

Network information theory for classical-quantum channels

Ivan Savov

School of Computer Science
McGill University, Montréal
July 2012

A thesis submitted to McGill University in partial fulfillment of the requirements of
the degree of Ph.D.

©Ivan Savov, 2012

Acknowledgements

This work would not have been possible without the continued support of my supervisor Patrick Hayden. He introduced me to many interesting mathematical research questions at the intersection of quantum physics and computer science. His outstanding abilities as a researcher, teacher and explainer have been an inspiration for me throughout the many years that I have known him. I am also very grateful to Mark M. Wilde for lending me his expertise on all aspects of quantum Shannon theory. I would like to thank Omar Fawzi, Pranab Sen, Mai Vu and Saikat Guha for the numerous discussions and their ability to distill complicated mathematical arguments into intuitive explanations. I want to thank Olivier Landon-Cardinal, Adriano Ferrari, Grant Salton and Benno Salwey for their help with the preparation of this manuscript. There are many other people who deserve an honourable mention and my gratitude for either directly or indirectly influencing me: Jan Florjanczyk, Andie Sigler, Eren Şaşoğlu, Gilles Brassard and Andreas Winter. I also want to thank my family for supporting me in my scientific endeavours.

Abstract

Network information theory is the study of communication problems involving multiple senders, multiple receivers and intermediate relay stations. The purpose of this thesis is to extend the main ideas of classical network information theory to the study of classical-quantum channels. We prove coding theorems for the following communication problems: quantum multiple access channels, quantum interference channels, quantum broadcast channels and quantum relay channels.

A quantum model for a communication channel describes more accurately the channel's ability to transmit information. By using physically faithful models for the channel outputs and the detection procedure, we obtain better communication rates than would be possible using a classical strategy. In this thesis, we are interested in the transmission of classical information, so we restrict our attention to the study of *classical-quantum* channels. These are channels with classical inputs and quantum outputs, and so the coding theorems we present will use classical encoding and quantum decoding.

We study the asymptotic regime where many copies of the channel are used in parallel, and the uses are assumed to be independent. In this context, we can exploit information-theoretic techniques to calculate the maximum rates for error-free communication for any channel, given the statistics of the noise on that channel. These theoretical bounds can be used as a benchmark to evaluate the rates achieved by practical communication protocols.

Most of the results in this thesis consider classical-quantum channels with finite dimensional output systems, which are analogous to classical discrete memoryless channels. In the last chapter, we will show some applications of our results to a practical optical communication scenario, in which the information is encoded in continuous quantum degrees of freedom, which are analogous to classical channels with Gaussian noise.

Résumé

La théorie de l'information multipartite étudie les problèmes de communication avec plusieurs émetteurs, plusieurs récepteurs et des stations relais. L'objectif de cette thèse est d'étendre les idées centrales de la théorie de l'information classique à l'étude des canaux quantiques. Nous allons nous intéresser aux scénarios de communication suivants: les canaux quantiques à accès multiples, les canaux quantiques à interférence, les canaux quantiques de diffusion et les canaux quantiques à relais. Dans chacun de ces scénarios de communication, nous caractérisons les taux de communication réalisables pour l'envoi d'information classique sur ces canaux quantiques.

La modélisation quantique des canaux de communication est importante car elle fournit une représentation plus précise de la capacité du canal à transmettre l'information. En utilisant des modèles physiquement réalistes pour les sorties du canal et la procédure de détection, nous obtenons de meilleurs taux de communication que ceux obtenus dans un modèle classique. En effet, l'utilisation de mesures quantiques collectives sur l'ensemble des systèmes physiques en sortie du canal permet une meilleure extraction d'information que des mesures indépendantes sur chaque sous-système. Nous avons choisi d'étudier les canaux à entrée classique et sortie quantique qui constituent une abstraction utile pour l'étude de canaux quantiques généraux où l'encodage est restreint au domaine classique.

Nous étudions le régime asymptotique où de nombreuses copies de du canal sont utilisées en parallèle, et les utilisations sont indépendantes. Dans ce contexte, il est possible de caractériser les limites absolues sur la transmission d'information d'un canal, si on connaît les statistiques du bruit sur ce canal. Ces résultats théoriques peuvent être utilisés comme un point de repère pour évaluer la performance des protocoles de communication pratiques.

Nous considérons surtout les canaux où les sorties sont des systèmes quantiques de dimension finie, analogues aux canaux classiques discrets. Le dernier chapitre présente des applications pratiques de nos résultats à la communication optique, où systèmes physiques auront des degrés de liberté continus. Ce contexte est analogue aux canaux classiques avec bruit gaussien.

Table of Contents

1	Introduction	1
1.1	Information theory	2
1.2	Network information theory	3
1.3	Quantum channels	4
1.4	Research contributions	5
1.5	Thesis overview	6
2	Background	9
2.1	Notation	9
2.2	Classical typicality	11
2.2.1	Typical sequences	11
2.2.2	Conditional typicality	12
2.2.3	Output-typical set	14
2.2.4	Joint typicality	15
2.3	Introduction to quantum information	15
2.3.1	Quantum states	16
2.3.2	Quantum channels	18
2.3.3	Quantum measurement	19
2.3.4	Quantum information theory	19
2.4	Quantum typicality	21
2.4.1	Quantum conditional typicality	22
2.5	Closing remarks	24
3	Point-to-point communication	25
3.1	Classical channel coding	25
3.1.1	Channel capacity	27
3.2	Quantum communication channels	29
3.2.1	Classical-quantum channel coding	31
3.3	Proof of HSW Theorem	34
3.4	Discussion	40
4	Multiple access channels	41
4.1	Introduction	41

4.1.1	Review of classical results	42
4.1.2	Quantum multiple access channels	44
4.1.3	Information processing task	46
4.1.4	Chapter overview	47
4.2	Successive decoding	48
4.3	Simultaneous decoding	52
4.3.1	Conjecture for three-sender simultaneous decoding	64
4.4	Rate-splitting	66
4.5	Example of a quantum multiple access channel	67
4.6	Discussion	69
5	Interference channels	71
5.1	Introduction	71
5.1.1	Applications	72
5.1.2	Review of classical results	73
5.1.3	Quantum interference channels	74
5.1.4	Information processing task	74
5.1.5	Chapter overview	77
5.2	Capacity results for special cases	78
5.2.1	Very strong interference case	78
5.2.2	Strong interference case	83
5.3	The quantum Han-Kobayashi rate region	85
5.4	The quantum Chong-Motani-Garg rate region	90
5.5	Quantum CMG rate region via two-sender simultaneous decoding	95
5.5.1	Geometry of the CMG rate region	98
5.5.2	Şaşoğlu argument	100
5.5.3	Two-message simultaneous decoding is sufficient for the rates of the facets a_i and c_i	101
5.6	Successive decoding strategies for interference channels	106
5.6.1	Time-sharing strategies	106
5.6.2	Split codebook strategies	107
5.7	Outer bound	110
5.8	Discussion	111
6	Broadcast channels	113
6.1	Introduction	114
6.1.1	Previous work	115
6.1.2	Quantum broadcast channels	115
6.1.3	Information processing task	116
6.1.4	Chapter overview	117
6.2	Superposition coding inner bound	117
6.3	Marton coding scheme	124
6.4	Discussion	128

7	Relay channels	131
7.1	Introduction	132
7.1.1	Classical relay channel coding strategies	133
7.1.2	Quantum relay channels	135
7.1.3	Chapter overview	135
7.2	Partial decode-and-forward strategy	136
7.3	Achievability proof	137
7.3.1	Decoding at the destination	139
7.3.2	Decoding at the relay	147
7.4	Discussion	150
8	Bosonic interference channels	153
8.1	Preliminaries	154
8.1.1	Gaussian channels	154
8.1.2	Introduction to quantum optics	156
8.1.3	Coherent states	158
8.2	Bosonic channels	158
8.2.1	Channel model	159
8.2.2	Encoding	160
8.2.3	Homodyne detection	161
8.2.4	Heterodyne detection	162
8.2.5	Joint detection	162
8.3	Free-space optical interference channels	163
8.3.1	Detection strategies	164
8.4	Very strong interference case	165
8.5	Strong interference case	167
8.6	Han-Kobayashi rate regions	168
8.7	Discussion	171
9	Conclusion	173
9.1	Summary	173
9.2	New results	174
A	Classical channel coding	177
A.1	Classical typicality	177
A.2	Classical packing lemma	179
B	Quantum channel coding	185
B.1	Quantum typicality	185
B.2	Quantum packing lemma	191
C	Miscellaneous proofs	195
C.1	Geometry of Chong-Motani-Garg rate region	195
C.2	Detailed explanation concerning moving points	198
C.3	Redundant inequality	201

Notation

CLASSICAL	QUANTUM
$y_a \in \mathcal{Y}$ symbol from a finite set	$ v\rangle^B \in \mathcal{H}^B$ vector in a Hilbert space
$p_Y \in \mathcal{P}(\mathcal{Y})$ probability distribution $p_Y(y) \geq 0, \forall y \in \mathcal{Y}$ $\sum_y p_Y(y) = 1$	$\rho^B \in \mathcal{D}(\mathcal{H}^B)$ density matrix \equiv quantum state $\langle v \rho^B v\rangle \geq 0, \forall v\rangle \in \mathcal{H}^B$ $\text{Tr}[\rho^B] = 1, (\rho^B)^\dagger = \rho^B$
$p_{Y X}$ conditional probability distribution \equiv classical-classical channel	$\{\rho_x^B\}, x \in \mathcal{X}$ conditional states \equiv classical-quantum channel
$p_{XY}(x, y) \equiv p_X(x)p_{Y X}(y x)$ joint input-output distribution	$\theta^{XB} \equiv \sum_x p_X(x) x\rangle\langle x ^X \otimes \rho_x^B$ joint input-output state
$p_{\bar{Y}} \equiv \mathbb{E}_X p_{Y X}$ average output distribution	$\bar{\rho}^B \equiv \mathbb{E}_X \rho_x^B$ average output state
$\mathbf{1}_{\{y^n \in \mathcal{T}_\delta^{(n)}(\bar{Y})\}}$ indicator function for the output-typical set	$\Pi_{\bar{\rho}} \equiv \Pi_{\bar{\rho}^{\otimes n}, \delta}^{B^n}$ projector onto the output-typical subspace
$\mathbf{1}_{\{y^n \in \mathcal{T}_\delta^{(n)}(Y x^n)\}}$ indicator function for the conditionally typical set	$\Pi_{x^n} \equiv \Pi_{\rho_{x^n}^B}^{B^n}$ conditionally typical projector for the state $\rho_{x^n}^{B^n}$

Chapter 1

Introduction

The central theme of this work is the transmission of information through noisy communication channels. The word *information* means different things to different people, so it is worthwhile to begin the discussion with a clear definition of the term. Statements like “Canada has an information-based economy” suggest that information is some kind of commodity that can be shipped on trains for export like oil or lumber. In the world of digital electronics, the word information is used as a synonym for the word *data* as in “How much information can you store on your USB memory stick?”. In that context, most people would say that a 7MB mp3 file contains just as much information as a 7MB file full of zeros.

In this work we will use the term *information* in the sense originally defined by Claude Shannon [Sha48]. Shannon realized that in order to study the problems of information storage and information transmission mathematically, we must step away from the *semantics* of the messages and focus on their *statistics*. Using the notions of entropy, conditional entropy and mutual information, we can *quantify* the information content of data sources and the information transmitting abilities of noisy communication channels.

We can arrive at an *operational* interpretation of the information content of a data source in terms of our ability to compress it. The more unpredictable the content of the data is, the more information it contains. Indeed, if we use WinZip to compress the mp3 file and the file full of zeros, we will see that the latter will result in a much smaller zip file, which is expected since a file full of zeros has less uncertainty and, by

extension, contains less information.

We can similarly give an operational interpretation of the information carrying capacity of a noisy communication channel in terms of our ability to convert it into a noiseless channel. Channels with more noise have a smaller capacity for carrying information. Consider a channel which allows us to send data at the rate of 1 MB/sec on which half of the packets sent get lost due to the effects of noise on the channel. It is not true that the capacity of such a channel is 1 MB/sec, because we also have to account for the need to retransmit lost packets. In order to correctly characterize the information carrying capacity of a channel, we must consider the rate of the end-to-end *protocol* which converts many uses of the noisy channel into an effectively noiseless communication channel.

1.1 Information theory

Information theory studies models of communication which are amenable to mathematical analysis. In order to model the effects of noise (ζ) in a point-to-point communication scenario, we represent the inputs and outputs of the channel probabilistically. We describe the channel as a triple $(\mathcal{X}, p_{Y|X}(y|x), \mathcal{Y})$, where \mathcal{X} is the set of possible

symbols that the Transmitter (Tx) can send, \mathcal{Y} is the set of possible outputs that the Receiver (Rx) can obtain and $p_{Y|X}(y|x)$ is a conditional probability distribution describing the channel's transition probabilities. This model is illustrated in Figure 1.1, where random variables are pictorially represented as small triangles (\blacktriangleright). For example, the noiseless binary channel is represented as the triple $(\{0, 1\}, p_{Y|X}(y|x) = \delta(x, y), \{0, 1\})$. Using this model of the channel, it is possible to calculate the optimal communication rates from Transmitter to Receiver in the limit of many independent uses of the channel [Sha48]. These theoretical results have wide-reaching applications in many areas of communication engineering but also in other fields like cryptography, computer science, neuroscience and even economics. So long as a probabilistic model for the channel at hand is available, we can use this model and the techniques of information theory to arrive at precise mathematical statements about its suitability for a given communication task in the limit of many uses of the channel.

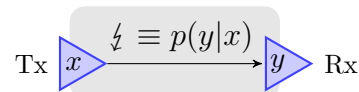


Figure 1.1: A point-to-point channel $\equiv p_{Y|X}(y|x)$.

1.2 Network information theory

Network information theory is the extension of Shannon’s model of noisy channels to communication scenarios with multiple senders and multiple receivers [EGC80, CT91, EGK10]. To model these channels probabilistically, we use multivariate conditional probability distributions. Some of the most important problems in network information theory are shown in Figure 1.2, and the relevant class of probability distributions is also indicated.

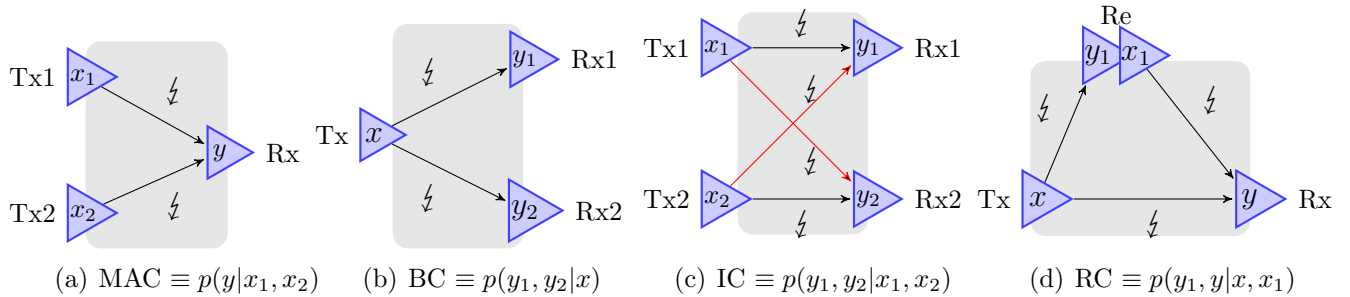


Figure 1.2: Classical network information theory studies communication channels with multiple senders and multiple receivers. These include, among others, (a) multiple access channels (MACs), (b) broadcast channels (BCs), (c) interference channels (ICs), and (d) relay channels (RCs).

Each of the above channels is a model for some practical communication scenario. In the multiple access channel, there are multiple transmitters trying to talk to a single base station, and we can describe the tradeoff between the communication rates that are achievable for the inbound communication links. The broadcast channel is the dual problem in which a single transmit antenna emits multiple information streams intended for different receivers. We can additionally have a *common* information stream intended for both receivers. Coding strategies for broadcast channels involve encodings that can “mix” the information streams to produce the transmit signal. Interference channels model situations where multiple independent transmissions are intended, but *crosstalk* occurs because the communication takes place in a shared medium. The relay channel is a *multi-hop* information network. The Relay is assumed to decode the message during one block of uses of the channel and re-transmit the information it has decoded during the next block. This allows the Receiver to collectively decode the information from both the Transmitter and the Relay and achieve better communication rates than what would be possible with point-to-point codes.

1.3 Quantum channels

Classical models are not adequate for the characterization of the information carrying capacity of communication systems in which the information carriers are quantum systems. Such systems need not be exotic: in optical communication links, the carriers are photons, which are properly described by quantum electrodynamics and only approximately described by Maxwell's equations. A more general model for communication channels is one which takes into account the underlying laws of physics concerning the encoding, transmission and decoding of information using quantum systems. Quantum decoding based on *collective* measurements of all the channel outputs in parallel can be shown to achieve higher communication rates compared to classical decoding strategies in which the channel outputs are measured individually.

Of particular interest are *classical-quantum* channel models, which model the sender's inputs as classical variables and the receiver's outputs as quantum systems. A classical-quantum channel $(\mathcal{X}, \mathcal{N}^{X \rightarrow B}(x) \equiv \rho_x^B, \mathcal{H}^B)$ is fully specified by the finite set of output states $\{\rho_x^B\}$ it produces for each of the possible inputs $x \in \mathcal{X}$. Figure 1.3 depicts a classical-quantum channel, in which the quantum output system is represented by a circle: \circ . Such channels form a useful abstraction for studying the transmission of classical data over quantum channels. The Holevo-Schumacher-Westmoreland (HSW) Theorem (see page 32) establishes the maximum achievable communication rates for classical-quantum channels [Hol98, SW97].

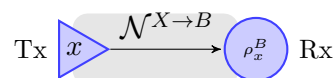


Figure 1.3: A point-to-point classical-quantum channel $\{\rho_x\}$.

Note that a classical-quantum (c - q) channel corresponds to the use of a quantum-quantum (q - q) channel in which the sender is restricted to selecting from a finite set of signalling states. Any code construction for a c - q channel can be augmented with an optimization over the choice of signal states to obtain a code for a q - q channel. For this reason, we restrict our study here to that of c - q channels.

The study of quantum channels finds practical applications in optical communications. Bosonic channels model the quantum aspects of optical communication links. It is known that optical receivers based on collective quantum measurements of the channel outputs outperform classical strategies, particularly in the low-photon-number regime [GGL⁺04, Guh11, WGTL12]. In other words, quantum measurements are *nec-*

essary to achieve their ultimate information carrying capacity. In [GGL⁺04] it is also demonstrated that classical encoding is *sufficient* to achieve the Holevo capacity of the lossy bosonic channel, giving further motivation for the theoretical study of classical-quantum models.

1.4 Research contributions

This thesis presents a collection of results for problems in network information theory for classical-quantum channels. As we stated before, the results here easily extend to quantum-quantum channels. The problems considered are illustrated in Figure 1.4.

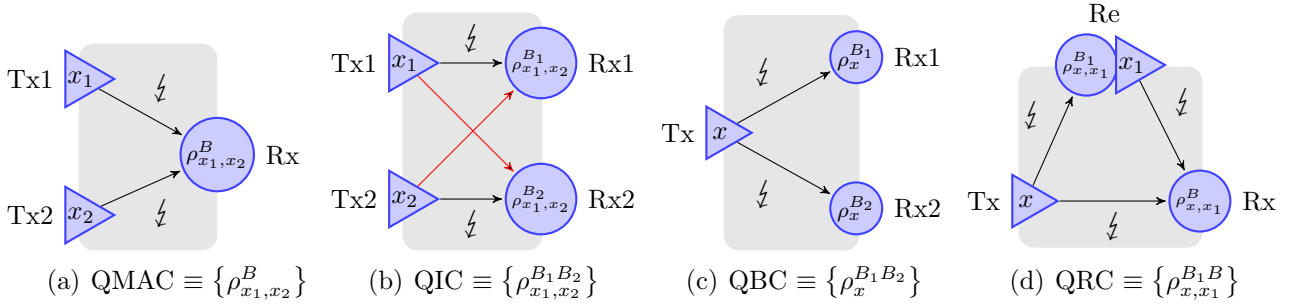


Figure 1.4: Network information theory can be extended to channels with quantum outputs. We call these “classical-quantum channels,” and consider the following communication scenarios: (a) multiple access channels (QMACs), (b) interference channels (QICs), (c) broadcast channels (QBCs), and (d) relay channels (QRCs).

Most of the results presented in this thesis have appeared in publication. The new results on the quantum multiple access channel and the quantum interference channel appeared in [FHS⁺12], which is a collaboration between Omar Fawzi, Patrick Hayden, Pranab Sen, Mark M. Wilde and the present author. That paper has been accepted for publication in the *IEEE Transactions on Information Theory*. A more compact version of the same results was presented by the author at the 2011 *Allerton conference* [FHS⁺11]. A follow-up paper on the bosonic quantum interference channel was presented by the author at the 2011 *International Symposium on Information Theory*, thanks to a collaboration with Saikat Guha and Mark M. Wilde [GSW11]. A further collaboration with Mark M. Wilde led to the publication of [SW12], which describes two coding strategies for the quantum broadcast channel. Finally, a collaboration with Mark M. Wilde and Mai Vu led to the development of the coding strategy for the

quantum relay channel presented in [SWV12]. The last two papers have been accepted for presentation at the 2012 *International Symposium on Information Theory*.

Our aim has been to present a comprehensive collection of the state-of-the-art of current knowledge in quantum network information theory analogous to the review paper by Cover and El Gamal [EGC80]. Indeed, the current work contains the classical-quantum extension of many of the results presented in that paper. Towards this aim, we have chosen to include in the text the statement of several important results by others. These include a proof of the capacity theorems of the point-to-point c-q channels different from the original ones due to Holevo, Schumacher and Westmoreland [Hol98, SW97] and the capacity result for quantum multiple access channel, originally due to Winter [Win01]. We will also present an alternate achievability proof of the quantum Chong-Motani-Garg rate region for the QIC, which was originally proved by Sen [Sen12a].

1.5 Thesis overview

Each of the communication problems covered in this thesis is presented in a separate chapter, and each chapter is organized in the same manner. The exposition in each chapter is roughly self-contained, but the ideas developed in Chapter 4 are of key importance to all other results in the thesis. Chapters 3 through 7 present results on classical-quantum ($c-q$) channels where the output systems are arbitrary quantum states in finite dimensional Hilbert spaces. This class of channels generalizes the class of classical discrete memoryless channels. The last chapter, Chapter 8, introduces the basic notions of quantum optics and studies bosonic quantum channels, for which the output system is a quantum system with continuous degrees of freedom.

Necessary background material on the notion of a classical typical set and its quantum analogue, the quantum typical subspace, is presented in Chapter 2. A more detailed discussion about typicality is presented in the appendix. Appendix A.1 concerns classical typical sets whereas Appendix B.1 reviews the properties of quantum typical subspaces, and quantum typical projectors. Of particular importance are conditionally typical projectors, which are used throughout the proofs in this work.

Our exploration of the classical-quantum world of communication channels begins in Chapter 3, where we discuss classical and classical-quantum models of point-to-point

communication. We will state and prove the capacity result for each class of channels: Shannon’s classical channel coding theorem (Theorem 3.1) and the Holevo-Schumacher-Westmoreland theorem (Theorem 3.2) concerning the capacity of the classical-quantum channel.

Chapter 4 presents results on the quantum multiple access channel (QMAC) and discusses the different coding strategies that can be employed. The capacity of the QMAC was established by Winter in [Win01] (Theorem 4.1) using a *successive decoding* strategy. Our contribution to the quantum multiple access channel problem is Theorem 4.2, which shows that the two-sender *simultaneous* decoding is possible. This result and the proof techniques used therein will form key building blocks for the results in subsequent chapters. The proof of Theorem 4.2 is the result of longstanding collaboration within our research group.

Chapter 5 will present results on quantum interference channels. These include the calculation of the capacity region for the quantum interference channel in two special cases and a description of the quantum Han-Kobayashi rate region [FHS⁺11, FHS⁺12]. In that chapter, we also provide an alternate proof of the achievability of the quantum Chong-Motani-Garg rate region, which was first established by Sen in [Sen12a]. This new proof is original to this thesis.

Chapter 6 is dedicated to the quantum broadcast channel problem. We prove two theorems: the superposition coding inner bound (Theorem 6.1), which was first proved in [YHD11] using a different approach, and the Marton inner bound with no common message (Theorem 6.2).

In Chapter 7, we will present Theorem 7.1 which is a proof of the *partial decode-and-forward* inner bound for the quantum relay channel. The *decode-and-forward* and *direct coding* strategies for the quantum relay channel are also established, since they are special cases of the more general Theorem 7.1.

Chapter 8 discusses the free-space optical communication interference channel in the presence of background thermal noise. This is a model for the crosstalk between two optical communication links. This chapter demonstrates the practical aspect of the ideas developed in this thesis.

We conclude with Chapter 9 wherein we state open problems and describe avenues for future research.



Chapter 2

Background

In this chapter we present all the necessary background material which is essential to the results presented in subsequent chapters.

2.1 Notation

We will denote the set $\{1, 2, \dots, n\}$ as $[1 : n]$ or with the shorthand $[n]$. A random variable X , defined over a finite set \mathcal{X} , is associated with a probability distribution $p_X(x) \equiv \Pr\{X = x\}$, where the lowercase x is used to denote particular values of the random variable. Furthermore, let $\mathcal{P}(\mathcal{X})$ denote the set of all probability mass functions on the finite set \mathcal{X} . Conditional probability distributions will be denoted as $p_{Y|X}(y|x)$ or simply $p_{Y|X}$.

In order to help distinguish between the classical systems (random variables) and the quantum systems in the equations, we use the following naming conventions. Classical random variables will be denoted by letters near to the end of the alphabet (U , W , X_1 , X_2) and denoted as small triangles, , in the diagrams of this thesis. The triangular shape was chosen in analogy to the 2-simplex $\equiv \mathcal{P}(\{1, 2, 3\})$. Quantum systems will be named with letters near the beginning of the alphabet (A , B_1 , B_2) and represented by circles, , in diagrams. The circular shape is chosen in analogy with the Bloch sphere [LS11].

Consider a communication scenario with one or more senders (female) and one or more receivers (male). In diagrams, a sender is denoted Tx (short for Transmitter)

2.2 Notation

and is associated with a random variable X . If there are multiple senders, then each of them will be referred to as Sender k and associated with a random variable X_k . Receivers will be denoted as Rx 1, Rx 2 and each is associated with a different output of the channel. The outputs of a *classical* channel will be denoted as Y_1, Y_2 , and the outputs of a *quantum* channel will be denoted as ρ^{B_1}, ρ^{B_2} .

The purpose of a communication protocol is to transfer bits of information from sender to receiver noiselessly. In this respect, the noiseless binary channel from sender to receiver is the standard unit resource for this task:

$$(\mathcal{X} = \{0, 1\}, p_{Y|X}(y|x) = \delta(x, y), \mathcal{Y} = \{0, 1\}) \equiv [c \rightarrow c], \quad (2.1)$$

where we have also defined the more compact notation $[c \rightarrow c]$. We will use $[c \rightarrow c]$ to denote the *communication resource* of being able to send one bit of classical information from the sender to the receiver [DHW08]. The square brackets indicate that the resource is noiseless. In order to describe multiuser communication scenarios, we extend this notation with superscripts indicating the sender and the receiver. Thus, in order to denote the noiseless classical communication of one bit from Sender k to Receiver z we will use the notation $[c^k \rightarrow c^z]$. The communication resource which corresponds to the sender being able to *broadcast* a message to Receiver 1 and Receiver 2 is denoted as $[c \rightarrow c^1 c^2]$. All the coding theorems presented in this work are protocols for converting many copies of some noisy channel resource into noiseless classical communication between a particular sender and a particular receiver as described above.

Codebooks $\{x^n(m)\}_{m \in \mathcal{M}}$ are lookup tables for codewords representing a discrete set of messages $\mathcal{M} = \{1, 2, 3, \dots, |\mathcal{M}|\}$ that could be transmitted. A communication rate R is a real number which describes our asymptotic ability to construct codes for a certain communication task. We will use the notation $|\mathcal{M}| = 2^{nR}$, and $\mathcal{M} = \{1, 2, 3, \dots, |\mathcal{M}|\} \equiv [1 : 2^{nR}]$, in which 2^{nR} should be interpreted to indicate $\lfloor 2^{nR} \rfloor$.

Let $\mathbb{R}_+^n \equiv \{\vec{v} \in \mathbb{R}^n \mid v_i \geq 0, \forall i \in [1 : n]\}$ be the non-negative subset of \mathbb{R}^n . We will denote a *rate region* as $\mathcal{R} \subseteq \mathbb{R}_+^n$ and the boundaries of regions as $\partial\mathcal{R}$. We denote points as $P \in \mathbb{R}^n$ and denote the *convex hull* of a set of points $\{P_i\}$ as $\text{conv}(\{P_i\})$.

2.2 Classical typicality

We present here a number of properties of typical sequences [CT91].

2.2.1 Typical sequences

Consider the random variable X with probability distribution $p_X(x)$ defined on a finite set \mathcal{X} . Denote by $|\mathcal{X}|$ the cardinality of \mathcal{X} . Let $H(X) \equiv H(p_X) \equiv -\sum_x p_X(x) \log_2 p_X(x)$ be the Shannon entropy of p_X , and it is measured in units of *bits*. The binary entropy function is denoted $H_2(p_0) \equiv -p_0 \log_2(p_0) - (1 - p_0) \log_2(1 - p_0) \equiv H_2(p_1)$, where $p_0 \equiv p_X(0)$ and $p_1 \equiv 1 - p_0$.

Denote by x^n a sequence $x_1 x_2 \dots x_n$, where each $x_i, i \in [n]$ belongs to the finite *alphabet* \mathcal{X} . To avoid confusion, we use $i \in [1 : n]$ to denote the index of a symbol x in the sequence x^n and $a \in [1, 2, \dots, |\mathcal{X}|]$ to denote the different symbols in the alphabet \mathcal{X} .

Define the probability distribution $p_{X^n}(x^n)$ on \mathcal{X}^n to be the n -fold product of p_X : $p_{X^n}(x^n) \equiv \prod_{i=1}^n p_X(x_i)$. The sequence x^n is drawn from p_{X^n} if and only if each letter x_i is drawn independently from p_X . For any $\delta > 0$, define the set of entropy δ -typical sequences of length n as:

$$\mathcal{T}_\delta^n(X) \equiv \left\{ x^n \in \mathcal{X}^n : \left| -\frac{\log p_{X^n}(x^n)}{n} - H(X) \right| \leq \delta \right\}. \quad (2.2)$$

Typical sequences enjoy many useful properties [CT91]. For any $\epsilon, \delta > 0$, and sufficiently large n , we have

$$\sum_{x^n \in \mathcal{T}_\delta^n(X)} p_{X^n}(x^n) \geq 1 - \epsilon, \quad (2.3)$$

$$2^{-n[H(X)+\delta]} \leq p_{X^n}(x^n) \leq 2^{-n[H(X)-\delta]} \quad \forall x^n \in \mathcal{T}_\delta^n(X), \quad (2.4)$$

$$[1 - \epsilon] 2^{n[H(X)-\delta]} \leq |\mathcal{T}_\delta^n(X)| \leq 2^{n[H(X)+\delta]}. \quad (2.5)$$

Property (2.3) indicates that a sequence X^n of random variables distributed according to $p_{X^n} = \prod^n p_X$ (identical and independently distributed), is very likely to be typical, since all but ϵ of the weight of the probability mass function is concentrated

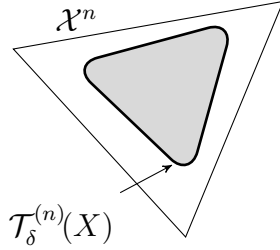


Figure 2.1: The typical set. Property (2.3) implies that draws of a random sequence $X^n \sim p_{X^n} \equiv \prod^n p_X$ are likely to fall inside the typical set $\mathcal{T}_\delta^{(n)}(X) \subset \mathcal{X}^n$ with high probability. If draws from $X^n \sim \prod^n p_X$ are represented as points, then after many draws the typical set will become darker as the shaded region in the diagram. The probability mass density on $\mathcal{T}_\delta^{(n)}(X)$ is approximately uniform: it varies between $2^{-n[H(X)+\delta]}$ and $2^{-n[H(X)-\delta]}$ (Property (2.4)), and the size of the shaded area will be at most $2^{n[H(X)+\delta]}$ (Property (2.5)). The non-typical set, $\mathcal{X}^n \setminus \mathcal{T}_\delta^{(n)}(X)$, will have at most ϵ weight in it (Property (2.3)).

on the typical set, which follows from the law of large numbers. Property (2.4) follows from the definition of the typical set (2.2). The lower bound on the probability of the typical sequences from (2.4) can be used to obtain an upper bound on the size of the typical set in (2.5). Similarly the upper bound from (2.4) and equation (2.3) can be combined to give the lower bound on the typical set in (2.5).

2.2.2 Conditional typicality

Consider now the conditional probability distribution $p_{Y|X}(y|x)$ associated with a communication channel. The induced joint input-output distribution is $(X, Y) \sim p_X(x)p_{Y|X}(y|x)$, when $p_X(x)$ is used as the input distribution.

The conditional entropy $H(Y|X)$ for this distribution is

$$H(Y|X) = H(X, Y) - H(X) = \sum_{x_a \in \mathcal{X}} p_X(x_a) H(Y|x_a). \quad (2.6)$$

where $H(Y|x_a) = -\sum_y p_{Y|X}(y|x_a) \log p_{Y|X}(y|x_a)$.

We define the x^n -conditionally typical set $\mathcal{T}_\delta^{(n)}(Y|x^n) \subseteq \mathcal{Y}^n$ to consist of all sequences y^n which are typically output when the input to the channel is x^n :

$$\mathcal{T}_\delta^{(n)}(Y|x^n) \equiv \left\{ y^n \in \mathcal{Y}^n : \left| -\frac{\log p_{Y^n|X^n}(y^n|x^n)}{n} - H(Y|X) \right| \leq \delta \right\}, \quad (2.7)$$

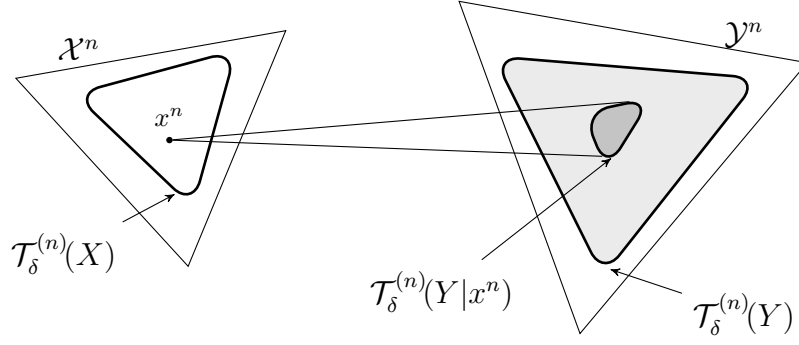


Figure 2.2: Illustration of the conditionally typical set $\mathcal{T}_\delta^{(n)}(Y|x^n)$ and the output-typical set $\mathcal{T}_\delta^{(n)}(Y)$. The “density” of $\mathcal{T}_\delta^{(n)}(Y)$, the lightly shaded area, is at least $2^{-n[H(Y)+\delta]}$, and the size of $\mathcal{T}_\delta^{(n)}(Y)$ is at most $2^{n[H(Y)+\delta]}$. The size of $\mathcal{T}_\delta^{(n)}(Y|x^n)$, the darker shaded region, is no greater than $2^{n[H(Y|X)+\delta]}$ for an x^n picked on average.

with $p_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n p_{Y|X}(y_i|x_i)$. The definition in (2.7) can be rewritten as:

$$2^{-n[H(Y|X)+\delta]} \leq p_{Y^n|X^n}(y^n|x^n) \leq 2^{-n[H(Y|X)-\delta]}, \quad \forall y^n \in \mathcal{T}_\delta^{(n)}(Y|x^n), \quad (2.8)$$

for any sequence x^n .

Suppose that a random input sequence $X^n \sim p_{X^n} = \prod^n p_X$ is passed through the channel $p_{Y^n|X^n}$. Then a conditionally typical sequence is likely to occur. More precisely, we have that for any $\epsilon, \delta > 0$, and sufficiently large n the statement is true under the expectation over the input sequence X^n :

$$\begin{aligned} \mathbb{E}_{X^n} \sum_{y^n \in \mathcal{T}_\delta^{(n)}(Y|X^n)} p_{Y^n|X^n}(y^n|X^n) &= \sum_{x^n} p_{X^n}(x^n) \sum_{y^n \in \mathcal{T}_\delta^{(n)}(Y|x^n)} p_{Y^n|x^n}(y^n|x^n) \\ &\geq 1 - \epsilon. \end{aligned} \quad (2.9)$$

We also have the following bounds on the expected size of the conditionally typical set:

$$[1 - \epsilon]2^{n[H(Y|X)-\delta]} \leq \mathbb{E}_{X^n} \left| \mathcal{T}_\delta^{(n)}(Y|X^n) \right| \leq 2^{n[H(Y|X)+\delta]}. \quad (2.10)$$

2.2.3 Output-typical set

Consider the distribution over symbols $y \in \mathcal{Y}$ induced by the channel $\mathcal{N} \equiv p_{Y|X}(y|x)$ whenever the input distribution is $p_X(x)$:

$$p_Y(y) \equiv \sum_x p_{Y|X}(y|x)p(x) = \mathbb{E}_X \mathcal{N}. \quad (2.11)$$

We define the output typical set as

$$\mathcal{T}_\delta^{(n)}(Y) \equiv \left\{ y^n \in \mathcal{Y}^n : \left| -\frac{\log p_{Y^n}(y^n)}{n} - H(Y) \right| \leq \delta \right\}, \quad (2.12)$$

where $p_{Y^n} = \prod^n p_Y$. Note that the output-typical set is just a special case of the general typical set shown in (2.2). The terminology *output-typical* is introduced to help with the exposition.

When the input sequences are chosen according to $X^n \sim p_{X^n} = \prod^n p_X$, then output sequences are likely to be output-typical:

$$\mathbb{E}_{X^n} \sum_{y^n \in \mathcal{T}_\delta^{(n)}(Y)} p_{Y^n|X^n}(y^n|X^n) \geq 1 - \epsilon. \quad (2.13)$$

An illustration and an intuitive interpretation of (2.9), (2.10) and (2.13) is presented in Figure 2.2. The expression in (2.9) for the property of the conditionally typical set $\mathcal{T}_\delta^{(n)}(Y|x^n)$ is the analogue of the typical property (2.3) for $\mathcal{T}_\delta^{(n)}(X)$. The interpretation is that the codewords of a random codebook are likely to produce output sequences that fall within their conditionally typical sets. This property will be used throughout this thesis to guarantee that the decoding strategies based on conditionally typical sets correctly recognize the channel outputs. On the other hand, (2.10) gives us both an upper bound and a lower bound on the size of the conditionally typical set for a random codebook. Finally, Property (2.13) tells us that the outputs of the channel which are not output-typical are not likely.

2.2.4 Joint typicality

Consider now the joint probability distribution $p_{XY}(x, y) \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$. Let (X^n, Y^n) be a pair of random variables distributed according to the product distribution $\prod^n p_{XY}$.

We define the jointly typical set $\mathcal{J}_\delta^{(n)}(X, Y) \subseteq \mathcal{X}^n \times \mathcal{Y}^n$ to be the set of sequences that are typical with respect to the joint probability distribution p_{XY} and with respect to the marginals p_X and p_Y .

$$\mathcal{J}_\delta^{(n)}(X, Y) \equiv \left\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \left| \begin{array}{l} x^n \in \mathcal{T}_\delta^{(n)}(X) \\ y^n \in \mathcal{T}_\delta^{(n)}(Y) \\ (x^n, y^n) \in \mathcal{T}_\delta^{(n)}(X, Y) \end{array} \right. \right\}. \quad (2.14)$$

A multi-variable sequence, therefore, is jointly typical if and only if all the sequences in the subsets of the variables are jointly typical.

The probability that two random sequences drawn from the marginals $X^n \sim \prod^n p_X$ and $Y^n \sim \prod^n p_Y$ are jointly typical can be bounded from above by $2^{-n[I(X;Y)-\delta]}$. This is straightforward to see from the definition in (2.14) and the properties of typical sets. If (x^n, y^n) is such that $x^n \in \mathcal{T}_{\delta'}^{(n)}(X)$ and $y^n \in \mathcal{T}_{\delta''}^{(n)}(Y)$ then $p_{X^n}(x^n) \leq 2^{-n[H(X)-\delta']}$ and $p_{Y^n}(y^n) \leq 2^{-n[H(Y)-\delta']}$. On the other hand, we know that the number of sequences that are typical according to the joint distribution is no larger than $2^{n[H(XY)+\delta']}$. Combining these two observations we get:

$$\begin{aligned} \sum_{(x^n, y^n) \in \mathcal{T}_{\delta''}^{(n)}(X, Y)} p_{X^n}(x^n) p_{Y^n}(y^n) &\leq \left| \mathcal{T}_{\delta''}^{(n)}(X, Y) \right| 2^{-n[H(X)-\delta']} 2^{-n[H(Y)-\delta']} \\ &\leq 2^{n[H(XY)+\delta']} 2^{-n[H(X)-\delta']} 2^{-n[H(Y)-\delta']} \\ &= 2^{-n[I(X;Y)-\delta]}. \end{aligned} \quad (2.15)$$

Note that the parameter $\delta = 2\delta' + \delta''$ is a function of our choice of typicality parameters for the typical sets.

2.3 Introduction to quantum information

The use of quantum systems for information processing tasks is no more mysterious than the use of digital technology for information processing. The use of an *analog*

to *digital converter* (ADC) to transform an analog signal to a digital representation and the use of a *digital to analog converter* (DAC) to transform from the digital world back into the analog world are similar to the *state preparation* and the *measurement* steps used in quantum information science. The *digital world* is sought after because of the computational, storage and communication benefits associated with manipulation of discrete systems instead of continuous signals. Similarly, there are benefits associated with using the *quantum world* (Hilbert space) in certain computation problems [Sho94, Sho95]. The use of digital and quantum technology can therefore both be seen operationally as a black box process with information encoding, processing and readout steps.

The focus of this thesis is the study of *quantum* aspects of communication which are relevant for *classical communication* tasks. In order to make the presentation more self-contained, we will present below a brief introduction to the subject which describes how quantum systems are represented, how information can be encoded and how information can be read out.

2.3.1 Quantum states

In order to describe the *state* of a quantum system B we use a density matrix ρ^B acting on a d -dimensional complex vector space \mathcal{H}^B (Hilbert space). To be a density matrix, the operator ρ^B has to be Hermitian, positive semidefinite and have unit trace. We denote the set of density matrices on a Hilbert space \mathcal{H}^B as $\mathcal{D}(\mathcal{H}^B)$.

A common choice of basis for \mathcal{H}^B is the standard basis $\{|0\rangle, |1\rangle, \dots, |d-1\rangle\}$:

$$|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad |1\rangle \equiv \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad |d-1\rangle \equiv \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad (2.16)$$

which is also known as the *computational* basis.

In two dimensions, another common basis is the *Hadamard* basis:

$$|+\rangle \equiv \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle, \quad (2.17)$$

$$|-\rangle \equiv \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle. \quad (2.18)$$

The eigen-decomposition of the density matrix ρ^B gives us another choice of basis in which to represent the state. Any density matrix can be written in the form:

$$\rho^B \equiv \begin{bmatrix} |e_{\rho;1}\rangle & |e_{\rho;2}\rangle & \cdots & |e_{\rho;d}\rangle \end{bmatrix} \begin{bmatrix} \lambda_{\rho;1} & 0 & \cdots & 0 \\ 0 & \lambda_{\rho;2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{\rho;d} \end{bmatrix} \begin{bmatrix} \langle e_{\rho;1}| \\ \langle e_{\rho;2}| \\ \vdots \\ \langle e_{\rho;d}| \end{bmatrix}, \quad (2.19)$$

where the eigenvalues $\lambda_{\rho;i}$ are all real and nonnegative. In our notation, column vectors are denoted as *kets* $|e_{\rho;i}\rangle$ and the dual (Hermitian conjugate) of a ket is the *bra*: $\langle e_{\rho;i}| \equiv |e_{\rho;i}\rangle^\dagger$ (a row vector). We say that ρ^B is a *pure state* if it has only a single non-zero eigenvalue: $\lambda_{\rho;1} = 1$, $\lambda_{\rho;i} = 0$, $\forall i > 1$.

Because the density matrix is positive semidefinite and has unit trace ($\sum_i \lambda_{\rho;i} = 1$), we can identify the eigenvalues of ρ^B with a probability distribution: $p_Y(y) \equiv \lambda_{\rho;y}$. A density matrix, therefore, corresponds to the probability distribution $p_Y(y)$ over the subspaces: $|e_{\rho;y}\rangle\langle e_{\rho;y}|$. This property will be important when we want to define the typical subspace for the tensor product state: $(\rho^B)^{\otimes n} \equiv \rho^{B_1} \otimes \rho^{B_2} \otimes \cdots \otimes \rho^{B_n}$.

Suppose that we have a two-party quantum state ρ^{AB} such that Alice has the subsystem A and Bob has the subsystem B . The state in Alice's lab is described by $\rho^A = \text{Tr}_B[\rho^{AB}]$, where Tr_B denotes a partial trace over Bob's degrees of freedom.

In order to describe the “distance” between two quantum states, we use the notion of *trace distance*. The trace distance between states σ and ρ is $\|\sigma - \rho\|_1 = \text{Tr}|\sigma - \rho|$, where $|X| = \sqrt{X^\dagger X}$. Two states that are similar have trace distance close to zero, whereas states that are perfectly distinguishable have trace distance equal to two.

Two quantum states can “substitute” for one another up to a penalty proportional to the trace distance between them:

Lemma 2.1. *Let $0 \leq \rho, \sigma, \Lambda \leq I$. Then*

$$\mathrm{Tr} [\Lambda \rho] \leq \mathrm{Tr} [\Lambda \sigma] + \|\rho - \sigma\|_1. \quad (2.20)$$

Proof. This follows from a variational characterization of trace distance as the distinguishability of the states under an optimal measurement operator M :

$$\begin{aligned} \|\rho - \sigma\|_1 &\equiv 2 \max_{0 \leq M \leq I} \mathrm{Tr} [M(\rho - \sigma)] \\ &\geq \max_{0 \leq M \leq I} \mathrm{Tr} [M(\rho - \sigma)] \\ &\stackrel{\textcircled{1}}{\geq} \mathrm{Tr} [\Lambda(\rho - \sigma)] \\ &\geq \mathrm{Tr} [\Lambda \rho] - \mathrm{Tr} [\Lambda \sigma]. \end{aligned}$$

Equation $\textcircled{1}$ follows since the operator Λ , $0 \leq \Lambda \leq 1$, is a particular choice of the measurement operator M . \square

Most of the quantum systems considered in this document are finite dimensional, but it is worth noting that there are also quantum systems with continuous degrees of freedom which are represented in infinite dimensional Hilbert spaces. We will discuss the infinite dimensional case in Chapter 8, where we consider the quantum aspects of optical communication.

2.3.2 Quantum channels

By convention we will denote the input state as σ (for sender) and the outputs of the channel as ρ (for receiver). A noiseless quantum channel is represented by a unitary operator U which acts on the input state σ by conjugation to produce the output state $\rho = U\sigma U^\dagger$. General quantum channels are represented by completely-positive trace-preserving (CPTP) maps $\mathcal{N}^{A \rightarrow B}$, which accept input states in A and produce output states in B : $\rho^B = \mathcal{N}^{A \rightarrow B}(\sigma^A)$.

If the sender wishes to transmit some classical message m to the receiver using a quantum channel, her encoding procedure will consist of a classical-to-quantum encoder $\mathcal{E}: m \rightarrow \sigma^A$, to prepare a message state $\sigma^A \in \mathcal{D}(\mathcal{H}^A)$ suitable as input for the channel. We call this the *state preparation* step.

If the sender's encoding is restricted to transmitting a finite set of orthogonal states $\{\sigma_x^A\}_{x \in \mathcal{X}}$, then we can consider the choice of the signal states $\{\sigma_x^A\}$ to be part of the channel. Thus we obtain a channel with classical inputs $x \in \mathcal{X}$ and quantum outputs: $\rho_x^B = \mathcal{N}^{X \rightarrow B}(x) \equiv \mathcal{N}^{A \rightarrow B}(\sigma_x^A)$. A classical-quantum channel, $\mathcal{N}^{X \rightarrow B}$, is represented by the set of $|\mathcal{X}|$ possible output states $\{\rho_x^B \equiv \mathcal{N}^{X \rightarrow B}(x)\}$, meaning that each classical input of $x \in \mathcal{X}$ leads to a different quantum output $\rho_x^B \in \mathcal{D}(\mathcal{H}^B)$.

2.3.3 Quantum measurement

The decoding operations performed by the receivers correspond to quantum measurements on the outputs of the channel. A quantum measurement is a positive operator-valued measure (POVM) $\{\Lambda_m\}_{m \in \{1, \dots, |\mathcal{M}|\}}$ on the system B , the output of which we denote M' . The probability of outcome $M' = m$ when the state ρ^B is measured is given by the Born rule:

$$\Pr\{M' = m\} \equiv \text{Tr}[\Lambda_m^B \rho^B]. \quad (2.21)$$

To be a valid POVM, the set of $|\mathcal{M}|$ operators Λ_m must all be positive semidefinite and sum to the identity: $\Lambda_m \geq 0$, $\sum_m \Lambda_m = I$.

A quantum instrument $\{\Upsilon_k\}^{A \rightarrow B}$ is a more general operation which consists of a collection of completely positive (CP) maps such that $\sum_k \Upsilon_k$ is trace preserving [DL70]. When applied to a quantum state σ^A , the different elements are applied with probability $p_k = \text{Tr}[\Upsilon_k(\sigma^A)]$ resulting in different normalized outcomes $\rho_k^B = \frac{1}{p_k} \Upsilon_k(\sigma^A)$.

2.3.4 Quantum information theory

Many of the fundamental ideas of quantum information theory are analogous to those of classical information theory. For example, we quantify the information content of quantum systems using the notion of entropy.

Definition 2.1 (von Neumann Entropy). Given the density matrix $\rho^A \in \mathcal{D}(\mathcal{H}^A)$, the expression

$$H(A)_\rho = -\text{Tr}(\rho^A \log \rho^A) \quad (2.22)$$

is known as the *von Neumann entropy* of the state ρ^A .

Note that the symbol H is used for both classical and quantum entropy. The

2.3 Introduction to quantum information

von Neumann entropy of quantum state ρ^A with spectral decomposition $\rho^A = \sum_i \lambda_i |e_i\rangle\langle e_i|$, is equal to the Shannon entropy of its eigenvalues.

$$H(A)_\rho = -\text{Tr}(\rho^A \log \rho^A) = -\sum_i \lambda_i \log \lambda_i = H(\{\lambda_i\}). \quad (2.23)$$

For bipartite states ρ^{AB} we can also define the quantum conditional entropy

$$H(A|B)_\rho \equiv H(AB)_\rho - H(B)_\rho, \quad (2.24)$$

where $H(B)_\rho = -\text{Tr}(\rho^B \log \rho^B)$ is the entropy of the reduced density matrix $\rho^B = \text{Tr}_A(\rho^{AB})$. In the same fashion we can also define the quantum mutual information

$$I(A; B)_\rho \equiv H(A)_\rho + H(B)_\rho - H(AB)_\rho, \quad (2.25)$$

and in the case of a tripartite system ρ^{ABC} we define the conditional mutual information as

$$I(A; B|C)_\rho \equiv H(A|C)_\rho + H(B|C)_\rho - H(AB|C)_\rho \quad (2.26)$$

$$= H(AC)_\rho + H(BC)_\rho - H(ABC)_\rho - H(C)_\rho. \quad (2.27)$$

It can be shown that $I(A; B|C)$ is non negative for any tripartite state ρ^{ABC} . The formula $I(A; B|C) \geq 0$ can also be written in the form

$$H(AC) + H(BC) \geq H(C) + H(ABC). \quad (2.28)$$

This inequality, originally proved in [LR73], is called the *strong subadditivity* of von Neumann entropy and forms an important building block of quantum information theory.

Consider the classical-quantum state ρ^{XB} given by:

$$\rho^{XB} = \sum_{x \in \mathcal{X}} p_X(x) |x\rangle\langle x|^X \otimes \rho_x^B. \quad (2.29)$$

The conditional entropy $H(B|X)$ of this state is equal to:

$$H(B|X) = \sum_{x \in \mathcal{X}} p_X(x) H(\rho_x^B) = \sum_{x \in \mathcal{X}} p_X(x) H(B)_{\rho_x}. \quad (2.30)$$

2.4 Quantum typicality

The notions of typical sequences and typical sets generalize to the quantum setting by virtue of the spectral theorem. Let \mathcal{H}^B be a d_B dimensional Hilbert space and let $\rho^B \in \mathcal{D}(\mathcal{H}^B)$ be the density matrix associated with a quantum state. We identify the eigenvalues of ρ^B with the probability distribution $p_Y(y) = \lambda_{\rho;y}$ and write the spectral decomposition as:

$$\rho^B = \sum_{y=1}^{d_B} p_Y(y) |e_{\rho;y}\rangle \langle e_{\rho;y}|^B \quad (2.31)$$

where $|e_{\rho;y}\rangle$ is the eigenvector of ρ^B corresponding to eigenvalue $p_Y(y)$.

Define the set of δ -typical eigenvalue labels according to the eigenvalue distribution p_Y as

$$\mathcal{T}_\delta^{(n)}(Y) \equiv \left\{ y^n \in \mathcal{Y}^n : \left| -\frac{\log p_{Y^n}(y^n)}{n} - H(Y) \right| \leq \delta \right\}. \quad (2.32)$$

For a given string $y^n = y_1 y_2 \dots y_i \dots y_n$ we define the corresponding eigenvector as

$$|e_{\rho;y^n}\rangle = |e_{\rho;y_1}\rangle \otimes |e_{\rho;y_2}\rangle \otimes \dots \otimes |e_{\rho;y_n}\rangle, \quad (2.33)$$

where for each symbol $y_i = b \in \{1, 2, \dots, d_B\}$ we select the b^{th} eigenvector $|e_{\rho;b}\rangle$.

The typical subspace associated with the density matrix ρ^B is defined as

$$A_{\rho,\delta}^n = \text{span} \left\{ |e_{\rho;y^n}\rangle : y^n \in \mathcal{T}_\delta^{(n)}(Y) \right\}. \quad (2.34)$$

The typical projector is defined as

$$\Pi_{\rho^B,\delta}^n = \sum_{y^n \in \mathcal{T}_\delta^{(n)}(Y)} |e_{\rho;y^n}\rangle \langle e_{\rho;y^n}|. \quad (2.35)$$

Note that the typical projector is linked twofold to the spectral decomposition of (2.31): the sequences y^n are selected according to p_Y and the set of typical vectors are built

2.4 Quantum typicality

from tensor products of orthogonal eigenvectors $|e_{\rho;y}\rangle$.

Properties analogous to (2.3) - (2.5) hold. For any $\epsilon, \delta > 0$, and all sufficiently large n we have

$$\text{Tr}\{\rho^{\otimes n}\Pi_{\rho,\delta}^n\} \geq 1 - \epsilon \quad (2.36)$$

$$2^{-n[H(B)_\rho+\delta]}\Pi_{\rho,\delta}^n \leq \Pi_{\rho,\delta}^n \rho^{\otimes n} \Pi_{\rho,\delta}^n \leq 2^{-n[H(B)_\rho-\delta]}\Pi_{\rho,\delta}^n, \quad (2.37)$$

$$[1 - \epsilon]2^{n[H(B)_\rho-\delta]} \leq \text{Tr}\{\Pi_{\rho,\delta}^n\} \leq 2^{n[H(B)_\rho+\delta]}. \quad (2.38)$$

Equation (2.36) tells us that most of the support of the state $\rho^{\otimes n}$ is within the typical subspace. The interpretation of (2.37) is that the eigenvalues of the state $\rho^{\otimes n}$ are bounded between $2^{-n[H(B)_\rho+\delta]}$ and $2^{-n[H(B)_\rho-\delta]}$ on the typical subspace $A_{\rho,\delta}^n$.

Signal states Consider now a set of quantum states $\{\rho_{x_a}^B\}$, $x_a \in \mathcal{X}$. We perform a spectral decomposition of each $\rho_{x_a}^B$ to obtain

$$\rho_{x_a}^B = \sum_{y=1}^{d_B} p_{Y|X}(y|x_a) |e_{\rho_{x_a};y}\rangle \langle e_{\rho_{x_a};y}|^B, \quad (2.39)$$

where $p_{Y|X}(y|x_a)$ is the y^{th} eigenvalue of $\rho_{x_a}^B$ and $|e_{\rho_{x_a};y}\rangle$ is the corresponding eigenvector.

We can think of $\{\rho_{x_a}^B\}$ as a classical-quantum (c - q) channel where the input is some $x_a \in \mathcal{X}$ and the output is the corresponding quantum state $\rho_{x_a}^B$. If the channel is memoryless, then for each input sequence $x^n = x_1 x_2 \cdots x_n$ we have the corresponding tensor product output state:

$$\rho_{x^n}^{B^n} = \rho_{x_1}^{B_1} \otimes \rho_{x_2}^{B_2} \otimes \cdots \otimes \rho_{x_n}^{B_n}. \quad (2.40)$$

2.4.1 Quantum conditional typicality

Conditionally typical projector Consider the ensemble $\{p_X(x_a), \rho_{x_a}\}$. The choice of distributions induces the following classical-quantum state:

$$\rho^{XB} = \sum_{x_a} p_X(x_a) |x_a\rangle \langle x_a|^X \otimes \rho_{x_a}^B. \quad (2.41)$$

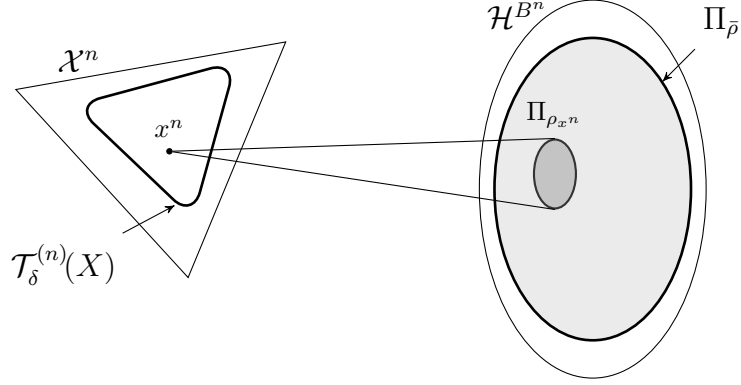


Figure 2.3: Illustration of a conditionally typical subspace for some sequence x^n , and the output-typical subspace.

We define $H(B|X)_\rho \equiv \sum_{x_a \in \mathcal{X}} p_X(x_a) H(\rho_{x_a})$ to be the conditional entropy of this state. Expressed in terms of the eigenvalues of the signal states, the conditional entropy becomes

$$H(B|X)_\rho \equiv H(Y|X) \equiv \sum_{x_a} p_X(x_a) H(Y|x_a), \quad (2.42)$$

where $H(Y|x_a) = -\sum_y p_{Y|X}(y|x_a) \log p_{Y|X}(y|x_a)$ is the entropy of the eigenvalue distribution shown in (2.39).

We define the x^n -conditionally typical projector as follows:

$$\Pi_{\rho_{x^n}, \delta}^n = \sum_{y^n \in \mathcal{T}_\delta^{(n)}(Y|x^n)} |e_{\rho_{x^n}; y^n}\rangle \langle e_{\rho_{x^n}; y^n}|, \quad (2.43)$$

where the set of conditionally typical eigenvalues $\mathcal{T}_\delta^{(n)}(Y|x^n)$ consists of all sequences y^n which satisfy:

$$\mathcal{T}_\delta^{(n)}(Y|x^n) \equiv \left\{ y^n : \left| -\frac{\log p_{Y^n|X^n}(y^n|x^n)}{n} - H(Y|X) \right| \leq \delta \right\}, \quad (2.44)$$

with $p_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n p_{Y|X}(y_i|x_i)$.

The states $|e_{\rho_{x^n}; y^n}\rangle$ are built from tensor products of eigenvectors for the individual signal states:

$$|e_{\rho_{x^n}; y^n}\rangle = |e_{\rho_{x_1}; y_1}\rangle \otimes |e_{\rho_{x_2}; y_2}\rangle \otimes \cdots \otimes |e_{\rho_{x_n}; y_n}\rangle, \quad (2.45)$$

where the string $y^n = y_1 y_2 \dots y_i \dots y_n$ varies over different choices of bases for \mathcal{H}^B . For each symbol $y_i = b \in \{1, 2, \dots, d_B\}$ we select $|e_{\rho_{x_a}; b}\rangle$: the b^{th} eigenvector from the

2.5 Closing remarks

eigenbasis of ρ_{x_a} corresponding to the letter $x_i = x_a \in \mathcal{X}$.

The following bound on the rank of the conditionally typical projector applies:

$$\text{Tr}\{\Pi_{\rho_{x^n, \delta}}^n\} \leq 2^{n[H(B|X)_\rho + \delta]}. \quad (2.46)$$

2.5 Closing remarks

In the next chapter, we will show how the properties of the typical sequences and typical subspaces can be used to construct coding theorems for classical and classical-quantum channels.

Chapter 3

Point-to-point communication

In this chapter we describe the point-to-point communication scenario in which there is a single sender and a single receiver. In Section 3.1, we review Shannon's channel coding theorem and give the details of the achievability proof in order to introduce the idea of *random coding* in its simplest form. Our presentation is somewhat unorthodox since we use only the properties of the conditionally typical sets and not the jointly typical sets. Though, following this approach allows us to directly generalize our proof techniques to the quantum case.

In Section 3.2.1 we will discuss the Holevo-Schumacher-Westmoreland (HSW) Theorem and show an achievability proof. We do so with the purpose of introducing important background material on the construction of quantum decoding operators. We show how to construct a decoding POVM defined in terms of the conditionally typical projectors. Readers interested only in the essential parts should consult Lemma 3.1 and Lemma 3.2, since they will be used throughout the remainder of the text.

3.1 Classical channel coding

The fundamental problem associated with communication channels is to calculate and formally prove their capacity for information transmission. We can think of the use of a channel \mathcal{N} as a *communication resource*, of which we have n instances. Each use of the channel is assumed to be independent, and modelled by the conditional probability distribution $p_{\mathcal{Y}|\mathcal{X}}(y|x)$, where x and y are elements from the finite sets \mathcal{X} , \mathcal{Y} . This is

called the discrete memoryless setting.

Our goal is to study the *rate* R at which the channel \mathcal{N} can be converted into copies of the noiseless binary channel $[c \rightarrow c] \equiv \delta(x, y)$, $x, y \in \{0, 1\}$, which represents the canonical unit resource of communication. This conversion can be expressed as follows:

$$n \cdot \mathcal{N} \xrightarrow{(1-\epsilon)} nR \cdot [c \rightarrow c]. \quad (3.1)$$

This equation describes a protocol in which n units of the noisy communication resource \mathcal{N} are transformed into nR bits of noiseless transmission, and the protocol succeeds with probability $(1 - \epsilon)$. Note that we allow the communication protocol to fail with probability ϵ , but ϵ is an arbitrarily small number for sufficiently large n . To prove that the rate R is *achievable*, one has to describe the coding strategy and prove that the probability of error for that strategy can be made arbitrarily small. Usually, the right hand side in equation (3.1) is measured as the number of different messages $\mathcal{M} \equiv \{1, 2, \dots, 2^{nR}\} \equiv [1 : 2^{nR}]$ that can be transmitted using n uses of the channel. One can think of the nR individual bits of the message as being noiselessly transmitted to the receiver. The channel coding pipeline can then be described as follows:

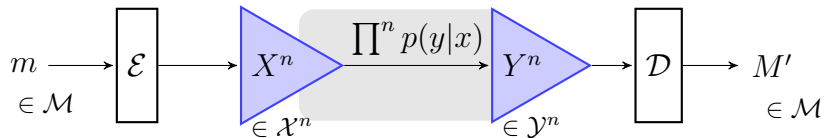


Figure 3.1: Classical channel coding setup. The diagram shows the encoding, transmission and decoding steps of a communication protocol that uses n copies of the classical channel $\mathcal{N} = (\mathcal{X}, p_{Y|X}(y|x), \mathcal{Y})$.

The probability of error when sending message m is defined as $p_e(m) \equiv \Pr\{M' \neq m\}$, where $M' \equiv \mathcal{D} \circ \mathcal{N}^n \circ \mathcal{E}(m)$ is the random variable associated with the output of the protocol. The average probability of error over all messages is

$$\bar{p}_e \equiv \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \Pr\{M' \neq m\}. \quad (3.2)$$

This is the quantity we have to bound when we perform an *error analysis* of some coding protocol.

Definition 3.1. An (n, R, ϵ) coding protocol consists of a message set \mathcal{M} , where $|\mathcal{M}| = 2^{nR}$, an encoding map $\mathcal{E} : \mathcal{M} \rightarrow \mathcal{X}^n$ described by a codebook $\{x^n(m)\}_{m \in \mathcal{M}}$, and a decoding map $\mathcal{D} : \mathcal{Y}^n \rightarrow \mathcal{M}$ such that the average probability of error is bounded

from above as $\bar{p}_e \leq \epsilon$.

A rate R is *achievable* if there exists an $(n, R - \delta, \epsilon)$ coding protocol for all $\epsilon, \delta > 0$ as $n \rightarrow \infty$.

3.1.1 Channel capacity

The capacity C of a channel is the maximum of the rates R that are achievable, and is established in Shannon's channel coding theorem.

Theorem 3.1 (Channel capacity [Sha48, Fei54]). *The communication capacity of a discrete memoryless channel $(\mathcal{X}, p_{Y|X}(y|x), \mathcal{Y})$ is given by*

$$C = \max_{p_X} I(X; Y), \quad (3.3)$$

where the optimization is taken over all possible input distributions $p_X(x)$. The mutual information is calculated on the induced joint probability distribution

$$(X, Y) \sim p_{XY}(x, y) = p_X(x)p_{Y|X}(y|x). \quad (3.4)$$

The proof of a capacity theorem usually contains two parts:

- A direct coding part that shows that for all $\epsilon, \delta > 0$, there exists a codebook $\mathcal{E}(m) \equiv \{x^n(m)\}$ of rate $R = C - \delta$ and a decoding map \mathcal{D} with average probability of error $\bar{p}_e \leq \epsilon$.
- A converse part that shows that the rate C is the maximum rate possible. A converse theorem establishes that the probability of error for a coding protocol $(n, C + \delta, \epsilon)$ is bounded away from zero (weak converse), or that the probability of error goes exponentially to 1 (strong converse).

Proof. We give an overview of the achievability proof of Theorem 3.1 in order to introduce key concepts, which will be used in the other proofs in this thesis.

We use a random codebook with $2^{nR} = |\mathcal{M}|$ codewords $x^n \in \mathcal{X}^n$ generated independently from the product distribution $p_{X^n}(x^n) = \prod^n p_X(x_i)$. When the sender wants to send the message $m \in \mathcal{M}$, she will input the m^{th} codeword, which we will denote as $x^n(m)$. Let Y^n denote the resulting output of the channel. The distribution on the output symbols induced by the input distribution is $p_Y(y) \equiv \sum_x p_{Y|X}(y|x)p(x)$, and

3.1 Classical channel coding

define the set of output-typical sequences $\mathcal{T}_\delta^{(n)}(Y)$ according to the distribution p_Y . For any sequence x^n , denote the set of conditionally typical output sequences $\mathcal{T}_\delta^{(n)}(Y|x^n)$.

Given the output of the channel y^n , the receiver will use the following algorithm:

1. If $y^n \notin \mathcal{T}_\delta^{(n)}(Y)$, then an error is declared.
2. Return m if y^n is an element of the conditionally typical set $\mathcal{T}_\delta^{(n)}(Y|x^n(m))$.
Report an error if no match or multiple matches are found.

We now define the three types of errors that may occur in the protocol when the message m is being sent.

(E0): The event that the channel output Y^n is not output-typical: $\{Y^n \notin \mathcal{T}_\delta^{(n)}(Y)\}$.

(E1): The event that the channel output sequence Y^n is not in the conditionally typical set $\{Y^n \notin \mathcal{T}_\delta^{(n)}(Y|x^n(m))\}$, which corresponds to the message m .

(E2): The event that Y^n is output-typical and it falls in the conditionally typical set for another message:

$$\{Y^n \in \mathcal{T}_\delta^{(n)}(Y)\} \cap \left(\bigcup_{m' \neq m} \{Y^n \in \mathcal{T}_\delta^{(n)}(Y|x^n(m')), m' \neq m\} \right). \quad (3.5)$$

We can bound the probability of all three events when a random codebook is used, that is, we will take the expectation over the random choices of the symbols for each codeword. We define the expectation of an event as the expectation of the associated indicated random variable.

The bound $\mathbb{E}_{X^n}(\mathbf{E0}) \leq \epsilon$ follows from (2.13). The crucial observation for the proof is to use the symmetry of the code construction: if the codewords for all the messages are constructed identically, then it is sufficient to analyze the probability of error for any one fixed message. We obtain a bound $\mathbb{E}_{X^n}(\mathbf{E1}) \leq \epsilon$ from (2.9).

In order to bound the probability of error event **(E2)**, we will use the *classical packing lemma*, Lemma A.1 in Appendix A.2. Using the packing lemma with $U = \emptyset$, we obtain a bound on the probability that the conditionally typical sets for different messages will overlap. We can thus bound the expectation of the probability of error

event **(E2)** as follows:

$$\mathbb{E}_{X^n} \Pr\{\mathbf{(E2)}\} \leq |\mathcal{M}| 2^{-n[I(X;Y)-\delta]}.$$

We can now use the union bound to bound the overall probability of error for our code as follows:

$$\begin{aligned} \mathbb{E}_{X^n} \{\bar{p}_e\} &= \mathbb{E}_{X^n} \Pr\{\mathbf{(E0)} \cup \mathbf{(E1)} \cup \mathbf{(E2)}\} \\ &\leq \mathbb{E}_{X^n} \Pr\{\mathbf{(E0)}\} + \mathbb{E}_{X^n} \Pr\{\mathbf{(E1)}\} + \mathbb{E}_{X^n} \Pr\{\mathbf{(E2)}\} \\ &\leq \epsilon + \epsilon + |\mathcal{M}| 2^{-n[I(X;Y)-\delta]} \\ &= \epsilon + \epsilon + 2^{-n[I(X;Y)-R-\delta]}. \end{aligned}$$

Thus, in the limit of many uses of the channel, we have:

$$\mathbb{E}_{X^n} \{\bar{p}_e\} \leq \epsilon', \quad (3.6)$$

provided the rate $R \leq I(X;Y) - 2\delta$.

The last step is called *derandomization*. If the expected probability of error of a random codebook can be bounded as above, then there must exist a particular codebook with $\bar{p}_e \leq \epsilon'$, which completes the proof. \square

Note that it is possible to use an *expurgation* step and throw out the worse half of the codewords in order to convert the bound on the average probability of error \bar{p}_e into a bound on the maximum probability of error $\bar{p}_e^{\max} = \max_m p_e(m)$ [CT91].

3.2 Quantum communication channels

A quantum channel $(\mathcal{H}^A, \mathcal{N}^{A \rightarrow B}, \mathcal{H}^B)$ is described as a completely positive trace-preserving map $\mathcal{N}^{A \rightarrow B}$ which takes a quantum system in state $\sigma^A \in \mathcal{D}(\mathcal{H}^A)$ as input and outputs a quantum system $\rho^B \in \mathcal{D}(\mathcal{H}^B)$. Figure 3.2 shows an example of such a channel. In recent years, the techniques

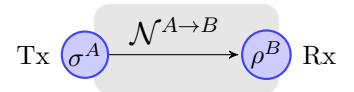


Figure 3.2: A point-to-point quantum channel $\mathcal{N}^{A \rightarrow B}$.

of classical information theory have been extended to the study of quantum channels. For a review of the subject see [Wil11].

In addition to the standard problem of *classical* transmission of information (denoted $[c \rightarrow c]$), for quantum channels we can study the transmission of quantum information (denoted $[q \rightarrow q]$). If pre-shared entanglement between Transmitter and Receiver is available, it can be used in order to improve the communication rates using an *entanglement-assisted* protocol. There are multiple communication tasks and different capacities associated with each task for any given quantum channel \mathcal{N} [BSST99]. Some of the possible communication tasks, along with their associated capacities are:

- Classical data capacity: $\mathcal{C}(\mathcal{N})$
- Quantum data capacity: $\mathcal{Q}(\mathcal{N})$
- Entanglement-assisted classical data capacity: $\mathcal{C}_{\text{E-A}}(\mathcal{N})$
- Entanglement-assisted quantum data capacity: $\mathcal{Q}_{\text{E-A}}(\mathcal{N})$

The latter two are actually equivalent up to a factor of 2, because we can use the *superdense coding* and *quantum teleportation* protocols to convert between them in the presence of free entanglement [BW92, BBC⁺93].

In the context of quantum information theory, pre-shared quantum entanglement between sender and receiver must be recognized as a communication resource. We denote this resource $[qq]$ and must take into account the rates at which it is consumed or generated as part of a communication protocol [DHW08]. It is interesting to note that shared randomness (denoted $[cc]$), which is the classical equivalent of shared entanglement, does not increase the capacity of point-to-point classical channels.

Classical-quantum channels

In the previous section we introduced some of the main communication problems of quantum information theory. The focus of this thesis will be the study of *classical communication* ($[c \rightarrow c]$) over quantum channels, with no entanglement assistance. For this purpose, we will use the *classical-quantum* (c-q) channel model, which corresponds to the use of a quantum channel where the Sender is restricted to sending a finite set of *signal states* $\{\sigma_x^A\}_{x \in \mathcal{X}}$. If we consider the choice of the signal states $\{\sigma_x^A\}$

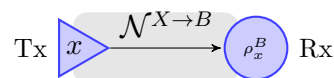


Figure 3.3: A point-to-point c-q channel $\{\rho_x\}$.

to be part of the channel, we obtain a channel with classical inputs $x \in \mathcal{X}$ and quantum outputs: $\mathcal{N}^{X \rightarrow B}(x) \equiv \mathcal{N}^{A \rightarrow B}(\sigma_x^A)$. Note that a classical-quantum channel $(\mathcal{X}, \mathcal{N}^{X \rightarrow B}(x) \equiv \rho_x^B, \mathcal{H}^B)$ is fully specified by the finite set of output states $\{\rho_x^B\}$ it produces for each of the possible inputs $x \in \mathcal{X}$. This channel model is a useful abstraction for studying the transmission of classical data over quantum channels. Any code construction for a c-q channel can be augmented with an optimization over the choice of signal states $\{\sigma_x^A\}_{x \in \mathcal{X}}$ to obtain a code for a quantum channel. The Holevo-Schumacher-Westmoreland Theorem establishes the classical capacity of the classical-quantum channel [Hol98, SW97]. The strong converse was later proved in [ON99].

3.2.1 Classical-quantum channel coding

The quantum channel coding problem for a point-to-point classical-quantum channel $(\mathcal{X}, \mathcal{N}^{X \rightarrow B}(x) \equiv \rho_x^B, \mathcal{H}^B)$ is studied in the following setting.

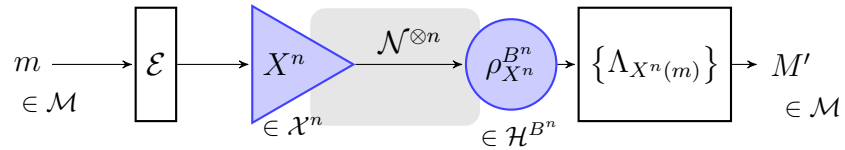


Figure 3.4: HSW coding setup.

Let $x^n(m) \equiv x_1 x_2 \cdots x_n \in \mathcal{X}^n$ be the codeword which is input to the channel when we want to send message m . The output of the channel will be the n -fold tensor product state:

$$\mathcal{N}^{\otimes n}(x^n(m)) \equiv \rho_{x^n(m)}^{B^n} \equiv \rho_{x_1(m)}^{B_1} \otimes \rho_{x_2(m)}^{B_2} \otimes \cdots \otimes \rho_{x_n(m)}^{B_n}. \quad (3.7)$$

To extract the classical information encoded into this state, we must perform a quantum measurement. The most general quantum measurement is described by a positive operator-valued measure (POVM) $\{\Lambda_m\}_{m \in \mathcal{M}}$ on the system B^n . To be a valid POVM, the set $\{\Lambda_m\}$ of $|\mathcal{M}|$ operators should all be positive semidefinite and sum to the identity: $\Lambda_m \geq 0$, $\sum_m \Lambda_m = I$.

In the context of our coding strategy, the decoding measurement aims to distinguish the $|\mathcal{M}|$ possible states of the form (3.7). The advantage of the quantum coding paradigm is that it allows for joint measurements on all the outputs of the channel, which is more powerful than measuring the systems individually.

We define the average probability of error for the end-to-end protocol as

$$\bar{p}_e \equiv \frac{1}{|\mathcal{M}|} \sum_m \text{Tr}\{(I - \Lambda_{x^n(m)}^{B^n}) \rho_{x^n(m)}^{B^n}\}, \quad (3.8)$$

where the operator $(I - \Lambda_{x^n(m)}^{B^n})$ corresponds to the complement of the correct decoding outcome.

Definition 3.2. An (n, R, ϵ) classical-quantum coding protocol consists of a message set \mathcal{M} , where $|\mathcal{M}| = 2^{nR}$, an encoding map $\mathcal{E} : \mathcal{M} \rightarrow \mathcal{X}^n$ described by a codebook $\{x^n(m)\}_{m \in \mathcal{M}}$, and a decoding measurement (POVM) $\{\Lambda_{x^n(m)}\}_{m \in \mathcal{M}}$ such that the average probability of error is bounded from above as $\bar{p}_e \leq \epsilon$.

Theorem 3.2 (HSW Theorem [Hol98, SW97]). *The classical communication capacity of a classical-quantum channel $(\mathcal{X}, \rho_x^B, \mathcal{H}^B)$ is given by:*

$$\mathcal{C}(\mathcal{N}) = \max_{p_X} I(X; B)_\theta \quad (3.9)$$

where the optimization is taken over all possible input distributions p_X , and where entropic quantities are calculated with respect to the following state:

$$\theta^{XB} = \sum_x p_X(x) |x\rangle\langle x|^X \otimes \rho_x^B. \quad (3.10)$$

The classical-quantum state θ^{XB} is the state with respect to which we will calculate mutual information quantities. We call this state the *code state* and it extends the classical joint probability distribution induced by a channel, when the input distribution p_X is used to construct the codebook: $p_X(x)p_{Y|X}(y|x)$. In the case of the classical-quantum channel, the outputs are quantum systems. Information quantities taken with respect to classical-quantum states are called ‘‘Holevo’’ quantities in honour of Alexander Holevo who was first to recognize the importance of this expression by proving that it is an upper bound to the accessible information of an ensemble [Hol73, Hol79]. Holevo quantities are expressed as a difference of two entropic terms:

$$I(X; B)_\theta \equiv H(B)_\theta - H(B|X)_\theta \equiv H\left(\sum_x p_X(x)\rho_x^B\right) - \sum_x p_X(x)H(\rho_x^B). \quad (3.11)$$

Holevo quantities are in some sense partially classical, since the entropies are with respect to quantum systems, but the conditioning is classical.

Quantum decoding

When devising coding strategies for classical-quantum channels, the main obstacle to overcome is the construction of a decoding POVM that correctly identifies the messages. Using the properties of quantum typical subspaces we can construct a set of positive operators $\{P_m\}_{m \in \mathcal{M}}$ which, analogously to the classical conditionally typical indicator functions, are good at detecting ($\text{Tr}[P_m \rho_m] \geq 1 - \epsilon$) and distinguishing ($\text{Tr}[P_m \rho_{m' \neq m}] \leq \epsilon$) the output states produced by each message. We can construct a valid POVM by *normalizing* these operators:

$$\Lambda_m \equiv \left(\sum_k P_k \right)^{-1/2} P_m \left(\sum_k P_k \right)^{-1/2}, \quad (3.12)$$

so that we will have $\sum_m \Lambda_m = I$. This is known as the *square root* measurement or the *pretty good* measurement [Hol98, SW97].

The achievability proof of Theorem 3.2 is based on the properties of typical subspaces and the square root measurement. We construct a set of unnormalized positive operators

$$P_m^{B^n} \equiv \Pi_{\bar{\rho}} \Pi_{x^n(m)} \Pi_{\bar{\rho}}, \quad (3.13)$$

where $\Pi_{x^n(m)} \equiv \Pi_{\rho_{x^n(m), \delta}}^{B^n}$ is the conditionally typical projector that corresponds to the input sequence $x^n(m)$ and $\Pi_{\bar{\rho}} \equiv \Pi_{\bar{\rho}^{\otimes n}, \delta}^{B^n}$ is the output-typical projector for the average output state $\bar{\rho} = \sum_x p_X(x) \rho_x^B$. The operator “sandwich” in equation (3.13) corresponds directly to the decoding criteria used in the classical coding theorem. We require the state to be in the output-typical subspace *and* inside the conditionally typical subspace for the correct codeword $x^n(m)$. The decoding POVM is then constructed as in (3.12).

By using the properties of the typical projectors, we can show that the probability of error of this coding scheme vanishes provided $R \leq I(X; B) - \delta$. An effort has been made to present the proofs of the classical and quantum coding theorems in a similar fashion in order to highlight similarities in the reasoning.

3.3 Proof of HSW Theorem

In this section we give the details of the POVM construction and the error analysis for the decoder used by the receiver in the HSW Theorem.

Recall the classical-quantum state (3.10), with respect to which our code is constructed:

$$\theta^{XB} = \sum_x p_X(x) |x\rangle\langle x|^X \otimes \rho_x^B. \quad (3.14)$$

For each input sequence x^n , there is a corresponding δ -conditionally typical projector: $\Pi_{x^n} \equiv \Pi_{\rho_{x^n}, \delta}^{B^n}$.

Define also the average output state $\bar{\rho} \equiv \sum_x p_X(x) \rho_x^B$, and the corresponding average-output-typical projector $\Pi_{\bar{\rho}} \equiv \Pi_{\bar{\rho}^{\otimes n}, \delta}^{B^n}$.

The Receiver constructs a decoding POVM $\{\Lambda_m\}_{m \in \mathcal{M}}$ by starting from the *projector sandwich*:

$$P_m^{B^n} \equiv \Pi_{\bar{\rho}} \Pi_{x^n(m)} \Pi_{\bar{\rho}}, \quad (3.15)$$

and *normalizing* the operators:

$$\Lambda_m \equiv \left(\sum_k P_k \right)^{-1/2} P_m \left(\sum_k P_k \right)^{-1/2}. \quad (3.16)$$

The error analysis of a square root measurement is greatly simplified by using the Hayashi-Nagaoka operator inequality.

Lemma 3.1 (Hayashi-Nagaoka [HN03]). *If S and T are operators such that $0 \leq T$ and $0 \leq S \leq I$, then*

$$I - (S + T)^{-\frac{1}{2}} S (S + T)^{-\frac{1}{2}} \leq 2(I - S) + 4T. \quad (3.17)$$

If we let $S = P_m$ and $T = \sum_{m' \neq m} P_{m'}$ in the above inequality we obtain

$$I - \Lambda_m \leq 2(I - P_m) + 4 \sum_{m' \neq m} P_{m'}, \quad (3.18)$$

which corresponds to the decomposition of the error outcome $(I - \Lambda_m)$ into two contributions:

- I. The probability that the correct detector does not “click”: $(I - P_m)$. This corresponds to the error events **(E0)** and **(E1)** in the classical coding theorem.
- II. The probability that a wrong detector “clicks”: $\sum_{m' \neq m} P_{m'}$. This corresponds to the error event **(E2)** in the classical case.

We will show that the average probability of error

$$\bar{p}_e \equiv \frac{1}{|\mathcal{M}|} \sum_m \text{Tr}\{(I - \Lambda_m^{B^n}) \rho_{x^n(m)}^{B^n}\},$$

will be small provided the rate $R \leq I(X; B) - \delta = H(B) - H(B|X) - \delta$. The bound follows from the following properties of typical projectors:

$$\text{Tr}[\Pi_{x^n(m)}] \leq 2^{n[H(B|X) + \delta]}, \quad (3.19)$$

$$\Pi_{\bar{\rho}} \bar{\rho}^{\otimes n} \Pi_{\bar{\rho}} \leq 2^{-n[H(B) - \delta]} \Pi_{\bar{\rho}}, \quad (3.20)$$

and reasoning analogous to that used in the classical coding theorem. Note that by the symmetry of both the codebook construction and the decoder we can study the error analysis for a fixed message m .

Consider the probability of error when the message m is sent, and let us apply the Hayashi-Nagaoka operator inequality (Lemma 3.1) to split the error into two terms:

$$\begin{aligned} \bar{p}_e &\equiv \text{Tr}[(I - \Lambda_m^{B^n}) \rho_{x^n(m)}^{B^n}] \\ &\leq 2 \underbrace{\text{Tr}[(I - P_m^{B^n}) \rho_{x^n(m)}^{B^n}]}_{\text{(I)}} + 4 \underbrace{\sum_{m' \neq m} \text{Tr}[P_{m'}^{B^n} \rho_{x^n(m)}^{B^n}]}_{\text{(II)}}. \end{aligned} \quad (3.21)$$

We bound the expectation of the average probability of error by bounding the individual terms.

We now state two useful results, which we need to bound the first error term. First, recall the inequality from Lemma 2.1 which states that:

$$\text{Tr}[\Lambda \rho] \leq \text{Tr}[\Lambda \sigma] + \|\rho - \sigma\|_1, \quad (3.22)$$

holds for all operators such that $0 \leq \rho, \sigma, \Lambda \leq I$.

The second ingredient is the gentle measurement lemma.

Lemma 3.2 (Gentle operator lemma for ensembles [Win99]). *Let $\{p(x), \rho_x\}$ be an ensemble and let $\bar{\rho} \equiv \sum_x p(x) \rho_x$. If an operator Λ , where $0 \leq \Lambda \leq I$, has high overlap with the average state, $\text{Tr}[\Lambda \bar{\rho}] \geq 1 - \epsilon$, then the subnormalized state $\sqrt{\Lambda} \rho_x \sqrt{\Lambda}$ is close in trace distance to the original state ρ_x on average: $\mathbb{E}_X \left\{ \left\| \sqrt{\Lambda} \rho_x \sqrt{\Lambda} - \rho_x \right\|_1 \right\} \leq 2\sqrt{\epsilon}$.*

We bound the expectation over the code randomness for the first term in (3.21) as follows:

$$\begin{aligned}
 \mathbb{E}_{X^n}(\text{I}) &= \mathbb{E}_{X^n} \text{Tr} \left[(I - P_m^{B^n}) \rho_{x^n(m)}^{B^n} \right] \\
 &= \mathbb{E}_{X^n} \text{Tr} \left[(I - \Pi_{\bar{\rho}} \Pi_{x^n(m)} \Pi_{\bar{\rho}}) \rho_{x^n(m)}^{B^n} \right] \\
 &= 1 - \mathbb{E}_{X^n} \left\{ \text{Tr} \left[\Pi_{x^n(m)} \Pi_{\bar{\rho}} \rho_{x^n(m)}^{B^n} \Pi_{\bar{\rho}} \right] \right\} \\
 &\stackrel{\textcircled{1}}{\leq} 1 - \mathbb{E}_{X^n} \left\{ \text{Tr} \left[\Pi_{x^n(m)} \rho_{x^n(m)}^{B^n} \right] + \left\| \Pi_{\bar{\rho}} \rho_{x^n(m)}^{B^n} \Pi_{\bar{\rho}} - \rho_{x^n(m)}^{B^n} \right\|_1 \right\} \\
 &= 1 - \mathbb{E}_{X^n} \text{Tr} \left[\Pi_{x^n(m)} \rho_{x^n(m)}^{B^n} \right] + \mathbb{E}_{X^n} \left\| \Pi_{\bar{\rho}} \rho_{x^n(m)}^{B^n} \Pi_{\bar{\rho}} - \rho_{x^n(m)}^{B^n} \right\|_1 \\
 &\stackrel{\textcircled{2}}{\leq} 1 - \mathbb{E}_{X^n} \text{Tr} \left[\Pi_{x^n(m)} \rho_{x^n(m)}^{B^n} \right] + 2\sqrt{\epsilon} \\
 &\stackrel{\textcircled{3}}{\leq} 1 - (1 - \epsilon) + 2\sqrt{\epsilon} = \epsilon + 2\sqrt{\epsilon}.
 \end{aligned}$$

The inequality $\textcircled{1}$ follows from equation (3.22). The inequality $\textcircled{2}$ follows from Lemma 3.2 and the property of the average output state $\text{Tr}[\Pi_{\bar{\rho}} \bar{\rho}^{\otimes n}] \geq 1 - \epsilon$. The inequality $\textcircled{3}$ follows from: $\mathbb{E}_{X^n} \text{Tr}[\Pi_{X^n(m)} \rho_{X^n(m)}] \geq 1 - \epsilon$.

The crucial Holevo information-dependent bound on the expectation of the second term in (3.21) can be obtained by using the quantum packing lemma. The quantum packing lemma (Lemma B.1) given in Appendix B.2, provides a bound on the amount of overlap between the conditionally typical subspaces for the codewords in our code construction and is analogous to the classical packing lemma (Lemma A.1), which we used to prove the classical channel coding theorem. Note that Lemma B.1 is less general than the quantum packing lemmas which appear in [HDW08] and [Wil11].

The overall probability of error is thus bounded as

$$\mathbb{E}_{X^n} \bar{p}_e \leq 2(\epsilon + 2\sqrt{\epsilon}) + 4 \left(2^{-n[I(X;B) - 2\delta - R]} \right), \quad (3.23)$$

and if we choose $R \leq I(X; B) - 3\delta$, the probability of error is bounded from above by ϵ in the limit $n \rightarrow \infty$.

Example 3.1 (Point-to-point channel). Consider the classical-quantum channel $\mathcal{N} \equiv (\{0, 1\}, \rho_x^B, \mathbb{C}^2)$, which takes a classical bit as input and outputs a qubit (a two-dimensional quantum system). Suppose the channel map is the following:

$$0 \rightarrow \rho_0 \equiv |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad 1 \rightarrow \rho_1 \equiv |+\rangle\langle +| = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \quad (3.24)$$

We calculate the channel capacity for three different measurement strategies: two classical strategies where the channel outputs are measured independently, and a quantum strategy that uses collective measurements on blocks of n channel outputs. Because the input is binary, it is possible to plot the achievable rates for all input distributions p_X . See Figure 3.5 for a plot of the achievable rates for these three strategies.

a) Basic classical decoding: A classical strategy for this channel corresponds to the channel outputs being individually measured in the computational basis:

$$\Lambda_0 = |0\rangle\langle 0|, \quad \Lambda_1 = |1\rangle\langle 1|, \quad \Lambda_{y^n}^{B^n} \equiv \Lambda_{y_1} \otimes \Lambda_{y_2} \otimes \cdots \otimes \Lambda_{y_n}. \quad (3.25)$$

Such a communication model for the channel is *classical* since we have $\text{Tr} [\Lambda_{y^n}^{B^n} \rho_{x^n}^{B^n}] \equiv p_{Y^n|X^n}(y^n|x^n)$. More specifically, $p_{Y^n|X^n}(y^n|x^n) = \prod^n p_{Y_i|X_i}^{(a)}(y_i|x_i)$, where $p_{Y_i|X_i}^{(a)}(y_i|x_i)$ is a classical Z -channel with transition probability $p_z \equiv p_{Y_i|X_i}^{(a)}(0|1) = \text{Tr}[\Lambda_0|+\rangle\langle +|] = 0.5$.

The capacity of the classical Z -channel is given by:

$$C^{(a)}(\mathcal{N}) = \max_{0 \leq p_0 \leq 1} H((1-p_0)(1-p_z)) - (1-p_0)H(p_z), \quad (3.26)$$

where we parametrize in terms of $p_0 = p_X(0)$. For this model, the capacity achieving input distribution has $p_0 = 0.6$ and the capacity is $C^{(a)} = H_2(0.2) - 0.4 \approx 0.3219$.

b) Aligned classical decoding: A better classical model is to use a “rotated” quantum measurement such that the measurement operators are symmetrically aligned with the channel outputs. The measurement directions $-\pi/8$ and $\pi/4 + \pi/8$ are symmetric around the output states $|0\rangle$ and $|+\rangle$. Define the notation $c_{\pi_8} = \cos(\pi/8)$ and $s_{\pi_8} = \sin(\pi/8)$. The measurement along the $-\pi/8$ and $\pi/4 + \pi/8$ directions corre-

3.4 Proof of HSW Theorem

sponds to the following POVM operators:

$$\Lambda_0 = (c_{\pi_8}|0\rangle - s_{\pi_8}|1\rangle)(c_{\pi_8}\langle 0| - s_{\pi_8}\langle 1|) = \begin{bmatrix} c_{\pi_8}^2 & -c_{\pi_8}s_{\pi_8} \\ -s_{\pi_8}c_{\pi_8} & s_{\pi_8}^2 \end{bmatrix}_{\{|0\rangle,|1\rangle\}}$$

$$\Lambda_1 = (c_{\pi_8}|+\rangle - s_{\pi_8}|-\rangle)(c_{\pi_8}\langle +| - s_{\pi_8}\langle -|) = \begin{bmatrix} c_{\pi_8}^2 & -c_{\pi_8}s_{\pi_8} \\ -s_{\pi_8}c_{\pi_8} & s_{\pi_8}^2 \end{bmatrix}_{\{|+\rangle,|-\rangle\}}$$

where the matrix representations are expressed in the basis indicated in subscript.

Using this measurement on channel outputs ρ_x^B induces a classical channel $p_{Y|X}^{(b)}$ with transition probabilities

$$p_{Y|X}^{(b)}(0|0) = c_{\pi_8}^2, \quad p_{Y|X}^{(b)}(1|0) = s_{\pi_8}^2, \quad p_{Y|X}^{(b)}(1|1) = c_{\pi_8}^2, \quad p_{Y|X}^{(b)}(0|1) = s_{\pi_8}^2, \quad (3.27)$$

which corresponds to a binary symmetric channel (BSC) with crossover probability $p_e = s_{\pi_8}^2 = \sin^2(\pi/8)$ and success probability $p_s = c_{\pi_8}^2$. The capacity of this BSC is given by:

$$C^{(b)}(\mathcal{N}) = 1 - H(p_s) = 1 - H(\cos^2(\pi/8)) \approx 0.3991. \quad (3.28)$$

c) Holevo limit: The HSW Theorem tells us the *ultimate* capacity of this channel is given by

$$\mathcal{C}^{(c)}(\mathcal{N}) \equiv \max_{p_X} H\left(\sum_x p_X(x)\rho_x^B\right) - \sum_x p_X(x)H(\rho_x^B). \quad (3.29)$$

In our case, the capacity is achieved using the uniform input distribution. The capacity for this channel using a quantum measurement is therefore:

$$\mathcal{C}^{(c)}(\mathcal{N}) = H_2(\cos^2(\pi/8)) \approx 0.6009. \quad (3.30)$$

In general, a collective measurement on blocks of n outputs of the channel are required to achieve the capacity. This means that the POVM operators $\{\Lambda_{x^n}^{B^n}\}$ cannot be written as a tensor product of measurement operators on the individual output systems. The channel capacity can be achieved using the random coding approach and the square root measurement based on conditionally typical projectors as shown in the proof of Theorem 3.2.

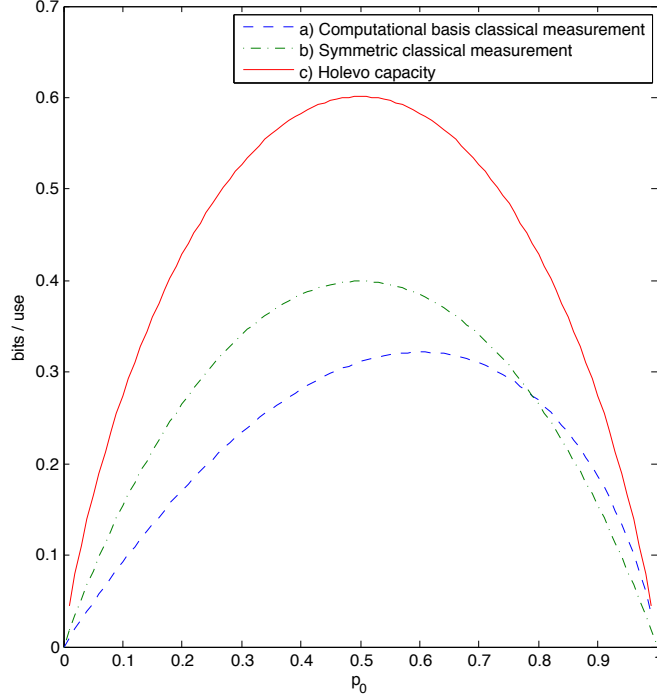


Figure 3.5: Plot of the achievable rates for the point-to-point channel ρ_x^B given by the map $0 \rightarrow |0\rangle\langle 0|^B, 1 \rightarrow |+\rangle\langle +|^B$ under three models. The horizontal axis corresponds to the parameter $p_0 = p_X(0)$ of the input distribution. The first model treats each output of the channel as a classical bit $Y^{(a)} \in \{0, 1\}$ corresponding to the output of a measurement in the computational basis: $\{\Lambda_y^{(a)}\}_{y \in \{0,1\}} = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. The mutual information $I(X; Y^{(a)})$ for all input distributions p_X is plotted as a dashed line. Under this model, the channel \mathcal{N} corresponds to a classical Z -channel. A better approach is to use a symmetric measurement with output denoted as $Y^{(b)}$, which corresponds to a classical binary symmetric channel. The mutual information $I(X; Y^{(b)})$ is plotted as a dot-dashed line. The best coding strategy is to use block measurements. The Holevo quantity $H(\sum_x p_X(x)\rho_x^B) - \sum_x p_X(x)H(\rho_x^B)$ for all input distributions is plotted as a solid line. The capacity of the channel under each model is given by the maximum of each function curve: $C^{(a)}(\mathcal{N}) \approx 0.3219$, $C^{(b)}(\mathcal{N}) \approx 0.3991$, and $C^{(c)}(\mathcal{N}) = H_2(\cos^2(\pi/8)) \approx 0.6009$. For this particular channel the quantum decoding strategy leads to a 50% improvement in the achievable communication rates relative to the best classical strategy.

3.4 Discussion

This chapter introduced the key concepts of the classical and quantum channel coding paradigms. The situation considered in Example 3.1 serves as an illustration of the potential benefits that exist for modelling communication channels using quantum mechanics.

The key take-away from this chapter is that collective measurements on blocks of channel outputs are necessary in order to achieve the *ultimate* capacity of classical-quantum communication channels, and that classical strategies which measure the channel outputs individually are suboptimal. The increased capacity is perhaps the most notable difference that exists between the classical and classical-quantum paradigms for communication [Gam].

In the remainder of this thesis, we will study multiuser classical-quantum communication models and see various coding strategies, measurement constructions and error analysis techniques which are necessary in order to prove coding theorems.

Chapter 4

Multiple access channels

The multiple access channel is a communication model for situations in which multiple senders are trying to transmit information to a single receiver. To fully solve the multiple access channel problem is to characterize all possible transmission rates for the senders which are decodable by the receiver. We will see that there is a natural tradeoff between the rates of the senders; the louder that one of the senders “speaks,” the more difficult it will be for the receiver to “hear” the other senders.

4.1 Introduction

The classical multiple access channel $\mathcal{N}^{X_1 X_2 \rightarrow Y}$ is a triple $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{N}(x_1, x_2) \equiv p_{Y|X_1 X_2}(y|x_1, x_2), \mathcal{Y})$, where \mathcal{X}_1 and \mathcal{X}_2 are the input alphabets for the two senders, \mathcal{Y} is the output alphabet and $p_{Y|X_1 X_2}(y|x_1, x_2)$ is a conditional probability distribution which describes the channel behaviour.

Our task is to characterize the communication rates (R_1, R_2) that are achievable from Sender 1 to the receiver and from Sender 2 to the receiver.

Example 4.1. Consider a situation in which two senders use laser light pulses to communicate to a distant receiver equipped with an optical instrument and a photodetector. In each time instant, Sender 1 can choose to send either a weak pulse of light or a strong pulse: $\mathcal{X}_1 = \{-, -\}$. Sender 2 similarly has two possible inputs

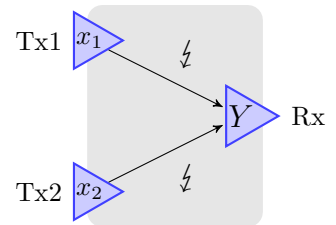


Figure 4.1: A classical multiple access channel.

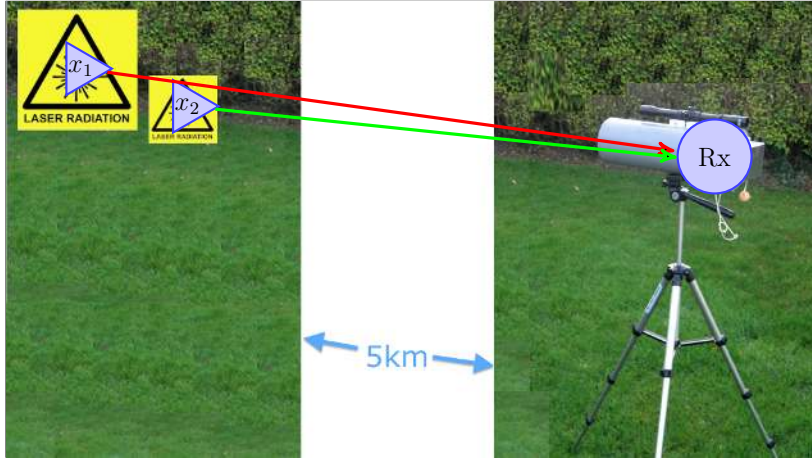


Figure 4.2: A real-world multiple access channel \mathcal{N}_1 .

$\mathcal{X}_2 = \{-, -\}$. The receiver measures the light intensity coming into the telescope, and we model his reading as the following output space $\mathcal{Y} = \{-, -, \text{---}\}$. The output signal is the sum of the incoming signals: $Y = X_1 + X_2$. We have $p_{Y|X_1X_2}(\cdot|\cdot, \cdot) = 1$, $p_{Y|X_1X_2}(-|\cdot, \cdot) = p_{Y|X_1X_2}(-|\cdot, -) = 1$ and $p_{Y|X_1X_2}(\text{---}|\cdot, -) = 1$.

The rate pair $(R_1, R_2) = (1, 0)$ is achievable if we force Sender 2 to always send a constant input. The resulting channel between Sender 1 and the receiver is a noiseless binary channel. The rate $(0, 1)$ is similarly achievable if we fix Sender 1's input. A natural question is to ask what other rates are achievable for this communication channel.

Note that the model used to describe the above communication scenario is very crude and serves only as a first approximation, which we use to illustrate the basic ideas of multiple access communication. In Section 4.1.2, we will consider more general models for multiple access channels, which allow the channel outputs to be quantum systems. In Chapter 8, we will refine the model further by taking into account certain aspects of quantum optics.

4.1.1 Review of classical results

The multiple access channel is one of the first multiuser communications problems ever considered [Sha61]. It is also one of the rare problems in network information theory where a full capacity result is known, i.e., the best known achievable rate region matches a proven outer bound. The multiple access channel plays an important role

as a building block for other network communication scenarios.

The capacity region of the classical discrete memoryless multiple access channel (DM-MAC) was established by Ahlswede [Ahl71, Ahl74a] and Liao [Lia72]. Consider the classical multiple access channel with two senders described by $\mathcal{N} = (\mathcal{X}_1 \times \mathcal{X}_2, p_{Y|X_1X_2}, \mathcal{Y})$. The capacity region for this channel is given by

$$\mathcal{C}_{\text{MAC}}(\mathcal{N}) \equiv \bigcup_{p_{X_1}, p_{X_2}} \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \left| \begin{array}{l} R_1 \leq I(X_1; Y|X_2) \\ R_2 \leq I(X_2; Y|X_1) \\ R_1 + R_2 \leq I(X_1X_2; Y) \end{array} \right. \right\},$$

where $p_{X_1} \in \mathcal{P}(\mathcal{X}_1)$, $p_{X_2} \in \mathcal{P}(\mathcal{X}_2)$ and the mutual information quantities are taken with respect to the joint input-output distribution

$$p_{X_1X_2Y}(x_1, x_2, y) \equiv p_{X_1}(x_1)p_{X_2}(x_2)p_{Y|X_1X_2}(y|x_1, x_2). \quad (4.1)$$

Note that the input distribution is chosen to be a product distribution $p_{X_1}p_{X_2}$, which reflects the assumption that the two senders are spatially separated and act independently. We can calculate the exact capacity region of any multiple access channel by evaluating the mutual information expressions for all possible input distributions and taking the union.

Example 4.1 (continued). The capacity region for the multiple access channel \mathcal{N}_1 described in Example 4.1 is given by:

$$\mathcal{C}_{\text{MAC}}(\mathcal{N}_1) = \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \left| \begin{array}{l} R_1 \leq 1 \\ R_2 \leq 1 \\ R_1 + R_2 \leq 1.5 \end{array} \right. \right\}. \quad (4.2)$$

To see how the rate pair $(1, 0.5)$ can be achieved consider an encoding strategy where each sender generates codebooks according to the uniform probability distribution and the receiver decodes the messages from Sender 2 first, followed by the messages from Sender 1. The effective channel from Sender 2 to the receiver when the input of Sender 1 is unknown corresponds to a symmetric binary erasure channel with erasure probability $\frac{1}{2}$. This is because when the receiver's output is “.” or “—” there is no ambiguity about what was sent. The output “-” could arise in two different ways, so we treat it as an erasure. The capacity of this channel is 0.5 bits per channel use

[CT91, Example 14.3.3]. Assuming the receiver correctly decodes the codewords from Sender 2, the resulting channel from Sender 1 to the receiver is a binary noiseless channel which has capacity one. To achieve the rate pair $(0.5, 1)$ we must generate codebooks at the appropriate rates and use the opposite decoding order. The capacity region is illustrated in the following figure.

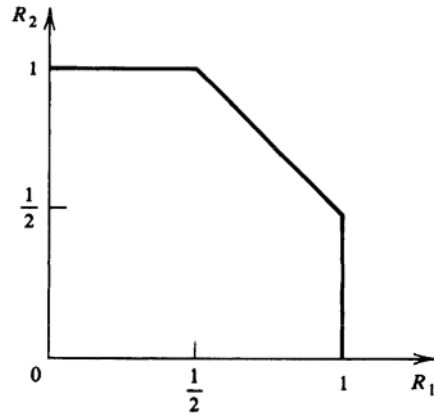


Figure 4.3: The capacity region of the adder channel.

The above example illustrates the key aspect of the multiple access channel problem: the trade off between the communication rates of the senders.

4.1.2 Quantum multiple access channels

The communication model used to evaluate the capacity in Example 4.1 is classical. We modelled the detection of light intensity in a classical way and ignored details of the quantum measurement process.

The capacity result of Ahlswede and Liao is therefore a result which depends on the classical model which we used. Better communication rates might be possible if we choose to model the quantum degrees of freedom in the communication channel. In Example 3.1, we saw how the *quantum* analysis of the detection aspects of the communication protocol can lead to improved communication rates for point-to-point channels. In this chapter, we pursue the study of quantum decoding strategies in the *multiple access* setting.

A classical-quantum multiple access channel is defined as the most general map with two classical inputs and one quantum output:

$$(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{N}^{X_1 X_2 \rightarrow B}(x_1, x_2) \equiv \rho_{x_1, x_2}^B, \mathcal{H}^B).$$

Our intent is to quantify the communication rates that are possible for classical communication from each of the two senders to the receiver. The main difference with the classical case is that the decoding operation we will use is a quantum measurement (POVM). We have to find the *rate region* for pairs (R_1, R_2) such that the following interconversion can be achieved:

$$n \cdot \mathcal{N}^{X_1 X_2 \rightarrow B} \xrightarrow{(1-\epsilon)} nR_1 \cdot [c^1 \rightarrow c] + nR_2 \cdot [c^2 \rightarrow c]. \quad (4.3)$$

The above expression states that n instances of the channel can be used to carry nR_1 classical bits from Sender 1 to the receiver (denoted $[c^1 \rightarrow c]$) and nR_2 bits from Sender 2 to the receiver (denoted $[c^2 \rightarrow c]$). The communication protocol succeeds with probability $(1 - \epsilon)$ for any $\epsilon > 0$ and sufficiently large n .

The problem of classical communication over a classical-quantum multiple-access channel was solved by Winter [Win01]. He provided single-letter formulas for the capacity region, which can be computed as an optimization over the choice of input distributions for the senders. We will discuss Winter's result and proof techniques in Section 4.2.

Note that there exist other quantum multiple access communication scenarios that can be considered. The bosonic multiple access channel was studied in [Yen05b]. The transmission of quantum information over a quantum multiple access channel was considered in [YDH05, Yar05, YHD08]. The quantum multiple access problem has also been considered in the entanglement-assisted setting [HDW08, XW11]. In this chapter, as in the rest of the thesis, we restrict our attention to the problem of classical communication over classical-quantum channels.

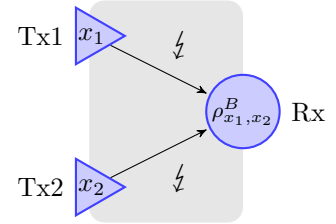


Figure 4.4: A quantum multiple access channel with two senders. The output of the channel are conditional quantum states $\mathcal{N}^B(x_1, x_2) \equiv \rho_{x_1, x_2}^B$.

4.1.3 Information processing task

To show that a certain rate pair (R_1, R_2) is achievable we must construct an end-to-end coding scheme that the two senders and the receiver can employ to communicate with each other. In this section we specify precisely the different steps involved in the transmission process.

Sender 1 will send a message m_1 chosen from the message set $\mathcal{M}_1 \equiv \{1, 2, \dots, |\mathcal{M}_1|\}$ where $|\mathcal{M}_1| = 2^{nR_1}$. Sender 2 similarly chooses a message m_2 from a message set $\mathcal{M}_2 \equiv \{1, 2, \dots, |\mathcal{M}_2|\}$ where $|\mathcal{M}_2| = 2^{nR_2}$. Senders 1 and 2 encode their messages as codewords $x_1^n(m_1) \in \mathcal{X}_1^n$ and $x_2^n(m_2) \in \mathcal{X}_2^n$, which are then input to the channel.

The output of the channel is an n -fold tensor product state of the form:

$$\mathcal{N}^{\otimes n}(x_1^n(m_1), x_2^n(m_2)) \equiv \rho_{x_1^n(m_1), x_2^n(m_2)}^{B^n} \in \mathcal{D}(\mathcal{H}^{B^n}). \quad (4.4)$$

In order to recover the messages m_1 and m_2 , the receiver performs a positive operator valued measure (POVM) $\{\Lambda_{m_1, m_2}\}_{m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2}$ on the output of the channel B^n . We denote the measurement outputs as M'_1 and M'_2 . An error occurs whenever the receiver measurement outcomes differ from the messages that were sent. The overall probability of error for message pair (m_1, m_2) is

$$\begin{aligned} p_e(m_1, m_2) &\equiv \Pr \{(M'_1, M'_2) \neq (m_1, m_2)\} \\ &= \text{Tr} \left[(I - \Lambda_{m_1, m_2}) \rho_{x_1^n(m_1), x_2^n(m_2)}^{B^n} \right], \end{aligned}$$

where the measurement operator $(I - \Lambda_{m_1, m_2})$ represents the complement of the correct decoding outcome.

Definition 4.1. An (n, R_1, R_2, ϵ) code for the multiple access channel consists of two codebooks $\{x_1^n(m_1)\}_{m_1 \in \mathcal{M}_1}$ and $\{x_2^n(m_2)\}_{m_2 \in \mathcal{M}_2}$, and a decoding POVM $\{\Lambda_{m_1, m_2}\}_{m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2}$, such that the average probability of error \bar{p}_e is bounded from above by ϵ :

$$\bar{p}_e \equiv \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} \sum_{m_1, m_2} p_e(m_1, m_2) \leq \epsilon. \quad (4.5)$$

A rate pair (R_1, R_2) is *achievable* if there exists an $(n, R_1 - \delta, R_2 - \delta, \epsilon)$ quantum multiple access channel code for all $\epsilon, \delta > 0$ and sufficiently large n . The *capacity region*

$\mathcal{C}_{\text{MAC}}(\mathcal{N})$ is the closure of the set of all achievable rates.

4.1.4 Chapter overview

Suppose we have a two-sender classical-quantum multiple access channel and the two messages m_1 and m_2 were sent. This chapter studies the different decoding strategies that can be used by the receiver in order to decode the messages.

The technique used by Winter to prove the achievability of the rates in the capacity region of the quantum multiple access channel is called *successive decoding*. In this approach, the receiver can achieve one of the corner points of the rate region by decoding the messages in the order “ $m_1 \rightarrow m_2|m_1$ ”. In doing so, the best possible rate R_2 is achieved, because the receiver will have the side information of m_1 , and by extensions $x_1^n(m_1)$, when decoding the message m_2 . This approach is also referred to as *successive cancellation* for channels with continuous variable inputs and additive white Gaussian noise (Gaussian channels) where the first decoded signal can be *subtracted* from the received signal. The other corner point can be achieved by decoding in the opposite order “ $m_2 \rightarrow m_1|m_2$ ”. These codes can be combined with *time-sharing* and *resource wasting* to achieve all other points in the rate region. We will discuss this strategy in further detail in Section 4.2 below.

Another approach is to use simultaneous decoding which requires no time-sharing. We denote the simultaneous decoding of the messages m_1 and m_2 as “ (m_1, m_2) ”. As far as the QMAC problem is concerned the two approaches yield equivalent achievable rate regions. However, if the QMAC code is to be used as part of a larger protocol (like a code for the interference channel for example) then the simultaneous decoding approach is much more powerful.

The main contribution in this chapter is Theorem 4.2 in Section 4.3, which shows that simultaneous decoding for the classical-quantum multiple access channel with two senders is possible. This result and the techniques developed for its proof will form the key building blocks for the subsequent chapters in this thesis. We will also comment on the difficulties in extending the simultaneous decoding approach to more than two senders (Conjecture 4.1). In Section 4.4, we will briefly discuss a third coding strategy for the QMAC called *rate-splitting*.

4.2 Successive decoding

Winter found a single-letter formula for the capacity of the classical-quantum multiple access channel with M senders [Win01]. We state the result here for two senders.

Theorem 4.1 (Theorem 10 in [Win01]). *The capacity region for the classical-quantum multiple access channel $(\mathcal{X}_1 \times \mathcal{X}_2, \rho_{x_1, x_2}^B, \mathcal{H}^B)$ is given by*

$$\mathcal{C}_{MAC} = \bigcup_{p_{X_1}, p_{X_2}} \{ (R_1, R_2) \in \mathbb{R}_+^2 \mid \text{Eqns. (4.7)-(4.9)} \} \quad (4.6)$$

$$R_1 \leq I(X_1; B | X_2)_\theta, \quad (4.7)$$

$$R_2 \leq I(X_2; B | X_1)_\theta, \quad (4.8)$$

$$R_1 + R_2 \leq I(X_1 X_2; B)_\theta, \quad (4.9)$$

where the information quantities are taken with respect to the classical-quantum state:

$$\theta^{X_1 X_2 B} \equiv \sum_{x_1, x_2} p_{X_1}(x_1) p_{X_2}(x_2) |x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_2} \otimes \rho_{x_1, x_2}^B. \quad (4.10)$$

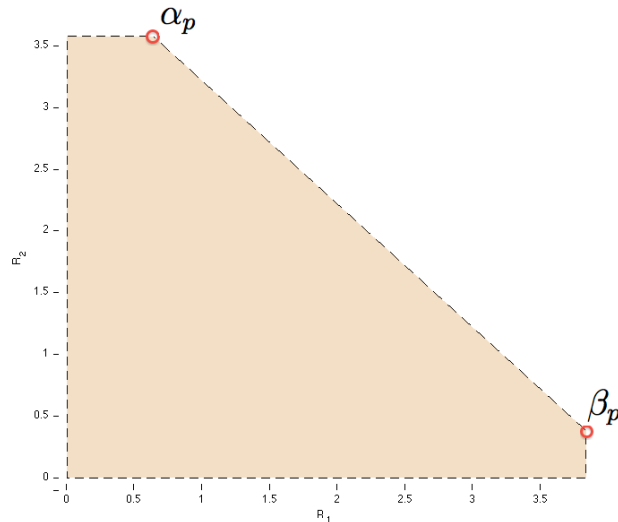


Figure 4.5: The rates achievable by successive decoding correspond to the dominant vertices of the rate region α_p and β_p . Rates in between these points can be achieved by *time-sharing* between the strategies for the two corners.

For a given choice of input probability distribution $p \equiv p_{X_1}, p_{X_2}$, the achievable rate region, $\mathcal{R}(\mathcal{N}, p)$, has the form of a pentagon bounded by the three in-

equalities in equations (4.7)-(4.9) and two rate positivity conditions. The two dominant vertices of this rate region have coordinates $\alpha_p \equiv (I(X_1; B)_\theta, I(X_2; B|X_1)_\theta)$ and $\beta_p \equiv (I(X_1; B|X_2)_\theta, I(X_2; B)_\theta)$ and correspond to two alternate successive decoding strategies. The portion of the line $R_1 + R_2 = I(X_1 X_2; B)_\theta$ which lies in between the points α_p and β_p will be referred to as the *dominant facet*.

In order to show achievability of the entire rate region, Winter proved that each of the corner points of the region is achievable. By the use of *time-sharing* we can achieve any point on the dominant facet of the region, and we can use *resource wasting* to achieve all the points on the interior of the region. It follows that the entire rate region is achievable. We show some of the details of Winter's proof below.

Proof sketch. We will use a random coding approach for the codebook construction and point-to-point decoding measurements based on the conditionally typical projectors.

Fix the input distribution $p = p_{X_1}(x_1)p_{X_2}(x_2)$ and choose the rates so that they correspond to the rate point α_p :

$$R_1 = I(X_1; B)_\theta - \delta, \quad R_2 = I(X_2; B|X_1)_\theta - \delta. \quad (4.11)$$

Codebook construction: Randomly and independently generate 2^{nR_1} sequences $x_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, according to $\prod_{i=1}^n p_{X_1}(x_{1i})$. Similarly generate randomly and independently the codebook $\{x_2^n(m_2)\}$, $m_2 \in [1 : 2^{nR_2}]$ according to $\prod_{i=1}^n p_{X_2}(x_{2i})$.

Decoding: When the message pair (m_1, m_2) is sent, the output of the channel will be $\rho_{x_1^n(m_1), x_2^n(m_2)}$. Let $\Pi_{\rho_{x_1^n(m_1), x_2^n(m_2)}, \delta}^n$ be the conditionally typical projector for that state. In order to define the other typical projectors necessary for the decoding, we define the

4.2 Successive decoding

following expectations of the output state:

$$\begin{aligned}\bar{\rho}_{x_1^n(m_1)} &\equiv \sum_{x_2^n} p_{X_2^n}(x_2^n) \rho_{x_1^n(m_1), x_2^n} = \bigotimes_{i=1}^n \left(\sum_{\mu} p_{X_2}(\mu) \rho_{x_{1i}(m_1), \mu} \right) \\ &= \mathbb{E}_{X_2^n} \left\{ \rho_{x_1^n(m_1), X_2^n} \right\}, \\ \bar{\rho}^{\otimes n} &\equiv \sum_{x_1^n, x_2^n} p_{X_1^n}(x_1^n) p_{X_2^n}(x_2^n) \rho_{x_1^n, x_2^n} = \bigotimes_{i=1}^n \left(\sum_{\tau, \mu} p_{X_1}(\tau) p_{X_2}(\mu) \rho_{\tau, \mu} \right) \\ &= \mathbb{E}_{X_1^n, X_2^n} \left\{ \rho_{X_1^n, X_2^n} \right\}.\end{aligned}$$

The state $\bar{\rho}_{x_1^n(m_1)}$ corresponds to the receiver's output if he treats the codewords of Sender 2 as noise to be averaged over. The state $\bar{\rho}^{\otimes n}$ corresponds to the average output state for a random code constructed according to $p_{X_1} p_{X_2}$. Let $\Pi_{\bar{\rho}_{x_1^n(m_1), \delta}}^n \equiv \Pi_{\bar{\rho}_{x_1^n(m_1), \delta}}^{B^n}$ be the conditionally typical projector for $\bar{\rho}_{x_1^n(m_1)}$ and let $\Pi_{\bar{\rho}}^n \equiv \Pi_{\bar{\rho}^{\otimes n}, \delta}^{B^n}$ be the typical projector for the state $\bar{\rho}^{\otimes n}$.

To achieve the rates of α_p , the receiver will decode the messages in the order “ $m_1 \rightarrow m_2 | m_1$ ” using a successive decoding procedure. The first step is to use a quantum instrument $\{\Upsilon_{m_1}^\alpha\}$ which acts as follows on any state defined on B^n :

$$\Upsilon^\alpha : \psi^{B^n} \longrightarrow \sum_{m_1} |m_1\rangle\langle m_1|^{M_1} \otimes \left(\frac{\sqrt{\Lambda_{m_1}^\alpha} \psi^{B^n} \sqrt{\Lambda_{m_1}^\alpha}}{\text{Tr}[\Lambda_{m_1}^\alpha \psi^{B^n}]} \right)^{B^n}. \quad (4.12)$$

The POVM operators $\{\Lambda_{m_1}^\alpha\}$ are constructed using the typical projector sandwich

$$\Pi_{\bar{\rho}}^n \Pi_{\bar{\rho}_{x_1^n(m_1), \delta}}^n \Pi_{\bar{\rho}}^n, \quad (4.13)$$

and normalized using the square root measurement approach in order to satisfy $\Lambda_{m_1}^\alpha \geq 0$, $\sum_{m_1} \Lambda_{m_1}^\alpha = I$. The purpose of the quantum instrument is to extract the message m_1 and store it in the register M_1 , but also leave behind a system in B^n which can be processed further.

An error analysis similar to that of the HSW theorem shows that the quantum instrument $\{\Upsilon_{m_1}^\alpha\}$ will correctly decode the message m_1 with high probability. This is because we chose the rate for the m_1 codebook to be $R_1 = I(X_1; B)_\theta - \delta$. Furthermore, it can be shown using the *gentle operator lemma for ensembles* (Lemma 3.2), that the state which remains in the system B^n is negligibly disturbed in the process.

The receiver will then perform a second measurement to recover the message m_2 . The second measurement is a POVM $\{\Lambda_{m_2|m_1}^\alpha\}$ constructed from the projectors

$$\Pi_{\bar{\rho}_{x_1^n(m_1),\delta}^n} \Pi_{\rho_{x_1^n(m_1),x_2^n(m_2)}^n} \Pi_{\bar{\rho}_{x_1^n(m_1),\delta}^n}, \quad (4.14)$$

and appropriately normalized. Note that this measurement is chosen conditionally on the codeword $X_1^n(m_1)$ that Sender 1 input to the channel. This is because, when the correct message m_1 is decoded in the first step, the receiver can infer the codeword which Sender 1 input to the channel. Thus, after the first step, the effective channel from Sender 2 to the receiver is

$$(X_1^n, x_2^n) \rightarrow (X_1^n, \rho_{X_1^n, x_2^n}^{B^n}), \quad (4.15)$$

where X_1^n is a random variable distributed according to $\prod_{i=1}^n p_{X_1}$. This is a setting in which the quantum packing lemma can be applied. By substituting $U^n = X_1^n$ and $X^n = X_2^n$ into Lemma B.1, we conclude that if we choose the rate to be $R_2 = I(X_2; B|X_1)_\theta - \delta$, then the message m_2 will be decoded correctly with high probability.

The rate point β_p corresponds to the alternate decode ordering where the receiver decodes the message m_2 first and m_1 second. All other rate pairs in the region can be obtained from the corner points α_p and β_p by using *time-sharing* and *resource wasting*. \square

Note that one of the key ingredients in the proof was the use of Lemma 3.2, which guarantees that the act of decoding m_1 does not disturb the state too much. This step of our quantum decoding procedure may be counterintuitive at a first glance, since quantum mechanical measurements are usually described as processes in which the quantum system is disturbed. Any retrieval of data from a quantum system inevitably disturbs the state of the system, so the second measurement, which the receiver performs on the system B^n , may fail if the first measurement has disturbed the state too much. The *gentle measurement lemma* guarantees that very little information disturbance to the state occurs when there is one measurement outcome that is very likely. When the state of the receiver is $\rho_{x_1^n, x_2^n}^{B^n}$, we can be almost certain that the outcome of the quantum instrument $\{\Upsilon_{m_1}^\alpha\}$ is going to be m_1 . Therefore, this process leaves the state in B^n only slightly disturbed.

The proof technique in Theorem 4.1 generalizes to the case of the M -sender MAC, which has $M!$ dominant vertices. Each vertex corresponds to one permutation of the decode ordering.

4.3 Simultaneous decoding

Another approach for achieving the capacity of the multiple access channel, which does not use time-sharing, is simultaneous decoding. In the classical version of this decoding strategy, the receiver will report (m_1, m_2) if he finds a unique pair of codewords $X_1^n(m_1)$ and $X_2^n(m_2)$ which are jointly typical with the output of the channel Y^n :

$$(X_1^n(m_1), X_2^n(m_2), Y^n) \in \mathcal{J}_\epsilon^{(n)}(X_1, X_2, Y). \quad (4.16)$$

Assuming the messages m_1 and m_2 are sent, we categorize the different kinds of *wrong message* decode errors that may occur.

error	\hat{M}_1	\hat{M}_2	
(E1)	*	m_2	
(E2)	m_1	*	
(E12)	*	*	(4.17)

The * in the above table denotes any message other than the one which was sent. The analysis of the *classical* simultaneous decoder uses the properties of the jointly typical sequences and the randomness in the codebooks. Recall that a multi-variable sequence is jointly typical if and only if all the sequences in the subsets of the variables are jointly typical. Thus, the condition $(X_1^n(m_1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)}(X_1, X_2, Y)$ implies that:

$$(X_1^n(m_1), Y^n) \in \mathcal{T}_\epsilon^{(n)}(X_1, Y), \quad (4.18)$$

$$(X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)}(X_2, Y), \quad (4.19)$$

$$Y^n \in \mathcal{T}_\epsilon^{(n)}(Y). \quad (4.20)$$

Starting from these conditions, it is straightforward to bound the probability of the different decoding error events using the properties of the jointly typical sequences

[EGK10].

In the quantum case, we can similarly identify three different error terms, the probabilities of which can be bounded by using the properties of the conditionally typical projectors. If we can construct a quantum measurement operator that “contains” all the typical projectors so that we can obtain the appropriate averages of the output state in the error analysis, then we would have a proof that simultaneous decoding is possible.

If only things were so simple! The construction of a simultaneous decoding POVM turns out to be a difficult problem. Despite being built out of the same typical projectors, the operator constructed according to

$$\Lambda_{m_1, m_2} \propto \Pi_{\bar{\rho}_{x_2^n(m_2)}}^n \Pi_{\bar{\rho}_{x_1^n(m_1)}}^n \Pi_{\rho_{x_1^n(m_1), x_2^n(m_2)}}^n \Pi_{\bar{\rho}_{x_1^n(m_1)}}^n \Pi_{\bar{\rho}_{x_2^n(m_2)}}^n, \quad (4.21)$$

is different from the operator

$$\Lambda'_{m_1, m_2} \propto \Pi_{\bar{\rho}_{x_1^n(m_1)}}^n \Pi_{\bar{\rho}_{x_2^n(m_2)}}^n \Pi_{\rho_{x_1^n(m_1), x_2^n(m_2)}}^n \Pi_{\bar{\rho}_{x_2^n(m_2)}}^n \Pi_{\bar{\rho}_{x_1^n(m_1)}}^n, \quad (4.22)$$

because the different typical projectors do not commute in general. In fact, there is very little we can say about the relationship between the subspaces spanned by the two averaged typical projectors: $\Pi_{\bar{\rho}_{x_1^n(m_1)}}^n$ and $\Pi_{\bar{\rho}_{x_2^n(m_2)}}^n$. This is a problem because, for one of the error terms in the analysis, we would like to have $\Pi_{\bar{\rho}_{x_2^n(m_2)}}^n$ on the “outside” as in (4.21) so that we can use Property 2.46 of typical projectors to obtain a factor $2^{nH(B|X_2)}$. For another error term, we want $\Pi_{\bar{\rho}_{x_1^n(m_1)}}^n$ to be on the outside as in (4.22) in order to be able to do the averaging in the alternate order to obtain a term of the form $2^{nH(B|X_1)}$. Thus it would seem, and originally it seemed so to my colleagues and me, that the construction of a simultaneous decoding POVM for which we can bound the probability of all error events might be a difficult task.

Quantum simultaneous decoding actually *is* possible, and this is what we will show in this section for the case of the multiple access channel with two senders. Our proof techniques do not generalize readily to quantum multiple access channels with more than two independent senders. At the end of this section we will formulate Conjecture 4.1 regarding the existence of a simultaneous decoder for three-sender multiple access channels, which will be required for the proof of Theorem 5.3 in the next chapter.

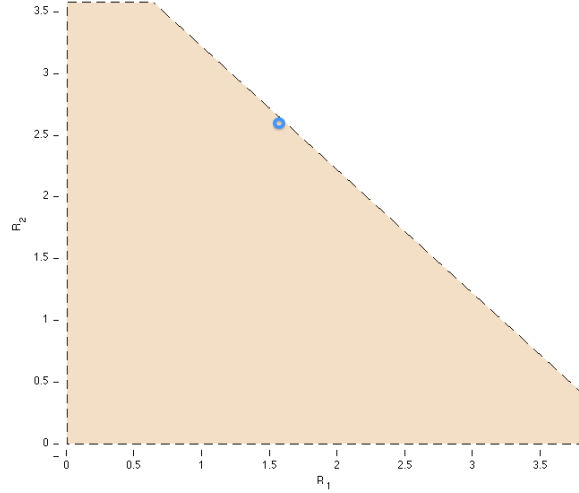


Figure 4.6: Simultaneous decoding strategy. Simultaneous decoding of the two messages is more powerful than successive decoding, because it allows us to achieve any rate pair (R_1, R_2) of the capacity region without the need for time-sharing.

Theorem 4.2 (Two-sender quantum simultaneous decoding). *Let $(\mathcal{X}_1 \times \mathcal{X}_2, \rho_{x_1, x_2}^B, \mathcal{H}^B)$ be a quantum multiple access channel with two senders and a single receiver, and let $p = p_{X_1} p_{X_2}$ be a choice for the input code distribution. Let $\{X_1^n(m_1)\}_{m_1 \in \{1, \dots, |\mathcal{M}_1|\}}$ and $\{X_2^n(m_2)\}_{m_2 \in \{1, \dots, |\mathcal{M}_2|\}}$ be random codebooks generated according to the product distributions $p_{X_1^n}$ and $p_{X_2^n}$. There exists a simultaneous decoding POVM $\{\Lambda_{m_1, m_2}\}_{m_1 \in \mathcal{M}_1, m_2 \in \mathcal{M}_2}$, with expected average probability of error bounded from above by ϵ for all $\epsilon, \delta > 0$ and sufficiently large n , provided the rates R_1, R_2 satisfy the inequalities*

$$R_1 \leq I(X_1; B | X_2)_\theta, \tag{4.23}$$

$$R_2 \leq I(X_2; B | X_1)_\theta, \tag{4.24}$$

$$R_1 + R_2 \leq I(X_1 X_2; B)_\theta, \tag{4.25}$$

where the state $\theta^{X_1 X_2 B}$ is defined in (4.10).

The main difference between the coding strategy employed by Winter in the proof of Theorem 4.1 and Theorem 4.2 above is that the latter does not require the use of time-sharing. Using the simultaneous decoding approach we can achieve any of the rates in the QMAC capacity region using a *single* codebook, whereas time-sharing requires us to switch between the two codebooks for the vertices. This distinction

is minor in the context of the multiple access channel problem, but it will become important in situations where there are multiple receivers as in the compound multiple access channel and the interference channel. Note that Sen gave an alternate proof of Theorem 4.2 using a different approach [Sen12a].

Proof of Theorem 4.2. The proof proceeds by random coding arguments using the properties of projectors onto the typical subspaces of the output states and the *square root* measurement.

Consider some choice $p = p_{X_1}(x_1)p_{X_2}(x_2)$ for the input distributions.

Codebook construction: Randomly and independently generate 2^{nR_1} sequences $x_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, according to $\prod_{i=1}^n p_{X_1}(x_{1i})$. Similarly, generate randomly and independently the codebook $\{x_2^n(m_2)\}$, $m_2 \in [1 : 2^{nR_2}]$, according to $\prod_{i=1}^n p_{X_2}(x_{2i})$.

POVM construction: In order to lighten the notation, the channel output will be denoted with the shorthand $\rho_{m_1, m_2} \equiv \rho_{x_1^n(m_1), x_2^n(m_2)}$, when the inputs to the channel are $x_1^n(m_1)$ and $x_2^n(m_2)$. Let $\Pi_{m_1, m_2}^n \equiv \Pi_{\rho_{x_1^n(m_1), x_2^n(m_2)}, \delta}^n$ be the conditionally typical projector for that state. Consider the following averaged output states:

$$\bar{\rho}_{x_1} \equiv \sum_{x_2} p_{X_2}(x_2) \rho_{x_1, x_2}, \quad (4.26)$$

$$\bar{\rho}_{x_2} \equiv \sum_{x_1} p_{X_1}(x_1) \rho_{x_1, x_2}, \quad (4.27)$$

$$\bar{\rho} \equiv \sum_{x_1, x_2} p_{X_1}(x_1) p_{X_2}(x_2) \rho_{x_1, x_2}. \quad (4.28)$$

Let $\Pi_{m_1}^n \equiv \Pi_{\bar{\rho}_{x_1^n(m_1)}, \delta}^n$ be the conditionally typical projector for the tensor product state $\bar{\rho}_{m_1} \equiv \bar{\rho}_{x_1^n(m_1)}$ defined by (4.26) for n uses of the channel. Let $\Pi_{m_2}^n \equiv \Pi_{\bar{\rho}_{x_2^n(m_2)}, \delta}^n$ be the conditionally typical projector for the tensor product state $\bar{\rho}_{m_2} \equiv \bar{\rho}_{x_2^n(m_2)}$ defined by (4.27) and finally let $\Pi_{\bar{\rho}, \delta}^n$ be the typical projector for the state $\bar{\rho}^{\otimes n}$ defined by (4.28).

The detection POVM $\{\Lambda_{m_1, m_2}\}$ has the following form:

$$\Lambda_{m_1, m_2} \equiv \left(\sum_{m'_1, m'_2} P_{m'_1, m'_2} \right)^{-\frac{1}{2}} P_{m_1, m_2} \left(\sum_{m'_1, m'_2} P_{m'_1, m'_2} \right)^{-\frac{1}{2}},$$

4.3 Simultaneous decoding

where

$$P_{m_1, m_2} \equiv \Pi_{\bar{\rho}, \delta}^n \Pi_{m_1}^n \Pi_{m_1, m_2}^n \Pi_{m_1}^n \Pi_{\bar{\rho}, \delta}^n, \quad (4.29)$$

is a positive operator which consists of three typical projectors “sandwiched” together. Observe that the layers of the sandwich go from the more general ones on the outside to the more specific ones on the inside. Observe also that the conditionally typical projector $\Pi_{m_2}^n$ is not included.

The average error probability of the code is given by:

$$\bar{p}_e \equiv \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} \sum_{m_1, m_2} \text{Tr}[(I - \Lambda_{m_1, m_2}) \rho_{m_1, m_2}]. \quad (4.30)$$

The first step in our error analysis is to make a substitution of the output state ρ_{m_1, m_2} with a smoothed version:

$$\tilde{\rho}_{m_1, m_2} \equiv \Pi_{m_2}^n \rho_{m_1, m_2} \Pi_{m_2}^n. \quad (4.31)$$

We do this to ensure that we will have the operator $\Pi_{m_2}^n$ inside the trace when we perform the averaging. The term *smoothing* refers to the fact that we are now coding for a different channel which has all of the $\Pi_{m_2}^n$ -atypical subspace removed, i.e., we remove the “spikes” (the large eigenvalues).

We can use the inequality

$$\text{Tr}[\Lambda \rho] \leq \text{Tr}[\Lambda \sigma] + \|\rho - \sigma\|_1 \quad (4.32)$$

from Lemma 2.1, which holds for all operators such that $0 \leq \rho, \sigma, \Lambda \leq I$, in order to bound the *smoothing penalty* which we incur as a result of the substitution.

After the substitution step (4.30) and the use of (4.32), we obtain the following bound on the probability of error:

$$\bar{p}_e \leq \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} \sum_{m_1, m_2} \left[\text{Tr}[(I - \Lambda_{m_1, m_2}) \tilde{\rho}_{m_1, m_2}] + \|\tilde{\rho}_{m_1, m_2} - \rho_{m_1, m_2}\|_1 \right]. \quad (4.33)$$

The next step is to use the Hayashi-Nagaoka operator inequality [HN03] (Lemma 3.1):

$$I - (S + T)^{-\frac{1}{2}} S (S + T)^{-\frac{1}{2}} \leq 2(I - S) + 4T.$$

Choosing $S = P_{m_1, m_2}$, $T = \sum_{(m'_1, m'_2) \neq (m_1, m_2)} P_{m'_1, m'_2}$, we apply the above operator inequality to bound the average error probability of the first term in (4.33) as:

$$\begin{aligned} \bar{p}_e \leq \frac{1}{|\mathcal{M}_1| |\mathcal{M}_2|} \sum_{m_1, m_2} & \left[2 \text{Tr}[(I - P_{m_1, m_2}) \tilde{\rho}_{m_1, m_2}] \right. \\ & \left. + 4 \sum_{(m'_1, m'_2) \neq (m_1, m_2)} \text{Tr}[P_{m'_1, m'_2} \tilde{\rho}_{m_1, m_2}] + \|\tilde{\rho}_{m_1, m_2} - \rho_{m_1, m_2}\|_1 \right]. \end{aligned} \quad (4.34)$$

The three terms in the summation have an intuitive interpretation. The first term corresponds to the case when the output state is non-typical, the second term describes the probability of a wrong message being decoded, and the third term accounts for the *smoothing penalty* which we have to pay for using a code designed for the channel $\tilde{\rho}_{m_1, m_2}$ on the channel ρ_{m_1, m_2} .

We apply a random coding argument to bound the expectation of the average error probability in (4.34). We compute the expected value of the error terms with respect to the random choice of codebook: $\{X_1^n(m_1)\}$, $\{X_2^n(m_2)\}$. Recall that in our shorthand notation, the codewords are not indicated. Thus when we say $\mathbb{E}_{X_1^n, X_2^n} \rho_{m_1, m_2}$, we really mean $\mathbb{E}_{X_1^n, X_2^n} \rho_{X_1^n(m_1), X_2^n(m_2)}$.

A bound on the first term in (4.34) follows from the following argument:

$$\begin{aligned} & \mathbb{E}_{X_1^n, X_2^n} \text{Tr}[P_{m_1, m_2} \tilde{\rho}_{m_1, m_2}] = \\ & = \mathbb{E}_{X_1^n, X_2^n} \text{Tr}[\Pi_{\tilde{\rho}, \delta}^n \Pi_{m_1}^n \Pi_{m_1, m_2}^n \Pi_{m_1}^n \Pi_{\tilde{\rho}, \delta}^n \Pi_{m_2}^n \rho_{m_1, m_2} \Pi_{m_2}^n] \\ & \geq \mathbb{E}_{X_1^n, X_2^n} \text{Tr}[\Pi_{m_1, m_2}^n \rho_{m_1, m_2}] \\ & \quad - \mathbb{E}_{X_1^n, X_2^n} \left\| \Pi_{m_2}^n \rho_{m_1, m_2} \Pi_{m_2}^n - \rho_{m_1, m_2} \right\|_1 \\ & \quad - \mathbb{E}_{X_1^n, X_2^n} \left\| \Pi_{\tilde{\rho}, \delta}^n \rho_{m_1, m_2} \Pi_{\tilde{\rho}, \delta}^n - \rho_{m_1, m_2} \right\|_1 \\ & \quad - \mathbb{E}_{X_1^n, X_2^n} \left\| \Pi_{m_1}^n \rho_{m_1, m_2} \Pi_{m_1}^n - \rho_{m_1, m_2} \right\|_1 \end{aligned}$$

4.3 Simultaneous decoding

$$\begin{aligned}
&\geq \mathbb{E}_{X_1^n, X_2^n} \text{Tr}[\Pi_{m_1, m_2}^n \rho_{m_1, m_2}] - 6\sqrt{\epsilon} \\
&\geq 1 - \epsilon - 6\sqrt{\epsilon}.
\end{aligned} \tag{4.35}$$

The first inequality follows from (4.32) (Lemma 2.1) applied three times. The second inequality follows from Lemma 3.2 and the properties of the conditionally typical projectors: (B.40), (B.41) and (B.42) given in Appendix B.1. The last inequality follows from equation (B.39).

The same reasoning is used to obtain a bound the expectation of the smoothing-penalty (the third term in (4.34)).

$$\begin{aligned}
\mathbb{E}_{X_1^n, X_2^n} \|\tilde{\rho}_{m_1, m_2} - \rho_{m_1, m_2}\|_1 &= \mathbb{E}_{X_1^n, X_2^n} \|\Pi_{m_2}^n \rho_{m_1, m_2} \Pi_{m_2}^n - \rho_{m_1, m_2}\|_1 \\
&\leq 2\sqrt{\epsilon}.
\end{aligned} \tag{4.36}$$

The main part of the error analysis consists of obtaining a bound on the second term in (4.34). This term corresponds to the probability that a wrong message pair is decoded by the receiver. We split this term into three parts, each representing a different type of decoding error:

$$\begin{aligned}
&\sum_{(m'_1, m'_2) \neq (m_1, m_2)} \text{Tr}[P_{m'_1, m'_2} \tilde{\rho}_{m_1, m_2}] = \\
&= \sum_{m'_1 \neq m_1} \text{Tr}[P_{m'_1, m_2} \tilde{\rho}_{m_1, m_2}]
\end{aligned} \tag{E1}$$

$$+ \sum_{m'_2 \neq m_2} \text{Tr}[P_{m_1, m'_2} \tilde{\rho}_{m_1, m_2}] \tag{E2}$$

$$+ \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \text{Tr}[P_{m'_1, m'_2} \tilde{\rho}_{m_1, m_2}]. \tag{E12}$$

We will bound each of these terms in turn.

Bound on (E1) : The expectation over the random choice of codebook for the error term (E1) is as follows:

$$\begin{aligned}
 \mathbb{E}_{X_1^n, X_2^n} \{(\text{E1})\} &= \mathbb{E}_{X_1^n, X_2^n} \left\{ \sum_{m'_1 \neq m_1} \text{Tr} [P_{m'_1, m_2} \tilde{\rho}_{m_1, m_2}] \right\} \\
 &\stackrel{\textcircled{1}}{=} \sum_{m'_1 \neq m_1} \mathbb{E}_{X_2^n} \left\{ \text{Tr} \left[\mathbb{E}_{X_1^n} \{P_{m'_1, m_2}\} \mathbb{E}_{X_1^n} \{\tilde{\rho}_{m_1, m_2}\} \right] \right\} \\
 &= \sum_{m'_1 \neq m_1} \mathbb{E}_{X_2^n} \left\{ \text{Tr} \left[\mathbb{E}_{X_1^n} \{P_{m'_1, m_2}\} \mathbb{E}_{X_1^n} \{\Pi_{m_2}^n \rho_{m_1, m_2} \Pi_{m_2}^n\} \right] \right\} \\
 &= \sum_{m'_1 \neq m_1} \mathbb{E}_{X_2^n} \left\{ \text{Tr} \left[\mathbb{E}_{X_1^n} \{P_{m'_1, m_2}\} \Pi_{m_2}^n \mathbb{E}_{X_1^n} \{\rho_{m_1, m_2}\} \Pi_{m_2}^n \right] \right\} \\
 &\stackrel{\textcircled{2}}{=} \sum_{m'_1 \neq m_1} \mathbb{E}_{X_1^n X_2^n} \left\{ \text{Tr} [P_{m'_1, m_2} \Pi_{m_2}^n \bar{\rho}_{m_2} \Pi_{m_2}^n] \right\} \\
 &\stackrel{\textcircled{3}}{\leq} 2^{-n[H(B|X_2) - \delta]} \sum_{m'_1 \neq m_1} \mathbb{E}_{X_1^n X_2^n} \left\{ \text{Tr} [P_{m'_1, m_2} \Pi_{m_2}^n] \right\}
 \end{aligned}$$

Equation ① follows because the codewords for m'_1 and m_1 are independent. Equality ② comes from the definition of the averaged code state $\bar{\rho}_{m_2} \equiv \bar{\rho}_{x_2^n(m_2)}$. The inequality ③ follows from the bound

$$\Pi_{m_2}^n \bar{\rho}_{m_2} \Pi_{m_2}^n \leq 2^{-n[H(B|X_2) - \delta]} \Pi_{m_2}^n.$$

We focus our attention on the expression inside the trace:

$$\begin{aligned}
 \text{Tr} [P_{m'_1, m_2} \Pi_{m_2}^n] &= \text{Tr} \left[\Pi_{\bar{\rho}, \delta}^n \Pi_{m'_1}^n \Pi_{m'_1, m_2}^n \Pi_{m'_1}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m_2}^n \right] \\
 &\stackrel{\textcircled{4}}{=} \text{Tr} \left[\Pi_{m'_1}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m_2}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m'_1}^n \Pi_{m'_1, m_2}^n \right] \\
 &\stackrel{\textcircled{5}}{\leq} \text{Tr} \left[\Pi_{m'_1, m_2}^n \right].
 \end{aligned}$$

In the first step we substituted the definition of P_{m_1, m_2} from equation (4.29). Equality ④ follows from the cyclicity of trace. Inequality ⑤ follows from

$$\Pi_{m'_1}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m_2}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m'_1}^n \leq \Pi_{m'_1}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m'_1}^n \leq \Pi_{m'_1}^n \leq I. \quad (4.37)$$

4.3 Simultaneous decoding

Next, we obtain the following bound on the expected probability of the term (E1):

$$\begin{aligned}
\mathbb{E}_{X_1^n, X_2^n} \{(\text{E1})\} &\leq 2^{-n[H(B|X_2)-\delta]} \sum_{m'_1 \neq m_1} \mathbb{E}_{X_1^n, X_2^n} \left\{ \text{Tr} \left[\Pi_{m'_1, m_2}^n \right] \right\} \\
&\stackrel{\textcircled{6}}{\leq} 2^{-n[H(B|X_2)-\delta]} \sum_{m'_1 \neq m_1} 2^{n[H(B|X_1 X_2)+\delta]} \\
&\leq |\mathcal{M}_1| 2^{-n[I(X_1; B|X_2)-2\delta]}. \tag{4.38}
\end{aligned}$$

Inequality $\textcircled{6}$ follows from the bound

$$\text{Tr} \{ \Pi_{m_1, m_2}^n \} \leq 2^{n[H(B|X_1 X_2)+\delta]}$$

on the rank of a conditionally typical projector.

Bound on (E2) : We employ a different argument to bound the probability of the second error term (E2) based on the following fact

$$\begin{aligned}
\Pi_{m_1, m_2}^n &\leq 2^{n[H(B|X_1 X_2)+\delta]} \Pi_{m_1, m_2}^n \rho_{m_1, m_2}^B \Pi_{m_1, m_2}^n \\
&= 2^{n[H(B|X_1 X_2)+\delta]} \sqrt{\rho_{m_1, m_2}^B} \Pi_{m_1, m_2}^n \sqrt{\rho_{m_1, m_2}^B} \\
&\leq 2^{n[H(B|X_1 X_2)+\delta]} \rho_{m_1, m_2}^B, \tag{4.39}
\end{aligned}$$

which we refer to as the *projector trick* [GLM12]. The first inequality is the standard lower bound on the eigenvalues of ρ_{m_1, m_2}^B expressed as an operator upper bound on the projector Π_{m_1, m_2}^n . The equality follows because the state and its typical projector commute. The last inequality follows from $0 \leq \Pi_{m_1, m_2}^n \leq I$.

We now proceed to bound the expectation of the error term (E2).

$$\begin{aligned}
\mathbb{E}_{X_1^n, X_2^n} \{(\text{E2})\} &= \mathbb{E}_{X_1^n, X_2^n} \left\{ \sum_{m'_2 \neq m_2} \text{Tr} [P_{m_1, m'_2} \tilde{\rho}_{m_1, m_2}] \right\} \\
&= \sum_{m'_2 \neq m_2} \mathbb{E}_{X_1^n} \left\{ \text{Tr} \left[\mathbb{E}_{X_2^n} \{P_{m_1, m'_2}\} \mathbb{E}_{X_2^n} \{\tilde{\rho}_{m_1, m_2}\} \right] \right\} \\
&= \sum_{m'_2 \neq m_2} \mathbb{E}_{X_1^n} \left\{ \text{Tr} \left[\mathbb{E}_{X_2^n} \left\{ \Pi_{\tilde{\rho}, \delta}^n \Pi_{m_1}^n \Pi_{m_1, m'_2}^n \Pi_{m_1}^n \Pi_{\tilde{\rho}, \delta}^n \right\} \mathbb{E}_{X_2^n} \{\tilde{\rho}_{m_1, m_2}\} \right] \right\}
\end{aligned}$$

$$= \sum_{m'_2 \neq m_2} \mathbb{E}_{X_1^n} \left\{ \text{Tr} \left[\Pi_{\bar{\rho}, \delta}^n \mathbb{E}_{X_2^n} \left\{ \Pi_{m_1}^n \Pi_{m_1, m'_2}^n \Pi_{m_1}^n \right\} \Pi_{\bar{\rho}, \delta}^n \mathbb{E}_{X_2^n} \left\{ \tilde{\rho}_{m_1, m_2} \right\} \right] \right\}$$

We focus our attention on the first expectation inside the trace:

$$\begin{aligned} \mathbb{E}_{X_2^n} \left\{ \Pi_{m_1}^n \Pi_{m_1, m'_2}^n \Pi_{m_1}^n \right\} &\stackrel{\textcircled{1}}{\leq} 2^{n[H(B|X_1 X_2) + \delta]} \mathbb{E}_{X_2^n} \left\{ \Pi_{m_1}^n \rho_{m_1, m'_2}^B \Pi_{m_1}^n \right\} \\ &= 2^{n[H(B|X_1 X_2) + \delta]} \Pi_{m_1}^n \mathbb{E}_{X_2^n} \left\{ \rho_{m_1, m'_2}^B \right\} \Pi_{m_1}^n \\ &= 2^{n[H(B|X_1 X_2) + \delta]} \Pi_{m_1}^n \bar{\rho}_{m_1} \Pi_{m_1}^n \\ &\stackrel{\textcircled{2}}{\leq} 2^{n[H(B|X_1 X_2) + \delta]} 2^{-n[H(B|X_1) - \delta]} \Pi_{m_1}^n \\ &= 2^{-n[I(X_2; B|X_1) - 2\delta]} \Pi_{m_1}^n. \end{aligned}$$

In inequality $\textcircled{1}$ we used the projector trick from (4.39). Inequality $\textcircled{2}$ follows from the properties of the conditionally typical projector $\Pi_{m_1}^n$.

Substituting back into the expression for the error bound, we obtain:

$$\begin{aligned} \mathbb{E}_{X_1^n, X_2^n} \left\{ (\text{E2}) \right\} &\leq 2^{-n[I(X_2; B|X_1) - 2\delta]} \sum_{m'_2 \neq m_2} \text{Tr} \left[\Pi_{\bar{\rho}, \delta}^n \Pi_{m_1}^n \Pi_{\bar{\rho}, \delta}^n \tilde{\rho}_{m_1, m_2} \right] \\ &= 2^{-n[I(X_2; B|X_1) - 2\delta]} \sum_{m'_2 \neq m_2} \text{Tr} \left[\Pi_{\bar{\rho}, \delta}^n \Pi_{m_1}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m_2}^n \rho_{m_1, m_2} \Pi_{m_2}^n \right] \\ &= 2^{-n[I(X_2; B|X_1) - 2\delta]} \sum_{m'_2 \neq m_2} \text{Tr} \left[\Pi_{m_2}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m_1}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m_2}^n \rho_{m_1, m_2} \right] \\ &\stackrel{\textcircled{3}}{\leq} 2^{-n[I(X_2; B|X_1) - 2\delta]} \sum_{m'_2 \neq m_2} \text{Tr} \left[\rho_{m_1, m_2} \right] \\ &\leq 2^{-n[I(X_2; B|X_1) - 2\delta]} |\mathcal{M}_2|. \end{aligned} \tag{4.40}$$

Inequality $\textcircled{3}$ follows from an argument analogous to (4.37).

Bound on (E12) : We use a slightly different argument in order to bound the probability of the third error term:

$$\mathbb{E}_{X_1^n, X_2^n} \left\{ (\text{E12}) \right\} = \mathbb{E}_{X_1^n, X_2^n} \left\{ \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \text{Tr} \left[P_{m'_1, m'_2} \tilde{\rho}_{m_1, m_2} \right] \right\}$$

$$\begin{aligned}
 & \stackrel{\textcircled{1}}{=} \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \mathbb{E}_{X_2^n} \left\{ \text{Tr} \left[\mathbb{E}_{X_1^n} \{ P_{m'_1, m'_2} \} \mathbb{E}_{X_1^n} \{ \tilde{\rho}_{m_1, m_2} \} \right] \right\} \\
 & \stackrel{\textcircled{2}}{=} \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \mathbb{E}_{X_2^n} \left\{ \text{Tr} \left[\mathbb{E}_{X_1^n} \{ P_{m'_1, m'_2} \} \Pi_{m_2}^n \bar{\rho}_{m_2} \Pi_{m_2}^n \right] \right\} \\
 & \stackrel{\textcircled{3}}{\leq} \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \mathbb{E}_{X_2^n} \left\{ \text{Tr} \left[\mathbb{E}_{X_1^n} \{ P_{m'_1, m'_2} \} \bar{\rho}_{m_2} \right] \right\} \\
 & = \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \text{Tr} \left[\mathbb{E}_{X_1^n, X_2^n} \{ P_{m'_1, m'_2} \} \mathbb{E}_{X_2^n} \{ \bar{\rho}_{m_2} \} \right] \\
 & = \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \text{Tr} \left[\mathbb{E}_{X_1^n, X_2^n} \{ P_{m'_1, m'_2} \} \bar{\rho}^{\otimes n} \right] \\
 & = \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \mathbb{E}_{X_1^n, X_2^n} \left\{ \text{Tr} \left[\Pi_{\bar{\rho}, \delta}^n \Pi_{m'_1}^n \Pi_{m'_1, m'_2}^n \Pi_{m'_1}^n \Pi_{\bar{\rho}, \delta}^n \bar{\rho}^{\otimes n} \right] \right\} \\
 & \stackrel{\textcircled{4}}{\leq} 2^{-n[H(B) - \delta]} \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \mathbb{E}_{X_1^n, X_2^n} \left\{ \text{Tr} \left[\Pi_{m'_1}^n \Pi_{m'_1, m'_2}^n \Pi_{m'_1}^n \Pi_{\bar{\rho}, \delta}^n \right] \right\} \\
 & \stackrel{\textcircled{5}}{\leq} 2^{-n[H(B) - \delta]} \sum_{m'_1 \neq m_1, m'_2 \neq m_2} \mathbb{E}_{X_1^n, X_2^n} \left\{ \text{Tr} \left[\Pi_{m_1, m'_2}^n \right] \right\} \\
 & \stackrel{\textcircled{6}}{\leq} 2^{-n[H(B) - \delta]} 2^{n[H(B|X_1 X_2) + \delta]} \sum_{m'_1 \neq m_1, m'_2 \neq m_2} 1 \\
 & \leq |\mathcal{M}_1| |\mathcal{M}_2| 2^{-n[I(X_1 X_2; B) - 2\delta]}. \tag{4.41}
 \end{aligned}$$

Equality ① follows from the independence of the codewords. To obtain equality ② we take the X_1^n expectation over the state. Inequality ③ follows from $\Pi_{m_2}^n \bar{\rho}_{m_2} \Pi_{m_2}^n = \sqrt{\bar{\rho}_{m_2}} \Pi_{m_2}^n \sqrt{\bar{\rho}_{m_2}} \leq \bar{\rho}_{m_2}$. Inequality ④ is obtained by using the cyclicity of trace to surround the state $\bar{\rho}^{\otimes n}$ by its typical projectors and then using the property $\Pi_{\bar{\rho}, \delta}^n \bar{\rho}^{\otimes n} \Pi_{\bar{\rho}, \delta}^n \leq 2^{-n[H(B) - \delta]} \Pi_{\bar{\rho}, \delta}^n$ of the average output-typical projector. Inequality ⑤ follows from $\Pi_{m'_1}^n \Pi_{\bar{\rho}, \delta}^n \Pi_{m'_1}^n \leq \Pi_{m'_1}^n \leq I$. Finally, inequality ⑥ follows from the bound on the rank of the conditionally typical projector.

Combining the bounds from equations (4.35), (4.38), (4.40), (4.41) and the smoothing penalty from (4.36), we get the following bound on the expectation of the average

error probability:

$$\begin{aligned} \mathbb{E}_{X_1^n, X_2^n} \left\{ \bar{p}_e \right\} &\leq 2(\epsilon + 6\sqrt{\epsilon}) + 2\sqrt{\epsilon} \\ &+ 4 \left[|\mathcal{M}_1| 2^{-n[I(X_1; B|X_2) - 2\delta]} + |\mathcal{M}_2| 2^{-n[I(X_2; B|X_1) - 2\delta]} \right. \\ &\quad \left. + |\mathcal{M}_1| |\mathcal{M}_2| 2^{-n[I(X_1 X_2; B) - 2\delta]} \right]. \end{aligned}$$

Thus, we can choose the message sets sizes to be $|\mathcal{M}_1| = 2^{n[R_1 - 3\delta]}$, and $|\mathcal{M}_2| = 2^{n[R_2 - 3\delta]}$, the expectation of the average error probability vanishes whenever the rates R_1 and R_2 obey the inequalities:

$$\begin{aligned} R_1 - \delta &< I(X_1; B|X_2), \\ R_2 - \delta &< I(X_2; B|X_1), \\ R_1 + R_2 - 4\delta &< I(X_1 X_2; B). \end{aligned}$$

If the probability of error of a random code vanishes, then there must exist a particular code with vanishing average error probability, and given that $\delta > 0$ is an arbitrarily small number, the bounds in the statement of the theorem follow. \square

We now state a corollary regarding the ‘‘coded time-sharing’’ approach to the MAC problem [HK81, EGK10]. The main idea is to introduce an auxiliary random variable Q distributed according to $p_Q(q)$ and use the probability distribution $p_Q(q)p_{X_1|Q}(x_1|q)p_{X_2|Q}(x_2)$ for the codebook construction. First we generate a random sequence $q^n \sim \prod_i^n p_Q(q_i)$, and then pick the codeword sequences x_1^n and x_2^n according to the distributions $p_{X_1^n|Q^n}(x_1^n|q^n) \equiv \prod_{i=1}^n p_{X_1|Q}(x_{1i}|q_i)$ and $p_{X_2^n|Q^n}(x_2^n|q^n) \equiv \prod_{i=1}^n p_{X_2|Q}(x_{2i}|q_i)$.

Corollary 4.1 (Coded time-sharing for QMAC). *Suppose that the rates R_1 and R_2 satisfy the following inequalities:*

$$R_1 \leq I(X_1; B|X_2Q)_\theta, \quad (4.42)$$

$$R_2 \leq I(X_2; B|X_1Q)_\theta, \quad (4.43)$$

$$R_1 + R_2 \leq I(X_1X_2; B|Q)_\theta, \quad (4.44)$$

where the entropies are with respect to a state $\theta^{QX_1X_2B}$ of the following form:

$$\sum_{x_1, x_2, q} p_Q(q) p_{X_1|Q}(x_1|q) p_{X_2|Q}(x_2|q) |q\rangle\langle q|^Q \otimes |x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_2} \otimes \rho_{x_1, x_2}^B. \quad (4.45)$$

Then there exists a corresponding simultaneous decoding POVM $\{\Lambda_{m_1, m_2}\}$ such that the expectation of the average probability of error is bounded above by ϵ for all $\epsilon > 0$ and sufficiently large n .

The proof of Corollary 4.1 proceeds exactly as the proof of Theorem 4.2, but all the typical projectors are chosen conditionally on Q^n , and we take the expectation over Q^n in the error analysis. The statement of the QMAC capacity rates using coded time-sharing will be important for the results in Chapter 5.

4.3.1 Conjecture for three-sender simultaneous decoding

We now state our conjecture regarding the existence of a quantum simultaneous decoder for a classical-quantum multiple access channel with three senders. We focus on the case of three senders, because this is the form that will be required in Section 5.3 for the achievability proof of the quantum Han-Kobayashi achievable rate region [HK81, Sen12a].

Conjecture 4.1 (Three-sender quantum simultaneous decoder).

Let $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3, \rho_{x_1, x_2, x_3}, \mathcal{H}^B)$ be a classical-quantum multiple access channel with three senders. Let p_{X_1}, p_{X_2} and p_{X_3} be distributions on the inputs. Define the following random code: let $\{X_1^n(m_1)\}_{m_1 \in \{1, \dots, |\mathcal{M}_1|\}}$ be an independent random codebook distributed according to the product distribution $p_{X_1^n}$ and similarly and independently let $\{X_2^n(m_2)\}_{m_2 \in \{1, \dots, |\mathcal{M}_2|\}}$ and $\{X_3^n(m_3)\}_{m_3 \in \{1, \dots, |\mathcal{M}_3|\}}$ be independent random codebooks distributed according to product distributions $p_{X_2^n}$ and $p_{X_3^n}$. Suppose that the rates of

the codebooks obey the following inequalities:

$$\begin{aligned}
 R_1 &\leq I(X_1; B|X_2X_3)_\rho, \\
 R_2 &\leq I(X_2; B|X_1X_3)_\rho, \\
 R_3 &\leq I(X_3; B|X_1X_2)_\rho, \\
 R_1 + R_2 &\leq I(X_1X_2; B|X_3)_\rho, \\
 R_1 + R_3 &\leq I(X_1X_3; B|X_2)_\rho, \\
 R_2 + R_3 &\leq I(X_2X_3; B|X_1)_\rho, \\
 R_1 + R_2 + R_3 &\leq I(X_1X_2X_3; B)_\rho,
 \end{aligned}$$

where the Holevo information quantities are with respect to the following classical-quantum state:

$$\rho^{X_1X_2X_3B} \equiv \sum_{x_1, x_2, x_3} p_{X_1}(x_1) p_{X_2}(x_2) p_{X_3}(x_3) \times \quad (4.46)$$

$$|x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_2} \otimes |x_3\rangle\langle x_3|^{X_3} \otimes \rho_{x_1, x_2, x_3}^B.$$

Then there exists a simultaneous decoding POVM $\{\Lambda_{m_1, m_2, m_3}\}_{m_1, m_2, m_3}$ such that the expectation of the average probability of error is bounded above by ϵ for all $\epsilon > 0$ and sufficiently large n :

$$\mathbb{E} \left\{ \frac{1}{|\mathcal{M}_1||\mathcal{M}_2||\mathcal{M}_3|} \sum_{m_1, m_2, m_3} \text{Tr}[(I - \Lambda_{m_1, m_2, m_3}) \rho_{X_1^n(m_1), X_2^n(m_2), X_3^n(m_3)}] \right\} \leq \epsilon,$$

where the expectation is with respect to X_1^n , X_2^n , and X_3^n .

The importance of this conjecture stems from the fact that it might be broadly useful for “quantizing” other results from classical multiuser information theory [FHS⁺12]. Indeed, many coding theorems in classical network information theory exploit a simultaneous decoding approach (sometimes known as jointly typical decoding) [EGK10]. Also, Dutil and Hayden have recently put forward a related conjecture known as the “multiparty typicality” conjecture [Dut11a], and it is likely that a proof of Conjecture 4.1 could aid in producing a proof of the multiparty typicality conjecture or vice versa. The notion of a multiparty quantum typicality also appears in the problem of universal state merging [BBJ11]. Recent progress towards the proof of this conjecture can be found in [Sen12b].

The conjecture naturally extends to M -senders, but we have described the three-sender case because this is the form that will be required for the Han-Kobayashi strategy discussed in Section 5.3.

4.4 Rate-splitting

Rate-splitting is another approach for achieving the rates of the classical multiple access channel capacity region [GRUW01] which generalizes readily to the quantum setting using the successive decoding approach in [Win01].

Lemma 4.1 (Quantum rate-splitting). *For a given $p = p_{X_1}, p_{X_2}$, any rate pair (R_1, R_2) that lies in between the two corner points of the MAC rate region α_p and β_p can be achieved if Sender 2 splits her message m_2 into two parts m_{2u} and m_{2v} and encodes them with a split codebook and a mixing function $(\{u^n(m_{2u})\}_{m_{2u}}, \{v^n(m_{2v})\}_{m_{2v}}, f)$. The receiver decodes the messages in the order $m_{2u} \rightarrow m_1|m_{2u} \rightarrow m_{2v}|m_1m_{2u}$ using successive decoding. The total rate for Sender 2 is the sum $R_2 = R_{2u} + R_{2v}$.*

The rate-split codebook consists of two random codebooks generated from p_U and p_V and a mixing function such that $f(U, V) = X_2$ [GRUW01]¹. The rate splitting coding strategy for the two sender quantum multiple access channel consists of a successive decoding strategy for the following three channels:

$$(U^n, V^n, X_1^n, X_2^n) \rightarrow \rho_{X_1^n, X_2^n}^{B^n}, \quad (4.47)$$

$$(U^n, V^n, X_1^n, X_2^n) \rightarrow (U^n, \rho_{X_1^n, X_2^n}^{B^n}), \quad (4.48)$$

$$(U^n, V^n, X_1^n, X_2^n) \rightarrow (U^n, X_1^n, \rho_{X_1^n, X_2^n}^{B^n}). \quad (4.49)$$

The codebooks are constructed with the following rates:

$$R_{2u} = I(U; B) - \delta, \quad (4.50)$$

$$R_1 = I(X_1; B|U) - \delta, \quad (4.51)$$

$$R_{2v} = I(V; B|UX_1) - \delta. \quad (4.52)$$

¹ Alternately, the mixing can be performed using a *switch random variable*, S , which is a *shared randomness* resource (denoted $[cc]$) between Sender 2 and the receiver [Rim01].

Observe that the resulting rate pair $(R_1, R_2) = (R_1, R_{2u} + R_{2v})$ is close to the *dominant facet* of the rate region, which is defined as $R_1 + R_2 = I(X_1 X_2 | B)$, since:

$$\begin{aligned} R_1 + R_2 &= R_{2u} + R_1 + R_{2v} \\ &= I(U; B) - \delta + I(X_1; B|U) - \delta + I(V; B|UX_1) - \delta \\ &= I(X_1 X_2 | B) - 3\delta. \end{aligned}$$

By varying the choice of the distributions p_U and p_V and choosing the rates rates of the split-codebooks appropriately, we can achieve all the rates of the dominant facet, and therefore all the rates of the region.

The choice of rate split $R_{2u} \leftrightarrow R_{2v}$ depends on the properties of the channel for which we are coding. This dependence limits the usefulness of the rate-splitting strategy in situations where there are multiple receivers. In general, we cannot choose the rates of the split codebooks such that they will be optimal for two receivers. Receiver 1 whose output is the system $\rho_{x_1, x_2}^{B_1}$ would want the rates of the codebooks to be set at $(R_{2u}, R_{2v}) = (I(U; B_1), I(V; B_1 | UX_1))$, whereas Receiver 2, with outputs $\rho_{x_1, x_2}^{B_2}$ would want to set $(R_{2u}, R_{2v}) = (I(U; B_2), I(V; B_2 | UX_1))$. We will comment on this further in the next chapter.

4.5 Example of a quantum multiple access channel

We now show an example of a simple quantum multiple access channel for which we can compute the capacity region.

Example 4.2. Consider the channel that takes two binary variables x_1 and x_2 as inputs and outputs one of the four “BB84” states. The following table shows the channel outputs for the different possible inputs.

	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$ 0\rangle^B$	$ +\rangle^B$
$x_2 = 1$	$ -\rangle^B$	$ 1\rangle^B$

4.6 Example of a quantum multiple access channel

The classical-quantum state on which we evaluate information quantities is

$$\rho^{X_1 X_2 B} \equiv \sum_{x_1, x_2=0}^1 p_{X_1}(x_1) p_{X_2}(x_2) |x_1\rangle \langle x_1|^{X_1} \otimes |x_2\rangle \langle x_2|^{X_2} \otimes \psi_{x_1, x_2}^B,$$

where ψ_{x_1, x_2}^B is one of $|0\rangle\langle 0|$, $|1\rangle\langle 1|$, $|+\rangle\langle +|$ or $|-\rangle\langle -|$ depending on the choice of the input bits x_1 and x_2 . The conditional entropy $H(B|X_1 X_2)_\rho$ vanishes for this state because the state is pure when conditioned on the classical registers X_1 and X_2 . We choose $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$ to be the uniform distribution. This gives the following state on X_1 , X_2 , and B :

$$\rho^{X_1 X_2 B} = \frac{1}{4} \left[|00\rangle\langle 00| \otimes |0\rangle\langle 0| + |01\rangle\langle 01| \otimes |-\rangle\langle -| + |10\rangle\langle 10| \otimes |+\rangle\langle +| + |11\rangle\langle 11| \otimes |1\rangle\langle 1| \right].$$

From this state we can calculate the reduced density matrix $\rho^{X_2 B} = \text{Tr}_{X_1}[\rho^{X_1 X_2 B}]$ by taking the partial trace over the X_1 system:

$$\rho^{X_2 B} = \frac{1}{2} \left[|0\rangle\langle 0|^{X_2} \otimes \frac{1}{2}(|0\rangle\langle 0| + |+\rangle\langle +|)^B + |1\rangle\langle 1|^{X_2} \otimes \frac{1}{2}(|-\rangle\langle -| + |1\rangle\langle 1|)^B \right],$$

from which we can determine that the conditional entropy $H(B|X_2)_\rho$ takes its maximum value of $H_2(\cos^2(\pi/8))$ when $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$ are uniform.

Taking the partial trace over X_2 we obtain the state

$$\rho^{X_1 B} = \frac{1}{2} \left[|0\rangle\langle 0|^{X_1} \otimes \frac{1}{2}(|0\rangle\langle 0| + |-\rangle\langle -|)^B + |1\rangle\langle 1|^{X_1} \otimes \frac{1}{2}(|+\rangle\langle +| + |1\rangle\langle 1|)^B \right],$$

from which we can observe that $H(B|X_1) = H_2(\cos^2(\pi/8))$.

Thus, the capacity region for this channel is:

$$\begin{aligned} R_1 &\leq H_2(\cos^2(\pi/8)) \approx 0.6009, \\ R_2 &\leq H_2(\cos^2(\pi/8)) \approx 0.6009, \\ R_1 + R_2 &\leq 1. \end{aligned}$$

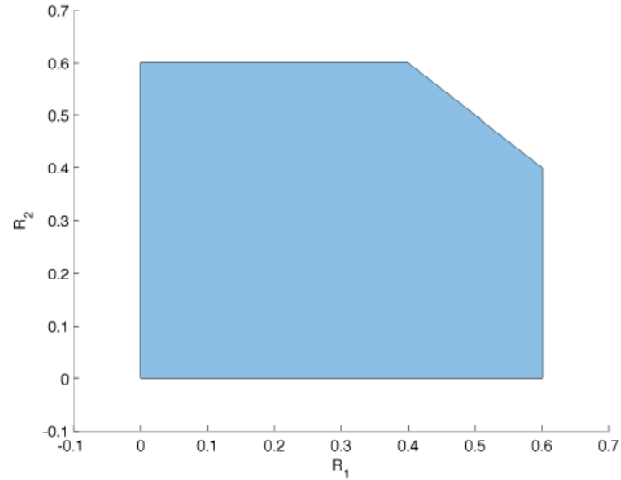


Figure 4.7: The capacity region for the multiple access channel in Example 4.2.

4.6 Discussion

This concludes our exposition on the quantum multiple access channel. The techniques used in the proof of Theorem 4.2 are the tools that will be used throughout the remainder of this thesis. We review them here for the convenience of the reader and in order to highlight them in isolation from the technicalities in the proof of Theorem 4.2.

The first idea is the POVM construction with layered typical projectors:

$$\Pi_{\tilde{\rho}, \delta}^n \Pi_{m_1}^n \Pi_{m_1, m_2}^n \Pi_{m_1}^n \Pi_{\tilde{\rho}, \delta}^n. \quad (4.53)$$

We call this a *projector sandwich*. Observe that the more specific projectors are on the inside. Each of the projectors seems to be necessary in some part of the proof, and this layering of the projectors ensures that the averaging can be performed.

The second idea that makes the quantum simultaneous decoder possible is the *state smoothing trick*, which is to perform the error analysis with the unnormalized state:

$$\tilde{\rho}_{m_1, m_2} \equiv \Pi_{m_2}^n \rho_{m_1, m_2} \Pi_{m_2}^n, \quad (4.54)$$

which is close to the original state, but has the $X_2^n(m_2)$ non-typical parts of it trimmed off.

4.6 Discussion

The third idea is to use equation (B.29) in order to obtain the bound

$$\Pi_{m_1, m_2}^n \leq 2^{n[H(B|X_1 X_2) + \delta]} \rho_{m_1, m_2}^B. \quad (4.55)$$

We will call this the *projector trick* [GLM12, Sen12a, FHS⁺12].

Because of the *ad hoc* nature of the proof of the two-sender simultaneous decoder, the ideas from the two-sender case cannot be applied to show that simultaneous decoding of three or more messages is possible. The techniques used in the proof are sufficiently general for the analysis of many problems of quantum network information theory: quantum interference channels (Chapter 5), quantum broadcast channels (Chapter 6), and quantum relay channels (Chapter 7).

Chapter 5

Interference channels

In an ideal world, when a sender and a receiver wish to communicate, the only obstacle they face is the presence of the background noise. Real-world communication scenarios, however, often involve multiple senders and multiple receivers sending information at the same time and in a shared communication medium. The receivers have to contend not only with the background noise but also with the interference caused by the other transmissions. The interference channel (IC) is a model for the effects of this *crosstalk*, which occurs whenever a communication channel is shared.

5.1 Introduction

Interference is a big problem for all modern multiuser communication systems. In order to avoid interference, techniques such as frequency division multiple access (FDMA) and time division multiple access (TDMA) can be used to ensure that the senders never transmit at the same time and in the same frequency band. Another approach is to use code division multiple access (CDMA) and allow users to transmit at the same time, but their signal power is randomly spread over large sections of the spectrum so as to make it look like white noise.

Rather than treating the interference as noise, a receiver could instead decode the interfering signal and then “subtract” it from the received signal in order to reduce (or even remove) the interference. We call this approach *interference cancellation*, and such strategies are the main theme of this chapter.

Note that the interference channel problem differs from the multiple access channel problem since in this case the multiple access communication is not *intended*. A receiver in the interference channel problem is not *required* to decode the interfering messages, but he will be able to achieve better communication rates if he does so. All the decoding strategies discussed in this chapter use some form of interference cancellation as part of the decoding strategy.

5.1.1 Applications

The interference channel is an excellent model for many practical communication scenarios where medium contention is an issue.

Example 5.1 (Next-generation WiFi routers). Consider two neighbours who want to connect to their respective WiFi routers. Suppose that the communication happens in the same frequency band (radio channel). Suppose further that the neighbours' laptops are located such that they are close to their neighbour's WiFi router and far from their own. In such a situation, the *interference* signal will be stronger than their own signal. Because the interference signal is “masking” the intended signal, it would be possible for the neighbours to decode it, and then *cancel* its effects. Thus, we see that it can be to a neighbour's advantage to decode wireless packets which are not intended for him. Decoding messages not intended for us can increase the communication rate from the intended sender. Note that to implement such a strategy in practice would require a re-engineering of the physical layer of transmission protocols.

Interference also plays an important role in digital subscriber line (DSL) internet connections. The twisted pair copper wires of the telephone system were not originally designed to carry high frequency and high bandwidth signals, and so there is a significant amount of crosstalk on the wires *en route* to the phone company premises. Cross-channel interference is in fact the current limiting factor which imposes speed limits on the order of 30Mb/s. The next generation VDSL technology includes the **G.vector** standard, which is essentially an interference cancellation scheme for a vector additive white Gaussian channel [GC02, OSC⁺10]. The use of the new **G.vector** VDSL standard for interference mitigation will allow speeds of up to 100Mb/s to the home.

Interestingly, Shannon's first paper on multiuser communication channels was on “Two-way communication channels”, which can model the simultaneous transmission

of information in both directions over a phone line [Sha61]. Shannon anticipated the importance of NEXT (near-end crosstalk) and FEXT (far-end crosstalk) to communication systems fifty years in advance. Clearly, he was a man ahead of his times!

5.1.2 Review of classical results

The seminal papers by Carleial [Car78] and Sato [Sat77] defined the interference channel problem in its present form and established many of the fundamental results. Finding the capacity region of the general discrete memoryless interference channel (DMIC) is still an open problem, but there are certain special cases where the capacity can be calculated. For channels with “strong” [Sat81] and “very strong” [Car75] interference, the full capacity region can be calculated. The capacity-achieving decoding strategies for both of the above special cases require the receivers to completely decode the interfering messages.

For an arbitrary interference channel, it may only be possible to *partially* decode the interfering signal. The Han-Kobayashi rate region \mathcal{R}_{HK} , which is achieved by using partial interference cancellation, is the best known achievable rate region for the general discrete memoryless interference channel [HK81]. Recently, Chong, Motani and Garg used a different encoding scheme to obtain an achievable rate region, \mathcal{R}_{CMG} , which contains the Han-Kobayashi rate region [CMG06]. Soon afterwards Kramer proposed a compact description of the Han-Kobayashi rate region, $\mathcal{R}_{\text{HK}}^c$, which involved fewer constraints [Kra06]. Han and Kobayashi published a comment regarding the Fourier-Motzkin elimination procedure used to derive the bounds [HK07], but the question remained whether the above rate regions are all equivalent or whether one is strictly larger than the others. The matter was finally settled by Chong, Motani, Garg and Hesham El Gamal, who showed that all three rate regions are in fact equivalent:

$$\mathcal{R}_{\text{HK}} \equiv \mathcal{R}_{\text{CMG}} \equiv \mathcal{R}_{\text{HK}}^c, \quad (5.1)$$

when the union is taken over all possible input distributions [CMGEG08].

There has been comparatively less work on proving outer bounds on the capacity region for general discrete memoryless interference channels [Sat77, Car83].

5.1.3 Quantum interference channels

In this chapter, we apply and extend insights from classical information theory to the study of the quantum interference channel (QIC):

$$(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{N}^{X_1 X_2 \rightarrow B_1 B_2}(x_1, x_2) \equiv \rho_{x_1, x_2}^{B_1 B_2}, \mathcal{H}^{B_1} \otimes \mathcal{H}^{B_2}), \quad (5.2)$$

which is a model for a general communication network with two classical inputs and a quantum state $\rho_{x_1, x_2}^{B_1 B_2}$ as output. The classical-quantum interference channel can model physical systems such as fibre-optic cables and free space optical communication channels [GSW11].

We fully specify a *cc-qq* interference channel by the set of output states it produces $\{\rho_{x_1, x_2}^{B_1 B_2}\}_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2}$ for each possible combination of inputs. Since Receiver 1 does not have access to the B_2 part of the state $\rho_{x_1, x_2}^{B_1 B_2}$, we model his state as $\rho_{x_1, x_2}^{B_1} = \text{Tr}_{B_2}[\rho_{x_1, x_2}^{B_1 B_2}]$, where Tr_{B_2} denotes the partial trace over Receiver 2's system. Similarly, the output state for Receiver 2 is given by $\rho_{x_1, x_2}^{B_2} = \text{Tr}_{B_1}[\rho_{x_1, x_2}^{B_1 B_2}]$.

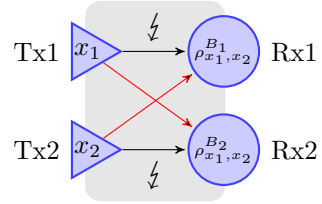


Figure 5.1: The quantum interference channel $\rho_{x_1, x_2}^{B_1 B_2}$.

A classical interference channel with transition probability function $p(y_1, y_2|x_1, x_2)$ is a special case of the *cc-qq* channel where the output states are of the form $\rho_{x_1, x_2}^{B_1 B_2} = \sum_{y_1, y_2} p(y_1, y_2|x_1, x_2) |y_1\rangle\langle y_1|^{B_1} \otimes |y_2\rangle\langle y_2|^{B_2}$ where $\{|y_1\rangle\}$ and $\{|y_2\rangle\}$ are orthonormal bases of \mathcal{H}^{B_1} and \mathcal{H}^{B_2} .

5.1.4 Information processing task

The task of communication over an interference channel can be described as follows. Using n independent uses of the channel, the objective is for Sender 1 to communicate with Receiver 1 at a rate R_1 and for Sender 2 to communicate with Receiver 2 at a rate R_2 .

If there exists an (n, R_1, R_2, ϵ) -code for the classical-quantum interference channel, then the following conversion is possible:

$$n \cdot \mathcal{N}^{X_1 X_2 \rightarrow B_1 B_2} \xrightarrow{(1-\epsilon)} nR_1 \cdot [c^1 \rightarrow c^1] + nR_2 \cdot [c^2 \rightarrow c^2].$$

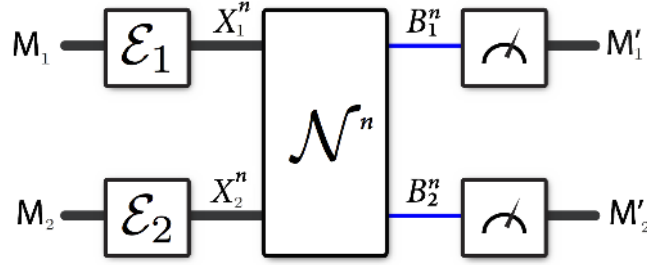


Figure 5.2: Diagram showing the parts of a classical-quantum interference channel *code* for n copies of the channel. Sender 1 selects a message m_1 to transmit (modeled by a random variable M_1), and Sender 2 selects a message m_2 to transmit (modeled by M_2). Each sender encodes their message as an n -symbol codeword suitable for transmission over the channel. The receivers each perform a quantum measurement in order to decode the messages that their partner sender transmitted.

Note that we are only interested in the communication rates from the sender to the intended receiver, and we ignore the communication capacity of the crosslinks: $[c^1 \rightarrow c^2]$ and $[c^2 \rightarrow c^1]$.

More specifically, Sender 1 chooses a message m_1 from a message set $\mathcal{M}_1 \equiv \{1, 2, \dots, |\mathcal{M}_1|\}$ where $|\mathcal{M}_1| = 2^{nR_1}$, and Sender 2 similarly chooses a message m_2 from a message set $\mathcal{M}_2 \equiv \{1, 2, \dots, |\mathcal{M}_2|\}$ where $|\mathcal{M}_2| = 2^{nR_2}$. Senders 1 and 2 encode their messages as codewords $x_1^n(m_1) \in \mathcal{X}_1^n$ and $x_2^n(m_2) \in \mathcal{X}_2^n$ respectively, which are then input to the channel. The output of the channel is an n -fold tensor product state of the form:

$$\mathcal{N}^{\otimes n}(x_1^n(m_1), x_2^n(m_2)) \equiv \rho_{x_1^n(m_1), x_2^n(m_2)}^{B_1^n B_2^n} \in \mathcal{D}(\mathcal{H}^{B_1^n B_2^n}). \quad (5.3)$$

To decode the message m_1 intended for him, Receiver 1 performs a positive operator-valued measure (POVM) $\{\Lambda_{m_1}\}_{m_1 \in \{1, \dots, |\mathcal{M}_1|\}}$ on the system B_1^n , the output of which we denote M'_1 . For all m_1 , Λ_{m_1} is a positive semidefinite operator and $\sum_{m_1} \Lambda_{m_1} = I$. Receiver 2 similarly performs a POVM $\{\Gamma_{m_2}\}_{m_2 \in \{1, \dots, |\mathcal{M}_2|\}}$ on the system B_2^n , and the random variable associated with this outcome is denoted M'_2 .

An error occurs whenever Receiver 1's measurement outcome is different from the message sent by Sender 1 ($M'_1 \neq m_1$) or Receiver 2's measurement outcome is different from the message sent by Sender 2 ($M'_2 \neq m_2$). The overall probability of error for

message pair (m_1, m_2) is

$$\begin{aligned} p_e(m_1, m_2) &\equiv \Pr \{(M'_1, M'_2) \neq (m_1, m_2)\} \\ &= \text{Tr} \left\{ (I - \Lambda_{m_1} \otimes \Gamma_{m_2}) \rho_{x_1^n(m_1)x_2^n(m_2)}^{B_1^n B_2^n} \right\}, \end{aligned}$$

where the measurement operator $(I - \Lambda_{m_1} \otimes \Gamma_{m_2})$ represents the complement of the correct decoding outcome.

Definition 5.1. An (n, R_1, R_2, ϵ) code for the interference channel consists of two codebooks $\{x_1^n(m_1)\}_{m_1 \in \mathcal{M}_1}$ and $\{x_2^n(m_2)\}_{m_2 \in \mathcal{M}_2}$, and two decoding POVMs $\{\Lambda_{m_1}\}_{m_1 \in \mathcal{M}_1}$ and $\{\Gamma_{m_2}\}_{m_2 \in \mathcal{M}_2}$, such that the average probability of error \bar{p}_e is bounded from above by ϵ :

$$\bar{p}_e \equiv \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} \sum_{m_1, m_2} p_e(m_1, m_2) \leq \epsilon. \quad (5.4)$$

A rate pair (R_1, R_2) is *achievable* if there exists an $(n, R_1 - \delta, R_2 - \delta, \epsilon)$ quantum interference channel code for all $\epsilon, \delta > 0$ and sufficiently large n . The channel's *capacity region* is the closure of the set of all achievable rates.

Interference channel as two disinterested MAC sub-channels

The quantum interference channel described by $(\mathcal{X}_1 \times \mathcal{X}_2, \rho_{x_1, x_2}^{B_1 B_2}, \mathcal{H}^{B_1} \otimes \mathcal{H}^{B_2})$ induces two quantum multiple access (QMAC) sub-channels. More specifically QMAC₁ is the channel to Receiver 1 given by $(\mathcal{X}_1 \times \mathcal{X}_2, \rho_{x_1, x_2}^{B_1} = \text{Tr}_{B_2} \{\rho_{x_1, x_2}^{B_1 B_2}\}, \mathcal{H}^{B_1})$, and QMAC₂ is the channel to Receiver 2 defined by $(\mathcal{X}_1 \times \mathcal{X}_2, \rho_{x_1, x_2}^{B_2}, \mathcal{H}^{B_2})$. Thus, one possible coding strategy for the interference channel is to build a codebook for each multiple access channel that is decodable for *both* receivers. For this reason, the coding theorems which we developed for quantum multiple access channels in Chapter 4 will play an important role in this chapter.

Note however that the IC *problem specification* does not require that Receiver 1 be able to decode m_2 correctly nor does it specify that Receiver 2 needs to be able to decode the message sent by Sender 1 correctly, though most interesting coding strategies involve at least partial decoding of the crosstalk messages. If we take the logical **and** of the two MAC subtasks, i.e., we require both receivers to be able to decode the messages from both senders, then this communication task is known as the

compound multiple access channel problem [Ahl74b].

5.1.5 Chapter overview

In this chapter, we use the theorems from Chapter 4 for quantum multiple access channels to prove coding theorems for quantum interference channels.

In Section 5.2, we prove capacity theorems for two special cases of the interference channel. In Theorem 5.1 we calculate the capacity region of the quantum interference channel with “very strong” interference (see Definition 5.2) using the successive decoding strategy from Theorem 4.1. In Theorem 5.2, we prove the capacity of the channels with “strong” interference (see Definition 5.3) using the simultaneous decoding strategy derived in Theorem 4.2.

In Section 5.3 we discuss the quantum Han-Kobayashi coding strategy, where the messages of the senders are split into two parts so that the receivers can perform partial interference cancelation [HK81]. The quantum Han-Kobayashi coding strategy (Theorem 5.3) requires the use of quantum simultaneous decoding for multiple access channels with three senders which we described in Conjecture 4.1.

The main contribution of this chapter is to show that the rates of the Han-Kobayashi rate region can be achieved without the need for Conjecture 4.1. We will show this in Section 5.4, where we present an achievability proof for the quantum Chong-Motani-Garg rate region which only uses the two-message simultaneous decoding technique from Theorem 4.2. Recall that the Chong-Motani-Garg region is equivalent to the Han-Kobayashi region.

Note that the achievability of the quantum Chong-Motani-Garg rate region was first proved by Sen in [Sen12a] using a different error analysis technique based on an *intersection projector* and a careful analysis of the geometric properties of the CMG rate region. The alternate proof given in Section 5.4 uses the simultaneous decoding techniques developed in Section 4.3 and an interesting geometric argument by Eren Şaşıoğlu [Sas08].

The arguments in Section 5.4 show that we can reduce the decoding requirements from three-message simultaneous decoding to two-message simultaneous decoding and still achieve all the rates in the Han-Kobayashi rate region. Perhaps, it might be

possible to remove the need for a simultaneous decoder altogether. Can the Han-Kobayashi rate region be achieved using only successive decoding? In Section 5.6, we discuss the difference between interference channel codes (both classical and quantum) based on successive decoding and those based on simultaneous decoding. In particular, we show that rate-splitting strategies based on successive decoding are not a good choice for interference channel codes, contrary to what has been claimed elsewhere [Sas08, YP11].

Finally, we obtain Theorem 5.8, which is a quantum analogue of Sato’s outer bound for the interference channel.

5.2 Capacity results for special cases

In this section, we consider decoding strategies where the receivers decode the messages from both senders. We show that this decoding strategy is optimal for the special cases of the interference channel with “very strong” and “strong” interference.

5.2.1 Very strong interference case

If we use a successive decoding strategy at both receivers, and calculate the best possible rates that are compatible with both receivers’ ability to decode, we obtain an achievable rate region. Consider the decoding strategy where Receiver 1 decodes in the decode order $m_2 \rightarrow m_1|m_2$ and Receiver 2 decodes in the order $m_1 \rightarrow m_2|m_1$. In this case, we know that the messages are decodable for Receiver 1 provided $R_1 \leq I(X_1; B_1|X_2)$ and $R_2 \leq I(X_2; B_1)$. Receiver 2 will be able to decode provided $R_1 \leq I(X_1; B_2)$ and $R_2 \leq I(X_2; B_2|X_1)$. Thus, the rate pair $R_1 \leq \min\{I(X_1; B_1|X_2), I(X_1; B_2)\}$, $R_2 \leq \min\{I(X_2; B_1), I(X_2; B_2|X_1)\}$ is achievable for the interference channel.

On the other hand, the rate $R_1 \leq I(X_1; B_1|X_2)$ is the optimal rate Receiver 1 could possibly achieve, since this rate corresponds the message m_1 being decoded second [Win01]. Similarly the rate $R_2 \leq I(X_2; B_2|X_1)$ is an upper bound on the rates achievable between Sender 2 and Receiver 2.

We now define a special class of interference channels, where the achievable rate region obtained using the above successive decoding strategy matches the outer bound.

Definition 5.2 (Very strong interference). An interference channel with *very strong* interference [Car75], is such that for all input distributions p_{X_1} and p_{X_2} ,

$$I(X_1; B_1|X_2) \leq I(X_1; B_2), \tag{5.5}$$

$$I(X_2; B_2|X_1) \leq I(X_2; B_1). \tag{5.6}$$

The information inequalities in (5.5)-(5.6) imply that the interference is so strong, that it is possible for each receiver to decode the other sender’s message before decoding the message intended for him. These conditions are a generalization of Carleial’s conditions for a classical Gaussian interference channel [Car75, EGK10].

Thus, we can calculate the exact capacity region for the special case of the classical-quantum interference channel with very strong interference.

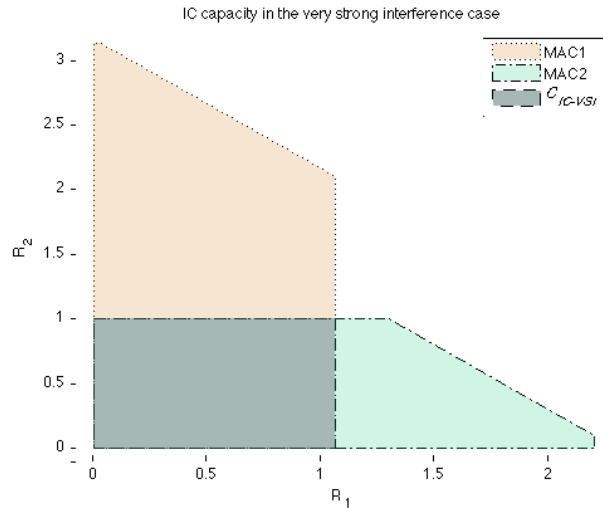


Figure 5.3: The capacity region for a *cc-qq* quantum interference channel which satisfies the “very strong” interference conditions (5.5) and (5.6). The figure also shows the capacity regions for the multiple access channel problems associated with each receiver: $QMAC_1$ and $QMAC_2$. The capacity region for the IC corresponds to their intersection.

Theorem 5.1 (Channels with very strong interference). *The channel's capacity region is given by:*

$$\bigcup_{p_Q, p_{X_1|Q}, p_{X_2|Q}} \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \left| \begin{array}{l} R_1 \leq I(X_1; B_1 | X_2 Q)_\theta, \\ R_2 \leq I(X_2; B_2 | X_1 Q)_\theta \end{array} \right. \right\}, \quad (5.7)$$

where the mutual information quantities are calculated with respect to a state $\theta^{QX_1X_2B}$ of the form:

$$\sum_{x_1, x_2, q} p_Q(q) p_{X_1|Q}(x_1|q) p_{X_2|Q}(x_2|q) |q\rangle\langle q|^Q \otimes |x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_2} \otimes \rho_{x_1, x_2}^B. \quad (5.8)$$

An intuitive interpretation of this result is the seemingly counterintuitive statement that, for channels with very strong interference, the capacity is the same as if there were no interference [Car75].

Proof. We require the receivers to decode the messages for both senders. The average probability of error for the interference channel code is given by:

$$\begin{aligned} \bar{p}_e &\equiv \frac{1}{|\mathcal{M}_1| |\mathcal{M}_2|} \sum_{m_1, m_2} p_e(m_1, m_2) \\ &\stackrel{\textcircled{1}}{=} p_e(m_1, m_2) \\ &= \text{Tr} \left[(I - \Lambda_{m_1, m_2}^{B_1^n} \otimes \Gamma_{m_1, m_2}^{B_2^n}) \rho_{x_2^n(m_1) x_1^n(m_2)}^{B_1^n B_2^n} \right], \end{aligned} \quad (5.9)$$

where equality $\textcircled{1}$ comes from the symmetry of the codebook construction: it is sufficient to perform the error analysis for a fixed message pair (m_1, m_2) .

Next, we use the following lemma, which is a kind of operator union bound [ADHW09].

Lemma 5.1. *For any operators $0 \leq P^A, Q^B \leq I$, we have:*

$$(I^{AB} - P^A \otimes Q^B) \leq (I^A - P^A) \otimes I^B + I^A \otimes (I^B - Q^B). \quad (5.10)$$

Proof of Lemma 5.1. Starting from $P^A \leq I$ and $Q^B \leq I$, we obtain $0 \leq (I - P^A)$ and $0 \leq (I - Q^B)$ which can be combined to obtain:

$$\begin{aligned} 0 &\leq (I - P^A) \otimes (I - Q^B) \\ &= I^{AB} - P^A \otimes I^B - I^A \otimes Q^B + P^A \otimes Q^B. \end{aligned}$$

The inequality (5.10) follows by moving the term $P^A \otimes Q^B$ to the left hand side and adding a term I^{AB} to both sides. \square

When applied to the current problem, the inequality (5.10) gives:

$$(I^{B_1^n B_2^n} - \Lambda_{m_1, m_2}^{B_1^n} \otimes \Gamma_{m_1, m_2}^{B_2^n}) \leq (I^{B_1^n} - \Lambda_{m_1, m_2}^{B_1^n}) \otimes I^{B_2^n} + I^{B_1^n} \otimes (I^{B_2^n} - \Gamma_{m_1, m_2}^{B_2^n}),$$

which in turn allows us to split expression (5.9) into two terms:

$$\begin{aligned} \bar{p}_e &= \text{Tr}_{B_1^n B_2^n} \left[(I - \Lambda_{m_1, m_2}^{B_1^n} \otimes \Gamma_{m_1, m_2}^{B_2^n}) \rho_{x_2^n(m_1)x_2^n(m_2)}^{B_1^n B_2^n} \right], \\ &\leq \text{Tr}_{B_1^n B_2^n} \left[(I - \Lambda_{m_1, m_2}^{B_1^n}) \rho_{x_2^n(m_1)x_2^n(m_2)}^{B_1^n B_2^n} \right] + \text{Tr}_{B_1^n B_2^n} \left[(I - \Gamma_{m_1, m_2}^{B_2^n}) \rho_{x_2^n(m_1)x_2^n(m_2)}^{B_1^n B_2^n} \right] \\ &= \text{Tr}_{B_1^n} \left[(I - \Lambda_{m_1, m_2}^{B_1^n}) \rho_{x_2^n(m_1)x_2^n(m_2)}^{B_1^n} \right] + \text{Tr}_{B_2^n} \left[(I - \Gamma_{m_1, m_2}^{B_2^n}) \rho_{x_2^n(m_1)x_2^n(m_2)}^{B_2^n} \right]. \end{aligned}$$

Each of the above error terms is associated with the probability of error for one of the receivers. The decoding problem for each receiver corresponds to a multiple access channel (MAC) problem. We can use the successive decoding techniques from Theorem 4.1 to show that the decoding at the rates $R_1 \leq I(X_1; B_1|X_2)$, $R_2 \leq I(X_2; B_2|X_1)$ will succeed.

Receiver 1 will decode in the order $m_2 \rightarrow m_1|m_2$. During the first decoding step Receiver 1 decodes the interfering message m_2 and we know that this is possible because the rate $R_2 \leq I(X_2; B_1)$, which is guaranteed by (5.5). In the second step, Receiver 1 now decodes the message from Sender 1 given full knowledge of the transmission of Sender 2, which is possible any rate $R_1 \leq I(X_1; B_1|X_2)$. Receiver 2 decodes in the order $m_1 \rightarrow m_2|m_1$ in order to use full interference cancellation and achieve the rate $R_2 \leq I(X_2; B_2|X_1)$.

The outer bound follows from the converse part of Theorem 4.1, since the individual rates are optimal in the two MAC sub-channels [Car75]. \square

Example 5.2. We now consider an example of a *cc-qq* quantum interference channel with two classical inputs and two quantum outputs and calculate its capacity region using Theorem 5.1 [FHS⁺12]. The “ θ -SWAP” channel $\mathcal{N} : \{0, 1\}^2 \rightarrow \mathbb{C}^4$ is described

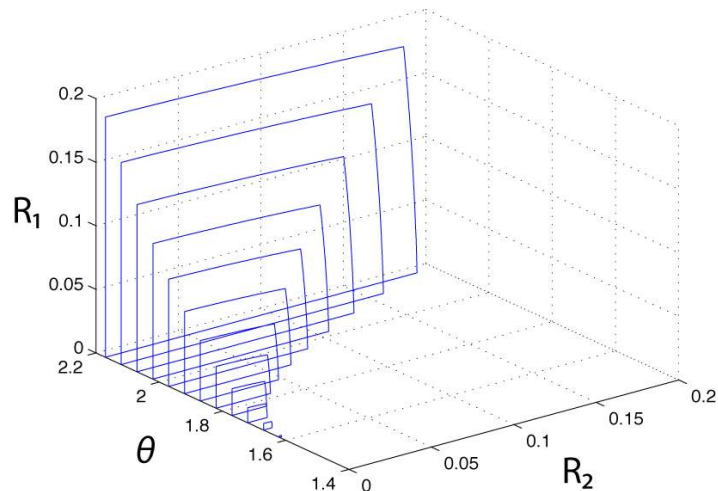


Figure 5.4: The capacity region of the “ θ -SWAP” interference channel for various values of θ such that the channel exhibits “very strong” interference. The capacity region is largest when θ gets closer to 2.18, and it vanishes when $\theta = \pi/2$ because the channel becomes a full SWAP (at this point, Receiver i gets no information from Sender i , where $i \in \{1, 2\}$).

by:

$$00 \rightarrow |00\rangle^{B_1 B_2}, \quad (5.11)$$

$$01 \rightarrow \cos(\theta) |01\rangle^{B_1 B_2} + \sin(\theta) |10\rangle^{B_1 B_2}, \quad (5.12)$$

$$10 \rightarrow -\sin(\theta) |01\rangle^{B_1 B_2} + \cos(\theta) |10\rangle^{B_1 B_2}, \quad (5.13)$$

$$11 \rightarrow |11\rangle^{B_1 B_2}. \quad (5.14)$$

We would like to determine an interval for the parameter θ for which the channel exhibits “very strong” interference. In order to do so, we need to consider classical-quantum states of the following form:

$$\rho^{X_1 X_2 B_1 B_2} \equiv \sum_{x_1, x_2=0}^1 p_{X_1}(x_1) p_{X_2}(x_2) |x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_2} \otimes \psi_{x_1, x_2}^{B_1 B_2}, \quad (5.15)$$

where $\psi_{x_1, x_2}^{B_1 B_2}$ is one of the pure output states in (5.11)-(5.14). We should then check whether the conditions in (5.5)-(5.6) hold for all distributions $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$. We can equivalently express these conditions in terms of von Neumann entropies as

follows:

$$\begin{aligned} H(B_1|X_2)_\rho - H(B_1|X_1X_2)_\rho &\leq H(B_2)_\rho - H(B_2|X_1)_\rho, \\ H(B_2|X_1)_\rho - H(B_2|X_1X_2)_\rho &\leq H(B_1)_\rho - H(B_1|X_2)_\rho, \end{aligned}$$

and thus, it suffices to calculate six entropies for states of the form in (5.15). After some straightforward calculations, we find that:

$$\begin{aligned} H(B_1|X_1X_2)_\rho &= H(B_2|X_1X_2)_\rho = (p_{X_1}(0)p_{X_2}(1) + p_{X_1}(1)p_{X_2}(0)) H_2(\cos^2(\theta)), \\ H(B_1)_\rho &= H_2(p_{X_1}(0) + (p_{X_1}(1)p_{X_2}(0) - p_{X_1}(0)p_{X_2}(1)) \sin^2(\theta)), \\ H(B_2)_\rho &= H_2(p_{X_2}(0) + (p_{X_1}(0)p_{X_2}(1) - p_{X_1}(1)p_{X_2}(0)) \sin^2(\theta)), \\ H(B_2|X_1)_\rho &= p_{X_1}(0) H_2(p_{X_2}(1) \cos^2(\theta)) + p_{X_1}(1) H_2(p_{X_2}(0) \cos^2(\theta)), \\ H(B_1|X_2)_\rho &= p_{X_2}(0) H_2(p_{X_1}(1) \cos^2(\theta)) + p_{X_2}(1) H_2(p_{X_1}(0) \cos^2(\theta)), \end{aligned}$$

where $H_2(p)$ is the binary entropy function. We numerically checked for particular values of θ whether the conditions (5.5)-(5.6) hold for all distributions $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$, and we found that they hold when $\theta \in [0.96, 2.18] \cup [4.10, 5.32]$ (the latter interval in the union is approximately a shift of the first interval by π). The interval $[0.96, 2.18]$ contains $\theta = \pi/2$, the value of θ for which the capacity should vanish because the transformation is equivalent to a full SWAP (the channel at this point has “too strong” interference). We compute the capacity region given in Theorem 5.1 for several values of θ in the interval $\theta \in [\pi/2, 2.18]$ (it is redundant to evaluate for other intervals because the capacity region is symmetric about $\pi/2$ and it is also equivalent for the two π -shifted intervals $[0.96, 2.18]$ and $[4.1, 5.32]$). Figure 5.4 plots these capacity regions for several values of θ in the interval $[\pi/2, 2.18]$.

5.2.2 Strong interference case

The simultaneous decoder from Theorem 4.2 allows us to calculate the capacity region for the following special case of the quantum interference channel.

Definition 5.3 (Strong interference). A quantum interference channel with *strong*

5.2 Capacity results for special cases

interference [Sat81, CEG87] is one for which the following conditions hold:

$$I(X_1; B_1|X_2) \leq I(X_1; B_2|X_2), \quad (5.16)$$

$$I(X_2; B_2|X_1) \leq I(X_2; B_1|X_1), \quad (5.17)$$

for all input distributions p_{X_1} and p_{X_2} .

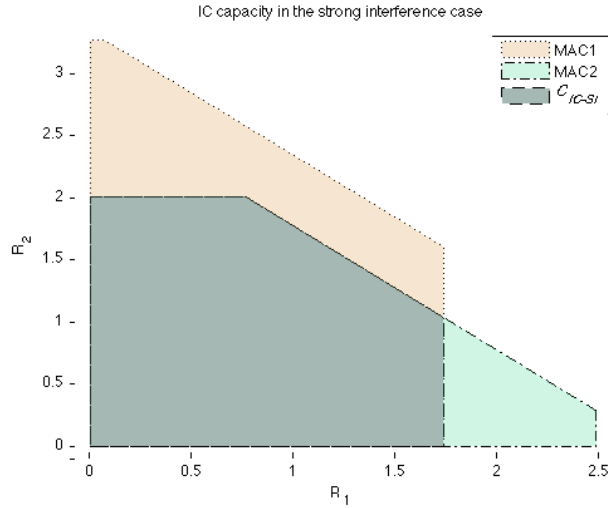


Figure 5.5: The capacity region for a cc - qq quantum interference channel which satisfies the “strong” interference conditions (5.16) and (5.17). The figure also shows the capacity regions for the multiple access channel problems associated with each receiver: QMAC₁ and QMAC₂. The capacity region corresponds to the intersection.

Theorem 5.2 (Channels with strong interference). *The channel’s capacity region is:*

$$\bigcup_{\substack{p_Q, p_{X_1|Q}, \\ p_{X_2|Q}}} \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \left| \begin{array}{l} R_1 \leq I(X_1; B_1|X_2Q)_\theta, \\ R_2 \leq I(X_2; B_2|X_1Q)_\theta, \\ R_1 + R_2 \leq \min \left\{ \begin{array}{l} I(X_1X_2; B_1|Q)_\theta \\ I(X_1X_2; B_2|Q)_\theta \end{array} \right\} \end{array} \right. \right\}, \quad (5.18)$$

where the mutual information quantities are calculated with respect to a state $\theta^{QX_1X_2B}$ of the form:

$$\sum_{x_1, x_2, q} p_Q(q) p_{X_1|Q}(x_1|q) p_{X_2|Q}(x_2|q) |q\rangle\langle q|^Q \otimes |x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_2} \otimes \rho_{x_1, x_2}^B. \quad (5.19)$$

The capacity region is the intersection of the MAC rate regions for the two receivers which corresponds to the condition that we choose the rates such that each receiver

can decode both m_1 and m_2 . See Figure 5.5.

Proof. The first part of the proof is analogous to the proof of Theorem 5.1 for the interference channel with very strong interference. We use Lemma 5.1 to split the error analysis for the interference channel decoding task into two multiple access channel decoding tasks, one for each receiver.

The key difference with Theorem 5.1 is that for the strong interference case, we require the decoders to use the simultaneous decoding approach from Theorem 4.2 and *coded time-sharing* codebooks as described in Corollary 4.1. The rate pairs described by the inequalities in (5.18) are decodable by both receivers. Therefore, these rates are achievable for the interference channel problem.

The proof of the outer bound for Theorem 5.2 follows from the outer bound in Theorem 4.1 and an argument similar to the one used in the classical case [CEG87] (see also [EGK10, page 6–13]). \square

5.3 The quantum Han-Kobayashi rate region

For general interference channels, the Han-Kobayashi coding strategy gives the best known achievable rate region [HK81] and involves *partial* decoding of the interfering signal. Instead of using a standard codebook to encode her message m_1 , Sender 1 splits her message into two parts: a *personal* message m_{1p} and a common message m_{1c} . Assuming that Receiver 1 is able to decode both of these messages, the net rate from Sender 1 to Receiver 1 will be the sum of the rates of the split codebooks: $R_1 = R_{1p} + R_{1c}$. The benefit of using a split codebook¹, is that Receiver 2 can decode Sender 1's common message m_{1c} and achieve a better communication rate by using interference cancellation. Because only part of the interfering message is used, we call this *partial* interference cancellation. Sender 2 will also split her message m_2 into two parts: m_{2p} and m_{2c} .

Codebook construction: Consider the auxiliary random variables Q, U_1, W_1, U_2, W_2 and the class of Han-Kobayashi probability distributions, \mathcal{P}_{HK} , which factorize as

¹ Note that the Han-Kobayashi strategy is also referred to as a *rate-splitting* in the literature. In this document we reserve this term *rate-splitting* for the use of a split codebook and successive decoding as in [GRUW01] and [Rim01].

5.3 The quantum Han-Kobayashi rate region

$p_{HK}(q, u_1, w_1, x_1, u_2, w_2, x_2) = p(q)p(u_1|q)p(w_1|q)p(x_1|u_1, w_1)p(u_2|q)p(w_2|q)p(x_2|u_2, w_2)$, where $p(x_1|u_1, w_1)$ and $p(x_2|u_2, w_2)$ are degenerate probability distributions that correspond to deterministic functions f_1 and f_2 , $f_i: \mathcal{U}_i \times \mathcal{W}_i \rightarrow \mathcal{X}_i$, which are used to combine the values of U and W to produce a symbol X suitable as input to the channel.

We generate the random codebooks in the following manner:

- Randomly and independently generate a sequence q^n according to $\prod_{i=1}^n p_Q(q_i)$.
- Randomly and independently generate $2^{nR_{1c}}$ sequences $w_1^n(m_{1c})$, $m_{1c} \in [1 : 2^{nR_{1c}}]$ conditionally on the sequence q^n according to $\prod_{i=1}^n p_{W_1|Q}(w_{1i}|q_i)$.
- Randomly and independently generate $2^{nR_{1p}}$ sequences $u_1^n(m_{1p})$, $m_{1p} \in [1 : 2^{nR_{1p}}]$ conditionally on the sequence q^n according to $\prod_{i=1}^n p_{U_1|Q}(u_{1i}|q_i)$.
- Apply the function f_1 symbol-wise to the codewords $w_1^n(m_{1c})$ and $u_1^n(m_{1p})$ to obtain the codeword $x_1^n(m_{1c}, m_{1p})$.
- We generate the common and personal codebooks for Sender 2 in a similar fashion and combine them using f_2 to obtain $x_2^n(m_{2c}, m_{2p})$.

Decoding: When the split codebooks are used for the interference channel, we are effectively coding for an interference network with four inputs and two outputs. We can think of the decoding performed by each of the receivers as two multiple access channel (MAC) decoding subproblems. We will denote the achievable rate regions for the MAC sub-problems as $\mathcal{R}_{HK}^{(o,1)}$ and $\mathcal{R}_{HK}^{(o,2)}$. The task for Receiver 1 is to decode the messages (m_{1p}, m_{1c}, m_{2c}) , and thus the sub-task $\mathcal{R}_{HK}^{(o,1)}$ corresponds to a three-sender multiple access channel, the rate region for which is described by seven inequalities on the rate triples (R_{1p}, R_{1c}, R_{2c}) . The decoding task for Receiver 2, $\mathcal{R}_{HK}^{(o,2)}$, is similarly described by seven inequalities on the rates (R_{1c}, R_{2c}, R_{2p}) .

We perform Fourier-Motzkin elimination on the inequalities of the MAC rate regions for the two receivers in order to eliminate the variables R_{1p} , R_{1c} , R_{2p} and R_{2c} and replacing them with the sum variables

$$R_1 = R_{1p} + R_{1c}, \quad R_2 = R_{2p} + R_{2c}. \quad (5.20)$$

At each step in the Fourier-Motzkin elimination process, we use the information the-

oretic properties in order to eliminate redundant inequalities. The result is the Han-Kobayashi rate region.

Theorem 5.3 (Quantum Han-Kobayashi rate region). *Consider the region:*

$$\mathcal{R}_{HK}^o(\mathcal{N}) \equiv \bigcup_{\substack{p_{HK} \in \mathcal{P}_{HK} \\ f_1, f_2}} \{(R_1, R_2) \in \mathbb{R}^2 \mid \text{Eqns. (HK1) - (HK9)}\}$$

$$R_1 \leq I(U_1 W_1; B_1 | W_2 Q) \quad (\text{HK1})$$

$$R_1 \leq I(U_1; B_1 | W_1 W_2 Q) + I(W_1; B_2 | U_2 W_2 Q) \quad (\text{HK2})$$

$$R_2 \leq I(U_2 W_2; B_2 | W_1 Q) \quad (\text{HK3})$$

$$R_2 \leq I(W_2; B_1 | U_1 W_1 Q) + I(U_2; B_2 | W_1 W_2 Q) \quad (\text{HK4})$$

$$R_1 + R_2 \leq I(U_1 W_1 W_2; B_1 | Q) + I(U_2; B_2 | W_1 W_2 Q) \quad (\text{HK5})$$

$$R_1 + R_2 \leq I(U_1; B_1 | W_2 W_1 Q) + I(U_2 W_2 W_1; B_2 | Q) \quad (\text{HK6})$$

$$R_1 + R_2 \leq I(U_1 W_2; B_1 | W_1 Q) + I(U_2 W_1; B_2 | W_2 Q) \quad (\text{HK7})$$

$$2R_1 + R_2 \leq I(U_1; B_1 | W_1 W_2 Q) + I(U_2 W_1; B_2 | W_2 Q) \\ + I(U_1 W_1 W_2; B_1 | Q) \quad (\text{HK8})$$

$$R_1 + 2R_2 \leq I(U_1 W_2; B_1 | W_1 Q) + I(U_2; B_2 | W_2 W_1 Q) \\ + I(U_2 W_2 W_1; B_2 | Q) \quad (\text{HK9})$$

where the information theoretic quantities are taken with respect to a state $\theta^{U_1 U_2 W_1 W_2 B_1 B_2}$ of the form:

$$\sum_{\substack{q, u_1, u_2, \\ w_1, w_2}} p_Q(q) p_{U_1|Q}(u_1|q) p_{U_2|Q}(u_2|q) p_{W_1|Q}(w_1|q) p_{W_2|Q}(w_2|q) |q\rangle\langle q|^Q \otimes \\ \otimes |u_1\rangle\langle u_1|^{U_1} \otimes |u_2\rangle\langle u_2|^{U_2} \otimes |w_1\rangle\langle w_1|^{W_1} \otimes |w_2\rangle\langle w_2|^{W_2} \otimes \rho_{f_1(u_1, w_1), f_2(u_2, w_2)}^{B_1 B_2}$$

is an achievable rate region provided Conjecture 4.1 holds.

Each of the inequalities (HK1)-(HK9) describes some limit imposed on the personal or common rates of the two senders. For example, (HK1) corresponds to the maximum rate at which m_{1p} and m_{1c} can be decoded by Receiver 1 *given* that he has already decoded m_{2c} . Other inequalities correspond to mixed bounds, in which one of the terms comes from a constraint on Receiver 1 and the other from a constraint on Receiver 2. An example of this is (HK2) which comes from the bound on Receiver 1's ability to decode m_{1p} (given m_{1c} and m_{2c}) and a bound from Receiver 2's ability to decode m_{1c}

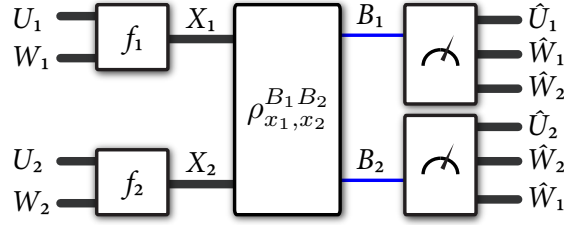


Figure 5.6: The random variables used in the Han-Kobayashi coding strategy. Sender 1 selects codewords according to a “personal” random variable U_1 and a “common” random variable W_1 . She then acts on U_1 and W_1 with some deterministic function f_1 that outputs a variable X_1 which serves as a classical input to the interference channel. Sender 2 uses a similar encoding. Receiver 1 performs a measurement to decode both variables of Sender 1 and the common random variable W_2 of Sender 2. Receiver 2 acts similarly. The advantage of this coding strategy is that it makes use of interference in the channel by having each receiver partially decode what the other sender is transmitting. Theorem 5.3 gives the rates that are achievable assuming that Conjecture 4.1 holds.

(given m_{2c} and m_{2p})².

Note that the original description of the rate region given by Han and Kobayashi in [HK81] and later in [HK07] contained two extra inequalities. Chong *et al.* showed that these extra inequalities are redundant, and so the best description of \mathcal{R}_{HK} involves only nine inequalities as above [CMGEG08].

Proof. The proof is in the same spirit as the original result of Han and Kobayashi [HK81]. The first step is to use the Lemma 5.1 to obtain:

$$\begin{aligned} & \left(I^{B_1^n B_2^n} - \Lambda_{m_{1p}, m_{1c}, m_{2c}}^{B_1^n} \otimes \Gamma_{m_{1c}, m_{2c}, m_{2p}}^{B_2^n} \right) \\ & \leq \left(I^{B_1^n} - \Lambda_{m_{1p}, m_{1c}, m_{2c}}^{B_1^n} \right) \otimes I^{B_2^n} + I^{B_1^n} \otimes \left(I^{B_2^n} - \Gamma_{m_{1c}, m_{2c}, m_{2p}}^{B_2^n} \right), \end{aligned}$$

which allows us to bound the error analysis for the interference channel task in terms of the error analysis for two MAC sub-channels. Our result is conditional on Conjecture 4.1 for the construction of the decoding POVMs for each MAC sub-channel: $\{\Lambda_{m_{1p}, m_{1c}, m_{2c}}\}$ for Receiver 1, and $\{\Gamma_{m_{1c}, m_{2c}, m_{2p}}\}$ for Receiver 2. \square

² Receiver 2 is not required to decode the common message of Sender 1, but the Han-Kobayashi strategy does require this condition despite the fact there could be no interference cancellation benefits for doing so, given that Receiver 2 has already decoded the messages m_{2c} and m_{2p} . This should serve as a hint that the Han-Kobayashi decoding requirements can be relaxed. We will discuss this further in the next section.

At the very least, observe that Theorem 5.3 depends on Conjecture 4.1 for its proof. While we do not doubt that the conjecture will ultimately turn out to be true, the fact remains that our result is conditional on an unproven conjecture, which is somewhat unsatisfactory.

In order to remedy this shortcoming, we searched for other approaches which could be used to prove that the rates of the quantum Han-Kobayashi rate region are achievable. First, we proved that the quantum Han-Kobayashi rate region is achievable for a special class of interference channels where the output states commute. We also derived an achievable rate region described in terms of min-entropies [Ren05, Tom12], which is in general smaller than the Han-Kobayashi rate region. These results are well documented in [FHS⁺12]. Another approach which we studied is the use of a rate-splitting and successive decoding approach in order to achieve the rates of the Han-Kobayashi rate region. We attempted to adapt the results of Şaşıoğlu in [Sas08], which claimed, erroneously, that the rate-splitting strategy can be used in order to achieve the Chong-Motani-Garg (CMG) rate region. Recall that the Chong-Motani-Garg rate region is equivalent to the Han-Kobayashi rate region [CMGEG08]. In fact, as we will see shortly, the Chong-Motani-Garg approach is simply a specific coding strategy to carry out the Han-Kobayashi partial interference cancellation idea.

The analysis in [Sas08] is in two parts. The first part is a geometric argument, henceforth referred to as the *Şaşıoğlu argument*, which shows that there is a many-to-one mapping between the rates of the split codebooks $(R_{1p}, R_{1c}, R_{2c}, R_{2p})$, and the resulting rates (R_1, R_2) for the interference channel task. In the second part of the analysis, Şaşıoğlu describes a strategy for the use of rate-splitting and successive decoding for the common message. The common-message codebook for one sender is split so as to accommodate one of the receivers assuming the common-message codebook of the other sender is not split. However, if both users split their common-message codebooks, the rates cannot be chosen, in general, so as to achieve all the rates of the Chong-Motani-Garg rate region. We will comment on this further in Section 5.6.

While rate-splitting and successive decoding turned out to be a dead end in our quest for the quantum Han-Kobayashi region, the *Şaşıoğlu argument* and the use of two-sender simultaneous decoding turns out to be sufficient in order to show the achievability of the quantum Chong-Motani-Garg rate region. This will be the subject of Section 5.5 below.

5.4 The quantum Chong-Motani-Garg rate region

The achievability of the quantum Chong-Motani-Garg (CMG) rate region was recently proved by Sen using novel geometric ideas for the “intersection subspace” of projectors and a “sequential decoding” technique [Sen12a]. In this section we will describe the CMG coding strategy and state Sen’s result in Theorem 5.4. In Section 5.4, we will provide an alternate proof of this result based on the Şaşoğlu argument [Sas08] and the two-sender simultaneous decoding techniques from Theorem 4.2.

The differences between the Chong-Motani-Garg coding strategy and the Han-Kobayashi coding strategy are: (1) the different way the senders’ codebooks are constructed and (2) the relaxed decoding requirements for the two receivers. We discuss these next.

Codebook construction: The codebooks are constructed using the *superposition coding* technique, which was originally developed by Cover in the context of the classical broadcast channel [Cov72]. The idea behind this encoding strategy is to first generate a set of *cloud centers* for each common message and then choose the satellite codewords for the personal messages relative to the cloud centers.

Let Q, W_1, W_2 be auxiliary random variables and let \mathcal{P}_{CMG} be the class of probability density functions which factorize as $p_{\text{CMG}}(q, w_1, x_1, w_2, x_2) = p(q) p(w_1|q) p(x_1|w_1, q) p(w_2|q) p(x_2|w_2, q)$. To construct the codebook we proceed as follows:

- First randomly and independently generate a sequence q^n according to $\prod_{i=1}^n p_Q(q_i)$.
- Randomly and independently generate $2^{nR_{1c}}$ sequences $w_1^n(m_{1c})$, $m_{1c} \in [1 : 2^{nR_{1c}}]$ conditionally on the sequence q^n according to $\prod_{i=1}^n p_{W_1|Q}(w_{1i}|q_i)$.
- Next, for each message m_{1c} , we randomly and independently generate $2^{nR_{1p}}$ conditional codewords $x_1^n(m_{1p}|m_{1c})$, $m_{1p} \in [1 : 2^{nR_{1p}}]$, $m_{1c} \in [2^{nR_{1c}}]$ according to the product conditional probability distribution $\prod_{i=1}^n p_{X_1|W_1Q}(x_{1i}|w_{1i}(m_{1c}), q_i)$.
- We generate the common and personal codebooks for Sender 2 in a similar fashion. First generate $\{w_2^n(m_{2c})\}$, $m_{2c} \in [2^{nR_{2c}}]$ according to $\prod^n p_{W_2|Q}$ and then generate $\{x_2^n(m_{2p}|m_{2c})\}$, $m_{2p} \in [2^{nR_{2p}}]$, $m_{2c} \in [2^{nR_{1c}}]$ conditionally on $w_2^n(m_{2c})$ according to $\prod^n p_{X_2|W_2Q}$.

Decoding for the MAC subproblems: The decoding task for each of the receivers is associated with a multiple access channel subproblem. We will denote the achievable rate regions for the MAC sub-problems for a fixed input distribution $p_{\text{CMG}} \in \mathcal{P}_{\text{CMG}}$ as $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ and $\mathcal{R}_{\text{CMG}}^2(\mathcal{N}, p_{\text{CMG}})$.

Consider the decoding task for Receiver 1. The messages to be decoded are (m_{1p}, m_{1c}, m_{2c}) , while the effects of the message m_{2p} superimposed on top of the codeword for m_{2c} are considered as noise to be averaged over. The desired achievable rate region $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ is defined as follows:

$$\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}}) \triangleq \bigcup_{\substack{p(x_1|w_1, q)p(w_1|q) \\ p(x_2|w_2, q)p(w_2|q)p(q)}} \{(R_{1p}, R_{1c}, R_{2c}) \in \mathbb{R}_+^3 \mid \text{Eqns (a1)-(d1) below}\}$$

$$R_{1p} \leq I(X_1; B_1 | W_1 W_2 Q) \triangleq I(a_1), \quad (\text{a1})$$

$$R_{1p} + R_{1c} \leq I(X_1; B_1 | W_2 Q) \triangleq I(b_1), \quad (\text{b1})$$

$$R_{1p} + R_{2c} \leq I(X_1 W_2; B_1 | W_1 Q) \triangleq I(c_1), \quad (\text{c1})$$

$$R_{1p} + R_{1c} + R_{2c} \leq I(X_1 W_2; B_1 | Q) \triangleq I(d_1). \quad (\text{d1})$$

The mutual information quantities are calculated with respect to the following state:

$$\sum_{\substack{q, w_1, \\ x_1, w_2}} p(q) p(w_1|q) p(x_1|w_1, q) p(w_2|q) \times \quad (5.21) \\ |q\rangle\langle q|^Q \otimes |w_1\rangle\langle w_1|^{W_1} \otimes |x_1\rangle\langle x_1|^{X_1} \otimes |w_2\rangle\langle w_2|^{W_2} \otimes \rho_{x_1, w_2}^{B_1},$$

where

$$\rho_{x_1, w_2}^{B_1} \equiv \sum_{x_2} p(x_2|w_2) \text{Tr}_{B_2} [\rho_{x_1, x_2}^{B_1 B_2}] \quad (5.22)$$

is the effective code state for Receiver 1. It is the average over the random variable X_2 (since we treat m_{2p} as noise) and the partial trace over the degrees of freedom associated with Receiver 2.

The rate region for Receiver 2 is similarly described by:

$$\mathcal{R}_{\text{CMG}}^2(\mathcal{N}, p_{\text{CMG}}) \triangleq \bigcup_{\substack{p(x_1|w_1, q)p(w_1|q) \\ p(x_2|w_2, q)p(w_2|q)p(q)}} \{(R_{2p}, R_{2c}, R_{1c}) \in \mathbb{R}_+^3 \mid \text{Eqns (a2)-(d2) below}\} \quad (5.23)$$

5.4 The quantum Chong-Motani-Garg rate region

$$R_{2p} \leq I(X_2; B_2 | W_1 W_2 Q) \quad \triangleq I(a_2), \quad (\text{a2})$$

$$R_{2p} + R_{2c} \leq I(X_2; B_2 | W_1 Q) \quad \triangleq I(b_2), \quad (\text{b2})$$

$$R_{2p} + R_{1c} \leq I(X_2 W_1; B_2 | W_2 Q) \quad \triangleq I(c_2), \quad (\text{c2})$$

$$R_{2p} + R_{2c} + R_{1c} \leq I(X_2 W_1; B_2 | Q) \quad \triangleq I(d_2), \quad (\text{d2})$$

with respect to a code state in which the variable X_1 is treated as noise and a partial trace over the system B_1 is performed.

Observe that the above MAC rate regions are described only by four inequalities, rather than by seven inequalities like the multiple access channel with three senders (cf. Conjecture 4.1). Two of the rate constraints do not appear because we are using the superposition encoding technique and always decode m_{1c} before m_{1p} . A third inequality can be dropped if we recognize that Receiver 1 is not *really* interested in decoding m_{2c} ; he is only decoding m_{2c} to serve as side information which will help him decode the messages m_{1c} and m_{1p} intended for him. This is called *relaxed decoding*, and allows us to drop the constraint associated the decoding of m_{2c} after m_{1c} and m_{1p} [CMG06]. The relaxed decoding approach cannot be applied directly to the quantum case, and so a different decoding strategy is required [Sen12a]. We postpone the discussion about the decoding strategies of the receivers until the end of this section.

We are now in a position to describe the Chong-Motani-Garg rate region \mathcal{R}_{CMG} , which is obtained by combining the constraints from $\mathcal{R}_{\text{CMG}}^1$ and $\mathcal{R}_{\text{CMG}}^2$. Recall that, for the interference channel problem, we are interested in the *total* rates achievable between each sender and the corresponding receiver. For Receiver 1, we have a net rate of $R_1 = R_{1c} + R_{1p}$ and similarly for Receiver 2 we have $R_2 = R_{2c} + R_{2p}$. Consider the projection $\mathbf{\Pi}$ which takes the 4-tuple of rates $(R_{1p}, R_{1c}, R_{2c}, R_{2p})$ to the space of net rates (R_1, R_2) :

$$\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = \begin{bmatrix} R_{1p} + R_{1c} \\ R_{2p} + R_{2c} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}}_{\mathbf{\Pi}} \begin{bmatrix} R_{1p} \\ R_{1c} \\ R_{2c} \\ R_{2p} \end{bmatrix}. \quad (5.24)$$

The Chong-Motani-Garg rate region for the interference channel is obtained by taking the union over all input distributions of the intersection between the two MAC rate

regions, followed by the projection $\mathbf{\Pi}$ to obtain:

$$\mathcal{R}_{\text{CMG}}(\mathcal{N}) \equiv \mathbf{\Pi} \left(\bigcup_{p_{\text{CMG}} \in \mathcal{P}_{\text{CMG}}} \mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}}) \cap \mathcal{R}_{\text{CMG}}^2(\mathcal{N}, p_{\text{CMG}}) \right). \quad (5.25)$$

Equivalently, it is possible to compute the intersection of the two MAC rate regions by performing Fourier-Motzkin elimination on the inequalities from equations (a1)-(d1) and (a2)-(d2). By taking all possible combinations of the inequalities in the two MAC subproblems, we obtain the equivalent set of inequalities in the two dimensional space (R_1, R_2) . The resulting achievable rate region has the following form:

Theorem 5.4 (Quantum Chong-Motani-Garg rate region [Sen12a]). *The following rate region is achievable for the quantum interference channel:*

$$\mathcal{R}_{\text{CMG}}(\mathcal{N}) \triangleq \bigcup_{\substack{p(x_1|w_1, q)p(w_1|q) \\ p(x_2|w_2, q)p(w_2|q)p(q)}} \{(R_1, R_2) \in \mathbb{R}_+^2 \mid \text{Eqns. (CMG1)-(CMG9) hold.}\} \quad (5.26)$$

$$R_1 \leq I(X_1; B_1|W_2Q) \quad (\text{CMG1})$$

$$R_1 \leq I(X_1; B_1|W_1W_2Q) + I(X_2W_1; B_2|W_2Q) \quad (\text{CMG2})$$

$$R_2 \leq I(X_2; B_2|W_1Q) \quad (\text{CMG3})$$

$$R_1 \leq I(X_1W_2; B_1|W_1Q) + I(X_2; B_2|W_1W_2Q) \quad (\text{CMG4})$$

$$R_1 + R_2 \leq I(X_1W_2; B_1Q) + I(X_2; B_2|W_1W_2Q) \quad (\text{CMG5})$$

$$R_1 + R_2 \leq I(X_1; B_1|W_1W_2Q) + I(X_2W_1; B_2Q) \quad (\text{CMG6})$$

$$R_1 + R_2 \leq I(X_1W_2; B_1|W_1Q) + I(X_2W_1; B_2|W_2Q) \quad (\text{CMG7})$$

$$2R_1 + R_2 \leq I(X_1W_2; B_1|Q) + I(X_1; B_1|W_1W_2Q) + I(X_2W_1; B_2|W_2Q) \quad (\text{CMG8})$$

$$R_1 + 2R_2 \leq I(X_2; B_2|W_1W_2Q) + I(X_2W_1; B_2|Q) + I(X_1W_2; B_1|W_1Q) \quad (\text{CMG9})$$

where the information theoretic quantities are taken with respect to a state of the form $\rho_{QW_1X_1W_2X_2B_1B_2} \equiv$

$$\sum_{\substack{q, w_1, w_2, \\ x_1, x_2}} p_Q(q) p_{W_1|Q}(w_1|q) p_{W_2|Q}(w_2|q) p_{X_1|W_1Q}(x_1|w_1, q) p_{X_2|W_2Q}(x_2|w_2, q) \\ |q\rangle\langle q|^Q \otimes |w_1\rangle\langle w_1|^{W_1} \otimes |w_2\rangle\langle w_2|^{W_2} \otimes |x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_2} \otimes \rho_{x_1, x_2}^{B_1B_2}.$$

The classical CMG rate region is known to be equivalent to the Han-Kobayashi

rate region [CMGEG08]. Thus, Sen’s achievability proof for the rates of the Chong-Motani-Garg rate region is also a proof of the quantum Han-Kobayashi rate region.

Quantum relaxed decoding

Let us consider more closely the relaxed decoding approach that is employed by Receiver 1 in the *classical case*. The decoding strategy for Receiver 1 is to use jointly typical decoding and search the codebooks $\{w_1^n(m_{1c})\}$, $\{x_1^n(m_{1p}|m_{1c})\}$ and $\{w_2^n(m_{2c})\}$ for messages $(m_{1c}, m_{1p}, \hat{m}_{2c})$ such that

$$\left(w_1^n(m_{1c}), x_1^n(m_{1p}|m_{1c}), w_2^n(\hat{m}_{2c}), Y_1^n \right) \in \mathcal{J}_\delta^{(n)}(W_1, X_1, W_2, Y_1).$$

If such messages are found, the decoder will output $m_1 = (m_{1c}, m_{1p})$. This decoding is *relaxed* because the above condition can be satisfied for some \hat{m}_{2c} which is not necessarily the correct m_{2c} transmitted by Sender 2.

The use of the relaxed decoding strategy allows us to drop the following constraint:

$$R_{2c} \leq I(W_2; B_1 | W_1 X_1), \tag{5.27}$$

which corresponds to the message m_{2c} being decoded last, given the side information of m_{1c} and m_{1p} .

The relaxed decoding strategy does not generalize readily to the case where a quantum decoding is to be performed [Sen12a]. For each message triple (m_{1c}, m_{1p}, m_{2c}) , we could define the measurement $\{\Lambda_{m_{1c}, m_{1p}, m_{2c}}\}$, but how does one combine the measurement operators $\{\Lambda_{m_{1c}, m_{1p}, \hat{m}_{2c}}\}$, $\hat{m}_{2c} \in [2^{nR_{2c}}]$ to form a “relaxed measurement”? Indeed, the usual quantum measurements we use are ones that “ask specific questions” and for which one outcome is more likely than the others. This allows us to use the gentle operator lemma which tells us that the our measurement disturbs the system only marginally.

Sen sidestepped the difficulty of asking a “vague” question by using two different decoding strategies depending on which rates we want to achieve. Receiver 1 will either decode m_{2c} or ignore it altogether. The set of achievable rates for Receiver 1

$(R_{1p}, R_{1c}, R_{2c}) \in \mathbb{R}_+^3$ obtained by Sen is described as follows:

$$\begin{aligned}
R_{2c} &\leq I(W_2; B_1|X_1), \\
R_{1p} &\leq I(X_1; B_1|W_1W_2), & R_{2c} &\geq I(W_2; B_1|X_1), \\
R_{1c} + R_{1p} &\leq I(X_1; B_1|W_2), & \text{OR} & & R_{1p} &\leq I(X_1; B_1|W_1), \\
R_{2c} + R_{1p} &\leq I(X_1W_2; B_1|W_1), & & & R_{1c} + R_{1p} &\leq I(X_1; B_1), \\
R_{1c} + R_{2c} + R_{1p} &\leq I(X_1W_2; B_1),
\end{aligned}$$

Note that the region is not convex. To achieve the rates on the left hand side, Sen developed a novel three-sender simultaneous decoding measurement. The rates on the right hand side correspond to a *disinterested MAC* problem, in which the message m_{2c} will not be decoded. After taking the intersection of the achievable rate regions for Receiver 1 and Receiver 2 and applying the projection as in (5.25), Sen obtained a region which is equivalent to the quantum CMG rate region [Sen12a].

In the next section we will describe another route to prove the achievability of the quantum CMG rate region. We will show that the use of three-sender simultaneous decoding is not necessary. Each of the receivers will use one of *three* different decoding strategies that only require two-sender simultaneous decoding and, in combination, these decoding strategies achieve all the rates $(R_1, R_2) \in \mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$.

5.5 Quantum CMG rate region via two-sender simultaneous decoding

In the original Han-Kobayashi paper [HK81] and the subsequent Chong-Motani-Garg papers [CMG06, CMGEG08], the decoding strategy is to use the three-sender simultaneous decoder. This strategy allows for *all possible* interference cancellation scenarios. An example of a *specific* decoding strategy would be to decode the interference message m_{2c} simultaneously with m_{1c} and then decode m_{1p} last using the side information from both common messages. We denote this $(m_{1c}, m_{2c}) \rightarrow m_{1p}|m_{1c}m_{2c}$. Another example would be to decode m_{1p} and m_{2c} simultaneously after having decoded m_{1c} first: $m_{1c} \rightarrow (m_{1p}|m_{1c}, m_{2c}|m_{1c})$. Simultaneous decoding is a catchall strategy that subsumes all of the above specific strategies. However, as we saw in Chapter 4, the existence of a simultaneous decoder for a general three-sender QMAC is still an open

problem (Conjecture 4.1). It would therefore be desirable to find some specific quantum decoding strategy (or a set of strategies like in [Sen12a]), which can be used to achieve all the rates of the quantum CMG rate region.

In this section, we will extend the geometrical argument presented in [Sas08], to do away with the need for the simultaneous decoding of three messages. We will show that the quantum two-sender simultaneous decoder from Theorem 4.2 is sufficient to achieve the quantum Han-Kobayashi rate region.

Observe that in equation (5.24) only the sum rate $R_{1c} + R_{1p}$ is of importance for Receiver 1. The relative values of R_{1c} and R_{1p} are not important — only their sum (provided that all the inequalities (a1)-(a4) are satisfied). This fact implies that we are allowed a certain freedom in the way we choose the rates of the codebooks for the interference channel. We define this freedom more formally as follows:

Definition 5.4 (Rate moving operation). Let p_{CMG} be the probability distribution used to construct CMG codebooks. Let \mathcal{C} and \mathcal{C}' be two codebooks with rates

$$\mathcal{C} : (R_{1p}, R_{1c}, R_{2c}, R_{2p}) \tag{5.28}$$

$$\mathcal{C}' : (R_{1p} + \delta_1, R_{1c} - \delta_1, R_{2c} - \delta_2, R_{2p} + \delta_2), \tag{5.29}$$

such that the rates of both codebooks satisfy all the inequalities (a1)-(d1) and (a2)-(d2), then they achieve the same rate pair $(R_1, R_2) \in \mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$. Such a transformation of rate tuples is called a *rate moving* operation.

In words, we say that to achieve the rate pair (R_1, R_2) for the interference channel, we are free to *move* the rate points so as to decrease the common rates and increase the personal rates. Intuitively, such a transformation is interesting because decreasing the common rates will make the decoding task easier overall, since *both* receivers have to decode the common messages whereas only a single receiver needs to decode the personal part. The idea for this *rate moving* operation is due to Eren Şaşoğlu [Sas08].

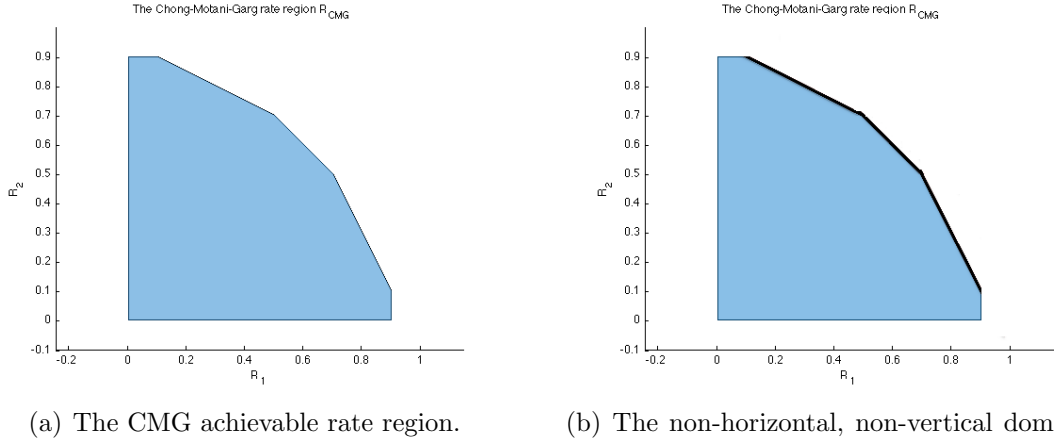
To show the achievability of the Chong-Motani-Garg rate region, $\mathcal{R}_{\text{CMG}}(\mathcal{N})$, it is sufficient to show that we can achieve points on the boundary of the region, which we will denote as $\partial\mathcal{R}_{\text{CMG}}(\mathcal{N})$. In fact, it is sufficient to achieve points on the non-vertical, non-horizontal boundary of the rate region which we will denote $\partial'\mathcal{R}_{\text{CMG}}(\mathcal{N}) \subseteq \partial\mathcal{R}_{\text{CMG}}(\mathcal{N})$. This region is illustrated in Figure 5.7 (b). We refer to the facets that

make up the $\partial'\mathcal{R}_{\text{CMG}}(\mathcal{N})$ as the *dominant facets* of the CMG rate region in analogy with the dominant facet of the multiple access channel capacity region.

We now state the main theorem of this section:

Theorem 5.5 (The dominant facets of the QCMG are achievable). *Any rate pair $(R_1, R_2) \in \partial'\mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$ of the non-horizontal, non-vertical facets of the CMG rate region is achievable for the quantum interference channel $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{N}(x_1, x_2) \equiv \rho_{x_1, x_2}^{B_1 B_2}, \mathcal{H}^{B_1} \otimes \mathcal{H}^{B_2})$.*

As a corollary of the above theorem, we can say that the quantum Chong-Motani-Garg rate region is achievable. Any point in the interior of the CMG rate region $\mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$, is *dominated* by some point on the non-vertical, non-horizontal dominant facets of the boundary $\partial'\mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$. Therefore, we can achieve all other points of the rate region by resource wasting.



(a) The CMG achievable rate region.

(b) The non-horizontal, non-vertical dominant facets of the CMG rate region, $\partial'\mathcal{R}_{\text{CMG}}$, which are achievable by two-sender simultaneous decoding, are shown in bold.

Figure 5.7: The CMG achievable rate region for a given input distribution $p(q)p(w_1, x_1|q)p(w_2, x_2|q)$ in general has the shape of a heptagon. The region is bounded by the two rate positivity conditions and each of the other facets corresponds to one of the inequalities (CMG1)-(CMG9).

The proof of Theorem 5.5 is somewhat long, so we have broken it up into several lemmas. Below we give a brief sketch of the steps involved:

- In Section 5.5.1, we will discuss the geometry of the achievable rate regions $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ and $\mathcal{R}_{\text{CMG}}^2(\mathcal{N}, p_{\text{CMG}})$ for the two receivers. We state Lemma 5.2, which identifies the relative placement of the inequalities (a1)-(d1) by using the

properties of mutual information quantities $I(a_1)$ through $I(d_1)$.

- In Section 5.5.2, we will show that any rate pair $(R_1, R_2) \in \partial' \mathcal{R}_{\text{CMG}}$ can be achieved using codebooks with rates that lie either on the (a) or (c) planes of the MAC rate regions. To show this statement, we will prove Lemma 5.3 which describes a procedure in which we use *rate moving* to transfer any rate point on the (b) or (d) planes to an equivalent rate point on the (a) or (c) planes.
- In Section 5.5.3, we prove that the receivers can use two-sender quantum simultaneous decoding to achieve any rate on the planes (a) and (c). More precisely, there are three possible decode orderings that may be used. Lemma 5.4 shows that the following three decoding strategies (shown for Receiver 1) are *sufficient* to achieve the rates in the CMG rate region:

Case a: $(m_{1c}, m_{2c}) \rightarrow m_{1p}|m_{1c}m_{2c}$,

Case c: $m_{1c} \rightarrow (m_{1p}|m_{1c}, m_{2c}|m_{1c})$,

Case c': $m_{1c} \rightarrow m_{1p}|m_{1c}$.

5.5.1 Geometry of the CMG rate region

For a general input distribution p_{CMG} , the CMG rate region $\mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$ and the two MAC subproblem rate regions could take on different shapes depending on the relative values of the mutual information quantities $I(a_1)$, $I(b_1)$, $I(c_1)$, $I(d_1)$, $I(a_2)$, $I(b_2)$, $I(c_2)$ and $I(d_2)$.

In his paper [Sas08], Şaşoğlu develops a powerful intuition for dealing with the polyhedra that describe their boundaries $\partial \mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$, $\partial \mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ and $\partial \mathcal{R}_{\text{CMG}}^2(\mathcal{N}, p_{\text{CMG}})$. Define the two-dimensional facets a_1, b_1, c_1, d_1 which make up the region boundary. Each facet is a subset of the plane in \mathbb{R}^3 associated with the equality condition of inequalities (a1), (b1), (c1) and (d1), which correspond to the rate constraints of Receiver 1. The boundary of the region $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ can be written as $\partial \mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}}) = a_1 \cup b_1 \cup c_1 \cup d_1$.

We can visualize the three dimensional rate region $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ as in Figure 5.8 below.

This shape of the rate region is governed by the information-theoretic quantities

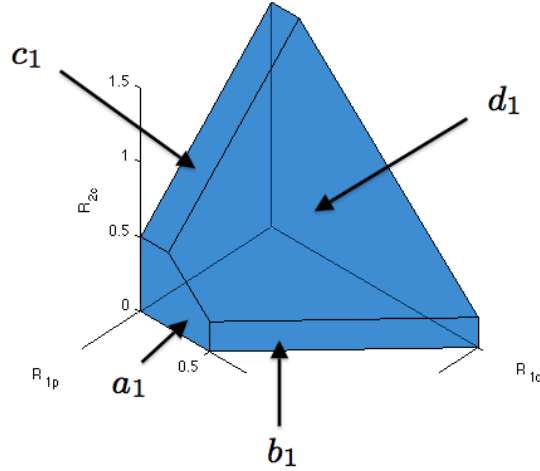


Figure 5.8: The achievable rate region $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ and its bounding facets $a_1, b_1, c_1,$ and d_1 . Each surface is associated with the equality condition in one of the equations (a1), (b1), (c1) and (d1) from page 91.

on the right hand side of equations (a1) through (d1). The following relations establish the geometry of the rate-region $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ which hold for any input distribution.

Lemma 5.2 (Geometry of $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$). *The information-theoretic quantities from equations (a1), (b1), (c1) and (d1) satisfy the following inequalities:*

$$I(a_1) \leq I(b_1) \leq I(d_1), \quad (5.30)$$

$$I(a_1) \leq I(c_1) \leq I(d_1), \quad (5.31)$$

$$I(a_1) + I(d_1) \leq I(b_1) + I(c_1). \quad (5.32)$$

Geometrically $I(a_1) \leq I(b_1)$ indicates that the plane containing b_1 intersects the plane containing a_1 in the positive octant. Similarly $I(b_1) \leq I(d_1)$ indicates that the plane containing d_1 intersects the plane containing b_1 inside \mathbb{R}_+^3 . Equation (5.31) dictates that the plane containing c_1 intersects the plane containing a_1 and that the plane containing d_1 intersects the plane of c_1 . Finally, equation (5.32) states that $I(a_1) + I(d_1) \leq I(b_1) + I(c_1)$, which means that the rate constraint on the sum $2R_{1p} + R_{1c} + R_{2c}$ obtained by adding (a1) and (d1) is tighter than the rate constraint obtained by adding (b1) and (c1). If we define the sets $A = \{1p, 1c\}$ and $B = \{1p, 2c\}$ and $\rho(X)$ to be the information-theoretic quantities of the right hand side, then equation (5.32) has a super-modular polymatroid structure $\rho(A \cap B) + \rho(A \cup B) \leq \rho(A) + \rho(B)$. The proof of Lemma 5.2 is given in Appendix C.1.

5.5.2 Şaşoğlu argument

Let the rate pair $(R_1, R_2) \in \partial' \mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$ be part of the non-horizontal, non-vertical boundary of the two dimensional rate region $\mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$. This rate pair is associated (non-uniquely) to a pair of points $P_1 = (R_{1p}, R_{1c}, R_{2c})$ and $P_2 = (R_{2p}, R_{2c}, R_{1c})$ on the boundaries of the respective regions $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ and $\mathcal{R}_{\text{CMG}}^2(\mathcal{N}, p_{\text{CMG}})$.

Claim 5.6. If the two-dimensional rate pair $(R_1, R_2) \in \partial' \mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$ is the projection of the points $P_1 = (R_{1p}, R_{1c}, R_{2c})$ and $P_2 = (R_{2p}, R_{2c}, R_{1c})$ via the mapping in (5.24), then $P_1 \in \partial \mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$ and $P_2 \in \partial \mathcal{R}_{\text{CMG}}^2(\mathcal{N}, p_{\text{CMG}})$.

Suppose that this were not the case — that is, we assume that at least one of the points, P_i is not on the boundary of its region $\partial \mathcal{R}_{\text{CMG}}^i(\mathcal{N}, p_{\text{CMG}})$. Suppose, for a contradiction, that P_i is in the interior of $\mathcal{R}_{\text{CMG}}^i(\mathcal{N}, p_{\text{CMG}})$, then there must exist a ball of achievable rates of size δ around P_i . This means that we would be able to increase the private rate to $R'_{ip} = R_{ip} + \delta$ for some $\delta > 0$. The resulting point P'_i will be still be achievable so long as we stay within the region $\mathcal{R}_{\text{CMG}}^i(\mathcal{N}, p_{\text{CMG}})$. However, such a δ displacement leads to an increase the sum rate $R'_i = R'_{ip} + R'_{ic} = R_{ip} + \delta + R_{ic} = R_i + \delta$. This contradicts our initial assumption that $(R_1, R_2) \in \partial' \mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$. Therefore, Claim 5.6 must be true, and this means that it is sufficient to show how to achieve all the rates on the boundary of the rate regions $\partial \mathcal{R}_{\text{CMG}}^i(\mathcal{N}, p_{\text{CMG}}) = a_i \cup b_i \cup c_i \cup d_i$.

A priori, we have to consider all possible starting combinations of the points $P_i \in a_i \cup b_i \cup c_i \cup d_i$. However, using the *rate moving* operation (Definition 5.4), we can move any point in $b_i \cup d_i \setminus a_i \cup c_i$ to an equivalent point in $a_i \cup c_i$ as illustrated in Figure 5.9.

Lemma 5.3 (Moving points [Sas08]). *Any point P_i that lies on one of the planes $b_i \cup d_i \setminus a_i \cup c_i$ can be converted to a different point P'_i on one of the planes $a_i \cup c_i$, while leaving the sum rate (R_1, R_2) unchanged.*

In order to be precise, we have to study the effects of the rate moving operation on both points P_1 and P_2 simultaneously. This is because the *same* rates R_{1c} and R_{2c} appear in the common coordinates of both P_1 and P_2 . The reasoning behind the proof of Lemma 5.3 is reminiscent of the argument used to prove Claim 5.6. The details are given in Appendix C.2.

Lemma 5.3 is important because in the next section we will show how to achieve the rates in the facets a_i and c_i using two-sender quantum simultaneous decoding.

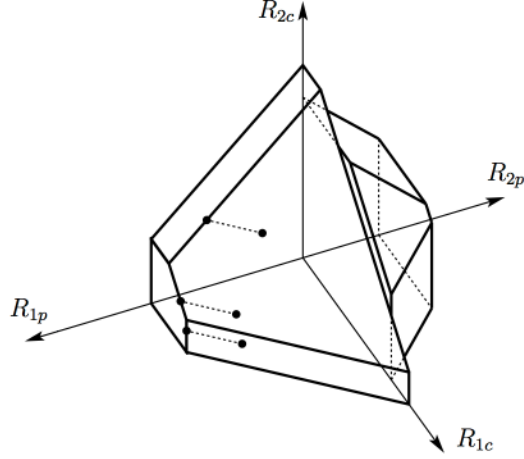


Figure 5.9: Moving points on the b_1 and d_1 facets to equivalent points on a_1 and c_1 .

This means that we can construct a decoder that achieves all the rates for the quantum Chong-Motani-Garg rate region without the need for a three sender simultaneous decoder from Conjecture 4.1.

5.5.3 Two-message simultaneous decoding is sufficient for the rates of the facets a_i and c_i

In this section we show how to achieve the rates on the a_1 and c_1 facets using only two-sender simultaneous decoding.

Lemma 5.4 (Two-simultaneous decoding for a and c planes). *Fix an input distribution $p_{CMG} \in \mathcal{P}_{CMG}$ and let the rate pair $(R_1, R_2) \in \partial\mathcal{R}_{CMG}(\mathcal{N}, p_{CMG})$ come from the rate triples $P_1 = (R_{1p}, R_{1c}, R_{2c}) \in \partial\mathcal{R}_{CMG}^1(\mathcal{N}, p_{CMG})$ and $P_2 = (R_{2p}, R_{2c}, R_{1c}) \in \partial\mathcal{R}_{CMG}^2(\mathcal{N}, p_{CMG})$ such that*

$$(P_1, P_2) \in a_1 \cup c_1 \times a_2 \cup c_2. \quad (5.33)$$

Then the rate (R_1, R_2) is achievable for the QIC using two-sender quantum simultaneous decoding.

Proof. Our analysis is similar to [Sas08], but we are not going to use a rate-splitting strategy.

Achieving points in a: Consider a point $P_1 \in a_1$, which implies

$$R_{1p} = I(X_1; B_1 | W_1 W_2 Q), \quad (5.34)$$

$$R_{1p} + R_{1c} \leq I(X_1; B_1 | W_2 Q), \quad (5.35)$$

$$R_{1p} + R_{2c} \leq I(X_1 W_2; B_1 | W_1 Q), \quad (5.36)$$

$$R_{1p} + R_{1c} + R_{2c} \leq I(X_1 W_2; B_1 | Q). \quad (5.37)$$

We can subtract equation (5.34) from the inequalities below it to obtain a new set of inequalities

$$R_{1p} = I(X_1; B_1 | W_1 W_2 Q), \quad (5.38)$$

$$R_{1c} \leq I(W_1; B_1 | W_2 Q) = I(X_1; B_1 | W_2 Q) - I(X_1; B_1 | W_1 W_2 Q), \quad (5.39)$$

$$R_{2c} \leq I(W_2; B_1 | W_1 Q) = I(X_1 W_2; B_1 | W_1 Q) - I(X_1; B_1 | W_1 W_2 Q), \quad (5.40)$$

$$R_{1c} + R_{2c} \leq I(W_1 W_2; B_1 | Q) = I(X_1 W_2; B_1 | Q) - I(X_1; B_1 | W_1 W_2 Q). \quad (5.41)$$

Looking at equations (5.39)-(5.41) we see that the rates (R_{1c}, R_{2c}) have the form of a MAC rate region with inputs $W_1 \sim p(w_1|q)$, $W_2 \sim p(w_2|q)$ and output B_1 . We will perform the decoding in the following order at Receiver 1: $(W_1, W_2) \rightarrow X_1 | W_1 W_2$.

Consider the quantum channel

$$w_1, w_2 \rightarrow \rho_{w_1, w_2}^{B_1}, \quad (5.42)$$

where $\rho_{w_1, w_2}^{B_1}$ is defined as the average output state assuming superposition encoding of the random variables x_1 and x_2 will be performed:

$$\rho_{w_1, w_2}^{B_1} \equiv \sum_{x_1} \sum_{x_2} p(x_1 | w_1) p(x_2 | w_2) \rho_{x_1, x_2}^{B_1}. \quad (5.43)$$

The decoding strategy for Receiver 1 when the rates are on the facet a_1 correspond to the use of the two-message simultaneous decoder (Theorem 4.2) on the channel shown in (5.42).

After the common parts have been decoded, Receiver 1 will use a conditional HSW decoder to decode the message encoded in X_1 .

Achieving points in c : Consider a point $P_1 \in c_1$, which implies that the constraint on the $R_{1p} + R_{2c}$ inequality is tight.

$$R_{1p} \leq I(X_1; B_1 | W_1 W_2 Q), \quad (5.44)$$

$$R_{1p} + R_{1c} \leq I(X_1; B_1 | W_2 Q), \quad (5.45)$$

$$R_{1p} + R_{2c} = I(X_1 W_2; B_1 | W_1 Q), \quad (5.46)$$

$$R_{1p} + R_{1c} + R_{2c} \leq I(X_1 W_2; B_1 | Q). \quad (5.47)$$

If we subtract (5.46) from (5.47) we obtain the following equivalent set of inequalities.

$$R_{1p} \leq I(X_1; B_1 | W_1 W_2 Q), \quad (5.48)$$

$$R_{1p} + R_{1c} \leq I(X_1; B_1 | W_2 Q), \quad (5.49)$$

$$R_{1p} + R_{2c} = I(X_1 W_2; B_1 | W_1 Q), \quad (5.50)$$

$$R_{1c} \leq I(W_1; B_1 | Q) = I(X_1 W_2; B_1 | Q) - I(X_1 W_2; B_1 | W_1 Q) \quad (5.51)$$

The constraint on the sum rate $R_{1p} + R_{1c}$ imposed by equation (5.49) is less tight than the sum rate constraint obtained by adding equations (5.48) and (5.51), therefore we will drop equation (5.49) from the remainder of the argument. The accuracy of this statement can be verified starting from $I(W_1; W_2 | B_1) \geq 0$ and rearranging the terms. See Appendix C.3 for the details.

The decoding strategy depends on the position of the point P_1 lying within the c_1 plane. We will treat two cases separately.

Case c : Suppose R_{1p} is such that:

$$I(X_1; B_1 | W_1 Q) \leq R_{1p}. \quad (5.52)$$

If we subtract this lower bound on R_{1p} from equation (5.50) we can obtain an upper bound on R_{2c} . We also have an upper bound on R_{1p} from (5.48) and a bound on the sum rate $R_{1p} + R_{2c}$ from (5.50). This gives us the following rate

constraints:

$$R_{1p} \leq I(X_1; B_1 | W_1 W_2 Q), \quad (5.48)$$

$$R_{2c} \leq I(W_2; B_1 | X_1 Q) = (5.50) - (5.52), \quad (5.53)$$

$$R_{1p} + R_{2c} = I(X_1 W_2; B_1 | W_1 Q). \quad (5.50)$$

$$R_{1c} \leq I(W_1; B_1 | Q) \quad (5.54)$$

Şaşoğlu recognizes the rate constraints on (R_{1p}, R_{2c}) in equations (5.48), (5.53) and (5.50) to correspond to the dominant facet of a MAC rate region for a channel with inputs $X_1 \sim p(x_1 | w_1, q)$, $W_2 \sim p(w_2 | q)$ and output (W_1, B_1) . In other words we have a special channel where W_1 is available as side information for Sender 1 and Receiver 1. The decode order is given by: $W_1 \rightarrow (X_1 | W_1, W_2 | W_1)$.

To achieve rates on the plane c_1 , Receiver 1 will first use a standard HSW decoder to decode the message m_{1c} encoded in W_1 and then apply the simultaneous decoding as stated in the following lemma:

Lemma 5.5 (Conditional simultaneous decoding). *Let $\{w_1^n(\ell_1)\}_{\ell_1 \in [2^{nR_{1\alpha}}]}$ be a codebook generated according to $\prod^n p_{W_1}$, and let $\{x_1^n(m_1|w_1^n(\ell_1))\}_{m_1 \in [2^{nR_{1\beta}}], \ell_1 \in [2^{nR_{1\alpha}}]}$ be a conditional codebook generated according to $\prod^n p_{X_1|W_1}$. Similarly for Sender 2, we define codebooks $\{w_2^n(\ell_2)\}_{\ell_2 \in [2^{nR_{2\alpha}}]}$ and another $\{x_2^n(m_2|w_2^n(\ell_2))\}_{m_2 \in [2^{nR_{2\beta}}], \ell_2 \in [2^{nR_{1\alpha}}]}$ generated according to $\prod^n p_{W_2}$ and $\prod^n p_{X_2|W_2}$. Suppose these codebooks are used on n copies of the quantum multiple access channel ρ_{x_1, x_2} , resulting in the map:*

$$(W_1^n, X_1^n, W_2^n, X_2^n) \longrightarrow \rho_{X_1^n|W_1^n, X_2^n|W_2^n}^n. \quad (5.55)$$

Consider the case where W_1^n is known to the receiver, and X_2^n is considered as noise (averaged over). This situation corresponds to the following map:

$$(W_1^n, X_1^n, W_2^n) \longrightarrow (W_1^n, \rho_{X_1^n|W_1^n, W_2^n}^n), \quad (5.56)$$

where we defined $\rho_{X_1^n|W_1^n, W_2^n}^n \equiv \mathbb{E}_{X_2^n} \rho_{X_1^n|W_1^n, X_2^n|W_2^n}^n$, or in terms of the channel outputs:

$$\rho_{X_1^n|W_1^n, W_2^n}^n = \bigotimes_{i=1}^n \left(\sum_{x_2} p_{X_2|W_2}(x_2|W_{2i}) \rho_{X_{1i}, x_2} \right). \quad (5.57)$$

An achievable rate region for the pair $(R_{1\beta}, R_{2\alpha})$ is described by:

$$R_{1\beta} \leq I(X_1; B|W_1W_2), \quad (5.58)$$

$$R_{2\alpha} \leq I(W_2; B|X_1W_1) = I(W_2; B|X_1), \quad (5.59)$$

$$R_{1\beta} + R_{2\alpha} \leq I(X_1W_2; B|W_1), \quad (5.60)$$

where the mutual information quantities are with respect to the state:

$$\theta^{W_1X_1W_2B} \equiv \sum_{w_1, x_1, w_2} p(w_1, x_1)p(w_2) |w_1\rangle\langle w_1|^{W_1} \otimes |x_1\rangle\langle x_1|^{X_1} \otimes |w_2\rangle\langle w_2|^{W_2} \otimes \rho_{x_1, w_2}^B. \quad (5.61)$$

Proof. The proof is similar to the two-sender MAC simultaneous decoding from Theorem 4.2. \square

Case c' : Now suppose that $R_{1p} \leq I(X_1; B_1|W_1Q)$, then the trivial successive decoding strategy is sufficient. Receiver 1 will decode in the order $W_1 \rightarrow X_1$.

The decoding for is done sequentially using HSW decoding. Receiver 1 decodes the message m_{1c} first, followed by m_{1p} . The decoding in this case is similar to

the successive decoding used in Theorem 4.1. The interfering messages m_{2c} and m_{2p} are treated as noise.

□

Thus we see that the combination of Lemma 5.2, Lemma 5.3, and Lemma 5.4 shows that the quantum Chong-Motani-Garg rate region is achievable using only two-sender simultaneous decoding.

5.6 Successive decoding strategies for interference channels

We report on some results concerning achievable rate regions for the interference channel that use the successive decoding approach.

5.6.1 Time-sharing strategies

In Section 4.2 on the multiple access channel, we saw that a successive decoding strategy can be used to achieve all the rates on the dominant vertices of the rate region. Recall that for a fixed choice of encoding distribution $p \equiv p_{X_1}(x_1)p_{X_2}(x_2)$, the two-sender QMAC capacity region has the shape of a pentagon with two extreme points $\alpha_p \equiv (I(X_1; B), I(X_2; B|X_1))$ and $\beta_p \equiv (I(X_1; B|X_2), I(X_2; B))$, which correspond to the rates achievable by successive decoding in two different orders. To achieve the rates in the convex hull of these points, we can use time-sharing between different codes achieving these rates.

Definition 5.5 (Time-sharing). Given two codebooks \mathcal{C}_1 and \mathcal{C}_2 with rates corresponding to rate points α_p and β_p and a desired rate point $P \in \text{conv}(\alpha_p, \beta_p)$, we will have

$$P = t\alpha_p + (1 - t)\beta_p, \quad (5.62)$$

for some $t \in \mathbb{R}$, which we call the *time-sharing* parameter. We can achieve the rates of a point $P^* \approx P$ if we use the rational time-sharing parameter $t^* \approx t$, $t^* \equiv \frac{M}{N} \in \mathbb{Q}$ and the following strategy: during each N block-uses of the channel, use codebook \mathcal{C}_1 during M of them and during the remaining $N - M$ uses of the channel, use codebook \mathcal{C}_2 .

The time-sharing strategy is not well-adapted for the interference channel. This is because the rates of the corner points of the achievable rate regions for the two receivers are not necessarily the same. The time-sharing strategy that works for one of the receivers might not work for the other one.

It is however possible to use successive decoding strategies for an interference channel in the following way. We start by considering a strategy where both receivers are asked to decode both messages, i.e., we are dealing with the compound multiple access channel. Such a strategy defines an achievable rate region known as the “successive decoding inner bound” for the interference channel (cf. page 6-7 of Ref. [EGK10]).

Consider all possible decode orderings that could be used by the two receivers:

$$\begin{aligned}
 \pi_1 : m_2 \rightarrow m_1 | m_2, & & \pi_2 : m_2, \\
 \pi_1 : m_2 \rightarrow m_1 | m_2, & & \pi_2 : m_1 \rightarrow m_2 | m_1, \\
 \pi_1 : m_1, & & \pi_2 : m_1 \rightarrow m_2 | m_1, \\
 \pi_1 : m_1, & & \pi_2 : m_2.
 \end{aligned} \tag{5.63}$$

Using each of these, we can achieve rates arbitrarily close to the following points:

$$P_1 = (I(X_1; B_1 | X_2), \min\{I(X_2; B_1), I(X_2; B_2)\}), \tag{5.64}$$

$$\begin{aligned}
 P_2 = (\min\{I(X_1; B_1 | X_2), I(X_1; B_2)\}, \\
 \min\{I(X_2; B_1), I(X_2; B_2 | X_1)\}), \tag{5.65}
 \end{aligned}$$

$$P_3 = (\min\{I(X_1; B_1), I(X_1; B_2)\}, I(X_2; B_2 | X_1)), \tag{5.66}$$

$$P_4 = (I(X_1; B_1), I(X_2; B_2)). \tag{5.67}$$

We can use time-sharing between these different codes for the interference channel to obtain all other rates in $\text{conv}(P_1, P_2, P_3, P_4)$. This achievable rate region is illustrated in the RHS of Figure 5.10.

5.6.2 Split codebook strategies

We can improve the successive decoding region described in Section 5.6 if we use split codebooks. Inspired by the Han-Kobayashi strategy we make the senders split their messages into two parts: the messages of Sender 1 will be m_{1p} and m_{1c} , and the messages of Sender 2 will be m_{2p} and m_{2c} . As in the Han-Kobayashi strategy, the use

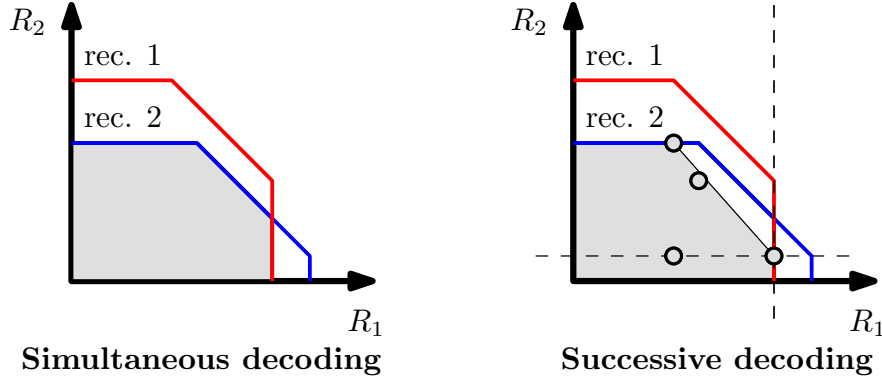


Figure 5.10: These plots show achievable rates regions for the interference channel for simultaneous decoding and successive decoding strategies with fixed input distributions. Using a simultaneous decoding strategy, it is possible to achieve the intersection of the two regions of the corresponding multiple access channels. Using a successive decoding strategy, we obtain four achievable rate points that correspond to the possible decoding orders for the two multiple access channels. The solid red and blue lines outline the different multiple access channel achievable rate regions, and the shaded gray areas outline the achievable rate regions for the two different decoding strategies.

of the split codebooks induces two three-sender multiple access channels. Receiver 1 is required to decode the set of messages m_{1p}, m_{1c} and m_{2c} using successive decoding, and there are six different decode orderings he can use.

Let the decoding ordering of Receiver 1 be represented by a permutation π_1 on the set three elements $\{1p, 1c, 2c\}$. For example, the successive decoding of the messages in the order $m_{2c} \rightarrow m_{1c}|m_{2c} \rightarrow m_{1p}|m_{1c}m_{2c}$ will be denoted as the permutation $\pi_1 = (2c, 1c, 1p)$.

We can naturally use all 6×6 pairs of decoding orders to obtain a set of achievable rate pairs.

Proposition 5.7. Consider the rate point P associated with the decode ordering π_1 for Receiver 1 and π_2 for Receiver 2:

$$P = \left(R_{1p}^{(1)} + \min\{R_{1c}^{(1)}, R_{1c}^{(2)}\}, \min\{R_{2c}^{(1)}, R_{2c}^{(2)}\} + R_{2p}^{(2)} \right),$$

where the rate constraints for Receiver j satisfy

$$R_{\pi_j(1)}^{(j)} \leq I(X_{\pi_j(1)}; B_j), \quad (5.68)$$

$$R_{\pi_j(2)}^{(j)} \leq I(X_{\pi_j(2)}; B_j | X_{\pi_j(1)}), \quad (5.69)$$

$$R_{\pi_j(3)}^{(j)} \leq \begin{cases} I(X_{\pi_j(3)}; B_j | X_{\pi_j(1)} X_{\pi_j(2)}) & \text{if } \pi_j(3) = jc \text{ or } \pi_j(3) = jp \\ \infty, & \text{otherwise} \end{cases} \quad (5.70)$$

The rate pair P is achievable for the quantum interference channel, for all permutations π_1 of the set of indices $(1p, 1c, 2c)$ and for all permutations π_2 of the set $(2p, 2c, 1c)$.

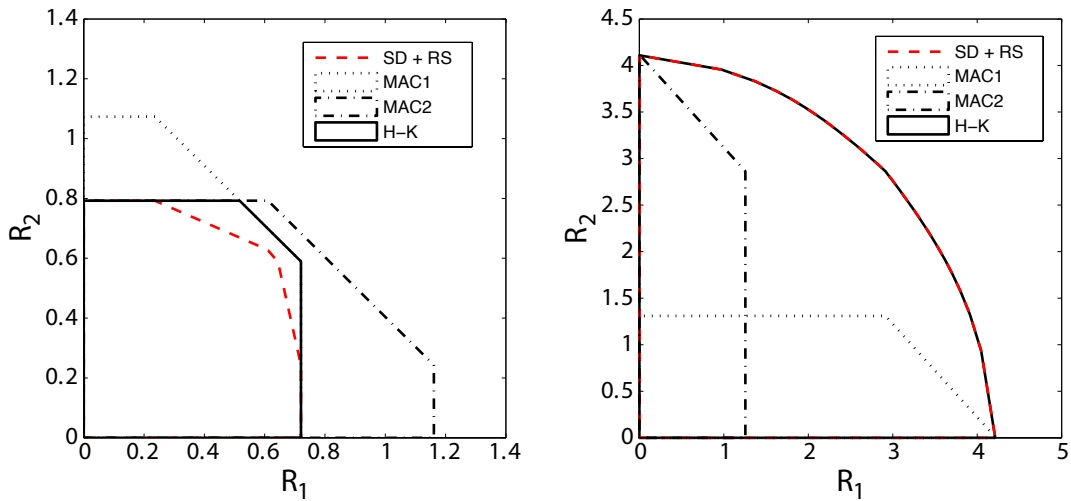


Figure 5.11: These two figures plot rate pairs that the senders and receivers in a classical Gaussian interference channel can achieve using successive decoding and rate-splitting (SD+RS). The figures compare these rates with those achievable by the Han-Kobayashi (HK) coding strategy, while also plotting the regions corresponding to the two induced multiple access channels to each receiver (MAC1 and MAC2). The LHS figure demonstrates that, for a particular choice of signal to noise (SNR) and interference to noise (INR) [ETW07] parameters (SNR1 = 1.7, SNR2 = 2, INR1 = 3.4, INR2 = 4), successive decoding with rate-splitting does not perform as well as the Han-Kobayashi strategy. The RHS figure demonstrates that, for a different choice of parameters (SNR1 = 343, SNR2 = 296, INR1 = 5, INR2 = 5), the two strategies perform equally well.

The rate region described by the convex hull of the points P is generally smaller than the Han-Kobayashi region as illustrated in Figure 5.11. Note that the split-codebook and successive decoding strategy works pretty well in the low interference regime. An interesting open problem is whether we can achieve all rates of the Han-Kobayashi region by splitting each sender's message into more than two parts and using only successive decoding.

In particular, we want to know whether the capacity of the interference channel with strong interference can be achieved using only successive decoding. Alternately, it would be interesting to prove that successive decoding is *not* sufficient in order to achieve all the capacity in the strong interference regime for any number of splits and any possible decode order.

We know that the time-sharing, rate-splitting [Sas08] and generalized time-sharing [YP11] strategies do not work for the interference channel, but is it possible to show a negative result for all successive decoding strategies? This question is explored further in [FS12].

5.7 Outer bound

We will close this chapter by giving a simple outer bound for the capacity of general quantum interference channels analogous to the classical result by Sato [Sat77].

Theorem 5.8 (Quantum Sato outer bound[Sav10]). *Consider the Sato region defined as follows:*

$$\mathcal{R}_{Sato}(\mathcal{N}) \triangleq \bigcup_{p_Q(q)p_1(x_1|q)p_2(x_2|q)} \{(R_1, R_2) \in \mathbb{R}_+^2 \mid \text{Eqns (5.72)-(5.74) below}\}, \quad (5.71)$$

$$R_1 \leq I(X_1; B_1 | X_2 Q)_\theta, \quad (5.72)$$

$$R_2 \leq I(X_2; B_2 | X_1 Q)_\theta, \quad (5.73)$$

$$R_1 + R_2 \leq I(X_1 X_2; B_1 B_2 | Q)_\theta. \quad (5.74)$$

The entropic quantities are with respect to the state $\theta^{QX_1X_2B_1B_2} \equiv$

$$\sum_{q, x_1, x_2} p_Q(q)p_1(x_1|q)p_2(x_2|q) |q\rangle\langle q|^Q \otimes |x_1\rangle\langle x_1|^{X_1} \otimes |x_2\rangle\langle x_2|^{X_2} \otimes \rho_{x_1x_2}^{B_1B_2}. \quad (5.75)$$

Then the region $\mathcal{R}_{Sato}(\mathcal{N})$ is an outer bound on the capacity region of the quantum interference channel.

This proof follows from the observation that any code for the quantum interference channel also gives codes for three quantum multiple access channel subproblems: one for Receiver 1, another for Receiver 2, and a third for the two receivers considered

together. We obtain the outer bound in Theorem 5.8 by using the outer bound on the quantum multiple access channel rates from Theorem 4.1 for each of these channels.

5.8 Discussion

In this chapter we saw how the coding techniques and theorems which we obtained in Chapter 4 can be applied to prove coding theorems for the quantum interference channel.

The key takeaway is that interference is not noise, and that it can be advantageous to the receivers to decode messages in which they are not interested. For Receiver 1, knowing the other user's transmissions allows him to increase the rate at which he can decode going from $I(X_1; B_1) = H(B_1) - H(B_1|X_1)$ to the improved rate of $I(X_1; B_1|X_2) = H(B_1|X_2) - H(B_1|X_1X_2)$.

Because some of our results concerned special cases of the interference channel problem, it is worthwhile to review our overall progress towards the characterization of the capacity region of the general quantum interference channel $\mathcal{C}_{\text{IC}}(\mathcal{N})$. For general interference channels we have:

$$\mathcal{R}_{\text{succ}}(\mathcal{N}) \subsetneq \mathcal{R}_{\text{sim}}(\mathcal{N}) \subsetneq \mathcal{R}_{\text{HK}}^o(\mathcal{N}) \equiv \mathcal{R}_{\text{CMG}}(\mathcal{N}) \subseteq \mathcal{C}_{\text{IC}}(\mathcal{N}) \subseteq \mathcal{R}_{\text{Sato}}(\mathcal{N}).$$

In the special case of the interference channel with very strong interference, the rate region achievable by successive decoding achieves the capacity $\mathcal{R}_{\text{succ}}(\mathcal{N}) = \mathcal{C}_{\text{IC}}(\mathcal{N})$. In the special case of strong interference, the rate region achievable by simultaneous decoding is optimal $\mathcal{R}_{\text{sim}}(\mathcal{N}) = \mathcal{C}_{\text{IC}}(\mathcal{N})$.

An interesting research question would be to investigate whether splitting the messages into more than two parts, that is, turning the two-user IC into a multiple-input multiple-output (MIMO) IC, can improve on the rates that are achievable using the Han-Kobayashi strategy.

In this chapter, we used the superposition coding technique to construct the codebooks for the CMG coding strategy. We will use this technique again in the next chapter in the context of the quantum broadcast channel.

Chapter 6

Broadcast channels

How can a broadcast station communicate separate messages to two receivers using a single antenna? The two message streams must somehow be “mixed” during the encoding process so that the transmitted codewords will contain the information intended for both receivers. In this chapter we apply two codebook construction ideas from the chapter on interference channels to build codebooks for the quantum broadcast channel.

The Chong-Motani-Garg construction used *superposition encoding* to encode a “personal” message (satellite codeword) on top of a “common” message (cloud center). In Section 6.2 we will use the superposition coding technique to encode a “personal” message for one of the receivers on top of a “common” message for both receivers. Such a choice of encoding is well suited for broadcast channels where one of the receivers’ signals is stronger than the other. We can pick the rate of the common message so as to be decodable by the receiver with the weaker reception, and use the left-over capacity to the better receiver to transmit a personal message for him. The superposition coding technique was originally developed in this context [Cov72].

Another approach to constructing the mixing of the information streams is to use two separate codebooks and an arbitrary mixing function that combines them as in the Han-Kobayashi coding strategy. The Marton coding scheme presented in Section 6.3 uses this approach.

6.1 Introduction

The general broadcast communication scenario with two receivers involves the transmission of up to three separate information streams. To illustrate the communication problem, consider the situation described in Figure 6.1 where the television station wants to transmit multiple streams of television programming to two separate receivers.

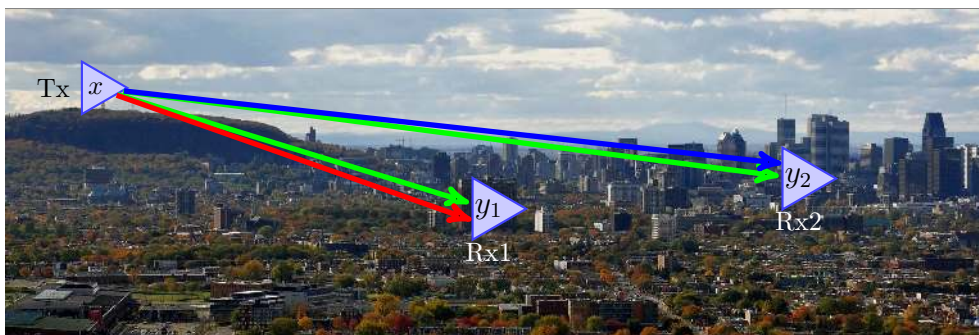


Figure 6.1: The broadcast channel. The sender wishes to transmit three separate information streams: an English language TV station for Receiver 1, a French language TV station for Receiver 2 and a weather TV station which is of interest to both receivers.

Suppose that in each block, the antenna has to transmit a common message $m \in [1 : 2^{nR}]$ intended for both receivers and personal messages $m_1 \in [1 : 2^{nR_1}]$ and $m_2 \in [1 : 2^{nR_2}]$ each intended for one of the receivers. The task is therefore described by the following resource transformation:

$$n \cdot \mathcal{N}^{X \rightarrow Y_1 Y_2} \xrightarrow{(1-\epsilon)} nR_1 \cdot [c \rightarrow c^1] + nR \cdot [c \rightarrow c^1 c^2] + nR_2 \cdot [c \rightarrow c^2].$$

What are the achievable rate triples (R_1, R, R_2) for this communication task?

Note that the everyday usage of the word *broadcast* presumes that only a common message is to be transmitted to all receivers. If only a common message is to be transmitted, that is, we are looking for rates of the form $(0, R, 0)$, the broadcast channel problem reduces to the compound point-to-point channel problem and the capacity is given by the minimum of the rates achievable for the receivers. In order to make the problem interesting from the information theory perspective, we have to consider the case where at least one personal message is to be transmitted.

6.1.1 Previous work

A wide body of research exists in classical information theory on the study of broadcast channels. An excellent review of this research is presented in [Cov98]. The broadcast channel is also covered in textbooks [CT91, EGK11, EGK10]. In the classical case, two of the best known strategies for transmitting information over broadcast channels are superposition coding [Cov72, Ber73, KM77] and Marton over-binning using correlated auxiliary random variables [Mar79]. Sections 6.2 and 6.3 of this chapter are dedicated to the generalization of these coding strategies to classical-quantum broadcast channels.

6.1.2 Quantum broadcast channels

Previous work on quantum broadcast channels includes [YHD11, GSE07, DHL10]. In [YHD11], the authors consider both classical and quantum communication over quantum broadcast channels and prove a superposition coding inner bound similar to our Theorem 6.1. There has also been research on quantum broadcast channels in two other settings: quantum-quantum channels [DHL10] and bosonic broadcast channels [GSE07]. The Marton rate region for the quantum-quantum broadcast channel was developed in [DHL10]. The authors use *decoupling techniques* [ADHW09, AHS08, Dup10] in order to show the Marton achievable rate region with no common message for *quantum* communication¹.

We define a classical-quantum-quantum broadcast channel as the triple:

$$(\mathcal{X}, \mathcal{N}(x) \equiv \rho_x^{B_1 B_2}, \mathcal{H}^{B_1 B_2}), \quad (6.1)$$

where x is a classical letter in an alphabet \mathcal{X} and $\rho_x^{B_1 B_2}$ is a density operator on the tensor product Hilbert space for systems B_1 and B_2 . The model is such that when the sender inputs a classical letter x , Receiver 1 obtains system B_1 , and Receiver 2 obtains system B_2 . Since Receiver 1 does not have access to the B_2 part of the state $\rho_x^{B_1 B_2}$, we model his state as $\rho_x^{B_1} = \text{Tr}_{B_2}[\rho_x^{B_1 B_2}]$, where Tr_{B_2} denotes the

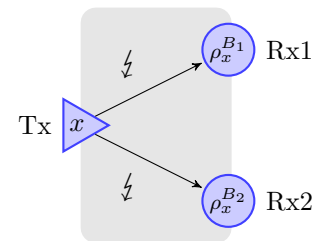


Figure 6.2: A quantum broadcast channel $\rho_x^{B_1 B_2}$.

¹ Note that the well known *no cloning* theorem of quantum information precludes the possibility of a quantum common message: $[q \rightarrow q^1 q^2]$, where the quantum information of some system controlled by the sender is faithfully transferred to *two* receivers. See [YHD11] for more comments on this issue.

partial trace over Receiver 2's system.

6.1.3 Information processing task

The task of communication over a broadcast channel is to use n independent instances of the channel in order to communicate classical information to Receiver 1 at a rate R_1 , to Receiver 2 at a rate R_2 , and to both receivers at a rate R . More specifically, the sender chooses a triple of messages $(m_1, m, m_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR}] \times [1 : 2^{nR_2}]$, and encodes these messages into an n -symbol codeword $x^n(m_1, m, m_2) \in \mathcal{X}^n$ suitable as input for the n channel uses.

The output of the channel is a quantum state of the form:

$$\mathcal{N}^{\otimes n}(x^n(m_1, m, m_2)) \equiv \rho_{x^n(m_1, m, m_2)}^{B_1^n B_2^n} \in \mathcal{D}(\mathcal{H}^{B_1^n B_2^n}), \quad (6.2)$$

where $\rho_{x^n}^{B_1^n B_2^n} \equiv \rho_{x_1}^{B_{11} B_{21}} \otimes \dots \otimes \rho_{x_n}^{B_{1n} B_{2n}}$. To decode the common message m and the message m_1 intended specifically for him, Receiver 1 performs a POVM $\{\Lambda_{m_1, m}\}$, $m_1 \in [1, \dots, |\mathcal{M}_1|]$, $m \in [1, \dots, |\mathcal{M}|]$, on the system B_1^n , the output of which we denote (M'_1, M') . Receiver 2 similarly performs a POVM $\{\Gamma_{m, m_2}\}$, $m_2 \in \{1, \dots, |\mathcal{M}_2|\}$, $m \in [1, \dots, |\mathcal{M}|]$ on the system B_2^n , and his outcome is denoted (M'', M''_2) .

An error occurs whenever either of the receivers decodes one of the messages incorrectly. The probability of error for a particular message triple (m_1, m, m_2) is

$$p_e(m_1, m, m_2) \equiv \text{Tr} \left\{ (I - \Lambda_{m_1, m} \otimes \Gamma_{m, m_2}) \rho_{x^n(m_1, m, m_2)}^{B_1^n B_2^n} \right\},$$

where the measurement operator $(I - \Lambda_{m_1, m} \otimes \Gamma_{m, m_2})$ represents the complement of the correct decoding outcome.

Definition 6.1. An $(n, R_1, R, R_2, \epsilon)$ classical-quantum broadcast channel code consists of a codebook $\{x^n(m_1, m, m_2)\}$, $m_1 \in \mathcal{M}_1$, $m \in \mathcal{M}$, $m_2 \in \mathcal{M}_2$ and two decoding POVMs $\{\Lambda_{m_1, m}\}_{m_1 \in \mathcal{M}_1, m \in \mathcal{M}}$ and $\{\Gamma_{m, m_2}\}_{m \in \mathcal{M}, m_2 \in \mathcal{M}_2}$ such that the average probability of error \bar{p}_e is bounded from above as

$$\bar{p}_e \equiv \frac{1}{|\mathcal{M}_1| |\mathcal{M}| |\mathcal{M}_2|} \sum_{m_1, m, m_2} p_e(m_1, m, m_2) \leq \epsilon. \quad (6.3)$$

We say that a rate pair (R_1, R, R_2) is *achievable* if there exists an $(n, R_1 - \delta, R - \delta, R_2 - \delta, \epsilon)$

quantum broadcast channel code for all $\epsilon, \delta > 0$ and sufficiently large n .

A broadcast channel code with *no common message* is a special case of the above communication task where the rate of the common message is set to zero: $(n, R_1, 0, R_2, \epsilon)$. Alternately, we could choose not to send a personal message for Receiver 2 and obtain codes of the form $(n, R_1, R, 0, \epsilon)$, which is known as the broadcast channel with a *degraded message set* [KM77].

6.1.4 Chapter overview

In this chapter, we derive two achievable rate regions for classical-quantum broadcast channels by exploiting the error analysis techniques developed in the context of quantum multiple access channels (Chapter 4) and quantum interference channels (Chapter 5).

In Section 6.2, we establish the achievability of the rates in the superposition coding rate region (Theorem 6.1). We use a quantum simultaneous decoder at one of the receivers. Yard *et al.* independently proved the quantum superposition coding inner bound [YHD11], but our proof is arguably simpler and more in the spirit of its classical analogue [EGK10].

In Section 6.3 we prove that the quantum Marton rate region with no common message is achievable (Theorem 6.2). The Marton coding scheme is based on the idea of *over-binning* and using correlated auxiliary random variables [Mar79]. The sub-channels to each receiver are essentially point-to-point, but it turns out that the *projector trick* technique seems to be necessary in our proof. The Marton coding scheme gives the best known achievable rate region for the classical-quantum broadcast channel.

6.2 Superposition coding inner bound

One possible strategy for the broadcast channel is to send a message at a rate that is low enough that both receivers are able to decode. Furthermore, if we assume that Receiver 1 has a better reception signal, then the sender can encode a further message *superimposed* on top of the common message that Receiver 1 will be able to decode

given the common message. The sender encodes the common message at rate R using a codebook generated from a probability distribution $p_W(w)$ and the additional message for Receiver 1 at rate R_1 using a conditional codebook with distribution $p_{X|W}(x|w)$. This is known as the superposition coding strategy [Cov72, Ber73].

Theorem 6.1 (Superposition coding inner bound). *Let W be an auxiliary random variable, let $p = p_{X|W}(x|w)p_W(w)$ be an arbitrary code distribution and let $(\mathcal{X}, \rho_x^{B_1B_2}, \mathcal{H}^{B_1B_2})$ be a classical-quantum broadcast channel. The superposition coding rate region $\mathcal{R}_{\text{SC}}(\mathcal{N}, p)$ consists of all rate pairs (R_1, R) such that:*

$$R_1 \leq I(X; B_1|W)_\theta, \quad (6.4)$$

$$R \leq I(W; B_2)_\theta, \quad (6.5)$$

$$R_1 + R \leq I(X; B_1)_\theta, \quad (6.6)$$

is achievable for the quantum broadcast channel. The information quantities are with respect to a state $\theta^{WXB_1B_2}$ of the form:

$$\sum_{w,x} p_W(w)p_{X|W}(x|w) |w\rangle\langle w|^W \otimes |x\rangle\langle x|^X \otimes \rho_x^{B_1B_2}. \quad (6.7)$$

The superposition coding strategy allows us to construct codes for the broadcast channel of the form $(n, R_1, R, 0, \epsilon)$, which have no personal message for Receiver 2. The task is therefore described as follows:

$$n \cdot \mathcal{N}^{X \rightarrow B_1B_2} \xrightarrow{(1-\epsilon)} nR_1 \cdot [c \rightarrow c^1] + nR \cdot [c \rightarrow c^1c^2], \quad (6.8)$$

where $[c \rightarrow c^1c^2]$ denotes the noiseless transmission of one bit to both receivers.

Proof. The new idea in the proof is to exploit superposition coding and a quantum simultaneous decoder for the decoding of the first receiver [Cov72, Ber73] instead of the quantum successive decoding used in [YHD11]. We use a standard HSW decoder for the second receiver [Hol98, SW97].

Codebook generation. We randomly and independently generate 2^{nR} sequences $w^n(m)$ according to the product distribution $\prod_{i=1}^n p_W(w_i)$. For each sequence $w^n(m)$, we then randomly and conditionally independently generate 2^{nR_1} sequences $x^n(m_1, m)$ according to the product distribution: $\prod_{i=1}^n p_{X|W}(x_i|w_i(m))$.

POVM Construction for Receiver 1. We now describe the POVM that Receiver 1 employs in order to decode the transmitted messages. First consider the state we obtain from (6.7) by tracing over the B_2 system:

$$\rho^{WXB_1} = \sum_{w,x} p_W(w) p_{X|W}(x|w) |w\rangle\langle w|^W \otimes |x\rangle\langle x|^X \otimes \rho_x^{B_1}.$$

Consider the following two averaged states:

$$\begin{aligned} \sigma_{w^n}^{B_1^n} &\equiv \sum_{x^n} p_{X^n|W^n}(x^n|w^n) \rho_{x^n}^{B_1^n} = \bigotimes_{i=1}^n \left(\sum_x p_{X|W}(x|w_i) \rho_x^{B_1} \right) = \mathbb{E}_{X^n|w^n} \left\{ \rho_{X^n}^{B_1^n} \right\}, \\ \bar{\rho}^{\otimes n} &\equiv \sum_{w_n, x^n} p_{W^n}(w^n) p_{X^n|W^n}(x^n|w^n) \rho_{x^n}^{B_1^n} = \bigotimes_{i=1}^n \left(\sum_{w,x} p(w)p(x|w) \rho_x^{B_1} \right) = \mathbb{E}_{W^n, X^n} \left\{ \rho_{X^n}^{B_1^n} \right\}. \end{aligned}$$

We now introduce the following shorthand notation to denote the conditionally typical projectors with respect to the output state $\rho_{X^n(m_1, m)}^{B_1^n}$ and the two averaged states defined above:

$$\Pi_{X^n(m_1, m)} \equiv \Pi_{\rho_{X^n(m_1, m)}^{B_1^n}, \delta}, \quad \Pi_{W^n(m)} \equiv \Pi_{\sigma_{W^n(m)}^{B_1^n}, \delta}, \quad \Pi \equiv \Pi_{\bar{\rho}, \delta}.$$

Receiver 1 will decode using a POVM $\{\Lambda_{m_1, m}\}$ defined as the square root measurement:

$$\Lambda_{m_1, m} \equiv \left(\sum_{k_1, k} P_{k_1, k} \right)^{-\frac{1}{2}} P_{m_1, m} \left(\sum_{k_1, k} P_{k_1, k} \right)^{-\frac{1}{2}}, \quad (6.9)$$

based on the following positive operators:

$$P_{m_1, m} \equiv \Pi \Pi_{W^n(m)} \Pi_{X^n(m_1, m)} \Pi_{W^n(m)} \Pi. \quad (6.10)$$

Note the *projector sandwich* structure with the more specific projectors on the inside. We have seen this previously in the construction of the simultaneous decoder POVM for the quantum multiple access channel.

POVM Construction for Receiver 2. Consider now the state in equation (6.7) from the point of view of Receiver 2. If we trace over the X and B_1 systems, we obtain

the following state:

$$\rho^{WB_2} = \sum_w p_W(w) |w\rangle\langle w|^W \otimes \sigma_w^{B_2},$$

where $\sigma_w^{B_2} \equiv \sum_x p_{X|W}(x|w) \rho_x^{B_2}$. Define also the state

$$\bar{\rho} \equiv \sum_{w,x} p_W(w) p_{X|W}(x|w) \rho_x^{B_2}. \quad (6.11)$$

The second receiver uses a standard square root measurement:

$$\Lambda_m \equiv \left(\sum_k P_k \right)^{-\frac{1}{2}} P_m \left(\sum_k P_k \right)^{-\frac{1}{2}}, \quad (6.12)$$

based on the following positive operators:

$$P_m^{B_2^n} = \Pi_{\bar{\rho}, \delta}^{B_2^n} \Pi_{\sigma_{W^n(m), \delta}^{B_2}} \Pi_{\bar{\rho}, \delta}^{B_2^n}, \quad (6.13)$$

where the above projectors are typical projectors defined with respect to the states $\sigma_{W^n(m)}^{B_2^n}$ and $\bar{\rho}^{\otimes n}$.

Error analysis for Receiver 1. We now analyze the expectation of the average error probability for the first receiver with the POVM defined in (6.9):

$$\begin{aligned} & \mathbb{E}_{x^n; W^n} \left\{ \frac{1}{M_1 M_2} \sum_{m_1, m} \text{Tr} \left\{ \left(I - \Gamma_{m_1, m}^{B_1^n} \right) \rho_{X^n(m_1, m)}^{B_1} \right\} \right\} \\ &= \frac{1}{M_1 M_2} \sum_{m_1, m} \mathbb{E}_{x^n; W^n} \left\{ \text{Tr} \left\{ \left(I - \Gamma_{m_1, m}^{B_1^n} \right) \rho_{X^n(m_1, m)}^{B_1} \right\} \right\}. \end{aligned}$$

Due to the above exchange between the expectation and the average and the symmetry of the code construction (each codeword is selected randomly and independently), it suffices to analyze the expectation of the average error probability for the first message pair $(m_1 = 1, m = 1)$, i.e., the last line above is equal to $\mathbb{E}_{x^n; W^n} \left\{ \text{Tr} \left\{ \left(I - \Gamma_{1,1}^{B_1^n} \right) \rho_{X^n(1,1)}^{B_1} \right\} \right\}$. Using the Hayashi-Nagaoka operator inequality (Lemma 3.1 on page 34), we obtain

the following upper bound on this term:

$$\begin{aligned} \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left[\left(I - \Gamma_{1,1}^{B_1^n} \right) \rho_{X^n(1,1)}^{B_1} \right] \right\} &\leq 2 \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left\{ \left(I - P_{1,1} \right) \rho_{X^n(1,1)}^{B_1} \right\} \right\} \\ &+ 4 \sum_{\substack{X^n, W^n \\ (m_1, m) \neq (1,1)}} \mathbb{E} \left\{ \text{Tr} \left\{ P_{m_1, m} \rho_{X^n(1,1)}^{B_1} \right\} \right\}. \end{aligned} \quad (6.14)$$

We begin by bounding the term in the first line above. Consider the following chain of inequalities:

$$\begin{aligned} \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left\{ \Pi'_{1,1} \rho_{X^n(1,1)}^{B_1} \right\} \right\} &= \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left\{ \Pi \Pi_{W^n(1)} \Pi_{X^n(1,1)} \Pi_{W^n(1)} \Pi \rho_{X^n(1,1)}^{B_1} \right\} \right\} \\ &\geq \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left\{ \Pi_{X^n(1,1)} \rho_{X^n(1,1)}^{B_1} \right\} \right\} \\ &\quad - \mathbb{E}_{X^n, W^n} \left\{ \left\| \rho_{X^n(1,1)}^{B_1} - \Pi \rho_{X^n(1,1)}^{B_1} \Pi \right\|_1 \right\} \\ &\quad - \mathbb{E}_{X^n, W^n} \left\{ \left\| \rho_{X^n(1,1)}^{B_1} - \Pi_{W^n(1)} \rho_{X^n(1,1)}^{B_1} \Pi_{W^n(1)} \right\|_1 \right\} \\ &\geq 1 - \epsilon - 4\sqrt{\epsilon}, \end{aligned}$$

where the first inequality follows from the inequality

$$\text{Tr} \{ \Lambda \rho \} \leq \text{Tr} \{ \Lambda \sigma \} + \|\rho - \sigma\|_1, \quad (6.15)$$

which holds for all ρ, σ , and Λ such that $0 \leq \rho, \sigma, \Lambda \leq I$. The second inequality follows from the *gentle operator lemma for ensembles* (see Lemma 3.2) and the properties of typical projectors for sufficiently large n .

We now focus on bounding the second term of (6.14). We can expand this term as follows:

$$\begin{aligned} &\sum_{\substack{X^n, W^n \\ (m_1, m) \neq (1,1)}} \mathbb{E} \left\{ \text{Tr} \left\{ P_{m_1, m} \rho_{X^n(1,1)}^{B_1} \right\} \right\} \\ &= \sum_{m_1 \neq 1} \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left\{ P_{m_1, 1} \rho_{X^n(1,1)}^{B_1} \right\} \right\} \end{aligned} \quad (\mathbf{E1})$$

$$+ \sum_{\substack{m_1, \\ m \neq 1}} \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left\{ P_{m_1, m} \rho_{X^n(1,1)}^{B_1} \right\} \right\}. \quad (\mathbf{E2})$$

We will now compute the expectation of the first the term, **(E1)**, with respect to

the code randomness:

$$\begin{aligned}
 \mathbb{E}_{X^n, W^n} \{(\mathbf{E1})\} &= \sum_{m_1 \neq 1} \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left\{ P_{m_1, 1} \rho_{X^n(1,1)}^{B_1} \right\} \right\} \\
 &= \sum_{m_1 \neq 1} \mathbb{E}_{X^n, W^n} \text{Tr} \left\{ \Pi \Pi_{W^n(1)} \Pi_{X^n(m_1,1)} \Pi_{W^n(1)} \Pi \rho_{X^n(1,1)}^{B_1} \right\} \\
 &\leq 2^{n[H(B_1|WX)+\delta]} \sum_{m_1 \neq 1} \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left[\Pi \Pi_{W^n(1)} \rho_{X^n(m_1,1)} \Pi_{W^n(1)} \Pi \rho_{X^n(1,1)}^{B_1} \right] \right\} \\
 &= 2^{n[H(B_1|WX)+\delta]} \sum_{m_1 \neq 1} \mathbb{E}_{W^n} \left\{ \text{Tr} \left[\Pi_{W^n(1)} \mathbb{E}_{X^n|W^n} \left\{ \rho_{X^n(m_1,1)} \right\} \Pi_{W^n(1)} \right. \right. \\
 &\quad \left. \left. \Pi \mathbb{E}_{X^n|W^n} \left\{ \rho_{X^n(1,1)}^{B_1} \right\} \Pi \right] \right\} \\
 &= 2^{n[H(B_1|WX)+\delta]} \sum_{m_1 \neq 1} \mathbb{E}_{W^n} \left\{ \text{Tr} \left\{ \Pi \Pi_{W^n(1)} \sigma_{W^n(1)} \Pi_{W^n(1)} \Pi \sigma_{W^n(1)} \right\} \right\} \\
 &\leq 2^{n[H(B_1|WX)+\delta]} 2^{-n[H(B_1|W)-\delta]} \sum_{m_1 \neq 1} \mathbb{E}_{W^n} \left\{ \text{Tr} \left\{ \Pi \Pi_{W^n(1)} \Pi \sigma_{W^n(1)} \right\} \right\} \\
 &\leq 2^{n[H(B_1|WX)+\delta]} 2^{-n[H(B_1|W)-\delta]} \sum_{m_1 \neq 1} \mathbb{E}_{W^n} \left\{ \text{Tr} \left\{ \sigma_{W^n(1)} \right\} \right\} \\
 &\leq 2^{-n[I(X;B_1|W)-2\delta]} |\mathcal{M}_1|,
 \end{aligned}$$

The first inequality is due to the *projector trick* inequality which states that:

$$\Pi_{X^n(m_1,1)} \leq 2^{n[H(B_1|WX)+\delta]} \rho_{X^n(m_1,1)}^{B_1}. \quad (6.16)$$

The second inequality follows from the properties of typical projectors:

$$\Pi_{W^n(1)} \sigma_{W^n(1)} \Pi_{W^n(1)} \leq 2^{-n[H(B_1|W)-\delta]} \Pi_{W^n(1)}. \quad (6.17)$$

We now consider the expectation of the second term (**E2**) with respect to the random choice of codebook.

$$\begin{aligned}
 \mathbb{E}_{X^n, W^n} \{(\mathbf{E2})\} &= \sum_{\substack{m_1, \\ m \neq 1}} \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left\{ P_{m_1, m} \rho_{X^n(1,1)}^{B_1} \right\} \right\} \\
 &= \sum_{\substack{m_1, \\ m \neq 1}} \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left[\Pi \Pi_{W^n(m)} \Pi_{X^n(m_1, m)} \Pi_{W^n(m)} \Pi \rho_{X^n(1,1)}^{B_1} \right] \right\}
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{\substack{m_1, \\ m \neq 1}} \text{Tr} \left[\mathbb{E}_{X^n, W^n} \left\{ \Pi_{W^n(m)} \Pi_{X^n(m_1, m)} \Pi_{W^n(m)} \right\} \right. \\
 &\quad \left. \Pi \mathbb{E}_{X^n, W^n} \left\{ \rho_{X^n(1,1)}^{B_1} \right\} \Pi \right] \\
 &= \sum_{\substack{m_1, \\ m \neq 1}} \text{Tr} \left\{ \mathbb{E}_{X^n, W^n} \left\{ \Pi_{W^n(m)} \Pi_{X^n(m_1, m)} \Pi_{W^n(m)} \right\} \Pi \bar{\rho}^{\otimes n} \Pi \right\} \\
 &\leq 2^{-n[H(B_1) - \delta]} \sum_{\substack{m_1, \\ m \neq 1}} \text{Tr} \left[\mathbb{E}_{X^n, W^n} \left\{ \Pi_{W^n(m)} \Pi_{X^n(m_1, m)} \Pi_{W^n(m)} \right\} \Pi \right] \\
 &= 2^{-n[H(B_1) - \delta]} \sum_{\substack{m_1, \\ m \neq 1}} \mathbb{E}_{X^n, W^n} \text{Tr} \left[\Pi_{X^n(m_1, m)} \Pi_{W^n(m)} \Pi \Pi_{W^n(m)} \right] \\
 &\leq 2^{-n[H(B_1) - \delta]} \sum_{m \neq 1, m_1} \mathbb{E}_{X^n, W^n} \left\{ \text{Tr} \left\{ \Pi_{X^n(m_1, m)} \right\} \right\} \\
 &\leq 2^{-n[H(B_1) - \delta]} 2^{n[H(B_1|WX) + \delta]} |\mathcal{M}_1| |\mathcal{M}_2| \\
 &= 2^{-n[I(WX; B_1) - 2\delta]} |\mathcal{M}_1| |\mathcal{M}_2| \\
 &= 2^{-n[I(X; B_1) - 2\delta]} |\mathcal{M}_1| |\mathcal{M}_2|.
 \end{aligned}$$

The equality $I(WX; B_1) = I(X; B_1)$ follows from the way the codebook is constructed (the quantum Markov chain $W - X - B$). This completes the error analysis for the first receiver.

Error analysis for Receiver 2. The proof for the second receiver is analogous to the point-to-point HSW theorem. The following bound holds for the expectation of the average error probability for the second receiver if n is sufficiently large:

$$\begin{aligned}
 &\mathbb{E}_{X^n, W^n} \left\{ \frac{1}{|\mathcal{M}_2|} \sum_m \text{Tr} \left\{ (I - \Lambda_m^{B_2^n}) \rho_{X^n(m_1, m)}^{B_2^n} \right\} \right\} \\
 &= \mathbb{E}_{W^n} \left\{ \frac{1}{|\mathcal{M}_2|} \sum_m \text{Tr} \left\{ (I - \Lambda_m^{B_2^n}) \mathbb{E}_{X^n|W^n} \left\{ \rho_{X^n(m_1, m)}^{B_2^n} \right\} \right\} \right\} \\
 &= \mathbb{E}_{W^n} \left\{ \frac{1}{|\mathcal{M}_2|} \sum_m \text{Tr} \left\{ (I - \Lambda_m^{B_2^n}) \sigma_{W^n(m)}^{B_2^n} \right\} \right\} \\
 &\leq 2(\epsilon + 2\sqrt{\epsilon}) + 4 \left[2^{-n[I(W; B_2) - 2\delta]} |\mathcal{M}_2| \right].
 \end{aligned}$$

Putting everything together, the joint POVM performed by both receivers is of

6.3 Marton coding scheme

the form: $\Gamma_{m_1, m}^{B_1^n} \otimes \Lambda_{m'}^{B_2^n}$, and the expectation of the average error probability for both receivers is bounded from above as

$$\begin{aligned}
& \mathbb{E}_{x^n, w^n} \frac{1}{|\mathcal{M}_1| |\mathcal{M}_2|} \sum_{m_1, m} \text{Tr} \left\{ (I - \Gamma_{m_1, m}^{B_1^n} \otimes \Lambda_m^{B_2^n}) \rho_{X^n(m_1, m)}^{B_1^n B_2^n} \right\} \\
& \leq \mathbb{E}_{x^n, w^n} \left\{ \frac{1}{|\mathcal{M}_1| |\mathcal{M}_2|} \sum_{m_1, m} \text{Tr} \left\{ (I - \Gamma_{m_1, m}^{B_1^n}) \rho_{X^n(m_1, m)}^{B_1^n} \right\} \right\} \\
& \quad + \mathbb{E}_{x^n, w^n} \left\{ \frac{1}{|\mathcal{M}_1| |\mathcal{M}_2|} \sum_{m_1, m} \text{Tr} \left\{ (I - \Lambda_m^{B_2^n}) \rho_{X^n(m_1, m)}^{B_2^n} \right\} \right\} \\
& \leq 4\epsilon + 12\sqrt{\epsilon} + 4 \left[2^{-n[I(W; B_2) - 2\delta]} |\mathcal{M}_2| \right] \\
& \quad 4 \left[2^{-n[I(X; B_1|W) - 2\delta]} |\mathcal{M}_1| + 2^{-n[I(X; B_1) - 2\delta]} |\mathcal{M}_1| |\mathcal{M}_2| \right],
\end{aligned}$$

where the first inequality uses the operator union bound from Lemma 5.1:

$$I^{B_1^n B_2^n} - \Gamma_{m_1, m}^{B_1^n} \otimes \Lambda_m^{B_2^n} \leq (I^{B_1^n B_2^n} - \Gamma_{m_1, m}^{B_1^n} \otimes I^{B_2^n}) + (I^{B_1^n B_2^n} - I^{B_1^n} \otimes \Lambda_m^{B_2^n}).$$

Thus, as long as the sender chooses the message sizes $|\mathcal{M}_1|$ and $|\mathcal{M}_2|$ such that $|\mathcal{M}_1| \leq 2^{n[I(X; B_1|W) - 3\delta]}$, $|\mathcal{M}_2| \leq 2^{n[I(W; B_2) - 3\delta]}$, and $|\mathcal{M}_1| |\mathcal{M}_2| \leq 2^{n[I(X; B_1) - 3\delta]}$, then there exists a particular code with asymptotically vanishing average error probability in the large n limit. \square

Taking the union over all possible choices of input distribution $p_{WX}(w, x)$ gives us the superposition coding inner bound: $\mathcal{R}_{\text{SC}}(\mathcal{N}) \equiv \bigcup_{p_{WX}} \mathcal{R}_{\text{SC}}(\mathcal{N}, p_{WX})$.

6.3 Marton coding scheme

We now prove that the Marton inner bound is achievable for quantum broadcast channels. The Marton scheme depends on auxiliary random variables U_1 and U_2 , *binning*, and the properties of strongly² typical sequences and projectors.

² The notion of *strong* typicality or *frequency* typicality differs from the entropy typicality we have used until now. See [Wil11, Section 14.2.3].

Theorem 6.2 (Marton inner bound). *Let $\{\rho_x^{B_1 B_2}\}$ be a classical-quantum broadcast channel and let $x = f(u_1, u_2)$ be a deterministic function. The following rate region is achievable:*

$$\begin{aligned} R_1 &\leq I(U_1; B_1)_\theta, \\ R_2 &\leq I(U_2; B_2)_\theta, \\ R_1 + R_2 &\leq I(U_1; B_1)_\theta + I(U_2; B_2)_\theta - I(U_1; U_2)_\theta, \end{aligned} \quad (6.18)$$

where the information quantities are with respect to the state:

$$\theta^{U_1 U_2 B_1 B_2} = \sum_{u_1, u_2} p(u_1, u_2) |u_1\rangle\langle u_1|^{U_1} \otimes |u_2\rangle\langle u_2|^{U_2} \otimes \rho_{f(u_1, u_2)}^{B_1 B_2}.$$

The coding scheme in Theorem 6.2 is a broadcast channel code with no common message: $(n, R_1, 0, R_2, \epsilon)$. The information processing task is described by:

$$n \cdot \mathcal{N}^{X \rightarrow B_1 B_2} \xrightarrow{(1-\epsilon)} nR_1 \cdot [c \rightarrow c^1] + nR_2 \cdot [c \rightarrow c^2]. \quad (6.19)$$

Proof. Consider the classical-quantum broadcast channel $\{\mathcal{N}(x) \equiv \rho_x^{B_1 B_2}\}$, and a deterministic mixing function: $f : \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow \mathcal{X}$. Using the mixing function as a pre-coder to the broadcast channel \mathcal{N} , we obtain a channel \mathcal{N}' defined as:

$$\mathcal{N}'(u_1, u_2) \equiv \rho_{f(u_1, u_2)}^{B_1 B_2} \equiv \rho_{u_1, u_2}^{B_1 B_2}. \quad (6.20)$$

Codebook construction. Define two auxiliary indices $\ell_1 \in [1 : L_1]$, $L_1 = 2^{n[I(U_1; B_1) - \delta]}$ and $\ell_2 \in [1 : L_2]$, $L_2 = 2^{n[I(U_2; B_2) - \delta]}$. For each ℓ_1 generate an i.i.d. random sequence $u_1^n(\ell_1)$ according to $p_{U_1^n}(u_1^n)$. Similarly we choose L_2 random i.i.d. sequences $u_2^n(\ell_2)$ according to $p_{U_2^n}(u_2^n)$. Partition the sequences $u_1^n(\ell_1)$ into 2^{nR_1} different bins B_{m_1} . Similarly, partition the sequences $u_2^n(\ell_2)$ into 2^{nR_2} bins C_{m_2} . For each message pair (m_1, m_2) , the sender selects a sequence $(u_1^n(\ell_1), u_2^n(\ell_2)) \in (B_{m_1} \times C_{m_2}) \cap \mathcal{A}_{p_{U_1 U_2}, \delta}^n$, such that each sequence is taken from the appropriate bin and the sender demands that they are strongly jointly typical and otherwise declares failure. The codebook $x^n(m_1, m_2)$ is deterministically constructed from $(u_1^n(\ell_1), u_2^n(\ell_2))$ by applying the function $x_i = f(u_{1i}, u_{2i})$.

Transmission. Let (ℓ_1, ℓ_2) denote the pair of indices of the joint sequence $(u_1^n(\ell_1), u_2^n(\ell_2))$ which was chosen as the codeword for message (m_1, m_2) . Expressed in terms of these

indices the output of the channel is

$$\rho_{u_1^n(\ell_1), u_2^n(\ell_2)}^{B_1^n B_2^n} = \bigotimes_{i \in [n]} \rho_{f(u_{1i}(\ell_1), u_{2i}(\ell_2))}^{B_1 B_2} \equiv \rho_{\ell_1, \ell_2}. \quad (6.21)$$

Define the following average states for Receiver 1:

$$\omega_{u_1}^{B_1} \equiv \sum_{u_2} p_{U_2|U_1}(u_2|u_1) \rho_{u_1, u_2}^{B_1}, \quad \bar{\rho} \equiv \sum_{u_1} p(u_1) \omega_{u_1}^{B_1}. \quad (6.22)$$

Decoding. The detection POVM for Receiver 1, $\{\Lambda_{\ell_1}\}_{\ell_1 \in [1, \dots, L_1]}$, is constructed by using the square-root measurement as in (3.12) based on the following combination of strongly typical projectors:

$$\Pi'_{\ell_1} \equiv \Pi_{\bar{\rho}, \delta}^n \Pi_{u_1^n(\ell_1)} \Pi_{\bar{\rho}, \delta}^n. \quad (6.23)$$

The outcome of the measurement will be denoted L'_1 . The projectors $\Pi_{u_1^n(\ell_1)}$ and $\Pi_{\bar{\rho}, \delta}^n$ are defined with respect to the states $\omega_{u_1^n(\ell_1)}$ and $\bar{\rho}^{\otimes n}$ given in (6.22). Note that we use *strongly* typical projectors in this case as defined in [Wil11, Section 14.2.3]. Knowing ℓ_1 and the binning scheme, Receiver 1 can deduce the message m_1 from the bin index. Receiver 2 uses a similar decoding strategy to obtain ℓ_2 and infer m_2 .

Error analysis. An error occurs if one (or more) of the following events occurs.

- (E0): An encoding error occurs whenever there is no jointly typical sequence in $B_{m_1} \times C_{m_2}$ for some message pair (m_1, m_2) .
- (E1): A decoding error occurs at Receiver 1 if $L'_1 \neq \ell_1$.
- (E2): A decoding error occurs at Receiver 2 if $L'_2 \neq \ell_2$.

The probability of an encoding error (E0) is bounded like in the classical Marton scheme [Mar79, EGK10, Cov98]. To see this, we use Cover's counting argument [Cov98]. The probability that two random sequences u_1^n, u_2^n chosen according to the marginals are jointly typical is $2^{-nI(U_1; U_2)}$ and since there are $2^{n[I(U_1; B_1) - R_1]}$ and $2^{n[I(U_2; B_2) - R_2]}$ sequences in each bin, the expected number of jointly typical sequences that can be constructed from each combination of bins is

$$2^{n[I(U_1; B_1) - R_1]} 2^{n[I(U_2; B_2) - R_2]} 2^{-nI(U_1; U_2)}. \quad (6.24)$$

Thus, if we choose $R_1 + R_2 + \delta \leq I(U_1; B_1) + I(U_2; B_2) - I(U_1; U_2)$, then the expected number of strongly jointly typical sequences in $B_{m_1} \times C_{m_2}$ is much larger than one.

To bound the probability of error event (**E1**), we use the Hayashi-Nagaoka operator inequality (Lemma 3.1):

$$\begin{aligned} \Pr(\mathbf{E1}) &= \frac{1}{L_1} \sum_{\ell_1} \text{Tr}[(I - \Lambda_{\ell_1})\rho_{\ell_1, \ell_2}] \\ &\leq \frac{1}{L_1} \sum_{\ell_1} \left(\underbrace{2 \text{Tr}[(I - \Pi_{\bar{\rho}, \delta}^n \Pi_{u_1^n(\ell_1)} \Pi_{\bar{\rho}, \delta}^n)\rho_{\ell_1, \ell_2}]}_{(T1)} \right. \\ &\quad \left. + 4 \underbrace{\sum_{\ell'_1 \neq \ell_1} \text{Tr}[\Pi_{\bar{\rho}, \delta}^n \Pi_{u_1^n(\ell'_1)} \Pi_{\bar{\rho}, \delta}^n \rho_{\ell_1, \ell_2}]}_{(T2)} \right). \end{aligned}$$

Consider the following lemma [Wil11, Property 14.2.7].

Lemma 6.1. *When $u_1^n(\ell_1)$ and $u_2^n(\ell_2)$ are strongly jointly typical, the state ρ_{ℓ_1, ℓ_2} is well supported by both the averaged and conditionally typical projector in the sense that: $\text{Tr}[\Pi_{\bar{\rho}, \delta}^n \rho_{\ell_1, \ell_2}] \geq 1 - \epsilon$, $\forall \ell_1, \ell_2$, and $\text{Tr}[\Pi_{u_1^n(\ell_1)} \rho_{\ell_1, \ell_2}] \geq 1 - \epsilon$, $\forall \ell_2$,*

To bound the first term (T1), we use the following argument:

$$\begin{aligned} 1 - (T1) &= \text{Tr}[\Pi_{\bar{\rho}, \delta}^n \Pi_{u_1^n(\ell_1)} \Pi_{\bar{\rho}, \delta}^n \rho_{\ell_1, \ell_2}] \\ &= \text{Tr}[\Pi_{u_1^n(\ell_1)} \Pi_{\bar{\rho}, \delta}^n \rho_{\ell_1, \ell_2} \Pi_{\bar{\rho}, \delta}^n] \\ &\geq \text{Tr}[\Pi_{u_1^n(\ell_1)} \rho_{\ell_1, \ell_2}] - \|\Pi_{\bar{\rho}, \delta}^n \rho_{\ell_1, \ell_2} \Pi_{\bar{\rho}, \delta}^n - \rho_{\ell_1, \ell_2}\|_1 \\ &\geq (1 - \epsilon) - 2\sqrt{\epsilon}, \end{aligned} \tag{6.25}$$

where the inequalities follow from (6.15) and Lemma 6.1. This use of Lemma 6.1 demonstrates why the Marton coding scheme selects the sequences $u_1^n(\ell_1)$ and $u_2^n(\ell_2)$ such that they are strongly jointly typical.

To bound the second term, we begin by applying a variant of the projector trick from (6.16). For what follows, note that the expectation \mathbb{E}_{U_1, U_2} over the random code

is with respect to the product distribution $p_{U_1^n}(u_1^n)p_{U_2^n}(u_2^n)$:

$$\begin{aligned} \mathbb{E}_{U_1, U_2} \{(T2)\} &= \mathbb{E}_{U_1, U_2} \left\{ \sum_{\ell'_1 \neq \ell_1} \text{Tr} [\Pi_{\bar{\rho}, \delta}^n \Pi_{U_1^n}(\ell'_1) \Pi_{\bar{\rho}, \delta}^n \rho_{\ell_1, \ell_2}] \right\} \\ &\leq 2^{n[H(B_1|U_1)+\delta]} \mathbb{E}_{U_1, U_2} \left\{ \sum_{\ell'_1 \neq \ell_1} \text{Tr} [\Pi_{\bar{\rho}, \delta}^n \omega_{\ell'_1} \Pi_{\bar{\rho}, \delta}^n \rho_{\ell_1, \ell_2}] \right\}. \end{aligned}$$

We continue the proof using averaging over the choice of codebook and the properties of typical projectors:

$$\begin{aligned} &= 2^{n[H(B_1|U_1)+\delta]} \mathbb{E}_{U_2} \sum_{\ell'_1 \neq \ell_1} \text{Tr} \left[\Pi_{\bar{\rho}, \delta}^n \mathbb{E}_{U_1} \{\omega_{\ell'_1}\} \Pi_{\bar{\rho}, \delta}^n \mathbb{E}_{U_1} \{\rho_{\ell_1, \ell_2}\} \right] \\ &= 2^{n[H(B_1|U_1)+\delta]} \mathbb{E}_{U_2} \sum_{\ell'_1 \neq \ell_1} \text{Tr} \left[\Pi_{\bar{\rho}, \delta}^n \bar{\rho} \Pi_{\bar{\rho}, \delta}^n \mathbb{E}_{U_1} \{\rho_{\ell_1, \ell_2}\} \right] \\ &\leq 2^{n[H(B_1|U_1)+\delta]} 2^{-n[H(B_1)-\delta]} \mathbb{E}_{U_1, U_2} \sum_{\ell'_1 \neq \ell_1} \text{Tr} [\Pi_{\bar{\rho}, \delta}^n \rho_{\ell_1, \ell_2}] \\ &\leq 2^{n[H(B_1|U_1)+\delta]} 2^{-n[H(B_1)-\delta]} \mathbb{E}_{U_1, U_2} \sum_{\ell'_1 \neq \ell_1} 1 \\ &\leq |\mathcal{L}_1| 2^{-n[I(U_1; B_1)-2\delta]}. \end{aligned}$$

Therefore, if we choose $2^{nR_1} = |\mathcal{L}_1| \leq 2^{n[I(U_1; B_1)-3\delta]}$, the probability of error will go to zero in the asymptotic limit of many channel uses. The analysis of the event **(E2)** is similar. \square

6.4 Discussion

We established two achievable rate regions for the classical-quantum broadcast channel. In each case a fundamentally different coding strategy was used.

The *superposition coding* strategy is a very powerful coding technique for encoding two “layers” of messages in the same codeword. Recall that the codebooks in the Chong-Motani-Garg coding strategy were also constructed using the superposition coding technique. In the next chapter, we will use this technique to build codes for the relay channel.

The *binning* strategy used in the Marton scheme is also applicable more widely. It can be used every time two uncorrelated messages must be encoded into a single codeword. From the point of view of Receiver 1, the messages intended for Receiver 2 are seen as random noise. By using the correlated variables $(U_1, U_2) \sim p(u_1, u_2)$ to construct the codebooks we can obtain better rates than would be possible if independent codebooks were used. This is because the “noise” codewords are now correlated with the messages for Receiver 1 and thus helping him with the communication task.

Note that the above two techniques can be combined to give the quantum Marton coding scheme with a common message [Tak12].

Chapter 7

Relay channels

Suppose that a source wishes to communicate with a remote destination and that a relay station is available which can decode the messages transmitted by the source during one time slot and *forward* them to the destination during the next time slot. With the relay's help, the source and the destination can improve communication rates because the destination can decode the intended messages in parallel from the channel outputs during two consecutive time slots. In this way, useful information is received both from the source and the relay.

The discrete memoryless relay channel is a probabilistic model for a communication scenario with a *source*, a *destination* and a cooperative *relay* station. The channel is modelled as a two-input two-output conditional probability distribution

$$p(y_1, y|x, x_1), \quad (7.1)$$

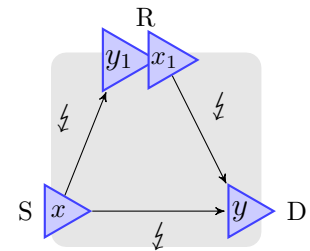


Figure 7.1: The classical relay channel.

where x is the input of the source, y_1 and x_1 are the received symbol and transmitted symbol of the relay, and y is the output at the destination. This relay channel model is very general and contains many of the other ideas presented in this thesis. The transmission of the source towards the relay and the destination is a kind of broadcast channel, whereas the decoding at the destination is an instance of the multiple access channel. These correspondences can inform our choice of coding strategies, but in order to take full advantage of the communication network we must build a *relay channel code* which aims to achieve the best overall rate from the source to the destination.

In this chapter, we will review some of the coding strategies for the classical relay channel and then show that the *partial decode-and-forward* strategy can be applied to the classical-quantum relay channel. Note that we depart from the usual naming conventions for senders and receivers. We do so because both the source and the relay act as senders in our scenario, so more specific identifiers are necessary.

7.1 Introduction

Consider two villages located in a valley that wish to establish a communication link between them using a direct link and also with the help of a radio tower on a nearby mountain peak. We can setup a relay station on the tower, which decodes the messages from the source village and retransmits them towards the destination village. Assuming the villagers only have access to point-to-point communication technologies, they now have two obvious options. Either they send information on the *direct transmission* link, or they use *full relaying*, where all their communication happens via the tower. In the first case, the tower is not used at all and in the second case the direct link is not used at all.

It is worthwhile to examine the exact timing associated with the information flow in the latter scenario, since it is the first appearance of a multi-hop communication protocol. Let us assume that the source wants to send the string “constitution” to the destination. Assume that we use codewords of size n , and that each character is encoded in a separate codeword. The source and the relay have transmit codebooks $\{X_s^n(a)\}, \{X_r^n(a)\}, a \in \mathcal{ASCII}$.

The *direct transmission* strategy will make $12n$ uses of the channel. The transmissions of the source will be $[X_s^n(\mathbf{c}), X_s^n(\mathbf{o}), X_s^n(\mathbf{n}), \dots, X_s^n(\mathbf{n})]$ in each block. The relay will transmit a fixed codeword during this time. The destination will simply use a point-to-point decoder to extract the messages. The rate achievable using this strategy is given by:

$$R \leq \sup_{p(x), x_1} I(X; Y | X_1 = x_1). \quad (7.2)$$

The *full relaying* strategy will use the channel $13n$ times, where the need for an extra block of transmission is introduced by the decoding delay at the relay. During the 13 blocks, the transmissions of the source will be $[X_s^n(\mathbf{c}), X_s^n(\mathbf{o}), X_s^n(\mathbf{n}), \dots, X_s^n(\mathbf{n}), \emptyset]$,

whereas the transmissions of the relay are one block behind: $[\emptyset, X_r^n(\mathbf{c}), X_r^n(\mathbf{o}), \dots, X_r^n(\mathbf{o}), X_r^n(\mathbf{n})]$. The source simply has no more messages to send during block 13, whereas the relay has no information to forward during the first block, so both parties will stay silent during these different times. The rates that are achievable by this approach are:

$$R \leq \sup_{p(x), p(x_1)} \min\{I(X_1; Y), I(X; Y_1 | X_1)\}. \quad (7.3)$$

This corresponds to the minimum of the point-to-point capacities of the two legs of the transmission. Note that the second mutual information term is conditional on X_1 , since the relay knows its own transmit signal.

Surely a better strategy must exist than the ones described above. How can we use both the direct link and the relayed link at the same time?

7.1.1 Classical relay channel coding strategies

Two important families of coding strategies exist for relay channels: *compress and forward* and *decode and forward* [CEG79, EGK10].

In compress-and-forward strategies, the relay does not try to decode the message from his received signal Y_1^n , but simply searches for a close sequence \hat{Y}_1^n chosen from a predetermined compression codebook. To continue the example from the previous section, suppose that the relay's decoding simply tries to determine whether the transmitted message is a vowel or a consonant. This partial information about the message is then forwarded to the destination during the next block, encoded into a codeword $x_1^n(\mathbf{s})$, $\mathbf{s} \in \{\text{consonant}, \text{vowel}\}$ to serve as side-information for the decoding at the destination.

Compress and forward strategies are appropriate in situations where the direct link between the source and the destination is stronger than the link from the source to the relay. In such a situation it would be disadvantageous to require that the messages from the source be fully decoded by the relay. Still, if the relay decodes *something* and forwards this information to the destination, better rates are achievable than if we simply chose to not use the relay as in the direct coding approach [EGK10].

In a decode-and-forward strategy, each of the transmitted messages is decoded by the relay and retransmitted during the next block. Using this strategy, the destination

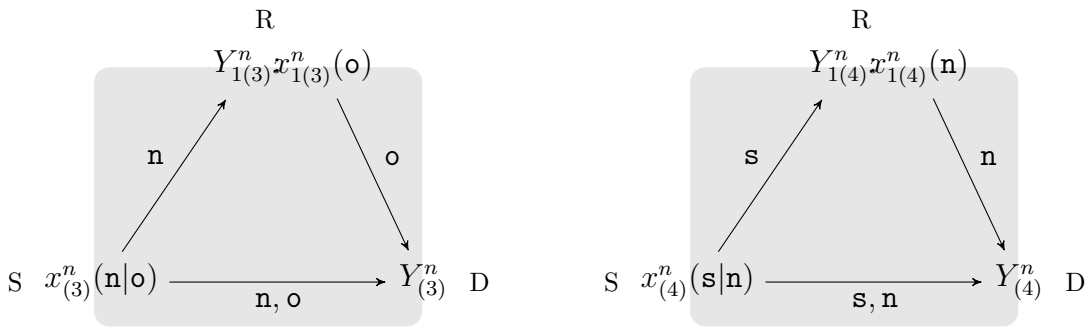
can decode useful information both from the source and the relay. In this way we could achieve the maximum possible throughput to the destination $I(X, X_1; Y)$.

There are at least three decoding strategies that can be used by the destination: backwards decoding, sequential decoding with binning at the relay, or collective decoding of consecutive output blocks of the channel (joint decoding). All three decoding techniques for the decode-and-forward strategy achieve the same rate:

$$R \leq \max_{p(x, x_1)} \min\{ I(X, X_1; Y), I(X; Y_1|X_1) \}. \tag{7.4}$$

We will focus on the collective decoding strategy.

To illustrate the collective decoding strategy let us consider again the situation in which the source village is transmitting the string “constitution” to the destination village. The transmission will take 13 block-uses of the channel. Figure 7.2 illustrates the flow of information for the character **n** which happens during the third and fourth block-uses of the channel. During the third and the fourth transmission blocks, the destination has collected the output variables $(Y_{(3)}^n, Y_{(4)}^n)$ and will perform a decoding operation on both outputs collectively. The rate $I(X, X_1; Y)$ is obtained from the decomposition $I(X, X_1; Y) = I(X; Y|X_1) + I(X_1; Y)$, where the second term will come from the probability of making a mistake when decoding $x_{1(4)}^n(\mathbf{n})$ from $Y_{(4)}^n$ and the first term comes from the probability of wrongly decoding $x_{(3)}^n(\mathbf{n})$ from $Y_{(3)}^n$.



(a) During block 3, the relay will transmit its codeword “o”, which we assume was received in the previous block. The source transmits a codeword $x^n(\mathbf{n}|\mathbf{o})$ which is chosen from a *coherent* codebook.

(b) During block 4, the relay will transmit its codeword for “n”, which we assume was received in the previous block. The source transmits a codeword $x^n(\mathbf{s}|\mathbf{n})$.

Figure 7.2: Information flow in the relay network during the third and fourth transmission blocks of the string “constitution”.

Observe that the optimization in (7.4) is taken over all joint input distributions $p_{X X_1}(x, x_1)$, which would seem to contradict the assumption that the source and the relay are different parties and cannot synchronize their encoding. Recall that in the multiple access channel problem, the assumption that the senders act independently translated to the optimization over all product distributions $p_{X_1}(x_1)p_{X_2}(x_2)$ in (4.6).

The change from $p_X(x)p_{X_1}(x_1)$ to $p_{X X_1}(x, x_1)$ is allowed because the source uses a *coherent* codebook. The codewords for the relay are chosen according to $p_{X_1}(x_1)$, whereas the codewords for the sender are chosen according to $p_{X|X_1}(x|x_1)$ conditional on the codeword of the relay. But how could the source possibly know what the relay will be transmitting during each time instant? No telepathic abilities are necessary — only optimism. The source knows what the relay will be transmitting because, if the protocol is working, it should be the codeword from the previous block.

The *partial* decode-and-forward strategy differs from the decode-and-forward strategy in that it requires the relay to decode only *part of* the message from the source [CEG79]. The idea is similar to the *partial* interference cancellation strategy used by Han and Kobayashi for the interference channel [HK81], which is its contemporary.

7.1.2 Quantum relay channels

A classical-quantum relay channel \mathcal{N} is a map with two classical inputs x and x_1 and two output quantum systems B_1 and B . For each pair of possible input symbols $(x, x_1) \in \mathcal{X} \times \mathcal{X}_1$, the channel prepares a density operator $\rho_{x, x_1}^{B_1 B}$ defined on the tensor-product Hilbert space $\mathcal{H}^{B_1} \otimes \mathcal{H}^B$:

$$\rho_{x, x_1}^{B_1 B} \equiv \mathcal{N}^{X X_1 \rightarrow B_1 B}(x, x_1), \quad (7.5)$$

where B_1 is the relay output and B is the destination output.

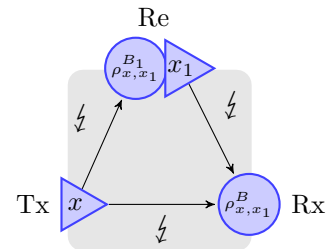


Figure 7.3: The quantum relay channel $\rho_{x, x_1}^{B_1 B}$.

7.1.3 Chapter overview

In this chapter we develop the partial decode-and-forward strategy for classical-quantum relay channels [SWV12]. This *partial* decoding at the relay is a more general strategy

than the *full* decode-and-forward strategy in the same way that the partial interference cancellation strategy for the interference channel (the Han-Kobayashi strategy) was more general than a full interference cancellation strategy.

Our results are the first extension of the quantum simultaneous decoding techniques used in [FHS⁺12, Sen12a] to multi-hop networks. The decoding is based on a novel “sliding-window” quantum measurement (see [Car82, XK05]) which involves a collective measurement on two consecutive blocks of the output in order to extract information from both the Sender and the relay.

The next section will describe the coding strategy in more detail and state our results. The proof is given in Section 7.3.

7.2 Partial decode-and-forward strategy

The idea for the code construction is to use a split codebook strategy where the source decomposes the message set into the Cartesian product of two different sets \mathcal{L} and \mathcal{M} . We can think of the set \mathcal{L} consisting of common messages that both the relay and the destination decode, while the set \mathcal{M} consists of personal messages that only the destination decodes.

In the context of our coding strategy, we analyze the average probability of error at the relay:

$$\bar{p}_e^R \equiv \frac{1}{|\mathcal{L}|} \sum_{\ell_j} \text{Tr} \left\{ \left(I - \Gamma_{\ell_j}^{B_{1(j)}^n} \right) \rho_{\ell_j}^{B_{1(j)}^n} \right\},$$

and the average probability of error at the destination:

$$\bar{p}_e^D \equiv \frac{1}{|\mathcal{M}||\mathcal{L}|} \sum_{m_j, \ell_j} \text{Tr} \left[\left(I - \Lambda_{m_j, \ell_j}^{B_{(j)}^n B_{(j+1)}^n} \right) \rho_{m_j, \ell_j}^{B_{(j)}^n B_{(j+1)}^n} \right]. \quad (7.6)$$

The operators $(I - \Gamma_{\ell_j})$ and $(I - \Lambda_{m_j, \ell_j})$ correspond to the complements of the correct decoding outcomes.

Definition 7.1. An (n, R, ϵ) partial decode-and-forward code for the quantum relay channel consists of two codebooks $\{x^n(m_j, \ell_j)\}_{m_j \in \mathcal{M}, \ell_j \in \mathcal{L}}$ and $\{x_1^n(\ell_j)\}_{\ell_j \in \mathcal{L}}$ and decoding POVMs $\{\Gamma_{\ell_j}\}_{\ell_j \in \mathcal{L}}$ (for the relay) and $\{\Lambda_{m_j, \ell_j}\}_{m_j \in \mathcal{M}, \ell_j \in \mathcal{L}}$ (for the destination), such that

the average probability of error is bounded from above as $\bar{p}_e = \bar{p}_e^R + \bar{p}_e^D \leq \epsilon$.

A rate R is *achievable* if there exists an $(n, R - \delta, \epsilon)$ quantum relay channel code for all $\epsilon, \delta > 0$ and sufficiently large n .

The theorem below captures the main result of this chapter.

Theorem 7.1 (Partial decode-and-forward inner bound). *Let $\{\rho_{x,x_1}\}$ be a cc-qq relay channel as in (7.5). Then a rate R is achievable, provided that the following inequality holds:*

$$R \leq \max_{p(u,x,x_1)} \min \left\{ \begin{array}{l} I(XX_1; B)_\theta, \\ I(U; B_1|X_1)_\theta + I(X; B|X_1U)_\theta \end{array} \right\}, \quad (7.7)$$

where the information quantities are with respect to the classical-quantum state

$$\theta^{UXX_1B_1B} \equiv \sum_{x,u,x_1} p(u,x,x_1) |u\rangle\langle u|^U \otimes |x\rangle\langle x|^X \otimes |x_1\rangle\langle x_1|^{X_1} \otimes \rho_{x,x_1}^{B_1B}. \quad (7.8)$$

Our code construction employs codebooks $\{x_1^n\}$, $\{u^n\}$, and $\{x^n\}$ generated according to the distribution $p(x_1)p(u|x_1)p(x|u, x_1)$. We split the message for each block into two parts $(m, \ell) \in \mathcal{M} \times \mathcal{L}$ such that we have $R = R_m + R_\ell$. The relay fully decodes the message ℓ and re-encodes it directly (without using binning) in the next block. The destination exploits a “sliding-window” decoding strategy [Car82, XK05] by performing a collective measurement on two consecutive blocks. In this approach, the message pair (m_j, ℓ_j) sent during block j is decoded from the outputs of blocks j and $j + 1$, using an “AND-measurement.”

7.3 Achievability proof

We divide the channel uses into many blocks and build codes in a randomized, block-Markov manner within each block. The channel is used for b blocks, each indexed by $j \in \{1, \dots, b\}$. Our error analysis shows that:

- The relay can decode the message ℓ_j during block j .
- The destination can simultaneously decode (m_j, ℓ_j) from a collective measurement on the output systems of blocks j and $j + 1$.

The error analysis at the relay is similar to that of the Holevo-Schumacher-Westmoreland theorem [Hol98, SW97]. The message ℓ_j can be decoded reliably if the rate R_ℓ obeys

the following inequality:

$$R_\ell \leq I(U; B_1 | X_1)_\theta. \quad (7.9)$$

The decoding at the destination is a variant of the quantum simultaneous decoder from Theorem 4.2. To decode the message (m_j, ℓ_j) , the destination performs a “sliding-window” decoder, implemented as an “AND-measurement” on the outputs of blocks j and $j + 1$. This coding technique does not require binning at the relay or backwards decoding at the destination [Car82, XK05].

In this section, we give the details of the coding strategy and analyze the probability of error for the destination and the relay.

Codebook construction. Fix a code distribution $p(u, x, x_1) = p(x_1)p(u|x_1)p(x|x_1, u)$ and independently generate a different codebook for each block j as follows:

- Randomly and independently generate 2^{nR_ℓ} sequences $x_1^n(\ell_{j-1})$, $\ell_{j-1} \in [1 : 2^{nR_\ell}]$, according to $\prod_{i=1}^n p(x_{1i})$.
- For each $x_1^n(\ell_{j-1})$, randomly and independently generate 2^{nR_ℓ} sequences $u^n(\ell_j, \ell_{j-1})$, $\ell_j \in [1 : 2^{nR_\ell}]$ according to $\prod_{i=1}^n p(u_i | x_{1i}(\ell_{j-1}))$.
- For each $x_1^n(\ell_{j-1})$ and each corresponding $u^n(\ell_j, \ell_{j-1})$, randomly and independently generate 2^{nR_m} sequences $x^n(m_j, \ell_j, \ell_{j-1})$, $m_j \in [1 : 2^{nR_m}]$, according to the distribution: $\prod_{i=1}^n p(x_i | x_{1i}(\ell_{j-1}), u_i(\ell_j, \ell_{j-1}))$.

Transmission. The transmission of (m_j, ℓ_j) to the destination happens during blocks j and $j + 1$ as illustrated in Figure 7.4. At the beginning of block j , we assume that the relay has correctly decoded the message ℓ_{j-1} . During block j , the source inputs the new messages m_j and ℓ_j , and the relay forwards the old message ℓ_{j-1} . That is, their inputs to the channel for block j are the codewords $x^n(m_j, \ell_j, \ell_{j-1})$ and $x_1^n(\ell_{j-1})$, leading to the following state at the channel outputs:

$$\rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \equiv \rho_{x^n(m_j, \ell_j, \ell_{j-1}), x_1^n(\ell_{j-1})}^{B_{1(j)}^{n} B_{(j)}^{n}}$$

During block $j + 1$, the source transmits (m_{j+1}, ℓ_{j+1}) given ℓ_j , whereas the relay sends ℓ_j , leading to the state:

$$\rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \equiv \rho_{x^n(m_{j+1}, \ell_{j+1}, \ell_j), x_1^n(\ell_j)}^{B_{1(j+1)}^n B_{(j+1)}^n}.$$

Our shorthand notation is such that the states are identified by the messages that they encode, and the codewords are implicit.

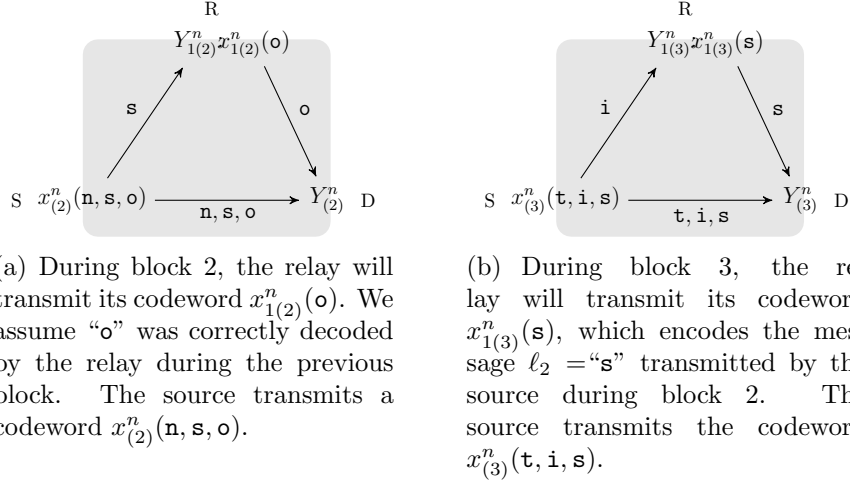


Figure 7.4: Information flow in the relay network during the second and third transmission blocks of the string “co ns ti tu ti on” when using the partial decode-and-forward strategy. The messages for each block (two characters) are encoded by the Sender using a codebook $x^n(m_j, \ell_j, \ell_{j-1})$ during block j . The messages pairs (m_j, ℓ_j) for the seven uses of the channel are: $\{(\text{c}, \text{o}), (\text{n}, \text{s}), (\text{t}, \text{i}), (\text{t}, \text{u}), (\text{t}, \text{i}), (\text{o}, \text{n}), (\emptyset, \emptyset)\}$. The source codebook depends on the current message pair (m_j, ℓ_j) as well as the message ℓ_{j-1} of the previous block, so the transmitted codewords during the seven blocks are: $\{x_{(1)}^n(\text{c}, \text{o}, \emptyset), x_{(2)}^n(\text{n}, \text{s}, \text{o}), x_{(3)}^n(\text{t}, \text{i}, \text{s}), x_{(4)}^n(\text{t}, \text{u}, \text{i}), x_{(5)}^n(\text{t}, \text{i}, \text{u}), x_{(6)}^n(\text{o}, \text{n}, \text{i}), x_{(7)}^n(\emptyset, \emptyset, \text{n})\}$ and $\{x_{1(1)}^n(\emptyset), x_{1(2)}^n(\text{o}), x_{1(3)}^n(\text{s}), x_{1(4)}^n(\text{i}), x_{1(5)}^n(\text{u}), x_{1(6)}^n(\text{i}), x_{1(7)}^n(\text{n})\}$.

7.3.1 Decoding at the destination

We now determine a decoding POVM that the destination can perform on the output systems spanning blocks j and $j + 1$. The destination is trying to recover messages ℓ_j and m_j given knowledge of ℓ_{j-1} .

First let us consider forming decoding operators for block $j + 1$. Consider the state

7.3 Achievability proof

obtained by tracing over the systems X , U , and B_1 in (7.8):

$$\theta^{X_1 B} = \sum_{x_1} p(x_1) |x_1\rangle\langle x_1|^{X_1} \otimes \tau_{x_1}^B,$$

where $\tau_{x_1}^B \equiv \sum_{u,x} p(x|x_1, u) p(u|x_1) \rho_{x,x_1}^B$. Also, let $\bar{\tau}^B$ denote the following state: $\bar{\tau}^B \equiv \sum_{x_1} p(x_1) \tau_{x_1}^B$. Corresponding to the above states are conditionally typical projectors of the following form:

$$\Pi_{\tau_{\ell_j}^{(j+1)}} \equiv \Pi_{\tau_{x_1^n(\ell_j), \delta}^{B_{(j+1)}^n}}, \quad \Pi_{\bar{\tau}^{(j+1)}} \equiv \Pi_{\bar{\tau}^{\otimes n, \delta}^{B_{(j+1)}^n}},$$

which we combine to form the positive operator:

$$P_{\ell_j}^{B_{(j+1)}^n} \equiv \Pi_{\bar{\tau}^{(j+1)}} \Pi_{\tau_{\ell_j}^{(j+1)}} \Pi_{\bar{\tau}^{(j+1)}}, \quad (7.10)$$

that acts on the output systems $B_{(j+1)}^n$ of block $j+1$.

Let us now form decoding operators for block j . Define the conditional typical projector for the state $\rho_{m_j, \ell_j, \ell_{j-1}}^{(j)}$ as

$$\Pi_{\rho_{m_j, \ell_j | \ell_{j-1}}^{(j)}} \equiv \Pi_{\rho_{x_1^n(m_j, \ell_j, \ell_{j-1}), x_1^n(\ell_{j-1}), \delta}^{B_{(j)}^n}}. \quad (7.11)$$

The state obtained from (7.8) by tracing over X and B_1 is

$$\theta^{UX_1 B} = \sum_{u, x_1} p(u|x_1) p(x_1) |u\rangle\langle u|^U \otimes |x_1\rangle\langle x_1|^{X_1} \otimes \bar{\rho}_{u, x_1}^B,$$

where $\bar{\rho}_{u, x_1}^B \equiv \sum_x p(x|x_1, u) \rho_{x, x_1}^B$. We can trace out over U as well to obtain the doubly averaged state $\bar{\bar{\rho}}_{x_1}^B \equiv \sum_{u, x} p(x|x_1, u) p(u|x_1) \rho_{x, x_1}^B$.

The following conditionally typical projectors will be used in the decoding:

$$\Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}^{(j)}} \equiv \Pi_{\bar{\rho}_{u^n(\ell_j, \ell_{j-1}), x_1^n(\ell_{j-1}), \delta}^{B_{(j)}^n}}, \quad \Pi_{\bar{\bar{\rho}}_{\ell_{j-1}}^{(j)}} \equiv \Pi_{\bar{\bar{\rho}}_{x_1^n(\ell_{j-1}), \delta}^{B_{(j)}^n}}.$$

We can then form a positive operator “sandwich”:

$$P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \equiv \Pi_{\bar{\bar{\rho}}_{\ell_{j-1}}^{(j)}} \Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}^{(j)}} \Pi_{\rho_{m_j, \ell_j | \ell_{j-1}}^{(j)}} \Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}^{(j)}} \Pi_{\bar{\bar{\rho}}_{\ell_{j-1}}^{(j)}}. \quad (7.12)$$

Finally, we combine the positive operators from (7.10) and (7.12) to form the “sliding-

window” positive operator:

$$P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} = P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \otimes P_{\ell_j}^{B_{(j+1)}^n}, \quad (7.13)$$

from which we can build the destination’s measurement $\Lambda_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n}$ using the square-root normalization. This measurement is what we call the “AND-measurement.”

Error analysis at the destination. In this section, we prove that the destination can correctly decode the message pair (m_j, ℓ_j) by employing the measurement $\{\Lambda_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n}\}$ on the output state $\rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)}$ spanning blocks j and $j+1$. The average probability of error for the destination is given in (7.6). For now, we consider the error analysis for a single message pair (m_j, ℓ_j) :

$$\begin{aligned} \bar{p}_e^D &\equiv \text{Tr} \left[\left(I - \Lambda_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} \right) \rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)} \right] \\ &\leq 2 \text{Tr} \left\{ \left(I - P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} \right) \rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)} \right\} \end{aligned} \quad (\text{I})$$

$$+ 4 \sum_{(\ell'_j, m'_j) \neq (\ell_j, m_j)} \text{Tr} \left\{ P_{m'_j, \ell'_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} \rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)} \right\}, \quad (\text{II})$$

where we used the Hayashi-Nagaoka inequality (Lemma 3.1) to decompose the error operator $(I - \Lambda_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n})$ into two components: (I) a term related to the probability that the correct detector does not “click”: $(I - P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n})$, and (II) another term related to the probability that a wrong detector “clicks”: $\sum_{(\ell'_j, m'_j) \neq (\ell_j, m_j)} P_{m'_j, \ell'_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n}$. These two errors are analogous to the classical error events in which an output sequence y^n is either not jointly typical with the transmitted codeword or happens to be jointly typical with another codeword.

We will bound the expectation of the average probability of error $\mathbb{E}_{U^n X^n X_1^n} \{\bar{p}_e^D\}$ by bounding the expectation of the average probability for the two error terms: $\mathbb{E}_{U^n X^n X_1^n} \{(\text{I})\}$ and $\mathbb{E}_{U^n X^n X_1^n} \{(\text{II})\}$.

The first term (I) is bounded by using the properties of typical projectors and the operator union bound from Lemma 5.1, which allows us to analyze the errors for the two blocks separately. Because $0 \leq P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \leq I$ and $0 \leq P_{\ell_j}^{B_{(j+1)}^n} \leq I$, we have:

$$\left(I - P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \otimes P_{\ell_j}^{B_{(j+1)}^n} \right) \leq \left(I - P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \right) + \left(I - P_{\ell_j}^{B_{(j+1)}^n} \right). \quad (7.14)$$

7.3 Achievability proof

We use the definition of $P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n}$ from (7.13) and the inequality (7.14) to obtain:

$$\begin{aligned}
& \text{Tr} \left[\left(I - P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} \right) \rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)} \right] \\
&= \text{Tr} \left[\left(I - P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \otimes P_{\ell_j}^{B_{(j+1)}^n} \right) \rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)} \right] \\
&\leq \underbrace{\text{Tr} \left[\left(I - P_{m_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \right) \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \right]}_{\alpha} \underbrace{\text{Tr} \left[\rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right]}_{=1} + \\
&\quad + \underbrace{\text{Tr} \left[\rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \right]}_{=1} \underbrace{\text{Tr} \left[\left(I - P_{\ell_j}^{B_{(j+1)}^n} \right) \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right]}_{\beta},
\end{aligned}$$

where we defined the error terms α and β associated with block j and block $(j+1)$.

We proceed to bound the term β as follows:

$$\begin{aligned}
\beta &= \text{Tr} \left[\left(I - P_{\ell_j}^{B_{(j+1)}^n} \right) \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right] \\
&= \text{Tr} \left[\left(I - \Pi_{\bar{\tau}}^{(j+1)} \Pi_{\tau_{\ell_j}}^{(j+1)} \Pi_{\bar{\tau}}^{(j+1)} \right) \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right] \\
&= 1 - \text{Tr} \left[\Pi_{\bar{\tau}}^{(j+1)} \Pi_{\tau_{\ell_j}}^{(j+1)} \Pi_{\bar{\tau}}^{(j+1)} \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right] \\
&\leq 1 - \text{Tr} \left[\Pi_{\tau_{\ell_j}}^{(j+1)} \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right] \\
&\quad + \left\| \Pi_{\bar{\tau}}^{(j+1)} \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \Pi_{\bar{\tau}}^{(j+1)} - \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right\|_1,
\end{aligned}$$

where the inequality follows from Lemma 2.20. We will analyze the terms labeled α and β separately.

By taking the expectation over the code randomness, we obtain the upper bound:

$$\begin{aligned}
\mathbb{E}_{U^n X^n X_1^n} \{\beta\} &= 1 - \mathbb{E}_{X_1^n} \text{Tr} \left[\Pi_{\tau_{\ell_j}}^{(j+1)} \mathbb{E}_{U^n X^n | X_1^n} \left\{ \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right\} \right] \\
&\quad + \mathbb{E}_{U^n X^n X_1^n} \left\| \Pi_{\bar{\tau}}^{(j+1)} \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \Pi_{\bar{\tau}}^{(j+1)} - \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right\|_1 \\
&\leq 1 - (1 - \epsilon) + 2\sqrt{\epsilon}.
\end{aligned}$$

The inequality follows from $\mathbb{E}_{U^n X^n | X_1^n} \left\{ \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right\} = \tau_{\ell_j}$, the properties of typical projectors: $\mathbb{E}_{X_1^n} \text{Tr}[\Pi_{\tau_{\ell_j}}^{(j+1)} \tau_{\ell_j}] \geq 1 - \epsilon$, $\text{Tr}[\Pi_{\bar{\tau}}^{(j+1)} \bar{\tau}] \geq 1 - \epsilon$ and Lemma 3.2.

The error term α is bounded in a similar fashion.

We can split the sum in the second error term (II) as follows:

$$\begin{aligned}
 & \sum_{(\ell'_j, m'_j) \neq (\ell_j, m_j)} \text{Tr} \left[P_{m'_j, \ell'_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} \rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)} \right] \\
 &= \underbrace{\sum_{m'_j \neq m_j} \text{Tr} \left[P_{m'_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} \rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)} \right]}_{\text{(A)}} \\
 & \quad + \underbrace{\sum_{\ell'_j \neq \ell_j, m'_j} \text{Tr} \left[P_{m'_j, \ell'_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} \rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)} \right]}_{\text{(B)}}.
 \end{aligned}$$

We now analyze the two terms (A) and (B) separately.

Matching ℓ_j , wrong m_j . Assuming ℓ_j is decoded correctly, we show that the message m_j will be decoded correctly provided $R_m < I(X; B|UX_1) = H(B|UX_1) - H(B|UX_1) - \delta$. We will use the following properties of typical projectors:

$$\Pi_{\rho_{m'_j, \ell_j | \ell_{j-1}}}^{(j)} \leq 2^{n[H(B|UX_1) + \delta]} \rho_{m'_j, \ell_j, \ell_{j-1}}^{(j)}, \quad (7.15)$$

$$\Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}}^{(j)} \bar{\rho}_{\ell_j, \ell_{j-1}}^{(j)} \Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}}^{(j)} \leq 2^{-n[H(B|UX_1) - \delta]} \Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}}^{(j)}. \quad (7.16)$$

Consider the first term:

$$\begin{aligned}
 \text{(A)} &= \sum_{m'_j \neq m_j} \text{Tr} \left[P_{m'_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} \rho_{m_j \ell_j \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1} \ell_{j+1} \ell_j}^{(j+1)} \right] \\
 &= \sum_{m'_j \neq m_j} \text{Tr} \left[\left(P_{m'_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \otimes P_{\ell_j}^{B_{(j+1)}^n} \right) \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right] \\
 &\leq \sum_{m'_j \neq m_j} \text{Tr} \left[P_{m'_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \otimes I^{B_{(j+1)}^n} \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right] \\
 &= \sum_{m'_j \neq m_j} \text{Tr} \left[P_{m'_j, \ell_j | \ell_{j-1}}^{B_{(j)}^n} \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \right] \\
 &= \sum_{m'_j \neq m_j} \text{Tr} \left[\underbrace{\Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}}^{(j)} \Pi_{\rho_{m'_j, \ell_j | \ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell_j | \ell_{j-1}}}^{(j)}}_{\text{(2)}} \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \right]_{\text{(1)}}
 \end{aligned}$$

7.3 Achievability proof

We now upper bound expression ① using (7.15) and take the conditional expectation with respect to X^n :

$$\mathbb{E}_{X^n|U^nX_1^n} \left\{ \rho_{m'_j, \ell_j, \ell_{j-1}}^{(j)} \right\} = \bar{\rho}_{\ell_j, \ell_{j-1}}^{(j)},$$

which is independent of the state $\rho_{m'_j, \ell_j, \ell_{j-1}}^{(j)}$ since $m'_j \neq m_j$. The resulting expression in ② has the state $\bar{\rho}_{\ell_j, \ell_{j-1}}^{(j)}$ sandwiched between its typical projector on both sides, and so we can use (7.16). After these steps, we obtain the upper bound:

$$\begin{aligned} \mathbb{E}_{X^n|U^nX_1^n} \{(\text{A})\} &\leq 2^{n[H(B|XUX_1)+\delta]} 2^{-n[H(B|UX_1)-\delta]} \times \\ &\quad \times \mathbb{E}_{X^n|U^nX_1^n} \sum_{m'_j \neq m_j} \text{Tr} \left[\Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell_j|\ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \rho_{m'_j, \ell_j, \ell_{j-1}}^{(j)} \right] \\ &\leq 2^{n[H(B|XUX_1)+\delta]} 2^{-n[H(B|UX_1)-\delta]} \sum_{m'_j \neq m_j} \text{Tr} \left[\rho_{m'_j, \ell_j, \ell_{j-1}}^{(j)} \right] \\ &\leq |\mathcal{M}| 2^{-n[I(X;B|UX_1)-2\delta]}. \end{aligned} \quad (7.17)$$

The second inequality follows because each operator inside the trace is positive semidefinite and less than or equal to the identity.

Wrong ℓ_j (and thus wrong m_j). We obtain the requirement $R \equiv R_\ell + R_m \leq I(XX_1; B) = I(X_1; B) + I(UX; B|X_1)$ from the ‘‘AND-measurement’’ and the following inequalities:

$$\mathbb{E}_{U^nX^nX_1^n} \text{Tr}[\Pi_{\bar{\tau}_{\ell_j}}^{(j+1)}] \leq 2^{n[H(B|X_1)+\delta]}, \quad (7.18)$$

$$\Pi_{\bar{\tau}}^{(j+1)} \bar{\tau} \Pi_{\bar{\tau}}^{(j+1)} \leq 2^{-n[H(B)-\delta]} \Pi_{\bar{\tau}}^{(j+1)}, \quad (7.19)$$

$$\mathbb{E}_{U^nX^nX_1^n} \text{Tr}[\Pi_{\rho_{m_j, \ell_j|\ell_{j-1}}}^{(j)}] \leq 2^{n[H(B|UXX_1)+\delta]}, \quad (7.20)$$

$$\Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \bar{\rho}_{\ell_{j-1}}^{(j)} \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \leq 2^{-n[H(B|X_1)-\delta]} \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)}. \quad (7.21)$$

Consider the following term:

$$\begin{aligned}
(\text{B}) &= \sum_{\ell'_j \neq \ell_j, m'_j} \text{Tr} \left[P_{m'_j, \ell'_j | \ell_{j-1}}^{B_{(j)}^n B_{(j+1)}^n} \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right] \\
&= \sum_{\ell'_j \neq \ell_j, m'_j} \text{Tr} \left[\left(P_{m'_j, \ell'_j | \ell_{j-1}}^{B_{(j)}^n} \otimes P_{\ell'_j}^{B_{(j+1)}^n} \right) \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \otimes \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right] \\
&= \sum_{\ell'_j \neq \ell_j, m'_j} \underbrace{\text{Tr} \left[P_{m'_j, \ell'_j | \ell_{j-1}}^{B_{(j)}^n} \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \right]}_{(\text{B1})} \underbrace{\text{Tr} \left[P_{\ell'_j}^{B_{(j+1)}^n} \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right]}_{(\text{B2})}.
\end{aligned}$$

We want to calculate the expectation of the term (B) with respect to the code randomness $\mathbb{E}_{U^n X^n X_1^n}$. The random variables in different blocks are independent, and so we can analyze the expectations of the factors (B1) and (B2) separately.

Consider first the calculation in block j , which leads to the following bound on the expectation of the factor (B1):

$$\begin{aligned}
\mathbb{E}_{U^n X^n X_1^n} \{(\text{B1})\} &= \mathbb{E}_{U^n X^n X_1^n} \left\{ \text{Tr} \left[P_{m'_j, \ell'_j | \ell_{j-1}}^{B_{(j)}^n} \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \right] \right\} \\
&= \mathbb{E}_{U^n X^n X_1^n} \text{Tr} \left[\begin{array}{c} \Pi_{\bar{\rho}_{\ell'_j | \ell_{j-1}}}^{(j)} \Pi_{\rho_{m'_j, \ell'_j | \ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell'_j | \ell_{j-1}}}^{(j)} \times \\ \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \end{array} \right] \\
&= \mathbb{E}_{X_1^n} \text{Tr} \left[\begin{array}{c} \mathbb{E}_{U^n X^n | X_1^n} \left\{ \Pi_{\bar{\rho}_{\ell'_j | \ell_{j-1}}}^{(j)} \Pi_{\rho_{m'_j, \ell'_j | \ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell'_j | \ell_{j-1}}}^{(j)} \right\} \times \\ \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \underbrace{\mathbb{E}_{U^n X^n | X_1^n} \left\{ \rho_{m_j, \ell_j, \ell_{j-1}}^{(j)} \right\}}_{\textcircled{3}} \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \end{array} \right] \\
&= \mathbb{E}_{X_1^n} \text{Tr} \left[\begin{array}{c} \mathbb{E}_{U^n X^n | X_1^n} \left\{ \Pi_{\bar{\rho}_{\ell'_j | \ell_{j-1}}}^{(j)} \Pi_{\rho_{m'_j, \ell'_j | \ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell'_j | \ell_{j-1}}}^{(j)} \right\} \times \\ \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \underbrace{\bar{\rho}_{\ell_{j-1}}^{(j)}}_{\textcircled{4}} \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \end{array} \right] \\
&\leq 2^{-n[H(B|X_1) - \delta]} \mathbb{E}_{U^n X^n X_1^n} \text{Tr} \left[\Pi_{\bar{\rho}_{\ell'_j | \ell_{j-1}}}^{(j)} \Pi_{\rho_{m'_j, \ell'_j | \ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell'_j | \ell_{j-1}}}^{(j)} \Pi_{\bar{\rho}_{\ell_{j-1}}}^{(j)} \right] \\
&\leq 2^{-n[H(B|X_1) - \delta]} \mathbb{E}_{U^n X^n X_1^n} \text{Tr} \left[\Pi_{\rho_{m'_j, \ell'_j | \ell_{j-1}}}^{(j)} \right] \\
&\leq 2^{-n[H(B|X_1) - \delta]} \mathbb{E}_{U^n X^n X_1^n} 2^{n[H(B|X_1 U X) + \delta]} \\
&= 2^{-n[I(U X; B|X_1) - 2\delta]}.
\end{aligned}$$

7.3 Achievability proof

The result of the expectation in ③ is $\bar{\rho}_{|\ell_{j-1}}^{(j)}$, and we can bound the expression in ④ using (7.21). The first inequality follows because all the other terms in the trace are positive semidefinite operators less than or equal to the identity. The final inequality follows from (7.20).

Now we consider the expectation of the second term:

$$\begin{aligned}
\mathbb{E}_{U^n X^n X_1^n} \{(\text{B2})\} &= \mathbb{E}_{U^n X^n X_1^n} \left\{ \text{Tr} \left[P_{\ell'_j}^{B^{(j+1)}} \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right] \right\} \\
&= \text{Tr} \left[\mathbb{E}_{U^n X^n X_1^n} \left\{ P_{\ell'_j}^{B^{(j+1)}} \right\} \mathbb{E}_{U^n X^n X_1^n} \left\{ \rho_{m_{j+1}, \ell_{j+1}, \ell_j}^{(j+1)} \right\} \right] \\
&= \text{Tr} \left[\mathbb{E}_{U^n X^n X_1^n} \left\{ P_{\ell'_j}^{B^{(j+1)}} \right\} \bar{\tau}^{\otimes n} \right] \\
&= \mathbb{E}_{U^n X^n X_1^n} \text{Tr} \left[\Pi_{\bar{\tau}}^{(j+1)} \Pi_{\tau_{\ell'_j}}^{(j+1)} \Pi_{\bar{\tau}}^{(j+1)} \bar{\tau}^{\otimes n} \right] \\
&= \mathbb{E}_{U^n X^n X_1^n} \text{Tr} \left[\Pi_{\tau_{\ell'_j}}^{(j+1)} \Pi_{\bar{\tau}}^{(j+1)} \bar{\tau}^{\otimes n} \Pi_{\bar{\tau}}^{(j+1)} \right] \\
&\leq 2^{-n[H(B)-\delta]} \mathbb{E}_{U^n X^n X_1^n} \text{Tr} \left[\Pi_{\tau_{\ell'_j}}^{(j+1)} \Pi_{\bar{\tau}}^{(j+1)} \right] \\
&\leq 2^{-n[H(B)-\delta]} 2^{n[H(B|X_1)+\delta]} = 2^{-n[I(X_1;B)-2\delta]}.
\end{aligned}$$

Combining the upper bounds on (B1) and (B2) gives our final upper bound:

$$\begin{aligned}
\mathbb{E}_{U^n X^n X_1^n} \{(\text{B})\} &= \mathbb{E}_{U^n X^n X_1^n} \sum_{\ell'_j \neq \ell_j, m'_j} (\text{B1}) \times (\text{B2}) \\
&\leq \sum_{\ell'_j \neq \ell_j, m'_j} 2^{-n[I(UX;B|X_1)-2\delta]} \times 2^{-n[I(X_1;B)-2\delta]} \\
&\leq |\mathcal{L}| |\mathcal{M}| 2^{-n[I(X_1;B)+I(UX;B|X_1)-4\delta]}. \tag{7.22}
\end{aligned}$$

By choosing the size of message sets to satisfy equations (7.17) and (7.22), the expectation of the average probability of error at the destination becomes arbitrarily small for n sufficiently large.

7.3.2 Decoding at the relay

In this section we give the details of the POVM construction and the error analysis for the decoding at the relay.

POVM Construction. During block j , the relay wants to decode the message ℓ_j encoded in $u^n(\ell_j, \ell_{j-1})$, given the knowledge of the message ℓ_{j-1} from the previous block. Consider the state obtained by tracing over the systems X and B in (7.8):

$$\theta^{UX_1B_1} = \sum_{u, x_1} p(u|x_1) p(x_1) |u\rangle\langle u|^U \otimes |x_1\rangle\langle x_1|^{X_1} \otimes \sigma_{u, x_1}^{B_1},$$

where $\sigma_{u, x_1}^{B_1} \equiv \sum_x p(x|x_1, u) \text{Tr}_B [\rho_{x, x_1}^{B_1B}]$. Further tracing over the system U leads to the state

$$\theta^{X_1B_1} = \sum_{x_1} p(x_1) |x_1\rangle\langle x_1|^{X_1} \otimes \bar{\sigma}_{x_1}^{B_1},$$

where $\bar{\sigma}_{x_1} \equiv \sum_u p(u|x_1) \sigma_{u, x_1}^{B_1}$. Corresponding to the above conditional states are conditionally typical projectors of the following form

$$\Pi_{\sigma_{\ell_j|\ell_{j-1}}} \equiv \Pi_{\sigma_{u^n(\ell_j, \ell_{j-1}), x_1^n(\ell_{j-1})}^{B_{1(j)}}}, \quad \Pi_{\bar{\sigma}_{\ell_{j-1}}} \equiv \Pi_{\bar{\sigma}_{x_1^n(\ell_{j-1})}^{B_{1(j)}}}.$$

The relay constructs a square-root measurement $\{\Gamma_{\ell_j|\ell_{j-1}}\}$ using the following positive operators:

$$P_{\ell_j|\ell_{j-1}}^{B_{1(j)}} \equiv \Pi_{\bar{\sigma}_{\ell_{j-1}}} \Pi_{\sigma_{\ell_j|\ell_{j-1}}} \Pi_{\bar{\sigma}_{\ell_{j-1}}}. \quad (7.23)$$

Error analysis. In this section we show that during block j the relay will be able to decode the message ℓ_j from the state $\rho_{x^n(m_j, \ell_j, \ell_{j-1}), x_1^n(\ell_{j-1})}^{B_{1(j)}}$, provided the rate $R_\ell < I(U; B_1|X_1) = H(B_1|X_1) - H(B_1|UX_1) - \delta$. The bound follows from the following properties of typical projectors:

$$\text{Tr}[\Pi_{\sigma_{\ell_j|\ell_{j-1}}}] \leq 2^{n[H(B_1|UX_1)+\delta]}, \quad (7.24)$$

$$\Pi_{\bar{\sigma}_{\ell_{j-1}}} \bar{\sigma} \Pi_{\bar{\sigma}_{\ell_{j-1}}} \leq 2^{-n[H(B_1|X_1)-\delta]} \Pi_{\bar{\sigma}_{\ell_{j-1}}}. \quad (7.25)$$

7.3 Achievability proof

Recall that the average probability of error at the relay is given by:

$$\bar{p}_e^R \equiv \frac{1}{|\mathcal{L}|} \sum_{\ell_j} \text{Tr} \left\{ \left(I - \Gamma_{\ell_j|\ell_{j-1}}^{B_{1(j)}^n} \right) \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right\}.$$

We consider the probability of error for a single message ℓ_j and begin by applying the Hayashi-Nagaoka operator inequality (Lemma 3.1) to split the error into two terms:

$$\begin{aligned} \bar{p}_e^R &\equiv \text{Tr} \left[\left(I - \Gamma_{\ell_j|\ell_{j-1}}^{B_{1(j)}^n} \right) \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right] \\ &\leq \underbrace{2 \text{Tr} \left[\left(I - P_{\ell_j|\ell_{j-1}}^{B_{1(j)}^n} \right) \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right]}_{\text{(I)}} + 4 \underbrace{\sum_{\ell'_j \neq \ell_j} \text{Tr} \left[P_{\ell'_j|\ell_{j-1}}^{B_{1(j)}^n} \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right]}_{\text{(II)}}. \end{aligned}$$

We will bound the expectation of the average probability of error by bounding the individual terms. We bound the first term as follows:

$$\begin{aligned} \text{(I)} &= \text{Tr} \left[\left(I - P_{\ell_j|\ell_{j-1}}^{B_{1(j)}^n} \right) \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right] \\ &= \text{Tr} \left[\left(I - \Pi_{\bar{\sigma}_{|\ell_{j-1}}} \Pi_{\sigma_{\ell_j|\ell_{j-1}}} \Pi_{\bar{\sigma}_{|\ell_{j-1}}} \right) \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right] \\ &= 1 - \text{Tr} \left[\Pi_{\bar{\sigma}_{|\ell_{j-1}}} \Pi_{\sigma_{\ell_j|\ell_{j-1}}} \Pi_{\bar{\sigma}_{|\ell_{j-1}}} \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right] \\ &\leq 1 - \text{Tr} \left[\Pi_{\sigma_{\ell_j|\ell_{j-1}}} \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right] + \left\| \Pi_{\bar{\sigma}_{|\ell_{j-1}}} \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \Pi_{\bar{\sigma}_{|\ell_{j-1}}} - \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right\|_1, \end{aligned}$$

where the inequality follows from Lemma 2.20.

By taking the expectation over the code randomness we obtain the bound

$$\begin{aligned} \mathbb{E}_{U^n X^n X_1^n} \text{(I)} &= 1 - \mathbb{E}_{U^n X_1^n} \text{Tr} \left[\Pi_{\sigma_{\ell_j|\ell_{j-1}}} \mathbb{E}_{X^n | U^n X_1^n} \left\{ \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right\} \right] \\ &\quad + \mathbb{E}_{U^n X^n X_1^n} \left\| \Pi_{\bar{\sigma}_{|\ell_{j-1}}} \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \Pi_{\bar{\sigma}_{|\ell_{j-1}}} - \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right\|_1 \\ &= 1 - \mathbb{E}_{U^n X_1^n} \text{Tr} \left[\Pi_{\sigma_{\ell_j|\ell_{j-1}}} \sigma_{\ell_j, \ell_{j-1}} \right] \\ &\quad + \mathbb{E}_{U^n X^n X_1^n} \left\| \Pi_{\bar{\sigma}_{|\ell_{j-1}}} \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \Pi_{\bar{\sigma}_{|\ell_{j-1}}} - \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right\|_1 \\ &\leq 1 - \mathbb{E}_{U^n X_1^n} \text{Tr} \left[\Pi_{\sigma_{\ell_j|\ell_{j-1}}} \sigma_{\ell_j, \ell_{j-1}} \right] + 2\sqrt{\epsilon} \\ &\leq 1 - (1 - \epsilon) + 2\sqrt{\epsilon} = \epsilon + 2\sqrt{\epsilon}. \end{aligned}$$

The first inequality follows from Lemma 3.2 and the property

$$\mathbb{E}_{U^n X_1^n} \text{Tr} \left[\Pi_{\bar{\sigma}_{|\ell_{j-1}}} \bar{\sigma} \right] \geq 1 - \epsilon. \quad (7.26)$$

The second inequality follows from:

$$\mathbb{E}_{U^n X_1^n} \text{Tr} \left[\Pi_{\sigma_{\ell_j|\ell_{j-1}}} \sigma_{\ell_j, \ell_{j-1}} \right] \geq 1 - \epsilon. \quad (7.27)$$

To bound the second term we proceed as follows:

$$\begin{aligned} \mathbb{E}_{U^n X^n X_1^n} \{(\text{II})\} &= \mathbb{E}_{U^n X^n X_1^n} \sum_{\ell'_j \neq \ell_j} \text{Tr} \left[P_{\ell'_j|\ell_{j-1}}^{B_{1(j)}^n} \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right] \\ &= \mathbb{E}_{X_1^n} \sum_{\ell'_j \neq \ell_j} \text{Tr} \left[\mathbb{E}_{U^n X^n | X_1^n} \left\{ P_{\ell'_j|\ell_{j-1}}^{B_{1(j)}^n} \right\} \mathbb{E}_{U^n X^n | X_1^n} \left\{ \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right\} \right] \\ &= \mathbb{E}_{X_1^n} \sum_{\ell'_j \neq \ell_j} \text{Tr} \left[\mathbb{E}_{U^n X^n | X_1^n} \left\{ P_{\ell'_j|\ell_{j-1}}^{B_{1(j)}^n} \right\} \bar{\sigma}_{|\ell_{j-1}} \right]. \end{aligned}$$

The expectation can be broken up because $\ell'_j \neq \ell_j$ and thus the U^n codewords are independent. We have also used

$$\mathbb{E}_{U^n X^n | X_1^n} \left\{ \rho_{m_j, \ell_j, \ell_{j-1}}^{B_{1(j)}^n} \right\} = \bar{\sigma}_{|\ell_{j-1}}. \quad (7.28)$$

We continue by expanding the operator $P_{\ell'_j|\ell_{j-1}}^{B_{1(j)}^n}$ as follows:

$$\begin{aligned} &= \mathbb{E}_{U^n X^n X_1^n} \sum_{\ell'_j \neq \ell_j} \text{Tr} \left[\Pi_{\bar{\sigma}_{|\ell_{j-1}}} \Pi_{\sigma_{\ell'_j|\ell_{j-1}}} \Pi_{\bar{\sigma}_{|\ell_{j-1}}} \bar{\sigma}_{|\ell_{j-1}} \right] \\ &= \mathbb{E}_{U^n X^n X_1^n} \sum_{\ell'_j \neq \ell_j} \text{Tr} \left[\Pi_{\sigma_{\ell'_j|\ell_{j-1}}} \underbrace{\Pi_{\bar{\sigma}_{|\ell_{j-1}}} \bar{\sigma}_{|\ell_{j-1}} \Pi_{\bar{\sigma}_{|\ell_{j-1}}}}_{\textcircled{5}} \right] \\ &\leq \mathbb{E}_{U^n X^n X_1^n} \sum_{\ell'_j \neq \ell_j} \text{Tr} \left[\Pi_{\sigma_{\ell'_j|\ell_{j-1}}} 2^{-n[H(B_1|X_1) - \delta]} \Pi_{\bar{\sigma}_{|\ell_{j-1}}} \right] \\ &\leq 2^{-n[H(B_1|X_1) - \delta]} \mathbb{E}_{U^n X^n X_1^n} \sum_{\ell'_j \neq \ell_j} \text{Tr} \left[\Pi_{\sigma_{\ell'_j|\ell_{j-1}}} \right] \end{aligned}$$

$$\begin{aligned}
&\leq 2^{-n[H(B_1|X_1)-\delta]} \mathbb{E}_{U^n X^n X_1^n} \sum_{\ell'_j \neq \ell_j} 2^{n[H(B_1|UX_1)+\delta]} \\
&\leq |\mathcal{L}| 2^{-n[I(U;B_1|X_1)-2\delta]}.
\end{aligned}$$

The first inequality follows from using (7.25) on the expression ⑤. The second inequality follows from the fact that $\Pi_{\bar{\sigma}|\ell_{j-1}}$ is a positive semidefinite operator less than or equal to the identity. More precisely we have

$$\begin{aligned}
\mathrm{Tr} \left[\Pi_{\sigma_{\ell'_j|\ell_{j-1}}} \Pi_{\bar{\sigma}|\ell_{j-1}} \right] &= \mathrm{Tr} \left[\Pi_{\sigma_{\ell'_j|\ell_{j-1}}} \Pi_{\bar{\sigma}|\ell_{j-1}} \Pi_{\sigma_{\ell'_j|\ell_{j-1}}} \right] \\
&\leq \mathrm{Tr} \left[\Pi_{\sigma_{\ell'_j|\ell_{j-1}}} I \Pi_{\sigma_{\ell'_j|\ell_{j-1}}} \right] \\
&= \mathrm{Tr} \left[\Pi_{\sigma_{\ell'_j|\ell_{j-1}}} \right].
\end{aligned}$$

The penultimate inequality follows from (7.24).

Thus if we choose $R_\ell \leq I(U;B_1|X_1) - 3\delta$, we can make the expectation of the average probability of error at the relay vanish in the limit of many uses of the channel.

Proof conclusion. Note that the *gentle operator lemma for ensembles* is used several times in the proof. First, it is used to guarantee that the effect of acting with one of the projectors from the “measurement sandwich” does not disturb the state too much. Furthermore, because each of the output blocks is operated on twice: we depend on the gentle operator lemma to guarantee that the disturbance to the state during the first decoding stage is asymptotically negligible if the correct messages are decoded.

7.4 Discussion

In this chapter, we established the achievability of the rates given by the partial decode-and-forward strategy, thus extending the study of classical-quantum channels to multi-hop scenarios.

The new techniques from this chapter are the use of the *coherent codebooks* and the AND-measurement, which collectively decodes messages from two blocks of the output of the channel.

We obtain the decoding-and-forward inner bound as a corollary of Theorem 7.1.

Corollary 7.1 (Decode-and-forward strategy for quantum relay channel). *The rates R satisfying*

$$R \leq \max_{p(x,x_1)} \min\{ I(X, X_1; B)_\theta, I(X; B_1|X_1)_\theta \} \quad (7.29)$$

where the mutual information quantities are taken with respect to the state

$$\theta^{XX_1B_1B} = \sum_{x,x_1} \underbrace{p_{X|X_1}(x|x_1)p_{X_1}(x_1)}_{p_{X,X_1}} |x\rangle\langle x|^X \otimes |x_1\rangle\langle x_1|^{X_1} \otimes \rho_{x,x_1}^{B_1B}. \quad (7.30)$$

are achievable for quantum relay channels by setting $X = U$ in Theorem 7.1.

Note also that setting the x_1 to a fixed input in Theorem 7.1 would give us a quantum direct coding inner bound similar to the one from equation (7.2).

An interesting open question is to determine a compress-and-forward strategy for the quantum setting. This could possibly involve combining results from quantum source coding and quantum channel coding [DHW11, WS12].

Another avenue for research would be to consider *quantum communication* and *entanglement distillation* scenarios on a quantum relay network. Further research in this area would have applications for the design of quantum repeaters [CGDR05, Dut11b].

Chapter 8

Bosonic interference channels

Optical communication links form the backbone of the information superhighway which is the Internet. A single optical fiber can carry hundreds of gigabits of data per second over long distances thanks to the excellent light-transmission properties of glass materials. Free-space optical communication is also possible at rates of hundreds of megabits per second [TNO02].

An optical communication system consists of a modulated source of photons, the optical channel (or more generally the *bosonic* channel, since photons are bosons), and an optical detector. Figure 4.2 on page 42 illustrates an example of such a communication system.

As information theorists, we are interested in determining the ultimate limits on the rates for communication over such channels. For each possible combination of the optical encoding and optical decoding strategies, we obtain a different communication model for which we can calculate the capacity. More generally, we are interested in the *ultimate* capacity of the bosonic channel as permitted by the laws of physics. For this purpose we must optimize over all possible encoding and decoding strategies, both practical and theoretical.

In this chapter we present a quantum treatment of a free-space optical interference channel. We consider the performance of laser-light encoding (coherent light) in conjunction with three detection strategies: (1) homodyne, (2) heterodyne, and (3) joint detection. In Section 8.1, we will introduce some basic notions of quantum optics which are required for the remainder of the chapter. In Section 8.2 we will discuss

previous results on bosonic quantum channels and describe the known capacity formulas for point-to-point free-space bosonic channels for the three detection strategies. In Section 8.3 we define the bosonic interference channel model and calculate the capacity region for the special cases of “strong” and “very strong” interference for each detection strategy. We also establish the Han-Kobayashi achievable rate regions for homodyne, heterodyne and joint detection.

8.1 Preliminaries

8.1.1 Gaussian channels

We begin by introducing some notation. Define the real-valued Gaussian probability density function with mean μ and variance σ^2 as follows:

$$\mathcal{N}_{\mathbb{R}}(x; \mu, \sigma^2) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \in \mathcal{P}(\mathbb{R}). \quad (8.1)$$

Define also the circularly symmetric complex-valued Gaussian distribution

$$\mathcal{N}_{\mathbb{C}}(z; \mu, \sigma^2) \equiv \frac{1}{2\pi\sigma^2} e^{-\frac{|z-\mu|^2}{2\sigma^2}} \equiv \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\text{Re}\{\mu\})^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\text{Im}\{\mu\})^2}{2\sigma^2}} \in \mathcal{P}(\mathbb{C}), \quad (8.2)$$

where we identify $z = x + iy$ and assume that the variance parameter is real-valued $\sigma^2 \in \mathbb{R}$. Note also that in the complex-valued case, the quantity σ^2 represents the *variance per real dimension*; a variable $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ will have variance $\text{Var}\{Z\} \equiv \mathbb{E}_Z[|Z - \mu|^2] = 2\sigma^2$.

The additive white Gaussian noise (AWGN) channel is a communication model where the input and output are continuous random variables and the noise is Gaussian. Let X be the random variable associated with the input of the channel. Then the output variable Y will be:

$$Y = X + Z, \quad (8.3)$$

where $Z \sim \mathcal{N}_{\mathbb{R}}(0, N)$ is a Gaussian random variable with zero-mean and variance N . As in the discrete memoryless case, we can use a codebook $\{x^n(m)\}$, $m \in [1 : 2^{nR}]$, with codewords generated randomly and independently according to a probability density function $\prod^n p_X(x)$. Furthermore we impose an *average power constraint* on the

codebook:

$$\mathbb{E}_{X^n} \left\{ \frac{1}{n} \sum_{i=1}^n X_i^2 \right\} \leq P. \quad (8.4)$$

The channel capacity is calculated using the *differential entropy*, $h : \mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}$, which plays the role of the Shannon entropy for continuous random variables. We know from Shannon's channel capacity theorem (Theorem 3.1) that a rate R is achievable provided it is less than the mutual information of the joint probability distribution induced by the input distribution and the channel: $(X, Y) \sim p_X p_{Y|X}$. For any choice of input distribution p_X , the following rate is achievable:

$$\begin{aligned} R \leq I(X; Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X + Z|X) \\ &= h(Y) - h(Z|X) \\ &= h(Y) - h(Z). \end{aligned} \quad (8.5)$$

The last equality follows because the noise Z is assumed to be independent of the input X . It can be shown that a Gaussian distribution with variance P is the optimal choice of input distribution [CT91]. Furthermore, when we choose $X \sim \mathcal{N}_{\mathbb{R}}(0, P)$ it is possible to compute the above expression exactly and obtain the capacity:

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{N} \right) \quad [\text{bits/use}]. \quad (8.6)$$

We will refer to the ratio P/N as the *signal to noise ratio*. We sometimes abbreviate this expression as: $\gamma(\text{SNR}) \equiv \frac{1}{2} \log_2(1 + \text{SNR})$. The above formula is one of the great successes of classical information theory.

The Gaussian multiple access channel is defined as:

$$Y = \sqrt{\alpha} X_1 + \sqrt{\beta} X_2 + Z, \quad (8.7)$$

where $\alpha, \beta \in \mathbb{R}$ are the *gain coefficients* and $Z \sim \mathcal{N}_{\mathbb{R}}(0, N)$ is an additive Gaussian noise term with average power N . When input power constraints $\mathbb{E}_{X_1^n} \left\{ \frac{1}{n} \sum_{i=1}^n X_{1i}^2 \right\} \leq P_1$

and $\mathbb{E}_{X_2^n} \left\{ \frac{1}{n} \sum_{i=1}^n X_{2i}^2 \right\} \leq P_2$ are imposed, the capacity region is given by:

$$C_{\text{MAC}} \equiv \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \left| \begin{array}{l} R_1 \leq I(X_1; Y|X_2) = \frac{1}{2} \log_2 \left(1 + \frac{\alpha P_1}{N} \right) \\ R_2 \leq I(X_2; Y|X_1) = \frac{1}{2} \log_2 \left(1 + \frac{\beta P_2}{N} \right) \\ R_1 + R_2 \leq I(X_1 X_2; Y) = \frac{1}{2} \log_2 \left(1 + \frac{\alpha P_1 + \beta P_2}{N} \right) \end{array} \right. \right\}.$$

Each of the constraints on the capacity region has an intuitive interpretation in terms of signal to noise ratios. In this context, we also have the expression $I(X_1; Y) = \frac{1}{2} \log_2 \left(1 + \frac{\alpha P_1}{N + \beta P_2} \right)$, in which the unknown codewords of the second transmitter are treated as contributing to the noise.

8.1.2 Introduction to quantum optics

Photons are excitations of the electromagnetic field. We say that photons are *bosons* because they obey Bose-Einstein statistics: they are indistinguishable particles that are symmetric under exchange¹. Multiple bosons with the same energy can occupy the same quantum state. This is in contrast with *fermions* which obey Pauli's exclusion principle. Bosonic channels are channels in which the inputs and the outputs are bosons.

In this section, we will introduce some background material on quantum optics which is needed for the rest of the presentation in this chapter. Recall that the states of quantum systems are described by density operators $\sigma, \rho \in \mathcal{D}(\mathcal{H})$, where \mathcal{H} is a Hilbert space. Unitary quantum operations act by conjugation, so that by applying U to σ we obtain $\rho = U\sigma U^\dagger$ as output. The expectation value of some operator \hat{A} when the system is in the state ρ is denoted $\langle \hat{A} \rangle = \text{Tr}[\hat{A}\rho]$.

Let $\rho_0 = |0\rangle\langle 0|$ be the *vacuum state* of one mode of the electromagnetic field. We define \hat{a}^\dagger to be the *creation operator* for that mode. Applying \hat{a}^\dagger to the vacuum state we obtain the first excited state:

$$|1\rangle\langle 1| = \hat{a}^\dagger |0\rangle\langle 0| \hat{a}, \quad (8.8)$$

and this process can be iterated to create further excitations in the field. The Hermitian conjugate of the creation operator is the *annihilation* operator which takes away

¹ The wave function describing two photons p_1 and p_2 is even under exchange of the two particles: $\psi(p_1, p_2) = \psi(p_2, p_1)$.

excitations from the field. More generally, we have

$$a|n\rangle = \sqrt{n} |n-1\rangle, \quad (8.9)$$

$$a^\dagger|n\rangle = \sqrt{n+1} |n+1\rangle. \quad (8.10)$$

$$(8.11)$$

The state space $|0\rangle\langle 0|, |1\rangle\langle 1|, |2\rangle\langle 2|, |3\rangle\langle 3|, \dots$ is known as *Fock space* and it is infinite dimensional. The creation and annihilation operators obey the commutation relation $[\hat{a}, \hat{a}^\dagger] = 1$.

The real part and the imaginary part of the operator \hat{a} are defined as the x quadrature and the p quadrature:

$$\hat{X} = \frac{\hat{a} + \hat{a}^\dagger}{\sqrt{2}}, \quad \hat{P} = \frac{\hat{a} - \hat{a}^\dagger}{i\sqrt{2}}, \quad (8.12)$$

and we have $[\hat{X}, \hat{P}] = i$.

If we want to measure how many excitations are in the field, we use the *number operator* $\hat{N} = \hat{a}^\dagger \hat{a}$. If the field is in excitation level n , the expected number of excitations will be:

$$\langle \hat{N} \rangle = \text{Tr} [\hat{a}^\dagger \hat{a} |n\rangle\langle n|] = n. \quad (8.13)$$

The Hamiltonian that describes one non-interacting mode of the electromagnetic field is given by:

$$\hat{H} = \hbar\omega \left(\hat{a}^\dagger \hat{a} + \frac{1}{2} \right). \quad (8.14)$$

The Hamiltonian is important because it gives the time evolution operator $U(t) \equiv e^{i\hat{H}t}$ and the energy of the system: $E_\rho \equiv \langle \hat{H} \rangle = \text{Tr}[\hat{H}\rho]$. Observe that the system has energy even when it is in the vacuum state:

$$E_0 = \text{Tr}[\hat{H}|0\rangle\langle 0|] = \langle 0|\hat{H}|0\rangle = \hbar\omega \langle 0| \left(\hat{a}^\dagger \hat{a} + \frac{1}{2} \right) |0\rangle = \frac{\hbar\omega}{2}. \quad (8.15)$$

This is known as the zero-point energy or vacuum energy.

8.1.3 Coherent states

A composite system exhibits coherence if all its components somehow coincide with each other. This could be either coincidence in time, space coherence, phase coherence or quantum coherence. An example of the latter is the process of *stimulated emission* of photons which occurs inside a laser. All new photons are created exactly “in phase” with the other photons inside the laser. Over time the number of photons in the laser will grow, but they will all have the same frequency, phase and polarization.

The coherent state $|\alpha\rangle$ describes an oscillation of the electromagnetic field. In general $\alpha \in \mathbb{C}$ and we have $\alpha = |\alpha|e^{i\phi}$, where $|\alpha|$ is the amplitude of the oscillation and ϕ is the initial phase. In the Fock basis, the coherent state $|\alpha\rangle$ is written as:

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \quad (8.16)$$

$$= e^{-\frac{|\alpha|^2}{2}} \left[|0\rangle + |\alpha|e^{i\phi}|1\rangle + \frac{|\alpha|^2}{\sqrt{2}}e^{2i\phi}|2\rangle + \frac{|\alpha|^3}{\sqrt{6}}e^{3i\phi}|3\rangle + \dots \right]. \quad (8.17)$$

The output of a laser is coherent light: the excitations at all energy levels will have the same phase. Coherent states remain coherent over time: $|\alpha(t)\rangle \equiv U(t)|\alpha\rangle = e^{i\omega t/2} |\alpha|e^{i(\phi-\omega t)}$.

A coherent state can also be defined in terms of the unitary *displacement operator* which acts as:

$$|\alpha\rangle = D(\alpha)|0\rangle = \exp(\alpha\hat{a}^\dagger - \alpha^*\hat{a})|0\rangle. \quad (8.18)$$

Note that in some respect $D(\alpha)$ is similar to the creation operator \hat{a}^\dagger , since it creates excited states from the vacuum state.

8.2 Bosonic channels

Point-to-point optical communication using laser-light modulation in conjunction with direct-detection and coherent-detection receivers has been studied in detail using the semiclassical theory of photodetection [GK95]. This approach treats light as a classical electromagnetic field, and the fundamental noise encountered in photodetection is the shot noise associated with the discreteness of the electron charge.

These semiclassical treatments for systems that exploit classical-light modulation and conventional receivers (direct, homodyne, or heterodyne) have had some success, but we should recall that electromagnetic waves are quantized, and the correct assessment of systems that use non-classical light sources and/or general optical measurements requires a full quantum-mechanical framework [Sha09]. There are several recent theoretical studies on the point-to-point [GGL⁺04, Guh11], broadcast [GSE07] and multiple-access [Yen05a] bosonic channels. These studies have shown that quantum communication rates (Holevo rates) surpass what can be obtained with conventional receivers. For the general quantum channel, attaining Holevo information rates may require collective measurements (a joint detection) across all the output systems of the channel.

Before stating our results on the bosonic interference channel, we will briefly review some results on point-to-point bosonic channels in the next subsection.

8.2.1 Channel model

The free-space optical communication channel is a physically realistic model for the propagation of photons from transmitter to receiver. We assume that a transmitter aperture of size A_t is placed at a distance L from a receiver aperture of size A_r , and that we are using λ -wavelength laser light for the transmission.

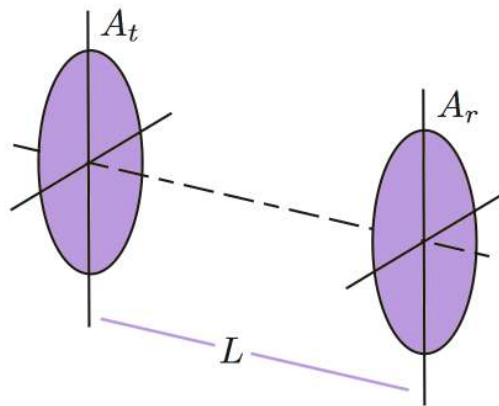


Figure 8.1: The free-space optical communication channel. Two apertures of area A_t and A_r are placed L distance apart. The channel decomposes into different modes of propagation. We model the channel as a transformation from an annihilation operator on the transmit side to an annihilation operator at the receiver side.

To analyze the communication capacity of the bosonic channel, we can decompose the problem into finding the capacity for each of the spatial modes of propagation, which will in general have different transmissivity coefficients η . In the far-field propagation regime, which is when we have $A_t A_r / (\lambda L)^2 \ll 1$, only two orthogonal spatial modes (one for each polarization degree of freedom) will have significant power transmissivity. We will analyze the channel for a single mode (one choice of polarization).

The channel input is an electromagnetic field mode with annihilation operator \hat{a} , and the channel output is another mode with annihilation operator \hat{b} . The channel map is described by:

$$\hat{b} = \sqrt{\eta} \hat{a} + \sqrt{1-\eta} \hat{\nu}, \quad (8.19)$$

in which $\hat{\nu}$ is associated with the noise of the environment and the parameter η , $0 \leq \eta \leq 1$, models the channel transmissivity.

We say that a channel is *pure-loss* if the environmental noise $\hat{\nu}$ is in the vacuum state $|0\rangle\langle 0|$. A channel has *thermal noise* if the mode $\hat{\nu}$ is in the thermal state:

$$\rho_t = \int d^2\alpha \frac{\exp(-|\alpha|^2/N_B)}{\pi N_B} |\alpha\rangle\langle\alpha|, \quad (8.20)$$

which is Gaussian mixture of coherent states with average photon number $N_B > 0$. One can also write the thermal state in the number basis as follows:

$$\rho_t = \frac{1}{N_B + 1} \sum_{n=0}^{\infty} \left(\frac{N_B}{N_B + 1} \right)^n |n\rangle\langle n|. \quad (8.21)$$

8.2.2 Encoding

We will use coherent state encoding of the information at the transmitter. The codebook consists of tensor products of vacuum states displaced randomly and independently by an amount drawn from a distribution p_α :

$$\alpha^n \sim \prod_{\alpha}^n p_\alpha \rightarrow |\alpha_1 \alpha_2 \cdots \alpha_n\rangle \equiv D(\alpha_1)|0\rangle \otimes D(\alpha_2)|0\rangle \otimes \cdots \otimes D(\alpha_n)|0\rangle.$$

This encoding strategy is chosen because it is simple to implement in practice, and also because it is known that it suffices to achieve the ultimate capacity of the bosonic channel [GGL⁺04].

When homodyne detection will be used at the receiver, we will encode the information using only the x quadrature. The displacements are chosen according to:

$$\alpha \sim \mathcal{N}_{\mathbb{R}}(0, N_S). \quad (8.22)$$

The distribution is chosen so that it satisfies the constraint on the average number of input photons $\langle |\alpha|^2 \rangle \leq N_S$, which is the quantum analogue of the input power constraint for the AWGN channel.

For heterodyne and joint detection, we will use both quadratures and choose the displacements according to a circularly-symmetric complex-valued Gaussian distribution:

$$\alpha \sim \mathcal{N}_{\mathbb{C}}(0, N_S/2). \quad (8.23)$$

8.2.3 Homodyne detection

Homodyne detection consists of combining on a beamsplitter the incoming light and a local oscillator signal and measuring the resulting difference of the intensities. By tuning the relative phase between the incoming signal and the local oscillator it is possible to measure the incoming photons in any quadrature.

When coherent state encoding is used with displacement values chosen as in (8.22) and homodyne detection is used, the resulting channel is Gaussian:

$$Y = \sqrt{\eta}\alpha + Z_{\text{hom}},$$

where $Z_{\text{hom}} \sim \mathcal{N}_{\mathbb{R}}(0, (2(1-\eta)N_B + 1)/4)$. The “+1” term in the noise variance arises physically from the zero-point fluctuations of the vacuum.

We can now use the general formula for the capacity of the AWGN channel from (8.6) to obtain the capacity with homodyne detection:

$$C_{\text{hom}} = \frac{1}{2} \log \left(1 + \frac{4\eta N_S}{2(1-\eta)N_B + 1} \right) \text{ bits/use.} \quad (8.24)$$

8.2.4 Heterodyne detection

The heterodyne detection strategy attempts to measure the incoming light in both quadratures. The sender inputs a coherent state $|\alpha\rangle$ with $\alpha \in \mathbb{C}$. Heterodyne detection of the channel output results in a classical complex Gaussian channel, where the receiver output is a complex random variable Y described by:

$$Y = \sqrt{\eta}\alpha + Z_{\text{het}}, \quad (8.25)$$

where $Z_{\text{het}} \sim \mathcal{N}_{\mathbb{C}}(0, ((1 - \eta)N_B + 1)/2)$. The capacity formula for this choice of detection strategy is given by:

$$C_{\text{het}} = \log \left(1 + \frac{\eta N_S}{(1 - \eta)N_B + 1} \right) \text{ bits/use.} \quad (8.26)$$

The factor of $1/2$ in the noise variances is due to the attempt to measure both quadratures of the field simultaneously [Sha09].

8.2.5 Joint detection

The capacity of the single-mode lossy bosonic channel with thermal background noise is thought to be equal to the channel's Holevo information:

$$\chi \equiv g(\eta N_S + (1 - \eta)N_B) - g((1 - \eta)N_B) \quad \text{bits/use,} \quad (8.27)$$

where N_S and N_B are the mean photon numbers per mode for the input signal and the thermal noise, and $g(N) \equiv (N + 1) \log(N + 1) - N \log(N)$ is the entropy of a thermal state with mean photon number N . The latter formula is easily obtained from (8.21):

$$\begin{aligned} h(\rho_t) &= -\text{Tr}[\rho_t \log \rho_t] \\ &= -\sum_{n=0}^{\infty} \frac{1}{N+1} \left(\frac{N}{N+1} \right)^n \log \left(\frac{1}{N+1} \left(\frac{N}{N+1} \right)^n \right) \\ &= \sum_{n=0}^{\infty} \frac{1}{N+1} \left(\frac{N}{N+1} \right)^n \left[-n \log N + (n+1) \log(N+1) \right] \\ &= (N+1) \log(N+1) - N \log N = g(N). \end{aligned}$$

This capacity formula from equation (8.27) assumes a long-standing conjecture regarding the minimum-output entropy of the thermal noise channel [GGL⁺04, GHLM10].

It is known that joint-detection (collective) measurements over long codeword blocks are necessary to achieve the rates in equation (8.27) for both the pure-loss and the thermal-noise lossy bosonic channel [Guh11, WGTL12]. Note, however, that quantum states of light are not necessary to achieve the rate χ ; coherent-state encoding is sufficient.

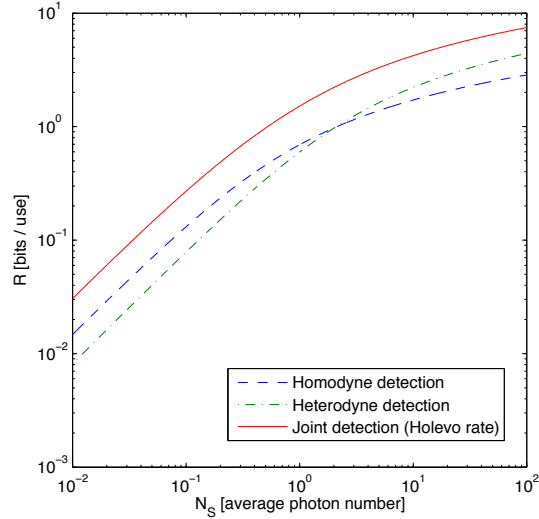


Figure 8.2: The achievable rates for the different decoding strategies: homodyne, heterodyne and joint detection in the low photon number regime $0.01 \leq \langle |\alpha|^2 \rangle = N_S \leq 100$. The channel has $\eta = 0.9$ and $N_B = 1$. The joint detection strategy outperforms the classical strategies in which the outputs of the channel are measured individually, cf. Figure 3.5.

The rates achievable by the three different detection strategies are illustrate in Figure 8.2, and on this we conclude our review of point-to-point bosonic communication. In the next section, we consider the bosonic interference channel with thermal-noise, particularly in the context of free-space terrestrial optical communications.

8.3 Free-space optical interference channels

Consider now a scenario similar to the one described in Figure 8.1, but now assume that there are two senders and two receivers. Sender 1 modulates her information on the first spatial mode of the transmitter-pupil, and Receiver 1 separates and demodulates information from the corresponding receiver-pupil spatial mode. With perfect

spatial-mode control at the transmitter and perfect mode separation at the receiver, the orthogonal spatial modes can be thought of as independent parallel channels with no crosstalk. However, imperfect (slightly non-orthogonal) mode generation or imperfect mode separation can result in crosstalk (interference) between the different channels.

We will model the bosonic interference channel as a passive linear mixing of the input modes along with a thermal environment adding zero-mean, isotropic Gaussian noise. The channel model is given by:

$$\hat{b}_1 = \sqrt{\eta_{11}}\hat{a}_1 + \sqrt{\eta_{21}}\hat{a}_2 + \sqrt{\bar{\eta}_1}\hat{\nu}_1, \quad (8.28)$$

$$\hat{b}_2 = \sqrt{\eta_{12}}\hat{a}_1 - \sqrt{\eta_{22}}\hat{a}_2 + \sqrt{\bar{\eta}_2}\hat{\nu}_2, \quad (8.29)$$

where $\eta_{11}, \eta_{12}, \eta_{21}, \eta_{22}, \bar{\eta}_1, \bar{\eta}_2 \in \mathbb{R}_+$, $\sqrt{\eta_{11}\eta_{12}} = \sqrt{\eta_{21}\eta_{22}}$, $\bar{\eta}_1 \equiv 1 - \eta_{11} - \eta_{21}$, and $\bar{\eta}_2 \equiv 1 - \eta_{12} - \eta_{22}$. The following conditions ensure that the network is passive:

$$\eta_{11} + \eta_{12} \leq 1, \quad \eta_{11} + \eta_{21} \leq 1, \quad \eta_{22} + \eta_{21} \leq 1, \quad \eta_{22} + \eta_{12} \leq 1.$$

We constrain the mean photon number of the transmitters \hat{a}_1 and \hat{a}_2 to be N_{S_1} and N_{S_2} photons per mode, respectively. The environment modes $\hat{\nu}_1$ and $\hat{\nu}_2$ are in statistically independent zero-mean thermal states with respective mean photon numbers N_{B_1} and N_{B_2} per mode [Sha09].

8.3.1 Detection strategies

For a coherent state encoding and coherent² detection at both receivers, the above model is a special case of the Gaussian interference channel, and we can study its capacity regions in various settings by applying the known classical results from [Car75, Sat81] and [HK81].

If the senders prepare their inputs in coherent states $|\alpha_1\rangle$ and $|\alpha_2\rangle$, with $\alpha_1, \alpha_2 \in \mathbb{R}$, and both receivers perform x -quadrature homodyne detection on their respective modes, the result is a classical Gaussian interference channel [Sha09], where Receivers 1 and

²We refer to both homodyne and heterodyne strategies as *coherent* strategies.

2 obtain respective conditional Gaussian random variables Y_1 and Y_2 distributed as

$$\begin{aligned} Y_1 &\sim \mathcal{N}_{\mathbb{R}}(\sqrt{\eta_{11}}\alpha_1 + \sqrt{\eta_{21}}\alpha_2, (2\bar{\eta}_1 N_{B_1} + 1)/4), \\ Y_2 &\sim \mathcal{N}_{\mathbb{R}}(\sqrt{\eta_{12}}\alpha_2 + \sqrt{\eta_{22}}\alpha_1, (2\bar{\eta}_2 N_{B_2} + 1)/4), \end{aligned}$$

where the “+1” term in the noise variances arises physically from the zero-point fluctuations of the vacuum. Suppose that the senders again encode their signals as coherent states $|\alpha_1\rangle$ and $|\alpha_2\rangle$, but this time with $\alpha_1, \alpha_2 \in \mathbb{C}$, and that the receivers both perform heterodyne detection. This results in a classical complex Gaussian interference channel [Sha09], where Receivers 1 and 2 detect respective conditional complex Gaussian random variables Z_1 and Z_2 , whose real parts are distributed as

$$\text{Re}\{Z_m\} \sim \mathcal{N}_{\mathbb{R}}(\mu_m, (\bar{\eta}_m N_{B_m} + 1)/2), \quad (8.30)$$

where $m \in \{1, 2\}$, $\mu_1 \equiv \sqrt{\eta_{11}} \text{Re}\{\alpha_1\} + \sqrt{\eta_{21}} \text{Re}\{\alpha_2\}$, $\mu_2 \equiv \sqrt{\eta_{12}} \text{Re}\{\alpha_1\} + \sqrt{\eta_{22}} \text{Re}\{\alpha_2\}$, and the imaginary parts of Z_1 and Z_2 are distributed with the same variance as their real parts, and their respective means are $\sqrt{\eta_{11}} \text{Im}\{\alpha_1\} + \sqrt{\eta_{21}} \text{Im}\{\alpha_2\}$ and $\sqrt{\eta_{12}} \text{Im}\{\alpha_1\} + \sqrt{\eta_{22}} \text{Im}\{\alpha_2\}$. The factor of 1/2 in the noise variances is due to the attempt to measure both quadratures of the field simultaneously [Sha09].

8.4 Very strong interference case

Recall the setting of the interference channel which we discussed in Section 5.2.1, where the crosstalk between the communication links is so strong that the receivers can fully decode the interfering signal and “subtract” it from the received signal to completely cancel its effects. The conditions in (5.5) and (5.6) translate to the following ones for the case of coherent-state encoding and coherent detection:

$$\begin{aligned} \frac{\eta_{21}}{\eta_{22}} &\geq \frac{4^i \eta_{11} N_{S_1} + 2^i \bar{\eta}_1 N_{B_1} + 1}{2^i \bar{\eta}_2 N_{B_2} + 1}, \\ \frac{\eta_{12}}{\eta_{11}} &\geq \frac{4^i \eta_{22} N_{S_2} + 2^i \bar{\eta}_2 N_{B_2} + 1}{2^i \bar{\eta}_1 N_{B_1} + 1}, \end{aligned}$$

8.4 Very strong interference case

and the capacity region becomes

$$R_1 \leq \frac{1}{2^i} \log \left(1 + \frac{4^i \eta_{11} N_{S_1}}{2^i \bar{\eta}_1 N_{B_1} + 1} \right), \quad (8.31)$$

$$R_2 \leq \frac{1}{2^i} \log \left(1 + \frac{4^i \eta_{22} N_{S_2}}{2^i \bar{\eta}_2 N_{B_2} + 1} \right), \quad (8.32)$$

where $i = 1$ for homodyne detection and $i = 0$ for heterodyne detection.

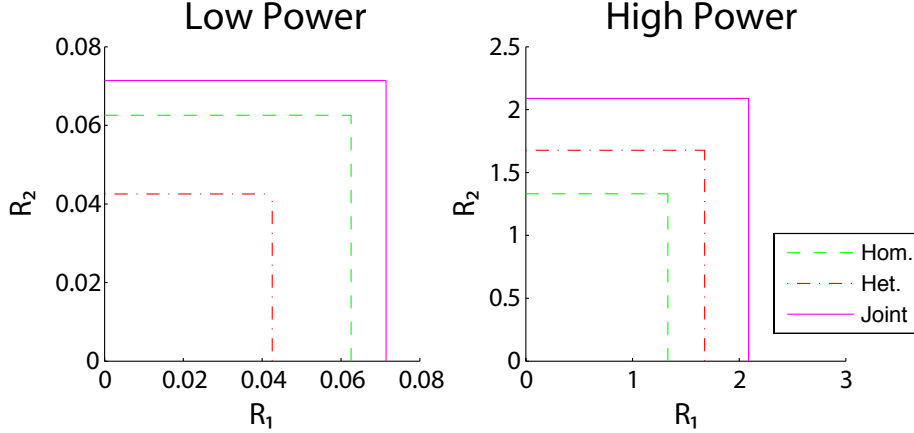


Figure 8.3: Capacity regions for coherent-state encodings and coherent detection, and achievable rate regions for coherent-state encodings and joint detection receivers—both with $\eta_{11} = \eta_{22} = 1/16$ and $\eta_{12} = \eta_{21} = 1/2$ (“very strong” interference for coherent detection). The LHS displays these regions in a low-power regime with $N_{S_1} = N_{S_2} = 1$ and $N_{B_1} = N_{B_2} = 1$, and the RHS displays these regions in a high-power regime where $N_{S_1} = N_{S_2} = 100$. Homodyne detection outperforms heterodyne detection in the low-power regime because it has a reduced detection noise, while heterodyne detection outperforms homodyne detection in the high-power regime because it has an increased bandwidth.

We can also consider the case when the senders employ coherent-state encodings and the receivers employ a joint detection strategy on all of their respective channel outputs. The conditions in (5.5) and (5.6) readily translate to this quantum setting where we now consider B_1 and B_2 to be quantum systems, and the information quantities in (5.5) and (5.6) become Holevo informations. The conditions in (5.5) and (5.6) when restricted to coherent-state encodings translate to:

$$\begin{aligned} g(\eta_{22} N_{S_2} + \bar{\eta}_2 N_{B_2}) - g(\bar{\eta}_2 N_{B_2}) &\leq g(\eta_{21} N_{S_2} + \eta_{11} N_{S_1} + \bar{\eta}_1 N_{B_1}) - g(\eta_{11} N_{S_1} + \bar{\eta}_1 N_{B_1}), \\ g(\eta_{11} N_{S_1} + \bar{\eta}_1 N_{B_1}) - g(\bar{\eta}_1 N_{B_1}) &\leq g(\eta_{12} N_{S_1} + \eta_{22} N_{S_2} + \bar{\eta}_2 N_{B_2}) - g(\eta_{22} N_{S_2} + \bar{\eta}_2 N_{B_2}). \end{aligned}$$

where $g(N) \equiv (N + 1) \log(N + 1) - N \log(N)$ is the entropy of a thermal state with

mean photon number N .

An achievable rate region is then

$$\begin{aligned} R_1 &\leq g(\eta_{11}N_{S_1} + \bar{\eta}_1N_{B_1}) - g(\bar{\eta}_1N_{B_1}), \\ R_2 &\leq g(\eta_{22}N_{S_2} + \bar{\eta}_2N_{B_2}) - g(\bar{\eta}_2N_{B_2}). \end{aligned}$$

These rates are achievable using a coherent-state encoding, but are not necessarily optimal (though they would be optimal if the minimum-output entropy conjecture from Refs. [GGL⁺04, GHLM10] were true). Nevertheless, these rates always beat the rates from homodyne and heterodyne detection. Figure 8.3 shows examples of the achievable rate regions for a bosonic interference channel with very strong interference. Both the low-power and high-power regimes are considered. Observe that the relative superiority of homodyne and heterodyne detection depend on power constraint and that the joint detection strategy always outperforms them.

8.5 Strong interference case

Sato [Sat81] determined the capacity of the classical Gaussian interference channel under “strong” interference. Theorem 5.2 from Chapter 5 gives us the capacity region for quantum interference channels with strong interference. We will now apply these results in the context of the bosonic interference channel.

The conditions for a channel to exhibit “strong” interference are given in equations (5.16) and (5.17), and they translate to the following ones for coherent-state encoding and coherent detection:

$$\frac{\eta_{21}}{\eta_{22}} \geq \frac{2^i \bar{\eta}_1 N_{B_1} + 1}{2^i \bar{\eta}_2 N_{B_2} + 1}, \quad \frac{\eta_{12}}{\eta_{11}} \geq \frac{2^i \bar{\eta}_2 N_{B_2} + 1}{2^i \bar{\eta}_1 N_{B_1} + 1},$$

and the capacity region becomes:

$$R_1 \leq \frac{1}{2^i} \log \left(1 + \frac{4^i \eta_{11} N_{S_1}}{2^i \bar{\eta}_1 N_{B_1} + 1} \right), \quad (8.33)$$

$$R_2 \leq \frac{1}{2^i} \log \left(1 + \frac{4^i \eta_{22} N_{S_2}}{2^i \bar{\eta}_2 N_{B_2} + 1} \right), \quad (8.34)$$

$$R_1 + R_2 \leq \frac{1}{2^i} \min \left\{ \begin{array}{l} \log \left(1 + 4^i \frac{\eta_{11} N_{S_1} + \eta_{21} N_{S_2}}{2^i \bar{\eta}_1 N_{B_1} + 1} \right), \\ \log \left(1 + 4^i \frac{\eta_{22} N_{S_2} + \eta_{12} N_{S_1}}{2^i \bar{\eta}_2 N_{B_2} + 1} \right) \end{array} \right\}, \quad (8.35)$$

where again $i = 1$ for homodyne detection and $i = 0$ for heterodyne detection.

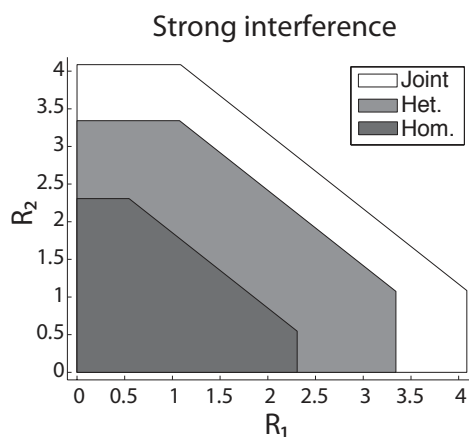


Figure 8.4: The figure depicts the “strong” interference capacity regions in the high-power regime for homodyne and heterodyne detection, and joint detection. The channel in the figure is in the high-power regime: $N_{B_1} = N_{B_2} = 1$, $\eta_{11} = \eta_{22} = 0.3$, $\eta_{21} = \eta_{12} = 0.6$, and $N_{S_1} = N_{S_2} = 100$. Heterodyne detection outperforms homodyne detection in this case.

We can also compute the achievable rate region using the joint detection strategy. Figure 8.4 displays the different capacity and achievable rate regions when a free-space interference channel exhibits “strong” interference.

8.6 Han-Kobayashi rate regions

The Han-Kobayashi rate region is the largest known achievable rate region for the classical interference channel [HK81]. The region was described in Theorem 5.3, and in Section 5.5 we established the achievability of the Chong-Motani-Garg, which is equivalent to the Han-Kobayashi rate region.

The Han-Kobayashi coding strategy readily translates into a strategy for coherent-state encoding and coherent detection. Sender m shares the total photon number N_{S_m} between her personal message and her common message. Let λ_m be the fraction of signal power that Sender m devotes to her personal message, and let $\bar{\lambda}_m \equiv (1 - \lambda_m)$ denote the remaining fraction of the signal power that Sender m devotes to her common message.

When Receiver 1 uses homodyne detection to decode the messages, we can identify the following components that are part of his received signal:

$$\lambda_1 \eta_{11} N_{S_1} = \text{power of own personal message}, \quad (8.36)$$

$$\bar{\lambda}_1 \eta_{11} N_{S_1} = \text{power of own common message}, \quad (8.37)$$

$$\eta_{11} N_{S_1} = \text{total own signal power}, \quad (8.38)$$

$$\eta_{21} N_{S_2} = \text{total interference power}, \quad (8.39)$$

$$\bar{\lambda}_2 \eta_{21} N_{S_2} = \text{useful part of interference (other's common)}, \quad (8.40)$$

$$\lambda_2 \eta_{21} N_{S_2} = \text{non-useful interference (other's personal)}, \quad (8.41)$$

$$N_1 = \frac{1}{4} (2\bar{\eta}_1 N_{B_1} + 1) = \text{noise power}, \quad (8.42)$$

Similar expressions exist for Receiver 2.

Consider now the inequalities (HK1)-(HK9) which define the Han-Kobayashi rate region (see page 87). When we evaluate each of the mutual informations for the signal and noise quantities (8.36) - (8.42), we obtain the Han-Kobayashi achievable rate region for the bosonic interference channel:

$$R_1 \leq \gamma \left(\frac{\eta_{11} N_{S_1}}{\lambda_2 \eta_{21} N_{S_2} + N_1} \right) \quad (\text{BHK1})$$

$$R_1 \leq \gamma \left(\frac{\lambda_1 \eta_{11} N_{S_1}}{\lambda_2 \eta_{21} N_{S_2} + N_1} \right) + \gamma \left(\frac{\bar{\lambda}_1 \eta_{12} N_{S_1}}{\lambda_1 \eta_{12} N_{S_1} + N_2} \right) \quad (\text{BHK2})$$

$$R_2 \leq \gamma \left(\frac{\eta_{22} N_{S_2}}{\lambda_1 \eta_{12} N_{S_1} + N_2} \right) \quad (\text{BHK3})$$

$$R_2 \leq \gamma \left(\frac{\lambda_2 \eta_{22} N_{S_2}}{\lambda_1 \eta_{12} N_{S_1} + N_2} \right) + \gamma \left(\frac{\bar{\lambda}_2 \eta_{21} N_{S_2}}{\lambda_2 \eta_{21} N_{S_2} + N_1} \right) \quad (\text{BHK4})$$

$$R_1 + R_2 \leq \gamma \left(\frac{\eta_{11} N_{S_1} + \bar{\lambda}_2 \eta_{21} N_{S_2}}{\lambda_2 \eta_{21} N_{S_2} + N_1} \right) + \gamma \left(\frac{\lambda_2 \eta_{22} N_{S_2}}{\lambda_1 \eta_{12} N_{S_1} + N_2} \right) \quad (\text{BHK5})$$

$$R_1 + R_2 \leq \gamma\left(\frac{\eta_{22}N_{S_2} + \bar{\lambda}_1\eta_{12}N_{S_1}}{\lambda_1\eta_{12}N_{S_1} + N_2}\right) + \gamma\left(\frac{\lambda_1\eta_{11}N_{S_1}}{\lambda_2\eta_{21}N_{S_2} + N_1}\right) \quad (\text{BHK6})$$

$$R_1 + R_2 \leq \gamma\left(\frac{\lambda_1\eta_{11}N_{S_1} + \bar{\lambda}_2\eta_{21}N_{S_2}}{\lambda_2\eta_{21}N_{S_2} + N_1}\right) + \gamma\left(\frac{\lambda_2\eta_{22}N_{S_2} + \bar{\lambda}_1\eta_{12}N_{S_1}}{\lambda_1\eta_{12}N_{S_1} + N_2}\right) \quad (\text{BHK7})$$

$$2R_1 + R_2 \leq \gamma\left(\frac{\eta_{11}N_{S_1} + \bar{\lambda}_2\eta_{21}N_{S_2}}{\lambda_2\eta_{21}N_{S_2} + N_1}\right) + \gamma\left(\frac{\lambda_1\eta_{11}N_{S_1}}{\lambda_2\eta_{21}N_{S_2} + N_1}\right) + \gamma\left(\frac{\lambda_2\eta_{22}N_{S_2} + \bar{\lambda}_1\eta_{12}N_{S_1}}{\lambda_1\eta_{12}N_{S_1} + N_2}\right) \quad (\text{BHK8})$$

$$R_1 + 2R_2 \leq \gamma\left(\frac{\eta_{22}N_{S_2} + \bar{\lambda}_1\eta_{12}N_{S_1}}{\lambda_1\eta_{12}N_{S_1} + N_2}\right) + \gamma\left(\frac{\lambda_2\eta_{22}N_{S_2}}{\lambda_1\eta_{12}N_{S_1} + N_2}\right) + \gamma\left(\frac{\lambda_1\eta_{11}N_{S_1} + \bar{\lambda}_2\eta_{21}N_{S_2}}{\lambda_2\eta_{21}N_{S_2} + N_1}\right) \quad (\text{BHK9})$$

Note the shorthand notation used $\gamma(x) = \frac{1}{2} \log_2(1 + x)$.

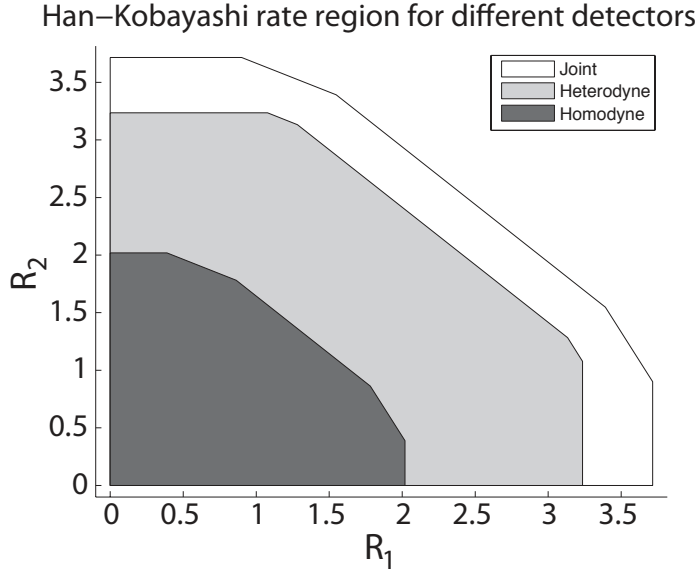


Figure 8.5: The figure depicts the achievable rate regions by employing a Han-Kobayashi coding strategy for homodyne and heterodyne detection. The channel parameters are $N_{S_1} = N_{S_2} = 100$, $N_{B_1} = N_{B_2} = 1$, $\eta_{11} = \eta_{22} = 0.8$, and $\eta_{21} = \eta_{12} = 0.1$. All of these regions are with respect to a 10%-personal, 90%-common Han-Kobayashi power split.

We can also calculate the shape of the Han-Kobayashi achievable rate region if the senders employ coherent-state encodings and the receivers exploit heterodyne or joint detection receivers. A statement of the inequalities for the other detection strategies has been omitted, because they are similar to (BHK1)-(BHK9). Figure 8.5 shows the relative sizes of the Han-Kobayashi rate regions achievable with coherent detection and joint detector for a particular choice of input power split: $\lambda_m = 0.1$, $\bar{\lambda}_m = 0.9$.

8.7 Discussion

The semiclassical models for free-space optical communication are not sufficient to understand the ultimate limits on reliable communication rates, for both point-to-point and multiuser bosonic channels. We presented a quantum-mechanical model for the free-space optical interference channel and determined achievable rate regions using three different decoding strategies for the receivers. We also determined the Han-Kobayashi inner bound for homodyne, heterodyne and joint detection.

Several open problems remain for this line of inquiry. We do not know if a coherent-state encoding is in fact optimal for the free-space interference channel—it might be that squeezed state transmitters could achieve higher communication rates as in [Yen05a]. One could also evaluate the ergodic and outage capacity regions based on the statistics of η_{ij} , which could be derived from the spatial coherence functions of the stochastic mode patterns under atmospheric turbulence.

Chapter 9

Conclusion

The time has come to conclude our inquiry into the problems of quantum network information theory. We will use this last chapter to summarize our results and highlight the specific contribution of this thesis. We will also discuss open problems and avenues for future research.

9.1 Summary

The present work demonstrates clearly that many of the problems of classical network information theory can be extended to the study of classical-quantum channels. Originally, we set out to investigate the network information theory problems discussed in [EGC80]. It is fair to say that we have been successful on that front, since we managed to develop coding strategies for multiple access channels (Chapter 4), interference channels (Chapter 5), broadcast channels (Chapter 6) and relay channels (Chapter 7), in the classical-quantum setting.

Our proof techniques are a mix of classical and quantum ideas. On the classical side we have the standard tools of information theory like averaging, conditional averaging and the use of the properties of typical sets. On the quantum side we saw how to build a *projector sandwich*, which contains many layers of conditionally typical projectors, how to incorporate *state smoothing*, which cuts out non-typical eigenvalues of a state, and the winning combination of the square root measurement and the Hayashi-Nagaoka operator inequality.

Above all, it is the quantum conditionally typical projectors that played the biggest role in all our results. Conditionally typical projectors are truly amazing constructs, since they not only give us a basis in terms of which to analyze the quantum outputs, but also tell us exactly in which subspace we are likely to find the output states on average.

9.2 New results

Some of the results presented in this thesis have previously appeared in publications and some are original to this thesis. We will use this section to highlight the new results.

The first contribution is the establishment of the classical/quantum packing lemmas using conditionally typical sets/projectors. While these packing lemmas are not new in themselves, the proofs presented highlight the correspondences between the indicator functions for the classical conditionally typical sets and, their quantum counterparts, the conditionally typical projectors. The quantum packing lemma is an effort to abstract away the details of the quantum decoding strategy into a reusable component as is done in [EGK10].

It is the author's hope that the classical and quantum packing lemmas presented in this work, along with their proofs, can serve as a bridge for classical information theorists to cross over to the quantum side. Alternately, we can say that there is only one side and interpret the move from classical Shannon theory to quantum Shannon theory as a type of system upgrade. Indeed, the change from indicator functions for the conditionally typical sets to conditionally typical projectors can be seen in terms of the OSI layered model for network architectures: quantum coding techniques are a change in *physical layer* (Layer 1) protocols while the random coding approach of the *data link layer* (Layer 2) stays the same. Note that this analogy only works for the *classical* communication problem, and that *quantum* communication and *entanglement-assisted* communication are completely new problems in quantum Shannon theory, which have no direct classical analogues.

The main original contribution of this thesis is the achievability proof for the quantum Chong-Motani-Garg rate region, which requires only two-sender simultaneous decoding. By the equivalence $\mathcal{R}_{\text{HK}}^o(\mathcal{N}) \equiv \mathcal{R}_{\text{CMG}}(\mathcal{N})$, we have established the

achievability of the quantum Han-Kobayashi rate region. We can therefore close the book on the original research question which prompted our investigation more than two years ago.

An interesting open problem is to prove Conjecture 4.1 on the simultaneous decoding for the three-sender quantum multiple access channels. This result would be a powerful building block for multiuser quantum Shannon theory.

Appendix A

Classical channel coding

This appendix contains the proof of the classical packing lemma (Section A.2) and a brief review on some of the properties of typical sets.

A.1 Classical typicality

In Section 2.2, we presented a number of properties of typical sequences and typical sets that were used in the proof of the classical coding theorem. The reader is invited to consult [CT91] and [Wil11] for the proofs.

In this section, we review the properties of conditionally typical sets in a more general setting where an additional random variable U^n is present. This is the setting of the classical packing lemma, which will be stated and proved in Section A.2.

Consider the probability distribution $p_U(u)p_{X|U}(x|u) \in \mathcal{P}(\mathcal{U}, \mathcal{X})$ and the channel $\mathcal{N} = (\mathcal{U} \times \mathcal{X}, p_{Y|XU}(y|x, u), \mathcal{Y})$. Let (U^n, X^n) be distributed according to the product distribution $\prod_{i=1}^n p_U(u_i)p_{X|U}(x_i|u_i)$. Let Y^n denote the random variable that corresponds to the output of the channel when the inputs are (U^n, X^n) .

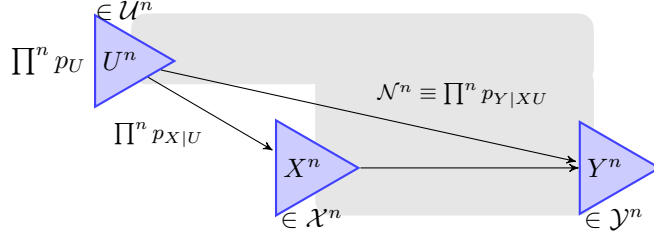


Figure A.1: An illustration of the conditional dependence between the random variables (U^n, X^n, Y^n) .

Conditionally typical sets

The input random variables $(U^n, X^n) \sim \prod_{i=1}^n p_U(u_i)p_{X|U}(x_i|u_i)$ and the channel \mathcal{N} induce the following joint distribution:

$$(U^n, X^n, Y^n) \sim \prod_{i=1}^n p_U(u_i)p_{X|U}(x_i|u_i)p_{Y|XU}(y_i|x_i, u_i). \quad (\text{A.1})$$

This corresponds to the assumption that the channel is memoryless, that is, the noise in the n uses of the channel is independent $p_{Y^n|X^n U^n} = \prod^n p_{Y|XU}$.

For any $\delta > 0$, define two sets of entropy conditionally typical sequences:

$$\mathcal{T}_\delta^{(n)}(Y|x^n, u^n) \equiv \left\{ y^n \in \mathcal{Y}^n : \left| -\frac{\log p_{Y^n|X^n U^n}(y^n|x^n, u^n)}{n} - H(Y|X, U) \right| \leq \delta \right\}, \quad (\text{A.2})$$

$$\mathcal{T}_\delta^{(n)}(Y|u^n) \equiv \left\{ y^n \in \mathcal{Y}^n : \left| -\frac{\log p_{Y^n|X^n}(y^n|u^n)}{n} - H(Y|U) \right| \leq \delta \right\}, \quad (\text{A.3})$$

where $H(Y|U) = -\sum_x p_U(u)p_{Y|U}(y|u) \log p_{Y|U}(y|u)$ is the conditional entropy of the distribution $p_{Y|U}(y|u) = \sum_x p_{X|U}(x|u)p_{Y|XU}(y|x, u)$.

By the definition of these typical sets, we have that the following bounds on the probability of the sequences within these sets:

$$2^{-n[H(Y|X,U)+\delta]} \leq p_{Y^n|X^n, U^n}(y^n|x^n, u^n) \leq 2^{-n[H(Y|X,U)-\delta]} \quad \forall y^n \in \mathcal{T}_\delta^{(n)}(Y|x^n, u^n),$$

$$2^{-n[H(Y|U)+\delta]} \leq p_{Y^n|U^n}(y^n|u^n) \leq 2^{-n[H(Y|U)-\delta]} \quad \forall y^n \in \mathcal{T}_\delta^{(n)}(Y|u^n), \quad (\text{A.4})$$

for any sequences u^n and x^n .

The channel outputs are likely to be conditionally typical sequences. More pre-

cisely, we have that for any $\epsilon, \delta > 0$, and sufficiently large n the expectations under U^n and $X^n|U^n$ obey the bounds:

$$\mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \sum_{y^n \in \mathcal{T}_\delta^{(n)}(Y|X^n, U^n)} p_{Y^n|X^n U^n}(y^n|X^n, U^n) \geq 1 - \epsilon, \quad (\text{A.5})$$

$$\mathbb{E}_{U^n} \sum_{y^n \in \mathcal{T}_\delta^{(n)}(Y|U^n)} p_{Y^n|U^n}(y^n|U^n) \geq 1 - \epsilon. \quad (\text{A.6})$$

Furthermore, we have the following bounds on the size of these conditionally typical sets:

$$\begin{aligned} \left| \mathcal{T}_\delta^{(n)}(Y|X^n, U^n) \right| &\leq 2^{n[H(Y|X, U) + \delta]}, \\ \left| \mathcal{T}_\delta^{(n)}(Y|U^n) \right| &\leq 2^{n[H(Y|U) + \delta]}. \end{aligned} \quad (\text{A.7})$$

Conditionally typical sets

Equations (A.4) and (A.7) will play a key role in the proof of the classical packing lemma in the next section. We restate these equations here in the language of indicator functions for the single and double conditionally typical sets:

$$p_{Y^n|U^n}(y^n|u^n) \mathbf{1}_{\{y^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n)\}} \leq 2^{-n[H(Y|U) - \delta]} \mathbf{1}_{\{y^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n)\}}, \quad (\text{A.4}')$$

and

$$\sum_{y^n \in \mathcal{Y}^n} \mathbf{1}_{\{y^n \in \mathcal{T}_\epsilon^{(n)}(Y|x^n, u^n)\}} \leq 2^{n[H(Y|X, U) + \delta]}. \quad (\text{A.7}')$$

A.2 Classical packing lemma

The packing lemma is a powerful tool for proving capacity theorems [EGK10]. We give a proof of a packing lemma which, instead of the usual jointly typical sequences argument, uses the properties of conditionally typical sets. This non-standard form of the packing lemmas is preferred because it highlights the similarities with its quantum analogue, the quantum conditional packing lemma stated in Appendix B.2.

Lemma A.1 (Classical conditional packing lemma). *Let $p_U(u)p_{X|U}(x|u) \in \mathcal{P}(\mathcal{U}, \mathcal{X})$ be an arbitrary code distribution, and let $\mathcal{N} = (\mathcal{U} \times \mathcal{X}, p_{Y|XU}(y|x, u), \mathcal{Y})$ be a channel. Let (U^n, X^n, \tilde{X}^n) be distributed according to $\prod_{i=1}^n p_U(u_i)p_{X|U}(x_i|u_i)p_{X|U}(\tilde{x}_i|u_i)$. Let \tilde{Y}^n denote the random variable that corresponds to the output of the channel when the inputs are (U^n, \tilde{X}^n) . Define \mathcal{E}_2 to be the event that the output \tilde{Y}^n will be part of the conditionally typical set $\mathcal{T}_\epsilon^{(n)}(Y|X^n, U^n)$, given that it is part of the output-typical set $\tilde{Y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n)$. We have that*

$$\begin{aligned} \mathbb{E}_{\substack{U^n, \\ X^n, \tilde{X}^n}} \Pr_{\tilde{Y}^n|X^n} \{\mathcal{E}_2\} &= \\ &= \mathbb{E}_{U^n} \mathbb{E}_{X^n} \mathbb{E}_{\tilde{X}^n} \Pr_{\tilde{Y}^n|X^n} \left\{ \left\{ \tilde{Y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X^n, U^n) \right\} \cap \left\{ \tilde{Y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n) \right\} \right\} \\ &\leq 2^{-n[I(X;Y|U) - \delta(\epsilon)]}. \end{aligned} \tag{A.8}$$

Consider the random codebook $\{X^n(m)\}$, $m \in [1 : 2^{nR}]$ generated randomly and independently according to $\prod_{i=1}^n p_{X|U}(x_i|u_i)$. There exists $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ such that the probability that the conditionally typical decoding will misinterpreting the channel output for $X^n(m)$ incorrectly as produced by $X^n(m')$ for some $m' \neq m$, that is,

$$Y^n(m) \equiv \mathcal{N}^n(U^n, X^n(m)), \quad Y^n(m) \in \mathcal{T}_\epsilon^{(n)}(Y|X^n(m'), U^n) \text{ and } Y^n(m) \in \mathcal{T}_\epsilon^{(n)}(Y|U^n),$$

vanishes as $n \rightarrow \infty$, if $R < I(X;Y|U) - \delta(\epsilon)$, where the mutual information is calculated on the induced joint probability distribution $(U, X, Y) \sim p_{UXY}(u, x, y) = p_{Y|XU}(y|x, u)p_{X|U}(x|u)p_U(u)$.

The description of the error event in the conditional packing lemma contains four sources of randomness. First we have $U^n \sim \prod^n p_U$, then there are two independent draws from $\prod^n p_{X|U}$ to produce X^n and \tilde{X}^n . Finally, the channel-randomness produces $\tilde{Y}^n = \mathcal{N}^n(U^n, \tilde{X}^n)$. The fact that \tilde{X}^n and X^n are conditionally independent given U^n implies that \tilde{Y}^n and X^n are also conditionally independent given U^n . The situation is illustrated in Figure A.2.

Proof. We give an argument based on the properties of the output-typical sequences and a cardinality bound on the conditionally typical sets. Assume that the output sequence $\tilde{Y}^n = \mathcal{N}^n(U^n, \tilde{X}^n)$ is output-typical ($\in \mathcal{T}_\epsilon^{(n)}(Y|U^n)$), and happens to also fall in the conditionally typical set for some other codeword $\mathcal{T}_\epsilon^{(n)}(Y|X^n U^n)$. This is

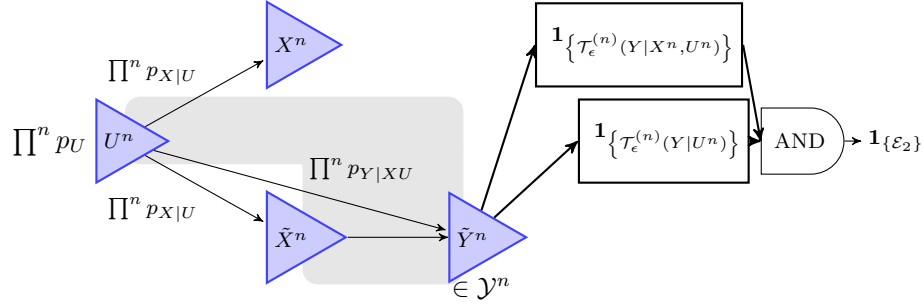


Figure A.2: The classical packing lemma. Two random codewords X^n and \tilde{X}^n are drawn randomly and independently conditional on a third random variable U^n . Assume that the random variable U^n is also available at the receiver. What is the chance that the output of the channel which corresponds to \tilde{X} and U^n will falsely be recognized to be in the set of outputs which are likely to come from inputs: X^n and U^n ? The receiver performs two tests on the output sequence \tilde{Y}^n : (1) test membership in $\mathcal{T}_\epsilon^{(n)}(Y|U^n)$ and (2) test membership in $\mathcal{T}_\epsilon^{(n)}(Y|X^n, U^n)$. If both these are successful, the outcome will be a *misidentification error* \mathcal{E}_2 .

described by the following event:

$$\mathcal{E}_2 = \left\{ \tilde{Y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X^n U^n) \right\} \cap \left\{ \tilde{Y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n) \right\}. \quad (\text{A.9})$$

Now consider the expectation of the probability of the event \mathcal{E}_2 under the code randomness:

$$\begin{aligned} & \mathbb{E}_{U^n} \mathbb{E}_{X^n} \mathbb{E}_{\tilde{X}^n} \Pr \{ \mathcal{E}_2 \} = \\ &= \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \mathbb{E}_{\tilde{X}^n|U^n} \Pr \left\{ \left\{ \tilde{Y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X^n U^n) \right\} \cap \left\{ \tilde{Y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n) \right\} \right\} \\ &= \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \mathbb{E}_{\tilde{X}^n|U^n} \mathbb{E}_{\tilde{Y}^n|\tilde{X}^n U^n} \mathbf{1}_{\{ \tilde{Y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X^n U^n) \}} \cdot \mathbf{1}_{\{ \tilde{Y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n) \}} \\ &= \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \sum_{\tilde{x}^n} \sum_{\tilde{y}^n} p_{X^n}(\tilde{x}^n|U^n) p_{Y^n|X^n U^n}(\tilde{y}^n|\tilde{x}^n, U^n) \mathbf{1}_{\{ \tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X^n U^n) \}} \mathbf{1}_{\{ \tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n) \}} \\ &\stackrel{\textcircled{1}}{=} \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \sum_{\tilde{y}^n} p_{Y^n|U^n}(\tilde{y}^n|U^n) \mathbf{1}_{\{ \tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X^n U^n) \}} \cdot \mathbf{1}_{\{ \tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n) \}} \\ &\stackrel{\textcircled{2}}{\leq} \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \sum_{\tilde{y}^n} 2^{-n[H(Y|U) - \delta'(\epsilon)]} \mathbf{1}_{\{ \tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X^n U^n) \}} \cdot \mathbf{1}_{\{ \tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|U^n) \}} \\ &\stackrel{\textcircled{3}}{\leq} 2^{-n[H(Y|U) - \delta'(\epsilon)]} \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \sum_{\tilde{y}^n} \mathbf{1}_{\{ \tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|X^n U^n) \}} \\ &= 2^{-n[H(Y|U) - \delta'(\epsilon)]} \sum_{u^n, x^n} p_{U^n X^n}(u^n, x^n) \sum_{\tilde{y}^n} \mathbf{1}_{\{ \tilde{y}^n \in \mathcal{T}_\epsilon^{(n)}(Y|x^n u^n) \}} \end{aligned}$$

A.2 Classical packing lemma

$$\begin{aligned}
&= 2^{-n[H(Y|U)-\delta'(\epsilon)]} \sum_{u^n, x^n} p_{U^n X^n}(u^n, x^n) \left| \mathcal{T}_\epsilon^{(n)}(Y|x^n u^n) \right| \\
&\stackrel{\textcircled{4}}{\leq} 2^{-n[H(Y|U)-\delta'(\epsilon)]} 2^{n[H(Y|XU)+\delta''(\epsilon)]} \\
&= 2^{-n[I(X;Y|U)-\delta(\epsilon)]}.
\end{aligned}$$

The equality ① follows from the definition of the conditional output distribution:

$$p_{Y^n|U^n}(\tilde{y}^n|u^n) = \sum_{\tilde{x}^n} p_{X^n|U^n}(\tilde{x}^n|u^n) p_{Y^n|X^n U^n}(\tilde{y}^n|\tilde{x}^n, u^n). \quad (\text{A.10})$$

Inequality ② follows from the fact that sequence \tilde{Y}^n is conditionally output-typical, which means that $p(y^n|u^n) \leq 2^{-n[H(Y|U)-\delta]}$. Inequality ③ is the consequence of dropping an indicator, since in this way we could only be enlarging the set. Inequality ④ follows from (A.4).

The second statement in the packing lemma follows from the independence of the codewords and the union bound. Let the random codebook $\{X^n(m)\}$, $m \in [1 : 2^{nR}]$ be generated randomly and independently according to $\prod_{i=1}^n p_{X|U}(x_i|U_i)$. Define $\{\mathcal{E}_2(m'|m)\}$ to be the event that the channel output when message m is sent, $Y^n(m) = \mathcal{N}^n(U^n, X^n(m))$ happens to fall in the conditionally typical set for some other codeword $\mathcal{T}_\epsilon^{(n)}(Y|X^n(m'), U^n)$ and is also output-typical ($\in \mathcal{T}_\epsilon^{(n)}(Y|U^n)$).

$$\mathcal{E}_2(m'|m) \equiv \{\{Y^n(m) \in \mathcal{T}_\epsilon^{(n)}(Y|X^n(m'), U^n)\} \cap \{Y^n(m) \in \mathcal{T}_\epsilon^{(n)}(Y|U^n)\}\}. \quad (\text{A.11})$$

If we define $(\mathbf{E2})$ to be the total probability of misidentifications of this kind, we get:

$$\begin{aligned}
\Pr\{(\mathbf{E2})\} &= \Pr\left\{ \bigcup_{m' \in \mathcal{M}, m' \neq m} \mathcal{E}_2(m'|m) \right\} \\
&\stackrel{\textcircled{5}}{\leq} \sum_{m' \in \mathcal{M}, m' \neq m} \Pr\{\mathcal{E}_2(m'|m)\} \\
&\stackrel{\textcircled{6}}{=} \sum_{m' \in \mathcal{M}, m' \neq m} \Pr\{\mathcal{E}_2\} \\
&\leq \sum_{m' \in \mathcal{M}, m' \neq m} 2^{-n[I(X;Y|U)-\delta(\epsilon)]} \\
&\leq |\mathcal{M}| 2^{-n[I(X;Y|U)-\delta(\epsilon)]}
\end{aligned}$$

$$= 2^{-n[I(X;Y|U)-R-\delta(\epsilon)]}.$$

Inequality ⑤ uses the union bound. Inequality ⑥ is true because the all the codewords of the codebook are picked independently.

Thus if we choose $R < I(X;Y) - \delta(\epsilon)$, the probability of error will tend to zero as $n \rightarrow \infty$. \square

The reader is now invited to review the Notation page (xi) in the beginning of the thesis. This table can be used as a bridge from classical information theory to the quantum information theory. In Appendix B, we will discuss the properties of conditionally typical projectors and prove a quantum packing lemma which follows *exactly* the same reasoning as in the classical packing lemma.

Appendix B

Quantum channel coding

The first part of this appendix defines the quantum typical subspaces and conditionally typical projectors associated with a quantum multiple access channel problem. The second part of the appendix is the statement of the *quantum packing lemma* which is a direct analogue of the classical packing lemma presented in Appendix A.2.

B.1 Quantum typicality

The concepts of entropy, and entropy-typical sets generalize to the quantum setting by virtue of the spectral theorem. Let \mathcal{H}^B be a d_B dimensional Hilbert space and let $\rho^B \in \mathcal{D}(\mathcal{H}^B)$ be the density matrix associated with a quantum state. The spectral decomposition of ρ^B is denoted $\rho^B = U\Lambda U^\dagger$ where Λ is a diagonal matrix of positive real eigenvalues that sum to one. We identify the eigenvalues of ρ^B with the probability distribution $p_Y(y) = \Lambda_{yy}$ and write the spectral decomposition as:

$$\rho^B = \sum_{y=1}^{d_B} p_Y(y) |e_{\rho;y}\rangle \langle e_{\rho;y}|^B \quad (\text{B.1})$$

where $|e_{\rho;y}\rangle$ is the eigenvector of ρ^B corresponding to eigenvalue $p_Y(y)$. The von Neumann entropy of the density matrix ρ^B is

$$H(B)_\rho = -\text{Tr}\{\rho^B \log \rho^B\} = H(p_Y). \quad (\text{B.2})$$

B.1 Quantum typicality

Define the set of δ -typical eigenvalues according to the eigenvalue distribution p_Y

$$\mathcal{T}_{p_Y, \delta}^n \equiv \left\{ y^n \in \mathcal{Y}^n : \left| -\frac{\log p_{Y^n}(y^n)}{n} - H(Y) \right| \leq \delta \right\}. \quad (\text{B.3})$$

For a given string $y^n = y_1 y_2 \dots y_i \dots y_n$ we define the corresponding eigenvector as

$$|e_{\rho; y^n}\rangle = |e_{\rho; y_1}\rangle \otimes |e_{\rho; y_2}\rangle \otimes \dots \otimes |e_{\rho; y_n}\rangle, \quad (\text{B.4})$$

where for each symbol where $y_i = b \in \{1, 2, \dots, d_B\}$ we select the b^{th} eigenvector $|e_{\rho; b}\rangle$.

The typical subspace associated with the density matrix ρ^B is defined as

$$A_{\rho, \delta}^n = \text{span}\{|e_{\rho; y^n}\rangle : y^n \in \mathcal{T}_{p_Y, \delta}^n\}. \quad (\text{B.5})$$

The typical projector is defined as

$$\Pi_{\rho^B, \delta}^n = \sum_{y^n \in \mathcal{T}_{p_Y, \delta}^n} |e_{\rho; y^n}\rangle \langle e_{\rho; y^n}|. \quad (\text{B.6})$$

Note that the typical projector is linked twofold to the spectral decomposition of (B.1): the sequences y^n are selected according to p_Y and the set of typical vectors are build from tensor products of orthogonal eigenvectors $|e_{\rho; y}\rangle$.

Properties analogous to (2.3) – (2.5) hold. For any $\epsilon, \delta > 0$, and all sufficiently large n we have

$$\text{Tr}\{\rho^{\otimes n} \Pi_{\rho, \delta}^n\} \geq 1 - \epsilon \quad (\text{B.7})$$

$$2^{-n[H(B)_\rho + \delta]} \Pi_{\rho, \delta}^n \leq \Pi_{\rho, \delta}^n \rho^{\otimes n} \Pi_{\rho, \delta}^n \leq 2^{-n[H(B)_\rho - \delta]} \Pi_{\rho, \delta}^n, \quad (\text{B.8})$$

$$[1 - \epsilon] 2^{n[H(B)_\rho - \delta]} \leq \text{Tr}\{\Pi_{\rho, \delta}^n\} \leq 2^{n[H(B)_\rho + \delta]}. \quad (\text{B.9})$$

The interpretation of (B.8) is that the eigenvalues of the state $\rho^{\otimes n}$ are bounded between $2^{-n[H(B)_\rho - \delta]}$ and $2^{-n[H(B)_\rho + \delta]}$ on the typical subspace $A_{\rho, \delta}^n$.

Signal states Consider now a set of quantum states $\{\rho_{x_a}\}$, $x_a \in \mathcal{X}$. We perform the spectral decomposition of each ρ_{x_a} to obtain

$$\rho_{x_a}^B = \sum_{y=1}^{d_B} p_{Y|X}(y|x_a) |e_{\rho_{x_a}; y}\rangle \langle e_{\rho_{x_a}; y}|^B, \quad (\text{B.10})$$

where $p_{Y|X}(y|x_a)$ is the y^{th} eigenvalue of $\rho_{x_a}^B$ and $|e_{\rho_{x_a};y}\rangle$ is the corresponding eigenvector.

We can think of $\{\rho_{x_a}\}$ as a classical-quantum (c - q) channel where the input is some $x_a \in \mathcal{X}$ and the output is the corresponding quantum state ρ_{x_a} . If the channel is memoryless, then for each input sequence $x^n = x_1 x_2 \cdots x_n$ we have the corresponding tensor product output state:

$$\rho_{x^n}^{B^n} = \rho_{x_1}^{B_1} \otimes \rho_{x_2}^{B_2} \otimes \cdots \otimes \rho_{x_n}^{B_n} = \bigotimes_{i=1}^n \rho_{x_i}^{B_i}. \quad (\text{B.11})$$

To avoid confusion with the indices, we use $i \in [n]$ to denote the index of a symbol x in the sequence x^n and $a \in [1, \dots, |\mathcal{X}|]$ to denote the different symbols in the alphabet \mathcal{X} .

Conditionally typical projector Consider the ensemble $\{p_X(x_a), \rho_{x_a}\}$. The choice of distributions induces the following classical-quantum state:

$$\rho^{XB} = \sum_{x_a} p_X(x_a) |x_a\rangle\langle x_a|^X \otimes \rho_{x_a}^B. \quad (\text{B.12})$$

We can now define the conditional entropy of this state as

$$H(B|X)_\rho \equiv \sum_{x_a \in \mathcal{X}} p_X(x_a) H(\rho_{x_a}), \quad (\text{B.13})$$

or equivalently, expressed in terms of the eigenvalues of the signal states, the conditional entropy becomes

$$H(B|X)_\rho \equiv H(Y|X) \equiv \sum_{x_a} p_X(x_a) H(Y|x_a), \quad (\text{B.14})$$

where $H(Y|x_a) = -\sum_y p_{Y|X}(y|x_a) \log p_{Y|X}(y|x_a)$ is the entropy of the eigenvalue distribution shown in (B.10).

We define the x^n -conditionally typical projector as follows:

$$\Pi_{\rho_{x^n}, \delta}^n = \sum_{y^n \in \mathcal{T}_{\rho_{x^n}, \delta}^n} |e_{\rho_{x^n}; y^n}\rangle\langle e_{\rho_{x^n}; y^n}|, \quad (\text{B.15})$$

where the set of conditionally typical eigenvalues $\mathcal{T}_{\rho_{x^n}, \delta}^n$ consists of all sequences y^n

B.1 Quantum typicality

which satisfy:

$$\mathcal{T}_{\rho_{x^n}^B, \delta}^n \equiv \left\{ y^n : \left| -\frac{\log p_{Y^n|X^n}(y^n|x^n)}{n} - H(Y|X) \right| \leq \delta \right\}, \quad (\text{B.16})$$

with $p_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n p_{Y|X}(y_i|x_i)$.

The states $|e_{\rho_{x^n}; y^n}\rangle$ are built from tensor products of eigenvectors for the individual signal states:

$$|e_{\rho_{x^n}; y^n}\rangle = |e_{\rho_{x_1}; y_1}\rangle \otimes |e_{\rho_{x_2}; y_2}\rangle \otimes \cdots \otimes |e_{\rho_{x_n}; y_n}\rangle,$$

where the string $y^n = y_1 y_2 \dots y_i \dots y_n$ varies over different choices of bases for \mathcal{H}^B . For each symbol $y_i = b \in \{1, 2, \dots, d_B\}$ we select $|e_{\rho_{x_a}; b}\rangle$: the b^{th} eigenvector from the eigenbasis of ρ_{x_a} corresponding to the letter $x_i = x_a \in \mathcal{X}$.

Analogous to the three properties (B.7), (B.8) and (B.9), the conditionally typical projector obeys:

$$\mathbb{E}_{X^n} \text{Tr} \left[\rho_{X^n}^B \Pi_{\rho_{X^n}, \delta}^n \right] \geq 1 - \epsilon \quad (\text{B.17})$$

$$2^{-n[H(B|X)_\rho + \delta]} \Pi_{\rho_{x^n}, \delta}^n \leq \Pi_{\rho_{x^n}, \delta}^n \rho_{x^n}^B \Pi_{\rho_{x^n}, \delta}^n \leq 2^{-n[H(B|X)_\rho - \delta]} \Pi_{\rho_{x^n}, \delta}^n, \quad (\text{B.18})$$

$$[1 - \epsilon] 2^{n[H(B|X)_\rho - \delta]} \leq \mathbb{E}_{X^n} \text{Tr} \left[\Pi_{\rho_{X^n}, \delta}^n \right] \leq 2^{n[H(B|X)_\rho + \delta]}. \quad (\text{B.19})$$

MAC code Consider now a quantum multiple access channel $(\mathcal{X}_1 \times \mathcal{X}_2, \rho_{x_1, x_2}^B, \mathcal{H}^B)$ and two input distributions p_{X_1} and p_{X_2} . Define the random codebooks $\{X_1^n(m_1)\}_{m_1 \in \mathcal{M}_1}$ and $\{X_2^n(m_2)\}_{m_2 \in \mathcal{M}_2}$ generated from the product distributions $p_{X_1^n}$ and $p_{X_2^n}$ respectively. The choice of distributions induces the following classical-quantum state $\rho^{X_1 X_2 B}$

$$\sum_{x_a, x_b} p_{X_1}(x_a) p_{X_2}(x_b) |x_a\rangle\langle x_a|^{X_1} \otimes |x_b\rangle\langle x_b|^{X_2} \otimes \rho_{x_a x_b}^B. \quad (\text{B.20})$$

and the averaged output states:

$$\bar{\rho}_{x_a} \equiv \sum_{x_b} p_{X_2}(x_b) \rho_{x_a, x_b}, \quad (\text{B.21})$$

$$\bar{\rho}_{x_b} \equiv \sum_{x_a} p_{X_1}(x_a) \rho_{x_a, x_b}, \quad (\text{B.22})$$

$$\bar{\rho} \equiv \sum_{x_a, x_b} p_{X_1}(x_a) p_{X_2}(x_b) \rho_{x_a, x_b}. \quad (\text{B.23})$$

The conditional quantum entropy $H(B|X_1X_2)_\rho$ is:

$$H(B|X_1X_2)_\rho = \sum_{x_a \in \mathcal{X}_1, x_b \in \mathcal{X}_2} p_{X_1}(x_a)p_{X_2}(x_b)H(\rho_{x_a, x_b}), \quad (\text{B.24})$$

and using the average states we define:

$$H(B|X_1)_\rho = \sum_{x_a \in \mathcal{X}_1} p_{X_1}(x_a)H(\bar{\rho}_{x_a}), \quad (\text{B.25})$$

$$H(B|X_2)_\rho = \sum_{x_b \in \mathcal{X}_2} p_{X_2}(x_b)H(\bar{\rho}_{x_b}), \quad (\text{B.26})$$

$$H(B)_\rho = H(\bar{\rho}). \quad (\text{B.27})$$

Similarly to equation (B.15) and for each message pair (m_1, m_2) we define the conditionally typical projector for the encoded state $\rho_{x_1^n(m_1)x_2^n(m_2)}^B$ to be $\Pi_{\rho_{x_1^n(m_1)x_2^n(m_2)}^B, \delta}^n$. From this point on, we will not indicate the messages m_1, m_2 explicitly, because the codewords are constructed identically for each message.

Analogous to (2.46), the following upper bound applies:

$$\mathbb{E}_{X_1^n X_2^n} \text{Tr}\{\Pi_{\rho_{X_1^n X_2^n}^B, \delta}^n\} \leq 2^{n[H(B|X_1X_2)_\rho + \delta]}, \quad (\text{B.28})$$

and we can also bound from below the eigenvalues of the state $\rho_{x_1^n x_2^n}^B$ as follows:

$$2^{-n[H(B|X_1X_2)_\rho + \delta]} \Pi_{\rho_{x_1^n x_2^n}^B, \delta}^n \leq \Pi_{\rho_{x_1^n x_2^n}^B, \delta} \rho_{x_1^n x_2^n}^B \Pi_{\rho_{x_1^n x_2^n}^B, \delta}^n \leq 2^{-n[H(B|X_1X_2)_\rho - \delta]} \Pi_{\rho_{x_1^n x_2^n}^B, \delta}^n. \quad (\text{B.29})$$

We define conditionally typical projectors for each of the averaged states:

$$\bar{\rho}_{x_1} \rightarrow \Pi_{\bar{\rho}_{x_1}^B, \delta}^n, \quad (\text{B.30})$$

$$\bar{\rho}_{x_2} \rightarrow \Pi_{\bar{\rho}_{x_2}^B, \delta}^n, \quad (\text{B.31})$$

$$\bar{\rho} \rightarrow \Pi_{\bar{\rho}^B, \delta}^n. \quad (\text{B.32})$$

These projectors obey the standard eigenvalue upper bounds when acting on the states

B.1 Quantum typicality

with respect to which they are defined:

$$2^{-n[H(B|X_1)_\rho+\delta]}\Pi_{\bar{\rho}_{x_1}^B,\delta}^n \leq \Pi_{\bar{\rho}_{x_1}^B,\delta}\bar{\rho}_{x_1}^n\Pi_{\bar{\rho}_{x_1}^B,\delta}^n \leq 2^{-n[H(B|X_1)_\rho-\delta]}\Pi_{\bar{\rho}_{x_1}^B,\delta}^n, \quad (\text{B.33})$$

$$2^{-n[H(B|X_2)_\rho+\delta]}\Pi_{\bar{\rho}_{x_2}^B,\delta}^n \leq \Pi_{\bar{\rho}_{x_2}^B,\delta}\bar{\rho}_{x_2}^n\Pi_{\bar{\rho}_{x_2}^B,\delta}^n \leq 2^{-n[H(B|X_2)_\rho-\delta]}\Pi_{\bar{\rho}_{x_2}^B,\delta}^n, \quad (\text{B.34})$$

$$2^{-n[H(B)_\rho+\delta]}\Pi_{\bar{\rho}^B,\delta}^n \leq \Pi_{\bar{\rho}^B,\delta}^n \bar{\rho}^B \Pi_{\bar{\rho}^B,\delta}^n \leq 2^{-n[H(B)_\rho-\delta]}\Pi_{\bar{\rho}^B,\delta}^n. \quad (\text{B.35})$$

We have the following bounds on the rank of the conditionally typical projectors:

$$\text{Tr}\{\Pi_{\bar{\rho}_{X_1}^B,\delta}^n\} \leq 2^{n[H(B|X_1)_\rho+\delta]}, \quad (\text{B.36})$$

$$\text{Tr}\{\Pi_{\bar{\rho}_{X_2}^B,\delta}^n\} \leq 2^{n[H(B|X_2)_\rho+\delta]}, \quad (\text{B.37})$$

$$\text{Tr}\{\Pi_{\bar{\rho}^B,\delta}^n\} \leq 2^{n[H(B)_\rho+\delta]}. \quad (\text{B.38})$$

The encoded state $\rho_{X_1^n X_2^n}^B$ is well supported by all the typical projectors on average:

$$\mathbb{E}_{X_1^n X_2^n} \left[\text{Tr}\{\Pi_{\bar{\rho}_{X_1^n X_2^n}^B,\delta}^n \rho_{X_1^n X_2^n}^B\} \right] \geq 1 - \epsilon, \quad (\text{B.39})$$

$$\mathbb{E}_{X_1^n X_2^n} \left[\text{Tr}\{\Pi_{\bar{\rho}_{X_1}^B,\delta}^n \rho_{X_1^n X_2^n}^B\} \right] \geq 1 - \epsilon, \quad (\text{B.40})$$

$$\mathbb{E}_{X_1^n X_2^n} \left[\text{Tr}\{\Pi_{\bar{\rho}_{X_2}^B,\delta}^n \rho_{X_1^n X_2^n}^B\} \right] \geq 1 - \epsilon, \quad (\text{B.41})$$

$$\mathbb{E}_{X_1^n X_2^n} \left[\text{Tr}\{\Pi_{\bar{\rho}^B,\delta}^n \rho_{X_1^n X_2^n}^B\} \right] \geq 1 - \epsilon. \quad (\text{B.42})$$

B.2 Quantum packing lemma

Lemma B.1. *Let $p_U(u)p_{X|U}(x|u) \in \mathcal{P}(\mathcal{U}, \mathcal{X})$ be an arbitrary code distribution, and let $\mathcal{N} = (\mathcal{U} \times \mathcal{X}, \rho_{u,x}, \mathcal{H}^B)$ be a classical-quantum channel. Let (U^n, X^n, \tilde{X}^n) be distributed according to $\prod_{i=1}^n p_U(u_i)p_{X|U}(x_i|u_i)p_{X|U}(\tilde{x}_i|u_i)$. Consider the channel \mathcal{N}' defined by the following map:*

$$\mathcal{N}' : (u^n, x^n) \rightarrow \left(u^n, \underbrace{\rho_{u_1, x_1}^{B_1} \otimes \rho_{u_2, x_2}^{B_2} \otimes \cdots \otimes \rho_{u_n, x_n}^{B_n}}_{\rho_{u^n, x^n}^{B^n}} \right), \quad (\text{B.43})$$

where u^n is available as side information to the receiver and the sender. Define the state $\bar{\rho}_{u^n} = \mathbb{E}_{X^n|u^n} \mathcal{N}'(u^n, X^n)$ and the conditionally typical projectors $\Pi_{\bar{\rho}_{u^n}}^{B^n}$ for the state $\bar{\rho}_{u^n}^{B^n}$ and $\Pi_{\rho_{u^n, x^n}^{B^n}}$ for the state $\rho_{u^n, x^n}^{B^n}$.

We want to measure the expectation of the overlap between $\rho_{U^n, \tilde{X}^n}^{B^n}$ and the operator $\Pi_{\bar{\rho}_{U^n}}^{B^n} \Pi_{\rho_{U^n, X^n}^{B^n}} \Pi_{\bar{\rho}_{U^n}}^{B^n}$ associated with some (U^n, X^n) . We define this quantity to be:

$$\mathcal{E}_2 = \text{Tr} \left[\Pi_{\bar{\rho}_{u^n}}^{B^n} \Pi_{\rho_{u^n, x^n}^{B^n}} \Pi_{\bar{\rho}_{u^n}}^{B^n} \rho_{U^n, \tilde{X}^n}^{B^n} \right]. \quad (\text{B.44})$$

Then \mathcal{E}_2 can be bounded as follows:

$$\mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \mathbb{E}_{\tilde{X}^n|U^n} \mathcal{E}_2 \leq 2^{-n[I(X;B|U) - \delta(\epsilon)]}. \quad (\text{B.45})$$

Let the random codebook $\{X^n(m)\}$, $m \in [1 : 2^{nR}]$ be generated randomly and independently according to $\prod_{i=1}^n p_{X|U}(x_i|U_i)$. Then there exists $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ such that the expectation of the total overlap between conditionally typical output spaces can be bounded from above as follows:

$$\begin{aligned} (\text{E2}) &\equiv \sum_{m' \in \mathcal{M}, m' \neq m} \mathbb{E}_{U^n} \mathbb{E}_{X^n(m)|U^n} \mathbb{E}_{X^n(m')|U^n} \text{Tr} \left[\Pi_{\bar{\rho}_{U^n}}^{B^n} \Pi_{\rho_{U^n, X^n(m')}}^{B^n} \Pi_{\bar{\rho}_{U^n}}^{B^n} \rho_{U^n, X^n(m)}^{B^n} \right] \\ &\leq |\mathcal{M}| 2^{-n[I(X;Y|U) - \delta(\epsilon)]}. \end{aligned} \quad (\text{B.46})$$

Thus if we choose $R < I(X;B|U) - \delta(\epsilon)$, the quantity (E2) will tend to zero as $n \rightarrow \infty$.

To bound the expectation of the second term, define $\tilde{X}(m)$ and $X^n(m')$ to be the two random codewords assigned to messages m and m' respectively.

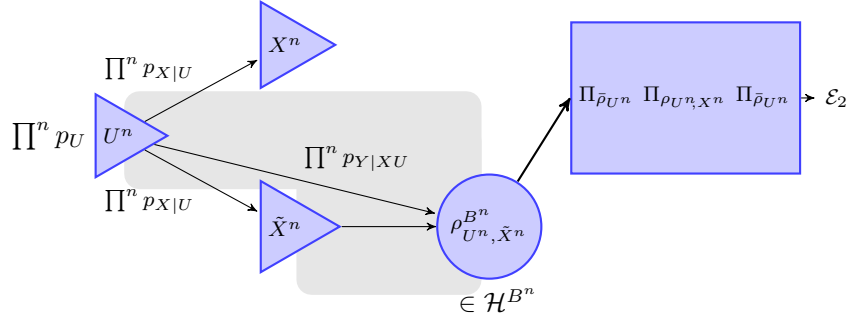


Figure B.1: The quantum packing lemma. Two random codewords X^n and \tilde{X}^n are drawn randomly and independently conditional on a third random variable U^n . Assume that the random variable U^n is also available at the receiver. What is the chance that the output of the channel which corresponds to \tilde{X} and U^n will falsely be recognized to be in the set of outputs which are likely to come from inputs X^n and U^n ?

$$\begin{aligned}
 \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \mathbb{E}_{\tilde{X}^n|U^n} \mathcal{E}_2 &= \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \mathbb{E}_{\tilde{X}^n|U^n} \text{Tr} \left[\Pi_{\bar{\rho}_{U^n}}^{B^n} \Pi_{\rho_{U^n, X^n}}^{B^n} \Pi_{\bar{\rho}_{U^n}}^{B^n} \rho_{U^n, \tilde{X}^n}^{B^n} \right] \\
 &= \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \text{Tr} \left[\Pi_{\bar{\rho}_{U^n}}^{B^n} \Pi_{\rho_{U^n, X^n}}^{B^n} \Pi_{\bar{\rho}_{U^n}}^{B^n} \mathbb{E}_{\tilde{X}^n|U^n} \{ \rho_{U^n, \tilde{X}^n}^{B^n} \} \right] \\
 &\stackrel{\textcircled{1}}{=} \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \text{Tr} \left[\Pi_{\bar{\rho}_{U^n}}^{B^n} \Pi_{\rho_{U^n, X^n}}^{B^n} \Pi_{\bar{\rho}_{U^n}}^{B^n} \bar{\rho}_{U^n} \right] \\
 &= \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \text{Tr} \left[\Pi_{\rho_{U^n, X^n}}^{B^n} \Pi_{\bar{\rho}_{U^n}}^{B^n} \bar{\rho}_{U^n} \Pi_{\bar{\rho}_{U^n}}^{B^n} \right] \\
 &\stackrel{\textcircled{2}}{\leq} 2^{-n[H(B|U)-\delta]} \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \text{Tr} \left[\Pi_{\rho_{U^n, X^n}}^{B^n} \Pi_{\bar{\rho}_{U^n}}^{B^n} \right] \\
 &\stackrel{\textcircled{3}}{\leq} 2^{-n[H(B|U)-\delta]} \mathbb{E}_{U^n} \mathbb{E}_{X^n|U^n} \text{Tr} \left[\Pi_{\rho_{U^n, X^n}}^{B^n} \right] \\
 &\stackrel{\textcircled{4}}{\leq} 2^{-n[H(B|U)-\delta]} 2^{n[H(B|U, X)+\delta]} \\
 &= 2^{-n[I(X; Y|U)-\delta(\epsilon)]}.
 \end{aligned}$$

Equation $\textcircled{1}$ is true by the definition $\mathbb{E}_{\tilde{X}^n|U^n} \{ \rho_{U^n, \tilde{X}^n}^{B^n} \} = \bar{\rho}_{U^n}$. The inequality $\textcircled{2}$ uses

the eigenvalue bound as in (B.18). The inequality ③ follows from

$$\begin{aligned} \mathrm{Tr} \left[\Pi_{\rho_{U^n, X^n}}^{B^n} \Pi_{\bar{\rho}_{U^n}}^{B^n} \right] &= \mathrm{Tr} \left[\Pi_{\rho_{U^n, X^n}}^{B^n} \Pi_{\bar{\rho}_{U^n}}^{B^n} \Pi_{\rho_{U^n, X^n}}^{B^n} \right] \\ &\leq \mathrm{Tr} \left[\Pi_{\rho_{U^n, X^n}}^{B^n} I \Pi_{\rho_{U^n, X^n}}^{B^n} \right] \\ &= \mathrm{Tr} \left[\Pi_{\rho_{U^n, X^n}}^{B^n} \right]. \end{aligned}$$

The inequality ④ follows from bound on the expected rank of the conditionally typical projector like in (B.19).

Applications

Holevo-Schumacher-Westmoreland (HSW) Theorem

Given a channel $(\mathcal{X}, \rho_x, \mathcal{H})$, if we set:

- $U = \emptyset$
- $p_U(u)p_{X|U}(x|u) = p_X(x)$
- $\rho_{u^n, x^n} = \rho_{x^n}$
- $\Pi_{\bar{\rho}_{u^n}}^{B^n} \Pi_{\rho_{u^n, x^n}}^{B^n} \Pi_{\bar{\rho}_{u^n}}^{B^n} = \Pi_{\bar{\rho}} \Pi_{\rho_{x^n}} \Pi_{\bar{\rho}}$,

then the quantum packing lemma tells us how many conditionally typical subspaces we can *pack* inside the output-typical subspace before they start to overlap too much.

Successive decoding for the quantum multiple access channel

Given a quantum multiple access channel $(\mathcal{X}_1 \times \mathcal{X}_2, \rho_{x_1, x_2}, \mathcal{H})$, we set:

- $U = X_1$
- $p_U(u)p_{X|U}(x|u) = p_{X_1}(x_1)p_{X_2}(x_2)$
- $\rho_{u^n, x^n} = \rho_{x_1^n, x_2^n}$
- $\Pi_{\bar{\rho}_{u^n}}^{B^n} \Pi_{\rho_{u^n, x^n}}^{B^n} \Pi_{\bar{\rho}_{u^n}}^{B^n} = \Pi_{\bar{\rho}_{x_1^n}} \Pi_{\rho_{x_1^n, x_2^n}} \Pi_{\bar{\rho}_{x_1^n}}$,

to obtain the bound on the rate R_2 when using the successive decoding $m_1 \rightarrow m_2|m_1$.

Superposition coding

Consider the situation in which superposition encoding is used to encode two messages ℓ and m in a codebook suitable for the channel $(\mathcal{X}, \rho_x, \mathcal{H})$:

$$\{W^n(\ell)\} \sim p_{W^n}(w^n), \quad \{X^n(\ell, m)\} \sim \prod_{i=1}^n p_{X|W}(x_i|w_i(\ell)).$$

Consider the following substitutions:

- $U = W$
- $p_U(u)p_{X|U}(x|u) = p_W(w)p_{X|W}(x|w)$
- $\rho_{u^n, x^n} = \rho_{x^n}$
- $\Pi_{\bar{\rho}_{u^n}}^{B^n} \Pi_{\rho_{u^n, x^n}}^{B^n} \Pi_{\bar{\rho}_{u^n}}^{B^n} = \Pi_{\bar{\rho}_{w^n}} \Pi_{\rho_{x^n}} \Pi_{\bar{\rho}_{w^n}}$.

The packing lemma gives us a bound on the error associated with decoding a wrong message m (the satellite message) given that we correctly decoded ℓ (the cloud center).

Appendix C

Miscellaneous proofs

This appendix contains a series of proofs which were omitted from the text in Section 5.4 in order to make it more readable.

C.1 Geometry of Chong-Motani-Garg rate region

We will now prove the inequalities from Lemma 5.2 on the geometry of $\mathcal{R}_{\text{CMG}}^1(\mathcal{N}, p_{\text{CMG}})$, the multiple access channel for Receiver 1 in the Chong-Motani-Garg coding strategy. This inequality structure is important for the geometrical observations of the Şaşoğlu argument.

Proof of Lemma 5.2. If we expand the shorthand notation of equations (5.30) through (5.32) we obtain the following inequalities.

$$I(X_1; B_1|W_1W_2Q) \leq I(X_1; B_1|W_2Q) \leq I(X_1W_2; B_1|Q), \quad (\text{C.1})$$

$$I(X_1; B_1|W_1W_2Q) \leq I(X_1W_2; B_1|W_1Q) \leq I(X_1W_2; B_1|Q), \quad (\text{C.2})$$

$$I(X_1; B_1|W_1W_2Q) + I(X_1W_2; B_1|Q) \leq I(X_1; B_1|W_2Q) + I(X_1W_2; B_1|W_1Q). \quad (\text{C.3})$$

Observe that W_2 is independent from W_1 and X_1 thus

$$H(X_1W_2) = H(X_1) + H(W_2), \quad H(W_1W_2) = H(W_1) + H(W_2). \quad (\text{C.4})$$

Also, since X_1 is obtained from W_1 , we have $H(X_1) = H(X_1W_1)$ and we can add or subtract the random variable W_1 next to X_1 as needed without changing the entropy.

To get the first part of the inequality (5.30), we observe

$$\begin{aligned}
 I(X_1; B_1|W_1W_2) &= I(X_1; B_1W_2|W_1) \\
 &= H(X_1W_1) + H(B_1W_2W_1) - H(X_1B_1W_2W_1) - H(W_1) \\
 &\quad - H(W_1W_2) + H(W_1W_2) \\
 &= H(X_1) + [H(B_1W_2W_1) - H(W_1W_2)] - H(X_1B_1W_2W_1) \\
 &\quad - H(W_1) + H(W_1) + H(W_2) \\
 &\leq H(X_1) + [H(B_1W_2) - H(W_2)] - H(X_1B_1W_2W_1) + H(W_2) \\
 &= [H(X_1) + H(W_2)] + H(B_1W_2) - H(X_1B_1W_2W_1) - H(W_2) \\
 &= I(X_1; B_1|W_2),
 \end{aligned}$$

where inequality follows from $H(B_1|W_1W_2) \leq H(B_1|W_2)$ (conditioning cannot increase entropy).

The second part of inequality (5.30), follows from a similar observation using $H(B_1|W_2) \leq H(B_1)$.

$$\begin{aligned}
 I(X_1; B_1|W_2) &= H(X_1W_2) + H(B_1W_2) - H(X_1B_1W_2) - H(W_2) \\
 &= H(X_1W_2) + [H(B_1W_2) - H(W_2)] - H(X_1B_1W_2) \\
 &\leq H(X_1W_2) + [H(B_1)] - H(X_1B_1W_2) \\
 &= I(X_1W_2; B_1).
 \end{aligned}$$

For the first part of (5.31) we repeat the above argument but with extra condi-

tioning on the W_1 system.

$$\begin{aligned}
I(X_1; B_1 | W_1 W_2) &= \\
&= H(X_1 W_1 W_2) + H(B_1 W_1 W_2) - H(X_1 B_1 W_1 W_2) - H(W_1 W_2) \\
&= H(X_1 W_1 W_2) + [H(B_1 W_1 W_2) - H(W_1 W_2)] - H(X_1 B_1 W_1 W_2) \\
&\leq H(X_1 W_2) + [H(B_1 | W_1)] - H(X_1 B_1 W_1 W_2) \\
&= H(X_1 W_2) + H(B_1 W_1) - H(X_1 B_1 W_1 W_2) - H(W_1) \\
&= H(X_1 W_1 W_2) + H(B_1 W_1) - H(X_1 B_1 W_1 W_2) - H(W_1) \\
&= I(X_1 W_2; B_1 | W_1).
\end{aligned}$$

For the second part of (5.31) we have

$$\begin{aligned}
I(X_1 W_2; B_1 | W_1) &= H(X_1 W_1 W_2) + H(B_1 W_1) - H(X_1 B_1 W_1 W_2) - H(W_1) \\
&= H(X_1 W_2) + [H(B_1 W_1) - H(W_1)] - H(X_1 B_1 W_2) \\
&\leq H(X_1 W_2) + H(B_1) - H(X_1 B_1 W_2) \\
&= I(X_1 W_2; B_1).
\end{aligned}$$

Finally for inequality (5.32) we need to use the strong subadditivity relation

$$H(B_1 W_1 W_2) + H(B_1) \leq H(B_1 W_1) + H(B_1 W_2). \quad (\text{C.5})$$

The steps are

$$\begin{aligned}
I(X_1; B_1 | W_1 W_2) + I(X_1 W_2; B_1) &= \\
&= H(X_1 W_1 W_2) + H(B_1 W_1 W_2) - H(X_1 B_1 W_1 W_2) - H(W_1 W_2) \\
&\quad + H(X_1 W_2) + H(B_1) - H(X_1 B_1 W_2) \\
&= [H(B_1 W_1 W_2) + H(B_1)] + H(X_1 W_1 W_2) - H(X_1 B_1 W_1 W_2) - H(W_1) - H(W_2) \\
&\quad + H(X_1 W_2) - H(X_1 B_1 W_2) \\
&\leq [H(B_1 W_1) + H(B_1 W_2)] + H(X_1 W_1 W_2) - H(X_1 B_1 W_1 W_2) - H(W_1) - H(W_2) \\
&\quad + H(X_1 W_2) - H(X_1 B_1 W_2) \\
&= H(X_1 W_1 W_2) + H(B_1 W_1) - H(X_1 B_1 W_1 W_2) - H(W_1) \\
&\quad + H(X_1 W_2) + H(B_1 W_2) - H(X_1 B_1 W_2) - H(W_2) \\
&= I(X_1 W_2; B_1 | W_1) + I(X_1; B_1 | W_2).
\end{aligned}$$

This completes the proof of Lemma 5.2. \square

C.2 Detailed explanation concerning moving points

In Section 5.5.2 we used Lemma 5.3 to show that we can move any point on the (b) or (d) planes to an equivalent point on the (a) or (c) planes. We now give the proof.

Proof. We have to show how to move any point in $b_i \cup d_i \setminus a_i \cup c_i$ to an equivalent point in $a_i \cup c_i$. Because the rates R_{1c} and R_{2c} appear in the coordinates of both P_1 and P_2 , we cannot move each point independently. Indeed Şaşoğlu points out that the points P_1 and P_2 are *coupled* by the common rates.

A priori, we have to consider all possible starting combinations the points However, using the following observations we can restrict the number of possibilities significantly.

1. If $P_1 \in b_1 \setminus a_1$, then $P_2 \in a_2 \cup b_2$.

The fact that $P_1 \in b_1 \setminus a_1$ implies that equation (b1) is tight

$$R_{1p} + R_{1c} = I(b_1), \quad (\text{C.6})$$

and (a1) is loose

$$R_{1p} < I(a_1). \quad (\text{C.7})$$

Then there exists $\delta > 0$ such that the point $P'_1 = (R_{1p} + \delta, R_{1c} - \delta, R_{2c}) \in \mathcal{R}_{\text{CMG}}^1(p)$. Suppose for a contradiction that P_2 was originally in $(c_2 \cup d_2) \setminus (a_1 \cup b_1)$. The decrease in R_{1c} associated with the move from P_1 to P'_1 , will have allowed us to increase the one of the rates for Receiver 2 which is a contradiction since we assumed the $R_2 = R_{2c} + R_{2p}$ was optimal. More specifically, if $P_2 \in c_2$, or $P_2 \in d_2$, then we would be allowed to increase R_{2p} by δ , to obtain $P'_2 = (R_{2p} + \delta, R_{2c}, R_{1c} - \delta)$, resulting in the operating point $(R_1, R_2 + \delta)$ which contradicts the assumption that the initial rate pair (R_1, R_2) was on the boundary of \mathcal{R}_{CMG} . Thus, if $P_1 \in b_1 \setminus a_1$, then P_2 must be in $a_2 \cup b_2$.

2. If $P_1 \in d_1 \setminus (a_1 \cup b_1 \cup c_1)$ then $P_2 \in a_2$.

Again consider moving the rates to obtain $P'_1 = (R_{1p} + \delta, R_{1c} - \delta, R_{2c}) \in d_1 \setminus (a_1 \cup$

$b_1 \cup c_1$), then if then if P_2 was originally in c_2 or d_2 , then the decrease in R_{1c} would allow us to move the point P_2 to a new rate triple $P'_2 = (R_{2p} + \delta, R_{2c}, R_{1c} - \delta)$, resulting in the operating point $(R_1, R_2 + \delta)$, which again leads to a contradiction. Therefore P_2 can only be in a_2 or b_2 . But if P_2 were in b_2 , then by observation 1 (with a change of roles between P_1 and P_2) we would have $P_1 \in (a_1 \cup b_1)$ which contradicts our assumption that $P_1 \in d_1 \setminus (a_1 \cup b_1 \cup c_1)$. Thus we see that if $P_1 \in d_1 \setminus (a_1 \cup b_1 \cup c_1)$, then $P_2 \in a_2$.

By the above reasoning we have restricted the possible combinations where the points (P_1, P_2) could lie initially. To prove Theorem 5.3, we have to show that we can deal with the following combinations: $b_1 \times a_2$, $a_1 \times b_2$, $b_1 \times b_2$, $d_1 \times a_1$ and $a_1 \times d_2$.

We now show that we can move any point $P_1 \in b_1 \cup d_1$ (on one of the bad planes) to an equivalent point lying in $a_1 \cup c_1$,

- **Case $(P_1, P_2) \in b_1 \times a_2$:**

In this case, equations (b1) and (a2) are tight which means that the rate pairs are of the form

$$\begin{aligned} P_1 &= (R_{1p}, R_{1c}, R_{2c}), \text{ such that } R_{1p} + R_{1c} = I(b_1), \\ P_2 &= (R_{2p}, R_{2c}, R_{1c}) = (I(a_2), R_{2c}, R_{1c}). \end{aligned}$$

If we apply a $R_{1c} \rightarrow R_{1p}$ rate moving operation to P_1 we can obtain a new point P'_1 with

$$P'_1 = (R'_{1p}, R'_{1c}, R_{2c}) = (I(a_1), I(b_1) - I(a_1), R_{2c}) \in a_1 \cap b_1.$$

As a result of the moving the point P_2 will be moved to

$$P'_2 = (R_{2p}, R_{2c}, R'_{1c}) = (I(a_2), R_{2c}, I(b_1) - I(a_1)),$$

which continues to lie in the a_2 plane. Observe that during this rate moving operation the sum rates remain unchanged $(R_{1p} + R_{1c}, R_{2p} + R_{2c}) = (R_1, R_2) = (R'_{1p} + R'_{1c}, R'_{2p} + R'_{2c})$.

The case when $(P_1, P_2) \in a_1 \times b_2$ is analogous.

- **Case $(P_1, P_2) \in b_1 \times b_2$:**

Our starting points are

$$\begin{aligned} P_1 &= (R_{1p}, R_{1c}, R_{2c}), \text{ such that } R_{1p} + R_{1c} = I(b_1), \\ P_2 &= (R_{2p}, R_{2c}, R_{1c}), \text{ such that } R_{2p} + R_{2c} = I(b_2). \end{aligned}$$

We will first do a $R_{1c} \rightarrow R_{1p}$ rate moving operation until we get to the plane a_1 . The points we obtain are

$$\begin{aligned} P'_1 &= (R'_{1p}, R'_{1c}, R_{2c}) = (I(a_1), I(b_1) - I(a_1), R_{2c}) \in a_1 \cap b_1, \\ P'_2 &= (R_{2p}, R_{2c}, R'_{1c}) = (R_{2p}, R_{2c}, I(b_1) - I(a_1)) \in b_2. \end{aligned}$$

We then perform second rate moving operation $R_{2c} \rightarrow R_{2p}$ in order to move to the plane a_2 .

$$\begin{aligned} P''_1 &= (R'_{1p}, R'_{1c}, R''_{2c}) = (I(a_1), I(b_1) - I(a_1), I(b_2) - I(a_2)) \in a_1 \cap b_1, \\ P''_2 &= (R''_{2p}, R''_{2c}, R'_{1c}) = (I(a_2), I(b_2) - I(a_2), I(b_1) - I(a_1)) \in a_1 \cap b_2. \end{aligned}$$

Thus we have managed to move the points $(P_1, P_2) \in b_1 \times b_2$ to equivalent points $(P''_1, P''_2) \in a_1 \times a_2$ while leaving the sum rate (R_1, R_2) unchanged.

- **Case** $(P_1, P_2) \in d_1 \times a_2$:

If $P_1 \in d_1$, it means that the triple sum inequality (d1) is tight. The starting rates are

$$\begin{aligned} P_1 &= (R_{1p}, R_{1c}, R_{2c}), \text{ such that } R_{1p} + R_{1c} + R_{2c} = I(d_1), \\ P_2 &= (I(a_2), R_{2c}, R_{1c}) \in a_2. \end{aligned}$$

To move P_1 away from the interior of the d_1 plane we will once again use a rate moving operation $R_{1c} \rightarrow R_{1p}$. This operation will increase the rate R_{1p} at the expense of the rate R_{1c} . We cannot increase the rate R_{1p} indefinitely – sooner or later one of the two other rate constraints on R_{1p} will saturate.

The other constraints on R_{1p} come from equations (a1) and (c1), so by rate moving we will eventually reach either the a_1 or the c_1 planes.

If the first case the resulting points will be

$$\begin{aligned} P'_1 &= (R'_{1p}, R'_{1c}, R_{2c}) = (I(a_1), R'_{1c}, R_{2c}) \in a_1 \cap d_1, \\ P'_2 &= (I(a_2), R_{2c}, R'_{1c}) \in a_2, \end{aligned}$$

where $R'_{1c} = I(d_1) - I(a_1) - R_{2c}$ because by rate moving we stayed in the d_1 plane.

In the latter case where moving the rates of $P_1 \in d_1$ puts us on the c_1 plane the resulting points will be

$$\begin{aligned} P'_1 &= (R'_{1p}, R'_{1c}, R_{2c}) \in c_1 \cap d_1, \text{ s.t. } R'_{1p} + R_{2c} = I(c_1) \\ P'_2 &= (I(a_2), R_{2c}, R'_{1c}) \in a_2. \end{aligned}$$

Once again, the sum rate (R_1, R_2) remains unchanged by the rate moving, but the moved points (P'_1, P'_2) are now either in $a_1 \times a_2$ or $c_1 \times a_2$ as claimed.

The case when $(P_1, P_2) \in a_1 \times d_2$ is analogous.

Therefore, given an arbitrary point $(R_1, R_2) \in \partial\mathcal{R}_{\text{CMG}}(\mathcal{N}, p_{\text{CMG}})$, there always exists a choice of common/private rates such that $(P_1, P_2) \in a_1 \cup c_1 \times a_2 \cup c_2$ with $(R_{1p} + R_{1c}, R_{2p} + R_{2c}) = (R_1, R_2)$.

□

C.3 Redundant inequality

In Section 5.5.3, we claimed that the inequality (5.49) is less tight than the sum rate constraint obtained by adding equations (5.48) and (5.51).

To that this is true, consider the following argument starting from the positivity of the mutual information $I(W_1; W_2|B_1) \geq 0$:

$$H(W_1W_2B_1) + H(B_1) \leq H(W_1B_1) + H(W_2B_1). \tag{C.8}$$

We now add $H(X_1W_1W_2)$ and subtract $-H(X_1W_1W_2B_1)$ on both sides of the equation:

$$\begin{aligned} H(W_1W_2B_1) + H(B_1) + H(X_1W_1W_2) &\leq H(W_1B_1) + H(W_2B_1) + H(X_1W_1W_2) \\ -H(X_1W_1W_2B_1) &\qquad\qquad\qquad -H(X_1W_1W_2B_1). \end{aligned}$$

We now use the fact that W_2 is independent from W_1 , so $H(W_1) - H(W_1W_2) = -H(W_2)$ to obtain:

$$\begin{aligned} H(W_1W_2B_1) + H(B_1) + H(X_1W_1W_2) &\leq H(W_1B_1) + H(W_2B_1) + H(X_1W_1W_2) \\ -H(X_1W_1W_2B_1) + H(W_1) - H(W_1W_2) &\leq -H(X_1W_1W_2B_1) - H(W_2). \end{aligned}$$

We move the term $H(W_1B_1)$ to the other side and rearrange the terms the final expression:

$$\begin{aligned} H(X_1W_1W_2) + H(W_1W_2B_1) - H(X_1W_1W_2B_1) - H(W_1W_2) \\ + H(W_1) + H(B_1) - H(W_1B_1) &\leq H(X_1W_1W_2) + H(W_2B_1) \\ &\leq -H(X_1W_1W_2B_1) - H(W_2) \end{aligned}$$

$$I(a_1) = I(X_1; B_1 | W_1W_2) + I(W_1; B_1) \leq I(X_1; B_1 | W_2) = I(b_1),$$

which shows that we can drop the constraint from equation (5.49).

Bibliography

- [ADHW09] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: Restructuring quantum information’s family tree. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 465(2108):2537–2563, 2009. arXiv:quant-ph/0606225.
- [Ahl71] R. Ahlswede. Multi-way communication channels. In *Proc. 2nd Int. Symp. Information Theory*, pages 23–52, 1971.
- [Ahl74a] R. Ahlswede. The capacity region of a channel with two senders and two receivers. *The Annals of Probability*, 2(5):805–814, 1974.
- [Ahl74b] R. Ahlswede. The capacity region of a channel with two senders and two receivers. *The Annals of Probability*, 2(5):805–814, 1974.
- [AHS08] D. Avis, P. Hayden, and I. Savov. Distributed compression and multiparty squashed entanglement. *Journal of Physics A: Mathematical and Theoretical*, 41:115301, 2008. arXiv:0707.2792.
- [BBC⁺93] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895–1899, March 1993.
- [BBJ11] I. Bjelakovic, H. Boche, and G. Janssen. Universal quantum state merging. In *IEEE International Symposium on Information Theory Proceedings.*, pages 99–103. IEEE, 2011.
- [Ber73] P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Transactions on Information Theory*, 19(2):197–207, March 1973.
- [BSST99] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.*, 83:3081, 1999. arXiv:quant-ph/9904023.
- [BW92] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69(20):2881–2884, November 1992.

- [Car75] A. B. Carleial. A case where interference does not reduce capacity. *IEEE Transactions on Information Theory*, 21:569, 1975.
- [Car78] A. B. Carleial. Interference channels. *IEEE Transactions on Information Theory*, 24(1):60–70, 1978.
- [Car82] A. B. Carleial. Multiple-access channels with different generalized feedback signals. *IEEE Trans. Inf. Theory*, 28(6):841–850, 1982.
- [Car83] A. B. Carleial. Outer bounds on the capacity of interference channels (Corresp.). *IEEE Transactions on information theory*, 29(4):602–606, 1983.
- [CEG79] T. Cover and A. El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory*, 25(5):572–584, 1979.
- [CEG87] M. H. M. Costa and A. El Gamal. The capacity region of the discrete memoryless interference channel with strong interference. *IEEE Transactions on Information Theory*, 33(5):710–711, 1987.
- [CGDR05] D. Collins, N. Gisin, and H. De Riedmatten. Quantum relays for long distance quantum cryptography. *Journal of Modern Optics*, 52(5):735–753, 2005.
- [CMG06] H.-F. Chong, M. Motani, and H. K. Garg. A comparison of two achievable rate regions for the interference channel. In *Proceedings of the USCD-ITA Workshop*, San Diego, California, USA, February 2006.
- [CMGEG08] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal. On the Han-Kobayashi region for the interference channel. *IEEE Transactions on Information Theory*, 54(7):3188–3195, 2008.
- [Cov72] T. M. Cover. Broadcast channels. *IEEE Trans. Inf. Theory*, 18(1):2–14, January 1972.
- [Cov98] T. M. Cover. Comments on broadcast channels. *IEEE Transactions on Information Theory*, 44(6):2524–2530, 1998.
- [CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.
- [DHL10] F. Dupuis, P. Hayden, and K. Li. A father protocol for quantum broadcast channels. *IEEE Trans. Inf. Theory*, 56(6):2946–2956, June 2010.
- [DHW08] I. Devetak, A. W. Harrow, and A. J. Winter. A resource framework for quantum shannon theory. *IEEE Transactions on Information Theory*, 54(10):4587–4618, 2008. arXiv:quant-ph/0512015.
- [DHW11] N. Datta, M. H. Hsieh, and M. M. Wilde. Quantum rate distortion, reverse shannon theorems, and source-channel separation. *arXiv:1108.4940*, 2011.
- [DL70] E. Davies and J. Lewis. An operational approach to quantum probability. *Commun. Math. Phys.*, 17:239–260, 1970.

- [Dup10] F. Dupuis. *The decoupling approach to quantum information theory*. PhD thesis, Université de Montréal, 2010. arXiv:1004.1641v1.
- [Dut11a] N. Dutil. *Multiparty quantum protocols for assisted entanglement distillation*. PhD thesis, McGill University, May 2011. arXiv:1105.4657.
- [Dut11b] N. Dutil. Multiparty quantum protocols for assisted entanglement distillation. *arXiv:1105.4657*, 2011.
- [EGC80] A. El Gamal and T. M. Cover. Multiple user information theory. *Proceedings of the IEEE*, 68(12):1466–1483, 1980.
- [EGK10] A. El Gamal and Y. H. Kim. Lecture notes on network information theory. January 2010. arXiv:1001.3404v4.
- [EGK11] A. El Gamal and Y. H. Kim. *Network Information Theory*. 2011.
- [ETW07] R. H. Etkin, D. N. C. Tse, and H. Wang. Gaussian interference channel capacity to within one bit: The general case. In *IEEE International Symposium on Information Theory*, pages 2181–2185, 2007.
- [Fei54] A. Feinstein. A new basic theorem of information theory. *Information Theory, IRE Professional Group on*, 4(4):2–22, 1954.
- [FHS⁺11] O. Fawzi, P. Hayden, I. Savov, P. Sen, and M. M. Wilde. Quantum interference channels. In *Proceedings of the Forty-Ninth Annual Allerton Conference*, pages 609–616. IEEE, 2011. arXiv:1102.2955.
- [FHS⁺12] O. Fawzi, P. Hayden, I. Savov, P. Sen, and M. M. Wilde. Classical communication over a quantum interference channel. *IEEE Transactions on Information Theory*, 58(6):3670–3691, June 2012. arXiv:1102.2624.
- [FS12] O. Fawzi and I. Savov. Rate-splitting in the presence of multiple receivers. June 2012. arXiv:1207.0543.
- [Gam] A. El Gamal. (personal communication at ISIT 2011).
- [GC02] G. Ginis and J. M. Cioffi. Vectored transmission for digital subscriber line systems. *IEEE Journal on Selected Areas in Communications*, 20(5):1085–1104, 2002.
- [GGL⁺04] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen. Classical capacity of the lossy bosonic channel: The exact solution. *Phys. Rev. Lett.*, 92(2):027902, January 2004.
- [GHLM10] V. Giovannetti, A. S. Holevo, S. Lloyd, and L. Maccone. Generalized minimal output entropy conjecture for one-mode Gaussian channels: definitions and some exact results. *Journal of Physics A: Mathematical and Theoretical*, 43(41):415305, 2010.
- [GK95] R. M. Gagliardi and S. Karp. *Optical Communications*. John Wiley and Sons, second edition, 1995.

- [GLM12] V. Giovannetti, S. Lloyd, and L. Maccone. Achieving the holevo bound via sequential measurements. *Phys. Rev. A*, 85:012302, Jan 2012. arXiv:1012.0386.
- [GRUW01] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting. Rate-splitting multiple access for discrete memoryless channels. *IEEE Transactions on Information Theory*, 47(3):873–890, 2001.
- [GSE07] S. Guha, J. H. Shapiro, and B. I. Erkmen. Classical capacity of bosonic broadcast communication and a minimum output entropy conjecture. *Physical Review A*, 76(3):032303, 2007.
- [GSW11] S. Guha, I. Savov, and M. M. Wilde. The free space optical interference channel. *IEEE International Symposium on Information Theory*, August 2011. arXiv:1102.2627.
- [Guh11] S. Guha. Structured optical receivers to attain superadditive capacity and the holevo limit. *Physical Review Letters*, 106(24):240502, 2011. arXiv:1101.1550.
- [HDW08] M.-H. Hsieh, I. Devetak, and A. Winter. Entanglement-assisted capacity of quantum multiple-access channels. *IEEE Transactions on Information Theory*, 54(7):3078–3090, 2008.
- [HK81] T.-S. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Transactions on Information Theory*, 27(1):49–60, Jan 1981.
- [HK07] T.-S. Han and K. Kobayashi. A further consideration on the HK and the CMG regions for the interference channel. In *Proc. Inf. Theory Applicat. Workshop*, 2007.
- [HN03] M. Hayashi and H. Nagaoka. General formulas for capacity of classical-quantum channels. *IEEE Transactions on Information Theory*, 49(7):1753–1768, 2003.
- [Hol73] A.S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973.
- [Hol79] A.S. Holevo. On capacity of a quantum communications channel. *Problemy Peredachi Informatsii*, 15(4):3–11, 1979.
- [Hol98] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory*, 44(1):269–273, 1998. arXiv:quant-ph/9611023.
- [KM77] J. Korner and K. Marton. General broadcast channels with degraded message sets. *IEEE Transactions on Information Theory*, 23(1):60–64, 1977.
- [Kra06] G. Kramer. Review of rate regions for interference channels. *International Zurich Seminar on Communications*, pages 162–165, 2006.
- [Lia72] H. Liao. *Multiple access channels*. PhD thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.
- [LR73] E. H. Lieb and M. B. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14:1938–1941, 1973.

- [LS11] M. S. Leifer and R. W. Spekkens. Formulating quantum theory as a causally neutral theory of bayesian inference, 2011. arXiv:1107.5849.
- [Mar79] K. Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, 25(3):306–311, 1979.
- [ON99] T. Ogawa and H. Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Transactions on Information Theory*, 45(7):2486–2489, 1999.
- [OSC⁺10] V. Oksman, H. Schenk, A. Clausen, J. M. Cioffi, M. Mohseni, G. Ginis, C. Nuzman, J. Maes, M. Peeters, K. Fisher, et al. The tu-t’s new g. vector standard proliferates 100 mb/s dsl. *Communications Magazine, IEEE*, 48(10):140–148, 2010.
- [Ren05] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH Zurich, September 2005. arXiv:quant-ph/0512258.
- [Rim01] B. Rimoldi. Generalized time sharing: a low-complexity capacity-achieving multiple-access technique. *IEEE Transactions on Information Theory*, 47(6):2432–2442, 2001.
- [Sas08] E. Sasoglu. Successive cancellation for cyclic interference channels. In *IEEE Information Theory Workshop, 2008. ITW’08*, pages 36–40, 2008.
- [Sat77] H. Sato. Two-user communication channels. *IEEE Transactions on Information Theory*, 23(3):295–304, 1977.
- [Sat81] H. Sato. The capacity of the Gaussian interference channel under strong interference (corresp.). *IEEE Transactions on Information Theory*, 27(6):786–788, 1981.
- [Sav10] I. Savov. Outer bounds on the quantum interference channel. Term project for *ECSE 612: Multiuser communications* with Prof. Mai Vu. McGill University, April 2010.
- [Sen12a] P. Sen. Achieving the Han-Kobayashi inner bound for the quantum interference channel by sequential decoding. *IEEE International Symposium on Information Theory*, 2012. arXiv:1109.0802.
- [Sen12b] P. Sen. A simultaneous decoder for the quantum multiple access channel and some applications. Mar 2012. (in preparation).
- [Sha48] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423,623–656, 1948.
- [Sha61] C. E. Shannon. Two-way communication channels. In J. Neyman, editor, *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 611–644, Berkeley, California, June 20–July 30 1961. University of California Press.
- [Sha09] J. H. Shapiro. The quantum theory of optical communications. *J. Special Topics in Quantum Elect.*, 15(6):1547–1569, 2009.

- [Sho94] P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings., 35th Annual Symposium on Foundations of Computer Science.*, pages 124–134. IEEE, 1994.
- [Sho95] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *arXiv:quant-ph/9508027*, 1995.
- [SW97] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, 1997. doi:10.1103/PhysRevA.56.131.
- [SW12] I. Savov and M. M. Wilde. Classical codes for quantum broadcast channels. *IEEE International Symposium on Information Theory*, 2012. arXiv:1111.3645.
- [SWV12] I. Savov, M. M. Wilde, and M. Vu. Partial decode-forward for quantum relay channels. *IEEE International Symposium on Information Theory*, 2012. arXiv:1201.0011.
- [Tak12] M. Takeoka. Marton’s inner bound with common messages. Mar 2012. (personal communication).
- [TNO02] T. Tolker-Nielsen and G. Oppenhauser. In-orbit test result of an operational optical intersatellite link between artemis and spot4, silex. In *Proc. SPIE*, volume 4635, pages 1–15, 2002.
- [Tom12] M. Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory*. PhD thesis, ETH Zurich, 2012.
- [WGTL12] M. M. Wilde, S. Guha, S.-H. Tan, and S. Lloyd. Explicit capacity-achieving receivers for optical communication and quantum reading. *IEEE International Symposium on Information Theory*, 2012. arXiv:1202.0518.
- [Wil11] M. M. Wilde. *From Classical to Quantum Shannon Theory*. 2011. arXiv:1106.1445.
- [Win99] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, 1999.
- [Win01] A. Winter. The capacity of the quantum multiple-access channel. *IEEE Transactions on Information Theory*, 47(7):3059–3065, 2001.
- [WS12] M. M. Wilde and I. Savov. Joint source-channel coding for a quantum multiple access channel. 2012. arXiv:1202.3467.
- [XK05] L. L. Xie and PR Kumar. An achievable rate for the multiple-level relay channel. *IEEE Trans. Inf. Theory*, 51(4):1348–1358, 2005.
- [XW11] S.C. Xu and M.M. Wilde. Sequential, successive, and simultaneous decoders for entanglement-assisted classical communication. *arXiv:1107.1347*, 2011.

- [Yar05] J. Yard. *Simultaneous classical-quantum capacities of quantum multiple access channels*. PhD thesis, Stanford University (arXiv:quant-ph/0506050), June 2005.
- [YDH05] J. Yard, I. Devetak, and P. Hayden. Capacity theorems for quantum multiple access channels. In *IEEE International Symposium on Information Theory*, pages 884–888, 2005.
- [Yen05a] B. J. Yen. *Multiple-User Quantum Optical Communication*. PhD thesis, Massachusetts Institute of Technology, 2005.
- [Yen05b] J. H. Yen, B. J. and Shapiro. Multiple-access bosonic communications. *Physical Review A*, 72(6):062312, 2005.
- [YHD08] J. Yard, P. Hayden, and I. Devetak. Capacity theorems for quantum multiple-access channels: Classical-quantum and quantum-quantum capacity regions. *IEEE Transactions on Information Theory*, 54(7):3091–3113, July 2008.
- [YHD11] J. Yard, P. Hayden, and I. Devetak. Quantum broadcast channels, 2011. arXiv:quant-ph/0603098.
- [YP11] H. Yagi and H. V. Poor. Multi-level rate-splitting for synchronous and asynchronous interference channels. In *IEEE International Symposium on Information Theory*, pages 2080–2084. IEEE, 2011.