

SIMULTANEOUS CLASSICAL-QUANTUM CAPACITIES  
OF QUANTUM MULTIPLE ACCESS CHANNELS

A DISSERTATION  
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING  
AND THE COMMITTEE ON GRADUATE STUDIES  
OF STANFORD UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

Jon Thomas Yard  
March, 2005

© Copyright by Jon Thomas Yard 2008  
All Rights Reserved

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

---

Thomas M. Cover  
(Principal Adviser)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

---

Yoshihisa Yamamoto

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

---

Abbas El Gamal

Approved for the University Committee on Graduate Studies.

*To my father*  
Carl A. Yard Jr.

# Acknowledgments

It is hard to express the gratitude I have for my advisor, Tom Cover. I thank him for not only giving me the freedom to learn about and work on any topic of my choosing, but also for his unending encouragement and belief that I would eventually figure out how to do what I needed to do in order to graduate.

I am particularly indebted to my collaborators Igor Devetak and Patrick Hayden. The assistance and guidance that I received from them has helped me to learn an extraordinary amount about quantum information theory, a task which would have been insurmountable without their help. In fact, this dissertation is based on my collaboration [53] with them. I would also like to give thanks to my reading committee, Yoshi Yamamoto, Abbas El Gamal and Alexander Fetter.

I thank the current members of the information theory group, Young-Han, Charles, Styrmir, George and Navid, as well as the past members Josh, David, Michael, Arak and Assaf, for years of stimulating conversations around the office and at our weekly group meetings.

Much thanks to my family for their love and support, especially Mom, who has been anticipating my graduation for quite some time. I also thank my many friends for being there for me over the years, most notably Glenn and Jeremy, who have helped me out of quite a few mathematical jams in which I have found myself over the years. I am gracious for the roommates I have had the pleasure of living with during graduate school, most notably Andy and Sophia, who lived with me during the writing of this dissertation. Thanks for dealing with me and my endless piles of paper.

I also thank Martin Morf, whose support and encouragement throughout my stay

at Stanford has certainly help to keep me going. I am appreciative of Denise Murphy for rescuing me numerous times from various administrative messes in which I found myself over the years. I am also grateful to John Preskill for hosting me for a week at the Caltech Institute for Quantum Information where I worked on completing this dissertation. Finally, a million thanks to Meagen for her love and support.

# Contents

	iv
<b>Acknowledgments</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background</b>	<b>6</b>
2.1 The basics . . . . .	6
2.1.1 Quantum mechanics . . . . .	6
2.1.2 Pure state ensembles . . . . .	8
2.1.3 Density matrices . . . . .	9
2.1.4 Trace norm . . . . .	11
2.1.5 Fidelity . . . . .	12
2.1.6 POVMs . . . . .	13
2.1.7 Classical systems . . . . .	14
2.2 Composite quantum systems . . . . .	15
2.2.1 Partial trace . . . . .	21
2.2.2 Purifications and extensions . . . . .	23
2.2.3 Classical-quantum (cq) systems . . . . .	23
2.3 Dynamics . . . . .	24
2.3.1 Classical channels . . . . .	24
2.3.2 Classical $\rightarrow$ quantum (c $\rightarrow$ q) channels . . . . .	25
2.3.3 Unitary quantum channels . . . . .	26
2.3.4 Quantum channels . . . . .	28

2.3.5	Representing quantum channels . . . . .	31
2.3.6	Complementary channels . . . . .	34
2.3.7	Controlled quantum channels and $cq \rightarrow q$ channels . . . . .	34
2.3.8	Quantum instruments ( $q \rightarrow cq$ channels) . . . . .	36
<b>3</b>	<b>Entropy and information quantities</b>	<b>38</b>
3.1	Entropy . . . . .	38
3.2	Conditional entropy . . . . .	42
3.3	Mutual Information . . . . .	43
3.4	Coherent Information . . . . .	46
3.5	Conditional coherent information . . . . .	49
<b>4</b>	<b>Single-user capacity theorems</b>	<b>51</b>
4.1	Classical capacities of classical channels . . . . .	52
4.2	Classical capacities of quantum channels . . . . .	58
4.3	Quantum capacities of quantum channels . . . . .	61
<b>5</b>	<b>Main results</b>	<b>66</b>
5.1	Quantum multiple access channels . . . . .	66
5.2	$\mathcal{S}(\mathcal{N})$ - the general problem . . . . .	67
5.3	$\mathcal{CQ}(\mathcal{N})$ - classical-quantum capacity region . . . . .	68
5.4	$\mathcal{Q}(\mathcal{N})$ - quantum-quantum capacity region . . . . .	70
<b>6</b>	<b>Supplementary results</b>	<b>72</b>
6.1	Further properties of distance measures . . . . .	72
6.2	Other useful lemmas . . . . .	76
6.3	Strong subadditivity and its consequences . . . . .	77
<b>7</b>	<b>Entanglement generation capacities</b>	<b>83</b>
7.1	The coding theorems . . . . .	85
7.2	The converse theorems . . . . .	100



<b>8</b>	<b>Transmission of quantum information</b>	<b>104</b>
8.1	Entanglement transmission . . . . .	104
8.2	Equivalence of ET and EG . . . . .	106
8.2.1	$\mathcal{CQ}_{\text{eg}} \subseteq \mathcal{CQ}_{\text{et}}$ and $\mathcal{Q}_{\text{eg}} \subseteq \mathcal{Q}_{\text{et}}$ . . . . .	106
8.2.2	$\mathcal{CQ}_{\text{et}} \subseteq \mathcal{CQ}_{\text{eg}}$ . . . . .	107
8.2.3	$\mathcal{Q}_{\text{et}} \subseteq \mathcal{Q}_{\text{eg}}$ . . . . .	108
8.3	Strong subspace transmission revisited . . . . .	109
8.3.1	classical-quantum scenario . . . . .	110
8.3.2	quantum-quantum scenario . . . . .	111
8.4	Equivalence of ET and SST . . . . .	112
8.4.1	$\mathcal{CQ}_{\text{et}} \subseteq \mathcal{CQ}$ . . . . .	112
8.4.2	$\mathcal{Q}_{\text{et}} \subseteq \mathcal{Q}$ . . . . .	118
<b>9</b>	<b>Single-letter examples</b>	<b>120</b>
9.1	Additivity of $\mathcal{CQ}$ for erasure channel . . . . .	120
9.2	Degradable channels . . . . .	124
9.3	Generalized dephasing channels . . . . .	126
9.4	Additivity of $\mathcal{Q}$ for phase-flip channel . . . . .	129
<b>10</b>	<b>Discussion</b>	<b>134</b>
<b>11</b>	<b>Appendix</b>	<b>139</b>
11.1	Quantum instruments and coherent information . . . . .	139
11.2	Proof of convexity of $\mathcal{CQ}$ and $\mathcal{Q}$ . . . . .	140
11.3	Proof of cardinality bound on $\mathcal{X}$ . . . . .	141
	<b>Bibliography</b>	<b>143</b>

# Chapter 1

## Introduction

Information is embodied in physical objects. The paint on the ceiling of the Sistine chapel, the groove of a record, a single strand of DNA, and the spin of an individual electron each reflect a configuration of a particular physical system which governs the way in which it interacts with the rest of the world. Michaelangelo's painting interacts with ambient light, emitting a spectrum of colors viewed by churchgoers and tourists alike. The spinning record induces vibrations in a needle which are converted to a flow of electrons which, when amplified, cause pressure waves to travel through the air. The structure of DNA encodes instructions for both self-replication and the construction of living beings. Interactions between individual electrons are mediated by photons and can be modeled with great precision using the tools of quantum electrodynamics. The physics of simplified models can be calculated using the rules of quantum mechanics. Quantum particles embody *quantum information*.

Claude Shannon, motivated by engineering problems in communication theory, initiated the study of information theory as an abstract discipline. On the first page of his seminal paper [44], he writes

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. . . . the actual message is one *selected from* a set of possible messages.

He later states “we wish to consider certain general problems involving communication systems.” These general problems, the outgrowth from which is referred to

nowadays as “Shannon theory,” concern characterizing the possibilities of reliably transmitting certain “information sources” over “information-bearing channels.” By making probabilistic assumptions regarding the behaviors of the sources and channels, a rich mathematical theory emerges which, in many cases, reasonably approximates the underlying physics. Specifically, Shannon showed that if a channel is modeled by a probability transition matrix  $p(y|x)$ , its capacity for the transmission of classical information is given by

$$C = \max_{p(x)} I(X; Y).$$

This formula will be discussed in Section 3.3, where we introduce the mutual information  $I(X; Y)$ , and in Section 4.1 where we discuss the proof of Shannon’s theorem. Shannon theory applies to network communication as well. A probability transition matrix  $p(z|x, y)$  models a situation where two senders transmit to a single receiver, subject to noise and interference. The rates at which the senders can transmit independent information were determined by Ahlswede [1] and Liao [31] to admit a single-letter characterization, given by the convex hull of the closure of the set of pairs of nonnegative rates  $(R_X, R_Y)$  satisfying

$$\begin{aligned} R_X &< I(X; Z|Y) \\ R_Y &< I(Y; Z|X) \\ R_X + R_Y &< I(XY; Z) \end{aligned}$$

for some  $p(x)p(y)$ . Further analysis by Cover, El Gamal and Salehi [8] gives single-letter characterizations of a set of correlated sources which can be reliably transmitted over a multiple access channel, generalizing the above, as well as Slepian-Wolf source coding and cooperative multiple access channel capacity. They also give a multi-letter expression for the capacity region, showing that an i.i.d. source  $(U, V)$  can be reliably

transmitted if and only if

$$\begin{aligned} H(U|V) &< \frac{1}{n}I(X^n; Z^n|U^nY^n) \\ H(V|U) &< \frac{1}{n}I(Y^n; Z^n|V^nX^n) \\ H(U, V) &< \frac{1}{n}I(X^nY^n; Z^n) \end{aligned}$$

for some  $n$  and  $p(x^n|u^n), p(y^n|v^n)$ , where by  $x^n$  we mean the sequence of symbols  $(x_1, \dots, x_n)$ . Here,  $H(U, V)$  and  $H(U|V)$  respectively denote the *entropy* and *conditional entropy* of the pair of random variables  $(U, V)$  which model the source. Such a characterization is of limited practical use, however, as it does not apparently lead to a finite computation for deciding whether or not a source can be transmitted.

The concept of the entropy of a physical system initially arose out of attempts to characterize the optimal efficiency of physical machines such as steam engines, as well as to rule out the possibility of such constructions as perpetual motion machines. The extensivity of entropy demands that it be additive for independent physical systems. Boltzmann defined the entropy of a physical system to be proportional to the logarithm of the number of microstates, or indistinguishable configurations of its constituents, a definition he was likely led to because  $\log(W_1W_2) = \log(W_1) + \log(W_2)$ , making additivity of entropy evident. In order to circumvent mathematical subtleties which arise due to coarse grainings of the system's configuration space, a probabilistic approach can be taken, allowing rigorous mathematical statements to be made about related systems which are essentially hidden Markov models [9]. A crucial philosophical step was taken by Boltzmann in his work; he assumed that things were made of atoms. In his framework, heat was not a fluid that flowed from warm to cold bodies; rather, vibrational energy of the constituents of a physical system induces similar behavior in neighboring systems. Without direct physical evidence to support the existence of atoms, Boltzmann provided a mechanism for the flow of heat which assumed such ingredients did in fact exist. The existence of atoms was experimentally verified soon after Boltzmann's untimely death.

In the years that followed, the structure of atoms was intensely investigated. The

assumption that atoms obey the laws of Newtonian mechanics quickly resulted in various logical inconsistencies in the form of predictions which did not agree with experimental results. Quantum mechanics was reluctantly developed as a collection of fundamental assumptions about the nature of the physics of atoms and their constituents. A collection of mathematical rules was thus constructed which allowed theoretical calculations of certain aspects of experimental results. A caveat was that the new theory introduced randomness as a fundamental assumption of the theory, a feature which was quite unsettling to even the creators of quantum mechanics, most notably Einstein, who thought that “God does not play dice with the universe.”

Today, we live in a quantum world. Progress in applied physics and engineering has begun to make manipulation of matter on the quantum scale a reality. It is a strange world, at least when viewed with a classical mind. From the other side of the fence, however, classical physics can be seen to be part of a quantum world. The emergence of classicality due to phase transitions in systems of many particles is one way this occurs. Mathematically, as we will see in Sections 2.1.7 and 2.2.3, the tools and language of quantum theory enable the expression of concepts from classical probability theory.

This opens up the possibility of analyzing communication scenarios in which the senders and receivers process quantum information. In this case, the medium quite literally is the message, whereas rather than sending information by selecting a message from a set, physical systems are suitably prepared for transmission to a receiver. The possible types of quantum communication range from transmitting particles from sender to receiver to generating entanglement between the users of a channel. Quite remarkably, certain basic components from the classical theory find a place in the quantum extension. The techniques used to separate possible quantum information processing tasks from the impossible are directly analogous to those used Shannon’s in original program. Possibility questions of this nature have much in common with the original motivations of thermodynamics. The ways in which entropy arises in characterizing the answers further deepens this connection.

In this sense, one aspect of quantum information theory involves generalizing existing classical results to include quantum resources. While network Shannon theory

is already a rich active area of research, its quantum extension has new aspects which do not fit into the former framework. This leads to a theory which, while including the old one as a special case, asks new questions leading to a deeper understanding of the physical nature of information. Apparently, quantum information is something new which cannot be properly analyzed with classical tools alone.

In this dissertation, we will analyze quantum channels with many senders and a single receiver, used in a variety of ways for the simultaneous transmission of classical and quantum information, representing an expanded version of the manuscript [53]. At a high level, the results and approaches contained within mirror those of classical Shannon theory. Yet, the mathematical tools utilized are distinctly quantum. Let us end this introduction by giving a quote from Asher Peres and Daniel Terno's paper [37] on quantum information and relativity, where it is written that "the goals of quantum information theory are the intersection of those of quantum mechanics and information theory, while its tools are the union of these two theories." Well said.

# Chapter 2

## Background

### 2.1 The basics

#### 2.1.1 Quantum mechanics

Let us briefly review some elements of quantum mechanics. The physical state of an isolated system with  $d$  quantum degrees of freedom is described by a complex unit vector  $|\psi\rangle \in \mathbb{C}^d$ . The notation  $|\psi\rangle$ , known as a *ket* or *ket vector*, refers to a normalized column vector with  $d$  complex components:

$$|\psi\rangle = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_d \end{pmatrix}, \text{ where } \sum_a |\alpha_a|^2 = 1.$$

The conjugate transpose of  $|\psi\rangle$  is a row vector

$$\langle\psi| = \left( \alpha_1^* \quad \alpha_2^* \quad \dots \quad \alpha_d^* \right).$$

$\langle\psi|$  is called a *bra* or *bra vector*. This notation (and nomenclature) was introduced by Dirac, partly to emphasize the inner product structure of  $\mathbb{C}^d$ . Indeed, the inner product between two state vectors  $|\phi\rangle$  and  $|\psi\rangle$  is written as a bra times a ket, or

bra-ket

$$\langle \phi | \psi \rangle = \langle \psi | \phi \rangle^*.$$

It is often useful to write a basis for  $\mathbb{C}^d$  by defining a collection of kets as

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, |2\rangle = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, |d\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Then, a state such as  $|\psi\rangle$  can be expanded in terms of this basis as

$$|\psi\rangle = \alpha_1|1\rangle + \dots + \alpha_d|d\rangle.$$

A measurement can be performed on the quantum system, obtaining classical information regarding the system's current quantum state. Quantum mechanics is only able to predict the probabilities of occurrence for each outcome of the measurement. Further, the state of the system will generally be disturbed by the act of obtaining this classical information. The simplest measurement to describe is a *pure state measurement*, which is completely described in terms of some orthogonal basis for  $\mathbb{C}^d$ . Such a basis will be called a *measurement basis*. Suppose that a pure state measurement in the measurement basis  $\{|1\rangle, \dots, |d\rangle\}$  is made on the state  $|\psi\rangle$ . Then,

- The measurement will return  $y$  with probability

$$p(y) \equiv \Pr\{\text{measure } |y\rangle\} = |\langle y | \psi \rangle|^2.$$

- If the measurement returns  $y$ , the post-measurement state is then  $|y\rangle$ .

In other words, the measurement result is modeled by a  $\mathcal{Y} = \{1, \dots, d\}$ -valued random variable  $Y$ , distributed as  $p(y) = |\langle y | \psi \rangle|^2 = |\alpha_y|^2$ , and the post-measurement state is a random vector  $|Y\rangle$ . If the same measurement is performed again, the same result  $Y$  is obtained with certainty, leaving the system in the same state  $|Y\rangle$  after the measurement.



### 2.1.2 Pure state ensembles

Here, let us fix a basis  $\{|y\rangle\}_{y=1}^d$  for  $\mathbb{C}^d$ . Imagine now a game with two parties, Alice and Bob. Assume that Alice has the ability to prepare any pure state from the finite collection of states  $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$ . Then, the probability that Bob obtains measurement result  $y$  given that Alice prepares state  $|\psi_x\rangle$  is given by

$$p(y|x) = |\langle y|\psi_x\rangle|^2. \quad (2.1)$$

Notice that this can be rewritten as

$$\begin{aligned} |\langle y|\psi_x\rangle|^2 &= \langle y|\psi_x\rangle(\langle y|\psi_x\rangle)^* \\ &= \langle y|\psi_x\rangle\langle\psi_x|y\rangle \\ &\equiv \langle y|(|\psi_x\rangle\langle\psi_x|)|y\rangle. \end{aligned} \quad (2.2)$$

We may interpret this as saying that if the 1-dimensional projection matrix  $|\psi_x\rangle\langle\psi_x|$  is written in the  $\{|y\rangle\}$  basis, then  $p(y|x)$  is equal to the diagonal matrix element corresponding to  $|y\rangle$ .

Now, suppose that Alice gives Bob a random state, choosing  $|\psi_x\rangle$  with probability  $p(x)$ . In this case, we say that Alice is preparing an *ensemble*  $\{p(x), |\psi_x\rangle\}$  of pure states. Together with elementary probability, (2.2) can be used to write the probability that Bob measures  $y$  as

$$\begin{aligned} p(y) &= \sum_x p(x)p(y|x) \\ &= \sum_x p(x)\langle y|\psi_x\rangle\langle\psi_x|y\rangle \\ &= \langle y|\left(\sum_x p(x)|\psi_x\rangle\langle\psi_x|\right)|y\rangle \\ &\equiv \langle y|\rho|y\rangle. \end{aligned}$$

where the third line is by linearity. The fourth line defines the *density matrix*

$$\rho = \sum_x p(x) |\psi_x\rangle \langle \psi_x|$$

of the ensemble  $\{p(x), |\psi_x\rangle\}$ , which contains all the data required to compute all probabilities associated with any possible measurement on the ensemble, under the assumption that Bob doesn't know the identities of the individual states. Note that  $\rho$  is Hermitian

$$\rho^\dagger = \sum_x p(x) (|\psi_x\rangle \langle \psi_x|)^\dagger = \sum_x p(x) |\psi_x\rangle \langle \psi_x| = \rho$$

and satisfies

$$\text{Tr } \rho = \sum_x p(x) \text{Tr } |\psi_x\rangle \langle \psi_x| = \sum_x p(x) = 1.$$

$\rho$ , as we've constructed it, is also nonnegative definite. This is because for any  $|\phi\rangle$ , we have

$$\langle \phi | \rho | \phi \rangle = \sum_x p(x) \langle \phi | \psi_x \rangle \langle \psi_x | \phi \rangle = \sum_x p(x) |\langle \phi | \psi_x \rangle|^2 \geq 0$$

where the last inequality is because each term in the sum is nonnegative.

### 2.1.3 Density matrices

We have now seen that if a quantum system is prepared in a random pure state, one can write down its density matrix. This contains all of the data necessary to compute the probabilities of the outcomes of any measurement that can be made on that system, provided that the identities of the random pure states are unknown to the measurer. For a system in a pure state  $|\psi\rangle$ , we will use the abbreviation  $\psi \equiv |\psi\rangle \langle \psi|$  for the density matrix corresponding to that pure state (this is just the matrix which projects onto the subspace spanned by  $|\psi\rangle$ ). Let us define here the collection of all density matrices of a  $d$ -level quantum system as

$$\mathcal{D}^d = \{\rho \in \mathbb{C}^{d \times d} : \rho = \rho^\dagger, \rho \geq 0, \text{Tr } \rho = 1\}.$$

In other words, a density matrix  $\rho \in \mathcal{D}^d$  is a Hermitian, nonnegative definite normalized matrix. We give the following facts about  $\mathcal{D}^d$  without proof, as they are proven in detail in many texts on quantum mechanics [39, 34, 36]:

**Property.**  $\mathcal{D}^d$  is convex.

**Property.** The extremal points of  $\mathcal{D}^d$  are the projections onto rank 1 subspaces of  $\mathbb{C}^d$ , corresponding to equivalence classes of pure states which are identified up to a global phase factor  $e^{i\theta}$ .

**Property.**  $\mathcal{D}^d$  is compact.

We may interpret the first fact as saying that if with probability  $p$ , one chooses to prepare a quantum system so that its density matrix is  $\rho$ , while with probability  $1 - p$ , it is instead prepared so that its density matrix is  $\sigma$ , someone who measures the resulting system (and is also ignorant about which preparation was made) computes measurement probabilities with the state  $p\rho + (1 - p)\sigma$ .

The second fact illustrates the fact that every density matrix can arise from some pure state ensemble. This can be seen more directly, since the Hermiticity of  $\rho$  implies that it is diagonalizable as

$$\rho = \sum_i \lambda_i |i\rangle\langle i|$$

for some orthogonal basis  $\{|i\rangle\}$  for  $\mathbb{C}^d$ . The positivity of  $\rho$  implies that  $\lambda_i \geq 0$ , and the fact that  $\rho$  is normalized implies that the  $\lambda_i$  may be interpreted as probabilities, implying the existence of the required pure state ensemble. Note that there is in fact an uncountable number of ways in which a density matrix can arise by probabilistically preparing pure states.

More importantly, the fact that the extremal points of  $\mathcal{D}^d$  are pure states implies that pure states are special, in that they cannot arise as nontrivial probabilistic preparations of other states. A quantum system in a pure state is in a *definite* state.

### 2.1.4 Trace norm

For an arbitrary  $M \in \mathbb{C}^{d \times d}$ , its *trace norm*  $|M|_1$  is defined as

$$|M|_1 = \text{Tr} \sqrt{MM^\dagger}.$$

This is easily seen to be equal to the sum of the singular values of  $M$ . Indeed, writing a singular value decomposition  $M = U\Lambda V^\dagger$ , it follows that

$$|M|_1 = \text{Tr} \sqrt{U\Lambda V^\dagger V\Lambda U^\dagger} = \text{Tr} U\sqrt{\Lambda^2}U^\dagger = \sum_i^d |\lambda_i|,$$

where  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$ . As  $|\cdot|_1$  is a norm (or rather, a *unitarily invariant matrix norm*), it satisfies the following properties:

**Property (Positivity).**  $|M|_1 \geq 0$ , while  $|M|_1 = 0$  if and only if  $M = 0$ .

**Property (Homogeneity).** for any  $c \in \mathbb{C}$ ,  $|cM|_1 = |c||M|_1$

**Property (Unitary invariance).**  $|M|_1 = |UMU^\dagger|_1$  for any unitary  $U$

**Property (Triangle inequality).**  $|M + N|_1 \leq |M|_1 + |N|_1$

**Property (Submultiplicativity).**  $|MN|_1 \leq |M|_1|N|_1$

Positivity follows because the singular values of any matrix  $M$  are always nonnegative, and are all equal to zero if and only if  $M = 0$ . Homogeneity is true because the singular values of  $cM$  equal  $|c|$  times those of  $M$ , and unitary invariance holds because  $UMU^\dagger$  and  $M$  have the same singular values. For proofs of the triangle inequality and submultiplicativity, the reader is referred to [26].

The trace norm gives a natural metric space structure to  $\mathbb{C}^{d \times d}$  which we will exploit considerably throughout this dissertation. Given two matrices  $M, N \in \mathbb{C}^{d \times d}$  their *trace distance* is thus defined as the trace norm of their difference  $|M - N|_1$ . For two density matrices  $\rho$  and  $\sigma$  of a  $d$ -level quantum system, their trace distance satisfies

$$0 \leq |\rho - \sigma|_1 \leq 2,$$

where the lower bound is saturated if and only if  $\rho = \sigma$ , while the upper bound is saturated if and only if  $\rho$  and  $\sigma$  are supported on orthogonal subspaces. Let us mention here the following alternative characterization of the trace distance between two density matrices [34]

$$|\rho - \sigma|_1 = 2 \max_{0 \leq \Lambda \leq 1} \text{Tr} \Lambda(\rho - \sigma).$$

The maximization above is over all nonnegative definite matrices  $\Lambda$  with spectrum bounded above by 1.

### 2.1.5 Fidelity

Given two density matrices  $\rho$  and  $\sigma$  of a  $d$ -level system, their *fidelity* is defined <sup>1</sup> as

$$F(\rho, \sigma) = \left( \text{Tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right)^2.$$

Fidelity can be expressed in terms of the trace norm as

$$F(\rho, \sigma) = \left| \sqrt{\rho} \sqrt{\sigma} \right|_1^2,$$

a form which makes apparent the symmetry of fidelity in its two arguments. The following bounds are always satisfied whenever the arguments are density matrices

$$0 \leq F(\rho, \sigma) \leq 1.$$

The lower bound is saturated if and only if  $\rho$  and  $\sigma$  have orthogonal support, while the upper bound is saturated if and only if  $\rho = \sigma$ . Contrary to the situation with the trace norm, a large value of the fidelity between two states signifies that they are close. Fidelity is not a norm, but it can be related to the trace norm in various ways which are summarized in Section 6.1.

---

<sup>1</sup>Note that many authors (such as [34]) define this quantity as the square root of our definition.

If one of the arguments of the fidelity is a pure state, (say  $\rho = |\phi\rangle\langle\phi|$ ), then

$$\begin{aligned} F(|\phi\rangle, \sigma) &= \left( \text{Tr} \sqrt{|\phi\rangle\langle\phi| \sigma |\phi\rangle\langle\phi|} \right)^2 \\ &= \left( \text{Tr} |\phi\rangle\langle\phi| \sqrt{\langle\phi| \sigma |\phi\rangle} \right)^2 \\ &= \langle\phi| \sigma |\phi\rangle. \end{aligned}$$

So  $F(|\phi\rangle, \sigma)$  is just the diagonal matrix element of  $\sigma$  corresponding to  $|\phi\rangle$ , when  $\sigma$  is written in a basis including  $|\phi\rangle$ . Note that this is the success probability for a pure state measurement which tests a system prepared in the state  $\sigma$  for the presence of the state  $|\phi\rangle$ . When both arguments are pure states, we obtain

$$F(|\phi\rangle, |\psi\rangle) = |\langle\phi|\psi\rangle|^2.$$

Finally observe the following easily verifiable property.

**Property (Linearity of fidelity).** *Fidelity is linear in each argument, i.e.*

$$F(c\rho, \sigma) = cF(\rho, \sigma) = F(\rho, c\sigma).$$

### 2.1.6 POVMs

We describe here a certain general type of measurement which can be performed on a  $d$ -level quantum system, called a positive operator valued measurement (POVM). A POVM is specified in terms of a finite collection of matrices  $\{\Lambda_x \in \mathbb{C}^{d \times d}\}_{x \in \mathcal{X}}$  which are positive ( $\Lambda_x \geq 0$ ) and sum to the  $d \times d$  identity matrix  $1_d$

$$\sum_y \Lambda_y = 1_d.$$

It is often said that the matrices  $\{\Lambda_x\}_{x \in \mathcal{X}}$  form a *partition of unity*. If the quantum system is in the state  $\rho$ , the probability of obtaining a measurement result  $y$  is given by

$$p(y) = \text{Pr}\{\text{measure } \Lambda_y\} = \text{Tr} \Lambda_y \rho.$$

Conditioned on having received the measurement result  $y$ , the post-measurement state after such a measurement is computed as

$$\rho \mapsto \rho_y = \frac{\sqrt{\Lambda_y} \rho \sqrt{\Lambda_y}}{p(y)}.$$

Here,  $\sqrt{\Lambda}$  is defined as the unique, positive operator which satisfies  $\sqrt{\Lambda} \sqrt{\Lambda} = \Lambda$ .<sup>2</sup> The measurement results in an ensemble of density matrices  $\{p(y), \rho_y\}$ . A pure state measurement in the basis  $\{|x\rangle\}$  can be expressed as the POVM  $\{|x\rangle\langle x|\}$  consisting of 1-dimensional projection matrices.

### 2.1.7 Classical systems

Let  $\mathcal{X}$  be a finite set and let  $X$  be an  $\mathcal{X}$ -valued random variable, distributed according to  $p(x)$ . We can define a vector space  $\mathbb{C}^{|\mathcal{X}|}$  with a fixed orthonormal basis  $\{|x\rangle^X\}_{x \in \mathcal{X}}$ , labeled by elements of the set  $\mathcal{X}$ . This sets up an identification  $|\cdot\rangle^X: \mathcal{X} \rightarrow \mathbb{C}^{|\mathcal{X}|}$  between the elements of  $\mathcal{X}$  and that particular basis. By this correspondence, the probability mass function  $p(x)$  can be mapped to a density matrix

$$\rho = \sum_{x \in \mathcal{X}} p(x) |x\rangle\langle x| \tag{2.3}$$

which is diagonal in the basis  $\{|x\rangle\}_{x \in \mathcal{X}}$ . Further, to every subset  $S \subseteq \mathcal{X}$  corresponds a projection matrix  $\Pi_S = \sum_{x \in S} |x\rangle\langle x|$  which commutes with  $\rho$ . In addition, the projections  $\Pi_S$  and  $\Pi_T$  corresponding to any two subsets  $S, T \subseteq \mathcal{X}$  commute. This way, we can express concepts from classical probability theory in the language of quantum probability. Consider the following translations from classical to quantum

---

<sup>2</sup>Note that some authors use a more general kind of measurement, described by matrices  $\{M_y\}$  satisfying  $\sum_y M_y^\dagger M_y = 1_d$ . This amounts to choosing a different square root of each  $\{\Lambda_y\}$ , giving post-measurement states which are unitarily equivalent to those of the convention above, conditioned on the measurement result. Such measurements can be modeled using the tools introduced in Section 2.3.8.

language:

$$\begin{aligned}
\Pr\{X \in S\} &= \text{Tr } \rho \Pi_S \\
\Pr\{X \notin S\} &= 1 - \text{Tr } \rho \Pi_S = \text{Tr } \rho(1^X - \Pi_S) \\
&\equiv \text{Tr } \rho \Pi_{S^c} \\
\Pr\{X \in S \text{ and } X \in T\} &= \text{Tr } \rho \Pi_S \Pi_T \\
&\equiv \text{Tr } \rho \Pi_{S \cap T} \\
\Pr\{X \in S \text{ or } X \in T\} &= 1 - \Pr\{X \notin S \text{ and } X \notin T\} \\
&= \text{Tr } \rho(1^X - (1^X - \Pi_S)(1^X - \Pi_T)) \\
&\equiv \text{Tr } \rho \Pi_{S \cup T}.
\end{aligned}$$

From the early development of quantum mechanics, noncommutativity has been seen to be the hallmark of quantum behavior. It is to be expected that classical probability, embedded in quantum theory's framework, is described entirely with commuting matrices.

## 2.2 Composite quantum systems

Let us begin by introducing a number of conventions which will be used when dealing with multiple quantum systems. We will use capital letters from the beginning of the alphabet  $A, B, C, \dots$  as labels for quantum systems. If  $A$  is a quantum system, we will abbreviate its level as  $|A|$  (which will *always* be finite), so that its pure states are unit vectors in  $\mathbb{C}^{|A|}$ . A generic pure state of  $A$  will then be written as  $|\psi\rangle^A$ , while a generic density matrix of  $A$  will be written  $\rho^A$ , to remind the reader to which system the state refers. Whenever we initially introduce a state, the superscript will identify the system it is describing, although later references to that state will not always include the superscript. This convention will not be cause for confusion, as different symbols will refer to different states. We will also write the  $|A| \times |A|$  identity matrix on  $\mathbb{C}^{|A|}$  as  $1^A$ .

If  $B$  is another quantum systems, then  $A$  and  $B$  may be combined to form a



*composite quantum system*  $AB$ . This new system has  $|A| \cdot |B| \equiv |AB|$  levels. The pure states of the new system are instead unit vectors in the *tensor product*  $\mathbb{C}^{|A|} \otimes \mathbb{C}^{|B|}$  vector space of the individual vector spaces. The simplest way to define  $\mathbb{C}^{|A|} \otimes \mathbb{C}^{|B|}$  is as follows. First, fix arbitrary bases  $\{|a\rangle^A\}_{a=1}^{|A|}$  and  $\{|b\rangle^B\}_{b=1}^{|B|}$  for  $\mathbb{C}^{|A|}$  and  $\mathbb{C}^{|B|}$ . Then,  $\mathbb{C}^{|A|} \otimes \mathbb{C}^{|B|}$  can be formally defined as the linear span of the basis vectors formed by the product of the two individual bases

$$\left\{ |a\rangle^A \otimes |b\rangle^B \right\}_{a,b=1}^{|A|,|B|}.$$

A convenient shorthand for the tensor product of pure states is to write

$$|a\rangle^A |b\rangle^B \equiv |a\rangle^A \otimes |b\rangle^B.$$

Then, any pure state of the quantum system can be written as

$$|\Psi\rangle^{AB} = \sum_{a=1}^{|A|} \sum_{b=1}^{|B|} c_{ab} |a\rangle^A |b\rangle^B. \quad (2.4)$$

Observe that this new vector space we have constructed has dimension  $|A| \cdot |B|$ . It is not difficult to show that this construction is universal, meaning that it is independent of the particular bases chosen for  $A$  and for  $B$ .

It will be useful here to describe a certain convention which can be used to write down the tensor product of two column vectors as a single column vector. This will amount to fixing a way to enumerate the components of the tensor. Suppose that  $\vec{v} \in \mathbb{C}^{|A|}$  and  $\vec{w} \in \mathbb{C}^{|B|}$  are arbitrary column vectors

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{|A|} \end{pmatrix} \quad \text{and} \quad \vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{|B|} \end{pmatrix}.$$

As  $\mathbb{C}^{|A|} \otimes \mathbb{C}^{|B|} \simeq \mathbb{C}^{|A| \cdot |B|}$ , we can “flatten”  $\vec{v} \otimes \vec{w}$  into a single column vector, organizing

its components according to the following convention

$$\text{flatten: } \vec{v} \otimes \vec{w} \mapsto \begin{pmatrix} v_1 \vec{w} \\ v_2 \vec{w} \\ \vdots \\ v_{|A|} \vec{w} \end{pmatrix}.$$

In this way, the earlier generic state (2.4) can be expressed as

$$\text{flatten: } |\Psi\rangle^{AB} \mapsto \begin{pmatrix} c_{11} \\ \vdots \\ c_{1|B|} \\ c_{21} \\ \vdots \\ c_{2|B|} \\ \vdots \\ c_{|A||B|} \end{pmatrix}$$

It is often the case that a pure state such as  $|\Psi\rangle^{AB}$  cannot be written as a tensor product of pure states of its constituent systems, i.e.

$$|\Psi\rangle^{AB} \neq |\psi\rangle^A |\phi\rangle^B$$

for any pure states  $|\psi\rangle^A$  and  $|\phi\rangle^B$ . If this is the case, then  $|\Psi\rangle^{AB}$  is said to be *entangled*. Nevertheless, for any pure state of the composite quantum system, there exists a pair of orthonormal bases  $\{|i\rangle^A\}$  and  $\{|i\rangle^B\}$  such that

$$|\Psi\rangle^{AB} = \sum_i d_i |i\rangle^A |i\rangle^B.$$

This form is called the *Schmidt decomposition* of  $|\Psi\rangle^{AB}$ . Together, the combination of the orthonormal bases  $\{|i\rangle^A |i\rangle^B\}$  is called the *Schmidt basis*, while the  $\{d_i\}$  are called the *Schmidt coefficients*. These are easily calculated from the singular value

decomposition of the matrix  $[c_{i,j}]$  of coefficients in (2.4), where the Schmidt basis consists of the left and right eigenvectors, while the Schmidt coefficients are the singular values themselves.

Just as the tensor product builds larger vector spaces out of pairs of smaller ones, it also builds larger matrices from pairs of smaller ones. Fix two matrices  $M \in \mathbb{C}^{|C| \times |A|}$  and  $N \in \mathbb{C}^{|D| \times |B|}$ . Recall that these are linear operators

$$M: \mathbb{C}^{|A|} \rightarrow \mathbb{C}^{|C|} \text{ and } N: \mathbb{C}^{|B|} \rightarrow \mathbb{C}^{|D|}.$$

Their tensor product  $M \otimes N$  is another linear operator

$$(M \otimes N): \mathbb{C}^{|A|} \otimes \mathbb{C}^{|B|} \rightarrow \mathbb{C}^{|C|} \otimes \mathbb{C}^{|D|}.$$

We will abbreviate this by writing

$$M: A \rightarrow C, \quad N: B \rightarrow D \quad \text{and} \quad (M \otimes N): AB \rightarrow CD.$$

This new object acts on the tensor product of vectors as

$$(M \otimes N)(|\psi\rangle^A \otimes |\phi\rangle^B) = (M|\psi\rangle^A) \otimes (N|\phi\rangle^B)$$

and linearity defines the action of  $M \otimes N$  on all of  $\mathbb{C}^{|A|} \otimes \mathbb{C}^{|B|}$ . The tensor product is also *bilinear*, i.e. for any  $c \in \mathbb{C}$ ,

$$c(M \otimes N) = (cM) \otimes N = M \otimes (cN).$$

In the same vein as the “flattened” representation  $\mathbb{C}^{|A|} \otimes \mathbb{C}^{|B|} \simeq \mathbb{C}^{|A| \cdot |B|}$  for the tensor product of vectors, there is more general mapping  $\mathbb{C}^{|A| \times |C|} \otimes \mathbb{C}^{|B| \times |D|} \simeq \mathbb{C}^{|A| \cdot |B| \times |C| \cdot |D|}$  given by

$$\text{flatten: } M \otimes N \mapsto \begin{pmatrix} m_{11}N & m_{12}N & \dots & m_{1|C|}N \\ m_{21}N & m_{22}N & & \\ \vdots & & \ddots & \\ m_{|A|1}N & & & m_{|A||C|}N \end{pmatrix}.$$

Note our convention, where the blocks are labelled by elements of the left-most component of the tensor product. We will use that convention throughout this dissertation. It is easy to see that calculations can be made in this representation, namely that

$$\text{flatten}\{M \otimes N\}\text{flatten}\{|\psi\rangle|\phi\rangle\} = \text{flatten}\{(M \otimes N)|\psi\rangle|\phi\rangle\}.$$

As the composite system  $AB$  is a quantum system itself, it includes a (strictly) larger collection of von Neumann measurements and unitary evolutions. Indeed, given any two bases  $\{|i\rangle^{AB}\}$  and  $\{|i'\rangle^{AB}\}$  for  $\mathbb{C}^{|A|} \otimes \mathbb{C}^{|B|}$ , they are related by a particular unitary matrix  $U$ , defined as

$$U = \sum_{i'i} |i'\rangle\langle i|.$$

It is not hard to see that any joint von Neumann measurement on the combined system  $AB$  can be performed using separate product measurements on  $A$  and  $B$ , provided that the unitary which takes intended measurement basis to the required product basis (and its inverse) are implementable.

Of particular interest is the subject of *local measurements* on a composite quantum system. Suppose that a measurement  $\{\Lambda_x\}_{x \in \mathcal{X}}$  is made on the  $A$  part of the bipartite state  $\rho^{AB}$ . New measurement operators  $\{\Lambda_x \otimes 1^B\}_{x \in \mathcal{X}}$  can be constructed, so that

$$p(x) = \text{Tr} \rho(\Lambda_x \otimes 1^B).$$

The post-measurement states are given as before

$$\rho \mapsto \rho_x = \frac{(\sqrt{\Lambda_x} \otimes 1^B)\rho(\sqrt{\Lambda_x} \otimes 1^B)}{p(x)}.$$

It is instructive to see what happens if a local pure state measurement is made on part of a bipartite pure state  $|\Psi\rangle^{AB}$ . Here,  $\Lambda_x = |x\rangle\langle x|$ , and we obtain

$$p(x) = \text{Tr} \rho(|x\rangle\langle x| \otimes 1^B).$$

As a first step, express  $|\Psi\rangle^{AB}$  in terms of the new basis for  $A$  as

$$|\Psi\rangle^{AB} = \sum_{xb} d_{xb} |x\rangle^A |b\rangle^B. \quad (2.5)$$

Note that the new coefficients  $d_{xb}$  are related to the old ones via

$$\sum_a U_{xa} c_{ab} = d_{xb},$$

where  $U: \{|a\rangle\} \mapsto \{|x\rangle\}$  is the unitary change of basis matrix. Then,

$$\begin{aligned} |\Psi\rangle^{AB} &= \sum_{xb} d_{xb} |x\rangle^A |b\rangle^B \\ &= \sum_x |x\rangle^A \left( \sum_b d_{xb} |b\rangle^B \right) \\ &\equiv \sum_x |x\rangle^A |\tilde{\psi}_x\rangle^B \\ &\equiv \sum_x \beta_x |x\rangle^A |\psi_x\rangle^B. \end{aligned}$$

The third step above defines the unnormalized vector  $|\tilde{\psi}_x\rangle^B$ , where in the last, the normalization constant  $\beta_x \equiv \sqrt{\langle \tilde{\psi}_x | \tilde{\psi}_x \rangle} = \sqrt{\sum_b |d_{xb}|^2}$  and normalized state  $|\psi_x\rangle \equiv \beta_x^{-1} |\tilde{\psi}_x\rangle$  are defined. Now, it is a simple task to compute

$$\begin{aligned} p(x) &= \text{Tr} (|x\rangle\langle x| \otimes 1^B) \Psi^{AB} \\ &= \langle \Psi |^{AB} (|x\rangle\langle x| \otimes 1^B) | \Psi \rangle^{AB} \\ &= \left( \sum_{x''} \beta_{x''}^* \langle x'' |^A \langle \psi_{x''} |^B \right) (|x\rangle\langle x| \otimes 1^B) \left( \sum_{x'} \beta_{x'} |x'\rangle^A |\psi_{x'}\rangle^B \right) \\ &= \sum_{x''x'} \beta_{x''}^* \beta_{x'} \langle x'' | x \rangle \langle x | x' \rangle \langle \psi_{x''} | \psi_{x'} \rangle \\ &= |\beta_x|^2. \end{aligned}$$

Conditioned on having received the measurement result  $x$ , the post-measurement

state is

$$\begin{aligned}
\Psi_x^{AB} &= \frac{(|x\rangle\langle x| \otimes 1^B) \Psi^{AB} (|x\rangle\langle x| \otimes 1^B)}{p(x)} \\
&= \frac{(|x\rangle\langle x| \otimes 1^B) \left( \sum_{x''x'} \beta_{x''} \beta_{x'}^* |x''\rangle\langle x'| \otimes |\psi_{x''}\rangle\langle\psi_{x'}| \right) (|x\rangle\langle x| \otimes 1^B)}{|\beta_x|^2} \\
&= |x\rangle\langle x| \otimes |\psi_x\rangle\langle\psi_x|.
\end{aligned}$$

Or rather,

$$|\Psi_x\rangle^{AB} = |x\rangle^A |\psi_x\rangle^B.$$

So, we see that a measurement on  $A$  causes the state of  $B$  to “collapse” as well. Rather, we see that the measurement on  $A$  creates a pure state ensemble  $\{p(x), |\psi_x\rangle^B\}$  on  $B$ . If an arbitrary POVM is performed on  $A$ , an ensemble of density matrices on  $B$  will generally result. To see this, we need to introduce the partial trace.

### 2.2.1 Partial trace

If we are instead concerned only with the measurement probabilities, and not with the post-measurement states, it is convenient to work with a density matrix on  $A$  to compute the measurement probabilities. This density matrix is defined in terms of the *partial trace* over  $B$ . Fixing a bipartite density matrix  $\Omega^{AB}$ , the partial trace over  $B$  of  $\Omega^{AB}$  can be defined as the unique density matrix  $\text{Tr}_B \Omega$  on  $A$  such that for every  $M \in \mathbb{C}^{|A| \times |A|}$ ,

$$\text{Tr} M (\text{Tr}_B \Omega) \equiv \text{Tr} (M \otimes 1^B) \Omega.$$

An equivalent way to define  $\text{Tr}_B \Omega$  is as follows. If we write  $\Omega_{a'ab'b} \equiv \langle a' | \langle b' | \Omega | b \rangle | a \rangle$  and  $(\text{Tr}_B \Omega)_{a'a} \equiv \langle a' | (\text{Tr}_B \Omega) | a \rangle$ , then

$$(\text{Tr}_B \Omega)_{a'a} = \sum_b \Omega_{aa'bb}.$$

With this in hand, we can express

$$\text{Tr} (\Lambda_x \otimes 1^B) \Omega = \text{Tr} \Lambda_x (\text{Tr}_B \Omega).$$

For any square matrix  $M$  on  $AB$ , the partial traces over  $A$  and  $B$  satisfy the following easily verifiable properties:

$$\mathrm{Tr} M = \mathrm{Tr}_{AB} M = \mathrm{Tr}_A \mathrm{Tr}_B M = \mathrm{Tr}_B \mathrm{Tr}_A M.$$

A perhaps more concrete definition of the partial trace is obtained by writing a bipartite density matrix in the flattened representation

$$\Omega^{AB} = \begin{pmatrix} \omega_{11} & \cdots & \omega_{1|A|} \\ \vdots & \ddots & \\ \omega_{|A|1} & & \omega_{|A||A|} \end{pmatrix}$$

where each  $\omega_{aa'} \in \mathbb{C}^{|B| \times |B|}$ . Then,  $\mathrm{Tr}_A \Omega^{AB}$  is obtained by summing the blocks on the diagonal

$$\mathrm{Tr}_A \Omega^{AB} = \sum_a \omega_{aa}$$

and  $\mathrm{Tr}_B \Omega^{AB}$  by taking the trace of each block separately

$$\mathrm{Tr}_B \Omega^{AB} = \begin{pmatrix} \mathrm{Tr} \omega_{11} & \cdots & \mathrm{Tr} \omega_{1|A|} \\ \vdots & \ddots & \\ \mathrm{Tr} \omega_{|A|1} & & \mathrm{Tr} \omega_{|A||A|} \end{pmatrix}.$$

In fact, this representation will allow us to define the following partial product

$$\langle a' | \Omega | a \rangle \equiv \omega_{a'a},$$

allowing the partial trace over  $A$  to be expressed in the same way as with the usual trace

$$\mathrm{Tr}_A \Omega = \sum_a \langle a | \Omega | a \rangle = \sum_a \omega_{aa}.$$

For the generic state  $\Psi^{AB}$  written in the form (2.5), let us compute

$$\begin{aligned}\mathrm{Tr}_A \Psi^{AB} &= \sum_x \langle x|^A \left( \sum_{x''x'} \beta_{x''}^* \beta_{x'} |x''\rangle \langle x'| \otimes |\psi_{x''}\rangle \langle \psi_{x'}| \right) |x\rangle^A \\ &= \sum_x |\beta_x|^2 \psi_x^B\end{aligned}$$

## 2.2.2 Purifications and extensions

Given an arbitrary density matrix  $\rho^B$ , it is easy to construct a pure state  $|\Psi\rangle^{AB}$  such that  $\mathrm{Tr}_A \Psi = \rho$ . The state  $|\Psi\rangle^{AB}$  is called a *purification* of  $\rho$ . The construction is as follows. First, choose any pure state ensemble  $\{p(x), |\psi_x\rangle^B\}$  giving rise to  $\rho^B$ , in the sense that

$$\sum_x p(x) \psi_x = \rho.$$

Then, the state

$$|\Psi\rangle^{AB} = \sum_x \sqrt{p(x)} |x\rangle^A |\psi_x\rangle^B$$

is a purification of  $\rho^B$ . This is easy to see by computing the partial trace over  $A$ , which was done for a pure state of the same form in the last subsection.

More generally we will speak of an *extension*  $\Omega^{AB}$  of a density matrix  $\rho^A$ , which is just any density matrix (not necessarily a pure state) for which  $\mathrm{Tr}_B \Omega = \rho$ . It is easy to see that any purification  $|\Psi\rangle^{ABC}$  of  $\Omega^{AB}$  is a purification of  $\rho^A$  as well, since

$$\mathrm{Tr}_{BC} \Psi = \mathrm{Tr}_B(\mathrm{Tr}_C \Psi) = \mathrm{Tr}_B \Omega = \rho.$$

to do: purifications! relate ensembles, purifications and measurements

## 2.2.3 Classical-quantum (cq) systems

Consider now a collection of density matrices  $\{\sigma_x^A\}_{x \in \mathcal{X}}$ , indexed by the finite set  $\mathcal{X}$ . If those states occur according to the probability mass function  $p(x)$ , we may speak of an *ensemble*  $\{p(x), \sigma_x^A\}$  of quantum states. In order to treat classical and quantum



probabilities in the same framework, a joint density matrix can be constructed

$$\sigma^{XA} = \sum_{x \in \mathcal{X}} p(x) |x\rangle \langle x|^X \otimes \sigma_x^A.$$

This is known as a *cq state*, and describes the classical and quantum aspects of the ensemble on the *extended Hilbert space*  $\mathbb{C}^{|\mathcal{X}|} \otimes \mathbb{C}^{|A|}$  [16]. The semiclassical nature of the ensemble is reflected in the embedding of a direct sum of Hilbert spaces  $\bigoplus_{x \in \mathcal{X}} \mathbb{C}^{|A|}$  into  $\mathbb{C}^{|\mathcal{X}|} \otimes \mathbb{C}^{|A|}$ . This should be compared with what was done in Section 2.1.7, where a direct sum of one-dimensional vector spaces  $\bigoplus_{x \in \mathcal{X}} \mathbb{C}$  was embedded into  $\mathbb{C}^{|\mathcal{X}|}$ . Just as the classical density matrix  $\rho$  from (2.3) was diagonal in a basis corresponding to elements of  $\mathcal{X}$ , the cq density matrix  $\sigma$  is *block-diagonal*, where the diagonal block corresponding to  $x$  contains the non-normalized density matrix  $p(x)\sigma_x$ . The classical state is recoverable as  $\rho = \text{Tr}_A \sigma$ , while the average quantum state is  $\text{Tr}_X \sigma = \sum_{x \in \mathcal{X}} \sigma_x$ . The classical-quantum formalism is not only of interest in its own right; information quantities evaluated on cq states play an important role in characterizing what is possible in quantum information theory.

## 2.3 Dynamics

We we have already seen an example of quantum dynamics; namely, the measurement process. In this section, we introduce the most general types of dynamical processes we will consider in this dissertation. The approach taken here will be to consider quantum channels whose inputs and/or outputs are classical-quantum systems. But first, let us review the notion of classical channels.

### 2.3.1 Classical channels

A discrete classical channel with input symbols belonging to a finite alphabet  $\mathcal{X}$  and output symbols from a finite alphabet  $\mathcal{Y}$  is modeled by a collection of *transition probabilities*  $p(y|x)$ . These probabilities comprise a *stochastic matrix*  $[p(y|x)]_{yx}$ , because

the following two conditions are satisfied:

$$p(y|x) \geq 0 \text{ for each } (x, y) \in \mathcal{X} \times \mathcal{Y}$$

and

$$\sum_y p(y|x) = 1 \text{ for each } x \in \mathcal{X}$$

ensuring that to each input symbol  $x$ , there corresponds a conditional probability mass function on the output symbols  $\mathcal{Y}$ .

Given an  $\mathcal{X}$ -valued random variable  $X$  with probability mass function  $p(x)$ , the action of the channel then defines another random variable  $Y$ , jointly distributed with  $X$  according to

$$p(x, y) = p(x)p(y|x).$$

Alternatively, we may view  $p(y|x)$  a linear map from the simplex of probability mass functions on  $\mathcal{X}$  to the simplex of probability mass functions on  $\mathcal{Y}$ , via

$$p(x) \mapsto p(y) = \sum_x p(x)p(y|x).$$

In this sense, a classical channel is a model for a device which allows a sender to “prepare probability mass functions” at the output. This way of looking at classical channels leads to our first “partial” quantum generalization, described in the next section.

### 2.3.2 Classical $\rightarrow$ quantum (c $\rightarrow$ q) channels

This generalization of classical channels consists of channels with a classical input and a quantum output. However, instead of preparing probability mass functions at the output, the sender prepares density matrices. A c  $\rightarrow$  q channel  $\mathcal{X} \rightarrow B$  is specified by a collection of *conditional density matrices*  $\{\rho_x^B\}_{x \in \mathcal{X}}$ , labeled by the elements of a finite set  $\mathcal{X}$ . As with classical channels, such maps extend to mappings from the simplex of probability mass functions on the input alphabet  $\mathcal{X}$  to the density matrices

on the output quantum system  $B$  via

$$p(x) \mapsto \sum_x \rho_x^B.$$

Such channels were implicitly considered in Section 2.2.3, where we saw that if the input is modeled by a random variable  $X$  distributed according to  $p(x)$ , the combined input-output is a cq system with cq state

$$\rho^{XB} = \sum_x p(x) |x\rangle\langle x| \otimes \rho_x^B.$$

The collection of  $c \rightarrow q$  channels with the same input set  $\mathcal{X}$  and output quantum system  $B$  has the structure of a compact convex set. Given two such channels with conditional density matrices  $\{\rho_x\}_{x \in \mathcal{X}}$  and  $\{\sigma_x\}_{x \in \mathcal{X}}$ , if  $0 \leq \lambda \leq 1$ , their corresponding convex combination has conditional density matrices  $\{\lambda\rho_x + (1 - \lambda)\sigma_x\}_{x \in \mathcal{X}}$ . The extremal points of this convex set have conditional density matrices which are extremal in the convex set of density matrices on  $B$ . In other words, the extremal points consist of channels which prepare pure states. This fact will be important when we discuss classical capacities of quantum channels in Section 4.2.

### 2.3.3 Unitary quantum channels

The simplest quantum channel is a unitary transformation. For a closed quantum system  $A$ , this is the kind of evolution predicted by the Schrodinger equation

$$|\psi\rangle^A \mapsto |\psi'\rangle^A = U|\psi\rangle^A.$$

We will write

$$U: A \rightarrow A$$

to reflect the fact that  $U \in \mathbb{C}^{|A| \times |A|}$  is a square matrix mapping

$$U: \mathbb{C}^{|A|} \rightarrow \mathbb{C}^{|A|}.$$

In this thesis, we will be exploring the consequences for processing quantum information which result from the ability to cause *any* unitary evolution to occur to a given quantum system. Ensuring that a quantum system undergoes a particular unitary evolution is generally a difficult engineering task, since it involves influencing the system in just the right way, from the outside, so as to inhibit its natural tendency to evolve in the way that it would have without any influence. To say that this will be of no concern to us here would be somewhat untrue. In fact, the central goal of this thesis is to show that, under that assumption that error-free processing of quantum information is possible, one can in fact *protect* and *correct* quantum information from this natural tendency to interfere with other quantum information and with the environment. Indeed, we will assume that it is possible to process quantum information *fault tolerantly*. If the state of  $A$  is specified by a density matrix  $\rho^A$ , the unitary channel acts as

$$U: \rho^A \mapsto \rho'^A = U\rho U^\dagger.$$

In other words,  $\rho$  transforms according to the adjoint map associated to  $U$ . We will frequently abbreviate this map as

$$U(\rho) \equiv U\rho U^\dagger.$$

It will often be useful for us to speak of unitaries *between* quantum systems. For example, we may think of a quantum system  $A$  at some time  $t$ , being turned into another quantum system  $B$  at a later time  $t'$ , where  $|A| = |B|$ . If this process acts unitarily, we will write

$$U: A \rightarrow B$$

for the associated unitary channel. As an example, consider a physical scenario in which an electron placed at a position  $x$  at time  $t$  is transferred to some other position  $x'$  by some later time  $t + T$ , after having been rotated by  $180^\circ$  about its  $z$  axis. The quantum system  $A$  thus represents the original preparation of the electron at  $x$ , while  $B$  represents the evolved electron,  $T$  seconds later, with its new state at position  $x'$ .

### 2.3.4 Quantum channels

Quantum channels represent a physical process which transfers quantum states forward in time. The state at the output of the channel will be some noisy version of what was put in. Examples include an optical fiber over which the polarization of an input photon may become corrupted by noise, or a quantum dot which will hold a single electron for an uncertain amount of time.

Here, we will give a precise mathematical definition of quantum channels as functions from the density matrices of an *input* quantum system to the density matrices of an *output* quantum system, generalizing the notion of discrete memoryless classical channels described in Section 2.3.1, which map probability mass functions on the input alphabet to probability mass functions on the output. The mathematical properties which we require a channel to satisfy are from the standard literature on open quantum systems and quantum information theory, so much of the content here is presented without proof. Some standard references for this material include [34, 38].

By a *quantum channel*  $\mathcal{N}: A \rightarrow B$ , we mean a mathematical object which maps density matrices on  $A$  to density matrices on  $B$ , while satisfying the following three physically motivated properties described below.

**Property (Linearity).**

$$\mathcal{N}: \mathbb{C}^{|A| \times |A|} \rightarrow \mathbb{C}^{|B| \times |B|}$$

is a linear map, so that

$$\mathcal{N}\left(\sum_i p_i \rho_i\right) = \sum_i p_i \mathcal{N}(\rho_i).$$

**Property (Trace preservation).**  $\mathcal{N}$  preserves the trace of the input density operator

$$\mathrm{Tr} \rho = \mathrm{Tr} \mathcal{N}(\rho).$$

This technical requirement will sometimes be relaxed to the requirement that  $\mathcal{N}$  only be *trace-non-increasing*

$$\mathrm{Tr} \rho \geq \mathrm{Tr} \mathcal{N}(\rho).$$

With a slight loss in pedantry, we will generally refer to such maps as *trace-reducing*.

In such a case,  $\mathcal{N}$  can be interpreted as a channel which is executed with some probability less than one. To introduce the third property, let us show that there is a unique way in which  $\mathcal{N}$  acts on the  $A$  part of a composite quantum system  $AC$ . It is sufficient to see what happens when acting upon part of a pure state

$$|\Psi\rangle^{AC} = \sum_{ac} d_{ac} |a\rangle^A |c\rangle^C.$$

Here, we obtain

$$\begin{aligned} (\mathcal{N} \otimes 1^C)(\Psi) &= (\mathcal{N} \otimes 1^C) \left( \sum_{a'a'c} d_{a'c} d_{ac}^* |a'\rangle \langle a| \otimes |c'\rangle \langle c| \right) \\ &= \sum_{a'a'c} d_{a'c} d_{ac}^* \mathcal{N}(|a'\rangle \langle a|) \otimes |c'\rangle \langle c|. \end{aligned}$$

The action of  $\mathcal{N} \otimes 1^C$  is then uniquely defined on any density matrix  $\omega^{AC}$  by first writing any pure state decomposition

$$\omega^{AC} = \sum_i p_i \Psi_i^{AC}.$$

Then by linearity,

$$(\mathcal{N} \otimes 1^C)(\omega) = \sum_i p_i (\mathcal{N} \otimes 1^C)(\Psi_i).$$

Now, we can mention the third characteristic property of a quantum channel.

**Property (Complete positivity).** *The channel must be completely positive, meaning that not only must  $\mathcal{N} : A \rightarrow B$  take nonnegative definite matrices on  $A$  to nonnegative definite matrices on  $B$ , but for any  $C$  it must take nonnegative definite matrices on  $AC$  to nonnegative definite matrices on  $BC$ .*

A physically satisfying consequence of these three properties is that if a quantum channel acts on part of a convex combination of density matrices, the resulting operator will be a density matrix. Quantum channels also obey the following locality properties, which can be derived from the above three. We will later invoke these (quite frequently) without reference.

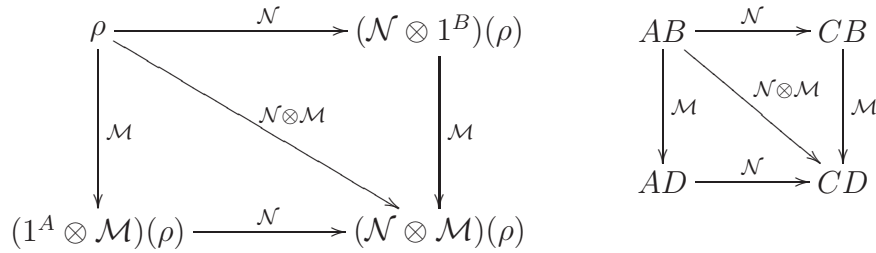
**Property (Locality I).** Given a bipartite density matrix  $\rho^{AB}$  and two quantum channels

$$\mathcal{N}: A \rightarrow C \text{ and } \mathcal{M}: B \rightarrow D,$$

the actions of  $\mathcal{N}$  and  $\mathcal{M}$  commute with one another, i.e.

$$(\mathcal{N} \otimes 1^B) \circ (1^A \otimes \mathcal{M}) = (1^A \otimes \mathcal{M}) \circ (\mathcal{N} \otimes 1^B) = \mathcal{N} \otimes \mathcal{M}.$$

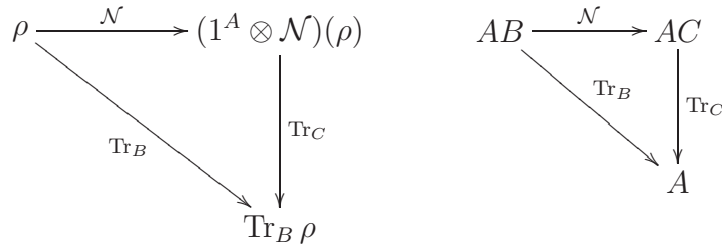
These equations are summarized by the leftmost commutative diagram below. The rightmost diagram is to remind the reader of the subsystems on which the corresponding states are defined.



**Property (Locality II).** Given a bipartite density matrix  $\rho^{AB}$ , a local operation on  $B$  will not affect the reduced density matrix on  $A$ , i.e. given a quantum channel  $\mathcal{N}: B \rightarrow C$ , we have

$$\text{Tr}_C(1^A \otimes \mathcal{N})(\rho) = \text{Tr}_B \rho.$$

This is summarized by the commutative diagram on the left below. On the right, we remind the reader of the subsystems involved.



This last property can be paraphrased as stating that  $\text{Tr}_B \rho^{AB}$  is independent of

any physical process which is carried out on  $B$ . Throughout this dissertation, we will often omit identity maps in expressions such as  $1^A \otimes \mathcal{N}$ , so that  $\mathcal{N}: B \rightarrow C$  will be interpreted as the map  $\mathcal{N}: AB \rightarrow AC$  whenever necessary. An advantage of this approach is that it allows long expressions to be simplified. This leaves no room for ambiguity, as the action of a channel on part of a larger system is always uniquely defined.

### 2.3.5 Representing quantum channels

In this section, we review two useful representation theorems for quantum channels. The first, due to Stinespring, shows how unitary processes can give rise to quantum channels. The second, due to Kraus, shows how quantum channels can be viewed as measuring devices which “forget”, or “keep secret”, the measurement result.

Suppose that a quantum system  $A$  is prepared and allowed to evolve unitarily with some extra system  $E$  which is promised to be prepared in some known pure state  $|1\rangle^E$  according to a unitary  $U: AE \rightarrow AE$ . Since the state of  $E$  is guaranteed to be in the same state before the application of  $U$ , some of the elements of  $U$  are irrelevant to the dynamics. For example, fixing bases  $\{|a\rangle^A\}$  and  $\{|e\rangle^E\}$ , suppose that  $U$  is given by

$$U = \sum_{ae} |\phi_{ae}\rangle^{AE} \langle a|^A \langle e|^E$$

for some other orthogonal basis  $\{|\phi_{ae}\rangle^{AE}\}$  of the combined system  $AE$ . Then, an arbitrary pure state of  $A$

$$|\phi\rangle^A = \sum_a \alpha_a |a\rangle^A$$

will be mapped to

$$\begin{aligned} U|\phi\rangle^A |1\rangle^E &= \sum_{ae} |\phi_{ae}\rangle^{AE} \langle a|^A \langle e|^E \left( \sum_{a'} \alpha_{a'} |a'\rangle^A \right) |1\rangle^E \\ &= \sum_{aa'e} \alpha_{a'} \langle a|a'\rangle \langle e|1\rangle |\phi_{ae}\rangle^{AE} \\ &= \sum_a \alpha_a |\phi_{a1}\rangle^{AE}. \end{aligned}$$



Thus, only the first  $|A|$  columns of  $U$  are relevant to this situation. Keeping only this “chunk” of the unitary  $U$  defines an *isometry*  $\mathcal{U}: A \rightarrow AE$ . Mathematically,  $\mathcal{V}: A \rightarrow B$  is an isometry if and only if it satisfies one (and thus both) of the following conditions:

$$\mathcal{V}^\dagger \circ \mathcal{V} = 1^A \text{ and } \mathcal{V} \circ \mathcal{V}^\dagger = \Pi_A.$$

Above,  $\Pi_A$  is a projection matrix on  $B$  satisfying  $\text{Tr } \Pi_A = |A|$ . In other words, an isometry is a length-preserving matrix whose range is a subspace of the target space, giving an image of the input space on the output space.

Returning to the isometry  $\mathcal{U}: A \rightarrow AE$ , consider what will happen if the extra system is disregarded. Given a density matrix  $\rho^A$ , a mapping  $\text{Tr}_E \mathcal{U} = \mathcal{N}: A \rightarrow A$  results. This map  $\mathcal{N}$  is a quantum channel, and the map  $\mathcal{U}$  will be called an *isometric extension* of  $\mathcal{N}$ . We will generally use a subscript to identify the channel which is being extended, saying that  $\mathcal{U}_{\mathcal{N}}$  *isometrically extends*  $\mathcal{N}$ . This way of representing a quantum channel is often referred to as the *Stinespring* representation, and we will use it almost exclusively throughout this dissertation. To be precise, we will often invoke the following proposition.

**Proposition (Isometric extension representation).** *A map  $\mathcal{N}: A \rightarrow B$  is a quantum channel if and only if there exists an isometric extension  $\mathcal{U}_{\mathcal{N}}: A \rightarrow BE$  of  $\mathcal{N}$ .*

**Remark.** In general, an isometric extension  $\mathcal{U}_{\mathcal{N}}$  of  $\mathcal{N}$  is *not unique*. This can be seen by defining  $\mathcal{U}'_{\mathcal{N}} = \mathcal{V} \circ \mathcal{U}_{\mathcal{N}}$ , where  $\mathcal{V}: E \rightarrow E'$  is any isometry into a (potentially different) environment  $E'$ . Since  $\text{Tr}_E \mathcal{U}_{\mathcal{N}} = \text{Tr}_{E'} \mathcal{V} \circ \mathcal{U}_{\mathcal{N}}$ , these extend the same channel  $\mathcal{N}$ .

Another way to represent a quantum channel is due to Kraus, and is called the operator sum representation (OSR). The following proposition was first proved in [?].

**Proposition (Operator sum representation).** *A map  $\mathcal{N}: A \rightarrow B$  is a quantum channel if and only if it can be written as*

$$\mathcal{N}(\rho) = \sum_{i=1}^k N_i \rho N_i^\dagger$$

for matrices  $\{N_i \in \mathbb{C}^{|B| \times |A|}\}$  which satisfy

$$\sum_{i=1}^k N_i^\dagger N_i = 1^A.$$

The matrices  $\{N_i\}$  are called the operator sum matrices (OSR matrices) of the representation. Such a representation of  $\mathcal{N}$  is generally not unique. It should be mentioned that this representation bears a strong resemblance to the measurement model of POVMs given in Section 2.1.6.

For a given channel, the two representations given above are intimately related, and having at hand one representation immediately gives the other as follows. If the action of  $\mathcal{N}$  can be expressed in terms of OSR matrices  $\{N_i\}_{i=1}^k$ , an isometric extension  $\mathcal{U}_{\mathcal{N}}: A \rightarrow BE$  into an environment of size  $|E| = k$  can be constructed as

$$\mathcal{U}_{\mathcal{N}} = \sum_{i=1}^k |i\rangle^E \otimes N_i.$$

This is perhaps easier expressed by writing  $\mathcal{U}_{\mathcal{N}}$  as a block matrix (in the flattened representation), with blocks given by the OSR matrices as

$$\mathcal{U}_{\mathcal{N}} = \begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ N_k \end{pmatrix}.$$

Note that the dimensions match up; namely,  $\mathcal{U}_{\mathcal{N}} \in \mathbb{C}^{|E| \cdot |B| \times |A|}$ . The reverse is also true, and the construction just involves identifying the OSR matrices with the corresponding blocks of a given isometric extension  $\mathcal{U}_{\mathcal{N}}$ .

**Remark.** As the nonuniqueness of isometric extensions is due to the isometric freedom in describing the environment, the operator sum representation inherits this freedom as well.

### 2.3.6 Complementary channels

Suppose that a channel  $\mathcal{N}: A \rightarrow B$  is given. Fixing an isometric extension  $\mathcal{U}_{\mathcal{N}}: A \rightarrow BE$  of  $\mathcal{N}$ , define the channel  $\mathcal{N}^c: A \rightarrow E$  via  $\mathcal{N}^c = \text{Tr}_B \mathcal{U}_{\mathcal{N}}$ . We will say that the channel  $\mathcal{N}^c$  is *complementary* to  $\mathcal{N}$ . If the channel acts on a density matrix  $\rho^A$ , the state  $\mathcal{N}^c(\rho)$  on  $E$  can be thought of as the disturbance induced into an initially pure environment by the action of the channel.

**Remark.** While the choice of complementary channel is generally not unique, it is unique up to isometries on  $E$ , inheriting this freedom from the choice of isometric extension.

### 2.3.7 Controlled quantum channels and cq $\rightarrow$ q channels

Consider a collection of quantum channels  $\{\mathcal{M}_x: A \rightarrow B\}_{x \in \mathcal{X}}$ , labeled by a finite set  $\mathcal{X}$ . Introducing a controlling classical system  $X$ , available at the input and output, the collection of channels can be represented by a *controlled channel*  $\mathcal{M}: XA \rightarrow XB$ . This channel acts on a cq state

$$\sigma^{XB} = \sum_{x \in \mathcal{X}} p(x) |x\rangle\langle x|^X \otimes \sigma_x^B$$

as

$$\mathcal{M}(\sigma) = \sum_{x \in \mathcal{X}} p(x) |x\rangle\langle x|^X \otimes \mathcal{M}_x(\sigma_x).$$

If the controlling system  $X$  is not available at the output, the action of the channel is modified to

$$\mathcal{M}'(\sigma) = \text{Tr}_X \mathcal{M}(\sigma) = \sum_{x \in \mathcal{X}} p(x) \mathcal{M}_x(\sigma_x).$$

We will show next that for *any* quantum channel  $\mathcal{N}: XB \rightarrow C$  which is only intended to act on cq states, less data is required to specify the action of the channel. In such a case, the channel can be represented in the same fashion as  $\mathcal{M}'$ , in the sense

that the action of  $\mathcal{N}$  on  $\sigma^{XB}$  decomposes as

$$\mathcal{N}(\sigma) = \sum_{x \in \mathcal{X}} p(x) \mathcal{N}_x(\sigma_x)$$

for some channels  $\{\mathcal{N}_x: A \rightarrow B\}_{x \in \mathcal{X}}$ . To see this, suppose that  $\mathcal{N}: XB \rightarrow C$  has an operator sum decomposition

$$\mathcal{N}: \tau \rightarrow \sum_{i=1}^d N_i \tau N_i^\dagger,$$

where the  $|C| \times |\mathcal{X}| \cdot |B|$ -dimensional matrices  $N_i$  satisfy  $\sum_{i=1}^d N_i^\dagger N_i = 1^{XB}$ . Consider each  $N_i$  to be composed of  $|\mathcal{X}|$  blocks of size  $|C| \times |B|$ , as

$$N_i = \begin{pmatrix} N_{i1} & N_{i2} & \cdots & N_{i|\mathcal{X}|} \end{pmatrix}.$$

The action of  $\mathcal{N}$  on  $\sigma$  then simplifies as

$$\begin{aligned} \mathcal{N}(\sigma) &= \sum_{i=1}^d N_i \sigma N_i^\dagger \\ &= \sum_{i=1}^d \sum_{x \in \mathcal{X}} p(x) N_{ix} \sigma_x N_{ix}^\dagger \\ &= \sum_{x \in \mathcal{X}} p(x) \sum_{i=1}^d N_{ix} \sigma_x N_{ix}^\dagger \\ &\equiv \sum_{x \in \mathcal{X}} p(x) \mathcal{N}_x(\sigma_x), \end{aligned}$$

where in the last step we identify, for each  $x$ , the matrices  $\{N_{ix}\}_{x \in \mathcal{X}}$  as the components of a trace preserving map  $\mathcal{N}_x$ .

### 2.3.8 Quantum instruments ( $\mathfrak{q} \rightarrow \mathfrak{cq}$ channels)

A *quantum instrument* [12]  $\mathcal{N}: A \rightarrow BX$  is a quantum channel whose output is a cq system. Mathematically, it is specified by collection of completely positive, trace-reducing channels  $\{\mathcal{N}_x: A \rightarrow B\}_{x \in \mathcal{X}}$ , labeled by a finite set  $\mathcal{X}$ , such that the sum  $\mathcal{N} = \sum_x \mathcal{N}_x$ , which acts on an arbitrary input state  $\rho^A$  as

$$\mathcal{N}(\rho) = \sum_x \mathcal{N}_x(\rho),$$

is trace preserving (and is thus a quantum channel). The action of the instrument on  $\rho^A$  is given by

$$\mathcal{N}(\rho) = \sum_x |x\rangle\langle x| \otimes \mathcal{N}_x(\rho).$$

The measurement process can be modeled by a quantum instrument as follows. Given a POVM  $\{\Lambda_x\}_{x \in \mathcal{X}}$ , consider the quantum instrument  $\mathcal{N}: A \rightarrow AX$  with components acting as

$$\mathcal{N}_x(\rho) = \sqrt{\Lambda_x} \rho \sqrt{\Lambda_x}.$$

Then, the action of  $\mathcal{N}$  is just

$$\begin{aligned} \mathcal{N}(\rho) &= \sum_x |x\rangle\langle x| \otimes \sqrt{\Lambda_x} \rho \sqrt{\Lambda_x} \\ &= \sum_x \text{Tr} \Lambda_x \rho |x\rangle\langle x| \otimes \frac{\sqrt{\Lambda_x} \rho \sqrt{\Lambda_x}}{\text{Tr} \Lambda_x \rho} \\ &\equiv \sum_x p(x) |x\rangle\langle x| \otimes \rho_x, \end{aligned}$$

where the  $\{\rho_x\}$  are the post-measurement states. In other words,  $\mathcal{N}_x(\rho)$  is an unnormalized density matrix satisfying

$$\text{Tr} \mathcal{N}_x(\rho) = p(x)$$

which is proportional to the post-measurement state. We will later utilize such a quantum instrument in order to simultaneously decode classical and quantum information which have been transmitted over a quantum multiple access channel.

It is also possible to use an instrument to model a measurement which ignores the post-measurement state. This is done with a *measuring instrument*  $\mathcal{M}: \rightarrow X$ , which can be defined in terms of the previous instrument as  $\text{Tr}_A \mathcal{N}$ . This simpler instrument acts as

$$\begin{aligned} \mathcal{M}(\rho) &= \text{Tr}_A \sum_x |x\rangle\langle x| \otimes \sqrt{\Lambda_x} \rho \sqrt{\Lambda_x} \\ &= \sum_x (\text{Tr}_A \Lambda_x) |x\rangle\langle x| \\ &= \sum_x p(x) |x\rangle\langle x|, \end{aligned}$$

which is exactly as one would expect a measuring device to act.

As an instrument is also a channel, it makes sense to speak of an isometric extension and complementary channel to an instrument. In the appendix (Section 11.1), we will demonstrate that any channel complementary to an instrument is another instrument with similar structure. Namely, the components of the complementary instrument are obtained as complements of the components of the original instrument.

# Chapter 3

## Entropy and information quantities

In this chapter, we review the notion of quantum entropy, as well as some related information theoretical quantities which characterize the capacities to be introduced later.

### 3.1 Entropy

Let  $\mathcal{X}$  be a finite set, and let  $X$  be a  $\mathcal{X}$ -valued random variable, distributed according to  $p(x)$ . The *Shannon entropy* of  $X$  is defined as

$$H(X) = - \sum_x p(x) \log p(x).$$

All logarithms in this dissertation will be to the base 2 ( $\log \equiv \log_2$ ). Also, note that we will always take  $0 \log 0 = 0$ , as  $\lim_{x \rightarrow 0} x \log x = 0$  by continuity. Further note that  $H(\cdot)$  does not depend on the values taken by  $X$ . Rather, it is a functional of the probability mass function  $p(x)$  of  $X$ . Indeed,  $\mathcal{X}$  is merely abstract set whose elements are merely labels for events. For example,  $X$  may be taken to represent the result of a fair coin flip, whereby  $\mathcal{X} = \{\text{heads}, \text{tails}\}$  and  $p(\text{heads}) = p(\text{tails}) = \frac{1}{2}$ . In this case,  $H(X) = 1$  bit. One interpretation to be gained from this example is that we obtain a bit of information by learning the result of a fair coin flip. In this sense, the coin flip example defines a “unit of information” equal to 1 bit.

$H(X)$  can also be interpreted as the number of bits, on average, required to represent the random variable  $X$ . Intuitively, entropy may be thought of as a measure of the amount of “information contained in” the random variable  $X$ . By definition, this is a statement concerning the asymptotic statistics of sequences of i.i.d. random variables  $X^n = (X_1, \dots, X^n)$ . Such an operational definition has its roots in the source coding theorem, which dates to Shannon’s original paper [44], where the entropy was established as the fundamental limit on the compressibility of information. As this dissertation will focus on the closely related problem of *channel coding*, we will not pursue this interpretation further.

Suppose that a quantum system is prepared with density matrix  $\rho$ . We define the *von Neumann entropy* of  $\rho$  as

$$H(\rho) \equiv -\text{Tr } \rho \log \rho.$$

Note that we overload the letter  $H$  to mean both Shannon and von Neumann entropy. Writing an eigendecomposition of  $\rho$  as

$$\rho = \sum_x p(x) |x\rangle\langle x|$$

we obtain an ensemble of orthogonal pure states  $\{p(x), |x\rangle\}$  which also gives rise to the density matrix  $\rho$ . The von Neumann entropy of  $\rho$  is then equal the Shannon entropy of the eigenvalues of  $\rho$ . Indeed,

$$\begin{aligned} H(\rho) &= -\text{Tr} \left( \sum_x p(x) |x\rangle\langle x| \right) \left( \sum_x \log p(x) |x\rangle\langle x| \right) \\ &= -\text{Tr} \left( \sum_x p(x) \log p(x) |x\rangle\langle x| \right) \\ &= -\sum_x p(x) \log p(x) = H(X) \end{aligned}$$

where  $X$  is a random variable with probability mass function  $p(x)$ .

If  $\rho^A$  is associated with system  $A$ , we will often write  $H(\rho) = H(A)_\rho$ , omitting the subscript when the state is apparent from the context. Given some multipartite



state  $\Omega^{AB}$ , the above notation gives a useful way to denote entropies of partial traces of  $\Omega$ . For example,  $H(A)_\Omega = H(\text{Tr}_B \Omega)$ , while  $H(AB)_\Omega = H(\Omega)$ . We now state the following elementary properties of entropy. These are proved in many introductory textbooks such as [34].

**Property (Entropy is nonnegative).**

$$H(A) \geq 0$$

This bound is saturated if and only if  $A$  is in a pure state  $|\phi\rangle^A$ .

**Property (Entropy is bounded).**

$$H(A) \leq \log |A|$$

This bound is saturated if and only if  $A$  is prepared in a *maximally mixed state*

$$\pi^A \equiv \frac{1}{|A|} 1^A.$$

**Property (Entropy is subadditive).**

$$H(AB) \leq H(A) + H(B)$$

This bound is saturated if and only if  $AB$  is prepared in a *product state*  $\rho^A \otimes \sigma^B$ .

**Property (Lieb's inequality).**

$$|H(A) - H(B)| \leq H(AB)$$

Let us compute the entropy of a generic cq state

$$\rho^{XA} = \sum_x p(x) |x\rangle\langle x| \otimes \rho_x^A. \quad (3.1)$$

To do so, we first diagonalize each  $\rho_x$  as

$$\rho_x = \sum_y p_x(y) |y_x\rangle \langle y_x|, \quad (3.2)$$

where for each  $x$ , the vectors  $\{|y_x\rangle\}_{y_x=1}^{|A|}$  form (generally) different orthonormal bases for  $A$ . Then, we write

$$\begin{aligned} H(XA) &= -\text{Tr} \left( \sum_x p(x) |x\rangle \langle x| \otimes \rho_x \right) \log \left( \sum_x p(x) |x\rangle \langle x| \otimes \rho_x \right) \\ &= -\text{Tr} \sum_x p(x) |x\rangle \langle x| \otimes \left( \rho_x \log (p(x) \rho_x) \right) \\ &= -\sum_x p(x) \text{Tr} \left( \rho_x \log (p(x) \rho_x) \right) \\ &= -\sum_{xy} p(x) p_x(y) \log(p(x) p_x(y)) \\ &= -\sum_x p(x) \left( \log p(x) + \sum_y p_y(x) \log p_y(x) \right) \\ &= H(X) + \sum_x p(x) H(\rho_x). \end{aligned}$$

Together with subadditivity, the calculation of the joint entropy of a cq state allows a simple proof of the convexity of entropy [34].

**Property (Convexity of entropy).**

$$\sum_x p_x H(\rho_x) \leq H\left(\sum_x p_x \rho_x\right).$$

*Proof.* Consider the cq state  $\rho^{XA}$  from (3.1). Beginning with subadditivity, we have

$$\begin{aligned} H(X) + \sum_x p(x) H(\rho_x) &= H(XA)_\rho \\ &\leq H(X) + H(A) \\ &= H(X) + H\left(\sum_x p_x(\rho_x)\right). \end{aligned}$$

Subtracting  $H(X)$  from each side completes the argument. ■

**Property (Invariance of entropy).** For any density matrix  $\rho^A$  and any isometry  $\mathcal{V}: A \rightarrow B$ ,

$$H(A)_\rho = H(B)_{\mathcal{V}(\rho)}.$$

*Proof.* The eigenvalues of  $\rho$  and of  $\mathcal{V}(\rho)$  are the same. ■

## 3.2 Conditional entropy

Let us begin by making the following formal definition for the *conditional entropy*

$$H(A|B) = H(AB) - H(B).$$

By the calculation of  $H(XA)$  for the cq state (3.1) from the previous section,

$$H(A|X) = \sum_x p(x)H(\rho_x).$$

Observe that  $H(A|X)$  is equal to the the average entropy of  $A$ , averaged over the classical part of the cq state. In classical information theory, conditional entropy is often *defined* as

$$H(Y|X) = - \sum_{xy} p(x, y) \log p(y|x).$$

It interesting to note that if we start with the cq state  $\rho^{XA}$  from (3.1), we may define a random variable  $Y$  which is jointly distributed with  $X$  in accordance with the conditional distribution  $p(y|x) \equiv p_x(y)$ , using the notation from (3.2). The equality  $H(A|X=x) = H(Y|X=x)$  holds, and thus  $H(A|X) = H(Y|X)$  holds as well.

However, for an arbitrary state on  $AB$ , this interpretation of  $H(A|B)$  as an average entropy is not valid. In particular, suppose that  $|A| = |B| = 2$ , and that  $AB$  is in a pure state

$$|\Psi\rangle^{AB} = \frac{1}{\sqrt{2}}(|0\rangle^A|0\rangle^B + |1\rangle^A|1\rangle^B).$$

Since  $\text{Tr}_A \Psi = \pi^B$ , it follows that for this state,

$$H(A|B) = H(\Psi) - H(\pi^B) = 0 - 1 = -1.$$

Defined in this formal way, conditional entropy can in fact be negative! As we will see in Sections 3.4 and 4.3, the negative of  $H(A|B)$ , referred to as the *coherent information*, plays a role in characterizing the quantum capacity of a quantum channel.

Let us conclude our discussion by noting the following property of conditional entropy. A proof can be found in [34].

**Property.**  $H(A|B)_\rho$  is concave as a function of  $\rho^{AB}$ .

### 3.3 Mutual Information

Given two random variables  $X$  and  $Y$ , jointly distributed according to  $p(x, y)$ , the *mutual information*  $I(X; Y)$  measures the amount of correlation between the two random variables.  $I(X; Y)$  is typically defined as an expected log likelihood ratio

$$I(X; Y) = \sum_{xy} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}.$$

Simple algebraic manipulations yield the following alternative formulas for  $I(X; Y)$ .

$$\begin{aligned} I(X; Y) &= H(X) + H(Y) - H(XY) \\ &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X). \end{aligned}$$

Given a stochastic matrix  $p(y|x)$  of conditional probabilities, further denotation of an input distribution  $p(x)$  determines a joint distribution  $p(x, y)$  for the random variables  $X$  and  $Y$ . In Section 4.1, we will see that the capacity of a classical channel with transition matrix  $p(y|x)$  is given by the expression

$$C = \max_{p(x)} I(X; Y).$$

A similar expression can be given for the capacity of a  $c \rightarrow q$  channel, in terms of the *quantum mutual information* evaluated on cq states.

Rather than define quantum mutual information in terms of a log-likelihood ratio, we opt here to give the following algebraic definition, valid for any composite quantum system  $AB$ .

$$I(A; B) = H(A) + H(B) - H(AB).$$

Using the formal definition of conditional quantum entropy from the previous section, we could have equivalently defined  $I(A; B)$  as

$$I(A; B) = H(A) - H(A|B)$$

or as

$$I(A; B) = H(B) - H(B|A).$$

Most relevant to this dissertation is the evaluation of mutual information on a cq state such as

$$\rho^{XB} = \sum_x p(x) |x\rangle\langle x|^X \otimes \rho_x^B.$$

With respect to  $\rho^{XB}$ , let us evaluate

$$\begin{aligned} I(X; B) &= H(B) - H(B|X) \\ &= H\left(\sum_x p(x)\rho_x\right) - \sum_x p(x)H(\rho_x). \end{aligned}$$

Together with the cq channel  $\mathcal{X} \rightarrow B$  defined by the conditional density matrices  $\{\rho_x^B\}$ , the cq state  $\rho^{XB}$  represents the joint distribution on the input and output of the channel, serving the same purpose that  $p(x, y) = p(x)p(y|x)$  did in the purely classical case. In fact, an analogous capacity formula is obtainable as well

$$C = \max_{p(x)} I(X; B).$$

This capacity is easily computable, as a consequence of the first of the following two convexity properties enjoyed by  $I(X; B)$ .

**Property.** For a fixed cq channel  $\mathcal{X} \rightarrow B$  defined by the conditional density matrices  $\{\rho_x^B\}$ ,  $I(X; B)$  is a concave function of  $p(x)$ .

*Proof.* As the  $\rho^B$  is linear in  $p(x)$ , and  $H(B)$  is concave in  $\rho^B$ ,  $H(B)$  is concave in  $p(x)$ . But  $H(B|X)$  is linear in  $p(x)$ , completing the argument. ■

**Property.** For a fixed input distribution  $p(x)$ ,  $I(X; B)$  is a convex function of the cq channel  $\mathcal{X} \rightarrow B$ .

*Proof.* This follows because  $H(X|B)$  is a concave function of  $\rho^{XB}$ , which is itself linear in the conditional density matrices  $\{\rho_x^B\}$ . ■

For an arbitrary quantum channel  $\mathcal{N} : A \rightarrow B$ , specification of a collection of input states  $\{\rho_x^A\}$ , or equivalently, of a cq channel with those conditional density matrices, yields a new cq channel  $\mathcal{X} \rightarrow B$  with conditional density matrices  $\{\mathcal{N}(\rho_x)\}$ . This channel is mathematically equivalent to the composed actions of the cq and quantum channels. By the discussion above, optimization over input distributions  $p(x)$  then gives the classical capacity of the newly constructed cq channel. However, the ultimate capacity of the quantum channel involves an optimization over collection of input states. Concavity of quantum mutual information in the input ensemble implies that extremal ensembles maximize capacity; such are ensembles of pure states. However, whether or not a single-letter converse can be obtained in this case remains a very important open problem in quantum information theory. As a result, the best known characterization of the capacity of a quantum channel for the transmission of classical information is

$$C(\mathcal{N}) = \lim_{k \rightarrow \infty} \frac{1}{k} \max_{XA^k} I_c(X; B^k)_\sigma$$

where for each  $k$ , the maximization is over all pure state ensembles  $\{p(x), |\phi_x\rangle^{A^k}\}$  consisting of  $|\mathcal{X}| \leq \min\{|A|, |B|\}^{2k} - 1$  states. The mutual information is evaluated with respect to the corresponding cq states

$$\sigma^{XB^k} = \sum_x p(x) |x\rangle\langle x| \otimes \mathcal{N}^{\otimes k}(\phi_x^{A^k}).$$

A state such as  $\sigma$  will be said to *arise from the channels*  $\mathcal{N}^{\otimes k}$  in the above sense.

### 3.4 Coherent Information

Suppose a channel  $\mathcal{N}: A' \rightarrow B$  is given. Fix an isometric extension  $\mathcal{U}_{\mathcal{N}}: A' \rightarrow BE$ , and let  $\mathcal{N}^c = \text{Tr}_B \mathcal{U}_{\mathcal{N}}$  the associated complementary channel. For a given input density operator  $\rho^{A'}$ , the *coherent information* is defined as

$$I_c(\rho, \mathcal{N}) = H(\mathcal{N}(\rho)) - H(\mathcal{N}^c(\rho)).$$

Since any two complementary channels are equivalent up to an isometry on  $E$ , and since isometries preserve entropy, this quantity is independent of the particular complementary channel  $\mathcal{N}^c$  chosen for the calculation.  $H(\mathcal{N}^c(\rho))$  is frequently referred to as the *entropy exchange* associated with sending a system with density matrix  $\rho$  over the channel  $\mathcal{N}$ .

Coherent information can be used to characterize the capacity of a quantum channel for transmitting quantum information as

$$Q(\mathcal{N}) = \lim_{k \rightarrow \infty} \frac{1}{k} \max_{\rho^{A'^k}} I_c(\rho, \mathcal{N}^{\otimes k}).$$

In Section 4.3 we will give an operational definition of quantum capacity, as well a discussion of the proof of this capacity formula. It should be noted that this multi-letter characterization is the most general expression known for an arbitrary quantum channel. However, as we illustrate in Sections 9.2 and 9.3, there are classes of channels for which a single-letter expression suffices.

Let us now explore other ways of writing  $I_c(\rho, \mathcal{N})$ . With respect to the joint output-environment state  $\mathcal{U}_{\mathcal{N}}(\rho)$  on  $BE$ , observe that

$$H(\mathcal{N}(\rho)) = H(B) \quad \text{and} \quad H(\mathcal{N}^c(\rho)) = H(E).$$

Then,

$$I_c(\rho, \mathcal{N}) = H(B) - H(E).$$

It is possible to write this quantity without making explicit mention of the environment. To do this, first fix any purification  $|\Psi\rangle^{AA'}$  of  $\rho^{A'}$ . Then, use this to write a global pure state

$$|\Omega\rangle^{ABE} = \mathcal{U}_{\mathcal{N}}|\Psi\rangle^{AA'}.$$

Since  $|\Omega\rangle^{ABE}$  is pure, it follows that  $H(E) = H(AB)$ . This allows us to rewrite

$$H(B) - H(E) = H(B) - H(AB) = -H(A|B).$$

**Remark.** Written this way, it is clear that coherent information can be positive or negative. However,  $Q(\mathcal{N}) \geq 0$  for every channel  $\mathcal{N}$ , as  $I_c(|\phi\rangle^{A'}, \mathcal{N}) = 0$  for every pure state  $|\phi\rangle^{A'}$ .

Observe that since any two purifications of  $\rho^{A'}$  are the same up to local unitaries on  $A$ , and such unitaries preserve  $H(AB)$ , this last expression is independent of the particular purification  $|\Psi\rangle^{AA'}$  chosen for  $\rho^{A'}$ . Further note that to compute  $-H(A|B)$ , it suffices to consider the joint state

$$\omega^{AB} = \text{Tr}_E \Omega^{ABE} = \mathcal{N}(\Psi^{AA'}).$$

It is common to write

$$I_c(A \rangle B)_\omega \equiv -H(A|B)_\omega$$

acknowledging the directionality of coherent information *from A to B*. While we will freely interchange the two notations for coherent information throughout this dissertation, we will generally write  $I_c(A \rangle B)$  when characterizing capacity regions and proving the converses for the main theorems, while the notation  $I_c(\rho, \mathcal{N})$  will be utilized more frequently in the coding theorems.

Let us review a few facts concerning coherent information.

**Property.** *For a maximally entangled state*

$$|\Phi\rangle^{AB} = \frac{1}{\sqrt{k}} \sum_{i=1}^k |i\rangle^A |i\rangle^B$$



we have

$$I_c(A \rangle B)_\Phi = \log k.$$

*Proof.*

$$H(B)_\Phi - H(AB)_\Phi = H(\pi_k) - H(\Phi) = \log k - 0.$$

■

**Property.** For any state on  $AB$ ,

$$I_c(A \rangle B) \leq \min\{H(A), H(B)\}.$$

*Proof.* We begin by observing that

$$\begin{aligned} I_c(A \rangle B) &= H(B) - H(AB) \leq H(B) \\ &\leq H(B). \end{aligned}$$

To see that  $I_c(A \rangle B) \leq H(A)$ , we start with Lieb's inequality

$$|H(B) - H(A)| \leq H(AB).$$

Getting rid of the absolute value and subtracting  $H(B)$  from each side yields

$$-H(A) \leq H(A|B).$$

Multiplying the sides by  $-1$  completes the argument. ■

**Property.** For any channel  $\mathcal{N}: A' \rightarrow B$  and any  $\rho^{A'}$ ,

$$I_c(\rho, \mathcal{N}) \leq \log |A'|.$$

*Proof.* Fix a purification  $|\Phi\rangle^{AA'}$  of  $\rho^{A'}$ . Then

$$\begin{aligned} I_c(\rho, \mathcal{N}) &= I_c(A|B)_{\mathcal{N}(\Phi)} \\ &\leq H(A)_\Phi \\ &= H(A')_\Phi \\ &\leq \log |A'|. \end{aligned}$$

■

**Property.** For fixed  $\rho^{A'}$ ,  $I_c(\rho, \mathcal{N})$  is a convex function of  $\mathcal{N}$ .

*Proof.* Fixing a purification  $|\Psi\rangle^{AA'}$  of  $\rho^{A'}$ , observe that the state  $\omega^{AB} = \mathcal{N}(\Psi)$  is linear function of  $\mathcal{N}$ . But  $H(A|B)$  is concave in  $\omega^{AB}$ , and thus in  $\mathcal{N}$ , so  $I_c(\rho, \mathcal{N}) = -H(A|B)$  is convex in  $\mathcal{N}$ . ■

**Remark.** This property is in close agreement to the corresponding statement that  $I(X; Y)$  is convex in  $p(y|x)$ . However,  $I(\rho, \mathcal{N})$  is not generally concave or convex in  $\rho$ , for a given fixed  $\mathcal{N}$ .

### 3.5 Conditional coherent information

In the appendix (Section 11.1), we show that if  $\mathcal{N}: A' \rightarrow BX$  is an instrument with components  $\{p(x)\mathcal{N}_x\}$ , then any isometric extension  $\mathcal{U}: A' \rightarrow BEX$  of  $\mathcal{N}$  can be expressed as

$$\mathcal{U} = \sum_x \sqrt{p(x)} |x\rangle^X |x\rangle^{X'} \otimes \mathcal{U}_x,$$

where the  $\mathcal{U}_x: A' \rightarrow BE'$  are isometric extensions of the  $\mathcal{N}_x$ , and  $E \equiv E'X'$ . We also verify there that  $\text{Tr}_E \mathcal{U} = \mathcal{N}$ , while  $\text{Tr}_{BX} \mathcal{U} = \mathcal{N}^c$ , where

$$\mathcal{N}^c = \sum_x p(x) |x\rangle \langle x|^{X'} \otimes \mathcal{N}_x^c.$$

Above, each component  $p(x)\mathcal{N}_x^c$  is formed from a complement  $\mathcal{N}_x^c: A' \rightarrow E'$  of the corresponding normalized component  $\mathcal{N}_x$  of  $\mathcal{N}$ . The main observation here is that the

environment  $E = E'X'$  of the instrument includes the common environment  $E'$  to the component channels  $\mathcal{N}_x^c$  as well as a part  $X'$  which purifies the classical component  $X$  of  $\mathcal{N}$ .

For any  $\rho^{A'}$ , the coherent information over  $\mathcal{N}$  can thus be expressed as

$$\begin{aligned}
I_c(\rho, \mathcal{N}) &= H(\mathcal{N}(\rho)) - H(\mathcal{N}^c(\rho)) \\
&= H\left(\sum_x p(x)|x\rangle\langle x|^X \otimes \mathcal{N}_x(\rho)\right) - H\left(\sum_x p(x)|x\rangle\langle x|^{X'} \otimes \mathcal{N}_x^c(\rho)\right) \\
&= H(X) + \sum_x p(x)H(\mathcal{N}_x(\rho)) - H(X) - \sum_x p(x)H(\mathcal{N}_x^c(\rho)) \\
&= \sum_x p(x)I_c(\rho, \mathcal{N}_x).
\end{aligned}$$

In the third line, we mirror the calculation of the entropy of a cq state performed in Section 3.2. The coherent information over  $\mathcal{N}$  is thus just the average of the coherent information over each  $\mathcal{N}_x$ . Another way to see this is to note that

$$\begin{aligned}
I_c(\rho, \mathcal{N}) &= H(\mathcal{N}(\rho)) - H(\mathcal{N}^c(\rho)) \\
&= H(BX) - H(E'X') \\
&= H(B|X) - H(E'|X).
\end{aligned}$$

A third derivation fixes a purification  $|\Psi\rangle^{AA'}$  of  $\rho^{A'}$  and defines the state

$$\begin{aligned}
|\Omega\rangle^{ABEX} &\equiv |\Omega\rangle^{ABE'X'X} \\
&= \mathcal{U}|\Psi^{AA'}\rangle,
\end{aligned}$$

noting that

$$\begin{aligned}
H(BX) - H(E) &= H(BX) - H(ABX) \\
&= -H(A|BX) \\
&= I_c(A)BX).
\end{aligned}$$

# Chapter 4

## Capacity theorems for single-user channels

In this chapter we recall various existing capacity theorems from the literature. After reviewing the proof of the capacity theorem for a classical channel, we will see that the main ingredients of that proof have counterparts for quantum channels, both for the transmission of classical and of quantum information. The common element to all of the situations is as follows. Each assumes that the sender and receiver are able to transmit an unlimited number of times over a collection of identical channels. It is useful to think of these channels as acting in parallel, as sequential transmissions can be thought of as parallel transmissions “in time”. After giving an operational definition of a set of rates at which the sender can communicate to the receiver arbitrarily well, the capacity is then *defined* to be the supremum, or least upper bound, of those achievable rates, representing the ultimate rate at which arbitrarily reliable communication can occur, provided that the channel can be used any number of times. The capacity is then described, or *characterized*, in terms of some optimization of entropic quantities over a well-defined collection of classical probabilities or quantum states. Later, when we characterize various capacity regions for quantum multiple access channels, we will invoke the single-user coding theorems for quantum channels introduced in this chapter.

## 4.1 Classical capacities of classical channels

Suppose that two parties, Alice and Bob, are connected by a large number of identical classical channels with probability transition matrix  $p(y|x)$ . This is to be interpreted as follows. At any given time, Alice can choose to send a symbol  $x \in \mathcal{X}$  to Bob. Because of noise, Bob “hears” a corrupted version of the symbol  $x$ . Specifically, he receives the symbol  $y \in \mathcal{Y}$  with the conditional probability  $p(y|x)$ . Fixing a probability distribution  $p(x)$  on Alice’s input symbols defines a random variable  $X$ . Together with the conditional probabilities  $p(y|x)$ , this yields a joint distribution  $p(x, y)$  of a pair of correlated random variables  $X$  and  $Y$ . The classical capacity of the channel  $p(y|x)$  is the logarithm of the number of distinguishable inputs, whereby Alice uses the channel many times to send Bob a message which he can ascertain arbitrarily well. Shannon [44] gave the following formula for the capacity:

$$C = \max_{p(x)} I(X; Y). \quad (4.1)$$

Mathematically, he proved that this expression equals a certain operationally defined capacity which we now review. Suppose Alice tries use the channel  $n$  times to send information to Bob at a rate of  $R$  bits per channel use. To this end, she selects a collection of *codewords*, consisting of  $2^{nR}$  sequences of input symbols  $x^n(m)$ , one sequence for each message she would like to send, and reveals them to Bob. This can be modeled by an *encoding function*

$$f: 2^{nR} \rightarrow \mathcal{X}^n.$$

Since the channel is noisy, Bob will receive a noisy version of Alice’s message, denoted  $Y^n(m)$ . Let the decoding function

$$g: \mathcal{Y}^n \rightarrow 2^{nR}$$

describe some scheme by which Bob attempts to decide which message Alice had intended for him to receive. Using this scheme, Alice and Bob have effectively created

a new channel

$$Q(\hat{m}|m) = \sum_{y^n \in g^{-1}(\hat{m})} p(y^n|f(m)),$$

whereby each message  $m \in 2^{nR}$  Alice may choose to send induces a distribution on the possible messages Bob may decode. We might allow Alice to use a stochastic encoder  $p(x^n|m)$ , in which case the effective channel would be

$$Q(\hat{m}|m) = \sum_{y^n \in g^{-1}(\hat{m})} \sum_{x^n} p(y^n|x^n)p(x^n|m).$$

If Alice sends the message  $m \in 2^{nR}$ , the probability Bob decodes the message incorrectly can be expressed in a number of ways:

$$\begin{aligned} P_e(m) &\equiv \Pr\{\widehat{M} \neq m | M = m\} \\ &= \Pr\{g(Y^n(m)) \neq m\} \\ &= 1 - Q(m|m) \\ &= \sum_{\substack{\hat{m} \in 2^{nR} \\ \hat{m} \neq m}} Q(\hat{m}|m). \end{aligned}$$

Associated to the coded channel  $Q(\hat{m}|m)$  is its *maximal probability of error*

$$P_{\max} = \max_{m \in 2^{nR}} P_e(m)$$

and its *average probability of error*

$$P_{\text{ave}} = 2^{-nR} \sum_{m \in 2^{nR}} P_e(m).$$

One may phrase the goal of successful communication as that of *simulating* a fictitious identity channel  $\text{id}: 2^{nR} \rightarrow 2^{nR}$  from Alice to Bob, where  $\text{id}(\hat{m}|m) = \delta_{\hat{m},m}$ . Perfect simulation would amount to using a zero-error code. Approximate simulation can be gauged in a number of ways. For example, one could require that either  $P_{\text{ave}}$  or  $P_{\max}$  is small. Clearly, the former will imply the latter.

Suppose that Alice chooses her message  $M$  randomly according to the distribution  $P(m)$ . If she sends her message through the identity channel to Bob, the two will hold a perfectly correlated pair of random variables  $(M, M)$ , distributed as

$$\text{dist}(M, M)_P(m, \hat{m}) = P(m)\delta_{m, \hat{m}}.$$

However, Alice will actually be sending through the coded channel  $Q(\hat{m}|m)$ , generating a pair of noisy correlated random variables  $(M, \widehat{M})$  distributed as

$$\text{dist}(M, \widehat{M})_P(m, \hat{m}) = P(m)Q(\hat{m}|m).$$

One way to judge the success of the simulation is to consider the the  $\ell_1$  norm  $\Delta(P)$  between the two distributions  $\text{dist}(M, \widehat{M})_P$  and  $\text{dist}(M, M)_P$ . This is calculated as

$$\begin{aligned} \Delta(P) &= \left| \text{dist}(M, M)_P - \text{dist}(M, \widehat{M})_P \right|_1 \\ &= \sum_{m, \hat{m}=1}^{\mu} |P(m)\delta_{m, \hat{m}} - P(m)Q(\hat{m}|m)| \\ &= \sum_{m, \hat{m}=1}^{\mu} P(m) |\delta_{m, \hat{m}} - Q(\hat{m}|m)| \\ &= \sum_{m=1}^{\mu} P(m) \left( (1 - Q(\hat{m}|m)) + \sum_{\substack{\hat{m}=1 \\ \hat{m} \neq m}}^{\mu} Q(\hat{m}|m) \right) \\ &= 2 \sum_{m=1}^{\mu} P(m) P_e(m) \\ &= 2 \mathbb{E}_P P_e(M). \end{aligned}$$

In other words, the  $\ell_1$  distance between the ideal and the actual joint distributions is precisely equal to twice the expected error probability. Observe that

$$\Delta(\text{unif}(2^{nR})) = 2P_{\text{ave}} \text{ and } \Delta(\delta_m) = 2P_e(m).$$

Further note that requiring that the maximal error probability be less than  $\epsilon$  is

equivalent to demanding that  $\Delta(\delta_m) \leq 2\epsilon$  for each  $m$ , where  $\delta_m$  is a point distribution at  $\{M = m\}$ . It is worth noting that the latter requirement is also equivalent to requiring that  $\Delta(P) \leq 2\epsilon$  for all distributions  $P(m)$ .

So, communication can be viewed in the light of generating near perfect common randomness over noisy quantum channels. We have phrased things in this way as it makes the road to quantum communication a bit easier. Rather than asking the sender and receiver to end up with classical correlations, we will see later in Section 4.3 that they attempt to build *quantum correlations*.

Any code  $(f, g)$  which encodes  $2^{nR}$  messages using  $n$  instances of a channel  $p(y|x)$  such that  $P_e(m) \leq \epsilon$  for all  $m \in 2^{nR}$  will be called an  $(R, n, \epsilon)$  maximal error code for the channel  $p(y|x)$ . A rate  $R$  is said to be *achievable* if there exists a sequence of  $(R, n, \epsilon_n)$  maximal error codes with  $\epsilon_n \rightarrow 0$ . The (operational) capacity of the channel  $p(y|x)$  is then defined to be the supremum of the set of achievable rates. Shannon's capacity theorem states that this operationally defined capacity is equal to the number  $C$ , defined in (4.1).

The channel capacity theorem is proved in two main parts. First, it is proven that for any rate  $R < C$ ,  $R$  is achievable. This is provided by a *coding theorem*, which is generally structured as follows. Given  $\epsilon > 0$  and some rate  $R < C$ , it is shown that there is a long enough blocklength  $n$  so that there exists an  $(R, n, \epsilon)$  code. As  $\epsilon$  was arbitrary, this immediately implies the existence of a sequence of such codes which achieves the rate  $R$ , corresponding to any sequence of error probabilities which go to zero. The second component is called the *converse*. In this part, it is shown that every achievable rate  $R$  satisfies  $R < C$ . These components are summarized in Figure 4.1.

One route to proving the coding theorem involves first showing that codes with a weaker error constraint exist. Rather than requiring that every message have a low error probability, it is sufficient to show that the error probability, averaged over all codewords  $m \in 2^{nR}$  is small. A code satisfying this weaker constraint will be called an *average error code*. A way to prove such a coding theorem is through the technique of random coding. For an arbitrary distribution  $p(x)$ , define the product distribution  $p(x^n) = \prod_{i=1}^n p(x_i)$ . A rate  $R$  random encoder is then defined by randomly selecting



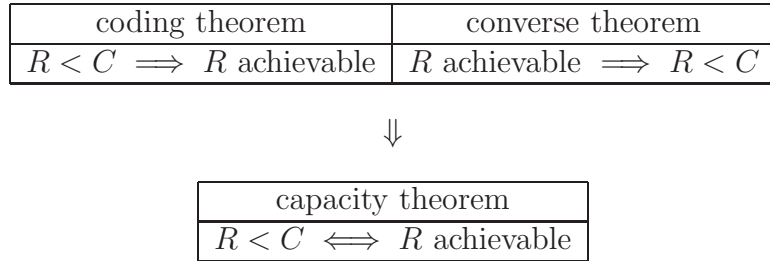


Figure 4.1: Components of a capacity theorem

$2^{nR}$  codewords

$$\mathcal{C} = \{X^n(1), \dots, X^n(2^{nR})\}$$

i.i.d. according to  $p(x^n)$ . The following coding proposition, or some variant thereof, is proved in many textbooks on information theory, such as in [10, 11].

**Proposition 0 (Classical channel coding theorem).** *Given is a channel  $p(y|x)$ , an input distribution  $p(x)$ , and a number  $0 \leq R < I(X; Y)$ , where  $I(X; Y)$  is computed with respect to  $p(x, y) = p(x)p(y|x)$ . For every  $\epsilon > 0$ , there is  $n$  sufficiently large so that if  $2^{nR}$  codewords  $\mathcal{C} = \{X^n(1), \dots, X^n(2^{nR})\}$  are chosen i.i.d. according to the product distribution  $p(x^n) = \prod_i p(x_i)$ , there exists a decoding function  $g: \mathcal{Y}^n \rightarrow 2^{nR}$  which depends on the random choice of codebook  $\mathcal{C}$  and correctly identifies the input message with expected average probability of error less than  $\epsilon$ , in the sense that*

$$\mathbb{E}_{\mathcal{C}} 2^{-nR} \sum_{m \in 2^{nR}} \Pr\{g(Y^n(m)) = m\} \geq 1 - \epsilon.$$

Observe that, because of the symmetry in the code construction, the expectation of each term in the above summation is the same. It is thus possible to reexpress that error condition as

$$\mathbb{E}_{\mathcal{C}} \Pr\{g(Y^n(1)) = 1\} \geq 1 - \epsilon,$$

showing that at the level of random codes, one may assume that the message  $m = 1$  has been sent without losing any generality.

It is a simple task to “derandomize” any code which is guaranteed to exist by Proposition 0. Suppose that Alice chooses a message uniformly distributed on the set

$\{1, \dots, 2^{nR}\}$ , represented by the random variable  $M$ , to send to Bob. Then

$$\begin{aligned} \mathbb{E}_C \Pr\{g(Y^n(M)) = M\} &= 2^{-nR} \sum_{m \in 2^{nR}} \mathbb{E}_C \Pr\{g(Y^n(m)) = m\} \\ &\geq 1 - \epsilon. \end{aligned}$$

It is then immediate that there must exist a particular deterministic code yielding an average probability of success at least as large as  $1 - \epsilon$ .

So far, this is enough to conclude that every input distribution  $p(x)$  yields a lower bound to the average error capacity of  $p(y|x)$ . This is because each  $p(x)$  corresponds to a set of achievable rates  $\{R : 0 \leq R < I(X; Y)\}$ , and the largest such set is given by optimizing over all  $p(x)$ .

Recall that we have defined the operational capacity  $C$  in terms of the maximal probability of error constraint. However, we have only outlined how to show that codes with low average error exist. By Markov's inequality from probability theory, if the average error probability is less than  $\epsilon$ , then at least half of the codewords have an error probability less than  $\sqrt{\epsilon}$ . By only using these codewords, a rate  $R - \frac{1}{n}$  code with maximal error probability  $\sqrt{\epsilon}$  is obtained, and thus every rate less than  $R$  is achievable with maximal error, showing that the maximal and average error capacities are the same.

While the coding proposition implies the existence of sequences of codes achieving any rate less than capacity, it remains to prove that no such sequences exist for rates above capacity. Rather than reproduce the entire converse theorem, we outline the basic structure of the theorem. First, one assumes that  $R$  is an achievable rate. This means that there should exist a sequence of  $(2^{nR}, n, \epsilon_n)$  codes with  $\epsilon_n \rightarrow 0$ . For any  $n$ , let  $p(x^n, y^n) = p(x^n) \prod_i p(y_i|x_i)$  be the joint distribution on  $X^n$  and  $Y^n$  induced by selecting codewords uniformly at random from the corresponding code in the sequence. An initial step in the proof shows that

$$R < \frac{1}{n} I(X^n; Y^n) + \epsilon'_n$$

where  $\epsilon'_n \rightarrow 0$  as  $\epsilon_n \rightarrow 0$ , and  $I(X^n; Y^n)$  is evaluated with respect to the induced

distribution  $p(x^n, y^n)$ . For *any* joint distribution on  $X^n$  and  $Y^n$ , the following can be easily proved:

$$\frac{1}{n}I(X^n; Y^n) \leq \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i) \leq \max_i I(X_i; Y_i).$$

If  $i^*$  achieves the maximum on the right hand side, the marginal distribution  $p(x_{i^*})$  provides a “witness” to the fact that the rate  $R$  is in fact achievable ( $R$  is thus less than the maximum mutual information over all input distributions). This proves that the capacity formula is *additive*, and thus that every achievable rate is upper bounded by the solution of a “single-letter” optimization problem. For this reason, this second conceptual step in the converse is known as single-letterization. Without it, one would only be able to write the capacity as

$$C = \lim_{k \rightarrow \infty} \frac{1}{k} \max_{p(x^k)} I(X^k; Y^k)$$

a result which follows by applying Proposition 0 to extensions of the channel

$$p(y^k | x^k) = \prod_{i=1}^k p(y_i | x_i).$$

Such an expression has become known as a “regularized” expression for the capacity. Actually, this is a persistent problem in quantum information theory. The best known expressions characterizing the capacities of an arbitrary quantum channel to transmit classical or quantum information are regularized maximizations of information quantities over appropriate sets of input states.

## 4.2 Classical capacities of quantum channels

Suppose that Alice and Bob are connected via some large number  $n$  of instances of a quantum channel  $\mathcal{N}$ , and that Alice wishes to transmit classical messages to Bob. The overall maximal rate at which this is possible is the *classical capacity*  $C(\mathcal{N})$  of the channel  $\mathcal{N}$ , which is the logarithm of the number of physical input preparations Alice can make, per channel use, so that Bob can distinguish them

arbitrarily well by measuring the induced states at the outputs of the channels. The best known expression for the classical capacity of a quantum channel, due to Holevo [25] Schumacher and Westmoreland [43], is the following regularized formula, known as the HSW Theorem:

$$C(\mathcal{N}) = \lim_{k \rightarrow \infty} \frac{1}{k} \max_{X^{A^k}} I(X; B^k)_\omega.$$

Here, the maximization is over all pure state input ensembles  $\{p(x), |\phi_x\rangle^{A^k}\}$  of states for Alice to prepare at the inputs to  $k$  parallel instances of the channel  $\mathcal{N}$ . For a given ensemble, the mutual information is computed relative to the corresponding cq state

$$\omega^{XB^k} = \sum_m |x\rangle\langle x|^X \otimes \mathcal{N}^{\otimes k}(\phi_x).$$

Operationally, the classical capacity of  $\mathcal{N}$  is defined in analogy to that of a classical channel. A  $(2^{nR}, n)$  code consists of  $2^{nR}$  message states  $\{|\phi_1\rangle^{A^k}, \dots, |\phi_{2^{nR}}\rangle^{A^k}\}$  for Alice and a corresponding measurement for Bob, mathematically modeled as POVM with  $2^{nR}$  outcomes  $\{\Lambda_m\}_{m \in 2^{nR}}$ . We call this code an  $(2^{nR}, n, \epsilon)$  code if the following constraint on success probability, averaged over all messages, is satisfied:

$$2^{-nR} \sum_{m \in 2^{nR}} \text{Tr} \Lambda_m \mathcal{N}^{\otimes n}(\phi_m) \geq 1 - \epsilon.$$

A rate  $R$  is achievable if there exists a sequence of  $(2^{nR}, n, \epsilon_n)$  codes with  $\epsilon_n \rightarrow 0$ , and the capacity  $C(\mathcal{N})$  is the supremum of all achievable rates.

As with the capacity of a classical channel, the proof that  $C(\mathcal{N})$  can be expressed in such a regularized form has two parts, a coding theorem and a converse. The following coding theorem is attributed to Holevo [25], Schumacher and Westmoreland [43].

**Proposition 1 (HSW Theorem).** *Given is a cq state  $\sigma^{XB} = \sum_x p(x) |x\rangle\langle x|^X \otimes \rho_x^B$  and a number  $0 \leq R < I(X; B)_\sigma$ . For every  $\epsilon > 0$ , there is  $n$  sufficiently large so that if  $2^{nR}$  codewords  $\mathcal{C} = \{X^n(m)\}$  are chosen i.i.d. according to the product distribution*

$p(x^n) = \prod_{i=1}^n p(x_i)$ , corresponding to input preparations

$$\rho_{x^n} = \rho_{x_1} \otimes \cdots \otimes \rho_{x_n},$$

there exists a decoding POVM  $\{\Lambda_m\}$  on  $B^n$  which depends on the random choice of codebook  $\mathcal{C}$  and correctly identifies the index  $m$  with average probability of error less than  $\epsilon$ , in the sense that

$$\mathbb{E}_{\mathcal{C}} 2^{-nR} \sum_{m=1}^{2^{nR}} \text{Tr} \rho_{X^n(m)} \Lambda_m \geq 1 - \epsilon. \quad (4.2)$$

Due to the symmetry of the distribution of  $\mathcal{C}$  under codeword permutations, it is clear that the expectations of each term in the above sum are equal. In other words,

$$\mathbb{E}_{\mathcal{C}} 2^{-nR} \sum_{m=1}^{2^{nR}} \text{Tr} \rho_{X^n(m)} \Lambda_m = \mathbb{E}_{\mathcal{C}} \text{Tr} \rho_{X^n(1)} \Lambda_1, \quad (4.3)$$

The arguments for derandomization and for obtaining a good maximal error code are identical to those used for classical channels in the previous section.

A proof of the converse begins, as before, by assuming that  $R$  is an achievable rate. Taking a cq state  $\omega^{XB^n}$  induced by an  $(R, n, \epsilon_n)$  code in the achieving sequence, Fano's inequality (Lemma 5) and the Holevo Bound (Lemma 7) are used<sup>1</sup> to show that

$$R < \frac{1}{n} I(X; B^n)_{\omega},$$

where again  $\epsilon'_n \rightarrow 0$  as  $\epsilon_n \rightarrow 0$ . However, it is an important open problem as to whether a single-letterization step can be proved. No counterexample to additivity is known, and it is widely believed that none exists.

<sup>1</sup>These details are given more explicitly in the converse proofs of the main theorems (Section 7.2).

### 4.3 Quantum capacities of quantum channels

The quantum capacity  $Q(\mathcal{N})$  of a quantum channel  $\mathcal{N}: A' \rightarrow B$  is the answer to a number of physical questions regarding the possibilities of performing various operational information processing tasks over many parallel instances of the channel  $\mathcal{N}$ .  $Q(\mathcal{N})$  is the logarithm of various quantities:

- the amount of entanglement that can be created (entanglement generation)
- the amount of entanglement that can be sent (entanglement transmission)
- the size of a Hilbert space all of whose states can be reliably transmitted (subspace transmission)
- the size of a Hilbert space all of whose entangled states can be reliably transmitted (strong subspace transmission).

All of these quantities have units of *qubits per channel use*, and as the rates at which these tasks are possible all coincide, it is justifiable to say that they all represent “sending quantum information,” and hence to speak of a single quantum capacity  $Q(\mathcal{N})$ . The best known characterization of the quantum capacity is a regularized maximization of the coherent information

$$Q(\mathcal{N}) = \lim_{k \rightarrow \infty} \frac{1}{k} \max_{XA'} I_c(A \rangle B^k)_\omega,$$

where for each  $k$ , the maximization is over all states of the form

$$\omega^{AB^k} = \mathcal{N}^{\otimes k}(\Psi^{AA'^k}).$$

Such a state  $\omega$  will be said to *arise from*  $\mathcal{N}^{\otimes k}$  or rather, to arise from the action of  $\mathcal{N}^{\otimes k}$  on the bipartite pure state  $|\Phi\rangle^{AA'^k}$ . Here, the regularization is known to be necessary for a general quantum channel, as opposed to the case with the classical capacity  $C(\mathcal{N})$ , where the existence of a single-letterization step in the converse is an open problem. The existence of a counterexample to additivity is known [46].

Of the different operational definitions of  $Q(\mathcal{N})$ , the simplest to describe is entanglement generation, since it can be defined without explicit mention of encodings. Suppose that a large number  $n$  of channels  $\mathcal{N}: A' \rightarrow B$  are available from Alice to Bob. Alice and Bob will use the channels to build a large maximally entangled state between degrees of freedom of some physical systems located in their respective laboratories. To this end, Alice prepares some bipartite pure state  $|\Upsilon\rangle^{AA'^n}$ , entangled between some system  $A$  of dimension  $|A| = 2^{nQ}$  in her laboratory, and the inputs  $A'^n$  of the channels. After the actions of the channels, Alice's system  $A$  is correlated with the outputs  $B^n$  of the channels quantum mechanically. Bob then performs some post-processing procedure, modeled by a quantum operation  $\mathcal{D}: B^n \rightarrow \hat{A}$ , to transfer the quantum correlations from the outputs  $B^n$  of the channels to an "output" physical system  $\hat{A}$ , also of dimension  $|\hat{A}| = 2^{nQ}$  in his laboratory. Their goal is to produce a state which is close to some target maximally entangled state  $|\Phi\rangle^{A\hat{A}}$ . More specifically, we say that they generate entanglement at rate  $Q$  if they produce a maximally entangled state of the form

$$|\Phi\rangle^{A\hat{A}} = \frac{1}{\sqrt{2^{nQ}}} \sum_{a \in 2^{nQ}} |a\rangle^A |a\rangle^{\hat{A}}.$$

We will call such a state a *rate  $Q$  maximally entangled state*. The blocklength  $n$  will always be apparent from the context.

$(|\Upsilon\rangle^{AA'^n}, \mathcal{D})$  will be called a  $(Q, n, \epsilon)$  *entanglement generation code* for the channel  $\mathcal{N}$  if, for the rate  $Q$  maximally entangled state  $|\Phi\rangle^{A\hat{A}}$ , we have

$$F(|\Phi\rangle^{A\hat{A}}, \mathcal{D} \circ \mathcal{N}^{\otimes n}(\Upsilon^{AA'^n})) \geq 1 - \epsilon.$$

A rate  $Q$  is an *achievable rate for entanglement generation* over the channel  $\mathcal{N}$  if there exists a sequence of  $(Q, n, \epsilon_n)$  entanglement generation codes with  $\epsilon_n \rightarrow 0$ . The *entanglement generating capacity*  $Q^{\text{eg}}(\mathcal{N})$  of  $\mathcal{N}$  is then defined operationally as the supremum of all such achievable rates.

We will now introduce a number of coding propositions from [13], each a more refined version of the previous one. While the first is sufficient to prove achievability

for single-user channels, the others have additional properties which we will need later when we characterize various capacity regions of quantum multiple access channels.

**Proposition (Entanglement generation coding theorem).** *Given is a channel  $\mathcal{N}: A' \rightarrow B$ , a density matrix  $\rho^{A'}$ , and a number  $0 \leq Q < I_c(\rho, \mathcal{N})$ . For every  $\epsilon > 0$ , there is  $n$  sufficiently large so that there is a  $(Q, n, \epsilon)$  entanglement generation code  $(|\Upsilon\rangle^{AA^n}, \mathcal{D})$  for  $\mathcal{N}$ .*

Recall the discussion in Section 3.4 regarding the two different ways of expressing coherent information. Given an input density operator  $\rho^{A'}$ , if  $|\Psi\rangle^{AA'}$  is any purification of  $\rho$ , then the identity

$$I_c(\rho, \mathcal{N}) = I_c(A \rangle B)_{\mathcal{N}(\Psi)}$$

holds. This proposition then guarantees that for every state  $\omega^{AB} = \mathcal{N}(\Psi)$  arising from the action of  $\mathcal{N}$  on a state  $|\Psi\rangle^{AA'}$ , every rate  $0 \leq Q < I_c(A \rangle B)_\omega$  is an achievable rate. This works by applying the coding theorem to the input state  $\rho^{A'} = \text{Tr}_A \Psi$ .

As with the classical capacity, it is also true that for each integer  $k > 0$ , if  $\omega'$  arises from  $\mathcal{N}^{\otimes k}$ , then every rate  $0 \leq Q < \frac{1}{k} I_c(A \rangle B^k)_{\omega'}$  is achievable as well. We then conclude that

$$Q(\mathcal{N}) \geq \lim_{k \rightarrow \infty} \frac{1}{k} \max I_c(A \rangle B^k).$$

The usual Shannon-theoretic prescription for converse theorems applies here as well, although as mentioned above, it is known that a single-letterization step cannot be proved for arbitrary  $\mathcal{N}$ . Suppose that  $Q$  is achievable, and fix a  $(Q, n, \epsilon_n)$  entanglement generation code  $(|\Upsilon\rangle^{AA^n}, \mathcal{D})$  in the achieving sequence of codes. The encoding  $|\Upsilon\rangle$  gives rise to the state  $\omega^{AB^n} = \mathcal{N}^{\otimes n}(\Upsilon)$ . It is a simple consequence of the quantum data processing inequality (Lemma 6) and continuity of coherent information in the input density operator (Lemma 3) that <sup>2</sup>

$$Q \leq \frac{1}{n} I_c(A \rangle B^n)_\omega + \epsilon'_n$$

---

<sup>2</sup>these details are given more explicitly in the converse proofs of the main theorems (Section 7.2).



where  $\epsilon'_n \rightarrow 0$ . By standard arguments we then conclude that

$$Q(\mathcal{N}) \leq \lim_{k \rightarrow \infty} \frac{1}{k} \max I_c(A \rangle B^k).$$

The state  $\text{Tr}_A \Upsilon$  which is induced by Alice's encoding at the inputs  $A^n$  of  $\mathcal{N}^{\otimes n}$  is called the *code density operator* of the entanglement generation code. With randomization, it is possible to make this operator arbitrarily close to the product state  $\rho^{\otimes n}$ , where  $\rho^{A'}$  is the input density matrix used when invoking the proposition. If Alice and Bob have access to a shared source of randomness, they may utilize an ensemble of codes to this end. This is very useful for our multiple access coding theorems, as it guarantees that if one sender codes randomly, the induced channel seen by the other sender is close to a product channel, allowing coding theorems for product channels to be invoked.

A  $(Q, n, \epsilon)$  *random entanglement generation code* consists of a collection of deterministic  $(Q, n, \epsilon)$  entanglement transmission codes  $(|\Upsilon^\beta\rangle^{AA^n}, \mathcal{D}^\beta)$  and a probability distribution  $P_\beta$ , corresponding to a source of shared common randomness available to both sender and receiver. We will often omit the subscript, once the randomness of the code has been clarified, and it will be understood that  $|\Upsilon\rangle$  and  $\mathcal{D}$  constitute a pair of classically correlated random objects. Associated to a random code is its expected, or average code density operator

$$\varrho^{A^n} = \mathbb{E}_\beta \text{Tr}_A \Upsilon = \sum_\beta P_\beta \text{Tr}_A \Upsilon^\beta$$

which is the expectation, over the shared randomness, of the state at the channel inputs  $A^n$ . The following extension of the previous coding proposition pertains to these random codes and is also proved in [13].

**Proposition (Random entanglement generation coding theorem).** *Given is a channel  $\mathcal{N}: A' \rightarrow B$ , a density matrix  $\rho^{A'}$ , and a number  $0 \leq Q < I_c(\rho, \mathcal{N})$ . For every  $\epsilon > 0$ , there is  $n$  sufficiently large so that there is a  $(Q, n, \epsilon)$  random entanglement*

generation code  $(P_\beta, |\Upsilon^\beta\rangle^{AA^n}, \mathcal{D}^\beta)$  for  $\mathcal{N}$  with average code density operator

$$\varrho^{A^n} = \mathbb{E}_\beta \text{Tr}_A \Upsilon = \sum_\beta P_\beta \text{Tr}_A \Upsilon^\beta$$

satisfying

$$|\varrho - \rho^{\otimes n}|_1 \leq \epsilon.$$

Finally, there are certain features of the decoder structure of random entanglement generation codes that are necessary for proofs which utilize quantum side information at the decoder. This final form of the coding proposition is the most powerful, utilizing features which are implicit from the proof of the coding theorem of [13]. This will be the proposition which is invoked later in the dissertation.

**Proposition 2.** *Given is a channel  $\mathcal{N}: A' \rightarrow B$ , a density matrix  $\rho^{A'}$ , and a number  $0 \leq Q < I_c(\rho, \mathcal{N})$ . For every  $\epsilon > 0$ , there is  $n$  sufficiently large so that there is a random  $(Q, n, \epsilon)$  entanglement generation code  $(P_\beta, |\Upsilon^\beta\rangle^{AA^n}, \mathcal{D}^\beta)$  for  $\mathcal{N}$  with average code density operator*

$$\varrho^{A^n} = \mathbb{E}_\beta \text{Tr}_A \Upsilon = \sum_\beta P_\beta \text{Tr}_A \Upsilon^\beta$$

satisfying

$$|\varrho - \rho^{\otimes n}|_1 \leq \epsilon.$$

Furthermore, given any particular isometric extension  $\mathcal{U}_\mathcal{N}: A' \rightarrow BE$  of  $\mathcal{N}$ , it is possible to choose isometric extensions  $\mathcal{U}_\mathcal{D}^\beta: B^n \rightarrow \widehat{A}F$  of the deterministic decoders so that

$$F(|\Phi\rangle^{A\widehat{A}}|\lambda\rangle^{E^n F}, \mathcal{U}_\mathcal{D}^\beta \circ \mathcal{U}_\mathcal{N}^{\otimes n} |\Upsilon^\beta\rangle^{AA^n}) \geq 1 - \epsilon$$

for every  $\ell$  and the same fixed pure state  $|\lambda\rangle^{E^n F}$ .

# Chapter 5

## Main results

### 5.1 Quantum multiple access channels

For this dissertation, a quantum multiple access channel will have two senders and a single receiver. While many-sender generalizations of the theorems which appear here are readily obtainable, we focus on the case with two senders for simplicity. Such a channel  $\mathcal{N}: AB' \rightarrow C$  will generally be one in which Alice and Bob simultaneously transmit to Charlie. We will assume throughout that no other resources are available to the three parties. Namely, none of the parties share any prior classical or quantum correlations between themselves, nor do they have access to any other auxiliary channels. If Alice inputs a physical system with density matrix  $\rho_1^{A'}$ , while Bob's input has density matrix  $\rho_2^{B'}$ , Charlie will receive the state  $\mathcal{N}(\rho_1 \otimes \rho_2)$ .

In the next section, we give an operational definition of the four-dimensional region  $\mathcal{S}(\mathcal{N})$ , which consists of the rates at which each sender can simultaneously send classical and quantum information to Charlie. Sections 5.3 and 5.4 state the main results of this dissertation. These results characterize the two-dimensional shadows of  $\mathcal{S}(\mathcal{N})$  corresponding to the situation where Alice sends classically while Bob sends quantum information (Theorem 1), and that where each sends quantum information (Theorem 2).

These theorems will be proved by first showing in Chapter 7 that the characterizations given in Sections 5.3 and 5.4 describe other sets of operationally defined rates,

corresponding to weaker constraints on good codes than those to be introduced in this chapter. In Chapter 8, it will ultimately be shown that the other sets of operationally defined rates equal those introduced in this chapter.

## 5.2 $\mathcal{S}(\mathcal{N})$ - the general problem

Assume that Alice and Bob are connected to Charlie by  $n$  instances of a multiple access channel  $\mathcal{N}: A'B' \rightarrow C$ , where Alice and Bob respectively have control over the  $A'^n$  and  $B'^n$  inputs. We will describe a scenario in which Alice wishes to transmit classical information at a rate of  $R_a$  bits per channel use, while simultaneously transmitting quantum information at a rate of  $Q_a$  qubits per channel use. At the same time, Bob will be transmitting classical and quantum information at rates of  $R_b$  and  $Q_b$  respectively. Alice attempts to convey any one of  $2^{nR_a}$  messages to Charlie, while Bob tries to send him one of  $2^{nR_b}$  such messages. We will also assume that the senders are presented with systems  $\tilde{A}$  and  $\tilde{B}$ , where  $|\tilde{A}| = 2^{nQ_a}$  and  $|\tilde{B}| = 2^{nQ_b}$ . Each will be required to complete the following two-fold task. Firstly, they must individually transfer the quantum information embodied in  $\tilde{A}$  and  $\tilde{B}$  to their respective inputs  $A'^n$  and  $B'^n$  of the channels, in such a way that it is recoverable by Charlie at the receiver. Second, they must simultaneously make Charlie aware of their independent messages  $M_a$  and  $M_b$ . Alice and Bob will encode with maps from the cq systems holding their classical and quantum messages to their respective inputs of  $\mathcal{N}^{\otimes n}$ , which we denote

$$\mathcal{E}_1: M_a \tilde{A} \rightarrow A'^n \quad \text{and} \quad \mathcal{E}_2: M_b \tilde{B} \rightarrow B'^n.$$

Charlie decodes with a quantum instrument

$$\mathcal{D}: C^n \rightarrow \widehat{M}_a \widehat{M}_b \widehat{A} \widehat{B}.$$

The output systems are assumed to be of the same sizes and dimensions as their respective input systems. For the quantum systems, we assume that there are pre-agreed upon unitary correspondences  $\text{id}_a: \tilde{A} \rightarrow \widehat{A}$  and  $\text{id}_b: \tilde{B} \rightarrow \widehat{B}$  between the

degrees of freedom in the quantum systems presented to Alice and Bob which embody the quantum information they are presented with and the target systems in Charlie's laboratory to which that information should be transferred. The goal for quantum communication will be to, in the strongest sense, simulate the actions of these corresponding identity channels. We similarly demand low error probability for each pair of classical messages. Formally,  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D})$  will be said to comprise an  $(R_a, R_b, Q_a, Q_b, n, \epsilon)$  *strong subspace transmission code* for the channel  $\mathcal{N}$  if for all  $m_a \in 2^{nR_a}$ ,  $m_b \in 2^{nR_b}$ ,  $|\Psi_1\rangle^{A\tilde{A}}$ ,  $|\Psi_2\rangle^{B\tilde{B}}$ , where  $A$  and  $B$  are purifying systems of arbitrary dimensions,

$$F\left(|m_a\rangle^{\widehat{M}_a}|m_b\rangle^{\widehat{M}_b}|\Psi_1\rangle^{A\widehat{A}}|\Psi_2\rangle^{B\widehat{B}}, \Omega_{m_a m_b}\right) \geq 1 - \epsilon$$

where

$$\Omega_{m_a m_b}^{\widehat{M}_a \widehat{M}_b A \widehat{A} B \widehat{B}} = \mathcal{D} \circ \mathcal{N}^{\otimes n} \left( \mathcal{E}_1(|m_a\rangle\langle m_a|^{M_a} \otimes \Psi_1^{A\tilde{A}}) \otimes \mathcal{E}_2(|m_b\rangle\langle m_b|^{M_b} \otimes \Psi_2^{B\tilde{B}}) \right).$$

We will say that a rate vector  $(R_a, R_b, Q_a, Q_b)$  is *achievable* if there exists a sequence of  $(R_a, R_b, Q_a, Q_b, n, \epsilon_n)$  strong subspace transmission codes with  $\epsilon_n \rightarrow 0$ . The simultaneous capacity region  $\mathcal{S}(\mathcal{N})$  is then defined as the closure of the collection of achievable rates. Setting various rate pairs equal to zero uncovers six two-dimensional rate regions. The next section contains our first theorem, which gives a multi-letter characterization of the two shadows relevant to the situation where one user only sends classical information, while the other only sends quantum information. The following section contains a theorem which describes the rates at which each sender can send quantum information via a multi-letter formula.

### 5.3 $\mathcal{CQ}(\mathcal{N})$ - classical-quantum capacity region

Suppose that Alice only wishes to send classical information at a rate of  $R$  bits per channel use, while Bob will only send quantum mechanically at  $Q$  qubits per use of the channel. The rate pairs  $(R, Q)$  at which this is possible comprise a classical-quantum (cq) region  $\mathcal{CQ}(\mathcal{N})$  consisting of rate vectors in  $\mathcal{S}(\mathcal{N})$  of the form  $(R, 0, 0, Q)$ . Our

first theorem gives a characterization of  $\mathcal{CQ}(\mathcal{N})$  as a regularized union of rectangles.

**Theorem 1.**  $\mathcal{CQ}(\mathcal{N}) =$  the closure of the union of pairs of nonnegative rates  $(R, Q)$  satisfying

$$\begin{aligned} R &\leq \frac{1}{k} I(X; C^k)_\omega \\ Q &\leq \frac{1}{k} I_c(B; C^k X)_\omega \end{aligned}$$

for some  $k$ , some pure state ensemble  $\{p(x), |\phi_x\rangle^{A^k}\}$  and some bipartite pure state  $|\Psi\rangle^{BB^k}$  giving rise to the state

$$\omega^{XBC^k} = \sum_x p(x) |x\rangle\langle x|^X \otimes \mathcal{N}^{\otimes k}(\phi_x \otimes \Psi). \quad (5.1)$$

Further, it is sufficient to consider ensembles for which

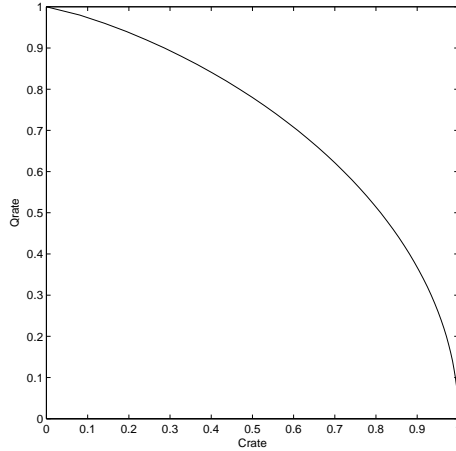
$$|\mathcal{X}| \leq \max\{|A'|, |C|\}^{2k}.$$

It should also be noted that this characterization does not apparently lead to a finite computation for determining the capacity regions, as it does not admit a single-letter characterization in general. However, as an application, the following example contains a channel for which this region is additive.

**Example.** Consider an erasure channel into which Alice inputs a classical bit (or rather, a qubit that will be dephased into the  $|0\rangle^{A'}, |1\rangle^{A'}$  basis), while Bob inputs a qubit. If Alice inputs  $|0\rangle^{A'}$ , Charlie receives Bob's qubit without error. If Alice inputs  $|1\rangle^{B'}$ , Charlie receives a pure erasure state  $|e\rangle^C$  which is orthogonal to the degrees of freedom of Bob's input state. The cq capacity region of this channel is equal to the collection of pairs of nonnegative cq rates  $(R, Q)$  which satisfy

$$\begin{aligned} R &\leq H(p) \\ Q &\leq 1 - 2p \end{aligned}$$

for some  $0 \leq p \leq \frac{1}{2}$ . This region is pictured in Figure 5.1.

Figure 5.1:  $\mathcal{CQ}$ (erasure channel)

*Proof.* In Section 9.1, we prove this for the more general case where Bob inputs a  $d$ -level quantum system. ■

**Remark.** It is also possible to characterize  $\mathcal{CQ}(\mathcal{N})$  as a regularized union of pentagons, a form which is analogous to the result of [1, 31] for classical multiple access channels. As we do not yet know an example of a channel for which this characterization is single-letter (and not equivalent to the rectangle region above), we defer further consideration of this characterization until Chapter 10.

**Remark.** The proof of the bound on  $|\mathcal{X}|$  is found in the appendix (Section 11.3).

## 5.4 $\mathcal{Q}(\mathcal{N})$ - quantum-quantum capacity region

The situation in which each sender only attempts to convey quantum information to Charlie is described by the quantum-quantum (qq) rate region  $\mathcal{Q}(\mathcal{N})$  which consists of rate vectors in  $\mathcal{S}(\mathcal{N})$  of the form  $(0, 0, Q_a, Q_b)$ . Our second theorem gives a characterization of  $\mathcal{Q}(\mathcal{N})$  as a regularized union of pentagons.

**Theorem 2.**  $\mathcal{Q}(\mathcal{N}) =$  the closure of the union of pairs of nonnegative rates  $(Q_a, Q_b)$

satisfying

$$\begin{aligned} Q_a &\leq \frac{1}{k} I_c(A \rangle BC^k)_\omega \\ Q_b &\leq \frac{1}{k} I_c(B \rangle AC^k)_\omega \\ Q_a + Q_b &\leq \frac{1}{k} I_c(AB \rangle C^k)_\omega \end{aligned}$$

for some  $k$  and some bipartite pure states  $|\Psi_1\rangle^{AA^k}$ ,  $|\Psi_2\rangle^{BB^k}$  giving rise to

$$\omega^{ABC^k} = \mathcal{N}^{\otimes k}(\Psi_1 \otimes \Psi_2). \quad (5.2)$$

**Example.** An example of a channel for which this region is single-letter is a channel into which Alice and Bob each input a qubit. With probability  $p$ , each of their qubits undergoes a phase flip, or  $180^\circ$  rotation about the  $z$ -axis, before being received by Charlie. Otherwise, Charlie receives both qubits without error. The qq capacity region of this channel is given by a single pentagon, consisting of the pairs of nonnegative qq rates  $(Q_a, Q_b)$  which satisfy

$$\begin{aligned} Q_a &\leq 1 \\ Q_b &\leq 1 \\ Q_a + Q_b &\leq 2 - H(p). \end{aligned}$$

*Proof.* See Section 9.4. ■

**Remark.** There does not appear to be any obstacle preventing application of the methods used in this paper to prove many-sender generalizations of Theorems 1 and 2. For simplicity, we have focused on the situations with two senders.

**Remark.** Contrary to the corresponding result for classical multiple access channels, the regions of Theorems 1 and 2 do not require convexification. That this follows from the multi-letter nature of the regions will be demonstrated in the appendix (Section 11.2).



# Chapter 6

## Supplementary results

In this chapter, we collect a number of auxiliary results which will be used to prove the main theorems. The first section contains some relationships satisfied by the distance measures of trace distance and fidelity which will comprise the machinery used to prove the coding theorems. The main novel contribution of that section is the statement and proof of Lemma 2. The next section contains other lemmas, proved elsewhere, which will be needed later. In the third section we review strong subadditivity of quantum entropy, and explore a number of its consequences. These include quantum versions of the classical data processing inequality, as well as the fact that conditioning decreases conditional quantum entropy or equivalently, increases coherent information. We also obtain a particularly elegant proof of the Holevo bound on the accessible information of an ensemble of quantum states.

### 6.1 Further properties of distance measures

We first collect some relevant results which will be used in what follows, starting with some relationships between our distance measures. If  $\rho$  and  $\sigma$  are density matrices defined on the same (or isomorphic) Hilbert spaces, set

$$F = F(\rho, \sigma) \text{ and } T = \|\rho - \sigma\|_1.$$

Then, the following inequalities hold (see e.g. [34])

$$1 - \sqrt{F} \leq T/2 \leq \sqrt{1 - F}, \quad (6.1)$$

$$1 - T \leq F \leq 1 - T^2/4. \quad (6.2)$$

From these inequalities, we can derive the following more useful relationships

$$F > 1 - \epsilon \Rightarrow T \leq 2\sqrt{\epsilon} \quad (6.3)$$

$$T \leq \epsilon \Rightarrow F > 1 - \epsilon, \quad (6.4)$$

which are valid for  $0 \leq \epsilon \leq 1$ . Uhlmann [47] has given the following characterization of fidelity

$$F(\rho, \sigma) = \max_{|\Psi_\rho\rangle, |\Phi_\sigma\rangle} |\langle \Psi_\rho | \Phi_\sigma \rangle|^2 = \max_{|\Psi_\rho\rangle} |\langle \Psi_\rho | \Phi_\sigma \rangle|^2$$

where the first maximization is over all purifications of each state, and the second maximization holds for any fixed purification  $|\Phi_\sigma\rangle$  of  $\sigma$ . This characterization is useful in two different ways. First, for any two states, it guarantees the existence of purifications of those states whose squared inner product equals the fidelity. Second, one can derive from that characterization the following monotonicity property [5] associated with an arbitrary trace-preserving channel  $\mathcal{N}$ ,

$$F(\rho, \sigma) \leq F(\mathcal{N}(\rho), \mathcal{N}(\sigma)) \quad (6.5)$$

An analogous property is shared by the trace distance [40],

$$|\rho - \sigma|_1 \geq |\mathcal{N}(\rho) - \mathcal{N}(\sigma)|_1, \quad (6.6)$$

which holds even if  $\mathcal{N}$  is trace-reducing. A simple proof for the trace-preserving case can be found in [34]. These inequalities reflect the fact that completely-positive maps are *contractive* and cannot improve the distinguishability of quantum states; the closer states are to each other, the harder it is to tell them apart. Another useful

property will be the multiplicativity of the fidelity under tensor products

$$F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = F(\rho_1, \sigma_1)F(\rho_2, \sigma_2). \quad (6.7)$$

Since the trace distance comes from a norm, it satisfies the triangle inequality. The fidelity does not come from a norm, but it is possible to derive the following analog by applying (6.1) and (6.2) to the triangle inequality for the trace distance

$$F(\rho_1, \rho_3) \geq 1 - 2\sqrt{1 - F(\rho_1, \rho_2)} - 2\sqrt{1 - F(\rho_2, \rho_3)}. \quad (6.8)$$

It will be possible to obtain a sharper triangle-like inequality as a consequence of the following lemma, which states that if a measurement succeeds with high probability on a state, it will also do so on a state which is close to that state in trace distance.

**Lemma 1.** *Suppose that  $\rho, \sigma, \Lambda \in \mathbb{C}^{d \times d}$ , where  $\rho$  and  $\sigma$  are density matrices, and  $0 \leq \Lambda \leq 1$ . Then,  $\text{Tr } \Lambda \sigma \geq \text{Tr } \Lambda \rho - |\rho - \sigma|_1$ .*

*Proof.*

$$\begin{aligned} \text{Tr } \Lambda \sigma &= \text{Tr } \Lambda \rho - \text{Tr } \Lambda(\rho - \sigma) \\ &\geq \text{Tr } \Lambda \rho - \max_{0 \leq \Lambda \leq 1} 2 \text{Tr } \Lambda(\rho - \sigma) \\ &= \text{Tr } \Lambda \rho - |\rho - \sigma|_1, \end{aligned}$$

where the last equality invokes a characterization of the trace distance between density matrices given in Section 2.1.4. ■

Since  $F(\phi, \rho) = \text{Tr } \phi \rho$  when  $\phi$  is a pure state, a corollary of Lemma 1 is a fact we will refer to as the “special triangle inequality.”

**Corollary (Special triangle inequality).**

$$F(\phi, \sigma) \geq F(\phi, \rho) - |\rho - \sigma|_1,$$

The following lemma can be thought of either as a type of transitivity property

inherent to any bipartite state with a component near a pure state, or as a partial converse to the monotonicity of fidelity.

**Lemma 2.** *For arbitrary quantum systems  $A$  and  $B$ , let  $|\phi\rangle^A$  be a pure state,  $\rho^B$  a density matrix, and  $\Omega^{AB}$  a density matrix of the composite system  $AB$  with partial traces  $\Omega^A = \text{Tr}_B \Omega$  and  $\Omega^B = \text{Tr}_A \Omega$ . Then*

$$F(\phi \otimes \rho, \Omega) \geq 1 - |\rho - \Omega^B|_1 - 3(1 - F(\phi, \Omega^A)).$$

*Proof.* We begin by defining the subnormalized density matrix  $\tilde{\omega}$  via the equation

$$(\phi \otimes 1)\Omega(\phi \otimes 1) = \phi \otimes \tilde{\omega}, \quad (6.9)$$

which we interpret as the upper-left block of  $\Omega$ , when the basis for  $\mathbb{C}^{|A|}$  is chosen in such a way that  $|\phi\rangle = (1, 0, \dots, 0)^T$ . Notice that  $F(\phi, \text{Tr}_B \Omega) = \text{Tr} \tilde{\omega} \equiv (1 - \epsilon)$ . Writing the normalized state  $\omega = \tilde{\omega}/(1 - \epsilon)$ , we see that it is close to  $\tilde{\omega}$  in the sense that

$$\begin{aligned} |\omega - \tilde{\omega}|_1 &\leq \epsilon |\tilde{\omega}|_1 \\ &\leq \epsilon. \end{aligned} \quad (6.10)$$

Now we write

$$\begin{aligned} \sqrt{F(\phi \otimes \rho, \Omega)} &= \text{Tr} \sqrt{\sqrt{(\phi \otimes \rho)} \Omega \sqrt{(\phi \otimes \rho)}} \\ &= \text{Tr} \sqrt{(1 \otimes \sqrt{\rho})(\phi \otimes 1)\Omega(\phi \otimes 1)(1 \otimes \sqrt{\rho})} \\ &= \text{Tr} \sqrt{(1 \otimes \sqrt{\rho})(\phi \otimes \tilde{\omega})(1 \otimes \sqrt{\rho})} \\ &= \text{Tr} \sqrt{\phi \otimes (\sqrt{\rho} \tilde{\omega} \sqrt{\rho})} \\ &= \text{Tr} \sqrt{\sqrt{\rho} \tilde{\omega} \sqrt{\rho}} \\ &= \sqrt{F(\tilde{\omega}, \rho)} \\ &= \sqrt{(1 - \epsilon)F(\omega, \rho)} \\ &\geq \sqrt{(1 - \epsilon)(1 - |\omega - \rho|_1)}. \end{aligned} \quad (6.11)$$

The first line is the definition of fidelity and the third follows from (6.9). The last equality relies on the fact that the fidelity, as we've defined it, is linear in either of its two inputs, while the inequality follows from (6.2).

Noting that  $\Omega^B \geq \tilde{\omega}$ , we define another positive operator  $\omega' = \Omega^B - \tilde{\omega}$ , which satisfies  $\text{Tr } \omega' \leq \epsilon$  and can be interpreted as the sum of the rest of the diagonal blocks of  $\Omega$ . The trace distance in the last line above can be bounded via double application of the triangle inequality as

$$\begin{aligned} |\rho - \omega|_1 &\leq |\rho - (\rho - \omega')|_1 + |(\rho - \omega') - \tilde{\omega}|_1 + |\tilde{\omega} - \omega|_1 \\ &\leq \text{Tr } \omega' + |\rho - \Omega^B|_1 + \epsilon \\ &\leq |\rho - \Omega^B|_1 + 2\epsilon, \end{aligned} \tag{6.12}$$

where the second line follows from (6.10). Combining (6.11) with (6.12), we obtain

$$\begin{aligned} F(\phi \otimes \rho, \Omega) &\geq (1 - \epsilon)(1 - |\rho - \Omega^B|_1 - 2\epsilon) \\ &\geq 1 - |\rho - \Omega^B|_1 - 3\epsilon. \end{aligned}$$

■

## 6.2 Other useful lemmas

This continuity lemma from [4] shows that if two bipartite states are close to each other, the difference between their associated coherent informations is small.

**Lemma 3 (Continuity of coherent information).** *Let  $\rho^{AB}$  and  $\sigma^{AB}$  be two states of a finite-dimensional bipartite system  $AB$  satisfying  $|\rho - \sigma|_1 \leq \epsilon$ . Then*

$$|I_c(A)B)_\rho - I_c(A)B)_\sigma| \leq 2H(\epsilon) + 4 \log |A| \epsilon,$$

where  $H(\epsilon)$  is the binary entropy function.

Next is Winter's "gentle measurement" lemma [48], which implies that a measurement which is likely to be successful in identifying a state tends not to significantly

disturb that state.

**Lemma 4 (Gentle measurement).** *Let a density matrix  $\rho^A$  be given, where  $|A|$  is finite. If  $\Lambda \in \mathbb{C}^{|A| \times |A|}$  is nonnegative with spectrum bounded above by 1, then*

$$\mathrm{Tr} \rho \Lambda \geq 1 - \epsilon$$

*implies*

$$\left| \sqrt{\Lambda} \rho \sqrt{\Lambda} - \rho \right|_1 \leq \sqrt{8\epsilon}.$$

We will also need a lemma from classical information theory which bounds the conditional entropy of two random variables with the same support in terms of the probability they are different.

**Lemma 5 (Fano's inequality).** *Let  $M, \widehat{M}$  be  $\mathcal{M}$ -valued random variables, and write  $P_e = \Pr\{M \neq \widehat{M}\}$ . Then*

$$H(M|\widehat{M}) \leq H(P_e) + P_e \log |\mathcal{M}|.$$

*Proof.* See [10]. ■

### 6.3 Strong subadditivity and its consequences

In this section, we recall an inequality which holds for *any* tripartite quantum system  $ABC$ . This inequality goes by the name *strong subadditivity*, and was originally proved in [32], stating that

$$H(AB) + H(BC) \geq H(B) + H(ABC). \quad (6.13)$$

As much has been written about the proof of strong subadditivity of quantum entropy (see e.g. [35]), we will not discuss the proof of the theorem here. Rather, we will endeavor to show here how strong subadditivity can be used as a mathematical “hammer of Thor,” enabling short and elegant proofs of many known entropy inequalities in quantum information theory. In fact, many of these results will turn out to be

*equivalent* to strong subadditivity, in the sense that the latter is easily derivable from many of them.

Begin by subtracting  $H(B) + H(BC)$  from either side of (6.13) to yield

$$H(A|B) \geq H(A|BC). \quad (6.14)$$

This inequality can be interpreted as a demonstration that *conditioning reduces entropy*. Collecting the terms on a single side yields the compact formula

$$I(A; B|C) \geq 0,$$

showing that the quantum conditional mutual information is always positive. Note that in the classical case, these inequalities are quite simple to prove [10]. For instance, positivity of  $I(X; Z|Y)$  follows from positivity of mutual information which, in turn, is a consequence of positivity of the Kullback-Leibler distance  $D(P||Q)$ .

Classically, it is simple to show that  $I(X; Z|Y) = 0$  if and only if  $X - Y - Z$  forms a Markov chain in that order. Necessary and sufficient conditions for saturation of quantum strong subadditivity were recently determined in [23], who showed that  $I(A; C|B) = 0$  if and only if

$$\rho^{ABC} = \bigoplus_x p_x \rho_x^{AB_x^A} \otimes \rho_x^{B_x^C C}$$

where  $B = \bigoplus_x B_x^A B_x^C$ . In other words, if and only if there is a local measurement that can be performed on  $B$  which determines  $x$  without disturbing the global state. Conditioned on knowing  $x$ , the global system is in a product state. Such a measurement is commonly known as a “which path measurement.”

Recall the definition of coherent information as

$$I_c(A)B) = -H(A|B).$$

Simply reexpressing (6.14) in terms of coherent information yields the inequality

$$I_c(A \rangle BC) \geq I_c(A \rangle C), \quad (6.15)$$

which can be interpreted in light of the quantum capacity theorem as saying that losing access to part of the output of a quantum channel can only decrease capacity. Observe that a similar property is obeyed by the classical mutual information, namely that

$$I(X; YZ) \geq I(X; Y).$$

More generally, coherent information can be shown to obey an analog of the classical data processing inequality (see e.g. [10]), which says that if  $X - Y - Z$  is a Markov chain, then

$$I(X; Y) \geq I(X; Z).$$

A quantum version of the data processing inequality [42] can be proved easily from strong subadditivity.

**Lemma 6 (Quantum data processing inequality).** *Let a bipartite density matrix  $\rho^{AB}$  and a channel  $\mathcal{N}: B \rightarrow C$  be given. Then*

$$I_c(A \rangle B)_\rho \geq I_c(A \rangle C)_{\mathcal{N}(\rho)}.$$

*Proof.* Choose any isometric extension  $\mathcal{U}_\mathcal{N}: B \rightarrow CE$  of  $\mathcal{N}$ . Then

$$\begin{aligned} I_c(A \rangle B)_\rho &= I_c(A \rangle CE)_{\mathcal{U}_\mathcal{N}(\rho)} \\ &\geq I_c(A \rangle C)_{\mathcal{N}(\rho)}, \end{aligned}$$

where the first step is because isometries preserve entropy, while the second is by (6.15). ■

It is thus apparent that post-processing of  $B$  can never increase coherence with  $A$ . It is also possible to derive strong subadditivity from data processing, by taking  $\mathcal{N} = \text{Tr}_C$ , so the data processing inequality is another equivalent way to express



strong subadditivity.

The quantum data processing inequality can be used to derive a more direct analog [2] of the classical data processing inequality, dealing with quantum mutual information rather than coherent information. A simple corollary of Lemma 6 is

**Corollary (Quantum mutual information data processing inequality).** *With the same conditions as in Lemma 6,*

$$I(A; B)_\rho \geq I(A; C)_{\mathcal{N}(\rho)}.$$

*Proof.* The conclusion of Lemma 6 can be rewritten in terms of conditional entropies as

$$-H(A|B)_\rho \geq -H(A|C)_{\mathcal{N}(\rho)}.$$

Adding  $H(A)$  to each side yields the required inequality. ■

As a simple consequence of this corollary, we obtain a completely elementary proof of the Holevo bound [24], an essential step in the converse part of the HSW capacity theorem.

**Lemma 7 (Holevo bound).** *Let a cq state*

$$\rho^{XB} = \sum_x p(x) |x\rangle\langle x|^X \otimes \rho_x^B$$

*be given. For any measurement on  $B$  with POVM  $\{\Lambda_y\}_{y \in \mathcal{Y}}$ , the following inequality holds.*

$$I(X; B)_\rho \geq I(X; Y).$$

*Proof.* Construct a measuring instrument  $\mathcal{N}: B \rightarrow Y$  (as in Section 2.3.8), acting as

$$\mathcal{N}(\tau^B) = \sum_y (\text{Tr } \Lambda_x \tau) |y\rangle\langle y|^Y.$$

Application of the previous version of the data processing inequality proves the result. ■

The following inequality from [34] will be useful in Section 9.2, where we give a proof that *degradable* quantum channels have single-letter quantum capacities.

**Lemma 8 (Joint subadditivity of conditional entropy).** *For any quadripartite state on  $ABCD$ , the following entropy inequality applies*

$$H(AB|CD) \leq H(A|C) + H(B|D). \quad (6.16)$$

*Proof.* Using the original formulation of strong subadditivity (6.13), we may write the following two inequalities:

$$\begin{aligned} H(ABCD) + H(C) &\leq H(AC) + H(BCD) \\ H(BCD) &\leq H(BD) + H(CD) - H(D). \end{aligned}$$

Combining these gives

$$H(ABCD) + H(C) \leq H(AC) + H(BD) + H(CD) - H(D).$$

Rearranging terms gives the required result. ■

Observe that this lemma can equivalently be expressed in terms of coherent information as

$$I_c(A \rangle C) + I_c(B \rangle D) \leq I_c(AB \rangle CD). \quad (6.17)$$

Note that if (6.17) is computed on a state of the form

$$\Omega^{ABCD} = |\phi\rangle\langle\phi|^B \otimes \omega^{ACD},$$

it follows that

$$I_c(B \rangle D) = H(D) - H(BD) = H(D) - H(D) = 0$$

and

$$I_c(AB \rangle CD) = H(CD) - H(ABCD) = H(CD) - H(ACD) = I_c(A \rangle CD),$$

implying that

$$I_c(A \rangle C) \leq I_c(A \rangle CD),$$

which is just the original strong subadditivity inequality we started with. So we see that strong subadditivity is equivalent to Lemma 8, as well as to the fact that coherent information is superadditive.

# Chapter 7

## Entanglement generation capacities

As a first step towards proving the theorems stated in Chapter 5, we introduce a less restrictive communication scenario, *entanglement generation*. While the criterion of strong subspace transmission is analogous to a classical requirement that the maximal error probability be small, the entanglement generation criterion will rather be related to an average error constraint on good codes.

**classical-quantum scenario** Alice sends classical information to Charlie at rate  $R$ , while Bob sends quantum information at rate  $Q$ . Rather than being required to transmit half of *any* quantum state Bob is presented with, Bob will only need to *create* near maximal quantum correlations with Charlie at rate  $Q$ . To this end, Bob begins by preparing a bipartite pure state  $|\Upsilon\rangle^{BB^n}$ , entangled between a physical system  $B$  located in his laboratory, and the  $B^n$  part of the inputs of  $\mathcal{N}^{\otimes n}$ .

At the same time, Charlie will only need to identify Alice's classical message with a low average error probability, averaged over all of Alice's classical messages. As with strong subspace transmission, Charlie's post-processing procedure will be modeled by a quantum instrument. While the outer bound provided by our converse theorem will apply to any decoding modeled by an instrument, the achievability proof will require a less general approach, consisting of the following steps.

In order to ascertain Alice's message  $M$ , Charlie first performs some measurement on  $C^n$ , whose statistics are given by a POVM  $\{\Lambda_m\}_{m \in 2^{nR}}$ . We let the result of that

measurement be denoted  $\widehat{M}$ , his declaration of the message sent by Alice. Based on the result of that measurement, he will perform one of  $2^{nR}$  decoding operations  $\mathcal{D}'_m: C^n \rightarrow \widehat{B}$ . These two steps can be mathematically combined to define a quantum instrument  $\mathcal{D}: C^n \rightarrow \widehat{M}\widehat{B}$  with (trace-reducing) components

$$\mathcal{D}_m: \tau \mapsto \mathcal{D}'_m(\sqrt{\Lambda_m}\tau\sqrt{\Lambda_m}).$$

The instrument acts as

$$\mathcal{D}: \tau \mapsto \sum_{m=1}^{2^{nR}} |m\rangle\langle m|^{\widehat{M}} \otimes \mathcal{D}_m(\tau),$$

and induces the trace preserving map  $\mathcal{D}: C^n \rightarrow \widehat{B}$ , acting according to

$$\mathcal{D}: \tau \mapsto \text{Tr}_{\widehat{M}} \mathcal{D}(\tau) = \sum_{m=1}^{2^{nR}} \mathcal{D}_m(\tau).$$

We again remark that this is the most general decoding procedure required of Charlie. Any situation in which he were to iterate the above steps by measuring, manipulating, measuring again, and so on, is asymptotically just as good as a single instance of the above mentioned protocol. This is because the inner and outer bounds provided by the coding theorem and converse coincide.  $(\{\phi_m\}_{m \in 2^{nR}}, \Upsilon^{BB'^n}, \mathcal{D})$  will be called an  $(R, Q, n, \epsilon)$  *cq entanglement generation code* for the channel  $\mathcal{N}$  if

$$2^{-nR} \sum_{m=1}^{2^{nR}} P_s^{\text{eg}}(m, \Upsilon) \geq 1 - \epsilon, \quad (7.1)$$

where

$$P_s^{\text{eg}}(m, \Upsilon) = F\left(|m\rangle\langle m|^{\widehat{B}}, \mathcal{D} \circ \mathcal{N}^{\otimes n}(\phi_m^{A'^n} \otimes \Upsilon^{BB'^n})\right). \quad (7.2)$$

We will say that  $(R, Q)$  is an *achievable cq rate pair for entanglement generation* if there exists a sequence of  $(R, Q, n, \epsilon_n)$  cq entanglement generation codes with  $\epsilon_n \rightarrow 0$ .

The capacity region  $\mathcal{C}\mathcal{Q}_{\text{eg}}(\mathcal{N})$  is defined to be the closure of the collection of all achievable cq rate pairs for entanglement generation.

**quantum-quantum scenario** As above, Alice and Bob are no longer required to transmit arbitrary quantum correlations with which they are presented. Rather, each has the goal of creating near-maximal entanglement with Charlie. For encoding, Alice and Bob respectively prepare the states  $|\Upsilon_1\rangle^{AA^n}$  and  $|\Upsilon_2\rangle^{BB^n}$ , entangled with the  $A^n$  and  $B^n$  parts of the inputs of  $\mathcal{N}^{\otimes n}$ . Their goal is to do this in such a way so that Charlie, after applying a suitable decoding operation  $\mathcal{D}: C^n \rightarrow \widehat{A}\widehat{B}$ , can hold the  $\widehat{A}\widehat{B}$  part of a state which is close to  $|\Phi_1\rangle^{A\widehat{A}}|\Phi_2\rangle^{B\widehat{B}}$ . Formally,  $(\Upsilon_1^{AA^n}, \Upsilon_2^{BB^n}, \mathcal{D})$  is an  $(R, Q, n, \epsilon)$  *qq entanglement generation code* for the channel  $\mathcal{N}$  if

$$F(\Phi_1 \otimes \Phi_2, \mathcal{D} \circ \mathcal{N}^{\otimes n}(\Upsilon_1 \otimes \Upsilon_2)) \geq 1 - \epsilon. \quad (7.3)$$

$(R, Q)$  is an *achievable qq rate pair for entanglement generation* if there is a sequence of  $(R, Q, n, \epsilon_n)$  qq entanglement generation codes with  $\epsilon_n \rightarrow 0$ . The capacity region  $\mathcal{Q}_{\text{eg}}(\mathcal{N})$  is the closure of the collection of all such achievable rates.

## 7.1 The coding theorems

For any quantum multiple access channel  $\mathcal{N}: A'B' \rightarrow C$ , we first prove that the single-letter regions  $\mathcal{C}\mathcal{Q}^{(1)}(\mathcal{N})$  and  $\mathcal{Q}^{(1)}(\mathcal{N})$ , defined as the restrictions to  $k = 1$  of the respective characterizations from Sections 5.3 and 5.4, are respectively contained in  $\mathcal{C}\mathcal{Q}_{\text{eg}}(\mathcal{N})$  and in  $\mathcal{Q}_{\text{eg}}(\mathcal{N})$ . It will then follow that

$$\bigcup_{k=1}^{\infty} \frac{1}{k} \mathcal{C}\mathcal{Q}^{(1)}(\mathcal{N}^{\otimes k}) \subseteq \mathcal{C}\mathcal{Q}_{\text{eg}}(\mathcal{N}) \quad \text{and} \quad \bigcup_{k=1}^{\infty} \frac{1}{k} \mathcal{Q}^{(1)}(\mathcal{N}^{\otimes k}) \subseteq \mathcal{Q}_{\text{eg}}(\mathcal{N})$$

by applying the coding theorems to extensions  $\mathcal{N}^{\otimes k}$  of  $\mathcal{N}$ .

*Proof of Theorem 1 (coding theorem).* Our method of proof for the coding theorem will work as follows. We will employ random HSW codes and random entanglement generation codes to ensure that the average state at the input of  $\mathcal{N}^{\otimes n}$  is close to

a product state. Each sender will utilize a code designed for the product channel induced by the other's random input, whereby existing coding theorems for product channels will be invoked. The quantum code used will be one which achieves the capacity of a modified channel, in which the classical input is copied, without error, to the output of the channel. As the random HSW codes will exactly induce a product state input, the existence of these quantum codes will follow directly from Proposition 2.

The random HSW codes will be those which exist for product channels. As random entanglement generation codes exist with average code density matrix arbitrarily close to a product state, this will ensure that the resulting output states are distinguishable with high probability. Furthermore, obtaining the classical information will be shown to cause but a small disturbance in the overall joint quantum state of the system. As we will show, it is possible to mimic the channel for which the quantum code is designed by placing the identities of the estimated classical message states into registers appended to the outputs of each channel in the product.

The decoder for the modified channel will then be shown to define a quantum instrument which satisfies the success condition for a cq entanglement transmission code, on average. This feature will then be used to infer the existence of a particular, deterministic code which meets the same requirement.

Fix a pure state ensemble  $\{p(x), |\phi_x\rangle^{A'}\}$  and a bipartite pure state  $|\Psi\rangle^{BB'}$  which give rise to the cq state

$$\omega^{XBC} = \sum_x p(x) |x\rangle\langle x|^X \otimes (1^B \otimes \mathcal{N})(\phi_x^{A'} \otimes \Psi^{BB'}),$$

which has the form of (5.1). Define  $\rho_1^{A'} = \sum_x p(x) \phi_x$  and  $\rho_2^{B'} = \text{Tr}_B \Psi$ . We will demonstrate the achievability of the corner point  $(I(X; C), I_c(B) \rangle CX)_\omega$  by showing that for every  $\epsilon, \delta > 0$ , if  $R = I(X; C)_\omega - \delta$  and  $Q = I_c(B) \rangle CX)_\omega - \delta$ , there exists an  $(R, Q, n, \epsilon)$  cq entanglement generation code for the channel  $\mathcal{N}$ , provided that  $n$  is sufficiently large. The rest of the region will follow by timesharing.

For encoding, Alice will choose  $2^{nR}$  sequences  $\mathcal{C} = \{X^n(m)\}_{m \in 2^{nR}}$ , i.i.d. according to the product distribution  $p(x^n) = \prod_{i=1}^n p(x_i)$ . As each sequence corresponds to a

preparation of channel inputs  $|\phi_m\rangle^{A'^n} = |\phi_{X_1(m)}\rangle \otimes \cdots \otimes |\phi_{X_n(m)}\rangle$ , the expected average density operator associated with Alice's input to the channel is precisely

$$\mathbb{E}_{\mathcal{C}} 2^{-nR} \sum_{m=1}^{2^{nR}} |\phi_m\rangle\langle\phi_m| = \sum_{x^n} p(x^n) |\phi_{x^n}\rangle\langle\phi_{x^n}| = \rho_1^{\otimes n}.$$

Define a new channel  $\mathcal{N}': B' \rightarrow C\hat{X}$  (which is also an instrument) by

$$\mathcal{N}': \rho \mapsto \sum_x p(x) \mathcal{N}(\phi_x \otimes \rho) \otimes |x\rangle\langle x|^{\hat{X}},$$

This can be interpreted as a channel which reveals the identity of Alice's input state to Charlie, with the added assumption that Alice chooses her inputs at random. Alternatively, one can view this as a channel with state information available to the receiver, where nature is randomly choosing the "state"  $x$  at Alice's input. By Proposition 2, there exists a  $(Q, n, \epsilon)$  random entanglement generation code  $\{q_\beta, |\Upsilon^\beta\rangle^{AA'^n}, \mathcal{D}^\beta\}$  for the channel  $\mathcal{N}'$  with average code density operator  $\varrho^{B'^n} = \sum_\beta q_\beta \text{Tr}_A \Upsilon^\beta$  satisfying

$$|\varrho - \rho_2^{\otimes n}|_1 \leq \epsilon.$$

In what follows, we will use the shorthand  $|\Upsilon\rangle$  for the random vector which takes the value  $|\Upsilon^\beta\rangle$  with probability  $q_\beta$ . We further abbreviate

$$\mathbb{E}_\beta \mathcal{M}(\Upsilon) \equiv \sum_\beta q_\beta \mathcal{M}(\Upsilon^\beta),$$

where  $\mathcal{M}$  is any function of the random vector  $\Upsilon$ .

Now, by Proposition 1, for the channel  $\mathcal{N}_1: \rho \mapsto \mathcal{N}(\rho \otimes \rho_2)$  which would result if Bob's average code density operator were *exactly* equal to  $\rho_2^{\otimes n}$ , there exists a decoding POVM  $\{\Lambda_m\}_{m \in 2^{nR}}$  which would identify Alice's index  $m$  with expected average probability of error less than  $\epsilon$ , in the sense that

$$\mathbb{E}_{\mathcal{C}} 2^{-nR} \sum_{m=1}^{2^{nR}} \text{Tr} \Lambda_m \tau'_m \geq 1 - \epsilon,$$



where

$$\tau'_m = \mathcal{N}^{\otimes n}(\phi_m \otimes \rho_2^{\otimes n}).$$

By the symmetry of the random code construction, we utilize (4.3) to write this as

$$\mathbb{E}_{\mathcal{C}} \text{Tr} \Lambda_1 \tau'_1 \geq 1 - \epsilon.$$

Define the *actual* output of the channel corresponding to  $M = m$  as

$$\tau_m = \mathcal{N}^{\otimes n}(\phi_m \otimes \text{Tr}_A \Upsilon),$$

as well as its extension

$$\xi_m^{BC^n} = \mathcal{N}^{\otimes n}(\phi_m \otimes \Upsilon),$$

where  $|\Phi\rangle^{B\tilde{B}}$  is the maximally entangled state which Bob is required to transmit. Note that

$$\mathbb{E}_{\beta} \tau_m = \mathbb{E}_{\beta} \text{Tr}_B \xi_m = \mathcal{N}^{\otimes n}(\phi_m \otimes \varrho).$$

It follows from monotonicity of trace distance that

$$|\mathbb{E}_{\beta} \tau_1 - \tau'_1|_1 \leq \epsilon,$$

which, together with Lemma 1, implies that

$$\mathbb{E}_{\mathcal{C}} 2^{-nR} \sum_{m=1}^{2^{nR}} \text{Tr} \Lambda_m \mathbb{E}_{\beta} \tau_m = \mathbb{E}_{\mathcal{C}\beta} \text{Tr} \Lambda_1 \tau_1 \geq 1 - 2\epsilon.$$

This allows us to bound the expected probability of correctly decoding Alice's message as

$$\mathbb{E}_{\mathcal{C}\beta} \text{Tr}(1 \otimes \Lambda_1) \xi_1 \geq 1 - 2\epsilon. \tag{7.4}$$

In order to decode, Charlie begins by performing the measurement  $\{\Lambda_m\}_{m \in 2^{nR}}$ . He declares Alice's message to be  $\widehat{M} = m$  if measurement result  $m$  is obtained. Charlie will then attempt to simulate the channel  $\mathcal{N}'^{\otimes n}$ , by associating a separate

classical register  $\widehat{X}_i$  to each channel  $\mathcal{N}: A'_i \rightarrow C_i$  in the product, preparing the states  $|X_i(m)\rangle^{\widehat{X}_i}$ , for each  $1 \leq i \leq n$ . Additionally, he stores the result of the measurement in the system  $\widehat{M}$ , his declaration of the message intended by Alice. This procedure results in the global state

$$\Gamma^{BC^n \widehat{X}^n \widehat{M}} = \sum_{m=1}^{2^{nR}} \left(1 \otimes \sqrt{\Lambda_m}\right) \xi_1 \left(1 \otimes \sqrt{\Lambda_m}\right) \otimes |X^n(m)\rangle \langle X^n(m)|^{\widehat{X}^n} \otimes |m\rangle \langle m|^{\widehat{M}}.$$

Let  $\Theta^{BC^n \widehat{X}^n} = \text{Tr}_{\widehat{M}} \Gamma$ . If Charlie was able to perfectly reconstruct Alice's classical message,  $\Gamma$  would instead be

$$\Gamma' = \xi_1 \otimes |X^n(1)\rangle \langle X^n(1)|^{\widehat{X}^n} \otimes |1\rangle \langle 1|^{\widehat{M}},$$

with  $\Theta' = \text{Tr}_{\widehat{M}} \Gamma'$ . When averaged over Alice's random choice of HSW code,  $\Theta'$  is precisely equal to the state which would arise via the action of the modified channel  $\mathcal{N}'$ . This is because

$$\begin{aligned} \mathbb{E}_{\mathcal{C}} \Theta' &= \sum_{x^n} p(x^n) \xi_{x^n} \otimes |x^n\rangle \langle x^n|^{\widehat{X}^n} \\ &= \mathcal{N}'^{\otimes n}(\Upsilon), \end{aligned} \tag{7.5}$$

where we have written the joint state which results when Alice prepares  $\phi_{x^n}$  as

$$\xi_{x^n}^{BC^n} = \mathcal{N}^{\otimes n}(\phi_{x^n} \otimes \Upsilon).$$

However, our choice of a good HSW code ensures that he can almost perfectly reconstruct Alice's message. A consequence of this will be that the two states  $\Theta$  and  $\Theta'$  are almost the same, as we will now demonstrate.

In what follows, we will need to explicitly keep track of the randomness in our

codes, by means of superscripts which are to be interpreted as indexing the deterministic codes which occur with the probabilities  $p_c$  and  $q_\beta$ . Rewriting (7.4) as

$$\sum_{c\beta} p_c q_\beta \operatorname{Tr} (1 \otimes \Lambda_1^c) \xi_1^{c\beta} \geq 1 - 2\epsilon,$$

it is clear that we may write

$$\operatorname{Tr} (1 \otimes \Lambda_1^c) \xi_1^{c\beta} \geq 1 - \epsilon_{c\beta},$$

for positive numbers  $\{\epsilon_{c\beta}\}$  chosen to satisfy

$$\sum_{c\beta} p_c q_\beta \epsilon_{c\beta} = 2\epsilon.$$

By the gentle measurement lemma,

$$\left| \left( 1 \otimes \sqrt{\Lambda_1^c} \right) \xi_1^{c\beta} \left( 1 \otimes \sqrt{\Lambda_1^c} \right) - \xi_1^{c\beta} \right|_1 \leq \sqrt{8\epsilon_{c\beta}},$$

and thus, by the concavity of the square root function,

$$\begin{aligned} & \mathbb{E}_{c\beta} \left| \left( 1 \otimes \sqrt{\Lambda_1} \right) \xi_1 \left( 1 \otimes \sqrt{\Lambda_1} \right) - \xi_1 \right|_1 \\ &= \sum_{c\beta} p_c q_\beta \left| \left( 1 \otimes \sqrt{\Lambda_1^c} \right) \xi_1^{c\beta} \left( 1 \otimes \sqrt{\Lambda_1^c} \right) - \xi_1^{c\beta} \right|_1 \\ &\leq 4\sqrt{\epsilon}. \end{aligned}$$

Along with (7.4) and monotonicity with respect to  $\text{Tr}_{\widehat{M}}$ , this estimate lets us write

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}\beta} |\Theta - \Theta'|_1 &\leq \mathbb{E}_{\mathcal{C}\beta} |\Gamma - \Gamma'|_1 \\
&= \mathbb{E}_{\mathcal{C}\beta} \left| \left( 1 \otimes \sqrt{\Lambda_1} \right) \xi_1 \left( 1 \otimes \sqrt{\Lambda_1} \right) - \xi_1 \right|_1 \\
&\quad + \mathbb{E}_{\mathcal{C}\beta} \sum_{m=2}^{2^{nR}} \left| \left( 1 \otimes \sqrt{\Lambda_m} \right) \xi_1 \left( 1 \otimes \sqrt{\Lambda_m} \right) \right|_1 \\
&= \mathbb{E}_{\mathcal{C}\beta} \left| \left( 1 \otimes \sqrt{\Lambda_1} \right) \xi_1 \left( 1 \otimes \sqrt{\Lambda_1} \right) - \xi_1 \right|_1 \\
&\quad + \mathbb{E}_{\mathcal{C}\beta} \sum_{m=2}^{2^{nR}} \text{Tr}(1 \otimes \Lambda_m) \xi_1 \\
&\leq 4\sqrt{\epsilon} + 2\epsilon \\
&\leq 5\sqrt{\epsilon},
\end{aligned} \tag{7.6}$$

provided that  $\epsilon \leq \frac{1}{2}$ . Since the the entanglement fidelity is linear in  $\mathcal{D}(\Theta)$ , which is itself linear in  $\Theta$ , we can also use the special triangle inequality to write

$$\begin{aligned}
F(|\Phi\rangle, \mathcal{D}(\mathbb{E}_{\mathcal{C}\beta} \Theta)) &= F(|\Phi\rangle, \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_{\mathcal{C}} \Theta)) \\
&\geq F(|\Phi\rangle, \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_{\mathcal{C}} \Theta')) - \left| \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_{\mathcal{C}} \Theta') - \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_{\mathcal{C}} \Theta) \right|_1.
\end{aligned}$$

Using our earlier observation from (7.5) and the definition of a  $(Q, n, \epsilon)$  entanglement transmission code, we can bound the first term as

$$\begin{aligned}
F(|\Phi\rangle, \mathcal{D}(\mathbb{E}_{\mathcal{C}} \Theta')) &= F(|\Phi\rangle, \mathcal{D} \circ \mathcal{N}'^{\otimes n} \circ \Upsilon) \\
&\geq 1 - \epsilon.
\end{aligned}$$

An estimate on the second term is obtained via

$$\begin{aligned}
\left| \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_{\mathcal{C}} \Theta) - \mathbb{E}_\beta \mathcal{D}(\mathbb{E}_{\mathcal{C}} \Theta') \right|_1 &\leq \mathbb{E}_\beta |\mathcal{D}(\mathbb{E}_{\mathcal{C}} \Theta) - \mathcal{D}(\mathbb{E}_{\mathcal{C}} \Theta')|_1 \\
&\leq \mathbb{E}_\beta |\mathbb{E}_{\mathcal{C}} \Theta - \mathbb{E}_{\mathcal{C}} \Theta'|_1 \\
&\leq \mathbb{E}_{\mathcal{C}\beta} |\Theta - \Theta'|_1 \\
&\leq 5\sqrt{\epsilon},
\end{aligned}$$

where first three lines are by convexity, monotonicity, and convexity once again of the trace norm. The last inequality follows from (7.7). Putting these together gives

$$\begin{aligned}\mathbb{E}_{\mathcal{C}\beta} F(|\Phi\rangle, \mathcal{D}(\Theta)) &\geq 1 - \epsilon - 5\sqrt{\epsilon} \\ &\geq 1 - 6\sqrt{\epsilon}.\end{aligned}\tag{7.8}$$

At last, observe that the final decoded state  $\Omega$  (which still depends on both sources of randomness  $\mathcal{C}$  and  $\beta$ ) is equal to

$$\Omega^{B\hat{B}\hat{M}} = \mathcal{D}(\Gamma^{BC^n} \hat{X}^n \hat{M}) \equiv \mathcal{D}(\xi_1^{BC^n}),$$

implicitly defining the desired decoding instrument  $\mathcal{D}: C^n \rightarrow \hat{B}\hat{M}$ . The expectation of (7.1) can now be bounded as

$$\begin{aligned}\mathbb{E}_{\mathcal{C}\beta} 2^{-nR} \sum_{m=1}^{2^{nR}} P_s^{\text{eg}}(m) &= \mathbb{E}_{\mathcal{C}\beta} P_s^{\text{eg}}(1) \\ &= F(|1\rangle|\Phi\rangle, \mathbb{E}_{\mathcal{C}\beta} \Omega) \\ &\geq 1 - |\text{Tr}_{B\hat{B}} \mathbb{E}_{\mathcal{C}\beta} \Gamma - |1\rangle\langle 1||_1 - 3(1 - F(|\Phi\rangle, \mathcal{D}(\Theta))) \\ &\geq 1 - 2\sqrt{2\epsilon} - 18\sqrt{\epsilon} \\ &\geq 1 - 21\sqrt{\epsilon}.\end{aligned}$$

The third line above is by Lemma 2. The first estimate in the fourth line follows from (7.4), while the second estimate is by (7.8), together with (6.3). We may now conclude that there are particular values of the randomness indices  $\beta$  and  $\mathcal{C}$  such that the same bound is satisfied for a deterministic code. We have thus proven that  $(\{\phi_m\}_{m \in 2^{nR}}, \mathcal{E}, \mathcal{D})$  comprises a  $(R, Q, n, 21\sqrt{\epsilon})$  entanglement generation code. This concludes the coding theorem.  $\blacksquare$

*Proof of Theorem 2 (coding theorem).* Begin by fixing bipartite pure states  $|\Psi_1\rangle^{A''A'}$  and  $|\Psi_2\rangle^{B''B'}$  which give rise to the state

$$\omega^{A''B''C} = (1^{A''B''} \otimes \mathcal{N}^{\otimes n})(\Psi_1 \otimes \Psi_2),$$

and defining  $\rho_1^{A'} = \text{Tr}_A \Psi_1$ ,  $\rho_2^{B'} = \text{Tr}_B \Psi_2$ . Letting  $\epsilon, \delta > 0$  be arbitrary, we will show that there exists a  $(Q_a, Q_b, n, \epsilon)$  qq entanglement transmission code where

$$Q_a = I_c(A'' \rangle C)_\omega - \delta \text{ and } Q_b = I_c(B'' \rangle A''C)_\omega - \delta$$

provided that  $Q_a, Q_b \geq 0$ . Note that the rates in Theorem 2 will be implied by taking the channel to be  $\mathcal{N}^{\otimes k}$ , with  $\omega^{ABC^k}$  defined similarly.

Let us begin by choosing an isometric extension  $\mathcal{U}_{\mathcal{N}} : A'B' \rightarrow CE$  of  $\mathcal{N}$ . Define the ideal channel  $\mathcal{N}_1 : A' \rightarrow C$  which would effectively be seen by Alice were Bob's average code density operator exactly equal to  $\rho_2^{\otimes n}$  as

$$\mathcal{N}_1 : \tau \mapsto \mathcal{N}(\tau \otimes \rho_2).$$

We now use  $\mathcal{U}_{\mathcal{N}}$  to define a particular isometric extension  $\mathcal{U}_{\mathcal{N}_1} : A' \rightarrow CE'$  of  $\mathcal{N}_1$ , where  $E' = B''E$ , as

$$\mathcal{U}_{\mathcal{N}_1} : \tau \mapsto \mathcal{U}_{\mathcal{N}}(\tau \otimes \Psi_2).$$

Observe that Bob's fake input  $B''$  is treated as part of the environment of Alice's ideal induced channel. We then further define the channel  $\mathcal{N}_2 : B' \rightarrow A''C$  by

$$\mathcal{N}_2 : \tau \mapsto \mathcal{N}(\Psi_1 \otimes \tau).$$

In contrast to the interpretation of  $\mathcal{N}_1$ , this may be viewed as the channel which would be seen by Bob if Alice were to input the  $A'$  part of the purification  $|\Psi_2\rangle^{A''A'}$  of  $\rho_2^{A'}$  to her input of the channel and then send the  $A''$  system to Charlie via a noiseless quantum channel. As in the proof of Theorem 1, Charlie will first decode Alice's information, after which he will attempt to simulate the channel  $\mathcal{N}_2$ , allowing a higher transmission rate for Bob than if Alice's information was treated as noise.

Since quantum information cannot be copied, showing that this is indeed possible will require different techniques than were utilized in the previous coding theorem. Although ensembles of random codes will be used in this proof, we introduce the technique of *coherent coding*, in which we pretend that the common randomness is purified. The main advantage of this approach will be that working with states in the enlarged Hilbert space allows monotonicity to be easily exploited in order to provide the estimates we require. Additionally, before we derandomize at the end of the proof, it will ultimately be only Bob who is using a random code. Alice will be able to use any deterministic code from her random ensemble, as Charlie will implement a decoding procedure which produces a global state which is close to that which would have been created had Alice coded with the coherent randomness. To show this, we will first analyze the state which would result if both senders used their full ensembles of codes. Then we show that if Alice uses any code from her ensemble, Charlie can create the proper global state himself, allowing him to effectively simulate  $\mathcal{N}_2$  and ultimately decode both states at the desired rates.

By Proposition 2, for large enough  $n$ , there exists a  $(Q_a, n, \epsilon)$  random entanglement generation code  $(p_\ell, |\Upsilon_1^\ell\rangle^{AA^n}, \mathcal{D}_1^\ell)$  for the channel  $\mathcal{N}_1$ , where  $Q_a = I_c(\rho_1, \mathcal{N}_1) - \delta = I_c(A'' \rangle C) - \delta$ . There similarly exists a  $(Q_b, n, \epsilon)$  random entanglement generation code  $(q_m, |\Upsilon_2^m\rangle^{BB^n}, \mathcal{D}_2^m)$  for  $\mathcal{N}_2$ , with  $Q_b = I_c(\rho_2, \mathcal{N}_2) - \delta = I_c(B'' \rangle A''C) - \delta$ . Proposition 2 further guarantees that these codes can be chosen so that their respective average code density operators

$$\varrho_1^{A^n} = \sum_{\ell} p_{\ell} \text{Tr}_A \Upsilon_1^{\ell} \quad \text{and} \quad \varrho_2^{B^n} = \sum_m q_m \text{Tr}_B \Upsilon_2^m$$

satisfy

$$|\varrho_i - \rho_i^{\otimes n}|_1 \leq \epsilon. \tag{7.9}$$

Recall that by Proposition 2 we may choose isometric extensions  $\mathcal{U}_{\mathcal{D}_1}^{\ell} : C^n \rightarrow \widehat{A}F$

implementing the  $\mathcal{D}_1^\ell$  from Alice's random code which satisfy

$$F\left(|\Phi_1\rangle^{A\hat{A}}|\lambda\rangle^{FE'n}, \mathcal{U}_{\mathcal{D}_1}^\ell \circ \mathcal{U}_{\mathcal{M}_1}^{\otimes n}(\Upsilon_1^\ell)\right) \geq 1 - \epsilon \quad (7.10)$$

for every random code index  $\ell$  and the same fixed state  $|\lambda\rangle^{FE'n}$ .

Let the code common randomness between Alice and Charlie be held between the systems  $L_A$  and  $L_C$ , represented by the state

$$\gamma_1^{L_AL_C} = \sum_{\ell} p_{\ell} |\ell\rangle\langle\ell|^{L_A} \otimes |\ell\rangle\langle\ell|^{L_C},$$

defining a similar state  $\gamma_2^{M_B M_C}$  for the Bob-Charlie common randomness. For convenience, let us further pretend that  $\gamma_1$  is part of a pure state

$$|\Gamma_1\rangle^{L_E L_A L_B} = \sum_{\ell} \sqrt{p_{\ell}} |\ell\rangle^{L_E} |\ell\rangle^{L_A} |\ell\rangle^{L_C}.$$

Similarly, let  $\gamma_2$  be purified by  $|\Gamma_2\rangle^{M_E M_B M_C}$ . Write controlled encoding isometries  $\mathcal{E}_1: L_A \rightarrow L_A A'^n$  and  $\mathcal{E}_2: M_B \rightarrow M_B B'^n$  as

$$\mathcal{E}_1 = \sum_{\ell} |\ell\rangle |\Upsilon_1^\ell\rangle\langle\ell| \quad \text{and} \quad \mathcal{E}_2 = \sum_m |m\rangle |\Upsilon_2^m\rangle\langle m|.$$

The states which would arise if Alice and Bob each encoded *coherently* are

$$\begin{aligned} |\Upsilon_1\rangle^{L_A A'^n} &\equiv \mathcal{E}_1 |\Gamma_1\rangle = \sum_{\ell} \sqrt{p_{\ell}} |\ell\rangle^L |\Upsilon_1^\ell\rangle \\ |\Upsilon_2\rangle^{M_B B'^n} &\equiv \mathcal{E}_2 |\Gamma_2\rangle = \sum_m \sqrt{q_m} |m\rangle^M |\Upsilon_2^m\rangle. \end{aligned}$$

Note that we have abbreviated  $L = L_E L_A L_C$  and  $M = M_E M_B M_C$ . As each  $|\Upsilon_i\rangle$  is a purification of  $\varrho_i$ , together with (7.9), Uhlmann's theorem tells us that there exist unitaries  $V_1: LA \rightarrow A'^n$  and  $V_2: MB \rightarrow B'^n$  such that

$$F(V_i |\Upsilon_i\rangle, |\Psi_i\rangle^{\otimes n}) \geq 1 - \epsilon. \quad (7.11)$$



Further define a corresponding controlled isometric decoder  $\mathcal{U}_{\mathcal{D}_1}: L_C C^n \rightarrow L_C \widehat{A}F$  for Alice's code as

$$\mathcal{U}_{\mathcal{D}_1} = \sum_{\ell} |\ell\rangle\langle\ell|^{L_C} \otimes \mathcal{U}_{\mathcal{D}_1}^{\ell}.$$

Let us now imagine that each of Alice and Bob encodes using the coherent common randomness, resulting in a joint pure state  $\mathcal{U}_{\mathcal{N}}^{\otimes n} |\Upsilon_1\rangle|\Upsilon_2\rangle$  on  $LAMBCE^n$ . If Charlie then applies the full controlled decoder from Alice's code, the resulting global pure state would be

$$|\Theta\rangle^{LA\widehat{A}MBFE^n} = \mathcal{U}_{\mathcal{D}_1} \circ \mathcal{U}_{\mathcal{N}}^{\otimes n} |\Upsilon_1\rangle|\Upsilon_2\rangle.$$

For each  $\ell$ , let us define an isometry  $\mathcal{O}^{\ell}: B^n \rightarrow A\widehat{A}FE^n$  as

$$\mathcal{O}^{\ell} = \mathcal{U}_{\mathcal{D}_1}^{\ell} \circ \mathcal{U}_{\mathcal{N}}^{\otimes n} (\Upsilon_1 \otimes \cdot)$$

which we use to define the pure states

$$|\theta_{\ell}\rangle^{A\widehat{A}MFBFE^n} = \mathcal{O}^{\ell} |\Upsilon_2\rangle.$$

These definitions allow us to express

$$|\Theta\rangle = \sum_{\ell} \sqrt{p_{\ell}} |\ell\rangle^L |\theta_{\ell}\rangle.$$

Further writing  $|\lambda'\rangle^{FMBE^n} \equiv V_2^{-1} |\lambda\rangle^{FB''E^n}$ , the following bound applies

$$\begin{aligned} F\left(|\Phi_1\rangle^{A\widehat{A}} |\lambda'\rangle^{FMBE^n}, |\theta_{\ell}\rangle\right) &= F\left(|\Phi_1\rangle |\lambda'\rangle^{FMBE^n}, \mathcal{O}^{\ell} |\Upsilon_2\rangle\right) \\ &= F\left(|\Phi_1\rangle |\lambda\rangle^{FB''E^n}, \mathcal{O}^{\ell} \circ V_2 |\Upsilon_2\rangle\right) \\ &\geq 1 - 2\sqrt{1 - F\left(|\Phi_1\rangle |\lambda\rangle^{FB''E^n}, \mathcal{O}^{\ell} |\Psi_2\rangle^{\otimes n}\right)} \\ &\quad - 2\sqrt{1 - F\left(V_2 |\Upsilon_2\rangle, |\Psi_2\rangle^{\otimes n}\right)} \\ &\geq 1 - 2\sqrt{1 - F\left(|\Phi_1\rangle |\lambda\rangle^{FE^n}, \mathcal{U}_{\mathcal{D}_1}^{\ell} \circ \mathcal{U}_{\mathcal{N}_1}^{\otimes n} \circ |\Upsilon_1^{\ell}\rangle\right)} - 2\sqrt{\epsilon} \\ &\geq 1 - 4\sqrt{\epsilon}. \end{aligned}$$

Above, the second equality is because the actions of  $\mathcal{O}^{\ell}$  and  $V_2$  commute, the first

inequality is by the triangle inequality and monotonicity with respect to  $\mathcal{O}^\ell$ , while for the second inequality, we have just rewritten the first term and used (7.11) for the second. The last bound is from (7.10). Observe that we are still free to specify the global phases of the outputs of the  $\mathcal{U}_{\mathcal{D}_1}^\ell$  so that the above bound further implies  $\langle \theta_\ell | \Phi_1 \rangle | \lambda' \rangle \geq (1 - 4\sqrt{\epsilon})^{1/2}$  for each  $\ell$ . Consequently,

$$\begin{aligned} F(|\Theta\rangle, |\Gamma_1\rangle|\Phi_1\rangle|\lambda'\rangle) &= \left| \sum_{\ell\ell'} \sqrt{p_\ell p_{\ell'}} \langle \ell | \ell' \rangle \langle \theta_\ell | \Phi_1 \rangle | \lambda' \rangle \right|^2 \\ &= \left| \sum_{\ell} p_\ell \langle \theta_\ell | \Phi_1 \rangle | \lambda' \rangle \right|^2 \\ &\geq 1 - 4\sqrt{\epsilon}. \end{aligned}$$

Essentially, the subsystems  $L$ ,  $A\hat{A}$  and  $MBFE^n$  of  $|\Theta\rangle$  are mutually decoupled.

As mentioned earlier, it will be sufficient for Alice to use *any* deterministic code from the random ensemble to encode. Without loss of generality, we assume that Alice chooses to use the first code ( $\ell = 1$ ) in her ensemble. Bob, on the other hand, will need to use randomness to ensure that Alice's effective channel is close to a product channel. The state on  $AMBC^n E^n$  which results from these encodings is  $\mathcal{U}_{\mathcal{N}}^{\otimes n} |\Upsilon_1^1\rangle | \Upsilon_2 \rangle$ .

We will now describe a procedure by which Charlie first decodes Alice's information, then produces a global state which is close to  $|\Theta\rangle$ , making it look like Alice had in fact utilized the coherent coding procedure. This will allow Charlie to apply local unitaries to effectively simulate the channel  $\mathcal{N}_2$  for which Bob's random code was designed, enabling him to decode Bob's information as well. These steps will constitute Charlie's decoding  $\mathcal{D} : M_C C^n \rightarrow M_C \hat{A} \hat{B}$ , which depends on the Bob-Charlie common randomness. The existence of a deterministic decoder will then be inferred.

Charlie first applies the isometric decoder  $\mathcal{U}_{\mathcal{D}_1}^1$ , placing all systems into the state  $|\theta_1\rangle$ . He then removes his local system  $\hat{A}$  (it is important that he keep  $\hat{A}$  in a safe place, as it represents the decoder output for Alice's quantum information) and replaces it with the corresponding parts of the locally prepared pure state  $|\Phi_1\rangle^{A^\circ \hat{A}^\circ}$ . Charlie also

locally prepares the state  $|\Gamma_1\rangle^L$ . The resulting state

$$\Theta' = \Gamma_1^L \otimes \Phi_1^{A^\circ \hat{A}^\circ} \otimes \text{Tr}_{A\hat{A}} \theta_1,$$

satisfies

$$\begin{aligned} F(\Theta', \Theta) &\geq 1 - |\text{Tr}_{A\hat{A}} \theta_1 - \lambda'|_1 - |\lambda' - \text{Tr}_{LA\hat{A}} \Theta|_1 \\ &\quad - 3(1 - F(|\Gamma\rangle|\Phi_1\rangle, \text{Tr}_{MBFE^n} \Theta)) \\ &\geq 1 - 2\sqrt{4\sqrt{\epsilon}} - 2\sqrt{4\sqrt{\epsilon}} - 12\sqrt{\epsilon} \\ &\geq 1 - 9\epsilon^{1/4} \end{aligned} \tag{7.12}$$

whenever  $\epsilon \leq 12^{-4}$ . The first line combines Lemma 2 and the triangle inequality. The first two estimates in the second line are from applying (6.3) and monotonicity with respect to  $\text{Tr}_{A\hat{A}}$  and  $\text{Tr}_{LA\hat{A}}$  to the previous two estimates. The last estimate in that line is from monotonicity with respect to the map  $\text{Tr}_{MBFE^n}$  applied to the previous estimate. Next, Charlie will apply  $V_1 \circ \mathcal{U}_{\mathcal{D}_1}^{-1}$  to  $\Theta'^1$  in order to simulate the channel  $\mathcal{N}_2$ . To see that this will work, define  $\mathcal{M}: LA\hat{A}FE^n \rightarrow A''^n C^n$  as  $\mathcal{M} \equiv \text{Tr}_{E^n} V_1 \circ \mathcal{U}_{\mathcal{D}_1}^{-1}$  and observe that by monotonicity with respect to  $\mathcal{N}^{\otimes n}(\cdot \otimes \Upsilon_2)$  and (7.11), the states on  $MBA''^n C^n$  satisfy

$$\begin{aligned} F(\mathcal{M}(\Theta), \mathcal{N}_2^{\otimes n}(\Upsilon_2)) &= F(V_1 \circ \mathcal{N}^{\otimes n}(\Upsilon_1 \otimes \Upsilon_2), \mathcal{N}^{\otimes n}(\Psi_1^{\otimes n} \otimes \Upsilon_2)) \\ &\geq F(V_1|\Upsilon_1\rangle, |\Psi_1\rangle^{\otimes n}) \\ &\geq 1 - \epsilon. \end{aligned}$$

We may now use the triangle inequality and monotonicity with respect to  $\mathcal{M}$  to

---

<sup>1</sup>This operation only acts on Charlie's local systems, i.e.  $V_1 \circ \mathcal{U}_{\mathcal{D}_1}^{-1}: LA^\circ \hat{A}^\circ F \rightarrow A''^n C^n$ .

combine our last two estimates, yielding

$$\begin{aligned}
F(\mathcal{M}(\Theta'), \mathcal{N}_2^{\otimes n}(\Upsilon_2)) &\geq 1 - 2\sqrt{1 - F(\mathcal{M}(\Theta'), \mathcal{M}(\Theta))} \\
&\quad - 2\sqrt{1 - F(\mathcal{M}(\Theta), \mathcal{N}_2^{\otimes n}(\Upsilon_2))} \\
&\geq 1 - 2\sqrt{9\epsilon^{1/4}} - 2\sqrt{\epsilon} \\
&\geq 1 - 7\epsilon^{1/8}
\end{aligned} \tag{7.13}$$

whenever  $\epsilon \leq 2^{-8/3}$ . We have thus far shown that Charlie's decoding procedure succeeds in simulating the channel  $\mathcal{N}_2^{\otimes n}$ , while simultaneously recovering Alice's quantum information. Charlie now uses the controlled decoder  $\mathcal{D}_2: M_C A'^n C^n \rightarrow M_C \widehat{B}$  defined as

$$\mathcal{D}_2 = \sum_m |m\rangle\langle m|^{M_C} \otimes \mathcal{D}_2^m$$

to decode Bob's quantum information. This entire procedure has defined our decoder  $\mathcal{D}: M_C C^n \rightarrow M_C \widehat{A} \widehat{B}$  which gives rise to a global state  $\Omega^{A \widehat{A} B \widehat{B}}$  representing the final output state of the protocol, averaged over Bob's common randomness. This state satisfies

$$\begin{aligned}
F(|\Phi_1\rangle, \text{Tr}_{B \widehat{B}} \Omega) &\geq F(\Theta, \Theta') \\
&\geq 1 - 9\epsilon^{1/4},
\end{aligned}$$

because of monotonicity with respect to  $\text{Tr}_{LMBFE^n}$  applied to the bound (7.12). By using the triangle inequality, the fact that Bob's codes are  $\epsilon$ -good for each  $m$ , and monotonicity of the estimate (7.13) with respect to  $\text{Tr}_M \mathcal{D}_2$ , the global state can

further be seen to obey

$$\begin{aligned}
F(|\Phi_2\rangle, \text{Tr}_{A\hat{A}}\Omega) &= F(|\Phi_2\rangle, \text{Tr}_M \mathcal{D}_2 \circ \mathcal{M}(\Theta')) \\
&\geq 1 - 2\sqrt{1 - F(|\Phi_2\rangle, \text{Tr}_M \mathcal{D}_2 \circ \mathcal{N}_2^{\otimes n}(\Upsilon_2))} \\
&\quad - 2\sqrt{1 - F(\text{Tr}_M \mathcal{D}_2 \circ \mathcal{N}_2^{\otimes n}(\Upsilon_2), \text{Tr}_M \mathcal{D}_2 \circ \mathcal{M}(\Theta'))} \\
&\geq 1 - 2\sqrt{\epsilon} - 2\sqrt{7\epsilon^{1/8}} \\
&\geq 1 - 7\epsilon^{1/16}
\end{aligned}$$

as long as  $\epsilon \leq 2^{-16/7}$ . Along with (6.3), a final application of Lemma 2 combines the above two bounds to give

$$\begin{aligned}
F(|\Phi_1\rangle|\Phi_2\rangle, \Omega) &\geq 1 - |\Phi_1 - \text{Tr}_{B\hat{B}}\Omega|_1 - 3\left(1 - F(|\Phi_2\rangle, \text{Tr}_{A\hat{A}}\Omega)\right) \\
&\geq 1 - 2\sqrt{9\epsilon^{1/4}} - 21\epsilon^{1/16} \\
&\geq 1 - 22\epsilon^{1/16},
\end{aligned}$$

provided that  $\epsilon \leq 6^{-16}$ . Since this estimate represents an average over Bob's common randomness, there must exist a particular value  $m^*$  of the common randomness so that the corresponding deterministic code is at least as good as the random one, thus concluding the coding theorem.  $\blacksquare$

## 7.2 The converse theorems

We will now demonstrate that

$$\mathcal{C}\mathcal{Q}_{\text{eg}}(\mathcal{N}) \subseteq \bigcup_{k=1}^{\infty} \frac{1}{k} \mathcal{C}\mathcal{Q}^{(1)}(\mathcal{N}^{\otimes k}) \quad \text{and} \quad \mathcal{Q}_{\text{eg}}(\mathcal{N}) \subseteq \bigcup_{k=1}^{\infty} \frac{1}{k} \mathcal{Q}^{(1)}(\mathcal{N}^{\otimes k}),$$

where the single-letter regions  $\mathcal{C}\mathcal{Q}^{(1)}(\mathcal{N})$  and  $\mathcal{Q}^{(1)}(\mathcal{N})$  are those defined at the beginning of the last section.

*Proof of Theorem 1 (converse).* Suppose there exists a sequence of  $(R, Q, n, \epsilon_n)$  entanglement generation codes with  $\epsilon_n \rightarrow 0$ . Fixing a blocklength  $n$ , let  $\{\phi_m\}, \Upsilon^{BB^n}, \mathcal{D}$

comprise the corresponding cq entanglement generation code. The state induced by the encoding is

$$\omega^{MBC^n} = 2^{-nR} \sum_{m=1}^{2^{nR}} |m\rangle\langle m|^M \otimes (1^B \otimes \mathcal{N}^{\otimes n})(\phi_m \otimes \Upsilon).$$

After application of the decoding instrument  $\mathcal{D}: C^n \rightarrow \widehat{B}\widehat{M}$ , this state becomes

$$\Omega^{M\widehat{M}B\widehat{B}} = (1^{MB} \otimes \mathcal{D})(\omega).$$

An upper bound on the classical rate of the code can be obtained as follows:

$$\begin{aligned} nR &= H(M)_\Omega \\ &= I(M; \widehat{M})_\Omega + H(M|\widehat{M})_\Omega \\ &\leq I(M; \widehat{M})_\Omega + H(\epsilon_n) + nR\epsilon_n \\ &\leq I(M; C^m)_\omega + n\epsilon'_n. \end{aligned}$$

The first inequality follows from Fano's inequality (Lemma 5) while in the second we use the Holevo bound (Lemma 7) and define  $\epsilon'_n = \frac{1}{n} + R\epsilon_n$ . The quantum rate of the code is upper bounded as

$$\begin{aligned} I_c(B \rangle C^n M)_\omega &\geq I_c(B \rangle \widehat{B}\widehat{M})_\Omega \\ &\geq I_c(B \rangle \widehat{B})_\Omega \\ &\geq I_c(B \rangle \widehat{B})_\Phi - 2H(\epsilon_n) - 8nQ\sqrt{\epsilon_n} \\ &= nQ - n\epsilon''_n. \end{aligned}$$

Above, the first two inequalities are consequences of the data processing inequality (Lemma 6), while the last inequality applies a combination of Lemma 3 and (6.3), along with the definition  $\epsilon''_n = \frac{2}{n} + nQ\sqrt{\epsilon_n}$ . Setting  $X = M$ , we have thus proven that

$$R \leq \frac{1}{n}I(X; C^m) + \epsilon'_n, \quad Q \leq \frac{1}{n}I_c(B \rangle C^m X) + \epsilon''_n$$

whenever  $(R, Q)$  is an achievable cq rate pair for entanglement generation, where  $\epsilon'_n, \epsilon''_n \rightarrow 0$ . It follows that for any achievable rate pair  $(R, Q)$  and any  $\delta > 0$ , we have

$$(R - \delta, Q - \delta) \in \frac{1}{n} \mathcal{CQ}^{(1)}(\mathcal{N}^{\otimes n}) \subseteq \mathcal{CQ}(\mathcal{N}).$$

Since  $\mathcal{CQ}(\mathcal{N})$  is closed by definition, this completes the proof.  $\blacksquare$

*Proof of Theorem 2 (converse).* Suppose that  $(Q_a, Q_b)$  is an achievable qq rate pair for entanglement generation. By definition, this means that there must exist a sequence of  $(Q_a, Q_b, n, \epsilon_n)$  entanglement generation codes with  $\epsilon_n \rightarrow 0$ . Fixing a block-length  $n$ , let  $|\Upsilon_1\rangle^{AA^n}, |\Upsilon_2\rangle^{BB^n}$  and  $\mathcal{D}: C^n \rightarrow \widehat{A}\widehat{B}$  comprise the corresponding encodings and decodings. Define

$$\omega^{ABC^n} = (1^{AB} \otimes \mathcal{N}^{\otimes n})(\Upsilon_1 \otimes \Upsilon_2)$$

to be the result of sending the respective  $A^n$  and  $B^n$  parts of  $\Upsilon_1$  and  $\Upsilon_2$  through the channel  $\mathcal{N}^{\otimes n}$ . Further defining

$$\Omega^{AB\widehat{A}\widehat{B}} = (1^{AB} \otimes \mathcal{D})(\omega)$$

as the corresponding state after decoding, the entanglement fidelity of the code is given by

$$F_{AB} = F(|\Phi_1\rangle \otimes |\Phi_2\rangle, \Omega) \geq 1 - \epsilon_n. \quad (7.14)$$

where  $|\Phi_1\rangle^{A\widehat{A}}$  and  $|\Phi_2\rangle^{B\widehat{B}}$  are the maximally entangled target states. The sum rate can be bounded as

$$\begin{aligned} I_c(AB)_{C^n} \omega &\geq I_c(AB)_{\widehat{A}\widehat{B}} \Omega \\ &\geq I_c(AB)_{\widehat{A}\widehat{B}}_{\Phi_1 \otimes \Phi_2} - 2H(\epsilon_n) - 8n(Q_a + Q_b)\sqrt{\epsilon_n} \\ &\geq n(Q_a + Q_b) - n\epsilon'_n. \end{aligned}$$

The first step is by the data processing inequality (Lemma 6). The second step uses

Lemma 3 and (6.3), along with monotonicity applied to (7.14). The last step has defined  $\epsilon'_n = \frac{2}{n} - 8(Q_a + Q_b)\sqrt{\epsilon_n}$  and holds because the binary entropy  $H(\cdot)$  is upper bounded by 1. We can bound Alice's rate  $Q_a$  by writing

$$\begin{aligned}
I_c(A \rangle BC^n)_\omega &\geq I_c(A \rangle C^n)_\omega \\
&\geq I_c(A \rangle \widehat{A}\widehat{B})_\Omega \\
&\geq I_c(A \rangle \widehat{A})_\Omega \\
&\geq I_c(A \rangle \widehat{A})_{\Phi_1} - 2H(\epsilon_n) - 8nQ_a\sqrt{\epsilon_n} \\
&\geq nQ_a - n\epsilon'_n.
\end{aligned}$$

The first three steps above are by data processing (Lemma 6). The remaining steps hold for the same reasons as in the previous chain of inequalities. Similarly, Bob's rate also must satisfy

$$nQ_b \leq I_c(B \rangle AC^n)_\omega + n\epsilon'_n.$$

Since  $\epsilon_n \rightarrow 0$  implies  $\epsilon'_n \rightarrow 0$ , this means that for every  $\delta > 0$ , any achievable qq rate pair  $(Q_a, Q_b)$  must satisfy

$$(Q_a - \delta, Q_b - \delta) \in \frac{1}{n}\mathcal{Q}^{(1)}(\mathcal{N}^{\otimes n}) \subseteq \mathcal{Q}(\mathcal{N}).$$

Since  $\mathcal{Q}(\mathcal{N})$  is closed by definition, this completes the proof. ■



# Chapter 8

## Transmission of quantum information

In the previous chapter, we have proven the main theorems for the restricted case in which all quantum communication has been in the sense of *generating* quantum correlations between senders and receiver. The results of this chapter will complete the proofs of the main theorems, by extending the weaker error criteria of entanglement generation (which incidentally, are analogous to a classical requirement on the *average* probability of error) to the stronger requirements of strong subspace transmission in the main theorem statements. As a first step, we demonstrate how the results of the last chapter immediately imply the ability to perform an intermediate task, *entanglement transmission*, where the senders are required to transmit preexisting maximal entanglement, while still adhering to an average error criterion on the classical error. We then show how to use a given entanglement transmission code to construct a strong subspace transmission codes achieving any rates less than those of the original code, while paying a negligible price in fidelity.

### 8.1 Entanglement transmission

**Classical-quantum scenario** In this scenario, rather than generating entanglement with Charlie, Bob will act to transmit *preexisting* entanglement to him. We

assume that Bob is presented with the  $\tilde{B}$  part of the maximally entangled state  $|\Phi\rangle^{B\tilde{B}}$ . It is assumed that he has complete control over  $\tilde{B}$ , while he has no access to  $B$ . He will perform a physical operation in order to transfer the quantum information embodied in his system  $\tilde{B}$  to the inputs  $B^n$  of the channel, modeled by an encoding operation  $\mathcal{E}: \tilde{B} \rightarrow B^n$ . The goal of this encoding will be to make it possible for Charlie, via post-processing of the information embodied in the system  $C^n$ , to hold the  $\hat{B}$  part of a state which is close to that which would have resulted if Bob had sent his system through a perfect quantum channel  $\text{id}: \tilde{B} \rightarrow \hat{B}$ . Here, we imagine that  $\tilde{B}$  and  $\hat{B}$  denote two distinct physical systems with the same number of quantum degrees of freedom. The role of the identity channel is to set up a unitary correspondence, or isomorphism, between the degrees of freedom of  $\tilde{B}$  in Bob's laboratory and those of  $\hat{B}$  in Charlie's. We will often tacitly assume that such an identity map has been specified ahead of time in order to judge how successful an imperfect quantum transmission has been. This convention will be taken for granted many times throughout the paper, wherein specification of an arbitrary state  $|\Psi\rangle^{B\tilde{B}}$  will immediately imply specification of the state  $|\Psi\rangle^{B\hat{B}} = (1^B \otimes \text{id})|\Psi\rangle^{B\tilde{B}}$ . Decoding is the same as it is for entanglement generation.

$(\{\phi_m\}_{m \in 2^{nR}}, \mathcal{E}, \mathcal{D})$  will be called an  $(R, Q, n, \epsilon)$  *cq entanglement transmission code* for the channel  $\mathcal{N}$  if

$$2^{-nR} \sum_{m=1}^{2^{nR}} P_s^{\text{et}}(m) \geq 1 - \epsilon, \quad (8.1)$$

where

$$P_s^{\text{et}}(m) = F\left(|m\rangle|\Phi\rangle^{B\hat{B}}, \mathcal{D} \circ \mathcal{N}^{\otimes n}(\phi_m^{A^n} \otimes \mathcal{E}(\Phi^{B\tilde{B}}))\right). \quad (8.2)$$

Achievable rate pairs and the capacity region  $\mathcal{C}_{\mathcal{Q}_{\text{et}}}(\mathcal{N})$  are defined analogous to those for entanglement generation.

**Quantum-quantum scenario** Alice and Bob each respectively have control over the  $\tilde{A}$  and  $\tilde{B}$  parts of the separate maximally entangled states  $|\Phi_1\rangle^{A\tilde{A}}, |\Phi_2\rangle^{B\tilde{B}}$ , while

neither has access to  $A$  or  $B$ . Alice transfers the correlations in her system to the  $A^n$  parts of the inputs of  $\mathcal{N}^{\otimes n}$  with an encoding operation  $\mathcal{E}_1: \tilde{A} \rightarrow A^n$ . Bob acts similarly with  $\mathcal{E}_2: \tilde{B} \rightarrow B^n$ . Their goal is to preserve the respective correlations, so that Charlie can apply a decoding operation  $\mathcal{D}: C^n \rightarrow \hat{A}\hat{B}$ , in order to end up holding the  $\hat{A}\hat{B}$  part of a state which is close to  $|\Phi_1\rangle^{A\hat{A}}|\Phi_2\rangle^{B\hat{B}}$ . Formally,  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D})$  is a  $(Q_a, Q_b, n, \epsilon)$  *qq entanglement transmission code* for the channel  $\mathcal{N}$  if

$$F(|\Phi_1\rangle|\Phi_2\rangle, \mathcal{D} \circ \mathcal{N}^{\otimes n} \circ (\mathcal{E}_1 \otimes \mathcal{E}_2)(\Phi_1 \otimes \Phi_2)) \geq 1 - \epsilon. \quad (8.3)$$

Achievable qq rate pairs for entanglement generation and the capacity region  $\mathcal{Q}_{\text{et}}(\mathcal{N})$  are defined as in the previous scenario.

## 8.2 Equivalence of entanglement transmission and entanglement generation

### 8.2.1 $\mathcal{CQ}_{\text{eg}} \subseteq \mathcal{CQ}_{\text{et}}$ and $\mathcal{Q}_{\text{eg}} \subseteq \mathcal{Q}_{\text{et}}$

*Proof.* This essentially follows as an artifact of the entanglement generation coding theorem from [13]. There, the input preparation  $|\Upsilon\rangle^{AA^n}$  for a  $(Q, n)$  entanglement generation code is constructed with the particular form

$$|\Upsilon\rangle^{AA^n} = \frac{1}{\sqrt{2^{nQ}}} \sum_{a \in 2^{nQ}} |a\rangle^A |\phi_a\rangle^{A^n},$$

where the  $\{|\phi_a\rangle\}$  are orthogonal. Observe that the if the encoder acts on the  $\tilde{A}$  part of the maximally entangled state

$$|\Phi\rangle^{A\hat{A}} = \frac{1}{\sqrt{2^{nQ}}} \sum_{a \in 2^{nQ}} |a\rangle^A |a\rangle^{\hat{A}}$$

with an encoding isometry  $\mathcal{E}: \tilde{A} \rightarrow A^n$  defined via

$$\mathcal{E} = \sum_{a \in 2^{nQ}} |\phi_a\rangle^{A^n} \langle a|^{\tilde{A}},$$

the identity  $\mathcal{E}|\Phi\rangle^{A\tilde{A}} = |\Upsilon\rangle^{AA^n}$  holds trivially. It is thus a simple task to modify the proofs of Chapter 7 to instead prove the existence of the entanglement transmission codes described in the previous section. Indeed, if  $(|\Upsilon\rangle, \{\phi_m\}, \mathcal{D})$  is a  $(R, Q, n, \epsilon)$  cq entanglement generation code, there then exists an encoder  $\mathcal{E}$  so that  $(\mathcal{E}, \{\phi_m\}, \mathcal{D})$  is a  $(R, Q, n, \epsilon)$  cq entanglement transmission code. Identical reasoning shows that to every qq entanglement generation code, there a qq entanglement transmission code with the same parameters.  $\blacksquare$

### 8.2.2 $\mathcal{CQ}_{\text{et}} \subseteq \mathcal{CQ}_{\text{eg}}$

*Proof.* Suppose there exists an  $(R, Q, n, \epsilon)$  cq entanglement transmission code, consisting of classical message states  $\{|\phi_m\rangle^{A^n}\}_{m \in 2^{nR}}$ , a quantum encoding map  $\mathcal{E}: \tilde{B} \rightarrow \hat{B}$ , and a decoding instrument  $\mathcal{D}: C^n \rightarrow \widehat{M}\hat{B}$ . Write any pure state decomposition of the encoded state

$$(1^B \otimes \mathcal{E})(\Phi) = \sum_i p_i |\Upsilon_i\rangle \langle \Upsilon_i|.$$

Then, the success condition (8.1) for a cq entanglement transmission code can be rewritten as

$$1 - \epsilon \leq 2^{-nR} \sum_{m=1}^{2^{nR}} P_s^{\text{et}}(m) \tag{8.4}$$

$$= 2^{-nR} \sum_{m=1}^{2^{nR}} F\left(|\Phi\rangle^{B\hat{B}}, \mathcal{D}_m \circ \mathcal{N}^{\otimes n}(\phi_m^{A^n} \otimes \left(\sum_i p_i \Upsilon_i\right))\right) \tag{8.5}$$

$$= \sum_i p_i \left( 2^{-nR} \sum_{m=1}^{2^{nR}} F\left(|\Phi\rangle^{B\hat{B}}, \mathcal{D}_m \circ \mathcal{N}^{\otimes n}(\phi_m^{A^n} \otimes \Upsilon_i)\right) \right) \tag{8.6}$$

$$= \sum_i p_i \left( 2^{-nR} \sum_{m=1}^{2^{nR}} P_s^{\text{eg}}(m, \Upsilon_i) \right), \tag{8.7}$$

so that there is a particular value  $i^*$  of  $i$  for which

$$2^{-nR} \sum_{m=1}^{2^{nR}} P_s^{\text{eg}}(m, \Upsilon_{i^*}) \geq 1 - \epsilon.$$

Hence,  $(\{|\phi_m\rangle\}_{m \in 2^{nR}}, |\Upsilon_{i^*}\rangle, \mathcal{D})$  comprises an  $(R, Q, n, \epsilon)$  cq entanglement generation code. ■

### 8.2.3 $\mathcal{Q}_{\text{et}} \subseteq \mathcal{Q}_{\text{eg}}$

*Proof.* Suppose there exists a  $(Q_a, Q_b, n, \epsilon)$  entanglement transmission code  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D})$  which transmits the maximally entangled states  $|\Phi_1\rangle, |\Phi_2\rangle$ . As in the cq case, the encoded states can be decomposed as

$$(1^A \otimes \mathcal{E}_1)(\Phi_1) = \sum_i p_i \Upsilon_{1i}$$

and

$$(1^B \otimes \mathcal{E}_2)(\Phi_2) = \sum_j q_j \Upsilon_{2j}.$$

The reliability condition (8.3) can then be rewritten as

$$\sum_{ij} p_i q_j F(|\Phi_1\rangle|\Phi_2\rangle, \mathcal{D} \otimes \mathcal{N}^{\otimes n}(\Upsilon_{1i} \otimes \Upsilon_{2j})) \geq 1 - \epsilon,$$

which implies the existence of a particular pair  $(i^*, j^*)$  of values of  $(i, j)$  such that

$$F(|\Phi_1\rangle|\Phi_2\rangle, \mathcal{D} \otimes \mathcal{N}^{\otimes n}(\Upsilon_{1i^*} \otimes \Upsilon_{2j^*})) \geq 1 - \epsilon.$$

Hence,  $(|\Upsilon_{1i^*}\rangle, |\Upsilon_{2j^*}\rangle, \mathcal{D})$  comprises a  $(Q_a, Q_b, n, \epsilon)$  qq entanglement generation code. ■

### 8.3 Strong subspace transmission revisited

The criteria of entanglement generation and transmission, both in the cq and qq cases, are directly analogous to the requirement in classical information theory that the average probability of error, averaged over all codewords, be small. However, the requirements imposed in Section 5.2 are analogous to the stronger classical condition that the *maximal* probability of error be small, or that the probability of error for *each* pair of codewords be small. There are examples of classical multiple access channels for which, when each encoder is a deterministic function from the set of the messages to the set of input symbols, the maximal error capacity region is *strictly* smaller than the average error region [18]. However, it is known that if stochastic encoders are allowed (see Problem 3.2.4 in [11]), the maximal and average error capacity regions are equal.

It is well-known that randomization is not necessary for such an equivalence to hold for single-user channels, as Markov's inequality implies that a fraction of the codewords with the worst probability of error can be purged, while incurring a negligible loss of rate. The obstacle to utilizing such an approach for classical multiple access channels, and hence for quantum ones as well, is that there is no guarantee that a large enough subset of bad pairs of codewords decomposes as the product of subsets of each sender's codewords.

A particularly attractive feature of the requirements of Section 5.2 is that they ensure *composability*; when combined with other protocols satisfying analogous criteria, the joint protocol will satisfy similar properties. As an example, recent work on organizing and classifying quantum Shannon-theoretic protocols by means of *resource inequalities* [22], makes heavy use of such concatenation of quantum information processing protocols.

In the next two subsections, we cast the requirements outlined in Section 5.2 into somewhat simpler forms which are specific to each of the cq and qq cases. We will use these forms in order to prove the equivalences of entanglement transmission and strong subspace transmission in both the cq and qq cases.

### 8.3.1 classical-quantum scenario

Strong subspace transmission can be considered a more ambitious version of entanglement transmission, whereby rather than requiring Bob to transmit half of a maximally entangled state  $|\Phi\rangle^{B\tilde{B}}$ , it is instead required that he faithfully transmit the  $\tilde{B}$  part, presented to him, of *any* bipartite pure state  $|\Psi\rangle^{B\tilde{B}}$ , where  $|B|$  can be any finite number. The reader should note that this constitutes a generalization of the usual subspace transmission [6], as whenever  $|\Psi\rangle^{B\tilde{B}} = |\psi\rangle^B|\varphi\rangle^{\tilde{B}}$ , this amounts to requiring that  $|\varphi\rangle$  be transmitted faithfully. We further demand that the maximal error probability for the classical messages be small.

As with entanglement transmission, Alice will send classical information at rate  $R$  by preparing one of  $2^{nR}$  pure states  $\{|\phi_m\rangle^{A^n}\}_{m \in 2^{nR}}$ . As previously discussed, our more restrictive information transmission constraints can only be met by allowing Alice to employ a stochastic encoding. We assume that Alice begins by generating some randomness, modeled by the random variable  $X$ . To send message  $M = m$ , she prepares a state  $\phi_{f(m)}$ , where  $f(m) \equiv f_X(m)$  is a random encoding function, depending on the randomness in  $X$ . In the language of Section 5.2, this amounts to the definition of a  $c \rightarrow q$  encoding function  $\mathcal{E}_1: M \rightarrow A^n$ . Observe that our definition there already allows for randomness to be part of the encoding process.

Bob will apply an encoding  $\mathcal{E}: \tilde{B} \rightarrow B^m$  (this is just his encoding  $\mathcal{E}_2$  from Section 5.2 without a classical input), and Charlie will employ a decoding instrument  $\mathcal{D}: C^n \rightarrow \widehat{M}\widehat{B}$ . These maps require a more complicated structure than was required for entanglement generation and transmission. Indeed, these will be constructed by means of a protocol, to be described below, out of the entanglement transmission codes which were proved to exist in Section 8.2.1. The success probability for the protocol, conditioned on  $m$  being sent and  $|\Psi\rangle^{B\tilde{B}}$  being presented, can be expressed as

$$P_s(m, \Psi) = F\left(|f(m)\rangle^{\widehat{M}}|\Psi\rangle^{B\widehat{B}}, \mathcal{D} \circ \mathcal{N}^{\otimes n}(\phi_{f(m)}^{A^n} \otimes \mathcal{E}(\Psi^{B\tilde{B}}))\right).$$

We will say that  $(f, X, \{|\phi_m\rangle\}_{m \in 2^{nR}}, \mathcal{E}, \mathcal{D})$  is an  $(R, Q, n, \epsilon)$  *cq strong subspace transmission code* for the channel  $\mathcal{N}$  if, for every  $m \in 2^{nR}$  and every  $|\Psi\rangle^{B\tilde{B}}$ ,

$$\mathbb{E}_X P_s(m, \Psi) \geq 1 - \epsilon. \quad (8.8)$$

The rate pair  $(R, Q)$  is an *achievable cq rate pair for strong subspace transmission* if there is a sequence of  $(R, Q, n, \epsilon_n)$  cq random strong subspace transmission codes with  $\epsilon_n \rightarrow 0$ , and the capacity region  $\mathcal{CQ}(\mathcal{N})$  is closure of the collection of all such achievable rates.

### 8.3.2 quantum-quantum scenario

This scenario is the obvious combination of the relevant concepts from the previous scenario and the qq entanglement transmission scenario. Alice and Bob are respectively presented with the  $\tilde{A}$  and  $\tilde{B}$  parts of some pure bipartite states  $|\Psi_1\rangle^{A\tilde{A}}$  and  $|\Psi_2\rangle^{B\tilde{B}}$ . As before, we place no restriction on  $|A|$  and  $|B|$ , other than that they are finite. They employ their respective encodings  $\mathcal{E}_1$  and  $\mathcal{E}_2$  (which are just the encodings from Section 5.2 without classical inputs), while Charlie decodes with  $\mathcal{D}$ . As in the above cq case, the structure of these maps will be more complicated than in the previous two scenarios.  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D})$  is then a  $(Q_a, Q_b, n, \epsilon)$  *qq strong subspace transmission code* if

$$F\left(|\Psi_1\rangle^{A\tilde{A}}|\Psi_2\rangle^{B\tilde{B}}, \mathcal{D} \circ \mathcal{N}^{\otimes n} \circ (\mathcal{E}_1 \otimes \mathcal{E}_2)(\Psi_1^{A\tilde{A}} \otimes \Psi_2^{B\tilde{B}})\right) \geq 1 - \epsilon, \quad (8.9)$$

for every pair of pure bipartite states  $|\Psi_1\rangle^{A\tilde{A}}$  and  $|\Psi_2\rangle^{B\tilde{B}}$ . Achievable rates and the capacity region  $\mathcal{Q}(\mathcal{N})$  are defined as in the cq case.



## 8.4 Equivalence of entanglement transmission and strong subspace transmission

Let us first prove the easy directions. To see that  $\mathcal{CQ} \subseteq \mathcal{CQ}_{\text{et}}$ , note that given a strong subspace transmission code, if Alice uses any deterministic value  $x$  for her locally generated randomness  $X$ , the average classical error will be equal to the expected maximal classical error of the randomized code. Since the ability to transmit any state includes the maximally entangled case, this completes the claim. The inclusion  $\mathcal{Q} \subseteq \mathcal{Q}_{\text{et}}$  follows trivially. As any states can be transmitted, this certainly includes the case of a pair of maximally entangled states.

### 8.4.1 $\mathcal{CQ}_{\text{et}} \subseteq \mathcal{CQ}$

*Proof.* Suppose there exists an  $(R, Q, n, \epsilon^2/2)$  entanglement transmission codes with classical message states  $\{|\phi_m\rangle^{A^m}\}_{m \in 2^{nR}}$ , quantum encoding  $\mathcal{E}: \tilde{B} \rightarrow \hat{B}$ , and decoding instrument  $\mathcal{D}: C^n \rightarrow \widehat{M}\hat{B}$  with trace-reducing components  $\{\mathcal{D}_m: C^n \rightarrow \hat{B}\}$ , which transmits a maximally entangled state  $|\Phi\rangle^{A\tilde{A}}$ , where  $|A| < \infty$  (although  $|\tilde{A}| = 2^{nQ}$ ).

We will initially prove the equivalence by constructing a code which requires two independent sources of shared common randomness  $X$  and  $Y$ .  $X$  is assumed to be available to Alice and to Charlie, while  $Y$  is available to Bob and to Charlie. Then, we will argue that it is possible to eliminate the dependence on the shared randomness, by using the channel to send a negligibly small “random seed”, which can be recycled to construct a code which asymptotically achieves the same performance as the randomized one.

We begin by demonstrating how shared common randomness between Alice and Charlie allows Alice to send any message with low probability of error. Setting  $\mu = 2^{nR}$ , let the random variable  $X$  be uniformly distributed on the set  $\{1, \dots, \mu\}$ . To send message  $M = m$ , Alice computes  $m' = m + X$  modulo  $\mu$ . She then prepares the state  $|\phi_{m'}\rangle$  for transmission through the channel. Bob encodes the  $\tilde{B}$  part of  $|\Phi\rangle^{B\tilde{B}}$  with  $\mathcal{E}$ , and each sends appropriately through the channel. Charlie decodes as usual with the instrument  $\mathcal{D}$ . Denoting the classical output as  $\widehat{M}'$ , his declaration of

Alice's message is then  $\widehat{M} = \widehat{M}' - X$  modulo  $\mu$ . Defining the trace-reducing maps  $\mathcal{M}_m: \widetilde{B} \rightarrow \widehat{B}$  by

$$\mathcal{M}_m: \tau \mapsto \mathcal{D}_m \circ \mathcal{N}^{\otimes n}(\phi_m \otimes \mathcal{E}(\tau)),$$

and the trace-reducing average map as

$$\mathcal{M}: \tau \rightarrow \frac{1}{\mu} \sum_{m=1}^{\mu} \mathcal{M}_m(\tau),$$

we can rewrite the success criterion (8.1) for entanglement transmission as

$$F(|\Phi\rangle, \mathcal{M}(\Phi)) \geq 1 - \epsilon^2/2,$$

which, together with (6.3), implies that for the identity map  $\text{id}: \widetilde{B} \rightarrow \widehat{B}$ ,

$$|(\mathcal{M} - \text{id})(\Phi)|_1 \leq \epsilon. \tag{8.10}$$

The above randomization of the classical part of the protocol can be mathematically expressed by replacing the  $\mathcal{M}_m$  with  $\mathcal{M}_{m+X}$ . As tracing over the common randomness  $X$  is equivalent to computing the expectation with respect to  $X$ , we see that  $\mathbb{E}_X \mathcal{M}_{m+X} = \mathcal{M}$ , or rather

$$\mathbb{E}_X F(|\Phi\rangle, \mathcal{M}_{m+X}(\Phi)) = F(|\Phi\rangle, \mathcal{M}(\Phi)).$$

It is thus clear that the maximal error criterion for the randomized protocol is equal to the average criterion for the original one.

We continue by randomizing the quantum part of the classically randomized protocol. Setting  $d = 2^{n_Q} = |\widetilde{B}|$ , let  $\{U_y\}_{y \in d^2}$  be the collection of Weyl unitaries, or generalized Pauli operators, on the  $d$ -dimensional input space. Observe that for any

$\rho$ , acting with a uniformly random choice of Weyl unitary has a completely randomizing effect, in the sense that

$$\frac{1}{d^2} \sum_{y=1}^{d^2} U_y \rho U_y^{-1} = \pi_d.$$

Let the random variable  $Y$  be uniformly distributed on  $\{1, \dots, d^2\}$ . It will be convenient to define the common randomness state

$$\Upsilon^{Y_B Y_C} = \frac{1}{d^2} \sum_{y=1}^{d^2} |y\rangle\langle y|^{Y_B} \otimes |y\rangle\langle y|^{Y_C},$$

where the system  $Y_B$  is in the possession of Bob, while  $Y_C$  is possessed by Charlie. Define now the controlled unitaries  $\mathcal{U}_B: Y_B \tilde{B} \rightarrow Y_B \tilde{B}$  and  $\mathcal{U}_C: Y_C \hat{B} \rightarrow Y_C \hat{B}$  by

$$\mathcal{U}_B = \sum_{y=1}^{d^2} |y\rangle\langle y|^{Y_B} \otimes U_y$$

and

$$\mathcal{U}_C = \sum_{y=1}^{d^2} |y\rangle\langle y|^{Y_C} \otimes U_y^{-1}.$$

Suppose Bob is given the  $\tilde{B}$  part of an arbitrary pure state  $|\Psi\rangle^{B\tilde{B}}$ , and Alice sends the classical message  $M = m$ . For encoding, Bob will apply  $\mathcal{E} \circ \mathcal{U}_B$  to the combined system  $\Upsilon \otimes \Psi$ . Charlie decodes with  $\mathcal{U}_C \circ \mathcal{D}$ . If  $\mathcal{M}$  were equal to the perfect quantum channel  $\text{id}: \tilde{B} \rightarrow \hat{B}$ , this procedure would result in the state

$$\frac{1}{d^2} \sum_{y=1}^{d^2} |y\rangle\langle y|^{Y_B} \otimes |y\rangle\langle y|^{Y_C} \otimes \Psi.$$

Note that the common randomness is still available for reuse. Abbreviating  $|y\rangle\langle y|^Y =$

$|y\rangle\langle y|^{Y_B} \otimes |y\rangle\langle y|^{Y_C}$ , and  $|\Psi_y\rangle^{B\tilde{B}} = (1^B \otimes U_y)|\Psi\rangle$ , we write

$$\sigma^{YB\tilde{B}} = \mathcal{U}_B(\Upsilon \otimes \Psi) \quad (8.11)$$

$$= \frac{1}{d^2} \sum_{y=1}^{d^2} |y\rangle\langle y|^Y \otimes \Psi_y. \quad (8.12)$$

Observe that  $\sigma$  is an extension of the maximally mixed state  $\pi^{\tilde{B}}$ , and can be seen to arise by storing in  $Y$  the result of a von Neumann measurement along the basis  $\{|y\rangle^F\}_{y \in d^2}$  on the  $F$  part of the pure state

$$|\Gamma\rangle^{FB\tilde{B}} = \frac{1}{d} \sum_{y=1}^{d^2} |y\rangle^F |\Psi_y\rangle^{B\tilde{B}}.$$

Since  $\text{Tr}_{R'R} \Gamma = \text{Tr}_{YR} \sigma = \pi^{\tilde{B}}$ ,  $|\Gamma\rangle$  is maximally entangled between  $FB$  and  $\tilde{B}$ . So, there exists an isometry  $V: B \rightarrow FB$  such that  $(V \otimes 1^{\tilde{B}})|\Phi\rangle^{B\tilde{B}} = |\Gamma\rangle$ . This implies that there is a quantum operation  $\mathcal{O}: B \rightarrow YB$  such that  $(\mathcal{O} \otimes 1^{\tilde{B}})(\Phi) = \sigma$ . Define the trace-reducing map  $\mathcal{T}: \tilde{B} \rightarrow \hat{B}$ , which represents the coded channel with common randomness accounted for, by

$$\mathcal{T}: \tau \mapsto \text{Tr}_Y \mathcal{U}_C \circ \mathcal{M} \circ \mathcal{U}_B(\Upsilon \otimes \tau).$$

Recalling our denotation of the noiseless quantum channel  $\text{id}: \tilde{B} \rightarrow \hat{B}$ , as well as our convention that  $\text{id}$  acts as the identity on any system which is not  $\tilde{B}$ , we now bound

$$\begin{aligned} 1 - F(|\Psi\rangle, \mathcal{T}(\Psi)) &\leq |(\mathcal{T} - \text{id})(\Psi)|_1 \\ &\leq |(\mathcal{U}_C \circ \mathcal{M} \circ \mathcal{U}_B - \text{id})(\Upsilon \otimes \Psi)|_1 \\ &= |(\mathcal{M} - \text{id}) \circ \mathcal{U}_B(\Upsilon \otimes \Psi)|_1 \\ &= |(\mathcal{M} - \text{id})(\sigma)|_1 \\ &\leq |(\mathcal{M} - \text{id})(\Phi)|_1 \\ &\leq \epsilon, \end{aligned}$$

where the first line is by (6.1) and the second by monotonicity with respect to  $\text{Tr}_Y$ . The third follows from unitary invariance of the trace. The second to last inequality is a consequence of monotonicity with respect to  $\mathcal{O}$ , while the last is by (8.10). Note that by monotonicity, this implies that any density matrix  $\Omega^{B\tilde{B}}$  satisfies

$$|\mathcal{T}(\Omega) - \Omega|_1 \leq \epsilon. \quad (8.13)$$

We have thus shown that if Alice and Charlie have access to a common randomness source of rate  $R$ , while Bob and Charlie can access one of rate  $2Q$ , the conditions for strong subspace transmission can be satisfied. Next, we will illustrate that, by modifying our protocol, it is possible to reduce the amount of shared randomness required. Using the previous blocklength- $n$  construction, we will concatenate  $N$  such codes, where each utilizes the *same* shared randomness, to construct a new code with blocklength  $nN$ . For an arbitrary  $|\Psi^{(N)}\rangle^{B\tilde{B}^N}$ , further define the commuting operations  $\{\mathcal{T}_i\}_{i \in N}$ , where  $\mathcal{T}_i: \tilde{B}_i \rightarrow \hat{B}_i$  is  $\mathcal{T}$  acting on the  $i$ 'th tensor factor of  $\Psi^{(N)}$ . Setting  $\xi_0 \equiv \Psi^{(N)}$ , we then recursively define the density operators  $\xi_i = \mathcal{T}_i(\xi_{i-1})$ , noting that  $\xi_N = \mathcal{T}_N \circ \dots \circ \mathcal{T}_1(\xi_0) = \mathcal{T}^{\otimes N}(\Psi^{(N)})$ . Because of (8.13),  $|\xi_{i+1} - \xi_i|_1 = |\mathcal{T}_{i+1}(\xi_i) - \xi_i|_1 \leq \epsilon$ , and we can use the triangle inequality to estimate

$$\begin{aligned} |\mathcal{T}^{\otimes N}(\Psi^{(N)}) - \Psi^{(N)}|_1 &= |\xi_N - \xi_0|_1 \\ &\leq \sum_{i=1}^N |\xi_i - \xi_{i-1}|_1 \\ &\leq N\epsilon. \end{aligned}$$

By choosing  $N = \frac{1}{\sqrt{\epsilon}}$ , it is clear that we have reduced Alice's and Bob's shared randomness rates respectively to  $\sqrt{\epsilon}R$  and  $2\sqrt{\epsilon}Q$ , while the error on the  $N$ -blocked protocol is now  $\sqrt{\epsilon}$ . Next, we argue that by using two more blocks of length  $n$ , it is possible to simulate the shared randomness by having Alice send  $nR$  random bits  $X$  using the first block, while Bob locally prepares two copies of  $\Phi$ ,  $\Phi^{B_1\tilde{B}_1} \otimes \Phi^{B_2\tilde{B}_2}$ , and transmits the  $\tilde{B}_1\tilde{B}_2$  parts over the channel using both blocks. Charlie decodes each block separately, obtaining a random variable  $\hat{X}$  and the  $\hat{B}_1$  and  $\hat{B}_2$  parts of the post-decoded states  $\Omega_1^{B_1\hat{B}_1}$  and  $\Omega_2^{B_2\hat{B}_2}$ . Bob and Charlie then measure their respective parts

of  $\Omega_1 \otimes \Omega_2$  in some previously agreed upon orthogonal bases to obtain a simulation  $\widehat{\Upsilon}$  of the perfect shared randomness state which, by monotonicity and telescoping, satisfies

$$\begin{aligned} |\Upsilon - \widehat{\Upsilon}|_1 &\leq |\Phi \otimes \Phi - \Omega_1 \otimes \Omega_2|_1 \\ &\leq \epsilon^2. \end{aligned}$$

Further, the noisy shared randomness for the classical messages can be shown to satisfy

$$\begin{aligned} |\text{dist}(X, X) - \text{dist}(X, \widehat{X})|_1 &= 2 \Pr\{X = \widehat{X}\} \\ &\leq \epsilon^2. \end{aligned}$$

By monotonicity of trace distance and the triangle inequality, using the noisy common randomness state  $\widehat{\Upsilon}$  increases the estimate for each block by  $2\epsilon^2$ . For identical reasons, the same increase is incurred by using the noisy common randomness  $(X, \widehat{X})$ . Thus, accounting for both sources of noisy common randomness, the estimate (8.13) is changed to  $2\epsilon$ , provided that  $\epsilon \leq \frac{1}{4}$ . The noisy common randomness thus increases the bound on the error of the  $N$ -blocked protocol to  $2\sqrt{\epsilon}$ , while costing each of Alice and Bob a negligible rate overhead of  $\frac{2}{N+2}$  in order to seed the protocol.

The above protocol can be considered as defining an encoding map  $\mathcal{E}' : \widetilde{B}^N \rightarrow B'^{(N+2)n}$  and decoding instrument  $\mathcal{D} : C^{(N+2)n} \rightarrow \widehat{B}^N \widehat{M}^N$ . Thus, the protocol takes an  $(R, Q, n, \epsilon_n)$  cq entanglement transmission code and constructs an  $(R', Q', n', \epsilon'_{n'})$  strong subspace transmission code with cq rate pair  $(R', Q') = \left(\frac{R}{1+\epsilon'_{n'}}, \frac{Q}{1+\epsilon'_{n'}}\right)$ , where  $n' = \left(2 + \frac{1}{\sqrt{\epsilon_n}}\right)n$ , and  $\epsilon'_{n'} = 2\sqrt{\epsilon_n}$ . Now, if the rates  $(R, Q)$  are achievable cq rates for entanglement transmission, there must exist a sequence of  $(R, Q, n, 2\epsilon_n^2)$  entanglement transmission codes with  $\epsilon_n \rightarrow 0$ . Since this means that  $\frac{1}{1+2\sqrt{\epsilon_n}}$  increases to unity, we have shown that for any  $\delta > 0$ , every rate pair  $(R - \delta, Q - \delta)$  is an achievable cq rate pair for strong subspace transmission. Since the capacity regions for each scenario are defined as the closure of the achievable rates, this completes the proof. ■

### 8.4.2 $\mathcal{Q}_{\text{et}} \subseteq \mathcal{Q}$

*Proof.* We will employ similar techniques as were used in the previous proof to obtain this implication. Suppose there exists a  $(Q_a, Q_b, n, \frac{1}{2}\epsilon^2)$  qq entanglement transmission code  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{D})$ , with  $\mathcal{E}_1: \tilde{A} \rightarrow A^n$ ,  $\mathcal{E}_2: \tilde{B} \rightarrow B^n$ , and  $\mathcal{D}: C^n \rightarrow \widehat{A}\widehat{B}$ . Setting  $a = |\tilde{A}| = 2^{nQ_a}$  and  $b = |\tilde{B}| = 2^{nQ_b}$ , define the common randomness states

$$\Upsilon_X^{X_A X_C} = \frac{1}{a^2} \sum_{x=1}^{a^2} |x\rangle\langle x|^{X_A} \otimes |x\rangle\langle x|^{X_C}$$

and

$$\Upsilon_Y^{Y_B Y_C} = \frac{1}{b^2} \sum_{y=1}^{b^2} |y\rangle\langle y|^{Y_B} \otimes |y\rangle\langle y|^{Y_C}$$

These states will be used as partial inputs to the controlled unitaries

$$\begin{aligned} \mathcal{U}_A &= \sum_{x=1}^{a^2} |x\rangle\langle x|^{X_A} \otimes U_x, \\ \mathcal{U}_C &= \sum_{x=1}^{a^2} |x\rangle\langle x|^{X_C} \otimes U_x^{-1}, \\ \mathcal{V}_B &= \sum_{y=1}^{b^2} |y\rangle\langle y|^{Y_B} \otimes V_y, \\ \mathcal{V}_C &= \sum_{y=1}^{b^2} |y\rangle\langle y|^{Y_C} \otimes V_y^{-1} \end{aligned}$$

where, as before, we have utilized the Weyl unitaries  $\{U_x\}_{x \in a^2}$  and  $\{V_y\}_{y \in b^2}$ , which respectively completely randomize any states on  $a$ -dimensional and  $b$ -dimensional spaces. Suppose Alice and Bob are respectively presented with the  $\tilde{A}$  and  $\tilde{B}$  parts of the arbitrary pure states  $|\Psi_1\rangle^{A\tilde{A}}$  and  $|\Psi_2\rangle^{B\tilde{B}}$ . Writing  $\mathcal{M} = \mathcal{D} \circ \mathcal{N}^{\otimes n} \circ (\mathcal{E}_1 \otimes \mathcal{E}_2)$ , and defining the map  $\mathcal{T}: \tilde{A}\tilde{B} \rightarrow \widehat{A}\widehat{B}$  by

$$\mathcal{T}: \tau \mapsto (\mathcal{U}_C \otimes \mathcal{V}_C) \circ \mathcal{M} \circ (\mathcal{U}_A \otimes \mathcal{V}_B)(\tau \otimes \Upsilon_1 \otimes \Upsilon_2),$$

the overall joint state of the randomized protocol is given by  $\mathcal{T}(\Psi_1 \otimes \Psi_2)$ . Abbreviating

$$|xy\rangle\langle xy|^{XY} = |x\rangle\langle x|^{XA} \otimes |x\rangle\langle x|^{XC} \otimes |y\rangle\langle y|^{YB} \otimes |y\rangle\langle y|^{YC}$$

and defining  $|\Psi_x\rangle^{A\tilde{A}} = (1^A \otimes U_x)|\Psi_1\rangle$ ,  $|\Psi_y\rangle^{B\tilde{B}} = (1^B \otimes V_y)|\Psi_2\rangle$ , we write

$$\sigma^{XYAB\tilde{A}\tilde{B}} = \frac{1}{a^2 b^2} \sum_{xy} |xy\rangle\langle xy| \otimes \Psi_x \otimes \Psi_y.$$

By similar arguments as in the cq case, there exists a map  $\mathcal{O}: AB \rightarrow ABQR$  so that

$$(\mathcal{O} \otimes 1^{\tilde{A}\tilde{B}})(\Phi_1 \otimes \Phi_2) = \sigma.$$

Again, for the same reasons as in the cq case, we have

$$\begin{aligned} |(\mathcal{T} - \text{id})(\Psi_1 \otimes \Psi_2)|_1 &\leq |(\mathcal{M} - \text{id})(\sigma)|_1 \\ &\leq |(\mathcal{M} - \text{id})(\Phi_1 \otimes \Phi_2)|_1 \\ &\leq \epsilon. \end{aligned}$$

The rest of the proof is nearly identical to that from the previous section, so we omit these details, so as not to have to repeat our previous arguments here. ■



# Chapter 9

## Single-letter examples

Due to the regularized form of our Theorems 1 and 2, the possibility of actually computing the capacity regions seems generally out of reach. Here we give some examples of channels whose capacity region does in fact admit a single-letter characterization, in the sense that no regularization is necessary. In the first section below, we show that a certain erasure quantum erasure multiple access channel has an additive cq capacity region. The next two sections describe classes of channels which have additive single-user capacities. The contents of these two sections are essentially an elaboration of results which appear elsewhere in [15]. The last section demonstrates that the qq capacity region of a certain collective phase-flip channel has an additive capacity region.

### 9.1 Proof of additivity of $\mathcal{CQ}$ for quantum erasure multiple access channel

Our first example is a multiple access erasure channel  $\mathcal{N}: A'B' \rightarrow C$ , where  $|A'| = 2$ ,  $|B'| = d$  and  $|C| = d + 1$ . Alice will send classical information while Bob will send quantum. Fixing bases  $\{|0\rangle^{A'}, |1\rangle^{A'}\}$ ,  $\{|1\rangle^{B'}, \dots, |d\rangle^{B'}\}$ ,  $\{|0\rangle^C, \dots, |d\rangle^C\}$ , the channel

has  $d + 1$  operation elements

$$\begin{aligned} N_0 &= \sum_{j=1}^d |0\rangle^C \langle 0|^{A'} \langle j|^{B'} \\ N_i &= |i\rangle^C \langle 1|^{A'} \langle i|^{B'}, \quad i = 1, \dots, d. \end{aligned}$$

The action of the channel can be interpreted as follows. First, a projective measurement of Alice's input along  $\{|0\rangle, |1\rangle\}$  is performed. If the result is 0, Charlie's output is prepared in a pure state  $|0\rangle$ . Otherwise, Bob's input is transferred perfectly to the remaining degrees of freedom in Charlie's output. Bob's input is "erased", or otherwise ejected into the environment, whenever Alice sends  $|0\rangle$ , and is perfectly preserved when she sends  $|1\rangle$ . Indeed, the action of  $\mathcal{N}$  on  $\tau^{A'} \otimes \rho^{B'}$  is given by

$$\mathcal{N}(\tau \otimes \rho) = \tau_{00}|0\rangle\langle 0| + \tau_{11}\rho.$$

We will show that the cq capacity region of this channel,  $\mathcal{CQ}(\mathcal{N}_{\text{erasure}})$ , has a single-letter characterization given by the collection of pairs of nonnegative classical-quantum rates  $(R, Q)$  such that

$$\begin{aligned} R &\leq H(p) \\ Q &\leq (1 - 2p) \log d \end{aligned}$$

for some  $0 \leq p \leq \frac{1}{2}$ , constituting a generalization of results in [7] on single-user erasure channels to a multiuser setting. Figure 5.1 contains a plot of this region for the case where  $d = 2$ .

In the sense of (5.1), any state  $\Omega^{XBC^k}$  which arises from  $\mathcal{N}^{\otimes k}$  can be specified by fixing some pure state ensemble  $\{p(x), |\phi_x\rangle^{A'^k}\}$  and a pure bipartite state  $|\Psi\rangle^{BB'^k}$ . We thus write

$$\Omega = \sum_x p(x) |x\rangle\langle x|^X \otimes (1^B \otimes \mathcal{N}^{\otimes k})(\phi_x \otimes \Psi).$$

For a binary string  $y^k$ , let  $|y^k\rangle^{A'^k} = |y_1\rangle^{A'} \dots |y_k\rangle^{A'}$  be the associated computational basis state. Writing  $p(y^k|x) = |\langle y^k|\phi_x\rangle|^2$  defines the random variable  $Y^k$ , which is

correlated with  $X$ , and can be interpreted as the erasure pattern associated with the state  $\Omega$ . We next define another state of the form (5.1),

$$\Omega^{XY^k BC^k} = \sum_{x, y^k} p(x)p(y^k|x)|x\rangle\langle x|^X \otimes |y^k\rangle\langle y^k|^{Y^k} \otimes \mathcal{N}^{\otimes k}(|y^k\rangle\langle y^k| \otimes \Phi),$$

for

$$|\Phi\rangle^{BB^k} = \sum_{j^k} |j^k\rangle^B |j_1\rangle^{B'_1} \cdots |j_k\rangle^{B'_k},$$

where the summation is over  $d$ -ary strings of length  $k$ ,  $j^k = (j_1, \dots, j_k)$ . Finally, for

$$\begin{aligned} q_i &= \Pr\{Y_i = 0\}, \\ q &= \frac{1}{k} \sum_{i=1}^k q_i, \\ |\varphi\rangle^{BC} &= \frac{1}{\sqrt{d}} \sum_{j=1}^d |j\rangle^B |j\rangle^C, \end{aligned}$$

define a third state

$$\omega^{UBC} = q|0\rangle\langle 0|^U \otimes \pi_d^B \otimes |0\rangle\langle 0|^C + (1-q)|1\rangle\langle 1|^U \otimes \varphi^{BC}.$$

The above states can easily be seen to satisfy the following chain of inequalities

$$\begin{aligned}
I(X; C^k)_\Omega &= I(X; C^k)_{\Omega'} \\
&= I(X; Y^k)_{\Omega'} \\
&\leq H(Y^k)_{\Omega'} \\
&\leq \sum_{i=1}^k H(Y_i)_{\Omega'} \\
&= \sum_{i=1}^k H(q_i) \\
&\leq kH(q) \\
&= kH(U)_\omega \\
&= kI(U; C)_\omega.
\end{aligned}$$

The only nontrivial step above is that we have used the concavity of the binary entropy function in the last inequality. Furthermore, it is not hard to see that

$$\begin{aligned}
I_c(B \rangle C^k X)_\Omega &\leq I_c(B \rangle C^k XY^k)_{\Omega'} \\
&= I_c(B \rangle C^k Y^k)_{\Omega'} \\
&= kI_c(B \rangle CU)_\omega.
\end{aligned}$$

Thus, we have shown that for any state  $\Omega^{XBC^k}$  arising from  $\mathcal{N}^{\otimes k}$  in the sense of (5.1), there is a state  $\omega^{UBC}$  arising from  $\mathcal{N}$  in the same sense, allowing the multi-letter information quantities to be bounded by single-letter information quantities; i.e.  $\mathcal{CQ}(\mathcal{N}) = \mathcal{CQ}^{(1)}(\mathcal{N})$ . ■

As it is clear that  $I(U; C)_\omega = H(q)$ , we focus on calculating

$$\begin{aligned}
I_c(B \rangle CU)_\omega &= q \left( H(|0\rangle\langle 0|^C) - H(\pi_d^B \otimes |0\rangle\langle 0|^C) \right) + (1 - q) \left( H(\pi_d^C) - H(\varphi^{BC}) \right) \\
&= q(0 - \log d) + (1 - q)(\log d - 0) \\
&= (1 - 2q) \log d.
\end{aligned}$$

Note that the above quantity is a weighted average of a positive and a negative coherent information. It is perhaps tempting to interpret these terms as follows. The positive term can be considered as resulting from a preservation of quantum information, while the negative term can be seen as signifying a complete loss of quantum information to the environment. The overall coherent information is positive only when  $q < \frac{1}{2}$ , a result which is in agreement with the result of Bennett et al. [7] on the quantum capacity of a binary erasure channel. Varying  $0 \leq q \leq \frac{1}{2}$ , the rate pairs

$$\begin{aligned} (R, Q) &= (I(U; C), I_c(B \rangle CU))_\omega \\ &= (H(q), (1 - 2q) \log d) \end{aligned}$$

can be seen to parameterize the outer boundary of  $\mathcal{CQ}(\mathcal{N})$ , as is pictured in figure 5.1 for the case  $d = 2$ .

As an aside, we remark that this calculation, together with the quantum channel capacity theorem from [13], gives a direct derivation of the quantum capacity of a quantum erasure channel, without relying on the no-cloning and hashing arguments used in [7].

## 9.2 Degradable channels

While for the single-user capacity  $Q(\mathcal{N})$  of an arbitrary quantum channel  $\mathcal{N}: A' \rightarrow B$  is known not to be additive in general, there is a certain class of channels for which additivity follows relatively easily. This is the class of so-called *degradable channels* [15]. A channel  $\mathcal{N}$  is degradable if its complement  $\mathcal{N}^c: A' \rightarrow E$  is a stochastically degraded version of  $\mathcal{N}$ , i.e. if there exists a *degrading channel*  $\mathcal{N}^d: B \rightarrow E$  such that

$$\mathcal{N}^c = \mathcal{N}^d \circ \mathcal{N}.$$

Below, we will give a version of the proof from [15] of the additivity of the quantum capacity of an arbitrary degradable channel. Then, we argue that the maximum sum

rate bound of the qq capacity region is additive for such channels.

Assume that  $\mathcal{N}_1: A'_1 \rightarrow C_1$  and  $\mathcal{N}_2: A'_2 \rightarrow C_2$  are degradable, with isometric extensions  $\mathcal{U}_i: A'_i \rightarrow C_i E_i$ . Fix an input state  $|\Psi\rangle^{AA'_1 A'_2}$  which gives rise to the global state  $|\Omega\rangle^{AC^2 E^2} = \mathcal{U}_1 \otimes \mathcal{U}_2(\Psi^{AA'^2})$ , where the  $\mathcal{U}_i$  are isometric extensions of the  $\mathcal{N}_i$ . By degradability, there exist  $\mathcal{N}_i^{d_i}$ 's so that  $\mathcal{N}_i^c = \mathcal{N}_i^{d_i} \circ \mathcal{N}_i$ , where  $\mathcal{N}_i^c = \text{Tr}_{C_i} \mathcal{U}_i$ . Letting  $\mathcal{V}_i: C_i \rightarrow E_i F_i$  isometrically extend each  $\mathcal{N}_i^{d_i}$ , define  $\Theta^{E^2 F^2} = \mathcal{V}_1 \otimes \mathcal{V}_2(\text{Tr}_{AE^2} \Omega)$ . Then

$$\begin{aligned}
I_c(A \rangle C^2)_\Omega &= H(C^2)_\Omega - H(E^2)_\Omega \\
&= H(F^2 E^2)_\Theta - H(E^2)_\Theta \\
&= H(F^2 | E^2)_\Theta \\
&\leq H(F_1 | E_1)_\Theta + H(F_2 | E_2)_\Theta \\
&= H(F_1 E_1)_\Theta - H(E_1)_\Theta + H(F_2 E_2)_\Theta - H(E_2)_\Theta \\
&= H(C_1)_\Omega - H(E_1)_\Omega + H(C_2)_\Omega - H(E_2)_\Omega \\
&= H(C_1)_\Omega - H(AC^2 E_2)_\Omega + H(C_2)_\Omega - H(AC^2 E_1)_\Omega \\
&= I_c(AC_2 E_2 \rangle C_1)_\Omega + I_c(AC_1 E_1 \rangle C_2)_\Omega \\
&= I_c(A_1 \rangle C_1)_{\omega_1} + I_c(A_2 \rangle C_2)_{\omega_2}
\end{aligned}$$

where the inequality is by Lemma 8. In the last line, we set  $\omega_i^{A_i C_i} = \mathcal{N}_i(\Psi)$ , identifying  $A_1 \equiv AA'_2$  and  $A_2 \equiv AA'_1$ . All other steps are either by the fact that isometries preserve entropy or by other trivial rewritings.

Now, if we are given  $k$  identical channels  $\mathcal{N}: A'_i \rightarrow C_i$  and we fix an input state  $|\Psi\rangle^{AA'^k}$  giving rise to  $|\Omega\rangle^{AC^n E^n} = \mathcal{U}^{\otimes k}(\Psi^{AA'^k})$ , recursive application of the above yields

$$I_c(A \rangle C^k)_\Omega \leq \sum_i I_c(A_i \rangle C_i)_{\omega_i}$$

where  $A_i = AA'_1 \cdots A'_{i-1} A'_{i+1} \cdots A'_k$ ,  $\omega_i^{A_i C_i} = \mathcal{N}_i(\Psi)$ , and  $\mathcal{N}_i$  is  $\mathcal{N}$  acting on the  $i$ th tensor factor. Choosing

$$i^* = \arg \max_i \{I_c(A_i \rangle C_i)_{\omega_i}\}$$

yields

$$\frac{1}{k} I_c(A \rangle C^k)_\Omega \leq I_c(A_{i^*} \rangle C_{i^*})_{\omega_{i^*}} \leq \max_{\omega \in \mathcal{AC}} I_c(A \rangle C)_\omega = Q^{(1)}(\mathcal{N}),$$

where the maximization is as over all  $\omega = \mathcal{N}(\phi^{AA'})$ .

Let us phrase this conclusion using different notation. Let  $\tau^{A^k B^k}$  be arbitrary, and define  $\tau_i^{A_i B_i} = \text{Tr}_{/A_i B_i} \tau$ , where  $\text{Tr}_{/A_i B_i}$  denotes the partial trace over all systems which are *not*  $A_i B_i$ . Then

$$I_c(\tau, \mathcal{N}^{\otimes k}) \leq k I_c(\tau_{i^*}, \mathcal{N}),$$

where

$$i^* = \arg \max_i I_c(\tau_i, \mathcal{N}).$$

Now, if  $\rho^{A^k}$  and  $\sigma^{B^k}$  are arbitrary, and we define  $\rho_i = \text{Tr}_{/A_i} \rho$  and  $\sigma_i = \text{Tr}_{/B_i} \sigma$ , observe that if  $\tau = \rho \otimes \sigma$ , then  $\tau_i = \rho_i \otimes \sigma_i$ . This immediately implies that

$$I_c(\rho \otimes \sigma, \mathcal{N}^{\otimes k}) \leq k I_c(\rho_{i^*} \otimes \sigma_{i^*}, \mathcal{N}),$$

where

$$i^* = \arg \max_i I_c(\rho_i \otimes \sigma_i, \mathcal{N}),$$

proving that the maximum sum rate of any degradable channel is additive, even when the inputs are restricted to be product states. This fact will be useful in Section 9.4, where we give a channel whose qq capacity region is single-letter.

### 9.3 Generalized dephasing channels

In this section we describe a certain subclass of the class of degradable channels. These are channels  $\mathcal{N}: A' \rightarrow B$  with  $|A| = |B| = d$  for which there is a particular orthogonal basis  $\{|x\rangle^{A'}\}$  which can be transmitted through the channel without error

$$\mathcal{N}(|x\rangle\langle x|) = |x\rangle\langle x|$$

although superpositions of these basis vectors are potentially subject to noise. Here,  $\{|x\rangle^B\}$  is a corresponding orthogonal basis for  $B$ . Such a channel has an isometric extension  $\mathcal{U}: A' \rightarrow BE$  given by

$$\mathcal{U} = \sum_x |x\rangle^B |\phi_x\rangle^E \langle x|^{A'},$$

where the states  $|\phi_x\rangle^E$  are not necessarily orthogonal. To see that these channels are degradable, observe that for any input state  $\rho^{A'}$ ,

$$\begin{aligned} \mathcal{N}^c(\rho) &= \text{Tr}_B \mathcal{U}(\rho) \\ &= \sum_x \langle x|^B \left( \sum_{x''x'} |x''\rangle^B |\phi_{x''}\rangle^E \langle x''|^{A'} \rho |x'\rangle^{A'} \langle x'|^B \langle \phi_{x'}|^E \right) |x\rangle^B \\ &= \sum_x \langle x|\rho|x\rangle \phi_x^E. \end{aligned}$$

Note that  $\mathcal{N}^c(\rho)$ , depends only on the diagonal matrix elements of  $\rho$  (when it is expressed in the dephasing basis). However, these are exactly the matrix elements which are unaffected by the action of  $\mathcal{N}$ , making degradability evident. In fact, the degrading channel is precisely  $\mathcal{N}^c$ , i.e.

$$\mathcal{N}^c = \mathcal{N}^c \circ \mathcal{N}.$$

It is interesting to relate the isometric extension  $\mathcal{U}_{\mathcal{N}}$  to the operator sum representation for  $\mathcal{N}$ . To do this, first express  $\mathcal{U}_{\mathcal{N}}$  in the flattened representation

$$\mathcal{U}_{\mathcal{N}} = \begin{pmatrix} |\phi_1\rangle & & & \\ & |\phi_2\rangle & & \\ & & \ddots & \\ & & & |\phi_d\rangle \end{pmatrix}.$$

Supposing that  $|E| = k$ , note that the matrix is “block diagonal”, with  $d$   $k \times 1$  blocks, where this is expressed as a map to the system  $EB$ . Regrouping the rows into to  $k$





basis while ignoring the result. This has the effect of setting all of the off-diagonal matrix elements of  $\rho$  equal to zero.  $\Delta$  obeys the following equations:

$$\begin{aligned}\mathcal{N}^c &= \mathcal{N}^c \circ \Delta \\ H(\Delta(\rho)) &\geq H(\rho).\end{aligned}$$

The first is because  $\mathcal{N}^c$  only depends on the diagonal components of  $\rho$ , while the second is proved in [34]. Observe that the inequality is saturated for diagonal  $\rho$ . Because of this, we may write

$$\begin{aligned}Q(\mathcal{N}) &= \max_{\rho} I_c(\rho, \mathcal{N}) \\ &= \max_{\rho} \left\{ H(\mathcal{N}(\rho)) - H(\mathcal{N}^c(\rho)) \right\} \\ &= \max_{\rho} \left\{ H(\mathcal{N} \circ \Delta(\rho)) - H(\mathcal{N}^c \circ \Delta(\rho)) \right\} \\ &= \max_{p(x)} \left\{ H(X) - H\left(\sum_x p(x)\phi_x\right) \right\}.\end{aligned}$$

## 9.4 Proof of additivity of $\mathcal{Q}$ for collective phase-flip channel

While the description of the capacity region  $\mathcal{Q}$  in Theorem 2 generally requires taking a many-letter limit, we give here an example of a quantum multiple access channel  $\mathcal{N}_p: A'B' \rightarrow C$  for which that description can be single-letterized. The channel  $\mathcal{N}_p$  takes as input two qubits, one from Alice and the other from Bob. With probability  $p$ , the channel causes each qubit to undergo a phase flip, by rotating each by  $180^\circ$  about its z-axis before it is received by the receiver Charlie. The action of  $\mathcal{N}_p$  on an input density operator  $\rho^{A'B'}$  is described in terms of the operator sum representation as

$$\mathcal{N}_p(\rho) = (1 - p)\rho + p(\sigma_z \otimes \sigma_z)\rho(\sigma_z \otimes \sigma_z),$$

where

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

is the Pauli phase flip matrix. We will demonstrate that  $\mathcal{Q}(\mathcal{N}_p)$  is equal to the collection of all pairs of nonnegative rates  $(Q_a, Q_b)$  which satisfy

$$\begin{aligned} Q_a &\leq 1 \\ Q_b &\leq 1 \\ Q_a + Q_b &\leq 2 - H(p). \end{aligned}$$

*Proof.* In order to prove this, we first recall that the maximum of the sum rate bound  $I_c(AB \rangle C)$  over all inputs of the form (5.2) is additive. Next, we calculate  $Q(\mathcal{N}_p)$ , the single-user capacity of the channel, and observe that it is achieved for inputs of the form (5.2), implying that the maximum sum rate bound equals the capacity. Then, we show that for the same inputs, the bounds  $I_c(A \rangle BC)$  and  $I_c(B \rangle AC)$  on the individual rates are as large as is possible. The characterization in terms of a single pentagon will then follow.

We first note that the the operator sum matrices  $\sqrt{p}\sigma_x \otimes \sigma_x$  and  $\sqrt{1-p}1_4$  commute. By results in the previous section, we conclude that  $\mathcal{N}_p$  is an example of a *generalized dephasing channel* and thus, the following two conditions are satisfied:

- for any state  $\Omega^{ABC^k} = \mathcal{N}^{\otimes k}(\Psi^{ABA^k B^k})$  arising from  $\mathcal{N}^{\otimes k}$  (where Alice and Bob can jointly prepare any state at the inputs), there is a state  $\omega^{ABC} = \mathcal{N}(\psi^{ABA'B'})$  for which

$$I_c(AB \rangle C)_\omega \geq \frac{1}{k} I_c(AB \rangle C^k)_\Omega.$$

Furthermore, the input density operator  $\rho^{A'B'} = \text{Tr}_{AB} \psi^{ABA'B'}$  is diagonal in the dephasing basis of  $\mathcal{N}_p$ .

- for any state  $\Omega'^{ABC^k} = \mathcal{N}^{\otimes k}(\Psi_1^{AA^k} \otimes \Psi_2^{BB^k})$  arising from  $\mathcal{N}^{\otimes k}$  in the sense of (5.2), there is a state  $\omega'^{ABC} = \mathcal{N}(\phi_1^{AA'} \otimes \phi_2^{BB'})$  arising from  $\mathcal{N}$  in the same sense for which

$$I_c(AB \rangle C)_{\omega'} \geq \frac{1}{k} I_c(AB \rangle C^k)_\Omega'.$$

The first condition above says that the single-user capacity  $Q(\mathcal{N}_p)$  is additive. It also guarantees that the relevant maximization is achieved by an input density operator  $\rho^{A'B'}$  which is diagonal in the dephasing basis. The second condition guarantees that the constrained single-user capacity of  $\mathcal{N}_p$ , when the users are constrained to preparing product input states, is additive.

In order to compute  $Q(\mathcal{N}_p)$ , let us first write an isometric extension  $\mathcal{U}: AB \rightarrow CE$  of  $\mathcal{N}_p$  as

$$\mathcal{U}|i\rangle^A|i\rangle^B = |ij\rangle^C|\phi_{ij}\rangle^E,$$

where

$$|\phi_{00}\rangle^E = |\phi_{11}\rangle^E = \sqrt{1-p}|0\rangle^E + \sqrt{p}|1\rangle^E \equiv |\phi_+\rangle^E$$

and

$$|\phi_{01}\rangle^E = |\phi_{10}\rangle^E = \sqrt{1-p}|0\rangle^E - \sqrt{p}|1\rangle^E \equiv |\phi_-\rangle^E.$$

A complementary channel  $\mathcal{N}_p^c$  is then defined as

$$\begin{aligned} \mathcal{N}_p^c(\rho) &= \text{Tr}_C \mathcal{U}(\rho) \\ &= \sum_{ij} |\phi_{ij}\rangle\langle i|\langle j|\rho|i\rangle|j\rangle\langle\phi_{ij}| \\ &= \sum_{ij} \rho_{ij} \phi_{ij} \\ &= (\rho_{00} + \rho_{11})\phi_+ + (\rho_{01} + \rho_{10})\phi_-. \end{aligned}$$

Observe that the output of the  $\mathcal{N}_p^c$  depends only on the diagonal elements of  $\rho$ , when  $\rho$  is written in the dephasing basis  $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ . Define  $\alpha = \rho_{00} + \rho_{11}$ . As  $Q(\mathcal{N}_p)$  is achieved when  $\rho$  is diagonal in this basis, let us calculate

$$\begin{aligned} H(C) &= H(A'B') \\ &= H(\{\rho_{00}, \rho_{01}, \rho_{10}, \rho_{11}\}) \\ &= H(\alpha) + \alpha H\left(\frac{\rho_{00}}{\alpha}\right) + (1-\alpha)H\left(\frac{\rho_{01}}{1-\alpha}\right) \\ &\leq H(\alpha) + 1, \end{aligned}$$

where the inequality is saturated when  $\rho_{00} = \rho_{11} = \frac{\alpha}{2}$  and  $\rho_{01} = \rho_{10} = \frac{1-\alpha}{2}$ . It thus suffices to optimize over the class of states

$$\rho^{A'B'} = \begin{pmatrix} \frac{\alpha}{2} & 0 & 0 & 0 \\ 0 & \frac{1-\alpha}{2} & 0 & 0 \\ 0 & 0 & \frac{1-\alpha}{2} & 0 \\ 0 & 0 & 0 & \frac{\alpha}{2} \end{pmatrix}$$

for which

$$H(C) = H(\rho) = 1 + H(\alpha),$$

Note that we may express

$$\phi_{\pm}^E = \frac{1}{2} \left( 1 \pm \sqrt{p(1-p)}\sigma_x - (1-2p)\sigma_z \right),$$

allowing us to write

$$\mathcal{N}_p^c(\rho) = \alpha\phi_+ + (1-\alpha)\phi_- = \frac{1}{2} \left( 1 + (2\alpha-1)\sqrt{p(1-p)}\sigma_x - (1-2p)\sigma_z \right),$$

so that  $H(E) = H\left(\frac{1}{2}(1 + \sqrt{p(1-p)}(2\alpha-1)^2 + (1-2p)^2)\right)$ . Thus,

$$\begin{aligned} I_c(\rho, \mathcal{N}) &= H(\mathcal{N}_p(\rho)) - H(\mathcal{N}_p^c(\rho)) \\ &= 1 + H(\alpha) - H\left(\frac{1}{2}(1 + \sqrt{p(1-p)}(2\alpha-1)^2 + (1-2p)^2)\right) \\ &\equiv h(\alpha). \end{aligned}$$

For fixed  $p$ ,  $h(\alpha)$  is symmetric about  $\alpha = \frac{1}{2}$ , and has a first derivative which is positive for  $0 \leq \alpha < \frac{1}{2}$  (and is thus negative for  $\frac{1}{2} < \alpha \leq 1$ ). Because  $h(\alpha)$  is continuous on  $0 \leq \alpha \leq 1$ , its maximum is attained when  $\alpha = \frac{1}{2}$ , so that

$$\max_{\rho} I_c(\rho, \mathcal{N}) = I_c(\pi^{A'B'}, \mathcal{N}) = 1 + H\left(\frac{1}{2}\right) - H(p) = 2 - H(p).$$

So we see that the maximum is already achieved for a product state  $\pi^{A'B'} = \pi^{A'} \otimes \pi^{B'}$ .

Define the Bell states

$$|\psi_{\pm}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle).$$

As  $|\psi_{+}\rangle$  purifies the maximally mixed state  $\pi_2$ , let us define the global state

$$\omega^{ABC} = \mathcal{N}(\psi_{+}^{AA'} \otimes \psi_{+}^{BB'}).$$

Identifying  $C = \widehat{A}\widehat{B}$  in the obvious way, let us reexpress

$$\omega^{A\widehat{A}B\widehat{B}} = (1-p)\psi_{+}^{A\widehat{A}} \otimes \psi_{+}^{B\widehat{B}} + p\psi_{-}^{A\widehat{A}} \otimes \psi_{-}^{B\widehat{B}}.$$

It is now a simple task to calculate

$$\begin{aligned} H(ABC) &= H(\omega) = H(p) \\ H(C) &= H(\pi^C) = 2 \\ H(AC) &= H(A\widehat{A}) + H(\widehat{B}) = H(p) + 1 = H(BC). \end{aligned}$$

Combining these gives the relevant coherent informations

$$\begin{aligned} I_c(AB \rangle C) &= H(C) - H(ABC) = 2 - H(p) \\ I_c(A \rangle BC) &= H(BC) - H(ABC) = 1 + H(p) - H(p) = 1 \\ I_c(B \rangle AC) &= H(AC) - H(ABC) = 1. \end{aligned}$$

As we saw in Section 3.4,  $I_c(A \rangle BC) \leq \log |A'| = 1$  and  $I_c(B \rangle AC) \leq \log |B'| = 1$  for any state arising from  $\mathcal{N}$ . The individual rate bounds are thus saturated and the claim follows. ■

# Chapter 10

## Discussion

There have been a number of results analyzing multiterminal coding problems in quantum Shannon theory. For an i.i.d. classical-quantum source  $XB$ , Devetak and Winter [14] have proved a Slepian-Wolf-like coding theorem achieving the cq rate pair  $(H(X|B), H(B))$  for classical data compression with quantum side information. Such codes extract classical side information from  $B^n$  to aid in compressing  $X^n$ . The extraction of side information is done in such a way as to cause a negligible disturbance to  $B^n$ . Our Theorem 1 is somewhat of this flavor. There, the quantum state of  $C^n$  is measured to extract Alice's classical message which, in turn, is used as side information for decoding Bob's quantum information. Analogous results to ours were obtained by Winter in his analysis of a multiple access channel with classical inputs and a quantum output, whereby the classical decoded message of one sender can be used as side information to increase the classical capacity of another sender.

We further mention the obvious connection between our coding theorems and the subject of channel codes with side information available to the receiver. The more difficult problem of classical and quantum capacities when side information is available at the *encoder* is analyzed by Devetak and Yard in [17], constituting quantum generalizations of results obtained by Gelfand and Pinsker [21] for classical channels with side information.

In an earlier draft of [53], we characterized  $\mathcal{Q}(\mathcal{N})$  as the closure of a regularized

union of rectangles

$$0 \leq R \leq \frac{1}{k} I_c(A) C^k$$

$$0 \leq S \leq \frac{1}{k} I_c(B) C^k.$$

This solution had been conjectured on the basis of a duality between classical Slepian-Wolf distributed source coding and classical multiple-access channels [11, 10], as well as on a purported no-go theorem for distributed data compression of so-called irreducible pure state ensembles that appeared in an early version of [3]. After the earlier preprint was made available, Andreas Winter announced [50] recent progress with Jonathan Oppenheim and Michal Horodecki [27] on the quantum Slepian-Wolf problem, offering a characterization identical in functional form to the classical one, while also supplying an interpretation of negative rates and apparently evading the no-go theorem. Motivated by the earlier mentioned duality, he informed us that the qq capacity region could also be characterized in direct analogy to the classical case. Subsequently, we found that we could modify our previous coding theorem to achieve the new region, provided that the rates are nonnegative. After those events unfolded, the authors of [3] found an error in the proof of their no-go theorem, leading to a revised version consistent with the newer developments. Our earlier characterization of  $\mathcal{Q}(\mathcal{N})$ , while correct, is contained in the rate region of Theorem 2 for any finite  $k$ , frequently strictly so. The newer theorem, therefore, gives a more accurate approximation to the rate region for finite  $k$ . In fact, for any state arising from the channel which does not saturate the strong subadditivity inequality [23], the corresponding pentagon and rectangle regions are distinct. As seen in Section 9.2, another beneficial feature of the new characterization is that for any channel which is *degradable*, the maximum sum rate bound  $R + S \leq \max I_c(AB)C$  is additive, where the maximization is over all states of the form (5.2). Furthermore, recall that in Section 9.4, the pentagon characterization was single-letterized for the collective phase flip channel. On the other hand, computer calculations have revealed that the rectangle region does *not* lead to a single-letter characterization of that channel. This seems to indicate that the newer characterization is the “correct” one, at least for that particular



channel.

More recently, we discovered that the same technique used to prove the new characterization of  $\mathcal{Q}(\mathcal{N})$  implies a new cq coding theorem, and thus a new characterization of  $\mathcal{CQ}(\mathcal{N})$ . By techniques nearly identical to those employed in the coding theorem for Theorem 2, it is possible to achieve the cq rate pair

$$(R, Q) = (I(X; BC), I_c(B \rangle C))$$

corresponding to Bob's quantum information being used as side information for decoding Alice's classical message. This is accomplished by having Charlie isometrically decode Bob's quantum information, then coherently decode to produce an effective channel  $\mathcal{N}_1: A' \rightarrow BC$  so that Alice can transmit classically at a higher rate. The new characterization is then a regularized union of pentagons, consisting of pairs of nonnegative rates  $(R, Q)$  satisfying

$$\begin{aligned} r &\leq I(X; BC) \\ S &\leq I_c(B \rangle CX) \\ r + S &\leq I(X; C) + I_c(B \rangle CX) = I(X; BC) + I_c(B \rangle C). \end{aligned}$$

Surprisingly, it is thus possible to characterize each of  $\mathcal{CQ}(\mathcal{N})$  and  $\mathcal{Q}(\mathcal{N})$  in terms of pentagons, in analogy to the original classical result. This situation makes apparent the dangers of being satisfied with regularized expressions for capacity regions. Without being able to prove single-letterization steps in the converses, it is hard to differentiate which characterization is the "right" one. While it is intuitively satisfying to see analogous formulae appear in both the classical and quantum theories, the regularized nature of the quantum results blurs the similarity. Indeed, the problems with single-letterization for single-user channels appear to be amplified when analyzing quantum networks (see e.g. [19]). While  $\mathcal{Q}$  is additive for the collective phase flip channel of Section 9.4, this behavior does not appear to be generic for the classes of degradable or generalized dephasing channels, as the saturation of the individual rate bounds for that example seem to be the source of additivity. Perhaps this indicates

that the necessity of understanding the capacities of single-user channels at a level beyond regularized optimizations is even more pressing than previously thought. It should be mentioned that for the erasure channel analyzed in Section 9.1, the newer description of  $\mathcal{CQ}(\mathcal{N})$  is not an issue, as the new corner point is contained in the old rectangle for any state arising from any number of parallel instances of the erasure channel.

Consider the full simultaneous classical-quantum region  $\mathcal{S}(\mathcal{N})$  defined in Section 5.2. This region can be characterized in a way that generalizes Theorems 1 and 2 as the regularization of the region  $\mathcal{S}^{(1)}(\mathcal{N})$ , defined as the vectors of nonnegative rates  $(R_a, R_b, Q_a, Q_b)$  satisfying

$$\begin{aligned} R_a &\leq I(X; C|Y) \\ R_b &\leq I(Y; C|X) \\ R_a + R_b &\leq I(XY; C) \\ Q_a &\leq I_c(A)BCXY) \\ Q_b &\leq I_c(B)ACXY) \\ Q_a + Q_b &\leq I_c(AB)CXY) \end{aligned}$$

for some state of the form

$$\sigma^{XYABC} = \sum_{x,y} p(x)p(y)|x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y \otimes \mathcal{N}(\psi_x^{AA'} \otimes \phi_y^{BB'}),$$

arising from the action of  $\mathcal{N}$  on the  $A'$  and  $B'$  parts of some pure state ensembles  $\{p(x), |\psi_x\rangle^{AA'}\}$ ,  $\{p(y), |\phi_y\rangle^{BB'}\}$ . Briefly, achievability of this region is obtained as follows. Using techniques introduced in [15], each sender “shapes” their quantum information into HSW codewords. Decoding is accomplished by first decoding all of the classical information, then using that information as side information for a quantum decoder. A formal proof of the achievability of this region is found in [52]. The main result of [15], the regularized optimization of the cq result from [49] over pairs of input ensembles, and our Theorems 1 and 2 follow as corollaries of

the corresponding capacity theorem. Indeed, the six two-dimensional “shadows” of the above region, obtained by setting pairs of rates equal to zero, reproduce those aforementioned results. This characterization, however, only utilizes the rectangle description of  $\mathcal{CQ}(\mathcal{N})$ . It is indeed possible to write a more accurate regularized description of  $\mathcal{S}(\mathcal{N})$  which generalizes the pentagon characterizations of  $\mathcal{CQ}(\mathcal{N})$  and  $\mathcal{Q}(\mathcal{N})$ , although we will not pursue that at this time.

# Chapter 11

## Appendix

### 11.1 Quantum instruments and coherent information

For some finite set  $\mathcal{S}$ , consider a labelled collection of channels  $\{\mathcal{N}_s\}_{s \in \mathcal{S}}$ , where  $\mathcal{N}_s: A' \rightarrow B$ . Define an instrument  $\mathcal{N}: A' \rightarrow SB$  to act as

$$\mathcal{N}: \tau \rightarrow \sum_s p(s) |s\rangle\langle s|^S \otimes \mathcal{N}_s(\tau).$$

An instrument channel such as  $\mathcal{N}$  may be interpreted as one with classical state information made available to the receiver. We will show that every channel  $\mathcal{N}^c: A' \rightarrow E$  which is complementary to  $\mathcal{N}$  is an instrument as well, as the environment  $E$  contains a copy of  $S$ . In other words, the classical state information is also available to an eavesdropper with full control of the environment.

An isometric extension  $\mathcal{U}$  of  $\mathcal{N}$  may be constructed as follows. First, fix isometric extensions  $\mathcal{U}_s: A' \rightarrow E'B$  for the individual  $\mathcal{N}_s$ 's. Then, define  $\mathcal{U}: A' \rightarrow SEB$  via

$$\mathcal{U} = \sum_s \sqrt{p(s)} |s\rangle^S |s\rangle^{E''} \otimes \mathcal{U}_s,$$

taking  $E = E'E''$ . That this is indeed an isometry is evident, because  $\mathcal{U}^\dagger \mathcal{U} =$

$\sum_s p(s) \mathcal{U}_s^\dagger \mathcal{U}_s = \sum_s p(s) 1^{A'} = 1^{A'}$ . We may further check that  $\mathcal{U}$  is in fact an extension of  $\mathcal{N}$ , by calculating

$$\begin{aligned} \mathrm{Tr}_E \mathcal{U}(\tau) &= \mathrm{Tr}_{E'} \mathrm{Tr}_{E''} \mathcal{U}(\tau) \\ &= \mathrm{Tr}_{E'} \sum_s p(s) |s\rangle \langle s|^S \otimes \mathcal{U}_s(\tau) \\ &= \sum_s p(s) |s\rangle \langle s|^S \otimes \mathcal{N}_s(\tau) \\ &= \mathcal{N}(\tau). \end{aligned}$$

Thus, the action of the complementary channel  $\mathcal{N}^c$  can be defined via  $\mathcal{U}$  as

$$\begin{aligned} \mathcal{N}^c(\tau) &= \mathrm{Tr}_{BS} \mathcal{U}(\tau) \\ &= \mathrm{Tr}_B \sum_s p(s) |s\rangle \langle s|^{E''} \otimes \mathcal{U}_s(\tau) \\ &= \sum_s p(s) |s\rangle \langle s|^{E''} \otimes \mathcal{N}_s^c(\tau), \end{aligned}$$

where the  $\mathcal{N}_s^c = \mathrm{Tr}_B \mathcal{U}_s$  are complementary channels to the  $\mathcal{N}_s$ 's.

## 11.2 Proof of convexity of $\mathcal{CQ}$ and $\mathcal{Q}$

Let  $\mathcal{N} : A'B' \rightarrow C$  be a quantum multiple access channel. We will prove that  $\mathcal{Q}(\mathcal{N})$  is convex, as the proof for  $\mathcal{CQ}$  is identical. Let  $k_0$  and  $k_1$  be positive integers, and fix any two states of the form (5.2),  $\sigma_0^{A_0 B_0 C^{k_0}}$  and  $\sigma_1^{A_1 B_1 C^{k_1}}$ . Then  $(R_0, S_0), (R_1, S_1) \in \mathcal{Q}(\mathcal{N})$ , where for  $i \in \{0, 1\}$ ,

$$\begin{aligned} R_i &= \frac{1}{k_i} I_c(A_i)_{C^{k_i} \sigma_i} \\ S_i &= \frac{1}{k_i} I_c(B_i)_{C^{k_i} \sigma_i}. \end{aligned}$$

We will now show that for any rational  $0 \leq \lambda \leq 1$ ,  $\lambda(R_0, S_0) + (1-\lambda)(R_1, S_1) \in \mathcal{Q}(\mathcal{N})$ . We first write  $\lambda = \frac{\alpha}{\beta}$ , for integers satisfying  $\beta > 0$ ,  $\beta \geq \alpha \geq 0$ . Setting  $p_0 = \alpha k_1$ ,  $p_1 = (\beta - \alpha)k_0$ , and  $k = p_0 k_0 + p_1 k_1$ , define the composite systems  $A = A_0^{p_0} A_1^{p_1}$  and

$B = B_0^{p_0} B_1^{p_1}$ , as well as the density matrix  $\sigma^{ABC^k} = \sigma_0^{\otimes p_0} \otimes \sigma_1^{\otimes p_1}$ , which is also of the form (5.2). Additivity of coherent information across product states and some simple algebra gives

$$\begin{aligned} \frac{1}{k} I_c(A \rangle C^k)_\sigma &= \frac{p_0}{k} I_c(A_0 \rangle C^{k_0})_{\sigma_0} + \frac{p_1}{k} I_c(A_1 \rangle C^{k_1})_{\sigma_1} \\ &= \frac{p_0 k_0 R_0 + p_1 k_1 R_1}{p_0 k_0 + p_1 k_1} \\ &= \lambda R_0 + (1 - \lambda) R_1. \end{aligned}$$

An identical calculation shows that  $\frac{1}{k} I_c(B \rangle C^k)_\sigma = \lambda S_0 + (1 - \lambda) S_1$ . As  $\mathcal{Q}(\mathcal{N})$  was defined as the topological closure of rate pairs corresponding to states which appropriately arise from the channel, the result follows because the set of previously considered  $\lambda$ 's comprises a dense subset of the unit interval. ■

### 11.3 Proof of cardinality bound on $\mathcal{X}$ .

Begin by fixing a finite set  $\mathcal{X}$ , a labelled collection of pure states  $\{|\phi_x\rangle^{A'}\}_{x \in \mathcal{X}}$ , and a pure bipartite state  $|\Psi\rangle^{BB'}$ . For each  $x$ , these define the states  $\sigma_x^{BC} = \mathcal{N}(\phi_x \otimes \Psi)$  and  $\omega_x^C = \text{Tr}_B \sigma_x$ . Assume for now that  $|A'| \geq |C|$ . Define a mapping  $f: \mathcal{X} \rightarrow \mathbb{R}^{|C|^2+1}$ , via

$$f: x \mapsto f_x \equiv (\omega_x, H(\omega_x), I_c(B \rangle C)_{\sigma_x}),$$

where we are considering  $\omega_x$  to be synonymous with its  $|C|^2 - 1$  dimensional parameterization. By linearity, this extends to a map from probability mass functions on  $\mathcal{X}$  to  $\mathbb{R}^{|C|^2+1}$ , where

$$f: p(x) \mapsto \sum_x p(x) f_x \equiv (\omega_p, H(C|X)_p, I_c(B \rangle CX)_p),$$

Our use of the subscript  $p$  should be clear from the context. The use of Caratheodory's theorem for bounding the support sizes of auxiliary random variables in information theory (see [11]) is well-known. Perhaps less familiar is the observation [51, 41] that

a better bound can often be obtained by use of a related theorem by Fenchel and Eggleston [20], which states that if  $S \subseteq \mathbb{R}^n$  is the union of at most  $n$  connected subsets, and if  $y$  is contained in the convex hull of  $S$ , then  $y$  is also contained in the convex hull of at most  $n$  points in  $S$ . As the map  $f$  is linear, it maps the simplex of distributions on  $\mathcal{X}$  into a single connected subset of  $\mathbb{R}^{|C|^2+1}$ . Thus, for any distribution  $p(x)$ , there is another distribution  $p'(x)$  which puts positive probability on at most  $|C|^2 + 1$  states, while satisfying  $f(p) = f(p')$ . If it is instead the case that  $|A'| < |C|$ , this bound can be reduced to  $|A|^2 + 1$  by replacing the first components of the map  $f$  with a parameterization of  $\phi_x^{A'}$ , as specification of a density matrix on  $A'$  is enough to completely describe the resulting state on  $C$ . It is therefore sufficient to consider  $|X| \leq \min\{|A'|, |C|\}^2 + 1$  when computing  $\mathcal{CQ}^{(1)}(\mathcal{N})$ .

■

# Bibliography

- [1] R. Ahlswede, “Multi-way communication channels,” Second Intern. Sympos. on Inf. Theory, Thakadsor, 1971, Publ. House of the Hungarian Acad. of Sciences, pp. 23–52, 1973.
- [2] R. Ahlswede, P. Löber, “Quantum data processing,” *IEEE Trans. Inform. Theory*, vol. 47, no. 1, pp. 474–478, January 2001.
- [3] C. Ahn, P. Doherty, P. Hayden, A. Winter, “On the distributed compression of quantum information,” to appear in *IEEE Trans. Inform. Theory*, quant-ph/0403042.
- [4] R. Alicki, M. Fannes, “Continuity of quantum conditional information,” *J. Phys. A*, vol. 37, pp. L55–L57, January 2004.
- [5] H. Barnum, C. M. Caves, C. A. Fuchs, R. Josza, B. Schumacher, “Noncommuting mixed states cannot be broadcast,” *Phys. Rev. Lett.*, vol. 76, no. 15, pp. 2818–2821, 1996.
- [6] H. Barnum, E. Knill, M. Nielsen, “On quantum fidelities and quantum capacities,” *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1317–1329, July 2000.
- [7] C. Bennett, D. DiVincenzo, J. Smolin, “Capacities of quantum erasure channels,” *Phys. Rev. Lett.*, vol. 78, pp. 3217–3220, April 1997.
- [8] T. Cover, A. El Gamal, M. Salehi, “Multiple-access channels with arbitrarily correlated sources,” *IEEE Trans. Inform. Theory*, vol. 26, no. 6, pp. 648–657, November 1980.



- [9] T. Cover, D. Julian, “Concavity of the Second Law of Thermodynamics,” *IEEE Intern. Symp. Inform. Theory*, Yokohama, Japan, June 2003.
- [10] T. Cover, J. A. Thomas, *Elements of Information Theory*, John–Wiley & Sons, Inc., 1991.
- [11] I. Cs szar, J. K rner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Akad miai Kiad , Budapest.
- [12] E. B. Davies, J. T. Lewis, “An operational approach to quantum probability,” *Comm. Math. Phys.*, vol. 17, no. 3, pp. 239–260, 1970.
- [13] I. Devetak, “The private classical information capacity and quantum information capacity of a quantum channel,” *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 44–55, January 2005.
- [14] I. Devetak, A. Winter, “Classical data compression with quantum side information,” *Phys. Rev. A*, vol. 68, pp. 042301-042306, October 2003.
- [15] I. Devetak, P. Shor, “The capacity of a quantum channel for simultaneous transmission of classical and quantum information,” [quant-ph/0311131](#).
- [16] I. Devetak, A. Winter, “Distilling common randomness from bipartite quantum states,” [quant-ph/0304196](#), 2003.
- [17] I. Devetak, J. Yard, “Quantum channels with side information,” in preparation.
- [18] G. Dueck, “Maximal error regions are strictly smaller than average error regions for multi-user channels,” *Problems of Control and Information Theory*, vol. 7, pp. 11–19, 1978.
- [19] W. D r, J. I. Cirac, P. Horodecki, “Non-additivity of quantum capacity for multiparty communication channels,” [quant-ph/0403068](#).
- [20] H. G. Eggleston, *Convexity*, Cambridge University Press, N.Y. 1963.

- [21] S. I. Gelfand, M. S. Pinsker, “Coding for a channel with random parameters,” *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [22] A. Harrow, I. Devetak, “A family of quantum protocols,” [quant-ph/0308044](#).
- [23] P. Hayden, R. Josza, D. Petz, A. Winter, “Structure of states which satisfy strong subadditivity of quantum entropy with equality,” *Commun. Math. Phys.*, vol. 246, no. 2, pp. 359–374, 2004.
- [24] A. S. Holevo, “Bounds for the quantity of information transmitted by a quantum channel,” *Probl. Inf. Transm.*, vol. 9, pp. 177–183, 1973.
- [25] A. S. Holevo, “The capacity of the quantum channel with general input states,” *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 269–273, January 1998.
- [26] R. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
- [27] M. Horodecki, J. Oppenheim, A. Winter, “Quantum information can be negative,” preprint available at <http://www.damtp.cam.ac.uk/user/jono/pub/merge.pdf>.
- [28] G. Klimovitch, “On the classical capacity of a quantum multiple access channel,” *IEEE Intern. Symp. Inform. Theory*, Washington D.C. June 2001, p. 278.
- [29] G. Klimovitch, A. Winter, “Classical capacity of quantum binary adder channels,” [quant-ph/0502055](#).
- [30] M. Koashi, A. Winter, “Monogamy of entanglement and other correlations,” *Phys. Rev. A* vol. 69, no. 2, pp. 022309–022314, February 2004.
- [31] Liao, “Multiple access channels”, Ph.D. dissertation, Dept. of Electrical Engineering, University of Hawaii, 1972.
- [32] E. H. Lieb, M. B. Ruskai, “Proof of the strong subadditivity of quantum-mechanical entropy,” *J. Math. Phys.*, vol. 14, pp. 1938–1941, December 1973.

- [33] S. Lloyd, “Capacity of the noisy quantum channel,” *Phys. Rev. A*, vol. 55, no. 3, pp. 1613–1622, March 1997.
- [34] M. Nielsen, I. Chuang, *Quantum Information and Quantum Computation*, Cambridge University Press, 2001.
- [35] M. Nielsen, D. Petz, “A simple proof the strong subadditivity inequality,” [quant-ph/0408130](https://arxiv.org/abs/quant-ph/0408130).
- [36] A. Peres, *Quantum theory: concepts and methods*, Kluwer Academic Publishers, 1995.
- [37] A. Peres, D. Terno, “Quantum information and relativity theory,” *Rev. Mod. Phys.*, vol. 76, pp. 93-123, January 2004.
- [38] Oyha, D. Petz, *Quantum Entropy and Its Use*, Springer, Berlin, 1993.
- [39] J. Preskill, *Lecture Notes for Physics 219, Quantum Computation and Information*, available at <http://www.theory.caltech.edu/~preskill/ph219/index.html#lecture>.
- [40] M. B. Ruskai, “Beyond strong subadditivity: improved bounds on the contraction of generalized relative entropy,” *Rev. Math. Phys.*, vol. 6, no. 5A, pp. 1147–1161, 1994.
- [41] M. Salehi, “Cardinality bounds on auxiliary random variables in multiple-user theory via the method of Ahlswede and Körner,” Technical Report No. 33, Dept. of Statistics, Stanford University, August 1978.
- [42] B. Schumacher, M. A. Nielsen, “Quantum data processing and error correction,” *Phys. Rev. A*, vol. 54, no. 4, pp. 2629–2635, October 1996.
- [43] B. Schumacher, M. D. Westmoreland, “Sending classical information via noisy quantum channels,” *Phys. Rev. A*, vol. 56, no. 1, pp. 131–138, July 1997.
- [44] C. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.

- [45] P. Shor, “The quantum channel capacity and coherent information,” lecture notes, MSRI Workshop on Quantum Computation, 2002. Available at <http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/>
- [46] P. W. Shor, J. A. Smolin, “Quantum error-correcting codes need not completely reveal the error syndrome,” `quant-ph/9604006`.
- [47] A. Uhlmann, “The ‘transition probability’ in the state space of a \*-algebra,” *Rep. Math. Phys.*, vol. 9, pp. 273-279, 1976.
- [48] A. Winter, “Coding theorem and strong converse for quantum channels,” *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2481–2485, November 1999.
- [49] A. Winter, “The capacity of the quantum multiple access channel,” *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 3059–3065, November 2001.
- [50] A. Winter, F. Verstraete, J. Smolin, J. Oppenheim, M. Horodecki, “Entanglement of assistance and applications to multiuser quantum information theory, or transmitting partial (quantum) information - which can be negative!,” talk given at QIP 2005, Cambridge, Massachusetts, January 2005.
- [51] A. Wyner, J. Ziv, “The rate-distortion function for source decoding with side-information at the decoder,” *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, January 1976.
- [52] J. Yard, “Capacity Theorems for Quantum Multiple Access Channels — Part II: Simultaneous Classical-Quantum Capacity Region”, in preparation.
- [53] J. Yard, I. Devetak, P. Hayden, “Capacity theorems for quantum multiple access channels — classical-quantum and quantum-quantum capacity regions,” `quant-ph/0501045`.
- [54] J. Yard, I. Devetak, P. Hayden, “Capacity theorems for quantum multiple access channels,” submitted to *IEEE Intern. Symp. Inform. Theory*, Adelaide, Australia, 2005.