar{N}\right), \tag{C.16}$$

and $\eta > 1/2$, we have that

$$\sum_k \frac{1}{2^{nR_C}} g\left((1-\eta)\beta_k \bar{N}_k\right) \geq g\left((1-\eta)\beta\bar{N}\right). \tag{C.17}$$

Proof — Substitute $x_k \triangleq \eta\beta_k\bar{N}_k$, $x_0 \triangleq \eta\beta\bar{N}$, $n \triangleq 1/2^{nR_C}$ and $\kappa \triangleq (1-\eta)/\eta$. As $\eta > 1/2$, it follows that $0 \leq \kappa \leq 1$. Using these substitutions, Eq. (C.17) follows from Theorem A.3, with $C = 0$.

Corollary A.5 — Given

$$\sum_k \frac{1}{2^{nR_C}} g\left(\eta\beta_k\bar{N}_k^A + (1-\eta)N\right) = g\left(\eta\beta\bar{N} + (1-\eta)N\right). \tag{C.18}$$

and $\eta > 1/2$, we have that

$$\sum_k \frac{1}{2^{nR_C}} g\left((1-\eta)\beta_k\bar{N}_k^A + \eta N\right) \geq g\left((1-\eta)\beta\bar{N} + \eta N\right). \tag{C.19}$$

Proof — Substitute $x_k \triangleq \eta\beta_k\bar{N}_k^A + (1-\eta)N$, $x_0 \triangleq \eta\beta\bar{N} + (1-\eta)N$, $n \triangleq 1/2^{nR_C}$ and $\kappa \triangleq (1-\eta)/\eta$. As $\eta > 1/2$, we have $0 \leq \kappa \leq 1$. Using these substitutions, Eq. (C.19) follows from Theorem A.3, with $C = (2\eta - 1)N/\eta > 0$.

Appendix D

Proofs of Weak Minimum Output Entropy Conjectures 2 and 3 for the Wehrl Entropy Measure

This appendix contains the proofs of the Wehrl-entropy versions of the weak conjectures 2 and 3 that do not draw upon the Entropy Power Inequality (EPI). As we pointed out in chapter 4, the EPI quickly leads to the Wehrl-entropy proofs for the strong forms of all the minimum output entropy conjectures. We still include the following proofs in the thesis for the sake of completeness, and because these proofs could be of mathematical interest in their own right.

Wehrl entropy is the Shannon differential entropy of the Husimi probability function $Q(\mu)$ for the state $\hat{\rho}$ [64], i.e., for a single mode we have

$$W(\hat{\rho}) \equiv -\int Q(\mu) \ln [\pi Q(\mu)] \mathrm{d}^2\mu, \tag{D.1}$$

where $Q(\mu) \equiv \langle \mu|\hat{\rho}|\mu\rangle/\pi$ with $|\mu\rangle$ a coherent state. The Wehrl entropy provides a measurement of the state $\hat{\rho}$ in phase space and its minimum value is achieved on coherent states [64].

D.1 Weak conjecture 2

The following single-mode version of conjecture 2 was stated in chapter 4:

Weak Conjecture 2 — *Let a lossless beam splitter have input $\hat{a}$ in its vacuum state, input $\hat{b}$ in a zero-mean state with von Neumann entropy $S(\hat{\rho}^B) = g(K)$, and output $\hat{c}$ from its transmissivity-η port. Then the von Neumann entropy of output $\hat{c}$ is minimized when input $\hat{b}$ is in a thermal state with average photon number K, and the minimum output entropy is given by $\mathbb{S}(\hat{\rho}^C) = g((1-\eta)K)$.*

The following is an analogous statement of the conjecture for the Wehrl entropy:

Weak Conjecture 2: Wehrl — *Let a lossless beam splitter have input $\hat{a}$ in its vacuum state, input $\hat{b}$ in a zero-mean state with Wehrl entropy $W(\hat{\rho}^B) = 1 + \ln(K+1)$, and output $\hat{c}$ from its transmissivity-η port. Then the Wehrl entropy of output $\hat{c}$ is minimized when input $\hat{b}$ is in a thermal state with average photon number K, and the minimum output entropy is given by $\mathbb{W}(\hat{\rho}^C) = 1 + \ln(K(1-\eta)+1)$.*

Proof — Before we begin the proof of the Wehrl-entropy conjecture, let us recall a few definitions. The antinormally ordered characteristic function $\chi_A^{\hat{\rho}}(\zeta)$ of a state $\hat{\rho}$ is given by:

$$\chi_A^{\hat{\rho}}(\zeta) = \mathrm{tr}\left(\hat{\rho} e^{-\zeta^* \hat{a}} e^{\zeta \hat{a}^\dagger}\right). \tag{D.2}$$

Also, the antinormally ordered characteristic function $\chi_A^{\hat{\rho}}(\zeta)$ and the Husimi function $Q_{\hat{\rho}}(\mu) \equiv \langle \mu | \hat{\rho} | \mu \rangle / \pi$ of a state $\hat{\rho}$ form a 2-D Fourier-Transform Inverse-Transform pair:

$$\chi_A^{\hat{\rho}}(\zeta) = \int Q_{\hat{\rho}}(\mu) e^{\zeta \mu^* - \zeta^* \mu} d^2\mu, \tag{D.3}$$

$$Q_{\hat{\rho}}(\mu) = \frac{1}{\pi^2} \int \chi_A^{\hat{\rho}}(\zeta) e^{-\zeta \mu^* + \zeta^* \mu} d^2\zeta \tag{D.4}$$

As the two input states to the beamsplitter are in a product state, Eq. D.2 implies that the output state characteristic function is a product of the input state characteristic functions with scaled arguments:

$$\chi_A^{\hat{\rho}_C}(\zeta) = \chi_A^{\hat{\rho}_A}(\sqrt{\eta}\zeta)\chi_A^{\hat{\rho}_B}(\sqrt{1-\eta}\zeta) \tag{D.5}$$

The input $\hat{a}$ is given to be in the vacuum state. Thus, the Husimi function and the Wehrl entropy of the input $\hat{a}$ are given by:

$$Q_{\hat{\rho}_A}(\mu) = \frac{1}{\pi}e^{-|\mu|^2}, \tag{D.6}$$

$$W(\hat{\rho}_A) = 1. \tag{D.7}$$

Equation D.5, and the multiplication-convolution property of Fourier transforms (FT) give us

$$\begin{aligned} Q_{\hat{\rho}_C}(\mu) &= \frac{1}{\eta}Q_{\hat{\rho}_A}\left(\frac{\mu}{\sqrt{\eta}}\right) \star \frac{1}{(1-\eta)}Q_{\hat{\rho}_B}\left(\frac{\mu}{\sqrt{1-\eta}}\right) \\ &= \frac{1}{\pi\eta}e^{-|\mu|^2/\eta} \star \frac{1}{(1-\eta)}Q_{\hat{\rho}_B}\left(\frac{\mu}{\sqrt{1-\eta}}\right) \end{aligned} \tag{D.8}$$

where, we used the scaling-property of FT: $\chi_A^{\hat{\rho}}(\sqrt{\eta}\zeta) \longleftrightarrow (1/\eta)Q_{\hat{\rho}}(\mu/\sqrt{\eta})$.

If the state of the input $\hat{b}$ is a thermal state with mean photon number K, i.e.,

$$\hat{\rho}_B = \frac{1}{\pi K}\int e^{-|\alpha|^2/K}|\alpha\rangle\langle\alpha|d^2\alpha,$$

we find that

$$W(\hat{\rho}_B) = 1 + \ln(K+1), \tag{D.9}$$

which satisfies the hypothesis of our Wehrl-entropy conjecture. Using Eq. D.9, we can then write out the Husimi function of the output state $\hat{c}$:

$$Q_{\hat{\rho}_C}(\mu) = \frac{1}{\pi(1+(1-\eta)K)}e^{-|\mu|^2/(1+(1-\eta)K)}, \tag{D.10}$$

obtaining

$$W(\hat{\rho}_C) = 1 + \ln(K(1-\eta)+1), \tag{D.11}$$

for the resulting Wehrl entropy, which provides us with an upper bound to the minimum output Wehrl entropy:

$$\mathbb{W}(\hat{\rho}_C) \leq 1 + \ln(K(1-\eta)+1). \tag{D.12}$$

To show that the expression in Eq. D.12 is also a lower bound for $\mathbb{W}(\hat{\rho}_C)$, we use Theorem 6 of [67], which states that for two probability distributions, $f(\mu)$ and $h(\mu)$ on $\mathbb{C}$, we have

$$W((f \star h)(\mu)) \geq \lambda W(f(\mu)) + (1-\lambda)W(h(\mu)) - \lambda \ln \lambda - (1-\lambda)\ln(1-\lambda) \tag{D.13}$$

for all $\lambda \in [0,1]$, where $f \star h$ is the convolution of f and h and where the Wehrl entropy of a probability distribution is found from Eq. 4.2 by replacing $Q(\mu)$ with the given probability distribution. Choosing

$$\begin{aligned} f(\mu) &\equiv \frac{1}{\eta} Q_{\hat{\rho}_A}\left(\frac{\mu}{\sqrt{\eta}}\right), \text{and} \\ h(\mu) &\equiv \frac{1}{1-\eta} Q_{\hat{\rho}_B}\left(\frac{\mu}{\sqrt{1-\eta}}\right), \end{aligned} \tag{D.14}$$

we get

$$W(\hat{\rho}_C) \geq \lambda(1+\ln\eta) + (1-\lambda)W\left(\frac{1}{1-\eta} Q_{\hat{\rho}_B}\left(\frac{\mu}{\sqrt{1-\eta}}\right)\right) - \lambda\ln\lambda - (1-\lambda)\ln(1-\lambda). \tag{D.15}$$

It is straightforward to show that the Wehrl entropy of a scaled distribution $(1/x)Q(\mu/\sqrt{x})$ is given by

$$W\left(\frac{1}{x} Q\left(\frac{\mu}{\sqrt{x}}\right)\right) = W\left(Q(\mu)\right) + \ln x, \tag{D.16}$$

for any $x \in \mathbb{R}$. From Equations D.16 and D.15, we obtain

$$
\begin{aligned}
W(\hat{\rho}_C) &\geq \lambda(1+\ln\eta) + (1-\lambda)\left(W(\hat{\rho}_B) + \ln(1-\eta)\right) \\
&\quad -\lambda\ln\lambda - (1-\lambda)\ln(1-\lambda) \\
&= \lambda(1+\ln\eta) + (1-\lambda)\left(1+\ln(K+1) + \ln(1-\eta)\right) \\
&\quad -\lambda\ln\lambda - (1-\lambda)\ln(1-\lambda) \\
&= 1 + \lambda\ln\left(\frac{\eta}{\lambda}\right) + (1-\lambda)\ln\left(\frac{(K+1)(1-\eta)}{(1-\lambda)}\right) \\
&= 1 + \ln(K(1-\eta)+1)
\end{aligned}
\tag{D.17}
$$

where the last equality uses $\lambda = \eta/(\eta + (K+1)(1-\eta)) \in [0,1], \forall \eta, K$. Therefore the minimum output Wehrl entropy of $\hat{c}$ must satisfy the lower bound

$$
\mathbb{W}(\hat{\rho}_C) \geq 1 + \ln(K(1-\eta)+1). \tag{D.18}
$$

The upper-bound (Eq. D.12) and the lower-bound (Eq. D.18) on the minimum output Wehrl entropy coincide, and thus we have the equality:

$$
\mathbb{W}(\hat{\rho}_C) = 1 + \ln(K(1-\eta)+1), \tag{D.19}
$$

which is achieved by a thermal-state $\hat{\rho}_B$ with mean photon number K (Eq. D.24), thus proving the conjecture for the minimum output Wehrl entropy.

D.2 Weak conjecture 3

The following single-mode version of conjecture 3 was stated in chapter 4:

Weak Conjecture 3 — *Let a lossless beam splitter have input $\hat{a}$ in a zero-mean thermal state with mean photon number N, input $\hat{b}$ in a zero-mean state with von Neumann entropy $S(\hat{\rho}^B) = g(K)$, and output $\hat{c}$ from its transmissivity-η port. Then the von Neumann entropy of output $\hat{c}$ is minimized when input $\hat{b}$ is in a thermal state with average photon number K, and the minimum output entropy is given by $\mathbb{S}(\hat{\rho}^C) = g(\eta N + (1-\eta)K)$.*

The following is an analogous statement of the conjecture for the Wehrl entropy:

Conjecture 3: Wehrl — *Let a lossless beam splitter have input $\hat{a}$ in a zero-mean thermal state with mean photon number N, input $\hat{b}$ in a zero-mean state with Wehrl entropy $W(\hat{\rho}^B) = 1 + \ln(K+1)$, and output $\hat{c}$ from its transmissivity-η port. Then the Wehrl entropy of output $\hat{c}$ is minimized when input $\hat{b}$ is in a thermal state with average photon number K, and the minimum output entropy is given by $\mathbb{W}(\hat{\rho}^C) = 1 + \ln(\eta N + (1-\eta)K + 1)$.*

Proof — Our proof of the Wehrl-entropy conjecture for the thermal-noise $\hat{\rho}^A$ parallels what we did for the vacuum-state $\hat{\rho}^A$. As before, we have that

$$\chi_A^{\hat{\rho}_C}(\zeta) = \chi_A^{\hat{\rho}_A}(\sqrt{\eta}\zeta)\chi_A^{\hat{\rho}_B}(\sqrt{1-\eta}\zeta) \tag{D.20}$$

Now, however, the input $\hat{a}$ is in a zero-mean thermal state with mean photon number N. Thus, its Husimi function and Wehrl entropy are given by:

$$Q_{\hat{\rho}_A}(\mu) = \frac{1}{\pi(N+1)} e^{-|\mu|^2/(N+1)}, \tag{D.21}$$

$$W(\hat{\rho}_A) = 1 + \ln(N+1). \tag{D.22}$$

From Eq. D.20, and the multiplication-convolution property of Fourier transforms (FT) we get

$$\begin{aligned} Q_{\hat{\rho}_C}(\mu) &= \frac{1}{\eta} Q_{\hat{\rho}_A}\left(\frac{\mu}{\sqrt{\eta}}\right) \star \frac{1}{(1-\eta)} Q_{\hat{\rho}_B}\left(\frac{\mu}{\sqrt{1-\eta}}\right) \\ &= \frac{1}{\pi\eta(N+1)} e^{-|\mu|^2/(\eta(N+1))} \star \frac{1}{(1-\eta)} Q_{\hat{\rho}_B}\left(\frac{\mu}{\sqrt{1-\eta}}\right). \end{aligned} \tag{D.23}$$

If the state of the input $\hat{b}$ is a thermal state with mean photon number K, i.e.,

$$\hat{\rho}_B = \frac{1}{\pi K} \int e^{-|\alpha|^2/K} |\alpha\rangle\langle\alpha| d^2\alpha,$$

we have

$$W(\hat{\rho}_B) = 1 + \ln(K+1), \tag{D.24}$$

which satisfies the hypothesis of our thermal-noise Wehrl-entropy conjecture. Using Eq. D.9, we can write out the Husimi function and the Wehrl entropy of the output $\hat{c}$:

$$\begin{aligned} Q_{\hat{\rho}_C}(\mu) &= \frac{1}{\pi(1+(1-\eta)K+\eta N)} e^{-|\mu|^2/(1+(1-\eta)K+\eta N)}, & \text{(D.25)} \\ W(\hat{\rho}_C) &= 1 + \ln(\eta N + K(1-\eta)+1), & \text{(D.26)} \end{aligned}$$

which gives us the upper bound

$$\mathbb{W}(\hat{\rho}_C) \leq 1 + \ln(\eta N + K(1-\eta)+1). \tag{D.27}$$

To show that the expression in Eq. D.12 is also a lower bound for $\mathbb{W}(\hat{\rho}_C)$, we use Eq. D.13, and definitions in Eq. D.15 to obtain:

$$W(\hat{\rho}_C) \geq \lambda(1+\ln(\eta(N+1)))+(1-\lambda)W\left(\frac{1}{1-\eta}Q_{\hat{\rho}_B}\left(\frac{\mu}{\sqrt{1-\eta}}\right)\right)-\lambda\ln\lambda-(1-\lambda)\ln(1-\lambda). \tag{D.28}$$

From equations D.16 and D.28, we find

$$\begin{aligned} W(\hat{\rho}_C) &\geq \lambda(1+\ln(\eta(N+1))) + (1-\lambda)\left(W(\hat{\rho}_B) + \ln(1-\eta)\right) & \text{(D.29)} \\ &\quad -\lambda\ln\lambda - (1-\lambda)\ln(1-\lambda) \\ &= \lambda(1+\ln(\eta(N+1))) + (1-\lambda)\left(1+\ln(K+1)+\ln(1-\eta)\right) \\ &\quad -\lambda\ln\lambda - (1-\lambda)\ln(1-\lambda) \\ &= 1 + \lambda\ln\left(\frac{\eta(N+1)}{\lambda}\right) + (1-\lambda)\ln\left(\frac{(K+1)(1-\eta)}{(1-\lambda)}\right) \\ &= 1 + \ln(\eta N + K(1-\eta)+1) \end{aligned}$$

where the last equality used $\lambda = \eta(N+1)/(\eta(N+1)+(K+1)(1-\eta)) \in [0,1], \forall \eta, K, N$. Therefore the minimum output Wehrl entropy of $\hat{c}$ must satisfy the lower bound

$$\mathbb{W}(\hat{\rho}_C) \geq 1 + \ln(\eta N + K(1-\eta) + 1). \tag{D.30}$$

The upper bound (Eq. D.27) and the lower bound (Eq. D.30) on the minimum output Wehrl entropy coincide, and thus we have the equality:

$$\mathbb{W}(\hat{\rho}_C) = 1 + \ln(\eta N + K(1-\eta) + 1). \tag{D.31}$$

which is achieved by a thermal-state $\hat{\rho}_B$ with mean photon number K (Equation D.24), thereby proving the thermal-noise Wehrl-entropy conjecture.

Bibliography

[1] Gagliardi, R. M. and Karp, S., *Optical Communications*, John Wiley & Sons, Inc. (1976).

[2] Shannon, C. E., "A mathematical theory of communications," *Bell System Technical Journal* **27**, 379 (part one), 623 (part two) (1948).

[3] Cover, T. M. and Thomas, J. A., *Elements of Information Theory*, John Wiley & Sons, Inc. (1991).

[4] Gallager, R. G., *Information Theory and Reliable Communication*, John Wiley & Sons, Inc. (1968).

[5] Holevo, A. S., "Coding theorems for quantum channels," `arXiv:quant-ph/9809023` v1 (1998).

[6] Nielsen, M. A. and Chuang, I. L., *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge (2000).

[7] Giovannetti, V., Guha, S., Lloyd, S., Maccone, L., Shapiro, J. H., and Yuen, H. P., "Classical capacity of the lossy bosonic channel: the exact solution," *Physical Review Letters* **92**, 027902 (2004).

[8] Martinez, A., "Spectral efficiency of optical direct detection," *J. Opt. Soc. Am. B* **24**, 735 (2007).

[9] Shapiro, J. H., Giovannetti, V., Guha, S., Lloyd, S., Maccone, L., and Yen, B. J., "Capacity of bosonic communications," in *Proceedings of the Seventh International Conference on Quantum Communication, Measurement and Computing*,

Barnett, S. M., Andersson, E., Jeffers, J., Ohberg, P., and Hirota, O., eds., AIP, Melville, NY (2004).

[10] Giovannetti, V., Guha, S., Lloyd, S., Maccone, L., and Shapiro, J. H., "Minimum output entropy of bosonic channels: a conjecture," *Physical Review A* **70**, 032315 (2004).

[11] Yen, B. J. and Shapiro, J. H., "Multiple-access bosonic communications," *Physical Review A* **72**, 062312 (2005).

[12] Guha, S., Shapiro, J. H., and Erkmen, B. I., "Classical capacity of bosonic broadcast communication and a minimum output entropy conjecture," *Physical Review A* **76**, 032303 (2007).

[13] Guha, S., Shapiro, J. H., and Erkmen, B. I., "Capacity of the bosonic wiretap channel and the entropy photon-number inequality," *Proceedings of International Symposium on Information Theory (ISIT)* (2008). `arXiv:quant-ph/0801.0841`.

[14] Louisell, W. H., *Quantum Statistical Properties of Radiation*, Wiley, New York (1973).

[15] Mandel, L. and E., W., *Optical Coherence and Quantum Optics*, Cambridge University Press, Cambridge (1995). Sections 10.1–10.3.

[16] Kingston, R. H., *Detection of Optical and Infrared Radiation*, Springer-Verlag, Berlin (1978).

[17] Gowar, J., *Optical Communication Systems*, Prentice Hall, Englewood Cliffs (1984).

[18] Yuen, H. P. and Shapiro, J. H., "Optical communication with two-photon coherent states—part III: quantum measurements realizable with photoemissive detectors," *IEEE Transactions on Information Theory* **26**, 78 (1980).

[19] Helstrom, C. W., *Quantum Detection and Estimation Theory*, Academic Press, New York (1976). Chapters 4, 6.

[20] Dolinar, S. J., "An optimum receiver for the binary coherent state quantum channel," tech. rep., M.I.T. Res. Lab. Electron. Quart. Prog. Rep. (1973).

[21] Shapiro, J. H., Yuen, H. P., and Machado Mata, J. A., "Optical communication with two-photon coherent states—part II: photoemissive detection and structured receiver performance," *IEEE Transactions on Information Theory* **25**, 179 (1979).

[22] Gordon, J. P., "Quantum effects in communications systems," *Proceedings of the Institute of Radio Engineers (IRE)* **50**, 1898 (1962).

[23] Davis, M. H. A., "Capacity and cutoff rate for poisson-type channels," *IEEE Transactions on Information Theory* **26**, 710 (1980).

[24] Pierce, J. R., Posner, E. C., and Rodemich, E. R., "The capacity of the photon-counting channel," *IEEE Transactions on Information Theory* **27**, 61 (1981).

[25] Wyner, A. D., "Capacity and error exponent for the direct detection photon channel–parts I and II," *IEEE Transactions on Information Theory* **34**, 1449 (1988).

[26] Shamai, S. and Lapidoth, A., "Bounds on the capacity of a spectrally constrained poisson channel," *IEEE Transactions on Information Theory* **39**, 19 (1993).

[27] Holevo, A. S., "The capacity of a quantum channel with general signal states," *IEEE Transactions on Information Theory* **44**, 269 (1998).

[28] Hausladen, P., Jozsa, R., Schumacher, B., Westmoreland, M., and Wootters, W. K., "Classical information capacity of a quantum channel," *Physical Review A* **54**, 1869 (1996).

[29] Schumacher, B. and Westmoreland, M. D., "Sending classical information via noisy quantum channels," *Physical Review A* **56**, 131 (1997).

[30] Yuen, H. P. and Ozawa, M., "Ultimate information carrying limit of quantum systems," *Physical Review Letters* **70**, 363 (1993).

[31] Caves, C. M. and Drummond, P. D., "Quantum limits on bosonic communication rates," *Review of Modern Physics* **66**, 481 (1994).

[32] Yuen, H. P. and Shapiro, J. H., "Optical communication with two-photon coherent states—part I: quantum state propagation and quantum noise reduction," *IEEE Transactions on Information Theory* **24**, 657 (1978).

[33] Caves, C. M., "Quantum limits on noise in linear amplifiers," *Physical Review D* **26**, 1817 (1982).

[34] Giovannetti, V., Lloyd, S., Maccone, L., Shapiro, J. H., and Yen, B. J., "Minimal rényi and wehrl entropies at the output of bosonic channels," *Physical Review A* **70**, 022328 (2004).

[35] Lapidoth, A. and Moser, S. M., "Bounds on the capacity of the discrete-time poisson channel," in [*Proceedings of the 41st Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL*], (2003). `http://www.isi.ee.ethz.ch/moser/publications.shtml`.

[36] Giovannetti, V. and Lloyd, S., "Additivity properties of a Gaussian channel," *Physical Review A* **69**, 062307 (2004).

[37] Tanaka, T., "A statistical-mechanics approach to large-system analysis of cdma multiuser detectors," *IEEE Transactions on Information Theory* **48**, 2888 (2002).

[38] Miller, R. R., "Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation," *IEEE Transactions on Signal Processing* **51**, 2821 (2003).

[39] Guha, S., *Classical capacity of the free-space quantum-optical channel*, Master's thesis, Massachusetts Institute of Technology (2004).

[40] Giovannetti, V., Guha, S., Lloyd, S., Maccone, L., Shapiro, J. H., Yen, B. J., and Yuen, H. P., "Classical capacity of free-space optical communication," *Quantum Information and Computation* **4**, 489 (2004).

[41] Slepian, D., "Prolate spheroidal wave functions, fourier analysis and uncertainty—iv: extensions to many dimensions; generalized prolate spheroidal functions," *Bell System Technical Journal* **43**, 3009 (1964).

[42] Slepian, D., "Analytic solution of two apodization problems," *Journal of the Optical Society of America* **55**, 1110 (1965).

[43] Shapiro, J. H., Guha, S., and Erkmen, B. I., "Ultimate channel capacity of free-space optical communications [invited]," *The Journal of Optical Networking: Special Issue* **4**, 501 (2005).

[44] Allen, L., Barnett, S. M., and Padgett, M. J., *Optical Angular Momentum*, Institute of Physics Publishing, Bristol (2004).

[45] Fujiwara, M., Takeoka, M., Mizuno, J., and Sasaki, M., "Exceeding classical capacity limit in quantum optical channel," *Physical Review A* **90**, 167906 (2003).

[46] Takeoka, M., M., F., Mizuno, J., and Sasaki, M., "Implementation of generalized quantum measurements: superadditive quantum coding, accessible information extraction, and classical capacity limit," *Physical Review A* **69**, 052329 (2004).

[47] Ishida, Y., Kato, K., and Sasaki, T. U., "Capacity of attenuated channel with discrete-valued input," *The Proceedings of the 8th International Conference on Quantum Communications, Measurement and Computing, Tsukuba, Japan* (2006).

[48] Cover, T. M., "Broadcast channels," *IEEE Transactions on Information Theory* **18**, 2 (1972).

[49] Bergmans, P., "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory* **19**, 197 (1973).

[50] Bergmans, P., "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Transactions on Information Theory* **20**, 279 (1974).

[51] Gallager, R. G., "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii (Problems of Information Transmission)* **16**(1), 17 (1980).

[52] Yard, J., Hayden, P., and Devetak, I., "Quantum broadcast channels," `arXiv:quant-ph/0603098` (2006).

[53] Jindal, N., Vishwanath, S., and Goldsmith, A., "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Transactions on Information Theory* **50**, 768 (2004).

[54] Borade, S., Zheng, L., and Trott, M., "Multilevel broadcast networks," *Proceedings of the IEEE International Symposium on Information Theory, Nice, France* (2007).

[55] Giovannetti, V., Guha, S., Lloyd, S., Maccone, L., Shapiro, J. H., Yen, B. J., and Yuen, H. P., *Quantum Information, Statistics, Probability*, Rinton Press, New Jersey (2004). Edited by Hirota, O.

[56] Weingarten, H., Steinberg, Y., and Shamai, S. S., "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory* **52** (2006).

[57] Wyner, A. D., "The wiretap channel," *Bell System Technical Journal* **54**, 1355 (1975).

[58] Csiszár, I. and Körner, J., "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory* **23**, 339 (1978).

[59] Devetak, I., "The private classical capacity and quantum capacity of a quantum channel," *IEEE Transactions on Information Theory* **51**, 44 (2005).

[60] Smith, G., "The private classical capacity with a symmetric side channel and its application to quantum cryptography," `arXiv:quant-ph/0705.3838` (2007).

[61] Wolf, M. M., Pérez-García, D., and Giedke, G., "Quantum capacities of bosonic channels," `arXiv:quant-ph/0606132` (2006).

[62] Shor, P. W., "Equivalence of additivity questions in quantum information theory," *Communications in Mathematical Physics* **246**, 473 (2004).

[63] Holevo, A. S. and Shirokov, M. E., "On shor's channel extension and constrained channels," `arXiv:quant-ph/0306196` (2003).

[64] Wehrl, A., "General properties of entropy," *Review of Modern Physics* **50**, 221 (1978).

[65] Yen, B., *Multiple-user Quantum Optical Communicaton*, PhD thesis, Massachusetts Institute of Technology (2004).

[66] Verdu, S. and Guo, D., "A simple proof of the entropy power inequality," *IEEE Transactions on Information Theory* **52**, 2165 (2006).

[67] Lieb, E. H., "Proof of an entropy conjecture of Wehrl," *Communications in Mathematical Physics* **62**, 35 (1978).

[68] Rioul, O., "Information theoretic proofs of entropy power inequalities," `arXiv:quant-ph/0704.1751` v1 (2007).

[69] Erkmen, B. I., *Phase-Sensitive Light: Coherence Theory and Applications to Optical Imaging*, PhD thesis, Massachusetts Institute of Technology (2008).

[70] Tan, S., Giovannetti, V., Guha, S., Erkmen, B. I., Lloyd, S., Maccone, L., Pirandola, S., and Shapiro, J. H., "Quantum illumination using Gaussian states," In preparation (2008).

[71] Artstein, S., Ball, K., Barthe, F., and Naor, A., "Solution of Shannon's problem on monotonicity of entropy," *Journal of American Mathematical Society* **17**, 975 (2004).

[72] Madiman, M. and Barron, A., "Generalized entropy power inequalities and monotonicity properties of information," *IEEE Transactions of Information Theory* **53**, 2317 (2007).

[73] Tulino, A. M. and Verdu, S., "Monotonic decrease of the non-gaussianness of the sum of independent random variables: A simple proof," *IEEE Transactions of Information Theory* **52**, 4295 (2006).

[74] Berrou, C., Glavieux, A., and Thitimajshima, P., "Near Shannon limit error-correcting coding and decoding: Turbo codes," *Proceedings of the IEEE International Conference on Communications, ICC, Geneva, Switzerland* , 1064 (1993).

[75] Shor, P. W., "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Journal on Computing* **26**, 1484 (1997).

[76] Shor, P. W., "The quantum channel capacity and coherent information," Lecture Notes, MSRI Workshop on Quantum Computation, San Francisco (2002).

[77] Lloyd, S., "Capacity of the noisy quantum channel," *Physical Review A* **55**, 1613 (1997).

[78] Weingarten, H., Steinberg, Y., and Shamai, S. S., "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Transactions on Information Theory* **52**, 3936 (2006).

[79] Garcia-Patron, R. and Cerf, N. J., "Unconditional optimality of gaussian attacks against continuous-variable qkd," *Physical Review Letters* **97**, 190503 (2006).

[80] Gottesman, D., Kitaev, A., and Preskill, J., "Encoding a qubit in an oscillator," *Physical Review A* **64**, 012310 (2001).

[81] Griffiths, D. J., *Introduction to Quantum Mechanics*, Prentice Hall; United States edition (1994). ISBN 0-13-124405-1.

[82] Sakurai, J. J., *Modern Quantum Mechanics*, Addison Wesley; 2 edition (1993). ISBN 0-20-153929-2.

[83] de Gosson, M., *Symplectic Geometry and Quantum Mechanics*, Birkhauser, Basel (2006). chapters 1, 2.

[84] Yuen, H. P., "Two-photon coherent states of the radiation field," *Physics Review A* **13**, 6 (1976).