Quantum Compression and Quantum Learning via Information Theory

by

Shima Bab Hadiashar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Combinatorics and Optimization (Quantum Information)

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Marco Tomamichel Associate Professor, Dept. of Electrical and Computer Engineering, Centre for Quantum Technologies, National University of Singapore
Supervisor(s):	Ashwin Nayak Professor, Dept. of Combinatorics & Optimization, Institute for Quantum Computing, University of Waterloo
Internal Member:	Jon Yard Associate Professor, Dept. of Combinatorics & Optimization, Institute for Quantum Computing, University of Waterloo, and Perimeter Institute for Theoretical Physics, Waterloo
	John Watrous Professor, Cheriton School of Computer Science, Institute for Quantum Computing, University of Waterloo
Internal-External Member:	Vern Paulsen Professor, Dept. of Pure Mathematics, Institute for Quantum Computing, University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis consists of two parts: quantum compression and quantum learning theory. A common theme between these problems is that we study them through the lens of information theory.

We first study the task of visible compression of an ensemble of quantum states with entanglement assistance in the one-shot setting. The protocols achieving the best compression use many more qubits of shared entanglement than the number of qubits in the states in the ensemble. Other compression protocols, with potentially higher communication cost, have entanglement cost bounded by the number of qubits in the given states. This motivates the question as to whether entanglement is truly necessary for compression, and if so, how much of it is needed. We show that an ensemble given by Jain, Radhakrishnan, and Sen (ICALP'03) cannot be compressed by more than a constant number of qubits without shared entanglement, while in the presence of shared entanglement, the communication cost of compression can be arbitrarily smaller than the entanglement cost.

Next, we study the task of quantum state redistribution, the most general version of compression of quantum states. We design a protocol for this task with communication cost in terms of a measure of distance from quantum Markov chains. More precisely, the distance is defined in terms of quantum max-relative entropy and quantum hypothesis testing entropy. Our result is the first to connect quantum state redistribution and Markov chains and gives an operational interpretation for a possible one-shot analogue of quantum conditional mutual information. The communication cost of our protocol is lower than all previously known ones and asymptotically achieves the well-known rate of quantum conditional mutual information.

In the last part, we focus on quantum algorithms for learning Boolean functions using quantum examples. We consider two commonly studied models of learning, namely, quantum PAC learning and quantum agnostic learning. We reproduce the optimal lower bounds by Arunachalam and de Wolf (JMLR'18) for the sample complexity of either of these models using information theory and spectral analysis. Our proofs are simpler than the previous ones and the techniques can be possibly extended to similar scenarios.

Acknowledgements

The research in this thesis is conducted under the supervision and in collaboration with my supervisor, Ashwin Nayak. I would like to deeply thank him for his continuous support and many constructive discussions without which this research would not have been possible. I would also like to thank my other collaborators in the work presented in Chapter 3, Anurag Anshu, Rahul Jain, and Dave Touchette. I am especially grateful to Anurag for lots of fruitful discussions. I would also like to thank members of the committee, Vern Paulsen, John Watrous, and Jon Yard for their time and feedback during the completion of my Ph.D. program. Thanks for his time and willingness to participate in the defense procedure are due as well to the external examiner, Marco Tomamichel.

I want to thank all my friends in Waterloo for their great company during these years. You all made this journey enjoyable. I am deeply grateful to my parents for their unfailing love and never-ending support. Without you, I would not be the person I am today. Above all, I would like to thank my love, Ala, for his unconditional support, and encouragement during all this time.

Dedication

This is dedicated to the one I love.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

Information theory is the field of studying the quantification, storage, and communication of information, founded in the 20-th century by Claude Shannon. In his seminal paper "A Mathematical Theory of Communication" [84], Shannon introduced the notion of *entropy* that is a measure of uncertainty in a source of data, and invented a framework for quantifying information. Using this framework, Shannon single-handedly formalized the optimal rate in lossless source coding as well as channel coding. Information theory has found applications in different fields, from physics, computer science, and electrical engineering to linguistics, neurobiology, and even musical composition.

Since the revelation of the idea of quantum computing in the last decades of the 20-th century, the notions of information theory have been widely extended to the quantum framework and utilized in studying quantum communication and computation. In 1995, Schumacher [81] derived the quantum analogue of the Shannon noiseless source coding, showing that the asymptotic rate for compressing a source of pure quantum states is captured by the *von Neumann entropy* of the average state, the quantum counterpart of Shannon entropy. Source compression (coding) can be considered as a distributed task between two parties, Alice and Bob, where Alice is given a data (a description of a quantum state) from the source (an ensemble of quantum states). Alice wants to send a compressed state to Bob, via a noiseless quantum channel, so that Bob can "approximately" reproduce Alice's input. Schumacher assumed that the source produces *pure* states and there is no initial quantum correlation between two parties. A more general scenario is when the states can be any *mixed* quantum state and parties may have access to initially shared *entangled*

states. This task is known in the literature as *visible quantum compression* or *remote state preparation* (if only noise-less classical channels are available) and has been studied extensively in both asymptotic [54, 23, 25, 66, 50, 24, 18, 17], and one-shot [58, 16] settings. As opposed to visible compression, *blind* compression [54] is the task in which the data given to Alice is a specimen of a quantum state from the ensemble rather than a description of the state. The most general quantum compression scenario is the task of *quantum state redistribution* [69, 38] in which the correlation of the quantum states with the environment as well as local registers of the parties are required to be preserved while transferring the input to Bob with minimal communication. This task is tightly related to the quantum information cost of interactive communication protocols and holds an important place in deriving a direct sum theorem for bounded-round quantum communication complexity [89].

Another field that has benefited from information theory is the *computational learning theory*, founded in the early 1980s by Leslie Valiant [91]. The main goal in this field is to mathematically understand and analyze practically successful machine learning algorithms, like deep learning and other heuristic methods. Information theory has been proved to be a powerful tool for analyzing and investigating the power and limitations of machine learning algorithms(see, e.g., Refs. [10, 45, 80, 12]). Recently, there has also been an increased interest in studying the intersection of (classical) learning theory and quantum computing, called *quantum learning theory*. Quantum machine learning investigates the power of quantum physics in improving machine learning algorithms, as well as the application of machine learning algorithms in quantum computing (see Refs. [11, 39] for a complete survey).

This thesis consists of two main parts. In Part I, we focus on quantum compression. We study the resources required for the tasks of visible quantum compression and quantum state redistribution. In Part II, using quantum information theory, we analyze the limitations of quantum algorithms for learning Boolean functions.

1.1.1 Visible quantum compression

First, we revisit the task of visible compression of an ensemble of quantum states with entanglement assistance in the one-shot setting. The protocols achieving the best compression use many more qubits of shared entanglement than the number of qubits in the states in the ensemble. Other compression protocols, with potentially larger communication cost, have entanglement cost bounded by the number of qubits in the given states. This motivates the question as to whether entanglement is truly necessary for compression, and if so, how much of it is needed.

In Chapter 2, we show that an ensemble of the form designed by Jain, Radhakrishnan, and Sen [60] is incompressible by more than a constant number of qubits without shared entanglement, even when constant error is allowed. Moreover, in the presence of shared entanglement, the communication cost of compression can be arbitrarily smaller than the entanglement cost. The ensemble can also be used to show the impossibility of reducing, via compression, the shared entanglement used in two-party protocols for computing Boolean functions.

1.1.2 Quantum state redistribution

In quantum state redistribution, there is a pure quantum state $|\psi\rangle^{RABC}$ distributed between parties, Alice, Bob, and the environment, called Referee. In this scenario, Register A is held with Alice, register B is held with Bob, register C is the one to be transmitted from Alice to Bob, and register R is the one that purifies registers ABC and is held with Referee. While being well understood in the asymptotic and i.i.d. setting [69, 38] (with the communication rate equal to the quantum conditional mutual information), an explicit near-optimal expression for the communication cost of this task in the one-shot setting is not known. Some achievable one-shot rates are given in Refs. [27, 36, 8], however, they are known to be not optimal for specific pure quantum states. Besides the important role of state redistribution in quantum communication complexity, an explicit expression for communication cost of one-shot state redistribution would result in an operational one-shot interpretation of quantum conditional mutual information.

In Chapter 3, we show a new achievable bound inspired by a representation of quantum conditional mutual information in terms of the minimum relative entropy distance from quantum Markov extensions of the original state. Our bound is tighter than the previously known bounds and, as opposed to those, achieves the near-optimal classical bound of Ref. [6] for the case that ψ^{RBC} is classical. Our key technique is a reduction procedure using embezzling quantum states that allows us to use the protocol of Ref. [8] as a subroutine.

1.1.3 Learning Boolean functions

In the last chapter of this thesis, we study two models of learning Boolean functions: the *probably approximately correct* (PAC) model [91] and the *agnostic* model [63]. Consider a Boolean function f and an unknown distribution D over bit strings of length n. In the PAC model, the learner receives a sequence of *labelled examples* $(x_1, f(x_1)), (x_2, f(x_2)), \ldots$ where $x_1, x_2, \ldots$ are random bit strings independently and identically distributed according

to distribution D. The learner's goal is to learn the function f with high probability, in a sense that given a random input $x \in \{0,1\}^n$ according to the unknown distribution D, the learner should be able to predict $f(x)$ with high probability. In reality, the examples could be possibly noisy. The agnostic model is a more realistic model addressing this problem. In this model, there is no underlying function, but instead labelled samples (x, l) are drawn according to an unknown distribution D over $\{0,1\}^{n+1}$, and the goal is to find a function (from a specific class of functions) predicting the data with minimum possible error. In the quantum setting, the learner receives *quantum examples* and has access to quantum computers for computing the corresponding function f. A quantum example is usually defined as the superposition of all possible labelled examples weighted according to the distribution D [31].

In any of the above-mentioned scenarios, the main quantity of interest is the *sample complexity* that is the minimum number of examples required to learn a specific class of function. It is well-known that the quantum and classical sample complexity of PAC learning and agnostic learning is characterized by a combinatorial parameter called *VC dimension* [29, 47, 94, 85]. In Ref. [12], Arunachalam and de Wolf reproved the optimal lower bounds in the classical learning models via an information-theoretic approach. However, they claimed that the information-theoretic argument is not sufficient for proving tight lower bounds in the quantum setting and used more complicated techniques to achieve this goal.

In Chapter 4, contrary to the claim in Ref. [12], we derive optimal lower bounds on quantum PAC sample complexity and quantum agnostic sample complexity using only information-theoretic arguments and spectral analysis. Our proof is simpler than the previously known ones and may be applied to other similar scenarios.

1.2 Preliminaries

1.2.1 Mathematical notation and background

For a positive integer d, we denote the set $\{0, 1, \ldots, d\}$ as $\mathbb{N}_d$ and the set $\{1, \ldots, d\}$ as $[d]$. Let $r \geq 1$ be a positive integer. The *Hamming weight* of a string $u \in \mathbb{N}_d^r$ is the number of non-zero symbols in the string and we denote it as $|u|$. For $u \in \mathbb{N}_d^r$ and $x \in \{0,1\}^r$, the sub-string of u corresponding to the non-zero bits of x is denoted by u_x. We define the *parity signature* of a string $u \in \mathbb{N}_d^r$ as a function $\mathrm{ps} : \mathbb{N}_d^r \to \{0,1\}^d$ such that the i-th bit of $\mathrm{ps}(u)$ is the parity of the number of times i occurs in the string u. As an example,

consider $d = r = 4$ and $a = (0, 1, 1, 2)$. Then, $\mathrm{ps}(a) = (0, 1, 0, 0)$. We denote the number of strings in the set $[d]^r$ with parity signature $b \in \{0,1\}^d$ as $n_{r,b}$. Note that $n_{r,b}$ only depends on $|b|$, not the actual value of b. For an integer $h \in \mathbb{N}_l$, we also use $n_{r,h}$ for the number of strings in $[d]^r$ that have the same parity signature with Hamming weight h.

We denote random variables with capital letters, like X, Y and Z. For a random variable X, the *Shannon entropy* of X is defined as

$$\mathrm{H}(X) \quad \coloneqq \quad \sum_x p(x) \log_2 \frac{1}{p(x)} \ ,$$

where p is the distribution corresponding to X. For jointly distributed random variables X and Y, the *conditional entropy* of X given Y is defined as

$$\begin{aligned} \mathrm{H}(X|\,Y) \quad &\coloneqq \quad \sum_y \Pr[Y = y]\, \mathrm{H}(X|Y = y) \\ &= \quad \sum_y \Pr[Y = y] \sum_x \Pr[X = x \,|\, Y = y] \log_2 \frac{1}{\Pr[X = x \,|\, Y = y]} \ . \end{aligned}$$

By definition, the conditional (Shannon) entropy is a non-negative quantity. Moreover, the Shannon entropy satisfies the following *chain rule* :

$$\mathrm{H}(XY) \quad = \quad \mathrm{H}(X) + \mathrm{H}(Y|\,X) \ .$$

Suppose X and Y are two random variables with the same sample space such that $\Pr[X \neq Y] \leq \epsilon$ for $\epsilon \in [0, 1]$. The *Fano inequality* [40] bounds the conditional entropy of Xgiven Y as

$$\mathrm{H}(X|\,Y) \quad \leq \quad \mathrm{H}(\epsilon) + \epsilon \log_2(|\mathcal{X}| - 1) \ , \tag{1.1}$$

where $\mathcal{X}$ is the support of X, and $|\mathcal{X}|$ denotes its size and $\mathrm{H}(\epsilon) \coloneqq \epsilon \log_2 \frac{1}{\epsilon} + (1-\epsilon) \log_2 \frac{1}{1-\epsilon}$.

Let $X_1, X_2, \ldots, X_n$ be n independent random variables over $\{0, 1\}$ and X denote their sum, i.e., $X = \sum_{i=1}^n X_i$. Let μ be the expected value of X. The *Chernoff bound* implies that the distribution corresponding to X is concentrated around μ. In particular, for $\delta \geq 0$, we have

$$\Pr[X \geq (1 + \delta)\mu] \quad \leq \quad \exp\left(-\frac{\delta^2 \mu}{2 + \delta}\right) \ . \tag{1.2}$$

The above equation is also known as the *multiplicative* Chernoff bound in the literature. In the case that X_i are all Bernoulli random variables with $\mathbb{E}X_i = p$, X is a random variable distributed according to the binomial distribution $\mathsf{B}(n, p)$ with $\mu = pn$, and we have

$$\Pr\left[X \geq \frac{3}{2}pn\right] \quad = \quad \sum_{r=\lceil 3pn/2 \rceil}^{n} \binom{n}{r} p^r (1-p)^{n-r} \quad \leq \quad \exp\left(-\frac{pn}{10}\right) \ . \tag{1.3}$$

For the special case $p = 1/2$, a tighter bound is

$$\Pr[X \leq k] \quad = \quad \Pr[X \geq n-k] \quad = \quad \sum_{r=0}^{k} \frac{\binom{n}{r}}{2^n} \quad \leq \quad 2^{-(1-\mathrm{H}(k/n))n} \qquad \forall k < \frac{n}{2} \ . \quad (1.4)$$

A simple proof of this inequality can be found in Ref. [42, Lemma 16.19, page 427].

We denote (finite dimensional) Hilbert spaces either by capital script letters like $\mathcal{H}$ and $\mathcal{K}$, or as $\mathbb{C}^m$ where m is the dimension. We denote the dimension of a Hilbert space corresponding to a register X as $|X|$. We use the Dirac notation, i.e., "ket" and "bra", for unit vectors and their adjoints, respectively. We denote the set of all unit vectors in a Hilbert space $\mathcal{H}$ by $\mathrm{Sphere}(\mathcal{H})$. For a Hilbert space $\mathcal{H} \coloneqq \mathbb{C}^S$ for some non-empty finite set S, we call $\{|x\rangle : x \in S\}$ its *canonical* basis.

A subset N of $\mathrm{Sphere}(\mathcal{H})$ is called ϵ-*dense* if for every vector $|u\rangle \in \mathrm{Sphere}(\mathcal{H})$, there exists a vector in the set N at Euclidean distance at most ϵ from $|u\rangle$. Such a set is also called an "ϵ-net" in the literature. The following proposition states that every finite dimensional Hilbert space has a relatively small ϵ-dense set.

Lemma 1.1 ([71], Lemma 13.1.1, Chapter 13)**.** *Let $\epsilon \in (0, 1]$, and m be a positive integer. The Hilbert space $\mathbb{C}^m$ has an ϵ-dense set N of size $|N| \leq \left(\frac{4}{\epsilon}\right)^{2m}$.*

A slightly better bound $\left(1 + \frac{2}{\epsilon}\right)^{2m}$ on the size of an ϵ-dense set is given in Ref. [73, Lemma 2.6].

We denote the set of all linear operators on Hilbert space $\mathcal{H}$ by $\mathsf{L}(\mathcal{H})$, the set of all positive semi-definite operators by $\mathsf{Pos}(\mathcal{H})$, and the set of all unitary operators by $\mathsf{U}(\mathcal{H})$. The identity operator on $\mathcal{H}$ is denoted by $\mathbb{1}_{\mathcal{H}}$. We denote the operator norm (*Schatten* ∞ *norm*) of an operator $M \in \mathsf{L}(\mathcal{H})$ by $\|M\|$, the Frobenius norm (*Schatten* 2 *norm*) by $\|M\|_{\mathrm{F}}$, and the trace norm (*Schatten* 1 *norm*) by $\|M\|_{\mathrm{tr}}$. Recall that $\|M\|_{\mathrm{tr}} \coloneqq \mathrm{Tr}\sqrt{M^*M}$ is the sum of the singular values of M, $\|M\|$ is the largest singular value, and $\|M\|_{\mathrm{F}} \coloneqq \sqrt{\mathrm{Tr}(M^*M)}$ is the ℓ_2-norm of the singular values with multiplicity. All of these norms are invariant under composition with a unitary operator.

Consider random unitary operators chosen according to the *Haar measure* η on $\mathsf{U}(\mathcal{H})$, where $\mathcal{H}$ is a finite dimensional Hilbert space. The Haar measure is the unique unitarily invariant probability measure over $\mathsf{U}(\mathcal{H})$.

Let $f : \mathsf{U}(\mathcal{H}) \to \mathbb{R}$ be a continuous function. Suppose f is κ-Lipschitz, i.e., for all unitary operators $U, V \in \mathsf{U}(\mathcal{H})$, we have

$$|f(U) - f(V)| \quad \leq \quad \kappa \, \|U - V\|_{\mathrm{F}} \ ,$$

for some $\kappa \geq 0$. If κ is small enough as compared to the dimension of $\mathcal{H}$, with high probability, the random variable $f(\boldsymbol{U})$ is close to its expectation, where $\boldsymbol{U} \in \mathsf{U}(\mathcal{H})$ is a Haar-random unitary operator. This *concentration of measure* property is formalized by the following theorem, which is a special case of Theorem 5.17 in Ref. [72].

Theorem 1.2 ([72], Theorem 5.17, page 159)**.** *Let η be the Haar measure on $\mathsf{U}(\mathcal{H})$, where $\mathcal{H}$ is a Hilbert space with finite dimension m, and let $\boldsymbol{U} \in \mathsf{U}(\mathcal{H})$ be a random unitary operator chosen according to η. For every function $f : \mathsf{U}(\mathcal{H}) \to \mathbb{R}$ that is κ-Lipschitz with respect to the Frobenius norm (with $\kappa > 0$), and every positive real number t, we have*

$$\eta\Big(\{U \in \mathsf{U}(\mathcal{H}) : f(U) - \mathbb{E}\left[f(\boldsymbol{U})\right] \geq t\}\Big) \quad \leq \quad \exp\left(-\frac{(m-2)t^2}{24\kappa^2}\right) .$$

1.2.2 Quantum information notation and background

For a thorough introduction to basics of quantum information, we refer the reader to the book by Watrous [96]. In this section, We briefly review the notation and some results used in this thesis.

We denote physical quantum systems ("registers") with capital letters, like X, Y and Z. The state space corresponding to a register is a finite-dimensional Hilbert space. We denote the set of all quantum states (or "density operators") over $\mathcal{H}$ by $\mathsf{D}(\mathcal{H})$. We denote quantum states or sub-normalized states (positive semi-definite operators with trace at most 1) by lowercase Greek letters like ρ, σ. We use notation such as ρ^X to indicate that register X is in state ρ, and may omit the superscript when the register is clear from the context. We say a register is *classical* if its state is diagonal in the canonical basis of the corresponding Hilbert space. A classical register corresponds to a random variable whose probability distribution is determined by the diagonal entries of the state of the register. For a non-trivial register B, we say ρ^{XB} is a *classical-quantum* state if X is classical in ρ^{XB}. We say a unitary operator $U^{AB} \in \mathsf{U}(\mathcal{H}^A \otimes \mathcal{H}^B)$ is *read-only* on register A if it is block-diagonal in the canonical basis of $\mathcal{H}^A$, i.e., $U^{AB} = \sum_a |a\rangle\langle a|^A \otimes U_a^B$ where each U_a^B is a unitary operator.

For registers B and B' with Hilbert space $\mathcal{H}$, the *swap operator* of B and B', denoted by $\mathrm{SWAP}_{B \rightleftarrows B'}$, is defined as

$$\mathrm{SWAP}_{B \rightleftarrows B'} \left(|v\rangle^B \otimes |u\rangle^{B'}\right) \quad = \quad |u\rangle^B \otimes |v\rangle^{B'}$$

for every $|u\rangle, |v\rangle \in \mathcal{H}$. For a state of the form $\rho^{AB} \otimes \tau^{B'}$, the swap operator $\mathrm{SWAP}_{B \rightleftarrows B'}$ transforms $\rho^{AB} \otimes \tau^{B'}$ to the state $\sigma^{ABB'}$ in which registers AB' are decoupled from register B, $\sigma^{AB'} = \rho^{AB}$ and $\sigma^B = \tau^{B'}$. To make the correlation between registers clear, we

sometimes change the order of registers, and denote $\mathrm{SWAP}_{B\rightleftarrows B'}\big(\rho^{AB}\otimes\tau^{B'}\big)\mathrm{SWAP}_{B\rightleftarrows B'}$ as $\rho^{AB'}\otimes\tau^{B}$.

A positive semi-definite operator $M\in\mathsf{Pos}(\mathcal{H})$ is called a *measurement operator* or *POVM operator* if $M\leq\mathbb{1}$. We usually denote quantum channels (also known as quantum operations), i.e., completely positive trace-preserving linear maps from the space of linear operators on a Hilbert space to another such space, by capital Greek letters like Ψ. The *partial trace* over a Hilbert space $\mathcal{K}$ is denoted as $\mathrm{Tr}_{\mathcal{K}}$. We say ρ^{AB} is an *extension* of σ^{A} if $\mathrm{Tr}_{B}[\rho^{AB}]=\sigma^{A}$. A *purification* of a quantum state ρ is an extension of ρ with rank one.

The *fidelity* between two sub-normalized states ρ and σ is defined as

$$\mathrm{F}(\rho,\sigma) \quad\coloneqq\quad \mathrm{Tr}\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}+\sqrt{(1-\mathrm{Tr}(\rho))(1-\mathrm{Tr}(\sigma))}\ .$$

Fidelity can be used to define a useful metric called the *purified distance* [46, 87] between sub-normalized states:

$$\mathrm{P}(\rho,\sigma) \quad\coloneqq\quad \sqrt{1-\mathrm{F}(\rho,\sigma)^2}\ .$$

For a quantum state $\rho\in\mathsf{D}(\mathcal{H})$ and $\epsilon\in[0,1]$, we define

$$\mathsf{B}^{\epsilon}(\rho) \quad\coloneqq\quad \{\widetilde{\rho}\in\mathsf{D}(\mathcal{H}):\mathrm{P}(\rho,\widetilde{\rho})\leq\epsilon\}$$

as the ball of quantum states that are within purified distance ϵ of ρ. Note that in some works, the states in the set $\mathsf{B}^{\epsilon}(\rho)$ are allowed to be sub-normalized. But here, we require the states in the ball to have trace equal to one.

Theorem 1.3 (Uhlmann [90])**.** *Let $\rho^{A},\sigma^{A}\in\mathsf{D}(\mathcal{H}^{A})$. Suppose $|\xi\rangle^{AB},|\theta\rangle^{AB}\in\mathsf{D}(\mathcal{H}^{A}\otimes\mathcal{H}^{B})$ are arbitrary purifications of ρ^{A} and σ^{A}, respectively. Then, there exists some unitary operator $V^{B}\in\mathsf{U}(\mathcal{H}^{B})$ such that*

$$\mathrm{P}\big(|\xi\rangle^{AB},\big(\mathbb{1}\otimes V^{B}\big)|\theta\rangle^{AB}\big) \quad=\quad \mathrm{P}(\rho^{A},\sigma^{A})\ .$$

The trace distance between quantum states is induced by the trace norm. The following property is well known (see, e.g., Ref. [96, Theorem 3.4, page 128]).

Theorem 1.4 (Holevo-Helstrom [52, 53])**.** *For any pair of quantum states $\rho,\sigma\in\mathsf{D}(\mathcal{H})$,*

$$\|\rho-\sigma\|_{\mathrm{tr}} \quad=\quad 2\ \max\ \{\,|\mathrm{Tr}(M\rho)-\mathrm{Tr}(M\sigma)|:M\ \textit{is a measurement operator on}\ \mathcal{H}\}\ .$$

Lemma 1.5 (Gentle Measurement [97, 77])**.** *Let $\epsilon\in[0,1]$, $\rho\in\mathsf{D}(\mathcal{H})$ and $\Pi\in\mathsf{Pos}(\mathcal{H})$ be a measurement operator, i.e., $\Pi\leq\mathbb{1}$, such that $\mathrm{Tr}[\Pi\rho]\geq1-\epsilon$. Then,*

$$\left\|\frac{\Pi\rho\Pi}{\mathrm{Tr}[\Pi\rho]}-\rho\right\|_{\mathrm{tr}} \quad\leq\quad 2\sqrt{\epsilon}\ .$$

Purified distance and trace distance are related to each other as follows (see, e.g., Ref. [96, Theorem 3.33, page 161]):

Theorem 1.6 (Fuchs and van de Graaf Inequalities [43])**.** *For any pair of quantum states* $\rho, \sigma \in \mathsf{D}(\mathcal{H})$*,*

$$1 - \sqrt{1 - \mathrm{P}(\rho, \sigma)^2} \quad \leq \quad \frac{1}{2} \left\| \rho - \sigma \right\|_{\mathrm{tr}} \quad \leq \quad \mathrm{P}(\rho, \sigma) \ .$$

Unless specified, we take the base of the logarithm function to be 2.

Let $\rho \in \mathsf{D}(\mathcal{H})$ be a quantum state of a register A over Hilbert space $\mathcal{H}$. The *von Neumann entropy* of ρ is denoted by $\mathrm{S}(A)_\rho$ or $\mathrm{S}(\rho)$, and defined as

$$\mathrm{S}(\rho) \quad := \quad -\mathrm{Tr}\left(\rho \log \rho\right) \ .$$

This coincides with the Shannon entropy of the spectrum of ρ. The *min-entropy* of ρ is defined as

$$\mathrm{S}_{\min}(\rho) \quad := \quad -\log \left\| \rho \right\| \ .$$

The *relative entropy* of two quantum states $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ is defined as

$$\mathrm{S}(\rho \| \sigma) \quad := \quad \mathrm{Tr}\left(\rho \log \rho - \rho \log \sigma\right) \ ,$$

when $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, and is ∞ otherwise. The *max-relative entropy* of ρ with respect to σ is defined as

$$\mathrm{D}_{\max}(\rho \| \sigma) \quad := \quad \min\{\lambda : \rho \leq 2^\lambda \sigma\} \ ,$$

when $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, and is ∞ otherwise. The following proposition bounds purified distance in terms of max-relative entropy. It is a special case of the monotonicity of *minimal quantum α-Rényi divergence* in α (see e.g., [86, Corollary 4.2, page 56]).

Proposition 1.7. *Let $\mathcal{H}$ be a Hilbert space, and let $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ be quantum states over $\mathcal{H}$. It holds that*

$$\mathrm{P}(\rho, \sigma) \quad \leq \quad \sqrt{1 - 2^{-\mathrm{D}_{\max}(\rho \| \sigma)}} \ .$$

The above property also implies the *Pinsker* inequality. For $\epsilon \in [0, 1]$, the *ϵ-smooth max-relative entopy* of ρ with respect to σ is defined as

$$\mathrm{D}^\epsilon_{\max}(\rho \| \sigma) \quad := \quad \min_{\rho' \in \mathsf{B}^\epsilon(\rho)} \mathrm{D}_{\max}(\rho' \| \sigma) \ .$$

For $\epsilon \in [0,1]$, the *ϵ-hypothesis testing relative entropy* of ρ with respect to σ is defined as

$$\mathrm{D}_{\mathrm{H}}^{\epsilon}(\rho\|\sigma) \quad := \quad \sup_{0\leq\Pi\leq\mathbb{1},\,\mathrm{Tr}(\Pi\rho)\geq 1-\epsilon} \log\left(\frac{1}{\mathrm{Tr}(\Pi\sigma)}\right) \ .$$

Smooth max-relative entropy and hypothesis relative entropy both converges to relative entropy in the asymptotic and i.i.d. setting. The following proposition gives upper and lower bounds tight up to second order terms for the convergence of these quantities for finite n.

Theorem 1.8 ([88],[67])**.** *Let $\epsilon \in (0,1)$, n be an integer and $\rho,\sigma \in \mathsf{D}(\mathcal{H})$ be quantum states. Define $\mathrm{V}(\rho\|\sigma) = \mathrm{Tr}(\rho(\log\rho - \log\sigma)^2) - (\mathrm{D}(\rho\|\sigma))^2$ and $\varphi(x) = \int_{-\infty}^{x} \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx$. It holds that*

$$\mathrm{D}_{\max}^{\epsilon}\big(\rho^{\otimes n}\|\sigma^{\otimes n}\big) \quad = \quad n\,\mathrm{D}(\rho\|\sigma) - \sqrt{n\,\mathrm{V}(\rho\|\sigma)}\varphi^{-1}(\epsilon) + \mathrm{O}(\log n) \ , \tag{1.5}$$

and

$$\mathrm{D}_{\mathrm{H}}^{\epsilon}\big(\rho^{\otimes n}\|\sigma^{\otimes n}\big) \quad = \quad n\,\mathrm{D}(\rho\|\sigma) + \sqrt{n\,\mathrm{V}(\rho\|\sigma)}\varphi^{-1}(\epsilon) + \mathrm{O}(\log n) \ . \tag{1.6}$$

We also need the following proposition due to Anshu, Berta, Jain and Tomamichel [2, Theorem 2].The original statement involves a minimization over all σ_B on both sides of the inequality, but the proof works for any σ_B.

Theorem 1.9 ([2], Theorem 2)**.** *Let $\epsilon,\delta \in (0,1)$ such that $0 \leq 2\epsilon + \delta \leq 1$. Consider quantum states $\sigma^B \in \mathsf{D}(\mathcal{H}^B)$ and $\rho^{AB} \in \mathsf{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$. We have*

$$\inf_{\substack{\overline{\rho}\in\mathsf{B}^{2\epsilon+\delta}(\rho^{AB}) \\ \overline{\rho}^A=\rho^A}} \mathrm{D}_{\max}\big(\overline{\rho}^{AB}\|\rho^A\otimes\sigma^B\big) \quad \leq \quad \mathrm{D}_{\max}^{\epsilon}\big(\rho^{AB}\|\rho^A\otimes\sigma^B\big) + \log\frac{8+\delta^2}{\delta^2} \ . \tag{1.7}$$

For $\epsilon \in [0,1]$ and $\rho^{AB} \in \mathsf{D}\big(\mathcal{H}^A \otimes \mathcal{H}^B\big)$, the conditional *$\epsilon$-smooth min-entropy of register A conditioned on register B* is defined as

$$\mathrm{S}_{\min}^{\epsilon}(A\,|\,B)_{\rho} \quad := \quad -\min_{\sigma^B\in\mathsf{D}(\mathcal{H}^B)} \mathrm{D}_{\max}^{\epsilon}\big(\rho^{AB}\|\mathbb{1}\otimes\sigma^B\big) \ ,$$

and the *ϵ-smooth max-entropy of register A conditioned on register B* as

$$\mathrm{S}_{\max}^{\epsilon}(A\,|\,B)_{\rho} \quad := \quad -\mathrm{S}_{\min}^{\epsilon}(A\,|\,C)_{\rho} \ ,$$

where $\rho^{ABC} \in \mathsf{D}\left(\mathcal{H}^A \otimes \mathcal{H}^B \otimes \mathcal{H}^C\right)$ is a purification of ρ^{AB}.

For a quantum state $\rho^{AB} \in \mathsf{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$, the *mutual information* of A and B is defined as

$$\mathrm{I}(A:B)_\rho \quad \coloneqq \quad \mathrm{S}\big(\rho^{AB} \| \rho^A \otimes \rho^B\big) \ .$$

When the state is clear from the context, the subscript ρ may be omitted from the notation. When ρ is a classical-quantum state, i.e., $\rho^{AB} = \sum_a p(a)|a\rangle\langle a|^A \otimes \rho_a^B$ with p being a probability distribution, $\{|a\rangle\}$ the canonical orthonormal basis for $\mathcal{H}^A$, and $\rho_a^B \in \mathsf{D}(\mathcal{H}^B)$, we have

$$\mathrm{I}(A:B) \quad = \quad \sum_a p(a) \ \mathrm{S}(\rho_a^B \| \rho^B) \ .$$

The mutual information of A and B is monotonic under the application of local quantum channels [95]. In particular, for a quantum channel $\Psi : \mathsf{L}(\mathcal{B}) \to \mathsf{L}(\mathcal{B}')$, we have

$$\mathrm{I}(A:B)_\rho \quad \geq \quad \mathrm{I}(A:B')_{(\mathbb{1}\otimes\Psi)(\rho)} \ , \tag{1.8}$$

which is also known as *Data Processing Inequality*. Suppose that registers A, B, C are in joint (tripartite) state $\rho^{ABC} \in \mathsf{D}(\mathcal{H}^A \otimes \mathcal{H}^B \otimes \mathcal{H}^C)$. The *conditional mutual information* of A and B given C is defined as

$$\mathrm{I}(A:B\,|\,C) \quad \coloneqq \quad \mathrm{I}(AC:B) - \mathrm{I}(C:B) \ .$$

When ρ^{ABC} is a tensor product of the states ρ^{AB} and ρ^C, we have

$$\mathrm{I}(A:B\,|\,C) \quad = \quad \mathrm{I}(AC:B) \quad = \quad \mathrm{I}(A:B) \ .$$

For any state $\rho^{AB} \in \mathsf{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$, the *max-information* register B has about register A [26] is defined as

$$\mathrm{I}_{\max}(A:B)_\rho \quad \coloneqq \quad \min_{\sigma \in \mathsf{D}(\mathcal{K})} \mathrm{D}_{\max}\big(\rho^{AB} \,\|\, \rho^A \otimes \sigma^B\big) \ .$$

For a parameter $\epsilon \in [0,1]$, the *smooth max-information* register B has about register A is defined as

$$\mathrm{I}^\epsilon_{\max}(A:B)_\rho \quad \coloneqq \quad \min_{\widetilde{\rho} \in \mathsf{B}^\epsilon(\rho)} \mathrm{I}_{\max}(A:B)_{\widetilde{\rho}} \ .$$

Part I

Quantum Compression

Chapter 2

Entanglement cost of compression of quantum ensembles [1]

2.1 Visible compression

Compression of quantum states is a fundamental task in information processing. In the simplest setting, we have two spatially separated parties, commonly called Alice and Bob, who can communicate with each other by exchanging quantum states. They have in mind an ensemble of m-dimensional quantum states

$$((p(x), \rho_x) : x \in S,\ \rho_x \in \mathsf{D}(\mathbb{C}^m)) \quad , \tag{2.1}$$

where S is some non-empty finite set, and p is a probability distribution over S. Alice gets an input $x \in S$ with probability $p(x)$, and would like to send a message, i.e., a quantum state $\sigma_x \in \mathsf{D}(\mathbb{C}^d)$ to Bob so that he can recover the state ρ_x, or even an approximation to it. Since the input x completely specifies the corresponding state ρ_x, this variant of the task is called *visible* compression. The *communication cost* of the protocol is $\log d$, the length of the message in qubits. Their goal is to accomplish this with as short a message as possible, i.e., to minimize the dimension d. A central question in quantum information theory is whether there is a simple characterization of the optimal communication cost in terms of the "information content" of the ensemble.

An additional resource that Alice and Bob may use in compression is a shared entangled state. In other words, the two parties may start with their qubits initialized to a fixed pure

[1] The work presented in this Chapter was previously published in Quantum, under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. [15]

quantum state independent of the input received by Alice. The local quantum operations performed for compression and decompression then also involve the respective parts of the shared state. This is depicted in Figure 2.1, and the protocol (or channel) is said to be *with shared entanglement* or *entanglement assisted*. As we may expect, the communication cost may decrease due to the availability of this additional resource. The *entanglement cost* of a protocol is the minimal dimension of the support of either party's share of the initial state (measured in qubits) required to achieve some communication cost. (We discuss the notion of entanglement cost in detail in Section 2.4.) We would also like to characterize the entanglement cost in this setting, in addition to the communication cost.

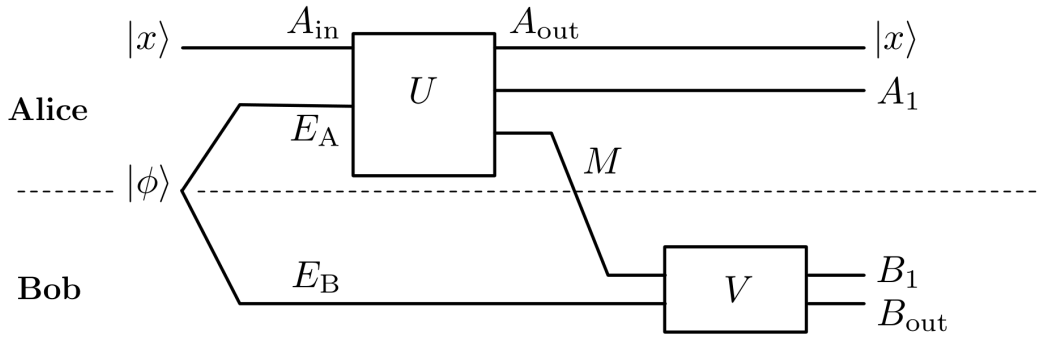

Figure 2.1: A one-message protocol for compression of quantum states, with shared entanglement. The register A_{in} holds the input given to Alice, and E_{A} contains Alice's workspace and her part of the initial shared state (the shared entanglement). The register E_{B} contains Bob's workspace and his part of the initial shared state. The compression is implemented by the isometry U, and the register M contains the compressed state and is sent as the message. The decompression is implemented by the isometry V. Bob's output is contained in the register B_{out}.

Compression problems similar to the one above have been studied extensively in quantum information theory, both in the *one-shot* setting (the one we described above), and in the *asymptotic setting* (where the sender's input consists of multiple samples picked independently from the same distribution). The problem has been studied in early works such as Ref. [21] in the setting of quantum communication without shared entanglement. It is known as *remote state preparation* when allowed one-way communication over a classical channel with shared entanglement. We refer the reader to Ref. [16, Table I] for a summary of the work on remote state preparation; we describe the most relevant results—in the one-shot setting—below.

Other tasks in the literature that come close to the one above are *state splitting* (see,

e.g., Ref. [26]), and that of channel simulation in the context of the *Quantum Reverse Shannon Theorem* [22, 26]. State splitting is the time reversal [37] of *state merging* [55, 56], and was called the "fully quantum reverse Shannon protocol" in Ref. [37]. We explain the connection to state splitting in detail in Section 2.2.2.

In both state splitting and channel simulation, the protocol is required to be "coherent" in specific ways. In particular, in compressing an ensemble of states as in Eq. (2.1), at the end of the protocol, Bob would be required to hold an approximation to the state ρ_x and Alice a purification of this state. In contrast to these tasks, we do not require that the compression protocol maintain such coherence. More precisely, the registers containing a purification of the output state may be shared by Alice and Bob. Such compression protocols are more relevant in the context of two-party communication protocols studied in complexity theory, especially in the context of *direct sum* and *direct product* results (see e.g., Refs. [60, 62, 89] and the references therein). In communication complexity, a typical goal is to compute a bi-variate Boolean function when the inputs are distributed between two parties. The parties communicate with each other, alternating messages with local computation, and at the end, one party produces the output of the protocol from the part of the final state in her possession. As a result, the output of the protocol does not depend on the part of the state held by the other party (i.e., on the purification of her part of the final joint state). A compression scheme for the final state then need only focus on the part being measured for the output.

2.1.1 Entanglement cost of compression

Jain, Radhakrishnan, and Sen [61, 62] gave a one-shot protocol for compressing an ensemble of states as in Eq. (2.1), and bounded its communication cost by $\mathrm{O}(\mathrm{I}(A:B)_\tau/\epsilon^3)$, where $\mathrm{I}(A:B)_\tau$ is the mutual information between registers A and B in the quantum state $\tau^{AB} := \frac{1}{n}\sum_{x\in[n]} |x\rangle\langle x|^A \otimes \rho_x^B$, and ϵ is the average approximation error (cf. Section 2.2.2 for a precise definition of average error). Using a more refined application of their technique, Bab Hadiashar, Nayak, and Renner [16] tightly characterized the communication cost of the task in terms of the *smooth max-information*, a one-shot entropic analogue of mutual information. Their results are stated for entanglement-assisted classical channels and use purified distance to quantify the approximation, but translate immediately to the setting here through the use of superdense coding [96, Section 6.3.1] and the Fuchs and van de Graaf Inequalities (Theorem 1.6). The upper bound so obtained is

$$\frac{1}{2}\,\mathrm{I}_{\max}^{\epsilon/\sqrt{2}}(A:B)_\tau + \mathrm{O}(\log\log(1/\epsilon)) \ .$$

This is slightly better than that derived from protocols for state splitting in terms of the approximation error; it has an additive term of $\mathrm{O}(\log\log\frac{1}{\epsilon})$ for average error ϵ versus the additive term of $\mathrm{O}(\log\frac{1}{\epsilon})$ in Ref. [4, Corollary 5]. However, both these protocols use shared entanglement that may be much longer than the message itself, namely $\mathrm{O}(k(\log\frac{1}{\epsilon})\log m)$ qubits and $\mathrm{O}((1+1/\epsilon^2)\log_2(m/\epsilon))$ qubits, respectively, where $\log_2 k = \mathrm{I}(A:B)_\tau$, and m is the dimension of the states in the ensemble. On the other hand, earlier protocols for state splitting [26, Lemma 3.5], with potentially larger communication cost, have entanglement cost bounded by $\log m$. Since sharing entanglement also entails some communication, in addition to the preparation and storage of a potentially delicate high dimensional state, this motivates the question as to whether shared entanglement is truly necessary for compression, and if so, how much of it is needed.

For the more restrictive task of state splitting, it follows from the proof of the converse bound for one-shot entanglement consumption due to Berta, Christandl, and Touchette [27, Proposition 10] that the sum of the communication and entanglement costs is at least the *min-entropy* $\mathrm{S}_{\min}(\rho)$ of the ensemble average state $\rho := \sum_x p(x)\rho_x$. (Although the proof is written assuming that the shared state consists of EPR pairs and some ancilla and an auxiliary error parameter, it may be modified to give a bound when an arbitrary state is shared and the auxiliary error is 0.) In this chapter, we show that there are ensembles for which the min-entropy bound equals the number of qubits in the states, and the bound holds up to an additive constant even with the more general compression protocols we allow.

Theorem 2.1. *Let $\epsilon \in (0,1)$, and $k \in \mathbb{N}$ with $k \geq 6/(1-\epsilon)$. For sufficiently large $m \in \mathbb{N}$ depending on k and ϵ and sufficiently large $n \in \mathbb{N}$ depending on k, m, and ϵ, there exists an ensemble $\left(\left(\frac{1}{n}, \rho_x\right) : x \in [n],\ \rho_x \in \mathsf{D}(\mathbb{C}^m)\right)$, such that*

(i) $\mathrm{I}(A:B)_\tau = \mathrm{I}_{\max}(A:B)_\tau = \log_2 k$, *where* $\tau := \frac{1}{n}\sum_{x\in[n]} |x\rangle\langle x|^A \otimes \rho_x^B$;

(ii) there is a one-way protocol with shared entanglement for the visible compression of the ensemble with average error $\epsilon/2$ and with communication cost $\frac{1}{2}\log k + \mathrm{O}(\log\log\frac{1}{\epsilon})$; and

(iii) the sum of communication and entanglement costs of any *one-way protocol with shared entanglement for visible compression of the ensemble, with average-error at most $\epsilon/2$, is at least*

$$\log m - 3\log\frac{1}{1-\epsilon} - \mathrm{O}(1) \ .$$

In particular, the theorem implies that in the absence of shared entanglement, the ensemble may only be compressed by a constant number of qubits (independent of m),

even if constant average error $\frac{\epsilon}{2} < 1/2$ is allowed. Note also that the straightforward protocol that prepares and sends the state ρ_x on input x has sum of entanglement and communication costs equal to $\log m$. So the lower bound in the theorem is optimal up to an additive universal constant term for constant $\epsilon \in (0, 1)$.

Proposition 2.5 and Corollary 2.6 in Section 2.3 contain more precise statements of the results stated in the theorem. As we explain in that section, $\mathrm{I}(A : B)_\tau$ may be interpreted as the "information content" of the ensemble; it is the *quantum information cost* [89] of the protocol in which Alice simply prepares the state ρ_x on input x and sends the state to Bob.

The compression task we study is a relaxation of oblivious (or *blind*) compression, in which the input to Alice is the state ρ_x, rather than x. It is also a relaxation of state splitting (more generally, of *state redistribution* [69, 38, 100]), and channel simulation. So the lower bound in Theorem 2.1(ii) holds for these tasks as well.

The ensemble mentioned in Theorem 2.1 is obtained via the probabilistic method, and is of a form devised by Jain, Radhakrishnan, and Sen [60]. They showed the incompressibility of such an ensemble when the decompression operation is unitary (i.e., via protocols as in Figure 2.1 in which the register B_1 is trivial). We adapt their proof method to protocols which allow a general quantum channel for decompression. A key step here is a technical lemma (Lemma 2.3 in Section 2.3) which allows us to reason about general quantum channels, and also yields a tighter lower bound on the sum of communication and entanglement costs.

2.1.2 Implications and related work

Jain *et al.* [61, 62], also used the same kind of ensemble as in Theorem 2.1 to design a two-party one-way communication protocol *with shared entanglement* for the Equality function. They showed that the initial shared state in the protocol cannot be replaced by one with polynomially smaller dimension in a "black-box fashion" (i.e., when the local operations of the two parties are not modified). Theorem 2.1 implies a similar impossibility result for protocols in which the sender and receiver can deviate from the original protocol arbitrarily, but they try to approximate the receiver's state in the original protocol after the message is sent. The impossibility holds even when the dimension of the initial shared entangled state is reduced only by a constant factor.

A remarkable property of the ensemble posited by Theorem 2.1 is that the communication cost of compression (with shared entanglement) may be arbitrarily smaller than the

entanglement cost. For constant error the communication cost is within an additive constant of the quantum information cost [89] of the protocol that simply prepares and sends the state. As a consequence, we infer that the quantum information cost of a protocol may be arbitrarily smaller than the communication cost of any protocol *without shared entanglement* for compressing its messages. Anshu, Touchette, Yao, and Yu [9] had previously proven a similar separation when the compression protocol *is* allowed to use shared entanglement. However, their separation is exponential: they exhibited an interactive protocol for a Boolean function with quantum information cost that is exponentially smaller than the communication cost of any interactive quantum protocol that computes the function. (Observe that a protocol for compressing the final state of the original protocol may also be used to compute the function.) In contrast to that protocol, the one we present *is* compressible to its quantum information cost, but requires an arbitrarily larger amount of shared entanglement to do so.

In another related work, Liu, Perry, Zhu, Koh, and Aaronson [68] show that one-way protocols cannot be compressed to their quantum information cost without using shared entanglement. They consider a certain one-way protocol in which Alice gets an n-bit input, Bob gets an m-bit input, with $m \in \mathrm{o}(n)$. The protocol has quantum information cost $\mathrm{O}(nm^{-2} \log m)$. They show that the protocol cannot be compressed by a one-way protocol without shared entanglement into a message of length $\mathrm{o}(\log n)$ with error at most $(n+1)^{-m}$. Thus the separation is limited, and only holds for exponentially small error (in the length of the inputs).

It is believed that the communication in any interactive quantum protocol which has a constant number of rounds and computes a function of classical inputs may be compressed, with constant error, to an amount proportional to the quantum information cost of the protocol. For one-way protocols such a result was shown by Jain, Radhakrishnan, and Sen [61, 62]. This was later re-proven by Anshu, Jain, Mukhopadhyay, Shayeghi, and Yao [5] using different techniques. A similar result for protocols with a larger constant number of rounds of communication was claimed by Touchette [89], but the proof has an error. The compression protocols achieving quantum information cost all rely on the presence of shared entanglement. Theorem 2.1 shows that even for the simplest protocols, such compression is not possible in the absence of shared entanglement. Moreover, it shows that the entanglement cost may be necessarily within an additive constant of the length of the message to be compressed, even when the quantum information cost is arbitrarily smaller than the message length.

In a recent independent work, Khanian and Winter [17] analyse the communication and entanglement costs of a variant of compression in the asymptotic setting. They study pure state ensembles with quantum side information in the form of pure states. In the case of vis-

ible compression with shared entanglement, they show that the asymptotic (per-instance) communication cost is at least $\frac{1}{2}\,\mathrm{S}(\rho)$, i.e., half the entropy of the ensemble average state ρ. So this cost may be at most a factor of $1/2$ smaller than that of compression *without* shared entanglement. Moreover, the asymptotic sum of communication and entanglement costs is at least the entropy $\mathrm{S}(\rho)$. Thus the kind of separation we show does not hold for pure states even in the asymptotic setting.

Organization. The rest of this chapter is organized as follows. In Section 2.2, we review basic concepts of quantum communication and compression. In section 2.3, we prove the main result and discuss its implications.

2.2 Preliminaries

2.2.1 Quantum communication protocols

We first describe a two-party quantum communication protocol informally and then give a formal definition for the special case of interest to us. We refer the reader to, e.g., Ref. [89] for a formal definition of the general case.

In a two-party quantum communication protocol, there are two parties, Alice and Bob, each of whom may get some input in registers designated for this purpose. Alice and Bob's inputs may be entangled with each other, and also with a "reference" system, which purifies it. Alice and Bob's goal is to accomplish an information processing task by communicating with each other.

Each party possesses some "work" (or "private") qubits (or registers) in addition to the input registers. The work qubits are initialized to a fixed pure state in tensor product with the input state. This fixed state may be entangled across the work registers of Alice and Bob, and may be used as a computational resource. In this case, we say the protocol or the channel is *with shared entanglement* or with *entanglement assistance*. If the fixed state is a tensor product state across Alice and Bob's registers, we say it is a protocol or channel *without shared entanglement* or simply *unassisted*.

The protocol proceeds in some number of "rounds". In each round, the sender applies an isometry to the qubits in her possession, and sends a sub-register (the message) to the other party. The length of the message (in qubits) is the base 2 logarithm of the dimension of the message register. After the last round, the recipient of the last message applies an

isometry to his registers. The output of the protocol is the state of a pair of designated registers of the two parties at the end.

We are often interested in minimizing the total length of the messages over all the rounds, i.e., the *communication cost* (or *complexity*) of the protocol. The idea is to accomplish the task at hand with minimum communication. In protocols with shared entanglement, we are also interested in the amount of shared entanglement needed in the protocol, i.e., the minimum dimension of the support of the *initial* state of either party's work space. This latter quantity, measured in number of qubits, is called the *entanglement cost* of the protocol.

In this chapter, we study only *one-way* protocols, i.e., protocols with one round, and therefore one message, (say) from Alice to Bob. We describe these more formally here. Alice and Bob initially hold registers $A_{\text{in}}E_{\text{A}}$ and $B_{\text{in}}E_{\text{B}}$, respectively. The input registers $A_{\text{in}}B_{\text{in}}$ are initialized to some state $\rho^{A_{\text{in}}B_{\text{in}}}$ whose purification is held in register R with a third party, called Referee. Alice and Bob's work registers E_{A} and E_{B} are initialized to a pure state $|\phi\rangle^{E_{\text{A}}E_{\text{B}}}$, which may be entangled across the partition $E_{\text{A}}E_{\text{B}}$. The local operations in the protocol are specified by two isometries U and V. The isometry U acts on registers $A_{\text{in}}E_{\text{A}}$ and maps them to registers $A_{\text{out}}A_1M$. The isometry V acts on registers $B_{\text{in}}E_{\text{B}}M$ and maps them to registers B_1B_{out}. First, Alice applies U to the registers A_{in} and E_{A} and sends the register M to Bob. Then, Bob applies V on his initial registers $B_{\text{in}}E_{\text{B}}$ and the message M. The output of the protocol is the state of Alice and Bob's registers $A_{\text{out}}B_{\text{out}}$. The communication cost of this protocol is $\log |M|$ and the entanglement cost is the logarithm of the Schmidt rank of the state $|\phi\rangle$ across the partition $E_{\text{A}}E_{\text{B}}$. We say it is a protocol *with shared entanglement* if the Schmidt rank of $|\phi\rangle$ is more than 1, and say that it is *without shared entanglement* otherwise. Such protocols are also called *entanglement-assisted* and *unassisted*, respectively, in the literature.

We say that the input is "classical" when there are non-empty finite sets $S_{\text{A}}, S_{\text{B}}$ (the sets of classical inputs) such that the Hilbert spaces corresponding to the input registers are $\mathbb{C}^{S_{\text{A}}}, \mathbb{C}^{S_{\text{B}}}$, respectively, and the initial joint quantum state in the input registers $A_{\text{in}}B_{\text{in}}$ is diagonal in the canonical basis $\{|x\rangle|y\rangle : x \in S_{\text{A}}, y \in S_{\text{B}}\}$. In the case that the inputs to Alice and Bob are classical, we assume without loss of generality that the input registers A_{in} and B_{in} are "read-only", i.e., the isometries U and V are of the form $\sum_{x\in S_{\text{A}}} |x\rangle\langle x|^{A_{\text{in}}} \otimes U_x^{E_{\text{A}}}$ and $\sum_{y\in S_{\text{B}}} |y\rangle\langle y|^{B_{\text{in}}} \otimes V_y^{ME_{\text{B}}}$, where $S_{\text{A}}, S_{\text{B}}$ are sets as above. A one-way protocol in which Alice gets a classical input and Bob does not have any input is depicted in Figure 2.1.

Let Π be a one-way quantum protocol (with or without shared entanglement) with a single message from Alice to Bob, in which Alice gets a classical input and Bob does not have any input. The register R with Referee purifies Alice's input so that the joint state

is $|\rho\rangle^{RA_{\text{in}}} \coloneqq \sum_{x \in S_{\text{A}}} \sqrt{p(x)} |xx\rangle^{RA_{\text{in}}}$, where $p(x)$ is a probability distribution over the input set S_{A}; see Fig. 2.2 for an illustration. Let M be the quantum register corresponding to the message in Π. The *quantum information cost* (or *quantum information complexity*) of the protocol Π is defined as

$$\text{QIC}(\Pi) \quad \coloneqq \quad \frac{1}{2} \, \text{I}(R : M \,|\, E_{\text{B}}) \ ,$$

where the registers are in the state immediately after Alice sends the message register M to Bob. This expression simplifies to $\text{I}(R : ME_{\text{B}})$ as the registers R, E_{B} are in a tensor product state at this point. It is intended to measure the information Bob gains about Alice's input from the message. This notion requires a nuanced definition for protocols with more general inputs and with multiple rounds of communication. As it is not central to our work, we refer the reader to Ref. [89] for the definition for general protocols.

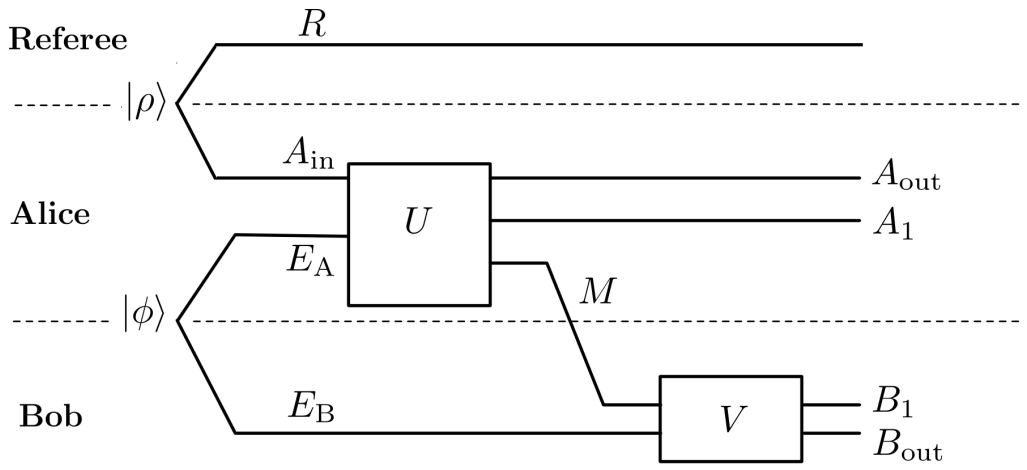

Figure 2.2: A one-way quantum communication protocol in which Bob has no input. The register A_{in} holds the input given to Alice, register R purifies Alice's input, and E_{A} contains Alice's workspace and her part of the possible shared entanglement. The register E_{B} contains Bob's workspace and his part of the possible shared entanglement. For a protocol without shared entanglement, the state $|\phi\rangle^{A_{in}B_{in}}$ is a product state. The isometry U is the operation performed by Alice, the register M corresponds to the message. The isometry V is the operation performed by Bob. Bob's output is contained in the register B_{out}.

2.2.2 Compression of quantum states

We study one-way protocols for *non-oblivious* or *visible* compression of quantum states, which is typical for tasks of this nature (see, e.g., Ref. [4]). The protocol may be with or without shared entanglement. Suppose we wish to compress states chosen from an ensemble $((p(x), \rho_x) : x \in S)$ for some finite set S, where p is a probability distribution over S and $\rho_x \in \mathsf{D}(\mathcal{H})$. The ensemble is known to both parties. The sender, say Alice, is given a classical input $x \in S$ chosen according to the distribution p. Alice and Bob execute a one-way protocol with a message from Alice to Bob in order to prepare an approximation of ρ_x on Bob's side. Following the notation from Section 2.2.1, we interpret the state of the message register M of this protocol as a compression of ρ_x. Suppose the state of the output register B_{out} is $\widetilde{\rho}_x$. We say that the average error of the compression protocol is $\epsilon \in [0, 2]$ if the output state $\widetilde{\rho}_x$ is ϵ-close in trace distance to the ideal state ρ_x on average over the inputs x:

$$\sum_x p(x) \left\| \rho_x - \widetilde{\rho}_x \right\|_{\text{tr}} \quad \leq \quad \epsilon \ .$$

It is sometimes desirable to express the error in terms of the purified distance. For simplicity, we state error bounds in terms of trace distance; we may express the bounds in terms of purified distance via Theorem 1.6.

Note that a protocol for visible compression without shared entanglement may be characterized by a sequence of quantum states $(\sigma_x : x \in S)$ and a quantum channel Ψ. We let σ_x be the state of the message register M sent by Alice to Bob on input x. We define Ψ as the channel resulting from the application of the isometry V followed by the tracing out of the register B_1. The average error of the protocol is then $\sum_x p(x) \left\| \rho_x - \Psi(\sigma_x) \right\|_{\text{tr}}$. Conversely, any choice of states $(\sigma_x : x \in S, \ \sigma_x \in \mathsf{D}(\mathcal{K}))$ and quantum channel $\Psi : \mathsf{L}(\mathcal{K}) \to \mathsf{L}(\mathcal{H})$ for some Hilbert space $\mathcal{K}$ defines a valid visible compression protocol.

An essentially equivalent formulation of the task of visible compression is the following (with the notation from Section 2.2.1). Consider the state τ over the registers RXA_1C:

$$\tau \quad := \quad \sum_{x \in S} \sqrt{p(x)} |x\rangle^R |x\rangle^X |\phi_x\rangle^{A_1 C} \ ,$$

where $|\phi_x\rangle^{A_1 C}$ is a purification of ρ_x, register R is held by Referee, and registers XA_1C together constitute Alice's input register A_{in}. Alice and Bob both know the full description of τ. Their goal is to run a one-way quantum communication protocol with a message from Alice to Bob, with or without shared entanglement, such that at the end, the state $\widetilde{\tau}$ of registers RB_{out} is close to τ^{RC}:

$$\left\| \widetilde{\tau}^{RB_{\text{out}}} - \tau^{RC} \right\|_{\text{tr}} \quad \leq \quad \epsilon \ .$$

The difference from state splitting is that for a fixed state $|x\rangle$ of register R, the purification of the state in register B_{out} may be shared arbitrarily between Alice and Bob (while in state splitting, it is required to be held by Alice, in register A_1). A protocol for state splitting can thus be used for this task, and conversely lower bounds on communication or entanglement costs derived for the above task applies to state splitting as well.

2.3 The main result

In this section, we prove Theorem 2.1, the main result of this chapter.

2.3.1 Two useful lemmas

We begin with two lemmas that we need for the result. The first allows us to focus on a finite number of subspaces of a finite dimensional Hilbert space, in the context of measurements. For an operator $M \in \mathsf{L}(\mathcal{H})$, and a subspace $\mathcal{A}$ of $\mathcal{H}$, define the semi-norm

$$\|M\|_{\mathcal{A}} \quad := \quad \max_{|w\rangle \in \text{Sphere}(\mathcal{A})} |\langle w|M|w\rangle| \quad .$$

Lemma 2.2 ([60], Lemma 6)**.** *Let d and q be positive integers with $q \geq d$, $\delta > 0$ be a real number, and $\mathcal{H}$ be an q-dimensional Hilbert space. There exists a set $\mathfrak{T}$ of subspaces of $\mathcal{H}$ of dimension at most d such that*

1. $|\mathfrak{T}| \leq \left(\frac{8\sqrt{d}}{\delta}\right)^{2qd}$, *and*
2. *for every d-dimensional subspace $\mathcal{A} \subseteq \mathcal{H}$, there is a subspace $\mathcal{B} \in \mathfrak{T}$ such that for every measurement operator $M \in \mathsf{Pos}(\mathcal{H})$,*

$$\left| \|M\|_{\mathcal{A}} - \|M\|_{\mathcal{B}} \right| \quad \leq \quad \delta \quad .$$

The set $\mathfrak{T}$ in the lemma is obtained as follows. We fix an ϵ-dense subset S of $\text{Sphere}(\mathcal{H})$ for a suitably small value of ϵ, as given by Lemma 1.1. For any d-dimensional subspace $\mathcal{A}$, we consider an orthonormal basis, and the d vectors in S closest to the respective elements in the basis. We include in $\mathfrak{T}$ the subspace $\mathcal{B}$ spanned by the d vectors from S so obtained.

By a uniformly random subspace of dimension ℓ of an m-dimensional Hilbert space $\mathcal{H}$, with $\ell \leq m$, we mean the image of a fixed ℓ-dimensional subspace under a Haar-random unitary operator on $\mathcal{H}$.

The next lemma is similar to Lemma 7 from Ref. [60], and is stronger in several respects. It enables the generalization of the incompressibility result in Ref. [60] that we prove, and helps us derive tighter bounds for compression. Informally, the lemma states that every state in a "small enough" subspace of a bipartite space has, with high probability, a small projection onto a "small enough" random subspace of one part.

Lemma 2.3. *Let m, d, ℓ, and p be positive integers such that $\ell \leq m$. Let $\mathcal{W}$ be a fixed d-dimensional subspace of $\mathbb{C}^m \otimes \mathbb{C}^p$. Let $\boldsymbol{\mathcal{Z}}$ be a uniformly random subspace of $\mathbb{C}^m$ of dimension ℓ, and $\boldsymbol{M}$ be the orthogonal projection operator onto $\boldsymbol{\mathcal{Z}}$. Then for any real number $\alpha > 2$, there is a real number $\alpha_1 > 0$ that depends only on α such that*

$$\Pr\left[\|\boldsymbol{M} \otimes \mathbb{1}_{\mathbb{C}^p}\|_{\mathcal{W}} \geq \frac{\alpha\ell}{m}\right] \quad \leq \quad \exp\left(-\frac{\alpha_1 \ell^2 (m-2)}{m^2}\right) \ ,$$

provided

$$(\alpha-2)^2 \ell^2 (m-2) \quad \geq \quad (4 \times 384) d m^2 \ln\left(\frac{8m}{\alpha\ell}\right) \ .$$

We may take $\alpha_1 := \frac{(\alpha-2)^2}{768}$ in the above statement.

Proof: The subspace $\mathcal{W}$ is isomorphic to $\mathbb{C}^d$ as it is d-dimensional. By Lemma 1.1, there is a set N with $|N| \leq \left(\frac{8m}{\alpha\ell}\right)^{2d}$ that is a $\frac{\alpha\ell}{2m}$-dense set of Sphere($\mathcal{W}$).

Note that for any two vectors $|u\rangle, |v\rangle \in \text{Sphere}(\mathbb{C}^m \otimes \mathbb{C}^p)$, we have

$$\begin{aligned}
|\langle u|(\boldsymbol{M} \otimes \mathbb{1})|u\rangle - \langle v|(\boldsymbol{M} \otimes \mathbb{1})|v\rangle| &= |\mathrm{Tr}\left((\boldsymbol{M} \otimes \mathbb{1})|u\rangle\langle u| - (\boldsymbol{M} \otimes \mathbb{1})|v\rangle\langle v|\right)| \\
&\leq \frac{1}{2}\left\||u\rangle\langle u| - |v\rangle\langle v|\right\|_{\mathrm{tr}} \qquad \text{(by Theorem 1.4)} \\
&\leq \frac{1}{2}\left\|(|u\rangle - |v\rangle)\langle v|\right\|_{\mathrm{tr}} + \frac{1}{2}\left\||u\rangle(\langle u| - \langle v|)\right\|_{\mathrm{tr}} \\
&= \||u\rangle\| \, \||u\rangle - |v\rangle\| \quad = \quad \||u\rangle - |v\rangle\| \ .
\end{aligned}$$

This implies that if $\|M \otimes \mathbb{1}_{\mathbb{C}^p}\|_{\mathcal{W}} \geq \frac{\alpha\ell}{m}$, there is a vector $|v\rangle \in N$ such that

$$\langle v|(\boldsymbol{M} \otimes \mathbb{1})|v\rangle \geq \tfrac{\alpha\ell}{2m} \ .$$

By the Union Bound, we get

$$\Pr\left[\|\boldsymbol{M} \otimes \mathbb{1}_{\mathbb{C}^p}\|_{\mathcal{W}} \geq \frac{\alpha\ell}{m}\right] \ \leq \ |N| \times \max_{|v\rangle \in N} \Pr\left[\langle v|(\boldsymbol{M} \otimes \mathbb{1})|v\rangle \geq \frac{\alpha\ell}{2m}\right] \ . \qquad (2.2)$$

Consider any fixed vector $|v\rangle \in N$ and let $P \in \mathsf{Pos}(\mathbb{C}^m)$ be a fixed orthogonal projection of rank ℓ. Consider the function $f : \mathsf{U}(\mathbb{C}^m) \to \mathbb{R}$ defined as

$$f(U) \quad \coloneqq \quad \langle v| \left(UPU^* \otimes \mathbb{1}_{\mathbb{C}^p}\right) |v\rangle \ .$$

For any $U, W \in \mathsf{U}(\mathbb{C}^m)$, we have

$$\begin{aligned} |f(U) - f(W)| &= \left| \mathrm{Tr}\left[\left((UPU^* - WPW^*) \otimes \mathbb{1} \right) |v\rangle\langle v| \ \right] \right| \\ &\leq \|UPU^* - WPW^*\| \\ &\leq \|UPU^* - WPU^*\| + \|WPU^* - WPW^*\| \\ &\leq \|U - W\| + \|U^* - W^*\| \\ &\leq 2\,\|U - W\|_{\mathrm{F}} \ , \end{aligned} \tag{2.3}$$

where the second last inequality holds since operator norm is sub-multiplicative, i.e., for linear operators A and B, we have $\|AB\| \leq \|A\|\,\|B\|$, and the last inequality holds since $\|A\| = \|A^*\|$ for every linear operator A. Eq. (2.3) implies that f is 2-Lipschitz.

Let $\boldsymbol{U} \in \mathsf{U}(\mathbb{C}^m)$ be a Haar-random unitary operation. The expectation of $f(\boldsymbol{U})$ is:

$$\begin{aligned} \mathbb{E}[f(\boldsymbol{U})] &= \langle v| \Big(\mathbb{E}[\boldsymbol{U}P\boldsymbol{U}^*] \otimes \mathbb{1} \Big) |v\rangle \\ &= \langle v| \left(\ell \frac{\mathbb{1}}{m} \otimes \mathbb{1} \right) |v\rangle \\ &= \frac{\ell}{m} \ . \end{aligned}$$

Since $\boldsymbol{U}P\boldsymbol{U}^*$ and $\boldsymbol{M}$ have the same distribution, by Theorem 1.2 we get

$$\begin{aligned} \Pr\left[\langle v| \left(\boldsymbol{M} \otimes \mathbb{1}\right) |v\rangle \geq \frac{\alpha\ell}{2m} \right] &= \Pr\left[\langle v| \left(\boldsymbol{U}P\boldsymbol{U}^* \otimes \mathbb{1}\right) |v\rangle \geq \frac{\alpha\ell}{2m} \right] \\ &\leq \exp\left(-\frac{(m-2)(\alpha-2)^2\ell^2}{384m^2} \right) \ . \end{aligned}$$

By Eq. (2.2), we get

$$\begin{aligned} \Pr\left[\|\boldsymbol{M} \otimes \mathbb{1}_{\mathbb{C}^p}\|_{\mathcal{W}} \geq \frac{\alpha\ell}{m} \right] &\leq \left(\frac{8m}{\alpha\ell} \right)^{2d} \exp\left(-\frac{(m-2)(\alpha-2)^2\ell^2}{384m^2} \right) \\ &\leq \exp\left(-\frac{(m-2)(\alpha-2)^2\ell^2}{768m^2} \right) \ , \end{aligned}$$

provided that m, ℓ, d, and α satisfy the stated condition. ∎

2.3.2 The ensemble and its compressibility

We study an ensemble of the same form as in Ref. [60]. For positive integers n, m, k such that k divides m and n, let $B_i = (|b_{i1}\rangle, |b_{i2}\rangle, \ldots, |b_{im}\rangle)$ be a suitably chosen orthonormal basis for $\mathbb{C}^m$, for each $i \in \left[\frac{n}{k}\right]$. Let $(B_{ij} : j \in [k])$ be a partition of B_i into k equal size sets. Define $\rho_{ij} \coloneqq \frac{k}{m} \sum_{|v\rangle \in B_{ij}} |v\rangle\langle v|$. We show that there is a choice of bases such that the ensemble

$$\left(\left(\tfrac{1}{n}, \rho_{ij}\right) : i \in \left[\frac{n}{k}\right], j \in [k] \right) \tag{2.4}$$

cannot be compressed significantly in the absence of shared entanglement. The following theorem, which we prove along the same lines as Theorem 5 in Ref. [60], contains the crux of the argument.

Theorem 2.4. *Let $\beta \in (0,1)$, and $\epsilon \in (0,1)$. Let k, m, n, d be positive integers such that k divides m and n. There exists an ensemble of n quantum states (ρ_{ij}) of the form in Eq. (2.4) such that for any sequence of quantum states $\left(\sigma_{ij} : \sigma_{ij} \in \mathsf{D}(\mathbb{C}^d), i \in \left[\frac{n}{k}\right], j \in [k]\right)$, and for all quantum channels $\Psi : \mathsf{L}(\mathbb{C}^d) \to \mathsf{L}(\mathbb{C}^m)$, we have*

$$\left| \left\{ (i,j) : \left\| \rho_{ij} - \Psi(\sigma_{ij}) \right\|_{\mathrm{tr}} > \epsilon \right\} \right| \quad > \quad \beta n \ , \tag{2.5}$$

when

$$\begin{aligned} k &\geq \frac{4}{1-\epsilon} \ , \\ m &> \max\left\{ \frac{3}{\gamma} \ln\left(\frac{\mathrm{e}}{1-\beta}\right), \ \frac{3}{\gamma} \ln k, \ \ 2 + \frac{d}{\gamma} \ln\left(\frac{16}{1-\epsilon}\right) \right\} \ , \qquad \textit{and} \\ n &> \frac{6kd^2m}{\gamma(1-\beta)} \ln\left(\frac{16\sqrt{d}}{\epsilon}\right) \ , \end{aligned}$$

where $\gamma \coloneqq \frac{(1-\epsilon)^2}{8\times 768}$.

Proof: We use the probabilistic method to show the existence of an ensemble with the claimed property. In particular, we show that for an ensemble of the form in Eq. (2.1) chosen at random, Eq. (2.5) holds with non-zero probability. We first derive a simpler property that suffices.

For $i \in \left[\frac{n}{k}\right]$ and $j \in [k]$, let $\tau_{ij} \in \mathsf{D}(\mathbb{C}^m)$ be m-dimensional quantum states and M_{ij} be the orthogonal projection onto the support of τ_{ij}. By Theorem 1.4, the condition

$$\left| \mathrm{Tr}\left(M_{ij}\tau_{ij}\right) - \mathrm{Tr}\left(M_{ij}\Psi(\sigma_{ij})\right) \right| \ > \ \frac{\epsilon}{2} \tag{2.6}$$

implies that $\|\tau_{ij} - \Psi(\sigma_{ij})\|_{\mathrm{tr}} > \epsilon$. Since $\mathrm{Tr}(M_{ij}\tau_{ij}) = 1$, Eq. (2.6) is equivalent to

$$\mathrm{Tr}\left(M_{ij}\Psi(\sigma_{ij})\right) \;<\; 1 - \frac{\epsilon}{2} \;. \tag{2.7}$$

Consider the following Stinespring representation [96, Corollary 2.27, Sec. 2.2] of the quantum channel $\Psi : \mathsf{L}(\mathbb{C}^d) \to \mathsf{L}(\mathbb{C}^m)$ in terms of a unitary operation $U \in \mathsf{U}(\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{C})$ and a fixed pure state $|\bar{0}\rangle \in \mathcal{B}\otimes\mathcal{C}$, with $\mathcal{A} = \mathbb{C}^d, \mathcal{B} = \mathcal{C} = \mathbb{C}^m$:

$$\Psi(\omega) \;=\; \mathrm{Tr}_{\mathcal{A}\otimes\mathcal{B}}\Big[U(\omega\otimes|\bar{0}\rangle\langle\bar{0}|)U^*\Big] \qquad\qquad \forall\omega\in\mathsf{L}(\mathbb{C}^d) \;.$$

So we have

$$\begin{aligned}\mathrm{Tr}\left(M_{ij}\Psi(\sigma_{ij})\right) &= \mathrm{Tr}\Big(M_{ij}\;\mathrm{Tr}_{\mathcal{A}\otimes\mathcal{B}}\big[U(\sigma_{ij}\otimes|\bar{0}\rangle\langle\bar{0}|)U^*\big]\Big)\\ &= \mathrm{Tr}\Big(\left(M_{ij}\otimes\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}}\right)U(\sigma_{ij}\otimes|\bar{0}\rangle\langle\bar{0}|)U^*\Big) \;,\end{aligned}$$

and Eq. (2.7) is equivalent to

$$\mathrm{Tr}\Big(\left(M_{ij}\otimes\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}}\right)U(\sigma_{ij}\otimes|\bar{0}\rangle\langle\bar{0}|)U^*\Big) \;<\; 1 - \frac{\epsilon}{2} \;. \tag{2.8}$$

For a fixed unitary operator U, for any i, j, the state $U(\sigma_{ij}\otimes|\bar{0}\rangle\langle\bar{0}|)U^*$ belongs to $\mathsf{D}(\mathcal{X})$ where $\mathcal{X} := U(\mathcal{A}\otimes|\bar{0}\rangle)$ is a fixed d-dimensional subspace of $\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{C}$. Thus, the expression on the left in Eq. (2.8) is bounded by $\|M_{ij}\otimes\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}}\|_{\mathcal{X}}$ for every i, j. So it suffices to exhibit an ensemble such that for all d-dimensional subspaces $\mathcal{W}\subseteq\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{C}$,

$$\left|\left\{(i,j) : \|M_{ij}\otimes\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}}\|_{\mathcal{W}} < 1 - \frac{\epsilon}{2}\right\}\right| \;>\; \beta n \;.$$

By Lemma 2.2, for any $\delta = \epsilon/2$, there is a collection $\mathfrak{T}$ of subspaces of $\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{C}$ of dimension at most d, such that size $|\mathfrak{T}| \leq (16\sqrt{d}/\epsilon)^{2d^2m^2}$, and for all subspaces $\mathcal{W}$ as above, there is a subspace $\mathcal{Y}\in\mathfrak{T}$ such that for all i, j,

$$\left|\,\|M_{ij}\otimes\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}}\|_{\mathcal{W}} - \|M_{ij}\otimes\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}}\|_{\mathcal{Y}}\,\right| \;\leq\; \frac{\epsilon}{2} \;.$$

So, it suffices to produce an ensemble such that for all subspaces $\mathcal{Y}\in\mathfrak{T}$,

$$\left|\left\{(i,j) : \|M_{ij}\otimes\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}}\|_{\mathcal{Y}} < 1 - \epsilon\right\}\right| \;>\; \beta n \;. \tag{2.9}$$

We pick bases $\boldsymbol{B}_i$ independently and uniformly at random, i.e., for each i, independently pick a Haar-random unitary operator on $\mathbb{C}^m$, and let $\boldsymbol{B}_i$ be the basis defined by its columns. Partition $\boldsymbol{B}_i$ into k sets $(\boldsymbol{B}_{ij} : j \in [k])$ of equal size. We then define an ensemble of the form in Eq. (2.4) with $\boldsymbol{\rho}_{ij} \coloneqq \frac{k}{m} \sum_{|v\rangle \in \boldsymbol{B}_{ij}} |v\rangle\langle v|$, and the corresponding projection operators $\boldsymbol{M}_{ij} \coloneqq \sum_{|v\rangle \in \boldsymbol{B}_{ij}} |v\rangle\langle v|$. We show that with non-zero probability, the operators $\boldsymbol{M}_{ij}$ satisfy Eq. (2.9) for all $\mathcal{Y} \in \mathfrak{T}$, by bounding the probability of the complementary event.

Suppose the operators $\boldsymbol{M}_{ij}$ do not satisfy Eq. (2.9) for some subspace $\mathcal{Y} \in \mathfrak{T}$. Then

$$\left| \left\{ (i,j) : \|\boldsymbol{M}_{ij} \otimes \mathbb{1}_{\mathcal{A}\otimes\mathcal{B}}\|_{\mathcal{Y}} < 1 - \epsilon \right\} \right| \quad \leq \quad \beta n \ . \tag{2.10}$$

Equivalently, there are at least $(1-\beta)n$ pairs i, j such that $\|\boldsymbol{M}_{ij} \otimes \mathbb{1}\|_{\mathcal{Y}} \geq 1 - \epsilon$. In particular, there are at least $(1-\beta)n/k$ indices i such that there is at least one $j \in [k]$ with $\|\boldsymbol{M}_{ij} \otimes \mathbb{1}\|_{\mathcal{Y}} \geq 1 - \epsilon$. For convenience, by $\boldsymbol{E}_i(\mathcal{Y})$ we denote the event that there is some $j \in [k]$ with $\|\boldsymbol{M}_{ij} \otimes \mathbb{1}\|_{\mathcal{Y}} \geq 1-\epsilon$, and by $\boldsymbol{I}(\mathcal{Y})$, we denote the subset of indices $i \in \left[\frac{n}{k}\right]$ such that $\boldsymbol{E}_i(\mathcal{Y})$ occurs.

Let $q \coloneqq \lceil (1-\beta)\frac{n}{k} \rceil$. By the above reasoning, it suffices to bound the probability that for some subspace $\mathcal{Y} \in \mathfrak{T}$, the subset $\boldsymbol{I}(\mathcal{Y})$ has at least q indices.

By Lemma 2.3, for a fixed subspace $\mathcal{Y}$ and pair i, j,

$$\Pr\left[\|\boldsymbol{M}_{ij} \otimes \mathbb{1}\|_{\mathcal{Y}} \geq 1 - \epsilon \right] \quad \leq \quad \exp\left(-\frac{((1-\epsilon)k - 2)^2 (m-2)}{768 k^2} \right) \quad \leq \quad \exp(-\gamma m) \ ,$$

with $\gamma \coloneqq \frac{(1-\epsilon)^2}{8 \times 768}$, when $(1-\epsilon)k \geq 4$ and

$$m - 2 \quad \geq \quad \frac{(16 \times 384) d}{(1-\epsilon)^2} \ \ln\left(\frac{8}{1-\epsilon} \right) \ .$$

So by the Union Bound

$$\Pr\left[\boldsymbol{E}_i(\mathcal{Y}) \right] \quad \leq \quad k \, \exp(-\gamma m) \ ,$$

and by the Union Bound and the independence of $\boldsymbol{M}_{ij}$ for distinct indices i,

$$\Pr\left[|\boldsymbol{I}(\mathcal{Y})| \geq q \right] \quad \leq \quad \binom{\frac{n}{k}}{q} \times (k \, \exp(-\gamma m))^q \ .$$

Finally, we get

$$\begin{aligned}\Pr\left[\exists \mathcal{Y} \in \mathfrak{T} : \text{Eq. (2.10) holds}\right] &\leq |\mathfrak{T}| \times \max_{\mathcal{Y}\in\mathfrak{T}} \Pr\left[\,|\boldsymbol{I}(\mathcal{Y})| \geq q\right] \\ &\leq \left(\frac{16\sqrt{d}}{\epsilon}\right)^{2d^2m^2} \binom{\frac{n}{k}}{q} (k\,\exp(-\gamma m))^q \\ &< 1\ ,\end{aligned}$$

when $m > \max\left\{\frac{3}{\gamma}\ln\left(\frac{\mathrm{e}}{1-\beta}\right), \frac{3}{\gamma}\ln k\right\}$, and

$$\gamma(1-\beta)n \quad > \quad 6kd^2m\,\ln\left(\frac{16\sqrt{d}}{\epsilon}\right)\ .$$

This proves the theorem. ∎

Note that the above proof considers an arbitrary choice of states σ_{ij} and quantum channel Ψ *after* the ensemble is chosen randomly. Together, the sequence (σ_{ij}) and the channel Ψ constitute a compression protocol. The proof shows that no matter how (σ_{ij}) and Ψ are chosen, the error due to the corresponding compression protocol is large if the dimension d is much smaller than m (provided n is chosen properly).

2.3.3 Application to entanglement cost

Consider a one-way protocol Π in which with probability $1/n$, Alice gets input (i,j), prepares state ρ_{ij} as in an ensemble given by Theorem 2.4, and sends it to Bob. The ensemble average ρ is the completely mixed state $\frac{\mathbb{1}}{m}$ over $\mathbb{C}^m$. By construction, $\mathrm{S}(\rho_{ij}\|\rho)$ equals $\log k$, and therefore $\mathrm{QIC}(\Pi) = \frac{1}{2}\log k$. In fact, we have $\mathrm{D}_{\max}(\rho_{ij}\|\rho) = \log k$. Theorem I.1(1) of Ref. [16] gives us a protocol for the visible compression of any such ensemble of states using classical communication and shared entanglement, with error ϵ. The communication cost of this protocol is

$$\mathrm{I}_{\max}^{\epsilon/\sqrt{2}}(A:B)_\tau + \mathrm{O}(\log\log(1/\epsilon))\ ,$$

where $\tau^{AB} := \frac{1}{n}\sum_{ij}|ij\rangle\langle ij|^A \otimes \rho_{ij}^B$ and we have used Theorem 1.6 to translate between purified and trace distance. This expression is bounded from above by $\log k + \mathrm{O}(\log\log\frac{1}{\epsilon})$, since $\mathrm{D}_{\max}(\rho_{ij}\|\rho)$ (and therefore $\mathrm{I}_{\max}(A:B)_\tau$) equals $\log k$. Using superdense coding [96, Section 6.3.1], we get a bound on the quantum communication cost of compressing the ensemble with entanglement assistance.

Proposition 2.5. *For any positive integers k, m, n such that k divides m and n, and error parameter $\epsilon > 0$, any ensemble of n equally likely quantum states in $\mathsf{D}(\mathbb{C}^m)$ of the form in Eq. (2.4) there is a one-shot one-way protocol* ***with shared entanglement*** *for compressing the states with quantum communication at most*

$$\tfrac{1}{2}\log k + \mathrm{O}(\log\log\tfrac{1}{\epsilon}) \ ,$$

with average error at most ϵ in trace distance.

This bound is an additive term of $\mathrm{O}(\log\log\frac{1}{\epsilon})$ more than QIC(Π). Theorem I.1(1) in Ref. [16] also gives a lower bound of $(1/2)\,\mathrm{I}_{\max}^{\sqrt{\epsilon}}(A:B)_\tau$ on the communication cost, which is at least $(1/2)\log k - 2$ for $\epsilon \le 1/81$ (see Proposition A.1 in the appendix). So for constant ϵ, the upper bound in Proposition 2.5 is close to optimal as a function of k. It is slightly better than those obtained from protocols for state splitting (see, e.g., Ref. [4, Corollary 5]), which have an additive term of order $\log\frac{1}{\epsilon}$. However, the protocol from Ref. [16] has entanglement cost of order $k(\log\frac{1}{\epsilon})\log m$, which is exponential in the communication cost, while the protocol for state splitting with the least known communication cost [4, Corollary 5] has entanglement cost of order $(1 + 1/\epsilon^2)\log(m/\epsilon)$.

Next we consider how small the entanglement cost of the visible compression of an ensemble (ρ_{ij}) given by Theorem 2.4 may be. By choosing the parameters in the statement of Theorem 2.4 appropriately, we get the following lower bound on the sum of communication and entanglement costs of any compression protocol.

Corollary 2.6. *There exist universal constants $c_1, c_2, c_3 > 0$ such that for any $\epsilon \in (0,1)$ and any positive integers k, m, n with $k \ge 6/(1-\epsilon)$, m and n divisible by k, $m \ge c_1(\ln k)/(1-\epsilon)^2$, and*

$$n \quad \ge \quad \frac{c_3}{(1-\epsilon)^2}\, k m^3 \ln\frac{16\sqrt{m}}{\epsilon} \ ,$$

there is an ensemble of n equally likely quantum states in $\mathsf{D}(\mathbb{C}^m)$ of the form in Eq. (2.4) for which any (one-shot) one-way protocol for compressing the states with average error at most $\frac{\epsilon}{2}$, the sum of the communication and entanglement costs is at least

$$\log m - 2\log\frac{1}{1-\epsilon} - \log\ln\frac{16}{1-\epsilon} - c_2 \ . \tag{2.11}$$

In particular, the entanglement cost of any such protocol with ***optimal*** *communication cost is at least*

$$\log m - \frac{1}{2}\log k - \mathrm{O}\Big(\log\frac{1}{1-\epsilon}\Big) - \mathrm{O}(1) \ ,$$

and the communication cost of any such protocol ***without entanglement*** *is at least the bound given in Eq. (2.11).*

Proof: We invoke Theorem 2.4 with $\epsilon \in (0,1)$, $\beta = 1/2$ and k, m, n satisfying the conditions stated in the corollary. Then γ as in Theorem 2.4 equals $(1-\epsilon)^2/(8 \times 768)$. We take $c_1 := (24 \times 768) + 1$, so that $m > (3/\gamma) \ln k$. Since $k \geq 6/(1-\epsilon)$, we have $k > 6 > 2\mathrm{e} = \mathrm{e}/(1-\beta)$, and $m > (3/\gamma) \ln(\mathrm{e}/(1-\beta))$. We take $c_3 := (6 \times 2 \times 8 \times 768) + 1$ so that $n > (6km^3/\gamma(1-\beta)) \ln(16\sqrt{m}/\epsilon)$.

Now we consider an ensemble (ρ_{ij}) given by Theorem 2.4. Let Π' be any one-way protocol, possibly with shared entanglement, for the visible compression of the ensemble (ρ_{ij}) with average error at most $\epsilon/2$. Following the notation from Section 2.2.1, suppose that Bob holds registers ME_{B} just after he receives the message M from Alice in Π'. If the entanglement cost of Π' is e, we may assume that the register E_{B} may be partitioned into sub-registers $E_{1\mathrm{B}}E_{2\mathrm{B}}$ with $|E_{1\mathrm{B}}| = e$, and that the state of register E_{B} is of the form $\omega \otimes |0\rangle\langle 0|$, where $E_{1\mathrm{B}}$ is in state ω and $E_{2\mathrm{B}}$ in state $|0\rangle\langle 0|$, and $|0\rangle$ is a pure state. (We may achieve this by applying a suitable isometry to register E_{B}.)

Let $d := |ME_{1\mathrm{B}}|$, so that the sum of the communication and entanglement costs of Π' is $\log d$, and let σ_{ij} be the state of the registers $ME_{1\mathrm{B}}$ with Bob when Alice is given input (i,j). If $d \geq m$, the bound in Eq. (2.11) holds, so consider the case when $d < m$. Then the choice of n above implies that $n > (6kd^2m/\gamma(1-\beta)) \ln(16\sqrt{d}/\epsilon)$.

Since the average error of Π' is at most $\epsilon/2$, by the Markov Inequality we have

$$\left| \left\{ (i,j) : \|\rho_{ij} - \Psi(\sigma_{ij})\|_{\mathrm{tr}} > \epsilon \right\} \right| \quad < \quad \frac{n}{2} \quad = \quad \beta n \ ,$$

where Ψ is the quantum channel corresponding to Bob's decompression operation in Π'. Theorem 2.4 then implies that

$$2 + \frac{d}{\gamma} \ln\left(\frac{16}{1-\epsilon}\right) \quad \geq \quad m \ .$$

Since $m - 2 \geq m/2$, this gives us the bound stated in Eq. (2.11) with $c_2 := \log(16 \times 768)$. ■

Note that the parameter m may be chosen arbitrarily larger than k, provided the number of states n in the ensemble is chosen large enough. Thus, we see that there are ensembles with m-dimensional states for which communication-optimal compression protocols with shared entanglement and with constant average error, say $1/4$, have entanglement cost almost as large as $\log m$. In particular, the number of qubits of shared entanglement needed may be arbitrarily larger than the quantum information cost of the original protocol. We also see that in the *absence* of shared entanglement, there are ensembles

with m-dimensional states that cannot be compressed to states with dimension smaller than cm with average error less than $1/4$, where c is a universal positive constant. In particular, the optimally compressed message may be arbitrarily longer than the quantum information cost of the protocol Π.

Corollary 2.6 shows that the number of qubits of shared entanglement used by protocol with the smallest known communication cost, due to Anshu and Jain [4, Corollary 5], is optimal up to a constant multiplicative factor and an additive $\log k$ term (for constant error in compression). The lower bound on entanglement cost given in the corollary may be achieved by protocols derived from those for state splitting, up to an additive term of $\frac{1}{2}\log k + \mathrm{O}(1)$, again for constant error (see, e.g., Ref. [26, Lemma 3.3]). However, the communication cost of these protocols may not be optimal.

The probabilistic construction in the results above gives us ensembles with a number of states n that is polynomial in m and k. Note that in the compression protocol Π', Alice may send the input (i, j) as her message, in which case the message register has dimension n. Similarly, she may send the state ρ_{ij} itself, and this has dimension m. So in order to study how much compression is truly possible (i.e., how much smaller the dimension of the message register may be as compared with m), we have to study ensembles with $n \geq m$ states, and compression protocols with message registers with dimension at most m. Further, consider any protocol Γ (similar to Π) in which Alice receives a random input x out of n possibilities according to some distribution, prepares a state ω_x and sends it to Bob. The quantum information cost of such a protocol Γ is at most $\frac{1}{2}\log n$. So the polynomial dependence of n on the dimension of the states in the ensemble (m in the construction above) and the exponential dependence of n on the quantum information cost of the corresponding protocol ($\frac{1}{2}\log k$ in the construction) is inevitable.

2.4 Concluding remarks

In this chapter, we revisited one-shot compression of an ensemble of quantum states. We proved that there are ensembles which cannot be compressed by more than a few qubits in the absence of shared entanglement, when allowed constant error. In the presence of shared entanglement, the ensemble can be compressed to many fewer qubits. However, the entanglement cost may not be smaller than the number of qubits being compressed by more than a constant, for constant error. Since we study compression protocols that are allowed to make some error, the bounds we establish are robust to perturbations to the shared entangled state that are sufficiently small relative to the error.

Entanglement and quantum communication are distinct resources in the context of information processing. Sharing entanglement involves the generation, distribution, and storage of a state that is independent of the input for the task at hand. Communication also involves the same steps, but may be dynamic, i.e., may depend on the input and the prior history of the communication protocol. Consequently, any physical implementation of these resources is likely to incur different costs for these steps. In this chapter, we focused on the cost of distributing quantum states, and as a first stab, assumed that the cost of distribution for shared entanglement or for communication is proportional to the number of qubits involved. Formally, this corresponds to the notion of *smooth 0-Rényi entropy*. The motivation for this focus comes largely from the area of communication complexity [65], in which the interaction between multiple processors takes centre stage, but shared entanglement is often taken for granted. Our result shows that entanglement plays a crucial role in important communication tasks and highlights the need for considering entanglement cost in addition to communication cost.

A question of interest, from a theoretical perspective, is the *degree* or *strength* of entanglement required for different information processing tasks. Several different measures of entanglement have been studied in the literature, depending on the context. Smooth 0-Rényi entropy is a very coarse measure in this respect, as it may be the same for states that are regarded as having widely different degrees of entanglement. A natural question is whether results such as the ones we derived also hold for other definitions of entanglement cost that capture the degree of entanglement more satisfactorily. We conjecture that analogous results hold also for other measures, and leave this to future work.

Many other questions surrounding compression remain open. For instance, we do not have tight characterizations for the entanglement cost of one-shot state redistribution. Even lesser is known for the one-shot compression of interactive quantum protocols. Progress on these questions might hold the key to resolving important questions in communication complexity as well.

Chapter 3

Quantum state redistribution

3.1 Introduction

Quantum state redistribution is a communication task between two parties, Alice and Bob, defined as follows: Initially, a pure quantum state $|\psi\rangle^{RABC}$ (known to both Alice and Bob) is shared between Alice (AC), Bob (B) and Referee (R). The goal is to transmit the register C to Bob using a protocol involving only Alice and Bob, in a way that all correlations, including those with Referee, are not affected; see figure 3.1 for an illustration of state redistribution. In this chapter, we assume that Alice and Bob have access to an arbitrary shared entangled state as well as unlimited local computational power.

State redistribution is a generalization of the well-studied task of state merging and its time reversal, state splitting, in which registers A and B are trivial, respectively. The communication cost of state merging and splitting is known to be characterized asymptotically by $\mathrm{I}(R:C)$ [56], given many independent and identically distributed (i.i.d.) copies of $|\psi\rangle$. If instead, parties have access to only one copy of the state $|\psi\rangle$ (one-shot setting), the quantity which captures the cost is the smooth max-information the register C has about the register R [26, 3]. For state redistribution, the asymptotic i.i.d. communication cost is captured by the conditional mutual information $\mathrm{I}(R:C\,|\,B)$; however, the optimal communication cost in the one-shot setting remains an open problem. In this chapter, we present a protocol for one-shot quantum state redistribution whose communication cost is lower than the cost of all previously known protocols. Our result is the first one connecting quantum state redistribution and Markov chains, and can be interpreted as an operational interpretation for a possible one-shot analogue of quantum conditional mutual information.

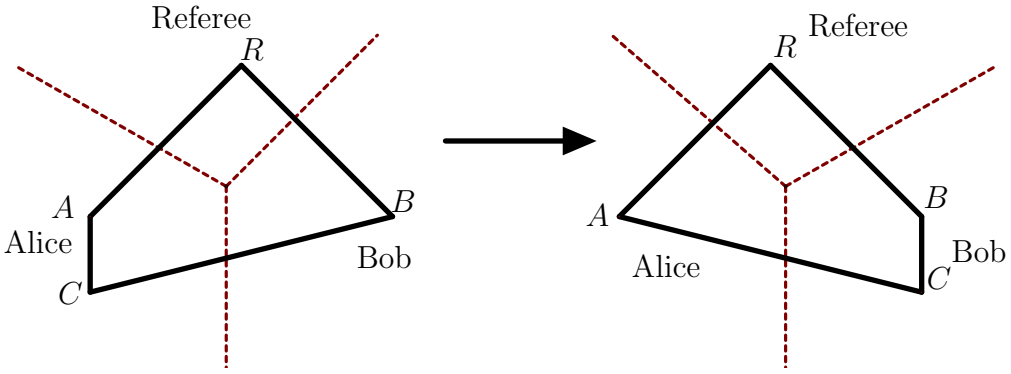

Figure 3.1: An illustration of quantum state redistribution

3.1.1 Previous works

Quantum state redistribution was first introduced by Luo and Devetak [69] who showed that $\frac{1}{2}\,\mathrm{I}(R:C\,|\,B)_\psi$ qubits of communication are necessary to attain this task in the asymptotic i.i.d. setting. Later, Yard and Devetak [38, 100] showed that this rate is also achievable and the conditional mutual information $\mathrm{I}(R:C\,|\,B)$ captures the optimal asymptotic cost of state redistribution. Clearly, an optimal state merging protocol can be used to redistribute a state $|\psi\rangle^{RABC}$ with asymptotic rate $\mathrm{I}(RA:C)$ assuming registers RA are with Referee. This protocol is not optimal, though, as $\mathrm{I}(R:C\,|\,B) = \mathrm{I}(R:C\,|\,A)$ can be much smaller than $\mathrm{I}(RA:C)$ for specific states. However, Oppenheim [78] and Ye, Bai and Wang [101], independently, showed that an optimal protocol for redistribution can be derived by combining two state merging protocols.

In the one-shot setting, where parties have access to only one copy of the state, Berta, Christandl and Touchette [27] and Datta, Hsieh and Oppenheim [36], independently obtained the following upper bound on the communication cost of state redistribution with error ϵ:

$$\frac{1}{2}\left[\mathrm{S}^{\epsilon}_{\max}(C\,|\,B)_\psi - \mathrm{S}^{\epsilon}_{\min}(C\,|\,BR)_\psi\right] + \mathrm{O}\left(\log(1/\epsilon)\right) \tag{3.1}$$

They used the aforementioned idea of combining state merging protocols. However, unlike the asymptotic setting, the resulting bound is not optimal. In particular, for trivial register B, there exists a quantum state for which $\mathrm{I}^{\epsilon}_{\max}(R:C)$ is much smaller than Eq. (3.1). Later, Anshu, Devabathiani and Jain [3] characterized the optimal cost of one-shot quantum state redistribution by the following expression:

$$\mathrm{Q}^{\epsilon}_{|\psi\rangle^{RABC}} \quad := \quad \frac{1}{2}\inf_{T,\sigma^T,U^{BCT}} \mathrm{I}^{\epsilon}_{\max}(RB:CT)_\kappa \ ,$$

where $U^{BCT} \in \mathsf{U}(BCT)$, $\sigma^T \in \mathsf{D}(T)$ subject to

$$\left(\mathbb{1}^R \otimes U^{BCT}\right) \kappa^{RBCT} \left(\mathbb{1}^R \otimes U^{BCT}\right)^\dagger \in \mathsf{B}^\epsilon(\psi^{RBC} \otimes \sigma^T)$$
$$\kappa^{RB} = \psi^{RB} \quad .$$

In the above definition, the size of the register T may be arbitrarily large. As a result, the given program does not give much information about the structure of an optimal protocol.

The best previously known one-shot protocol for quantum state redistribution is due to Anshu, Jain and Warsi [8]. Using *convex-split* and *position-based decoding* techniques, they designed a protocol for state redistribution with error at most 9ϵ and communication cost

$$\frac{1}{2} \inf_{\sigma^C} \inf_{\psi' \in \mathsf{B}^\epsilon(\psi^{RBC})} \left(\mathrm{D}_{\max}\left(\psi'^{RBC} \| \psi'^{RB} \otimes \sigma^C\right) - \mathrm{D}_{\mathrm{H}}^{\epsilon^2}\left(\psi'^{BC} \| \psi'^{B} \otimes \sigma^C\right) \right) + \log \frac{1}{\epsilon^2} \quad . \tag{3.2}$$

Although this cost is lower than the quantity in Eq. (3.1) (See Ref. [8], Theorem 4), it is still sub-optimal. In particular, when R is trivial, the cost in Eq. (3.2) can be as large as $\frac{1}{2}\log|C|$ while there is a zero cost protocol for redistributing state $|\psi\rangle^{ABC}$ in which Alice and Bob simply share $|\psi\rangle^{ABC}$, (A) with Alice and (BC) with Bob, as the initial entanglement.

A near-optimal bound on the one-shot rate for a classical version of this task was only recently provided by Anshu, Jain and Warsi [6], where the rate was shown to be

$$\mathrm{D}_{\max}^{\epsilon}(P_{XYZ} \| Q_{XYZ}) + \mathrm{O}\left(\log \frac{1}{\epsilon}\right), \tag{3.3}$$

where registers XZ are with Alice, register Y is with Bob, P_{XYZ} is the joint distribution of XYZ, Z is the random variable to be redistributed and Q_{XYZ} is the Markov chain given by the distribution $Q_{XYZ}(xyz) \coloneqq P_{X|Y}(x)P_Y(y)P_{Z|Y}(z)$.

3.1.2 Quantum Markov states

A tripartite quantum state $\sigma^{RBC} \in \mathsf{D}(\mathcal{H}^{RBC})$ is called a *quantum Markov state* if there exists a quantum operation $\Lambda : \mathsf{L}(\mathcal{H}^B) \to \mathsf{L}(\mathcal{H}^{BC})$ such that $(\mathbb{1} \otimes \Lambda)(\sigma^{RB}) = \sigma^{RBC}$, equivalently, if $\mathrm{I}(R : C \,|\, B) = 0$. This definition coincides with the notion of *Markov chains* for classical registers. Classical registers YXM form a *Markov chain* in this order (denoted as $Y{-}X{-}M$) if registers Y and M are independent given X. Hayden, Josza, Petz, and Winter [51] showed that an analogous property holds for quantum Markov states. In

particular, they showed that a state $\sigma^{RBC} \in \mathsf{D}(\mathcal{H}^R \otimes \mathcal{H}^B \otimes \mathcal{H}^C)$ is a Markov state if and only if there is a decomposition of the space $\mathcal{H}^B$ into a direct sum of tensor products as

$$\mathcal{H}^B \quad = \quad \bigoplus_j \mathcal{H}^{B_j^R} \otimes \mathcal{H}^{B_j^C} \quad , \tag{3.4}$$

such that

$$\sigma^{RBC} \quad = \quad \bigoplus_j p(j)\, \sigma_j^{RB_j^R} \otimes \sigma_j^{B_j^C C} \quad , \tag{3.5}$$

where $\sigma_j^{RB_j^R} \in \mathsf{D}\left(\mathcal{H}^R \otimes \mathcal{H}^{B_j^R}\right)$, $\sigma_j^{B_j^C C} \in \mathsf{D}\left(\mathcal{H}^{B_j^C} \otimes \mathcal{H}^C\right)$ and p is a probability distribution.

For a state ϕ^{RBC}, we say that σ^{RBC} is a *Markov extension* of ϕ^{RB} if $\sigma^{RB} = \phi^{RB}$ and σ^{RBC} is a quantum Markov state. We denote the set of all Markov extensions of ϕ^{RB} by $\mathsf{QMC}^{\phi}_{R-B-C}$. Note that $\mathsf{QMC}^{\phi}_{R-B-C}$ is non-empty, as we may take $\sigma^{RBC} \coloneqq \phi^{RB} \otimes \phi^C$.

For a Markov extension $\sigma \in \mathsf{QMC}^{\phi}_{R-B-C}$, let Π_j^σ be the orthogonal projection operator onto the j-th subspace of the register B given by the decomposition corresponding to the Markov state σ as described above. In other words, Π_j^σ is the projection onto the Hilbert space $\mathcal{H}^{B_j^R} \otimes \mathcal{H}^{B_j^C}$ in Eq. (3.4). For a quantum state ϕ^{RBC}, we define

$$\mathsf{ME}^{\epsilon,\phi}_{R-B-C} \coloneqq \left\{ \sigma \in \mathsf{QMC}^{\phi}_{R-B-C} \;\middle|\; \text{for all } j,\ \sigma_j^{B_j^C C} \in \mathsf{B}^\epsilon\left(\mathrm{Tr}_{B_j^R}\left[(\Pi_j^\sigma \otimes \mathbb{1})\phi^{BC}(\Pi_j^\sigma \otimes \mathbb{1})\right]\right) \right\} \quad .$$

Informally, this is the subset of Markov extensions σ of ϕ such that the restrictions of σ and ϕ to the j-th subspace in the decomposition of σ agree well on the registers $B_j^C C$. Again, the state $\sigma^{RBC} \coloneqq \phi^{RB} \otimes \phi^C$ belongs to $\mathsf{ME}^{\epsilon,\phi}_{R-B-C}$ for every $\epsilon \geq 0$, so the set is non-empty.

3.1.3 Our result

In Eq. (3.2), for a fixed ψ'^{RBC}, the minimization is over the extensions of ψ'^{RB} of the product form $\psi'^{RB} \otimes \sigma^C$. Notice that such product states satisfy $\mathrm{I}(R : C \,|\, B) = 0$, and hence, are *Markov states*. In this chapter, we derive an achievability bound similar to Eq. (3.2) for the one-shot communication cost of state redistribution where the minimization is instead over the larger subset $\mathsf{ME}^{\epsilon,\psi'}_{R-B-C}$ of *Markov extensions* of ψ'^{RB}. A simplified version of our result is stated in the following theorem. A more detailed statement is presented in Section 3.3, Theorem 3.9.

Theorem 3.1. *For any pure quantum state $|\psi\rangle^{RABC}$, the quantum communication cost of redistributing the register C from Alice (who initially holds AC) to Bob (who initially holds B) with error $10\sqrt{\epsilon}$ is at most*

$$\frac{1}{2} \inf_{\psi' \in \mathsf{B}^{\epsilon}(\psi^{RBC})} \inf_{\sigma^{RBC} \in \mathsf{ME}^{\epsilon^2/4,\psi'}_{R-B-C}} \left[\mathrm{D}_{\max}\left(\psi'^{RBC} \| \sigma^{RBC}\right) - \mathrm{D}^{\epsilon}_{\mathrm{H}}\left(\psi'^{BC} \| \sigma^{BC}\right) \right] + \log \frac{1}{\epsilon} + 1 \ .$$

The difference between minimizing over the set $\mathsf{ME}^{\epsilon,\psi'}_{R-B-C}$ versus QMC_{R-B-C} appears to be minor. We believe the above result can be stated in terms of a minimization over all of QMC_{R-B-C}. Note that $\sigma^C := \psi'^C$ is a nearly optimal solution for Eq. (3.2) as discussed in Ref. [34], and the product state $\psi'^{RB} \otimes \psi'^C$ is a Markov state in the set $\mathsf{ME}^{\epsilon,\psi'}_{R-B-C}$. So, our bound in Theorem 3.1 is tighter than Eq. (3.2) in the sense that the minimization is over a larger set. The protocol we design also recovers the near-optimal classical result in Ref. [6] when ψ^{RBC} is classical. Moreover, if register R is trivial, our protocol achieves the optimal cost of zero qubits of communication by choosing $\sigma^{BC} := \psi'^{BC}$, whereas the cost of the protocol in Ref. [8], stated in Eq. (3.2), may be as large as $(1/2)\log|C|$.

3.1.4 Motivations and implications

The connection between conditional mutual information and Markov chains has led to a rich body of results in classical computer science and information theory. It is well known that for any tripartite distribution P^{RBC},

$$\mathrm{I}(R:C\,|\,B)_P \quad = \quad \min_{Q^{RBC} \,\in\, \mathsf{MC}_{R-B-C}} \mathrm{D}\left(P^{RBC} \middle\| Q^{RBC}\right) \ ,$$

where MC_{R-B-C} is the set of Markov distributions Q, i.e., those that satisfy $\mathrm{I}(R:C\,|\,B)_Q = 0$. In fact, one can choose a distribution Q achieving the minimum above with $Q^{RB} = P^{RB}$ and $Q^{BC} = P^{BC}$. In the quantum case, the above identity fails drastically. For an example presented in Ref. [33] (see also Ref. [57, Section VI]), the right-hand side is a constant, whereas the left-hand side approaches zero as the system size increases. Given this, it is natural to ask if there is an extension of the classical identity to the quantum case. This is shown to be true in a sense that for any tripartite quantum state ψ^{RBC}, it holds that

$$\mathrm{I}(R:C|B)_{\psi} \quad = \quad \min_{\sigma^{RBC} \in \mathsf{QMC}_{R-B-C}} \left(\mathrm{D}\left(\psi^{RBC} \| \sigma^{RBC}\right) - \mathrm{D}\left(\psi^{BC} \| \sigma^{BC}\right)\right) \ . \tag{3.6}$$

This is implicitly proved in [28, Lemma 1]; we provide a proof in Appendix A for completeness. The difference between the quantum and the classical expressions can now be

understood as follows. For the classical case, the closest Markov chain Q to a distribution P (in relative entropy) satisfies the aforementioned relations $Q^{RB} = P^{RB}$ and $Q^{BC} = P^{BC}$. Thus, the second relative entropy term vanishes in Eq. (3.6). In quantum case, due to monogamy of entanglement we cannot in general ensure that $\sigma^{BC} = \rho^{BC}$. Thus, the relative entropy distance to Markov chains can be bounded away from the conditional mutual information.

Theorem 3.1 proves a one-shot analogue of Eq. (3.6). This is achieved in an operational manner, by showing that a one-shot analogue of the right-hand side in Eq. (3.6) is the achievable communication cost in quantum state redistribution of $|\psi\rangle^{RABC}$, a purification of ψ^{RBC}. Our bound also satisfies some desirable properties for a one-shot analogue of $\mathrm{I}(R : C \,|\, B)$ including non-negativity and monotonicity under local operations applied to register R. However, it is not clear whether it is also monotone under local quantum operations applied to register C.

In addition, the protocol we design is reversible. So, to redistribute C from Alice to Bob, Alice and Bob can instead run the time-reversal of the protocol in which register C is initially with Bob and he wants to send it to Alice. This implies the following corollary.

Corollary 3.2. *For any pure quantum state $|\psi\rangle^{RABC}$, the quantum communication cost of redistributing the register C from Alice (who initially holds AC) to Bob (who initially holds B) with error $10\sqrt{\epsilon}$ is at most the minimum of*

$$\frac{1}{2} \inf_{\psi' \in \mathsf{B}^{\epsilon}(\psi^{RBC})} \inf_{\sigma^{RBC} \in \mathsf{ME}^{\epsilon^2/4,\psi'}_{R-B-C}} \left[\mathrm{D}_{\max}\left(\psi'^{RBC} \| \sigma^{RBC}\right) - \mathrm{D}^{\epsilon}_{\mathrm{H}}\left(\psi'^{BC} \| \sigma^{BC}\right) \right] + \log \frac{1}{\epsilon} + 1$$

and

$$\frac{1}{2} \inf_{\psi' \in \mathsf{B}^{\epsilon}(\psi^{RAC})} \inf_{\sigma^{RAC} \in \mathsf{ME}^{\epsilon^2/4,\psi'}_{R-A-C}} \left[\mathrm{D}_{\max}\left(\psi'^{RAC} \| \sigma^{RAC}\right) - \mathrm{D}^{\epsilon}_{\mathrm{H}}\left(\psi'^{AC} \| \sigma^{AC}\right) \right] + \log \frac{1}{\epsilon} + 1 \ .$$

Recall that quantum conditional mutual information satisfies the duality property that $\mathrm{I}(R : C \,|\, B) = \mathrm{I}(R : C \,|\, A)$ for a pure state $|\psi\rangle^{RABC}$. The bound in the above corollary satisfies this property as well as non-negativity and monotonicity under local quantum operations applied to R.

The other motivation for studying state redistribution comes from the communication complexity setting. An important open problem in communication complexity is the *direct sum* problem which studies whether computing n copies of a Boolean function simultaneously requires as much communication as computing each independently. It has been

shown that direct sum holds for one-way [60] and bounded-round [30] *randomized communication complexity* as well as one-way *quantum communication complexity* [61, 62, 5]. Although it is known that direct sum does not hold if there is no restriction on the number of rounds in both classical [44] and quantum [9] settings, it is believed that a direct sum result must hold for quantum communication complexity when there are a constant number of rounds. A promising approach to prove this is to compress the communication in an interactive protocol to its *information content*. In fact, Braverman and Rao [30] showed that the problem of one-shot compression of interactive protocols is equivalent to the direct sum problem.

In an interactive protocol, parties know the description of their average joint state (averaged over all possible inputs) in each round of communication. Hence, the communication in each round can be compressed by performing quantum state redistribution. Inspired by this idea and the asymptotic cost of quantum state redistribution, Touchette [89] proposed a notion of *quantum information complexity* inspired by the asymptotic cost of state redistribution, and showed that a quantum state redistribution protocol with cost $\mathrm{O}(\mathrm{I}(R : C \,|\, B))$ suffices to derive a direct sum theorem for bounded-round quantum communication complexity. He also claimed such a direct sum result by bounding the cost in Eq. (3.1) in terms of $\mathrm{I}(R : C \,|\, B)$, but the proof has an error.

Although our protocol achieves a one-shot analogue of conditional mutual information in the sense described earlier, it is not clear whether the communication cost of our protocol is $\mathrm{O}(\mathrm{I}(R : C \,|\, B))$. We believe that the techniques used here shed light on a better understanding of quantum state redistribution and possibly leads to a near-optimal protocol.

3.1.5 Techniques

The protocol we design is most easily understood by considering a folklore protocol for redistributing quantum Markov states. In the case that ψ^{RBC} is a Markov state, its purification $|\psi\rangle^{RABC}$ can be transformed through local isometry operators $V_A : A \to A^R J' A^C$ and $V_B : B \to B^R J B^C$ as follows:

$$V_A \otimes V_B |\psi\rangle^{RABC} \quad = \quad \sum_j \sqrt{p(j)} |\psi_j\rangle^{RA^R B^R} \otimes |jj\rangle^{JJ'} \otimes |\psi_j\rangle^{A^C B^C C} \ . \tag{3.7}$$

The existence of isometry operators V_A and V_B is a consequence of the special structure of quantum Markov states explained in Section 3.1.2, Eq. (3.5). Note that after the above transformation, conditioned on registers J and J', systems $RA^R B^R$ are decoupled from

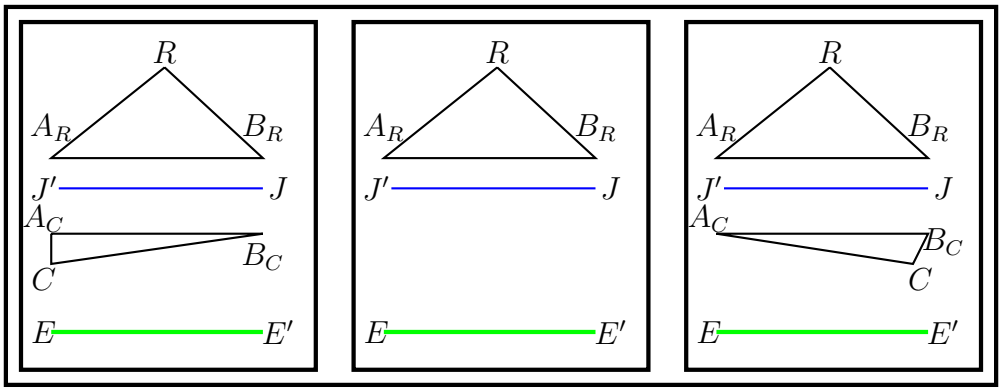

Figure 3.2: An illustration of the zero-cost protocol for redistributing Markov states. Left: Registers $RA^RB^RJJ'A^CCB^C$ are in the state given in Eq. (3.7) and registers E and E' contain Alice and Bob's share of an embezzling state, respectively. Middle: Using embezzling registers, Alice and Bob jointly embezzled out registers A^CCB^C via local unitary operations. Right: Using embezzling registers, conditioned on J and J', Alice and Bob embezzled in $|\psi_j\rangle^{A^CCB^C}$ such that registers C and B^C are with Bob and register A^C is with Alice. This step also only contain local unitary operations and no communication.

systems A^CCB^C. So using the embezzling technique due to van Dam and Hayden [92], conditioned on J and J', Alice and Bob can first embezzle-out systems A^CCB^C and then embezzle-in the same systems such that at the end the global state is close to the state in Eq. (3.7) and system C is with Bob. This protocol incurs no communication; see Fig. 3.2 for an illustration.

The protocol we design is a more sophisticated version of the above protocol. The key technique underlying our protocol is a reduction procedure using embezzling quantum states, that allows us to use a protocol due to Anshu, Jain, and Warsi [8] as a subroutine. Let σ^{RBC} be a quantum Markov extension of ψ^{RB}. The reduction is a unitary procedure which decouples C from RB when applied to σ^{RBC}, while preserving ψ^{RB} when applied to ψ^{RBC}.

To elaborate further, consider the special case where ψ^{RBC} is the GHZ state

$$\frac{1}{\sqrt{d}} \sum_{j=1}^{d} |j\rangle^R |j\rangle^B |j\rangle^C \ .$$

In this case, the closest Markov extension σ^{RBC} of ψ^{RB} is $\frac{1}{d}\sum_{j=1}^{d} |j\rangle\langle j|^R \otimes |j\rangle\langle j|^B \otimes |j\rangle\langle j|^C$. A naive way to decouple register C from registers RB in σ^{RBC} is to coherently erase register C

conditioned on register B. However, the same operation applied to ψ^{RBC} changes ψ^{RB}. To overcome this problem, first, we coherently measure register B by adding a maximally entangled state $|\Phi\rangle^{TT'}$ and making another "copy" of $|j\rangle^B$ in Φ^T. The copying is done by applying a distinct Heisenberg-Weyl operator to the state Φ^T, for each $j \in [d]$. This operation measures register B in ψ^{RBC}, keeps σ^{RBC} unchanged, and and leaves Φ^T in tensor product with registers RB in both ψ and σ. Then, conditioned on register B, we can coherently erase register C in σ^{RBC}; this operation applied to ψ does not change the state ψ^{RB}.

For a general state ψ^{RBC} with quantum Markov extension σ^{RBC}, the isometry operator V_B can be used to transform σ^{RBC} to the classical-quantum state

$$\sum_j p(j)\sigma_j^{RB^R} \otimes |j\rangle\langle j|^J \otimes \sigma_j^{B^C C} \quad .$$

However, we encounter an additional issue here: for a given j, $\sigma_j^{B^C C}$ does not necessarily have a flat spectrum. So we first flatten each $\sigma_j^{B^C C}$ through a unitary procedure. This task can be achieved via the technique of coherent flattening via embezzlement due to Anshu and Jain [4]. After flattening, the dimension of the support of systems $B^C C$ no more depends on j and so the states in registers $B^C C$ can be all rotated to a flat state over a fixed subspace. Hence, $B^C C$ gets decoupled from $RB^R J$ in the state σ. Finally, to keep ψ^{RB} unchanged, we regenerate the system B^C via a standard embezzling technique similar to the protocol in Fig. 3.2.

Organization The rest of this chapter is organized as follows. In Section 3.2, we review the notions and techniques required for the proof of the main theorem. In Section 3.3, we state our result formally in Theorem 3.3 and prove it. Then, in Section 3.4, we discuss the optimality of the derived bound in the asymptotic i.i.d. setting. Finally, we summarize our result and explain some interesting open questions in Section 3.5.

3.2 Preliminaries

3.2.1 Quantum state redistribution

Consider a pure state $|\psi\rangle^{RABC}$ shared between Referee (R), Alice (AC) and Bob (B). In an ϵ-error *quantum state redistribution* protocol, Alice and Bob share an entangled state $|\theta\rangle^{E_A E_B}$, register E_A with Alice and register E_B with Bob. Alice applies an encoding

operation $\mathcal{E} : \mathsf{L}(\mathcal{H}^{ACE_A}) \to \mathsf{L}(\mathcal{H}^{AQ})$, and sends the register Q to Bob. Then, Bob applies a decoding operation $\mathcal{D} : \mathsf{L}(\mathcal{H}^{QBE_B}) \to \mathsf{L}(\mathcal{H}^{BC})$. The output of the protocol is the state ϕ^{RABC} with the property that $\mathrm{P}(\psi^{RABC}, \phi^{RABC}) \leq \epsilon$, and the communication cost of the protocol is $\log |Q|$.

To derive the upper bound in Theorem 3.1, we use an existing protocol due to Anshu, Jain and Warsi [8] which we call the AJW protocol in the sequel. The bound is derived by combining the AJW protocol with a decoupling technique via embezzling described in Section 3.2.2.

AJW protocol

The AJW protocol is based on two powerful and remarkable techniques, *convex-split lemma* (introduced in Ref. [3]) and *position-based decoding* (introduced in Ref. [7]).

Let n be an integer, $\rho^{AB} \in \mathsf{D}(\mathcal{H}^{AB})$ and $\sigma^B \in \mathsf{D}(\mathcal{H}^B)$. Consider the quantum state $\tau^{AB_1\ldots B_n}$ derived by adding $n-1$ independent copies of σ^B in tensor product with ρ^{AB} and swapping j-th copy of σ^B with ρ^B for uniformly random $j \in [n-1]$. In particular, we define

$$\tau^{AB_1B_2\ldots B_n} \quad := \quad \frac{1}{n}\sum_{j=1}^{n} \rho^{AB_j} \otimes \sigma^{B_1} \otimes \cdots \otimes \sigma^{B_{j-1}} \otimes \sigma^{B_{j+1}} \otimes \cdots \otimes \sigma^{B_n} \ , \tag{3.8}$$

on $n+1$ registers $A, B_1, B_2, \ldots, B_n$, where $\rho^{AB_j} \otimes \sigma^{B_1} \otimes \cdots \otimes \sigma^{B_{j-1}} \otimes \sigma^{B_{j+1}} \otimes \cdots \otimes \sigma^{B_n}$ denotes the state $\mathrm{SWAP}_{B_1 \rightleftarrows B_j} \left(\rho^{AB_1} \otimes \sigma^{B_2} \otimes \cdots \otimes \sigma^{B_j} \otimes \cdots \otimes \sigma^{B_n}\right) \mathrm{SWAP}^{\dagger}_{B_1 \rightleftarrows B_j}$ as explained in Section 1.2.2. The convex-split lemma states that the state $\tau^{AB_1\ldots B_n}$ is almost indistinguishable from the product state $\rho^A \otimes (\sigma^B)^{\otimes n}$, provided that n is large enough.

Lemma 3.3 (Convex-split lemma [3])**.** *Let $\rho^{AB} \in \mathsf{D}(\mathcal{H}^{AB})$ and $\sigma^B \in \mathsf{D}(\mathcal{H}^B)$ be quantum states with $\mathrm{D}_{\max}(\rho^{AB} \| \rho^A \otimes \sigma^B) = k$ for some finite number k. Let $\delta > 0$ and $n = \lceil \frac{2^k}{\delta} \rceil$. For the state $\tau^{AB_1\ldots B_n}$ defined in Eq. (3.8), we have*

$$\mathrm{P}\left(\tau^{AB_1\ldots B_n}, \tau^A \otimes \sigma^{B_1} \otimes \cdots \otimes \sigma^{B_n}\right) \quad \leq \quad \sqrt{\delta} \ . \tag{3.9}$$

The above lemma provides the condition under which the correlation in between registers A and B in ρ is lost in a certain convex combination of quantum states. A dual problem is to find conditions sufficient for identifying the location of the desired correlation in a convex combination. This task is achievable via position-based decoding technique using quantum hypothesis testing.

Lemma 3.4 (Position-based decoding [7]). *Let $\epsilon > 0$, and $\rho^{AB} \in \mathsf{D}(\mathcal{H}^{AB})$ and $\sigma^B \in \mathsf{D}(\mathcal{H}^B)$ be quantum states such that* $\mathrm{supp}(\rho^B) \subseteq \mathrm{supp}(\sigma^B)$. *Let* $n := \lceil \epsilon\, 2^{\mathrm{D}_{\mathrm{H}}^{\epsilon}(\rho^{AB}\|\rho^A\otimes\sigma^B)} \rceil$, *and for every* $j \in [n]$,

$$\tau_j^{AB_1\ldots B_n} \quad := \quad \rho^{AB_j} \otimes \sigma^{B_1} \otimes \cdots \otimes \sigma^{B_{j-1}} \otimes \sigma^{B_{j+1}} \otimes \cdots \otimes \sigma^{B_n} \ .$$

There exists a set of POVM operators $\{\Lambda_j : j \in [n+1]\}$ *on registers* $AB_1B_2\ldots B_n$ *such that*

$$\sum_{j=1}^{n+1} \Lambda_j \quad = \quad \mathbb{1}$$

and for all $j \in [n]$,

$$\mathrm{Tr}\big[\Lambda_j \tau_j^{AB_1\ldots B_n}\big] \quad \geq \quad 1 - 6\epsilon \ .$$

The above statement is slightly different from the one in Ref. [7] because of a minor difference in defining quantum hypothesis testing relative entropy.

Let $|\psi\rangle^{RABC}$ be the quantum state shared between Alice (AC), Bob (B) and Referee (R), and $\psi'^{RBC} \in \mathsf{B}^{\epsilon}(\psi^{RBC})$. The AJW protocol works as follows.

AJW protocol: Alice and Bob initially share $m = \lceil 2^{\beta}/\epsilon^2 \rceil$ copies of a purification $|\sigma\rangle^{LC}$ of σ^C where $\beta = \mathrm{D}_{\max}\big(\psi'^{RBC}\|\psi'^{RB}\otimes\sigma^C\big)$.So their global state is

$$|\psi\rangle^{RABC} \otimes |\sigma\rangle^{L_1C_1} \otimes \ldots \otimes |\sigma\rangle^{L_mC_m} \ .$$

Let b be the smallest integer such that $\log b \geq \mathrm{D}_{\mathrm{H}}^{\epsilon^2}(\psi'^{BC}\|\psi'^{B}\otimes\sigma^C) - \log\frac{1}{\epsilon^2}$. By performing a proper unitary operator, Alice transforms the global state into a state close to the state

$$\frac{1}{m}\sum_{j=1}^{m} |\lfloor j/b \rfloor\rangle^{J_1}|j \pmod b\rangle^{J_2}|0\rangle^{L_j}|\psi\rangle^{RABC_j} \otimes |\sigma\rangle^{L_1C_1} \otimes \ldots \otimes |\sigma\rangle^{L_{j-1}C_{j-1}} \otimes |\sigma\rangle^{L_{j+1}C_{j+1}} \otimes \ldots \otimes |\sigma\rangle^{L_nC_n} \ . \quad (3.10)$$

This is possible due to the Uhlmann theorem, the convex-split lemma and the choice of m. Alice sends register J_1 to Bob with communication cost at most $1/2(\log m - \log b)$ using superdense coding. Then, for each $j_2 \leq b$, Bob swaps registers C_{j_2} and $C_{j_2+bj_1}$, conditioned on $J_1 = j_1$. At this point, registers $RBC_1\ldots C_b$ are in a state close to

$$\frac{1}{b}\sum_{j_2=1}^{b} \psi^{RBC_{j_2}} \otimes \sigma^{C_1} \otimes \ldots \otimes \sigma^{C_{j_2-1}} \otimes \sigma^{C_{j_2+1}} \otimes \ldots \otimes \sigma^{C_b} \ . \quad (3.11)$$

Then, Bob uses position-based decoding to determine the index j_2 for which register C_{j_2} is correlated with registers RB. This is possible by the choice of b. Since the state over registers RBC_{j_2} is close to ψ^{RBC} and it is independent of the state over registers $C_1 \dots C_{j_2-1}, C_{j_2+1}, \dots, C_b$, the register purifying registers RBC_{j_2} is with Alice and she can transform it to the register A such that the final state over registers $RABC_{j_2}$ is close to ψ^{RABC}.

The following theorem states the communication cost and the error in the final state of the above protocol.

Theorem 3.5 ([8]). *Let $\epsilon \in (0,1)$, and $|\psi\rangle^{RABC}$ be a pure quantum state shared between Referee (R), Alice (AC) and Bob (B). There exists an entanglement-assisted one-way protocol operated by Alice and Bob which starts in the state $|\psi\rangle^{RABC}$, and outputs a state $\phi^{RABC} \in \mathsf{B}^{9\epsilon}(\psi^{RABC})$, and the number of qubits communicated by Alice and Bob is upper bounded by*

$$\frac{1}{2} \inf_{\sigma^C} \inf_{\psi' \in \mathsf{B}^{\epsilon}(\psi^{RBC})} \left(\mathrm{D}_{\max}\left(\psi'^{RBC} \| \psi'^{RB} \otimes \sigma^C\right) - \mathrm{D}_{\mathrm{H}}^{\epsilon^2}\left(\psi'^{BC} \| \psi'^{B} \otimes \sigma^C\right) \right) + \log \frac{1}{\epsilon^2} \ .$$

For a complete proof (including correctness and error analysis), see the proof of Theorem 1 in Ref. [8].

3.2.2 Decoupling classical-quantum states via embezzlement

Embezzlement refers to a process introduced by van Dam and Hayden [92] in which any bipartite quantum state, possibly entangled, can be produced from a bipartite catalyst, called the *embezzling quantum state*, using only local unitary operations. For an integer n and registers D and D' with $|D| = |D'| \geq n$, the embezzling state is defined as

$$|\xi\rangle^{DD'} \quad := \quad \frac{1}{\sqrt{S(n)}} \sum_{i=1}^{n} \frac{1}{\sqrt{i}} |i\rangle^D |i\rangle^{D'} \ , \tag{3.12}$$

where $S(n) := \sum_{i=1}^{n} \frac{1}{i}$. Van Dam and Hayden showed that for every bipartite state $|\phi\rangle^{AB}$ with Schmidt rank m, there exists local isometries $V_A : \mathcal{H}^D \to \mathcal{H}^{DA}$ and $V_B : \mathcal{H}^{D'} \to \mathcal{H}^{D'B}$ such that

$$\mathrm{P}((V_A \otimes V_B)|\xi\rangle, |\xi\rangle \otimes |\phi\rangle) \quad \leq \quad \delta \ , \tag{3.13}$$

provided that $n \geq m^{2/\delta^2}$. Therefore, by using n large enough, any accuracy in embezzlement can be achieved.

For a fixed $a \leq n$, a close variant of the above embezzling state is defined as

$$|\xi_{a:n}\rangle^{DD'} \quad := \quad \frac{1}{\sqrt{S(a,n)}} \sum_{i=a}^{n} \frac{1}{\sqrt{i}} |i\rangle^D |i\rangle^{D'} \ . \tag{3.14}$$

Using these states, Lemma 3.6 achieves the embezzling of the uniform distribution where the closeness is guaranteed in max-relative entropy.

Lemma 3.6 ([4]). *Let $\delta \in (0, \frac{1}{15})$, and $a, b, n \in \mathbb{Z}$ be positive integers such that $a \geq b$ and $n \geq a^{1/\delta}$. Let D and E be registers with $|D| \geq n$ and $|E| \geq b$. Let W_b be a unitary operation that acts as*

$$W_b |i\rangle^D |0\rangle^E \quad = |\lfloor i/b \rfloor\rangle^D |i \pmod b\rangle^E \qquad \forall i \in \{0, \ldots |D| - 1\} \ , \tag{3.15}$$

and $\Pi_b \in \mathsf{Pos}\big(\mathcal{H}^{DE}\big)$ be the projection operator onto the support of $W_b \left(\xi^D_{a:n} \otimes |0\rangle\langle 0|^E\right) W_b^\dagger$. It holds that

$$W_b \left(\xi^D_{a:n} \otimes |0\rangle\langle 0|^E\right) W_b^\dagger \quad \leq \quad (1 + 15\delta)\ \xi^D_{1:n} \otimes \mu_b^E \ , \tag{3.16}$$

and

$$\Pi_b \big(\xi^D_{1:n} \otimes \mu_b^E\big) \Pi_b \quad \leq \quad 2 \cdot W_b \left(\xi^D_{a:n} \otimes |0\rangle\langle 0|^E\right) W_b^\dagger \ . \tag{3.17}$$

where $\mu_b^E = \frac{1}{b} \sum_{e=0}^{b-1} |e\rangle\langle e|$.

The proof of Eq. (3.16) is due to Anshu and Jain [4, Claim 1], and Eq. (3.17) follows from a similar argument. For completeness, we provide a proof for Lemma 3.6 below.
Proof: Let W_b be a unitary operator satisfying Eq. (3.15). We have

$$\begin{aligned}
W_b \left(\xi^D_{a:n} \otimes |0\rangle\langle 0|^E\right) W_b^\dagger \quad &= \quad \frac{1}{S(a,n)} \sum_{i=1}^{n} \frac{1}{i} W_b \left(|i\rangle\langle i|^D \otimes |0\rangle\langle 0|^E\right) W_b^\dagger \\
&= \quad \frac{1}{S(a,n)} \sum_{i=1}^{n} \frac{1}{i} \left||\lfloor i/b \rfloor\rangle\langle \lfloor i/b \rfloor|^D \otimes |i \pmod b\rangle\langle i \pmod b|^E \right. \\
&= \quad \frac{1}{S(a,n)} \sum_{i'=\lfloor \frac{a}{b} \rfloor}^{\lfloor \frac{n}{b} \rfloor} \sum_{e=0}^{\min\{b-1, n-i'b\}} \frac{1}{bi'+e} |i'\rangle\langle i'|^D \otimes |e\rangle\langle e|^E \qquad (3.18) \\
&\leq \quad \frac{1}{S(a,n)} \sum_{i'=\lfloor \frac{a}{b} \rfloor}^{\lfloor \frac{n}{b} \rfloor} \sum_{e=0}^{b-1} \frac{1}{bi'} |i'\rangle\langle i'|^D \otimes |e\rangle\langle e|^E \\
&\leq \quad \frac{S(1,n)}{S(a,n)} \xi^D_{1:n} \otimes \mu_b^E \ . \qquad (3.19)
\end{aligned}$$

In Ref. [70], it is shown that $\left|S(a,n) - \log \frac{n}{a}\right| \leq 4$. Since $n \geq a^{1/\delta}$, we have

$$\frac{S(1,n)}{S(a,n)} \quad \leq \quad \frac{\log n + 4}{\log n - \log a - 4} \quad \leq \quad \frac{1+4\delta}{1-5\delta} \quad \leq \quad 1 + 15\delta \ . \tag{3.20}$$

Now, Eq. (3.19) and Eq. (3.20) together imply Eq. (3.16). It remains to prove Eq. (3.17). Let $\Pi_b \in \mathsf{Pos}\big(\mathcal{H}^{DE}\big)$ be the projection operator onto the support of $W_b \left(\xi^D_{a:n} \otimes |0\rangle\langle 0|^E\right) W_b^\dagger$. Eq. (3.18) implies that

$$\Pi_b \quad = \quad \sum_{i'=\lfloor \frac{a}{b} \rfloor}^{\lfloor \frac{n}{b} \rfloor} \ \sum_{e=0}^{\min\{b-1, n-i'b\}} |i'\rangle\langle i'|^D \otimes |e\rangle\langle e|^E \ .$$

Thus,

$$\begin{aligned}
\Pi_b \big(\xi^D_{1:n} \otimes \mu^E_b\big) \Pi_b \quad &= \quad \frac{1}{S(1,n)} \sum_{i'=\lfloor \frac{a}{b} \rfloor}^{\lfloor \frac{n}{b} \rfloor} \ \sum_{e=0}^{\min\{b-1, n-i'b\}} \frac{1}{bi'} \, |i'\rangle\langle i'|^D \otimes |e\rangle\langle e|^E \\
&\leq \quad \frac{1}{S(1,n)} \sum_{i'=\lfloor \frac{a}{b} \rfloor}^{\lfloor \frac{n}{b} \rfloor} \ \sum_{e=0}^{\min\{b-1, n-i'b\}} \frac{2}{bi'+e} \, |i'\rangle\langle i'|^D \otimes |e\rangle\langle e|^E \\
&= \quad \frac{2 \cdot S(a,n)}{S(1,n)} \, W_b \left(\xi^D_{a:n} \otimes |0\rangle\langle 0|^E\right) W_b^\dagger \qquad \text{(by Eq. (3.16))} \\
&\leq \quad 2 \cdot W_b \left(\xi^D_{a:n} \otimes |0\rangle\langle 0|^E\right) W_b^\dagger \ ,
\end{aligned}$$

where the first inequality holds since $bi' + e \leq 2\, bi'$ for $i' \geq 1$ and $0 \leq e \leq b-1$, and the second inequality holds since $S(a,n) \leq S(1,n)$. ∎

As a corollary of the above lemma, Anshu and Jain [4] showed that the embezzling state $\xi^D_{a:n}$ can be used almost catalytically to *flatten* any quantum state using unitary operations.

Corollary 3.7 ([4], Eq. (6))**.** *Let $\rho \in \mathsf{D}(\mathcal{H}^C)$ be a quantum state with spectral decomposition $\rho^C = \sum_c q(c) |v_c\rangle\langle v_c|^C$. Let $\delta \in (0, \frac{1}{15})$ and $\gamma \in (0,1)$ such that $\frac{|C|}{\gamma}$ is an integer and all eigenvalues $q(c)$ are integer multiples of $\frac{\gamma}{|C|}$. Let $a := \frac{|C|}{\gamma} \max_c q(c)$, $n = a^{1/\delta}$, and D and E be quantum registers with $|D| \geq n$ and $|E| = a$. Let $W \in \mathsf{U}(\mathcal{H}^{CED})$ be the unitary operator defined as*

$$W \quad := \quad \sum_c |v_c\rangle\langle v_c|^C \otimes W_{b(c)}$$

and $\Pi \in \mathsf{Pos}(\mathcal{H}^{CED})$ *be the projection operator defined as*

$$\Pi \quad := \quad \sum_c |v_c\rangle\langle v_c|^C \otimes \Pi_{b(c)} \ ,$$

where $W_{b(c)}$ *and* $\Pi_{b(c)}$ *are the operators defined in Lemma 3.6 with* $b(c) := \frac{q(c)|C|}{\gamma}$. *Then, we have*

$$W\left(\rho^C \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) W^\dagger \quad \leq \quad (1+15\delta)\ \rho^{CE} \otimes \xi^D_{1:n} \tag{3.21}$$

and

$$\Pi\big(\rho^{CE} \otimes \xi^D_{1:n}\big)\Pi \quad \leq \quad 2 \cdot W\left(\rho^C \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) W^\dagger \ , \tag{3.22}$$

where $\rho^{CE} := \frac{\gamma}{|C|}\sum_c |v_c\rangle\langle v_c|^C \otimes \sum_{e=0}^{b(c)-1} |e\rangle\langle e|^E$ *is an extension of* ρ^C *with flat spectrum.*

The proof of Eq. (3.21) is provided in Ref. [4, Eq. (6)], and Eq. (3.22) follows from a Eq. (3.17). For completeness, we provide a proof for Corollary 3.7 below.

Proof: Let W be the unitary operator defined in the statement of the corollary . We have

$$\begin{aligned}
&W\left(\rho^C \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) W^\dagger \\
&= \quad \sum_c q(c)|v_c\rangle\langle v_c|^C \otimes W_{b(c)}\big(|0\rangle\langle 0|^E \otimes \xi^D_{a:n}\big)W^\dagger_{b(c)} \\
&\leq \quad (1+15\delta)\ \sum_c q(c)|v_c\rangle\langle v_c|^C \otimes \frac{\gamma}{q(c)|C|}\ \sum_{e=0}^{b(c)-1} |e\rangle\langle e|^E \otimes \xi^D_{a:n} \\
&= \quad (1+15\delta)\ \rho^{CE} \otimes \xi^D_{a:n} \ ,
\end{aligned}$$

where the inequality follows from Lemma 3.6. So, it remains to prove Eq. (3.22). Let Π be the projection operator defined in the statement of the corollary. We have

$$\begin{aligned}
\Pi\left(\rho^{CE} \otimes \xi^D_{1:n}\right)\Pi \quad &= \quad \frac{\gamma}{|C|}\sum_c b(c)|v_c\rangle\langle v_c|^C \otimes \Pi_{b(c)}\big(\mu^E_{b(c)} \otimes \xi^D_{a:n}\big)\Pi_{b(c)} \\
&\leq \quad 2\sum_c q(c)|v_c\rangle\langle v_c|^C \otimes W_{b(c)}\big(|0\rangle\langle 0|^E \otimes \xi^D_{a:n}\big)W^\dagger_{b(c)} \\
&= \quad 2 \cdot W\big(\rho^C \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\big)W^\dagger \ ,
\end{aligned}$$

where the inequality is a consequence of Lemma 3.6. ∎

We use the above flattening procedure to decouple the quantum register in a classical-quantum state.

Corollary 3.8. *Consider a classical-quantum state* $\rho^{JC} = \sum_j p(j)|j\rangle\langle j|^J \otimes \rho_j^C$*, where* p *is a probability distribution and* $\rho_j^C \in \mathsf{D}\left(\mathcal{H}^C\right)$*. Let* $\delta \in (0, \frac{1}{15})$ *and* $\gamma \in (0,1)$ *such that* $a \coloneqq \frac{|C|}{\gamma}$ *is an integer and eigenvalues of all* ρ_j^C *are integer multiples of* $\frac{\gamma}{|C|}$*. Let* $n = a^{1/\delta}$*,* D *and* E *be quantum registers with* $|D| \geq n$ *and* $|E| = a$*. Then, there exists a unitary operator* $U \in \mathsf{U}(\mathcal{H}^{JCED})$*, read-only on register* J*, and a projection operator* $\widetilde{\Pi} \in \mathsf{Pos}(\mathcal{H}^{JCED})$ *such that*

$$U\left(\rho^{JC} \otimes |0\rangle\langle 0|^E \otimes \xi_{a:n}^D\right) U^\dagger \quad \leq \quad (1+15\delta)\ \rho^J \otimes \nu^{CE} \otimes \xi_{1:n}^D \ , \tag{3.23}$$

$$\widetilde{\Pi}\left(\rho^J \otimes \nu^{CE} \otimes \xi_{1:n}^D\right)\widetilde{\Pi} \quad \leq \quad 2 \cdot U\left(\rho^{JC} \otimes |0\rangle\langle 0|^E \otimes \xi_{a:n}^D\right) U^\dagger \ , \tag{3.24}$$

and

$$\mathrm{Tr}\left[\widetilde{\Pi} U\left(\rho^{JC} \otimes |0\rangle\langle 0|^E \otimes \xi_{a:n}^D\right) U^\dagger\right] \quad = \quad 1 \ , \tag{3.25}$$

where $\nu^{CE} = \frac{1}{a}\sum_{s=0}^{a-1} |s\rangle\langle s|^{CE}$*.*

Proof: Notice that integers a and n and registers D and E satisfy required properties in Corollary 3.7. For each j, let $W^{(j)}$ be the unitary operator given by Corollary 3.7 for flattening $\rho_j^C = \sum_c q_j(c)|v_{j,c}\rangle\langle v_{j,c}|$. Hence, we can flatten all ρ_j^C simultaneously using the unitary operator $U_1 = \sum_j |j\rangle\langle j| \otimes W^{(j)}$, and we get

$$U_1\left(\rho^{JC} \otimes |0\rangle\langle 0|^E \otimes \xi_{a:n}^D\right) U_1^\dagger \quad \leq \quad (1+15\delta)\sum_j p(j)|j\rangle\langle j|^J \otimes \rho_j^{CE} \otimes \xi_{1:n}^D \ ,$$

where $\rho_j^{CE} = \frac{\gamma}{|C|}\sum_c |v_{j,c}\rangle\langle v_{j,c}|^C \otimes \sum_{e=0}^{q_j(c)|C|/\gamma} |e\rangle\langle e|^E$ is an extension of ρ^C with flat spectrum. For each j, the support of ρ_j^C has dimension $\sum_c q_j(c)\frac{|C|}{\gamma} = a$, which is independent of j. Hence, there exists a unitary operator $V^{(j)}$ mapping ρ_j^{CE} to ν^{CE}. Define the unitary operator $U_2 \coloneqq \sum_j |j\rangle\langle j| \otimes V^{(j)}$ on registers JCE. Then, the unitary operator $U \coloneqq U_2 U_1$ satisfies Eq. (3.23).

Now, for each j, let $\Pi^{(j)} \in \mathsf{Pos}(\mathcal{H}^{CED})$ be the projection operator given by Corollary 3.7.

Define $\Pi' \coloneqq \sum_j |j\rangle\langle j| \otimes \Pi^{(j)}$ and $\widetilde{\Pi} \coloneqq U_2 \Pi' U_2^\dagger$. We have

$$\begin{aligned}
\widetilde{\Pi}\left(\rho^J \otimes \nu^{CE} \otimes \xi^D_{1:n}\right)\widetilde{\Pi} &= U_2\Pi' U_2^\dagger \left(\rho^J \otimes \nu^{CE} \otimes \xi^D_{1:n}\right) U_2 \Pi' U_2^\dagger \\
&= U_2 \Pi' \left(\sum_j p(j) |j\rangle\langle j|^J \otimes \rho_j^{CE} \otimes \xi^D_{1:n}\right) \Pi' U_2^\dagger \\
&= U_2 \left(\sum_j p(j) |j\rangle\langle j|^J \otimes \Pi^{(j)} \left(\rho_j^{CE} \otimes \xi^D_{1:n}\right) \Pi^{(j)}\right) U_2^\dagger \\
&\leq 2 \cdot U_2 \left(\sum_j p(j) |j\rangle\langle j|^J \otimes W^{(j)} \left(\rho_j^C \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) W^{(j)\dagger}\right) U_2^\dagger \\
&= 2 \cdot U_2 U_1 \left(\sum_j p(j) |j\rangle\langle j|^J \otimes \rho_j^C \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) U_1^\dagger U_2^\dagger \\
&= 2 \cdot U \left(\rho^{JC} \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) U^\dagger \ ,
\end{aligned}$$

where the inequality follows from Corollary 3.7, Eq. (3.22).

Moreover, by the construction in Lemma 3.6 and Corollary 3.7, for each j, the operator $\Pi^{(j)}$ is the projection operator onto the support of $W^{(j)}\left(\rho_j^C \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) W^{(j)\dagger}$. Hence, we have

$$\begin{aligned}
\mathrm{Tr}\left[\widetilde{\Pi} U \left(\rho^{JC} \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) U^\dagger\right] &= \mathrm{Tr}\left[\Pi' U_1 \left(\rho^{JC} \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) U_1^\dagger\right] \\
&= \sum_j p(j) \mathrm{Tr}\left[\Pi^{(j)} W^{(j)} \left(\rho_j^C \otimes |0\rangle\langle 0|^E \otimes \xi^D_{a:n}\right) W^{(j)\dagger}\right] \\
&= 1 \ .
\end{aligned}$$

This completes the proof. ∎

Remark: In the above corollary, there is an implicit assumption that the eigenvalues of ρ_j^C are rational. However, we can assume without loss of generality that this is always the case since the set of rational numbers is dense and the error due to this assumption can be made arbitrarily close to zero.

3.3 Main result

In this section, we prove our main result.

Theorem 3.9. *Let $|\psi\rangle^{RABC}$ be a pure quantum state shared between Referee (R), Alice (AC) and Bob (B). For every $\epsilon_1, \epsilon_2 \in (0,1)$ satisfying $\epsilon_1 + 9\epsilon_2 \leq 1$, there exists an entanglement-assisted one-way protocol operated by Alice and Bob which starts in the state $|\psi\rangle^{RABC}$, and outputs a state $\phi^{RABC} \in \mathsf{B}^{\epsilon_1+9\epsilon_2}(\psi^{RABC})$ where registers A, BC and R are held by Alice, Bob and Referee, respectively. The communication cost of this protocol is upper bounded by*

$$\frac{1}{2} \inf_{\psi' \in \mathsf{B}^{\epsilon_1}(\psi^{RBC})} \inf_{\sigma \in \mathsf{ME}^{\epsilon_2^4/4,\psi'}_{R-B-C}} \left[\mathrm{D}_{\max}\left(\psi'^{RBC} \| \, \sigma^{RBC}\right) - \mathrm{D}_{\mathrm{H}}^{\epsilon_2^2}\left(\psi'^{BC} \| \, \sigma^{BC}\right) \right] + \log \frac{1}{\epsilon_2^2} + 1 \ . \quad (3.26)$$

Proof: Fix $\psi'^{RBC} \in \mathsf{B}^{\epsilon_1}(\psi^{RBC})$ and $\sigma^{RBC} \in \mathsf{ME}^{\epsilon_2^4/4,\psi'}_{R-B-C}$. As explained in Section 3.1.2, there exists a decomposition of register B as $\mathcal{H}^B = \bigoplus_j \mathcal{H}^{B_j^R} \otimes \mathcal{H}^{B_j^C}$ such that

$$\psi'^{RB} \quad = \quad \sigma^{RB} \quad = \quad \bigoplus_j p(j) \, \psi'^{RB_j^R}_j \otimes \psi'^{B_j^C}_j \ , \quad (3.27)$$

and

$$\sigma^{RBC} \quad = \quad \bigoplus_j p(j) \, \sigma_j^{RB_j^R} \otimes \sigma_j^{B_j^C C} \ , \quad (3.28)$$

where $\sigma_j^{RB_j^R} = \psi'^{RB_j^R}_j$, $\sigma_j^{B_j^C C} \in \mathsf{B}^{\epsilon_2^4/4}\left(\mathrm{Tr}_{B_j^R}\left((\Pi_j \otimes \mathbb{1})\psi'^{BC}(\Pi_j \otimes \mathbb{1})\right)\right)$ and Π_j is the projection operator over the j-th subspace of register B. If $\sigma^{RBC} = \psi'^{RB} \otimes \psi'^{C}$, Alice and Bob can redistribute ψ^{RABC} with error $9\epsilon_2 > 0$ and communication cost bounded by Eq. (3.26) using the AJW protocol as explained in Theorem 3.5. However, in general, σ^{RBC} is not necessarily a product state. In that case, our broad strategy is to let Alice and Bob transform ψ'^{RBC} through a local unitary procedure which maps σ^{RBC} to a product state. Then, they can use the AJW protocol to redistribute this new state. Before achieving this, Alice and Bob will perform some pre-processing on their shared state, as follows.

i. Viewing σ^{RBC} as a classical-quantum state: Let B^R and B^C be two registers with $\left|B^R\right| = \max_j \left|B_j^R\right|$ and $\left|B^C\right| = \max_j \left|B_j^C\right|$. By Eq. (3.28), there exists an isometry operator $U_{\mathrm{i}} : \mathcal{H}^B \to \mathcal{H}^{B^R J B^C}$ which takes σ^{RBC} to the state $\sigma_1^{RB^R J B^C C}$ defined as

$$\sigma_1^{RB^R J B^C C} \quad \coloneqq \quad \sum_j p(j) \, \sigma_j^{RB^R} \otimes |j\rangle\langle j|^J \otimes \sigma_j^{B^C C} \ . \quad (3.29)$$

Let $|\psi'\rangle^{RABC}$ be a purification of ψ'^{RBC} satisfying $\mathrm{P}(|\psi\rangle^{RABC}, |\psi'\rangle^{RABC}) \leq \epsilon_1$, as guaranteed by the Uhlmann theorem. Define

$$|\psi_1\rangle^{RAB^RJB^CC} := U_{\mathrm{i}}|\psi'\rangle^{RABC} = \sum_{j,j'} |j\rangle\langle j'|^J \otimes \psi_{j,j'}^{RAB^RB^CC} \quad .$$

ii. Transferring B^C from Bob to Alice without communication: Recall that register J is classical in $\psi_1^{RB^RJB^C}$ and conditioned on J, RB^R and B^C are decoupled, that is

$$\psi_1^{RB^RJB^C} \quad = \quad \sigma_1^{RB^RJB^C} \quad = \quad \sum_j p(j)\, \sigma_j^{RB^R} \otimes |j\rangle\langle j|^J \otimes \sigma_j^{B^C}.$$

This implies that $\mathrm{I}(RB^R : B^C \,|\, J)_{\psi_1} = 0$. So, Alice and Bob can use the folklore protocol for redistributing quantum Markov states explained in Fig. 3.2 and transfer B^C to Alice, as follows:

Since Alice holds registers AC, she can prepare the following purification of $\psi_1^{RB^RJB^C}$:

$$|\psi_1'\rangle^{RB^RJJ'B^CGH} = \sum_j \sqrt{p(j)}|\sigma_j\rangle^{RB^RG} \otimes |j,j\rangle^{JJ'} \otimes |\sigma_j\rangle^{B^CH} \quad ,$$

where registers $J'GH$ are held by Alice. Let $\delta_1 \in (0,1)$, $n_1 := \left|B^CH\right|^{2/\delta_1^2}$, and D_1, D_1' be registers with $|D_1| = |D_1'| = n_1$. Conditioned on register J, Alice and Bob use the embezzling state $|\xi\rangle^{D_1D_1'}$ (as defined in Eq. (3.12)) and the inverse of the van Dam-Hayden protocol [92] to embezzle out $|\sigma_j\rangle^{B^CH}$ in superposition and obtain a state $\widetilde{\psi}_1$ such that

$$\mathrm{P}\left(\widetilde{\psi}_1^{RB^RGJJ'D_1D_1'}, \sum_j \sqrt{p(j)}|\sigma_j\rangle^{RB^RG} \otimes |j,j\rangle^{JJ'} \otimes |\xi\rangle^{D_1D_1'}\right) \quad \leq \quad \delta_1 \quad .$$

Finally, conditioned on register J, Alice locally generates $|\sigma_j\rangle^{B^CH}$ in superposition with registers B^CH on her side, and applies an Uhlmann unitary operator to her registers in order to prepare the purification $|\psi_1\rangle^{RAB^RJB^CC}$. Let $U_{\mathrm{ii},A}$ and $U_{\mathrm{ii},B}$ denote the overall unitary operators applied by Alice and Bob, respectively, in this step. After applying $U_{\mathrm{ii},A}$ and $U_{\mathrm{ii},B}$, the global state is $|\psi_2\rangle$ satisfying

$$\mathrm{P}\left(\psi_2^{RAB^RJB^CCD_1D_1'}, |\psi_1\rangle\langle\psi_1|^{RAB^RJB^CC} \otimes |\xi\rangle\langle\xi|^{D_1D_1'}\right) \quad \leq \quad \delta_1 \quad ,$$

where registers AB^CC are with Alice, registers B^RJ are with Bob and register R is with Referee. Thus, the problem reduces to the case where the global state is $|\psi_1\rangle$ and the

register B^C belongs to Alice, up to a purified distance δ_1. We will now assume that this is indeed the case and later, we will account for the inaccuracy introduced by this assumption in the error analysis of our protocol, using the data processing inequality.

Suppose the global state is $|\psi_1\rangle^{RAB^RJB^CC}$ such that registers AB^CC, B^RJ and R are held by Alice, Bob and Referee, respectively, and Alice wants to transfer B^CC to Bob. To achieve this, we introduce a two-step unitary procedure which decouples registers RB^RJ and B^CC in $\sigma_1^{RB^RJB^CC}$ while keeping the state of registers RB^RJ unchanged. This operation transforms σ_1 to a product state and allows us to use the AJW protocol as a subroutine to achieve the redistribution.

To decouple RB^RJ from B^CC in σ_1, we use embezzlement and the unitary operator, given by Corollary 3.8. This unitary operator acts on registers JB^CC and is read-only on register J. However, since register J is not classical in $\psi_1^{RB^RJB^CC}$, it may disturb the marginal state $\psi_1^{RB^RJ}$. This issue can be resolved by first coherently measuring register J using an additional maximally entangled state. This operation transforms $\psi_1^{RB^RJB^CC}$ to a classical-quantum state, classical in register J, while keeps $\sigma_1^{RB^RJB^CC}$ intact. The following two steps contain the detailed construction of these two unitary procedures.

1. Coherent measurement of register J: Let F be a register with $|F| = |J|$, and $\{|f\rangle\}_{f=0}^{|F|-1}$ be a basis for $\mathcal{H}^F$. For $a, b \in \{0, \ldots, |F|-1\}$, let $P_{a,b} \in \mathsf{U}\big(\mathcal{H}^F\big)$ be the *Heisenberg-Weyl* operator defined as $P_{a,b} \coloneqq \sum_f \exp\Big(\frac{2\pi i f b}{|F|}\Big)|f+a\rangle\langle f|^F$. Let $U_1 \in \mathsf{U}(\mathcal{H}^{JF})$ be a unitary operator defined as $U_1 \coloneqq \sum_j |j\rangle\langle j|^J \otimes P_{j,0}^F$. Define

$$|\kappa_1\rangle^{RAB^RJB^CCFF'} \quad \coloneqq \quad U_1\left(|\psi_1\rangle^{RAB^RJB^CC} \otimes |\Phi\rangle^{FF'}\right) \ ,$$

and

$$\tau_1^{RB^RJB^CCF} \quad \coloneqq \quad U_1\left(\sigma_1^{RB^RJB^CC} \otimes \frac{\mathbb{1}^F}{|F|}\right)U_1^\dagger \ , \tag{3.30}$$

where $|\Phi\rangle^{FF'}$ is the maximally entangled state over registers F and F'. Notice that the set of Heisenberg-Weyl operators is closed under multiplication and each $P_{a,b}$ is traceless unless $a = b = 0$. Therefore, the unitary operator U_1 acts trivially on σ_1 while it measures register J in $\psi_1^{RB^RJB^CC}$ coherently. In particular,

$$\tau_1^{RB^RJB^CCF} \quad = \quad \sigma_1^{RB^RJB^CC} \otimes \frac{\mathbb{1}^F}{|F|} \ , \tag{3.31}$$

and

$$\kappa_1^{RB^RJB^CC} \quad = \quad \sum_j |j\rangle\langle j|^J \otimes \psi_{j,j}^{RB^RB^CC} \ . \tag{3.32}$$

2. Decoupling registers B^CC from RB^RJ in τ_1: By Eqs. (3.29) and (3.31), register J is classical in $\tau_1^{RB^RJB^CC}$ and conditioned on J, registers RB^R are decoupled from B^CC. Hence, we can decouple registers B^CC from registers RB^RJ in τ_1 using embezzling states and applying the unitary operator given in Corollary 3.8.

For $\gamma_2 \in (0,1)$ chosen as in Corollary 3.8, let $a_2 = |B^CC|/\gamma_2$, $n_2 = a_2^{1/\delta_2^2}$, D_2, D_2' and E_2 be quantum registers with $|D_2| = |D_2'| \geq n_2$ and $|E_2| = a_2$. By Corollary 3.8, there exists a unitary operator $U_2 \in \mathsf{U}(\mathcal{H}^{JB^CCE_2D_2})$, read-only on register J, and a projection operator $\widetilde{\Pi} \in \mathsf{Pos}(\mathcal{H}^{JB^CCE_2D_2})$ such that

$$\mathrm{D}_{\max}\left(U_2\left(\tau_1^{RB^RJB^CC} \otimes |0\rangle\langle 0|^{E_2} \otimes \xi_{a_2:n_2}^{D_2}\right)U_2^\dagger \middle\| \tau_1^{RB^RJ} \otimes \nu_2^{B^CCE_2} \otimes \xi_{1:n_2}^{D_2}\right) \leq \log(1+15\delta_2^2) \ , \quad (3.33)$$

$$\widetilde{\Pi}\left(\tau_1^{B^RJ} \otimes \nu_2^{B^CCE_2} \otimes \xi_{1:n_2}^{D_2}\right)\widetilde{\Pi} \leq 2 \cdot U_2\left(\tau_1^{RB^RJB^CC} \otimes |0\rangle\langle 0|^{E_2} \otimes \xi_{a_2:n_2}^{D_2}\right)U_2^\dagger \ , \quad (3.34)$$

and

$$\mathrm{Tr}\left[\widetilde{\Pi}U_2\left(\tau_1^{RB^RJB^CC} \otimes |0\rangle\langle 0|^{E_2} \otimes \xi_{a_2:n_2}^{D_2}\right)U_2^\dagger\right] = 1 \ , \quad (3.35)$$

where $\nu_2^{B^CCE_1} = \frac{1}{a_2}\sum_{r=1}^{a_2}|r\rangle\langle r|^{B^CCE_2}$. Define

$$\tau_2^{RB^RJB^CCE_2D_2} := U_2\left(\tau_1^{RB^RJB^CC} \otimes |0\rangle\langle 0|^{E_2} \otimes \xi_{a_2:n_2}^{D_2}\right)U_2^\dagger \ ,$$

and

$$|\kappa_2\rangle^{RAB^RJB^CCE_2D_2D_2'FF'} := U_2\left(|\kappa_1\rangle^{RAB^RJB^CC} \otimes |0\rangle^{E_2} \otimes |\xi_{a_2:n_2}\rangle^{D_2D_2'}\right) \ .$$

Note that by construction, U_2 is read-only on register J and since J is classical in the state $\kappa_1^{RB^RJB^CC}$, the unitary operator U_2 keeps $\kappa_1^{RB^RJ}$ intact. So, we have

$$\kappa_2^{RB^RJ} = \kappa_1^{RB^RJ} = \psi_1^{RB^RJ} \ . \quad (3.36)$$

Moreover, by Eq. (3.33), τ_2 is close to a product state in max-relative entropy and therefore, we can claim the following statement, proved towards the end.

Claim 3.10. *For the state κ_2 defined above, we have*

$$\mathrm{D}_{\max}\left(\kappa_2^{RB^RJB^CCE_2D_2F} \middle\| \kappa_2^{RB^RJ} \otimes \nu_2^{B^CCE_2} \otimes \xi_{1:n_2}^{D_2} \otimes \frac{\mathbb{1}^F}{|F|}\right) \leq \mathrm{D}_{\max}\left(\psi'^{RBC} \| \sigma^{RBC}\right) + 5\delta_2 \quad (3.37)$$

and

$$\mathrm{D}_{\mathrm{H}}^{\epsilon_2^2}\left(\kappa_2^{B^RJB^CCE_2D_2F}\middle\|\ \kappa_2^{B^RJ}\otimes\nu_2^{B^CCE_2}\otimes\xi_{1:n_2}^{D_2}\otimes\frac{\mathbb{1}^F}{|F|}\right)\quad\geq\quad\mathrm{D}_{\mathrm{H}}^{\epsilon_2^4/4}\left(\psi'^{BC}\|\ \sigma^{BC}\right)-1\ . \tag{3.38}$$

Using the claim, we proceed as follows. Let

$$\beta\quad:=\quad\mathrm{D}_{\max}\left(\psi'^{RBC}\|\ \sigma^{RBC}\right)+5\delta_2\ ,$$

and S and T be quantum registers such that $|S|=|T|=\left|B^CCE_2D_2\right|$. Let $|\eta\rangle^{ST}$ be a purification of $\nu_2^{B^CCE_2}\otimes\xi_{1:n_2}^{D_2}\otimes\frac{\mathbb{1}^F}{|F|}$ such that $\eta^T=\nu_2^{B^CCE_2}\otimes\xi_{1:n_2}^{D_2}\otimes\frac{\mathbb{1}^F}{|F|}$.

To redistribute registers B^CC in the state ψ_1 with the desired cost, Claim 3.10 suggests that it would be sufficient for parties to transform their joint state ψ_1 to κ_2 through the unitary operators U_2U_1, then use the AJW protocol to redistribute registers $B^CCE_2D_2F$, and finally, transform back κ_2 to the state ψ_1 by applying $U_1^{-1}U_2^{-1}$. However, in order to apply U_2U_1, one needs to have access to all the registers J,B^C, and C, but initially registers B^CC are with Alice and register J is with Bob. This problem can be resolved using the Uhlmann theorem. Recall that $\kappa_2^{RB^RJ}=\psi_1^{RB^RJ}$ as mentioned in Eq. (3.36). Therefore, there exists an Uhlmann isometry $V:\mathcal{H}^{AB^CC}\to\mathcal{H}^{ACB^CE_2D_2D_2'FF'}$ such that

$$V|\psi_1\rangle^{RAB^RJB^CC}\quad=\quad|\kappa_2\rangle^{RAB^RJB^CCE_2D_2D_2'FF'}\ . \tag{3.39}$$

Notice that V only acts on registers AB^CC which are initially with Alice and so Alice can apply the isometry V locally to transform ψ_1 to κ_2.

Now, we have all the ingredients for the protocol and we proceed to construct the protocol as follows:

The protocol. In order to redistribute $|\psi\rangle^{RABC}$, Alice and Bob run the following protocol.

1. Initially, Alice and Bob start in the state $|\psi\rangle^{RABC}$ and share quantum states $|\xi'\rangle^{X'X}$, $|\xi_{a_2:n_2}\rangle^{D_2'D_2}$ and $m:=\lceil\frac{2^\beta}{\epsilon_2^2}\rceil$ copies of the state $|\eta\rangle^{ST}$. Hence, the initial joint quantum state of Referee, Alice and Bob is

$$|\psi\rangle^{RABC}\otimes|\xi\rangle^{D_1'D_1}\otimes|\xi_{a_2:n_2}\rangle^{D_2'D_2}\bigotimes_{i=1}^{m}|\eta\rangle^{S_iT_i}\ ,$$

   such that register R is held by Referee, registers $(ACS_1\dots S_mD_1'D_2')$ are held by Alice and registers $(BT_1\dots T_mD_1D_2)$ are held by Bob.

2. Bob prepares ancilla qubits $|0\rangle^{E_1}$ and applies the isometry $U_{\mathrm{ii},B}U_{\mathrm{i}}$ on his registers, and Alice applies the isometry $VU_{\mathrm{ii},A}$ on her registers. Hence, their joint state transforms into a quantum state, say ω, which has purified distance at most δ_1 from $|\kappa_2\rangle^{RAB^RJB^CCE_2D_2D_2'FF'}$ such that registers $(AB^CCD_1'E_2D_2D_2'FF')$ are with Alice, registers $(B^RJB^CE_1D_1)$ are with Bob and register (R) is with Referee.

3. Running the protocol given by Theorem 3.5, parties redistribute their registers assuming their joint state is $|\kappa_2\rangle^{RAB^RJB^CCE_2D_2D_2'FF'}$ and using the shared entangled state $\bigotimes_{i=1}^{m}|\eta\rangle^{S_iT_i}$, and end up in the state $\hat{\omega}^{RAB^RJB^CCE_2D_2D_2'FF'}$ such that register (R) is held with Referee, $(AD_2'F')$ are held with Alice and $(B^RJB^CCE_2D_2F)$ are held with Bob.

4. Bob applies the operator $(U_2U_1U_{\mathrm{i}})^{-1}$ on registers $B^RJB^CCE_2D_2F$.

5. The final state is obtained in registers $RABC$.

According to Theorem 3.5, the communication cost of this protocol is

$$\frac{1}{2}\left[\mathrm{D}_{\max}\left(\kappa_2^{RB^RJB^CCE_2D_2F}\middle\|\ \kappa_2^{RB^RJ}\otimes\nu_2^{B^CCE_2}\otimes\xi_{1:n_2}^{D_2}\otimes\frac{\mathbb{1}^F}{|F|}\right)\right.$$
$$\left.-\ \mathrm{D}_{\mathrm{H}}^{\epsilon_2^2}\left(\kappa_2^{B^RJB^CCE_2D_2F}\middle\|\ \kappa_2^{B^RJ}\otimes\nu_2^{B^CCE_2}\otimes\xi_{1:n_2}^{D_2}\otimes\frac{\mathbb{1}^F}{|F|}\right)\right]+\log\frac{1}{\epsilon_2^2}$$

which is at most

$$\frac{1}{2}\left[\mathrm{D}_{\max}\left(\psi'^{RBC}\|\ \sigma^{RBC}\right)-\mathrm{D}_{\mathrm{H}}^{\epsilon_2^4/4}\left(\psi'^{BC}\|\ \sigma^{BC}\right)\right]+2.5\,\delta_2+\log\frac{1}{\epsilon_2^2}+1\ ,$$

by Claim 3.10.

Correctness of the protocol. Let ϕ be the final joint state of parties in the above protocol. We have

$$\begin{aligned}
&\mathrm{P}\left(\phi^{RABC}, \psi^{RABC}\right) \\
&\leq \quad \mathrm{P}\left(\phi^{RABC}, \psi'^{RABC}\right) + \mathrm{P}\left(\psi'^{RABC}, \psi^{RABC}\right) \\
&\leq \quad \mathrm{P}\left(\phi^{RABCE_2D_2D_2'FF'}, \psi'^{RABC} \otimes |0\rangle\langle 0|^{E_2} \otimes \xi_{a_2:n_2}^{D_2D_2'} \otimes |\Phi\rangle\langle\Phi|^{FF'}\right) + \epsilon_1 \\
&\leq \quad \mathrm{P}\left(\hat{\omega}^{RAB^RJB^CCE_2D_2D_2'FF'}, \kappa_2^{RAB^RJB^CCE_2D_2D_2'FF'}\right) + \epsilon_1 \\
&\leq \quad \mathrm{P}\left(\hat{\omega}^{RAB^RJB^CCE_2D_2D_2'FF'}, \omega^{RAB^RJB^CCE_2D_2D_2'FF'}\right) \\
&\qquad + \mathrm{P}\left(\omega^{RAB^RJB^CCE_2D_2D_2'FF'}, \kappa_2^{RAB^RJB^CCE_2D_2D_2'FF'}\right) + \epsilon_1 \\
&\leq \quad \epsilon_1 + 9\epsilon_2 + \delta_1 \quad .
\end{aligned}$$

where the second and third inequalities follows from monotonicity of purified distance under quantum operations and the last inequality holds since $\hat{\omega} \in \mathsf{B}^{9\epsilon_2}(\omega)$ by Theorem 3.5, and $\omega \in \mathsf{B}^{\delta_1}(\kappa_2)$.

By construction of embezzling states, we can choose δ_1 and δ_2 arbitrarily small by allowing arbitrary large shared entanglement between Alice and Bob. Hence, the statement of the theorem follows. ∎

It only remains to prove Claim 3.10.

Proof of Claim 3.10: Consider the states and operators defined in the proof of Theorem 3.9. Since register J is classical in both $\kappa_1^{RB^RJB^CC}$ and $\tau_1^{RB^RJB^CC}$ and U_2 is read-only on J, we have that $\kappa_2^{RB^RJ} = \tau_2^{RB^RJ} = \tau_1^{RB^RJ}$, and $\kappa_2^{JB^CCE_2D_2} \in \mathsf{B}^{\epsilon_2^4/4}\left(\tau_2^{JB^CCE_2D_2}\right)$. Therefore, we get

$$\begin{aligned}
&\mathrm{D}_{\max}\left(\kappa_2^{RB^RJB^CCE_2D_2F} \middle\| \kappa_2^{RB^RJ} \otimes \nu_2^{B^CCE_2} \otimes \xi_{1:n_2}^{D_2} \otimes \frac{\mathbb{1}^F}{|F|}\right) \\
&\leq \quad \mathrm{D}_{\max}\left(\kappa_2^{RB^RJB^CCE_2D_2F} \middle\| \tau_2^{RB^RJB^CCE_2D_2} \otimes \frac{\mathbb{1}^F}{|F|}\right) \\
&\qquad + \mathrm{D}_{\max}\left(\tau_2^{RB^RJB^CCE_2D_2} \middle\| \tau_2^{RB^RJ} \otimes \nu_2^{B^CCE_2} \otimes \xi_{1:n_2}^{D_2}\right) \\
&\leq \quad \mathrm{D}_{\max}\left(\psi'^{RBC} \| \sigma^{RBC}\right) + \log(1 + 15\delta_2^2) \quad ,
\end{aligned}$$

where the last inequality is a consequence of Eq. (3.33) and the fact that $\kappa_2^{RB^RJB^CCE_2D_2F}$ and $\tau_2^{RB^RJB^CCE_2D_2F}$ are unitary transformations of ψ'^{RBC} and σ^{RBC}, respectively. The above equation implies Eq. (3.37) since $\log_2(1 + 15x^2) \leq 5x$ for all $x \geq 0$.

In the rest of the proof, we show that

$$\begin{aligned} &\mathrm{D}_{\mathrm{H}}^{\epsilon_2^2}\left(\kappa_2^{B^R J B^C C E_2 D_2 F} \middle\| \kappa_2^{B^R J} \otimes \nu_2^{B^C C E_2} \otimes \xi_{1:n_2}^{D_2} \otimes \frac{\mathbb{1}^F}{|F|}\right) \\ &\geq \quad \mathrm{D}_{\mathrm{H}}^{\epsilon_2^4/4}\left(\kappa_2^{B^R J B^C C E_2 D_2 F} \middle\| \tau_2^{B^R J B^C C E_2 D_2} \otimes \frac{\mathbb{1}^F}{|F|}\right) - 1 \ . \end{aligned} \tag{3.40}$$

Then, Eq. (3.38) follows since $\kappa_2^{RB^R J B^C C E_2 D_2 F}$ and $\tau_2^{RB^R J B^C C E_2 D_2 F}$ are unitary transformations of ${\psi'}^{RBC}$ and σ^{RBC}, respectively. Let $\lambda := \mathrm{D}_{\mathrm{H}}^{\epsilon_2^4/4}\left(\kappa_2^{B^R J B^C C E_2 D_2 F} \middle\| \tau_2^{B^R J B^C C E_2 D_2 F}\right)$, and Π' be the POVM operator achieving λ, i.e.,

$$\mathrm{Tr}\left[\Pi' \kappa_2^{B^R J B^C C E_2 D_2 F}\right] \quad \geq \quad 1 - \frac{\epsilon_2^4}{4} \tag{3.41}$$

and

$$\mathrm{Tr}\left[\Pi' \left(\tau_2^{B^R J B^C C E_2 D_2} \otimes \frac{\mathbb{1}^F}{|F|}\right)\right] \quad = \quad 2^{-\lambda} \ . \tag{3.42}$$

Recall that $\kappa_2^{B^R J} = \tau_2^{B^R J} = \tau_1^{B^R J}$. So, Eq. (3.34) implies that

$$\widetilde{\Pi}\left(\kappa_2^{B^R J} \otimes \nu_2^{B^C C E_2} \otimes \xi_{1:n_2}^{D_2}\right)\widetilde{\Pi} \quad \leq \quad 2 \cdot \tau_2^{B^R J B^C C E_2 D_2} \ . \tag{3.43}$$

Since $\sigma^{RBC} \in \mathsf{ME}^{\epsilon,\psi'}_{R-B-C}$, the state $\kappa_2^{JB^C C E_2 D_2}$ is $(\epsilon_2^4/4)$-close to $\tau_2^{JB^C C E_2 D_2}$ in purified distance. This implies that

$$\mathrm{Tr}\left[\widetilde{\Pi}\kappa_2^{B^R J B^C C E_2 D_2 F}\right] \quad \geq \quad \mathrm{Tr}\left[\widetilde{\Pi}\tau_2^{B^R J B^C C E_2 D_2 F}\right] - \frac{\epsilon_2^4}{4} \quad = \quad 1 - \frac{\epsilon_2^4}{4} \ , \tag{3.44}$$

using Theorem 1.4, Theorem 1.6 and Eq. (3.35). So, the Gentle Measurement lemma, Lemma 1.5, implies that

$$\left\| \frac{\widetilde{\Pi}\kappa_2^{B^R J B^C C E_2 D_2 F}\widetilde{\Pi}}{\mathrm{Tr}\left[\widetilde{\Pi}\kappa_2^{B^R J B^C C E_2 D_2 F}\right]} - \kappa_2^{B^R J B^C C E_2 D_2 F} \right\|_{\mathrm{tr}} \quad \leq \quad \epsilon_2^2 \ .$$

Define the POVM operator $\Pi := \widetilde{\Pi}\Pi'\widetilde{\Pi}$. By Eq. (3.10) and Eq. (3.44), we have

$$\begin{aligned} \mathrm{Tr}\left[\Pi\kappa_2^{B^R J B^C C E_2 D_2 F}\right] \quad &= \quad \mathrm{Tr}\left[\Pi'\widetilde{\Pi}\kappa_2^{B^R J B^C C E_2 D_2 F}\widetilde{\Pi}\right] \\ &\geq \quad \left(1 - \frac{\epsilon_2^4}{4}\right)\left(\mathrm{Tr}\left[\Pi'\kappa_2^{B^R J B^C C E_2 D_2 F}\right] - \frac{\epsilon_2^2}{2}\right) \quad \text{(By Theorem 1.4)} \\ &\geq \quad 1 - \epsilon_2^2 \ , \end{aligned}$$

and by Eq. (3.43), we get

$$\begin{aligned}\mathrm{Tr}\left[\Pi\left(\kappa_2^{B^RJ}\otimes\nu_2^{B^CCE_2}\otimes\xi_{1:n_2}^{D_2}\otimes\frac{\mathbb{1}^F}{|F|}\right)\right] &\leq 2\cdot\mathrm{Tr}\left[\Pi'\left(\tau_2^{B^RJB^CCE_2D_2}\otimes\frac{\mathbb{1}^F}{|F|}\right)\right]\\ &= 2^{-\lambda+1},\end{aligned}$$

which imply Eq. (3.40), as desired. ∎

3.4 Asymptotic and i.i.d. analysis

Suppose that the state $|\psi\rangle^{R^nA^nB^nC^n}=(|\psi\rangle^{RABC})^{\otimes n}$ is shared between Referee (R^n), Alice (A^nC^n) and Bob (B^n) where R^n, A^n, B^n and C^n denote n-fold tensor product of registers R, A, B and C, respectively. Let $\epsilon:=\epsilon_1=\epsilon_2^4/4$. By Theorem 3.9 and choosing $\sigma^{R^nB^nC^n}=\psi'^{R^nB^n}\otimes\psi^{C^n}$, there exists an entanglement-assisted one-way protocol which outputs a state $\phi^{R^nA^nB^nC^n}\in\mathsf{B}^{14\epsilon^{1/4}}(\psi^{R^nA^nB^nC^n})$ with communication cost $Q(n,\epsilon)$ at most

$$\begin{aligned}&\frac{1}{2}\inf_{\psi'\in\mathsf{B}^\epsilon(\psi^{R^nB^nC^n})}\left[\mathrm{D}_{\max}\left(\psi'^{R^nB^nC^n}\|\,\psi'^{R^nB^n}\otimes\psi^{C^n}\right)-\mathrm{D}_{\mathrm{H}}^{\epsilon}\left(\psi'^{B^nC^n}\|\,\psi'^{B^n}\otimes\psi^{C^n}\right)\right]+\log\frac{1}{2\sqrt{\epsilon}}\\ &\quad\leq\frac{1}{2}\inf_{\substack{\psi'\in\mathsf{B}^\epsilon(\psi^{R^nB^nC^n})\\ \psi'^{R^nB^n}=\psi^{R^nB^n}}}\left[\mathrm{D}_{\max}\left(\psi'^{R^nB^nC^n}\|\,\psi^{R^nB^n}\otimes\psi^{C^n}\right)-\mathrm{D}_{\mathrm{H}}^{\epsilon}\left(\psi'^{B^nC^n}\|\,\psi^{B^n}\otimes\psi^{C^n}\right)\right]\\ &\qquad+\log\frac{1}{2\sqrt{\epsilon}}\\ &\quad\leq\frac{1}{2}\inf_{\substack{\psi'\in\mathsf{B}^\epsilon(\psi^{R^nB^nC^n})\\ \psi'^{R^nB^n}=\psi^{R^nB^n}}}\left[\mathrm{D}_{\max}\left(\psi'^{R^nB^nC^n}\|\,\psi^{R^nB^n}\otimes\psi^{C^n}\right)-\mathrm{D}_{\mathrm{H}}^{2\epsilon}\left(\psi^{B^nC^n}\|\,\psi^{B^n}\otimes\psi^{C^n}\right)\right]\\ &\qquad+\log\frac{1}{2\sqrt{\epsilon}}\\ &\quad\leq\frac{1}{2}\left[\mathrm{D}_{\max}^{\epsilon/3}\left(\psi^{R^nB^nC^n}\|\,\psi^{R^nB^n}\otimes\psi^{C^n}\right)-\mathrm{D}_{\mathrm{H}}^{2\epsilon}\left(\psi^{B^nC^n}\|\,\psi^{B^n}\otimes\psi^{C^n}\right)\right]+\log\frac{1}{2\sqrt{\epsilon}}\\ &\qquad+\log\frac{72+\epsilon^2}{\epsilon^2},\end{aligned}$$

where the first inequality follows from Eq. (3.26), the third inequality follows from the definition of Hypothesis testing entropy, and the last inequality follows from Theorem 1.9 for the choice of $\epsilon,\delta\leftarrow\epsilon/3$, $\rho^{AB}\leftarrow\psi^{R^nB^nC^n}$, $\rho^A\leftarrow\psi^{R^nB^n}$ and $\sigma^B\leftarrow\psi^{C^n}$. Therefore,

using Theorem 1.8, the asymptotic communication rate of redistributing n copies of a pure state $|\psi\rangle^{RABC}$ is

$$\lim_{n\to\infty}\frac{1}{n}Q(n,\epsilon) \quad \leq \quad \frac{1}{2}\,\mathrm{I}(R:C\,|\,B)_{\psi}\ . \tag{3.45}$$

3.5 Conclusion and outlook

In this chapter, we revisited the task of one-shot quantum state redistribution, and introduced a new protocol achieving this task with communication cost

$$\frac{1}{2}\min_{\psi'\in\mathsf{B}^{\epsilon}(\psi^{RBC})}\ \min_{\sigma^{RBC}\in\mathsf{ME}^{\epsilon^2/4,\psi'}_{R-B-C}}\left[\mathrm{D}_{\max}\left(\psi'^{RBC}\|\sigma^{RBC}\right)-\mathrm{D}^{\epsilon}_{\mathrm{H}}\left(\psi'^{BC}\|\sigma^{BC}\right)\right]+\log\frac{1}{\epsilon}+1\ , \tag{3.46}$$

with error parameter ϵ. Our result is the first to operationally connect one-shot quantum state redistribution and quantum Markov chains and provides an operational interpretation for a one-shot representation of quantum conditional mutual information as explained in Sec 3.1. In the special case where ψ^{RBC} is a quantum Markov chain, our protocol leads to near-zero communication which was not known for the previous protocols. Moreover, the communication cost of our protocol is lower than all previously known one-shot protocols and we show that it achieves the optimal cost of $\mathrm{I}(R:C\,|\,B)$ in the asymptotic i.i.d. setting. Our protocol also achieves the near-optimal result of Ref. [6] for the case that ψ^{RBC} is classical.

A question of interest is whether the communication cost of our one-shot protocol can be bounded with $\mathrm{I}(R:C\,|\,B)$. In the quantum communication complexity setting, such a bound would imply the possibility of compressing the communication of bounded-round quantum protocols to their information content which also results in a direct-sum theorem for bounded-round quantum communication complexity [89].

Another question that we have not addressed in this thesis is whether our bound is optimal. In other words, it is interesting to find out if it is possible to show that the same quantity is also a lower bound for the communication cost of quantum state redistribution. There have been several lower bounds in the literature for the communication cost of entanglement-assisted quantum state redistribution including the ones in Ref. [27] and Ref. [64]. However, it is not clear if our bound achieves any of them.

Part II

Quantum Learning Theory

Chapter 4

Optimal quantum bounds for learning Boolean functions

4.1 Introduction

Learning Boolean functions has been studied widely in the theoretical machine learning. In this framework, a *concept class* $\mathcal{C}$ is a subset of the set of all Boolean functions $f : \{0,1\}^n \to \{0,1\}$. A rigorous definition of learning a concept class was firstly developed by Leslie Valiant [91] who introduced the notion of *probably approximately correct* (PAC) learning. An (ϵ, δ)-PAC learner is a learning algorithm which for every *target* concept $c \in \mathcal{C}$ and distribution D over $\{0,1\}^n$, given some i.i.d. random *labelled examples*, with probability at least $1-\delta$, outputs an ϵ-approximation h of $c \in \mathcal{C}$ such that $\Pr_{x\sim D}[h(x) \neq c(x)] \leq \epsilon$. For a concept $c \in \mathcal{C}$ and distribution D, a labelled example is a random pair $(x, c(x))$ where x is distributed according to distribution D. Later, Bshouty and Jackson [31] extended the notion of PAC learning to the quantum setting. For a target concept $c \in \mathcal{C}$ and distribution D, they defined a *quantum example* as a superposition of labelled examples $(x, c(x))$ weighted according to the distribution D. Then, a *PAC quantum learner* is a quantum learning algorithm which with high probability, outputs an approximation of the concept c given some quantum examples.

In reality, the labelled examples may be noisy or there may not be necessarily an underlying target concept at all. A more realistic model is the *agnostic* learning which was introduced by Haussler [49], and Kearns, Schapire and Sellie [63]. In the agnostic model, examples are random pairs $(x, l) \in \{0,1\}^n \times \{0,1\}$ distributed according to a distribution D over $\{0,1\}^{n+1}$, and the goal is to find a concept in a specific concept class $\mathcal{C}$ with minimum

error. In this framework, the error of a concept $c \in \mathcal{C}$ with respect to D is defined as $\text{err}_D = \Pr_{(x,l)\sim D}[c(x) \neq l]$. An (ϵ, δ)-*agnostic learner* is a learning algorithm which for every distribution D, with probability at least $1 - \delta$, outputs a concept $c \in \mathcal{C}$ with error at most an additive ϵ worse than the lowest possible error. Similar to the PAC model, the agnostic model can be extended to the quantum setting where a quantum example for a distribution D is defined as the superposition of all pairs $(x, l) \in \{0,1\}^{n+1}$ weighted according to the distribution D.

In any of the above models, the goal is to find "efficient" learners. Here, we are interested in the measure of *sample complexity* which is the minimum number of examples required to learn any concept in a concept class $\mathcal{C}$. Besides ϵ and δ, the sample complexity of $\mathcal{C}$ depends on a combinatorial parameter called the *VC dimension* of $\mathcal{C}$. The VC dimension of a concept class $\mathcal{C}$ is the size of the largest subset $S \subseteq \{0,1\}^n$ which can be labelled with all $2^{|S|}$ possible bit strings of length $|S|$ by concepts from $\mathcal{C}$. The notions of sample complexity and VC dimension are defined formally in Section 4.2.

In a series of works [29, 47], it has been shown that the (ϵ, δ)-*PAC (classical) sample complexity* of a concept class with VC dimension d is

$$\Theta\left(\frac{d}{\epsilon} + \frac{\log(1/\delta)}{\epsilon}\right) \quad . \tag{4.1}$$

For the agnostic model, the (ϵ, δ)-*agnostic (classical) sample complexity* is

$$\Theta\left(\frac{d}{\epsilon^2} + \frac{\log(1/\delta)}{\epsilon^2}\right) \quad . \tag{4.2}$$

The lower bound was shown by Vapnik and Chervonenkis [94] and Talagrand [85] showed that this bound is indeed achievable.

Although quantum learning of a concept class can be more efficient than classical learning in certain scenarios (see e.g. Ref. [31] and Ref. [82]), PAC quantum sample complexity and agnostic quantum sample complexity cannot be better than their classical counterparts by more than a constant factor. In particular, Eq. (4.1) and Eq. (4.2) also hold for (ϵ, δ)-*PAC quantum sample complexity* and (ϵ, δ)-*agnostic quantum sample complexity*, respectively. The upper bounds are carried over directly from the classical bounds since a (classical) labelled example can be derived by measuring a quantum example. The lower bounds are due to Arunachalam and de Wolf [12] who used two different proof approaches. First, they used a simple information-theoretic approach to reprove the classical lower bounds in Eq. (4.1) and Eq. (4.2), and similar bounds for quantum sample complexities, but smaller by a factor of $\log(d/\epsilon)$. They claimed that the $\log(d/\epsilon)$ is inherent in

the information-theoretic proof, and to remove the $\log(d/\epsilon)$ factor, they presented a much more complicated proof via analysis of quantum state identification and Fourier analysis.

In this chapter, we show that, as opposed to the claim in Ref. [12], the information-theoretic approach can be used to derive the optimal lower bound for both PAC quantum sample complexity and agnostic quantum sample complexity. Our proof is much simpler than the state identification based proof in Ref. [12] and the same technique potentially can be applied to derive optimal bounds in other related problems.

4.1.1 Proof techniques

Our proof is built up on the information theoretic proof in Ref. [12] which consisted of three steps. In the first step, they showed that, at the end of the learning process, the information obtained by a PAC (or agnostic) learner about the target concept (or minimal-error concept) is $\Omega(d)$. In the second step, they showed that this information is upper bounded by the entropy of the average of a copy of the quantum example for different possible concepts multiplied by the sample complexity t of the learner, using the *subadditivity* property of entropy. By bounding the entropy of this averaged state in the third step, they showed that the information is $\mathrm{O}(T\epsilon\log(d/\epsilon))$. These steps together implies the $\Omega(\frac{d}{\epsilon\log(d/\epsilon)})$ bound for the sample complexity t.

It turns out that the second step, i.e., the use of subadditivity, is the origin of the sub-optimality in their proof. In our proof, we combine steps two and three and use spectral analysis and concentration of measure bounds in order to derive a tighter upper bound on the information obtained by the learner. We also improve the constants in the $\Omega(d)$ lower bound by using the well-known Fano's inequality, and therefore, we succeed to obtain the optimal lower bound for both PAC and agnostic quantum learning.

4.2 Quantum learning theory

We refer the readers to the book [83] for an introduction to machine learning theory and the survey [11] for an introduction to quantum learning theory. Here, we briefly review the notation related to the PAC model and agnostic model that we require in this chapter.

In this work, we study learning Boolean functions over n-dimensional cube. We refer to a Boolean function $c : \{0,1\}^n \to \{0,1\}$ as a *concept*. We can also think of a concept as a bit-string in $\{0,1\}^N$ for $N \coloneqq 2^n$ which contains the value of c over all possible n-bit

strings. A *concept class* is a subset $\mathcal{C} \subseteq \{0,1\}^N$ of Boolean functions. For a concept c, we refer to $c(x)$ as the *label* of $x \in \{0,1\}^n$, and the tuple $(x, c(x))$ as a *labelled example*.

A crucial combinatorial quantity in learning Boolean functions is the *VC dimension* of a concept class which was introduced by Vapnik and Chervonenkis [93]. We say a set $S = \{s_1, \ldots, s_d\} \subseteq \{0,1\}^n$ is *shattered* by a concept class $\mathcal{C}$ if for every $a \in \{0,1\}^d$, there exists a concept $c \in \mathcal{C}$ such that $(c(s_1), \ldots, c(s_d)) = a$. The *VC dimension* of $\mathcal{C}$ is the size of the largest set shattered by $\mathcal{C}$, and we denote it as VC-dim($\mathcal{C}$).

PAC model

Consider a concept class $\mathcal{C} \subseteq \{0,1\}^N$. The PAC (*probably approximately correct*) model for learning concepts was introduced in the classical setting by Valiant [91] and was extended to the quantum setting by Bshouty and Jackson [31]. In the quantum PAC model, a learning algorithm is given a *quantum PAC example oracle* QPEX(c, D) for an unknown concept $c \in \mathcal{C}$ and an unknown distribution D over $\{0,1\}^n$. The oracle QPEX(c, D) has no inputs; when invoked, it outputs a superposition of labelled examples of c distributed according to distribution D, namely,

$$\sum_{x \in \{0,1\}^n} \sqrt{D(x)}\, |x, c(x)\rangle \ .$$

This is one possible generalization of classical random examples to the quantum setting. Although such quantum examples may be difficult to generate and store, many quantum learning applications involve data that are naturally provided as coherent quantum states.

We say a Boolean function h, commonly called a *hypothesis*, is an ϵ-*approximation* of c with respect to distribution D, if

$$\Pr_{x \sim D} [h(x) \neq c(x)] \quad \leq \quad \epsilon \ . \tag{4.3}$$

Given access to QPEX(c, D) oracle, the goal of a quantum PAC learner is to find an ϵ-approximation hypothesis h of c with high success probability.

Definition 4.1. *For $\epsilon, \delta \in [0,1]$, we say an algorithm $\mathcal{A}$ is an (ϵ, δ)-PAC quantum learner for concept class $\mathcal{C}$ if for every $c \in \mathcal{C}$ and distribution D, given access to* QPEX(c, D)*, with probability at least $1 - \delta$, $\mathcal{A}$ outputs a hypothesis $h \in \{0,1\}^N$ which is an ϵ-approximation of c*

The *sample complexity* of a quantum learner $\mathcal{A}$ is the maximum number of times $\mathcal{A}$ invokes the oracle $\mathrm{QPEX}(c, D)$ for any concept $c \in \mathcal{C}$ and any distribution D over $\{0,1\}^n$. The (ϵ, δ)-*PAC quantum sample complexity* of a concept class $\mathcal{C}$ is the minimum sample complexity of a (ϵ, δ)-PAC quantum learner for $\mathcal{C}$.

Agnostic model

In the PAC model, it is assumed that examples are generated perfectly according to some unknown concept $c \in \mathcal{C}$. The *agnostic* model is a more realistic model where examples correspond to random labelled pairs (x, l) distributed according to an unknown distribution D over $\{0,1\}^{n+1}$, not necessarily derived from a specific concept $c \in \mathcal{C}$. This model was introduced in the classical setting by Haussler [49], and Kearn, Schapire, and Sellie [63] and was first studied in the quantum setting by Arunachalam and de Wolf [12]. In the agnostic model, the learning algorithm has access to an agnostic quantum oracle $\mathrm{QAEX}(D)$ for an unknown distribution D over $\{0,1\}^{n+1}$. When invoked, the oracle $\mathrm{QAEX}(D)$ outputs the superposition

$$\sum_{(x,l)\in\{0,1\}^{n+1}} \sqrt{D(x,l)}\,|x,l\rangle \ .$$

The *error* of a hypothesis $h \in \{0,1\}^N$ under distribution D is defined as

$$\mathrm{err}_D(h) \quad \coloneqq \quad \Pr_{(x,l)\sim D}[h(x) \neq l] \ . \tag{4.4}$$

For a concept class $\mathcal{C}$, the minimal error achievable is defined as

$$\mathrm{opt}_D(\mathcal{C}) \quad \coloneqq \quad \min_{c\in\mathcal{C}} \quad \mathrm{err}_D(c)$$

Given access to a $\mathrm{QAEX}(D)$ oracle, the goal of quantum agnostic learner is to find a hypothesis $h \in \mathcal{C}$ with error not "much larger" than $\mathrm{opt}_D(\mathcal{C})$.

Definition 4.2. *For $\epsilon, \delta \in [0,1]$, we say a learning algorithm $\mathcal{A}$ is an (ϵ, δ)-agnostic quantum learner for $\mathcal{C}$ if for every distribution D, given access to $\mathrm{QAEX}(D)$, with probability at least $1-\delta$, $\mathcal{A}$ outputs a hypothesis $h \in \mathcal{C}$ such that $\mathrm{err}_D(h) \leq \mathrm{opt}_D(\mathcal{C}) + \epsilon$.*

Similar to the PAC model, we define the *sample complexity* of a quantum learner $\mathcal{A}$ as the maximum number of times $\mathcal{A}$ invokes the oracle $\mathrm{QAEX}(D)$ for any distribution D over $\{0,1\}^{n+1}$. The (ϵ, δ)-*agnostic quantum sample complexity* of a concept class $\mathcal{C}$ is the minimum sample complexity of an (ϵ, δ)-agnostic quantum learner for $\mathcal{C}$.

4.3 A lower bound on sample complexity of PAC learning

Let $\mathcal{C}$ be a concept class with VC-dim$(\mathcal{C}) = d$ and D be a distribution over n-bit strings. In this section, we use information-theoretic arguments to show that, for $\epsilon \in (0, 1/4)$ and $\delta \in (0, 1/2)$, the (ϵ, δ)-PAC quantum sample complexity of $\mathcal{C}$ is

$$\Omega\left(\frac{d}{\epsilon} + \frac{\log(1/\delta)}{\epsilon}\right) \ .$$

The above bound contains two parts, one depends on the VC dimension and the other one is independent of the VC dimension. The VC-independent part was shown by Atici and Servadio [14], and for completeness, we provide their proof in Lemma 4.1.

Lemma 4.1. *Let $\mathcal{C}$ be a non-trivial concept class. For every $\delta \in (0, 1/2)$ and $\epsilon \in (0, 1/4)$, an (ϵ, δ)-PAC quantum learner for $\mathcal{C}$ has quantum sample complexity at least $\Omega\big(\frac{1}{\epsilon}\log\frac{1}{\delta}\big)$.*

Proof: Since $\mathcal{C}$ is non-trivial, there exists concepts $c_1, c_2 \in \mathcal{C}$ and inputs $x_1, x_2 \in \{0,1\}^n$ such that $c_1(x_1) = c_2(x_1)$ and $c_1(x_2) \neq c_2(x_2)$. Consider the distribution D defined as follows:

$$D(x_1) = 1 - 2\epsilon \qquad \text{and} \qquad D(x_2) = 2\epsilon \ .$$

Let $|\psi_i\rangle := \sqrt{1-2\epsilon}|x_1, c_i(x_1)\rangle + \sqrt{2\epsilon}|x_2, c_i(x_2)\rangle$ for $i \in \{1, 2\}$. Under this distribution, no hypothesis can simultaneously ϵ-approximates c_1 and c_2. So, an (ϵ, δ)-PAC quantum learner for $\mathcal{C}$ can be used to distinguish $|\psi_1\rangle^{\otimes t}$ from $|\psi_2\rangle^{\otimes t}$ with success probability at least $1-\delta$, where t is the sample complexity of the learner. On the other hand, by Holevo-Helstrom theorem (see e.g. Ref. [96, Theorem 3.4]), $|\psi_1\rangle^{\otimes t}$ and $|\psi_2\rangle^{\otimes t}$ are distinguishable with probability at most $\frac{1}{2} + \frac{1}{2}\sqrt{1 - |\langle\psi_1|\psi_2\rangle^t|^2}$. So, we have $\langle\psi_1|\psi_2\rangle^t \leq 2\sqrt{\delta(1-\delta)}$. Since $\langle\psi_1|\psi_2\rangle = 1 - 2\epsilon$, it follows that $t = \Omega\big(\frac{1}{\epsilon}\log(1/\delta)\big)$. ∎

Now, we are ready to prove the VC-dependent part of the lower bound.

Theorem 4.2. *Let $n \geq 2$, $d \geq 2512$, $N = 2^n$ and $\mathcal{C} \subseteq \{0,1\}^N$ be a concept class with* VC-dim$(\mathcal{C}) = d + 1$. *For $\epsilon \in (0, 1/4)$ and $\delta \in (0, 1/2)$, every (ϵ, δ)-PAC quantum learner for $\mathcal{C}$ has sample complexity $\Omega(d/\epsilon)$.*

Proof: Suppose there is an (ϵ, δ)-PAC quantum learner $\mathcal{A}$ for $\mathcal{C}$ with quantum sample complexity t. Let $S = \{s_0, \ldots, s_d\} \subseteq \{0,1\}^n$ be a set of size VC-dim$(\mathcal{C}) = d + 1$ shattered

by the concept class $\mathcal{C}$. Let D be a probability distribution over the set $\{0,1\}^n$ such that $D(s_0) = 1-4\epsilon$, and $D(s_i) = 4\epsilon/d$ for $i \in [d]$. Since D is only supported over the set S, in the rest of the proof, we write i instead of s_i for the sake of brevity. Since $x = 0$ has the largest weight in distribution D, concepts with equal value at $x = 0$ are hard to be identified by a learner. So we consider a set of concepts defined as follows. For each $a \in \{0,1\}^d$, let $c_a \in \mathcal{C}$ be a concept such that $c_a(0) = 0$ and $c_a(i) = a_i$ for all $i \in [d]$. Such a concept exists in $\mathcal{C}$ since the set S is shattered by the concept class $\mathcal{C}$. The PAC quantum example for c_a according to D is written as $|\psi_a\rangle = \sqrt{1-4\epsilon}|0,0\rangle + \sqrt{4\epsilon/d}\sum_{i=1}^{d}|i,a_i\rangle$. Let ρ_a denote the state corresponding to t copies of the example $|\psi_a\rangle\langle\psi_a|$ and let $\rho^{AB_1\ldots B_t}$ be the following classical-quantum state

$$\rho^{AB_1\ldots B_t} \quad \coloneqq \quad \frac{1}{2^d}\sum_{a\in\{0,1\}^d}|a\rangle\langle a|^A \otimes \rho_a^{B_1\ldots B_t} \quad = \quad \frac{1}{2^d}\sum_{a\in\{0,1\}^d}|a\rangle\langle a| \otimes |\psi_a\rangle\langle\psi_a|^{\otimes t} \ . \tag{4.5}$$

Let $B \coloneqq B_1 \ldots B_t$. First, we show that $\mathrm{I}(A:B)_\rho \geq \Omega(d)$ using the hypothesis that $\mathcal{C}$, and therefore, its subset $\{c_a : a \in \{0,1\}^d\}$, is (ϵ,δ)-PAC quantum learnable using t quantum examples. Then, we show that $\mathrm{I}(A:B)_\rho \geq \Omega(d)$ implies that t is $\Omega(d/\epsilon)$.

Suppose the learner $\mathcal{A}$ is given t quantum examples in the state ρ_a^B. Let E be the a binary random variable corresponding to the event that $\mathcal{A}$ outputs a desirable hypothesis, i.e., a hypothesis $f \in \{0,1\}^N$ satisfying $\Pr_{i\sim D}[f(s_i) \neq c_a(s_i)] \leq \epsilon$. The event E occurs with probability at least $1-\delta$. Let F be the random variable corresponding to the output of $\mathcal{A}$. Using the data processing inequality (DPI) and the chain rule for mutual information, we have

$$\begin{aligned}
\mathrm{I}(A:B) &\geq \mathrm{H}(A) - \mathrm{H}(A\,|\,F) \\
&\geq \mathrm{H}(A) - \mathrm{H}(AE\,|\,F) \\
&= \mathrm{H}(A) - \mathrm{H}(A\,|\,FE) - \mathrm{H}(E\,|\,F) \qquad \text{(by chain rule for } \mathrm{H}(AE\,|\,F)) \\
&\geq \mathrm{H}(A) - \mathrm{H}(A\,|\,FE) - \mathrm{H}(E) \\
&= \mathrm{H}(A) - \Pr[E=0]\,\mathrm{H}(A\,|\,F,E=0) - \Pr[E=1]\,\mathrm{H}(A\,|\,F,E=1) - \mathrm{H}(E) \ ,
\end{aligned}$$

where the second inequality follows from the fact that Shannon entropy is monotonic under taking marginals, and the third inequality holds since conditioning on classical registers does not increase the entropy. Note that $\mathrm{H}(A\,|\,F,E=1) \leq \mathrm{H}(A) \leq d$ and $\Pr[E=0] \geq \delta$. Also, given $E = 1$, the probability that F is equal to A is at least $1-\epsilon$. So, the Fano inequality implies that

$$\mathrm{I}(A:B) \quad \geq \quad d - \delta d - \mathrm{H}(\epsilon) - \epsilon d - \mathrm{H}(\delta) \quad \geq \quad \frac{1}{4}d - 2 \ , \tag{4.6}$$

where the last inequality follows from the assumption that $\epsilon \leq 1/4$ and $\delta \leq 1/2$.

Now assume $t < d/\epsilon$; otherwise, we are done. Suppose $t = \nu d/\epsilon$ for some $\nu \in (0,1)$. (Note that $t = 0$ is not possible.) We show that there are positive universal constants α, d_0 such that if $\nu < \alpha$ and $d \geq d_0$, the mutual information $\mathrm{I}(A:B)$ violates Eq. (4.6).

Notice that ρ^{AB} and ρ^A have the same spectrum. Hence, $\mathrm{I}(A:B)_\rho = \mathrm{S}(B)_\rho$ with

$$\rho^B \quad = \quad \frac{1}{2^d} \sum_{a\in\{0,1\}^d} |\psi_a\rangle\langle\psi_a|^{\otimes t} \quad = \quad \frac{1}{2^d} \sum_{a\in\{0,1\}^d} \sum_{u,v\in\mathbb{N}_d^t} \sqrt{D(u)D(v)}\, |u\rangle\langle v|^I \otimes |a_u\rangle\langle a_v|^L \ , \tag{4.7}$$

where register B is partitioned into registers I and L, storing indices and labels, respectively. Let σ^B be the quantum state after applying Hadamard operator $H^{\otimes t}$ on register L, i.e.,

$$\begin{aligned} \sigma^B \quad &\coloneqq \quad \frac{1}{2^d} \sum_{a\in\{0,1\}^d} \sum_{u,v\in\mathbb{N}_d^t} \sqrt{D(u)D(v)}\, |u\rangle\langle v|^I \otimes H^{\otimes t} |a_u\rangle\langle a_v|^L H^{\otimes t} \\ &= \quad \frac{1}{2^t 2^d} \sum_{u,v\in\mathbb{N}_d^t} \sqrt{D(u)D(v)}\, |u\rangle\langle v|^I \otimes \sum_{a\in\{0,1\}^d} \sum_{x,y\in\{0,1\}^t} (-1)^{x\cdot a_u + y\cdot a_v} |x\rangle\langle y|^L \ . \end{aligned}$$

For fixed $x, y \in \{0,1\}^d$ and $u, v \in \mathbb{N}_d^t$, we have

$$\sum_{a\in\{0,1\}^d} (-1)^{x\cdot a_u + y\cdot a_v} \quad = \quad \begin{cases} 2^d & \text{if } \mathrm{ps}(u_x) = \mathrm{ps}(v_y) \ , \\ 0 & \text{otherwise.} \end{cases}$$

where $\mathrm{ps} : \mathbb{N}_d^t \to \{0,1\}^d$ is the parity signature function defined in Section 1.2.1. So, we get

$$\begin{aligned} \sigma^B \quad &= \quad \frac{1}{2^t} \sum_{x,y\in\{0,1\}^t} \sum_{\substack{u,v\in\mathbb{N}_d^t \\ \mathrm{ps}(u_x)=\mathrm{ps}(v_y)}} \sqrt{D(u)D(v)}\, |u\rangle\langle v|^I \otimes |x\rangle\langle y|^L \\ &= \quad \frac{1}{2^t} \sum_{b\in\{0,1\}^d} \left(\sum_{x\in\{0,1\}^t} \sum_{\substack{u\in\mathbb{N}_d^t \\ \mathrm{ps}(u_x)=b}} \sqrt{D(u)}\, |u\rangle^I |x\rangle^L \right) \left(\sum_{y\in\{0,1\}^t} \sum_{\substack{v\in\mathbb{N}_d^t \\ \mathrm{ps}(v_y)=b}} \sqrt{D(v)}\, \langle v|^I \langle y|^L \right) . \end{aligned}$$

For each $b \in \{0,1\}^d$, let

$$|\phi_b\rangle \quad \coloneqq \quad \sum_{x\in\{0,1\}^t} \sum_{\substack{u\in\mathbb{N}_d^t \\ \mathrm{ps}(u_x)=b}} \sqrt{D(u)}\, |u\rangle|x\rangle \ .$$

Note that for each $x \in \{0,1\}^t$ and $u \in \mathbb{N}_d^t$, the value of $\mathrm{ps}(u_x)$ is unique. So, $\langle \phi_b | \phi_{b'} \rangle = 0$ for distinct d-bit strings b and b'. Moreover, for each $b \in \{0,1\}^d$, $|\phi_b\rangle$ is an eigenvector of σ^B with corresponding eigenvalue

$$\begin{aligned}
\lambda_b &\coloneqq \frac{1}{2^t} \sum_{x \in \{0,1\}^t} \sum_{\substack{u \in \mathbb{N}_d^t \\ \mathrm{ps}(u_x)=b}} D(u) \\
&= \frac{1}{2^t} \sum_{x \in \{0,1\}^t} \sum_{\substack{u \in \mathbb{N}_d^{|x|} \\ \mathrm{ps}(u)=b}} D(u) \\
&= \sum_{r=0}^{t} \frac{\binom{t}{r}}{2^t} \sum_{\substack{u \in \mathbb{N}_d^r \\ \mathrm{ps}(u)=b}} (1-4\epsilon)^{r-|u|} \left(\frac{4\epsilon}{d}\right)^{|u|} \\
&= \sum_{r=0}^{t} \frac{\binom{t}{r}}{2^t} \sum_{l=0}^{r} \binom{r}{l} (1-4\epsilon)^{r-l} \left(\frac{4\epsilon}{d}\right)^{l} n_{l,|b|} \ ,
\end{aligned}$$

where $n_{l,|b|}$ denotes the number of strings in $[d]^l$ that have the same parity signature with Hamming weight $|b|$ (see Section 1.2.1). Note that the eigenvalue λ_b only depends on $|b|$. Thus, its multiplicity is $\binom{d}{h}$. In the rest of the proof, we write λ_h instead of λ_b for a string b with Hamming weight h.

Since the Hadamard operator is unitary, we have $\mathrm{S}(B)_\rho = \mathrm{S}(B)_\sigma$ and therefore,

$$\begin{aligned}
\mathrm{S}(B)_\rho &= \sum_{h=0}^{d} \binom{d}{h} \lambda_h \log \frac{1}{\lambda_h} \\
&\leq \log d + \sum_{h=0}^{d} \binom{d}{h} \lambda_h \log \binom{d}{h} \qquad \text{(since } \textstyle\sum_{h=0}^{d} \binom{d}{h} \lambda_h = 1\text{)} \\
&= \log d + \sum_{r=0}^{t} \frac{\binom{t}{r}}{2^t} \sum_{h=0}^{d} p_r(h) \log \binom{d}{h} \ , \qquad (4.8)
\end{aligned}$$

where

$$p_r(h) \coloneqq \binom{d}{h} \sum_{l=0}^{r} \binom{r}{l} (1-4\epsilon)^{r-l} \left(\tfrac{4\epsilon}{d}\right)^{l} n_{l,h} \ .$$

For each r, p_r is a probability distribution over $\mathbb{N}_d$ to be described shortly. So, $p_r(h)\binom{t}{r}/2^t$ forms a probability distribution over $r \in \mathbb{N}_t$ and $h \in \mathbb{N}_d$. In the rest of the proof, we use

concentration of measure to bound $\mathrm{S}(B)_\rho$. In particular, we first use the concentration of the binomial distribution over $r \in \mathbb{N}_t$ around $t/2$ to bound the terms with $r \leq t/4$ or $r \geq 3t/4$. We then bound the mean of the distribution $p_r(h)$ by $8\epsilon r$, which is bounded by $6\nu d$ for $r \in [t/4, 3t/4]$. Finally, we use the concentration of p_r around its mean to bound the terms corresponding to $h \geq 7\nu d$.

For an integer $m \in [t]$, let X be a random string in $\mathbb{N}_d^m$ chosen according to the distribution $D^{\otimes m}$, and Z_j be the binary random variable corresponding to the parity of the number of occurrences of the symbol $j \in [d]$ in X. Hence, $Z \coloneqq \sum_j Z_j$ is the random variable corresponding to the Hamming weight of the parity signature of X, and $p_m(h)$ is the probability that the parity signature of X has Hamming weight h. In other words, $p_m(h) = \Pr[Z = h]$. For a fixed symbol $j \in [d]$, the parity of the number of occurrences of j in X is even with probability α_0 and is odd with probability α_1 where $\alpha_0 + \alpha_1 = 1$ and

$$\alpha_0 - \alpha_1 = \sum_{w=0}^{r} (-1)^w \binom{r}{w} \left(\frac{4\epsilon}{d}\right)^w \left(1 - \frac{4\epsilon}{d}\right)^{r-w} \quad = \quad \left(1 - \frac{8\epsilon}{d}\right)^m \ .$$

Therefore, the expected value of Z_j equals $\mathbb{E}[Z_j] = \alpha_1 = \frac{1}{2}\left(1 - \left(1 - \frac{8\epsilon}{d}\right)^m\right)$. Moreover, linearity of expectation implies that

$$\mathbb{E}[Z] \quad = \quad \sum_{j=1}^{d} \mathbb{E}[Z_j] \quad = \quad \frac{d}{2}\left(1 - \left(1 - \frac{8\epsilon}{d}\right)^m\right) \quad \leq \quad \frac{d}{2}\left(1 - \left(\frac{1}{4}\right)^{\frac{8\epsilon m}{d}}\right) \quad \leq \quad 8\epsilon m \ , \tag{4.9}$$

where the inequalities follow from the fact that $1 - 2x \leq \frac{1}{4^x} \leq 1 - x$ for all $x \in [0, 1/2]$. Since $p_m(h) = 0$ for $h > m$, we have

$$\sum_{h=0}^{d} p_m(h) \log \binom{d}{h} \quad \leq \quad \max_{0 \leq h \leq m} \log \binom{d}{h} \quad \leq \quad d\,\mathrm{H}\left(\min\left\{\frac{1}{2}, \frac{m}{d}\right\}\right) \ .$$

This implies that

$$\begin{aligned}
&\sum_{r=0}^{t/4} \frac{\binom{t}{r}}{2^t} \sum_{h=0}^{d} p_r(h) \log \binom{d}{h} + \sum_{r=3t/4}^{t} \frac{\binom{t}{r}}{2^t} \sum_{h=0}^{d} p_r(h) \log \binom{d}{h} \\
&\qquad \leq \quad 2 \cdot 2^{-(1-\mathrm{H}(1/4))t} \cdot \mathrm{H}\left(\min\left\{\frac{1}{2}, \frac{\nu}{\epsilon}\right\}\right) \cdot d \qquad \text{(By Eq. (1.4))} \\
&\qquad \leq \quad \frac{2\,\epsilon}{(1 - \mathrm{H}(1/4))\nu}\, \mathrm{H}\left(\min\left\{\frac{1}{2}, \frac{\nu}{\epsilon}\right\}\right) \qquad \text{(since } x2^{-x} \leq 1 \quad \forall x \geq 0 \text{)} \\
&\qquad \leq \quad \max\left\{\frac{3\,\mathrm{H}(4\nu)}{\nu}, \frac{4}{1 - \mathrm{H}(1/4)}\right\} \ . \qquad (4.10)
\end{aligned}$$

The last inequality holds since $\mathrm{H}(x)/x$ is a non-increasing function of x and $\epsilon \leq 1/4$. So it remains to bound the terms in Eq. (4.8) with $r \in [t/4, 3t/4]$.

For $r \in [t/4, 3t/4]$, let X and Z be the random variables defined above with $m = r$. By Eq. (4.9), we have $\mathbb{E}[Z] \leq 6\nu d$, and therefore, we have

$$\sum_{r=t/4}^{3t/4} \frac{\binom{t}{r}}{2^t} \sum_{h=0}^{d} p_r(h) \log \binom{d}{h} \quad \leq \quad d\,\mathrm{H}(7\nu) + \sum_{r=t/4}^{3t/4} \frac{\binom{t}{r}}{2^t} \sum_{h=7\nu d}^{d} p_r(h) \log \binom{d}{h} \tag{4.11}$$

since $\log \binom{d}{h} \leq d\,\mathrm{H}(7\nu)$ for $h \leq 7\nu d$. Notice that the Hamming weight of X is at least as large as the Hamming weight of its parity signature, and has mean value $\mathbb{E}|X| = 4\epsilon r \leq 3\nu d$ for $r \in [t/4, 3t/4]$. Hence, the multiplicative Chernoff bound implies that

$$\sum_{h=7\nu d}^{d} p_r(h) \quad \leq \quad \Pr[\,|X| \geq 7\nu d\,] \quad \leq \quad \exp(-\nu d)\ . \tag{4.12}$$

Combining Eqs. (4.8), (4.10), (4.11) and (4.12), and using the fact that $x\,\mathrm{e}^{-x} \leq 0.5$ for $x \geq 0$ and $\epsilon \leq 1/4$, we get

$$\begin{aligned} \mathrm{I}(A:B)_\rho \quad &\leq \quad \log d + d\,\mathrm{H}(7\nu) + d\,\exp(-\nu d) + \max\left\{\frac{3\,\mathrm{H}(4\nu)}{\nu}, \frac{4}{(1-\mathrm{H}(1/4))}\right\} \\ &\leq \quad \log d + d\,\mathrm{H}(7\nu) + \frac{1}{2\nu} + \max\left\{\frac{3\,\mathrm{H}(4\nu)}{\nu}, \frac{4}{1-\mathrm{H}(1/4)}\right\} \\ &< \quad \frac{d}{4} - 2\ , \end{aligned}$$

for $\nu = 0.0023$ and $d \geq 2512$. ∎

4.4 A lower bound on sample complexity of agnostic learning

In this section, we show that the method used in Section 4.3 can also be utilized to bound agnostic quantum sample complexity. In particular, let $\mathcal{C}$ be a concept class with VC-dim$(\mathcal{C}) = d$ and D be a distribution over n-bit strings. We use information-theoretic arguments to show that, for $\epsilon \in (0, 1/4)$ and $\delta \in (0, 1/2)$, the (ϵ, δ)-agnostic quantum sample complexity of $\mathcal{C}$ is

$$\Omega\left(\frac{d}{\epsilon^2} + \frac{\log(1/\delta)}{\epsilon^2}\right)\ .$$

Similar to the PAC model, the lower bound contains two parts, one depends on the VC dimension and the other one is independent of the VC dimension. For completeness, we provide the proof of the VC-independent part, due to Arunachalam and de Wolf [12], in the following lemma.

Lemma 4.3. *Let $\mathcal{C}$ be a non-trivial concept class. For every $\delta \in (0, 1/2)$, $\epsilon \in (0, 1/4)$, an (ϵ, δ)-agnostic quantum learner for $\mathcal{C}$ has quantum sample complexity at least $\Omega(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$.*

Proof: For a non-trivial $\mathcal{C}$, there exists two concepts $c_1, c_2 \in \mathcal{C}$ and an input $x \in \{0,1\}^n$ such that $c_1(x) \neq c_2(x)$. Define distributions D_+ and D_- as follows:

$$D_\pm(x, c_1(x)) = \frac{1 \pm 2\epsilon}{2} \qquad \text{and} \qquad D_\pm(x, c_2(x)) = \frac{1 \mp 2\epsilon}{2} \ .$$

Let $|\psi_\pm\rangle := \sqrt{(1 \pm 2\epsilon)/2}\,|x, c_1(x)\rangle + \sqrt{(1 \mp 2\epsilon)/2}\,|x, c_2(x)\rangle$ be the outputs of agnostic quantum oracles $\mathrm{QAEX}(D_\pm)$. Under these distributions, $\mathrm{opt}_{D_\pm}(\mathcal{C}) = (1 - 2\epsilon)/2$ and no hypothesis can simultaneously have error at most ϵ larger than $\mathrm{opt}_{D_+}(\mathcal{C})$ and $\mathrm{opt}_{D_-}(\mathcal{C})$. So, an (ϵ, δ)-agnostic quantum learner for $\mathcal{C}$ can be used to distinguish $|\psi_+\rangle^{\otimes t}$ from $|\psi_-\rangle^{\otimes t}$ with probability at least $1 - \delta$, where t is the sample complexity of the learner. As in Lemma 4.1, we can conclude that $\langle\psi_+|\psi_-\rangle^t \leq 2\sqrt{\delta(1-\delta)}$ while $\langle\psi_+|\psi_-\rangle = \sqrt{1 - 4\epsilon^2}$. This implies that $t = \Omega\big(\frac{1}{\epsilon^2} \log \frac{1}{\delta}\big)$. ∎

Next, we prove the VC-dependent part of the lower bound using the same approach as in the proof of Theorem 4.2.

Theorem 4.4. *Let $n \geq 2$, $d \geq 2884$, $N = 2^n$ and $\mathcal{C} \subseteq \{0,1\}^N$ be a concept class with $\mathrm{VC\text{-}dim}(\mathcal{C}) = d$. For $\epsilon \in (0, 1/4)$ and $\delta \in (0, 1/2)$, every (ϵ, δ)-agnostic quantum learner for $\mathcal{C}$ has sample complexity $\Omega(d/\epsilon^2)$.*

Proof: Suppose there is an (ϵ, δ)-agnostic quantum learner $\mathcal{A}$ for $\mathcal{C}$ using t quantum examples. Let $S = \{s_1, \ldots, s_d\} \subseteq \{0,1\}^n$ be a set of size d shattered by the concept class $\mathcal{C}$. For $a \in \{0,1\}^d$, let $c_a \in \mathcal{C}$ be a concept such that $c_a(s_i) = a_i$ for all $i \in [d]$, and D_a be a distribution over $\{0,1\}^n \times \{0,1\}$ such that $D_a(s_i, l) = \big(1 + (-1)^{a_i + l} 4\epsilon\big)/2d$ for $i \in [d]$ and $l \in \{0,1\}$, and $D_a(x, l) = 0$ otherwise. Since distributions D_a are only supported over the set $S \times \{0,1\}$, in the rest of the proof, we use i and s_i interchangeably for the sake of brevity. Similar to Theorem 4.2, we define

$$\rho^{AB_1 \ldots B_t} \quad := \quad \frac{1}{2^d} \sum_{a \in \{0,1\}^d} |a\rangle\langle a|^A \otimes \rho_a^{B_1 \ldots B_t} \quad = \quad \frac{1}{2^d} \sum_{a \in \{0,1\}^d} |a\rangle\langle a| \otimes |\psi_a\rangle\langle\psi_a|^{\otimes t} \ , \quad (4.13)$$

where $|\psi_a\rangle \coloneqq \sum_{i=1}^{d} \sum_{l=0}^{1} \sqrt{D_a(i,l)}|i,l\rangle$ is the agnostic quantum example for c_a according to distribution D_a. Suppose the learner $\mathcal{A}$ is given t quantum examples in the state ρ_a^B. With probability at least $1-\delta$, $\mathcal{A}$ outputs a hypothesis $h_a \in \{0,1\}^N$ such that $\mathrm{err}_{D_a}(h_a)$ is at most $\mathrm{opt}_{D_a}(\mathcal{C}) + \epsilon$. By construction, c_a is the minimal-error concept from $\mathcal{C}$ with respect to the distribution D_a and h_a has additional error $\Pr_{i\sim D_a}[h_a(s_i) \neq c_a(s_i)] \cdot 4\epsilon$. Therefore, we can conclude that $\Pr_{i\sim D_a}[h_a(s_i) \neq c_a(s_i)] \leq 1/4$. Using the same argument as in Eq. (4.6), we get

$$\mathrm{I}(A:B)_\rho \quad \geq \quad (1-\delta-1/4)\,d - \mathrm{H}(\delta) - \mathrm{H}(1/4) \quad \geq \quad \frac{d}{4} - 2 \ . \tag{4.14}$$

Now assume $t < d/(1-\sqrt{1-16\epsilon^2})$; otherwise, we are done. Suppose $t = \nu d/(1-\sqrt{1-16\epsilon^2})$ for some $\nu \in (0,1)$. (Note that $t = 0$ is not possible.) We show that there are positive universal constants α, d_0 such that if $\nu < \alpha$ and $d \geq d_0$, the mutual information $\mathrm{I}(A:B)$ violates Eq. (4.14).

Notice that $\mathrm{I}(A:B)_\rho = \mathrm{S}(B)_\rho$ with

$$\rho^B \quad = \quad \frac{1}{2^d} \sum_{a\in\{0,1\}^d} \sum_{u,v\in[d]^t} \sum_{l,k\in\{0,1\}^t} \sqrt{D_a(u,l)D_a(v,k)}\,|u\rangle\langle v|^I \otimes |l\rangle\langle k|^L \ , \tag{4.15}$$

where register B is partitioned into registers I and L, storing indices and labels, respectively. We define

$$\begin{aligned} \sigma^B \quad &\coloneqq \quad \left(\mathbb{1}^L \otimes H^{\otimes t}\right)\rho^B\left(\mathbb{1}^L \otimes H^{\otimes t}\right) \\ &= \quad \frac{1}{2^d 2^t} \sum_{u,v\in[d]^t} |u\rangle\langle v|^I \otimes \sum_{\substack{l,k\in\{0,1\}^t \\ x,y\in\{0,1\}^t}} \sum_{a\in\{0,1\}^d} (-1)^{x\cdot l + y\cdot k} \sqrt{D_a(u,l)D_a(v,k)}\,|x\rangle\langle y|^L \ . \end{aligned}$$

By definition of D_a, $\sqrt{D_a(j,0)} + (-1)^y\sqrt{D_a(j,1)} = \frac{(-1)^{ya_j}}{\sqrt{2d}}\left(\sqrt{1+4\epsilon} + (-1)^y\sqrt{1-4\epsilon}\right)$ for every $j \in [d]$ and $y \in \{0,1\}$. So, we have

$$\sum_{l,k\in\{0,1\}^t} \sum_{a\in\{0,1\}^d} (-1)^{x\cdot l + y\cdot k} \sqrt{D_a(u,l)D_a(v,k)} \quad = \quad \begin{cases} 2^d\beta_x\beta_y & \text{if } \mathrm{ps}(u_x) = \mathrm{ps}(v_y) \\ 0 & \text{otherwise} \end{cases} \ ,$$

where $\beta_x \coloneqq \frac{1}{(2d)^{t/2}} \left(\sqrt{1+4\epsilon} - \sqrt{1-4\epsilon}\right)^{|x|} \left(\sqrt{1+4\epsilon} + \sqrt{1-4\epsilon}\right)^{t-|x|}$. This implies that

$$\begin{aligned}
\sigma^B &= \frac{1}{2^t} \sum_{x,y\in\{0,1\}^t} \beta_x\beta_y \sum_{\substack{u,v\in[d]^t \\ \mathrm{ps}(u_x)=\mathrm{ps}(v_y)}} |u\rangle\langle v|^I \otimes |x\rangle\langle y|^L \\
&= \frac{1}{2^t} \sum_{b\in\{0,1\}^d} \left(\sum_{x\in\{0,1\}^t} \beta_x \sum_{\substack{u\in\mathbb{N}_d^t \\ \mathrm{ps}(u_x)=b}} |u\rangle^I |x\rangle^L \right) \left(\sum_{y\in\{0,1\}^t} \beta_y \sum_{\substack{v\in\mathbb{N}_d^t \\ \mathrm{ps}(v_y)=b}} \langle v|^I \langle y|^L \right) .
\end{aligned}$$

For each $b \in \{0,1\}^d$, let

$$|\phi_b\rangle \coloneqq \sum_{x\in\{0,1\}^t} \beta_x \sum_{\substack{u\in[d]^t \\ \mathrm{ps}(u_x)=b}} |u\rangle|x\rangle .$$

For each $x \in \{0,1\}^t$ and $u \in \mathbb{N}_d^t$, the value of $\mathrm{ps}(u_x)$ is unique, and so $\langle\phi_b|\phi_{b'}\rangle = 0$ for distinct d-bit strings b and b'. Moreover, for each $b \in \{0,1\}^d$, $|\phi_b\rangle$ is an eigenvector of σ^B with corresponding eigenvalue

$$\begin{aligned}
\lambda_b &\coloneqq \sum_{x\in\{0,1\}^t} \sum_{\substack{u\in[d]^t \\ \mathrm{ps}(u_x)=b}} \frac{1}{2^t}\,\beta_x^2 && (4.16)\\
&= \sum_{x\in\{0,1\}^t} \frac{d^{t-|x|}}{2^t}\,\beta_x^2 \cdot n_{|x|,|b|} && (4.17)\\
&= \sum_{x\in\{0,1\}^t} \frac{1}{2^t d^{|x|}} n_{|x|,|b|} \left(1-\sqrt{1-16\epsilon^2}\right)^{|x|} \left(1+\sqrt{1-16\epsilon^2}\right)^{t-|x|} && (4.18)\\
&= \frac{1}{2^t} \sum_{r=0}^{t} \binom{t}{r} \frac{n_{r,|b|}}{d^r} \left(1-\sqrt{1-16\epsilon^2}\right)^{r} \left(1+\sqrt{1-16\epsilon^2}\right)^{t-r} . && (4.19)
\end{aligned}$$

Notice that the eigenvalue λ_b only depends on $|b|$. Thus its multiplicity is $\binom{d}{h}$, and we write λ_h instead of λ_b for a string b with Hamming weight h. Since the Hadamard operator

is unitary, we have $\mathrm{S}(B)_\rho = \mathrm{S}(B)_\sigma$ and therefore, we have

$$\begin{aligned}\mathrm{S}(B)_\rho &\leq \sum_{h=0}^{d} \binom{d}{h} \lambda_h \log \binom{d}{h} + \log d \\ &= \sum_{r=0}^{t} \binom{t}{r} \left(\frac{1-\sqrt{1-16\epsilon^2}}{2}\right)^r \left(\frac{1+\sqrt{1-16\epsilon^2}}{2}\right)^{t-r} \sum_{h=0}^{d} \binom{d}{h} \frac{n_{r,h}}{d^r} \log \binom{d}{h} \\ &\quad + \log d \ . \end{aligned} \tag{4.20}$$

Let $p \coloneqq \frac{1-\sqrt{1-16\epsilon^2}}{2}$. The term $\binom{t}{r} p^r (1-p)^{t-r}$ corresponds to the binomial distribution $\mathsf{B}(p,t)$ over $r \in \mathbb{N}_t$. In the rest of the proof, we use the concentration of measure to bound $S(B)_\rho$. In particular, we use the concentration of $\mathsf{B}(p,t)$ around $r = pt$ to bound the terms with $r \geq 3pt/2$. Then, we use the fact that $3pt/2 < d/2$ to bound the terms corresponding to $r < 3pt/2$.

We partition the interval $0 \leq r \leq t$ into two intervals, $0 \leq r < 3pt/2$ and $3pt/2 \leq r \leq t$. For $r \geq 3pt/2$, since the binomial distribution $\mathsf{B}(t,p)$ is concentrated around $r = pt$, Eq. (1.3) implies that

$$\sum_{r=\frac{3pt}{2}}^{t} \binom{t}{r} p^r (1-p)^{t-r} \sum_{h=0}^{d} \binom{d}{h} \frac{n_{r,h}}{d^r} \log \binom{d}{h} \leq d \exp\left(-\frac{pt}{10}\right) . \tag{4.21}$$

Note that the Hamming weight of the parity signature of a string of length r can be at most r, i.e., $n_{r,h} = 0$ if $h > r$. Moreover, $\log \binom{d}{h}$ is an increasing function of h for $h \leq d/2$. Hence, for each $r \leq \frac{d}{2}$,

$$\sum_{h=0}^{d} \binom{d}{h} \tfrac{n_{r,h}}{d^r} \log \binom{d}{h} = \sum_{h=0}^{r} \binom{d}{h} \tfrac{n_{r,h}}{d^r} \log \binom{d}{h} \leq \log \binom{d}{r} .$$

Since $3pt/2 = 3\nu d/4 < d/2$, the above equation implies that

$$\sum_{r=0}^{\frac{3pt}{2}} \binom{t}{r} p^r (1-p)^{t-r} \sum_{h=0}^{d} \binom{d}{h} \frac{n_{r,h}}{d^r} \log \binom{d}{h} \leq \log \binom{d}{\frac{3\nu}{4}d} . \tag{4.22}$$

Combining Eqs. (4.20), (4.21) and (4.22), since $x\, \mathrm{e}^{-x} \leq 0.4$ for $x \geq 0$ and $\log \binom{n}{k} \leq \mathrm{H}(k/n)n$

for $k \in [0, n]$, we get

$$\begin{aligned} \mathrm{I}(A:B)_\rho &\leq \log d + \log \binom{d}{\frac{3\nu}{4}d} + d \exp\left(-\frac{pt}{10}\right) \\ &\leq \log d + \mathrm{H}\left(\frac{3\nu}{4}\right) d + \frac{8}{\nu} \\ &< \frac{d}{4} - 2 \ , \end{aligned}$$

where the last inequality holds for $\nu = 0.025$ and $d \geq 2884$. ■

4.5 Conclusion and outlook

In this chapter, we considered two commonly studied models for learning Boolean functions: PAC learning and agnostic learning. We used an information-theoretic approach to show that classical PAC learners and classical agnostic learners are as powerful as quantum PAC learners and quantum agnostic learners, respectively. A similar result was previously proved in Ref. [12] using a quantum state identification argument and Fourier analysis. However, their proof is more complicated than the one we provide. Also, we believe that our method can possibly result in optimal bounds for similar scenarios like learning *noisy* quantum samples [45, 12] and the *quantum coupon collector* [13] problem.

References

[1] Anura Abeyesinghe, Igor Devetak, Patrick Hayden, and Andreas Winter. The mother of all protocols: Restructuring quantum information's family tree. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 465(2108):2537–2563, June 2009.

[2] Anurag Anshu, Mario Berta, Rahul Jain, and Marco Tomamichel. Partially smoothed information measures. *IEEE Transactions on Information Theory*, 66(8):5022–5036, August 2020.

[3] Anurag Anshu, Vamsi Krishna Devabathini, and Rahul Jain. Quantum communication using coherent rejection sampling. *Physical Review Letters*, 119:120506, September 2017.

[4] Anurag Anshu and Rahul Jain. Efficient methods for one-shot quantum communication. Technical Report arXiv:1809.07056 [quant-ph], arXiv.org, `https://arxiv.org/abs/1809.07056`, September 2018.

[5] Anurag Anshu, Rahul Jain, Priyanka Mukhopadhyay, Ala Shayeghi, and Penghui Yao. New one shot quantum protocols with application to communication complexity. *IEEE Transactions on Information Theory*, 62(12):7566–7577, December 2016.

[6] Anurag Anshu, Rahul Jain, and Naqueeb A. Warsi. A unified approach to source and message compression. Technical Report arXiv:1707.03619 [quant-ph], arXiv.org, `https://arxiv.org/pdf/1410.3031.pdf`, July 2017.

[7] Anurag Anshu, Rahul Jain, and Naqueeb A. Warsi. Building blocks for communication over noisy quantum networks. *IEEE Transactions on Information Theory*, 65(2):1287–1306, February 2019.

[8] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. A one-shot achievability result for quantum state redistribution. *IEEE Transactions on Information Theory*, 64(3):1425–1435, March 2017.

[9] Anurag Anshu, Dave Touchette, Penghui Yao, and Nengkun Yu. Exponential separation of quantum communication and classical information. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, STOC 2017, pages 277–288, New York, NY, USA, 2017. ACM.

[10] Bruno Apolloni and Claudio Gentile. Sample size lower bounds in PAC learning by algorithmic complexity theory. *Theoretical Computer Science*, 209(1–2):141–162, December 1998.

[11] Srinivasan Arunachalam and Ronald de Wolf. Guest column: A survey of quantum learning theory. *SIGACT News*, 48(2):41–67, June 2017.

[12] Srinivasan Arunachalam and Ronald de Wolf. Optimal quantum sample complexity of learning algorithms. *Journal of Machine Learning Research*, 19(1):2879–2878, January 2018.

[13] Srinivasan Arunachalan, Aleksandrs Belovs, Andrew M. Childs, Robin Kothari, Ansis Rosmanis, and Ronald de Wolf. Quantum coupon collector. Technical Report arXiv:2002.07688 [quant-ph], arXiv.org, `https://arxiv.org/abs/2002.07688`, February 2020.

[14] Alp Atici and Rocco A. Servedio. Improved bounds on quantum learning algorithms. *Quantum Information Processing*, 4(5):355–386, November 2005.

[15] Shima Bab Hadiashar and Ashwin Nayak. On the Entanglement Cost of One-Shot Compression. *Quantum*, 4:286, June 2020.

[16] Shima Bab Hadiashar, Ashwin Nayak, and Renato Renner. Communication complexity of one-shot remote state preparation. *IEEE Transactions on Information Theory*, 64(7):4709–4728, July 2018.

[17] Zahra Baghali Khanian and Andreas Winter. Entanglement-assisted quantum data compression. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 1147–1151, 2019.

[18] Zahra Baghali Khanian and Andreas Winter. General mixed state quantum data compression with and without entanglement assistance. Technical Report

arXiv:1912.08506 [quant-ph], arXiv.org, https://arxiv.org/abs/1912.08506, December 2019.

[19] Ziv Bar-Yossef, Thathachar S. Jayram, Ravi Kumar, and Dandapani Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[20] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[21] Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. On quantum coding for ensembles of mixed states. *Journal of Physics A: Mathematical and General*, 34(35):6767–6785, August 2001.

[22] Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels. *IEEE Transactions on Information Theory*, 60(5):2926–2959, May 2014.

[23] Charles H. Bennett, David P. DiVincenzo, Peter W. Shor, John A. Smolin, Barbara M. Terhal, and William K. Wootters. Remote state preparation. *Physical Review Letters*, 87:077902, July 2001.

[24] Charles H. Bennett, Patrick Hayden, Debbie W. Leung, Peter W. Shor, and Andreas Winter. Remote preparation of quantum states. *IEEE Transactions on Information Theory*, 51(1):56–74, January 2005.

[25] Dominic W. Berry and Barry C. Sanders. Optimal remote state preparation. *Physical Review Letters*, 90:057901, February 2003.

[26] Mario Berta, Matthias Christandl, and Renato Renner. The quantum reverse Shannon theorem based on one-shot information theory. *Communications in Mathematical Physics*, 306(3):579–615, August 2011.

[27] Mario Berta, Matthias Christandl, and Dave Touchette. Smooth entropy bounds on one-shot quantum state redistribution. *IEEE Transactions on Information Theory*, 62(3):1425–1439, March 2016.

[28] Mario Berta, Kaushik P. Seshadreesan, and Mark M. Wilde. Rényi generalizations of the conditional quantum mutual information. *Journal of Mathematical Physics*, 56(2):022205, 2015.

[29] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, October 1989.

[30] Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, October 2014.

[31] Nader H. Bshouty and Jeffrey C. Jackson. Learning DNF over the uniform distribution using a quantum example oracle. *SIAM Journal on Computing*, 28(3):1136–1153, 1998.

[32] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 270–278, Washington, DC, USA, October 2001. IEEE Computer Society.

[33] Matthias Christandl, Norbert Schuch, and Andreas Winter. Entanglement of the antisymmetric state. *Communications in Mathematical Physics*, 311(2):397–422, April 2012.

[34] Nikola Ciganovic, Normand J. Beaudry, and Renato Renner. Smooth max-information as one-shot generalization for mutual information. *IEEE Transactions on Information Theory*, 60(3):1573–1581, March 2014.

[35] Richard Cleve and Harry Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201, 1997.

[36] Nilanjana Datta, Min-Hsiu Hsieh, and Jonathan Oppenheim. An upper bound on the second order asymptotic expansion for the quantum communication cost of state redistribution. *Journal of Mathematical Physics*, 57(5):052203, 2016.

[37] Igor Devetak. Triangle of dualities between quantum communication protocols. *Physical Review Letters*, 97:140503, October 2006.

[38] Igor Devetak and Jon Yard. Exact cost of redistributing multipartite quantum states. *Physical Review Letters*, 100(230501), June 2008.

[39] Vedran Dunjko and Hans J Briegel. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Reports on Progress in Physics*, 81(7):074001, June 2018.

[40] Robert M. Fano. *The transmission of Information : A Statistical Theory of Communications*. Technical report. - Research Laboratory of Electronics, Massachusetts Institute of Technology ; no. 65, 149. Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Mass, 1961.

[41] William Feller. *An Introduction to Probability Theory and its Applications*. John Wiley & Sons New York, January 1968. Volume 1, 3rd Edition.

[42] Jörg Flum and Martin Grohe. *Parameterized Complexity Theory*. Texts in theoretical computer science: an EATCS series. Springer Verlag, DE, 2006.

[43] Christopher A. Fuchs and Jeroen van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, May 1999.

[44] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. *Journal of the ACM*, 63(5), November 2016.

[45] Claudio Gentile and David P. Helmbold. Improved lower bounds for learning from noisy examples: An information-theoretic approach. *Information and Computation*, 166(2):133 – 155, 2001.

[46] Alexei Gilchrist, Nathan K. Langford, and Michael A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71:062310, June 2005.

[47] Steve Hanneke. The optimal sample complexity of PAC learning. *Journal of Machine Learning Research*, 17(1):13190–1333, January 2016.

[48] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 1(56):438–449, January 2010.

[49] David Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100(1):78 – 150, 1992.

[50] Akihisa Hayashi, Takaaki Hashimoto, and Minoru Horibe. Remote state preparation without oblivious conditions. *Physical Review A*, 67:052302, May 2003.

[51] Patrick Hayden, Richard Jozsa, Dénes Petz, and Andreas Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in Mathematical Physics*, 246(2):359–374, April 2004.

[52] Carl W. Helstrom. Detection theory and quantum mechanics. *Information and Control*, 10(3):254–291, 1967.

[53] Alexander S. Holevo. An analogue of statistical decision theory and noncommutative probability theory. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 26:133–149, 1972.

[54] Michal Horodecki. Optimal compression for mixed signal states. *Physical Review A*, 61:052309, April 2000.

[55] Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. Partial quantum information. *Nature*, 436(7051):673–676, August 2005.

[56] Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269(1):107–136, January 2007.

[57] Ben Ibinson, Noah Linden, and Andreas Winter. Robustness of quantum Markov chains. *Communications in Mathematical Physics*, 277(2):289–304, January 2008.

[58] Rahul Jain. Communication complexity of remote state preparation with entanglement. *Quantum Information & Computation*, 6(4):461–464, July 2006.

[59] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for two-party bounded-round public-coin communication complexity. *Algorithmica*, 76(3):720–748, November 2016.

[60] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In Jos C. M. Baeten, Jan Karel Lenstra, Joachim Parrow, and Gerhard J. Woeginger, editors, *Automata, Languages and Programming*, volume 2719 of *Lecture Notes in Computer Science*, pages 300–315, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[61] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, pages 285–296. IEEE Computer Society, 2005.

[62] Rahul Jain, Pranab Sen, and Jaikumar Radhakrishnan. Optimal direct sum and privacy trade-off results for quantum and classical communication complexity. Technical Report arXiv:0807.1267v1 [cs.DC], arXiv.org, `https://arxiv.org/abs/0807.1267`, July 2008.

[63] Michael J. Kearns, Robert E. Schapire, and Linda M. Sellie. Toward efficient agnostic learning. *Machine Learning*, 17(2):115–141, 1994.

[64] Felix Leditzky, Mark M. Wilde, and Nilanjana Datta. Strong converse theorems using Rényi entropies. *Journal of Mathematical Physics*, 57(8):082202, 2016.

[65] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–399, 2009.

[66] Debbie W. Leung and Peter W. Shor. Oblivious remote state preparation. *Physical Review Letters*, 90:127905, March 2003.

[67] Ke Li. Second-order asymptotics for quantum hypothesis testing. *Annals of Statistics*, 42(1):171–189, February 2014.

[68] Zi-Wen Liu, Christopher Perry, Yechao Zhu, Dax Enshan Koh, and Scott Aaronson. Doubly infinite separation of quantum information and communication. *Physical Review A*, 93:012347, January 2016.

[69] Zhicheng Luo and Igor Devetak. Channel simulation with quantum side information. *IEEE Transactions on Information Theory*, 55(3):1331–1342, March 2009.

[70] Lorenzo Mascheroni. *Adnotationes ad calculum integralem Euleri: in quibus nonnulla problemata*. Galeatii, 1790.

[71] Jiří Matoušek. *Lectures on Discrete Geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 1st edition, 2002.

[72] Elizabeth S. Meckes. *The Random Matrix Theory of the Classical Compact Groups*, volume 218 of *Cambridge Tracts in Mathematics*. Cambridge University Press, July 2019.

[73] Vitali D. Milman and Gideon Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces*, volume 1200 of *Lecture notes in mathematics*. Springer-Verlag Berlin Heidelberg, 1986.

[74] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[75] John von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Springer DE, 1996.

[76] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, New York, NY, USA, 2011. 10th Anniversary Edition.

[77] Tomohiro Ogawa and Hiroshi Nagaoka. Making good codes for classical-quantum channel coding via quantum hypothesis testing. *IEEE Transactions on Information Theory*, 53(6):2261–2266, June 2007.

[78] Jonathan Oppenheim. State redistribution as merging: introducing the coherent relay. Technical Report arXiv:0805.1065 [quant-ph], arXiv.org, `https://arxiv.org/abs/0805.1065`, May 2008.

[79] Christopher Perry, Rahul Jain, and Jonathan Oppenheim. Communication tasks with infinite quantum-classical separation. *Physical Review Letters*, 115:030504, July 2015.

[80] Maxim Raginsky and Alexander Rakhlin. Lower bounds for passive and active learning. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 1026–1034. Curran Associates, Inc., 2011.

[81] Benjamin Schumacher. Quantum coding. *Physical Review A*, 51:2738–2747, April 1995.

[82] Rocco A. Servedio and Steven J. Gortler. Equivalences and separations between quantum and classical learnability. *SIAM Journal on Computing*, 33(5):1067–1092, 2004.

[83] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

[84] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[85] Michel Talagrand. Sharper bounds for Gaussian and empirical processes. *The Annals of Probability*, 22(1):28–76, 1994.

[86] Marco Tomamichel. *Quantum Information Processing with Finite Resources: Mathematical Foundations*, volume 5. Springer, 2015.

[87] Marco Tomamichel, Roger Colbeck, and Renato Renner. Duality between smooth min- and max-entropies. *IEEE Transactions on Information Theory*, 56(9):4674–4681, September 2010.

[88] Marco Tomamichel and Masahito Hayashi. A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Transactions on Information Theory*, 59(11):7693–7710, November 2013.

[89] Dave Touchette. Quantum information complexity. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 317–326, New York, NY, USA, 2015. ACM.

[90] Armin Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9(2):273 – 279, 1976.

[91] Leslie G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[92] Wim van Dam and Patrick Hayden. Universal entanglement transformations without communication. *Physical Review A*, 67:060302, June 2003.

[93] Vladimir N. Vapnik and Alexey Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280, 1971.

[94] Vladimir N. Vapnik and Alexey Y. Chervonenkis. *Theory of pattern recognition*. Nauka, USSR, 1974.

[95] John Von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete ; Bd. 38. Springer, Berlin, 1968.

[96] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, May 2018.

[97] Andreas Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, November 1999.

[98] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, STOC '79, pages 209–213, New York, NY, USA, 1979. ACM.

[99] Andrew Chi-Chih Yao. Quantum circuit complexity. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 352–361, November 1993.

[100] Jon T. Yard and Igor Devetak. Optimal quantum source coding with quantum side information at the encoder and decoder. *IEEE Transactions on Information Theory*, 55(11):5339–5351, November 2009.

[101] Ming-Yong Ye, Yan-Kui Bai, and Z. D. Wang. Quantum state redistribution based on a generalized decoupling. *Physical Review A*, 78:030302, September 2008.

APPENDICES

Appendix A

Proofs of some claims

Here, we include the proofs of some statements from the main body of the thesis.

Proposition A.1. *Let (ρ_{ij}) be an ensemble of the form in Eq. (2.4), and let the state τ^{AB} be defined as $\frac{1}{n}\sum_{ij}|ij\rangle\langle ij|^A \otimes \rho_{ij}^B$. For any $\zeta \in [0, 1/8)$, we have*

$$\mathrm{I}_{\max}^{\zeta}(A:B)_{\tau} \quad \geq \quad \log k - \log\left(\frac{3-12\zeta}{1-8\zeta}\right) \ .$$

Proof: As shown in Ref. [16, Proposition II.5], there is a classical-quantum state τ' within purified distance ζ of τ such that $\mathrm{I}_{\max}^{\zeta}(A:B)_{\tau} = \mathrm{I}_{\max}(A:B)_{\tau'}$. Let $\tau' \coloneqq \sum_{ij} q_{ij}|ij\rangle\langle ij| \otimes \widetilde{\rho}_{ij}$.

By Proposition 1.6, we have

$$\|\tau - \tau'\|_{\mathrm{tr}} \quad \leq \quad 2\zeta \ . \tag{A.1}$$

Let $\xi \coloneqq 2\zeta$. By monotonicity of trace distance under measurements [96, Proposition 3.5], we further get

$$\sum_{ij} |q_{ij} - p_{ij}| \quad \leq \quad \xi \ .$$

If $q_{ij} > 3/2n$ or $q_{ij} < 1/2n$, we have $|q_{ij} - p_{ij}| > 1/2n$. So for at least $(1-2\xi)n$ pairs (i,j), we have $1/2n \leq q_{ij} \leq 3/2n$, and we call such pairs (i,j) *typical*.

Eq. (A.1) may be written as

$$\sum_{ij} \|q_{ij}\widetilde{\rho}_{ij} - p_{ij}\rho_{ij}\|_{\mathrm{tr}} \quad \leq \quad \xi \ ,$$

so, by monotonicity of trace distance,

$$\sum_{ij} \sum_{|v\rangle \in B_{ij}} \left| q_{ij} \langle v|\widetilde{\rho}_{ij}|v\rangle - \frac{k}{nm} \right| \quad \leq \quad \xi \ ,$$

where B_{ij} is as in the definition of the ensemble (ρ_{ij}). In particular,

$$\sum_{\text{typical } ij} \sum_{|v\rangle \in B_{ij}} \left| q_{ij} \langle v|\widetilde{\rho}_{ij}|v\rangle - \frac{k}{nm} \right| \quad \leq \quad \xi \ . \tag{A.2}$$

There are at least $(1-2\xi)n/k$ indices $i \in [n/k]$ such that there is a typical pair (i,j) for some $j \in [k]$. Let S be the set of such indices i. Let $\eta \in (0,1)$. If for all indices $i \in S$, there are less than $(1-\eta)m$ pairs (j,v) with (i,j) typical, $|v\rangle \in B_{ij}$, and

$$\frac{k}{2nm} \quad \leq \quad q_{ij} \langle v|\widetilde{\rho}_{ij}|v\rangle \quad \leq \quad \frac{3k}{2nm} \ , \tag{A.3}$$

then we would have

$$\sum_{\text{typical } ij} \sum_{|v\rangle \in B_{ij}} \left| q_{ij} \langle v|\widetilde{\rho}_{ij}|v\rangle - \frac{k}{nm} \right| \quad > \quad (1-2\xi)\frac{n}{k} \times \eta m \times \frac{k}{2nm} \quad = \quad (1-2\xi)\frac{\eta}{2} \ .$$

Taking $\eta \coloneqq 2\xi/(1-2\xi)$, we see that this is in contradiction with Eq. (A.2). So there is an index $i \in S$ such that there are at least $(1-\eta)m$ pairs (j,v) with $j \in [k]$ and $|v\rangle \in B_{ij}$ such that (i,j) is typical, and (i,j,v) satisfy Eq. (A.3). Denote such an index i by i_0, and let

$$T \quad \coloneqq \quad \Big\{ (j,v) : j \in [k],\ |v\rangle \in B_{i_0 j},\ (i_0,j) \text{ typical} \ ,\ (i,j,v) \text{ satisfy Eq. (A.3)} \Big\} \ .$$

We have that for all the pairs $(j,v) \in T$,

$$\frac{k}{2nm} \quad \leq \quad q_{i_0 j} \langle v|\widetilde{\rho}_{i_0 j}|v\rangle \quad \leq \quad \frac{3}{2n} \langle v|\widetilde{\rho}_{i_0 j}|v\rangle \ ,$$

so that

$$\frac{k}{3m} \quad \leq \quad \langle v|\widetilde{\rho}_{i_0 j}|v\rangle \ . \tag{A.4}$$

Let $\sigma \in \mathsf{D}(\mathbb{C}^m)$ be a state that achieves $\mathrm{I}_{\max}(A:B)_{\tau'}$, and let λ denote this max-information. For typical pairs (i,j), since $q_{ij} > 0$, we have $\widetilde{\rho}_{ij} \leq 2^{\lambda}\sigma$. By Eq. (A.4), we

also have $k/3m \leq 2^{\lambda}\langle v|\sigma|v\rangle$ for all pairs $(j,v) \in T$. Summing up over all pairs $(j,v) \in T$, we get $(1-\eta)k/3 \leq 2^{\lambda}$, as the sets $B_{i_0 j}$ are a partition of an orthonormal basis, and σ has trace at most 1. So $\lambda \geq \log k - \log(3/(1-\eta))$. ∎

Next, we provide a proof of Eq. (3.6) from Section 3.1, formally stated in Lemma A.2. For the sake of clarity, in the proof below, we suppress tensor products with the identity in expressions involving sums or products of quantum states over different sequences of registers. For example, we write $\omega^{XY} + \tau^{YZ}$ to represent the sum $\omega^{XY} \otimes \mathbb{1}^{Z} + \mathbb{1}^{X} \otimes \tau^{YZ}$, and $\omega^{XY}\tau^{YZ}$ to represent the product $\left(\omega^{XY} \otimes \mathbb{1}^{Z}\right)\left(\mathbb{1}^{X} \otimes \tau^{YZ}\right)$. All the expressions involving entropy and mutual information are with respect to the state ψ.

Lemma A.2. *For any tripartite quantum state ψ^{RBC}, and any quantum Markov extension $\sigma^{RBC} \in \mathsf{QMC}^{\psi}_{R-B-C}$, it holds that*

$$\mathrm{I}(R:C|B)_{\psi} \quad = \quad \mathrm{D}\left(\psi^{RBC} \| \sigma^{RBC}\right) - \mathrm{D}\left(\psi^{BC} \| \sigma^{BC}\right) \ .$$

The proof of this lemma is implicit in Ref. [28, Lemma 1], but we provide a proof here for completeness.

Proof: Consider any quantum Markov chain σ^{RBC} satisfying $\sigma^{RB} = \psi^{RB}$. From Equation 3.5, we have

$$\log \sigma^{RBC} \quad = \quad \bigoplus_{j} \left(\log\left(p(j)\sigma_j^{RB_j^R}\right) + \log \sigma_j^{B_j^C C} \right) \ ,$$

and similarly,

$$\log \sigma^{BC} \quad = \quad \bigoplus_{j} \left(\log\left(p(j)\sigma_j^{B_j^R}\right) + \log \sigma_j^{B_j^C C} \right) \ .$$

Thus, we can evaluate

$$\begin{aligned}
&\mathrm{D}\left(\psi^{RBC} \| \sigma^{RBC}\right) - \mathrm{D}\left(\psi^{BC} \| \sigma^{BC}\right) \\
&\quad = \quad \mathrm{Tr}\left(\psi^{RBC} \log \psi^{RBC}\right) - \mathrm{Tr}\left(\psi^{RBC} \log \sigma^{RBC}\right) - \mathrm{Tr}\left(\psi^{BC} \log \psi^{BC}\right) \\
&\qquad + \mathrm{Tr}\left(\psi^{BC} \log \sigma^{BC}\right) \\
&\quad = \quad \mathrm{S}(BC) - \mathrm{S}(RBC) - \sum_j \mathrm{Tr}\left(\psi^{RBC} \log\left(p(j)\sigma_j^{RB_j^R}\right)\right) - \sum_j \mathrm{Tr}\left(\psi^{RBC} \log \sigma_j^{B_j^C C}\right) \\
&\qquad + \sum_j \mathrm{Tr}\left(\psi^{BC} \log\left(p(j)\sigma_j^{B_j^R}\right)\right) + \sum_j \mathrm{Tr}\left(\psi^{BC} \log \sigma_j^{B_j^C C}\right) \ .
\end{aligned}$$

Since $\mathrm{Tr}\Big(\psi^{RBC}\log\sigma_j^{B_j^C C}\Big) = \mathrm{Tr}\Big(\psi^{BC}\log\sigma_j^{B_j^C C}\Big)$, the above equation can be simplified to obtain

$$\begin{aligned}
&\mathrm{D}\big(\psi^{RBC}\|\sigma^{RBC}\big) - \mathrm{D}\big(\psi^{BC}\|\sigma^{BC}\big) \\
&\quad= \mathrm{S}(BC) - \mathrm{S}(RBC) - \sum_j \mathrm{Tr}\Big(\psi^{RBC}\log\Big(p(j)\sigma_j^{RB_j^R}\Big)\Big) \\
&\qquad + \sum_j \mathrm{Tr}\Big(\psi^{BC}\log\Big(p(j)\sigma_j^{B_j^R}\Big)\Big) \\
&\quad= \mathrm{S}(BC) - \mathrm{S}(RBC) - \mathrm{Tr}\left(\psi^{RBC}\log\left(\bigoplus_j p(j)\sigma_j^{RB_j^R}\right)\right) \\
&\qquad + \mathrm{Tr}\left(\psi^{BC}\log\left(\bigoplus_j p(j)\sigma_j^{B_j^R}\right)\right) \\
&\quad= \mathrm{S}(BC) - \mathrm{S}(RBC) - \mathrm{Tr}\left(\psi^{RBC}\log\bigoplus_j\Big(p(j)\sigma_j^{RB_j^R}\otimes\sigma_j^{B_j^C}\Big)\right) \\
&\qquad + \mathrm{Tr}\left(\psi^{BC}\log\bigoplus_j\Big(p(j)\sigma_j^{B_j^R}\otimes\sigma_j^{B_j^C}\Big)\right) \ ,
\end{aligned}$$

where the last equality above follows by noting that

$$\mathrm{Tr}\Big(\psi^{RBC}\log\sigma_j^{B_j^C}\Big) \quad = \quad \mathrm{Tr}\Big(\psi^{BC}\log\sigma_j^{B_j^C}\Big) \ .$$

Since $\psi^{RB} = \sigma^{RB}$, we get that

$$\begin{aligned}
\mathrm{D}\big(\psi^{RBC}\|\sigma^{RBC}\big) - \mathrm{D}\big(\psi^{BC}\|\sigma^{BC}\big) &= \mathrm{S}(BC) - \mathrm{S}(RBC) - \mathrm{Tr}\big(\psi^{RBC}\log\sigma^{RB}\big) \\
&\quad + \mathrm{Tr}\big(\psi^{BC}\log\sigma^{B}\big) \\
&= \mathrm{S}(BC) - \mathrm{S}(RBC) - \mathrm{Tr}\big(\psi^{RB}\log\psi^{RB}\big) \\
&\quad + \mathrm{Tr}\big(\psi^{B}\log\psi^{B}\big) \\
&= \mathrm{S}(BC) - \mathrm{S}(RBC) + \mathrm{S}(RB) - \mathrm{S}(B) \\
&= \mathrm{I}(R:C\,|\,B) \ .
\end{aligned}$$

This completes the proof. ∎

Next, we restate Theorem 1.9 and provide a proof for it. The proof is reproduced from Ref. [2] verbatim for completeness. In the following proof, for two (possibly sub-normalized)

states $\rho, \sigma \in \mathsf{D}(\mathcal{H})$, we define

$$\overline{\mathrm{F}}(\rho,\sigma) \quad := \quad \mathrm{Tr}\left[\left|\sqrt{\rho}\sqrt{\sigma}\right|\right] \ .$$

Theorem A.3 ([2], Theorem 2)**.** *Let $\epsilon, \delta \in (0,1)$ such that $0 \leq 2\epsilon + \delta \leq 1$. Consider quantum states $\sigma^B \in \mathsf{D}(\mathcal{H}^B)$ and $\rho^{AB} \in \mathsf{D}(\mathcal{H}^A \otimes \mathcal{H}^B)$. We have*

$$\inf_{\substack{\rho' \in \mathsf{B}^{2\epsilon+\delta}(\rho^{AB}) \\ \rho'^A = \rho^A}} \mathrm{D}_{\max}\big(\rho'^{AB} \| \rho^A \otimes \sigma^B\big) \quad \leq \quad \mathrm{D}^{\epsilon}_{\max}\big(\rho^{AB} \| \rho^A \otimes \sigma^B\big) + \log \frac{8+\delta^2}{\delta^2} \ . \tag{A.5}$$

Proof: Let $\tilde{\rho}^{AB} \in \mathsf{D}(\mathcal{H}^{AB})$ be the optimizer on the right-hand side of Eq. (A.5). Moreover, for some $\gamma > 0$, let

$$\Pi^A_\gamma \quad := \quad \left\{\frac{1}{\gamma}\tilde{\rho}^A - \rho^A\right\}_+ \qquad \text{and} \qquad \overline{\rho}^{AB} \quad := \quad \Pi^A_\gamma \tilde{\rho}^{AB} \Pi^A_\gamma \ ,$$

where $\{X\}_+$ denotes the projection operator onto the positive part of any Hermitian operator X. Let V^A be the unitary operator from the polar decomposition of $\rho^{A\frac{1}{2}}\overline{\rho}^{A\frac{1}{2}}$ such that

$$\mathrm{F}\big(\rho^A, \overline{\rho}^A\big) \quad = \quad \mathrm{Tr}\left[\left|\rho^{A\frac{1}{2}}\overline{\rho}^{A\frac{1}{2}}\right|\right] = \quad \mathrm{Tr}\left(\rho^{A\frac{1}{2}}\overline{\rho}^{A\frac{1}{2}} V^A\right) \ .$$

For $\gamma = \frac{\delta^2}{8}$, define the bipartite quantum state

$$\hat{\rho}^{AB} \quad := \underbrace{\rho^{A\frac{1}{2}}\, V^A\, \overline{\rho}^{A-\frac{1}{2}}\, \overline{\rho}^{AB}\, \overline{\rho}^{A-\frac{1}{2}}\, V^{A\dagger}\, \rho^{A\frac{1}{2}}}_{=:\tau^{AB}} + \underbrace{\left(\rho^{A\frac{1}{2}}\Big(\mathbb{1}^A - V^A \Pi^A_\gamma V^{A\dagger}\Big)\rho^{A\frac{1}{2}}\right) \otimes \sigma^B}_{=:\sigma^{AB}} \ ,$$

which by inspection has $\hat{\rho}^A = \rho^A$. We calculate

$$\begin{aligned} \hat{\rho}^{AB} \quad \leq \quad & \left\|\big(\rho^A \otimes \sigma^B\big)^{-\frac{1}{2}} \tilde{\rho}^{AB} \big(\rho^A \otimes \sigma^B\big)^{-\frac{1}{2}}\right\| \left(\rho^{A\frac{1}{2}}\, V^A\, \overline{\rho}^{A-\frac{1}{2}}\, \Pi^A_\gamma \rho^A \Pi^A_\gamma\, \overline{\rho}^{A-\frac{1}{2}}\, V^A\, \rho^{A\frac{1}{2}}\right) \otimes \sigma^B \\ & + \left(\rho^{A\frac{1}{2}}\Big(\mathbb{1}^A - V^A \Pi^A_\gamma V^{A\dagger}\Big)\rho^{A\frac{1}{2}}\right) \otimes \sigma^B \ , \end{aligned}$$

and by definition of Π^A_γ, we have $\Pi^A_\gamma \rho^A \Pi^A_\gamma \leq \frac{8}{\delta^2}$ as well as $\mathbb{1}^A - P^A_\gamma \leq \mathbb{1}^A$ leading to

$$\hat{\rho}^{AB} \quad \leq \quad \left(\frac{8}{\delta^2} \cdot \left\|\big(\rho^A \otimes \sigma^B\big)^{-\frac{1}{2}} \tilde{\rho}^{AB} \big(\rho^A \otimes \sigma^B\big)^{-\frac{1}{2}}\right\| + 1\right) \cdot \rho^A \otimes \sigma^B \ .$$

Using that $\mathrm{D}_{\max}\big(\tilde{\rho}^{AB}\|\rho^A\otimes\sigma^B\big)\geq 0$, we get

$$\frac{8}{\delta^2}\cdot\left\|\big(\rho^A\otimes\sigma^B\big)^{-\frac{1}{2}}\tilde{\rho}^{AB}\big(\rho^A\otimes\sigma^B\big)^{-\frac{1}{2}}\right\|+1 \quad\leq\quad \frac{8+\delta^2}{\delta^2}\left\|\big(\rho^A\otimes\sigma^B\big)^{-\frac{1}{2}}\tilde{\rho}^{AB}\big(\rho^A\otimes\sigma^B\big)^{-\frac{1}{2}}\right\| \ .$$

Hence, the claim follows as soon as we establish that $\hat{\rho}^{AB}$ is close enough to ρ^{AB} in purified distance. Now, notice that $\mathrm{Tr}[\sigma^{AB}]=1-\mathrm{Tr}[\tau^{AB}]$ and hence the states $\overline{\tau}^{AB}:=\frac{\tau^{AB}}{\mathrm{Tr}[\tau^{AB}]}$ and $\overline{\sigma}^{AB}:=\frac{\sigma^{AB}}{\mathrm{Tr}[\mathbb{1}-\tau^{AB}]}$ are normalized. We can then write

$$\hat{\rho}^{AB} \quad=\quad \mathrm{Tr}[\tau^{AB}]\cdot\overline{\tau}^{AB}+(1-\mathrm{Tr}[\tau^{AB}])\cdot\overline{\sigma}^{AB} \ .$$

Since the fidelity $\overline{\mathrm{F}}^2(\rho,\sigma)$ is concave in each argument (this follows from the operator concavity of the logarithm), we can estimate

$$\begin{aligned}\overline{\mathrm{F}}^2(\hat{\rho}^{AB},\overline{\rho}^{AB}) &\geq \mathrm{Tr}[\tau^{AB}]\cdot\mathrm{F}^2(\overline{\tau}^{AB},\overline{\rho}^{AB})+(1-\mathrm{Tr}[\tau^{AB}])\cdot\overline{\mathrm{F}}^2(\overline{\sigma}^{AB},\overline{\rho}^{AB})\\ &\geq \mathrm{Tr}[\tau^{AB}]\cdot\overline{\mathrm{F}}^2(\overline{\tau}^{AB},\overline{\rho}^{AB})\\ &= \overline{\mathrm{F}}^2(\tau^{AB},\overline{\rho}^{AB}) \ . \end{aligned} \tag{A.6}$$

By the triangle inequality for the purified distance, we get for the quantity of interest

$$\mathrm{P}(\hat{\rho}^{AB},\rho^{AB}) \quad\leq\quad \mathrm{P}(\hat{\rho}^{AB},\overline{\rho}^{AB})+\mathrm{P}(\overline{\rho}^{AB},\rho^{AB}) \ , \tag{A.7}$$

and since $\hat{\rho}^{AB}$ is normalized, we get for the first term on the right-hand side that

$$\mathrm{P}(\hat{\rho}^{AB},\overline{\rho}^{AB}) \quad=\quad \sqrt{1-\overline{\mathrm{F}}^2(\hat{\rho}^{AB},\overline{\rho}^{AB})} \ .$$

We continue with

$$\overline{\mathrm{F}}^2(\hat{\rho}^{AB},\overline{\rho}^{AB}) \quad\geq\quad \overline{\mathrm{F}}^2(\tau^{AB},\overline{\rho}^{AB}) \quad\geq\quad \overline{\mathrm{F}}^2(\tau^{ABC},\overline{\rho}^{ABC}) \ ,$$

where the first step is Eq. (A.6) and the second step follows since the fidelity is monotone under partial trace (this holds for general non-negative operators) together with choosing τ^{ABC} as an extension of τ^{AB} and $\overline{\rho}^{ABC}$ as an extension of $\overline{\rho}^{AB}$. We choose the purification of $\overline{\rho}^{AB}$ on ABC defined through the pure state

$$|\overline{\rho}\rangle^{ABC} \quad:=\quad \overline{\rho}^{A\frac{1}{2}}|\Phi\rangle^{A:BC} \ ,$$

where $|\Phi\rangle^{A:BC}$ denotes the non-normalized maximally entangled pure state vector in the "cut" $A:BC$ (on the subspace on A spanned by the projector Π^A_γ). Furthermore, we take the purification of τ^{AB} on ABC given by

$$|\tau\rangle^{ABC} \quad:=\quad \rho^{A\frac{1}{2}}\,V^A\,\overline{\rho}^{A\frac{1}{2}}|\overline{\rho}\rangle^{ABC}$$

which is fine since

$$\tau^{AB} \quad = \quad \mathrm{Tr}_C\left[|\tau\rangle\langle\tau|^{ABC}\right] \quad = \quad \rho^{A\frac{1}{2}}\, V^A\, \overline{\rho}^{A^{-\frac{1}{2}}}\, \overline{\rho}^{AB}\, \overline{\rho}^{A^{-\frac{1}{2}}}\, V^{A^\dagger}\, \rho^{A\frac{1}{2}} \ .$$

We calculate

$$\begin{aligned}
\overline{\mathrm{F}}^2(\tau^{ABC}, \overline{\rho}^{ABC}) \quad &= \quad \left|\langle\overline{\rho}^{ABC}|\tau^{ABC}\rangle\right|^2 \quad = \quad \left|\langle\Phi^{A:BC}|\overline{\rho}^{A\frac{1}{2}}|\tau^{ABC}\rangle\right|^2 \\
&= \quad \left|\langle\Phi^{A:BC}|\overline{\rho}^{A\frac{1}{2}}\rho^{A\frac{1}{2}}V^A\Pi^A_\delta|\Phi^{A:BC}\rangle\right|^2 = \left|\mathrm{Tr}\left[\overline{\rho}^{A\frac{1}{2}}\rho^{A\frac{1}{2}}V^A\Pi^A_\delta\right]\right|^2 \\
&= \quad \left|\mathrm{Tr}\left[\Pi^A_\delta\overline{\rho}^{A\frac{1}{2}}\rho^{A\frac{1}{2}}V^A\right]\right|^2 \quad = \quad \left|\mathrm{Tr}\left[\overline{\rho}^{A\frac{1}{2}}\rho^{A\frac{1}{2}}V^A\right]\right|^2 \\
&= \quad \overline{\mathrm{F}}^2(\overline{\rho}^A, \rho^A) \quad = \quad \mathrm{F}^2(\overline{\rho}^A, \rho^A) \ .
\end{aligned}$$

Hence, together with Eq. (A.7), we arrive at

$$\mathrm{P}(\hat{\rho}^{AB}, \rho^{AB}) \quad \leq \quad \mathrm{P}(\overline{\rho}^A, \rho^A) + \mathrm{P}(\overline{\rho}^{AB}, \rho^{AB}) \quad \leq \quad 2\cdot\mathrm{P}(\overline{\rho}^{AB}, \rho^{AB}) \ , \tag{A.8}$$

where the last step follows from the monotonicity of the purified distance under partial trace. Using again the triangle inequality for the purified distance, we then bound

$$\begin{aligned}
\mathrm{P}(\overline{\rho}^{AB}, \rho^{AB}) \quad &\leq \quad \mathrm{P}(\overline{\rho}^{AB}, \tilde{\rho}^{AB}) + \mathrm{P}(\tilde{\rho}^{AB}, \rho^{AB}) \\
&\leq \quad \mathrm{P}(\Pi^A_\gamma\tilde{\rho}^{AB}\Pi^A_\gamma, \tilde{\rho}^{AB}) + \epsilon \\
&\leq \quad \sqrt{2\cdot\mathrm{Tr}\left[(\mathbb{1}^A - \Pi^A_\gamma)\tilde{\rho}^{AB}\right]} + \epsilon \\
&\leq \quad \sqrt{2\cdot\frac{\delta^2}{8}} + \epsilon \quad = \quad \frac{\delta}{2} + \epsilon \ .
\end{aligned}$$

Together with Eq. (A.8), we conclude that $\mathrm{P}(\rho^{\hat{A}B}, \rho^{AB}) \leq 2\epsilon + \delta$. ∎