Lower Bound on Expected Communication Cost of Quantum Huffman Coding

Anurag Anshu[*1], **Ankit Garg**[†2], **Aram W. Harrow**[‡3], **and Penghui Yao**[§4]

1 **Centre for Quantum Technologies, National University of Singapore, Singapore**
a0109169@u.nus.edu

2 **Microsoft Research New England, USA**
garga@microsoft.com

3 **Center for Theoretical Physics, Massachusetts Institute of Technology, USA**
aram@mit.edu

4 **Institute for Quantum Computing, University of Waterloo, Canada; and Department of Combinatorics and Optimization, University of Waterloo, Canada**
phyao1985@gmail.com

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 them 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 the goal is to minimize its expected length.

For classical messages this problem has a well-known solution given by Huffman coding. In this scheme, the expected length of the message is equal to the Shannon entropy of the source (with a constant additive factor) and the scheme succeeds with zero error. This is a single-shot result which implies the asymptotic result, viz. Shannon's source coding theorem, by encoding each 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 when the cost of a quantum task is significantly different than the classical analogue. Our result has implications for direct sum theorems in quantum communication complexity and one-shot formulations of Quantum Reverse Shannon theorem.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases Quantum information, quantum communication, expected communication cost, huffman coding

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.3

* A.A. is supported by Core Grants of Centre for Quantum Technologies.
† This work was done when A.G. was a student at Princeton University and his research was partially supported by NSF grants CCF-1149888 and CCF-1525342, a Simons fellowship for graduate students in theoretical computer science and a Siebel scholarship.
‡ A.W.H. was funded by NSF grants CCF-1111382 and CCF-1452616.
§ P.Y. is supported by NSERC and CIFAR.

1 Introduction

The central theme of information theory is compression of messages up to their *information content*. The celebrated work of Shannon [19] 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 [14] 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 Huffman coding scheme can easily be illustrated in the following way (adapted from the reference [11]). Alice and Bob share infinitely many copies of the joint random variable XY (X with Alice and Y with Bob), such that $p(x,y)=\delta_{x,y}p(x)$. These copies are arranged in a sequence known to both parties. If Alice gets an input x, she measures her half of the copies in this sequence, and sends to Bob the address of the first location where she finds her input x. The average length of the message is can easily be computed to be at most $\log(\frac{1}{p(x)})+1$. Thus, average length of the message in overall protocol is at most $\sum_x p(x)(\log(\frac{1}{p(x)})+1)=H(p)+1$.

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 [11], the following task was considered (inspired by a result of Wyner [25]): 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 one-shot setting (different from the one already given by Shannon [19] 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. [11] used their compression result to prove the following theorem:

▶ **Theorem** (Informal, [11]). *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.*

The message compression in the presence of side information was first studied in the asymptotic setting by Slepian and Wolf [20]. The work by Braverman and Rao [8] gave its one-shot analogue in the following manner. Given a probability distribution P with Alice and Q with Bob, they constructed an interactive protocol (assisted by shared randomness) that allowed both Alice and Bob to output a 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 relative entropy between P and Q. This work thus provided an operational interpretation to *relative entropy* [1] and extended the above theorem to general distributions. The holy grail for such

[1] The work [11] given 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.

direct sum theorems is to remove the dependence on the number of rounds, and the above mentioned results ([11],[8]) along with [4] 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 [8, 4], see also references therein) with the aid of shared randomness and consideration of expected communication cost as communication measure.

On the other hand, while many of the above results have their quantum counterpart, 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 the quantum state merging, introduced in [13] as the quantum analogue of Slepian-Wolf protocol [20] (in the asymptotic setting). The most general form of coherent quantum protocols, involving two parties and one Referee, is known as the 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 [9, 26] to give an operational meaning of the quantum conditional mutual information in the asymptotic setting. Furthermore, as shown by Touchette [22], 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 [6], and the notion of quantum information complexity, Touchette [22] obtains the following direct sum result for entanglement assisted quantum communication complexity.

▶ **Theorem** (Informal, [22]). *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 [11]), that comes from the consideration of the worst case quantum communication cost for the quantum state redistribution in the work [6]. Furthermore, it has been shown recently in [1] 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 the coherence with the Referee. The works which exhibit one-shot quantum compression protocols in the non-coherent setting, include [15, 16] (which also show direct sum theorems for entanglement assisted one-way quantum communication complexity) and [3] (which is an extension of Braverman-Rao protocol [8] 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** (Quantum Huffman task)**.** Alice (A) receives an input x and an associated quantum pure state $|\Psi_x\rangle$ with probability $p(x)$. For a given $\eta > 0$, which we shall henceforth identify as '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 the classical one-shot source coding. The expected communication cost in the asymptotic setting is lower bounded by $S(\sum_x p(x)\Psi_x)$ due to [12], which is also the quantum information cost of this task. 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.*[7] had considered our question and had noted several issues in generalizing directly the techniques of 'classical' Huffman coding to 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 introduction, we would like to compare the expected communication cost of any protocol achieving quantum Huffman task with the von-Neumann entropy of average state with Alice : $S(\sum_x p(x)\,|\Psi_x\rangle\langle\Psi_x|)$. Our main result is a large gap between the two quantities, that we state below.

▶ **Theorem 2.** *Fix a positive integer $d > 10^{12}$ and real δ that satisfy $\frac{16}{\sqrt{d}} < \delta < \frac{1}{100}$. There exist a collection of $N \stackrel{\text{def}}{=} (\frac{3}{\delta^2})^d$ 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 $S(\sum_x p(x)\,|\Psi_x\rangle\langle\Psi_x|) \leq 4\delta\log(d) + H(\delta) + 1$.*
- *For any one-way protocol achieving quantum Huffman coding of above ensemble with error parameter $\eta < \frac{\delta}{16}$, the expected communication cost is lower bounded by $(1-\eta)\cdot\log(d\delta) - 6$.*
- *For any r-round protocol achieving quantum Huffman coding of above ensemble with error parameter $\eta < \frac{\delta}{16}$, the expected communication cost is lower bounded by $\Omega(\frac{\log(d\delta)}{\log r})$.*

The one-way part of this theorem is proved in Section 5, as a special case of Theorem 19. The r-round part follows argument similar to that of one-way part, and its technical details can be found in the arXiv eprint of this work [2].

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

▶ **Theorem 3.** *Fix a positive integer $d > 10^{12}$, real δ that satisfies $\frac{\sqrt{768}}{\log(d)} < \delta < 1$ and a monotonically increasing function $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) \geq x^2$. There exist a collection of $N \stackrel{\text{def}}{=} 2^{f(d)}$ 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 the following holds for the ensemble $\{(p(x), \Psi_x)\}_{x=1}^N$.*

- *The von-Neumann entropy of the average state satisfies $S(\sum_x p(x)\,|\Psi_x\rangle\langle\Psi_x|) \leq 4\delta\log(d) + H(\delta) + 1$.*
- *For any interactive protocol with error parameter $\eta \stackrel{\text{def}}{=} \frac{1}{\log^2(d)} = \frac{4}{\log^2(f^{-1}(\log(N)))}$, the expected communication cost is lower bounded by $\Omega(\frac{\log(d\delta)}{\log\log(d)})$.*

The proof of this theorem can again be found in the arXiv eprint of this work [2]. It may be noted that the dependence of error parameter η on input size $\log N$ can be made as weak as desired, by choosing an appropriate function f which increases sufficiently fast.

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. Section 3 for one-way protocols is 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 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 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 not much smaller than expected communication cost.

We remedy this problem by introducing a free variable 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 $d-1$ dimensional subspace orthogonal to $|0\rangle$. We choose $|x\rangle$ according to 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 4, where epsilon nets have been used to make the input size finite.

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 $\|X\|_1 \stackrel{\text{def}}{=} \mathrm{Tr}\sqrt{X^\dagger X}$ and ℓ_2 norm is $\|X\|_2 \stackrel{\text{def}}{=} \sqrt{\mathrm{Tr} X X^\dagger}$. 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 $|A| \stackrel{\text{def}}{=} \dim(\mathcal{H}_A)$. 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

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

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 4.** We shall consider the following information theoretic quantities. Let $\varepsilon \geq 0$.

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

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

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 $\rho_A \in \mathcal{D}(A)$,

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

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

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

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

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

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

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

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

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

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

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

We will use the following facts.

▶ **Fact 5** (Monotonicity of quantum operations). [[18, 5], [21], 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 6** (Joint concavity of fidelity). [[24], 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 7** (Fannes inequality). [[10]] 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 8** (Levy's concentration lemma). [[17]] Let $f : S^d \to \mathbb{R}$ be Lipschitz continuous function with Lipschitz constant ℓ, defined as

$$\ell \stackrel{\text{def}}{=} \max_{x,y} \frac{|f(x) - f(y)|}{\|x - y\|_2}.$$

Let $\mathbb{E}(f)$ be expectation value of f with respect to uniform measure over S^d. Then

$$\mathrm{Prob}(|f - \mathbb{E}(f)| \geq \alpha) \leq 2e^{-\frac{d\alpha^2}{18\pi^3\ell^2}}.$$

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 $\{M_1^x, M_2^x \ldots\}$ on her side and sends the outcome i to Bob. Let

$$p_i^x \stackrel{\text{def}}{=} \mathrm{Tr}(M_i^x \theta_A), \quad \rho_i^x \stackrel{\text{def}}{=} \frac{\mathrm{Tr}_A(M_i^x \theta_{AB})}{p_i^x}.$$

- 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, proof of which has been given in Appendix A.

▶ **Lemma 9.** *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 $a(1-\sqrt{\eta})^2 - 1$.*

4 Example separating expected communication and information

4.1 An epsilon net over S^d

We will use the epsilon net over S^d, as defined below.

▶ **Definition 10** (Epsilon-nets, [23])**.** Fix an $\varepsilon > 0$. There exists an integer N and a set of vectors $\{|x_1\rangle, |x_2\rangle \dots |x_N\rangle\}$ on S^d such that the following properties hold:

- $N \leq (\frac{2}{\varepsilon})^d$.
- For any two vectors $|x_i\rangle, |x_j\rangle$ it holds that $\|(|x_i\rangle - |x_j\rangle)(\langle x_i| - \langle x_j|)\|_2 \leq \varepsilon$.
- For any vector $|y\rangle \in S^d$, there exists j such that $\|(|y\rangle - |x_j\rangle)(\langle y| - \langle x_j|)\|_2 \leq \varepsilon$

Let the set be denoted as $\mathcal{N}_\varepsilon$.

We recall that μ is a uniform measure over S^d. For every vector $|x_i\rangle \in \mathcal{N}_\varepsilon$, we let $S_i \subset S^d$ be the set of all vectors $|y\rangle \in S^d$ such that $|x_i\rangle$ is one of the closest (in euclidean distance) to $|y\rangle$ among all vectors in $\mathcal{N}_\varepsilon$. Let $\mu(S_i)$ be the measure associated to S_i. $\mu(S_i)$ can also be interpreted as the volume of S_i. Due to the fact that set of vectors in S^d which are equidistant to two or more vectors in $\mathcal{N}_\varepsilon$ have measure zero, we obtain the relation:

$$\sum_i \mu(S_i) = 1, \quad \mu(S_i \cap S_j) = 0 \tag{1}$$

Let λ be a distribution over $\mathcal{N}_\varepsilon$, such that $\lambda(i) \stackrel{\text{def}}{=} \mu(S_i)$. Let $\mathbb{E}_i$ denote the expectation over the set $\mathcal{N}_\varepsilon$ with vectors chosen according to λ. That is, for any function $f(.)$ on $\mathcal{N}_\varepsilon$, we define

$$\mathbb{E}_i f(|x_i\rangle) \stackrel{\text{def}}{=} \sum_i \lambda(i) f(|x_i\rangle).$$

The following lemma follows from the the above definition.

▶ **Lemma 11.** *It holds that*

$$\|(\mathbb{E}_i |x_i\rangle)(\mathbb{E}_i \langle x_i|)\|_1 \leq \varepsilon, \quad \|\mathbb{E}_i |x_i\rangle\langle x_i| - \frac{I}{d}\|_1 \leq 2\sqrt{\varepsilon}.$$

Proof. For the first part, we use the identities

$$\int_y \mu(y) dy\, |y\rangle = 0, \quad \int_y \mu(y) dy\, |y\rangle = \sum_i \mu(S_i) \frac{\int_{y \in S_i} \mu(y) dy\, |y\rangle}{\mu(S_i)},$$

where the second identity follows from Equation 1. Now we notice from the definition of set S_i that

$$\|(|x_i\rangle - \frac{\int_{y \in S_i} \mu(y) dy\, |y\rangle}{\mu(S_i)})(\langle x_i| - \frac{\int_{y \in S_i} \mu(y) dy\, \langle y|}{\mu(S_i)})\|_2 \leq \varepsilon.$$

Applying expectation $\mathbb{E}_i$ to both sides and then using the triangle inequality, we immediately obtain

$$\|(\mathbb{E}_i |x_i\rangle)(\mathbb{E}_i \langle x_i|)\|_2 = \|(\sum_i \mu(S_i) |x_i\rangle)(\sum_i \mu(S_i) \langle x_i|)\| \leq \varepsilon.$$

For the second part, we again notice the identities

$$\int_y \mu(y)dy\, |y\rangle\langle y| = \frac{I}{d}, \quad \int_y \mu(y)dy\, |y\rangle\langle y| = \sum_i \mu(S_i) \frac{\int_{y\in S_i} \mu(y)dy\, |y\rangle\langle y|}{\mu(S_i)}.$$

From the definition of the set S_i, we have that for every $|y\rangle \in S_i$, $|\langle x_i|y\rangle|^2 \geq 1 - 2\varepsilon$. Thus, $\mathrm{F}(|x\rangle\langle x|, \frac{\int_{y\in S_i} \mu(y)dy|y\rangle\langle y|}{\mu(S_i)}) \geq 1 - 2\varepsilon$, which translates to $\| |x\rangle\langle x| - \frac{\int_{y\in S_i} \mu(y)dy|y\rangle\langle y|}{\mu(S_i)}\|_1 \leq 2\sqrt{\varepsilon}$. Now the proof follows along the same lines as first part. ◀

4.2 Our construction

Our construction now proceeds as follows, recalling the quantum Huffman task in Definition 1. Fix a $\delta > 0$. Alice is given the input i with probability $\lambda(i)$, which corresponds to the vector $|x_i\rangle \in \mathcal{N}_\varepsilon$. We embed C^d in a $d+1$ dimensional space C^{d+1} and let P be a projector onto the original space C^d. We define $|\Psi_i\rangle \stackrel{\text{def}}{=} \sqrt{1-\delta}\,|0\rangle + \sqrt{\delta}\,|x_i\rangle$, where $|0\rangle$ is a vector satisfying $P|0\rangle = 0$.

We have the following lemma.

▶ **Lemma 12.** *The von-Neumann entropy of the average state $\mathbb{E}_i\Psi_i = \sum_i \lambda(i)\Psi_i$ satisfies $S(\mathbb{E}_i\Psi_i) \leq (\delta + 3\sqrt{\varepsilon})\log(d) + H(\delta) + 1$.*

Proof. Consider,

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

From Lemma 11, it follows that

$$\|\mathbb{E}_i\Psi_i - (1-\delta)\,|0\rangle\langle 0| + \delta\frac{P}{d}\|_1 \leq 3\sqrt{\varepsilon}.$$

Now we use Fannes inequality (Fact 7) to conclude that $S(\mathbb{E}_i\Psi_i)$ is at most $(\delta + 3\sqrt{\varepsilon})\log(d) + H(\delta) + 1$. ◀

4.3 A property of smooth relative max entropy

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

▶ **Lemma 13.** *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

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

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

▶ **Claim 14.** *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 15.** *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. ◀

4.4 Final lower bound

Let μ be uniform measure over S^d. For any vector $|y\rangle$ belonging to subspace of P, let $|\Psi_y\rangle = \sqrt{1-\delta}\,|0\rangle + \sqrt{\delta}\,|y\rangle$ be a vector in C^{d+1}. We have the following claims, the first of which computes the expectation value and the second computes the Lipschitz constant.

▶ **Claim 16.** *It holds that*

$$\int_y \mu(y)dy\,\langle\Psi_y| Q^- |\Psi_y\rangle = (1-\delta-\frac{\delta}{d})\langle 0| Q |0\rangle + \delta(\frac{d+1-k}{d}).$$

Proof. Consider the following analysis, from which the statement follows.

$$\begin{aligned}
\int_y \mu(y)dy\,\langle\Psi_y| Q^- |\Psi_y\rangle &= (1-\delta)\langle 0| Q^- |0\rangle + \delta\int_y \mu(y)dy\,\langle y| Q^- |y\rangle \\
&= (1-\delta)\langle 0| Q^- |0\rangle + \frac{\delta}{d}\mathrm{Tr}(PQ) \\
&= (1-\delta)\langle 0| Q^- |0\rangle + \frac{\delta}{d}(\mathrm{Tr}(Q) - \langle 0| Q |0\rangle) \\
&= (1-\delta-\frac{\delta}{d})\langle 0| Q |0\rangle + \delta(\frac{d+1-k}{d})
\end{aligned}$$

This proves the claim. ◀

▶ **Claim 17.** *Let Q be a projector and $|y\rangle, |y'\rangle$ be any two vectors in S^d . Then it holds that*

$$|\langle\Psi_y| Q^- |\Psi_y\rangle - \langle\Psi_{y'}| Q^- |\Psi_{y'}\rangle| \le (2\sqrt{2\delta(1-\delta)}+2\delta)\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_2 \le 4\sqrt{\delta}\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_2.$$

Proof. Consider the analysis

$$\begin{aligned}
& |\langle\Psi_y| Q^- |\Psi_y\rangle - \langle\Psi_{y'}| Q^- |\Psi_{y'}\rangle| \le \|\Psi_y - \Psi_{y'}\|_1 \\
\le\ & 2\sqrt{\delta(1-\delta)}\|\,|0\rangle(\langle y| - \langle y'|)\|_1 + \delta\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_1 \\
=\ & 4\sqrt{\delta(1-\delta)}(1 - F(|y\rangle\langle y|, |y'\rangle\langle y'|)) + \delta\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_1 \\
\le\ & (2\sqrt{\delta(1-\delta)} + \delta)\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_1 \\
\le\ & (2\sqrt{\delta(1-\delta)} + \delta)\sqrt{2}\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_2
\end{aligned}$$

This proves the claim. ◀

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

▶ **Lemma 18.** *Assume the conditions $\delta > \frac{16}{\sqrt{d}}$ and $d > 10^{12}$. Let η, ε be such that $\eta < \frac{\delta}{16}$ and $\varepsilon < \frac{\delta}{100}$. Let a be the largest real that satisfies $\mathbb{E}_i 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} \le 2^{-a}$. Then it holds that $a \ge \log(\frac{d\delta(1-2\eta)^2}{50})$*

5 Proof of main result

We provide the proof of Theorem 2 in this section. It can easily be obtained from the following more general result by setting $\varepsilon = \delta^2$ and letting $\delta < \frac{1}{100}$.

▶ **Theorem 19.** *Fix a positive integer $d > 10^{12}$ and reals δ, ε that satisfy $\frac{16}{\sqrt{d}} < \delta < 1$ and $\varepsilon < \frac{\delta}{100}$. There exist a collection of $N \stackrel{\text{def}}{=} (\frac{3}{\varepsilon})^d$ states $\{|\Psi_x\rangle\}_{x=1}^N$ that depend on δ and belong to 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 $S(\sum_x p(x)\Psi_x) \leq (\delta + 3\sqrt{\varepsilon}) \log(d) + H(\delta) + 1$*
- *For any one-way protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta}{16}$, the expected communication cost is lower bounded by $(1 - \sqrt{\eta})^2 \log(\frac{d\delta}{300})$.*
- *For any r-round protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta}{16}$, the expected communication cost is lower bounded by*

$$\frac{1}{20} \cdot \frac{\log(\frac{d\delta}{400})}{(\log r)}.$$

Proof. We use the construction as given in Subsection 4.2.

For the first part of the theorem, we combine Lemma 9 and Lemma 18 to obtain a lower bound on expected communication cost as

$$(1 - \sqrt{\eta})^2 \log(\frac{d\delta(1 - 2\sqrt{\eta})^2}{50}) - 1 > (1 - \sqrt{\eta})^2 \log(\frac{d\delta}{300}).$$

The proof of second part of the theorem follows from the Reference [2] (Theorem 6.1). ◀

The proof of Theorem 3 is given in Reference [2] (Lemma 6.2).

6 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). 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 2 has a dependence on the number of rounds. We get rid of this dependence in Theorem 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 2 itself. For comparison, it may be noted that the results in [1] have no dependence on the number of rounds.
- Our lower bounds on expected communication cost and the quantum information complexity of the quantum Huffman tasks that we construct are doubly-logarithmically small in input size N, that is $\mathcal{O}(\log\log(N))$ (see Theorem 2). Can we have examples where the dependence on input size is better?
- 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 [22]?

Acknowledgements. 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.

References

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

2 Anurag Anshu, Ankit Garg, Aram Harrow, and Penghui Yao. Lower bound on expected communication cost of quantum huffman coding, 2016. URL: http://arxiv.org/abs/1605.04601.

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

4 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.

5 Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. Noncommuting mixed states cannot be broadcast. *Phys. Rev. Lett.*, 76:2818–2821, Apr 1996. doi:10.1103/PhysRevLett.76.2818.

6 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. doi:10.1109/TIT.2016.2516006.

7 S. L. Braunstein, C. A. Fuchs, D. Gottesman, and Hoi-Kwong Lo. A quantum analog of huffman coding. In *Information Theory, 1998. Proceedings. 1998 IEEE International Symposium on*, pages 353–, Aug 1998. doi:10.1109/ISIT.1998.708958.

8 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.

9 Igor Devetak and Jon Yard. Exact cost of redistributing multipartite quantum states. *Phys. Rev. Lett.*, 100:230501, Jun 2008. doi:10.1103/PhysRevLett.100.230501.

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

11 P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, Jan 2010. doi:10.1109/TIT.2009.2034824.

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

13 Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269:107–136, 2007. doi:10.1007/s00220-006-0118-x.

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

15 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. doi:10.1109/CCC.2005.24.

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

17 Michel Ledoux. The concentration of measure phenomenon. Mathematical Surveys and Monographs. American Mathematical Society, 2005.

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

19 Claude Elwood Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

20 D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, Jul 1973. doi:10.1109/TIT.1973.1055037.

21 Marco Tomamichel. A framework for non-asymptotic quantum information theory, 2012. PhD Thesis, ETH Zurich. URL: arXiv:1203.2142.

22 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. doi:10.1145/2746539.2746613.

23 Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing, Theory and Applications*. Cambridge University Press, 2012.

24 John Watrous. Theory of Quantum Information, lecture notes, 2011. URL: https://cs.uwaterloo.ca/~watrous/LectureNotes.html.

25 A. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, Mar 1975. doi:10.1109/TIT.1975.1055346.

26 J. T. Yard and I. Devetak. Optimal quantum source coding with quantum side information at the encoder and decoder. *IEEE Transactions on Information Theory*, 55(11):5339–5351, Nov 2009.

A Proof of Lemma 9

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 5), to obtain

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

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, $\mathrm{D}_{\max}^{\eta}(\Psi^x \| \mathcal{E}_i(\theta_B)) < \mathrm{D}_{\max}(\sigma_i^x \| \mathcal{E}_i(\theta_B))$ for all $i \in \mathcal{G}_x$. Using Equation 2, we observe that for all $x \in \mathcal{G}$ it holds that

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

This implies

$$\begin{aligned} s_i &= \sum_x p'(x) q_i^x \\ &\leq \frac{1}{1-\sqrt{\eta}} \sum_x p'(x) 2^{-\mathrm{D}_{\max}^{\eta}(\Psi^x \| \mathcal{E}_i(\theta_B))} \\ &\leq \frac{1}{(1-\sqrt{\eta})^2} \sum_x p(x) 2^{-\mathrm{D}_{\max}^{\eta}(\Psi^x \| \mathcal{E}_i(\theta_B))} \\ &= \frac{t_i}{(1-\sqrt{\eta})^2} < \frac{2^{-a}}{(1-\sqrt{\eta})^2} \end{aligned} \tag{3}$$

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 3, 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

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

B Proof of Claims 14 and 15

Proof of Claim 14 . For a fixed i, let ρ_i be the state that achieves the infimum in the definition of $\mathrm{D}_{\max}^{\eta}(\Psi_i \| \sigma)$. 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 15. 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:

$$|a|^2\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + |b|^2\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle + |c|^2 = 1, \quad |a\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + b\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle\,| > \sqrt{1-\eta} \tag{4}$$

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

$$\langle\lambda_i|\,Q^-\,|\lambda_i\rangle = \langle\lambda_i|\,(aQ^-\,|\Psi_i\rangle + cQ^-\,|\theta\rangle) = |a|^2\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + |c|^2\,\langle\theta|\,Q^-\,|\theta\rangle \tag{5}$$

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

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

Thus, changing the complex coefficients a, b to $|a|, |b|$ does not change the objective function (Equation 5) and ensures that the constraints (Equation 4) are still satisfied. Thus, we can restrict ourselves to real variables a, b.

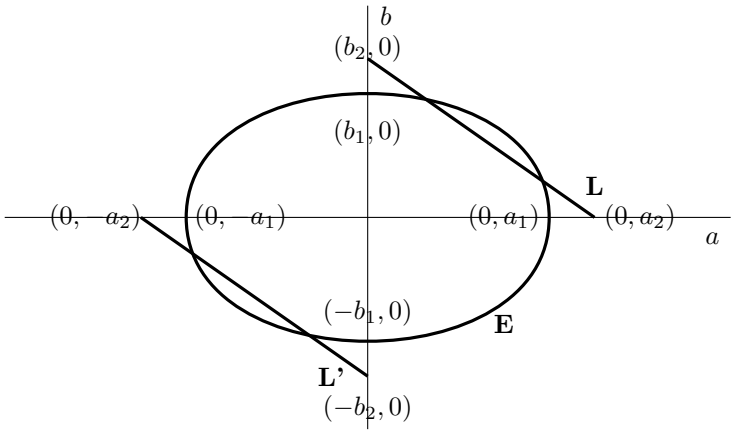

Figure 1 Plot of the constraints.

To find the optimal solution for equations 4 and 5, 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

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

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

$$\begin{aligned} 1 - \eta &< (a \langle\Psi_i| Q^- |\Psi_i\rangle + b \langle\Psi_i| Q^+ |\Psi_i\rangle)^2 \\ &\leq (\langle\Psi_i| Q^- |\Psi_i\rangle + \langle\Psi_i| Q^+ |\Psi_i\rangle)(a^2 \langle\Psi_i| Q^- |\Psi_i\rangle + b^2 \langle\Psi_i| Q^+ |\Psi_i\rangle) \\ &= (a^2 \langle\Psi_i| Q^- |\Psi_i\rangle + b^2 \langle\Psi_i| Q^+ |\Psi_i\rangle) = 1 - c^2. \end{aligned}$$

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. ◀

C Proof of Lemma 18

Proof. Proof shall proceed in two stages.

1. **Concentration of measure for Epsilon-nets:**

   From Claim 17, we infer that the Lipschitz constant of the function $f(|y\rangle) \stackrel{\text{def}}{=} \langle\Psi_y| Q^- |\Psi_y\rangle$ is upper bounded by $4\sqrt{\delta}$. From Lemma 16, we have that $\int \mu(y) dy f(y) = (1 - \delta - \frac{\delta}{d}) \langle 0| Q^- |0\rangle + \delta(1 - \frac{k-1}{d})$.

   Let α be a positive real to be chosen later. It now follows from Levy's concentration lemma (Fact 8) that

   $$\Pr_\mu(\langle\Psi_y| Q^- |\Psi_y\rangle < \int \mu(y) dy f(y) - \alpha) \leq e^{-\frac{d\alpha^2}{18\pi^3 \cdot 16\delta}} = e^{-\frac{d\alpha^2}{288\pi^3\delta}} \tag{6}$$

   In other words,

   $$\Pr_\mu(\langle\Psi_y| Q^- |\Psi_y\rangle < \delta(1 - \frac{k-1}{d}) - \alpha) \leq e^{-\frac{d\alpha^2}{288\pi^3\delta}}.$$

   Now, let S be the set of all $|y\rangle \in S^d$ for which $\langle\Psi_y| Q^- |\Psi_y\rangle > \delta(1 - \frac{k-1}{d}) - \alpha$. Let $\mathcal{G}$ be the set of all i such that S_i has an intersection with S. Let $T \stackrel{\text{def}}{=} \cup_{i\in\mathcal{G}} S_i$. Then T contains S, except for some points of measure zero, and furthermore from Claim 17, any $|z\rangle \in T$ satisfies

   $$\langle\Psi_z| Q^- |\Psi_z\rangle \geq \delta(1 - \frac{k-1}{d}) - \alpha - 4\sqrt{\delta}\varepsilon > \delta(1 - \frac{k-1}{d}) - \alpha - 2\varepsilon.$$

   Since $\mu(T) > (1 - e^{-\frac{d\alpha^2}{288\pi^3\delta}})$, and T is a union of S_i with $i \in \mathcal{G}$, it holds that for an i drawn according to $\lambda(i)$, probability that $i \in \mathcal{G}$ is equal to $\mu(T)$ and hence at least $(1 - e^{-\frac{d\alpha^2}{288\pi^3\delta}})$. Thus we have show the following inequality

   $$\Pr_\lambda(\langle\Psi_i| Q^- |\Psi_i\rangle \geq \delta(1 - \frac{k-1}{d}) - \alpha - 2\varepsilon) \geq 1 - 2e^{-\frac{d\alpha^2}{288\pi^3\delta}}, \tag{7}$$

2. Using concentration of measure for upper bound:
Now, to evaluate $\mathbb{E}_i 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)}$, we divide the expectation into two parts. For all i for which $\langle\Psi_i| Q^- |\Psi_i\rangle < \delta(1-\frac{k}{d}) - \alpha - 2\varepsilon$, we upper bound $2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} < 1$. For the rest of i, we use Lemma 13 to obtain

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

Note that for this to hold, we need $\delta(1-\frac{k}{d}) - \alpha - 2\varepsilon > 2\eta$. For this, we set $k = \frac{d}{4}, \alpha^2 = \frac{\delta}{\sqrt{d}} < \frac{\delta^2}{16}$ and we can upper bound $\delta(1-\frac{k}{d}) - \alpha - 2\varepsilon > \frac{\delta}{4} > 2\eta$, using $\varepsilon < \frac{\delta}{100}, \eta < \frac{\delta}{16}$ (assumptions of theorem). Then we have

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

Thus we get

$$\begin{aligned}
\mathbb{E}_i 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} &< 2e^{-\frac{d\alpha^2}{144\pi^3\delta}} + \frac{40}{d(1-2\eta)^2\delta} \\
&= 2e^{-\frac{\sqrt{d}}{288\pi^3}} + \frac{40}{d(1-2\eta)^2\delta} \\
&< \frac{50}{d\delta(1-2\eta)^2} = 2^{-\log(\frac{d\delta(1-2\eta)^2}{50})}
\end{aligned}$$

Last inequality holds for $d > 10^{12}$. This proves the theorem. ◀