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

Matthew Coudron* Aram W. Harrow†

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 $\epsilon$-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 [15]. 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 $\ell_\infty$ Earth Mover’s distance, and we show that the communication cost of converting between $|\chi\rangle$ and $|\upsilon\rangle$ is upper bounded, up to a constant multiplicative factor, by $d_\infty(|\chi\rangle, |\upsilon\rangle)$. 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 $\epsilon$-Smoothed $\ell_\infty$-Earth Mover’s Distance, which is a natural smoothing of the $\ell_\infty$-Earth Mover’s Distance that we will define via a connection with optimal transport theory.

* Institute for Quantum Computing, University of Waterloo mcoudron@uwaterloo.ca.
† Center for Theoretical Physics, MIT. aram@mit.edu
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 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 $\epsilon$. (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 [9, 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 $R$ whose goal is to compute a joint function $f(x, y) \in \{0, 1\}$. Suppose that $R$ 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 $R'$ which simulates $R$ with error $\epsilon$, 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 $R$ computes $f$ with error $\epsilon'$ it follows that $R'$ 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 [10, 14], unitary gate simulation [6], and communication tasks involving quantum communication between referees and players [11]. 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.
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 $P$ using $Q$ qubits of communication and a shared entangled state $\ket{\psi}$, we can replace $\ket{\psi}$ with a “consolidated” state $\rho$ at the cost of error $\epsilon$. Moreover, $\rho$ can be prepared from a maximally entangled state using $O(Q/\epsilon + \log(1/\epsilon)/\epsilon)$ communication. Taking $\epsilon$ 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 $P$ to achieve this result, except to pre-compose it with a pre-processing protocol which starts with only EPR pairs, and prepares the state $\rho$ using only $O(Q/\epsilon + \log(1/\epsilon)/\epsilon)$ communication. $P$ can then be run on $\rho$ 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 [9, 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 $\ket{\chi}^{AB}$ and $\ket{\nu}^{AB}$ are bipartite pure quantum states, with vectors of Schmidt coefficients denoted respectively by $\chi$ and $\nu$. In this setting it is known that $\ket{\chi}$ can be exactly converted into $\ket{\nu}$ using LOCC if and only if $\chi$ is majorized by $\nu$ [13]. But the communication cost of this transformation is known only in a few special cases. If $\ket{\chi} = \ket{\chi_0}^{\otimes n}$ and $\ket{\nu} = \ket{\nu_0}^{\otimes n}$ for some states $\ket{\chi_0}$, $\ket{\nu_0}$, then this cost is $O(\sqrt{n})$ or less in some special cases (e.g. $\ket{\nu_0}$ 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 $\ell_\infty$ Earth Mover’s (or Wasserstein) Distance, which tells us approximately how much quantum communication is required to transform $\ket{\chi}$ to $\ket{\nu}$. Given its simple form, we believe that this quantity may be a useful tool in quantum information theory.

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

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

We can think of $\chi$ as corresponding to placing $\chi_i$ mass at position $\log(\chi_i)$ for each $i$, and similarly for $\nu$. Then $d_\infty(\ket{\chi}, \ket{\nu})$ is the $\ell_\infty$ 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 $|\psi\rangle^{AB}$ be two bipartite shared states. There is a protocol $\mathcal{M}_{\chi \rightarrow \psi}$ which can prepare $|\psi\rangle$ from $|\chi\rangle$, using only $4\lceil d_{\infty}(|\chi\rangle, |\psi\rangle) \rceil + 8$ qubits of communication.

In Section 3 we establish a complementary lower bound, showing that a “$\epsilon$-smoothed” version of the $\ell_\infty$ Earth Mover’s Distance, denoted by $d_{\infty}^\epsilon(|\psi\rangle, |\phi\rangle)$, gives a lower bound on the cost of state transformation. That is:

Theorem 4. Given any two bipartite shared states $|\psi\rangle^{AB} = \sum_i \sqrt{\mu_i} |i\rangle^A \otimes |i\rangle^B$ and $|\phi\rangle^{AB} = \sum_i \sqrt{\pi_i} |i\rangle^A \otimes |i\rangle^B$, shared between two parties $A$ and $B$, together with a unitary $U_P$ 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 $\epsilon$:

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

In words: If two shared states cannot be brought within small $\ell_\infty$ Earth Mover’s Distance of each other by moving an $\epsilon$ 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 $\Omega(d_{\infty}^\epsilon(|\psi\rangle, |\phi\rangle))$ qubits of communication (for sufficiently large values of $d_{\infty}^\epsilon(|\psi\rangle, |\phi\rangle)$). Thus, the $\epsilon$-smoothed $\ell_\infty$ Earth Mover’s Distance provides a lower bound on the communication cost of state transformation. On the other hand, from the definition of $d_{\infty}^\epsilon(|\psi\rangle, |\phi\rangle)$, stated in Definition 10, we note here that one can use Theorem 3 to move $|\psi\rangle$ to within $1 - \epsilon$ fidelity of $|\phi\rangle$ using only $O(d_{\infty}^\epsilon(|\psi\rangle, |\phi\rangle))$ qubits of communication. To do this, omit the $\epsilon$ mass of Schmidt coefficients on which the two states have large $\epsilon$-smoothed $\ell_\infty$ distance, and apply Theorem 3 as one would do with the regular $\ell_\infty$ Earth Mover’s Distance. In this sense $d_{\infty}^\epsilon(|\psi\rangle, |\phi\rangle)$ 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 [12], 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, 8]. These lower bounds could be applied to general state conversion but relied on Rényi 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 1. In theorem statements above, and where appropriate, we have made use of superscripts $A$ and $B$, as in $|\psi\rangle^{AB} = \sum_i \sqrt{\mu_i} |i\rangle^A \otimes |i\rangle^B$ 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 2. When considering a bipartite state $|\psi\rangle$, we will assume that the state has a Schmidt decomposition of the form $|\psi\rangle = \sum_i \sqrt{\mu_i} |i\rangle \otimes |i\rangle$ 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.

## 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 6 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 5.

**Lemma 5.** Given two quantum states $|\psi\rangle$ and $|\nu\rangle$ on $\mathcal{H}_A \otimes \mathcal{H}_B$, such that the Schmidt coefficients of $\psi$ are upper bounded by $\lambda_{\max}$, and those of $\nu$ are upper bounded by $\nu_{\max}$, and further given a unitary transformation $U$ 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:

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

**Proof.** If $U$ is a unitary transform using $Q$ qubits of communication, then $rk_{\text{Schmidt}}(U| \nu \rangle) \leq 2^Q rk_{\text{Schmidt}}(|\nu\rangle)$ [8]. We also know that the Schmidt coefficients of $U| \nu \rangle$ are bounded above by $2^Q \nu_{\max}$ [8]. The desired result now follows by Lemma 6.

**Lemma 6.** Given two quantum states $|\psi\rangle$ and $|\nu\rangle$ on $\mathcal{H}_A \otimes \mathcal{H}_B$, such that the Schmidt coefficients of $\psi$ are upper bounded by $\lambda_{\max}$, and those of $\nu$ are upper bounded by $\nu_{\max}$, we have:

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

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

$$\langle \psi |\nu \rangle = \sum_{i=0}^{r-1} \sum_j \sqrt{\lambda_i} \lambda_j \langle i\rangle_A \otimes \langle j\rangle_B = \sum_{i=0}^{r-1} \sum_j \sqrt{\lambda_i} \lambda_j \langle i\rangle_A \otimes \langle j\rangle_B = \sum_{i=0}^{r-1} \sum_j \sqrt{\lambda_i} \lambda_j \langle i\rangle_A \otimes \langle j\rangle_B \otimes |i\rangle_B^*$$

Now, by definition of a Schmidt Decomposition, we know that the maximum singular value of $M_\nu$ is $\sqrt{\nu_{\max}}$. Thus, for all $i$ we have that $|\langle i\rangle_A |M_\nu |i\rangle_B^*| \leq \sqrt{\nu_{\max}}$ (since $|i\rangle_A$ and $|i\rangle_B$ are normalized vectors by definition). It then follows that:
\[
|\langle \psi |v \rangle| = \left| \sum_{i=0}^{r-1} \sqrt{\lambda_i} \langle i_A | M_v | i_B^* \rangle \right| \leq \sqrt{\lambda_{\text{max}}} \sum_{i=0}^{r-1} | \langle i_A | M_v | i_B^* \rangle | \leq r \sqrt{\lambda_{\text{max}}} v_{\text{max}} = r k_{\text{Schmidt}}(|\psi\rangle) \sqrt{\lambda_{\text{max}}} v_{\text{max}}
\]

Theorem 1 is the main result of this work. The proof is long enough that a high-level outline may be valuable. 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 E of the Appendix.

**Theorem (Restatement of Theorem 1).** Consider a quantum communication protocol \( \mathcal{R} \) whose goal 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 \( \epsilon \), 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 \( \epsilon' \) it follows that \( \mathcal{R}' \) computes \( f \) with error \( \epsilon + \epsilon' \).

**Outline of the Proof of Theorem 1:** The proof of Theorem 1 has three main parts. First, the initial entangled state used by the protocol can be converted using a small amount of communication to a state \( \varphi \) in which the Schmidt coefficients are grouped into evenly spaced groups. This is achieved using Theorem 3.

Second, we show that our new “grouped” entangled state can be divided into three “pieces” (more precisely termed subset-matrices in Definition 20), one piece which has small trace norm and can therefore be omitted, one piece called \( \varphi_{\text{far}} \) which only has non-zero terms which are far from the diagonal in the appropriate basis, and one piece called \( \varphi_{\text{block}} \) which is a block-diagonal mixed state that can be produced with small error and low communication cost from a maximally entangled state.

To show that, if one starts with a quantum communication protocol with an arbitrary shared entangled state, then that protocol can be modified, using a small amount of additional communication, to instead use an entangled state, \( \varphi \), (a property which will be useful later in the proof). Once we have reduced, without loss of generality, to appropriately “grouped” entangled state \( \varphi \) in this way, the proof proceeds in two halves. In the first half, which is summed up in Lemma 22, we show that

In the second half of the proof, which is summed up in Lemma 23, we show that the \( \varphi_{\text{far}} \) piece of \( \varphi \) has very little effect on the outcome of the quantum communication protocol in question. This means that \( \varphi \) can be replaced by \( \varphi_{\text{block}} \) alone while incurring very little error in the outcome of the quantum communication protocol. Since \( \varphi_{\text{block}} \) can be produced with low cost from a maximally entangled state, this then achieves the desired result. The full proof of Theorem 1 is included in Section E of the Appendix. The role of Lemma 5 in the proof is within this step for controlling the terms far from the diagonal, in Lemma 23.

### 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 \( \epsilon \)-Smoothed \( \ell_\infty \) 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 $\ell_\infty$ 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 $\epsilon$-smoothed $\ell_\infty$ 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 $\ell_\infty$ Earth Mover’s Distance which is derived by establishing the relationship between the $\ell_\infty$ 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 9, the generalization to the $\epsilon$-smoothed $\ell_\infty$ Earth Mover’s Distance in Definition 10 is straightforward and natural.

**Definition 7.** Given two probability distributions $\mu$ and $\nu$ on the real line, and a function $c : \mathbb{R} \times \mathbb{R} \to [0, \infty]$ the corresponding Monge-Kantorovich distance, $d_{MK}(\mu, \nu)$ between $\mu$ and $\nu$ is defined as:

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

Where $\Gamma(\mu, \nu)$ is defined to be the collection of all probability distributions on $X \times Y \equiv \mathbb{R} \times \mathbb{R}$ which have marginal on $X$ equal to $\mu$ and marginal on $Y$ equal to $\nu$.

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

**Definition 8.** Given a bipartite shared state $|\psi\rangle = \sum_{i \in X} \sqrt{p_i} |i\rangle \otimes |i\rangle$ let us define a random variable $V_\psi$ which takes value $\log(p_i)$ with probability $p_i$ (note that, since the $p_i$ sum to one, this is a well defined random variable). We now define $p_\psi$ to be the probability distribution of this random variable.

It is clear that, for every $\psi$, $p_\psi$ is a probability distribution on the real line. One may note the following simple relationship between Monge-Kantorovich distance and $\ell_\infty$ Earth Mover’s Distance:

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

Given this concrete connection between $\ell_\infty$ 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 $\mu$ and $\nu$ be probability distributions supported on the real line, and let $F_\mu$ and $F_\nu$ 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 $\ell_\infty$ Earth Mover’s Distance is given by:

**Definition 9.**

$$d_{\infty}(|\psi\rangle, |\phi\rangle) \equiv \max_{q \in [0,1]} |F_\psi^{-1}(q) - F_\phi^{-1}(q)|$$
In the context of this equivalent formulation of \( \ell_\infty \) Earth Mover’s Distance, we can succinctly introduce a “smoothed” version of the same distance. The reader may note that, since the above definition of \( d_\infty (|\psi \rangle, |\phi \rangle) \) is evidently not robust against tiny changes of either distribution in the total variation distance it would be impossible to prove a lower bound of the form of Theorem 4 if stated using that definition. Hence the motivation for introducing a “smoothed” version of the distance measure, which has built-in robustness by definition.

**Definition 10.** \( \epsilon \)-Smoothed \( \ell_\infty \)-Earth Mover’s Distance

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

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

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

\[
\left| \langle \phi |^{AB} U_P |\psi \rangle^{AB} \right| \leq 1 - \frac{1}{4} \epsilon^2 + 24 \cdot 2^{-\frac{1}{4} (d_\infty^\epsilon (|\psi \rangle, |\phi \rangle) - 3Q)}
\]

Intuitively, Theorem 4 states that two bipartite shared states which are far apart in the \( \ell_\infty \)-Earth Mover’s Distance, cannot be made equal via a quantum communication protocol unless it uses at least \( \epsilon \cdot d_\infty^\epsilon (|\psi \rangle, |\phi \rangle) \) qubits of communication (for a particular constant \( \epsilon \) which can be computed from the statement of Theorem 4).

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

\[
\min_{r \in [p-\epsilon,p+\epsilon]} |F^{-1}_{pq}(p) - F^{-1}_{pq}(r)| = d
\]  
(1)

Suppose that \( F^{-1}_{pq}(p) < F^{-1}_{pq}(r) \) (if the opposite is true then we simply switch the roles of \( \psi \) and \( \phi \) and continue with the same proof). Define \( x \equiv F^{-1}_{pq}(p) \). Further define \( |\psi \rangle \equiv \sum_{i} \sqrt{\varphi_i} |i\rangle \otimes |i\rangle \), and \( |\phi \rangle \equiv |\psi \rangle - |\psi \rangle \). Similarly define \( |\phi \rangle \equiv \sum_{i} \sqrt{\psi_i} |i\rangle \otimes |i\rangle \), and \( |\phi \rangle \equiv |\phi \rangle - |\phi \rangle \). Note that \( |\phi \rangle \) and \( |\phi \rangle \) are orthogonal, as are \( |\phi \rangle \) and \( |\phi \rangle \).

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

The main idea in the proof of this theorem is that we can now partition \( |\psi \rangle, |\phi \rangle \) into three nearly orthogonal parts, depending on \( U_P \), as follows:

**Definition 11.**

\[
|\psi^1 \rangle \equiv U_P |\psi \rangle \ \text{for} \ \langle \psi^1 | \leq x, \quad |\phi^1 \rangle \equiv |\psi \rangle_\geq x+d, \quad |\psi^2 \rangle \equiv (I - |\phi^3 \rangle \langle \phi^3 |) U_P |\phi \rangle_\geq x \\
|\psi^3 \rangle \equiv |\phi^3 \rangle \langle \phi^3 | U_P |\psi \rangle_\geq x, \quad |\phi^1 \rangle \equiv |\psi^1 \rangle \langle \psi^1 | \langle \phi^1 | \leq x+d, \quad |\phi^2 \rangle \equiv (I - |\phi^1 \rangle \langle \phi^1 |) |\phi \rangle_\leq x+d
\]
Lemma 12. 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 \( h(Q, d) \equiv 4 \cdot 2^{30 - d} \).

The proof of Lemma 12 is given separately in the appendix. Within that proof is the key use of Lemma 5 which is the primary conceptual step in proving Theorem 4. Understanding the proof of Lemma 12 is also the best way of understanding the motivation behind Definition 11 above.

It follows from the definitions that:

\[
U_P |\psi\rangle = |\psi^1\rangle + |\psi^2\rangle + |\psi^3\rangle
\]

(2)

\[
|\phi\rangle = |\phi^1\rangle + |\phi^2\rangle + |\phi^3\rangle
\]

(3)

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 \( |\phi^3\rangle \perp |\phi^2\rangle + |\phi^3\rangle \). Together these imply

\[
1 = \left\| |\psi^1\rangle \right\|^2 + \left\| |\psi^2\rangle \right\|^2 + \left\| |\psi^3\rangle \right\|^2
\]

(4a)

\[
1 = \left\| |\phi^1\rangle \right\|^2 + \left\| |\phi^2\rangle \right\|^2 + \left\| |\phi^3\rangle \right\|^2
\]

(4b)

From Lemma 12 it follows that:

\[
|\langle \phi | P(\psi) \rangle| = |\langle \phi | U_P |\psi\rangle| = \left| \left( |\langle \phi^1 | \psi^1 \rangle| + |\langle \phi^2 | \psi^2 \rangle| + |\langle \phi^3 | \psi^3 \rangle| \right) \right|
\]

\[
\leq |\langle \phi^1 | \psi^1 \rangle| + |\langle \phi^2 | \psi^2 \rangle| + |\langle \phi^3 | \psi^3 \rangle| + 6 \cdot h(Q, d)
\]

\[
\leq \left\| |\psi^1\rangle \right\| \left\| |\psi^1\rangle \right\| + \left\| |\phi^2\rangle \right\| \left\| |\phi^2\rangle \right\| + \left\| |\phi^3\rangle \right\| \left\| |\phi^3\rangle \right\| + 6 \cdot h(Q, d)
\]

(5)

Now recall that

\[
\left\| |\psi^1\rangle \right\| = \left\| U_P |\psi\rangle \right\|_{\leq x} = \left\| |\psi\rangle \right\|_{\leq x} = \sqrt{p}
\]

\[
\left\| |\phi^2\rangle \right\| = \left\| |\phi\rangle \right\|_{\geq x + d} \geq \sqrt{1 - p + \epsilon}
\]

We now return to Equation 5. Setting \( x_i = \| |\psi^i\rangle \| \) and \( y_i = \| |\phi^i\rangle \| \) for \( i = 1, 2, 3 \) we have

\[
|\langle \phi | P(\psi) \rangle| \leq x_1 y_1 + x_2 y_2 + x_3 y_3 + 6 \cdot h(Q, d)
\]

(6)

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 \( \sqrt{p} y_1 + x_2 y_2 \leq x_1 y_1 x_1 y_1 \) where \( x_1 y_1 = \sqrt{x_1^2 + x_2^2} \) and \( y_1 y_1 = \sqrt{y_1^2 + y_2^2} \). Now define \( x_1 = \cos(\alpha), x_2 = \sin(\alpha), y_2 = \cos(\beta), y_3 = \sin(\beta) \) and we have

\[
x_1 y_1 + x_2 y_2 + x_3 y_3 \leq \cos(\alpha - \beta).
\]

(7)
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

\[ |\langle \phi|\mathcal{P}(\psi)\rangle| \leq \sqrt{p-\epsilon}\sqrt{p} + \sqrt{1-p}\sqrt{1-p+\epsilon} + 6 \cdot h(Q,d). \quad (8) \]

Finally we would like an upper bound independent of \(p\). This maximization is performed in the proof of Fact 18 from Section B of the Appendix and yields the following.

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

### 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 \(\ell_\infty\) 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 \(\ell_\infty\) 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 15, and Lemma 16 together with Corollary 17. At a high level Lemma 15 tells us that, given bipartite states \(|\chi\rangle\) and \(|v\rangle\), one can map the Schmidt coefficients of \(|\chi\rangle\) directly onto the Schmidt coefficients of \(|v\rangle\) using a series of bipartite “flows” that have small degree (where degree is a quantity defined below). Lemma 16 and Corollary 17 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 15 and 16 which, together, prove the desired theorem. We begin with a couple definitions establishing the concept of flows, as we use it here.

**Definition 13** (Right (Left) Index-1 Flow). Fix two states \(|\chi\rangle = \sum_{i \in X} \sqrt{\chi_i} |i\rangle \otimes |i\rangle\) and \(|v\rangle = \sum_{j \in Y} \sqrt{v_j} |j\rangle \otimes |j\rangle\). A Right Index-1 Flow from \(|\chi\rangle\) to \(|v\rangle\) is a bipartite graph \(G_{X,Y}\) with vertices given by \(X \cup Y\), and edge set \(E_{X,Y}\), such that:

- Each vertex in \(j \in Y\) has index 1 in \(G_{X,Y}\).
- For all \(i \in X\), 
  \[ \chi_i = \sum_{j \in Y : (i,j) \in E_{X,Y}} v_j \]
  If the roles of \(|\chi\rangle\) and \(|v\rangle\) are reversed in the above, then we say that there is a Left Index-1 Flow from \(|v\rangle\) to \(|\chi\rangle\). Equivalently, there is a Left Index-1 Flow from \(|v\rangle\) to \(|\chi\rangle\) exactly when there is a a Right Index-1 Flow from \(|\chi\rangle\) to \(|v\rangle\).

**Definition 14** (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 \(|v\rangle = \sum_{j \in Y} \sqrt{v_j} |j\rangle \otimes |j\rangle\) to be the maximum index 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 \(\ell_\infty\) 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 16. The main step in the proof of Lemma 15 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 15. 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 $2^{\lceil d_{\infty}(|\chi\rangle, |\upsilon\rangle) \rceil} + 4$, 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 15 is included in the Appendix, section G.

Lemma 15, 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 $\ell_\infty$ Earth Mover’s Distance between them. The next step is to establish that every flow can be implemented via a quantum communication protocol. Lemma 16 and Corollary 17, 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 16. 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 $P$, 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 xC.

Corollary 17 establishes the same result as Lemma 16, but in the reverse direction.

Corollary 17. 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 $P$, which uses $Q$ qubits of communication, and converts the shared state $|\kappa\rangle$ to the shared state $|\tau\rangle$.

The proof of Corollary 17 is straightforward and appears in Appendix D.

Theorem (Restatement of Theorem 3). Let $|\chi\rangle^{AB}$ and $|\upsilon\rangle^{AB}$ be two bipartite shared states. There is a protocol $\mathcal{M}_{\chi \rightarrow \upsilon}$ 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 15, followed by Lemma 16 and Corollary 17. \qed
A Proof of Lemma 12

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 \( | \rho \rangle \equiv | \psi \rangle \leq \chi \), and let \( \rho_j \) be the non-zero Schmidt coefficients of \( | \rho \rangle \) (which are just a renamed version of the non-zero Schmidt coefficients of \( | \psi \rangle \leq \chi \)). Therefore, we know that, for all \( j, i \geq \rho_j \geq 2^{-x} \), and \( | \psi \rangle \leq \chi = | \rho \rangle = \sum_j \sqrt{\rho_j} | j \rangle \otimes | j \rangle \). The purpose of this renaming convention is that we can now cleanly make the following definition. For integers \( i \) define \( | \rho \rangle_i \equiv \sum_{j : \rho_j < \frac{\log 1}{\rho_j} \leq i + 1} \sqrt{\rho_j} | j \rangle \otimes | j \rangle \), so that we have \( | \psi \rangle \leq \chi = | \rho \rangle = \sum_{i = -1}^{[\chi]} | \rho \rangle_i \), and \( \langle \rho | \rho_i \rangle = 0 \) whenever \( k \neq i \). So,

\[
\sum_{i = -1}^{[\chi]} \| | \rho \rangle_i \|^2 \leq \| | \rho \rangle \|^2 \leq 1
\]

(9)

By definition, for any \( 1 \leq i \leq [\chi] \), 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 5 that:

\[
\| \langle \phi \geq x + d | U_P | \rho \rangle \| \leq 2^{3Q} rk_{Schmidt}(| \rho \rangle_i) \sqrt{2^{-(x + d)2^{-i}}} \leq 2^{3Q} \cdot 2^{i+1} \| | \rho \rangle \|^2 \cdot \sqrt{2^{-(x + d)2^{-i}}} = 2 \cdot 2^{3Q} \| | \rho \rangle \|^2 \cdot \sqrt{2^{2^{-x + d} - 2^{-i} - 1}}
\]

(10)

where the final inequality follows because \( i \leq [\chi] \) by assumption. Thus,

\[
\| \langle \phi^3 | \psi^1 \rangle \| = \| \langle \phi \geq x + d | U_P | \psi \rangle \leq \chi \| = \sum_{i = -1}^{[\chi]} \langle \phi \geq x + d | U_P | \rho \rangle_i \leq \sum_{i = -1}^{[\chi]} \| \langle \phi \geq x + d | U_P | \rho \rangle_i \|
\]

\[
\leq 4 \cdot 2^{3Q - d} \sum_{i = 1}^{[\chi]} \| | \rho \rangle_i \|^2 \leq 4 \cdot 2^{3Q - d} \| | \psi \rangle \leq \chi \|^2 \leq 4 \cdot 2^{3Q - d} = h(Q, d),
\]

(11)

where the second inequality follows by Equation 10 and the subsequent equality follows by Equation 9. Having established this upper bound on \( \langle \phi^3 | \psi^1 \rangle \) we now proceed with bounding the other inner products in the Lemma statement:

\[
\| \langle \psi^3 | \psi^1 \rangle \| = \| \langle \psi \geq x | U_P^* | \phi^3 \rangle \langle \phi^3 | \psi^1 \rangle \| \leq \| \langle \psi \geq x | U_P^* | \phi^3 \rangle \| \| \langle \phi^3 | \psi^1 \rangle \| \leq h(Q, d),
\]

(12)

\[
\| \langle \psi^2 \psi^1 \rangle \| \leq \| \langle \psi \geq x | U_P^* | \phi^3 \rangle \| \| \langle \phi^3 | \psi^1 \rangle \| \leq \| \langle \psi \geq x | U_P^* \| \phi^3 \rangle \| \langle \phi^3 | \psi^1 \rangle \| = \| \langle \psi \geq x | U_P^* | \psi \rangle \leq \chi \| + \| \langle \phi^3 | \psi^1 \rangle \| \leq h(Q, d),
\]

where both of the inequality steps follow by Equation 12 (the first of which also uses the triangle inequality).
\[ |\langle \phi^3|\phi^1 \rangle| = |\langle \phi^3|\psi^1 \rangle \langle \psi^1|\phi_{<x+d} \rangle| = |\langle \phi^3|\psi^1 \rangle| |\langle \psi^1|\phi_{<x+d} \rangle| \leq |\langle \phi^3|\psi^1 \rangle| \leq h(Q, d), \]

\[ |\langle \phi^3|\phi^2 \rangle| = |\langle \phi^3|(I - |\psi^1 \rangle \langle \psi^1|)|\phi_{<x+d} \rangle| \leq |\langle \phi^3|\phi_{<x+d} \rangle| + |\langle \phi^3|\psi^1 \rangle \langle \psi^1|\phi_{<x+d} \rangle| \]

\[ = |\langle \phi_{>x+d}|\phi_{<x+d} \rangle| + |\langle \phi^3|\psi^1 \rangle| |\langle \psi^1|\phi_{<x+d} \rangle| = |\langle \phi^3|\psi^1 \rangle| |\langle \psi^1|\phi_{<x+d} \rangle| \leq |\langle \phi^3|\psi^1 \rangle| \leq h(Q, d) \]

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

\[ |\langle \phi^1|\phi^2 \rangle| = |\langle \phi^2|\phi^1 \rangle| = |\langle \phi^2|\psi^1 \rangle \langle \psi^1|\phi_{<x+d} \rangle| = |\langle \phi^2|\psi^1 \rangle| |\langle \psi^1|\phi_{<x+d} \rangle| \leq |\langle \phi^2|\psi^1 \rangle| \leq h(Q, d), \]

\[ |\langle \phi^1|\psi^3 \rangle| = |\langle \phi^3|\phi^1 \rangle| = |\langle \phi^3|\psi^1 \rangle \langle \psi^1|\phi_{<x+d} \rangle| = |\langle \phi^3|\psi^1 \rangle| |\langle \psi^1|\phi_{<x+d} \rangle| \leq |\langle \phi^3|\psi^1 \rangle| \leq h(Q, d), \]

where the last inequality follows from Equation 12. And, since we already have \( |\langle \phi^3|\psi^1 \rangle| \leq h(Q, d) \) from Equation 11, the final inner product to bound is:

\[ |\langle \phi^2|\psi^3 \rangle| = |\langle \phi_{<x+d}|(I - |\psi^1 \rangle \langle \psi^1|)|\psi^3 \rangle| \]

\[ \leq |\langle \phi_{<x+d}|\psi^3 \rangle| + |\langle \phi_{<x+d}|\psi^1 \rangle \langle \psi^1|\psi^3 \rangle| \]

\[ = |\langle \phi_{<x+d}|\phi^3 \rangle \langle \phi^3|U_P|\psi_{>x} \rangle| + |\langle \phi_{<x+d}|\psi^1 \rangle \langle \psi^1|\psi^3 \rangle| \]

\[ = |\langle \phi_{<x+d}|\phi^3 \rangle| |\langle \phi^3|U_P|\psi_{>x} \rangle| + |\langle \phi_{<x+d}|\psi^1 \rangle| |\langle \psi^1|\psi^3 \rangle| \]

\[ \leq 0 + |\langle \psi^1|\psi^3 \rangle| \leq h(Q, d), \]

where the last inequality follows by Equation 12.

\[ \square \]

### B Fact 18

**Fact 18.** For \( p \in [0, 1] \) and \( 0 \leq \varepsilon \leq p \cdot \sqrt{p - \varepsilon} + \sqrt{1 - p}\sqrt{1 - p + \varepsilon} \leq 1 - \frac{1}{8}\varepsilon^2 \)

**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

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

12
for all $p \in [0, 1]$ and $0 \leq x \leq p$. It follows by integration that:

$$f(x) = 1 + \int_{0}^{x} \int_{0}^{x} f''(y)dydz \leq 1 + \int_{0}^{x} \int_{0}^{x} (-1/4)dydz = 1 - \frac{1}{8}x^2$$

So,

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

\[\Box\]

C Proof of Lemma 16

\textbf{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 vertices given by $X \cup Y$, and edge set $E_{X,Y}$, such that:

- Each vertex in $j \in Y$ has index 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:

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

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 $|0^\otimes Q\rangle_{C_1} \otimes |0^\otimes Q\rangle_{C_2}$ 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 $|0^\otimes Q\rangle_{C_1} \otimes |0^\otimes Q\rangle_{C_2}$ 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:

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

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

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.

\[ \square \]

### D Proof of Corollary 17

**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 16). One can check that, in the proof Lemma 16, 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 16, 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.

\[ \square \]

### E 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 19** (Spread). For a finite dimensional bipartite entangled state $|\psi\rangle^{AB} = \sum_i \sqrt{\psi_i} |i\rangle^A \otimes |i\rangle^B$ let $\lambda_{\text{max}}$ be the maximum of the Schmidt coefficients of $\psi$, and let $\lambda_{\text{min}}$ be the minimum Schmidt coefficient. We define the spread of $|\psi\rangle$ to be the quantity $\log(\lambda_{\text{max}} / \lambda_{\text{min}})$. 

14
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 $\epsilon$-smoothed variant of the spread of a state [8, 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, 8].

**Theorem** (Restatement of Theorem 1). Consider a quantum communication protocol $R$ whose goal is to compute a joint function $g(x, y) \in \{0, 1\}$. Suppose that $R$ 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 $R'$ which simulates $R$ with error $\epsilon$, 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 $R$ computes $f$ with error $\epsilon'$ it follows that $R'$ computes $f$ with error $\epsilon + \epsilon'$.

**Proof.** Given $R$, $g$, and $|\psi\rangle$ as in the theorem statement, Schmidt decompose $|\psi\rangle$ as $\sum_i \sqrt{\lambda_i} |i, i\rangle$ (see Remark 2 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 $f : [0, 1] \rightarrow \{0, 1, \ldots, N\}$ given by

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

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

$$\left| \log \left( \frac{\nu_{ij}}{2^{-\left\lfloor \frac{\log(1/\lambda_i)}{N} \right\rfloor} N} \right) \right| = \left| \log (\lambda_i) - \log(f(\lambda_i)) + \left\lfloor \frac{\log(1/\lambda_i)}{N} \right\rfloor N \right|$$

$$= \left| \log (\lambda_i) - \log \left( 2 \left\lceil \frac{\log(1/\lambda_i)}{N} \right\rceil \right) \right| + \left\lfloor \frac{\log(1/\lambda_i)}{N} \right\rfloor N$$

$$\leq 1$$

Next, we can upper bound $d_\infty(|\psi\rangle, |\varphi\rangle) \leq N$ by considering the coupling in which each $\nu_{ij}$ is moved to $\lambda_i$. 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 $M$ by which Alice and Bob can prepare $|\varphi\rangle$ from $|\psi\rangle$, using $4 \left[ d_\infty(|\chi\rangle, |\nu\rangle) \right] + 8 \leq 4N + 8$ qubits of communication. (For this special case, of course a simpler protocol could also be used.)

Define $C \equiv R \circ M$ to be the composed protocol in which Alice and Bob start with shared state $|\varphi\rangle$, first use protocol $M$ to convert $|\varphi\rangle$ to $|\psi\rangle$, and then perform protocol $R$ using shared state $|\psi\rangle$ and inputs $x$ and $y$, to compute the joint function $g(x, y)$. It is evident that $C$ has exactly the same success probability as $R$. Since $M$ uses at most $4N + 8$ qubits of communication and $R$ uses $Q$ qubits of communication, $C$ can be performed with $Q + 4N + 8$ qubits of communication.
For $j$ a nonnegative integer, define $I_j := \{i : 2^{-j}N + 1 \geq \lambda_i > 2^{-j+1}N - 1 \}$ and define the subnormalized state

$$|\varphi_j\rangle \equiv \sum_{i \in I_j} \sqrt{\lambda_i} |i, i\rangle.$$  \hspace{1cm} (16)

From Equation (15) and the surrounding discussion, we have that $|\varphi\rangle = \sum_j |\varphi_j\rangle$. Furthermore, by the definition of $I_j$, it follows that $|\varphi_j\rangle$ has spread at most 2; note that the spread of $|\varphi_j\rangle$ does not depend on whether the state is normalized or not.

The idea of the proof is that different $|\varphi_j\rangle$ are not only orthogonal, but must remain approximately orthogonal even after a small amount of quantum communication. In particular, note that for any $j$, $\text{rk}_{\text{Schmidt}}(|\varphi_j\rangle) \leq 2^{j+1} ||\varphi_j||^2$. Furthermore, for all $j$ we have, by definition, that the Schmidt coefficients of $|\varphi_{rk}\rangle$ are bounded above by $2^{-j}N + 1$. Therefore, if $U$ is a unitary transform using $M$ qubits of communication, then, it follows by Lemma 5, that $\forall j, k$,

$$|\langle \varphi_k | U | \varphi_j \rangle| \leq 2^{2j+2\min(j,k)} \|\varphi_{\min(j,k)}\|^2 \sqrt{2^{-j}N + 1 \cdot 2^{-j}N + 1} \leq 2^{2j+2\min(j,k)} \|\varphi_{\min(j,k)}\|^2 \hspace{1cm} (17)$$

To apply this to our problem, we first note that the protocol $C$ depends, a priori, on the inputs $x, y$ to the function $g(x, y)$ that we wish to compute (just like the the protocol $R$). 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 $C$ (which we have already established is equal to the success probability of the original protocol $R$) can be expressed WLOG by performing $C$ and then computing the probability of outcomes when measuring the first qubit in the computational basis. The probability that such a measurement on protocol $C$ outputs $b \in \{0, 1\}$ is

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

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

$$\Pr[0] - \Pr[1] = \langle \varphi | P | \varphi \rangle = \sum_{j,k} \langle \varphi_j | P | \varphi_k \rangle \hspace{1cm} (18)$$

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

The proof will proceed as follows: In Lemma 22 we show that the density matrix $\rho = |\varphi\rangle \langle \varphi|$ can be divided into three “pieces” (in a manner that does not depend on the inputs $x, y$), one piece which has small trace norm and can therefore be omitted, one piece called $\varphi_{\text{far}}$ which only has non-zero terms which are far from the diagonal in the appropriate basis, and one piece called $\varphi_{\text{block}}$ 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 23, we show that the $\varphi_{\text{far}}$ piece of $\varphi$ has very little effect on the protocol $C$. This means
that \( \varphi \) can be replaced by \( \varphi_{\text{block}} \) alone while incurring very little error in the success probability of \( \mathcal{C} \). Stated equivalently, via the equality in Equation 18 above, Lemma 23 shows that the quantity \( |\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))| \) is small. Since we know from Lemma 22 that \( \varphi_{\text{block}} \) can be produced with low cost from a maximally entangled state, this leads us to the desired result. Since \( \varphi_{\text{block}} \) does not depend on the inputs \( x, y \) this same statement holds for every pair of inputs \( x, y \). From this point forward we will no longer specify the fixed inputs \( x, y \), as it will be clear that the state substitutions do not depend on these inputs, and thus that the argument holds for every input as discussed in this paragraph.

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

**Definition 20** (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 21** (Non-Zero Set). For an operator \( \theta \) on the Hilbert space which is the span of the \( |\varphi_j\rangle \), define the non-zero set of \( \theta \) to be \( T_{\theta} = \{ (l, k) : \langle \varphi_{k} | \theta | \varphi_{l} \rangle \neq 0 \} \).

**Lemma 22.** Consider the density matrix \( \varphi \equiv \sum_{k} |\varphi_{k}\rangle \langle \varphi_{k}|. \) For any \( \epsilon > 0 \), there exist subset-matrices, \( \varphi_{\text{blocks}}, \varphi_{\text{fars}} \), of \( \varphi \), such that

1. \( \| \varphi - (\varphi_{\text{block}} + \varphi_{\text{far}}) \|_1 \leq 2\epsilon \)

2. \( T_{\varphi_{\text{far}}} \subseteq \{ (l, k) : |k - l| > B \} \), where \( B \equiv 30 + 2 \left[ \frac{\log(1/\epsilon)}{\log(2)} \right] \).

3. The bipartite shared state \( \varphi_{\text{block}} \) can be prepared starting from EPR pairs with \( O(N/\epsilon + \log(1/\epsilon)/\epsilon) \) bits of communication.

The proof of Lemma 22 is included in Section F of the Appendix.

We can now bound the difference between the protocol \( \mathcal{C} \) acting on \( \varphi \) versus \( \mathcal{C} \) acting on \( \varphi_{\text{block}} \), following equation 18 as follows:

\[
|\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))| = |\text{Tr}(\mathcal{P}(\varphi - \varphi_{\text{block}}))|
\]

Setting \( N = 2Q \) and recalling from the Theorem statement that \( Q \geq 15 \) by assumption, it follows by Lemma 23, stated below, that:

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

(19)

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

We know from Lemma 22 that there is a quantum communication protocol, call it \( \mathcal{K} \), which prepares the shared state \( \varphi_{\text{block}} \) 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 19), 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 \( \varphi_{\text{block}} \) as an entangled resource, which, by Equation 19 above and the original definition \( \mathcal{C} \equiv \mathcal{R} \circ \mathcal{M} \), is within \( 3\epsilon \) 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. \( \square \)
Lemma 23. For $\phi_{\text{block}}$ as constructed in Lemma 22, and for $N, Q$ as defined in the proof of Theorem 1 we have, $|\text{Tr}(\mathcal{P}(\phi - \phi_{\text{block}}))| \leq 3\varepsilon$ whenever $N \geq 2Q \geq 30$.

Proof. Following Lemma 22, we define $B \equiv 30 + 2 \left\lceil \log\left(\frac{1}{\varepsilon}N\right) \right\rceil$. Now, letting $\phi_{\text{block}}$ and $\phi_{\text{far}}$ be as in Lemma 22, and recalling that $\|\phi - (\phi_{\text{block}} + \phi_{\text{far}})\|_1 \leq 2\varepsilon$, we have:

$$|\text{Tr}(\mathcal{P}(\phi - \phi_{\text{block}}))| \leq |\text{Tr}(\mathcal{P}((\phi_{\text{block}} + \phi_{\text{far}}) - \phi_{\text{block}}))| + 2\varepsilon = |\text{Tr}(\mathcal{P}\phi_{\text{far}})| + 2\varepsilon$$

$$= \sum_{(k,l) \in T_{\phi_{\text{far}}}} |\langle \phi_k | \mathcal{P} | \phi_l \rangle| + 2\varepsilon \leq \sum_{(k,l) \in T_{\phi_{\text{far}}}} |\langle \phi_k | \mathcal{P} | \phi_l \rangle| + 2\varepsilon$$

where the final inequality follows because $T_{\phi_{\text{far}}} \subseteq \{(l, k) : |k - l| > B\}$ by Lemma 22. Recalling that the unitary $\mathcal{P}$ can be implemented using $2Q + 8N + 16$ qubits of communication, and applying equation 17 then gives that:

$$|\text{Tr}(\mathcal{P}(\phi - \phi_{\text{block}}))| - 2\varepsilon \leq \sum_{k,l : |k-l| > B} \min(1, 2^{3/2} \cdot \left(2Q + 8N + 16\right) \cdot 2^{-N \cdot \frac{|k-l|}{2} + 4}) \cdot \left\| \phi_{\text{min}}(k,l) \right\|_2^2$$

$$= 2 \sum_l \|\phi_l\|_2^2 \sum_{k > l + B} \min(1, 2^{3Q + 12N + 242N \cdot \frac{|k-l|}{2} + 4})$$

$$= 2 \sum_{n > B} \min(1, 2^{3Q + 12N + 242N \cdot \frac{1}{2} + 4})$$

$$\leq 2 \cdot 2^{3Q + 12N + 242N \cdot \frac{1}{2} - 8N / 2 + 4} \sum_{k=0}^{\infty} 2^{-N \frac{k}{2}}$$

$$= 2 \cdot 2^{3Q + 12N + 242N \cdot \frac{1}{2} - 8N / 2 + 4} \left(1 + \frac{2^{-\frac{N}{2}}}{1 - 2^{-\frac{N}{2}}} \right)$$

$$\leq 4 \cdot 2^{3Q + 12N + 242N \cdot \frac{1}{2} - 8N / 2 + 4}$$

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

$$|\text{Tr}(\mathcal{P}(\phi - \phi_{\text{block}}))| - 2\varepsilon \leq 4 \cdot 2^{3Q + 12N + 242N \cdot \frac{1}{2} - 8N / 2 + 4}$$

$$\leq 4 \cdot 2^{28 \cdot 14N \cdot \frac{15N - \left\lceil \log(1/\varepsilon) \right\rceil}{N}}$$

$$\leq 2^{30N - \log(1/\varepsilon)}$$

$$\leq \varepsilon$$

So,

$$|\text{Tr}(\mathcal{P}(\phi - \phi_{\text{block}}))| \leq 3\varepsilon$$
Lemma 22 there (Appendix E).

Proof. Note: The terminology used in this proof is defined in the proof of Theorem 1 preceding the use of $S$.

Note, in the pre-processing step in the proof of Theorem 1, and again at a point within the proof of Lemma 22 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, 8]. 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.

F Proof of Lemma 22

Lemma (Restatement of Lemma 22). Consider the density matrix $\varphi \equiv \sum_{k,l} |\varphi_k\rangle \langle \varphi_l|$. For any $\epsilon > 0$, there exist subset-matrices, $\varphi_{\text{block}}, \varphi_{\text{far}}$, of $\varphi$, such that

1. $\| \varphi - (\varphi_{\text{block}} + \varphi_{\text{far}}) \|_1 \leq 2\epsilon$

2. $T_{\varphi_{\text{far}}} \subseteq \{(l,k): |k-l| > B\}$, where $B \equiv 30 + 2 \left[ \frac{\log(1/\epsilon)}{N} \right]$.

3. The bipartite shared state $\varphi_{\text{block}}$ 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 22 there (Appendix E).

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 $2\epsilon$).

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

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

Now, for $k \in [1,\ldots,[1/\epsilon]]$ define

$$S_k \equiv \sum_{i=0}^{\infty} M_i |1/\epsilon| + k.$$ 

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

$$1 = \text{tr} \left[ \sum_{k=1}^{[1/\epsilon]} S_k \right].$$

Choose $k'$ such that $\text{tr}[S_{k'}] \leq 1/[1/\epsilon] \leq \epsilon$. 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 $\varphi_{\text{block}} + \varphi_{\text{far}}$ 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 $\varphi$, would make the proof and statement of Lemma 22 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
\[ D_i \equiv P_{2i-1} \varphi P_{2i-1} + P_{2i} \varphi P_{2i} A_i \equiv P_{2i-1} \varphi P_{2i} + P_{2i} \varphi P_{2i-1} \]  
(20)

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

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

Thus,
\[ \| K_{k'} \|_1 = \| S_{k'} - \sum_{i=0}^{\infty} D_{i,[1/\epsilon]+k'} \|_1 \leq \| S_{k'} \|_1 + \sum_{i=0}^{\infty} D_{i,[1/\epsilon]+k'} \|_1 = 2 \| S_{k'} \|_1 \leq 2 \epsilon \]

We now define a “cut down” version of \( \varphi \) by \( \tilde{\varphi} \equiv \varphi - K_{k'} \). From this definition we have:
\[ \| \varphi - \tilde{\varphi} \|_1 = \| K_{k'} \|_1 \leq 2 \epsilon. \]  
(21)

Further, we define the projectors
\[ \eta_j \equiv \sum_{2((j-1)[1/\epsilon]+k') \leq l < 2(j[1/\epsilon]+k')} P_{ij} \]  
(22)

and define the block diagonal matrix \( \varphi_{\text{block}} \) as:
\[ \varphi_{\text{block}} = \sum_j Q_j \tilde{\varphi} Q_j = \sum_j Q_j (\varphi - K_{k'}) Q_j = \sum_j Q_j \varphi Q_j \]  
(23)

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 \( A_{i,[1/\epsilon]+k'} \) which lie outside of the \( Q_j \). Note that \( \varphi_{\text{block}} \) is a subset-matrix of \( \tilde{\varphi} \) according to Definition 20. Now define \( \varphi_{\text{far}} \) by:
\[ \varphi_{\text{far}} \equiv \tilde{\varphi} - \varphi_{\text{block}} \]  
(24)

Therefore, \( \varphi_{\text{far}} \) is also a subset-matrix of \( \varphi \) according to Definition 20. Furthermore, it follows immediately using Equation 21 that:
\[ \| \varphi - (\varphi_{\text{far}} + \varphi_{\text{block}}) \|_1 = \| \varphi - \tilde{\varphi} \|_1 \leq 2 \epsilon \]  
(25)

**Second Claim:** To establish the second claim in Lemma 22 we now show that \( T_{\varphi_{\text{far}}} \subseteq \{ (l, k) : |k - l| > B \} \) (recall that \( B \equiv 30 + 2 \left\lfloor \frac{\log(1/\epsilon)}{N} \right\rfloor \)). 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:
\[ 2B(2(j-1)[1/\epsilon] + 2k' - 1) < l, \ k \leq 2B(2j[1/\epsilon] + 2k' - 1) \]  
(26)
or \( \exists j \) such that:
4B(j\left[1/e\right] + k') - 3B \leq l \leq 2B(2j\left[1/e\right] + 2k' - 1) \leq k \leq 4B(j\left[1/e\right] + k') - B \quad (27)

In the first case, denoted by Equation 26, we have that the coordinates \((l,k)\) lie within the subset-matrix \(\phi_{\text{block}}\) of \(\phi\), and thus that either \((l,k) \in T_{\phi_{\text{block}}}\) or \((l,k) \notin T_{\phi}\) by definition. In particular, either \((l,k) \in T_{\phi}Q_{\phi}Q_l \subseteq T_{\phi_{\text{block}}}\) as follows by Equation 23 and the definition of \(Q_l\) in Equation 22, or \((l,k) \notin T_{\phi}\). If \((l,k) \in T_{\phi_{\text{block}}}\) then we note that \(T_{\phi_{\text{block}}} \cap T_{\phi_{\text{full}}} = \emptyset\) by definition (Equation 24), and this implies that \((l,k) \notin T_{\phi_{\text{full}}}\). If \((l,k) \notin T_{\phi}\), then \((l,k) \notin T_{\phi_{\text{full}}}\) because \(T_{\phi_{\text{full}}} \subseteq T_{\phi}\).

On the other hand, in the case denoted by Equation 27, we have the coordinates \((l,k)\) lie within the subset-matrix \(K_{\phi}\) of \(\phi\), and thus that either \((l,k) \in T_{K_{\phi}}\) or \((l,k) \notin T_{\phi}\). The reason for this is that we know that, in this case, the coordinates \((l,k)\) are within the subset-matrix \(M_{j[1/e]+k'}\) of \(\phi\). Furthermore, since we have already ruled out the case of Equation 26, we know that \((l,k)\) is not in \(D_{j[1/e]+k'}\), the block diagonal portion of \(M_{j[1/e]+k'}\). Therefore, the coordinates \((l,k)\) must lie in the block-anti-diagonal portion \(A_{j[1/e]+k'} = M_{j[1/e]+k'} - D_{j[1/e]+k'}\) (this can also be determined directly from Equation 27 itself, and the definition of \(A_{j[1/e]+k'}\)).

Since \(K_{\phi} \equiv \sum_{l=0}^{\infty} A_{j[1/e]+k'}\) we know that the coordinates \((l,k)\) lie within the \(K_{\phi}\), or more precisely, either \((l,k) \in T_{K_{\phi}}\), or \((l,k) \notin T_{\phi}\). Just as before, if \((l,k) \notin T_{\phi}\), then \((l,k) \notin T_{\phi_{\text{full}}}\).

On the other hand, in the case denoted by Equation 27, we know that \(T_{K_{\phi}} \cap T_{\phi_{\text{full}}} = \emptyset\) because \(T_{\phi_{\text{full}}} \subseteq T_{\phi}\) by Equation 24, and \(T_{\phi} \cap T_{K_{\phi}} = \emptyset\) as follows from the definition \(\phi \equiv \phi - K_{\phi}\).

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

**Third Claim:** To establish the third claim in Lemma 22, and complete the proof, we will show that \(\phi_{\text{block}}\) is a mixture of states of spread at most \(O(N/e + \log(1/e)/e)\), which means that \(\phi_{\text{block}}\) can be produced from a shared maximally entangled state with at most \(O(N/e + \log(1/e)/e)\) bits of communication.

Recalling the definition of \(\phi_{\text{block}}\) in Equation 23, let us define \(\rho'_j \equiv Q_j\phi Q_j\), so that it is clear that \(\phi_{\text{block}} = \sum_j \rho'_j\). It is also clear that \(\rho'_j\) is not only PSD, but also an un-normalized pure state, because

\[
\rho'_j \equiv Q_j\phi Q_j = Q_j \varphi \langle \varphi | Q_j.
\]

From the definition of \(Q_j\) in Equation 22 we have that:

\[
Q_j \varphi = \sum_{B_b < l \leq B_b} |\varphi_l\rangle,
\]

Where the index limits are

\[
B_a \equiv 2(2(j - 1) \cdot [1/e] + k') - 1) \cdot B
\]

\[
B_b \equiv 2(2(j \cdot [1/e] + k') - 1) \cdot B.
\]

We know from the definition in Equation 16 that \(\{|\varphi_l\rangle\}\) are orthogonal to each other, and that each \(|\varphi_l\rangle\) has Schmidt coefficients bounded by \(2^{-I_N+1} \geq \lambda_l > 2^{-I_N-1}\). Thus, it is immediate that \(\rho'_j\) has spread at most \((B_b - B_a) N + 4 = 2[1/e]BN + 4 = O(N/e + \log(1/e)/e)\), where the last equality follows because \(B = 30 + 2 \left[\log(1/e)/N\right]\). Therefore \(\phi_{\text{block}}\) is a normalized mixture of states with spread at most \(O(N/e + \log(1/e)/e)\).

Consider the normalized version of \(\rho'_j\), which is still a pure state of spread at most \(O(N/e + \log(1/e)/e)\) it is clear that this state has Earthmover distance at most \(O(N/e + \log(1/e)/e)\) 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 $\rho_i'$ from EPR pairs, with only $O(N/e + \log(1/e)/e)$ bits of communication (we note that this line of the proof could also have been established using result from [7, 8], for example). Now the state $q_{\text{block}} \equiv \sum_i \rho_i'$ 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 $q_{\text{block}}$ can be prepared starting from EPR pairs with $O(N/e + \log(1/e)/e)$ bits of communication.

□

G Proof of Lemma 15

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

Define $|\rho\rangle \equiv \sum_{j \in \mathcal{Y}} \sum_{k \in [2^{d_\infty(|\chi\rangle, |\nu\rangle)}+2]} \sqrt{\gamma_{j,k,r}} |j\rangle \otimes |k\rangle \otimes |j\rangle \otimes |k\rangle$, where

$$\rho_{j,k} \equiv \nu_j / 2^{[d_\infty(|\chi\rangle, |\nu\rangle)]+2}.$$  

We now define the intermediate state

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

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 $|\nu\rangle$ as follows:

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

Fact 24. If $|m - l| > d_\infty(|\chi\rangle, |\nu\rangle) + 1$, then $\omega(X_m, Y_l) = 0$

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

We will now specify an iterative, “greedy” procedure to define the Schmidt coefficients $\gamma_{j,k,r}$ 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 $\omega(X_m, Y_l) > 0$ we first note that by Fact 24 that $|m - l| < d_\infty(|\chi\rangle, |\nu\rangle) + 1$. Thus, for each $(i, j) \in X_m \times Y_l$,

$$\chi_i \geq 2^{-(m+1)} \geq 2^{l - d_\infty(|\chi\rangle, |\nu\rangle)-2} \geq 2^{-l/2^{[d_\infty(|\chi\rangle, |\nu\rangle)]+2}} \geq \nu_j / 2^{[d_\infty(|\chi\rangle, |\nu\rangle)]+2} \equiv \rho_{j,k}$$

for all $k \in [2^{d_\infty(|\chi\rangle, |\nu\rangle)]+2]$. One may check that Algorithm 1 defines Schmidt coefficients $\gamma_{j,k,r}$, satisfying
Algorithm 1

1: For all \( i \) set \( \text{temp}_i = X_i \)
2: Set \( i_m = \min\{X_m\} \) for all \( m \)
3: for \( l \in \mathbb{N} \cup \{0\} \) do
4: \( j := \min\{Y_l\} \);
5: \( k = 0 \);
6: Set overflow = 0
7: for \( m \in \mathbb{N} \cup \{0\} \) do
8: \( \text{if } \omega(X_m, Y_l) > 0 \) then
9: \( \text{temp}_\omega = \omega(X_m, Y_l) \)
10: \( \text{while } \text{temp}_\omega > 0 \) do
11: \( \text{if } \sum_{r < \text{overflow}} \gamma_{j,k,r} < \rho_{j,k} \) then
12: \( \text{while } \text{temp}_\omega \geq \rho_{j,k} - \sum_{r < \text{overflow}} \gamma_{j,k,r} \) do
13: \( \text{if } k = 2^{\lceil d_\omega(|x|,|y|) \rceil} + 2 \) then
14: \( j = j + 1 \)
15: \( \text{Set overflow = 0} \)
16: \( \text{Set } k = 0 \)
17: \( \text{if } \text{temp}_{i_m} \leq \rho_{j,k} - \sum_{r < \text{overflow}} \gamma_{j,k,r} \) then
18: \( \text{Set } \gamma_{j,k,\text{overflow}+1} = \text{temp}_{i_m} \)
19: \( \text{Set } \text{temp}_\omega = \text{temp}_\omega - \text{temp}_{i_m} \)
20: \( \text{Set temp}_{i_m} = 0 \)
21: \( \text{Add an edge in the flow graph from } i_m \text{ to } (j, k, \text{overflow} + 1) \)
22: \( \text{Set } i_m = i_m + 1 \)
23: \( \text{Set overflow = overflow + 1} \)
24: \( \text{if } \text{temp}_{i_m} \geq \rho_{j,k} - \sum_{r < \text{overflow}} \gamma_{j,k,r} \) and \( \text{temp}_\omega \geq \rho_{j,k} - \sum_{r < \text{overflow}} \gamma_{j,k,r} \) then
25: \( \text{Set } \gamma_{j,k,\text{overflow}+1} = \rho_{j,k} - \sum_{r < \text{overflow}} \gamma_{j,k,r} \)
26: \( \text{Set } \text{temp}_\omega = \text{temp}_\omega - \gamma_{j,k,\text{overflow}+1} \)
27: \( \text{Set temp}_{i_m} = \text{temp}_{i_m} - \gamma_{j,k,\text{overflow}+1} \)
28: \( \text{Add an edge in the flow graph } G_{X,Z} \text{ from } i_m \text{ to } (j, k, \text{overflow} + 1) \)
29: \( \text{Set } k = k + 1 \)
30: \( \text{Set overflow = 0} \)
Algorithm 1 (continued)

31: \( \textbf{if} \ k = 2^{|d_\omega(|\chi\rangle,\nu\rangle)|+2} \ \textbf{then} \)
32: \quad \text{Set } j = j + 1
33: \quad \text{Set overflow } = 0
34: \quad \text{Set } k = 0
35: \textbf{if } \text{temp}_\omega < \rho_{j,k} - \sum_{r \leq \text{overflow}} \gamma_{j,k,r} \ \textbf{then}
36: \quad \textbf{if } \text{temp}_{i_m} \leq \text{temp}_\omega \ \textbf{then}
37: \quad \quad \text{Set } \gamma_{j,k,\text{overflow}+1} = \text{temp}_{i_m}
38: \quad \quad \text{Set } \text{temp}_\omega = \text{temp}_\omega - \text{temp}_{i_m}
39: \quad \quad \text{Set } \text{temp}_{i_m} = 0
40: \quad \quad \text{Add an edge in the flow graph } G_{X,Z} \text{ from } i_m \text{ to } (j,k,\text{overflow} + 1)
41: \quad \quad \text{Set } i_m = i_m + 1
42: \quad \quad \text{Set overflow } = \text{overflow} + 1
43: \quad \textbf{if } \text{temp}_{i_m} \geq \text{temp}_\omega \ \textbf{then}
44: \quad \quad \text{Set } \gamma_{j,k,\text{overflow}+1} = \text{temp}_\omega
45: \quad \quad \text{Set } \text{temp}_\omega = 0
46: \quad \quad \text{Set } \text{temp}_{i_m} = \text{temp}_{i_m} - \text{temp}_\omega
47: \quad \quad \text{Add an edge in the flow graph } G_{X,Z} \text{ from } i_m \text{ to } (j,k,\text{overflow} + 1)
48: \quad \quad \text{Set overflow } = \text{overflow} + 1
49: \textbf{if } k = 2^{|d_\omega(|\chi\rangle,\nu\rangle)|+2} \ \textbf{then}
50: \quad \text{Set } j = j + 1
51: \quad \text{Set overflow } = 0
52: \quad \text{Set } k = 0
\[
\sum_{j \in Y} \sum_{k \in [2^{\lceil d_{\infty}(\chi,\nu) \rceil}]+2} \gamma_{j,k,r} = \sum_{i \in X} \chi_i = 1,
\]

as well as a Right Index-1 Flow from \(|\chi\rangle\) to \(|\gamma\rangle\), with degree at most \(2^{\lceil d_{\infty}(\chi,\nu) \rceil} + 2 \cdot 2^{\lceil d_{\infty}(\chi,\nu) \rceil} + 2 = 2^{2\lceil d_{\infty}(\chi,\nu) \rceil} + 2\). 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 \(Z \equiv (Y, [2^{\lceil d_{\infty}(\chi,\nu) \rceil}]+2, [2^{\lceil d_{\infty}(\chi,\nu) \rceil}]+2)\).

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,

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

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 \((j, k, r) \in Y \times [2^{\lceil d_{\infty}(\chi,\nu) \rceil}]+2 \times [2^{\lceil d_{\infty}(\chi,\nu) \rceil}]+2\), 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 \(2^{\lceil d_{\infty}(\chi,\nu) \rceil} + 2\).

Finally, recall that,

\[
\sum_{k \in [2^{\lceil d_{\infty}(\chi,\nu) \rceil}]+2} \rho_{j,k} = \sum_{k \in [2^{\lceil d_{\infty}(\chi,\nu) \rceil}]+2} v_j / 2^{\lceil d_{\infty}(\chi,\nu) \rceil} + 2 = v_j
\]

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

AWH was funded by NSF grants CCF-1452616, CCF-1729369, PHY-1818914 and ARO contract W911NF-17-1-0433. MC was supported at MIT by an Akamai Fellowship, and 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.

References

[1] D. Aharonov, A. W. Harrow, Z. Landau, D. Nagaj, M. Szegedy, and U. 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, Oct 2014, arXiv:1410.0951.

[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, arXiv:0912.5537.

[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 A. Winter. On the communication cost of entanglement transformations. Phys. Rev. A, 67:012306, 2003, arXiv:quant-ph/0204092.

[9] R. Jain, J. Radhakrishnan, and P. Sen. Optimal direct sum and privacy trade-off results for quantum and classical communication complexity, 2008, arXiv:0807.1267.

[10] M. Junge and C. Palazuelos. Large violation of bell inequalities with low entanglement. Communications in Mathematical Physics, 306(3):695–746, 2011, arXiv:1007.3043.

[11] D. Leung, B. Toner, and J. Watrous. Coherent state exchange in multi-prover quantum interactive proof systems. Chicago Journal of Theoretical Computer Science, 11:1–18, 2013, arXiv:0804.4118.

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

[13] M. A. Nielsen. Conditions for a class of entanglement transformations. 83:436–439, 1999, arXiv:quant-ph/9811053.
[14] O. Regev. Bell violations through independent bases games. *Quantum Info. Comput.*, 12(1-2):9–20, Jan. 2012, arXiv:1101.0576.

[15] T. Vidick and S. Wehner. More nonlocality with less entanglement. *Phys. Rev. A*, 83:052310, May 2011.