arXiv:1605.04601v2 [quant-ph] 31 Oct 2016

Lower bound on expected communication cost of quantum Huffman coding

Anurag Anshu
Centre for Quantum Technologies, National University of Singapore
a0109169@u.nus.edu

Ankit Garg
Princeton University
garg@cs.princeton.edu

Aram W. Harrow
Center for Theoretical Physics, Massachusetts Institute of Technology
aram@mit.edu

Penghui Yao
Institute for Quantum Computing and Department of Combinatorics and Optimization, University of Waterloo
phyao1985@gmail.com

November 1, 2016

Abstract

Data compression is a fundamental problem in quantum and classical information theory. A typical version of the problem is that the sender Alice receives a (classical or quantum) state from some known ensemble and needs to transmit it to the receiver Bob with average error below some specified bound. We consider the case in which the message can have a variable length and goal is to minimize its expected length.

For classical case this problem has a well-known solution given by the Huffman coding. In this scheme, expected length of the message is equal to the Shannon entropy of the source (with a constant additive factor) and the protocol succeeds with zero error. This is a single-shot scheme from which the asymptotic result, viz. Shannon's source coding theorem, can be recovered by encoding each input state sequentially.

For the quantum case, the asymptotic compression rate is given by the von-Neumann entropy. However, we show that there is no one-shot scheme which is able to match this rate, even if interactive communication is allowed. This is a relatively rare case in quantum information theory where the cost of a quantum task is significantly different from its classical analogue. Our result has implications for direct sum theorems in quantum communication complexity and one-shot formulations of Quantum Reverse Shannon theorem.

1 Introduction

The central theme of information theory is compression of messages up to their *information content*. The celebrated work of Shannon [Sha] initiated this idea by showing that in the asymptotic setting, compression could be achieved up to the *Shannon entropy* of the message source. Subsequently, it was shown by Huffman [Huf52] that by encoding each message into a codeword of different length based on the probability of the occurrence $p(x)$ of message x, one can construct a code whose expected length is at most $H(p)+1$, where $H(\cdot)$ is the Shannon entropy. This led to an operational interpretation of the Shannon entropy of a source in the *one-shot* setting.

The study of compression of messages in terms of *expected communication cost*, rather than *worst case communication cost* has been very fruitful in information theory, both in operational interpretation of fundamental quantities and in applications in communication complexity. In the work [HJMR10], the following task was considered (inspired by a result of Wyner [Wyn75]): Alice is given an input x with probability $p(x)$ and she needs to send a message to Bob so that Bob can output a y distributed according to $p(y|x)$. This is a joint sampling task of the probability distribution $p(x,y) \stackrel{\text{def}}{=} p(x)p(y|x)$. The authors showed that in the presence of shared randomness, the expected communication cost of jointly sampling $p(x,y)$ is upper and lower bounded by $\mathrm{I}(X:Y) + 2\log(\mathrm{I}(X:Y)) + \mathcal{O}(1)$ and $\mathrm{I}(X:Y)$, respectively. This served as a natural characterization of mutual information in a one-shot setting (different from the one already given by Shannon [Sha] in terms of channel capacity). Huffman coding can be seen as a special case of the above task by setting $p(y|x) = \delta_{y,x}$. This result also has applications in proving direct sum theorems for communication complexity. The direct sum problem asks whether computing N copies of a function (or a task in general) requires N times as much communication as computing a single copy. [HJMR10] used their compression result to prove the following theorem:

Theorem (Informal, [HJMR10])**.** *The minimum expected communication cost of an r-round protocol, w.r.t. N iid copies of a product distribution μ, required to compute N copies of a function $f(x,y)$ is at least $N \cdot (CC_r(f) - O(r))$, where $CC_r(f)$ is the minimum expected communication cost (w.r.t μ) of an r-round protocol required to compute a single copy of f.*

Message compression in the presence of side information was first studied in the asymptotic setting by Slepian and Wolf [SW73]. The work by Braverman and Rao [BR11] gave its one-shot analogue in the following manner. Given a probability distribution P known to Alice and Q known to Bob, they constructed an interactive protocol (assisted by shared randomness) that allowed both Alice and Bob to output a sample from distribution P' satisfying $\|P' - P\|_1 \leq \varepsilon$, with expected communication cost $\mathrm{D}(P\|Q) + \mathcal{O}(\sqrt{\mathrm{D}(P\|Q)}) + 2\log(\frac{1}{\varepsilon})$. Here $\mathrm{D}(P\|Q)$ is the relative entropy between P and Q. This work thus provided an operational and non-asymptotic interpretation to *relative entropy* [1] and extended the above theorem to general distributions. The holy grail for such direct sum theorems is to remove the dependence on the number of rounds, and the above mentioned results ([HJMR10],[BR11]) along with [BBCR10] are important steps in this direction.

The aforementioned discussion points to a generic principle: it is possible to compress communication protocols up to their *Information Cost* (formally introduced in [BR11, BBCR10], see also references therein) with the aid of shared randomness and consideration of expected communication cost as communication measure.

[1]The work [HJMR10] gives an operational interpretation of relative entropy as well, but for the task where Alice knows the distribution P and both Alice and Bob know the distribution Q.

On the other hand, while many of the above results have quantum counterparts, a similar principle for entanglement-assisted quantum communication protocols has not yet been well established, as we discuss now. Quantum communication protocols typically fall into two classes: non-coherent protocols and coherent protocols.

In the case of coherent quantum protocols, Alice and Bob share a tripartite quantum state with the Referee and their objective is to perform a task while maintaining quantum coherence with the Referee. An example of coherent quantum protocols is quantum state merging, introduced in [HOW07] as the quantum analogue of the Slepian-Wolf protocol [SW73] (in the asymptotic setting). The most general form of coherent quantum protocols, involving two parties and one Referee, is known as quantum state redistribution. It is defined as follows: Alice (AC), Bob (B) and Referee (R) share a pure quantum state Ψ_{RABC} and Alice needs to transfer the register C to Bob. This task was originally introduced in [DY08, YD09] to give an operational meaning of the quantum conditional mutual information in the asymptotic setting. Furthermore, as shown by Touchette [Tou15], it nicely captures interactive quantum communication protocols within the framework of quantum communication complexity and leads to a formulation of *quantum information complexity*.

Using the one-shot quantum protocols for quantum state redistribution developed in [BCT16], and the notion of quantum information complexity, Touchette [Tou15] obtains the following direct sum result for entanglement-assisted quantum communication complexity.

Theorem (Informal, [Tou15])**.** *The minimum worst-case quantum communication cost of an r-round quantum protocol required to compute N copies of a (classical) function $f(x, y)$ is at least $N \cdot (\frac{QCC_r(f)}{r^2} - O(r))$, where $QCC_r(f)$ is the worst-case communication cost of an r-round quantum protocol required to compute a single copy of f.*

The above result has a strong dependence on number of rounds (as opposed to a weaker dependence in the the direct sum result by [HJMR10]), that comes from the consideration of the worst-case quantum communication cost for the quantum state redistribution in the work [BCT16]. Furthermore, it has been shown recently in [Ans15] that the expected quantum communication cost of a protocol achieving quantum state redistribution cannot be substantially better than its worst-case quantum communication cost. This leads to a bottleneck in the improvement of the direct sum results for the quantum case within the framework of coherent quantum protocols.

In non-coherent protocols, Alice and Bob perform a task on their inputs without maintaining coherence with the Referee. The works which exhibit one-shot quantum compression protocols in the non-coherent setting include [JRS05, JRS08] (which also show direct sum theorems for entanglement-assisted one-way quantum communication complexity) and [AJM$^+$14] (which is an extension of the Braverman-Rao protocol [BR11] to the quantum domain). All of these results take into consideration only the worst-case quantum communication cost, and it is not clear if the expected communication cost of these message compression task can be substantially improved (to the information cost) over the worst-case cost.

In this work, we explore the possibility of having quantum protocols with better expected communication cost in the non-coherent framework. Towards this, we define the following *quantum Huffman task*.

Definition 1.1 (Quantum Huffman task)**.** Alice (A) receives an input x corresponding to a quantum pure state $|\Psi_x\rangle$ with probability $p(x)$. For a given $\eta > 0$, which we shall henceforth identify as the 'error parameter', Alice needs to transfer the state $|\Psi_x\rangle$ to Bob, such that the final state Φ_x with Bob satisfies $\sum_x p(x)\mathrm{F}^2(\Psi_x, \Phi_x) \geq 1 - \eta^2$. Here, $\mathrm{F}(\cdot, \cdot)$ is fidelity and η^2 is average error of the protocol.

The above task is a quantum version of classical one-shot source coding. The expected communication cost in the asymptotic setting is equal to $S(\sum_x p(x)\Psi_x)$ [Hol73, Sch95]. The main question that we address is whether there exists a communication protocol that achieves the above task with expected communication cost close to $S(\sum_x p(x)\Psi_x)$.

A prior work by Braunstein *et. al.* [BFGL98] had considered our question and had noted several issues in generalizing directly the techniques of classical Huffman coding to the quantum case. In present work, we show that no such compression scheme is possible.

Our results

We refer to the collection of pairs $\{(p(x), \Psi_x)\}_x$ as an *ensemble* of states and associated probabilities. Following the discussion in the introduction, we would like to compare the expected communication cost of any protocol achieving the quantum Huffman task with the von-Neumann entropy of Alice's average state: $S(\sum_x p(x) |\Psi_x\rangle\langle\Psi_x|)$. Our main result is a large gap between the two quantities, described below.

Theorem 1.2. *Fix a positive integer $d > 4$ and real δ that satisfies $\delta < \frac{1}{4}$. There exist a collection of $N \stackrel{\text{def}}{=} 8d^7$ states $\{|\Psi_x\rangle\}_{x=1}^N$ that depend on δ and belong to a d-dimensional Hilbert space, and a probability distribution $\{p(x)\}_{x=1}^N$, such that following holds for the ensemble $\{(p(x), \Psi_x)\}_{x=1}^N$.*

- *The von-Neumann entropy of the average state satisfies $S(\sum_x p(x)\Psi_x) \leq \delta\log(d)+H(\delta)+2$*
- *For any one-way protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta}{8}$, the expected communication cost is lower bounded by $(1 - \sqrt{\eta})\log(d\delta) - 6$.*
- *For any r-round protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta}{8}$, the expected communication cost is lower bounded by*

$$\Omega(\frac{\log(d\delta)}{(\log r)}).$$

This theorem is proved in Section 6, in the more formal version of Theorem 6.1.

For the interactive case, we also give a round-independent statement for small enough η.

Theorem 1.3. *Fix a positive integer $d > 4$ and a real δ that satisfies $\delta < \frac{1}{4}$. There exist a collection of $N \stackrel{\text{def}}{=} 8d^7$ states $\{|\Psi_x\rangle\}_{x=1}^N$ that depend on δ and belong to a d-dimensional Hilbert space, and a probability distribution $\{p(x)\}_{x=1}^N$ such that following holds for the ensemble $\{(p(x), \Psi_x)\}_{x=1}^N$.*

- *The von-Neumann entropy of the average state satisfies $S(\sum_x p(x)\Psi_x) \leq \delta\log(d) + H(\delta) + 2$.*
- *For any interactive protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta^2}{192}$, the expected communication cost is lower bounded by*

$$\Omega(\frac{\log(d\delta)}{(\log\log(d) - 2\log\eta)}).$$

This theorem is proved in Section 6,in the more formal version of Theorem 6.2.

Our techniques

Our proof follows in two main steps, which we illustrate here for the case of one-way protocols for simplicity. All the quantum states appearing below are assumed to belong to a Hilbert space of dimension d. We first show that for every message i sent from Alice to Bob, there exists a quantum state σ_i, such that the probability p_i of this message is upper bounded by $p_i \leq \sum_x p(x) 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_x \| \sigma_i)}$, where η is the error parameter and $\mathrm{D}^{\eta}_{\max}(\cdot\|\cdot)$ is smooth relative max-entropy. This upper bound crucially uses the fact that the quantum states Ψ_x are pure. Sections 3 (for one-way protocols) and 4 (for interactive protocols) are built upon this idea. Our aim now is to find an ensemble $\{p(x), \Psi_x\}$ for which the quantity $\sum_x p(x) 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_x \| \sigma_i)}$ is small, as a result of which the expected communication cost must be large.

Our second step is based upon the observation that given the quantum state σ_i (as mentioned above), and a pure state Ψ chosen according to Haar measure, the smooth relative max-entropy ($= \mathrm{D}^{\eta}_{\max}(\Psi \| \sigma_i)$) must attain a large value ($\approx \log(d)$) with high probability. This suggests that the ensemble $\{(p(x), \Psi_x)\}_x$ should be constructed by choosing vectors from Haar measure, making the quantity $\sum_x p(x) 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_x \| \sigma_i)}$ close to $\mathcal{O}(1) \cdot 2^{-\log(d)}$. This gives the upper bound $p_i \leq \frac{\mathcal{O}(1)}{d}$ and hence the expected communication cost is at least $\log(d) - \mathcal{O}(1)$. Unfortunately, this choice of ensemble makes the von-Neumann entropy of the average state $\sum_x p(x) \Psi_x$ equal to $\log(d)$, which is larger than our lower bound on the expected communication cost.

We remedy this problem by introducing a parameter δ and letting $|\Psi_x\rangle = \sqrt{1-\delta}\,|0\rangle + \sqrt{\delta}\,|x\rangle$, where $|0\rangle$ is some fixed vector and $|x\rangle$ belongs to the $d-1$ dimensional subspace orthogonal to $|0\rangle$. We choose $|x\rangle$ according to the Haar measure in the $d-1$ dimensional subspace and show that the smooth relative max entropy $\mathrm{D}^{\eta}_{\max}(\Psi_x \| \sigma)$ is still large ($\approx \log(d\delta)$ with high probability) as long as $\eta < \delta/16$. Interestingly, now the von-Neumann entropy of the average state $\sum_x p(x) \Psi_x$ is $\approx \delta \log(d)$, which is much smaller than expected communication cost. Details have been discussed in Section 5, where we also make the input size of the order of d^7, and hence the number of bits in input of the size $7 \log(d)$.

Application to one-shot Quantum Reverse Shannon theorem

Given a quantum channel $\mathcal{E} : A \to B$ that takes as input a state from Alice's side and outputs a state on Bob's side, the entanglement-assisted classical capacity of $\mathcal{E}$ captures the number of bits that can be sent (via encoding and decoding) through $\mathcal{E}$ in the setting of asymptotically many uses of the channel and vanishing error. The Quantum Reverse Shannon Theorem [BDH+14, BCR11] concerns the simulation of the channel: it states that using shared entanglement, Alice and Bob can simulate the action of n copies of $\mathcal{E}$ using a number of bits equal to n times the entanglement-assisted classical capacity of $\mathcal{E}$, in the limit of $n \to \infty$. For a classical channel the shared entanglement can be replaced by shared randomness [BSST02].

An interesting question is to study the Reverse Shannon theorem in the one-shot setting. It was shown in [HJMR10] that a classical channel E with classical capacity $\mathcal{C}(E)$ can be simulated using shared randomness and expected communication cost (for every input to channel) of $\mathcal{C}(E) + 2\log(\mathcal{C}(E) + 1) + \mathcal{O}(1)$. We shall consider the possibility of extending this result to the one-shot quantum setting.

A one-way protocol P for sending an m-bit classical message over a quantum channel $\mathcal{E} : A \to B$ is as follows. Alice and Bob share an entangled state $|\theta\rangle_{A'B'}$. Based on input $x \in \{1, 2 \ldots 2^m\}$, Alice applies a map $\mathcal{A}_x : A' \to A$ on her part of the shared entanglement and sends the resulting state through the channel $\mathcal{E}$. Upon receiving the output from the channel $\mathcal{E}$, Bob applies a decoding map $\mathcal{B}$ on his registers B, B' and outputs a distribution P_x over $\{1, 2 \ldots 2^m\}$. The error of the protocol, given input x, is $\sum_{y \neq x} P_x(y)$. The worst-case error of protocol is defined as $\eta_{\mathrm{P}} \stackrel{\mathrm{def}}{=} \max_x \sum_{y\neq x} P_x(y)$.

The one-shot η-error entanglement-assisted classical capacity of channel $\mathcal{E}$ is the largest m such that there exists a protocol P that sends n-bit classical message over $\mathcal{E}$ with worst-case error $\eta_{\mathrm{P}} \leq \eta$.

The simulation of a quantum channel $\mathcal{E}$ can be regarded as a converse to the transmission of a message using the channel. A general (possibly two-way) protocol $\mathcal{Q}$ for simulating a quantum channel $\mathcal{E}$ is as follows. Alice and Bob share an entangled pure state $|\theta\rangle_{A'B'}$. Alice receives a input ρ_A in register A. Then Alice and Bob perform an interactive quantum communication protocol (see Section 4) and Bob outputs a quantum state σ. The error of $\mathcal{Q}$ for input ρ is defined as $\sqrt{1-\mathrm{F}^2(\mathcal{E}(\rho),\sigma)}$. The error of the protocol is $\eta_{\mathcal{Q}} \stackrel{\mathrm{def}}{=} \max_\rho \sqrt{1-\mathrm{F}^2(\mathcal{E}(\rho),\sigma)}$. Given an input ρ, we shall consider the expected communication cost of the protocol $\mathcal{Q}$ running on input ρ and define the *simulation cost* of $\mathcal{Q}$ as the maximum expected communication cost over all inputs ρ.

We shall show the following negative result.

Theorem 1.4. *Fix a positive integer $d > 4$ and a real δ that satisfies $\delta < \frac{1}{4}$. There exists a register A with dimension $8d^7$ and classical-quantum channel $\mathcal{E} : A \to B$ with one-shot η-error entanglement assisted classical capacity upper bounded by*

$$\frac{\delta \log(d) + H(\delta) + H(\eta) + 2}{1-\eta},$$

such that for any protocol achieving the simulation of above channel with error at most η, following holds:

- *If the protocol is one-way and $\eta < \frac{\delta}{8}$, then simulation cost of the protocol is at least $(1-\sqrt{\eta})\log(d\delta) - 6$.*
- *If the protocol is interactive with r-rounds and $\eta < \frac{\delta}{8}$, the simulation cost is lower bounded by $\Omega(\frac{\log(d\delta)}{(\log r)})$.*
- *If $\eta < \frac{\delta^2}{192}$ and the protocol is interactive, the simulation cost of the protocol is lower bounded by*

$$\Omega(\frac{\log(d\delta)}{(\log\log(d) - 2\log\eta)}).$$

Proof of theorem is given in Section 7, in the more formal version of Theorem 7.2.

2 Preliminaries

In this section we present some notations, definitions, facts and lemmas that we will use in our proofs.

Information theory

For a natural number n, let $[n]$ represent the set $\{1, 2, \ldots, n\}$. For a set S, let $|S|$ be the size of S. A *tuple* is a finite collection of positive integers, such as $(i_1, i_2 \ldots i_r)$ for some finite r. We let log represent logarithm to the base 2 and ln represent logarithm to the base e. The ℓ_1 norm of an operator X is $\|X\|_1 \stackrel{\mathrm{def}}{=} \mathrm{Tr}\sqrt{X^\dagger X}$ and the ℓ_2 norm is $\|X\|_2 \stackrel{\mathrm{def}}{=} \sqrt{\mathrm{Tr} X X^\dagger}$. A quantum state (or just a state) is a positive semi-definite matrix with trace equal to 1. It is called *pure* if and only if the rank is 1. Let $|\psi\rangle$ be a unit vector. We use ψ to represent the state and also the density matrix $|\psi\rangle\langle\psi|$, associated with $|\psi\rangle$.

A sub-normalized state is a positive semi-definite matrix with trace less than or equal to 1. A *quantum register* A is associated with some Hilbert space $\mathcal{H}_A$. Define $|A| \stackrel{\mathrm{def}}{=} \dim(\mathcal{H}_A)$. We denote by $\mathcal{D}(A)$, the set of quantum states in the Hilbert space $\mathcal{H}_A$ and by $\mathcal{D}_{\leq}(A)$, the set of all sub-normalized states on register A. State ρ with subscript A indicates $\rho_A \in \mathcal{D}(A)$.

For two quantum states ρ and σ, $\rho\otimes\sigma$ represents the tensor product (Kronecker product) of ρ and σ. Composition of two registers A and B, denoted AB, is associated with Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. If two registers A, B are associated with the same Hilbert space, we shall denote it by $A \equiv B$. Let ρ_{AB} be a bipartite quantum state in registers AB. We define

$$\rho_B \stackrel{\mathrm{def}}{=} \mathrm{Tr}_A(\rho_{AB}) \stackrel{\mathrm{def}}{=} \sum_i (\langle i| \otimes \mathbb{1}_B)\rho_{AB}(|i\rangle \otimes \mathbb{1}_B),$$

where $\{|i\rangle\}_i$ is an orthonormal basis for the Hilbert space A and $\mathbb{1}_B$ is the identity matrix in space B. The state ρ_B is referred to as the marginal state of ρ_{AB} in register B. Unless otherwise stated, a missing register from subscript in a state will represent partial trace over that register. A quantum map $\mathcal{E} : A \to B$ is a completely positive and trace preserving (CPTP) linear map (mapping states from $\mathcal{D}(A)$ to states in $\mathcal{D}(B)$). A completely positive and trace non-increasing linear map $\tilde{\mathcal{E}} : A \to B$ maps quantum states to sub-normalized states. The identity operator in Hilbert space $\mathcal{H}_A$ (and associated register A) is denoted I_A. A *unitary* operator $U_A : \mathcal{H}_A \to \mathcal{H}_A$ is such that $U_A^\dagger U_A = U_A U_A^\dagger = I_A$. An *isometry* $V : \mathcal{H}_A \to \mathcal{H}_B$ is such that $V^\dagger V = I_A$. The set of all unitary operations on register A is denoted by $\mathcal{U}(A)$.

We denote a unit ball in space $\mathbb{R}^d$ as S^d. An element of S^d is a unit vector in $\mathbb{R}^d$. We shall represent an element $x \in S^d$ using the bra-ket notation as $|x\rangle$. Euclidean norm of $|x\rangle$ is $\| \, |x\rangle\langle x| \, \|_1$. Given two vectors $|x\rangle, |y\rangle \in S^d$, the *Euclidean distance* between them is $\|(|x\rangle - |y\rangle)(\langle x| - \langle y|)\|_1$.

Definition 2.1. We shall consider the following information theoretic quantities. Let $\varepsilon \geq 0$.

1. **generalized fidelity** For $\rho, \sigma \in \mathcal{D}_{\leq}(A)$,

$$\mathrm{F}(\rho, \sigma) \stackrel{\mathrm{def}}{=} \|\sqrt{\rho}\sqrt{\sigma}\|_1 + \sqrt{(1 - \mathrm{Tr}(\rho))(1 - \mathrm{Tr}(\sigma))}.$$

2. **purified distance** For $\rho, \sigma \in \mathcal{D}_{\leq}(A)$,

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

3. **ε-ball** For $\rho_A \in \mathcal{D}(A)$,

$$\mathcal{B}^\varepsilon(\rho_A) \stackrel{\mathrm{def}}{=} \{\rho'_A \in \mathcal{D}(A) | \ \mathrm{F}(\rho_A, \rho'_A) \geq 1 - \varepsilon\}.$$

4. **entropy** For $\rho_A \in \mathcal{D}(A)$,

$$\mathrm{H}(A)_\rho \stackrel{\mathrm{def}}{=} -\mathrm{Tr}(\rho_A \log \rho_A).$$

5. **relative entropy** For $\rho_A, \sigma_A \in \mathcal{D}(A)$,

$$\mathrm{D}(\rho_A \| \sigma_A) \stackrel{\mathrm{def}}{=} \mathrm{Tr}(\rho_A \log \rho_A) - \mathrm{Tr}(\rho_A \log \sigma_A).$$

6. **max-relative entropy** For $\rho_A, \sigma_A \in \mathcal{D}(A)$,

$$\mathrm{D}_{\max}(\rho_A \| \sigma_A) \stackrel{\mathrm{def}}{=} \inf\{\lambda \in \mathbb{R} : 2^\lambda \sigma_A \geq \rho_A\}.$$

7. **smooth max-relative entropy** For $\rho_A, \sigma_A \in \mathcal{D}(A)$,

$$\mathrm{D}^{\eta}_{\max}(\rho_A \| \sigma_A) \stackrel{\mathrm{def}}{=} \inf_{\rho'_A \in \mathcal{B}^{\eta}(\rho_A)} \mathrm{D}_{\max}(\rho'_A \| \sigma_A) \, .$$

8. **mutual information** For $\rho_{AB} \in \mathcal{D}(AB)$,

$$\mathrm{I}(A:B)_\rho \stackrel{\mathrm{def}}{=} \mathrm{D}(\rho_{AB} \| \rho_A \otimes \rho_B) = \mathrm{H}(A)_\rho + \mathrm{H}(B)_\rho - \mathrm{H}(AB)_\rho \, .$$

We will use the following facts.

Fact 2.2 (Monotonicity of quantum operations)**.** [[Lin75, BCF$^+$96], [Tom12], Theorem 3.4] For states ρ, $\sigma \in \mathcal{D}(A)$, and quantum map $\mathcal{E}(\cdot)$,

$$\|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1 \leq \|\rho - \sigma\|_1 , \mathrm{F}(\rho, \sigma) \leq \mathrm{F}(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \text{ and } \mathrm{D}_{\max}(\rho \| \sigma) \geq \mathrm{D}_{\max}(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) \, .$$

Fact 2.3 (Joint concavity of fidelity)**.** [[Wat11], Proposition 4.7] Given quantum states $\rho_1, \rho_2 \ldots \rho_k, \sigma_1, \sigma_2 \ldots \sigma_k \in \mathcal{D}(A)$ and positive numbers $p_1, p_2 \ldots p_k$ such that $\sum_i p_i = 1$. Then

$$\mathrm{F}(\sum_i p_i \rho_i, \sum_i p_i \sigma_i) \geq \sum_i p_i \mathrm{F}(\rho_i, \sigma_i).$$

Fact 2.4 (Fannes inequality)**.** [[Fan73]] Given quantum states $\rho_1, \rho_2 \in \mathcal{D}(A)$, such that $|A| = d$ and $\mathrm{P}(\rho_1, \rho_2) = \varepsilon \leq \frac{1}{2\mathrm{e}}$,

$$|S(\rho_1) - S(\rho_2)| \leq \varepsilon \log(d) + 1.$$

Fact 2.5 (Matrix Hoeffding Bound)**.** [[Tro12]] Let $Z_1, Z_2 \ldots Z_r$ be independent and identically distributed random $d \times d$ Hermitian matrices with $\mathbb{E}(Z_i) = 0$ and $\|Z_i\|_\infty \leq \lambda$. Then

$$\mathrm{Prob}(\|\frac{1}{r}\sum_i Z_i\|_\infty \geq \varepsilon) \leq d \cdot e^{-\frac{n\varepsilon^2}{8\lambda}}$$

and

$$\mathrm{Prob}(\|\frac{1}{r}\sum_i Z_i\|_1 \geq \varepsilon) \leq d \cdot e^{-\frac{n\varepsilon^2}{8d^2\lambda}}.$$

3 One-way communication

A one-way quantum communication protocol P for quantum Huffman coding with error η^2 is described as follows.

Input: Alice gets an input x with probability $p(x)$ and she needs to send the state $|\Psi_x\rangle$ to Bob.

Pre-shared entanglement: They have a pre-shared entanglement $|\theta\rangle_{AB}$.

- Conditioned on the input x, Alice applies a measurement $\{M^x_1, M^x_2 \ldots\}$ on her side and sends the outcome i to Bob. Let

$$p^x_i \stackrel{\mathrm{def}}{=} \mathrm{Tr}(M^x_i \theta_A), \quad \rho^x_i \stackrel{\mathrm{def}}{=} \frac{\mathrm{Tr}_A(M^x_i \theta_{AB})}{p^x_i}.$$

- Receiving message i from Alice, Bob applies a quantum channel $\mathcal{E}_i$ based on the message i, to obtain a state σ_i^x in his output register.
- The final state in the output register is $\sum_i p_i^x \sigma_i^x$ and it follows that

$$\sum_x p(x) \sum_i p_i^x \langle \Psi_x | \sigma_i^x | \Psi_x \rangle > 1 - \eta^2$$

due to correctness of protocol.

The expected communication cost of P is $\sum_x p(x) \sum_i p_i^x \lceil \log(i) \rceil$ which can be lower bounded by $\sum_x p(x) \sum_i p_i^x \log(i)$. Since we are interested in lower bounding the expected communication cost, we shall consider the latter quantity.

Define the quantity $t_i \stackrel{\text{def}}{=} \sum_x p(x) 2^{-\mathrm{D}_{\max}^{\eta}(\Psi^x \| \mathcal{E}_i(\theta_B))}$.

We have the following lemma.

Lemma 3.1. *Let a be the largest integer such that $t_i \leq 2^{-a}$ for all i. Then expected communication cost of* P *is lower bounded by $a(1-\sqrt{\eta})^2 - 1$.*

Proof. Our proof shall proceed in the steps outlined below.

1. **Pruning away x with low fidelity:**

   Let $\mathcal{G}$ be the set of all x such that $\sum_i p_i^x \langle \Psi_x | \sigma_i^x | \Psi_x \rangle \geq 1 - \eta^{3/2}$. Let $\mathcal{B}$ be the set of rest of x. Then we have that $\sum_{x \in \mathcal{G}} p(x) \geq 1 - \sqrt{\eta}$ and equivalently $\sum_{x \in \mathcal{B}} p(x) \leq \sqrt{\eta}$.

   Define a new probability distribution $p'(x)$ which is 0 whenever $x \in \mathcal{B}$ and equal to $\frac{p(x)}{\sum_{x \in \mathcal{G}} p(x)}$ for $x \in \mathcal{G}$. Since $\sum_{x \in \mathcal{G}} p(x) \geq 1 - \sqrt{\eta}$, it holds that $p'(x) \leq \frac{p(x)}{1-\sqrt{\eta}}$ for all x.

2. **Upper bound on probabilities p_i^x:**

   We upper bound the probabilities p_i^x in the following way. Consider,

$$\theta_B = \mathrm{Tr}_A(M_i^x \theta_{AB}) + \mathrm{Tr}_A((\mathrm{I} - M_i^x)\theta_{AB}) > \mathrm{Tr}_A(M_i^x \theta_{AB}).$$

   Thus,

$$p_i^x \rho_i^x < \theta_B \implies \rho_i^x < \frac{1}{p_i^x} \theta_B.$$

   By definition of max-entropy, this means $2^{\mathrm{D}_{\max}(\rho_i^x \| \theta_B)} < \frac{1}{p_i^x}$. Now we use monotonicity of max-entropy under quantum operations (Fact 2.2), to obtain

$$p_i^x < 2^{-\mathrm{D}_{\max}(\rho_i^x \| \theta_B)} < 2^{-\mathrm{D}_{\max}(\sigma_i^x \| \mathcal{E}_i(\theta_B))}. \tag{1}$$

3. **Upper bound on probability of each message:**

   For every $x \in \mathcal{G}$, let $\mathcal{B}_x$ be set of i such that $\langle \Psi_x | \sigma_i^x | \Psi_x \rangle < 1 - \eta$. Let $\mathcal{G}_x$ be rest of the indices. Using the relation

$$\sum_i p_i^x (1 - \langle \Psi_x | \sigma_i^x | \Psi_x \rangle) < \eta^{3/2},$$

   we obtain that $\sum_{i \in \mathcal{B}_x} p_i^x < \sqrt{\eta}$. Define a new probability distribution q_i^x which is 0 whenever $i \in \mathcal{B}_x$ and equal to $\frac{p_i^x}{\sum_{i \in \mathcal{G}_x}}$ otherwise.

Define $s_i \stackrel{\text{def}}{=} \sum_x p'(x)q_i^x$. Note that by definition, $\mathrm{D}^{\eta}_{\max}(\Psi^x \| \mathcal{E}_i(\theta_B)) < \mathrm{D}_{\max}(\sigma_i^x \| \mathcal{E}_i(\theta_B))$ for all $i \in \mathcal{G}_x$. Using Equation 1, we observe that for all $x \in \mathcal{G}$ it holds that

$$q_i^x < \frac{1}{1-\sqrt{\eta}} 2^{-\mathrm{D}^{\eta}_{\max}(\Psi^x \| \mathcal{E}_i(\theta_B))}.$$

This implies

$$\begin{aligned} s_i &= \sum_x p'(x) q_i^x \\ &\leq \frac{1}{1-\sqrt{\eta}} \sum_x p'(x) 2^{-\mathrm{D}^{\eta}_{\max}(\Psi^x \| \mathcal{E}_i(\theta_B))} \\ &\leq \frac{1}{(1-\sqrt{\eta})^2} \sum_x p(x) 2^{-\mathrm{D}^{\eta}_{\max}(\Psi^x \| \mathcal{E}_i(\theta_B))} \\ &= \frac{t_i}{(1-\sqrt{\eta})^2} < \frac{2^{-a}}{(1-\sqrt{\eta})^2} \end{aligned} \tag{2}$$

where in first inequality, we have used the fact that for $x \in \mathcal{B}$, $p'(x) = 0$.

4. **Lower bound on expected communication:**

Since $p_i^x > (1-\sqrt{\eta})q_i^x$ for all pair (x, i) such that $x \in \mathcal{G}$, the expected communication cost is lower bounded by

$$\sum_x p(x) \sum_i p_i^x \log(i) > (1-\sqrt{\eta}) \sum_{x \in \mathcal{G}} p(x) \sum_i q_i^x \log(i) > (1-\sqrt{\eta})^2 \sum_x p'(x) \sum_i q_i^x \log(i).$$

From Equation 2, we have $s_i \leq \frac{2^{-a}}{(1-\eta)^2}$ and $\sum_i s_i = 1$. Thus, the quantity $\sum_i s_i \log(i)$ is minimized if $s_i = \frac{2^{-a}}{(1-\sqrt{\eta})^2}$ for all $i \leq 2^a(1-\sqrt{\eta})^2$. This gives following lower bound on expected communication cost

$$(1-\sqrt{\eta})^2 \cdot \frac{2^{-a}}{(1-\sqrt{\eta})^2} 2^a (1-\sqrt{\eta})^2 \log(2^a(1-\sqrt{\eta})^2/e) > (1-\sqrt{\eta})^2 \cdot a - 1.$$

□

4 Interactive communication

In this section, we extend the formalism of previous section to interactive communication setting.

Let an input x be given to Alice with probability $p(x)$. A r-round interactive protocol $\mathcal{P}$ (where r is an odd number) with error η^2 and expected communication cost C is described below. Since the protocol is entanglement assisted, it has been assumed that Alice and Bob only use classical communication (through quantum teleportation). For the convenience of notations, we drop the label x from every operation of Alice, but it is implicit that Alice's operations depend on x. Please note that Bob's operations do not depend on x.

Input: Quantum state $|\Psi_x\rangle$ with probability $p(x)$ and error parameter $\eta < 1$.

Shared entanglement: $|\theta\rangle_{AB}$.

- Alice performs a measurement $\mathcal{M} = \{M_A^1, M_A^2 \ldots\}$. The probability of outcome i_1 is

$p^x_{i_1} \stackrel{\text{def}}{=} \text{Tr}(M_A^{i_1}\theta_A)$. Let ϕ^{x,i_1}_{AB} be the global normalized quantum state, conditioned on this outcome. She sends message i_1 to Bob.

- Upon receiving the message i_1 from Alice, Bob performs a measurement
$$\mathcal{M}^{i_1} = \{M_B^{1,i_1}, M_B^{2,i_1}\ldots\}.$$
The probability of outcome i_2 is $p^x_{i_2|i_1} \stackrel{\text{def}}{=} \text{Tr}(M_B^{i_2,i_1}\phi_B^{x,i_1})$. Let ϕ^{x,i_2,i_1}_{AB} be the global normalized quantum state conditioned on this outcome i_2 and previous outcome i_1. Bob sends message i_2 to Alice.

- Consider any odd round $1 < k \leq r$. Let the measurement outcomes in the past rounds be $i_1, i_2 \ldots i_{k-1}$ and the global normalized state be $\phi^{x,i_{k-1},i_{k-2}\ldots i_1}_{AB}$. Alice performs the measurement $\mathcal{M}^{i_{k-1},i_{k-2}\ldots i_2,i_1} = \{M_A^{1,i_{k-1},i_{k-2}\ldots i_2,i_1}, M_A^{2,i_{k-1},i_{k-2}\ldots i_2,i_1}\ldots\}$ and obtains outcome i_k with probability
$$p^x_{i_k|i_{k-1},i_{k-2}\ldots i_2,i_1} \stackrel{\text{def}}{=} \text{Tr}(M_A^{i_k,i_{k-1},i_{k-2}\ldots i_2,i_1}\phi_A^{x,i_{k-1},i_{k-2}\ldots i_1}).$$
Let the global normalized state after outcome i_k be $\phi^{x,i_k,i_{k-1},i_{k-2}\ldots i_1}_{AB}$. Alice sends the outcome i_k to Bob.

- Consider an even round $2 < k \leq r$. Let the measurement outcomes in the past rounds be $i_1, i_2 \ldots i_{k-1}$ and global normalized state be $\phi^{x,i_{k-1},i_{k-2}\ldots i_1}_{AB}$. Bob performs the measurement
$$\mathcal{M}^{i_{k-1},i_{k-2}\ldots i_2,i_1} = \{M_B^{1,i_{k-1},i_{k-2}\ldots i_2,i_1}, M_B^{2,i_{k-1},i_{k-2}\ldots i_2,i_1}\ldots\}$$
and obtains outcome i_k with probability
$$p^x_{i_k|i_{k-1},i_{k-2}\ldots i_2,i_1} \stackrel{\text{def}}{=} \text{Tr}(M_B^{i_k,i_{k-1},i_{k-2}\ldots i_2,i_1}\phi_B^{x,i_{k-1},i_{k-2}\ldots i_1}).$$
Let the global normalized state after outcome i_k be $\phi^{x,i_k,i_{k-1},i_{k-2}\ldots i_1}_{AB}$. Bob sends the outcome i_k to Alice.

- After receiving message i_r from Alice at the end of round r, Bob applies a unitary $U_{i_r,i_{r-1}\ldots i_1} : B \to B'C$ such that $B \equiv B'C$. Define
$$\left|\tau^{x,i_r,i_{r-1}\ldots i_1}\right\rangle_{AB'C} \stackrel{\text{def}}{=} U_{i_r,i_{r-1}\ldots i_1}\left|\phi^{x,i_r,i_{r-1}\ldots i_1}\right\rangle_{AB}.$$

- For every $k \leq r$, define
$$p^x_{i_1,i_2\ldots i_k} \stackrel{\text{def}}{=} p^x_{i_1}\cdot p^x_{i_2|i_1}\cdot p^x_{i_3|i_2,i_1}\cdots p^x_{i_k|i_{k-1},i_{k-2}\ldots i_1}.$$
The joint state in register C, after Bob's final unitary and averaged over all messages is $\Phi^x_C \stackrel{\text{def}}{=} \sum_{i_r,i_{r-1}\ldots i_1} p^x_{i_1,i_2\ldots i_r}\tau_C^{x,i_r,i_{r-1}\ldots i_1}$. It satisfies $\sum_x p(x)\text{F}^2(\Phi^x_C,\Psi_x) \geq 1-\eta^2$.

Remark 4.1. In above protocol, it is possible that there exists a k such that the POVM $\mathcal{M}^{i_{k-1},i_{k-2}\ldots i_1}$ has just one outcome, which means that it is a local unitary operation. But then this local operation can be postponed to take place after the next message by other party, as local operations of Alice and Bob commute. Thus, we assume without loss of generality that if either Alice or Bob do not abort the protocol in some round, then they send at least one bit of message to the other party.

Remark 4.2. In our description of interactive protocol, it has not been stated what happens if either Alice or Bob abort the protocol. If Alice aborts a protocol, Bob can learn of this event as he does not receive any message from Alice, and vice-versa. We accommodate the 'abort' event in our framework in the following way. Given an x, let $(i_1, i_2 \dots i_k)$ be a sequence of messages after which Alice aborts (if k is odd) or Bob aborts (if k is even). Then, we define $p^x_{i_r,i_{r-1}\dots i_1} = p^x_{i_k,i_{k-1},\dots i_1}$ iff $(i_r, i_{r-1}, \dots i_1) = (1, 1, \dots 1, i_k, i_{k-1} \dots i_1)$ and 0 otherwise. We define $\phi^{x,i_r,i_{r-1}\dots i_1}_{AB} = \phi^{x,i_k,i_{k-1},\dots i_1}_{AB}$ iff $(i_r, i_{r-1}, \dots i_1) = (1, 1, \dots 1, i_k, i_{k-1} \dots i_1)$ and $|0\rangle_{AB}$ otherwise. That is, any sequence of messages $(i_k, i_{k-1}, \dots i_1)$ that leads to abort automatically extends to $(1, 1, \dots i_k, i_{k-1}, \dots i_1)$. Below, we shall find that this convention naturally captures the expected communication cost of the protocol.

The following lemma lower bounds the expected communication cost of the protocols achieving the task.

Lemma 4.3. *The expected communication cost of $\mathcal{P}$ is lower bounded by*

$$\sum_x p(x) \sum_{i_1,i_2\dots i_r} p^x_{i_1,i_2\dots i_r} \log(i_1 \cdot i_2 \dots i_r)$$

Proof. The expected communication cost is the expected length of the messages over all probability outcomes. Fix an x. If message sent by either party is i, then number of bits sent is $\lceil \log(i) \rceil > \log(i)$. If there is an abort, then number of bits sent is 0, but we have set $i = 1$ for the messages exchanged after such an event (following Remark 4.2). For this, $\log(i) = 0$. Hence the expected communication cost is lower bounded by

$$\sum_x p(x) \Big(\sum_{i_1} p^x_{i_1} \log(i_1) + \sum_{i_1,i_2} p^x_{i_1} p^x_{i_2|i_1} \log(i_2) + \dots \sum_{i_1,i_2\dots i_r} p^x_{i_1,i_2\dots i_{r-1}} p^x_{i_r|i_{r-1},i_{r-2}\dots i_1} \log(i_r) \Big)$$

$$= \sum_x p(x) \sum_{i_1,i_2\dots i_r} p^x_{i_1,i_2\dots i_r} (\log(i_1) + \log(i_1) + \dots \log(i_r)).$$

This proves the fact. ☐

Now, we shall provide upper bounds on probabilities $p^x_{i_1,i_2\dots i_r}$. Before proceeding to the main statement, we illustrate with the simplest case $r = 3$.

Example with $r = 3$: Consider particular measurement outcomes (i, j, k), where i is first outcome is Alice, j is outcome of Bob given message i from Alice and k is last outcome of Alice given message j from Bob. The post selected state corresponding to it is $\left|\phi^{x,k,j,i}\right\rangle = \frac{M^{k,j,i}_A M^{j,i}_B M^i_A |\theta\rangle_{AB}}{\sqrt{p^x_{i,j,k}}}$. Now consider the following chain of Lowner inequalities, recalling that only the measurements by Alice have dependence on x.

$$
\begin{aligned}
\mathrm{Tr}_A(\phi^{x,k,j,i}) &= \frac{\mathrm{Tr}_A(M_A^{k,j,i} M_B^{j,i} M_A^i \theta_{AB} M_A^{i\dagger} M_B^{j,i\dagger} M_A^{k,j,i\dagger})}{p^x_{k,j,i}} \\
&< \frac{\sum_k \mathrm{Tr}_A(M_A^{k,j,i} M_B^{j,i} M_A^i \theta_{AB} M_A^{i\dagger} M_B^{j,i\dagger} M_A^{k,j,i\dagger})}{p^x_{k,j,i}} \quad \text{(Adding positive matrices)} \\
&= \frac{\sum_k \mathrm{Tr}_A(M_A^{k,j,i\dagger} M_A^{k,j,i} M_B^{j,i} M_A^i \theta_{AB} M_A^{i\dagger} M_B^{j,i\dagger})}{p^x_{k,j,i}} \quad \text{(Partial trace is cyclic on Alice's side)} \\
&= \frac{\mathrm{Tr}_A(M_B^{j,i} M_A^i \theta_{AB} M_A^{i\dagger} M_B^{j,i\dagger})}{p^x_{k,j,i}} \quad (\{M^{k,j,i}\}_k \text{ form POVM for fixed } (i,j)) \\
&< \frac{\sum_l \mathrm{Tr}_A(M_B^{j,i} M_A^l \theta_{AB} M_A^{l\dagger} M_B^{j,i\dagger})}{p^x_{k,j,i}} \quad \text{(Adding positive matrices indexed by } l\text{. Includes } l = i) \\
&= \frac{\sum_l \mathrm{Tr}_A(M_B^{j,i} M_A^{l\dagger} M_A^l \theta_{AB} M_B^{j,i\dagger})}{p^x_{k,j,i}} \quad \text{(operators of Alice and Bob commute)} \\
&= \frac{\mathrm{Tr}_A(M_B^{j,i} \theta_{AB} M_B^{j,i\dagger})}{p^x_{k,j,i}} = \frac{M_B^{j,i} \theta_B M_B^{j,i\dagger}}{p_{k,j,i}} < \frac{\sum_j M_B^{j,i} \theta_B M_B^{j,i\dagger}}{p^x_{k,j,i}}
\end{aligned}
$$

Noting that $\sigma^i \stackrel{\text{def}}{=} \sum_j M_B^{j,i} \theta_B M_B^{j,i\dagger}$ is a quantum state independent of x, we get $p^x_{k,j,i} \leq 2^{-\mathrm{D}_{\max}\left(\mathrm{Tr}_A(\phi^{x,k,j,i}) \middle\| \sigma^i\right)}$. Now, Bob performs unitary $U_{k,j,i}$ depending on messages i, j, k on his registers $B'C$, and traces out B' to obtain the state $\tau_C^{x,k,j,i}$. We then have $p^x_{k,j,i} \leq 2^{-\mathrm{D}_{\max}\left(\tau_C^{x,k,j,i} \middle\| \omega^{k,j,i}\right)}$, where the state $\omega^{k,j,i} \stackrel{\text{def}}{=} \mathrm{Tr}_{B'}(U_{k,j,i} \sigma^i U^\dagger_{k,j,i})$ is independent of x.

Now we formally extend the argument to general r as follows, proof of which is similar to above example (for $r = 3$) and is deferred to Appendix A

Lemma 4.4. *For each $(i_1, i_2 \ldots i_r)$, there exists a quantum state $\omega^{i_r, i_{r-1} \ldots i_1}$ independent of x, such that $p^x_{i_r, i_{r-1} \ldots i_1} < 2^{-\mathrm{D}_{\max}\left(\tau_C^{x, i_r, i_{r-1} \ldots i_1} \middle\| \omega^{i_r, i_{r-1} \ldots i_1}\right)}$.*

We are now in a position to prove the following lemma. The proof follows outline similar to that of Lemma 3.1 and has been proved in Appendix B.

Lemma 4.5. *For each $(i_1, i_2 \ldots i_r)$, define $t_{i_r, i_{r-1} \ldots i_1} \stackrel{\text{def}}{=} \sum_x p(x) 2^{-\mathrm{D}^\eta_{\max}\left(\Psi_x \middle\| \omega^{i_r, i_{r-1} \ldots i_1}\right)}$. Let a be the largest integer such that $t_{i_r, i_{r-1} \ldots i_1} \leq 2^{-a}$ for all $(i_1, i_2 \ldots i_r)$. Then expected communication cost of the protocol is lower bounded by $(1 - \sqrt{\eta})^2 \cdot \frac{a + 2\log(1-\sqrt{\eta})}{2(\log r + 4)}$.*

5 Example separating expected communication and information

5.1 Our construction

Given the vector space $\mathbb{R}^d$ and associated unit ball S^d, we embedd it in $\mathbb{R}^{d+1}$ and let P denote the projector onto the original space $\mathbb{R}^d$. Let $|0\rangle$ be the unit vector satisfying $P|0\rangle = 0$. From the uniform measure μ over S^d, we draw m samples $\{|x_1\rangle, |x_2\rangle \ldots |x_m\rangle\}$ independently from μ, where m is to be chosen later. Let $\delta > 0$ be a parameter less than $1/4$. For each i, define the following random hermitian matrices

$$
Z_i^1 \stackrel{\text{def}}{=} |x_i\rangle\langle x_i|, \quad Z_i^2 \stackrel{\text{def}}{=} \sqrt{\delta - \delta^2}(|x_i\rangle\langle 0| + |0\rangle\langle x_i|) + \delta\,|x_i\rangle\langle x_i|, \quad Z_i^3 \stackrel{\text{def}}{=} |x_i\rangle\langle x_i| \otimes |x_i\rangle\langle x_i|.
$$

We have that $\|Z_i^1\|_\infty \leq 1, \|Z_i^2\|_\infty \leq 2\sqrt{\delta} < 1, \|Z_i^3\|_\infty \leq 1$. Furthermore, it holds that

$$\begin{aligned}
\mathbb{E}(Z_i^1) &= \frac{P}{d}, \\
\mathbb{E}(Z_i^2) &= \delta\frac{P}{d}, \\
\mathbb{E}(Z_i^3) &= \frac{P \otimes P + F}{d(d+1)}
\end{aligned}$$

where F is the swap operator on $\mathbb{R}^d \times \mathbb{R}^d$

From Matrix Hoeffding bound (Fact 2.5), following three relations hold for any $\varepsilon > 0$:

$$\begin{aligned}
\Pr(\|\frac{\sum_i Z_i^1}{m} - \frac{P}{d}\|_1 \geq \varepsilon) &\leq d \cdot e^{-\frac{m\varepsilon^2}{8 \cdot d^2}}, \\
\Pr(\|\frac{\sum_i Z_i^2}{m} - \delta\frac{P}{d}\|_1 \geq \varepsilon) &\leq d \cdot e^{-\frac{m\varepsilon^2}{8 \cdot d^2}}, \\
\Pr(\|\frac{\sum_i Z_i^3}{m} - \frac{P \otimes P + F}{d(d+1)}\|_1 \geq \varepsilon) &\leq d \cdot e^{-\frac{m\varepsilon^2}{8 \cdot d^4}},
\end{aligned}$$

Setting $m = \frac{8d^5}{\varepsilon^2}$, we find that for $d > 4$, all the upper bounds are less than $1/3$. Thus there exists a set of vectors $\{|x_1\rangle, |x_2\rangle \dots |x_m\rangle\}$, which we label with $\mathcal{N}_\varepsilon$, such that all three relations hold

$$\begin{aligned}
&\|\frac{\sum_i Z_i^1}{m} - \frac{P}{d}\|_1 \leq \varepsilon, \\
&\|\frac{\sum_i Z_i^2}{m} - \delta\frac{P}{d}\|_1 \leq \varepsilon, \\
&\|\frac{\sum_i Z_i^3}{m} - \frac{P \otimes P + F}{d(d+1)}\|_1 \leq \varepsilon
\end{aligned} \tag{3}$$

Our construction now proceeds as follows, recalling the quantum Huffman task in Definition 1.1. With δ as defined above, Alice is given the input i with probability $\lambda(i) \stackrel{\text{def}}{=} \frac{1}{m}$, which corresponds to the vector $|x_i\rangle \in \mathcal{N}_\varepsilon$. We define $|\Psi_i\rangle \stackrel{\text{def}}{=} \sqrt{1-\delta}\,|0\rangle + \sqrt{\delta}\,|x_i\rangle$.

We have the following lemma.

Lemma 5.1. *The von-Neumann entropy of the average state* $\mathbb{E}_i\Psi_i = \sum_i \lambda(i)\Psi_i$ *satisfies* $S(\mathbb{E}_i\Psi_i) \leq (\delta + \varepsilon)\log(d) + H(\delta) + 1$.

Proof. Consider,

$$\mathbb{E}_i |\Psi_i\rangle\langle\Psi_i| = (1-\delta)|0\rangle\langle 0| + \mathbb{E}_i(\sqrt{\delta(1-\delta)}(|0\rangle\langle x_i| + |x_i\rangle\langle 0|) + \delta|x_i\rangle\langle x_i|).$$

From Equation 3, it follows that

$$\|\mathbb{E}_i\Psi_i - (1-\delta)|0\rangle\langle 0| - \delta\frac{P}{d}\|_1 \leq \varepsilon.$$

Now we use Fannes inequality (Fact 2.4) to conclude that

$$S(\mathbb{E}_i\Psi_i) \leq S((1-\delta)|0\rangle\langle 0| + \delta\frac{P}{d}) + \varepsilon\log(d) + 1 = (\delta+\varepsilon)\log(d) + H(\delta) + 1.$$

□

5.2 A property of smooth relative max entropy

Following lower bound on smooth relative entropy shall be crucial for our argument.

Lemma 5.2. *Let σ be any quantum state belonging to C^{d+1}. Let $k < d$ be an integer and Q^- (Q^+) be projector onto subspace where σ has eigenvalues less than (greater than) $\frac{1}{k}$. For any i and $\eta > 0$ such that $\langle\Psi_i| Q^- |\Psi_i\rangle > 2\eta$, it holds that*

$$2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} \leq \frac{1}{k(1-\eta)(\sqrt{(1-\eta)\langle\Psi_i| Q^- |\Psi_i\rangle} - \sqrt{\langle\Psi_i| Q^+ |\Psi_i\rangle\,\eta})^2}.$$

Proof. Since $\dim(Q^+) \leq k$, it holds that $\dim(Q^-) \geq d + 1 - k$. Define the quantity

$$S^{\eta}(\Psi_i||Q^-) \stackrel{\mathrm{def}}{=} \inf_{|\lambda\rangle:|\langle\lambda|\Psi_i\rangle|^2>1-\eta} \langle\lambda| Q^- |\lambda\rangle .$$

The lemma follows from the following two claims, which have been proved in Appendix C.

Claim 5.3. For any i, it holds that

$$2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} < \frac{1}{k(1-\eta)S^{2\eta}(\Psi_i||Q^-)}.$$

We now calculate an explicit expression for $S^{\eta}(\Psi_i||Q^-)$ in the following claim.

Claim 5.4. If $\langle\Psi_i| Q^- |\Psi_i\rangle > \eta$, then we have

$$S^{\eta}(\Psi_i||Q^-) = (\sqrt{(1-\eta)\langle\Psi_i| Q^- |\Psi_i\rangle} - \sqrt{\langle\Psi_i| Q^+ |\Psi_i\rangle\,\eta})^2.$$

Else $S^{\eta}(\Psi_i||Q^-) = 0$.

Combining the two claims, our lemma follows. □

5.3 Final lower bound

Let Q^- be the projector onto an arbitrary subspace of dimension $d - k + 1$. Consider the vector $Q^- |0\rangle$, if it is non-zero. It is possible to find a basis of $d - k$ vectors in the subspace corresponding to Q^-, all of which are orthogonal to $Q^- |0\rangle$. This implies that these vectors are orthogonal to $|0\rangle$. Even if $Q^- |0\rangle$ is a zero vector, we can find such a basis of $d - k$ vectors in the subspace corresponding to Q^-, all of which are orthogonal to $|0\rangle$. Let Q be the subspace formed by these vectors. We have that $\dim(Q) = d - k$.

We show the following two claims.

Claim 5.5. Let i be drawn according to $\lambda(i)$. Then

$$\frac{\delta(d-k)}{d} + \delta\varepsilon > \mathbb{E}_{i\leftarrow\lambda}\langle\Psi_i| Q |\Psi_i\rangle > \frac{\delta(d-k)}{d} - \delta\varepsilon.$$

Proof. Since Q is orthogonal to $|0\rangle$, we have that

$$\mathbb{E}_{i\leftarrow\lambda}\langle\Psi_i| Q |\Psi_i\rangle = \delta\mathbb{E}_{i\leftarrow\lambda}\langle x_i| Q |x_i\rangle .$$

Using Equation 3, we find that

$$\frac{\dim(Q)}{d} + \varepsilon = \mathrm{Tr}(Q\frac{P}{d}) + \varepsilon > \mathbb{E}_{i\leftarrow\lambda}\langle x_i| Q |x_i\rangle > \mathrm{Tr}(Q\frac{P}{d}) - \varepsilon = \frac{\dim(Q)}{d} - \varepsilon.$$

Using the value of $\dim(Q)$, the claim follows.

□

Claim 5.6. Let i be drawn according to $\lambda(i)$. Then

$$\delta^2\frac{(d-k)(d-k+1)}{d(d+1)} + \delta^2\varepsilon > \mathbb{E}_{i\leftarrow\lambda}(\langle\Psi_i|\, Q\,|\Psi_i\rangle)^2 > \delta^2\frac{(d-k)(d-k+1)}{d(d+1)} - \delta^2\varepsilon.$$

Proof. Since Q is orthogonal to $|0\rangle$, we have that

$$\mathbb{E}_{i\leftarrow\lambda}(\langle\Psi_i|\, Q\,|\Psi_i\rangle)^2 = \delta^2\mathbb{E}_{i\leftarrow\lambda}(\langle x_i|\, Q\,|x_i\rangle)^2 = \delta^2\mathbb{E}_{i\leftarrow\lambda}\mathrm{Tr}((Q\otimes Q)(|x_i\rangle\langle x_i|\otimes|x_i\rangle\langle x_i|)).$$

Using Equation 3, we find that

$$\frac{(\dim(Q))^2+\dim(Q)}{d(d+1)} + \varepsilon = \mathrm{Tr}((Q\otimes Q)\frac{P\otimes P+F}{d(d+1)}) + \varepsilon > \mathbb{E}_{i\leftarrow\lambda}\mathrm{Tr}((Q\otimes Q)(|x_i\rangle\langle x_i|\otimes|x_i\rangle\langle x_i|))$$

$$> \mathrm{Tr}((Q\otimes Q)\frac{P\otimes P+F}{d(d+1)}) - \varepsilon = \frac{(\dim(Q))^2+\dim(Q)}{d(d+1)} - \varepsilon.$$

Using the value of $\dim(Q)$, the claim follows.

□

These claims together imply the following lemma.

Lemma 5.7. *For every $\alpha > 0$, it holds that*

$$Pr_\lambda(\langle\Psi_i|\, Q\,|\Psi_i\rangle < \delta\frac{d-k}{d} + \delta\varepsilon - \alpha) \le \frac{\delta^2}{\alpha^2}(3\varepsilon + \frac{3}{d}).$$

Proof. The variance of $\langle\Psi_i|\, Q\,|\Psi_i\rangle$ can be upper bounded using Claims 5.5 and 5.6 as

$$\mathbb{E}_{i\leftarrow\lambda}(\langle\Psi_i|\, Q\,|\Psi_i\rangle)^2 - (\mathbb{E}_{i\leftarrow\lambda}(\langle\Psi_i|\, Q\,|\Psi_i\rangle))^2 \le \delta^2\frac{(d-k)(d-k+1)}{d(d+1)} + \delta^2\varepsilon - (\frac{\delta(d-k)}{d} - \delta\varepsilon)^2$$

$$\le \delta^2(\varepsilon - \varepsilon^2 + 2\varepsilon + \frac{(d-k)(d-k+1)}{d(d+1)} - \frac{(d-k)^2}{d^2}) \le \delta^2(3\varepsilon + \frac{3}{d}).$$

Now, using Chebysev's inequality, we find that

$$\mathrm{Pr}_\lambda(\langle\Psi_i|\, Q\,|\Psi_i\rangle < \mathbb{E}_{i\leftarrow\lambda}(\langle\Psi_i|\, Q\,|\Psi_i\rangle) - \alpha) \le \frac{\delta^2}{\alpha^2}(3\varepsilon + \frac{3}{d}).$$

Using Claim 5.5, the lemma follows.

□

We now proceed to the main lemma of this section, proof of which is deferred to Appendix D.

Lemma 5.8. *Assume the conditions $\eta < \frac{\delta}{8}$ and $\delta < 1/4$. Let $k = d/4$ and $\varepsilon = \frac{1}{d}$. Let a be the largest real that satisfies $\mathbb{E}_{i\leftarrow\lambda}2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} \le 2^{-a}$. Then it holds that $a \ge \log(d\delta) - 6$*

6 Proof of main results

We begin with restatement and proof of Theorem 1.2.

Theorem 6.1. *Fix a positive integer $d > 4$ and real δ that satisfies $\delta < \frac{1}{4}$. There exist a collection of $N \stackrel{\text{def}}{=} 8d^7$ states $\{|\Psi_x\rangle\}_{x=1}^N$ that depend on δ and belong to a d-dimensional Hilbert space, and a probability distribution $\{p(x)\}_{x=1}^N$, such that following holds for the ensemble $\{(p(x), \Psi_x)\}_{x=1}^N$.*

- *The von-Neumann entropy of the average state satisfies $S(\sum_x p(x)\Psi_x) \leq \delta\log(d)+H(\delta)+2$*
- *For any one-way protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta}{8}$, the expected communication cost is lower bounded by $(1-\sqrt{\eta})^2\log(\frac{d\delta}{128})$.*
- *For any r-round protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta}{8}$, the expected communication cost is lower bounded by*

$$\frac{1}{20}\cdot\frac{\log(\frac{d\delta}{128})}{(\log r)}.$$

Proof. We use the construction as given in Subsection 5.1 with $\varepsilon = \frac{1}{d}$ as chosen in lemma 5.8. Thus the set $\mathcal{N}_{\frac{1}{d}}$ has $8d^7$ elements.

From lemma 5.1, the von-Neumann entropy of the average state is upper bounded by

$$(\delta+\frac{1}{d})\log(d)+H(\delta)+1 < \delta\log(d)+H(\delta)+2.$$

For the 'one-way' part of the theorem, we combine Lemma 3.1 and Lemma 5.8 to obtain a lower bound on expected communication cost as

$$(1-\sqrt{\eta})^2\log(\frac{d\delta}{64})-1 > (1-\sqrt{\eta})^2\log(\frac{d\delta}{128}).$$

For the 'round dependent' part of the theorem, we combine Lemma 4.5 and Lemma 5.8 to obtain a lower bound on expected communication cost as

$$(1-\sqrt{\eta})^2\frac{\log(\frac{d\delta(1-2\sqrt{\eta})}{64})}{2(\log r+4)} > \frac{1}{20}\cdot\frac{\log(\frac{d\delta}{128})}{(\log r)}.$$

□

In last part of this section, we restate and prove lemma 1.3.

Theorem 6.2. *Fix a positive integer $d > 4$ and a real δ that satisfies $\delta < \frac{1}{4}$. There exist a collection of $N \stackrel{\text{def}}{=} 8d^7$ states $\{|\Psi_x\rangle\}_{x=1}^N$ that depend on δ and belong to a d-dimensional Hilbert space, and a probability distribution $\{p(x)\}_{x=1}^N$ such that following holds for the ensemble $\{(p(x), \Psi_x)\}_{x=1}^N$.*

- *The von-Neumann entropy of the average state satisfies $S(\sum_x p(x)\Psi_x) \leq \delta\log(d)+H(\delta)+2$.*

- *For any interactive protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta^2}{192}$, the expected communication cost is lower bounded by*

$$\frac{1}{20} \cdot \frac{\log(\frac{d\delta}{128})}{(\log\log(d) - 2\log\eta)}.$$

Proof. Fix a communication protocol P with r rounds and error parameter η. Define an odd number $\ell \stackrel{\text{def}}{=} 2\lceil\frac{\log(d)}{\eta^2}\rceil + 1 > \frac{\log(d)}{\eta^2}$. Let $\mathcal{B}$ denote the set of all instances $(x, i_1, i_2 \ldots i_r)$ (input x and messages exchanged) in which the protocol aborts before round ℓ. From Remark 4.2, such instances are of the form $(x, i_1, i_2, \ldots, 1, 1, \ldots 1)$. Let $\mathcal{G}$ be the remaining set of instances, that is, the protocol aborts at or after round ℓ. It is easy to infer that if $(x, i_r, i_{r-1} \ldots i_1) \in \mathcal{G}$, then number of bits exchanged in this instance is at least ℓ (as at least one bit must be exchanged in each round till round number ℓ).

Now we consider two cases. First case is that $\sum_{(x,i_1,i_2\ldots i_r)\in\mathcal{G}} p(x)p^x_{i_r,i_{r-1}\ldots i_1} > \eta^2$. Then expected communication cost is lower bounded by $\eta^2\frac{\log(d)}{\eta^2} = \log(d)$.

Second case is $\sum_{(i_1,i_2\ldots i_r)\in\mathcal{G}} p(x)p^x_{i_r,i_{r-1}\ldots i_1} \leq \eta^2$. Consider a protocol P′ which proceeds as P till round ℓ. If round ℓ is reached then parties abort and Bob considers $|0\rangle_C$ as his output. If protocol P′ aborts before round ℓ, then Bob outputs according to P.

If the protocol P′ aborts after exchanging messages $(i_1, i_2 \ldots i_k)$, for $k < \ell$, then following Remark 4.2, we define $p^x_{i_r,i_{r-1}\ldots i_1} = p^x_{i_k,i_{k-1},\ldots i_1}$ iff $(i_r, i_{r-1}, \ldots i_1) = (1, 1, \ldots 1, i_k, i_{k-1} \ldots i_1)$ and 0 otherwise. We define $\phi^{x,i_r,i_{r-1}\ldots i_1}_{AB} = \phi^{x,i_k,i_{k-1},\ldots i_1}_{AB}$ iff $(i_r, i_{r-1}, \ldots i_1) = (1, 1, \ldots 1, i_k, i_{k-1} \ldots i_1)$ and $|0\rangle_{AB}$ otherwise.

Let $\tilde{\Phi}^x_C$ be the final state with Bob, conditioned on input x. We have that

$$\tilde{\Phi}^x_C = \sum_{(x,i_1,i_2\ldots i_r)\in\mathcal{B}} p^x_{i_r,i_{r-1},\ldots i_1}\tau^{x,i_r,i_{r-1},\ldots i_1}_C + \beta\,|0\rangle\langle 0|_C$$

with $\beta \leq \eta^2$. On the other hand, the final state Φ^x_C of original protocol is

$$\Phi^x_C = \sum_{(x,i_1,i_2\ldots i_r)\in\mathcal{B}} p^x_{i_r,i_{r-1},\ldots i_1}\tau^{x,i_r,i_{r-1},\ldots i_1}_C + \sum_{(x,i_1,i_2\ldots i_r)\in\mathcal{B}} p^x_{i_r,i_{r-1},\ldots i_1}\tau^{x,i_r,i_{r-1},\ldots i_1}_C.$$

Using joint concavity of fidelity (Fact 2.3), we thus obtain that $\mathrm{F}(\tilde{\Phi}^x_C, \Psi^x_C) \geq 1 - \eta^2$. This implies

$$\sum_x p(x)\mathrm{F}^2(\Psi_x, \tilde{\Phi}^x_C) \geq \sum_x p(x)\mathrm{F}^2(\Psi_x, \Phi^x_C) - \sum_x p(x)\|\tilde{\Phi}^x_C - \Psi^x_C\|_1 \geq 1 - \eta^2 - 2\eta \geq 1 - 3\eta.$$

Thus, P′ is a protocol with ℓ rounds and error parameter $\sqrt{3\eta} < \frac{\delta}{8}$. The expected communication cost of P′ is lower bounded by (Theorem 6.1):

$$\frac{1}{20} \cdot \frac{\log(\frac{d\delta}{128})}{(\log\ell)} = \frac{1}{20} \cdot \frac{\log(\frac{d\delta}{128})}{(\log\log(d) - 2\log\eta)}.$$

This is also the lower bound on expected communication cost of P. This proves the corollary. □

7 Example for one-shot Quantum Reverse Shannon theorem

In this section, we construct an example of a quantum channel which requires large amount of expected communication for its simulation. It uses the construction presented in Section 5 and we carry over the notations from Section 5.

As assumed in Lemma 5.8, we have set $\varepsilon = \frac{1}{d}$, which means that $m = \mathcal{N}_\varepsilon = 8d^7$. Let A be a register corresponding to Hilbert space spanned by $|j\rangle$, where j ranges over $\{1, 2 \ldots m\}$. We consider the following classical-quantum channel $\mathcal{E} : A \to B$,

$$\mathcal{E}(\rho) = \sum_j \langle j| \rho |j\rangle |\Psi_j\rangle\langle\Psi_j| .$$

We shall compute the entanglement assisted classical capacity of this channel, which is given by ([BSST02])

$$\mathcal{C}(\mathcal{E}) \stackrel{\text{def}}{=} \max_{\rho_A} \mathrm{I}(R : A)_{\mathcal{E}(\rho_{RA})} ,$$

where ρ_{RA} is a purification of the state ρ_A on register R.

Lemma 7.1. *Entanglement assisted classical capacity $\mathcal{C}(\mathcal{E})$ of the channel $\mathcal{E}$ is upper bounded by $\delta \log(d) + H(\delta) + 2$.*

Proof. Since $\mathcal{E}$ is a classical-quantum channel, the quantum state $\mathcal{E}(\rho_{RA})$ is a classical quantum state. More specifically, we have

$$\mathcal{E}(\rho_{RA}) = \sum_j \langle j|_A |\rho\rangle_{RA} \langle\rho|_{RA} |j\rangle_A \otimes |\Psi_j\rangle\langle\Psi_j|_B = \sum_j \rho^{\frac{1}{2}} |j\rangle\langle j|_R \rho^{\frac{1}{2}} \otimes |\Psi_j\rangle\langle\Psi_j|_B .$$

Thus, we can use the relation $\mathrm{I}(R : A)_{\mathcal{E}(\rho_{RA})} \leq S(\mathcal{E}(\rho_A))$. This implies

$$\mathcal{C}(\mathcal{E}) \leq \max_{\rho_A} S(\sum_j \langle j| \rho |j\rangle |\Psi_j\rangle\langle\Psi_j|) = \max_{\mu(j)} S(\sum_j \mu(j) |\Psi_j\rangle\langle\Psi_j|).$$

The equality above holds since the quantity $\langle j| \rho |j\rangle$ can be viewed as a probability distribution over indices in $\{1, 2 \ldots m\}$. Now, let $\mu(j), \mu'(j)$ be any two distributions over $\{1, 2 \ldots m\}$. Then it holds that

$$\begin{aligned} & S(\sum_j (\alpha\mu(j) + (1-\alpha)\mu'(j)) |\Psi_j\rangle\langle\Psi_j|) \\ = \; & S(\alpha \sum_j \mu(j) |\Psi_j\rangle\langle\Psi_j| + (1-\alpha) \sum_j \mu'(j) |\Psi_j\rangle\langle\Psi_j|) \\ > \; & \alpha \cdot S(\sum_j \mu(j) |\Psi_j\rangle\langle\Psi_j|) + (1-\alpha) \cdot S(\sum_j \mu'(j) |\Psi_j\rangle\langle\Psi_j|) \end{aligned}$$

This implies that the desired distribution achieving the maximum is the uniform distribution over $\{1, 2 \ldots K\}$. But for such a distribution, we have

$$\max_{\mu(j)} S(\sum_j \mu(j) |\Psi_j\rangle\langle\Psi_j|) = S(\sum_j \frac{1}{m} |\Psi_j\rangle\langle\Psi_j|).$$

As computed in Lemma 5.1, this entropy is upper bounded by $\delta \log(d) + H(\delta) + 2$. This proves the lemma. ∎

Now we are in a position to prove theorem 1.4. We restate it below.

Theorem 7.2. *Fix a positive integer $d > 4$ and a real δ that satisfies $\delta < \frac{1}{4}$. There exists a register A with dimension $8d^7$ and classical-quantum channel $\mathcal{E} : A \to B$ with one-shot η-error entanglement assisted classical capacity upper bounded by*

$$\frac{\delta \log(d) + H(\delta) + H(\eta) + 2}{1 - \eta},$$

such that for any protocol achieving the simulation of above channel with error at most η, following holds:

- *If the protocol is one-way and* $\eta < \frac{\delta}{8}$*, then simulation cost of the protocol is at least* $(1-\sqrt{\eta})^2 \log(\frac{d\delta}{128})$.
- *If the protocol is interactive with* r*-rounds and* $\eta < \frac{\delta}{8}$*, the simulation cost is lower bounded by* $\frac{1}{20} \cdot \frac{\log(\frac{d\delta}{128})}{(\log r)}$.
- *If* $\eta < \frac{\delta^2}{192}$ *and the protocol is interactive, the simulation cost of the protocol is lower bounded by*

$$\frac{1}{20} \cdot \frac{\log(\frac{d\delta}{128})}{(\log \log(d) - 2\log \eta)}.$$

Proof. From Lemma 30 in [WM14], the one-shot entanglement assisted classical capacity of channel $\mathcal{E}$ with error η is upper bounded by

$$\frac{\mathcal{C}(\mathcal{E}) + H(\eta)}{1-\eta} < \frac{\delta \log(d) + H(\delta) + H(\eta) + 2}{1-\eta}.$$

On the other hand, consider a protocol that simulates the action of the channel $\mathcal{E}$ with error η. Simulation cost of the protocol is maximum expected cost over all inputs to the channel. Thus, it is lower bounded by expected communication cost when inputs are given according to a fixed distribution. We consider a distribution over inputs as follows: Alice receives a $|j\rangle\langle j|$ with probability $\frac{1}{m}$. The channel outputs the state $|\Psi_j\rangle\langle\Psi_j|$ and the protocol must simulate this output with error at most η. It is now easy to observe that the lower bounds of Theorems 6.1 and 6.2 apply for respective choice of parameters, which proves the theorem. □

8 Conclusion

In this work, we have shown a large gap between the quantum information complexity and the average/expected communication complexity of the quantum Huffman task (Definition 1.1). As an application of our main results, we show that in one-shot setting, quantum channels cannot be simulated with a cost as good as their entanglement assisted classical capacity.

We have following questions that we leave open.

- The interactive part of our main theorem, Theorem 1.2 has a dependence on the number of rounds. We get rid of this dependence in Lemma 1.3, but at the expense of weaker lower bound on expected communication cost. Can we get rid of dependence on number of rounds in Theorem 1.2 itself. For comparison, it may be noted that the results in [Ans15] have no dependence on the number of rounds.
- What is the correct way to operationally understand fundamental quantum information theoretic quantities in one-shot setting? Our result says that expected communication cost is not the right notion, but naturally we cannot rule out other notions.
- Is there a way to improve the direct sum result for bounded-round entanglement assisted quantum information complexity of [Tou15]?

Acknowledgment

A.A., A.G. and P.Y. would like to thank the Institute for Mathematical Science, Singapore for their hospitality and their organized workshop "Semidefinite and Matrix Methods for Optimization and Communication". A.A would like to thank the Institute for Quantum Computing, University of Waterloo for their hospitality, where part of this work was done. We thank Dave Touchette for helpful comments on the manuscript. A.A. thanks Rahul Jain and Guo Yalei for helpful discussions. A.G. thanks Mohammed Bavarian and Henry Yuen for helpful discussions.

P.Y. is supported by NSERC and CIFAR. A.A. is supported by Core Grants of Centre for Quantum Technologies. A.G. is supported by NSF grants CCF-1149888 and CCF-1525342, a Simons fellowship for graduate students in theoretical computer science and a Siebel scholarship. AWH was funded by NSF grants CCF-1111382 and CCF-1452616.

References

[AJM+14] Anurag Anshu, Rahul Jain, Priyanka Mukhopadhyay, Ala Shayeghi, and Penghui Yao. A new operational interpretation of relative entropy and trace distance between quantum states. http://arxiv.org/abs/1404.1366, 2014.

[Ans15] Anurag Anshu. A lower bound on expected communication cost of quantum state redistribution. http://arxiv.org/abs/1506.06380, 2015.

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the forty-second ACM symposium on Theory of computing*, STOC '10, pages 67–76, New York, NY, USA, 2010. ACM.

[BCF+96] Howard Barnum, Carlton M. Cave, Christopher A. Fuch, Richard Jozsa, and Benjamin Schmacher. Noncommuting mixed states cannot be broadcast. *Phys. Rev. Lett.*, 76:2818–2821, 1996.

[BCR11] Mario Berta, Matthias Christandl, and Renato Renner. The Quantum Reverse Shannon Theorem based on one-shot information theory. *Commun. Math. Phys.*, 306:579–615, 2011.

[BCT16] M. Berta, M. Christandl, and D. Touchette. Smooth entropy bounds on one-shot quantum state redistribution. *IEEE Transactions on Information Theory*, 62(3):1425–1439, March 2016.

[BDH+14] Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. Quantum reverse shannon theorem. *IEEE Transactions on Information Theory*, 60(5):2926–2959, 2014.

[BFGL98] Samuel L. Braunstein, Christopher A. Fuchs, Daniel Gottesman, and Hoi-Kwong Lo. A quantum analog of huffman coding. In *1998 IEEE International Symposium on Information Theory*, 1998.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *Proceedings of the 52nd Symposium on Foundations of Computer Science*, FOCS '11, pages 748–757, Washington, DC, USA, 2011. IEEE Computer Society.

[BSST02] Charles H. Bennett, Peter W. Shor, John A. Smolin, and A. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem. *IEEE Transactions on Information Theory*, 48:2637 – 2655, 2002.

[DY08] Igor Devetak and Jon Yard. Exact cost of redistributing multipartite quantum states. *Phys. Rev. Lett.*, 100, 2008.

[Fan73] M. Fannes. A continuity property of the entropy density for spin lattice systems. *Communications in Mathematical Physics*, 31:291–294, 1973.

[HJMR10] Prahladh Harsha, Rahul Jain, David Mc.Allester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transcations on Information Theory*, 56:438–449, 2010.

[Hol73] Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, (9):177–183, 1973.

[HOW07] Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269:107–136, 2007.

[Huf52] David Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of IRE*, 40(9):1098–1101, 1952.

[JRS05] 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, Washington, DC, USA, 2005. IEEE Computer Society.

[JRS08] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Optimal direct sum and privacy trade-off results for quantum and classical communication complexity. http://arxiv.org/abs/0807.1267, 2008.

[Lin75] G. Lindblad. Completely positive maps and entropy inequalities. *Commun. Math. Phys.*, 40:147–151, 1975.

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

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

[SW73] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19:471–480, 1973.

[Tom12] Marco Tomamichel. A framework for non-asymptotic quantum information theory, 2012. PhD Thesis, ETH Zurich.

[Tou15] Dave Touchette. Quantum information complexity. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 317–326, New York, NY, USA, 2015. ACM.

[Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

[Wat11] John Watrous. Theory of Quantum Information, lecture notes, 2011. https://cs.uwaterloo.ca/ watrous/LectureNotes.html.

[WM14] Stephanie Wehner William Matthews. Finite blocklength converse bounds for quantum channels. *IEEE Transcations on Information Theory*, 60:7317–7329, 2014.

[Wyn75] Aaron D Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21:163–179, 1975.

[YD09] Jon T. Yard and Igor Devetak. Optimal quantum source coding with quantum side information at the encoder and decoder. *IEEE Transactions on Information Theory*, 55:5339–5351, 2009.

A Proof of Lemma 4.4

Proof. We consider following chain of Lowener inequalities

$$\mathrm{Tr}_A(\phi^{x,i_r,i_{r-1}\dots i_1}) = \frac{\mathrm{Tr}_A(M_A^{i_r,i_{r-1}\dots i_1} M_B^{i_{r-1}\dots i_1} \dots M_A^{i_1} \theta_{AB} M_A^{i_1\dagger} \dots M_B^{i_{r-1}\dots i_1\dagger} M_A^{i_r,i_{r-1}\dots i_1\dagger})}{p^x_{i_r,i_{r-1}\dots i_1}}$$

$$< \frac{\sum_{i_r} \mathrm{Tr}_A(M_A^{i_r,i_{r-1}\dots i_1} M_B^{i_{r-1}\dots i_1} \dots M_A^{i_1} \theta_{AB} M_A^{i_1\dagger} \dots M_B^{i_{r-1}\dots i_1\dagger} M_A^{i_r,i_{r-1}\dots i_1\dagger})}{p^x_{i_r,i_{r-1}\dots i_1}}$$

$$= \frac{\mathrm{Tr}_A(M_B^{i_{r-1}\dots i_1} M_A^{i_{r-2}\dots i_1} \dots M_A^{i_1} \theta_{AB} M_A^{i_1\dagger} \dots M_A^{i_{r-2}\dots i_1\dagger} M_B^{i_{r-1}\dots i_1\dagger})}{p^x_{i_r,i_{r-1}\dots i_1}}$$

$$< \frac{\sum_j \mathrm{Tr}_A(M_B^{i_{r-1}\dots i_1} M_A^{j\dots i_1} \dots M_A^{i_1} \theta_{AB} M_A^{i_1\dagger} \dots M_A^{j\dots i_1\dagger} M_B^{i_{r-1}\dots i_1\dagger})}{p^x_{i_r,i_{r-1}\dots i_1}}$$

(Adding positive matrices such that $j = i_{r-2}$ is included)

$$= \frac{\sum_j \mathrm{Tr}_A(M_B^{i_{r-1}\dots i_1} M_A^{j\dots i_1\dagger} M_A^{j\dots i_1} \dots M_A^{i_1} \theta_{AB} M_A^{i_1\dagger} \dots M_B^{i_{r-1}\dots i_1\dagger})}{p^x_{i_r,i_{r-1}\dots i_1}}$$

(Alice and Bob have commuting POVMs, and partial trace is cyclic for Alice)

$$= \frac{\mathrm{Tr}_A(M_B^{i_{r-1}\dots i_1} M_B^{i_{r-3}\dots i_1} \dots M_A^{i_1} \theta_{AB} M_A^{i_1\dagger} \dots M_B^{i_{r-3}\dots i_1\dagger} M_B^{i_{r-1}\dots i_1\dagger})}{p^x_{i_r,i_{r-1}\dots i_1}}$$

(Continuing the same way for all POVMs of Alice)

$$< \frac{\mathrm{Tr}_A(M_B^{i_{r-1}\dots i_1} M_B^{i_{r-3}\dots i_1} \dots M_B^{i_2,i_1} \theta_{AB} M_B^{i_2,i_1\dagger} \dots M_B^{i_{r-3}\dots i_1\dagger} M_B^{i_{r-1}\dots i_1\dagger})}{p^x_{i_r,i_{r-1}\dots i_1}}$$

$$= \frac{M_B^{i_{r-1}\dots i_1} M_B^{i_{r-3}\dots i_1} \dots M_B^{i_2,i_1} \theta_B M_B^{i_2,i_1\dagger} \dots M_B^{i_{r-3}\dots i_1\dagger} M_B^{i_{r-1}\dots i_1\dagger}}{p^x_{i_r,i_{r-1}\dots i_1}}$$

$$< \frac{\sum_l M_B^{l,i_{r-2}\dots i_1}(\dots(\sum_k M_B^{k,i_3,i_2,i_1}(\sum_j M_B^{j,i_1} \theta_B M_B^{j,i_1\dagger}) M_B^{k,i_3,i_2,i_1\dagger})\dots) M_B^{l,i_{r-2}\dots i_1\dagger}}{p^x_{i_r,i_{r-1}\dots i_1}}$$

(Adding positive operators to make numerator a quantum state)

Now define

$$\sigma^{i_{r-2}\dots i_1} \stackrel{\mathrm{def}}{=} \sum_l M_B^{l,i_{r-2}\dots i_1}(\dots(\sum_k M_B^{k,i_3,i_2,i_1}(\sum_j M_B^{j,i_1} \theta_B M_B^{j,i_1\dagger}) M_B^{k,i_3,i_2,i_1\dagger})\dots) M_B^{l,i_{r-2}\dots i_1\dagger}$$

which is clearly independent of x. Then we have $p^x_{i_r,i_{r-1}\dots i_1} < 2^{-\mathrm{D}_{\max}\left(\mathrm{Tr}_A(\phi^{x,i_r,i_{r-1}\dots i_1}) \| \sigma^{i_{r-2}\dots i_1}\right)}$. Applying Bob's final unitary and tracing out register B', we obtain a state

$$\omega^{i_r,i_{r-1}\dots i_1} \stackrel{\mathrm{def}}{=} \mathrm{Tr}_{B'}(U_{i_r,i_{r-1}\dots i_1} \sigma^{i_{r-2}\dots i_1} U^\dagger_{i_r,i_{r-1}\dots i_1})$$

which is independent of x. Now using monotonicity of max-entropy under quantum operations (Fact 2.2), we find that

$$p^x_{i_r,i_{r-1}\dots i_1} < 2^{-\mathrm{D}_{\max}\left(\tau_C^{x,i_r,i_{r-1}\dots i_1} \middle\| \omega^{i_r,i_{r-1}\dots i_1}\right)}.$$

This proves the lemma.

□

B Proof of Lemma 4.5

Proof. Our proof follows in the following steps.

1. **Pruning out x with low fidelity:**

   From correctness of the protocol, we know that $\sum_x p(x)\mathrm{F}^2(\Phi^x_C, \Psi_x) \geq 1 - \eta^2$. Let $\mathcal{G}$ be the set of all indices x such that $\mathrm{F}^2(\Phi^x_C, \Psi_x) > 1 - \eta^{3/2}$. Let $\mathcal{B}$ be rest of the indices. Then by Markov's inequality, we have $\sum_{x\in\mathcal{G}} p(x) \geq 1 - \sqrt{\eta}$. Define a new probability distribution $p'(x)$ which is 0 whenever $x \in \mathcal{B}$ and equal to $\frac{p(x)}{\sum_{x\in\mathcal{G}} p(x)}$ for $x \in \mathcal{G}$. It holds that $p'(x) \leq \frac{p(x)}{1-\sqrt{\eta}}$ for all x.

2. **Upper bound on probabilities $p^x_{i_r,i_{r-1}\dots i_1}$:**

   From lemma 4.4, it follows that $p^x_{i_r,i_{r-1}\dots i_1} < 2^{-\mathrm{D}_{\max}\left(\tau_C^{x,i_r,i_{r-1}\dots i_1} \middle\| \omega^{i_r,i_{r-1}\dots i_1}\right)}$.

   Now for each $x \in \mathcal{G}$, we define $\mathcal{B}_x$ to be the set of tuples $(i_r, i_{r-1}\dots i_1)$ for which $\langle\Psi_x|\, \tau^{x,i_r,i_{r-1}\dots i_1}\, |\Psi_x\rangle < 1 - \eta$. Let $\mathcal{G}_x$ be rest of the indices. Then we have $\sum_{(i_r,i_{r-1}\dots i_1)\in\mathcal{B}_x} p^x_{i_r,i_{r-1}\dots i_1} < \sqrt{\eta}$. And hence for all $(i_r, i_{r-1}\dots i_1) \notin \mathcal{B}_x$, we obtain $p^x_{i_r,i_{r-1}\dots i_1} < 2^{-\mathrm{D}^{\eta}_{\max}\left(\Psi_x \middle\| \omega^{i_r,i_{r-1}\dots i_1}\right)}$.

3. **Upper bound on average probability of a message:**

   We define new probability distribution $q^x_{i_r,i_{r-1}\dots i_1}$ which is 0 whenever $(i_r, i_{r-1}\dots i_1) \in \mathcal{B}_x$ and equal to $\frac{p^x_{i_r,i_{r-1}\dots i_1}}{\sum_{(i_r,i_{r-1},\dots i_1)\in\mathcal{G}_x} p^x_{i_r,i_{r-1},\dots i_1}}$ otherwise. It follows that $q^x_{i_r,i_{r-1}\dots i_1} < \frac{1}{1-\sqrt{\eta}} 2^{-\mathrm{D}^{\eta}_{\max}\left(\Psi_x \middle\| \omega^{i_r,i_{r-1}\dots i_1}\right)}$.

   Define $s_{i_r i_{r-1}\dots i_1} \stackrel{\mathrm{def}}{=} \sum_x p'(x) q^x_{i_r,i_{r-1}\dots i_1}$. We have $\sum_{i_r,i_{r-1}\dots i_1} s_{i_r,i_{r-1}\dots i_1} = 1$. Furthermore,

$$\begin{aligned} s_{i_r,i_{r-1}\dots i_1} &\leq \frac{1}{1-\sqrt{\eta}} \sum_x p'(x) 2^{-\mathrm{D}^{\eta}_{\max}\left(\Psi_x \middle\| \omega^{i_r,i_{r-1}\dots i_1}\right)} \\ &\leq \frac{1}{(1-\sqrt{\eta})^2} \sum_x p(x) 2^{-\mathrm{D}^{\eta}_{\max}\left(\Psi_x \middle\| \omega^{i_r,i_{r-1}\dots i_1}\right)} \\ &= \frac{t_{i_r,i_{r-1}\dots i_1}}{(1-\sqrt{\eta})^2} < \frac{2^{-a}}{(1-\sqrt{\eta})^2} \end{aligned} \tag{4}$$

4. **Lower bound on expected communication cost:**

The expected communication cost of the protocol is lower bounded by

$$\begin{aligned}
& \sum_x p(x) \sum_{i_r,i_{r-1}\ldots i_1} p^x_{i_r,i_{r-1}\ldots i_1}(1-\sqrt{\eta}) \\
\geq & \sum_{x\in\mathcal{G}} p(x) \sum_{i_r,i_{r-1}\ldots i_1} q^x_{i_r,i_{r-1}\ldots i_1}\log(i_r i_{r-1}\ldots i_1) \\
\geq & (1-\sqrt{\eta})^2 \sum_x p'(x) \sum_{i_r,i_{r-1}\ldots i_1} q^x_{i_r,i_{r-1}\ldots i_1}\log(i_r i_{r-1}\ldots i_1) \\
= & (1-\sqrt{\eta})^2 \sum_{i_r,i_{r-1}\ldots i_1} s_{i_r,i_{r-1}\ldots i_1}\log(i_r i_{r-1}\ldots i_1)
\end{aligned}$$

where the last equality follows from definition of $s_{i_r,i_{r-1}\ldots i_1}$. Defining $b \overset{\text{def}}{=} a+2\log(1-\sqrt{\eta})$, invoking Equation 4 and applying the claim B.1 shown below, the lemma follows.

□

Claim B.1. Suppose $s_{i_r,i_{r-1},\ldots i_1} < 2^{-b}$ for all $(i_1, i_2 \ldots i_r)$. Then we have

$$\sum_{i_1,i_2\ldots i_1} s_{i_r,i_{r-1},\ldots i_1}\log(i_r i_{r-1}\ldots i_1) > \frac{b}{2(\log r+4)}.$$

Proof. For an integer k let $N(k)$ be the number of *ordered* tuples $(i_1, i_2 \ldots i_r)$ such that $k = i_1 \cdot i_2 \cdot \ldots i_r$. Let $M(k) = \sum_{k'=1}^{k} N(k')$. The quantity $\sum_{i_1,i_2\ldots i_1} s_{i_r,i_{r-1},\ldots i_1}\log(i_r i_{r-1}\ldots i_1)$ is minimized when all $s_{i_r,i_{r-1},\ldots i_1}$ with *smallest possible values* of the product $i_1 \cdot i_2 \cdot \ldots i_r$ have taken the value 2^{-b}. Let k^* be the largest integer such that $M(k^*) < 2^b$. Let $N'(k^*+1) \overset{\text{def}}{=} 2^b - M(k^*)$. Then

$$\sum_{i_1,i_2\ldots i_1} s_{i_r i_{r-1}\ldots i_1}\log(i_r i_{r-1}\ldots i_1) = 2^{-b}\Big(\sum_{k=1}^{k^*} N(k)\log(k) + N'(k^*+1)\log(k^*+1)\Big).$$

Our lower bound shall now proceed by evaluating $N(k)$. Let $k = 2^{a_1}3^{a_2}\ldots p_t^{a_t}$ be prime decomposition of k. Each of the r integers that multiply to give k can be written as $n_f = 2^{a_1^f}3^{a_2^f}\ldots p_t^{a_t^f}$ where $f \in [r]$. Since $n_1 \cdot n_2 \ldots n_r = k$, we have that

$$a_1^1 + a_1^2 + \ldots a_1^r = a_1, \quad a_2^1 + a_2^2 + \ldots a_2^r = a_2, \quad a_t^1 + a_t^2 + \ldots a_t^r = a_t.$$

We need to compute the number of ways of selecting the ordered tuple $(a_1^1, a_1^2 \ldots a_1^r, a_2^1 \ldots a_2^r, \ldots a_t^1, a_t^2, \ldots a_t^r)$ that satisfy above constraints. Note that the order matters. Looking at the first constraint, the number of ways is well known to be $\binom{a_1+r-1}{r-1}$. Similar argument holds for rest of the constraints, and each being independent, we obtain that the number of ways is :

$$\binom{a_1+r-1}{r-1}\binom{a_2+r-1}{r-1}\ldots\binom{a_t+r-1}{r-1}.$$

Since

$$\binom{a_1+r-1}{r-1} = (a_1+1)(\frac{a_1}{2}+1)\ldots(\frac{a_1}{r-1}+1) < 2^{a_1+\frac{a_1}{2}+\ldots\frac{a_1}{r-1}} < 2^{a_1(\log r+\gamma)},$$

where γ is Euler-Mascheroni constant, we have that

$$N(k) < 2^{(a_1+a_2+\ldots a_t)(\log r+\gamma)} < k^{(\log r+\gamma)}.$$

Thus, $M(k) < k^{(\log r+\gamma+1)}$.

Now, k^* is the integer such that $M(k^*) < 2^b < M(k^*+1)$. This means, $k^* > 2^{\frac{b}{\log r+\gamma+1}} - 1$. Now we are in a position to lower bound expected communication cost. Consider,

$$\begin{aligned}
& \sum_{i_1,i_2\ldots i_1} s_{i_r i_{r-1}\ldots i_1} \log(i_r i_{r-1}\ldots i_1) \\
= \; & 2^{-b}(\sum_{k=1}^{k^*} N(k)\log(k) + N'(k^*+1)\log(k^*+1)) \\
> \; & 2^{-b}(\sum_{k=\sqrt{k^*}}^{k^*} N(k)\log(k) + N'(k^*+1)\log(k^*+1)) \\
> \; & \frac{\log(k^*)}{2}\cdot 2^{-b}(\sum_{k=\sqrt{k^*}}^{k^*} N(k) + N'(k^*+1)) \\
= \; & \frac{\log(k^*)}{2}\cdot(1 - 2^{-b}\sum_{k=1}^{\sqrt{k^*}-1} N(k)) \\
> \; & \frac{\log(k^*)}{2}\cdot(1-2^{-b}(\sqrt{k^*})^{\log r+\gamma+1}) > \frac{b(1-2^{-b/2})}{2(\log r+\gamma+1)}
\end{aligned}$$

This proves the claim. □

C Proof of Claims 5.3 and 5.4

Proof of Claim 5.3 . For a fixed i, let ρ_i be the state that achieves the infimum in the definition of $\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)$. It satisfies $\langle\Psi_i|\,\rho_i\,|\Psi_i\rangle \geq 1-\eta$. This means the largest eigenvalue of ρ_i is at least $1-\eta$. Thus, consider the eigen-decomposition $\rho_i = \lambda_1\,|\lambda_1\rangle\langle\lambda_1| + \sum_{j>1}\lambda_j\,|\lambda_j\rangle\langle\lambda_j|$. We have $\lambda_1 > 1-\eta$ or equivalently $\sum_{j>1}\lambda_j < \eta$. Thus,

$$1-\eta < \langle\Psi_i|\,\rho_i\,|\Psi_i\rangle = \lambda_1|\langle\Psi_i|\lambda_1\rangle|^2 + \sum_{j>1}\lambda_j|\langle\Psi_i|\lambda_j\rangle|^2 < |\langle\Psi_i|\lambda_1\rangle|^2 + \sum_{j>1}\lambda_j < |\langle\Psi_i|\lambda_1\rangle|^2 + \eta.$$

Hence, $|\langle\Psi_i|\lambda_1\rangle|^2 > 1-2\eta$. Moreover,

$$2^{\mathrm{D}_{\max}(\rho_i\|\sigma)} = \|\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\|_\infty > (1-\eta)\|\sigma^{-\frac{1}{2}}\,|\lambda_1\rangle\langle\lambda_1|\,\sigma^{-\frac{1}{2}}\|_\infty = (1-\eta)\,\langle\lambda_1|\,\sigma^{-1}\,|\lambda_1\rangle\,,$$

where σ^{-1} is the pseudo-inverse of σ. From the definition of the projector Q^-, the following inequality easily follows:

$$\langle\lambda_1|\,\sigma^{-1}\,|\lambda_1\rangle \geq k\,\langle\lambda_1|\,Q^-\,|\lambda_1\rangle\,.$$

Thus we get

$$2^{\mathrm{D}_{\max}(\rho_i\|\sigma)} > k(1-\eta)\,\langle\lambda_1|\,Q^-\,|\lambda_1\rangle\,.$$

Inverting and using $|\langle\Psi_i|\lambda_1\rangle|^2 > 1-2\eta$, we have

$$2^{-\mathrm{D}_{\max}(\rho_i\|\sigma)} < \frac{1}{k(1-\eta)\,\langle\lambda_1|\,Q^-\,|\lambda_1\rangle} < \frac{1}{k(1-\eta)S^{2\eta}(\Psi_i||Q^-)}.$$

This proves the claim.

□

Proof of Claim 5.4. Let $|\lambda_i\rangle$ be the state that achieves the infimum in the definition of $S^\eta(\Psi_i||Q^-)$. We know that $|\lambda_i\rangle$ has fidelity at least $1-\eta$ with $|\Psi_i\rangle$ and also minimizes the overlap with the subspace Q^-. Intuitively, this state must lie in the span of two vectors $\{Q^-|\Psi_i\rangle, Q^+|\Psi_i\rangle\}$. This we shall find to be true below.

Let us expand

$$|\lambda_i\rangle = aQ^-|\Psi_i\rangle + bQ^+|\Psi_i\rangle + c|\theta\rangle,$$

where $|\theta\rangle$ is normalized vector orthogonal to $\{Q^-|\Psi_i\rangle, Q^+|\Psi_i\rangle\}$. Then we have the conditions:

$$|a|^2\langle\Psi_i|Q^-|\Psi_i\rangle + |b|^2\langle\Psi_i|Q^+|\Psi_i\rangle + |c|^2 = 1, \quad |a\langle\Psi_i|Q^-|\Psi_i\rangle + b\langle\Psi_i|Q^+|\Psi_i\rangle| > \sqrt{1-\eta} \tag{5}$$

where the first condition is normalization condition and second condition says that overlap between $|\lambda_i\rangle$ and $|\Psi_i\rangle$ is at least $\sqrt{1-\eta}$. We would like to minimize the function

$$\langle\lambda_i|Q^-|\lambda_i\rangle = \langle\lambda_i|(aQ^-|\Psi_i\rangle + cQ^-|\theta\rangle) = |a|^2\langle\Psi_i|Q^-|\Psi_i\rangle + |c|^2\langle\theta|Q^-|\theta\rangle \tag{6}$$

Note that $\langle\Psi_i|Q^-Q^-|\theta\rangle = 0$, hence the above expression.

First we shall show that a, b, c can be chosen to be real. Clearly c can be chosen real as it only appears as $|c|^2$. Only place where a, b appear as complex is in the constraint $|a\langle\Psi_i|Q^-|\Psi_i\rangle + b\langle\Psi_i|Q^+|\Psi_i\rangle| > \sqrt{1-\eta}$. Let $a = a_R + ia_I, b = b_R + ib_I$. Then

$$\begin{aligned}
& |a\langle\Psi_i|Q^-|\Psi_i\rangle + b\langle\Psi_i|Q^+|\Psi_i\rangle|^2 \\
= \; & (a_R\langle\Psi_i|Q^-|\Psi_i\rangle + b_R\langle\Psi_i|Q^+|\Psi_i\rangle)^2 + (a_I\langle\Psi_i|Q^-|\Psi_i\rangle + b_I\langle\Psi_i|Q^+|\Psi_i\rangle)^2 \\
= \; & |a|^2\langle\Psi_i|Q^-|\Psi_i\rangle^2 + |b|^2\langle\Psi_i|Q^+|\Psi_i\rangle^2 + 2(a_Rb_R + a_Ib_I)\langle\Psi_i|Q^-|\Psi_i\rangle\langle\Psi_i|Q^+|\Psi_i\rangle \\
\le \; & |a|^2\langle\Psi_i|Q^-|\Psi_i\rangle^2 + |b|^2\langle\Psi_i|Q^+|\Psi_i\rangle^2 + 2(\sqrt{a_R^2 + a_I^2}\sqrt{a_I^2 + b_I^2})\langle\Psi_i|Q^-|\Psi_i\rangle\langle\Psi_i|Q^+|\Psi_i\rangle \\
= \; & |a|^2\langle\Psi_i|Q^-|\Psi_i\rangle^2 + |b|^2\langle\Psi_i|Q^+|\Psi_i\rangle^2 + 2|a||b|\langle\Psi_i|Q^-|\Psi_i\rangle\langle\Psi_i|Q^+|\Psi_i\rangle \\
= \; & (|a|\langle\Psi_i|Q^-|\Psi_i\rangle + |b|\langle\Psi_i|Q^+|\Psi_i\rangle)^2
\end{aligned}$$

Thus, changing the complex coefficients a, b to $|a|, |b|$ does not change the objective function (Equation 6) and ensures that the constraints (Equation 5) are still satisfied. Thus, we can restrict ourselves to real variables a, b.

To find the optimal solution for equations 5 and 6, we fix a c and minimize a^2 with the constraints

$$a^2\langle\Psi_i|Q^-|\Psi_i\rangle + b^2\langle\Psi_i|Q^+|\Psi_i\rangle = 1 - c^2, \quad |a\langle\Psi_i|Q^-|\Psi_i\rangle + b\langle\Psi_i|Q^+|\Psi_i\rangle| > \sqrt{1-\eta}.$$

We plot these constraints on (a, b) plane in figure 1. The ellipse

$$E \stackrel{\text{def}}{=} a^2\langle\Psi_i|Q^-|\Psi_i\rangle + b^2\langle\Psi_i|Q^+|\Psi_i\rangle = 1 - c^2$$

intersects a-axis at $|a_1| = \sqrt{\frac{1-c^2}{\langle\Psi_i|Q^-|\Psi_i\rangle}}$ and intersects b-axis at $|b_1| = \sqrt{\frac{1-c^2}{\langle\Psi_i|Q^+|\Psi_i\rangle}}$. The lines

$$L \stackrel{\text{def}}{=} a\langle\Psi_i|Q^-|\Psi_i\rangle + b\langle\Psi_i|Q^+|\Psi_i\rangle = \sqrt{1-\eta}, \quad L' \stackrel{\text{def}}{=} a\langle\Psi_i|Q^-|\Psi_i\rangle + b\langle\Psi_i|Q^+|\Psi_i\rangle = -\sqrt{1-\eta}$$

intersect a-axis at $|a_2| = \frac{\sqrt{1-\eta}}{\langle\Psi_i|Q^-|\Psi_i\rangle}$ and intersects b-axis at $|b_2| = \frac{\sqrt{1-\eta}}{\langle\Psi_i|Q^+|\Psi_i\rangle}$.

First note that if $c^2 > \eta$, then there is no solution. For this, consider

$$1-\eta < (a\langle\Psi_i|Q^-|\Psi_i\rangle + b\langle\Psi_i|Q^+|\Psi_i\rangle)^2 \le (\langle\Psi_i|Q^-|\Psi_i\rangle + \langle\Psi_i|Q^+|\Psi_i\rangle)(a^2\langle\Psi_i|Q^-|\Psi_i\rangle + b^2\langle\Psi_i|Q^+|\Psi_i\rangle)$$

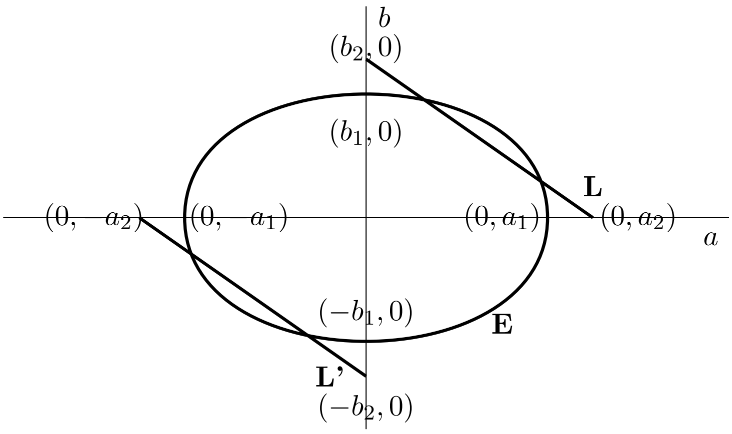

Figure 1: Plot of the constraints

$$= (a^2 \langle \Psi_i | Q^- | \Psi_i \rangle + b^2 \langle \Psi_i | Q^+ | \Psi_i \rangle) = 1 - c^2.$$

So we assume that $c^2 \leq \eta$. Now lets focus on first quadrant. We can easily observe from the plot that we get $a = 0$ as minimum value of a^2 whenever ellipse E intersects b-axis above the line L. This occurs when

$$\sqrt{\frac{1 - c^2}{\langle \Psi_i | Q^+ | \Psi_i \rangle}} > \frac{\sqrt{1-\eta}}{\langle \Psi_i | Q^+ | \Psi_i \rangle} \rightarrow \langle \Psi_i | Q^+ | \Psi_i \rangle > \frac{1-\eta}{1-c^2}.$$

But this is obvious, since the condition implies $\langle \Psi_i | Q^+ | \Psi_i \rangle > 1 - \eta$ in which case there is a vector in Q^+ with high overlap with $|\Psi_i\rangle$ and hence the objective function is 0.

So lets assume that $\langle \Psi_i | Q^+ | \Psi_i \rangle < 1 - \eta$, in which case, for all c, the ellipse E intersects b-axis below the line L. To find the point of intersection, we simultaneously solve the equations for line and ellipse, that is

$$a^2 \langle \Psi_i | Q^- | \Psi_i \rangle + b^2 \langle \Psi_i | Q^+ | \Psi_i \rangle = 1 - c^2, \quad a \langle \Psi_i | Q^- | \Psi_i \rangle + b \langle \Psi_i | Q^+ | \Psi_i \rangle = \sqrt{1-\eta}.$$

The value of a, b thus obtained is

$$a = \sqrt{1-\eta} - \sqrt{\frac{\langle \Psi_i | Q^+ | \Psi_i \rangle (\eta - c^2)}{\langle \Psi_i | Q^- | \Psi_i \rangle}}, \quad b = \sqrt{1-\eta} + \sqrt{\frac{\langle \Psi_i | Q^- | \Psi_i \rangle (\eta - c^2)}{\langle \Psi_i | Q^+ | \Psi_i \rangle}}.$$

It is easy to verify that the solution satisfies above equations. The other solution is with signs reversed.

Thus, we have the result that whenever $\langle \Psi_i | Q^+ | \Psi_i \rangle < 1-\eta$, the minimum $|a|^2 \langle \Psi_i | Q^- | \Psi_i \rangle + |c|^2 \langle \theta | Q^- | \theta \rangle$ is

$$(\sqrt{1-\eta} - \sqrt{\frac{\langle \Psi_i | Q^+ | \Psi_i \rangle (\eta - c^2)}{\langle \Psi_i | Q^- | \Psi_i \rangle}})^2 \langle \Psi_i | Q^- | \Psi_i \rangle + c^2 \langle \theta | Q^- | \theta \rangle.$$

This quantity is monotonically increasing with c. Hence above expression is minimized when $c = 0$. This justifies our intuition that the optimal vector lies in the plane $\{Q^+ |\Psi_i\rangle, Q^- |\Psi_i\rangle\}$. With this, we have found an overall minimum to be

$$(\sqrt{1-\eta} - \sqrt{\frac{\langle \Psi_i | Q^+ | \Psi_i \rangle \eta}{\langle \Psi_i | Q^- | \Psi_i \rangle}})^2 \langle \Psi_i | Q^- | \Psi_i \rangle = (\sqrt{(1-\eta) \langle \Psi_i | Q^- | \Psi_i \rangle} - \sqrt{\langle \Psi_i | Q^+ | \Psi_i \rangle \eta})^2.$$

This proves the claim. $\square$

D Proof of Lemma 5.8

Proof. In Lemma 5.7, we set $k = \frac{d}{4}$, $\varepsilon = \frac{1}{d}$ and $\alpha = \frac{\delta}{4}$. Then we obtain that

$$\Pr_{\lambda}(\langle \Psi_i | \, Q \, | \Psi_i \rangle < \delta/2 + \delta/d) \leq \frac{96}{d}.$$

Now, to evaluate $\mathbb{E}_{i \leftarrow \lambda} 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i \| \sigma)}$, we divide the expectation into two parts. For all i for which $\langle \Psi_i | \, Q \, | \Psi_i \rangle < \frac{\delta}{2}$, we upper bound $2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i \| \sigma)} < 1$. For the rest of i, also recalling that $Q^- > Q$ in Lowner order, we use lemma 5.2 to obtain

$$2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i \| \sigma)} < \frac{1}{k(1-\eta)(\sqrt{(1-2\eta)(\frac{\delta}{2})} - \sqrt{2(1-\frac{\delta}{2})\eta})^2}$$

Note that for this to hold, we need $\frac{\delta}{2} > 2\eta$ (as assumed in lemma 5.2). This is satisfied as we have set $\eta < \frac{\delta}{8}$.

Then we have

$$2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i \| \sigma)} \leq \frac{4}{d(1-\eta)(\sqrt{(1-2\eta)(\delta/4)} - \sqrt{2(1-\delta/4)\eta})^2} \leq \frac{40}{d\delta}.$$

Thus we get

$$\begin{aligned} \mathbb{E}_{i \leftarrow \lambda} 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i \| \sigma)} &< \frac{96}{d} + \frac{40}{d\delta} \\ &\leq \frac{50}{d\delta} \leq 2^{-\log(d\delta)+6} \end{aligned}$$

This proves the Lemma. ∎