Université de Montréal

The decoupling approach to quantum information theory

par
Frédéric Dupuis

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée à la Faculté des arts et des sciences
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en informatique

Novembre, 2009

Université de Montréal
Faculté des arts et des sciences

Cette thèse intitulée:

The decoupling approach to quantum information theory

présentée par:

Frédéric Dupuis

a été évaluée par un jury composé des personnes suivantes:

Pierre McKenzie,	président-rapporteur
Gilles Brassard,	directeur de recherche
Patrick Hayden,	codirecteur
Louis Salvail,	membre du jury
Renato Renner,	examinateur externe
François Lalonde,	représentant du doyen de la FES

Thèse acceptée le: ....

RÉSUMÉ

La théorie de l'information quantique étudie les limites fondamentales qu'imposent les lois de la physique sur les tâches de traitement de données comme la compression et la transmission de données sur un canal bruité. Cette thèse présente des techniques générales permettant de résoudre plusieurs problèmes fondamentaux de la théorie de l'information quantique dans un seul et même cadre. Le théorème central de cette thèse énonce l'existence d'un protocole permettant de transmettre des données quantiques que le receveur connaît déjà partiellement à l'aide d'une seule utilisation d'un canal quantique bruité. Ce théorème a de plus comme corollaires immédiats plusieurs théorèmes centraux de la théorie de l'information quantique.

Les chapitres suivants utilisent ce théorème pour prouver l'existence de nouveaux protocoles pour deux autres types de canaux quantiques, soit les canaux de diffusion quantiques et les canaux quantiques avec information supplémentaire fournie au transmetteur. Ces protocoles traitent aussi de la transmission de données quantiques partiellement connues du receveur à l'aide d'une seule utilisation du canal, et ont comme corollaires des versions asymptotiques avec et sans intrication auxiliaire. Les versions asymptotiques avec intrication auxiliaire peuvent, dans les deux cas, être considérées comme des versions quantiques des meilleurs théorèmes de codage connus pour les versions classiques de ces problèmes.

Le dernier chapitre traite d'un phénomène purement quantique appelé *verrouillage* : il est possible d'encoder un message classique dans un état quantique de sorte qu'en lui enlevant un sous-système de taille logarithmique par rapport à sa taille totale, on puisse s'assurer qu'aucune mesure ne puisse avoir de corrélation significative avec le message. Le message se trouve donc « verrouillé » par une clé de taille logarithmique. Cette thèse présente le premier protocole de verrouillage dont le critère de succès est que la distance trace entre la distribution jointe du message et du résultat de la mesure et le produit de leur marginales soit suffisamment petite.

Mots clés: Théorie de l'information, information quantique

ABSTRACT

Quantum information theory studies the fundamental limits that physical laws impose on information processing tasks such as data compression and data transmission on noisy channels. This thesis presents general techniques that allow one to solve many fundamental problems of quantum information theory in a unified framework. The central theorem of this thesis proves the existence of a protocol that transmits quantum data that is partially known to the receiver through a single use of an arbitrary noisy quantum channel. In addition to the intrinsic interest of this problem, this theorem has as immediate corollaries several central theorems of quantum information theory.

The following chapters use this theorem to prove the existence of new protocols for two other types of quantum channels, namely quantum broadcast channels and quantum channels with side information at the transmitter. These protocols also involve sending quantum information partially known by the receiver with a single use of the channel, and have as corollaries entanglement-assisted and unassisted asymptotic coding theorems. The entanglement-assisted asymptotic versions can, in both cases, be considered as quantum versions of the best coding theorems known for the classical versions of these problems.

The last chapter deals with a purely quantum phenomenon called *locking*. We demonstrate that it is possible to encode a classical message into a quantum state such that, by removing a subsystem of logarithmic size with respect to its total size, no measurement can have significant correlations with the message. The message is therefore "locked" by a logarithmic-size key. This thesis presents the first locking protocol for which the success criterion is that the trace distance between the joint distribution of the message and the measurement result and the product of their marginals be sufficiently small.

Keywords: Information theory, quantum information

CONTENTS

LIST OF APPENDICES

NOTATION

General

log	Logarithm base 2.
ln	Natural logarithm.
$\mathbb{R}$	Real numbers.
$\mathbb{C}$	Complex numbers.
c^*	Complex conjugate of c.
$\mathbb{E}_U[f(U)]$	Expectation value of $f(U)$ over the random variable U.

Linear Algebra and Quantum Systems

$A, B, C, \ldots$	Labels for quantum systems, or linear operators between Hilbert spaces. (Should be clear from context.)
$\mathsf{A}, \mathsf{B}, \mathsf{C}, \ldots$	Hilbert spaces associated with the systems $A, B, C, \ldots$
$\lvert A\rvert$	Dimension of A.
AB	Composite quantum system whose associated Hilbert space is $\mathsf{A} \otimes \mathsf{B}$.
A^n	Quantum system composed of n copies of A.
$\mathrm{L}(\mathsf{A}, \mathsf{B})$	The space of linear operators from A to B
$\mathrm{L}(\mathsf{A})$	$\mathrm{L}(\mathsf{A}, \mathsf{A})$
$M^{A \to B}$	Indicates that the operator M is in $\mathrm{L}(\mathsf{A}, \mathsf{B})$.
$M^\dagger$	Adjoint of M
$M_T^{A \to B}$	Transpose of M with respect to the canonical bases of A and B. This has lower priority than matrix multiplication: $AB \cdot C = (AB)C(AB)^\dagger$
$M \cdot N$	$MNM^\dagger$
$\mathrm{Herm}(\mathsf{A})$	The set of Hermitian operators from A to A
$\mathrm{Pos}(\mathsf{A})$	The subset of $\mathrm{Herm}(\mathsf{A})$ consisting of positive semidefinite matrices
$M \leqslant N$	If $M, N \in \mathrm{Herm}(\mathsf{A})$, this means that $N - M \in \mathrm{Pos}(\mathsf{A})$.

Linear Algebra and Quantum Systems, continued

$\mathrm{D}(\mathsf{A})$	The set of all density operators on A; i.e. $\mathrm{D}(\mathsf{A}) = \{\rho : \rho \in \mathrm{Pos}(\mathsf{A}), \mathrm{Tr}[\rho] = 1\}$
$\mathcal{N}^{A\to B}, \mathcal{T}^{A\to B}, \dots$	Superoperators (completely positive linear maps from $\mathrm{L}(\mathsf{A})$ to $\mathrm{L}(\mathsf{B})$)
$\mathbb{I}^A$	Identity operator on A or identity superoperator on $\mathrm{L}(\mathsf{A})$. (Should be clear from context.)
$\lvert\psi\rangle^A, \lvert\varphi\rangle^A, \dots$	Vectors in A.
$\psi^A, \varphi^A, \dots$	The "unketted" versions denote their associated density matrices: $\psi^A = \lvert\psi\rangle\langle\psi\rvert$. Furthermore, if we have defined a state ψ^{AB}, then $\psi^A = \mathrm{Tr}_B[\psi^{AB}]$.
$\mathrm{op}_{A\to B}(\lvert\psi\rangle^{AB})$	Turns a vector into an operator. See Section 2.6.
$\mathrm{vec}(M^{A\to B})$	Turns an operator into a vector. See Section 2.6.
$\sqrt{M}$	If $M \in \mathrm{Pos}(\mathsf{A})$ has spectral decomposition $M = \sum_i \lambda_i \lvert\psi_i\rangle\langle\psi_i\rvert$, then $\sqrt{M} = \sum_i \sqrt{\lambda_i}\lvert\psi_i\rangle\langle\psi_i\rvert$.
$\lvert\Phi\rangle^{AA'}$	$\frac{1}{\sqrt{\lvert A\rvert}} \sum_{i=1}^{\lvert A\rvert} \lvert i\rangle^A \lvert i\rangle^{A'}$, where $\lvert i\rangle^A$ and $\lvert i\rangle^{A'}$ are fixed canonical bases for A and A', and $\mathsf{A} \cong \mathsf{A}'$.
π^A	The maximally mixed state $\frac{\mathbb{I}^A}{\lvert A\rvert}$

Norms and Distance measures

$\left\lVert M^{A\to B}\right\rVert_1$	$\mathrm{Tr}\sqrt{M^\dagger M}$
$\lVert\lvert\psi\rangle\rVert_2$	$\sqrt{\lvert\langle\psi\vert\psi\rangle\rvert}$
$\left\lVert M^{A\to B}\right\rVert_2$	$\sqrt{\mathrm{Tr}[M^\dagger M]}$
$\left\lVert M^{A\to B}\right\rVert_\infty$	Largest singular value of M.
$\left\lVert \mathcal{N}^{A\to B}\right\rVert_\diamond$	Diamond norm; see Section 2.3.1.
$F(\rho^A, \sigma^A)$	$\left\lVert\sqrt{\rho}\sqrt{\sigma}\right\rVert_1$. This is called the *fidelity*.
$d_F(\rho^A, \sigma^A)$	$\sqrt{1 - F(\rho,\sigma)^2}$. This is called the *fidelity distance*.

Entropies

$H(A\vert B)_\rho$	Conditional von Neumann entropy of A given B on ρ^{AB}, see Definition 2.4.
$H_2(A\vert B)_\rho$	Conditional 2-entropy of A given B, defined as $-\log\min_{\sigma^B\in\mathrm{D(B)}}\mathrm{Tr}\left[\left((\sigma^B\otimes\mathbb{I}^A)^{-1/2}\rho^{AB}\right)^2\right]$
$H_2^\varepsilon(A\vert B)_\rho$	Smooth 2-entropy of A given B, defined as $\max_{\sigma^{AB},d_F(\rho,\sigma)\leqslant\varepsilon} H_2(A\vert B)_\sigma$
$H_{\min}(A\vert B)_\rho$	Conditional min-entropy, see Definition 2.10.
$H_{\max}(A\vert B)_\rho$	Conditional max-entropy, see Definition 2.12.
$H_{\min}^\varepsilon(A\vert B)_\rho$	ε-smooth conditional min-entropy, see Definition 2.13.
$H_{\max}^\varepsilon(A\vert B)_\rho$	ε-smooth conditional max-entropy, see Definition 2.15.
$I(A\rangle B)_\rho$	Coherent information, see Definition 2.8.
$I(A;B)_\rho$	Mutual information, see Definition 2.6.
$I(A;B\vert C)_\rho$	Conditional mutual information, see Definition 2.7.

First names

Alice	The sender in all the protocols.
Bob	The receiver in all the protocols.

REMERCIEMENTS

Je tiens d'abord à remercier mes deux directeurs, Gilles Brassard et Patrick Hayden de qui j'ai tant appris, pour leur soutien constant et sans qui cette thèse n'aurait jamais vu le jour.

Je dois aussi beaucoup à mes coauteurs, avec lesquels j'ai eu énormément de plaisir à collaborer : Simon-Pierre Desrosiers, Patrick Hayden, Ke Li, Debbie Leung, André Méthot, Nicolas Gisin, Avinatan Hassidim et Haran Pilpel.

Je tiens aussi à remercier Andreas Winter, Renato Renner et Aram Harrow pour leur hospitalité ainsi que pour les discussions enrichissantes que j'ai eues avec eux durant les quelques semaines où une grosse partie de la rédaction de cette thèse a été accomplie.

Je remercie aussi Mario Berta, qui m'a envoyé une copie de sa thèse de « Diplom », d'où j'ai appris un des trucs utilisés dans la preuve du théorème principal.

Je voudrais aussi remercier les gens qui ont été (ou qui sont toujours) au LITQ à l'Université de Montréal ou au LaCIQ à McGill durant mon doctorat et qui ont contribué à rendre mon environnement de travail si agréable : Ashton Anderson, Somshubhro Bandyopadhay, Laurent Beaudou, Guido Berlín, Hugue Blier, Anne Broadbent, Félix Bussières, Michaël Cadilhac, Claude Crépeau, Omar Fawzi, Jan Florjanczyk, Sébastien Gambs, Claude Gravel, Charles Hélou, Kassem Kalach, Olivier Landon-Cardinal, Adrien Lemaître, Abubakr Muhammad, Roger Müller, Nicolas Dutil, Éric Paquette, Paul Chouha, David Pouliot, Louis Salvail, Ivan Savov, Jean-Raymond Simard, Alain Tapp et Jürg Wullschleger.

Puisqu'il faut d'abord se loger et se nourrir pour être en mesure de déchiffrer les mystères de l'univers, je tiens à remercier le CRSNG ainsi que l'ICRA pour leur soutien financier tout au long de mon doctorat.

Finalement, je tiens tout particulièrement à remercier ma famille, qui a de tout temps cru en moi et dont le soutien n'a jamais faibli depuis le début de ma vie.

CHAPTER 1

INTRODUCTION

The origins of information theory go back to 1948, when Claude Shannon published “A mathematical theory of communication” [Sha48], in which he proposed a mathematical framework to study information processing tasks such as data compression and data transmission over noisy channels. Data compression is the following task: we have a large amount of digital data, and we would like to shrink it down to a smaller size for efficient storage or transmission. If the data is sufficiently redundant, then it is possible to do this with a very small probability of decompressing it incorrectly. Data transmission over noisy channels involves the following problem: one has a communication channel in which the transmitter can select an input and the receiver receives an output that has been corrupted by noise in the channel. A concrete example of this would be the phone line between a house and the telephone central, or the radio link between a cellphone tower and the handsets. One would then like to use this channel to send a message and make sure that, with high probability, the receiver will be able to reconstruct it exactly.

Since our universe is governed by the laws of quantum mechanics, the physical limits imposed on these problems are themselves quantum mechanical. It also turns out that information can behave in counterintuitive ways under the laws of quantum mechanics: for example, one can know precisely the state of a two-particle quantum system while remaining ignorant of the state of either of the two particles separately. Furthermore, measurements made on two particles that are kept very far apart can exhibit correlations that could not be explained classically without assuming that information was transmitted faster than the speed of light. This is why a quantum version of information theory is so interesting: it is our attempt at taming these apparent paradoxes and counterintuitive facts. In this thesis, we will be concerned specifically with coding for various different types of quantum channels. In the last chapter, we will also look at the phenomenon of *information locking*, in which a small key can "unlock" an amount of information far beyond what would be possible classically.

1.1 Decoupling

One of the most bizarre features of quantum information theory turns out to be extremely useful for solving channel coding problems. It is the notion of *purification*: given any quantum system A whose state is random, one can find a bigger system AB such that the state on A is the same as before, but where the global state on AB is completely deterministic. This is impossible classically: if the state of a system is random, considering it together with another system only adds the potential of having more randomness globally. This, however, will help us tremendously. In a channel coding problem, we want to ensure that the output of the channel is strongly correlated (or "coupled") with the input. When we look at the purification of the final state that we want between the input and output, however, it turns out that this is equivalent to requiring that the input to the channel be completely decorrelated with the entire universe minus the channel output. This helps us because we can achieve it by *destroying* correlations—and, as in other areas of life, destruction is easier to achieve than construction.

This "decoupling" approach—we use the term "decouple" to mean "decorrelate"—has therefore become a staple of quantum information theory. It was already used to some extent in [Dev05], the first general coding theorem for the quantum capacity of quantum channels, and was used more systematically in [HOW07] and [ADHW06], which derived basic quantum protocols from which a large number of other, previously known protocols could be derived. In [HHYW08], the results of [Dev05] were revisited using a "purer" decoupling approach.

While we have some sense that these last three papers use the same "trick", they are nonetheless proven separately, and while they can be used to derive other protocols, one sometimes needs to work quite a bit to accomodate the particular forms of the theorems (using, for instance, typical projectors to limit the dimensions of various quantum systems). One of the main contributions of this thesis is to give a general decoupling theorem, from which all of the known ones can be derived very easily, and which is much more flexible. We then go on to give quantum coding theorems for different varieties of quantum channels, including quantum broadcast channels and quantum channels with side information at the transmitter. In both cases, no prior results exist regarding the particular tasks considered. Finally, we also use the main decoupling theorem to prove a result on information locking.

1.2 Contributions

This thesis is broken down into the following chapters:

Chapter 2 (Preliminaries): This chapter contains the concepts and definitions necessary to understand the rest of the thesis. It does not contain original material.

Chapter 3 (The decoupling theorem): This chapter is devoted to the main decoupling theorem. We state it and prove it along with several variants, including a new one-shot coding theorem for quantum channels, in which Bob potentially knows part of the state before the start of the protocol. We then use it to rederive the main results of [HOW07], [ADHW06], [GPW05] and [HHYW08] in a more straightforward manner. The contents of this chapter will be published as a paper at a later date.

Chapter 4 (Quantum channels with side information at the transmitter): This chapter derives new results on quantum channels with side information at the transmitter. A channel with side information at the transmitter is a channel in which the transmitter has access ahead of time to information about the noise in the channel, but where the receiver does not have access to this information.

We give a one-shot coding theorem for them similar to the one for regular channels in Chapter 3, and show that applying it to entanglement-assisted coding for memoryless channels yields an optimal protocol. In particular, we show that the entanglement-assisted capacity of these channels admits a single-letter formula that parallels the solution to the classical version of this problem given in [GP80]. Part of the work in this section was presented in a different form at the 2009 International Symposium on Information Theory [Dup09].

Chapter 5 (Quantum broadcast channels): This chapter contains a coding theorem for quantum broadcast channels, namely channels with one input but two outputs going to two physically separated receivers. Again, we give a general one-shot coding theorem, and we then derive from it an entanglement-assisted coding scheme for memoryless channels that parallels the best known classical coding theorem for broadcast channels given in [Mar79]. These are the first coding theorems given for these tasks. A different version of this work was accepted for publication in IEEE Transactions on Information Theory and is joint work with Patrick Hayden and Ke Li [DHL09].

Chapter 6 (Locking classical information in quantum states): This chapter deals with the purely quantum phenomenon of information locking. We show that there exists a unitary such that if we encode a classical message into a quantum state, apply this unitary to it, and remove a very small part (logarithmic in the total size), then one can get almost no information about the message by measuring the remaining part. This is done by showing that the statistical distance between the joint distribution of the message and the measurement result and a product distribution can be made very small. This is slightly stronger than what was done in prior information locking results, in which upper bounds on the mutual information between the measurement result and the message were derived. Furthermore, this is the first locking protocol in which one uses a single unitary and a quantum key instead of applying one of several unitaries and using the choice of unitary as the key. We also show that this scheme can be used to construct a quantum key distribution protocol that guarantees that the eavesdropper can gain almost no information about the key by making a measurement immediately after the execution of the protocol, but where the eavesdropper only needs to learn a very small portion of the key to be able to recover the rest. This underscores much more spectacularly than before [KRBM07] the need to take into account the fact that an eavesdropper might keep *quantum* information after the protocol and use it only when making his actual attack. This will be published at a later date and is joint work with Patrick Hayden and Debbie Leung.

Chapter 7 (Conclusion): This chapter concludes the thesis with a recapitulation of what was done, and speculates on what the future might hold.

CHAPTER 2

PRELIMINARIES

This chapter explains the notation used throughout the thesis and presents some concepts one needs to understand this document.

2.1 Notation

Linear algebra is the language of quantum mechanics; we therefore start by introducing the notation we will use for linear algebraic concepts. One can find explanations of all the concepts below in any linear algebra textbook, or, to be introduced to these concepts in the setting of quantum information, in [Wat08]. Note that a condensed version of this appears on pages ix–xi so the reader can refer back to it more easily. We will denote by sans-serif capital letters (such as $\mathsf{A}, \mathsf{B}, \ldots$) complex finite-dimensional inner product vector spaces (which we will usually simply call Hilbert spaces following the usual quantum information convention—these are the only Hilbert spaces that we will ever consider in this thesis), and we will use regular capital letters $A, B, \ldots$ to label the quantum systems associated with the spaces $\mathsf{A}, \mathsf{B}, \ldots$. We will denote the dimension of A by $|A|$. Vectors in A are denoted by "kets" $|\psi\rangle^A$ with the superscript omitted when it causes no confusion. Furthermore, we will denote by $\mathrm{L}(\mathsf{A}, \mathsf{B})$ the space of linear operators from A to B, and we will use the shorthand $\mathrm{L}(\mathsf{A})$ for $\mathrm{L}(\mathsf{A}, \mathsf{A})$. Elements of the dual space $\mathrm{L}(\mathsf{A}, \mathbb{C})$ of A are written as "bras"; for instance the dual of $|\psi\rangle$ is written $\langle\psi|$. We use † to designate the adjoint of an operator, $\mathrm{Herm}(\mathsf{A})$ is the set of all Hermitian (self-adjoint) operators on A, and $\mathrm{Pos}(\mathsf{A}) \subseteq \mathrm{Herm}(\mathsf{A})$ is the set of all positive semidefinite operators on A. Given two operators $M, N \in \mathrm{Herm}(\mathsf{A})$, we say that $M \leqslant N$ if $N - M \in \mathrm{Pos}(\mathsf{A})$. Given an operator $M \in \mathrm{L}(\mathsf{A}, \mathsf{B})$, we will use the superscript $M^{A \to B}$ to indicate its input and output spaces. We will use the symbol $\cdot$ to denote conjugation: given two operators $M^{A \to B}$ and N^A, we define $M \cdot N = MNM^\dagger$.

We will denote by the calligraphic letters $\mathcal{N}^{A\to B}, \mathcal{T}^{A\to B}, \mathcal{S}^{A\to B}, \ldots$ completely positive linear maps from $\mathrm{L}(\mathsf{A})$ to $\mathrm{L}(\mathsf{B})$; we will call these "superoperators". We will also write $\mathbb{I}^A$ for either the identity operator on A, or the identity superoperator on $\mathrm{L}(\mathsf{A})$; which one is meant should be clear from the context.

In superscripts, we will simply concatenate letters to indicate the tensor product: for instance, $M^{AB\to CD} \in \mathrm{L}(\mathsf{A}\otimes\mathsf{B}, \mathsf{C}\otimes\mathsf{D})$. When applying an operator $M^{A\to B}$ to a vector $|\psi\rangle^{AC}$, we will usually omit the implicit identity: for instance, $M^{A\to B}|\psi\rangle^{AC} = (M^{A\to B}\otimes\mathbb{I}^C)|\psi\rangle^{AC}$.

Given an operator $M^{AB} \in \mathrm{L}(\mathsf{A}\otimes\mathsf{B})$ on a composite Hilbert space, we can define its *partial trace on B*, denoted either as $\mathrm{Tr}_B[M^{AB}]$ or simply by M^A, omitting the B in the superscript, as the unique operator N in $\mathrm{L}(\mathsf{A})$ such that $\mathrm{Tr}[ZN] = \mathrm{Tr}[(Z\otimes\mathbb{I}^B)M^{AB}]$ for all $Z \in \mathrm{L}(\mathsf{A})$. In other words, the partial trace is defined as the adjoint of the superoperator $\mathcal{F}^{A\to AB}, \mathcal{F}(N^A) = N^A\otimes\mathbb{I}^B$ under the Hilbert-Schmidt inner product $\langle X, Y\rangle := \mathrm{Tr}[X^\dagger Y]$.

We will also need the concept of *partial isometries*. A partial isometry is an operator $V^{A\to B}$ whose singular values are all either 1 or 0. Equivalently, they can be defined as any operator $V^{A\to B}$ such that $V^\dagger V$ and $VV^\dagger$ are projectors. A *full-rank partial isometry* is a partial isometry $V^{A\to B}$ whose rank is $\min\{|A|, |B|\}$.

2.2 Quantum mechanics: an extremely short introduction

Since one cannot hope to cover basic quantum mechanics in a few paragraphs, the author strongly recommends the interested reader to consult [NC00] or [Wat08] for a more complete introduction. Nonetheless, a short introduction to the basic concepts using the notation that we will use is given here for the sake of completeness.

A quantum system A is represented by a Hilbert space A; a state of the system is a positive semidefinite operator $\rho^A \in \mathrm{Pos}(\mathsf{A})$ such that $\mathrm{Tr}[\rho^A] = 1$. We also call these states *density operators* or *density matrices* and denote the set of all density operators on A as $\mathrm{D}(\mathsf{A})$. A state ρ is considered *pure* if $\operatorname{rank}\rho = 1$, in which case there exists a $|\psi\rangle \in \mathsf{A}$ of norm 1 and such that $\rho = |\psi\rangle\langle\psi|$. We sometimes also call general states *mixed states* when we want to emphasize the fact that the state is not necessarily pure. Let A and B be Hilbert spaces corresponding to quantum systems A and B; we can then consider them as a single composite system AB with $\mathsf{A} \otimes \mathsf{B}$ as its associated Hilbert space. By convention in this document, we will write systems on which a quantum state is defined as a superscript; for instance, $\rho^{AB} \in \mathrm{D}(\mathsf{A}\otimes\mathsf{B})$. The same convention will apply to all operators. If the input and output spaces of an operator are different, we will write an arrow in the superscript to indicate this: for example, $M^{A\to B} \in \mathrm{L}(\mathsf{A},\mathsf{B})$. When we want to consider only part of a composite system, we take its partial trace on the system we want to eliminate.

The operations that can be applied to a quantum system without making it interact with other systems correspond to the unitary operators on the associated Hilbert space, namely all transformations of the form $\rho \to U\rho U^\dagger$, where U is unitary. Since conjugation will be used so often in this thesis, we will use the notation $A \cdot B$ to denote $ABA^\dagger$. Transformations involving interactions with other systems can be simulated by adding an ancillary system, applying a unitary on the composite system, and then tracing out part of the remaining system.

Such a transformation can also be represented by a trace-preserving superoperator (sometimes called CPTP map, which stands for “completely positive trace-preserving” map). It can be shown that a linear map $\mathcal{N}^{A\to B}$ is completely positive if and only if it can be written as $\mathcal{N}(\rho) = \sum_i N_i \rho N_i^\dagger$, where $N_i \in \mathrm{L}(\mathsf{A},\mathsf{B})$; furthermore, any such linear map is trace-preserving (i.e. $\mathrm{Tr}[\mathcal{N}(M)] = \mathrm{Tr}[M] \;\; \forall M$) if $\sum_i N_i^\dagger N_i = \mathbb{I}^A$. We will sometimes call trace-preserving superoperators “quantum channels” when we want to emphasize that this is a transformation over which we have no control and wish to view as a noisy channel.

There is also a class of operations that leaves the quantum system intact but changes its underlying Hilbert space. For instance, suppose we have a state ρ^A

and want to embed the information it contains into the system B. An operation that does this is a partial isometry $V^{A\to B}$ such that $\mathrm{Tr}[V\rho V^\dagger] = 1$ (i.e. the image of $V^\dagger$ must contain the support of ρ). We will sometimes call these simply "isometries" since they act as isometries on the part of A in which ρ lies. Such an operation can be implemented by the superoperator $\mathcal{V}^{A\to B}(\rho^A) = V\rho V^\dagger + \sum_i N_i \rho N_i^\dagger$, where the N_i are such that $V^\dagger V + \sum_i N_i^\dagger N_i = \mathbb{I}^A$ and $\mathrm{Tr}[N_i \rho N_i^\dagger] = 0$ for all i.

Quantum systems can also be measured, yielding a classical output. In addition to the measurement result, a measurement can also have a quantum residue, in case the measurement does not completely measure the state. To represent this, we will use a special type of trace-preserving superoperator that we will call a *measurement superoperator*. A measurement superoperator is a superoperator of the form $\mathcal{M}^{A\to BX}(\sigma^A) = \sum_x |x\rangle\langle x|^X \otimes N_x \sigma^A {N_x}^\dagger$, where the $|x\rangle$ are all part of the same orthonormal basis for X, and the $N_x^{A\to B}$ are arbitrary operators such that $\mathcal{M}$ is CPTP. The interpretation for this is that we get the measurement result x with probability $\mathrm{Tr}[N_x \sigma N_x^\dagger]$, X is a classical register that holds the measurement result, and, if the measurement result was x, the B register gets the state $N_x \sigma N_x^\dagger / \mathrm{Tr}[N_x \sigma N_x^\dagger]$. If we are not interested in the quantum residue and only care about the classical result, we only need to describe the set of positive semidefinite operators $\{N_x^\dagger N_x\}$ to describe the measurement, which is then called a *positive operator valued measure*, or POVM. We call a measurement superoperator *complete* if all of the N_x are of rank 1, in which case the quantum residue is superfluous since it can be reconstructed from the classical result only.

One particularly strange and interesting feature of quantum mechanics is the concept of entanglement. We say that a bipartite state ρ^{AB} is *entangled* if it cannot be written in the form $\rho^{AB} = \sum_i \alpha_i \sigma_i^A \otimes \omega_i^B$. In other words, a state on AB in entangled if it cannot be expressed as a probabilistic mixture of separate states on A and B. An example of such a state is the following pure maximally entangled state that will be of great importance throughout the thesis: $\Phi^{AB} = |\Phi\rangle\langle\Phi|^{AB}$ with $|\Phi\rangle^{AB} = \frac{1}{\sqrt{|A|}} \sum_i |i\rangle^A \otimes |i\rangle^B$, where $|A| = |B|$ and the $|i\rangle^A$ and $|i\rangle^B$ are standard orthonormal bases for A and B. When $|A| = |B| = 2$, we will call this state an *EPR pair* [EPR35], after Einstein, Podolski and Rosen who first noticed the phenomenon of entanglement and defined this state. With some abuse of terminology, we will call higher-dimensional instances of this state "EPR pairs" even when the dimension is not a power of two.

Of central importance to this thesis is the concept of *purification*. Given a mixed state ρ^A, it is always possible to find a pure state ω^{AB} on a larger system such that $\rho^A = \mathrm{Tr}_B[\omega^{AB}]$. Note that this is also a purely quantum phenomenon: if one has a probability distribution p over a set $\mathfrak{X}$, it is impossible to find a single element (x, y) of $\mathfrak{X} \times \mathfrak{Y}$ which then somehow ends up being distributed as p when we stop looking at the y part of it!

An analogous fact holds for quantum channels: given a completely positive superoperator $\mathcal{N}^{A \to B}$, it is possible to find a partial isometry $U_{\mathcal{N}}^{A \to BE}$ such that $\mathcal{N}(X) = \mathrm{Tr}_E[U_{\mathcal{N}} \cdot X]$ for every $X \in \mathrm{L}(\mathsf{A})$. In other words, one can find a deterministic operation that takes the input A to two output systems: the actual output of the channel B, and an environment system E. When we ignore the environment system, we get exactly the same channel. We call such a partial isometry a *Stinespring dilation* of $\mathcal{N}$.

2.3 Distance measures

We will often need a notion of distance between quantum states, usually to state that the result of a particular protocol that we developed is "close" to some ideal output state that we would like to get. The distance we will use most of the time is called the *trace distance*; the trace distance between two states ρ and σ is $\|\rho - \sigma\|_1$, where $\|M\|_1 := \mathrm{Tr}\sqrt{M^\dagger M}$ for any $M^{A \to B}$. In other words, it is equal to the sum of the absolute values of the eigenvalues of the matrix $\rho - \sigma$. The reason for which this is a meaningful measure of distance is that it characterizes how easy it is for someone to determine through a measurement whether an unknown state is ρ or σ, as was discovered by Helstrom:

Theorem 2.1 (Helstrom's theorem [Hel69]). *Let ρ^A and σ^A be two density operators* *on A, and suppose one holds ρ^A with probability $\frac{1}{2}$ and σ^A with probability $\frac{1}{2}$, and one tries to determine which one it is by performing a measurement on A. Then, the best possible measurement will give the correct answer with probability $\frac{1}{2} + \frac{1}{4}\|\rho - \sigma\|_1$.*

Proof. Let $\{M_\rho^A, M_\sigma^A\}$ be a POVM used to guess which state we have (since there are only two possible answers, one only needs two POVM operators). Then, the probability of guessing correctly is

$$\begin{aligned}
\frac{1}{2}\operatorname{Tr}[M_\rho \rho] + \frac{1}{2}\operatorname{Tr}[M_\sigma \sigma] &= \frac{1}{2}\operatorname{Tr}[M_\rho \rho + (\mathbb{I} - M_\rho)\sigma] \\
&= \frac{1}{2}\operatorname{Tr}[M_\rho \rho + \sigma - M_\rho \sigma] \\
&= \frac{1}{2} + \frac{1}{2}\operatorname{Tr}[M_\rho(\rho - \sigma)] \\
&\leqslant \frac{1}{2} + \frac{1}{2}\operatorname{Tr}[M_\rho P_+(\rho - \sigma)P_+] \\
&\leqslant \frac{1}{2} + \frac{1}{2}\operatorname{Tr}[P_+(\rho - \sigma)P_+] \\
&= \frac{1}{2} + \frac{1}{4}\|\rho - \sigma\|_1
\end{aligned}$$

where P_+ is a projector onto the eigenspaces of $\rho - \sigma$ corresponding to positive eigenvalues. The first inequality is due to the operator inequality $P_+(\rho - \sigma)P_+ \geqslant \rho - \sigma$, and the second inequality, to the operator inequality $M_\rho \leqslant \mathbb{I}$. The last equality is due to the fact that, since $\rho - \sigma$ has zero trace, $P_+(\rho - \sigma)P_+$ must correspond to exactly half of the trace distance. Of course, equality can be attained if $M_\rho = P_+$. □

This means that, if the trace distance between two states is very small, someone trying to determine which of the states an unknown state is in will be scarcely better off by doing the optimal measurement than by guessing randomly. In particular, if the output of a quantum protocol is ε-close in trace distance to the output of an ideal protocol, then, regardless of what we use the protocol for, we will almost never be able to tell the difference.

A related notion is the *fidelity* between quantum states:

Definition 2.1 (Fidelity). *Given two states ρ and σ, their fidelity is defined as $F(\rho, \sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1$.*

One can easily see that the fidelity approaches one when two states get closer together; in fact, $F(\rho, \rho) = \|\rho\|_1 = 1$. An important property of the fidelity is that it is stable under purifications: given two states ρ^A and σ^A, and a purification ρ^{AB} of ρ^A, then $F(\rho^A, \sigma^A) = \max_{\sigma^{AB}} F(\rho^{AB}, \sigma^{AB})$, where we maximize over all purifications of σ^A. This is due to Uhlmann's theorem [Uhl76] and will be proven in the next chapter as Theorem 3.1. One can also define a distance measure based on the fidelity:

Definition 2.2 (Fidelity distance). *Let ρ and σ be two density operators. Then, their* fidelity distance *is defined as*

$$d_F(\rho, \sigma) := \sqrt{1 - F(\rho, \sigma)^2}.$$

The fidelity distance is essentially equivalent to the trace distance, as shown by the Fuchs-van de Graaf inequalities [FvdG99]:

Lemma 2.2 (Fuchs-van de Graaf inequalities). *Let $\rho \in \mathsf{D}(\mathsf{A})$ and $\sigma \in \mathsf{D}(\mathsf{A})$ be density operators on A. Then,*

$$1 - \frac{1}{2}\|\rho - \sigma\|_1 \leqslant F(\rho, \sigma) \leqslant \sqrt{1 - \frac{1}{4}\|\rho - \sigma\|_1^2}.$$

This implies that

$$\frac{1}{2}\|\rho - \sigma\|_1 \leqslant d_F(\rho, \sigma) \leqslant \sqrt{\|\rho - \sigma\|_1}.$$

2.3.1 The diamond norm

It will also be convenient on a few occasions to be able to compare two superoperators. To do this, we introduce the so-called *diamond norm*:

Definition 2.3 (Diamond norm). *Let $\mathcal{N} : \mathsf{L}(\mathsf{A}) \to \mathsf{L}(\mathsf{B})$ be any linear operator from*

$\mathrm{L}(\mathsf{A})$ *to* $\mathrm{L}(\mathsf{B})$. *Then, we define its diamond norm to be*

$$\|\mathcal{N}\|_\diamond := \max_{\sigma^{AA'} \in \mathrm{D}(\mathsf{A} \otimes \mathsf{A}')} \left\| (\mathcal{N}^{A \to B} \otimes \mathbb{I}^{A' \to A'})(\sigma^{AA'}) \right\|_1$$

where the maximization is taken over all mixed states $\sigma^{AA'}$*, and where* $\mathsf{A}' \cong \mathsf{A}$.

This norm is usually called the *completely bounded trace norm* in operator theory and has been an object of study in that field for many years (see, for example, [Pau02] for an introduction to the area), but it was introduced to quantum information theory by Kitaev [Kit97] as the "diamond norm".

The main reason for using the diamond norm to define a notion of distance on quantum channels is essentially the same as for using the trace norm on quantum states: it characterizes the optimal probability of successfully distinguishing two channels. Just as Theorem 2.1 shows that the optimal probability of distinguishing the quantum states ρ and σ is $\frac{1}{2} + \frac{1}{4}\|\rho - \sigma\|_1$, it is possible to show that the optimal probability of distinguishing the quantum channels $\mathcal{N}$ and $\mathcal{M}$ is given by $\frac{1}{2} + \frac{1}{4}\|\mathcal{N} - \mathcal{M}\|_\diamond$.

2.4 Information measures

2.4.1 von Neumann entropy and derived quantities

To be able to give solutions to information theory problems, we must have ways of measuring amounts of information. The fundamental quantity is the *von Neumann entropy* of a quantum state:

Definition 2.4 (von Neumann entropy [vN32]). *The von Neumann entropy of a quantum state* ρ^A *is defined as* $H(A)_\rho := -\operatorname{Tr}[\rho^A \log \rho^A]$ *(where, if* $\rho^A = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|^A$ *is a spectral decomposition of* ρ^A, $\log \rho^A = \sum_i \log(\lambda_i)|\psi_i\rangle\langle\psi_i|^A$*, and we interpret* $0 \log 0$ *as 0).*

The von Neumann entropy measures the amount of information present in sequences of many copies of the same state, i.e. in $\rho^{\otimes n}$. More specifically, it has been shown by Schumacher that an i.i.d. state $\rho^{A^{\otimes n}}$ can be compressed into $n[H(A)_\rho - \delta]$ qubits with an error rate going to zero as $n \to \infty$ for any $\delta > 0$ [Sch95]. Hence, the higher the entropy, the less certain we are about the state and the more space we need to store it.

Many other information measures are derived from the von Neumann entropy. The first one is the conditional von Neumann entropy:

Definition 2.5 (Conditional von Neumann entropy)**.** *Given a state ρ^{AB}, the conditional von Neumann entropy of A given B is defined as*

$$H(A|B)_\rho := H(AB)_\rho - H(B)_\rho.$$

This is meant to describe the amount of uncertainty that we have about A if we already possess B. This interpretation was problematic for a long time, however, given that it can be negative (for instance, $H(A|A')_{\Phi^{AA'}} = -1$, where $|A| = |A'| = 2$ and $\Phi^{AA'} = \frac{1}{2}\sum_{i,j=1}^{2} |ii\rangle\langle jj|$). However, it turns out to give the solution to the following problem: given a state $\rho^{AB^{\otimes n}}$ between Alice and Bob, how many EPR pairs are required between Alice and Bob to teleport Alice's n shares to Bob (with free classical communication)? This task, called *state merging* [HOW07], is possible if we have $n[H(A|B)_\rho + \delta]$ EPR pairs, and the error goes to zero as $n \to \infty$ for every $\delta > 0$. When $H(A|B)_\rho$ is negative, we can teleport while *generating* EPR pairs at this rate.

Another information measure derived from the von Neumann entropy is the quantum mutual information:

Definition 2.6 (Quantum mutual information)**.** *Let ρ^{AB} be a quantum state. Then, the mutual information between A and B is defined as*

$$\begin{aligned} I(A;B)_\rho &:= H(A)_\rho + H(B)_\rho - H(AB)_\rho \\ &= H(A)_\rho - H(A|B)_\rho \\ &= H(B)_\rho - H(B|A)_\rho. \end{aligned}$$

Without going into details, this quantity gives the entanglement-assisted classical capacity of memoryless quantum channels (channels of the form $\mathcal{N}^{\otimes n}$) [BSST02]. It is always nonnegative.

We can also define a conditional version of quantum mutual information:

Definition 2.7 (Conditional quantum mutual information). *Let ρ^{ABC} be a quantum state. Then, the mutual information between A and B given C is defined as*

$$I(A;B|C)_\rho := I(A;BC)_\rho - I(A;C)_\rho.$$

This is also never negative [LR73], a fact that turns out to be highly nontrivial to prove. While we will only use this quantity in a technical proof in Chapter 4, it nonetheless has a natural interpretation through the task of *state redistribution* [DY06].

The coherent information is another measure that is important for channel coding problems. Unlike the previous one, this one has no classical analogue:

Definition 2.8 (Coherent information). *Let ρ^{AB} be a quantum state. Then, the coherent information from A to B is defined as*

$$I(A\rangle B)_\rho := -H(A|B)_\rho.$$

This is only positive when conditional entropy is negative, which only happens when a state is entangled. This quantity gives the best known general rate for unassisted transmission of quantum data through i.i.d. quantum channels.

2.4.2 Properties of the von Neumann entropy

The family of entropic quantities defined above have a number of useful properties. In all of the statements below, let ψ^{ABC} be any pure state with respect to which all entropic quantities are computed:

- $H(A) = H(BC)$

- $H(AB) = H(A) + H(B|A)$
- $H(A) = \frac{1}{2}I(A;B) + \frac{1}{2}I(A;C)$
- $I(A\rangle B) = \frac{1}{2}I(A;B) - \frac{1}{2}I(A;C)$
- $H(A|B) = -H(A|C)$

All of the above can be easily proven from the definitions. One can also show that, on a *mixed* state ρ^{ABC}, the following holds:

- $I(A;BC) \geqslant I(A;B)$.

In other words, the mutual information is monotonic under the addition of more subsystems; by taking into consideration an additional system C in addition to B, one cannot lose information about A. This comes from the strong subadditivity of the von Neumann entropy [LR73] and its proof is rather involved compared to the previously stated properties.

2.4.3 One-shot information measures

All of the above quantities were relevant for tasks involving n copies of a state, or n uses of a quantum channel, and where we then take the limit as $n \to \infty$. However, in this thesis, we will generally start from protocols involving a single use of an arbitrary channel on a given arbitrary state, and then derive this special case by considering a "single use" of the channel $\mathcal{N}^{\otimes n}$. We will therefore need information measures that are relevant for the one-shot case and that reduce to the above quantities in the case of multiple copies.

The first one is the min-entropy of a quantum state:

Definition 2.9 (Quantum min-entropy)**.** *Let ρ^A be a quantum state. Then, its min-entropy is defined as*

$$H_{\min}(A)_\rho := -\log \min_{\lambda \in \mathbb{R}} \{\lambda : \rho^A \leqslant \lambda \mathbb{I}^A\}.$$

In other words, the min-entropy is the negative logarithm of the largest eigenvalue. Classically, this definition goes back to Chor and Goldreich [CG88], and was generalized to quantum information by Renner [Ren05].

Renner also defined a conditional version of the quantum min-entropy:

Definition 2.10 (Quantum conditional min-entropy)**.** *Let* $\rho^{AB} \in \mathrm{Pos}(\mathsf{A}\otimes\mathsf{B})$*. Then, the conditional min-entropy of* A *given* B *is defined as*

$$H_{\min}(A|B)_\rho := -\log\min\left\{\mathrm{Tr}[\sigma^B] : \sigma^B \in \mathrm{Pos}(\mathsf{B}), \rho^{AB} \leqslant \mathbb{I}^A \otimes \sigma^B\right\}.$$

This quantity measures how much uniform and private randomness we can extract from a random variable that is correlated with a quantum state that an attacker might possess, as shown in [Ren05]. It is also the quantity that governs how many bits of key must be used to encrypt the A part of a quantum state ρ^{AB} against an adversary that knows B [DD].

While much more is known about the min-entropy, the following slightly more unwieldy quantity is used in many proofs:

Definition 2.11 (Quantum conditional 2-entropy)**.** *Let* $\rho^{AB} \in \mathrm{Pos}(\mathsf{A}\otimes\mathsf{B})$*. Then, the conditional 2-entropy of* A *given* B *is defined as*

$$H_2(A|B)_\rho := -\log \inf_{\sigma^B \in \mathrm{D}(\mathsf{B}), \sigma^B > 0} \mathrm{Tr}\left[\left((\sigma^B \otimes \mathbb{I}^A)^{-1/2}\rho^{AB}\right)^2\right].$$

Note that the conditional 2-entropy is always lower-bounded by the conditional min-entropy:

Lemma 2.3. *Let* $\rho^{AB} \in \mathrm{Pos}(\mathsf{A}\otimes\mathsf{B})$*; then* $H_{\min}(A|B)_\rho \leqslant H_2(A|B)_\rho$*.*

Proof. Let $\lambda = 2^{-H_{\min}(A|B)_\rho}$, and let σ^B be a normalized density operator such that $\rho^{AB} \leqslant \lambda\mathbb{I}^A \otimes \sigma^B$; assume without loss of generality that σ^B is positive definite (otherwise redefine B as the support of ρ^B). Also, let $P^{AB} = \mathbb{I}^{AB} - (\rho^{AB})^0$ (i.e. P is a projector onto the kernel of ρ^{AB}). Then, using the fact that $X \leqslant Y \Rightarrow X^{-1/2} \geqslant Y^{-1/2}$ (which one can derive from Propositions V.I.6 and V.I.8 in [Bha96]), we have that $\lambda^{1/2}(\rho^{AB} + \varepsilon P)^{-1/2} \geqslant (\mathbb{I}^A \otimes \sigma^B + \varepsilon P)^{-1/2}$, and therefore

$$\begin{aligned} \operatorname{Tr}\left[\left((\mathbb{I}^A \otimes \sigma^B + \varepsilon P)^{-1/2} \rho^{AB}\right)^2\right] &\leqslant \lambda \operatorname{Tr}\left[\left((\rho^{AB} + \varepsilon P)^{-1/2} \rho^{AB}\right)^2\right] \\ &= \lambda \operatorname{Tr}[\rho^{AB}] \\ &= \lambda. \end{aligned}$$

Taking the limit as $\varepsilon \to 0$ yields the lemma. □

One can also define a "max-entropy" as done in [KRS09] in the following manner:

Definition 2.12 (Quantum conditional max-entropy)**.** *Let ψ^{ABC} be a pure state. Then, the conditional max-entropy of A given B is defined as*

$$H_{\max}(A|B)_\psi := -H_{\min}(A|C)_\psi.$$

Since $H_{\min}(A|C)_\psi$ is invariant under unitaries on C, this does not depend on the particular choice of purification.

Note that there are at present two competing definitions of the max-entropy in circulation, at least in the non-conditional case. The other one is simply the logarithm of the rank of a state. However, the author feels that Definition 2.12 is more compelling given the various results in [KRS09] and [TCR09], as well as the results in this thesis.

In [KRS09], the authors give a nice direct interpretation of both the min- and the max-entropy: given a state ρ^{AB} the conditional min-entropy $H_{\min}(A|B)_\rho$ quantifies how close to a maximally entangled state we can make ρ^{AB} by applying an arbitrary CPTP map on B:

$$2^{-H_{\min}(A|B)_\rho} = |A| \max_{\mathcal{F}^{B \to A'}} F\left((\mathbb{I}^A \otimes \mathcal{F})(\rho^{AB}), \Phi^{AA'}\right)^2$$

where $\mathcal{F}$ ranges over all CPTP maps from $\mathrm{L}(\mathsf{B})$ to $\mathrm{L}(\mathsf{A}')$, and A' is a quantum system of the same dimension as A. Likewise, the max-entropy $H_{\max}(A|B)_\rho$ characterizes how close the state is to being decoupled and uniform on A:

$$2^{H_{\max}(A|B)_\rho} = |A| \max_{\sigma^B \in \mathrm{D}(\mathsf{B})} F\left(\rho^{AB}, \pi^A \otimes \sigma^B\right).$$

It can also be shown that $H_{\min}(A|B)_\rho \leqslant H(A|B) \leqslant H_{\max}(A|B)_\rho$ ([TCR08], Lemma 2).

One problem with all of the above quantities is that they are very sensitive to small variations in the state on which they are defined, whereas most of the quantities that we are bounding with them are not. Hence, if we use these quantities directly, we can end up with very poor bounds in certain cases. For this reason, we define "smooth" versions of these entropies. Instead of computing the entropic quantities directly on the state we are given, we optimize them over an ε-ball around the state; this idea was introduced by Renner and Wolf in [RW04]. For any $\rho^{AB} \in \mathrm{D}(\mathsf{A} \otimes \mathsf{B})$, define

$$\mathfrak{B}(\rho, \varepsilon) := \{\tilde{\rho}^{AB} : \mathrm{Tr}[\tilde{\rho}] \leqslant 1, d_F(\rho, \tilde{\rho}) \leqslant \varepsilon\}.$$

We then define the following quantities:

Definition 2.13 (Smooth conditional min-entropy). *Let ρ^{AB} be a quantum state. Then, the ε-smooth conditional min-entropy of A given B is defined as*

$$H_{\min}^{\varepsilon}(A|B)_\rho := \max_{\sigma^{AB} \in \mathfrak{B}(\rho,\varepsilon)} H_{\min}(A|B)_\sigma.$$

Definition 2.14 (Smooth conditional 2-entropy). *Let ρ^{AB} be a quantum state. Then, the ε-smooth conditional 2-entropy of A given B is defined as*

$$H_2^{\varepsilon}(A|B)_\rho := \max_{\sigma^{AB} \in \mathfrak{B}(\rho,\varepsilon)} H_2(A|B)_\sigma.$$

Definition 2.15 (Smooth conditional max-entropy). *Let ρ^{AB} be a quantum state.*

Then, the ε-smooth conditional max-entropy of A given B is defined as

$$H_{\max}^{\varepsilon}(A|B)_{\rho} := \min_{\sigma^{AB} \in \mathfrak{B}(\rho,\varepsilon)} H_{\max}(A|B)_{\sigma}.$$

As mentioned before, these quantities reduce to von Neumann quantities in the i.i.d. case. This is formalized in the following theorem, called the *fully quantum asymptotic equipartition property* [TCR08] by Tomamichel, Colbeck and Renner:

Theorem 2.4 (Fully Quantum Asymptotic Equipartition Property)**.** *Let ρ^{AB} be a density operator, $\varepsilon > 0$, $\eta \leqslant 2^{-\frac{1}{2}H_{\min}(A|B)_{\rho}} + 2^{\frac{1}{2}H_{\max}(A|B)_{\rho}} + 1 \leqslant 2\sqrt{|A|} + 1$, and $n \in \mathbb{N}$. Then, if $n \geqslant \frac{8}{5}\log\frac{2}{\varepsilon^2}$,*

$$\frac{1}{n}H_{\min}^{\varepsilon}(A^n|B^n)_{\rho^{\otimes n}} \geqslant H(A|B)_{\rho} - 4\log\eta\sqrt{\frac{\log(2/\varepsilon^2)}{n}}.$$

Alternatively, if $\varepsilon = 2^{-kn}$ with $k > \frac{4\log\eta}{\log 3}$, we get

$$\frac{1}{n}H_{\min}^{\varepsilon}(A^n|B^n)_{\rho^{\otimes n}} \geqslant H(A|B)_{\rho} - 2k^2 - \frac{1}{n}k - \frac{4}{k}(\log\eta)^2.$$

2.5 Quantum channel capacities

There are many variants of quantum channel capacities that can be defined, reflecting the large number of possible data transmission scenarios in which quantum channels can be useful. The two main ones are the classical capacity (the best rate at which we can send classical data through a quantum channel) and the quantum capacity (the best rate at which we can send arbitrary qubits through the channel). We can also define entanglement-assisted capacities of these two problems, in which the sender and the receiver share an arbitrary number of EPR pairs that they can use for free to help them transmit either classical or quantum data through the quantum channel, as the case may be. We will not be concerned with the classical capacity of quantum channels in this thesis, however, and the entanglement-assisted classical capacity is simply twice the entanglement-assisted quantum capacity. We shall therefore only talk about the unassisted and entanglement-assisted quantum capacities.

Furthermore, we can either consider what we can do with a single use of an arbitrary channel (which is the most general version of the problem), or we can restrict ourselves to i.i.d. channels (i.e. n copies of a relatively small channel $\mathcal{N}$). The i.i.d. case, in addition to being a practically relevant special case, is also typically much easier to solve. We will consider both problems in this thesis: for all of the problems that we will consider, we will first prove a theorem for a single use of an arbitrary channel, and we will then apply it to an i.i.d. channel. The goal of this section is to define these problems and say a few words about them.

2.5.1 One-shot capacities

We first consider the simpler case of one-shot capacity. (Simpler to define, not to solve!) By the term "one-shot", we mean that we will consider protocols involving a single use of a channel, as opposed to using the same channel n times. Suppose Alice would like to send an arbitrary quantum system M to Bob using the channel $\mathcal{N}^{A'\to C}$ a single time. Alice will therefore encode her message M into the channel input A' using some encoding CPTP map $\mathcal{E}^{M\to A'}$. Upon receiving the channel output C, Bob will attempt to recover M using a decoding CPTP map $\mathcal{D}^{C\to M}$. One would like to make sure that, regardless of the actual state of the message system M, Bob gets that same state at the output. When we consider every possible state of M, we must also include cases in which the contents of M are entangled with another system. While M can be entangled with an arbitrarily large system, it is mathematically equivalent to consider only entanglement with another system R of dimension $|R| = |M|$.

Of course, since we are only using the channel once, one cannot hope in general to have no error whatsoever. We must therefore decide on an error level that we are willing to tolerate, and then look at how big a message we can transmit given this constraint.

Taking all this into consideration, our goal is to find an encoder-decoder pair that satisfies

$$\left\|(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})(\psi^{RM}) - \psi^{RM}\right\|_1 \leqslant \varepsilon$$

for every pure state ψ^{RM}. An alternative way of writing this is via the *diamond norm* on superoperators [Kit97]:

$$\left\|\mathcal{D} \circ \mathcal{N} \circ \mathcal{E} - \mathbb{I}^M\right\|_\diamond \leqslant \varepsilon. \tag{2.1}$$

In other words, the composition of the encoder, channel, and decoder, must be nearly indistinguishable from the identity channel.

This quantity, however, is rather difficult to bound directly because of the optimization over the input state. Fortunately, there exists an essentially equivalent criterion which is much easier to establish for a given protocol. Instead of considering the worst case input, one can consider only a fixed maximally entangled state Φ^{RM} between R and M:

$$\left\|(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})(\Phi^{RM}) - \Phi^{RM}\right\|_1 \leqslant \varepsilon. \tag{2.2}$$

Requiring that an encoder and decoder fulfill this condition is weaker, but it turns out that by slightly reducing the dimension of the input system of the channel, one can turn an encoder-decoder pair that fulfills (2.2) into one that fulfills (2.1) [KW03].

Alice and Bob might also have EPR pairs at their disposal to help them increase the transmission rate; we call this the *entanglement-assisted* capacity. The setting is the same as above, except that Alice and Bob start out with an additional state $\Phi^{\widetilde{A}B}$ that they are allowed to consume at will to help them in their task. We now have an encoder $\mathcal{E}^{M\widetilde{A}\to A'}$ and a decoder $\mathcal{D}^{CB\to M}$, and we want to ensure that

$$\left\|(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E})(\Phi^{RM} \otimes \Phi^{\widetilde{A}B}) - \Phi^{RM}\right\|_1 \leqslant \varepsilon.$$

Classical one-shot capacities were first considered by Han and Verdú [HV94], who pioneered the so-called information-spectrum approach to the capacity of general non-i.i.d. channels. Using an approach that is much closer to the one used in this thesis, Renner, Wolf and Wullschleger [RWW06] used classical versions of min- and max-entropies to derive bounds for the one-shot capacity of classical channels. One the quantum side, Buscemi and Datta [BD09] consider the one-shot capacity using different tools from the ones used here.

2.5.2 Capacities of memoryless channels

We now wish to consider coding for channels of the form $\mathcal{N}^{\otimes n}$, where n grows arbitrarily. In this case, we will want to find the best rate (number of qubits sent divided by number of channel uses) at which we can send quantum data such that the error rate goes to zero as $n \to \infty$. We call such channels *memoryless channels*, since the channel behaves exactly the same way from one use to the next without "remembering" previous inputs; we also sometimes call this the "i.i.d. case". The definitions in this case are slightly more involved due to the fact that we need a series of encoders and decoders that grows with n. We begin by defining the unassisted quantum capacity:

Definition 2.16 (Quantum code). *An (n, R)-code for a quantum channel $\mathcal{N}^{A'\to C}$ is an encoding superoperator $\mathcal{E}^{M\to A'^n}$ and a decoding superoperator $\mathcal{D}^{C^n\to M}$, where M is a 2^{nR}-dimensional quantum system.*

Definition 2.17 (Achievable rate). *A rate R is said to be achievable for a channel $\mathcal{N}^{A'\to C}$ if there exists a sequence of (n, R)-codes $(\mathcal{E}_n, \mathcal{D}_n)$ such that*

$$\lim_{n\to\infty} \left\| \left(\mathcal{D}_n \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}_n\right) - \mathbb{I}^M \right\|_\diamond = 0. \tag{2.3}$$

Definition 2.18 (Quantum capacity). *The quantum capacity $Q(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is the supremum of all achievable rates for this channel.*

Despite considerable efforts, we do not yet have a satisfactory characterization of the quantum capacity. We do have a general lower bound for the capacity [Llo96, Sho02, Dev05] which is given by the coherent information (see Theorem 3.15). This bound is known not to be tight, however [DSS98], and a very strange phenomenon appears: this capacity is not additive. More specifically, there exist pairs of quantum channels $\mathcal{N}$ and $\mathcal{M}$ such that the capacity of both $\mathcal{N}$ and $\mathcal{M}$ is zero, while the capacity of $\mathcal{N} \otimes \mathcal{M}$ is strictly positive [SY08].

We now turn to the definitions relevant for entanglement-assisted capacity:

Definition 2.19 (Quantum entanglement-assisted code)**.** *An (n, R, E)-code for a quantum channel $\mathcal{N}^{A' \to C}$ is an encoding superoperator $\mathcal{E}^{M\widetilde{A} \to A'^n}$, and an associated decoding superoperator $\mathcal{D}^{C^n \widetilde{B} \to M}$, such that $|M| = 2^{nR}$ and $|\widetilde{A}| = |\widetilde{B}| = 2^{nE}$.*

Definition 2.20 (Achievable rate)**.** *A rate R is said to be achievable for a channel $\mathcal{N}^{A' \to C}$ if there exists a sequence of (n, R, E)-codes $(\mathcal{C}_n, \mathcal{D}_n)$ (for arbitrary finite E) such that*

$$\lim_{n \to \infty} \left\| \mathcal{M}_n - \mathbb{I}^M \right\|_\diamond = 0. \tag{2.4}$$

where $\mathcal{M}_n$ is the superoperator $\mathcal{M}_n(\rho) = (\mathcal{D}_n \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}_n)(\Phi^{\widetilde{A}\widetilde{B}} \otimes \rho)$.

Definition 2.21 (Entanglement-assisted quantum capacity)**.** *The entanglement-assisted quantum capacity $Q_E(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is the supremum of all achievable rates for this channel.*

2.5.3 Regularization and single-letter converses

When we set out to characterize the capacity of a type of memoryless channel, we ultimately want to get an expression that can be efficiently computed from the description of the channel. Unfortunately, in quantum information theory, we seldom achieve this ideal. What usually happens is that we are able to give an easily computable *achievable rate region*, meaning a set of transmission rates that we know can be achieved, and we can often give an uncomputable expression for the true capacity. In some cases, such as for the unassisted transmission of quantum data through quantum channels, we know that there is a gap between the two expressions [DSS98]. The same is true for the transmission of classical data through quantum channels [Has09]. In other cases, such as the entanglement-assisted transmission through quantum multiple-access channels [HDW08], we do not know whether this is the case. Only in a few rare instances can we show that the two coincide; the main example is the entanglement-assisted quantum (and classical) capacities of quantum channels.

The achievable rate region and the uncomputable expression for the capacity usually take particular forms. For the sake of concreteness, we will consider the case of the transmission of quantum data through quantum channels, but the situation tends to be very similar in other settings. The best known achievable rate for this task is expressed in the following theorem ([Llo96, Sho02, Dev05], see also Theorem 3.15 for a proof):

Theorem 2.5. *Let $\mathcal{N}^{A'\to C}$ be a quantum channel, let $\sigma^{AA'}$ be any pure state with $\mathsf{A}' \cong \mathsf{A}$, and let $\rho^{AC} = \mathcal{N}^{A'\to C}(\sigma)$. Then, any rate $R < I(A\rangle C)_\rho$ is achievable for the transmission of quantum data through $\mathcal{N}$.*

The main feature of this theorem is that it states the existence of protocols that send quantum data using the channel n times, but whose rates can be computed by looking at a single instance of $\mathcal{N}$. Indeed, the state ρ on which we compute $I(A\rangle C)_\rho$ is a state produced by a single application of $\mathcal{N}$. Furthermore, the proof of the theorem shows that these protocols can be constructed by choosing codes that "look" like the state $(\sigma^{A'})^{\otimes n}$ at the channel input. It therefore gives us some information about the structure of codes that achieve the rates advertised in the theorem statement.

The main question at this point is whether this theorem is optimal or not: is it possible to create codes that go beyond the highest rate this theorem can give? Since the above theorem holds for any channel, it is certainly possible to look at the rates we obtain for channels of the form $\mathcal{N}^{\otimes k}$ (i.e. if we regard k uses of $\mathcal{N}$ as a single channel). If the above theorem were optimal, then doing this should never give us a higher rate than simply looking at $\mathcal{N}$ alone. In [DSS98], the authors show that it is in fact possible to get a higher rate this way, thereby showing Theorem 2.5 to be suboptimal.

This raises a further question: can we in fact get the optimal rate only by using the above theorem on some large number of copies of the same channel, or do we need to do something altogether different? The answer is that taking many copies is sufficient, and is expressed in the following theorem:

Theorem 2.6. *Let $\mathcal{N}^{A' \to C}$ be a quantum channel. Then, the capacity of $\mathcal{N}$ is given by*

$$C = \sup_{n, \sigma^{AA'^n}} \frac{1}{n} I(A\rangle C^n)_\rho \qquad (2.5)$$

where $\sigma^{AA'^n}$ ranges over all pure states, $\mathsf{A} \cong (\mathsf{A}')^{\otimes n}$, $\rho^{AC^n} = \mathcal{N}^{\otimes n}(\sigma)$ and n ranges over all positive integers.

This is what we call a *regularized converse* or *multiletter converse*: it is a converse of Theorem 2.5, provided that we "regularize" it by considering many copies of the channel. This is not a very strong characterization of the capacity. One reason for this is that we cannot compute it: we have no bound on how large n has to be to get within a given factor of the capacity. Another perhaps even more depressing reason is the way we prove this last theorem: we look at an arbitrary code achieving quantum transmission, use it as the state σ in the above theorem, and show that the resulting $\frac{1}{n} I(A\rangle C)_\rho$ is lower-bounded by the rate of the code. Since there exists a code for every rate below the capacity (by definition), the right-hand side of Equation (2.5) can never be lower than the capacity. This makes the above theorem nearly tautological: if we choose the best possible code, then we reach the capacity. It says nothing whatsoever about the structure of capacity-achieving codes, which is perhaps the main motivation for studying channel capacity problems.

As mentioned earlier, however, it is sometimes possible to prove that regularization is not necessary, and that a theorem that considers only one copy (like Theorem 2.5) is optimal. When this is the case, we say that we have a *single-letter converse*. The main example in quantum information theory is the entanglement-assisted quantum and classical capacities (see again Theorem 3.15). A further example is the entanglement-assisted quantum and classical capacities of quantum channels with side-information at the transmitter, which is studied in Chapter 4, the single-letter converse being given as Theorem 4.5. When we have a single-letter converse, it means that the code structure used in the proof of the corresponding theorem is in fact the optimal way to code for this type of channels. We can they say that we have a good grasp on how the channel carries information.

Finding expressions for the various capacities that are easily computable and that give us information about the structure of optimal codes is one of the main goals of information theory, and it has generally been rather difficult to achieve in the quantum setting. Finding such expressions for the most basic quantum capacities (such as the unassisted quantum and classical capacities of quantum channels) is one of the most central open problems in the field today.

2.6 The duality between vectors and operators

Periodically throughout this thesis it will be extremely useful to turn multipartite pure states into operators, and vice versa. This is simply a generalization of turning a "ket" into a "bra": if we have a vector $|\psi\rangle \in \mathsf{A}$, then we can turn it into an operator $\langle\psi| \in \mathrm{L}(\mathsf{A}, \mathbb{C})$ from vectors to the complex numbers, by defining $\langle\psi|$ as the only operator in $\mathrm{L}(\mathsf{A}, \mathbb{C})$ such that $\langle\psi|\varphi\rangle = \langle|\psi\rangle, |\varphi\rangle\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the inner product in A. We can turn multipartite states into more interesting operators, however. Endow A and B with standard orthonormal bases $\{|a_i\rangle^A\}$ and $\{|b_i\rangle^B\}$ respectively, and let $\mathrm{op}_{A\to B} : \mathsf{A}\otimes\mathsf{B} \to \mathrm{L}(\mathsf{A}, \mathsf{B})$ be defined as

$$\mathrm{op}_{A\to B}(|a_i\rangle|b_j\rangle) = |b_j\rangle\langle a_i| \qquad \forall i, j.$$

This operation depends on the choice of standard basis; therefore, whenever it is used, a particular choice of basis is implied. Since this choice will never matter in this thesis, we shall not explicitly define these bases.

The following properties of the op transformation will be needed:

Lemma 2.7. *Let $|\psi\rangle^{AB}$ and $|\varphi\rangle^{AC}$ be any vectors in $\mathsf{A}\otimes\mathsf{B}$ and $\mathsf{A}\otimes\mathsf{C}$ respectively. Then,* $\mathrm{op}_{A\to B}(|\psi\rangle^{AB})|\varphi\rangle^{AC} = \mathrm{op}_{A\to C}(|\varphi\rangle^{AC})|\psi\rangle^{AB}$.

Proof. Let $\{|a_i\rangle\}$, $\{|b_i\rangle\}$, and $\{|c_i\rangle\}$ be the canonical bases for A, B and C respectively, and let

$$\begin{aligned}|\psi\rangle^{AB} &= \sum_{ij} \alpha_{ij}|a_i\rangle|b_j\rangle \\ |\varphi\rangle^{AC} &= \sum_{ij} \beta_{ij}|a_i\rangle|c_j\rangle.\end{aligned}$$

Then,

$$\begin{aligned}\mathrm{op}_{A\to B}(|\psi\rangle^{AB})|\varphi\rangle^{AC} &= \sum_{ijkl} \alpha_{ij}\beta_{kl}|b_j\rangle\langle a_i|a_k\rangle|c_l\rangle \\ &= \sum_{ijl} \alpha_{ij}\beta_{il}|b_j\rangle|c_l\rangle \\ \mathrm{op}_{A\to C}(|\varphi\rangle^{AC})|\psi\rangle^{AB} &= \sum_{ijkl} \alpha_{ij}\beta_{kl}|c_l\rangle\langle a_k|a_i\rangle|b_j\rangle \\ &= \sum_{ijl} \alpha_{ij}\beta_{il}|b_j\rangle|c_l\rangle.\end{aligned}$$

□

Lemma 2.8. *Let $|\psi\rangle^{AB}$ be any vector in $\mathsf{A}\otimes\mathsf{B}$, let A' be a system of equal dimension to A, and let $|\Phi\rangle^{AA'} = \frac{1}{\sqrt{|A|}}\sum_i |a_i\rangle|a_i'\rangle$, where the $|a_i\rangle$'s and $|a_i'\rangle$'s are the canonical bases of A and A' respectively. Then,*

$$\sqrt{|A|}\,\mathrm{op}_{A\to B}(|\psi\rangle^{AB})|\Phi\rangle^{AA'} = |\psi\rangle^{A'B}.$$

Proof. Let $|\psi\rangle^{AB} = \sum_{ij} \alpha_{ij} |a_i\rangle|b_j\rangle$; we then get that

$$\begin{aligned}
\sqrt{|A|}\,\mathrm{op}_{A\to B}(|\psi\rangle^{AB})|\Phi\rangle^{AA'} &= \sum_{ijk} \alpha_{ij}|b_j\rangle\langle a_i|a_k\rangle|a'_k\rangle \\
&= \sum_{ij} \alpha_{ij}|a'_i\rangle^{A'}|b_j\rangle^{B} \\
&= |\psi\rangle^{A'B}.
\end{aligned}$$

□

Lemma 2.9. *For any $|\psi\rangle \in \mathsf{A}\otimes\mathsf{B}$ and any $M^{A\to C}$, we have that*

$$\begin{aligned}
\mathrm{op}_{B\to C}(M|\psi\rangle) &= M\,\mathrm{op}_{B\to A}(|\psi\rangle) \\
\mathrm{op}_{C\to B}(M|\psi\rangle) &= \mathrm{op}_{A\to B}(|\psi\rangle)M_T
\end{aligned}$$

where the T subscript denotes transposition.

Proof. Let $|\psi\rangle = \sum_{ij}\alpha_{ij}|a_i\rangle|b_j\rangle$ and $M = \sum_{kl}\gamma_{kl}|c_k\rangle\langle a_l|$. Then,

$$\begin{aligned}
\mathrm{op}_{B\to C}(M|\psi\rangle) &= \mathrm{op}_{B\to C}\left(\sum_{ijkl}\alpha_{ij}\gamma_{kl}|c_k\rangle\langle a_l|a_i\rangle|b_j\rangle\right) \\
&= \mathrm{op}_{B\to C}\left(\sum_{ijk}\alpha_{ij}\gamma_{ki}|c_k\rangle|b_j\rangle\right) \\
&= \sum_{ijk}\alpha_{ij}\gamma_{ki}|c_k\rangle\langle b_j|.
\end{aligned}$$

Likewise,

$$\begin{aligned}
M\,\mathrm{op}_{B\to A}(|\psi\rangle) &= \sum_{ijkl}\alpha_{ij}\gamma_{kl}|c_k\rangle\langle a_l|a_i\rangle\langle b_j| \\
&= \sum_{ijk}\alpha_{ij}\gamma_{ki}|c_k\rangle\langle b_j|.
\end{aligned}$$

The other statement is proven in the same manner:

$$\begin{aligned}\mathrm{op}_{C\to B}(M|\psi\rangle) &= \sum_{ijkl} \alpha_{ij}\gamma_{kl}\,\mathrm{op}_{C\to B}\left(|c_k\rangle\langle a_l|a_i\rangle|b_j\rangle\right)\\ &= \sum_{ijk} \alpha_{ij}\gamma_{ki}|b_j\rangle\langle c_k|\end{aligned}$$

and

$$\begin{aligned}\mathrm{op}_{A\to B}(|\psi\rangle)M_T &= \sum_{ijkl} \alpha_{ij}\gamma_{kl}|b_j\rangle\langle a_i|a_l\rangle\langle c_k|\\ &= \sum_{ijk} \alpha_{ij}\gamma_{ki}|b_j\rangle\langle c_k|.\end{aligned}$$

□

Lemma 2.10. *Let* $|\psi\rangle \in \mathsf{A}\otimes\mathsf{B}$*. Then,* $\mathrm{Tr}_B[\psi^{AB}] = \mathrm{op}_{B\to A}(|\psi\rangle)\,\mathrm{op}_{B\to A}(|\psi\rangle)^\dagger$*.*

Proof. Let $|\psi\rangle = \sum_i \alpha_i|\psi_i\rangle^A|\varphi_i\rangle^B$ be the Schmidt decomposition of $|\psi\rangle$.

$$\mathrm{Tr}_B[\psi^{AB}] = \sum_i \alpha_i^2|\psi_i\rangle\langle\psi_i|^A$$

and

$$\begin{aligned}\mathrm{op}_{B\to A}(|\psi\rangle)\,\mathrm{op}_{B\to A}(|\psi\rangle)^\dagger &= \sum_{ij} \alpha_i\alpha_j|\psi_i\rangle\langle\varphi_i^*|\varphi_j^*\rangle\langle\psi_j|\\ &= \sum_i \alpha_i^2|\psi_i\rangle\langle\psi_i|^A.\end{aligned}$$

□

We will also need to turn operators into vectors through the same process. For any pair of systems A and B, define $\mathrm{vec} : \mathrm{L}(\mathsf{A},\mathsf{B}) \to \mathsf{A}\otimes\mathsf{B}$ as the transformation:

$$\mathrm{vec}(|b_j\rangle\langle a_i|) = |a_i\rangle|b_j\rangle.$$

It is simply the inverse of op.

We will need the following property of the vec transformation:

Lemma 2.11. *Let* $M^{A\to B}$ *and* $N^{A\to B}$ *be arbitrary operators. Then,* $\mathrm{Tr}[N^\dagger M] = \mathrm{vec}(N)^\dagger \, \mathrm{vec}(M)$.

Proof. Let $M = \sum_{ij} m_{ij} |b_j\rangle\langle a_i|$ and $N = \sum_{ij} n_{ij} |b_j\rangle\langle a_i|$. Then,

$$\begin{aligned}\mathrm{Tr}[N^\dagger M] &= \sum_{ijkl} \mathrm{Tr}[m_{ij} n_{kl}^* |a_k\rangle\langle b_l|b_j\rangle\langle a_i|] \\ &= \sum_{ij} m_{ij} n_{ij}^*\end{aligned}$$

and

$$\begin{aligned}\mathrm{vec}(N)^\dagger \, \mathrm{vec}(M) &= \sum_{ijkl} m_{ij} n_{kl}^* \langle a_k|a_i\rangle\langle b_l|b_j\rangle \\ &= \sum_{ij} m_{ij} n_{ij}^*.\end{aligned}$$

□

CHAPTER 3

THE DECOUPLING THEOREM

One peculiar feature of quantum information theory is that some of the simplest coding theorems that we know come from theorems that tell us how to *remove* correlations, even though the goal of an error-correcting code is to establish correlations between the sender and the receiver. The basic idea is the following: to prove a coding theorem, we generally need to assert the existence of a decoder of some sort; this decoder must be able to reproduce a particular state with good fidelity given only partial or noisy information. By purifying all systems, we can consider all subsystems that are not held by the decoder. These will generally include a subsystem purifying the state that the decoder needs to produce, as well as systems considered as part of the environment or that we otherwise don't care about. It turns out that, in such a case, a decoder exists if and only if the system purifying the desired state and the "environment" are close to a product state. The theorem that ensures this is called Uhlmann's theorem, and is the subject of the next section. Of course, for this approach to work, we need a way to ensure that two systems are close to a product state. Section 3.2 will present a very general decoupling theorem with which we will prove all of the coding theorems in this thesis.

Although some elements of this approach were already used earlier, this method came into its own with the discovery of the *state merging* protocol [HOW07], and later, the *Fully Quantum Slepian-Wolf (FQSW)* [ADHW06] protocol. A whole array of results, including the "mother" and "father" [DHW03], can be easily derived from either of these protocols, such as the quantum reverse Shannon theorem [BDH$^+$06], the Lloyd-Shor-Devetak (LSD) theorem [Llo96] [Sho02] [Dev05], one-way entanglement distillation [DW05], and distributed compression [ADHW06]. This chapter will present a generalization of both FQSW and state merging that is much more flexible and which can therefore be used in more diversified contexts.

3.1 Uhlmann's theorem

Before starting, we will need to formally define what we mean by *purification*:

Definition 3.1 (Purification). *Let $\rho^A \in \mathrm{D}(\mathsf{A})$ be any normalized density operator. Then, a purification of ρ^A is any normalized vector $|\psi\rangle \in \mathsf{A} \otimes \mathsf{B}$, with B an arbitrary quantum system, such that $\mathrm{Tr}_B[|\psi\rangle\langle\psi|^{AB}] = \rho^A$. We then call B the* purifying system.

For any density operator, a purification exists, and is unique up to isometries on the purifying system.

Uhlmann's theorem was first shown in [Uhl76]; the proof given here essentially follows the one in [Wat08].

Theorem 3.1 (Uhlmann). *Let ρ^A and σ^A be two quantum states, and let $|\psi\rangle^{AB}$ and $|\varphi\rangle^{AC}$ be purifications of ρ^A and σ^A respectively (the purifying systems B and C need not be isomorphic). Then,*

$$F(\rho^A, \sigma^A) = \max_{V^{B\to C}} \left|\langle\psi|V^\dagger|\varphi\rangle\right| \tag{3.1}$$

where the maximization is over all partial isometries from B to C.

Proof. Let $U^{A\to B}$ and $W^{A\to C}$ be partial isometries such that $|\psi\rangle^{AB} = \mathrm{vec}(U\sqrt{\rho})$ and $|\varphi\rangle^{AC} = \mathrm{vec}(W\sqrt{\sigma})$. Then,

$$\begin{aligned}
F(\rho^A, \sigma^A) &= \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1 && (3.2)\\
&= \left\|U\sqrt{\rho}\sqrt{\sigma}W^\dagger\right\|_1 && (3.3)\\
&= \max_{V^{B\to C}} \left|\mathrm{Tr}\left[VU\sqrt{\rho}\sqrt{\sigma}W^\dagger\right]\right| && (3.4)\\
&= \max_{V^{B\to C}} \left|\mathrm{vec}(VU\sqrt{\rho})^\dagger\,\mathrm{vec}(W\sqrt{\sigma})\right| && (3.5)\\
&= \max_{V^{B\to C}} \left|\langle\psi|V^\dagger|\varphi\rangle\right| && (3.6)
\end{aligned}$$

where we have used Lemma I.6 from the appendix on line (3.4) and 2.11 on line (3.5). □

The main use of this theorem for coding purposes is that it often gives us a decoder "for free". Indeed, assume that, at the end of the execution of a channel coding protocol, we have a tripartite pure state $|\psi\rangle^{BER}$, with the three subsystems representing the shares of Bob, the environment, and a "reference" system which purifies the qubits that Alice wanted to send to Bob. Now, suppose that we were able to show that the environment is nearly uncorrelated with the reference: $F\left(\psi^{RE}, \rho^R \otimes \sigma^E\right) \geqslant 1 - \varepsilon$. Then, given a product purification $|\varphi\rangle^{R\bar{B}} \otimes |\xi\rangle^{E\hat{B}}$ of $\rho^R \otimes \sigma^E$, there exists a partial isometry $V^{B\to\bar{B}\hat{B}}$ such that $F\left(V|\psi\rangle^{BER}, |\varphi\rangle^{R\bar{B}} \otimes |\xi\rangle^{E\hat{B}}\right) \geqslant 1 - \varepsilon$.

Since we generally use the trace distance rather than the fidelity, the following corollary of Uhlmann's theorem (Lemma 2.2 in [DHW05]) will be very useful to us:

Corollary 3.2. *Let $|\psi\rangle^{AB}$ and $|\varphi\rangle^{AC}$ be two quantum states such that $\left\|\psi^A - \varphi^A\right\|_1 \leqslant \varepsilon$. Then there exists an isometry $U^{B\to C}$ such that $\left\|\left(U^{B\to C} \cdot \psi^{AB}\right) - \varphi^{AC}\right\|_1 \leqslant 2\sqrt{\varepsilon}$.*

Proof. If $\|\psi^A - \varphi^A\|_1 \leqslant \varepsilon$, then by the Fuchs-van de Graaf inequalities [FvdG99] (Lemma 2.2) we have that $F(\psi^A, \varphi^A) \geqslant 1 - \frac{1}{2}\epsilon$. By Uhlmann's theorem, this means that there exists a partial isometry $U^{B\to C}$ such that $F(U^{B\to C} \cdot \psi^{AB}, \varphi^{AC}) \geqslant 1 - \frac{1}{2}\varepsilon$. A second application of the Fuchs-van de Graaf inequalities concludes the proof. □

3.2 The decoupling theorem

To be able to use Uhlmann's theorem to derive a coding scheme, we need a way to ensure that two quantum systems are nearly uncorrelated. The main theorem of this section will achieve this for us.

Suppose Alice holds the A share of a mixed state ρ^{AR}. We would like to perform an operation on Alice's system to ensure that her share is decoupled from the reference. We will consider a very general operation: a fixed unitary transformation followed by an arbitrary completely positive superoperator $\mathcal{T}^{A\to E}$. We will show that if we choose the unitary transformation randomly according to the Haar measure (which can essentially be viewed as the uniform distribution over all unitaries), then the resulting protocol will on average perform well. This generalizes all of the decoupling theorems in the literature that the author is aware of, including the Fully Quantum Slepian-Wolf theorem [ADHW06], which corresponds to the special case in which $\mathcal{T}$ traces out part of the system, as well as the state merging [HOW07] theorem, in which $\mathcal{T}^{A\to EX}$ corresponds to making a rank-$|E|$ measurement and then storing the measurement result in the classical register X and the residual quantum state in E. One advantage of this generalization is that it allows us to choose $\mathcal{T}$ to be a very complex operation; one especially interesting example is to pick $\mathcal{T}$ to be the complementary channel (the channel to the environment) of a channel we are interested in coding for. Another advantage is the use of (smooth) conditional 2-entropies rather than purities and dimension bounds as was done in all of these theorems (although, in the case of state merging, this was already done in [Ber08] and [BCR09], and, in the case of FQSW, by Hayden in [Hay06]). This theorem allows to show directly that the environment is decoupled from any system of interest, which is usually what we need to show.

We will calculate how close the remaining state on ER is to a product state in the main theorem of this section (Theorem 3.7). To get to it, however, we will first need the following four technical lemmas. The first one is simply a trick that we will use to compute the trace of the square of a matrix:

Lemma 3.3 (Swap trick). *Given two operators $M \in \mathrm{L}(\mathsf{A})$ and $N \in \mathrm{L}(\mathsf{A})$, then $\mathrm{Tr}[MN] = \mathrm{Tr}[(M \otimes N)F]$, where F swaps the two copies of the A subsystem.*

Proof. Write M and N in the standard basis for A: $M = \sum_{ij} m_{ij}|i\rangle\langle j|$ and $N =$

$\sum_{kl} n_{kl}|k\rangle\langle l|$. Then,

$$\mathrm{Tr}[(M \otimes N)F] = \mathrm{Tr}\left[\left(\sum_{ijkl} m_{ij}n_{kl}|i\rangle\langle j| \otimes |k\rangle\langle l|\right) F\right] \tag{3.7}$$

$$= \mathrm{Tr}\left[\sum_{ijkl} m_{ij}n_{kl}|i\rangle\langle l| \otimes |k\rangle\langle j|\right] \tag{3.8}$$

$$= \sum_{ij} m_{ij}n_{ji} \tag{3.9}$$

$$= \mathrm{Tr}[MN]. \tag{3.10}$$

□

The second lemma involves averaging over Haar-distributed unitaries. While it would take us too far afield to formally introduce the Haar measure, it can simply be thought of as the uniform probability distribution over the set of all unitaries on a Hilbert space. The following then tells us the expected value of $U^{\otimes 2} \cdot M$ (with $M \in \mathrm{L}(\mathsf{A})$) when U is selected "uniformly at random":

Lemma 3.4. *Given an operator $M \in \mathrm{L}(\mathsf{A}^{\otimes 2})$, we have that*

$$\mathbb{E}(M) := \int_{\mathbb{U}(A)} (U^{\otimes 2} \cdot M)dU = \alpha \mathbb{I}^{AA'} + \beta F^A \tag{3.11}$$

where α and β are such that $\mathrm{Tr}[M] = \alpha|A|^2 + \beta|A|$ and $\mathrm{Tr}[MF] = \alpha|A| + \beta|A|^2$, and where dU is the normalized Haar measure on $\mathbb{U}(A)$.

Proof. This is a standard result in Schur-Weyl duality. This is a special case of, for instance, Proposition 2.2 in [CS06]. To see this, note that Proposition 2.2 states that $\mathbb{E} : \mathrm{L}(\mathsf{A}^{\otimes 2}) \to \mathrm{L}(\mathsf{A}^{\otimes 2})$ is an orthogonal projection onto $\mathrm{span}\{\mathbb{I}, F\}$ under the inner product $\langle A, B\rangle = \mathrm{Tr}[A^\dagger B]$. Hence, $\mathbb{E}(M)$ can be written as $\alpha \mathbb{I}^{AA'} + \beta F^A$ as claimed, and the conditions $\mathrm{Tr}[\mathbb{I}\mathbb{E}(M)] = \mathrm{Tr}[M]$ and $\mathrm{Tr}[F\mathbb{E}(M)] = \mathrm{Tr}[FM]$ must be fulfilled, and these lead to the two conditions on α and β. □

The following bounds the ratio of the purity of a bipartite state and the purity of the reduced state on one subsystem:

Lemma 3.5. *Let $\xi^{AB} \in \mathrm{Pos}(\mathsf{A} \otimes \mathsf{B})$ be any positive semidefinite operator. Then*

$$\frac{1}{|A|} \leqslant \frac{\mathrm{Tr}\left[\xi^{AB^2}\right]}{\mathrm{Tr}\left[\xi^{B^2}\right]} \leqslant |A|. \tag{3.12}$$

Proof. Letting A' be a system isomorphic to A, we first prove the left-hand side:

$$\begin{aligned}
\mathrm{Tr}\left[\xi^{B^2}\right] &= \mathrm{Tr}\left[\mathrm{Tr}_A\left[\xi^{AB}\right]^2\right] && (3.13)\\
&= \mathrm{Tr}\left[\mathrm{Tr}_A\left[\xi^{AB}\right]\mathrm{Tr}_{A'}\left[\xi^{A'B}\right]\right] && (3.14)\\
&= \mathrm{Tr}\left[\xi^{AB}\left(\mathrm{Tr}_{A'}\left[\xi^{A'B}\right]\otimes \mathbb{I}^A\right)\right] && (3.15)\\
&= \mathrm{Tr}\left[(\xi^{AB}\otimes \mathbb{I}^{A'})(\xi^{A'B}\otimes \mathbb{I}^A)\right] && (3.16)\\
&\leqslant \sqrt{\mathrm{Tr}\left[(\xi^{AB}\otimes \mathbb{I}^{A'})^2\right]\mathrm{Tr}\left[(\xi^{A'B}\otimes \mathbb{I}^{A})^2\right]} && (3.17)\\
&= \mathrm{Tr}\left[\xi^{AB^2}\otimes \mathbb{I}^{A'}\right] && (3.18)\\
&= |A|\,\mathrm{Tr}\left[\xi^{AB^2}\right] && (3.19)
\end{aligned}$$

where the inequality is due to an application of Cauchy-Schwarz. The right-hand side follows from the fact that $\xi^{AB} \leqslant |A|\mathbb{I}^A \otimes \xi^B$. This can in turn be seen from the fact that $|A|\mathbb{I}^A \otimes \xi^B = \sum_{i=1}^{|A|^2} U_i^A \cdot \xi^{AB}$, where the U_i's are Weyl operators with $U_1 = \mathbb{I}$. □

In the main proof, we will need to bound the trace distance between two states using the 2-norm. The following lemma will allow us to do this:

Lemma 3.6. *Let $M \in \mathrm{L}(\mathsf{A})$ be any operator and let $\sigma \in \mathrm{Pos}(\mathsf{A})$ be a positive definite operator. Then,*

$$\|M\|_1 \leqslant \sqrt{\mathrm{Tr}[\sigma]\,\mathrm{Tr}[\sigma^{-1/4}M\sigma^{-1/2}M^\dagger\sigma^{-1/4}]}. \tag{3.20}$$

In particular, if M is Hermitian, then

$$\|M\|_1 \leqslant \sqrt{\mathrm{Tr}[\sigma]\,\mathrm{Tr}[(\sigma^{-1/4}M\sigma^{-1/4})^2]}. \tag{3.21}$$

This is a slight generalization of Lemma 5.1.3 in [Ren05]; we give a different proof here for completeness:

Proof.

$$
\begin{aligned}
\|M\|_1 &= \max_U |\mathrm{Tr}[UM]| && (3.22)\\
&= \max_U \left|\mathrm{Tr}[(\sigma^{1/4}U\sigma^{1/4})(\sigma^{-1/4}M\sigma^{-1/4})]\right| && (3.23)\\
&\leqslant \max_U \sqrt{\mathrm{Tr}[(\sigma^{1/4}U\sigma^{1/4})(\sigma^{1/4}U^\dagger(\sigma^{1/4})]\,\mathrm{Tr}\left[\sigma^{-1/4}M\sigma^{-1/2}M^\dagger\sigma^{-1/4}\right]} && (3.24)\\
&= \sqrt{\max_U \mathrm{Tr}[\sigma^{1/2}U\sigma^{1/2}U^\dagger]\,\mathrm{Tr}\left[\sigma^{-1/4}M\sigma^{-1/2}M^\dagger\sigma^{-1/4}\right]} && (3.25)\\
&= \sqrt{\mathrm{Tr}[\sigma]\,\mathrm{Tr}\left[\sigma^{-1/4}M\sigma^{-1/2}M^\dagger\sigma^{-1/4}\right]} && (3.26)
\end{aligned}
$$

where the first equality is an application of Lemma I.6 and the inequality results from an application of Cauchy-Schwarz, and the maximizations are over all unitaries on A. The last equality follows from

$$
\begin{aligned}
\max_U \mathrm{Tr}[\sigma^{1/2}U\sigma^{1/2}U^\dagger] &\leqslant \max_U \sqrt{\mathrm{Tr}[\sigma]\,\mathrm{Tr}[U\sigma^{1/2}U^\dagger U\sigma^{1/2}U^\dagger]}\\
&= \mathrm{Tr}[\sigma]\\
&\leqslant \max_U \mathrm{Tr}[\sigma^{1/2}U\sigma^{1/2}U^\dagger].
\end{aligned}
$$

□

We are now ready to prove the main theorem:

Theorem 3.7. *Let ρ^{AR} be a density operator, $\mathcal{T}^{A\to E}$ be any completely positive superoperator, and define $\omega^{A'E} := (\mathcal{T}\otimes\mathbb{I}^{A'})(\Phi^{AA'})$. Then,*

$$
\int_{\mathbb{U}(A)} \left\|\mathcal{T}(U\cdot\rho^{AR}) - \omega^E\otimes\rho^R\right\|_1 dU \leqslant 2^{-\frac{1}{2}H_2(A'|E)_\omega - \frac{1}{2}H_2(A|R)_\rho} \qquad (3.27)
$$

where $\int \cdot dU$ denotes the integral over the Haar measure over unitaries U^A acting on A.

Proof. Throughout the proof, we will denote with a prime the "twin" subsystems used when we take tensor copies of operators, and F^S denotes a swap between S and S'.

We first use Lemma 3.6. Letting σ^E and ζ^R be any normalized, positive definite density matrices on E and R respectively, we get:

$$\begin{aligned}&\left\|\mathcal{T}(U\cdot\rho^{AB})-\omega^E\otimes\rho^R\right\|_1\\&\qquad\leqslant\sqrt{\operatorname{Tr}\left[\left((\sigma^E\otimes\zeta^R)^{-1/4}(\mathcal{T}(U\cdot\rho^{AR})-\omega^E\otimes\rho^R)(\sigma^E\otimes\zeta^R)^{-1/4}\right)^2\right]}.\end{aligned}\tag{3.28}$$

Define the CP map $\tilde{\mathcal{T}}^{A\to E}$ as $\tilde{\mathcal{T}}(\xi)=\sigma^{E^{-1/4}}\mathcal{T}(\xi)\sigma^{E^{-1/4}}$, the state $\tilde{\rho}^{AR}$ as $\tilde{\rho}^{AR}=\zeta^{R^{-1/4}}\rho^{AR}\zeta^{R^{-1/4}}$, and the state $\tilde{\omega}^{A'E}$ as $\tilde{\omega}^{A'E}=\tilde{\mathcal{T}}(\Phi^{A'A})$. We then rewrite the above as

$$\left\|\mathcal{T}(U\cdot\rho^{AR})-\omega^E\otimes\rho^R\right\|_1\leqslant\sqrt{\operatorname{Tr}\left[\Big((\tilde{\mathcal{T}}(U\cdot\tilde{\rho}^{AR})-\tilde{\omega}^E\otimes\tilde{\rho}^R)\Big)^2\right]}.\tag{3.29}$$

Using Jensen's inequality, we can get

$$\int\left\|\mathcal{T}(U\cdot\rho^{AR})-\omega^E\otimes\rho^R\right\|_1 dU\leqslant\sqrt{\int\operatorname{Tr}\left[\Big((\tilde{\mathcal{T}}(U\cdot\tilde{\rho}^{AR})-\tilde{\omega}^E\otimes\tilde{\rho}^R)\Big)^2\right]dU}.\tag{3.30}$$

We now simplify the integral:

$$\begin{aligned}&\int\operatorname{Tr}\left[\Big(\tilde{\mathcal{T}}(U\cdot\tilde{\rho}^{AR})-\tilde{\omega}^E\otimes\tilde{\rho}^R\Big)^2\right]dU\\&=\int\operatorname{Tr}\left[\Big(\tilde{\mathcal{T}}(U\cdot\tilde{\rho}^{AR})\Big)^2\right]dU-2\int\operatorname{Tr}\left[\tilde{\mathcal{T}}(U\cdot\tilde{\rho}^{AR})\left(\tilde{\omega}^E\otimes\tilde{\rho}^R\right)\right]dU+\operatorname{Tr}\left[\left(\tilde{\omega}^E\otimes\tilde{\rho}^R\right)^2\right]\\&=\int\operatorname{Tr}\left[\Big(\tilde{\mathcal{T}}(U\cdot\tilde{\rho}^{AR})\Big)^2\right]dU-2\operatorname{Tr}\left[\tilde{\mathcal{T}}\left(\int U\cdot\tilde{\rho}^{AR}dU\right)\left(\tilde{\omega}^E\otimes\tilde{\rho}^R\right)\right]+\operatorname{Tr}\left[\left(\tilde{\omega}^E\otimes\tilde{\rho}^R\right)^2\right]\\&=\int\operatorname{Tr}\left[\Big(\tilde{\mathcal{T}}(U\cdot\tilde{\rho}^{AR})\Big)^2\right]dU-\operatorname{Tr}\left[(\tilde{\omega}^E)^2\right]\operatorname{Tr}\left[(\tilde{\rho}^R)^2\right].\end{aligned}\tag{3.31}$$

We attack the first term as follows:

$$\begin{aligned}
&\int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}(U\cdot\tilde{\rho}^{AR})\right)^2\right]dU \\
&\qquad\qquad = \int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}(U\cdot\tilde{\rho}^{AR})\right)^{\otimes 2} F^{ER}\right]dU \\
&\qquad\qquad = \int \mathrm{Tr}\left[\left(\tilde{\mathcal{T}}^{\otimes 2}(U^{\otimes 2}\cdot(\tilde{\rho}^{AR})^{\otimes 2})\right) F^{ER}\right]dU \\
&\qquad\qquad = \int \mathrm{Tr}\left[(\tilde{\rho}^{AR})^{\otimes 2}\left(\left\{U^{\dagger\otimes 2}\cdot(\tilde{\mathcal{T}}^{\dagger})^{\otimes 2}(F^E)\right\}\otimes F^R\right)\right]dU \\
&\qquad\qquad = \mathrm{Tr}\left[(\tilde{\rho}^{AR})^{\otimes 2}\left(\int\left\{U^{\dagger\otimes 2}\cdot(\tilde{\mathcal{T}}^{\dagger})^{\otimes 2}(F^E)\right\}dU\otimes F^R\right)\right].
\end{aligned} \tag{3.32}$$

where we have used Lemma 3.3 in the first equality, and the definition of the adjoint of a superoperator in the third equality. We now compute the integral using Lemma 3.4:

$$\int U^{\dagger\otimes 2}\cdot(\tilde{\mathcal{T}}^{\dagger})^{\otimes 2}(F^E)dU = \alpha\mathbb{I}^{AA'} + \beta F^A \tag{3.33}$$

where α and β satisfy the following equations:

$$\alpha|A|^2 + \beta|A| = \mathrm{Tr}\left[(\tilde{\mathcal{T}}^{\dagger})^{\otimes 2}(F^E)\right] \tag{3.34}$$

$$= \mathrm{Tr}\left[F^E(\tilde{\mathcal{T}})^{\otimes 2}(\mathbb{I}^{AA'})\right] \tag{3.35}$$

$$= |A|^2\,\mathrm{Tr}\left[F^E(\tilde{\omega}^E)^{\otimes 2}\right] \tag{3.36}$$

$$= |A|^2\,\mathrm{Tr}\left[(\tilde{\omega}^E)^2\right] \tag{3.37}$$

and

$$\begin{aligned}
\alpha|A| + \beta|A|^2 &= \operatorname{Tr}\left[(\tilde{\mathcal{T}}^{\dagger})^{\otimes 2}(F^E)F^A\right] \qquad (3.38)\\
&= \operatorname{Tr}\left[F^E(\tilde{\mathcal{T}})^{\otimes 2}(F^A)\right] \qquad (3.39)\\
&= |A|^2 \operatorname{Tr}\left[F^E \operatorname{Tr}_{AA'}\left[(\tilde{\omega}^{AE})^{\otimes 2}(F^A \otimes \mathbb{I}^{EE'})\right]\right] \qquad (3.40)\\
&= |A|^2 \operatorname{Tr}\left[(\mathbb{I}^{AA'} \otimes F^E)(\tilde{\omega}^{AE})^{\otimes 2}(F^A \otimes \mathbb{I}^{EE'})\right] \qquad (3.41)\\
&= |A|^2 \operatorname{Tr}\left[F^{AE}(\tilde{\omega}^{AE})^{\otimes 2}\right] \qquad (3.42)\\
&= |A|^2 \operatorname{Tr}\left[(\tilde{\omega}^{A'E})^2\right]. \qquad (3.43)
\end{aligned}$$

where, $\tilde{\omega}^{AE}$ is simply $\tilde{\omega}^{A'E}$ with A and A' reversed. In the second equality, we have used the fact that $|A|\tilde{\omega}^{AE}$ is a Choi-Jamiołkowski [Cho75, Jam72] representation of $\tilde{\mathcal{T}}$; the fourth equality is due to the fact that the adjoint of the partial trace is tensoring with the identity.

Solving this system of equations yields

$$\alpha = \operatorname{Tr}\left[(\tilde{\omega}^E)^2\right]\left(\frac{|A|^2 - \frac{|A|\operatorname{Tr}\left[(\tilde{\omega}^{A'E})^2\right]}{\operatorname{Tr}[(\tilde{\omega}^E)^2]}}{|A|^2-1}\right) \tag{3.44}$$

$$\beta = \operatorname{Tr}\left[(\tilde{\omega}^{A'E})^2\right]\left(\frac{|A|^2 - \frac{|A|\operatorname{Tr}\left[(\tilde{\omega}^{E})^2\right]}{\operatorname{Tr}\left[(\tilde{\omega}^{A'E})^2\right]}}{|A|^2-1}\right). \tag{3.45}$$

By applying Lemma 3.5, we can simplify this to $\alpha \leqslant \operatorname{Tr}\left[(\tilde{\omega}^E)^2\right]$ and $\beta \leqslant \operatorname{Tr}\left[(\tilde{\omega}^{A'E})^2\right]$. Substituting this into (3.32) and using Lemma 3.3 twice, and then substituting into (3.30) yields

$$\int \left\|\mathcal{T}(U\cdot\rho^{AR}) - \omega^E \otimes \rho^R\right\|_1 dU \leqslant \sqrt{\operatorname{Tr}\left[(\tilde{\omega}^{A'E})^2\right]\operatorname{Tr}\left[(\tilde{\rho}^{AR})^2\right]}. \tag{3.46}$$

We then get the theorem by using the definitions of $\tilde{\omega}$, $\tilde{\rho}$ and the definition of H_2. □

We now prove a version of the theorem that allows us to replace the H_2 in the upper bound by the smoothed versions of H_2. Among other things, this allows us to use the fully quantum AEP (Theorem 2.4) and therefore to use the theorem directly on i.i.d. states and channels.

Theorem 3.8. *Let ρ^{AR} be a density operator, $\mathcal{T}^{A\to E}$ be any completely positive, trace-preserving superoperator, let $\omega^{A'E} = (\mathcal{T} \otimes \mathbb{I}^{A'})(\Phi^{AA'})$, and let $\varepsilon > 0$. Then,*

$$\int_{\mathbb{U}(A)} \left\|\mathcal{T}(U \cdot \rho^{AR}) - \omega^E \otimes \rho^R\right\|_1 dU \leqslant 2^{-\frac{1}{2}H_2^\varepsilon(A'|E)_\omega - \frac{1}{2}H_2^\varepsilon(A|R)_\rho + 1} + 12\varepsilon \tag{3.47}$$

where $\int \cdot dU$ denotes the integral over the Haar measure on all unitaries U^A.

Proof. Let $U_{\mathcal{T}}^{A\to CE}$ be a Stinespring extension of $\mathcal{T}$, and let $\widehat{\omega}^{A'E}$ be such that $d_F(\widehat{\omega}, \omega) \leqslant \varepsilon$ and $H_2(A'|E)_{\widehat{\omega}} = H_2^\varepsilon(A'|E)_\omega$. Also, let $\widehat{\rho}^{AR}$ be such that $d_F(\widehat{\rho}, \rho) \leqslant \varepsilon$ and $H_2(A|R)_{\widehat{\rho}} = H_2^\varepsilon(A|R)_\rho$. Write $\widehat{\omega} - \omega = \Delta_+ - \Delta_-$ where $\Delta_\pm \in \mathrm{Pos}(\mathsf{A}' \otimes \mathsf{E})$ have disjoint support. Since $d_F(\widehat{\omega}, \omega) \leqslant \varepsilon$, $\|\widehat{\omega} - \omega\|_1 \leqslant 2\varepsilon$ (see Lemma 2.2) and $\|\Delta_\pm\|_1 \leqslant 2\varepsilon$. We now define $\widehat{\omega}' := \omega - \Delta_-$. By the definition of H_2 and the fact that $\widehat{\omega}' \leqslant \widehat{\omega}$, we have that $H_2(A'|E)_{\widehat{\omega}'} \geqslant H_2(A'|E)_{\widehat{\omega}}$.

Let $P^C \leqslant \mathbb{I}^C$ be a positive semidefinite operator such that $\mathrm{Tr}_C[PU_{\mathcal{T}} \cdot \Phi^{AA'}] = \widehat{\omega}'$ (whose existence is guaranteed by Lemma I.3, since $\widehat{\omega}' \leqslant \omega$) and define $\widehat{\mathcal{T}}(\xi) = \mathrm{Tr}_C[PU_{\mathcal{T}} \cdot \xi]$. Then, using the previous theorem, we get

$$\begin{aligned} 2^{-\frac{1}{2}H_2^\varepsilon(A'|E)_\omega - \frac{1}{2}H_2^\varepsilon(A|R)_\rho} &\geqslant \int_{\mathbb{U}(A)} \left\|\widehat{\mathcal{T}}(U \cdot \widehat{\rho}^{AR}) - \widehat{\omega}'^E \otimes \widehat{\rho}^R\right\|_1 dU \\ &\geqslant \int_{\mathbb{U}(A)} \left\|\widehat{\mathcal{T}}(U \cdot \rho^{AR}) - \omega^E \otimes \rho^R\right\|_1 dU - 6\varepsilon. \end{aligned}$$

Now, note that by Lemma I.2, $\widehat{\mathcal{T}}(\xi) \leqslant \mathcal{T}(\xi)$ for any $\xi \in \mathrm{Pos}(\mathsf{A})$; note also that $\widehat{\mathcal{T}}$ is trace non-increasing while $\mathcal{T}$ is trace-preserving. Hence, Lemma I.1 applies to the above trace distance to give us:

$$\int_{\mathbb{U}(A)} \left\|\mathcal{T}(U \cdot \rho^{AR}) - \omega^E \otimes \rho^R\right\|_1 dU \leqslant 2^{-\frac{1}{2}H_2^\varepsilon(A'|E)_\omega - \frac{1}{2}H_2^\varepsilon(A|R)_\rho + 1} + 12\varepsilon$$

which concludes the proof. □

It is also possible to show that, with very high probability, the value of the left-hand side in Theorem 3.7 is very close to its expected value. This is shown in the next theorem. First, however, we must define the *Lipschitz constant* of a function:

Definition 3.2 (Lipschitz constant). *Let $f : \mathfrak{X} \to \mathfrak{Y}$ be a function from the set $\mathfrak{X}$ to the set $\mathfrak{Y}$ endowed with distance measures $d_{\mathfrak{X}}$ and $d_{\mathfrak{Y}}$. Then, the Lipschitz constant of f is defined as*

$$\sup_{x_1, x_2 \in \mathfrak{X}} \frac{d_{\mathfrak{Y}}(f(x_1), f(x_2))}{d_{\mathfrak{X}}(x_1, x_2)}.$$

If the above quantity is not bounded, the constant is not defined.

Theorem 3.9. *In the scenario described in the statement of Theorem 3.7, we have that*

$$\Pr\left\{\left\|\mathcal{T}(U \cdot \rho^{AR}) - \omega^E \otimes \rho^R\right\|_1 \geqslant 2^{-\frac{1}{2}H_2(A'|E)_\omega - \frac{1}{2}H_2(A|R)_\rho} + r\right\} \leqslant 2e^{-\frac{|A|r^2}{16K^2\|\rho^A\|_\infty}} \tag{3.48}$$

where $K = \max\{\|\mathcal{T}(X)\|_1 : X \in \mathrm{Herm}(\mathsf{A}), \|X\|_1 \leqslant 1\}$, and where the probability is computed over the choice of U.

Proof. This is a corollary of Corollary 4.4.28 in [AGZ09], which states that, for a c-Lipschitz function $f : \mathbb{U}(A) \to \mathbb{R}$,

$$\Pr_U\left\{|f(U) - \mathbb{E}f| \geqslant \delta\right\} \leqslant 2e^{-|A|\delta^2/4c^2}. \tag{3.49}$$

We are interested in the function $f(U) = \left\|\mathcal{T}(U \cdot \rho^{AR}) - \omega^E \otimes \rho^R\right\|_1$; we therefore need to bound its Lipschitz constant. Let $|\rho\rangle^{ABR}$ be a purification of ρ and, without loss of generality, assume $f(U) \geqslant f(V)$, and

$$\begin{aligned} f(U) - f(V) &= \left\|\mathcal{T}(U \cdot \rho^{AR}) - \omega^E \otimes \rho^R\right\|_1 - \left\|\mathcal{T}(V \cdot \rho^{AR}) - \omega^E \otimes \rho^R\right\|_1 \\ &\leqslant \left\|\mathcal{T}(U \cdot \rho^{AR}) - \mathcal{T}(V \cdot \rho^{AR})\right\|_1 \\ &\leqslant K \left\|U \cdot \rho^{ABR} - V \cdot \rho^{ABR}\right\|_1 \\ &\leqslant 2K \left\|(U - V)|\rho\rangle^{ABR}\right\|_2 \\ &= 2K \left\|\mathrm{op}_{BR\to A}((U - V)|\rho\rangle^{ABR})\right\|_2 \\ &= 2K \left\|(U - V)\,\mathrm{op}_{BR\to A}(|\rho\rangle^{ABR})\right\|_2 \\ &\leqslant 2K \left\|U - V\right\|_2 \left\|\mathrm{op}_{BR\to A}(|\rho\rangle^{ABR})\right\|_\infty \\ &= 2K \left\|U - V\right\|_2 \sqrt{\|\rho^A\|_\infty}. \end{aligned}$$

where the third inequality comes from Lemma I.4, the last inequality is an application of Lemma I.5, and $\|\cdot\|_\infty$ denotes the largest singular value of a matrix. Hence, the Lipschitz constant of f is upper bounded by $2K\sqrt{\|\rho^A\|_\infty}$ and the theorem follows. □

All of the constructions given in this section involve selecting unitary operators randomly according to the Haar measure. This wouldn't be very practical in real life: there is no guarantee that a matrix chosen this way could be implemented efficiently by a quantum circuit; in fact there is a extremely high chance that it wouldn't be. However, for most of the above theorem, there is a way out of this: apart from Theorem 3.9, the Haar measure can be replaced in all of the above theorems by a *unitary 2-design*, which is defined as follows:

Definition 3.3 (Unitary 2-design). *We call a finite set of unitaries $\mathfrak{D} \subset \mathbb{U}(A)$ a* unitary 2-design *if*

$$\frac{1}{|\mathfrak{D}|}\sum_{U\in\mathfrak{D}} U^{\otimes 2} \cdot M = \int_{\mathbb{U}(A)} (U^{\otimes 2} \cdot M) dU$$

for every $M \in \mathrm{L}(\mathsf{A}^{\otimes 2})$, where the integral is taken over the Haar measure on $\mathbb{U}(A)$.

Since all of the Theorems of this section with the exception of Theorem 3.9

only involve the Haar measure in integrals of this type, the Haar measure can be replaced by a unitary 2-design without affecting the rest of the theorem statements. An example of a unitary 2-design is the Clifford group (the group of unitaries that take Pauli operators to Pauli operators) [DLT02].

For Theorem 3.9, which makes us of the concentration properties of the Haar measure, we do not yet know how to replace the Haar measure by something constructive.

3.3 Corollaries of the decoupling theorem

As mentioned previously, many well-known results can be shown to be special cases of this theorem, including the Fully Quantum Slepian-Wolf theorem [ADHW06], as well as state merging [HOW07]. We present them here for completeness:

Corollary 3.10 (FQSW [ADHW06]). *Let ρ^{AR} be a mixed state, and $\mathsf{A} = \mathsf{A}_1 \otimes \mathsf{A}_2$. Then, we have that*

$$\int \left\| \mathrm{Tr}_{A_2}\left[U \cdot \rho^{AR}\right] - \pi^{A_1} \otimes \rho^{R} \right\|_1 dU \leqslant \sqrt{\frac{|A_1|}{|A_2|} 2^{-H_2(A|R)_\rho}} \tag{3.50}$$

where the integral is over the Haar measure on $\mathbb{U}(A)$, and π^{A_1} denotes the completely mixed state on A_1.

Proof. Consider the superoperator $\mathcal{T}^{A\to A_1}(\xi) = \mathrm{Tr}_{A_2}[\xi]$ and define $\omega^{A'A_1} = \mathrm{Tr}_{A_2}[\Phi^{AA'}]$, and then apply Theorem 3.7. It is easy to show that $\mathrm{Tr}[{\omega^{A'A_1}}^2] = 1/|A_2|$, from which the result follows. □

Corollary 3.11 (State merging). *Let ρ^{AR} be a mixed state, and let $\{M_i^{A\to E} : M_i \in$ $\mathrm{L}(\mathsf{A},\mathsf{E}), i \in \{1,\ldots,n\}\}$ be a set of measurement operators (i.e. $\sum_i M_i^\dagger M_i = \mathbb{I}^A$) such that each M_i is a rank-$|E|$ partial isometry. Then,*

$$\int \sum_i \left\| M_i U \cdot \rho^{AR} - \frac{\pi^E}{n} \otimes \rho^R \right\|_1 dU \leqslant \sqrt{|E| 2^{-H_2(A|R)_\rho}} \tag{3.51}$$

where we integrate over $\mathbb{U}(A)$.

For simplicity, we do not consider the case where $|A|$ is not divisible by $|E|$; the extension to the general case is straightforward.

Proof. Let X be an n-dimensional subsystem, and let $\mathcal{T}^{A\to EX}$ be a superoperator such that

$$\mathcal{T}(\sigma^A) = \sum_i |i\rangle\langle i|^X \otimes (M_i^{A\to E} \cdot \sigma^A) \tag{3.52}$$

and define the state $\omega^{A'EX} = \mathcal{T}(\Phi^{AA'})$. It can easily be shown that $\mathrm{Tr}[\omega^{A'EX^2}] = 1/n$, from which we get

$$\int \left\| \sum_i |i\rangle\langle i|^X \otimes (M_i U \cdot \rho^{AR}) - \pi^{EX} \otimes \rho^R \right\|_1 dU \leqslant \sqrt{\frac{|E||X|}{n} 2^{-H_2(A|R)_\rho}} \tag{3.53}$$

and the result follows. □

The next corollary is an intermediate lemma from the state merging paper [HOW07], and also forms the basis of [HHYW08]:

Corollary 3.12 (Random subspaces)**.** *Let* ρ^{AR} *be a mixed state and let* $V^{A\to E}$ *be a fixed rank-*$|E|$ *partial isometry. Then,*

$$\int \left\| \frac{|A|}{|E|} VU \cdot \rho^{AR} - \pi^E \otimes \rho^R \right\|_1 dU \leqslant \sqrt{|E| 2^{-H_2(A|R)_\rho}}. \tag{3.54}$$

Proof. Consider the superoperator $\mathcal{T}^{A\to E}$ such that $\mathcal{T}(\sigma^A) = \frac{|A|}{|E|} V \cdot \sigma^A$. The result follows immediately from Theorem 3.7 and the fact that $\mathcal{T}(\Phi^{AA'})$ is a pure state. □

One can also come up with interesting blends of the above. For instance, one can mix FQSW and the random subspaces theorem above. The operation $\mathcal{T}$ we will consider is the following: we apply a fixed unitary operator on Alice's share, then restrict her system to only a subspace by applying a fixed projector, and then trace out part of the remaining state. We call the post-restriction system

E, which is divided into two shares E_1 and E_2; we trace out E_2 and Alice is left with only E_1. The result is the following protocol, which will be presented and shown to be essentially optimal in [BCR09]:

Corollary 3.13. *Let ρ^{AR} be a density operator, and let $\mathcal{T}^{A\to E_1}$ be a completely positive superoperator such that*

$$\mathcal{T}(\sigma^A) = \frac{|A|}{|E|}\operatorname{Tr}_{E_2}[V^{A\to E_1E_2}\cdot\sigma^A]$$

where V is a partial isometry, and $\mathsf{E} = \mathsf{E}_1 \otimes \mathsf{E}_2$. Then,

$$\int_{\mathbb{U}(A)} \left\|\mathcal{T}(U\cdot\rho^{AR}) - \pi^{E_1}\otimes\rho^R\right\|_1 dU \leqslant \sqrt{\frac{|E_1|}{|E_2|}2^{-H_2(A|R)_\rho}} \tag{3.55}$$

where $\int \cdot dU$ denotes the integral over the Haar measure on all unitaries U^A.

Proof. Follows trivially from Theorem 3.7 and the calculation of $H_2(A'|E_1)_{\mathcal{T}(\Phi^{AA'})} = \log|E_2| - \log|E_1|$. □

3.4 Quantum coding theorems via decoupling

In this section, we use the theorems from Section 3.2 to derive a one-shot coding theorem for quantum channels. As explained earlier, our strategy will be to show that the complementary channel (i.e. the channel to the environment) completely breaks all correlations with a system that purifies the input. As a result, we get that Bob is able to reconstruct the message.

We will consider the following problem: Alice and Bob share a pure state ψ^{ABR}; Alice holds A, Bob holds B, and R is a reference system which purifies the state. Alice would like to send her share of the state to Bob through a single use of the quantum channel $\mathcal{N}^{A'\to C}$. To accomplish this, we need to find encoding and decoding superoperators $\mathcal{E}^{A\to A'}$ and $\mathcal{D}^{CB\to AB}$ such that

$$\|(\mathcal{D}\circ\mathcal{N}\circ\mathcal{E})(\psi) - \psi\|_1 \leqslant \varepsilon \tag{3.56}$$

Note that Buscemi and Datta [BD09] have considered a similar problem but without any B system already at Bob's. The following generalizes their result to the case where Bob already has a share of the system:

Theorem 3.14. *Let ψ^{ABR} be a pure state, $\mathcal{N}^{A'\to C}$ be any completely positive trace-preserving superoperator with Stinespring dilation $U_{\mathcal{N}}^{A'\to CE}$ and complementary channel $\bar{\mathcal{N}}^{A'\to E}$, let $\omega^{A''CE} = U_{\mathcal{N}} \cdot \sigma^{A''A'}$, where σ is any pure state and $\mathsf{A}'' \cong \mathsf{A}'$, and let $\varepsilon > 0$. Then, there exists an encoding partial isometry $V^{A\to A'}$ and a decoding superoperator $\mathcal{D}^{CB\to AB}$ such that*

$$\left\|\bar{\mathcal{N}}(V \cdot \psi^{AR}) - \omega^{E} \otimes \psi^{R}\right\|_1 \leqslant 2\sqrt{\delta_1} + \delta_2$$

and

$$\left\|(\mathcal{D} \otimes \mathcal{N})(V \cdot \psi^{ABR}) - \psi^{ABR}\right\|_1 \leqslant 2\sqrt{2\sqrt{\delta_1} + \delta_2}$$

where

$$\delta_1 = 3 \times 2^{\frac{1}{2}H^{\varepsilon}_{\max}(A)_{\psi} - \frac{1}{2}H^{\varepsilon}_{2}(A'')_{\omega} + 1} + 36\varepsilon$$
$$\delta_2 = 3 \times 2^{-\frac{1}{2}H^{\varepsilon}_{2}(A''|E)_{\omega} - \frac{1}{2}H^{\varepsilon}_{2}(A|R)_{\psi} + 1} + 36\varepsilon.$$

Here, δ_1 determines how closely the input ψ^A can be made to fit the target input distribution $\omega^{A''}$, whereas δ_2 depends on the difference between the amount of information that must be transmitted ($-H^{\varepsilon}_2(A|R)_{\psi}$) and the information-carrying capability of the channel ($H^{\varepsilon}_2(A''|E)_{\omega}$). See Figures 3.1 and 3.2 for an illustration of the theorem.

Proof. Let $W^{A\to A''}$ be any full-rank partial isometry, and consider the superoperator $\mathcal{T}^{A''\to E}$ defined as $\mathcal{T}(\xi) = |A''|\bar{\mathcal{N}}(\mathrm{op}_{A''\to A'}(|\sigma\rangle) \cdot \xi)$. Theorem 3.8 then tells us

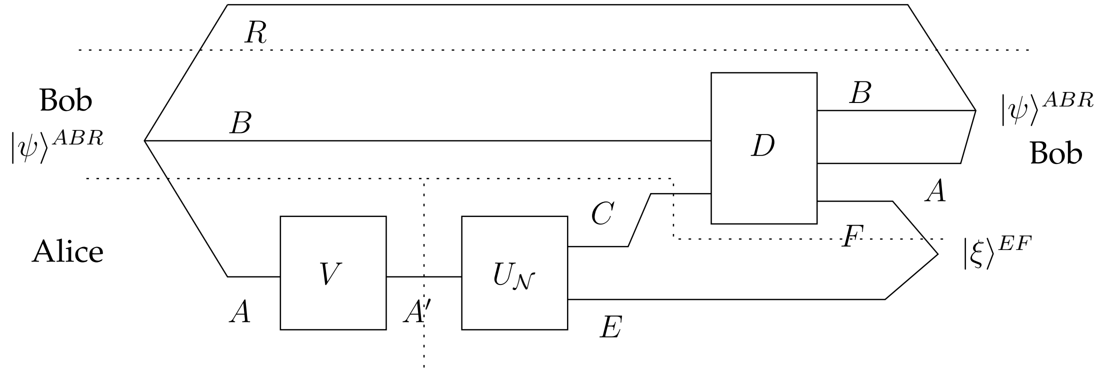

Figure 3.1: Diagram illustrating Theorem 3.14. Each line represents a quantum system, boxes represent isometries, and the horizontal axis represents the passage of time. Lines joined together at either end of the diagram represent pure states. Alice used V to encode her message A into the input to the channel A', and Bob uses the channel output C together with the B that he had since the beginning to decode A (and B) back. The decoder also produces a system F that purifies the environment.

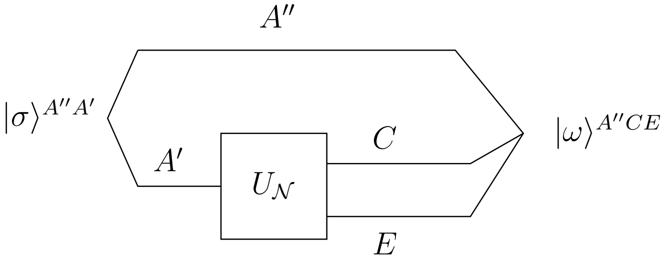

Figure 3.2: Diagram illustrating the states σ and ω in Theorem 3.14. Each line represents a quantum system, boxes represent isometries, and the horizontal axis represents the passage of time. Lines joined together at either end of the diagram represent pure states.

that

$$\begin{aligned}\int \left\| |A''| \bar{\mathcal{N}}(\mathrm{op}_{A''\to A'}(|\sigma\rangle) U^{A''} W \cdot \psi^{AR}) - \omega^E \otimes \psi^R \right\|_1 dU \\ = \int \left\| \mathcal{T}(UW \cdot \psi^{AR}) - \omega^E \otimes \psi^R \right\|_1 dU \\ \leqslant 2^{-\frac{1}{2}H_2^\varepsilon(A''|E)_\omega - \frac{1}{2}H_2^\varepsilon(A|R)_\psi + 1} + 12\varepsilon.\end{aligned}$$

We must now prove that there exists a U such that conjugating ψ by $\sqrt{|A''|}\,\mathrm{op}_{A''\to A'}(|\sigma\rangle)UW$ can be approximated by an isometry and for which the above inequality holds. For this, we use Theorem 3.8 again, with $\mathcal{E}^{A''\to G}$, $\mathcal{E}(\xi) = |A''|\,\mathrm{Tr}[\mathrm{op}_{A''\to A'}(|\sigma\rangle) \cdot \xi]$ (G is a dummy 1-dimensional system):

$$\begin{aligned}\int \left\| |A''|\,\mathrm{Tr}_{A'}[\mathrm{op}_{A''\to A'}(|\sigma\rangle) UW \cdot \psi^{ABR}] - \psi^{BR} \right\|_1 dU &\leqslant 2^{-\frac{1}{2}H_2^\varepsilon(A|BR)_\psi - \frac{1}{2}H_2^\varepsilon(A''|G)_{\mathcal{E}(\Phi)} + 1} + 12\varepsilon \\ &\leqslant 2^{\frac{1}{2}H_{\max}^\varepsilon(A)_\psi - \frac{1}{2}H_2^\varepsilon(A'')_\omega + 1} + 12\varepsilon\end{aligned}$$

where we have used Lemma 2.8 to establish that $\mathcal{E}(\Phi) = \omega$. Now, by Markov's inequality (Lemma I.7), we can choose a U such that

$$\begin{aligned}\int \left\| |A''|\,\mathrm{Tr}_{A'}[\mathrm{op}_{A''\to A'}(|\sigma\rangle) UW \cdot \psi^{ABR}] - \psi^{BR} \right\|_1 dU &\leqslant 3 \times 2^{\frac{1}{2}H_{\max}^\varepsilon(A)_\psi - \frac{1}{2}H_2^\varepsilon(A'')_\omega + 1} + 36\varepsilon \\ \int \left\| |A''| \bar{\mathcal{N}}(\mathrm{op}_{A''\to A'}(|\sigma\rangle) U^{A''} W \cdot \psi^{AR}) - \omega^E \otimes \psi^R \right\|_1 dU &\leqslant 3 \times 2^{-\frac{1}{2}H_2^\varepsilon(A''|E)_\omega - \frac{1}{2}H_2^\varepsilon(A|R)_\psi + 1} + 36\varepsilon.\end{aligned}$$

The first of these two inequalities allows us to use Uhlmann's theorem (Theorem 3.1) to find the encoding isometry: there exists a $V^{A\to A'}$ such that

$$\begin{aligned}\left\| |A''|\,\mathrm{op}_{A''\to A'}(|\sigma\rangle) UW \cdot \psi^{ABR} - V \cdot \psi^{ABR} \right\|_1 &\leqslant 2\sqrt{3 \times 2^{\frac{1}{2}H_{\max}^\varepsilon(A)_\psi - \frac{1}{2}H_2^\varepsilon(A'')_\omega + 1} + 36\varepsilon} \\ &= 2\sqrt{\delta_1}.\end{aligned}$$

Using the triangle inequality and the monotonicity of trace distance under CPTP maps, we get that

$$\left\|\bar{\mathcal{N}}(V\cdot\psi^{AR}) - \omega^{E}\otimes\psi^{R}\right\|_1 \leqslant 2\sqrt{\delta_1}+\delta_2. \tag{3.57}$$

To finish, we use Uhlmann's theorem again on this last inequality to get the decoder: there exists a partial isometry $D^{CB\to FAB}$ such that

$$\left\|DU_{\bar{\mathcal{N}}}V\cdot\psi^{ABR} - \xi^{EF}\otimes\psi^{ABR}\right\|_1 \leqslant 2\sqrt{2\sqrt{\delta_1}+\delta_2}$$

for some state ξ^{EF}. Finally, we trace over EF to get the theorem. □

We now turn to the case of memoryless channels used to transmit arbitrary quantum information with some entanglement assistance. This consists of the following special situation: the state ψ^{ABR} has the form $\Phi^{RM}\otimes\Phi^{\widetilde{A}B}$ where $M\widetilde{A}$ plays the role of A from the previous theorem, and the channel is $\left(\mathcal{N}^{A'\to C}\right)^{\otimes n}$. We will say that a pair (Q,E) is achievable if there exists a sequence of codes of length n, with encoders $\mathcal{E}_n^{M_n\widetilde{A}_n\to A'^n}$ and decoders $\mathcal{D}_n^{C^nB_n\to M_n}$, with $|M_n| = |R_n| = 2^{nQ}$, $|\widetilde{A}_n| = |B_n| = 2^{nE}$, such that

$$\lim_{n\to\infty}\left\|\left(\mathcal{D}_n\circ\mathcal{N}^{\otimes n}\circ\mathcal{E}_n\right)\left(\Phi^{R_nM_n}\otimes\Phi^{\widetilde{A}_nB_n}\right) - \Phi^{R_nM_n}\right\|_1 = 0.$$

A rate Q is achievable for entanglement-assisted transmission if there exists an $E\geqslant 0$ such that (Q,E) is achievable, and it is achievable for unassisted transmission if $(Q,0)$ is achievable. The capacity region is the closure of the convex hull of all achievable points.

The achievability of the coherent information for unassisted transmission was proven with increasing standards of rigour by Lloyd [Llo96], Shor [Sho02], and Devetak [Dev05]. Since then, several other proofs have been published; the proof given below shares some similarities with the one by Hayden, Horodecki, Yard and Winter [HHYW08]. The entanglement-assisted capacity was first given by Bennett, Shor, Smolin and Thapliyal [BSST02]. Theorem 3.15, which interpolates between those two results, can also be obtained by time-sharing between the completely assisted and unassisted protocols.

Theorem 3.15. *For any quantum channel $\mathcal{N}^{A'\to C}$, any pure state $\sigma^{AA'}$ with $\mathsf{A}' \cong \mathsf{A}$, the rate pair (Q, E) is achievable for quantum transmission with rate-limited entanglement assistance through $\mathcal{N}$ if*

$$Q + E < H(A)_\sigma \qquad\qquad \textit{and} \qquad\qquad Q - E < I(A\rangle C)_{\mathcal{N}(\sigma)}.$$

As a corollary, if we do not limit the rate of entanglement assistance, $Q < \frac{1}{2}I(A;C)_{\mathcal{N}(\sigma)}$ is achievable.

The first condition (that $Q + E < H(A)_\sigma$) says that both the quantum information to be transmitted and Alice's share of the EPR pairs must fit into the input to the channel, and the second condition says that the channel can carry $I(A\rangle C)$ qubits per transmission when no entanglement is used, but the rate can be "boosted" at the rate of one ebit per qubit until the first condition is saturated. If we saturate the first condition, then we get the entanglement-assisted capacity of $\frac{1}{2}I(A;C)_{\mathcal{N}(\sigma)}$.

Proof. The proof essentially consists of using the previous theorem on $\mathcal{N}^{\otimes n}$ and using the fully quantum AEP (Theorem 2.4) to bound the various conditional entropies. Let $U_{\mathcal{N}}^{A'\to CE}$ be a Stinespring dilation of $\mathcal{N}$, and let R and M be subsystems of dimension 2^{nQ}, M storing the quantum message Alice wants to transmit, and R being its purifying system. Likewise, let $\widetilde{A}$ and B be systems storing Alice's and Bob's part of the shared entanglement respectively, both of dimension 2^{nE}. Now, the input state we will consider is $\psi^{M\widetilde{A}BR} = \Phi^{RM} \otimes \Phi^{\widetilde{A}B}$, where $M\widetilde{A}$ play the role of A from the previous theorem. We are now in a position to use the previous theorem with ψ as the input state, $\mathcal{N}^{\otimes n}$ as the channel, and $\omega^{A^nC^nE^n} = U_{\mathcal{N}}^{\otimes n} \cdot \sigma^{\otimes n}$ to conclude that there exists an isometry $V^{M\widetilde{A}\to A'^n}$ and a

CPTP map $\mathcal{D}^{C^n B \to M}$ such that

$$\left\| (\mathcal{D} \circ \mathcal{N}^{\otimes n})(V \cdot \psi^{M\widetilde{A}BR}) - \psi^{ABR} \right\|_1 \leqslant 2\sqrt{2\sqrt{\delta_1} + \delta_2}$$

where

$$\delta_1 = 3 \times 2^{\frac{1}{2}H^{\varepsilon}_{\max}(M\widetilde{A})_{\psi} - \frac{1}{2}H^{\varepsilon}_{2}(A^n)_{\omega} + 1} + 36\varepsilon$$
$$\delta_2 = 3 \times 2^{-\frac{1}{2}H^{\varepsilon}_{2}(A^n|E^n)_{\omega} - \frac{1}{2}H^{\varepsilon}_{2}(M\widetilde{A}|R)_{\psi} + 1} + 36\varepsilon$$

Now we simply need to ensure that both δ_1 and δ_2 go down to zero as $n \to \infty$. Letting $\varepsilon = 2^{-kn}$, with k chosen according to the requirements of Theorem 2.4, we get

$$H^{\varepsilon}_{\max}(M\widetilde{A})_{\psi} \leqslant nQ + nE \tag{3.58}$$
$$H^{\varepsilon}_{2}(A^n)_{\omega} \geqslant nH(A)_{\sigma} - n\Delta_1 \tag{3.59}$$
$$H^{\varepsilon}_{2}(A^n|E^n)_{\omega} \geqslant nH(A|E)_{U_{\mathcal{N}}\cdot\sigma} - n\Delta_2 = nI(A\rangle C)_{\mathcal{N}(\sigma)} - n\Delta_2 \tag{3.60}$$
$$H^{\varepsilon}_{2}(M\widetilde{A}|R)_{\psi} \geqslant -nQ + nE. \tag{3.61}$$

where the Δ depend polynomially on k (and can be computed from the statement of the fully quantum AEP (Theorem 2.4)). Hence, we get that

$$\delta_1 = 3 \times 2^{\frac{n}{2}[Q+E-H(A)_{\sigma}+\Delta_1]} + 36\varepsilon$$
$$\delta_2 = 3 \times 2^{\frac{n}{2}[Q-E-I(A\rangle C)_{\mathcal{N}(\sigma)}+\Delta_2]} + 36\varepsilon.$$

Hence, for any pair (Q, E) such that $Q+E < H(A)_{\sigma}$ and $Q-E < I(A\rangle C)_{\mathcal{N}(\sigma)}$, there exists a protocol for which the error goes down to zero as $n \to \infty$.

To get the corollary on fully entanglement-assisted transmission, we simply add the two constraints to get $2Q < H(A)_{\sigma} + I(A\rangle C)_{\mathcal{N}(\sigma)} = I(A;C)_{\mathcal{N}(\sigma)}$ and hence $Q < \frac{1}{2}I(A;C)_{\mathcal{N}(\sigma)}$. □

3.5 Destroying correlations by adding classical randomness

In [GPW05], the authors discuss the following question: given a quantum state $\rho^{AB \otimes n}$, how many bits of classical randomness must be added to it to turn it into a product state? By "adding k bits of randomness" to a quantum state, we mean applying one of 2^k unitaries uniformly at random to either A^n or B^n. They find a method such that, as long as $k \geqslant n[I(A;B)_\rho + \delta]$, the distance to a decoupled state goes to zero as $n \to \infty$. This theorem constitutes one of the first direct operational intepretations of the quantum mutual information for arbitrary density operators. We can recover both this result and a one-shot version of it from Theorem 3.14:

Theorem 3.16. *Let $\rho^{AB} \in \mathrm{D}(\mathsf{A} \otimes \mathsf{B})$ be any quantum state. Then, for any $\varepsilon > 0$, there exists a set of 2^k unitaries U_i^A, with $k \leqslant 2H_{\max}^{\varepsilon}(A)_\rho + 4\log(1/\varepsilon) + 4$, and a $\xi^A \in \mathrm{D}(\mathsf{A})$ such that*

$$\left\| 2^{-k} \sum_{i=1}^{2^k} U_i^A \cdot \rho^{AB} - \xi^A \otimes \rho^B \right\|_1 \leqslant 3 \times 2^{\frac{1}{2}[H_{\max}^{\varepsilon}(A)_\rho - H_2^{\varepsilon}(A|B)_\rho - k + 2\log(1/\varepsilon) + 3]} + 2\sqrt{39\varepsilon} + 36\varepsilon.$$

Proof. Let P^A be a projector onto a subspace of A of dimension $D \geqslant \sqrt{2^k}$, and let $\{V_i^A\}_{i=1}^{2^k}$ be a set of Weyl operators (unitaries such that $\mathrm{Tr}[V_i^\dagger V_j] = 0$ for every $i \neq j$) on the support of P^A. Now, define the superoperator $\mathcal{T}^{A \to A}$ as

$$\mathcal{T}(\xi) = 2^{-k} \sum_{i=1}^{2^k} V_i \cdot \xi.$$

We can now apply Theorem 3.14 with $\mathcal{T}$ playing the role of $\bar{\mathcal{N}}$, ρ^{AB} as the input state and with input distribution $\sigma^{A''A} = \frac{|A|}{D} P^A \Phi^{A''A} P^A$ to get that there exists a unitary U^A such that

$$\left\| \mathcal{T}(U^A \cdot \rho^{AB}) - \omega^A \otimes \rho^B \right\|_1 \leqslant 2\sqrt{\delta_1} + \delta_2$$

where $\omega^{A''A} = \mathcal{T}(\sigma^{A''A})$, and

$$\delta_1 = 3 \times 2^{\frac{1}{2}H^{\varepsilon}_{\max}(A)_\rho - \frac{1}{2}\log D + 1} + 36\varepsilon$$
$$\delta_2 = 3 \times 2^{-\frac{1}{2}H^{\varepsilon}_2(A|B)_\rho + \frac{1}{2}\log D - \frac{1}{2}k + 1} + 36\varepsilon$$

since $H_2(A'')_\omega = \log D$ and $H_2(A''|A)_\omega = -\log D + k$. In other words,

$$\left\| 2^{-k} \sum_i V_i U^A \cdot \rho^{AB} - \omega^A \otimes \rho^B \right\|_1 \leqslant 2\sqrt{\delta_1} + \delta_2.$$

We can now define $U_i^A := V_i^A U^A$ to get our desired set of unitaries. All that is left to do is to let $\log D = H^{\varepsilon}_{\max}(A)_\rho + 2\log(1/\varepsilon) + 2$ to get that $\delta_1 = 39\varepsilon$, and hence, the theorem. □

One can also show using the fully quantum AEP (Theorem 2.4) that, for an i.i.d. state $\rho^{AB\otimes n}$, $H^{\varepsilon}_{\max}(A^n)_{\rho^{\otimes n}} - H^{\varepsilon}_2(A^n|B^n)_{\rho^{\otimes n}} \rightarrow n[H(A)_\rho - H(A|B)_\rho] = nI(A;B)_\rho$, which allows us to recover the theorem of [GPW05].

CHAPTER 4

QUANTUM CHANNELS WITH SIDE INFORMATION AT THE TRANSMITTER

4.1 Introduction

Consider the following problem: we have a noisy quantum memory device that can store n qubits and that contains a certain fraction of defective cells. The cells that do work can be modelled as depolarizing channels, but the defective ones always output $|0\rangle$. We can test which cells are defective before writing to the memory device, but this information is not necessarily available when reading from it. What is the best asymptotic rate at which we can store qubits reliably on this device? This problem can be generalized to any channel where the transmitter has access to side information about the channel state while the receiver does not.

The corresponding classical problem has been solved by Gel'fand and Pinsker in [GP80]. They consider channels modelled as a conditional probability distribution $p_{Y|XS}(y|x,s)$, $x \in \mathcal{X}, s \in \mathcal{S}, y \in \mathcal{Y}$, where x, y and s represent the input, output and state of the channel respectively. The channel state is i.i.d. and distributed according to $p_S(s)$. The encoder has access to the entire sequence of channel states ahead of time whereas the decoder does not. They have shown that the capacity of such a channel is given by

$$C = \max_{q_{USX} \in \mathcal{P}} [I(U;Y) - I(U;S)] \tag{4.1}$$

where $\mathcal{P}$ is the set of all probability distributions on $\mathcal{U} \times \mathcal{X} \times \mathcal{S}$ such that the marginal on $\mathcal{S}$ is equal to $p_S(s)$; $\mathcal{U}$ is an arbitrary set that can be chosen such that $|\mathcal{U}| \leqslant |\mathcal{X}| + |\mathcal{S}|$. The mutual informations are computed for the distribution $p_{Y|XS} \cdot q_{USX}$.

Here we shall generalize this result to quantum channels and potentially quantum side-information using the methods developed in Chapter 3. Namely, we will prove that the entanglement-assisted quantum capacity of quantum channels with side information at the transmitter has the same form as (4.1) and, a relatively rare fact in quantum information theory, has a single-letter converse. Along the way, we will prove a one-shot coding theorem as well as a coding theorem for quantum transmission with rate-limited entanglement assistance for such channels, both in the same spirit as those proven at the end of the last chapter.

4.2 Definition of quantum channels with side information at the transmitter

A quantum channel with side information at the transmitter is defined by a superoperator $\mathcal{N}^{A'S\to C}$ and a quantum state $|\phi\rangle^{SS'}$; this quantum state represents the side information. Alice has access to S' and can input a state of her choice into A'. One way to view this is to say that Alice shares entanglement with the channel itself. This framework allows us to consider both quantum and classical side information about the channel in a unified manner.

To illustrate this, consider the example of the depolarizing channel with defects given in the introduction. For this case, we can choose $|\phi\rangle$ to be $\sqrt{p}|00\rangle + \sqrt{1-p}|11\rangle$. The superoperator $\mathcal{N}$ then measures the S subsystem, and outputs $|0\rangle$ if the outcome is 0. If the outcome is 1, it applies the depolarizing channel to A' and sends the output to Bob.

In this chapter, we will first be interested in the following one-shot task: Alice and Bob initially share the A and B parts of the state ψ^{ABR}, and Alice would like to use the channel $(\mathcal{N}^{A'S\to C}, |\phi\rangle^{SS'})$ to send A to Bob. Hence, we want to ascertain the existence of an encoder $\mathcal{E}^{AS'\to A'}$ and decoder $\mathcal{D}^{CB\to AB}$ such that

$$\left\|(\mathcal{D}\circ\mathcal{N}\circ\mathcal{E})\left(\psi^{ABR}\otimes\phi^{SS'}\right)-\psi^{ABR}\right\|_1\leqslant\varepsilon$$

with a ε small enough for our purposes.

We will then specialize our one-shot theorem to the i.i.d. case, in which the state $\psi^{M\widetilde{A}BR}$ has the form $\Phi^{RM} \otimes \Phi^{\widetilde{A}B}$ where $M\widetilde{A}$ plays the role of A, and the channel is $\mathcal{N}^{\otimes n}$. We will say that a pair (Q, E) is achievable if there exists a sequence of codes of length n, with encoders $\mathcal{E}_n^{M_n\widetilde{A}_n S'^n \to A'^n}$ and decoders $\mathcal{D}_n^{C^n B_n \to M_n}$, with $|M_n| = |R_n| = 2^{nQ}$, $|\widetilde{A}_n| = |B_n| = 2^{nE}$, such that

$$\lim_{n\to\infty} \left\| \left(\mathcal{D}_n \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}_n\right) \left(\Phi^{R_n M_n} \otimes \Phi^{\widetilde{A}_n B_n} \otimes (\phi^{SS'})^{\otimes n}\right) - \Phi^{R_n M_n} \right\|_1 = 0.$$

A rate Q is achievable for entanglement-assisted transmission if there exists a $E \geqslant 0$ such that (Q, E) is achievable, and it is achievable for unassisted transmission if $(Q, 0)$ is achievable.

The capacity region is the closure of the convex hull of all achievable points.

The goal of this chapter is to establish the following theorems:

Theorem 4.1. *Let $(\mathcal{N}^{A'S\to C}, |\phi\rangle^{SS'})$ be a quantum channel with side-information at the transmitter, and let $\sigma^{AA'S}$ be any mixed state with $\sigma^S = \phi^S$. Then, any rate point (Q, E) such that*

$$Q + E < H(A|S)_\sigma \qquad \textit{and} \qquad Q - E < I(A\rangle C)_{\mathcal{N}(\sigma)}$$

is achievable for transmission with rate-limited entanglement assistance. As a corollary, any rate Q such that $Q < \frac{1}{2}[I(A;C)_{\mathcal{N}(\sigma)} - I(A;S)_\sigma]$ is achievable for entanglement-assisted transmission.

Theorem 4.2. *The entanglement-assisted quantum capacity of a quantum channel with side information at the transmitter $(\mathcal{N}^{A'S\to C}, |\phi\rangle^{SS'})$ is*

$$C = \sup_\sigma \left\{ \frac{1}{2} I(A;C)_\omega - \frac{1}{2} I(A;S)_\sigma \right\}. \tag{4.2}$$

The supremum is taken over all mixed states of the form $\sigma^{AA'S}$ where $\sigma^S = \phi^S$ and $\omega^{AC} = \mathcal{N}^{A'S\to C}(\sigma^{AA'S})$. In other words, the previous theorem with an i.i.d. input distribution is optimal for coding for entanglement-assisted i.i.d. channels.

This theorem also entails that the entanglement-assisted classical capacity of quantum channels with side information at the transmitter is

$$C = \sup_{\sigma} \left\{ I(A;C)_{\mathcal{N}(\sigma)} - I(A;S)_{\sigma} \right\} \tag{4.3}$$

via super-dense coding.

4.3 Direct coding theorem

We begin with the one-shot coding theorem:

Theorem 4.3. *Let ψ^{ABR} be a pure state, $(\mathcal{N}^{A'S\to C}, |\phi\rangle^{SS'})$ be any channel with side-information at the transmitter with $U_{\mathcal{N}}^{A'S\to CE}$ as Stinespring dilation, and let $\omega^{A''CED} = U_{\mathcal{N}} \cdot \sigma^{A''A'SD}$, where σ is any pure state with $\sigma^S = \phi^S$. Then, there exists a encoding CPTP map $\mathcal{E}^{AS'\to A'}$ and a decoding CPTP map $\mathcal{D}^{CB\to AB}$ such that*

$$\left\| (\mathcal{D}\circ\mathcal{N}\circ\mathcal{E})(\psi^{ABR}\otimes\phi^{SS'}))^{ABR} - \psi^{ABR} \right\|_1 \leqslant 2\sqrt{2\sqrt{\delta_1}+\delta_2}$$

where

$$\begin{aligned}
\delta_1 &= 3\times 2^{\frac{1}{2}H_{\max}^{\varepsilon}(A)_{\psi} - \frac{1}{2}H_2^{\varepsilon}(A''|S)_{\sigma}+1} + 36\varepsilon \\
\delta_2 &= 3\times 2^{-\frac{1}{2}H_2^{\varepsilon}(A''|ED)_{\omega} - \frac{1}{2}H_2^{\varepsilon}(A|R)_{\psi}+1} + 36\varepsilon.
\end{aligned}$$

Hence, to have a good code, one must ensure that both δ_1 and δ_2 are sufficiently small. Both of these quantities have natural interpretations: δ_1 characterizes the difference between how "big" the message is ($H_{\max}(A)_{\psi}$) and how much space there is in the input to the channel ($H_2^{\varepsilon}(A''|S)_{\sigma}$), and δ_2 depends on the difference between how hard the state is to transmit ($-H_2^{\varepsilon}(A|R)_{\psi}$) and how good the channel to the environment is at destroying correlations ($H_2^{\varepsilon}(A''|ED)_{\omega}$). See Figures 4.1 and 4.2 for illustrations of the protocol.

Proof. Let $W^{A\to A''}$ any full-rank partial isometry, and consider the superoperator

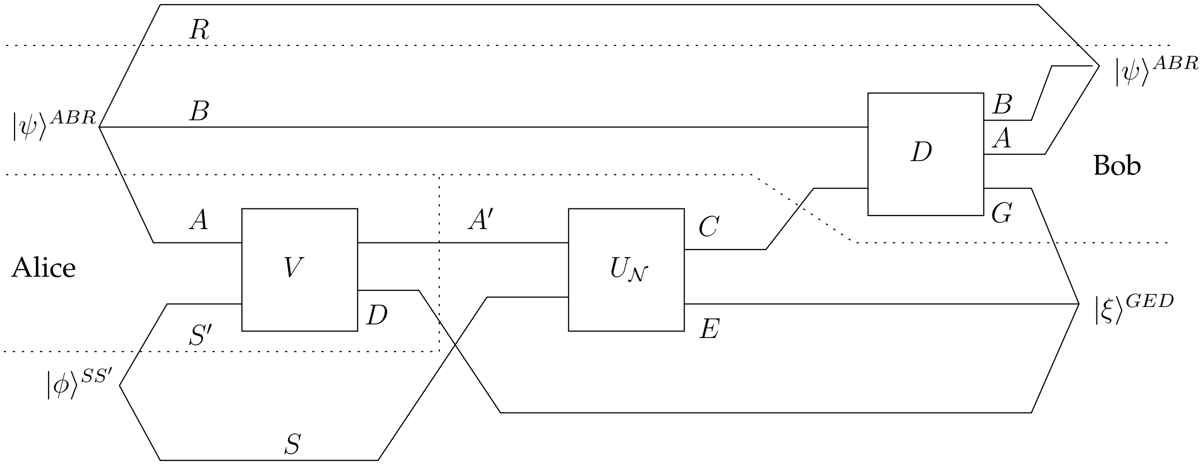

Figure 4.1: Diagram illustrating Theorem 4.3, with encoder, channel and decoder purified. Each line represents a quantum system, boxes represent isometries, and the horizontal axis represents the passage of time. Lines joined together at either end of the diagram represent pure states. V represents Alice's encoder: she uses the side information S' to encode the message A into the channel input A' and discards a system D. The decoder D takes the channel output C together with Bob's initial system B and produces A and B as output; the result being close to the initial state ψ.

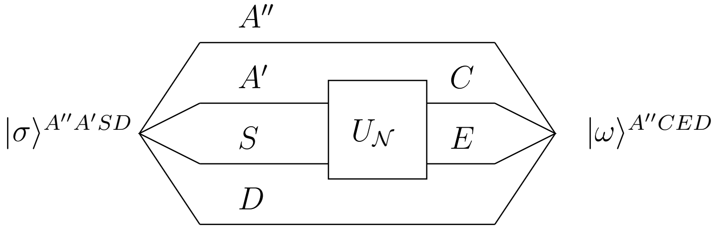

Figure 4.2: Diagram illustrating the states ω and σ which define the input distribution in Theorem 4.3. Each line represents a quantum system, boxes represent isometries, and the horizontal axis represents the passage of time. Lines joined together at either end of the diagram represent pure states.

$\mathcal{T}^{A''\to ED}$ defined as

$$\mathcal{T}(\xi) = |A''| \operatorname{Tr}_C[U_{\mathcal{N}} \operatorname{op}_{A''\to A'SD}(|\sigma\rangle) \cdot \xi].$$

Theorem 3.8 then tells us that

$$\int \left\|\mathcal{T}(UW \cdot \psi^{AR}) - \omega^{ED} \otimes \psi^{R}\right\|_1 dU \leqslant 2^{-\frac{1}{2}H_2^{\varepsilon}(A''|ED)_{\omega} - \frac{1}{2}H_2^{\varepsilon}(A|R)_{\psi}+1} + 12\varepsilon.$$

We must now prove that there exists a U such that conjugating ψ by $\sqrt{|A''|}\operatorname{op}_{A''\to A'SD}(|\sigma\rangle)UW$ can be approximated by an isometry of the form $V^{AS'\to A'D}$ acting on $\psi^{ABR} \otimes \phi^{SS'}$ and for which the above inequality holds. For this, we use Theorem 3.8 again, with $\mathcal{E}^{A''\to S}$, $\mathcal{E}(\xi) = |A''| \operatorname{Tr}_{A'D}[\operatorname{op}_{A''\to A'SD}(|\sigma\rangle) \cdot \xi]$:

$$\begin{aligned}\int \left\||A''| \operatorname{Tr}_{A'D}[\operatorname{op}_{A''\to A'SD}(|\sigma\rangle)UW \cdot \psi^{ABR}] - \psi^{BR} \otimes \phi^{S}\right\|_1 dU& \\ &\leqslant 2^{-\frac{1}{2}H_2^{\varepsilon}(A|BR)_{\psi} - \frac{1}{2}H_2^{\varepsilon}(A''|S)_{\omega}+1} + 12\varepsilon \\ &\leqslant 2^{\frac{1}{2}H_{\max}^{\varepsilon}(A)_{\psi} - \frac{1}{2}H_2^{\varepsilon}(A''|S)_{\omega}+1} + 12\varepsilon\end{aligned}$$

where we have used Lemma 2.8 to deduce that $\mathcal{E}(\Phi) = \omega$. We would like to have a $U^{A''}$ that satisfies both inequalities. We can do this using Markov's inequality (Lemma I.7): there exists a $U^{A''}$ such that:

$$\begin{aligned}\int \left\|\mathcal{T}(UW \cdot \psi^{AR}) - \omega^{ED} \otimes \psi^{R}\right\|_1 dU &\leqslant 3 \times 2^{-\frac{1}{2}H_2^{\varepsilon}(A''|E)_{\omega} - \frac{1}{2}H_2^{\varepsilon}(A|R)_{\psi}+1} + 36\varepsilon \\ &= \delta_2\end{aligned}$$

and

$$\begin{aligned}\int \left\||A''| \operatorname{Tr}_{A'D}[\operatorname{op}_{A''\to A'SD}(|\sigma\rangle)U \cdot \psi^{ABR}] - \psi^{BR} \otimes \phi^{S}\right\|_1 dU& \\ &\leqslant 3 \times 2^{\frac{1}{2}H_{\max}^{\varepsilon}(A)_{\psi} - \frac{1}{2}H_2^{\varepsilon}(A''|S)_{\omega}+1} + 36\varepsilon \\ &= \delta_1.\end{aligned}$$

This last condition allows us to use Uhlmann's theorem (Theorem 3.1) to find the encoding isometry: there exists a $V^{AS'\to A'D}$ such that

$$\left\| |A''| \operatorname{op}_{A''\to A'SD}(|\sigma\rangle) UW \cdot \psi^{ABR} - V \cdot (\psi^{ABR} \otimes \phi^{SS'}) \right\|_1 \leqslant 2\sqrt{\delta_1}$$

Using the triangle inequality and the monotonicity of trace distance under superoperators, we get that

$$\left\| (U_{\mathcal{N}} V \cdot (\psi^{AR} \otimes \phi^{SS'}))^{EDR} - \omega^{ED} \otimes \psi^R \right\|_1 \leqslant 2\sqrt{\delta_1} + \delta_2.$$

Finally, we use Uhlmann's theorem a second time to get a decoding partial isometry $D^{CB\to ABG}$:

$$\left\| \left(DU_{\mathcal{N}} V \cdot \left(\psi^{ABR} \otimes \phi^{SS'} \right) \right) - \xi^{GED} \otimes \psi^{ABR} \right\|_1 \leqslant 2\sqrt{2\sqrt{\delta_1} + \delta_2}$$

for some state ξ^{GED}. We then take a partial trace over GED inside the trace distance to get the theorem. □

We now move on to the memoryless case:

Theorem 4.4. *Let $(\mathcal{N}^{A'S\to C}, |\phi\rangle^{SS'})$ be a quantum channel with side-information at the transmitter, and let $\sigma^{AA'S}$ be any mixed state with $\sigma^S = \phi^S$. Then, any rate point (Q, E) such that*

$$Q + E < H(A|S)_\sigma \qquad \textit{and} \qquad Q - E < I(A\rangle C)_{\mathcal{N}(\sigma)}$$

is achievable for transmission with rate-limited entanglement assistance. As a corollary, any rate Q such that $Q < \frac{1}{2}[I(A;S)_\sigma - I(A;C)_{\mathcal{N}(\sigma)}]$ is achievable for entanglement-assisted transmission.

Again, the first condition corresponds to how closely we can make the input fit the target input distribution σ, and the second one is the limit imposed by the channel noise. Once again, we can trade ebits for qubits at a one-to-one ratio until we reach the limit imposed by the first condition.

Proof. The proof essentially consists of using the previous theorem on $\mathcal{N}^{\otimes n}$ and using the fully quantum AEP (Theorem 2.4) to bound the various conditional entropies. Let R and M be subsystems of dimension 2^{nQ}, M storing the quantum message Alice wants to transmit, and R being its purifying system. Likewise, let $\widetilde{A}$ and B be systems storing Alice's and Bob's parts of the shared entanglement respectively, both of dimension 2^{nE}. Now consider the input state $\psi^{M\widetilde{A}BR} = \Phi^{RM} \otimes \Phi^{\widetilde{A}B}$, where $M\widetilde{A}$ plays the role of A in the one-shot theorem. We now use the previous theorem with ψ as the input state, $\mathcal{N}^{\otimes n}$ as the channel, and $\omega^{A^nC^nE^nD^n} = \mathcal{N}^{\otimes n}(\sigma^{\otimes n})$ to conclude that there exist encoding and decoding CPTP maps $\mathcal{E}^{M\widetilde{A}S' \to A'^n}$ and $\mathcal{D}^{C^nB \to AB}$ such that

$$\left\| \left(\mathcal{D} \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}\right) \left(\psi^{M\widetilde{A}BR} \otimes (\phi^{SS'})^{\otimes n}\right) - \psi^{ABR} \right\|_1 \leqslant 2\sqrt{2\sqrt{\delta_1} + \delta_2}$$

with

$$\begin{aligned} \delta_1 &= 3 \times 2^{\frac{1}{2}H^{\varepsilon}_{\max}(M\widetilde{A})_{\psi} - \frac{1}{2}H^{\varepsilon}_{2}(A^n|S^n)_{\sigma^{\otimes n}} + 1} + 36\varepsilon \\ \delta_2 &= 3 \times 2^{-\frac{1}{2}H^{\varepsilon}_{2}(A^n|E^nD^n)_{\omega} - \frac{1}{2}H^{\varepsilon}_{2}(M\widetilde{A}|R)_{\psi} + 1} + 36\varepsilon. \end{aligned}$$

We can bound all the entropic terms above. Let $\varepsilon = 2^{-kn}$, with k chosen according to Theorem 2.4, and we get

$$H^{\varepsilon}_{\max}(M\widetilde{A})_{\psi} \leqslant nQ + nE \tag{4.4}$$

$$H^{\varepsilon}_{2}(A^n|S^n)_{\sigma^{\otimes n}} \geqslant nH(A|S)_{\sigma} - n\Delta_1 \tag{4.5}$$

$$H^{\varepsilon}_{2}(A^n|E^nD^n)_{\omega} \geqslant nH(A|ED)_{\mathcal{N}(\sigma)} - n\Delta_2 = nI(A\rangle C)_{\mathcal{N}(\sigma)} - n\Delta_2 \tag{4.6}$$

$$H^{\varepsilon}_{2}(M\widetilde{A}|R)_{\psi} \geqslant -nQ + nE \tag{4.7}$$

where the Δ depend polynomially on k (and can be computed from the statement of the fully quantum AEP (Theorem 2.4)).

Hence, as long as $Q+E < H(A|S)_{\sigma} - \Delta_1 - 2k$ and $Q-E < I(A\rangle C)_{\mathcal{N}(\sigma)} - \Delta_2 - k$, both δ_1 and δ_2 go down exponentially with n. The first condition corresponds to the fact that the message qubits and Alice's share of the entanglement must fit in the input, and the second condition means that the transmission rate minus the amount of entanglement must not exceed the coherent information.

To get the entanglement-assisted rate, we can simply add the two inequalities so as to eliminate E. The result is that $2Q < H(A|S)_\sigma + I(A\rangle C)_{\mathcal{N}(\sigma)}$ and a simple calculation reveals this to be equivalent to $Q < \frac{1}{2}[I(A;C)_{\mathcal{N}(\sigma)} - I(A;S)_\sigma]$.

□

4.4 Optimality for entanglement-assisted coding

We shall now prove that the previous theorem is optimal for entanglement-assisted coding. In other words, for any achievable rate Q for entanglement-assisted transmission, there exists a state $\sigma^{AA'S}$ as in Theorem 4.2 for which $Q = \frac{1}{2}I(A;C)_{\mathcal{N}(\sigma)} - \frac{1}{2}I(A;S)_\sigma$. This part is essentially the same as in [GP80], with a few adaptations to the quantum case. In particular, one must pay close attention to which state the various mutual informations are defined on, since we will be dealing with states where only some fraction of the n instances of the channel has been applied.

Theorem 4.5. *The entanglement-assisted quantum capacity of a quantum channel with side information at the transmitter* $(\mathcal{N}^{A'S\to C}, |\phi\rangle^{SS'})$ *is*

$$C = \sup_\sigma \left\{ \frac{1}{2}I(A;C)_\omega - \frac{1}{2}I(A;S)_\sigma \right\}. \tag{4.8}$$

The supremum is taken over all mixed states of the form $\sigma^{AA'S}$ *where* $\sigma^S = \phi^S$ *and* $\omega^{AC} = \mathcal{N}^{A'S\to C}(\sigma^{AA'S})$.

Proof. The achievability of this rate follows directly from Theorem 4.4. We therefore now need to prove that one cannot go above this rate. First, let $\mathcal{E}^{M_n\widetilde{A}_nS'^n\to A'_n}$ and $\mathcal{D}^{C^nB_n\to M_nB_n}$ be the encoder and the decoder respectively of an arbitrary code of block size n with $\log|R_n| = \log|M_n| = nQ$ such that

$$\left\| \left(\mathcal{D} \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}\right) \left(\Phi^{R_n M_n} \otimes \Phi^{\widetilde{A}_n B_n} \otimes (\phi^{SS'})^{\otimes n}\right) - \Phi^{R_n M_n} \right\|_1 \leqslant \varepsilon,$$

let $\sigma^{R_n B_n A'^n S^n} = \mathcal{E}(\Phi^{R_n M_n} \otimes \Phi^{\widetilde{A} B_n} \otimes (\phi^{SS'})^{\otimes n})$ and $\omega^{R_n B_n C^n} = \mathcal{N}^{\otimes n}(\sigma)$. Then, by Fannes' inequality (Theorem I.9) and the monotonicity of the mutual information (see Section 2.4.2) we must have that

$$I(R_n; C^n B_n)_\omega \geqslant 2n(Q - d(\varepsilon, n)) \tag{4.9}$$

where $d(\varepsilon, n) := \frac{3\varepsilon Q}{2} + \frac{3\varepsilon \log \varepsilon}{n}$. Notice that

$$\begin{aligned} I(R_n; B_n C^n)_\omega &= I(B_n; R_n)_\omega + I(R_n; C^n | B_n)_\omega && (4.10)\\ &= I(R_n; C^n | B_n)_\omega && (4.11)\\ &\leqslant I(R_n B_n; C^n)_\omega && (4.12) \end{aligned}$$

where (4.11) is due to the fact that R_n and B_n are independent. Combining this with $I(R_n B_n; S^n)_\sigma = 0$, we have

$$I(R_n B_n; C^n)_\omega - I(R_n B_n; S^n)_\sigma \geqslant 2n(Q - d(\varepsilon, n)). \tag{4.13}$$

We will now introduce a few shorthands which will make the notation considerably less cumbersome: we will write C^i instead of $C_1, \ldots, C_i$ and C_i^j instead of $C_i, \ldots, C_j$, and likewise for S. Define also

$$\begin{aligned} X(i) &:= R_n B_n C^{i-1} S_{i+1}^n && (4.14)\\ Y(i) &:= R_n B_n S_{i+1}^n && (4.15) \end{aligned}$$

Note that these are nothing more than groupings of subsystems. We also define the following sequence of states:

$$\omega(i) := (\mathcal{N}^{\otimes i} \otimes \mathbb{I}^{\otimes n-i})(\sigma) \tag{4.16}$$

In other words, $\omega(i)$ is the result of applying the first i instances of the channel to the state σ.

We shall now prove the inequality

$$\begin{aligned} I(R_nB_n; C^n)_\omega - I(R_nB_n; S^n)_\sigma & \\ & \leqslant \sum_{i=1}^{n} \left\{ I(X(i); C_i)_{\omega(i)} - I(X(i); S_i)_{\omega(i-1)} \right\}. \end{aligned} \tag{4.17}$$

Since each term in this sum is of the form $I(A;C)_{\mathcal{N}(\sigma)} - I(A;S)_\sigma$ for some $\sigma^{AA'S}$, the highest term is achievable by the direct coding theorem and therefore there exists a state for which $Q \leqslant I(A;C)_{\mathcal{N}(\sigma)} - I(A;S)_\sigma$. This allows us to conclude the theorem.

We now proceed in exactly the same way as in [GP80] to establish (4.17): we consider the inequality

$$\begin{aligned} I(Y(i); C^i)_{\omega(i)} - I(Y(i); S^i)_{\omega(i-1)} & \\ \leqslant \left[I(Y(i-1); C^{i-1})_{\omega(i-1)} - I(Y(i-1); S^{i-1})_{\omega(i-2)} \right] & \\ + \left[I(X(i); C_i)_{\omega(i)} - I(X(i); S_i)_{\omega(i-1)} \right]. & \end{aligned} \tag{4.18}$$

Summing up all these inequalities from $i = 2$ to $i = n$, we obtain (4.17), since $Y(n) = R_nB_n$ and $Y(1) = X(1)$.

Now, to prove (4.18), we use the following identities which follow from the definitions of $X(i)$ and $Y(i)$ and from basic properties of the mutual information:

$$I(Y(i);C^i)_{\omega(i)} = I(Y(i);C^{i-1})_{\omega(i)} + I(Y(i);C_i|C^{i-1})_{\omega(i)} \tag{4.19}$$

$$I(Y(i);S^i)_{\omega(i-1)} = I(Y(i);S_i)_{\omega(i-1)} + I(Y(i);S^{i-1}|S_i)_{\omega(i-1)} \tag{4.20}$$

$$I(Y(i-1);S^{i-1})_{\omega(i-1)} = I(Y(i);S^{i-1}|S_i)_{\omega(i-1)} \tag{4.21}$$

$$I(Y(i-1);C^{i-1})_{\omega(i-1)} = I(S_i;C^{i-1})_{\omega(i-1)} + I(Y(i);C^{i-1}|S_i)_{\omega(i-1)} \tag{4.22}$$

$$I(X(i);C_i)_{\omega(i)} = I(C^{i-1};B_i)_{\omega(i)} + I(Y(i);C_i|C^{i-1})_{\omega(i)} \tag{4.23}$$

$$I(X(i);S_i)_{\omega(i-1)} = I(C^{i-1};S_i)_{\omega(i-1)}d + I(Y(i);S_i|C^{i-1})_{\omega(i-1)}. \tag{4.24}$$

Substituting these into (4.18) and using the identity

$$I(A;B) - I(A;B|C) = I(A;C) - I(A;C|B) \tag{4.25}$$

which holds on any mixed state ρ^{ABC}, we get that the difference between the right-hand side and the left-hand side of (4.18) is $I(C^{i-1};C_i)_{\omega(i)}$, which is always nonnegative. This concludes the proof. □

4.5 Discussion

This result further strengthens the parallel between classical information theory problems and their entanglement-assisted quantum counterparts. Indeed, the capacity formula (4.2) has the same form as the classical version (4.1); the same phenomenon arises in the case of the entanglement-assisted capacities of regular point-to-point channels [BSST02], multiple-access channels [HDW08], and, for the best coding theorem we know, broadcast channels (see Chapter 5). It would be particularly interesting to have a systematic way in which classical coding theorems could be transformed into entanglement-assisted quantum protocols as it would enable us to import much larger classes of results from classical information theory into the quantum world.

Returning to our result, there is one remaining issue that one would like to solve in order to have a fully satisfactory characterization of the achievable rate region: we currently have no upper bound on the dimension of the A system needed to achieve the capacity in expression (4.2). Thus, despite having a single-letter converse, we unfortunately do not have a way to compute the capacity. In the classical case, it is possible to use Carathéodory's theorem [Car07] to bound the cardinality of $\mathcal{U}$ in the optimal input distribution. However, in the quantum case, this approach fails due to the fact that the quantum conditional entropy cannot in general be expressed as $H(A|B) = \sum_b p(b)H(A|B = b)$. On the other hand, there is little reason to believe that large dimensions are necessary to achieve the optimal rate, but we do not know how to prove that this is not the case. In fact, one encounters a very similar difficulty when trying to calculate the squashed entanglement [CW04] of a particular state since we have no bound on the size of the subsystem we need to condition on. We therefore leave this issue as an open problem.

One might also wonder about a related problem: whether the capacity can in general be achieved by optimizing only over pure states $\sigma^{AA'S}$. This would imply an upper bound on $|A|$. However, one can show that this cannot be the case: take, for example, a qubit-to-qubit channel which applies one of the four Pauli operations with equal probability, but where S tells the transmitter which one of the four operations is applied. The capacity of such a channel is clearly one qubit per transmission. Suppose that this rate is achievable using a pure state $\sigma^{AA'S}$. Then, we must have $\frac{1}{2}I(A;C)_{\mathcal{N}(\sigma)} = 1$ (since C is two-dimensional) and therefore $\frac{1}{2}I(A;S)_\sigma = 0$. However, this last equation together with the fact that σ is pure implies that the purification of S must be entirely in A'. This is impossible since S is maximally mixed over a four-dimensional system whereas A' is two-dimensional, and hence the optimal σ cannot be pure.

CHAPTER 5

QUANTUM BROADCAST CHANNELS

5.1 Introduction

Discrete memoryless broadcast channels are channels with one sender and n receivers, modelled using an input set $\mathfrak{X}$, output sets $\mathfrak{Y}_1, \ldots, \mathfrak{Y}_n$, and a probability transition matrix $p(y_1, \ldots, y_n|x)$. When the transmitter selects the input symbol $x_0 \in \mathfrak{X}$, the output at the receivers is distributed according to $p(y_1, \ldots, y_n|x = x_0)$. These can represent, for instance, a radio tower broadcasting a signal to many receivers, each of whom experiences different signal corruption due being closer or further away from the tower, or due to the proximity of buildings. There are many natural tasks that one may want to perform using these channels, such as sending common messages to all the users, sending separate information to each user, sending data to each user privately, or some combination of these tasks. Here we shall focus only on sending separate data to two different receivers that we will call Bob 1 and Bob 2.

One should note in passing that while this definition of broadcast channels is standard in electrical engineering, it may strike computer scientists (and particularly cryptographers) as bizarre. Indeed, computer scientists are used to defining broadcast channels as a *task* to be performed: send the same message to multiple parties, with no notion of noise. Here we think of broadcast channels more as physical objects: a physical channel with one input and multiple outputs, with which we may want to perform a number of different tasks.

Broadcast channels were first introduced by Cover in [Cov72], where he suggested that it may be possible to use them more efficiently than by timesharing between the different users. Since then, several results concerning broadcast channels have been found, such as the capacity of degraded broadcast channels (see, for example, [CT91]). Furthermore, these results form the basis of many protocols that are currently used in real multiuser systems, such as cellphone networks.

The best achievable rate region known for general classical broadcast channels is due to Marton [Mar79]: given a probability distribution $p(x, u_1, u_2) = p(u_1, u_2)p(x|u_1, u_2)$, the following rate region is achievable for the general two-user broadcast channel $p(y_1, y_2|x)$:

$$\begin{aligned} 0 \leqslant R_1 &\leqslant I(U_1; Y_1) \\ 0 \leqslant R_2 &\leqslant I(U_2; Y_2) \\ R_1 + R_2 &\leqslant I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2) \end{aligned} \tag{5.1}$$

It is conjectured that this characterizes the capacity region of general two-receiver broadcast channels, but despite considerable efforts, no one has been able to prove a converse theorem.

The quantum generalization of broadcast channels was first studied in [AS00] and [YHD06] as part of a recent effort to develop a network quantum information theory [DH06a, DH06b, YDH05, LOW06, HIN$^+$06, Win01, Kli01, SVW05, GS07]. In [YHD06], the authors derived three classes of results, the first one about channels with a classical input and quantum outputs, the second one about sending a common classical message while sending quantum information to one receiver, and the third about sending qubits to one receiver while establishing a GHZ state with the two receivers.

In this chapter, we study quantum broadcast channels using the general techniques developed in this thesis. We look at the case where Alice initially shares a tripartite state $\psi_1^{A_1B_1R_1}$ with Bob 1 and a reference, and would like to transfer her share A_1 to Bob 1 using the broadcast channel. She would simultaneously like to do the same with $\psi_2^{A_2B_2R_2}$ and Bob 2. We first give a one-shot theorem for this task, and then specialize it to the i.i.d. case (i.e. the channel has the form $\mathcal{N}^{\otimes n}$) in which ψ_1 consists of a maximally entangled pair between Alice and a reference, and separate maximally entangled pairs between Alice and Bob 1;

the same goes for ψ_2 and Bob 2. This corresponds to transmitting qubits with rate-limited entanglement assistance to Bob 1 and Bob 2 simultaneously over the broadcast channel. When we use the maximum possible amount of entanglement assistance, we recover a quantum version of Marton's region. On the other hand, when no entanglement assistance at all is used, the rate region does not appear to have any independent constraint on the sum rate; the information going to Bob 1 and to Bob 2 appear to "talk past each other". Interestingly, it turns out that the same phenomenon occurs in the classical scenario of Gaussian multiple-antenna broadcast channels (a particular type of classical broadcast channels) with confidential messages [LLPS09]. This is perhaps not so surprising, since private classical communication tends to be the closest classical parallel to quantum communication, in which one must inherently keep the information private from the environment.

We then prove a regularized converse for the fully entanglement-assisted case, and give an example of a channel for which the single-letter region is optimal.

5.2 Definitions

Here we define the various concepts needed for this chapter.

Definition 5.1 (Quantum broadcast channel)**.** *A quantum broadcast channel is a CPTP map with more than one subsystem as its output, and whose outputs are held by separate receivers.*

In the one-shot case, we will be interested in the following situation: the initial state is $\psi_1^{A_1B_1R_1} \otimes \psi_2^{A_2B_2R_2}$, where A_1 and A_2 are held by Alice, B_1 and B_2 by Bob 1 and Bob 2 respectively, and R_1 and R_2 are reference systems making the states pure. Alice wants to use the broadcast channel $\mathcal{N}^{A'\to C_1C_2}$ to send A_1 to Bob 1 and A_2 to Bob 2 (of course, Bob 1 gets C_1 and Bob 2 gets C_2). Hence, we will need to assert the existence of an encoding superoperator $\mathcal{E}^{A_1A_2\to A'}$ and decoders $\mathcal{D}_1^{B_1C_1\to A_1B_1}$ and $\mathcal{D}_2^{B_2C_2\to A_2B_2}$ such that

$$\left\|((\mathcal{D}_1 \otimes \mathcal{D}_2) \circ \mathcal{N} \circ \mathcal{E}) \left(\psi_1^{A_1B_1R_1} \otimes \psi_2^{A_2B_2R_2}\right) - \psi_1^{A_1B_1R_1} \otimes \psi_2^{A_2B_2R_2}\right\|_1 \leqslant \delta$$

for some δ that we find suitably small. Note here that the two decoders $\mathcal{D}_1$ and $\mathcal{D}_2$ commute and can be applied in parallel.

In the i.i.d. case, we will want to use the broadcast channel $\mathcal{N}^{A'\to C_1C_2}$ n times, to transmit separate arbitrary quantum data to Bob 1 and to Bob 2, with separate preshared entanglement with Bob 1 and Bob 2. In other words, $\psi_1^{M_1\widetilde{A}_1B_1R_1} = \Phi^{R_1M_1} \otimes \Phi^{\widetilde{A}_1B_1}$ where $M_1\widetilde{A}_1$ plays the role A_1; likewise for ψ_2. Now, for a given protocol for this task, we define the transmission rate to Bob 1 (resp. Bob 2) Q_1 (resp. Q_2) as $\frac{1}{n}\log|M_1|$ (resp. $\frac{1}{n}\log|M_2|$) and the entanglement consumption rate to Bob 1 (resp. Bob 2) as $E_1 = \frac{1}{n}\log|\widetilde{A}_1|$ (resp. $E_2 = \frac{1}{n}\log|\widetilde{A}_2|$).

We say that a four-tuple (Q_1, Q_2, E_1, E_2) is achievable if there exists a sequence of encoders $\mathcal{E}_n^{M_{1,n}\widetilde{A}_{1,n}\to A'^n}$ and decoders $\mathcal{D}_{1,n}^{C_1^nB_{1,n}\to M_{1,n}}$ and $\mathcal{D}_{2,n}^{C_2^nB_{2,n}\to M_{2,n}}$, such that

$$\lim_{n\to\infty}\left\|\left((\mathcal{D}_{1,n} \otimes \mathcal{D}_{2,n}) \circ \mathcal{N}^{\otimes n} \circ \mathcal{E}_n\right)\left(\Phi^{R_1M_1} \otimes \Phi^{\widetilde{A}_1B_1} \otimes \Phi^{R_2M_2} \otimes \Phi^{\widetilde{A}_2\otimes B_2}\right) - \Phi^{R_1M_1} \otimes \Phi^{R_2M_2}\right\|_1 = 0$$

with

$$\begin{aligned}
Q_1 &= \lim_{n\to\infty}\frac{1}{n}\log|M_{1,n}| = \lim_{n\to\infty}\frac{1}{n}\log|R_{1,n}|\\
Q_2 &= \lim_{n\to\infty}\frac{1}{n}\log|M_{2,n}| = \lim_{n\to\infty}\frac{1}{n}\log|R_{2,n}|\\
E_1 &= \lim_{n\to\infty}\frac{1}{n}\log|\widetilde{A}_{1,n}| = \lim_{n\to\infty}\frac{1}{n}\log|B_{1,n}|\\
E_2 &= \lim_{n\to\infty}\frac{1}{n}\log|\widetilde{A}_{2,n}| = \lim_{n\to\infty}\frac{1}{n}\log|B_{2,n}|.
\end{aligned}$$

5.3 Direct coding theorem

We start by proving the one-shot version of the protocol. First, however, we prove a simple technical lemma:

Lemma 5.1. *If we have density operators $\rho^{ABC}, \sigma^A, \omega^{BC}, \tau^{AB}, \eta^C$ such that*

$$\begin{aligned}\left\|\rho^{ABC} - \sigma^A \otimes \omega^{BC}\right\|_1 &\leqslant \varepsilon_1 \\ \left\|\rho^{ABC} - \tau^{AB} \otimes \eta^C\right\|_1 &\leqslant \varepsilon_2\end{aligned}$$

then $\left\|\rho^{ABC} - \sigma^A \otimes \tau^B \otimes \eta^C\right\|_1 \leqslant 2\varepsilon_1 + \varepsilon_2$.

Proof.

$$\begin{aligned}\left\|\rho^{ABC} - \sigma^A \otimes \tau^B \otimes \eta^C\right\|_1 &\leqslant \left\|\rho^{ABC} - \sigma^A \otimes \omega^{BC}\right\|_1 \\ &\quad + \left\|\sigma^A \otimes \omega^{BC} - \sigma^A \otimes \tau^B \otimes \eta^C\right\|_1 \\ &= \varepsilon_1 + \left\|\omega^{BC} - \tau^B \otimes \eta^C\right\|_1 \\ &\leqslant \varepsilon_1 + \left\|\omega^{BC} - \rho^{BC}\right\|_1 + \left\|\rho^{BC} - \tau^B \otimes \eta^C\right\|_1 \\ &\leqslant 2\varepsilon_1 + \varepsilon_2\end{aligned}$$

where the first two inequalities are applications of the triangle inequality, and the equality is due to the fact that $\|A\|_1 = \|\sigma \otimes A\|_1$ for any operator A and density matrix σ. □

Theorem 5.2. *For any quantum broadcast channel $\mathcal{N}^{A' \to C_1C_2}$, any pair of pure quantum states $\psi_1^{A_1B_1R_1}$ and $\psi_2^{A_2B_2R_2}$, any pure quantum state $\sigma^{A_1''A_2''A'D}$ and any $\varepsilon > 0$, there exists an encoding superoperator $\mathcal{E}^{A_1A_2 \to A'}$ and decoding superoperators $\mathcal{D}_1^{C_1B_1 \to A_1B_1}$ and $\mathcal{D}_2^{C_2B_2 \to A_2B_2}$ such that*

$$\begin{aligned}&\left\|((\mathcal{D}_1 \otimes \mathcal{D}_2) \circ \mathcal{N} \circ \mathcal{E})(\psi_1^{A_1B_1R_1} \otimes \psi_2^{A_2B_2R_2}) - \psi_1^{A_1B_1R_1} \otimes \psi_2^{A_2B_2R_2}\right\|_1 \\ &\qquad\qquad \leqslant 4\sqrt{2\sqrt{\delta_{\text{enc}}} + \delta_1} + 2\sqrt{2\sqrt{\delta_{\text{enc}}} + \delta_2}\end{aligned}$$

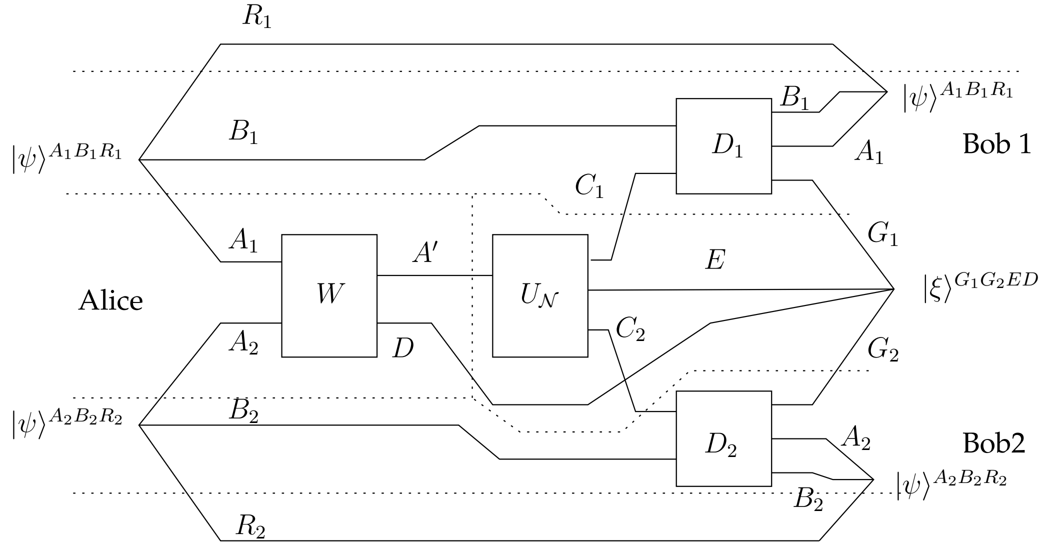

Figure 5.1: Diagram illustrating Theorem 5.2, with encoder, channel and decoders purified. Each line represents a quantum system, boxes represent isometries, and the horizontal axis represents the passage of time. Lines joined together at either end of the diagram represent pure states. W represents Alice's encoder: she encodes the messages A_1 and A_2 into the channel input A' and discards a system D. The decoders D_1 and D_2 take the channel outputs C_1 and C_2 together with Bob 1 and 2's initial systems B_1 and B_2 and produce A_1B_1 and A_2B_2 as output; the result being close to the initial state ψ.

where

$$\delta_{\text{enc}} = 4 \times 2^{\frac{1}{2}H^{\varepsilon}_{\max}(A_1)_{\psi_1} - \frac{1}{2}H^{\varepsilon^2/20}_{\min}(A_1''|A_2'')_\sigma + 1} \\ + 5 \times 2^{\frac{1}{2}H^{\varepsilon}_{\max}(A_2)_{\psi_2} - \frac{1}{2}H^{\varepsilon}_{\min}(A_2'')_\sigma + 1} + 108\varepsilon$$

and

$$\delta_1 \leqslant 4 \times 2^{-\frac{1}{2}H^{\varepsilon^2/20}_{\min}(A_1''|EDA_2''C_2)_{U_{\mathcal{N}}\cdot\sigma} - \frac{1}{2}H^{\varepsilon}_{\min}(A_1|R_1)_{\psi_1} + 1} + 48\varepsilon$$

$$\delta_2 \leqslant 5 \times 2^{-\frac{1}{2}H^{\varepsilon^2/16}_{\min}(A_2''|EDA_1''C_1)_{U_{\mathcal{N}}\cdot\sigma} - \frac{1}{2}H^{\varepsilon}_{\min}(A_2|R_2)_{\psi_2} + 1} + 60\varepsilon.$$

See Figure 5.1 for an illustration of the purified version of the protocol.

Proof. Let $U_{\mathcal{N}}^{A' \to C_1C_2E}$ and $W^{A_1A_2 \to A'D}$ be Stinespring dilations of $\mathcal{N}$ and $\mathcal{E}$ respectively (it will turn out later that the encoder indeed only needs to discard a system of size D). To be able to assert the existence of the two decoders, we need to ensure that the two associated decoupling conditions are fulfilled. Those two conditions stipulate that there must exist states ξ_1 and ξ_2 such that

$$\left\| (U_{\mathcal{N}} W \cdot (\psi_1 \otimes \psi_2))^{R_1 R_2 C_2 B_2 E D} - \psi^{R_1} \otimes \xi_1^{R_2 C_2 B_2 E D} \right\|_1$$
$$\left\| (U_{\mathcal{N}} W \cdot (\psi_1 \otimes \psi_2))^{R_2 R_1 C_1 B_1 E D} - \psi^{R_2} \otimes \xi_2^{R_1 C_1 B_1 E D} \right\|_1$$

are both appropriately small.

To ensure this, let $V_1^{A_1 \to A_1''}$ and $V_2^{A_2 \to A_2''}$ be any full-rank partial isometries, $|\tilde{\psi}_1\rangle^{A_1'' B_1 R_1} = V_1 |\psi_1\rangle^{A_1 B_1 R_1}$ and $|\tilde{\psi}_2\rangle^{A_2'' B_2 R_2} = V_2 |\psi_2\rangle^{A_2 B_2 R_2}$, and define the states

$$\begin{aligned} |\omega_1(U_2)\rangle^{A_1'' A' D R_2 B_2} &= \sqrt{|A_2''|} \left(\mathrm{op}_{A_2'' \to A_1'' A' D}(|\sigma\rangle) U_2^{A_2''} |\tilde{\psi}_2\rangle^{A_2'' B_2 R_2} \right) \\ |\omega_2(U_1)\rangle^{A_2'' A' D R_1 B_1} &= \sqrt{|A_1''|} \left(\mathrm{op}_{A_1'' \to A_2'' A' D}(|\sigma\rangle) U_1^{A_1''} |\tilde{\psi}_1\rangle^{A_1'' B_1 R_1} \right) \end{aligned}$$

We now use Theorem 3.8 to get

$$\begin{aligned} \int \left\| \left(|A_1''| U_{\mathcal{N}} \, \mathrm{op}_{A_1'' \to A' D R_2 B_2}(|\omega_1(U_2)\rangle) U_1 \cdot \tilde{\psi}_1^{A_1'' R_1 B_1} \right)^{R_1 E D R_2 B_2 C_2} - \psi_1^{R_1} \otimes \omega_1(U_2)^{C_2 E D R_2 B_2} \right\|_1 dU_1 \\ \leqslant 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A_1'' | E D R_2 B_2 C_2)_{U_{\mathcal{N}} \cdot \omega_1(U_2)} - \frac{1}{2} H_{\min}^{\varepsilon}(A_1 | R_1)_{\psi_1} + 1} + 12\varepsilon \quad (5.2) \end{aligned}$$

and

$$\begin{aligned} \int \left\| \left(|A_2''| U_{\mathcal{N}} \, \mathrm{op}_{A_2'' \to A' D R_1 B_1}(|\omega_2(U_1)\rangle) U_2 \cdot \tilde{\psi}_2^{A_2'' R_2 B_2} \right)^{R_2 E D R_1 B_1 C_1} - \psi_2^{R_2} \otimes \omega_2(U_1)^{C_1 E D R_1 B_1} \right\|_1 dU_2 \\ \leqslant 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A_2'' | E D R_1 B_1 C_1)_{U_{\mathcal{N}} \cdot \omega_2} - \frac{1}{2} H_{\min}^{\varepsilon}(A_2 | R_2)_{\psi_2} + 1} + 12\varepsilon \quad (5.3) \end{aligned}$$

Note that the first states on the left-hand side of both inequalities are actually the same state written differently, namely $|A_1''||A_2''|(U_{\mathcal{N}} \, \mathrm{op}_{A_1'' A_2'' \to A' D}(|\sigma\rangle)(U_1 \otimes U_2) \cdot (\tilde{\psi}_1 \otimes \tilde{\psi}_2))$ (see Lemma 2.7). This is close to what we need, but there are still two problems: the encoder in the above is not an isometry, and the smooth-min-entropies should be in terms of $U_{\mathcal{N}} \cdot \sigma$ rather than $U_{\mathcal{N}} \cdot \omega_1$ and $U_{\mathcal{N}} \cdot \omega_2$. To solve the first problem (and temporarily exacerbate the second!), we use Theorem 3.8 again to get

$$\int \left\| |A_1''| \left(\mathrm{op}_{A_1'' \to R_2 B_2 A' D}(|\omega_1\rangle) U_1 \cdot \tilde{\psi}_1^{A_1 B_1 R_1} \right)^{R_2 B_2 R_1 B_1} - \psi_1^{R_1 B_1} \otimes \omega_1(U_2)^{R_2 B_2} \right\|_1 dU_1$$
$$\leqslant 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A_1''|R_2 B_2)_{\omega_1(U_2)} - \frac{1}{2} H_{\min}^{\varepsilon}(A_1|R_1 B_1)_{\psi_1} + 1} + 12\varepsilon$$

and

$$\int \left\| |A_2''| \left(\mathrm{op}_{A_2'' \to A_1'' A' D}(|\sigma\rangle) U_2 \cdot \tilde{\psi}_2^{A_2'' B_2 R_2} \right)^{R_2 B_2} - \psi_2^{R_2 B_2} \right\|_1 dU_2$$
$$\leqslant 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A_2'')_{\sigma} - \frac{1}{2} H_{\min}^{\varepsilon}(A_2|R_2 B_2)_{\psi_2} + 1} + 12\varepsilon.$$

Note that, in this last inequality, the first state on the left inside the trace norm is simply ω_1; we can therefore use the triangle inequality to get

$$\int \left\| |A_1''| \left(\mathrm{op}_{A_1'' \to R_2 B_2 A' D}(|\omega_1\rangle) U_1 \cdot \tilde{\psi}_1^{A_1'' B_1 R_1} \right)^{R_2 B_2 R_1 B_1} - \psi_1^{R_1 B_1} \otimes \psi_2^{R_2 B_2} \right\|_1 dU_1 dU_2$$
$$\leqslant \int 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A_1''|R_2 B_2)_{\omega_1(U_2)} - \frac{1}{2} H_{\max}^{\varepsilon}(A_1)_{\psi_1} + 1} dU_2 + 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A_2'')_{\sigma} - \frac{1}{2} H_{\max}^{\varepsilon}(A_2)_{\psi_2} + 1} + 24\varepsilon$$

This will allow us to solve our first problem: we will use Uhlmann's theorem on this last inequality to obtain our encoding isometry $W^{A_1 A_2 \to A' D}$. But before doing this, we will turn our attention to the second problem, namely that of bounding the min-entropies on $\omega_1(U_2)$ and $\omega_2(U_1)$.

There are three such problematic min-entropies: $H_{\min}^{\varepsilon}(A_1''|R_2 B_2)_{\omega_1(U_2)}$ in this latest inequality, as well as $H_{\min}^{\varepsilon}(A_1''|EDR_2 B_2 C_2)_{U_{\mathcal{N}} \cdot \omega_2(U_1)}$ and $H_{\min}^{\varepsilon}(A_2''|EDR_1 B_1 C_1)_{U_{\mathcal{N}} \cdot \omega_1(U_2)}$ in Equations (5.2) and (5.3) respectively. We will first deal with the first one explicitly; the same technique applies to the other two.

Let $\tilde{\sigma}^{A_1'' A_2'' A' D}$ be a state such that $\|\tilde{\sigma} - \sigma\|_1 \leqslant 2\varepsilon$ and $H_{\min}(A_1''|A_2'')_{\tilde{\sigma}} = H_{\min}^{\varepsilon}(A_1''|A_2'')_{\sigma}$, and let $\mathcal{T}^{A_2'' \to R_2 B_2}$ be defined as $\mathcal{T}(\xi) = |A_2''|(\mathrm{op}_{A_2'' \to R_2 B_2}(|\tilde{\psi}_2\rangle) \cdot \xi)$

(and hence, $\omega_1(U_2) = \mathcal{T}(\sigma)$). Furthermore, let $\theta^{A_2''}$ be a positive semi-definite operator such that $\tilde{\sigma}^{A_1''A_2''} \leqslant \mathbb{I}^{A_1''} \otimes \theta^{A_2''}$, with $\mathrm{Tr}[\theta] = 2^{-H_{\min}^{\varepsilon}(A_1''|A_2'')_\sigma}$. Then, since $\mathcal{T}$ is completely positive, we have that, for any $U_2^{A_2''}$,

$$\mathcal{T}(U_2^{A_2''} \cdot \tilde{\sigma})^{A_1''R_2B_2} \leqslant \mathbb{I}^{A_1''} \otimes \mathcal{T}(U_2^{A_2''} \cdot \theta)^{R_2B_2}.$$

Hence, if we could be certain that $\mathcal{T}(U_2 \cdot \tilde{\sigma})^{A_1''R_2B_2}$ is within ε in fidelity distance to $\mathcal{T}(U_2 \cdot \sigma)$, we would have shown that $H_{\min}^{\varepsilon}(A_1''|R_2B_2)_{\omega_1(U_2)} \geqslant H_{\min}^{\varepsilon}(A_1''|A_2'')_\sigma$ for any U_2. Unfortunately, things are not so easy for us: we will instead have to show that, when averaging over U_2, the fidelity distance is not too bad. Note that first that $\int \mathcal{T}(U_2 \cdot \xi) dU_2 = \mathrm{Tr}[\xi]\mathcal{T}(\pi^{A_2''})$. Likewise, letting $\tilde{\sigma} - \sigma = \Delta_+ - \Delta_-$ with $\Delta_\pm$ positive semi-definite and having disjoint support, and with $\mathrm{Tr}[\Delta_\pm] \leqslant 2\varepsilon$,

$$\begin{aligned}\int \|\mathcal{T}(U_2 \cdot \tilde{\sigma}) - \mathcal{T}(U_2 \cdot \sigma)\|_1 dU_2 &= \int \|\mathcal{T}(U_2 \cdot (\Delta_+ - \Delta_-)\|_1 dU_2 \\ &\leqslant \int \mathrm{Tr}[\mathcal{T}(U_2 \cdot \Delta_+)] + \mathrm{Tr}[\mathcal{T}(U_2 \cdot \Delta_-)] dU_2 \\ &\leqslant 4\varepsilon\end{aligned}$$

and therefore

$$\int d_F(\mathcal{T}(U_2 \cdot \tilde{\sigma}), \mathcal{T}(U_2 \cdot \sigma)) dU_2 \leqslant 2\sqrt{\varepsilon}. \tag{5.4}$$

Hence, on average, the fidelity distance is not too bad and we conclude that, on average,

$$H_{\min}^{2\sqrt{\varepsilon}}(A_1''|R_2B_2)_{\omega_1(U_2)} \geqslant H_{\min}^{\varepsilon}(A_1''|A_2'')_\sigma.$$

Since we want a bound on $H_{\min}^{\varepsilon}(A_1''|R_2B_2)$, we can state this as

$$H_{\min}^{\varepsilon}(A_1''|R_2B_2) \geqslant H_{\min}^{\varepsilon^2/4}(A_1''|A_2'')_\sigma.$$

We can also use the same trick on the other two smooth min-entropies. We now have three inequalities that we want U_1 to satisfy and four more inequalities that we need U_2 to satisfy. We can use Markov's inequality (see Lemma I.7) on these to show that there exist a $U_1^{A''_1}$ and a $U_2^{A''_2}$ that satisfy all of them. The inequalities that must be satisfied by U_1 are the following:

$$\left\| |A''_1| \left(U_{\mathcal{N}} \operatorname{op}_{A''_1 \to A' D R_2 B_2}(|\omega_1(U_2)\rangle\rangle) U_1 \cdot \tilde{\psi}_1^{A''_1 R_1 B_1} \right)^{R_1 E D R_2 B_2 C_2} - \psi_1^{R_1} \otimes \omega_1^{E D R_2 B_2 C_2} \right\|_1 \leqslant 4 \times 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A''_1|EDR_2B_2C_2)_{U_{\mathcal{N}} \cdot \omega_1(U_2)} - \frac{1}{2} H_{\min}^{\varepsilon}(A_1|R_1)_{\psi_1} + 1} + 48\varepsilon \quad (5.5)$$

$$\left\| |A''_1| \left(\operatorname{op}_{A''_1 \to R_2 B_2 A' D}(|\omega_1(U_2)\rangle\rangle) U_1 \cdot \tilde{\psi}_1^{A''_1 B_1 R_1} \right)^{R_2 B_2 R_1 B_1} - \psi_1^{R_1 B_1} \otimes \omega_1^{R_2 B_2} \right\|_1 \leqslant 4 \times 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A''_1|R_2B_2)_{\omega_1(U_2)} - \frac{1}{2} H_{\min}^{\varepsilon}(A_1|R_1B_1)_{\psi_1} + 1} + 48\varepsilon \quad (5.6)$$

and

$$2^{-H_{\min}^{\varepsilon}(A''_2|EDR_1B_1C_1)_{U_{\mathcal{N}} \cdot \omega_2(U_1)}} \leqslant 2^{-H_{\min}^{\varepsilon^2/16}(A''_2|EDA_1C_1)_{U_{\mathcal{N}} \cdot \sigma}}$$

(The last one is actually a bound on a fidelity distance as in (5.4)).

Likewise, there must exist a $U_2^{A''_2}$ such that

$$\int \left\| \left(|A''_2| U_{\mathcal{N}} \operatorname{op}_{A''_2 \to A' D R_1 B_1}(|\omega_2(U_1)\rangle\rangle) U_2 \cdot \tilde{\psi}_2^{A''_2 R_2 B_2} \right)^{R_2 E D R_1 B_1 C_1} - \psi_2^{R_2} \otimes \omega_2^{C_1 E D R_1 B_1} \right\|_1 dU_2 \leqslant 5 \times 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A''_2|EDR_1B_1C_1)_{U_{\mathcal{N}} \cdot \omega_2} - \frac{1}{2} H_{\min}^{\varepsilon}(A_2|R_2)_{\psi_2} + 1} + 60\varepsilon \quad (5.7)$$

$$\int \left\| |A''_2| \left(\operatorname{op}_{A''_2 \to A''_1 A' D}(|\sigma\rangle\rangle) U_2 \cdot \tilde{\psi}_2^{A''_2 B_2 R_2} \right)^{R_2 B_2} - \psi_2^{R_2 B_2} \right\|_1 dU_2 \leqslant 5 \times 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A''_2)_{\sigma} - \frac{1}{2} H_{\min}^{\varepsilon}(A_2|R_2B_2)_{\psi_2} + 1} + 60\varepsilon. \quad (5.8)$$

$$2^{-H_{\min}^{\varepsilon}(A''_1|R_2B_2)_{\omega_1(U_2)}} \leqslant 2^{-H_{\min}^{\varepsilon^2/20}(A''_1|A''_2)_{\sigma}} \quad (5.9)$$

$$2^{-H_{\min}^{\varepsilon}(A''_1|EDR_2B_2C_2)_{U_{\mathcal{N}} \cdot \omega_1(U_2)}} \leqslant 2^{-H_{\min}^{\varepsilon^2/20}(A''_1|EDA''_2C_2)_{U_{\mathcal{N}} \cdot \sigma}}. \quad (5.10)$$

Now, we can combine Equations (5.6), (5.8), and (5.9) to get

$$\left\| |A_1''| \left(\mathrm{op}_{A_1''\to R_2B_2A'D}(|\omega_1(U_2)\rangle)U_1 \cdot \tilde{\psi}_1^{A_1''B_1R_1}\right)^{R_2B_2R_1B_1} - \psi_1^{R_1B_1} \otimes \psi_2^{R_2B_2} \right\|_1 \leqslant 4 \times 2^{-\frac{1}{2}H_{\min}^{\varepsilon^2/20}(A_1''|A_2'')_\sigma - \frac{1}{2}H_{\min}^{\varepsilon}(A_1|R_1B_1)_{\psi_1}+1} + 5 \times 2^{-\frac{1}{2}H_{\min}^{\varepsilon}(A_2'')_\sigma - \frac{1}{2}H_{\min}^{\varepsilon}(A_2|R_2B_2)_{\psi_2}+1} + 108\varepsilon. \quad (5.11)$$

Using Uhlmann's theorem, we finally get our encoding isometry $W^{A_1A_2\to A'D}$:

$$\left\| |A_1''| \left(\mathrm{op}_{A_1''\to R_2B_2A'D}(|\omega_1\rangle)U_1 \cdot \tilde{\psi}_1^{A_1''B_1R_1}\right)^{R_2B_2R_1B_1A'D} - W \cdot (\psi_1^{A_1R_1B_1} \otimes \psi_2^{A_2R_2B_2}) \right\|_1 \leqslant 2\sqrt{\delta_{\mathrm{enc}}} \quad (5.12)$$

where δ_{enc} is defined as the right-hand side of (5.11). Finally, the two decoupling conditions (5.5) and (5.7) together with Uhlmann's theorem and Lemma 5.1 yield the existence of the two decoders. □

We can now use this to prove an i.i.d. version for both entanglement-assisted and unassisted coding:

Theorem 5.3. *Let $\mathcal{N}^{A'\to C_1C_2}$ be a quantum broadcast channel with Stinespring extension $U_{\mathcal{N}}^{A'\to C_1C_2E}$, let $\sigma^{A_1A_2A'D}$ be any pure state, and define $\rho^{A_1A_2C_1C_2ED} = U_{\mathcal{N}} \cdot \sigma$. Then, all rates satisfying*

$$\begin{aligned} 0 \leqslant Q_1 + E_1 &< H(A_1)_\rho & Q_1 - E_1 &< I(A_1\rangle C_1)_\rho \\ 0 \leqslant Q_2 + E_2 &< H(A_2)_\rho & Q_2 - E_2 &< I(A_2\rangle C_2)_\rho \\ Q_1 + E_1 + Q_2 + E_2 &< H(A_1A_2)_\rho \end{aligned} \quad (5.13)$$

are achievable for quantum transmission with rate-limited entanglement assistance through $\mathcal{N}$. In particular, if we allow E_1 and E_2 to be maximized (corresponding to *fully entanglement-assisted coding), we get a quantum version of Marton's region:*

$$\begin{aligned} Q_1 &< \frac{1}{2} I(A_1;C_1)_\rho \\ Q_2 &< \frac{1}{2} I(A_2;C_2)_\rho \\ Q_1 + Q_2 &< \frac{1}{2} I(A_1;C_1)_\rho + \frac{1}{2} I(A_2;C_2)_\rho - \frac{1}{2} I(A_1;A_2)_\rho. \end{aligned}$$

Proof. The proof is little more than applying the previous theorem together with the fully quantum AEP (Theorem 2.4). Consider using the previous theorem on $\mathcal{N}^{\otimes n}$ with input distribution $\sigma^{\otimes n}$ and with transmission and entanglement consumption rates Q_1, Q_2, E_1 and E_2. Let R_1 and M_1 be systems of dimension 2^{nQ_1}, with M_1 representing the quantum information that Alice wants to send to Bob 1, with R_1 being the system that purifies it. Furthermore, let $\widetilde{A}_1$ and B_1 be systems of dimension 2^{nE_1} representing Alice's and Bob 1's halves of the preshared entanglement. Replicate all these definitions with subscript 2 for Bob 2. Then, we define $\psi_1 = \Phi^{R_1 M_1} \otimes \Phi^{\widetilde{A}_1 B_1}$ and $\psi_2 = \Phi^{R_2 M_2} \otimes \Phi^{\widetilde{A}_2 B_2}$, where $M_1 \widetilde{A}_1$ and $M_2 \widetilde{A}_2$ play the roles of A_1 and A_2 from the previous theorem.

To get an error that goes down to zero as $n \to \infty$, we need to ensure that δ_{enc}, δ_1 and δ_2 all go down to zero as $n \to \infty$. By the fully quantum AEP and using the fact that $H_{\max}(A_1)_{\psi_1} = n(Q_1 + E_1)$ and $H_{\max}(A_2)_{\psi_2} = n(Q_2 + E_2)$, δ_{enc} goes down to zero if

$$\begin{aligned} Q_1 + E_1 &< H(A_1|A_2)_\rho \\ Q_2 + E_2 &< H(A_2)_\rho. \end{aligned}$$

Likewise, using the fact that $H_{\min}(A_1|R_1)_{\psi_1} = n(E_1 - Q_1)$ and $H_{\min}(A_2|R_2)_{\psi_2} = n(E_2 - Q_2)$, we get that δ_1 goes down to zero if

$$Q_1 - E_1 < H(A_1|EDA_2C_2)_\rho = I(A_1\rangle C_1)_\rho$$

and δ_2 goes to zero if

$$Q_2 - E_2 < H(A_2|EDA_1C_1)_\rho = I(A_2\rangle C_2)_\rho.$$

By switching the roles of Bob 1 and Bob 2, we can also get any rate in the region

$$\begin{aligned} Q_1 + E_1 &< H(A_1)_\rho & Q_1 - E_1 &< I(A_1\rangle C_1)_\rho \\ Q_2 + E_2 &< H(A_2|A_1)_\rho & Q_2 - E_2 &< I(A_2\rangle C_2)_\rho. \end{aligned}$$

Taking convex combinations of points in these two regions (which corresponds to timesharing between different protocols) yields the region in the theorem statement.

To get the fully entanglement-assisted region, we simply take linear combinatinons of the various inequalities to get constraints only on Q_1 and Q_2:

$$\begin{aligned} Q_1 &< \frac{1}{2}H(A_1)_\rho + \frac{1}{2}I(A_1\rangle C_1)_\rho = \frac{1}{2}I(A_1;C_1)_\rho \\ Q_2 &< \frac{1}{2}H(A_2)_\rho + \frac{1}{2}I(A_2\rangle C_2)_\rho = \frac{1}{2}I(A_2;C_2)_\rho \\ Q_1 + Q_2 &< \frac{1}{2}\left[H(A_1A_2)_\rho + I(A_1\rangle C_1)_\rho + I(A_2\rangle C_2)_\rho\right] \\ &= \frac{1}{2}\left[H(A_1A_2)_\rho - H(A_1|C_1)_\rho - H(A_2|C_2)_\rho\right] \\ &= \frac{1}{2}\left[H(A_1)_\rho + H(A_2)_\rho - I(A_1;A_2)_\rho - H(A_1|C_1)_\rho - H(A_2|C_2)_\rho\right] \\ &= \frac{1}{2}\left[I(A_1;C_1)_\rho + I(A_2;C_2)_\rho - I(A_1;A_2)_\rho\right]. \end{aligned}$$

□

5.4 Regularized converse

The rate region for the case given in Theorem 5.3 is indeed the capacity for quantum transmission with rate-limited entanglement assistance of quantum broadcast channels provided we regularize over many uses of the channel. It is important to remember, however, that regions defined by very different formulas can nonetheless agree after regularization, so the following theorem should be understood to be only a very weak characterization of the capacity.

Theorem 5.4. *The capacity region for rate-limited quantum transmission of a quantum broadcast channel $\mathcal{N}^{A'\to C_1C_2}$ is the convex hull of the union of all rate points (Q_1, Q_2, E_1, E_2) satisfying*

$$\begin{aligned} Q_1 + E_1 &\leqslant \frac{1}{n}H(A_1)_\psi & \qquad Q_1 - E_1 &\leqslant \frac{1}{n}I(A_1\rangle C_1^n)_\psi \\ Q_2 + E_2 &\leqslant \frac{1}{n}H(A_2)_\psi & \qquad Q_2 - E_1 &\leqslant \frac{1}{n}I(A_2\rangle C_2^n)_\psi \\ Q_1 + Q_2 + E_1 + E_2 &\leqslant \frac{1}{n}H(A_1A_2)_\psi \end{aligned} \tag{5.14}$$

for some state of the form $|\psi\rangle^{A_1A_2C_1^nC_2^nDE^n} = U_{\mathcal{N}}^{\otimes n}|\phi\rangle^{A_1A_2A'^nD}$, where $|\phi\rangle$ is a pure state.

Proof. It is immediate from Theorem 5.3 that the region is achievable. We now prove the converse.

Suppose that (Q_1, Q_2, E_1, E_2) is an achievable four-tuple. That means that there exists a sequence of codes of length n with these rates and with error rate going to 0 as $n \to \infty$. Consider the code of block size n in this sequence. Let $\psi = \Phi^{R_1M_1}\otimes\Phi^{\widetilde{A}_1B_1}\otimes\Phi^{R_2M_2}\otimes\Phi^{\widetilde{A}_2B_2}$ be the input state as in Theorem 5.3, $\mathcal{E}^{M_1M_2\widetilde{A}_1\widetilde{A}_2\to A'^n}$ be the encoding superoperator, and let $\rho^{R_1R_2C_1^nC_2^nB_1B_2E^n} = U_{\mathcal{N}}^{\otimes n}\cdot\mathcal{E}(\psi)$. We will evaluate entropic quantities with respect to ρ.

Given that Bob 1 must be able to recover a system which purifies R_1 from C_1^n and B_1, we have by Fannes' inequality (Theorem I.9) and the monotonicity of the mutual information (see Section 2.4.2) that $I(R_1; C_1^nB_1) \geqslant 2nQ_1 - n\delta_n$, where

$\delta_n \to 0$ as $n \to \infty$, and likewise for Bob 2. We also have

$$\begin{aligned}
2nQ_1 - n\delta_n &\leqslant I(R_1; C_1^n B_1) \\
&= H(R_1) + H(C_1^n B_1) - H(R_1 C_1^n B_1) \\
&\leqslant H(R_1) + H(C_1^n) + H(B_1) - H(R_1 C_1^n B_1) \\
&= nQ_1 + nE_1 + H(C_1^n) - H(R_1 C_1^n B_1) \\
&= nQ_1 + nE_1 + I(R_1 B_1 \rangle C_1^n)
\end{aligned}$$

where the second line follows from subadditivity, and the third line from the definition of R_1 and B_1. Hence, if we identify $R_1 B_1$ as A_1 and likewise for Bob 2, we get

$$Q_1 - E_1 \leqslant \frac{1}{n} I(A_1 \rangle C_1^n) + \delta_n \tag{5.15}$$

$$Q_2 - E_2 \leqslant \frac{1}{n} I(A_2 \rangle C_2^n) + \delta_n \tag{5.16}$$

where $\delta_n \to 0$ as $n \to \infty$. Since $Q_1 + E_1 = \frac{1}{n} H(A_1)$, $Q_2 + E_2 = \frac{1}{n} H(A_2)$, and $H(A_1 A_2) = H(A_1) + H(A_2)$ by construction, this rate point is clearly inside the region in Equation (5.14), and it follows that this is indeed the capacity of the channel. □

While one might conjecture that Theorem 5.4 characterizes the capacity region of a broadcast channel for quantum transmission with rate-limited entanglement assistance even with the restriction $n = 1$, this is false even in the special case of unassisted quantum transmission through a channel with a single receiver [DSS98]. It may however be the true capacity for the fully entanglement-assisted case, but there is no reason to believe that this would be any easier to prove than to prove that Marton's region is optimal in the classical case.

5.5 Single-letter example

In the classical case, the simplest example of a broadcast channel for which Marton's region is optimal is a deterministic channel, i.e. a channel where the outputs are completely determined by the inputs. Similarly, we can show that our rate region is optimal for entanglement-assisted quantum transmission through classical deterministic channels. This is perhaps unsurprising since entanglement would be highly unlikely to help classical transmission through a classical channel, but it nonetheless provides an example for which our theorem is optimal.

We say that $\mathcal{N}^{A'\to C_1C_2}$ is a classical deterministic broadcast channel if there exist two deterministic functions $f_1 : \{1,\dots,|A'|\} \to \{1,\dots,|C_1|\}$ and $f_2 : \{1,\dots,|A'|\} \to \{1,\dots,|C_2|\}$ such that $U_{\mathcal{N}}|i\rangle = |f_1(i)\rangle^{C_1} \otimes |f_2(i)\rangle^{C_2} \otimes |i\rangle^E$ for some fixed orthonormal bases on A', C_1, C_2 and E. We claim that any rate point that can be achieved for such a channel is a convex combination of rates that can be achieved via our coding method with input states of the form $\varphi^{A_1A_2A'} = \sum_{i=1}^{|A'|} p_i |f_1(i)\rangle\langle f_1(i)|^{A_1} \otimes |f_2(i)\rangle\langle f_2(i)|^{A_2} \otimes |i\rangle\langle i|^{A'}$ for some probability distribution $\{p_i\}$. To prove this, we first need the following observation:

Lemma 5.5. *Let $f : \{1,\dots,|D|\} \to \{1,\dots,|B|\}$ be a function, and $|\xi\rangle^{ABCD}$ be $\sum_i \alpha_i |\mu_i\rangle^A \otimes |f(i)\rangle^B \otimes |\nu_i\rangle^C \otimes |i\rangle^D$, where $|\mu_i\rangle$ and $|\nu_i\rangle$ are any pure states, and $|i\rangle$ and $|f(i)\rangle$ represent i and $f(i)$ encoded in a standard bases on D and B respectively. Then, $I(A;B)_\xi \leqslant H(B)_\xi$.*

Proof. The lemma follows from the observation that because of the structure of ξ^{AB} and strong subadditivity (see Section 2.4.2), $H(B|A) \geqslant H(B|D)$. The latter is a classical conditional entropy and is therefore never negative, which means that $I(A;B)_\xi = H(B)_\xi - H(B|A)_\xi \leqslant H(B)_\xi$. □

Armed with this, we can now show the following:

Theorem 5.6. *Let $\mathcal{N}^{A'\to C_1C_2}$ be a classical deterministic channel. Then, the capacity region for entanglement-assisted quantum transmission on this channel is the same as* *the achievable rate region given by Theorem 5.3.*

Proof. According to the regularized converse theorem (Theorem 5.4), for any achievable rate point (Q_1, Q_2), there exists a state $|\psi\rangle^{A_1A_2C_1^nC_2^nE^nD} = U_{\mathcal{N}}^{\otimes n}|\varphi\rangle^{A_1A_2A'^nD}$ such that $Q_1 = \frac{1}{2n}I(A_1;C_1^n)_\psi + \delta_n$, $Q_2 = \frac{1}{2n}I(A_2;C_2^n)_\psi + \delta_n$, where $\delta_n \geqslant 0$, and $I(A_1;A_2)_\psi = 0$. Let $C_{1,i}$ and $C_{2,i}$ be the ith copies of C_1 and C_2 in C_1^n and C_2^n, and, for each i, let $\psi_i^{A_1A_2C_1C_2} = \sum_{jk} |jkjk\rangle\langle jk|\psi^{C_{1,i}C_{2,i}}|jk\rangle\langle jkjk|$, where the $\langle jkjk||jk\rangle$ are defined in the classical basis on $C_{1,i}$ and $C_{2,i}$ and in some fixed basis on A_1, A_2, C_1 and C_2. Then, we can bound the individual rates as follows:

$$\begin{aligned}
Q_1 &\leqslant \frac{1}{2n}I(A_1;C_1^n)_\psi + \delta_n && (5.17)\\
&\leqslant \frac{1}{2n}H(C_1^n)_\psi + \delta_n && (5.18)\\
&\leqslant \frac{1}{2n}\sum_i H(C_{1,i})_\psi + \delta_n && (5.19)\\
&= \frac{1}{2n}\sum_i H(C_1)_{\psi_i} + \delta_n && (5.20)\\
&= \frac{1}{n}\sum_i \frac{1}{2}I(A_1;C_1)_{\psi_i} + \delta_n && (5.21)
\end{aligned}$$

and likewise for Q_2. The second inequality is due to Lemma 5.5, with the roles of the B and D subsystems in the lemma played by C_1^n and E^n respectively, and the third inequality makes use the subadditivity of the von Neumann entropy.

We can now do the same thing for the sum rate:

$$\begin{aligned}
Q_1 + Q_2 &= \frac{1}{2n}\left\{I(A_1;C_1^n)_\psi + I(A_2;C_2^n)_\psi\right\} + 2\delta_n \\
&= \frac{1}{2n}\left\{H(A_1)_\psi + H(A_2)_\psi - H(A_1|C_1^n)_\psi - H(A_1;C_2^n)_\psi\right\} + 2\delta_n \\
&\leqslant \frac{1}{2n}\left\{H(A_1A_2)_\psi - H(A_1A_2|C_1^nC_2^n)_\psi\right\} + 2\delta_n \\
&= \frac{1}{2n} I(A_1A_2;C_1^nC_2^n)_\psi + 2\delta_n \\
&\leqslant \frac{1}{2n} H(C_1^nC_2^n)_\psi + 2\delta_n \qquad (5.22) \\
&\leqslant \frac{1}{2n}\sum_i H(C_{1,i}C_{2,i})_\psi + 2\delta_n \\
&= \frac{1}{2n}\sum_i H(C_1C_2)_{\psi_i} + 2\delta_n \\
&= \frac{1}{n}\sum_i \frac{1}{2}\left\{H(C_1)_{\psi_i} + H(C_2)_{\psi_i} - I(C_1;C_2)_{\psi_i}\right\} + 2\delta_n \\
&= \frac{1}{n}\sum_i \frac{1}{2}\left\{I(A_1;C_1)_{\psi_i} + I(A_2;C_2)_{\psi_i} - I(A_1;A_2)_{\psi_i}\right\} + 2\delta_n
\end{aligned}$$

where, in the first inequality, we have made use of the fact that A_1 and A_2 are independent and of the standard inequality $H(AB|CD) \leqslant H(A|C) + H(B|D)$, and the last equality follows from the special form of the ψ_i's.

Since every i in equations (5.21) and (5.22) corresponds to a rate which is achievable via Theorem 5.3, this concludes the proof. □

5.6 Discussion

We have exhibited and analyzed a new protocol for quantum communication with rate-limited entanglement assistance through quantum broadcast channels.

Our protocol achieves the following rate region for every mixed state $\sigma^{A_1 A_2 A'}$:

$$\begin{aligned} 0 \leqslant Q_1 &\leqslant \frac{1}{2} I(A_1; C_1)_\rho \\ 0 \leqslant Q_2 &\leqslant \frac{1}{2} I(A_2; C_2)_\rho \\ Q_1 + Q_2 &\leqslant \frac{1}{2} \left[I(A_1; C_1)_\rho + I(A_2; C_2)_\rho - I(A_1; A_2)_\rho \right] \end{aligned} \tag{5.23}$$

where $\rho^{A_1 A_2 C_1 C_2 E} = U_{\mathcal{N}}^{A' \to C_1 C_2 E} \cdot \sigma^{A_1 A_2 A'}$.

The corresponding rate region (Equation (5.13)) is very similar to Marton's region for classical broadcast channels (Equation (5.1)) [Mar79]; except for the factors of $1/2$, the two expressions are formally identical. In fact, for classical channels, the rates for entanglement-assisted quantum communication found here can be achieved directly using teleportation between the senders and the receiver, with the classical communication required by teleportation transmitted using Marton's protocol. From this point of view, our results can be viewed as a direct generalization of Marton's region to quantum channels.

Therefore, once again, it is the entanglement-assisted version of the quantum capacity that bears the strongest resemblance to its classical counterpart. The same is true for both the regular point-to-point quantum channel [BSST02] and the quantum multiple-access channel [HDW08, HOW07] and, of course, the quantum channels with side information at the transmitter that were discussed in the last chapter. In both those cases, the known achievable rate regions for entanglement-assisted quantum communication are identical to their classical counterparts. This collection of similarities suggests a fundamental question. To what extent does the addition of free entanglement make quantum information theory similar to classical information theory?

Of course, the lack of a single-letter converse for Marton's region and, by extension, for our region, leaves open the possibility that the analogy might break down for a new, better broadcast region that remains to be discovered. A first step towards eliminating that uncertainty could be to find a better characterization of the quantum regions we have presented here. The presence of the "discarded" system D in Theorem 5.3 is equivalent to optimizing over all mixed states $\phi^{A_1A_2A'}$ rather than only over pure states. This is not required for most theorems in quantum information theory, but we have not found a way to prove the regularized converse without allowing for the possibility of mixed states. We leave it as an open problem to determine whether it is possible to demonstrate a converse theorem that does not require allowing mixed states.

Finally, for the unassisted case, it is very interesting to note the absence of an independent constraint on the sum-rate. However, we already know that this region is suboptimal even for channels with a single receiver. It would therefore be desirable to know whether this holds for the true capacity region and whether there is an underlying principle that explains this phenomenon.

CHAPTER 6

LOCKING CLASSICAL INFORMATION IN QUANTUM STATES

One particularly shocking feature of quantum information is the "information locking" effect that one sometimes observes. At the general level, it consists of a system in which one encodes classical data into a quantum system with two parts, one part being a very large "cyphertext", and the other being a very small key. The strange phenomenon is that it is possible to set up the system in such a way that, given the large portion, one can get almost no information about the classical data by measuring the cyphertext, whereas the key allows one to "unlock" this information. This may seem at first somewhat unsurprising, since this is what classical cryptographic systems aim to do, but it must be stressed that this is at the information theory level: the distribution on the classical data given the large portion is almost the same as the prior distribution even if the key is much smaller than the message. Classical encryption cannot achieve this at all: the distribution on the message given the cyphertext is vastly different from the prior distribution unless the key is as large as the message.

Information locking schemes have already been shown to exist by [DHL$^+$04] and [HLSW04]. In [DHL$^+$04], the authors construct a scheme by encoding the classical information in one of two mutually unbiased bases, and the one-bit classical key simply tells which basis it's encoded in. Without the key, the Shannon entropy about the message is approximately half of the entropy of the message; the key therefore increases the information the receiver has by the same amount.

In [HLSW04], the authors look at a protocol where one encodes classical information in the computational basis, and then applies one of a few (logarithmic in the number of possible messages) fixed unitaries. The classical key tells which unitary was applied. If the unitaries are chosen according to the Haar measure, then locking occurs with high probability.

In both of these papers, locking was defined in terms of the *accessible information* between the cyphertext and the message, which defined as follows:

Definition 6.1 (Accessible information)**.** *Let $\rho^{AB} \in \mathrm{D}(\mathsf{A} \otimes \mathsf{B})$ be a quantum state. Then, the accessible information $I_{\mathrm{acc}}(A; B)$ is defined as*

$$I_{\mathrm{acc}}(A;B)_\rho := \sup_{\mathcal{A},\mathcal{B}} I(X;Y)_{(\mathcal{A}\otimes\mathcal{B})(\rho)},$$

where $\mathcal{A}^{A\to X}$ and $\mathcal{B}^{B\to Y}$ are measurement superoperators, and the supremum is taken over all possible superoperators. In other words, the accessible information is the largest possible mutual information between the results of measurements made on A and B.

Locking was said to occur when the difference in accessible information with and without the key was larger than the size of the key. Here we will instead use the trace distance between the joint distribution of the measurement results and the message and the product of their marginals. This will imply a bound on the mutual information via the Alicki-Fannes inequality (Lemma I.10).

We now give the formal definition of locking that we will use:

Definition 6.2. *Let C and K be two quantum systems. We call a set of quantum states $\{\rho_m^{CK} : m \in \{1, \ldots, N\}\}$ an ε-locking scheme if $\|\rho_i^{CK} - \rho_j^{CK}\|_1 = 2$ whenever $i \neq j$, and for any complete measurement superoperator $\mathcal{M}^{C\to X}$, we have that*

$$\left\|\mathcal{M}(\omega^{MC}) - \mathcal{M}(\pi^C) \otimes \omega^M\right\|_1 \leqslant \varepsilon$$

where $\omega^{MC} = \frac{1}{N}\sum_{i=1}^N |i\rangle\langle i|^M \otimes \rho_i^C$ and π^C denotes the completely mixed state on C.

In other words, a set of states is a locking scheme if the states are perfectly distinguishable when one has both the cyphertext C and the key K, whereas a measurement on C alone yields practically no information about which state was present. Note that the restriction to complete measurement is a natural one, since, if the goal is to maximize the information about the message, keeping a quantum residue is of no use: it can never hurt to measure it until nothing is left.

Note that it is impossible to achieve this classically without making the key almost as long as the message. One can see this by considering that fact that, if one only needs to know an extra $\log K$ bits to reconstruct the message, our probability distribution of the message given the cyphertext must always be supported on at most K distinct messages; such a distribution must necessary be far away from the uniform distribution over all messages unless K is nearly equal to the number of messages.

The scheme we will construct here is a special case of this model. We consider a scheme where we encode classical information in the computational basis of a quantum system, apply a fixed unitary, and split the system into two components, a large one (C) that becomes the cyphertext, and a small one (K) that becomes the key.

Note also that an ε-locking scheme also automatically implies locking of the accessible information:

Lemma 6.1. *Let* $\{\rho_m^{CK} : m \in \{1, \ldots, N\}\}$ *be an* ε*-locking scheme, and let* $\omega^{MC} = \frac{1}{N}\sum_{i=1}^{N} |i\rangle\langle i|^M \otimes \rho_i^C$. *Then,*

$$I_{\mathrm{acc}}(M; C)_\omega \leqslant \varepsilon \log N + 2\eta(1-\varepsilon) + 2\eta(\varepsilon),$$

where $\eta(x) := -x \log x$ *and* $\eta(0) = 0$.

Proof. Direct application of the Alicki-Fannes inequality (Lemma I.10). □

6.1 The locking scheme

Our information locking scheme is straightforward. To encode N equiprobable classical messages, we embed them via a random partial isometry into a system CK of total dimension at least N; C constitutes the cyphertext and K constitutes the key. The key is therefore itself a quantum state; if one prefers a scheme with a classical key as was done in [DHL$^+$04, HLSW04], one can simply perfectly encrypt the quantum key with a $2 \log K$-bit classical key and make the encrypted quantum key part of the cyphertext; the classical key is then the locking key.

To prove that this works, let $\{|\psi_m\rangle : 1 \leqslant m \leqslant N\}$ be any set of orthonormal pure states in CK. We would like to prove that there exists a U^{CK} such that $\{U^{CK}|\psi_m\rangle\}$ is a good locking scheme. To do this, we will consider the state $\rho^{MCK} = \frac{1}{N}\sum_{x=1}^{N} |m\rangle\langle m|^M \otimes |\psi_m\rangle\langle\psi_m|^{CK}$ and the expression

$$\left\|\mathcal{M}(\mathrm{Tr}_K[U \cdot \rho^{MCK}]) - \mathcal{M}(\pi^C) \otimes \rho^M\right\|_1$$

for a U^{CK} chosen according to the Haar measure and an arbitrary $\mathcal{M}$. We will show that the average is sufficiently small and that the distribution is sufficiently concentrated around the mean value to ensure that there exists a U that makes this expression small for every $\mathcal{M}$.

Theorem 6.2. *Let ρ^{MCK} be a state of the form $\rho^{MCK} = \sum_{m=1}^{N} |m\rangle\langle m|^M \otimes \rho_m^{CK}$. Then, there exists a U^{CK} such that for every measurement superoperator $\mathcal{M}^{C\to X}$,*

$$\left\|\mathcal{M}(\mathrm{Tr}_K[U \cdot \rho^{MCK}]) - \mathcal{M}(\pi^C) \otimes \rho^M\right\|_1 \leqslant 7\varepsilon$$

as long as $N \geqslant \frac{8\sqrt{2}}{\varepsilon}$, $\varepsilon \leqslant e^{-2}$, and $|K| \geqslant \frac{32}{\varepsilon}\sqrt{\log\left(\frac{4N^2}{\varepsilon}\right)\ln(1/\varepsilon)}$.

To prove this, we will first consider $\mathcal{M}$'s of a very specific form that will allow us to take a union bound:

Definition 6.3 (Quasi-measurement)**.** *We call a superoperator $\mathcal{M}^{C\to X}$ an (n,k)-quasi-measurement if it is of the form $\mathcal{M}(\sigma) = \frac{|C|}{n}\sum_{x=1}^{n} |x\rangle\langle\psi_x|\sigma|\psi_x\rangle\langle x|$ where the $|x\rangle$ are orthonormal, and $\frac{|C|}{n}\sum_{x=1}^{n} |\psi_x\rangle\langle\psi_x| \leqslant k\mathbb{I}^C$.*

The starting point will be the following concentration of measure result:

Lemma 6.3. *Let $\mathcal{M}$ be an (n,k)-quasi-measurement. Then,*

$$\Pr_U\left\{\left\|\mathcal{M}(\mathrm{Tr}_K[U \cdot \rho^{MCK}]) - \mathcal{M}(\pi^C) \otimes \rho^M\right\|_1 \geqslant 2^{\frac{1}{2}\log k - \frac{1}{2}\log|K|} + r\right\} \leqslant 2e^{-N^2r^2/16k^2}$$

Proof. The lemma is a direct application of Theorem 3.9. We first show that $\max\{\|\mathcal{M}(\mathrm{Tr}_K[X])\|_1 : X \in \mathrm{Herm}(\mathsf{A}), \|X\|_1 \leqslant 1\} \leqslant k$. Let $X \in \mathrm{Herm}(\mathsf{A})$; then,

$$\begin{aligned}
\|\mathcal{M}(\mathrm{Tr}_K(X))\|_1 &= \frac{|C|}{n}\left\|\sum_x |x\rangle\langle\psi_x|\,\mathrm{Tr}_K(X)|\psi_x\rangle\langle x|\right\|_1 \\
&= \frac{|C|}{n}\sum_x |\langle\psi_x|\,\mathrm{Tr}_K(X)|\psi_x\rangle| \\
&\leqslant \frac{|C|}{n}\sum_x \langle\psi_x||\,\mathrm{Tr}_K(X)||\psi_x\rangle \\
&= \frac{|C|}{n}\mathrm{Tr}\left[\sum_x |\psi_x\rangle\langle\psi_x||\,\mathrm{Tr}_K(X)|\right] \\
&\leqslant k\|\,\mathrm{Tr}_K(X)\|_1 \\
&\leqslant k\|X\|_1
\end{aligned}$$

where the first inequality follows from the matrix inequality $-|Y| \leqslant Y \leqslant |Y|$, which holds for any Hermitian Y. Next, let $\omega^{C'K'X} = \mathcal{M}(\mathrm{Tr}_K[\Phi^{C'K'CK}])$; we will show that $H_2(C'K'|X)_\omega \geqslant -\log k + \log|K|$. Since $\mathcal{M}$ is a quasi-measurement, ω has the form $\omega^{C'K'X} = \sum_{x=1}^n \alpha_x |x\rangle\langle x|^X \otimes \pi^{K'} \otimes (\psi_x^T)^{C'}$, where $\alpha_x = \frac{|C|}{n}\langle\psi_x|\pi|\psi_x\rangle$. Then, $(\omega^X)^{-1/4} = \sum_x \alpha_x^{-1/4}|x\rangle\langle x|$, and we have that

$$\begin{aligned}
2^{-H_2(C'K'|X)_\omega} &\leqslant \mathrm{Tr}\left[\left((\omega^X)^{-1/4}\omega^{C'K'X}(\omega^X)^{-1/4}\right)^2\right] \\
&= \mathrm{Tr}\left[\sum_x \alpha_x |x\rangle\langle x| \otimes (\pi^{K'})^2 \otimes (\psi_x^{C'})^T\right] \\
&= \frac{1}{|K|}\sum_x \frac{|C|}{n}\langle\psi_x|\pi|\psi_x\rangle \\
&= \frac{1}{|K|n}\mathrm{Tr}\left[\sum_x |\psi_x\rangle\langle\psi_x|\right] \\
&= \frac{k}{|K|}
\end{aligned}$$

We combine this with the fact that $H_2(CK|M)_\rho = 0$ to get the lemma. □

At this point, we would like to take a union bound over all possible quasi-measurements to be able to say that there is a nonzero probability that the trace distance above is small for every quasi-measurement $\mathcal{M}$. We do this by introducing an ε-net (see Definition I.1) $\mathfrak{N}$ over C, with $|\mathfrak{N}| \leqslant \left(\frac{5}{\varepsilon}\right)^{2|C|}$. For any (n,k)-quasi-measurement $\mathcal{M}^{C\to X}$ of the form $\mathcal{M}(\sigma) = \sum_x \alpha_x |x\rangle\langle\psi_x|\sigma|\psi_x\rangle\langle x|$, define $\mathcal{M}_{\mathfrak{N}}$ as $\mathcal{M}_{\mathfrak{N}}(\sigma) = \sum_x \alpha_x |x\rangle\langle\psi'_x|\sigma|\psi'_x\rangle\langle x|$, where $|\psi'_x\rangle$ is the state in $\mathfrak{N}$ closest to $|\psi_x\rangle$.

Given a sequence $(|\varphi_x\rangle : 1 \leqslant x \leqslant n, |\varphi_x\rangle \in \mathfrak{N})$, we say that is it ε-close to an (n,k)-quasi-measurement if there exists an (n,k)-quasi-measurement $\mathcal{M}(\sigma) = \frac{|A|}{n}\sum_{x=1}^n |x\rangle\langle\psi_x|\sigma|\psi_x\rangle\langle x|$ such that $\|\psi_x - \varphi_x\|_1 \leqslant \varepsilon$ for all x. Furthermore, given a sequence $q = \{|\psi_x\rangle : 1 \leqslant x \leqslant n\}$, we define $\mathcal{M}_q^{C\to X}$ as $\mathcal{M}_q(\sigma) = \frac{|C|}{n}\sum_{x=1}^n |x\rangle\langle\psi_x|\sigma|\psi_x\rangle\langle x|$.

We can now take the desired union bound:

Lemma 6.4. *Let $\mathfrak{Q} \subseteq \mathfrak{N}^n$ be the set of all sequences of n elements of $\mathfrak{N}$ that are ε-close to an (n,k)-quasi-measurement. Then,*

$$\Pr_U\left\{\exists q \in \mathfrak{Q} : \left\|\mathcal{M}_q(\mathrm{Tr}_K[U\cdot\rho^{MCK}]) - \mathcal{M}_q(\pi^C)\otimes\rho^M\right\|_1 \geqslant 2^{\frac{1}{2}\log k - \frac{1}{2}\log|K|} + 2\varepsilon + r\right\}$$
$$\leqslant 2e^{2n\ln(5/\varepsilon)|C| - N^2r^2/16k^2}$$

and therefore, as long as $2n\ln(5/\varepsilon)|C| - N^2r^2/16k^2 < -\ln 2$, there exists a U such that all $q \in \mathfrak{Q}$ satisfy the above inequality.

Proof. Let $\mathcal{M}_q(\sigma) = \frac{|C|}{n}\sum_x |x\rangle\langle\psi_x|\sigma|\psi_x\rangle\langle x|$ and let $\mathcal{M}'(\sigma) = \frac{|C|}{n}\sum_x |x\rangle\langle\psi'_x|\sigma|\psi'_x\rangle\langle x|$ be an (n,k)-quasi-measurement that is ε-close to q. Furthermore, let $\xi_x = \psi_x - \psi'_x$; clearly, for each x, $\|\xi_x\|_1 \leqslant \varepsilon$. Then, given any cq-state ζ^{MC} with $\zeta^C = \pi^C$, we have that

$$\begin{aligned}
&\left\|\mathcal{M}(\zeta^{MC} - \zeta^M \otimes \zeta^C)\right\|_1 \\
&\leqslant \frac{|C|}{n} \sum_{x=1}^{n} \left\|\mathrm{Tr}_C[\psi_x(\zeta^{MC} - \zeta^M \otimes \zeta^C)]\right\|_1 \\
&= \frac{|C|}{n} \sum_{x=1}^{n} \left\|\mathrm{Tr}_C[(\psi'_x + \xi_x)(\zeta^{MC} - \zeta^M \otimes \zeta^C)]\right\|_1 \\
&\leqslant \frac{|C|}{n} \sum_{x=1}^{n} \left(\left\|\mathrm{Tr}_C[\psi'_x(\zeta^{MC} - \zeta^M \otimes \zeta^C)]\right\|_1 + \left\|\mathrm{Tr}_C[\xi_x(\zeta^{MC} - \zeta^M \otimes \zeta^C)]\right\|_1\right) \\
&\leqslant \frac{|C|}{n} \sum_{x=1}^{n} \left(\left\|\mathrm{Tr}_C[\psi'_x(\zeta^{MC} - \zeta^M \otimes \zeta^C)]\right\|_1 + \left\|\mathrm{Tr}_C[\xi_x\zeta^{MC}]\right\|_1 + \left\|\mathrm{Tr}_C[\xi_x(\zeta^M \otimes \zeta^C)]\right\|_1\right) \\
&\leqslant \frac{|C|}{n} \sum_{x=1}^{n} \left(\left\|\mathrm{Tr}_C[\psi'_x(\zeta^{MC} - \zeta^M \otimes \zeta^C)]\right\|_1 + \frac{2\varepsilon}{|C|}\right) \\
&= \left\|\mathcal{M}'(\zeta^{MC} - \zeta^M \otimes \zeta^C)\right\|_1 + 2\varepsilon
\end{aligned}$$

Now, we have that $|\mathfrak{Q}| \leqslant |\mathfrak{N}^n| \leqslant \left(\frac{5}{\varepsilon}\right)^{2n|C|}$. Hence, by the union bound and Lemma 6.4, we get the lemma. □

We need to use this to get a bound on general measurement superoperators. The idea will be to imagine that, given any measurement operator, we perform n independent measurements on n i.i.d. copies of ρ. The operator Chernoff bound (Lemma I.8) will then ensure that the resulting sequence of measurement results is an (n, k)-quasi-measurement with high probability.

Lemma 6.5. *Let $\mathcal{M}^{C\to X}$ be any complete measurement superoperator, with $\mathcal{M}(\pi) = \sum_x \alpha_x |x\rangle\langle\psi_x|\pi|\psi_x\rangle\langle x|$, and consider the operator-valued random variable Y which takes the value $|\psi_x\rangle\langle\psi_x|$ with probability $\alpha_x\langle\psi_x|\pi|\psi_x\rangle = \alpha_x/|C|$. Then, n i.i.d. copies of Y will fail to be an (n, k)-quasi-measurement with probability at most $2|C|e^{-n(k-1)^2/|C|2\ln 2}$.*

Proof. Y fulfills all the conditions for the operator Chernoff bound (Lemma I.8) to apply, with $\mathbb{E}Y = \pi^C$. This yields

$$\Pr\left\{\frac{1}{n}\sum_{j=1}^{n} Y_j \nleqslant k\pi\right\} \leqslant 2|C|e^{-n(k-1)^2/|C|2\ln 2}$$

and the probability on the left is an upper bound on the probability that the sequence $Y_1, \ldots, Y_n$ is not an (n,k)-quasi-measurement. $\square$

Putting all the pieces together, we finally get the main theorem of this section:

Theorem 6.6. *There exists a U^{CK} such that for all measurement operators $\mathcal{M}^{C\to X}$,*

$$\left\|\mathcal{M}(\mathrm{Tr}_K[U\cdot\rho^{MCK}]) - \mathcal{M}(\pi^C)\otimes\rho^M\right\|_1 \leqslant 7\varepsilon$$

as long as $N \geqslant \frac{8\sqrt{2}}{\varepsilon}$, $\varepsilon \leqslant e^{-2}$, and $|K| \geqslant \frac{32}{\varepsilon}\sqrt{\log\left(\frac{4N^2}{\varepsilon}\right)\ln(1/\varepsilon)}$.

Proof. Let $\mathcal{M}^{C\to X}$ be any complete measurement superoperator of the form $\mathcal{M}(\sigma) = \sum_x \alpha_x |x\rangle\langle\psi_x|\sigma|\psi_x\rangle\langle x|$, and define Y to be the operator-valued RV which takes value ψ_x with probability $\alpha_x/|C|$. Let Q be the event that $Y_1,\ldots,Y_n$ is an (n,k)-quasi-measurement, where the Y_i are i.i.d. with the same distribution as Y. Now, assuming U fulfills the requirements of Lemma 6.4, we have that

$$\begin{aligned}
&\left\|\mathcal{M}(\mathrm{Tr}_K[U\cdot\rho^{MCK}]) - \mathcal{M}(\pi^C)\otimes\rho^M\right\|_1 \\
&\quad= \sum_x \alpha_x \left\|\mathrm{Tr}_C[\psi_x(\mathrm{Tr}_K[U\cdot\rho^{MCK}] - \pi^C\otimes\rho^M)]\right\|_1 \\
&\quad= |C|\mathbb{E}_Y \left\|\mathrm{Tr}_C[Y(\mathrm{Tr}_K[U\cdot\rho^{MCK}] - \pi^C\otimes\rho^M)]\right\|_1 \\
&\quad= \frac{|C|}{n}\mathbb{E}_{Y_1,\ldots,Y_n}\sum_{i=1}^{n} \left\|\mathrm{Tr}_C[Y_i(\mathrm{Tr}_K[U\cdot\rho^{MCK}] - \pi^C\otimes\rho^M)]\right\|_1 \\
&\quad= \frac{|C|}{n}\Pr\{Q\}\mathbb{E}\left[\sum_{i=1}^{n} \left\|\mathrm{Tr}_C[Y_i(\mathrm{Tr}_K[U\cdot\rho^{MCK}] - \pi^C\otimes\rho^M)]\right\|_1 \middle| Q\right] \\
&\qquad+ \frac{|C|}{n}\Pr\{\bar{Q}\}\mathbb{E}\left[\sum_{i=1}^{n} \left\|\mathrm{Tr}_C[Y_i(\mathrm{Tr}_K[U\cdot\rho^{MCK}] - \pi^C\otimes\rho^M)]\right\|_1 \middle| \bar{Q}\right] \\
&\quad\leqslant 2^{\frac{1}{2}\log k - \frac{1}{2}\log|K|} + 4\varepsilon + r + 4|C|^2 e^{-n(k-1)^2/|C|2\ln 2}
\end{aligned}$$

In the above, we have bounded the first conditional expectation using Lemma 6.4, with the 2ε going to 4ε due to the fact that, by definition, any (n, k)-quasi-measurement is ε-close to element of $\mathfrak{Q}$. The second conditional expectation was simply upper bounded by $2n$ (i.e. each trace distance in the sum cannot exceed 2) and we used Lemma 6.5 to bound $\Pr\{\bar{Q}\}$.

All that is left to do is to choose the various constants such that $2n \ln(5/\varepsilon)|C| - N^2r^2/16k^2 < -\ln 2$ as imposed by Lemma 6.4, and such that $4|C|^2 e^{-n(k-1)^2/|C|2\ln 2} \leqslant \varepsilon$. Setting $k = 2$ and $r = \varepsilon$ and doing a few simple computations yields that this is possible as long as

$$n \leqslant \frac{N^2\varepsilon^2}{512|C|\ln(1/\varepsilon)}$$

and

$$n \geqslant 2|C| \log \frac{4|C|^2}{\varepsilon}$$

given that $N \geqslant \frac{8\sqrt{2}}{\varepsilon}$ and $\varepsilon \leqslant e^{-2}$. It follows that choosing K such that

$$|K| \geqslant \frac{32}{\varepsilon}\sqrt{\log\left(\frac{4N^2}{\varepsilon}\right)\ln(1/\varepsilon)}$$

suffices to ensure that there exists a U^{CK} such that

$$\left\|\mathcal{M}(\mathrm{Tr}_K[U \cdot \rho^{MCK}]) - \mathcal{M}(\pi^C) \otimes \rho^M\right\|_1 \leqslant 7\varepsilon$$

□

6.2 Implications for the security of quantum protocols against quantum adversaries

When designing quantum cryptographic protocols, it is often necessary to show that a quantum adversary ("Eve") is left with only a negligible amount of information on some secret string. An initial attempt at formalizing this idea is to say that, at the end of the protocol, regardless of what measurement Eve makes on her quantum system, the mutual information between her measurement result and the secret string is at most ε (in other words, her accessible information about the message is at most ε). This was often taken as the security definition for quantum key distribution, usually implicitly by simply not considering that the adversary might keep quantum data at the end of the protocol [LC99, SP00, NC00, GL03, LCA05] (see also discussion in [BOHL$^+$05, RK05, KRBM07]). In [KRBM07], it is shown that this definition of security is inadequate, precisely because of possible locking effects. Indeed, this security definition does not exclude the possibility that Eve, upon gaining partial knowledge of S after the end of the protocol, could then gain more by making a measurement on her quantum register that depends on the partial information that she has learned. In [KRBM07], the authors exhibit a (somewhat contrived) quantum key distribution protocol which generates a secret n-bit key such that, if Eve learns the first $n-1$ bits, she can then learn the remaining bit by measuring her own quantum register.

The locking scheme presented above allows us to demonstrate a much more spectacular failure of this security definition. We will show that there exists a quantum key distribution protocol that ensures that an adversary has negligible accessible information about the final key, but with which an adversary can recover the entire key upon learning only a very small fraction of it.

6.2.1 Description of the protocol

We will derive this protocol by taking a protocol that is truly secure, and then making Alice send a locked version of the secret string directly to Eve. We will be able to prove that regardless of what measurement Eve makes on her state, she will learn essentially no information on the string, but of course, she only needs to learn a tiny amount of information to unlock what Alice sent her. More precisely, let P be a quantum key distribution protocol such that, at the end of its execution, Alice and Bob share an n-bit string, and Eve has a quantum state representing everything that she has managed to learn about the string. We will also assume that P is a truly secure protocol: the string together with Eve's quantum state can be represented as a quantum state σ^{SE} such that $\|\sigma^{SE} - \pi^S \otimes \sigma^E\|_1 \leqslant \varepsilon$, where S is a quantum register holding the secret string, and E is Eve's quantum register. Now, we will define the protocol P' to be the following quantum key distribution protocol: Alice and Bob first run P to generate a string s of length n, and then Alice splits s into two parts: the first part s_k is of size $\log\left(32/\varepsilon\sqrt{\log\left(\frac{4\cdot 2^{2n}}{\varepsilon}\right)\ln(1/\varepsilon)}\right)$, and the second part s_c contains the rest of the key. Alice then uses the classical key s_k to create a quantum state in register C that contains a locked version of s_c and sends the system C to Eve.

How secure is P'? It is clearly very insecure, since, if Eve ever ends up learning s_k (via a known plaintext attack, for instance), she can then completely recover s_c. However, the next theorem shows that, right after the execution of P', Eve cannot make any measurement that will reveal information about the key. In particular, P' satisfies the requirement that Eve's accessible information on the key be very low.

Theorem 6.7. *Let P and P' be quantum key distribution protocols as defined as above, and let ρ^{CES} be the state at the end of the execution of P': S contains the n-bit string s, E is Eve's quantum register after the execution of P, and C contains the locked version of s_c that Alice sent to Eve. Then, for any measurement superoperator $\mathcal{M}^{CE\to X}$, there exists a state ξ^X such that*

$$\left\|\mathcal{M}(\rho^{CES}) - \xi^X \otimes \pi^S\right\|_1 \leqslant 2\varepsilon.$$

This also entails that

$$I_{\text{acc}}(S;CE) \leqslant 2\varepsilon n + 2\eta(1-2\varepsilon) + 2\eta(2\varepsilon)$$

via the Alicki-Fannes inequality (Lemma I.10).

Proof. From the definition of P, we have that

$$\left\|\rho^{ES} - \pi^S \otimes \rho^E\right\|_1 \leqslant \varepsilon. \tag{6.1}$$

Now, let $\mathcal{C}^{S\to CS}$ be a superoperator that takes a classical string in S, splits it into s_k and s_c, creates a locked version of s_c with s_k as the key into the quantum system C, and leaves the classical string in S unchanged; this is simply the operation that Alice performs when preparing C for Eve. The above inequality, combined with the monotonicity of the trace distance under CPTP maps yields

$$\left\|\rho^{CES} - \mathcal{C}(\pi^S) \otimes \rho^E\right\|_1 \leqslant \varepsilon \tag{6.2}$$

and hence, for any measurement superoperator $\mathcal{M}^{CE\to X}$,

$$\left\|\mathcal{M}(\rho^{CES}) - \mathcal{M}(\mathcal{C}(\pi^S) \otimes \rho^E)\right\|_1 \leqslant \varepsilon \tag{6.3}$$

Consider now the expression $\mathcal{M}^{CE\to X}(\mathcal{C}(\pi^S) \otimes \rho^E)$: it can be viewed as a measurement on the C system of $\mathcal{C}^{S\to CS}(\pi^S)$ alone that is implemented by creating the state ρ^E and then measuring $\mathcal{M}^{CE\to X}$. Furthermore, note that, by the definition of an ε-locking scheme, we have that, for every measurement superoperator $\mathcal{N}^{C\to X}$,

$$\left\|\mathcal{N}(\mathcal{C}(\pi^S)) - \mathcal{N}(\mathrm{Tr}_S[\mathcal{C}(\pi^S)]) \otimes \pi^S\right\|_1 \leqslant \varepsilon. \tag{6.4}$$

Applying this to $\mathcal{M}^{CE\to X}(\mathcal{C}(\pi^S) \otimes \rho^E)$, we get that

$$\left\|\mathcal{M}(\mathcal{C}(\pi^S) \otimes \rho^E) - \mathcal{M}(\mathrm{Tr}_S[\mathcal{C}(\pi^S)] \otimes \rho^E) \otimes \pi^S\right\|_1 \leqslant \varepsilon. \tag{6.5}$$

We now use the triangle inequality on Equations (6.3) and (6.5) to obtain

$$\left\|\mathcal{M}(\rho^{CES}) - \mathcal{M}(\mathrm{Tr}_S[\mathcal{C}(\pi^S)] \otimes \rho^E) \otimes \pi^S\right\|_1 \leqslant 2\varepsilon \tag{6.6}$$

which yields the theorem with $\xi^X := \mathcal{M}(\mathrm{Tr}_S[\mathcal{C}(\pi^S)] \otimes \rho^E)$. □

Hence, we have shown that requiring that Eve's accessible information on the generated key be low is not an adequate definition of security for quantum key distribution. We have exhibited a protocol P' which guarantees low accessible information and yet is clearly insecure due to locking effects.

6.3 Discussion

The essence of the locking phenomenon is that it is possible to possess *purely* quantum information about a classical message: the cyphertext by itself must contain a lot of information about the message, since only a tiny key is required to get the message, but none of it can be considered classical, since no measurement succeeds in extracting this information. This phenomenon has particular importance in cryptography: it highlights the need to consider an adversary having access to quantum memory, since it is possible for a protocol to ensure that no adversary has any classical information about a particular string while having a lot of quantum information about it. The adversary then needs only a very small amount of additional information to unlock his quantum information. This essentially means that security definitions in cryptography must take quantum information into account to be composable in the physical world.

The main improvement of this work over previous locking schemes is the fact that locking is defined in terms of a trace distance between measurement outputs rather than in terms of accessible information. This is strictly stronger, and has a more compelling interpretation: measurements made on a locked message cannot be distinguished with more than negligible probability from data generated independently of the message. Furthermore, it demonstrates the failure of cryptographic security definitions based on measurement results even more flagrantly: previous results [KRBM07] showed that there exists a quantum key distribution protocol that produces an n-bit key about which no adversary can obtain significant information through a measurement, but for which there can exist a quantum adversary who, upon learning the first $n-1$ bits of the key, can then learn the last one by measuring his quantum data. In this work, the quantum adversary only needs to get $\mathrm{polylog}(n)$ bits on the key before being able to reconstruct the entire key, rather than $n-1$ bits.

CHAPTER 7

CONCLUSION

In this thesis, we have developed a set of mathematical tools to solve quantum information theory problems within a unified framework. These tools are based on the idea of *decoupling*: in the quantum world, ensuring that two systems are uncorrelated implies that both of these systems are completely correlated with a third system that purifies the state that they are holding. Hence, the problem of information transmission, which can be viewed as the problem of establishing perfect correlation between a sender and a receiver, can be solved by *destroying* correlation between the sender and an "environment" system that purifies the global state. Chapter 3 presents this concept in detail and gives a general theorem that allows us to ensure that two systems are decorrelated. This theorem analyzes the following situation: we have a quantum channel $\mathcal{T}^{A\to E}$ and a quantum state ρ^{AR}, we apply a unitary U on the A system of ρ (a unitary chosen at random uniformly over the set of all unitaries works on average) and then we send A into the input of $\mathcal{T}$. The result is that the quality of decorrelation only depends on two parameters: one that indicates how easy it is to decorrelate the state and the other that measures how good the channel is at decorrelating. Several different versions of this theorem are presented to adapt it to different uses.

The rest of the thesis then goes on to apply these tools to more concrete information theory problems, allowing us to obtain new theorems as well as many of the most important theorems in the field, often in a more general form. These include the best known achievable rate for quantum transmission through quantum channels and the entanglement-assisted capacities of quantum channels for classical and quantum transmission. It also allowed us to come up with hitherto unknown coding theorems on quantum channels with side-information at the transmitter, as well as quantum broadcast channels.

In all of these cases, the coding theorems followed the same pattern: we first obtain a theorem that applies to a single use of a channel, with the quality of transmission depending on various min- and max-entropies. We then specialize these theorems to the case where the "single channel" in question is actually n copies of the same channel, yielding an asymptotic result. In this process, the min- and max-entropies are bounded using the fully quantum asymptotic equipartition property, and turn out to become von Neumann entropies.

We end the thesis in Chapter 6 with an application of decoupling of a slightly different flavour: locking classical information in quantum states. This involves encoding a classical message into two quantum systems: a large one (the cyphertext) that is almost as large as the message itself, and a very small one (the key). The encoding has the property that, given only the cyphertext, no measurement can yield any significant amount of information about the message, even though the cyphertext and the key together provide full information about the message. In contrast with previous work on locking, the definition of locking used here involves a trace distance between two classical distributions: results of a measurement made on a locked message, and measurement results generated independently of the message; this is both a stronger condition and has the clear operational interpretation that a locked message is virtually indistinguishable from a random state when a measurement is made.

7.1 Open problems and future research directions

There are several open problems and possible research projects that arise out of the results presented in this thesis. Here are some of them:

A constructive version of the locking scheme: The results presented in this thesis involve a random unitary chosen according to the Haar measure in some way. This yields proofs that certain protocols exist, but does not directly give a way of actually constructing them. For almost all of the results in this thesis, however, one can replace the Haar measure by a unitary 2-design, since only the second moment of the Haar measure is necessary for the proofs. But there is one exception: the information locking scheme of Chapter 6. This is because its proof relies not only on the second moment of the Haar distribution, but also on its concentration properties (Theorem 3.9). Indeed, Theorem 3.9 states that, in the main decoupling theorem (Theorem 3.7), not only do we get good decoupling on average when choosing a unitary randomly, but also that this holds with overwhelmingly high probability. Statements of this nature abound in quantum information theory and its applications are far from being limited to information locking: it can be used to show the existence of completely entangled subspaces [HLW06], to prove the existence of counterexamples to the additivity conjecture [Has09]. In all of these cases (as well as locking), it would be of great interest to have explicit, constructive examples. Finding a constructive version of Theorem 3.9 (or perhaps something slightly more general) would most likely achieve this for all of the problems mentioned.

Min-entropy bounds for larger classes of states: For all of the channel coding problems shown in this document, we have proceeded as follows: we first gave a general one-shot coding theorem, and then we used it to give a theorem for memoryless channels. To do this, we used the fully quantum asymptotic equipartition property (Theorem 2.4, [TCR08]) to get a bound on the smooth min-entropy of i.i.d. states. If we had a way to similarly bound the smooth min-entropy of a larger class of states, we could apply it to a larger class of channels, such as various types of channels with memory.

Optimality of the one-shot coding theorems: The various one-shot coding theorems presented were left without converses. However, it seems likely that they are, in fact, optimal, at least for some particular input distributions. Some special cases have already been shown to be optimal, such as state merging [BCR09].

Systematically relating classical information theory and quantum information theory with free entanglement: In quantum Shannon theory, it has very often proven to be the case that the quantum problems that bear the strongest resemblance to their classical counterparts are those in which the various participants share entanglement before the protocol starts and are allowed to use it to improve the performance of the protocol. This is the case for information transmission through a regular channel: Shannon showed that the mutual information gives the capacity of a classical channel; and it turns out that the quantum mutual information characterizes the *entanglement-assisted* capacity of quantum channels. This is also true for channels with side-information at the transmission (see Chapter 4) as well as broadcast channels (see Chapter 5). In all of these cases, we get essentially identical expressions for the capacities (or achievable rate regions) in the classical and in the quantum case. This suggests that there might be a general principle at work relating the two. Such a principle would allow us to automatically import large classes of results from the extremely vast body of work in classical information theory directly into quantum information theory. It is not clear at this point, however, to what extent this principle would apply, or what the most appropriate definitions would be.

BIBLIOGRAPHY

[ADHW06] Anura Abeyesinghe, Igor Devetak, Patrick Hayden, and Andreas Winter. The mother of all protocols: Restructuring quantum information's family tree. 2006. quant-ph/0606225.

[AF04] Robert Alicki and Mark Fannes. Continuity of quantum mutual information. *Journal of Physics A: Mathematical and General*, 37(5):L55–L57, 2004. quant-ph/0312081.

[AGZ09] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2009. http://www.wisdom.weizmann.ac.il/ zeitouni/cupbook.pdf.

[AS00] Armen E. Allahverdyan and David B. Saakian. Broadcast of classical information through a quantum channel. *Europhysics Letters*, 6(50):718–723, 2000.

[AW02] Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002. quant-ph/0012127.

[BCR09] Mario Berta, Matthias Christandl, and Renato Renner. A conceptually simple proof of the quantum reverse shannon theorem. 2009. arXiv:0912.3805.

[BD09] Francesco Buscemi and Nilanjana Datta. The quantum capacity of channels with arbitrarily correlated noise. 2009. arXiv:0902.0158.

[BDH$^+$06] Charles Bennett, Igor Devetak, Aram Harrow, Peter Shor, and Andreas Winter. The quantum reverse Shannon theorem. In preparation, 2006.

[Ber08] Mario Berta. Single-shot quantum state merging. Diploma Thesis, ETH Zurich, 2008.

[Bha96] Rajendra Bhatia. *Matrix Analysis*. Springer-Verlag, 1996.

[BOHL+05] Michael Ben-Or, Michal Horodecki, Debbie Leung, Dominic Mayers, and Jonathan Oppenheim. The universal composable security of quantum key distribution. *Second Theory of Cryptography Conference, TCC 2005*, 3378:386–406, 2005. quant-ph/0409078.

[BSST02] Charles H. Bennett, Peter W. Shor, John A. Smolin, and Ashish V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Trans. Info. Theory*, 48:10:2637–2655, 2002. quant-ph/0106052.

[Car07] Constantin Carathéodory. Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. *Math. Ann.*, 64:95–115, 1907.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17:230–261, April 1988.

[Cho75] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear algebra and its applications*, (10):285, 1975.

[Cov72] Thomas Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18:2–14, 1972.

[CS06] Benoît Collins and Piotr Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264:773–795, 2006. math-ph/0402073.

[CT91] Thomas Cover and Joy Thomas. *Elements of Information Theory*. John-Wiley and Sons, 1991.

[CW04] Matthias Christandl and Andreas Winter. Squashed entanglement - an additive entanglement measure. *J. Math. Phys.*, 45(3):829–840, 2004.

[DD] Simon Pierre Desrosiers and Frédéric Dupuis. Quantum entropic security and approximate quantum encryption. *IEEE Transactions on Information Theory*. to appear. arXiv:0707.0691.

[Dev05] Igor Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Info. Theory*, (51(1):44), 2005. quant-ph/0304127.

[DH06a] Maciej Demianowicz and Pawel Horodecki. Capacity regions for multiparty quantum channels. 2006.

[DH06b] Maciej Demianowicz and Pawel Horodecki. Quantum channel capacities - multiparty communication. 2006.

[DHL$^+$04] David P. DiVincenzo, Michał Horodecki, Debbie W. Leung, John A. Smolin, and Barbara M. Terhal. Locking classical correlation in quantum state. *Phys. Rev. Lett.*, (92, 067902), 2004. quant-ph/0303088.

[DHL09] Frédéric Dupuis, Patrick Hayden, and Ke Li. A father protocol for quantum broadcast channels. 2009.

[DHW03] Igor Devetak, Aram Harrow, and Andreas Winter. A family of quantum protocols. *Phys. Rev. Lett.*, 93(230504), 2003.

[DHW05] Igor Devetak, Aram Harrow, and Andreas Winter. A resource framework for quantum Shannon theory. 2005. quant-ph/0512015.

[DLT02] David DiVincenzo, Debbie Leung, and Barbara Terhal. Quantum data hiding. *IEEE Trans. Inf. Theory*, 48(3):580–598, 2002. quant-ph/0103098.

[DSS98] David P. DiVincenzo, Peter W. Shor, and John A. Smolin. Quantum-chJohn A. Smolinof very noisy channels. *Phys. Rev. A*, 57:830–839, February 1998. quant-ph/9706061.

[Dup09] Frédéric Dupuis. The capacity of quantum channels with side information at the transmitter. *Proceedings of the 2009 IEEE International Symposium on Information Theory*, 2009. arXiv:0805.3352.

[DW05] Igor Devetak and Andreas Winter. Distillation of secret key and entanglement from quantum states. *Proc. R. Soc. Lond. A*, (461):207–237, 2005. quant-ph/0306078.

[DY06] Igor Devetak and Jon Yard. The operational meaning of quantum conditional information. 2006. quant-ph/0612050.

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, (47):777, 1935.

[Fan73] Mark Fannes. A continuity property of the entropy density for spin lattices. *Commun. Math. Phys.*, 31:291–294, 1973.

[FvdG99] Christopher Fuchs and Jeroen van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Info. Theory*, 45(4):1216–1227, May 1999. quant-ph/9712042.

[GL03] Daniel Gottesman and Hoi-Kwong Lo. Proof of security of quantum key distribution with two-way classical communications. *IEEE Transactions on Information Theory*, 49(2):457–475, 2003. quant-ph/0105121.

[GP80] Sergei I. Gel'fand and Mark S. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9:19–31, 1980.

[GPW05] Berry Groisman, Sandu Popescu, and Andreas Winter. On the quantum, classical and total amount of correlations in a quantum state. *Phys Rev A*, 72(032317), 2005. quant-ph/0410091.

[GS07] Saikat Guha and Jeffrey H. Shapiro. Classical information capacity of the bosonic broadcast channel. 2007.

[Has09] Matthew B. Hastings. A counterexample to the additivity of minimum output entropy. *Nature Physics*, 5(255), 2009. arXiv:0809.3972.

[Hay06] Patrick Hayden. Unpublished, 2006.

[HDW08] Min-Hsiu Hsieh, Igor Devetak, and Andreas Winter. Entanglement-assisted capacity of quantum multiple-access channels. *IEEE Transactions on Information Theory*, 54–7:3078–3090, July 2008. quant-ph/0511228.

[Hel69] Carl W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2):231–252, June 1969.

[HHYW08] Patrick Hayden, Michał Horodecki, Jon Yard, and Andreas Winter. A decoupling approach to the quantum capacity. *Open Syst. Inf. Dyn.*, (15):7–19, 2008. quant-ph/0702005.

[HIN$^+$06] Masahito Hayashi, Kazuo Iwama, Harumichi Nishimura, Rudy Raymond, and Shigeru Yamashita. Quantum network coding. 2006.

[HLSW04] Patrick Hayden, Debbie Leung, Peter Shor, and Andreas Winter. Randomizing quantum states: Constructions and applications. *Comm. Math. Phys.*, 250(2):371–391, 2004.

[HLW06] Patrick Hayden, Debbie Leung, and Andreas Winter. Aspects of generic entanglement. *Comm. Math. Phys.*, 265(1):95–117, 2006. quant-ph/0407049.

[HOW07] Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Comm. Math. Phys.*, (269, 107), 2007. quant-ph/0512247.

[HV94] Te Sun Han and Sergio Verdú. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40(4):1147–1157, 1994.

[Jam72] Andrzej Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972.

[Kit97] Alexei Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[Kli01] Gleb Klimovitch. On the classical capacity of a quantum multiple access channel. *Proc. IEEE Intern. Sympos. Info. Theory*, page 278, 2001.

[KRBM07] Robert König, Renato Renner, Andor Bariska, and Ueli Maurer. Locking of accessible information and implications for the security of quantum cryptography. *Phys. Rev. Lett.*, 98(140502), 2007. quant-ph/0512021.

[KRS09] Robert König, Renato Renner, and Christian Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55(9), 2009. arXiv:0807.1338.

[KW03] Dennis Kretschmann and Reinhard F. Werner. Tema con variazioni: quantum channel capacity. 2003. quant-ph/0311037.

[LC99] Hoi-Kwong Lo and Hoi-Fung Chau. Unconditional security of quantum key distribution over arbitrarily long distances. *Science*, 283(5410):2050–2056, 1999. quant-ph/9803006.

[LCA05] Hoi-Kwong Lo, Hoi-Fung Chau, and M. Ardehali. Efficient quantum key distribution scheme and proof of its unconditional security. *Journal of Cryptology*, 18(133), 2005. quant-ph/0011056.

[Llo96] Seth Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, (55:1613), 1996. quant-ph/9604015.

[LLPS09] Ruoheng Liu, Tie Liu, H. Vincent Poor, and Shlomo Shamai. Multiple-input multiple-output Gaussian broadcast channels with confidential messages. 2009. arXiv:0903.3786.

[LOW06] Debbie Leung, Jonathan Oppenheim, and Andreas Winter. Quantum network communication – the butterfly and beyond. 2006.

[LR73] Elliot H. Lieb and Mary Beth Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14:1938–1941, 1973.

[Mar79] Katalin Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Transactions on Information Theory*, IT-25:306–311, 1979.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, New York, NY, USA, 2000.

[Pau02] Vern Paulsen. *Completely bounded maps and operator algebras*. Cambridge University Press, 2002.

[Ren05] Renato Renner. *Security of quantum key distribution*. PhD thesis, ETH Zurich, 2005.

[RK05] Renato Renner and Robert König. Universally composable privacy amplification against quantum adversaries. *Second Theory* *of Cryptography Conference, TCC 2005*, 3378:407–425, 2005. quant-ph/0403133.

[RW04] Renato Renner and Stefan Wolf. Smooth Rényi entropy and applications. *Proceedings of the 2004 International Symposium on Information Theory*, 2004.

[RWW06] Renato Renner, Stefan Wolf, and Jürg Wullschleger. The single-serving channel capacity. *Proceedings of the 2006 International Symposium on Information Theory*, 2006.

[Sch95] Benjamin Schumacher. Quantum coding. *Phys. Rev. A*, (51):2738–2747, 1995.

[Sha48] Claude Elwood Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[Sho02] Peter Shor. The quantum channel capacity and coherent information. *Lecture notes, MSRI workshop on quantum computation*, 2002. Available online at http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/.

[SP00] Peter Shor and John Preskill. Simple proof of security of the bb84 quantum key distribution protocol. *Phys. Rev. Lett.*, 85, 2000. quant-ph/0003004.

[SVW05] John A. Smolin, Frank Verstraete, and Andreas Winter. Entanglement of assistance and multipartite state distillation. *Phys. Rev. A*, 72(5):052317–+, November 2005.

[SY08] Graeme Smith and Jon Yard. Quantum communication with zero-capacity channels. 2008. arXiv:0807.4935.

[TCR08] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. 2008. arXiv:0811.1221.

[TCR09] Marco Tomamichel, Roger Colbeck, and Renato Renner. Duality between smooth min- and max-entropies. 2009. arXiv:0907.5238.

[Uhl76] Armin Uhlmann. The 'transition probability' in the state space of a *-algebra. *Rep. Math. Phys.*, (9:273), 1976.

[vN32] John von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Springer, 1932.

[Wat08] John Watrous. *Theory of Quantum Information—Lecture notes from Fall 2008*. 2008. www.cs.uwaterloo.ca/˜watrous/quant-info/.

[Win01] Andreas Winter. The capacity of the quantum multiple access channel. *IEEE Trans. Info. Theory*, 47:3059–3065, 2001.

[YDH05] Jon Yard, Igor Devetak, and Patrick Hayden. Capacity theorems for quantum multiple access channels: Classical-quantum and quantum-quantum capacity regions. 2005.

[YHD06] Jon Yard, Patrick Hayden, and Igor Devetak. Quantum broadcast channels. 2006.

Appendix I

Various technical lemmas

In this section, we state (and usually prove) various technical lemmas used at various points throughout the thesis.

The first lemma is a simple application of the triangle inequality:

Lemma I.1. *Let ρ, ρ' and σ be positive semidefinite operators on A such that $\|\rho-\sigma\|_1 \leqslant \varepsilon$, $\mathrm{Tr}[\rho'] \leqslant \mathrm{Tr}[\sigma]$, and $\rho' \geqslant \rho$. Then, $\|\rho' - \sigma\|_1 \leqslant 2\varepsilon$.*

Proof. We have that

$$\begin{aligned} \|\rho' - \rho\|_1 &= \mathrm{Tr}[\rho' - \rho] && \text{(I.1)} \\ &\leqslant \mathrm{Tr}[\sigma - \rho] && \text{(I.2)} \\ &\leqslant \varepsilon && \text{(I.3)} \end{aligned}$$

and hence

$$\|\rho' - \sigma\|_1 \leqslant \|\rho - \sigma\|_1 + \|\rho' - \rho\|_1 \leqslant 2\varepsilon. \tag{I.4}$$

□

We then prove the following operator inequalities:

Lemma I.2. *Let ρ^{AB} be positive semidefinite, and let $0 \leqslant P^B \leqslant \mathbb{I}^B$. Then,*

$$\mathrm{Tr}_B[P^B \rho^{AB} P^B] \leqslant \rho^A$$

Proof. Let M^A be any positive semidefinite operator. Then,

$$\begin{aligned}\operatorname{Tr}[M^A \operatorname{Tr}_B[P^B \rho^{AB} P^B]] &= \operatorname{Tr}[(M^A \otimes \mathbb{I}^B)(P^B \rho^{AB} P^B)] \\ &= \operatorname{Tr}[(M^A \otimes P^{B^2})\rho^{AB}] \\ &\leqslant \operatorname{Tr}[(M^A \otimes \mathbb{I}^B)\rho^{AB}] \\ &= \operatorname{Tr}[M^A \rho^A]\end{aligned}$$

where we have used the fact that tensoring with the identity is the adjoint of the trace superoperator, as well as the fact that $P^{B^2} \leqslant \mathbb{I}^B$. Since this is true for every positive semidefinite M^A, the lemma follows. □

Lemma I.3. *Let* $|\psi\rangle^{AB} \in \mathsf{A} \otimes \mathsf{B}$, $\rho^A \in \operatorname{Pos}(\mathsf{A})$ *such that* $\rho^A \leqslant \psi^A$*. Then, there exists a* $P^B \in \operatorname{Pos}(\mathsf{B})$ *such that* $P^B \leqslant \mathbb{I}^B$ *and* $\operatorname{Tr}_B[P^B \cdot \psi^{AB}] = \rho^A$*.*

Proof. Without loss of generality, let A and B be equal to the support of ψ^A and ψ^B respectively. Define the partial isometry $V^{B\to A} = \psi^{A^{-1/2}} \operatorname{op}_{B\to A}(|\psi\rangle) = \operatorname{op}_{B\to A}(|\psi\rangle)\psi_T^{B^{-1/2}}$ where the T subscript denotes transposition. Now,

$$\begin{aligned}\rho^A &= VV^\dagger \rho V V^\dagger \\ &= \operatorname{op}_{B\to A}(|\psi\rangle)\psi_T^{B^{-1/2}} V^\dagger \rho V \psi_T^{B^{-1/2}} \operatorname{op}_{B\to A}(|\psi\rangle) \\ &= \operatorname{op}_{B\to A}(|\psi\rangle) V^\dagger \psi^{A^{-1/2}} \rho \psi^{A^{-1/2}} V \operatorname{op}_{B\to A}(|\psi\rangle) \\ &= \operatorname{op}_{B\to A}(|\psi\rangle) P_T^{B^2} \operatorname{op}_{B\to A}(|\psi\rangle)^\dagger \\ &= \operatorname{op}_{B\to A}(P^B|\psi\rangle) \operatorname{op}_{B\to A}(P^B|\psi\rangle)^\dagger \\ &= \operatorname{Tr}_B[P^B \cdot \psi^{AB}]\end{aligned}$$

where we have defined $P_T^{B^2} = V^\dagger \psi^{A^{-1/2}} \rho \psi^{A^{-1/2}} V \in \operatorname{Pos}(\mathsf{B})$ and the T subscript denotes transposition. We can now easily check that $P_T^{B^2} \leqslant \mathbb{I}^B$ since $\rho \leqslant \psi^A$ implies that $\psi^{A^{-1/2}} \rho \psi^{A^{-1/2}} \leqslant \mathbb{I}^A$. □

The following lemma comes from Lemma II.4 from [HLSW04]:

Lemma I.4. *Given two normalized vectors $|\psi\rangle$ and $|\varphi\rangle$ in A, we have that*

$$\|\psi - \varphi\|_1 \leqslant 2\,\||\psi\rangle - |\varphi\rangle\|_2$$

Proof. By Lemma 3.6 with σ as the projector onto the 2-dimensional support of $\psi - \varphi$, we have that

$$\begin{aligned}
\|\psi - \varphi\|_1 &\leqslant \sqrt{2\operatorname{Tr}[(\psi-\varphi)^2]} \\
&= 2\sqrt{1 - \operatorname{Tr}[\varphi\psi]} \\
&= 2\sqrt{1 - |\langle\psi|\varphi\rangle|^2} \\
&= 2\sqrt{(1 - |\langle\psi|\varphi\rangle|)(1 + |\langle\psi|\varphi\rangle|)} \\
&\leqslant 2\sqrt{2 - 2|\langle\psi|\varphi\rangle|} \\
&\leqslant 2\sqrt{2 - \langle\psi|\varphi\rangle - \langle\varphi|\psi\rangle} \\
&= 2\sqrt{(\langle\psi| - \langle\varphi|)(|\psi\rangle - |\varphi\rangle)} \\
&= 2\||\psi\rangle - |\varphi\rangle\|_2
\end{aligned}$$

□

The next two lemmas are simple inequalities regarding operator norms:

Lemma I.5. *Let $M^{A\to B}$ and $N^{B\to C}$ be arbitrary matrices. Then,*

$$\|NM\|_2 \leqslant \|N\|_2\|M\|_\infty$$

Proof. Let $U^{B\to A}$ be an isometry such that $P^B := MU$ is positive semidefinite (such an isometry can be seen to exist by taking the singular-value decomposition of M). Then, we have that

$$\begin{aligned} \|NM\|_2 &= \|NP\|_2 && \text{(I.5)} \\ &= \sqrt{\mathrm{Tr}[NP^2N^\dagger]} && \text{(I.6)} \\ &\leqslant \|P\|_\infty \sqrt{\mathrm{Tr}[NN^\dagger]} && \text{(I.7)} \\ &= \|M\|_\infty \|N\|_2 && \text{(I.8)} \end{aligned}$$

where the inequality comes from the matrix inequality $P^2 \leqslant \|P\|_\infty^2 \mathbb{I}$. □

Lemma I.6. *Let $M^{A\to B}$ be an arbitrary matrix. Then,*

$$\|M\|_1 = \max_{V^{B\to A}} |\,\mathrm{Tr}[VM]|$$

where the maximization is taken over all partial isometries $V^{B\to A}$.

Proof. Let us decompose M as $M = \sum \alpha_j |\psi_j\rangle\langle\varphi_j|$ where the $|\psi_j\rangle^B$ are orthonormal, as are the $|\varphi_j\rangle^A$, and the α_j are the singular values of M. Furthermore, let $W^{B\to A}$ be a partial isometry such that $W|\psi_j\rangle = |\varphi_j\rangle$. Then,

$$\begin{aligned} \|M\|_1 &= |\,\mathrm{Tr}[WM]| \\ &\leqslant \max_{V^{B\to A}} |\,\mathrm{Tr}[VM]| \\ &= \max_V \left| \sum_j \alpha_j \,\mathrm{Tr}[V|\psi_j\rangle\langle\varphi_j|] \right| \\ &\leqslant \max_V \sum_j \alpha_j \,|\langle\varphi_j|V|\psi_j\rangle| \\ &\leqslant \sum_j \alpha_j \\ &= \|M\|_1 \end{aligned}$$

□

The next lemma is simply Markov's inequality, which we use several times to assert the existence of a unitary satisfying many conditions at once:

Lemma I.7 (Markov's inequality)**.** *Let X be a random variable which is always positive. Then,*

$$\Pr\{X \geqslant k\mathbb{E}X\} \leqslant \frac{1}{k}$$

Hence, for example, if $f_1, \ldots f_k : \mathbb{U} \to \mathbb{R}_+$, then, there exists a U such that

$$\begin{aligned} f_1(U) &\leqslant (k+1)\mathbb{E}f_1(U) \\ &\vdots \\ f_k(U) &\leqslant (k+1)\mathbb{E}f_k(U) \end{aligned}$$

by the union bound.

The next lemma is known as the operator Chernoff bound and was first proven in [AW02]:

Lemma I.8 (Operator Chernoff bound)**.** *Let $X_1, \ldots, X_M$ be i.i.d. random variables taking values in the operators* $\mathrm{Pos}(\mathsf{A})$*, with $0 \leqslant X_j \leqslant \mathbb{I}$, with $A = \mathbb{E}X_j \geqslant \alpha\mathbb{I}$, and let $0 < \eta \leqslant 1/2$. Then*

$$\Pr\left\{\frac{1}{M}\sum_{j=1}^{M} X_j \nleqslant (1+\eta)A\right\} \leqslant 2|A| \exp\left(-M\frac{\alpha\eta^2}{2\ln 2}\right). \tag{I.9}$$

We also need Fannes's inequality [Fan73] as well as its relative, the Alicki-Fannes inequality [AF04]:

Lemma I.9 (Fannes's inequality [Fan73])**.** *Let ρ and σ be density operators on A such that $\|\rho - \sigma\|_1 \leqslant 1/e$. Then,*

$$|H(A)_\rho - H(A)_\sigma| \leqslant \|\rho - \sigma\|_1 \log|A| + \eta\left(\|\rho - \sigma\|_1\right)$$

where $\eta(x) := -x\log x$ and e is the base of the natural logarithm.

Lemma I.10 (Alicki-Fannes inequality [AF04]). *Given two states $\rho^{AB} \in \mathrm{D}(\mathsf{A} \otimes \mathsf{B})$ and $\sigma^{AB} \in \mathrm{D}(\mathsf{A} \otimes \mathsf{B})$, with $\|\rho^{AB} - \sigma^{AB}\|_1 = \varepsilon$, the following holds:*

$$|H(A|B)_\rho - H(A|B)_\sigma| \leqslant 4\varepsilon \log |A| + 2\eta(1-\varepsilon) + 2\eta(\varepsilon)$$

where η is defined as above.

The locking chapter needs the concept of ε-nets. The following definition and lemma were taken from [HLSW04], but these concepts are used rather extensively in other areas of mathematics, particularly in random matrix theory.

Definition I.1 (ε-net). *A set of pure states $\mathfrak{N} \subseteq \mathsf{A}$ is called an ε-net if, for every normalized $|\psi\rangle \in \mathsf{A}$, there exists a $|\varphi\rangle \in \mathfrak{N}$ such that $\||\psi\rangle - |\varphi\rangle\|_2 \leqslant \varepsilon/2$ and $\|\psi - \varphi\|_1 \leqslant \varepsilon$.*

Lemma I.11 (Existence of small nets). *For any Hilbert space A of dimension $|A|$, there exists an ε-net $\mathfrak{N} \subseteq \mathsf{A}$ of size $|\mathfrak{N}| \leqslant \left(\frac{5}{\varepsilon}\right)^{2|A|}$.*