Universality of EPR Pairs in Entanglement-Assisted Communication Complexity, and the Communication Cost of State Conversion

Matthew Coudron
Institute for Quantum Computing, University of Waterloo, Canada
mcoudron@uwaterloo.ca

Aram W. Harrow
Center for Theoretical Physics, MIT, Cambridge, MA, USA
https://web.mit.edu/aram/www/
aram@mit.edu

Abstract

In this work we consider the role of entanglement assistance in quantum communication protocols, focusing, in particular, on whether the type of shared entangled state can affect the quantum communication complexity of a function. This question is interesting because in some other settings in quantum information, such as non-local games, or tasks that involve quantum communication between players and referee, or simulating bipartite unitaries or communication channels, maximally entangled states are known to be less useful as a resource than some partially entangled states. By contrast, we prove that the bounded-error entanglement-assisted quantum communication complexity of a partial or total function cannot be improved by more than a constant factor by replacing maximally entangled states with arbitrary entangled states. In particular, we show that every quantum communication protocol using Q qubits of communication and arbitrary shared entanglement can be ϵ-approximated by a protocol using $O(Q/\epsilon+\log(1/\epsilon)/\epsilon)$ qubits of communication and *only* EPR pairs as shared entanglement. This conclusion is opposite of the common wisdom in the study of non-local games, where it has been shown, for example, that the I3322 inequality has a non-local strategy using a non-maximally entangled state, which surpasses the winning probability achievable by any strategy using a maximally entangled state of any dimension [17]. We leave open the question of how much the use of a shared maximally entangled state can reduce the quantum communication complexity of a function.

Our second result concerns an old question in quantum information theory: How much quantum communication is required to approximately convert one pure bipartite entangled state into another? We give simple and efficiently computable upper and lower bounds. Given two bipartite states $|\chi\rangle$ and $|\upsilon\rangle$, we define a natural quantity, $d_\infty(|\chi\rangle, |\upsilon\rangle)$, which we call the ℓ_∞ Earth Mover's distance, and we show that the communication cost of converting between $|\chi\rangle$ and $|\upsilon\rangle$ is upper bounded by a constant multiple of $d_\infty(|\chi\rangle, |\upsilon\rangle)$. Here $d_\infty(|\chi\rangle, |\upsilon\rangle)$ may be informally described as the minimum over all transports between the log of the Schmidt coefficients of $|\chi\rangle$ and those of $|\upsilon\rangle$, of the maximum distance that any amount of mass must be moved in that transport. A precise definition is given in the introduction. Furthermore, we prove a complementary lower bound on the cost of state conversion by the ϵ-Smoothed ℓ_∞-Earth Mover's Distance, which is a natural smoothing of the ℓ_∞-Earth Mover's Distance that we will define via a connection with optimal transport theory.

2012 ACM Subject Classification Theory of computation → Quantum communication complexity; Theory of computation → Quantum information theory

Keywords and phrases Entanglement, quantum communication complexity

Digital Object Identifier 10.4230/LIPIcs.CCC.2019.20

Related Version A full version is available at arXiv:1902.07699.

34th Computational Complexity Conference (CCC 2019).
Editor: Amir Shpilka; Article No. 20; pp. 20:1–20:25

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

COMPUTATIONAL COMPLEXITY CONFERENCE

Funding *Matthew Coudron*: Supported at the IQC by Canada's NSERC and the Canadian Institute for Advanced Research (CIFAR), and through funding provided to IQC by the Government of Canada and the Province of Ontario.
Aram W. Harrow: NSF grants CCF-1452616, CCF-1729369, PHY-1818914 and ARO contract W911NF-17-1-0433.

1 Introduction

1.1 Entanglement-assisted communication complexity

Imagine that two cooperating players, Alice and Bob, are given the task of evaluating a function $f(x, y)$ $(x, y \in \{0, 1\}^n)$, where x is known only to Alice and y is known only to Bob. The communication complexity of f is the number of bits that Alice and Bob need to exchange in order to compute f. Popular variations of this framework include allowing a small probability of error, allowing qubits to be communicated instead of classical bits, and allowing extra resources such as shared randomness or entanglement.

In classical communication complexity, Newman's theorem [14] states that arbitrarily large amounts of shared randomness in a protocol can be replaced by a distribution with $O(\log(n/\epsilon))$ bits of entropy while only reducing the success probability of that protocol by ϵ. (Here n is the input size of each party.) Is there a quantum analogue to this result?

In one sense the answer is "no". Given a two-party entanglement-assisted protocol for, say, computing the value of some function, we cannot replace the shared entanglement with some different, less entangled, state, without causing large errors [10, 1]. It is an open question whether it is possible to replace a large entangled state with a less entangled one while also changing the communication protocol.

However, while it remains a challenge to characterize the *dimension* of shared entanglement required for optimal entanglement-assisted quantum communication protocols, in this work we show that the *type* of shared entanglement required by such protocols can be neatly characterized. In Theorem 1 below, we establish that the bounded-error entanglement-assisted quantum communication complexity of a partial or total function cannot be improved by more than a constant factor by replacing maximally entangled states with arbitrary entangled states. This is accomplished by constructing an explicit protocol which allows two parties, who only share maximally entangled states, to simulate any entanglement-assisted quantum communication task regardless of the shared state that that task originally required.

▶ **Theorem 1.** *Consider a quantum communication protocol $\mathcal{R}$ whose goal it is to compute a joint function $f(x, y) \in \{0, 1\}$. Suppose that $\mathcal{R}$ uses an arbitrary bipartite entangled state $|\psi\rangle^{AB}$ (of unbounded dimension), as well as Q qubits of communication total, in either direction (for sufficiently large $Q \geq 15$). Then, for every $\epsilon > 0$, there exists a quantum communication protocol $\mathcal{R}'$ which simulates $\mathcal{R}$ with error ϵ, while using only a maximally entangled state as an entangled resource (rather than $|\psi\rangle^{AB}$ or any other state), and using $O(Q/\epsilon + \log(1/\epsilon)/\epsilon)$ qubits of communication. Thus, if $\mathcal{R}$ computes f with error ϵ' it follows that $\mathcal{R}'$ computes f with error $\epsilon + \epsilon'$.*

Theorem 1 shows that, although the role of shared entanglement in quantum communication complexity is still not well understood, the *type* of shared entanglement does not drastically change communication complexity. This is true regardless of input size or promise, as long as we are in the constant-error regime and some communication is allowed between players (unlike, say, the simultaneous-message-passing model). This result sets quantum communication complexity apart from settings such as channel simulation [3], nonlocal

games [11, 16], unitary gate simulation [6], and communication tasks involving quantum communication between referees and players [12]. In each of those cases the ratio between the EPR-assisted costs and the (unrestricted) entanglement-assisted costs can be made arbitrarily large. This suggests that the role of shared entanglement in quantum communication complexity may be fundamentally different than in these other settings. Furthermore, the result achieved in Theorem 1 may be useful in future work attempting to further bound the role of entanglement in quantum communication complexity, as it restricts the problem to the case of shared EPR pairs, without loss of generality.

Previously it was known that a universal form of entanglement existed: the embezzling state [8]. Our Theorem 1 can be viewed as showing that these states can be replaced by the much simpler family of maximally entangled states.

It may be worth noting that the proof of Theorem 1 is nearly oblivious to the entanglement-assisted protocol being considered in the following sense: Given a protocol $\mathcal{P}$ using Q qubits of entanglement and a shared entangled state $|\psi\rangle$, we can replace $|\psi\rangle$ with a "consolidated" state ρ at the cost of error ϵ. Moreover, ρ can be prepared from a maximally entangled state using $O(Q/\epsilon + \log(1/\epsilon)/\epsilon)$ communication. Taking ϵ constant implies that the EPR-assisted communication complexity of a function is at most $O(1)$ times the (unrestricted) entanglement-assisted communication complexity of that function. It was not necessary to modify the protocol $\mathcal{P}$ to achieve this result, except to pre-compose it with a pre-processing protocol which starts with only EPR pairs, and prepares the state ρ using only $O(Q/\epsilon + \log(1/\epsilon)/\epsilon)$ communication. $\mathcal{P}$ can then be run on ρ directly. Such a protocol-agnostic preprocessing should not be taken for granted, since it is known that reducing the number of EPR pairs may in some cases require more than just pre-processing [10, 1].

1.2 Communication cost of state transformations

Our second contribution, which is related at the level of techniques to Theorem 1, is to provide upper and lower bounds for an old quantity studied in quantum information theory, the communication cost of state transformation.

Suppose that $|\chi\rangle^{AB}$ and $|\nu\rangle^{AB}$ are bipartite pure quantum states, with vectors of Schmidt coefficients denoted respectively by χ and ν. In this setting it is known that $|\chi\rangle$ can be exactly converted into $|\nu\rangle$ using LOCC if and only if χ is majorized by ν [15]. But the communication cost of this transformation is known only in a few special cases. If $|\chi\rangle = |\chi_0\rangle^{\otimes n}$ and $|\nu\rangle = |\nu_0\rangle^{\otimes n}$ for some states $|\chi_0\rangle, |\nu_0\rangle$, then this cost is $O(\sqrt{n})$ or less in some special cases (e.g. $|\nu_0\rangle$ is maximally entangled). More generally there is, in principle, an exact characterization of the communication cost (either LOCC, or quantum communication) of state transformation using the Schubert calculus due to Daftuar and Hayden [4], but in practice it is difficult to extract concrete bounds from their main theorem.

In this work we identify a simple and efficiently computable quantity, which we call the ℓ_∞ Earth Mover's (or Wasserstein) Distance, which tells us approximately how much quantum communication is required to transform $|\chi\rangle$ to $|\nu\rangle$. Given its simple form, we believe that this quantity may be a useful tool in quantum information theory.

▶ **Definition 2** (ℓ_∞ Earth Mover's Distance). *Let* $|\chi\rangle^{AB} = \sum_{i\in X} \sqrt{\chi_i}\,|i\rangle^A \otimes |i\rangle^B$ *and* $|\upsilon\rangle^{AB} = \sum_{j\in Y} \sqrt{\upsilon_j}\,|j\rangle^A \otimes |j\rangle^B$ *be two states. We define* $d_\infty(|\chi\rangle, |\upsilon\rangle)$ *to be the* ℓ_∞ *Earth Mover's distance between* $|\chi\rangle$ *and* $|\upsilon\rangle$*, which is equal to the minimum* $\mu \geq 0$ *for which there exists a joint distribution* $\omega(x, y) : X \times Y \to \mathbb{R}_{\geq 0}$ *such that:*

- $\sum_{j\in Y} \omega(i, j) = \chi_i \ \forall i \in X$
- $\sum_{i\in X} \omega(i, j) = \upsilon_j \ \forall j \in Y$
- $\omega(i, j) = 0$ *whenever* $|\log(\chi_i) - \log(\upsilon_j)| > \mu$

We can think of χ as corresponding to placing χ_i mass at position $\log(\chi_i)$ for each i, and similarly for υ. Then $d_\infty(|\chi\rangle, |\upsilon\rangle)$ is the ℓ_∞ EMD (Earth Mover's distance) between these distributions.

In Section 4 we will show that this quantity gives an intuitive upper bound on the amount of quantum communication required to transform one bipartite shared state into another. In particular we prove the following theorem.

▶ **Theorem 3.** *Let $|\chi\rangle^{AB}$ and $|\upsilon\rangle^{AB}$ be two bipartite shared states. There is a protocol $\mathcal{M}_{\chi\to\upsilon}$ which can prepare $|\upsilon\rangle$ from $|\chi\rangle$, using only $4\lceil d_\infty(|\chi\rangle, |\upsilon\rangle)\rceil + 8$ qubits of communication.*

In Section 3 we establish a complementary lower bound, showing that a "ϵ-smoothed" version of the ℓ_∞ Earth Mover's Distance, denoted by $d^\epsilon_\infty(|\chi\rangle, |\upsilon\rangle)$, gives a lower bound on the cost of state transformation. That is:

▶ **Theorem 4.** *Given any two bipartite shared states $|\psi\rangle^{AB} = \sum_i \sqrt{\psi_i}\,|i\rangle^A \otimes |i\rangle^B$ and $|\phi\rangle^{AB} = \sum_i \sqrt{\phi_i}\,|i\rangle^A \otimes |i\rangle^B$, shared between two parties A and B, together with a unitary $U_\mathcal{P}$ which can be performed on the state $|\psi\rangle^{AB}$ via a quantum communication protocol $\mathcal{P}$, that uses Q qubits of communication between A and B, we have that, for every ϵ:*

$$\left|\langle\phi|^{AB} U_\mathcal{P} |\psi\rangle^{AB}\right| \leq 1 - \frac{1}{4}\epsilon^2 + 24 \cdot 2^{-\frac{1}{2}(d^\epsilon_\infty(|\psi\rangle,|\phi\rangle)-3Q)}$$

In words: If two shared states cannot be brought within small ℓ_∞ Earth Mover's Distance of each other by moving an ϵ quantity of mass of their Schmidt coefficients, then they also cannot be brought closer than $1-O(\epsilon^2)$ fidelity with each other without using $\Omega(d^\epsilon_\infty(|\psi\rangle, |\phi\rangle))$ qubits of communication (for sufficiently large values of $d^\epsilon_\infty(|\psi\rangle, |\phi\rangle)$). Thus, the ϵ-smoothed ℓ_∞ Earth Mover's Distance provides a lower bound on the communication cost of state conversion. On the other hand, from the definition of $d^\epsilon_\infty(|\psi\rangle, |\phi\rangle)$, stated in Definition 13, we note here that one can use Theorem 3 to move $|\psi\rangle$ to within $1-\epsilon$ fidelity of $|\phi\rangle$ using only $O(d^\epsilon_\infty(|\psi\rangle, |\phi\rangle))$ qubits of communication. To do this, omit the ϵ mass of Schmidt coefficients on which the two states have large ϵ-smoothed ℓ_∞ distance, and apply Theorem 3 as one would do with the regular ℓ_∞ Earth Mover's Distance. In this sense $d^\epsilon_\infty(|\psi\rangle, |\phi\rangle)$ gives both an upper and lower bound on the communication cost of state conversion.

To put these bounds in context: One could consider entanglement concentration and dilution to be the starting point for the study of state conversion. The original paper on entanglement concentration and dilution [2] concerned the many-copy limit and did not attempt to bound the amount of classical communication used. The first time the classical communication cost of state conversion was considered explicitly seems to have been in [13], which could be said to establish a version of our upper bound in the case where the starting state is maximally entangled. (Their result is not quite that general but contains many of the key ideas.) A version of our lower bound was established, again for the case of starting with maximally entangled states, in [7, 9]. These lower bounds could be applied to general state conversion but relied on Rènyi entropy inequalities that are clearly not tight in many cases. Finally, as noted earlier, a full characterization of the communication cost of general state conversion was given in [4] but the resulting formula is complicated and there is not an efficient algorithm known to evaluate it.

We conclude the section with two remarks about notation.

▶ Remark 5. In theorem statements above, and where appropriate, we have made use of superscripts A and B, as in $|\psi\rangle^{AB} = \sum_i \sqrt{\psi_i}\,|i\rangle^A \otimes |i\rangle^B$ to explicitly denote the two halves of the bipartite division of a state. However, since all of the shared entangled states considered in this paper are bipartite, and since the two components of the bipartite division are generally clear from context, we will usually omit this notation.

▶ Remark 6. When considering a bipartite state $|\psi\rangle$, we will assume that the state has a Schmidt decomposition of the form $|\psi\rangle = \sum_i \sqrt{\psi_i}\,|i\rangle \otimes |i\rangle$ across the implicit bipartite division. This is done in the theorem statements above and everywhere in the paper. We can assume this WLOG because any state that has the same Schmidt coefficients as $|\psi\rangle$ can be moved to this canonical form (and vice versa) using only local unitary transformations, which can be implemented with no quantum communication between the two components of the bipartite division. Thus our analysis of communication costs is unaffected by assuming WLOG that, in any quantum communication protocol, shared entangled states start and end in this form.

▶ Remark 7. Given a quantum state $|\nu\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_B$, we will use $rk_{Schmidt}(|\nu\rangle)$ to denote the Schmidt rank of $|\nu\rangle$ across the bipartite division between $\mathcal{H}_A$ and $\mathcal{H}_B$.

2 Entanglement-Assisted Communication Complexity

In this section we will discuss the proof of our main result, Theorem 1, which shows that arbitrary entanglement-assisted quantum communication protocols can be simulated by quantum communication protocols that use only the maximally entangled state as an entangled resource. A basic fact we will need is that two bipartite pure states which are sufficiently different in the distribution of mass across their Schmidt coefficients must be nearly orthogonal. This fact is stated for our specific purposes in Lemma 9 below. Crucially, such states *remain* nearly orthogonal even after one of them is acted on by any unitary which can be implemented with a small amount of quantum communication, as we detail in Lemma 8.

▶ **Lemma 8.** *Given two quantum states $|\psi\rangle$ and $|\nu\rangle$ on $\mathcal{H}_A \otimes \mathcal{H}_B$, such that the Schmidt coefficients of ψ are upper bounded by $\lambda_{\max}$, and those of ν are upper bounded by $\nu_{\max}$, and further given a unitary transformation $\mathcal{U}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ which can be implemented using at most Q qubits of communication between the $\mathcal{H}_A$ and $\mathcal{H}_B$ components of the Hilbert space, it follows that:*

$$|\langle\psi|\mathcal{U}|\nu\rangle| \leq 2^{\frac{3}{2}Q} \cdot rk_{Schmidt}(|\psi\rangle)\sqrt{\lambda_{\max}\nu_{\max}}$$

Proof. If $\mathcal{U}$ is a unitary transform using Q qubits of communication, then $rk_{Schmidt}(\mathcal{U}|\nu\rangle) \leq 2^Q rk_{Schmidt}(|\nu\rangle)$ [9]. We also know that the Schmidt coefficients of $\mathcal{U}|\nu\rangle$ are bounded above by $2^Q\nu_{max}$ [9]. The desired result now follows by Lemma 9. ◀

▶ **Lemma 9.** *Given two quantum states $|\psi\rangle$ and $|\nu\rangle$ on $\mathcal{H}_A \otimes \mathcal{H}_B$, such that the Schmidt coefficients of ψ are upper bounded by $\lambda_{\max}$, and those of ν are upper bounded by $\nu_{\max}$, we have:*

$$|\langle\psi|\nu\rangle| \leq rk_{Schmidt}(|\psi\rangle)\sqrt{\lambda_{\max}\nu_{\max}}$$

Proof. For brevity let $r = rk_{Schmidt}(|\psi\rangle)$. Schmidt decompose $|\psi\rangle$ and $|\nu\rangle$ as $|\psi\rangle = \sum_{i=0}^{r-1}\sqrt{\lambda_i}\,|i\rangle_A \otimes |i\rangle_B$, as $|\nu\rangle = \sum_j \sqrt{\nu_j}\,|j\rangle_A \otimes |j\rangle_B$. Define the matrix $M_\nu = \sum_j \sqrt{\nu_j}\,|j\rangle_A\,\langle j|_B^*$, and note that

$$\begin{aligned}\langle\psi|\nu\rangle &= \sum_{i=0}^{r-1}\sum_{j}\sqrt{\lambda_i\nu_j}\,\langle i_A|j_A\rangle\cdot\langle i_B|j_B\rangle \\ &= \sum_{i=0}^{r-1}\sqrt{\lambda_i}\,\langle i_A|\left(\sum_j\sqrt{\nu_j}\,|j\rangle_A\,\langle j|_B^*\right)|i_B^*\rangle \\ &= \sum_{i=0}^{r-1}\sqrt{\lambda_i}\,\langle i_A|\,M_\nu\,|i_B^*\rangle\end{aligned}$$

Now, by definition of a Schmidt Decomposition, we know that the maximum singular value of M_ν is $\sqrt{\nu_{max}}$. Thus, for all i we have that $|\,\langle i_A|\,M_\nu\,|i_B^*\rangle\,| \leq \sqrt{\nu_{max}}$ (since $|i_A\rangle$ and $|i_B\rangle$ are normalized vectors by definition). It then follows that:

$$|\,\langle\psi|\nu\rangle\,| \leq r\sqrt{\lambda_{\max}\nu_{\max}} = rk_{Schmidt}(|\psi\rangle)\sqrt{\lambda_{\max}\nu_{\max}}$$ ◀

Although Theorem 1 is the main result of this work, the proof is too long to fit in a 10 page abstract. Therefore will now give a brief, intuitive outline of the proof of Theorem 1, restated below for the reader's convenience, and include the complete proof in Section A of the Appendix.

▶ **Theorem** (Restatement of Theorem 1). *Consider a quantum communication protocol $\mathcal{R}$ whose goal it is to compute a joint function $g(x,y) \in \{0,1\}$. Suppose that $\mathcal{R}$ uses an arbitrary bipartite entangled state $|\psi\rangle^{AB}$ (of unbounded dimension), as well as Q qubits of communication total, in either direction (for sufficiently large $Q \geq 15$). Then, for every $\epsilon > 0$, there exists a quantum communication protocol $\mathcal{R}'$ which simulates $\mathcal{R}$ with error ϵ, while using only a maximally entangled state as an entangled resource (rather than $|\psi\rangle^{AB}$ or any other state), and using $O(Q/\epsilon + \log(1/\epsilon)/\epsilon)$ qubits of communication. Thus, if $\mathcal{R}$ computes f with error ϵ' it follows that $\mathcal{R}'$ computes f with error $\epsilon + \epsilon'$.*

Proof Sketch of Theorem 1. Suppose that we have a communication protocol using Q qubits of communication and a pure entangled state $|\varphi\rangle$. If we can prepare $|\varphi\rangle$ from EPR pairs using $O(Q)$ qubits of communication then we are done. Thus we can assume that $|\varphi\rangle$ has entanglement spread that is $\gg Q$. This will be defined more precisely below (see also [9, 7, 5]) but roughly speaking it means that we can write $|\varphi\rangle$ as a superposition of varying numbers of EPR pairs, say from $E_{\min}$ to $E_{\max}$, with $E_{\max} - E_{\min} \gg Q$. (Technically we need to use Theorem 3 to show that with a small amount of communication $|\varphi\rangle$ can be mapped to a superposition of maximally entangled states of different sizes.)

For simplicity, suppose that $|\varphi\rangle = |\alpha\rangle + |\beta\rangle$ where $|\alpha\rangle, |\beta\rangle$ each have norm $1/\sqrt{2}$ and, up to normalization, $|\alpha\rangle$ is locally equivalent to $E_{\min}$ EPR pairs and $|\beta\rangle$ is locally equivalent to $E_{\max}$ EPR pairs. This difference in entanglement means that $|\alpha\rangle$ and $|\beta\rangle$ must be nearly orthogonal: specifically their overlap can be at most $1/\sqrt{2}^{E_{\max}-E_{\min}}$. Moreover, if we apply a unitary communication protocol $\mathcal{P}$ using Q qubits of communication to one of them, say β, then that will not be enough to bridge the gap. If we perform $\mathcal{P}$ and then measure the first qubit, this is equivalent to measuring the observable $\mathcal{P}^\dagger\sigma_z^{(1)}\mathcal{P}$, which conveniently is also a unitary using $2Q$ qubits of communication. The bias of the protocol (i.e. $\Pr[1] - \Pr[0]$) is then

$$\langle\varphi|\,\mathcal{P}^\dagger\sigma_z^{(1)}\mathcal{P}\,|\varphi\rangle = \langle\alpha|\,\mathcal{P}^\dagger\sigma_z^{(1)}\mathcal{P}\,|\alpha\rangle + \langle\beta|\,\mathcal{P}^\dagger\sigma_z^{(1)}\mathcal{P}\,|\beta\rangle + \tag{1}$$

$$\langle\alpha|\,\mathcal{P}^\dagger\sigma_z^{(1)}\mathcal{P}\,|\beta\rangle + \langle\beta|\,\mathcal{P}^\dagger\sigma_z^{(1)}\mathcal{P}\,|\alpha\rangle \tag{2}$$

Because of the limited communication used by $\mathcal{P}$ the terms in (2) are negligible, on the order of $1/\sqrt{2}^{E_{\max}-E_{\min}-4Q}$.

This means that $|\varphi\rangle$ behaves effectively like an incoherent mixture of $|\alpha\rangle$ and $|\beta\rangle$. The phase of the superposition between $|\alpha\rangle$ and $|\beta\rangle$ cannot be observed without using more communication. Similar arguments were used in [6] to argue that projecting onto $|\varphi\rangle$ required a lot of communication even given the assistance of unlimited EPR pairs. Indeed applying a phase of -1 to $|\alpha\rangle + |\beta\rangle$ but not $|\alpha\rangle - |\beta\rangle$ is equivalent to the unitary which maps $|\alpha\rangle \leftrightarrow |\beta\rangle$, and so must require quantum communication if the two states have different amounts of entanglement. (The only exception would be if an embezzling state is used.)

If $|\alpha\rangle$ and $|\beta\rangle$ are locally equivalent to maximally entangled states of different sizes then their mixture can be prepared using no communication starting with a large number of EPR pairs and a shared random bit (which can also be obtained from an EPR pair). Since $\mathcal{P}$ cannot distinguish φ from this mixture, we can run the protocol successfully starting with EPR pairs which we map for free (or almost for free, given our above use of Theorem 3) to the mixture of $|\alpha\rangle$ and $|\beta\rangle$.

To prove the full theorem we need to consider more general superpositions and new mathematical subtleties arise. But the key principle is still that a low-communication protocol cannot observe phases of superpositions between states whose degrees of entanglement are too far apart. ◀

3 The Cost of State Transformation: A Lower Bound

It is natural at this point to discuss the background and proof for Theorem 4, which establishes a lower-bound on the cost of State Transformation by the ϵ-Smoothed ℓ_∞ Earth Mover's Distance, and to postpone the discussion of Theorem 3 until Section 4, for two reasons. First, the proof of Theorem 4 in this section shares key techniques in common with the proof of Theorem 1 in Section 2 above, and so this progression may provide the reader with some continuity of thought while also reiterating the usefulness of the techniques. Second, Theorem 4 in this section motivates the notion of the ℓ_∞ Earth Mover's Distance by highlighting its, perhaps surprising, relevance to *lower* bounding the cost of state transformation. This prepares the reader with some motivation for why the *upper* bound proven in Theorem 3, in Section 4 below, is interesting and potentially useful. Thus, covering Theorem 4 at this point may provide the reader with a reason to accept the ϵ-smoothed ℓ_∞ Earth Mover's Distance as a useful proxy for the cost of State Transformation.

Whereas the proof of Theorem 3 in the next section will make direct use of Definition 2, the proof of Theorem 4 in this section is elucidated by first establishing an equivalent formulation of the ℓ_∞ Earth Mover's Distance which is derived by establishing the relationship between the ℓ_∞ Earth Mover's Distance as defined in Definition 2, and the Monge-Kantorovich Transportation distance on the real line, as shown below. After translating to this equivalent definition, stated in Definition 12, the generalization to the ϵ-smoothed ℓ_∞ Earth Mover's Distance in Definition 13 is straightforward and natural.

▶ **Definition 10.** *Given two probability distributions μ and ν on the real line, define $\Gamma(\mu,\nu)$ to be the set of probability distributions on $\mathbb{R}\times\mathbb{R}$ whose marginals are μ and ν, respectively. Given a cost function $c : \mathbb{R}\times\mathbb{R} \to [0,\infty]$ the corresponding Monge-Kantorovich distance, $d_{MK}(\mu,\nu)$ between μ and ν is defined as:*

$$d_{MK}(\mu,\nu) = \inf\left\{\int_{\mathbb{R}\times\mathbb{R}} c(x,y)d\gamma(x,y) | \gamma \in \Gamma(\mu,\nu)\right\}.$$

We can interpret $\Gamma(\mu,\nu)$ as flows mapping μ to ν and $c(x,y)$ as the cost of moving one unit of mass from position x to position y. The Monge-Kantorovich distance is then the minimum cost flow using this cost function and boundary conditions.

In order to translate into a statement about quantum states, we make the following definition in a similar style to Definition 2:

▶ **Definition 11.** *Given a bipartite shared state* $|\psi\rangle = \sum_{i\in X}\sqrt{\psi_i}\,|i\rangle\otimes|i\rangle$ *let us define a random variable* V_ψ *which takes value* $\log(\psi_i)$ *with probability* ψ_i *(note that, since the* ψ_i *sum to one, this is a well defined random variable). We now define* p_ψ *to be the probability distribution of this random variable.*

It is clear that, for every ψ, p_ψ is a probability distribution on the real line. One may note the following simple relationship between Monge-Kantorovich distance and ℓ_∞ Earth Mover's Distance:

For any $\theta > 0$, consider the Monge-Kantorovich distance, $d_{MK(\theta)}$ where the function $c : \mathbb{R}\times\mathbb{R} \to [0,\infty]$ is defined by $c(x,y) = 1$ if $|x-y| \geq \theta$ and $c(x,y) = 0$ if $|x-y| < \theta$. Then, for any two quantum states $|\psi\rangle$ and $|\phi\rangle$, we have that $d_\infty(|\psi\rangle, |\phi\rangle) < \theta$ if and only if $d_{MK(\theta)}(p_\psi, p_\phi) = 0$.

Given this concrete connection between ℓ_∞ Earth Mover's Distance and the Monge-Kantorovich distance, we can now make use of the following characterization of Monge-Kantorovich distance for distributions on the real line, which is well known in optimal transport theory:

▶ **Fact.** *Let* μ *and* ν *be probability distributions supported on the real line, and let* F_μ *and* F_ν *be their cumulative distribution functions, respectively. Then, for any* $c : \mathbb{R}\times\mathbb{R} \to [0,\infty]$ *:*

$$d_{MK}(\mu,\nu) \equiv \inf_{\gamma\in\Gamma(\mu,\nu)}\left\{\int_{\mathbb{R}\times\mathbb{R}} c(x,y)d\gamma(x,y)\right\} = \int_0^1 c(F_\mu^{-1}(s), F_\nu^{-1}(s))ds$$

It follows from this Fact, combined with the discussion above, that an equivalent definition of the ℓ_∞ Earth Mover's Distance is given by:

▶ **Definition 12.**

$$d_\infty(|\psi\rangle, |\phi\rangle) \equiv \max_{q\in[0,1]} |F_{p_\psi}^{-1}(q) - F_{p_\phi}^{-1}(q)|$$

A drawback of $d_\infty(|\psi\rangle, |\phi\rangle)$ is that it is not robust against tiny changes of either distribution in the total variation distance. Concretely, it is lower but not upper semi-continuous, since adding an infinitesimal amount of mass far away can cause $d_\infty(,\text{ t})$o increase by an unbounded amount. This is acceptable for an upper bound (i.e. protocols using communication scaling with $d_\infty(,\,)$) but it would be impossible to prove a lower bound (or no-go theorem) of the form of Theorem 4 if stated using that definition. For this reason, we find it convenient to introduce a "smoothed" version of the distance measure.

▶ **Definition 13.** *ϵ-Smoothed ℓ_∞-Earth Mover's Distance*

$$d_\infty^\epsilon(|\psi\rangle, |\phi\rangle) \equiv \max_{q\in[0,1]} \min_{r\in[q-\epsilon,q+\epsilon]} |F_{p_\psi}^{-1}(q) - F_{p_\phi}^{-1}(r)|$$

With this definition in place we can now state the lower bound.

▶ **Theorem** (Restatement of Theorem 4)**.** *Given any two bipartite shared states* $|\psi\rangle^{AB} = \sum_i \sqrt{\psi_i}\,|i\rangle^A\otimes|i\rangle^B$ *and* $|\phi\rangle^{AB} = \sum_i\sqrt{\phi_i}\,|i\rangle^A\otimes|i\rangle^B$*, shared between two parties A and*

B, together with a unitary $U_{\mathcal{P}}$ which can be performed on the state $|\psi\rangle^{AB}$ via a quantum communication protocol $\mathcal{P}$, that uses Q qubits of communication between A and B, we have that, for every ϵ:

$$\left|\langle\phi|^{AB} U_{\mathcal{P}} |\psi\rangle^{AB}\right| \leq 1 - \frac{1}{4}\epsilon^2 + 24 \cdot 2^{-\frac{1}{2}(d^{\epsilon}_{\infty}(|\psi\rangle,|\phi\rangle)-3Q)}$$

Intuitively, Theorem 4 states that two bipartite shared states which are far apart in the ϵ-Smoothed ℓ_∞-Earth Mover's Distance, cannot be made equal via a quantum communication protocol unless it uses at least $c \cdot d^{\epsilon}_{\infty}(|\psi\rangle, |\phi\rangle)$ qubits of communication (for a particular constant c which can be computed from the statement of Theorem 4).

Proof. Suppose that two bipartite shared states $|\psi\rangle$ and $|\phi\rangle$ have $d^{\epsilon}_{\infty}(|\psi\rangle, |\phi\rangle) = d$. By definition $\exists p \in [0,1]$ such that

$$\min_{r\in[p-\epsilon,p+\epsilon]} |F^{-1}_{p_\psi}(p) - F^{-1}_{p_\phi}(r)| = d \tag{3}$$

Suppose that $F^{-1}_{p_\psi}(p) < F^{-1}_{p_\phi}(r)$ (if the opposite is true then we simply switch the roles of ψ and ϕ and continue with the same proof). Define $x \equiv F^{-1}_{p_\psi}(p)$. Further define $|\psi\rangle_{\leq x} \equiv \sum_{\{i:|\log 1/\psi_i|\leq x\}} \sqrt{\psi_i}\,|i\rangle \otimes |i\rangle$, and $|\psi\rangle_{>x} \equiv |\psi\rangle - |\psi\rangle_{\leq x}$. Similarly define $|\phi\rangle_{\geq x+d} \equiv \sum_{\{i:|\log 1/\phi_i|\geq x+d\}} \sqrt{\phi_i}\,|i\rangle \otimes |i\rangle$, and $|\phi\rangle_{<x+d} \equiv |\phi\rangle - |\phi\rangle_{\geq x+d}$. Note that $|\psi\rangle_{\leq x}$, and $|\psi\rangle_{>x}$ are orthogonal, as are $|\phi\rangle_{<x+d}$ and $|\phi\rangle_{\geq x+d}$.

Since we have $x \equiv F^{-1}_{p_\psi}(p)$ it follows from the definitions that $||\,|\psi\rangle_{\leq x}\,||^2 = p$. Since $F_{p_\psi}(x) = p$, and $F^{-1}_{p_\psi}(p) < F^{-1}_{p_\phi}(r)$, it follows from Equation 3 that $F_{p_\phi}(x+d) \leq p - \epsilon$. Therefore, $||\,|\phi\rangle_{<x+d}\,||^2 \leq p - \epsilon$ and thus $||\,|\phi\rangle_{\geq x+d}\,||^2 = 1 - ||\,|\phi\rangle_{<x+d}\,||^2 \geq 1 - p + \epsilon$.

The main idea in the proof of this theorem is that we can now decompose $U_{\mathcal{P}}|\psi\rangle$ and $|\phi\rangle$ each into three nearly orthogonal parts as follows:

▶ **Definition 14.**

$$\begin{aligned}
&|\psi^1\rangle \equiv U_{\mathcal{P}}|\psi\rangle_{\leq x}, && |\phi^3\rangle \equiv |\phi\rangle_{\geq x+d}, && |\psi^2\rangle \equiv \left(I - |\phi^3\rangle\langle\phi^3|\right) U_{\mathcal{P}}|\psi\rangle_{>x}\\
&|\psi^3\rangle \equiv |\phi^3\rangle\langle\phi^3| U_{\mathcal{P}}|\psi\rangle_{>x}, && |\phi^1\rangle \equiv |\psi^1\rangle\langle\psi^1|\phi\rangle_{<x+d}, && |\phi^2\rangle \equiv \left(I - |\psi^1\rangle\langle\psi^1|\right)|\phi\rangle_{<x+d}
\end{aligned}$$

It follows from this definition that:

$$U_{\mathcal{P}}|\psi\rangle = |\psi^1\rangle + |\psi^2\rangle + |\psi^3\rangle \tag{4}$$

$$|\phi\rangle = |\phi^1\rangle + |\phi^2\rangle + |\phi^3\rangle \tag{5}$$

The motivation and key property of the particular decomposition described in Definition 14 is best illustrated by the discussion of Lemma 15 below and the remainder of the proof of Theorem 4, which follows that.

▶ **Lemma 15.** *For $i,j \in \{1,2,3\}$ with $i \neq j$, we have that $|\langle\phi^i|\psi^j\rangle| \leq h(Q,d)$, $|\langle\psi^i|\psi^j\rangle| \leq h(Q,d)$, and $|\langle\phi^i|\phi^j\rangle| \leq h(Q,d)$, where $h(Q,d) \equiv 4 \cdot 2^{\frac{3Q-d}{2}}$.*

The proof of Lemma 15 is given separately in the appendix. Within that proof is the key use of Lemma 8 which is the primary conceptual step in proving Theorem 4. Understanding the proof of Lemma 15 is also the best way of understanding the motivation behind Definition 14 above.

While the individual $|\psi^i\rangle$ and $|\phi^i\rangle$ are not necessarily all orthogonal we do have $|\psi^2\rangle \perp |\psi^3\rangle$ and $|\psi^1\rangle \perp |\psi^2\rangle + |\psi^3\rangle$. Likewise $|\phi^1\rangle \perp |\phi^2\rangle$ and $|\phi^3\rangle \perp |\psi^1\rangle + |\psi^2\rangle$. Together with equations (4) and (5), these imply

$$1 = \left\||\psi^1\rangle\right\|^2 + \left\||\psi^2\rangle\right\|^2 + \left\||\psi^3\rangle\right\|^2 \tag{6a}$$

$$1 = \left\||\phi^1\rangle\right\|^2 + \left\||\phi^2\rangle\right\|^2 + \left\||\phi^3\rangle\right\|^2 \tag{6b}$$

From Lemma 15 it follows that:

$$\begin{aligned}|\langle\phi|\mathcal{P}(\psi)\rangle| \equiv |\langle\phi|U_{\mathcal{P}}|\psi\rangle| &= \left|\left(\langle\phi^1| + \langle\phi^2| + \langle\phi^3|\right)\left(|\psi^1\rangle + |\psi^2\rangle + |\psi^3\rangle\right)\right| \\ &\leq \left|\langle\phi^1|\psi^1\rangle\right| + \left|\langle\phi^2|\psi^2\rangle\right| + \left|\langle\phi^3|\psi^3\rangle\right| + 6\cdot h(Q,d) \\ &\leq \||\phi^1\rangle\|\,\||\psi^1\rangle\| + \||\phi^2\rangle\|\,\||\psi^2\rangle\| + \||\phi^3\rangle\|\,\||\psi^3\rangle\| + 6\cdot h(Q,d) \end{aligned} \tag{7}$$

Now recall that

$$\begin{aligned}\||\psi^1\rangle\| &= \left\|U_{\mathcal{P}}|\psi\rangle_{\leq x}\right\| = \left\||\psi\rangle_{\leq x}\right\| = \sqrt{p} \\ \||\phi^3\rangle\| &= \left\||\phi\rangle_{\geq x+d}\right\| \geq \sqrt{1-p+\epsilon}\end{aligned}$$

We now return to Equation 7. Setting $x_i = \||\psi^i\rangle\|$ and $y_i = \||\phi^i\rangle\|$ for $i = 1,2,3$ we have

$$|\langle\phi|\mathcal{P}(\psi)\rangle| \leq x_1y_1 + x_2y_2 + x_3y_3 + 6\cdot h(Q,d) \tag{8}$$

where $x_1 = \sqrt{p}$, $y_3 \geq \sqrt{1-p+\epsilon}$ and $(x_1,x_2,x_3),(y_1,y_2,y_3)$ are unit vectors. We claim that this quantity is maximized by setting $x_2 = y_2 = 0$ and $y_3 = \sqrt{1-p+\epsilon}$. Indeed we can upper bound $\sqrt{p}y_1 + x_2y_2 \leq x_{12}y_{12}$ where $x_{12} \equiv \sqrt{x_1^2+x_2^2}$ and $y_{12} \equiv \sqrt{y_1^2+y_2^2}$. Now define $x_{12} = \cos(\alpha), x_3 = \sin(\alpha), y_{12} = \cos(\beta), y_3 = \sin(\beta)$ and we have

$$x_1y_1 + x_2y_2 + x_3y_3 \leq \cos(\alpha-\beta). \tag{9}$$

This is maximized by taking $(x_1,x_2,x_3) = (\sqrt{p},0,\sqrt{1-p})$ and $(y_1,y_2,y_3) = (\sqrt{p-\epsilon},0,\sqrt{1-p+\epsilon})$. Thus

$$|\langle\phi|\mathcal{P}(\psi)\rangle| \leq \sqrt{p-\epsilon}\sqrt{p} + \sqrt{1-p}\sqrt{1-p+\epsilon} + 6\cdot h(Q,d). \tag{10}$$

Finally we would like an upper bound independent of p. This maximization is performed in the proof of Fact 27 from Section G of the Appendix and yields the following.

$$|\langle\phi|\mathcal{P}(\psi)\rangle| \leq 1 - \frac{1}{4}\epsilon^2 + 6\cdot h(Q,d). \qquad \blacktriangleleft$$

4 The Cost of State Transformation: An Upper Bound

In this section we will give a proof of Theorem 3, which states that the quantum communication cost of converting between two bipartite entangled states is upper bounded by the ℓ_∞ Earth Mover's Distance between those states. This upper bound represents the second half of our two sided argument (employing both Theorem 3 and Theorem 4) that the ℓ_∞ Earth Mover's Distance is a simple and efficiently computable proxy for the cost of state conversion. The proof is divided into two parts which are proved separately in Lemma 18, and Lemma 19 together with Corollary 20. At a high level Lemma 18 tells us that, given bipartite states $|\chi\rangle$ and $|\upsilon\rangle$, one can map the Schmidt coefficients of $|\chi\rangle$ directly onto the Schmidt coefficients of $|\upsilon\rangle$ using a series of bipartite "flows" that have small degree (where degree is a quantity defined below). Lemma 19 and Corollary 20 then tell us that any such "flow" which has small degree, can be implemented as an actual bipartite state transformation, with correspondingly small communication required.

Here we establish Lemmas 18 and 19 which, together, prove the desired theorem. We begin with a couple definitions establishing the concept of flows, as we use it here.

▶ **Definition 16** (Right (Left) Index-1 Flow). *Fix two states* $|\chi\rangle = \sum_{i\in X}\sqrt{\chi_i}\,|i\rangle\otimes|i\rangle$ *and* $|\upsilon\rangle = \sum_{j\in Y}\sqrt{\upsilon_j}\,|j\rangle\otimes|j\rangle$*. A Right Index-1 Flow from* $|\chi\rangle$ *to* $|\upsilon\rangle$ *is a bipartite graph* $G_{X,Y}$ *with vertex set* $X\cup Y$*, and edge set* $E_{X,Y}$ *(where* X, Y *represents the bipartition of the vertices) such that:*

- *Each vertex in* $j\in Y$ *has degree 1 in* $G_{X,Y}$*.*
- *For all* $i\in X$*,* $\chi_i = \sum_{j\in Y:(i,j)\in E_{X,Y}}\upsilon_j$

If the roles of $|\chi\rangle$ *and* $|\upsilon\rangle$ *are reversed in the above, then we say that there is a Left Index-1 Flow from* $|\chi\rangle$ *to* $|\upsilon\rangle$*. Equivalently, there is a Left Index-1 Flow from* $|\upsilon\rangle$ *to* $|\chi\rangle$ *exactly when there is a a Right Index-1 Flow from* $|\chi\rangle$ *to* $|\upsilon\rangle$*.*

▶ **Definition 17** (Degree of a Right (Left) Index-1 Flow). *We define the degree of a Right (Left) Index-1 Flow from* $|\chi\rangle = \sum_{i\in X}\sqrt{\chi_i}\,|i\rangle\otimes|i\rangle$ *to* $|\upsilon\rangle = \sum_{j\in Y}\sqrt{\upsilon_j}\,|j\rangle\otimes|j\rangle$ *to be the maximum degree of any vertex in the bipartite graph* $G_{X,Y}$*.*

The following lemma, which a key step in proving Theorem 3, establishes that bipartite states which are close to each other in the ℓ_∞ Earth Mover's Distance of Definition 2, can be mapped to each other through a series of flows of bounded degree. This series of flows intuitively establishes a map for converting one bipartite state to the other using bounded quantum communication, in a manner that will be made rigorous in Lemma 19. The main step in the proof of Lemma 18 involves constructing a flow through a type of greedy algorithm whose analysis has a number of subtle cases. In order to concretely exhibit these cases the entire greedy algorithm, including every case, is written out in pseudocode in Algorithm 1.

▶ **Lemma 18.** *Given two states* $|\chi\rangle$ *and* $|\upsilon\rangle$*, there exist two "intermediate" states* $|\gamma\rangle$ *and* $|\rho\rangle$*, such that there is a Right Index-1 Flow from* $|\chi\rangle$ *to* $|\gamma\rangle$ *of degree at most* $2^{2\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+4}$*, a Left Index-1 Flow from* $|\gamma\rangle$ *to* $|\rho\rangle$ *of degree at most* $2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}$*, and a Left Index-1 Flow from* $|\rho\rangle$ *to* $|\upsilon\rangle$ *of degree at most* $2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}$*.*

The Proof of Lemma 18 is included in the Appendix, section D.

Lemma 18, above, shows that two bipartite entangled states can be connected to each other by a series of flows which have a degree which is bounded in terms of the ℓ_∞ Earth Mover's Distance between them. The next step is to establish that every flow can be implemented via a quantum communication protocol. Lemma 19 and Corollary 20, below, accomplish this by showing that, if two bipartite states can be connected by flows of small degree, then one state can be converted to the other (and vice versa) using a quantum communication protocol which only requires small amounts of communication.

▶ **Lemma 19.** *Given two states* $|\tau\rangle$ *and* $|\kappa\rangle$ *such that there is a Right Index-1 Flow from* $|\tau\rangle$ *to* $|\kappa\rangle$ *with degree at most* 2^Q*, there exists a quantum communication protocol* $\mathcal{P}$*, which uses* Q *qubits of communication, and converts the shared state* $|\tau\rangle$ *to the shared state* $|\kappa\rangle$*.*

The idea of the proof is that if $|\tau\rangle = \sum_i\sqrt{\tau_i}\,|i\rangle\otimes|i\rangle$ then it suffices to define separately protocols for each $|i\rangle\otimes|i\rangle$ term. These protocols simply use quantum communication to create a shared entangled state, resulting in the state $\sum_i\tau_i\,|i\rangle_A\otimes|i\rangle_B\otimes|\psi_i\rangle_{A'B'}$. Choosing the Schmidt coefficients according to the given Right Index-1 Flow yields the result. The details of this argument are in the Appendix xE.

Corollary 20 establishes the same result as Lemma 19, but in the reverse direction.

▶ **Corollary 20.** *Given two states* $|\tau\rangle$ *and* $|\kappa\rangle$ *such that there is a Left Index-1 Flow from* $|\kappa\rangle$ *to* $|\tau\rangle$ *with degree at most* 2^Q*, then, for two parties sharing entangled state* $|\kappa\rangle$*, there exists a quantum communication protocol* $\mathcal{P}$*, which uses* Q *qubits of communication, and converts the shared state* $|\kappa\rangle$ *to the shared state* $|\tau\rangle$*.*

The proof of Corollary 20 is straightforward and appears in Appendix F.

▶ **Theorem** (Restatement of Theorem 3). *Let $|\chi\rangle^{AB}$ and $|\upsilon\rangle^{AB}$ be two bipartite shared states. There is a protocol $\mathcal{M}_{\chi\to\upsilon}$ which can prepare $|\upsilon\rangle$ from $|\chi\rangle$, using only $4\lceil d_\infty(|\chi\rangle, |\upsilon\rangle)\rceil + 8$ qubits of communication.*

Proof. The proof follows by applying Lemma 18, followed by Lemma 19 and Corollary 20. ◀

References

1 Dorit Aharonov, Aram W. Harrow, Zeph Landau, Daniel Nagaj, Mario Szegedy, and Umesh Vazirani. Local Tests of Global Entanglement and a Counterexample to the Generalized Area Law. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 246–255, October 2014. doi:10.1109/FOCS.2014.34.

2 C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53:2046–2052, 1996. arXiv:quant-ph/9511030.

3 C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter. The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels. *IEEE Trans. Inf. Theory*, 60(5):2926–2959, May 2014. doi:10.1109/TIT.2014.2309968.

4 S. Daftuar and P. Hayden. Quantum state transformations and the Schubert calculus. *Annals of Physics*, 315:80–122, 2005. arXiv:quant-ph/0410052.

5 A. W. Harrow. Entanglement spread and clean resource inequalities. In P. Exner, editor, *XVIth Int. Cong. on Math. Phys.*, pages 536–540. World Scientific, 2009. arXiv:0909.1557.

6 A. W. Harrow and D. W. Leung. A communication-efficient nonlocal measurement with application to communication complexity and bipartite gate capacities. *IEEE Trans. Inf. Theory*, 57(8):5504–5508, 2011. arXiv:0803.3066.

7 A. W. Harrow and H.-K. Lo. A tight lower bound on the classical communication cost of entanglement dilution. *IEEE Trans. Inf. Theory*, 50(2):319–327, 2004. arXiv:quant-ph/0204096.

8 P. Hayden and W. van Dam. Universal entanglement transformations without communication. *pra*, 67:060302(R), 2003. arXiv:quant-ph/0201041.

9 P. Hayden and A.J. Winter. On the communication cost of entanglement transformations. *Phys. Rev. A*, 67:012306, 2003. arXiv:quant-ph/0204092.

10 Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Optimal Direct Sum and Privacy Tradeoff Results for Quantum and Classical Communication Complexity, 2008. arXiv:0807.1267.

11 M. Junge and C. Palazuelos. Large Violation of Bell Inequalities with Low Entanglement. *Communications in Mathematical Physics*, 306(3):695–746, 2011. doi:10.1007/s00220-011-1296-8.

12 Debbie Leung, Ben Toner, and John Watrous. Coherent state exchange in multi-prover quantum interactive proof systems. *Chicago Journal of Theoretical Computer Science*, 11:1–18, 2013. arXiv:0804.4118.

13 H.-K. Lo and S. Popescu. The classical communication cost of entanglement manipulation: Is entanglement an inter-convertible resource? *Phys. Rev. Lett.*, 83:1459–1462, 1999. arXiv:quant-ph/9902045.

14 Ilan Newman. Private vs. common random bits in communication complexity. *Inf. Process. Lett.*, 39(2):67–71, 1991. doi:10.1016/0020-0190(91)90157-D.

15 M. A. Nielsen. Conditions for a Class of Entanglement Transformations. *Phys. Rev. Lett.*, 83:436–439, 1999. arXiv:quant-ph/9811053.

16 Oded Regev. Bell Violations Through Independent Bases Games. *Quantum Info. Comput.*, 12(1-2):9–20, January 2012. arXiv:1101.0576.

17 Thomas Vidick and Stephanie Wehner. More nonlocality with less entanglement. *Phys. Rev. A*, 83:052310, May 2011. doi:10.1103/PhysRevA.83.052310.

A Proof of Theorem 1

A concept which will be useful in the proof of Theorem 1 is the notion of the spread of a state:

▶ **Definition 21** (Spread). *For a finite dimensional bipartite entangled state* $|\psi\rangle^{AB} = \sum_i \sqrt{\psi_i}\,|i\rangle^A \otimes |i\rangle^B$ *let* λ_{max} *be the maximum of the Schmidt coefficients of* ψ*, and let* λ_{min} *be the minimum Schmidt coefficient. We define the spread of* $|\psi\rangle$ *to be the quantity* $\log(\lambda_{max}/\lambda_{min})$.

We note that the above definition of spread is given in the case of finite dimensional $|\psi\rangle$, which is the only case we will need. There is also an ϵ-smoothed variant of the spread of a state [9, 5], but it will not be needed for this proof. Within the proof of Theorem 1 the spread of a bipartite state will be used as a proxy for the amount of communication required to create that state from a maximally entangled state. This intuition is formalized, for example, by Theorem 3, but in this case of converting from a maximally entangled state, is also an implication of earlier works, such as [7, 9].

▶ **Theorem** (Restatement of Theorem 1). *Consider a quantum communication protocol* $\mathcal{R}$ *whose goal it is to compute a joint function* $g(x, y) \in \{0, 1\}$*. Suppose that* $\mathcal{R}$ *uses an arbitrary bipartite entangled state* $|\psi\rangle^{AB}$ *(of unbounded, but finite, dimension), as well as* Q *qubits of communication total, in either direction (for sufficiently large* $Q \geq 15$*). Then, for every* $\epsilon > 0$*, there exists a quantum communication protocol* $\mathcal{R}'$ *which simulates* $\mathcal{R}$ *with error* ϵ*, while using only a maximally entangled state as an entangled resource (rather than* $|\psi\rangle^{AB}$ *or any other state), and using* $O(Q/\epsilon + \log(1/\epsilon)/\epsilon)$ *qubits of communication. Thus, if* $\mathcal{R}$ *computes* f *with error* ϵ' *it follows that* $\mathcal{R}'$ *computes* f *with error* $\epsilon + \epsilon'$.

Proof. Given $\mathcal{R}$, g, and $|\psi\rangle$ as in the theorem statement, Schmidt decompose $|\psi\rangle$ as $\sum_i \sqrt{\lambda_i}\,|i, i\rangle$ (see Remark 6 for why we may assume WLOG that $|\psi\rangle$ has this form).

Let $N \geq 2$ be an integer, which will be specified later. Define a function $f : [0, 1] \to \{0, 1, \ldots, 2^N\}$ given by

$$f(\lambda) = 2^{\left\lceil \left\lceil \frac{\log(1/\lambda)}{N} \right\rceil N - \log(1/\lambda) \right\rceil} \in \{1, 2, 4, \ldots, 2^N\},$$

and define a new state $|\varphi\rangle \equiv \sum_i \sum_{j \in \{1,\ldots,f(\lambda_i)\}} \sqrt{\nu_{i,j}}\,|(i, j), (i, j)\rangle$, where $\nu_{i,j} \equiv \frac{\lambda_i}{f(\lambda_i)}$. Note that $\sum_{i,j} \nu_{i,j} = 1$, so that $|\varphi\rangle$ is a normalized pure state. Furthermore, every Schmidt coefficient $\nu_{i,j}$ of $|\varphi\rangle$ is within a multiple of 2 of the integer power $2^{-\left\lceil \frac{\log(1/\lambda_i)}{N} \right\rceil N}$. This follows because

$$\begin{aligned}
\left| \log \left(\frac{\nu_{i,j}}{2^{-\left\lceil \frac{\log(1/\lambda_i)}{N} \right\rceil N}} \right) \right| &= \left| \log(\lambda_i) - \log(f(\lambda_i)) + \left\lceil \frac{\log(1/\lambda_i)}{N} \right\rceil N \right| \\
&= \left| \log(\lambda_i) - \log\left(2^{\left\lceil \left\lceil \frac{\log(1/\lambda_i)}{N} \right\rceil N - \log(1/\lambda_i) \right\rceil} \right) + \left\lceil \frac{\log(1/\lambda_i)}{N} \right\rceil N \right| \\
&= \left| \left\lceil \frac{\log(1/\lambda_i)}{N} \right\rceil N - \log(1/\lambda_i) - \left\lceil \left\lceil \frac{\log(1/\lambda_i)}{N} \right\rceil N - \log(1/\lambda_i) \right\rceil \right| \\
&\leq 1 \qquad (11)
\end{aligned}$$

Next, we can upper bound $d_\infty(|\psi\rangle, |\varphi\rangle) \leq N$ by considering the coupling in which each $\nu_{i,j}$ is moved to λ_i. The largest distance obtained here is the maximum $\log f(\lambda_i)$ for which $\lambda_i > 0$, and this in turn is $\leq N$. Therefore, by Theorem 3, there is a protocol $\mathcal{M}$ by which Alice and Bob can prepare $|\psi\rangle$ from $|\varphi\rangle$, using $4\lceil d_\infty(|\chi\rangle, |\upsilon\rangle)\rceil + 8 \leq 4N + 8$ qubits of communication. (For this special case, of course a simpler protocol could also be used.)

Define $\mathcal{C} \equiv \mathcal{R} \circ \mathcal{M}$ to be the composed protocol in which Alice and Bob start with shared state $|\varphi\rangle$, first use protocol $\mathcal{M}$ to convert $|\varphi\rangle$ to $|\psi\rangle$, and then perform protocol $\mathcal{R}$ using shared state $|\psi\rangle$ and inputs x and y, to compute the joint function $g(x, y)$. It is evident that $\mathcal{C}$ has exactly the same success probability as $\mathcal{R}$. Since $\mathcal{M}$ uses at most $4N + 8$ qubits of communication and $\mathcal{R}$ uses Q qubits of communication, $\mathcal{C}$ can be performed with $Q + 4N + 8$ qubits of communication.

For k a nonnegative integer, define $I_k := \{i : 2^{-kN+1} \geq \lambda_i > 2^{-kN-1}\}$ and define the subnormalized state

$$|\varphi_k\rangle \equiv \sum_{i \in I_k} \sqrt{\lambda_i} \, |i, i\rangle \, . \tag{12}$$

From Equation (11) and the surrounding discussion, we have that $|\varphi\rangle = \sum_k |\varphi_k\rangle$. Furthermore, by the definition of I_k, it follows that $|\varphi_k\rangle$ has spread at most 2; note that the spread of $|\varphi_k\rangle$ does not depend on whether the state is normalized or not.

The idea of the proof is that different $|\varphi_k\rangle$ are not only orthogonal, but must remain approximately orthogonal even after a small amount of quantum communication. In particular, note that for any l, $rk_{Schmidt}(|\varphi_l\rangle) \leq 2^{lN+1} \| \, |\varphi_l\rangle \, \|^2$. Furthermore, for all l we have, by definition, that the Schmidt coefficients of $|\varphi_l\rangle$ are bounded above by 2^{-lN+1}. Therefore, if U is a unitary transform using M qubits of communication, then, it follows by Lemma 8, that $\forall j, k$,

$$\begin{aligned} |\langle \varphi_k | \, U \, |\varphi_j\rangle| &\leq 2^{\frac{3}{2}M} 2^{\min(j,k)N+1} \left\| \left| \varphi_{\min(j,k)} \right\rangle \right\|^2 \sqrt{2^{-jN+1} \cdot 2^{-kN+1}} \\ &\leq 2^{\frac{3}{2}M} 2^{-N\frac{|j-k|}{2}+2} \left\| \left| \varphi_{\min(j,k)} \right\rangle \right\|^2 \end{aligned} \tag{13}$$

To apply this to our problem, we first note that the protocol $\mathcal{C}$ depends, a priori, on the inputs x, y to the function $g(x, y)$ that we wish to compute (just like the the protocol $\mathcal{R}$). We now fix any input pair x, y and for the remainder of the proof of this theorem we will perform only transformations of the shared state which do not depend on the value of x, y. We will therefore establish that our transformation to a maximally entangled shared state does not significantly impact the success probability of the quantum communication protocol *regardless* of the value of x, y. The desired Theorem then follows.

With the input x, y now fixed, we observe that the success probability of protocol $\mathcal{C}$ (which we have already established is equal to the success probability of the original protocol $\mathcal{R}$) can be expressed WLOG by performing $\mathcal{C}$ and then computing the probability of outcomes when measuring the first qubit in the computational basis. The probability that such a measurement on protocol $\mathcal{C}$ outputs $b \in \{0, 1\}$ is

$$\Pr[b] = \langle \varphi | \, \mathcal{C}^\dagger (|b\rangle \langle b| \otimes I) \mathcal{C} \, |\varphi\rangle \, ,$$

where I acts on all qubits except for the first, which is being measured. Define $\mathcal{P} \equiv \mathcal{C}^\dagger(\sigma_z \otimes I)\mathcal{C} = \mathcal{C}^\dagger(|0\rangle \langle 0| \otimes I)\mathcal{C} - \mathcal{C}^\dagger(|1\rangle \langle 1| \otimes I)\mathcal{C}$. Then

$$\Pr[0] - \Pr[1] = \langle \varphi | \, \mathcal{P} \, |\varphi\rangle = \sum_{j,k} \langle \varphi_j | \, \mathcal{P} \, |\varphi_k\rangle \tag{14}$$

Observe, for later, that $\mathcal{P}$ is a unitary operator that can be implemented using $2Q+8N+16$ qubits of communication.

The proof will proceed as follows: In Lemma 24 we show that the density matrix $\varphi = |\varphi\rangle \langle \varphi|$ can be divided into three "pieces", one piece which has small trace norm and can therefore be omitted, one piece called φ_{far} which only has non-zero terms which are far from

the diagonal in the appropriate basis, and one piece called φ_{block} which is a block-diagonal mixed state that can be produced with small error and low communication cost from a maximally entangled state. Then, in Lemma 25, we show that the φ_{far} piece of φ has very little effect on the protocol $\mathcal{C}$. This means that φ can be replaced by φ_{block} alone while incurring very little error in the success probability of $\mathcal{C}$. Stated equivalently, via the equality in Equation 14 above, Lemma 25 shows that the quantity $|\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))|$ is small. Since we know from Lemma 24 that φ_{block} can be produced with low cost from a maximally entangled state, this leads us to the desired result.

We now establish some notation which will be useful throughout the rest of the proof:

▶ **Definition 22** (subset-matrix). *Consider operators on the Hilbert space which is the span of the $|\varphi_j\rangle$. We say that an operator M' is a* subset-matrix *of an operator M, if it is the case that for all l, k either $\langle\varphi_l| M' |\varphi_k\rangle = \langle\varphi_l| M |\varphi_k\rangle$, or $\langle\varphi_l| M' |\varphi_k\rangle = 0$.*

▶ **Definition 23** (Non-Zero Set). *For an operator θ on the Hilbert space which is the span of the $|\varphi_j\rangle$, define the non-zero set of θ to be $T_\theta = \{(l,k) : \langle\varphi_k| \theta |\varphi_l\rangle \neq 0\}$.*

▶ **Lemma 24.** *Consider the density matrix $\varphi \equiv \sum_{k,l} |\varphi_k\rangle \langle\varphi_l|$. For any $\epsilon > 0$, there exist subset-matrices, $\varphi_{block}, \varphi_{far}$, of φ, such that*

1. $\|\varphi - (\varphi_{block} + \varphi_{far})\|_1 \leq 2\epsilon$
2. $T_{\varphi_{far}} \subseteq \{(l,k) : |k-l| > B\}$, *where* $B \equiv 30 + 2\left\lceil \frac{\log(1/\epsilon)}{N} \right\rceil$.
3. *The bipartite shared state φ_{block} can be prepared starting from EPR pairs with $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$ bits of communication.*

The proof of Lemma 24 is included in Section B of the Appendix.

We can now bound the difference between the protocol $\mathcal{C}$ acting on φ versus $\mathcal{C}$ acting on φ_{block}, following equation 14 as follows:

$$|(\text{Pr}_\varphi[0] - \text{Pr}_\varphi[1]) - (\text{Pr}_{\varphi_{\text{block}}}[0] - \text{Pr}_{\varphi_{\text{block}}}[1])| = |\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))|$$

Setting $N = 2Q$ and recalling from the Theorem statement that $Q \geq 15$ by assumption, it follows by Lemma 25, stated below, that:

$$|(\text{Pr}_\varphi[0] - \text{Pr}_\varphi[1]) - (\text{Pr}_{\varphi_{\text{block}}}[0] - \text{Pr}_{\varphi_{\text{block}}}[1])| = |\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))| \leq 3\epsilon \tag{15}$$

This completes the proof of the Theorem as we now describe.

We know from Lemma 24 that there is a quantum communication protocol, call it $\mathcal{K}$, which prepares the shared state φ_{block} starting from just a maximally entangled state using at most $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$ bits of communication. Now define the protocol $\mathcal{R}' \equiv \mathcal{C} \circ \mathcal{K}$. Since $\mathcal{C}$ uses at most $Q + 4N + 8$ qubits of communication, and since we have chosen to set $N = 2Q$ (in the line above Equation 15), it follows that $\mathcal{R}'$ uses at most $O(N/\epsilon + \log(1/\epsilon)/\epsilon) = O(Q/\epsilon + \log(1/\epsilon)/\epsilon)$ qubits of communication. Furthermore, the success probability of $\mathcal{R}'$ with only the maximally entangled state as an entangled resource is the same, by construction, as the success probability of $\mathcal{C}$ with φ_{block} as an entangled resource, which, by Equation 15 above and the original definition $\mathcal{C} \equiv \mathcal{R} \circ \mathcal{M}$, is within 3ϵ of the success probability of the original protocol $\mathcal{R}$ from the theorem statement when using the original shared state $|\psi\rangle$ as an entangled resource. This is the desired result. ◀

▶ **Lemma 25.** *For φ_{block} as constructed in Lemma 24, and for N, Q as defined in the proof of Theorem 1 we have, $|\text{Tr}(\mathcal{P}(\varphi - \varphi_{block}))| \leq 3\epsilon$ whenever $N \geq 2Q \geq 30$.*

Proof. Following Lemma 24, we define $B \equiv 30 + 2\left\lceil \frac{\log(1/\epsilon)}{N} \right\rceil$. Now, letting φ_{block} and φ_{far} be as in Lemma 24, and recalling that $\|\varphi - (\varphi_{\text{block}} + \varphi_{\text{far}})\|_1 \leq 2\epsilon$, we have:

$$\begin{aligned}
|\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))| &\leq |\text{Tr}(\mathcal{P}((\varphi_{\text{block}} + \varphi_{\text{far}}) - \varphi_{\text{block}}))| + 2\epsilon = |\text{Tr}(\mathcal{P}\varphi_{\text{far}})| + 2\epsilon \\
&= \left| \sum_{(k,l) \in T_{\varphi_{\text{far}}}} \langle \varphi_k | \mathcal{P} | \varphi_l \rangle \right| + 2\epsilon \leq \sum_{(k,l) \in T_{\varphi_{\text{far}}}} |\langle \varphi_k | \mathcal{P} | \varphi_l \rangle| + 2\epsilon \\
&\leq \sum_{k,l:|k-l|>B} |\langle \varphi_k | \mathcal{P} | \varphi_l \rangle| + 2\epsilon
\end{aligned}$$

where the final inequality follows because $T_{\varphi_{\text{far}}} \subseteq \{(l,k) : |k-l| > B\}$ by Lemma 24. Recalling that the unitary $\mathcal{P}$ can be implemented using 2Q+8N+16 qubits of communication, and applying equation 13 then gives that:

$$\begin{aligned}
|\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))| - 2\epsilon &\leq \sum_{k,l:|k-l|>B} \min(1, 2^{3/2\cdot(2Q+8N+16)} 2^{-N\frac{|k-l|}{2}+4}) \left\| |\varphi_{\min(k,l)}\rangle \right\|^2 \\
&= 2 \sum_l \||\varphi_l\rangle\|^2 \sum_{k>l+B} \min(1, 2^{3Q+12N+24} 2^{-N\frac{|k-l|}{2}+4}) \\
&= 2 \sum_{n>B} \min(1, 2^{3Q+12N+24} 2^{-N\frac{n}{2}+4}) \\
&\leq 2 \cdot 2^{3Q+12N+24} 2^{-BN/2+4} \sum_{k=0}^{\infty} 2^{-N\frac{k}{2}} \\
&= 2 \cdot 2^{3Q+12N+24} 2^{-BN/2+4} \left(1 + \frac{2^{-\frac{N}{2}}}{1 - 2^{-\frac{N}{2}}}\right) \\
&\leq 4 \cdot 2^{3Q+12N+24} 2^{-BN/2+4}
\end{aligned}$$

So, recalling from the Lemma statement that $N \geq 2Q \geq 30$ by assumption:

$$\begin{aligned}
|\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))| - 2\epsilon &\leq 4 \cdot 2^{3Q+12N+24} 2^{-BN/2+4} \\
&\leq 4 \cdot 2^{28} 2^{14N} 2^{-15N - \left\lceil \frac{\log(1/\epsilon)}{N} \right\rceil N} \\
&\leq 2^{30} 2^{-N - \log(1/\epsilon)} \\
&\leq \epsilon
\end{aligned}$$

So,

$$|\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))| \leq 3\epsilon$$ ◀

Note, in the pre-processing step in the proof of Theorem 1, and again at a point within the proof of Lemma 24 we use our Theorem 3 in a setting where either the starting or ending state is very close to a maximally entangled state. It is helpful to observe, to avoid confusion, that in such cases Theorem 3 is not strictly necessary and could be replaced with previously known results from, for example, [7, 9]. In this manuscript we will use Theorem 3 in these cases in order to remain self-contained, and for the convenience of the reader, but we emphasize that the lines of the proof of Theorem 1 in which we use Theorem 3 could be replaced with known results.

B Proof of Lemma 24

▶ **Lemma** (Restatement of Lemma 24). *Consider the density matrix $\varphi \equiv \sum_{k,l} |\varphi_k\rangle \langle\varphi_l|$. For any $\epsilon > 0$, there exist subset-matrices, $\varphi_{block}, \varphi_{far}$, of φ, such that*

1. $\|\varphi - (\varphi_{block} + \varphi_{far})\|_1 \leq 2\epsilon$
2. $T_{\varphi_{far}} \subseteq \{(l,k) : |k-l| > B\}$, *where* $B \equiv 30 + 2\left\lceil \frac{\log(1/\epsilon)}{N} \right\rceil$.
3. *The bipartite shared state φ_{block} can be prepared starting from EPR pairs with $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$ bits of communication.*

Proof. Note: The terminology used in this proof is defined in the proof of Theorem 1 preceding the use of Lemma 24 there (Appendix A).

Fixing an $\epsilon > 0$ we will now show how to "cut" $\varphi \equiv \sum_{k,l} |\varphi_k\rangle \langle\varphi_l|$ down into a mixture of states of small spread such that the cut only removes subset-matrices of the operator which are either far from the diagonal or small in the trace norm (less than 2ϵ).

Define a sequence of mutually orthogonal projectors $\{P_i\}$, where each P_i is the projection onto the span of $\{|\varphi_l\rangle\}_{2(i-1)B<l\leq 2i\cdot B}$. Let

$$M_i \equiv (P_{2i-1} + P_{2i})\varphi(P_{2i-1} + P_{2i}).$$

Now, for $k \in [1, ..., \lceil 1/\epsilon \rceil]$ define

$$S_k \equiv \sum_{i=0}^{\infty} M_{i\cdot\lceil 1/\epsilon\rceil + k}.$$

The S_k are block-diagonal subset-matrices of φ, which are disjoint in the sense that $T_{S_k} \cap T_{S_{k'}} = \emptyset$ when $k \neq k'$. Additionally, $\sum_{k=1}^{\lceil 1/\epsilon\rceil} S_k = \sum_i M_i$ is a subset-matrix of φ which contains the entire diagonal of φ. Indeed $\sum_{k=1}^{\lceil 1/\epsilon\rceil} S_k$ can be obtained from φ via the "pinching" TPCP which has Kraus operators given by the $\{P_{2i-1} + P_{2i}\}$. Thus

$$1 = \mathrm{tr} \sum_{k=1}^{\lceil 1/\epsilon\rceil} S_k.$$

Choose k' such that $\mathrm{tr}[S_{k'}] \leq 1/\lceil 1/\epsilon\rceil \leq \epsilon$. Since the S_k are all PSD we also have $\|S_{k'}\|_1 \leq \epsilon$.

Our strategy now is to use something like $\varphi - S_{k'}$ as a candidate for $\varphi_{\mathrm{block}} + \varphi_{\mathrm{far}}$ in the Lemma statement. However, subtracting all of $S_{k'}$ removes some terms close to the diagonal, which, even though it is not a large fraction of all entries in φ, would make the proof and statement of Lemma 24 somewhat awkward. So, in order to make the Lemma statement as clean as possible we will only subtract the "anti-diagonal" parts of $S_{k'}$, and leave the "diagonal" parts of $S_{k'}$ in a manner made precise below.

Define the block matrices

$$D_i \equiv P_{2i-1}\varphi P_{2i-1} + P_{2i}\varphi P_{2i} \qquad A_i \equiv P_{2i-1}\varphi P_{2i} + P_{2i}\varphi P_{2i-1}$$

D_i and A_i are, respectively, the diagonal and off-diagonal blocks of M_i.

Further define $K_{k'} \equiv \sum_{i=0}^{\infty} A_{i\cdot\lceil 1/\epsilon\rceil + k'}$. We have that $K_{k'} = S_{k'} - \sum_{i=0}^{\infty} D_{i\cdot\lceil 1/\epsilon\rceil + k'}$, and that $\|\sum_{i=0}^{\infty} D_{i\cdot\lceil 1/\epsilon\rceil + k'}\|_1 = \|S_{k'}\|_1$ since $\sum_{i=0}^{\infty} D_{i\cdot\lceil 1/\epsilon\rceil + k'}$ is a block-diagonal subset-matrix of $S_{k'}$ containing the entire diagonal of $S_{k'}$. Thus,

$$\|K_{k'}\|_1 = \|S_{k'} - \sum_{i=0}^{\infty} D_{i\cdot\lceil 1/\epsilon\rceil + k'}\|_1 \leq \|S_{k'}\|_1 + \|\sum_{i=0}^{\infty} D_{i\cdot\lceil 1/\epsilon\rceil + k'}\|_1 = 2\|S_{k'}\|_1 \leq 2\epsilon$$

We now define a "cut down" version of φ by $\tilde{\varphi} \equiv \varphi - K_{k'}$. From this definition we have:

$$\|\varphi - \tilde{\varphi}\|_1 = \|K_{k'}\|_1 \leq 2\epsilon. \tag{16}$$

Further, we define the projectors

$$\eta_j \equiv \sum_{2((j-1)\cdot\lceil 1/\epsilon\rceil + k') \leq l < 2(j\cdot\lceil 1/\epsilon\rceil + k')} P_l, \tag{17}$$

and define the block diagonal matrix φ_{block} as:

$$\varphi_{\text{block}} \equiv \sum_j Q_j \tilde{\varphi} Q_j = \sum_j Q_j(\varphi - K_{k'})Q_j = \sum_j Q_j \varphi Q_j \tag{18}$$

where the last equality follows because $\sum_j Q_j K_{k'} Q_j = 0$ because $K_{k'}$ consists only of the "anti-diagonal" components $A_{i\cdot\lceil 1/\epsilon\rceil + k'}$ which lie outside of the Q_j. Note that φ_{block} is a subset-matrix of $\tilde{\varphi}$ according to Definition 22. Now define φ_{far} by:

$$\varphi_{\text{far}} \equiv \tilde{\varphi} - \varphi_{\text{block}} \tag{19}$$

Therefore, φ_{far} is also a subset-matrix of $\tilde{\varphi}$ according to Definition 22. Furthermore, it follows immediately using Equation 16 that:

$$\|\varphi - (\varphi_{\text{far}} + \varphi_{\text{block}})\|_1 = \|\varphi - \tilde{\varphi}\|_1 \leq 2\epsilon \tag{20}$$

Second Claim: To establish the second claim in Lemma 24 we now show that $T_{\varphi_{\text{far}}} \subseteq \{(l,k) : |k-l| > B\}$ (recall that $B \equiv 30 + 2\left\lceil \frac{\log(1/\epsilon)}{N} \right\rceil$). To see this, we consider the case that $|k-l| \leq B$ and show that in this case $(l,k) \notin T_{\varphi_{\text{far}}}$. Assume WLOG that $k \geq l$. When $|k-l| \leq B$ we know that either $\exists j$ such that:

$$2B(2(j-1)\lceil 1/\epsilon\rceil + 2k' - 1) < l, k \leq 2B(2j\lceil 1/\epsilon\rceil + 2k' - 1) \tag{21}$$

or $\exists j$ such that:

$$4B(j\lceil 1/\epsilon\rceil + k') - 3B \leq l \leq 2B(2j\lceil 1/\epsilon\rceil + 2k' - 1) \leq k \leq 4B(j\lceil 1/\epsilon\rceil + k') - B \tag{22}$$

In the first case, denoted by Equation 21, we have that the coordinates (l,k) lie within the subset-matrix φ_{block} of φ, and thus that either $(l,k) \in T_{\varphi_{\text{block}}}$ or $(l,k) \notin T_\varphi$ by definition. In particular, either $(l,k) \in T_{Q_j \varphi Q_j} \subseteq T_{\varphi_{\text{block}}}$ as follows by Equation 18 and the definition of Q_j in Equation 17, or $(l,k) \notin T_\varphi$. If $(l,k) \in T_{\varphi_{\text{block}}}$ then we note that $T_{\varphi_{\text{block}}} \cap T_{\varphi_{\text{far}}} = \emptyset$ by definition (Equation 19), and this implies that $(l,k) \notin T_{\varphi_{\text{far}}}$. If $(l,k) \notin T_\varphi$, then $(l,k) \notin T_{\varphi_{\text{far}}}$ because $T_{\varphi_{\text{far}}} \subseteq T_\varphi$.

On the other hand, in the case denoted by Equation 22, we have the coordinates (l,k) lie within the subset-matrix $K_{k'}$ of φ, and thus that either $(l,k) \in T_{K_{k'}}$, or $(l,k) \notin T_\varphi$. The reason for this is that we know that, in this case, the coordinates (l,k) are within the subset-matrix $M_{j\lceil 1/\epsilon\rceil + k'}$ of φ. Furthermore, since we have already ruled out the case of Equation 21, we know that (l,k) is not in $D_{j\lceil 1/\epsilon\rceil + k'}$, the block diagonal portion of $M_{j\lceil 1/\epsilon\rceil + k'}$. Therefore, the coordinates (l,k) must lie in the block-anti-diagonal portion $A_{j\lceil 1/\epsilon\rceil + k'} = M_{j\lceil 1/\epsilon\rceil + k'} - D_{j\lceil 1/\epsilon\rceil + k'}$ (this can also be determined directly from Equation 22 itself, and the definition of $A_{j\lceil 1/\epsilon\rceil + k'}$). Since $K_{k'} \equiv \sum_{i=0}^{\infty} A_{i\lceil 1/\epsilon\rceil + k'}$ we know that the coordinates (l,k) lie within the $K_{k'}$, or more precisely, either $(l,k) \in T_{K_{k'}}$, or $(l,k) \notin T_\varphi$. Just as before, if $(l,k) \notin T_\varphi$, then $(l,k) \notin T_{\varphi_{\text{far}}} \subseteq T_\varphi$. On the other hand, in the case

that $(l,k) \in T_{K_{k'}}$ we know that $T_{K_{k'}} \cap T_{\varphi_{\text{far}}} = \emptyset$ because $T_{\varphi_{\text{far}}} \subseteq T_{\tilde{\varphi}}$ by Equation 19, and $T_{\tilde{\varphi}} \cap T_{K_{k'}} = \emptyset$ as follows from the definition $\tilde{\varphi} \equiv \varphi - K_{k'}$.

This establishes that $T_{\varphi_{\text{far}}} \subseteq \{(l,k) : |k-l| > B\}$.

Third Claim: To establish the third claim in Lemma 24, and complete the proof, we will show that φ_{block} is a mixture of states of spread at most $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$, which means that φ_{block} can be produced from a shared maximally entangled state with at most $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$ bits of communication.

Recalling the definition of φ_{block} in Equation 18, let us define $\rho'_j \equiv Q_j \varphi Q_j$, so that it is clear that $\varphi_{\text{block}} = \sum_j \rho'_j$. It is also clear that ρ'_j is not only PSD, but also an un-normalized pure state, because

$$\rho'_j \equiv Q_j \varphi Q_j = Q_j |\varphi\rangle \langle\varphi| Q_j.$$

From the definition of Q_j in Equation 17 we have that:

$$Q_j |\varphi\rangle = \sum_{B_s < l \leq B_b} |\varphi_l\rangle ,$$

Where the index limits are

$$B_s \equiv 2(2((j-1)\cdot \lceil 1/\epsilon \rceil + k') - 1)\cdot B$$
$$B_b \equiv 2(2(j\cdot \lceil 1/\epsilon \rceil + k') - 1)\cdot B.$$

We know from the definition in Equation 12 that the $|\varphi_l\rangle$ are orthogonal to each other, and that each $|\varphi_l\rangle$ has Schmidt coefficients bounded by $2^{-lN+1} \geq \lambda_i > 2^{-lN-1}$. Thus, it is immediate that ρ'_j has spread at most $(B_b - B_s)N + 4 = 2\lceil 1/\epsilon \rceil BN + 4 = O(N/\epsilon + \log(1/\epsilon)/\epsilon)$, where the last equality follows because $B = 30 + 2\left\lceil \frac{\log(1/\epsilon)}{N} \right\rceil$. Therefore φ_{block} is a normalized mixture of states with spread at most $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$.

Consider the normalized version of ρ'_j, which is still a pure state of spread at most $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$ it is clear that this state has Earthmover distance at most $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$ from the nearest maximally entangled state (simply move all of the weight onto Schmidt coefficients of the size of the smallest Schmidt coefficient, which can be done by moving all the weight a distance less than or equal to the spread). It follows, by using Theorem 3 that there is a protocol which prepares the normalized version of ρ'_i from EPR pairs, with only $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$ bits of communication (we note that this line of the proof could also have been established using result from [7, 9], for example). Now the state $\varphi_{\text{block}} \equiv \sum_i \rho'_i$ can be prepared by applying this same protocol in superposition over i (with the probability $\text{tr}(\rho'_i)$ assigned to each i), and then tracing out over the i register. Thus φ_{block} can be prepared starting from EPR pairs with $O(N/\epsilon + \log(1/\epsilon)/\epsilon)$ bits of communication. ◀

C Proof of Lemma 15

Proof. First note that it is immediate from the definitions that $\langle \phi^2 | \psi^1 \rangle = \langle \phi^3 | \psi^2 \rangle = 0$, so the conditions of the lemma are automatically satisfied in those cases.

To bound the remaining inner products we will first prove a bound on the inner product $|\langle \phi^3 | \psi^1 \rangle|$ and note that the remaining inner products are bounded as a consequence of this first bound. For notational convenience, while establishing the bound on $|\langle \phi^3 | \psi^1 \rangle|$, we set $|\rho\rangle \equiv |\psi\rangle_{\leq x}$, and let ρ_j be the non-zero Schmidt coefficients of $|\rho\rangle$ (which are just a renamed version of the non-zero Schmidt coefficients of $|\psi\rangle_{\leq x}$). Therefore, we know that, for all j, $1 \geq \rho_j \geq 2^{-x}$, and $|\psi\rangle_{\leq x} = |\rho\rangle = \sum_j \sqrt{\rho_j} |j\rangle \otimes |j\rangle$. The purpose of this renaming

convention is that we can now cleanly make the following definition. For integers i define $|\rho\rangle_i \equiv \sum_{\{j:i<|\log 1/\rho_j|\leq i+1\}} \sqrt{\rho_j}\,|j\rangle \otimes |j\rangle$, so that we have $|\psi\rangle_{\leq x} = |\rho\rangle = \sum_{i=-1}^{\lceil x\rceil} |\rho\rangle_i$, and $\langle\rho_k|\rho_i\rangle = 0$ whenever $k \neq i$. So,

$$\sum_{i=-1}^{\lceil x\rceil} \||\rho\rangle_i\|^2 = \||\rho\rangle\|^2 \leq 1 \tag{23}$$

By definition, for any $1 \leq i \leq \lceil x\rceil$, the Schmidt coefficients of $|\rho\rangle_i$ are upper bounded by 2^{-i}, and lower bounded by $2^{-(i+1)}$, and from the latter we have $rk_{Schmidt}(|\rho\rangle_i) \leq 2^{i+1}\||\rho\rangle\|^2$. Furthermore, the Schmidt coefficients of $|\phi_{\geq x+d}\rangle$ are upper bounded by $2^{-(x+d)}$, and thus, we have by Lemma 8 that:

$$\begin{aligned}&|\langle\phi_{\geq x+d}|\,U_{\mathcal{P}}\,|\rho\rangle_i| \leq 2^{\frac{3}{2}Q} rk_{Schmidt}(|\rho\rangle_i)\sqrt{2^{-(x+d)}2^{-i}} \leq 2^{\frac{3}{2}Q}\cdot 2^{i+1}\,\||\rho\rangle_i\|^2 \cdot \sqrt{2^{-(x+d)}2^{-i}}\\ &= 2\cdot 2^{\frac{3}{2}Q}\,\||\rho\rangle_i\|^2\sqrt{2^{i-x-d}} \leq 2\cdot 2^{\frac{3}{2}Q}\,\||\rho\rangle_i\|^2\cdot 2\cdot 2^{-d/2} = 4\cdot 2^{\frac{3Q-d}{2}}\,\||\rho\rangle_i\|^2,\end{aligned} \tag{24}$$

where the final inequality follows because $i \leq \lceil x\rceil$ by assumption. Thus,

$$\begin{aligned}&\left|\left\langle\phi^3\middle|\psi^1\right\rangle\right| = \left|\langle\phi_{\geq x+d}|\,U_{\mathcal{P}}\,|\psi\rangle_{\leq x}\right| = \left|\sum_{i=-1}^{\lceil x\rceil}\langle\phi_{\geq x+d}|\,U_{\mathcal{P}}\,|\rho\rangle_i\right| \leq \sum_{i=-1}^{\lceil x\rceil}|\langle\phi_{\geq x+d}|\,U_{\mathcal{P}}\,|\rho\rangle_i|\\ &\leq 4\cdot 2^{\frac{3Q-d}{2}}\sum_{i=-1}^{\lceil x\rceil}\||\rho\rangle_i\|^2 = 4\cdot 2^{\frac{3Q-d}{2}}\left\||\psi\rangle_{\leq x}\right\|^2 \leq 4\cdot 2^{\frac{3Q-d}{2}} = h(Q,d),\end{aligned} \tag{25}$$

where the second inequality follows by Equation 24 and the subsequent equality follows by Equation 23. Having established this upper bound on $\left|\left\langle\phi^3\middle|\psi^1\right\rangle\right|$ we now proceed with bounding the other inner products in the Lemma statement:

$$\left|\left\langle\psi^3\middle|\psi^1\right\rangle\right| = \left|\langle\psi_{>x}|\,U_{\mathcal{P}}^\dagger\left|\phi^3\right\rangle\left\langle\phi^3\middle|\psi^1\right\rangle\right| = \left|\langle\psi_{>x}|\,U_{\mathcal{P}}^\dagger\left|\phi^3\right\rangle\right|\left|\left\langle\phi^3\middle|\psi^1\right\rangle\right| \leq \left|\left\langle\phi^3\middle|\psi^1\right\rangle\right| \leq h(Q,d), \tag{26}$$

$$\begin{aligned}&\left|\left\langle\psi^2\middle|\psi^1\right\rangle\right| = \left|\langle\psi_{>x}|\,U_{\mathcal{P}}^\dagger\left(I-\left|\phi^3\right\rangle\left\langle\phi^3\right|\right)\left|\psi^1\right\rangle\right| \leq \left|\langle\psi_{>x}|\,U_{\mathcal{P}}^\dagger\left|\psi^1\right\rangle\right| + \left|\langle\psi_{>x}|\,U_{\mathcal{P}}^\dagger\left|\phi^3\right\rangle\left\langle\phi^3\middle|\psi^1\right\rangle\right|\\ &= \left|\langle\psi_{>x}|\,U_{\mathcal{P}}^\dagger U_{\mathcal{P}}\,|\psi\rangle_{\leq x}\right| + \left|\left\langle\psi^3\middle|\psi^1\right\rangle\right| = |\langle\psi_{>x}|\psi_{\leq x}\rangle| + \left|\left\langle\psi^3\middle|\psi^1\right\rangle\right| = \left|\left\langle\psi^3\middle|\psi^1\right\rangle\right| \leq h(Q,d),\end{aligned}$$

where both of the inequality steps follow by Equation 26 (the first of which also uses the triangle inequality).

$$\left|\left\langle\phi^3\middle|\phi^1\right\rangle\right| = \left|\left\langle\phi^3\middle|\psi^1\right\rangle\left\langle\psi^1\middle|\phi_{<x+d}\right\rangle\right| = \left|\left\langle\phi^3\middle|\psi^1\right\rangle\right|\left|\left\langle\psi^1\middle|\phi_{<x+d}\right\rangle\right| \leq \left|\left\langle\phi^3\middle|\psi^1\right\rangle\right| \leq h(Q,d),$$

$$\begin{aligned}&\left|\left\langle\phi^3\middle|\phi^2\right\rangle\right| = \left|\left\langle\phi^3\right|\left(I-\left|\psi^1\right\rangle\left\langle\psi^1\right|\right)|\phi\rangle_{<x+d}\right| \leq \left|\left\langle\phi^3\middle|\phi_{<x+d}\right\rangle\right| + \left|\left\langle\phi^3\middle|\psi^1\right\rangle\left\langle\psi^1\middle|\phi_{<x+d}\right\rangle\right|\\ &= |\langle\phi_{>x+d}|\phi_{<x+d}\rangle| + \left|\left\langle\phi^3\middle|\psi^1\right\rangle\right|\left|\left\langle\psi^1\middle|\phi_{<x+d}\right\rangle\right| = \left|\left\langle\phi^3\middle|\psi^1\right\rangle\right|\left|\left\langle\psi^1\middle|\phi_{<x+d}\right\rangle\right| \leq \left|\left\langle\phi^3\middle|\psi^1\right\rangle\right| \leq h(Q,d)\end{aligned}$$

Now, as noted earlier, $\left\langle\phi^2\middle|\psi^1\right\rangle = \left\langle\phi^3\middle|\psi^2\right\rangle = 0$. Continuing with the cross terms we have:

$$\left|\left\langle\phi^1\middle|\psi^2\right\rangle\right| = \left|\left\langle\psi^2\middle|\phi^1\right\rangle\right| = \left|\left\langle\psi^2\middle|\psi^1\right\rangle\left\langle\psi^1\middle|\phi_{<x+d}\right\rangle\right| = \left|\left\langle\psi^2\middle|\psi^1\right\rangle\right|\left|\left\langle\psi^1\middle|\phi_{<x+d}\right\rangle\right| \leq \left|\left\langle\psi^2\middle|\psi^1\right\rangle\right| \leq h(Q,d),$$

$$\left|\left\langle\phi^1\middle|\psi^3\right\rangle\right| = \left|\left\langle\psi^3\middle|\phi^1\right\rangle\right| = \left|\left\langle\psi^3\middle|\psi^1\right\rangle\left\langle\psi^1\middle|\phi_{<x+d}\right\rangle\right| = \left|\left\langle\psi^3\middle|\psi^1\right\rangle\right|\left|\left\langle\psi^1\middle|\phi_{<x+d}\right\rangle\right| \leq \left|\left\langle\psi^3\middle|\psi^1\right\rangle\right| \leq h(Q,d),$$

where the last inequality follows from Equation 26. And, since we already have $\left|\langle\phi^3|\psi^1\rangle\right| \leq h(Q,d)$ from Equation 25, the final inner product to bound is:

$$\begin{aligned}
&\left|\langle\phi^2|\psi^3\rangle\right| = \left|\langle\phi|_{<x+d}\left(I - |\psi^1\rangle\langle\psi^1|\right)|\psi^3\rangle\right| \\
&\leq \left|\langle\phi_{<x+d}|\psi^3\rangle\right| + \left|\langle\phi_{<x+d}|\psi^1\rangle\langle\psi^1|\psi^3\rangle\right| \\
&= \left|\langle\phi_{<x+d}|\phi^3\rangle\langle\phi^3|U_{\mathcal{P}}|\psi\rangle_{>x}\right| + \left|\langle\phi_{<x+d}|\psi^1\rangle\langle\psi^1|\psi^3\rangle\right| \\
&= \left|\langle\phi_{<x+d}|\phi^3\rangle\right|\left|\langle\phi^3|U_{\mathcal{P}}|\psi\rangle_{>x}\right| + \left|\langle\phi_{<x+d}|\psi^1\rangle\right|\left|\langle\psi^1|\psi^3\rangle\right| \\
&= \left|\langle\phi_{<x+d}|\phi_{>x+d}\rangle\right|\left|\langle\phi^3|U_{\mathcal{P}}|\psi\rangle_{>x}\right| + \left|\langle\phi_{<x+d}|\psi^1\rangle\right|\left|\langle\psi^1|\psi^3\rangle\right| \\
&\leq 0 + \left|\langle\psi^1|\psi^3\rangle\right| \leq h(Q,d),
\end{aligned}$$

where the last inequality follows by Equation 26. ◀

D Proof of Lemma 18

Proof. Given two states $|\chi\rangle = \sum_{i\in X}\sqrt{\chi_i}\,|i\rangle\otimes|i\rangle$ and $|\upsilon\rangle = \sum_{j\in Y}\sqrt{\upsilon_j}\,|j\rangle\otimes|j\rangle$, let $\omega(i,j) : X\times Y \to \mathbb{R}_{\geq 0}$ be the joint distribution on $X\times Y$ which satisfies the ℓ_∞ Earth Mover conditions for $|\chi\rangle$ and $|\upsilon\rangle$, and acheives the optimal earth mover bound $d_\infty(|\chi\rangle, |\upsilon\rangle)$. That is, for all $i\in X$, $\sum_{j\in Y}\omega(i,j) = \chi_i$, for all $j\in Y$, $\sum_{i\in X}\omega(i,j) = \upsilon_j$, and $\omega(i,j) = 0$ whenever $|\log(\chi_i) - \log(\upsilon_j)| > d_\infty(|\chi\rangle, |\upsilon\rangle)$.

Define $|\rho\rangle \equiv \sum_{j\in Y}\sum_{k\in[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]}\sqrt{\rho_{j,k}}\,|j\rangle\otimes|k\rangle\otimes|j\rangle\otimes|k\rangle$, where

$$\rho_{j,k} \equiv \upsilon_j / 2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}.$$

We now define the intermediate state

$$|\gamma\rangle \equiv \sum_{j\in Y}\sum_{k\in[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]}\sum_{r\in[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]}\sqrt{\gamma_{j,k,r}}\,|j\rangle\otimes|k\rangle\otimes|r\rangle\otimes|j\rangle\otimes|k\rangle\,|r\rangle\,,$$

where the Schmidt coefficients $\gamma_{j,k,r}$ are left unspecified for now.

In order to specify the Schmidt coefficients of the intermediate state $|\gamma\rangle$ as well as the Right Index-1 Flow from $|\chi\rangle$ to $|\gamma\rangle$, and the Left Index-1 Flow from $|\gamma\rangle$ to $|\rho\rangle$ we will first define "bins" for the Schmidt coefficients of $|\upsilon\rangle$ as follows:

For $l\in\mathbb{N}\cup\{0\}$ let $\Upsilon_l \equiv \{j\in Y : 2^{-l} \geq \upsilon_j \geq 2^{-(l+1)}\}$, and $X_l \equiv \{i\in X : 2^{-l} \geq \chi_i \geq 2^{-(l+1)}\}$. Define $\omega(X_m,\Upsilon_l) \equiv \sum_{(i,j)\in X_m\times\Upsilon_l}\omega(i,j)$.

▶ **Fact 26.** *If $|m-l| > d_\infty(|\chi\rangle, |\upsilon\rangle) + 1$, then $\omega(X_m,\Upsilon_l) = 0$*

Proof. Given $i\in X_m$, and $j\in\Upsilon_l$ we have by definition that $2^{-l}\geq\upsilon_j\geq 2^{-(l+1)}$, and $2^{-m}\geq\chi_i\geq 2^{-(m+1)}$, and therefore that $|\log(\chi_i) - \log(\upsilon_j)| \geq |m-l| - 1 > d_\infty(|\chi\rangle, |\upsilon\rangle)$, where the last equality follows by assumption. It follows by definition of $d_\infty(|\chi\rangle, |\upsilon\rangle)$ and of ω, that $\omega(i,j) = 0$. Since this is true for all $(i,j)\in X_m\times\Upsilon_l$, the claim follows. ◀

We will now specify an iterative, "greedy" procedure to define the Schmidt coefficients $\gamma_{j,k,c}$ as a function of the $|\chi\rangle$ and $|\rho\rangle$.

For each $(m,l)\in\mathbb{N}\cup\{0\}\times\mathbb{N}\cup\{0\}$ such that $\omega(X_m,\Upsilon_l) > 0$ we first note that by Fact 26 that $|m-l| < d_\infty(|\chi\rangle, |\upsilon\rangle) + 1$. Thus, for each $(i,j)\in X_m\times\Upsilon_l$,

$$\chi_i \geq 2^{-(m+1)} \geq 2^{-l-d_\infty(|\chi\rangle,|\upsilon\rangle)-2} \geq 2^{-l}/2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2} \geq \upsilon_j/2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2} \equiv \rho_{j,k}$$

for all $k\in[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]$.

Algorithm 1.

For all i set $\text{temp}_i = \chi_i$
Set $i_m = \min\{X_m\}$ for all m
for $l \in \mathbb{N} \cup \{0\}$ **do**
  Set $j := \min\{Y_l\}$;
  Set $k = 0$;
  Set overflow $= 0$
  for $m \in \mathbb{N} \cup \{0\}$ **do**
    if $\omega(X_m, \Upsilon_l) > 0$ **then**
      Set $\text{temp}_\omega = \omega(X_m, \Upsilon_l)$
      while $\text{temp}_\omega > 0$ **do**
        if $\sum_{r \leq \text{overflow}} \gamma_{j,k,r} < \rho_{j,k}$ **then**
          while $\text{temp}_\omega \geq \rho_{j,k} - \sum_{r \leq \text{overflow}} \gamma_{j,k,r}$ **do**
            if $k = 2^{\lceil d_\infty(|\chi\rangle, |\upsilon\rangle)\rceil + 2}$ **then**
              Set $j = j + 1$
              Set overflow $= 0$
              Set $k = 0$
            if $\text{temp}_{i_m} < \rho_{j,k} - \sum_{r \leq \text{overflow}} \gamma_{j,k,r}$ **then**
              Set $\gamma_{j,k,\text{overflow}+1} = \text{temp}_{i_m}$
              Set $\text{temp}_\omega = \text{temp}_\omega - \text{temp}_{i_m}$
              Set $\text{temp}_{i_m} = 0$
              Add an edge in the flow graph from i_m to $(j, k, \text{overflow} + 1)$
              Set $i_m = i_m + 1$
              Set overflow $=$ overflow $+ 1$
            if $\text{temp}_{i_m} \geq \rho_{j,k} - \sum_{r \leq \text{overflow}} \gamma_{j,k,r}$ and $\text{temp}_\omega \geq \rho_{j,k} - \sum_{r \leq \text{overflow}} \gamma_{j,k,r}$ **then**
              Set $\gamma_{j,k,\text{overflow}+1} = \rho_{j,k} - \sum_{r \leq \text{overflow}} \gamma_{j,k,r}$
              Set $\text{temp}_\omega = \text{temp}_\omega - \gamma_{j,k,\text{overflow}+1}$
              Set $\text{temp}_{i_m} = \text{temp}_{i_m} - \gamma_{j,k,\text{overflow}+1}$
              Add an edge in the flow graph $G_{X,Z}$ from i_m to $(j, k, \text{overflow} + 1)$
              Set $k = k + 1$
              Set overflow $= 0$
          if $k = 2^{\lceil d_\infty(|\chi\rangle, |\upsilon\rangle)\rceil + 2}$ **then**
            Set $j = j + 1$
            Set overflow $= 0$
            Set $k = 0$
          if $\text{temp}_\omega < \rho_{j,k} - \sum_{r \leq \text{overflow}} \gamma_{j,k,r}$ **then**
            if $\text{temp}_{i_m} \leq \text{temp}_\omega$ **then**
              Set $\gamma_{j,k,\text{overflow}+1} = \text{temp}_{i_m}$
              Set $\text{temp}_\omega = \text{temp}_\omega - \text{temp}_{i_m}$
              Set $\text{temp}_{i_m} = 0$
              Add an edge in the flow graph $G_{X,Z}$ from i_m to $(j, k, \text{overflow} + 1)$
              Set $i_m = i_m + 1$
              Set overflow $=$ overflow $+ 1$
            if $\text{temp}_{i_m} \geq \text{temp}_\omega$ **then**
              Set $\gamma_{j,k,\text{overflow}+1} = \text{temp}_\omega$
              Set $\text{temp}_\omega = 0$
              Set $\text{temp}_{i_m} = \text{temp}_{i_m} - \text{temp}_\omega$
              Add an edge in the flow graph $G_{X,Z}$ from i_m to $(j, k, \text{overflow} + 1)$
              Set overflow $=$ overflow $+ 1$
            if $k = 2^{\lceil d_\infty(|\chi\rangle, |\upsilon\rangle)\rceil + 2}$ **then**
              Set $j = j + 1$
              Set overflow $= 0$
              Set $k = 0$

One may check that Algorithm 1 defines Schmidt coefficients $\gamma_{j,k,r}$, satisfying

$$\sum_{j\in Y}\sum_{k\in[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]}\sum_{r\in[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]}\gamma_{j,k,r}=\sum_{i\in X}\chi_i=1,$$

as well as a Right Index-1 Flow from $|\chi\rangle$ to $|\gamma\rangle$, with degree at most $2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}\cdot 2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}=2^{2\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+4}$. In particular the Right Index-1 Flow from $|\chi\rangle$ to $|\gamma\rangle$ is constructed in Algorithm 1 by iteratively adding edges to form the bipartite flow-graph $G_{X,Z}$ where $Z\equiv(Y,[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}],[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}])$. Each line in the pseudocode which reads "Add an edge in the flow graph from i_m to $(j,k,\text{overflow}+1)$", or similar, adds a single edge to the graph $G_{X,Z}$ and the union of all these edges forms the bipartite flow $G_{X,Z}$ between X and Z. Furthermore, for the $\gamma_{j,k,r}$ defined by Algorithm 1,

$$\sum_{r\in[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]}\gamma_{j,k,r}=\rho_{j,k},$$

so that there is a Left Index-1 flow from $|\gamma\rangle$ to $|\rho\rangle$ defined by a bipartite graph between the Schmidt coefficients of $|\gamma\rangle$ and $|\rho\rangle$ respectively, in which, for every $(j,k,r)\in Y\times[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]\times[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]$, there is an edge from $\gamma_{j,k,r}$ to $\rho_{j,k}$ of weight $\gamma_{j,k,r}$. This Left Index-1 flow then clearly has degree $2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}$.

Finally, recall that,

$$\sum_{k\in[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]}\rho_{j,k}=\sum_{k\in[2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}]}\upsilon_j/2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}=\nu_j$$

So, by very similar reasoning, there is a Left Index-1 flow from $|\rho\rangle$ to $|\nu\rangle$ with degree exactly $2^{\lceil d_\infty(|\chi\rangle,|\upsilon\rangle)\rceil+2}$. ◀

E Proof of Lemma 19

Proof. By assumption there is a Right Index-1 Flow from $|\tau\rangle$ to $|\kappa\rangle$ with degree at most 2^Q, so there exists a bipartite graph $G_{X,Y}$ with vertex set $X\cup Y$, and edge set $E_{X,Y}$ (where X,Y represents the bipartition of the vertices), such that:

- Each vertex in $j\in Y$ has degree 1 in $G_{X,Y}$.
- For all $i\in X$, $\tau_i=\sum_{j\in Y:(i,j)\in E_{X,Y}}\kappa_j$.
- The maximum degree of any vertex $i\in X$ in $G_{X,Y}$ is 2^Q.

The protocol for Alice and Bob to start with shared state $|\tau\rangle$ and end up with shared state $|\kappa\rangle$ will proceed as follows: Beginning with the state $|\tau\rangle$ shared between Alice and Bob, we will refer to the register containing the Alice half of $|\tau\rangle$ as A, and the register containing the Bob half as B. Alice will append two additional registers, of Q qubits each, and initialize each of them to the all zeros state. We will call these two new registers C_1 and C_2 respectively. Alice will then perform a controlled unitary operation between A and the registers C_1 and C_2. She will then pass the register C_2 to Bob using Q qubits of quantum communication to do so. Bob will then perform a controlled unitary between B and C_2, Alice will perform a controlled unitary between A and C_1, and after that Alice and Bob will share the state $|\kappa\rangle$.

To describe the protocol more precisely we will define the specific controlled unitaries performed by Alice and Bob at each step. Beginning with a shared state $|\tau\rangle$, after Alice appends the two additional Q-qubit registers to her side of $|\tau\rangle$, the shared state looks as follows:

$$|\tau\rangle=\sum_{i\in X}\sqrt{\tau_i}\left|0^{\otimes Q}\right\rangle_{C_1}\otimes\left|0^{\otimes Q}\right\rangle_{C_2}\otimes|i\rangle_A\otimes|i\rangle_B$$

Where, initially, Alice holds the registers A, C_1, and C_2. Alice now performs a controlled unitary operation, acting on registers C_1 and C_2 and controlled on register A. To describe this controlled unitary concisely we will need to imagine that there is some total order on the elements $j \in Y$ (any total order will do, one can simply imagine that the j's are indexed by bit strings which encode integers), and we will define $s_{ij} \equiv |\{j' \in Y : j' < j, \text{ and } (i,j') \in E_{X,Y}\}|$. Note that, since every $i \in X$ has degree at most 2^Q, s_{ij} is always an integer between 0 and 2^Q, so it can always be expressed in binary as a Q-bit binary number. We will take this convention in the following argument.

Now to define Alice's controlled unitary: When controlled on $|i\rangle_A$ Alice's unitary moves the state $\left|0^{\otimes Q}\right\rangle_{C_1} \otimes \left|0^{\otimes Q}\right\rangle_{C_2}$ to the state $|i\text{-controlled}\rangle_{C_1C_2} \equiv \sum_{j \in Y:(i,j)\in E_{X,Y}} \sqrt{\kappa_j/\tau_i}\,|s_{ij}\rangle_{C_1} \otimes |s_{ij}\rangle_{C_2}$. Note that since s_{ij} is always a Q-bit binary string, it can always be contained in the Q-qubit registers C_1 and C_2. Further note that, since $\tau_i = \sum_{j \in Y:(i,j)\in E_{X,Y}} \kappa_j$ by assumption, $|i\text{-controlled}\rangle_{C_1C_2}$ is a normalized pure state. Thus there exists a unitary operation that moves $\left|0^{\otimes Q}\right\rangle_{C_1} \otimes \left|0^{\otimes Q}\right\rangle_{C_2}$ to $|i\text{-controlled}\rangle_{C_1C_2}$ and Alice need only perform this specific unitary when the control register is in state $|i\rangle_A$. So, when Alice applies this controlled unitary to her registers C_1, C_2 and A (where A is the controlling register), the resulting new shared state between Alice and Bob is:

$$|\tau\rangle = \sum_{i \in X} |i\text{-controlled}\rangle_{C_1C_2} \otimes |i\rangle_A \otimes |i\rangle_B \tag{27}$$

$$= \sum_{i \in X} \sum_{j \in Y:(i,j)\in E_{X,Y}} \sqrt{\tau_i} \cdot \sqrt{\kappa_j/\tau_i}\,|s_{ij}\rangle_{C_1} \otimes |s_{ij}\rangle_{C_2} \otimes |i\rangle_A \otimes |i\rangle_B \tag{28}$$

$$= \sum_{i \in X} \sum_{j \in Y:(i,j)\in E_{X,Y}} \sqrt{\kappa_j}\,|s_{ij}\rangle_{C_1} \otimes |s_{ij}\rangle_{C_2} \otimes |i\rangle_A \otimes |i\rangle_B \tag{29}$$

At this point Alice uses Q qubits of communication to pass the Q-qubit register C_2 to Bob. The resulting shared state is:

$$\sum_{i \in X} \sum_{j \in Y:(i,j)\in E_{X,Y}} \sqrt{\kappa_j}\,|s_{ij}\rangle_{C_1} \otimes |i\rangle_A \otimes |i\rangle_B \otimes |s_{ij}\rangle_{C_2}$$

Where Alice owns registers C_1 and A, and Bob owns registers C_2 and B. Now it is not hard to see from the definition of s_{ij} and the fact that every $j \in Y$ has degree exactly 1 in the graph $G_{X,Y}$, that there is a bijection mapping each $j \in Y$ to the tuple (i, s_{ij}). Alice and Bob both know this bijection since they know the description of $G_{X,Y}$, and since bijections are invertible, Alice and Bob can now both apply a local unitary which relabels the basis element $|i\rangle \otimes |s_{ij}\rangle$ to the basis element j. The resulting shared state is:

$$\sum_{i \in X} \sum_{j \in Y:(i,j)\in E_{X,Y}} \sqrt{\kappa_j}\,|j\rangle_A \otimes |j\rangle_B = \sum_{j \in Y} \sqrt{\kappa_j}\,|j\rangle_A \otimes |j\rangle_B \equiv |\kappa\rangle$$

Where the first equality follows because each $j \in Y$ appears in the initial sum exactly once (because j has degree exactly one in $G_{X,Y}$).

This completes the protocol. ◀

F Proof of Corollary 20

Proof. By definition, if there is a Left Index-1 Flow from $|\kappa\rangle$ to $|\tau\rangle$, then there is a Right Index-1 Flow from $|\tau\rangle$ to $|\kappa\rangle$ (which is the starting assumption of Lemma 19). One can check that, in the proof Lemma 19, every operation performed by Alice and Bob was reversible.

Therefore, the proof of this corollary is simply to start at the end of the proof of Lemma 19, and "reverse" every step of the proof in order from end to beginning (including the communication step...now communication goes from Bob to Alice rather than Alice to Bob). The result is the desired quantum communication protocol, which converts the shared state $|\kappa\rangle$ to the shared state $|\tau\rangle$ using Q qubits of communication. ◀

G Fact 27

▶ **Fact 27.** *For $p \in [0,1]$ and $0 \leq \epsilon \leq p$, $\sqrt{p-\epsilon}\sqrt{p} + \sqrt{1-p}\sqrt{1-p+\epsilon} \leq 1 - \frac{1}{8}\epsilon^2$*

Proof. Define $f(x) \equiv \sqrt{p-x}\sqrt{p}+\sqrt{1-p}\sqrt{1-p+x}$. Note that $f'(x) = -\frac{\sqrt{p}}{2\sqrt{p-x}} + \frac{\sqrt{1-p}}{2\sqrt{1-p+x}}$, and $f''(x) = -1/4\left(\frac{\sqrt{p}}{(p-x)^{3/2}} + \frac{\sqrt{1-p}}{(1-p+x)^{3/2}}\right)$. So, $f(0) = 1$, $f'(0) = 0$, and

$$f''(x) = -1/4\left(\frac{\sqrt{p}}{(p-x)^{3/2}} + \frac{\sqrt{1-p}}{(1-p+x)^{3/2}}\right) \leq -1/4\frac{\sqrt{p}}{(p-x)^{3/2}} \leq -1/4\frac{1}{p} \leq -1/4$$

for all $p \in [0,1]$ and $0 \leq x \leq p$. It follows by integration that:

$$f(x) = 1 + \int_0^x \int_0^z f''(y)dydz \leq 1 + \int_0^x \int_0^z (-1/4)dydz = 1 - \frac{1}{8}x^2$$

So,

$$\sqrt{p-\epsilon}\sqrt{p} + \sqrt{1-p}\sqrt{1-p+\epsilon} = f(\epsilon) \leq 1 - \frac{1}{8}\epsilon^2.$$ ◀