Interaction in Quantum Communication

Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma and David Zuckerman

Abstract

In some scenarios there are ways of conveying information with many fewer, even exponentially fewer, qubits than possible classically [1], [2], [3]. Moreover, some of these methods have a very simple structure— they involve only few message exchanges between the communicating parties. It is therefore natural to ask whether *every* classical protocol may be transformed to a "simpler" quantum protocol—one that has similar efficiency, but uses fewer message exchanges.

We show that for any constant k, there is a problem such that its $k+1$ message classical communication complexity is exponentially smaller than its k message quantum communication complexity. This, in particular, proves a round hierarchy theorem for quantum communication complexity, and implies, via a simple reduction, an $\Omega(N^{1/k})$ lower bound for k message quantum protocols for Set Disjointness for constant k.

Enroute, we prove information-theoretic lemmas, and define a related measure of *correlation*, the *informational distance*, that we believe may be of significance in other contexts as well.

I. Introduction

A recurring theme in quantum information processing has been the idea of exploiting the exponential resources afforded by quantum states to encode information in very non-obvious ways. One representative result of this kind is due to Ambainis, Schulman, Ta-Shma, Vazirani, and Wigderson [2]. They show that two players can deal a random set of $\sqrt{N}$ cards each, from a pack of N cards, by the exchange of $O(\log N)$ quantum bits between them. Another example is given by Raz [3] who shows that a natural geometric promise problem that has an efficient quantum protocol, is hard to solve via classical communication. Both are examples of problems for which exponentially fewer quantum bits are required to accomplish a communication task, as compared to classical bits. A third example is the $O(\sqrt{N} \log N)$ qubit protocol for Set Disjointness due to Buhrman, Cleve, and Wigderson [1], which represents quadratic savings in the communication cost over classical protocols.

The protocols presented by Ambainis *et al.* [2] and Raz [3] share the feature that they require minimal *interaction* between the communicating players. For example, in the protocol of Ambainis *et al.* [2] one player prepares a set of qubits in a certain state and sends half of the qubits across as the message, after which both players measure their qubits to obtain the result. In contrast, the protocol of Buhrman, Cleve and Wigderson [1] for

checking set disjointness (DISJ) requires $\Omega(\sqrt{N})$ messages. This raises a natural question: Can we exploit the features of quantum communication and always reduce interaction while maintaining the same communication cost? In particular, are there efficient quantum protocols for DISJ that require only a few messages?

Kitaev and Watrous [4] show that every efficient quantum interactive proof can be transformed into a protocol with only three messages of similar total length. This suggests that it might be possible to reduce interaction in other protocols as well. In this paper we show that for any constant k, there is a problem such that its $k + 1$ message classical communication complexity is exponentially smaller than its k message quantum communication complexity, thus answering the above question in the negative. This, in particular, proves a round hierarchy theorem for quantum communication complexity, and implies, via a simple reduction, polynomial lower bounds for constant round quantum protocols for Set Disjointness.

Our Separation Results

The role of interaction in *classical* communication is well-studied, especially in the context of the Pointer Jumping function [5], [6], [7], [8], [9]. Our first result is for a subproblem S_k of Pointer Jumping that is singled out in Miltersen *et al.* [10] (see Section V-A for a formal definition of S_k). We show:

Theorem I.1: For any constant k, there is a problem S_{k+1} such that any quantum protocol with only k messages and constant probability of error requires $\Omega(N^{1/(k+1)})$ communication qubits, whereas it can be solved with $k + 1$ messages by a deterministic protocol with $O(\log N)$ bits.

A more precise version of this theorem is given in Section V-D and implies a round hierarchy even when the number of messages k grows as a function of input size N, up to $k = \Theta(\log N / \log \log N)$. Our analysis of S_k follows the same intuition as that behind the result of Miltersen *et al.* [10], but relies on entirely new ideas from quantum information theory. The resulting lower bound is optimal for a constant number of rounds.

Next, we study the Pointer Jumping function itself. Let f_k denote the Pointer Jumping function with path length $k + 1$ on graphs with $2n$ vertices, as defined in Section VI. The input length for the Pointer Jumping function f_k is $N = 2n \log n$, independent of k,

whereas the input length for S_k is exponential in k. The function f_k is thus usually more appropriate for studying the effect of rounds on communication when k grows rapidly as a function of the input length.

We first show an improved upper bound on the classical complexity of Pointer Jumping, further closing the gap between the known classical upper and lower bounds. We then turn into proving a quantum lower bound. We prove:

Theorem I.2: For any constant k, there is a classical deterministic protocol with k message exchanges, that computes f_k with $O(\log n)$ bits of communication, while any $k - 1$ round quantum protocol with constant error for f_k needs $\Omega(n)$ qubits communication.

The lower bound of Theorem I.2 decays exponentially in k, and leads only to separation results for $k = O(\log N)$. We believe it is possible to improve this dependence on k, but leave it as an open problem. Note that in the preliminary version of this paper [11] this decay was even doubly exponential, and the improvement here is obtained by using a quantum version of the Hellinger distance.

Our lower bounds for S_k and Pointer Jumping also have implications for Set Disjointness. The problem of determining the quantum communication complexity of DISJ has inspired much research in the last few years, yet the best known lower bound prior to this work was $\Omega(\log n)$ [2], [12]. We mentioned earlier the protocol of Buhrman *et al.* [1] which solves DISJ with $O(\sqrt{N} \log N)$ qubits and $\Omega(\sqrt{N})$ messages. Buhrman and de Wolf [12] observed (based on a lower bound for random access codes [13], [14]) that any one message quantum protocol for DISJ has linear communication complexity. We describe a simple reduction from Pointer Jumping in a bounded number of rounds to DISJ and prove:

Corollary I.3: For any constant k, the communication complexity of any k-message quantum protocol for Set Disjointness is $\Omega(N^{1/k})$.

A model of quantum communication complexity that has also been studied in the literature is that of communication with prior entanglement (see, e.g., Refs. [15], [12]). In this model, the communicating parties may hold an arbitrary input-independent entangled state in the beginning of a protocol. One can use superdense coding [16] to transmit n classical bits of information using only $\lceil n/2 \rceil$ qubits when entanglement is allowed. The players may also use measurements on EPR-pairs to create a shared classical random key.

While the first idea often decreases the communication complexity by a factor of two, the second sometimes saves $\log n$ bits of communication. It is unknown if shared entanglement may sometimes decrease the communication more than that. Currently no general methods for proving super-logarithmic lower bounds on the quantum communication complexity with prior entanglement and unrestricted interaction are known. Our results all hold in this model as well.

Our interest in the role of interaction in quantum communication also springs from the need to better understand the ways in which we can access and manipulate information encoded in quantum states. We develop information-theoretic techniques that expose some of the limitations of quantum communication. We believe our information-theoretic results are of independent interest.

The paper is organized as follows. In Section II we give some background on classical and quantum information theory. We recommend Preskill's lecture notes [17] or Nielsen and Chuang's book [18] as thorough introductions into the field. In Section III we present new lower bounds on the quantum relative entropy function (Section III-A) and introduce the informational distance (Section III-B). In Section IV we explain the communication complexity model, followed by Section V where we prove our separation results and the reduction to Set Disjointness (Section V-C). In Section VI we give our new upper bound (Section VI-B) and quantum lower bound (Section VI-C) for the pointer-jumping problem.

Subsequent Results

Subsequent to the publication of the preliminary version of this paper [11] several new related results have appeared. First, Razborov proves in Ref. [19] that the quantum communication complexity of the Set Disjointness problem is indeed $\Omega(\sqrt{N})$, no matter how many rounds are allowed. An upper bound of $O(\sqrt{N})$ is given by Aaronson and Ambainis [20]. A result by Jain, Radhakrishnan, and Sen in Ref. [21] shows that the complexity of protocols solving this problem in k rounds is at least $\Omega(n/k^2)$. The same authors show in Ref. [22] that quantum protocols with $k-1$ rounds for the Pointer Jumping function f_k have complexity $\Omega(n/k^4)$, but this result seems to hold only for the case of protocols *without* prior entanglement. The same authors [23] also consider the complexity of quantum protocols for the version of the Pointer Jumping function, in which not only

one bit of the last vertex has to be computed, but its full name. Several papers ([24], [25], [21], [22], [26]) have used the information theoretic techniques developed in the present paper.

In this paper, we improve the dependence of communication complexity lower bounds on the number of rounds, as compared to our results in Ref. [11]. To achieve this, we use a different information-theoretic tool based on the quantum Hellinger distance. The version of our Average Encoding Theorem based on Hellinger distance was independently found by Jain *et al.* [21].

II. Information Theory Background

The quantum mechanical analogue of a random variable is a probability distribution over superpositions, also called a *mixed state*. For the mixed state $X = \{p_i, |\phi_i\rangle\}$, where $|\phi_i\rangle$ has probability p_i, the *density matrix* is defined as $\rho_X = \sum_i p_i |\phi_i\rangle\langle\phi_i|$. Density matrices are Hermitian, positive semi-definite, and have trace 1. I.e., a density matrix has an eigenvector basis, all the eigenvalues are real and between zero and one, and they sum up to one.

A. Trace Norm And Fidelity

The *trace norm* of a matrix A is defined as $\| A \|_{\mathrm{t}} = \mathrm{Tr} \sqrt{A^\dagger A}$, which is the sum of the magnitudes of the singular values of A. Note that if ρ is a density matrix, then it has trace norm one. If ϕ_1, ϕ_2 are pure states then:

$$\| |\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2| \|_{\mathrm{t}} = 2\sqrt{1 - |\langle\phi_1|\phi_2\rangle|^2}.$$

We will need the following consequence of Kraus representation theorem (see for example Preskill's lecture notes [17]):

Lemma II.1: For each Hermitian matrix ρ and each trace-preserving completely positive superoperator T: $\|T(\rho)\|_{\mathrm{t}} \leq \|\rho\|_{\mathrm{t}}$.

A useful alternative to the trace metric as a measure of closeness of density matrices is *fidelity*. Let ρ be a mixed state with support in a Hilbert space $\mathcal{H}$. A *purification* of ρ is any pure state $|\phi\rangle$ in an extended Hilbert space $\mathcal{H} \otimes \mathcal{K}$ such that $\mathrm{Tr}_{\mathcal{K}} |\phi\rangle\langle\phi| = \rho$. Given

two density matrices ρ_1, ρ_2 on the same Hilbert space $\mathcal{H}$, their *fidelity* is defined as

$$F(\rho_1, \rho_2) \;=\; \sup \, |\langle \phi_1 | \phi_2 \rangle|^2 \,,$$

where the supremum is taken over all purifications $|\phi_i\rangle$ of ρ_i in the same Hilbert space. Jozsa [27] gave a simple proof, for the finite dimensional case, of the following remarkable equivalence first established by Uhlmann [28].

Fact II.2 (Jozsa) For any two density matrices ρ_1, ρ_2 on the same finite dimensional space $\mathcal{H}$,

$$F(\rho_1, \rho_2) \;=\; \left[\mathrm{Tr} \left(\sqrt{ \rho_1^{1/2} \, \rho_2 \, \rho_1^{1/2} } \right) \right]^2 \;=\; \| \sqrt{\rho_1} \sqrt{\rho_2} \|_{\mathrm{t}}^2 \,.$$

Using this equivalence, Fuchs and van de Graaf [29] relate fidelity to the trace distance.

Fact II.3 (Fuchs, van de Graaf) For any two mixed states ρ_1, ρ_2,

$$1 - \sqrt{F(\rho_1, \rho_2)} \;\leq\; \frac{1}{2} \, \| \, \rho_1 - \rho_2 \, \|_{\mathrm{t}} \;\leq\; \sqrt{1 - F(\rho_1, \rho_2)}.$$

While the definition of fidelity uses purifications of the mixed states and relates them via the inner product, fidelity can also be characterized via measurements (see Nielsen and Chuang [18]).

Fact II.4: For two probability distributions p, q on finite sample spaces, let $F(p, q) = (\sum_i \sqrt{p_i q_i})^2$ denote their fidelity. Then, for any two mixed states ρ_1, ρ_2,

$$F(\rho_1, \rho_2) \;=\; \min_{\{E_m\}} F(p_m, q_m),$$

where the minimum is over all POVMs $\{E_m\}$, and $p_m = \mathrm{Tr}(\rho_1 E_m), q_m = \mathrm{Tr}(\rho_2 E_m)$ are the probability distributions created by the measurement on the states.

A useful property of the trace distance $\| \, \rho_1 - \rho_2 \, \|_{\mathrm{t}}$ as a measure of distinguishability is that it is a metric, and hence satisfies the triangle inequality. This is not true for fidelity $F(\rho_1, \rho_2)$ or for $1 - F(\rho, \rho_2)$. Fortunately, a variant of fidelity is actually a metric. Denote by

$$h\,(\rho_1, \rho_2) \;=\; \sqrt{ 1 - \sqrt{F(\rho_1, \rho_2)} }$$

the quantum *Hellinger distance*. Clearly $h\,(\rho_1, \rho_2)$ inherits most of the desirable properties of fidelity, like unitary invariance, definability as a maximum over all measurements of the classical Hellinger distance of the resulting distributions, and so on. To see that $h\,(\rho_1, \rho_2)$

is actually a metric one can simply use Fact II.4 to reduce this problem to showing that the classical Hellinger distance is a metric, which is well known.

Analogously to Lemma II.1, due to the monotonicity of fidelity [18], we have:

Lemma II.5: For all density matrices ρ_1, ρ_2 and each trace-preserving completely positive superoperator T: $h\left(T(\rho_1), T(\rho_2)\right) \leq h\left(\rho_1, \rho_2\right)$.

Let us also note the following relation between the Hellinger distance and the trace norm that follows directly from Fact II.3.

Lemma II.6: For any two mixed states ρ_1, ρ_2,

$$h^2\left(\rho_1, \rho_2\right) \ \leq \ \frac{1}{2}\left\|\rho_1 - \rho_2\right\|_{\mathrm{t}} \ \leq \ \sqrt{2} \cdot h\left(\rho_1, \rho_2\right).$$

We will sometimes work with $h^2\left(\cdot, \cdot\right)$ instead of $h\left(\cdot, \cdot\right)$. This is not a metric, but it is true that for all density matrices ρ_1, ρ_2, ρ_3:

$$h^2\left(\rho_1, \rho_2\right) \leq \left(h\left(\rho_1, \rho_3\right) + h\left(\rho_3, \rho_2\right)\right)^2 \leq 2h^2\left(\rho_1, \rho_3\right) + 2h^2\left(\rho_3, \rho_2\right).$$

B. Local Transition Between Bipartite States

Jozsa [27] proved:

Theorem II.7 (Jozsa) Suppose $\left|\phi_1\right\rangle, \left|\phi_2\right\rangle \in \mathcal{H} \otimes \mathcal{K}$ are the purifications of two density matrices ρ_1, ρ_2 in $\mathcal{H}$. Then, there is a local unitary transformation U on $\mathcal{K}$ such that $F(\rho_1, \rho_2) = \left|\left\langle\phi_1\right|\left(I \otimes U\right)\left|\phi_2\right\rangle\right|^2$.

As noticed by Lo and Chau [30] and Mayers [31], Theorem II.7 immediately implies that if two states have close reduced density matrices, than there exists a *local* unitary transformation transforming one state close to the other. Formally,

Lemma II.8: (Local Transition Lemma, based on Refs. [30], [31], [27], [29]) Let ρ_1, ρ_2 be two mixed states with support in a Hilbert space $\mathcal{H}$. Let $\mathcal{K}$ be any Hilbert space of dimension at least $\dim(\mathcal{H})$, and $\left|\phi_i\right\rangle$ any purifications of ρ_i in $\mathcal{H} \otimes \mathcal{K}$.

Then, there is a local unitary transformation U on $\mathcal{K}$ that maps $\left|\phi_2\right\rangle$ to $\left|\phi_2'\right\rangle = I \otimes U \left|\phi_2\right\rangle$ such that

$$h\left(\left|\phi_1\right\rangle\!\left\langle\phi_1\right|, \left|\phi_2'\right\rangle\!\left\langle\phi_2'\right|\right) \ = \ h\left(\rho_1, \rho_2\right).$$

Furthermore,

$$\left\|\left|\phi_1\right\rangle\!\left\langle\phi_1\right| - \left|\phi_2'\right\rangle\!\left\langle\phi_2'\right|\right\|_{\mathrm{t}} \ \leq \ 2\left\|\rho_1 - \rho_2\right\|_{\mathrm{t}}^{\frac{1}{2}}.$$

Proof: (Of Lemma II.8): By Theorem II.7, there is a (local) unitary transformation U on $\mathcal{K}$ such that $(I \otimes U) |\phi_2\rangle = |\phi_2'\rangle$, a state which achieves fidelity: $F(\rho_1, \rho_2) = |\langle \phi_1 | \phi_2' \rangle|^2$. Hence the statement about the Hellinger distance holds.

By Lemma II.6

$$
\begin{aligned}
\| \, |\phi_1\rangle\langle\phi_1| &- |\phi_2'\rangle\langle\phi_2'| \, \|_{\mathrm{t}} \\
&\leq \;\; 2\sqrt{2} \cdot h\left(|\phi_1\rangle\langle\phi_1|, |\phi_2'\rangle\langle\phi_2'|\right) \\
&= \;\; 2\sqrt{2} \cdot h\left(\rho_1, \rho_2\right) \\
&\leq \;\; 2 \cdot \| \, \rho_1 - \rho_2 \, \|_{\mathrm{t}}^{\frac{1}{2}} .
\end{aligned}
$$

∎

C. Entropy, Mutual Information, And Relative Entropy.

$H(\cdot)$ denotes the binary entropy function $H(p) = p \log(\frac{1}{p}) + (1-p) \log(\frac{1}{1-p})$. The *Shannon entropy* $S(X)$ of a classical random variable X on a finite sample space is $\sum_x p_x \log(\frac{1}{p_x})$ where p_x is the probability the random variable X takes value x. The *mutual information* $I(X : Y)$ of a pair of random variables X, Y is defined to be $I(X : Y) = H(X) + H(Y) - H(X, Y)$. For other equivalent definitions, and more background on the subject see, e.g., the book by Cover and Thomas [32].

We use a simple form of Fano's inequality.

Fact II.9 (Fano's inequality) Let X be a uniformly distributed Boolean random variable, and let Y be a Boolean random variable such that $\mathrm{Prob}(X = Y) = p$. Then $I(X : Y) \geq 1 - H(p)$.

The Shannon entropy and the mutual information functions have natural generalizations to the quantum setting. The *von Neumann entropy* $S(\rho)$ of a density matrix ρ is defined as $S(\rho) = -\mathrm{Tr}\,\rho \log \rho = -\sum_i \lambda_i \log \lambda_i$, where $\{\lambda_i\}$ is the multi-set of all the eigenvalues of ρ. Notice that the eigenvalues of a density matrix form a probability distribution. In fact, we can think of the density matrix as a mixed state that takes the i'th eigenvector with probability λ_i. The von Neumann entropy of a density matrix ρ is, thus, the entropy of the classical distribution ρ defines over its eigenstates.

The mutual information $I(X : Y)$ of two disjoint quantum systems X, Y is defined to

be $I(X:Y) = S(X) + S(Y) - S(XY)$, where XY is the density matrix of the system that includes the qubits of both systems. Then

$$
\begin{align}
I(X:YZ) &= I(X:Y) + I(XY:Z) - I(Y:Z), \tag{1}\\
I(X:YZ) &\geq I(X:Y), \tag{2}
\end{align}
$$

Equation (2) is in fact equivalent to the *strong sub-additivity property* of von Neumann entropy.

We need the following slight generalization of Theorem 2 in Cleve *et al.* [15].

Lemma II.10: Let Alice own a state ρ_A of a register A. Assume Alice and Bob communicate and apply local transformations, and at the end register A is measured in the standard basis. Assume Alice sends Bob at most k qubits, and Bob sends Alice arbitrarily many qubits. Further assume all these local transformations do not change the state of register A, if A is in a classical state. Let ρ_{AB} be the final state of A and Bob's private qubits B. Then $I(A:B) \leq 2k$.

Proof: Considering the joint state of register A and Bob's qubits, there cannot be any interference between basis states differing on A. Thus we can assume that ρ_A is measured in the beginning, i.e., that ρ_A is classical. In this case the result directly follows from Theorem 2 in Ref. [15]. ■

Note that in the above lemma Alice and Bob can use Bob's free communication to set up an arbitrarily large amount of entanglement independent of ρ_A.

The *relative* von Neumann entropy of two density matrices, defined by $S(\rho\|\sigma) = \operatorname{Tr}\rho\log\rho - \operatorname{Tr}\rho\log\sigma$. One useful fact to know about the relative entropy function is that $I(A:B) = S(\rho_{AB}\|\rho_A \otimes \rho_B)$. For more properties of this function see Refs. [17], [18].

III. Informational Distance And New Lower Bounds On Relative Entropy

A. New Lower Bounds On Relative Entropy

We now prove that the relative entropy $S(\rho_1\|\rho_2)$ is lower bounded by $\Omega(\|\rho_1 - \rho_2\|_t^2)$ and by $\Omega(h^2(\rho_1, \rho_2))$. We believe these results are of independent interest. A classical

version of the theorem can be found in, e.g., Cover and Thomas' book on Information Theory [32].

Theorem III.1: For all density matrices ρ_1, ρ_2:

$$S(\rho_1\|\rho_2) \ \geq \ \frac{1}{2\ln 2}\,\|\,\rho_1 - \rho_2\,\|_{\mathrm{t}}^2\,.$$

Although this relationship has appeared in the literature [33], it was rediscovered by several authors, including us. Below we give a proof of this theorem for completeness. The earlier version of our paper [11] contained a more complicated proof.

Proof: (Theorem III.1) The proof goes by reduction to the classical case. Consider the classical distributions $\tilde{\rho}_1, \tilde{\rho}_2$ obtained by measuring ρ_1, ρ_2 in the basis diagonalizing their difference $\rho_1 - \rho_2$. It is known [17], [18] that

$$\|\,\tilde{\rho}_1 - \tilde{\rho}_2\,\|_1 \ = \ \|\,\rho_1 - \rho_2\,\|_{\mathrm{t}}\,.$$

Due to Lindblad-Uhlmann monotonicity of relative von Neumann entropy [17], [18],

$$S(\rho_1\|\rho_2) \ \geq \ S(\tilde{\rho}_1\|\tilde{\rho}_2).$$

The classical version of the theorem [32] now gives

$$\begin{aligned}
S(\tilde{\rho}_1\|\tilde{\rho}_2) \ &\geq \ \frac{1}{2\ln 2}\,\|\,\tilde{\rho}_1 - \tilde{\rho}_2\,\|_1^2 \\
&= \ \frac{1}{2\ln 2}\,\|\,\rho_1 - \rho_2\,\|_{\mathrm{t}}^2\,.
\end{aligned}$$

This completes the proof. ∎

Now we show an analogous result for the quantum Hellinger distance.

Theorem III.2: For all density matrices ρ_1, ρ_2:

$$S(\rho_1\|\rho_2) \ \geq \ \frac{2}{\ln 2}\,h^2\,(\rho_1, \rho_2)\,.$$

This theorem has also been shown independently by Jain *et al.* [21].

Proof: We first show that the theorem holds when ρ_1 and ρ_2 are classical distributions, and then generalize this to the quantum case.

In the classical case we first show $S(\rho_1\|\rho_2) \geq -2\log(1 - h^2\,(\rho_1, \rho_2))$. This was shown by Dacunha-Castelle in Ref. [34].

$$
\begin{aligned}
\log(1 - h^2\left(\rho_1, \rho_2\right)) &= \log(\sqrt{F(\rho_1, \rho_2)}) \\
&= \log\left(\sum_i \sqrt{\rho_1(i)\rho_2(i)}\right) \\
&= \log\left(\sum_i \rho_1(i)\frac{\sqrt{\rho_2(i)}}{\sqrt{\rho_1(i)}}\right) \\
&\geq \sum_i \rho_1(i)\log\left(\frac{\sqrt{\rho_2(i)}}{\sqrt{\rho_1(i)}}\right) \\
&= -\frac{1}{2}S(\rho_1\|\rho_2).
\end{aligned}
$$

The first equation is by definition of h, the second by definition of the classical fidelity function, and the inequality is by an application of Jensen's inequality.

Having that, $S(\rho_1\|\rho_2) \geq \frac{2}{\ln 2}h^2\left(\rho_1, \rho_2\right)$ using $-\ln(1-x) \geq x$ for all $0 \leq x \leq 1$ and so the theorem holds in the classical case.

To show the quantum case recall that both $h\left(\cdot, \cdot\right)$ and $S(\cdot\|\cdot)$ can be defined as the maximum over all POVM measurements of the classical versions of these functions on the distributions obtained by the measurements. Fix a POVM $\{E_m\}$ that maximizes $h\left(p, q\right)$ for the distributions p, q obtained from ρ_1, ρ_2. Then $S(\rho_1\|\rho_2) \geq S(p\|q)$ by Lindblad-Uhlmann monotonicity, and $S(p\|q) \geq \frac{2}{\ln 2}h^2\left(p, q\right) = \frac{2}{\ln 2}h^2\left(\rho_1, \rho_2\right)$ because $h\left(p, q\right) = h\left(\rho_1, \rho_2\right)$. The result follows. ∎

B. Informational Distance

From Theorem III.2 follows that for a bipartite state ρ_{AB},

$$
I(A:B) = S(\rho_{AB}\|\rho_A \otimes \rho_B) \geq \frac{2}{\ln 2}h^2\left(\rho_{AB}, \rho_A \otimes \rho_B\right).
$$

Thus the distance between the tensor product state and the "real" (possibly entangled) bipartite state can be bounded in terms of the Hellinger distance. We call the quantity $D(A:B) = h\left(\rho_{AB}, \rho_A \otimes \rho_B\right)$ the "informational distance." $D(A:B)$ measures the amount of correlation between the quantum registers A and B, and can be positive even when the system is classical or not entangled. Later we state some of its properties and use it for proving the quantum communication lower bound on the pointer jumping problem.

The next lemma collects a few immediate properties of informational distance.

Lemma III.3: For all states ρ_{XYZ} the following hold:

1. $D(X:Y) = D(Y:X)$,

2. $0 \leq D(X:Y) \leq 1$,

3. $D(X:Y) \geq h\left(T(\rho_{XY}), T(\rho_X \otimes \rho_Y)\right)$ for all completely positive, trace-preserving superoperators T,

4. $D(XY:Z) \geq D(X:Z)$,

5. $D(X:Y) \leq \sqrt{I(X:Y)}$.

Proof: (1) is true by definition, (2) follows from the definition and the triangle inequality, (3,4) follow from Lemma II.5 and (5) from Theorem III.2. ∎

We now examine the informational distance in the special case where ρ_{QX} is block diagonal, with classical ρ_X. We denote by $\rho_Q^{(x)}$ the density matrix obtained by fixing X to some classical value x and normalizing. $\Pr(x)$ is the probability of $X = x$.

Lemma III.4: For all block diagonal ρ_{QX}, where ρ_X corresponds to a classical distribution,

1. $D^2(Q:X) = \mathbf{E}_x \, h^2\left(\rho_Q^{(x)}, \rho_Q\right)$.

2. Further assume X is Boolean with $\Pr(X=1) = \Pr(X=0) = 1/2$. Let there be a measurement acting on the Q system only, yielding a Boolean random variable Y with $\Pr(X = Y) \geq 1 - \epsilon$ and $\Pr(X \neq Y) \leq \epsilon$. Then $D^2(Q:X) \geq 1/8 - \epsilon/2$.

The first item is true because ρ_{QX} is block-diagonal with respect to X. In the second item, notice that the same measurement applied to $\rho_X \otimes \rho_Q$ yields a distribution with $\Pr(X = Y) = \Pr(X \neq Y) = 1/2$, because Q is independent of X, and X is uniform. Observe that $\| \rho_{XQ} - \rho_X \otimes \rho_Q \|_t \geq \| \rho_{XY} - \rho_X \otimes \rho_Y \|_t \geq 1 - 2\epsilon$ and then apply Lemma II.6. Note that this is a rather crude estimate, since $D(Q:X)$ approaches $1 - 1/\sqrt{2}$ when ϵ goes to zero.

C. The Average Encoding Theorem

A corollary of Theorems III.1,III.2 is the following "Average encoding theorem":

Theorem III.5 (Average encoding theorem) Let $x \mapsto \rho_x$ be a quantum encoding mapping an m bit string $x \in \{0,1\}^m$ into a mixed state with density matrix ρ_x. Let X be distributed over $\{0,1\}^m$, where $x \in \{0,1\}^m$ has probability p_x, let Q be the encoding of X

according to this map, and let $\bar{\rho} = \sum_x p_x \rho_x$. Then,

$$\sum_x p_x \left\| \bar{\rho} - \rho_x \right\|_{\mathrm{t}} \;\; \leq \;\; \left[(2 \ln 2) \, I(Q:X) \right]^{1/2}$$

and

$$\sum_x p_x \; h^2 \left(\bar{\rho}, \rho_x \right) \;\; \leq \;\; \frac{\ln 2}{2} \, I(Q:X).$$

In other words, if an encoding Q is only weakly correlated to a random variable X, then the "average encoding" $\bar{\rho}$ is in expectation (over a random string) a good approximation of any encoded state. Thus, in certain situations, we may dispense with the encoding altogether, and use the single state $\bar{\rho}$ instead. The preliminary version of our paper [11] did not include the second statement. The present stronger version was also observed independently by Jain *et al.* [21].

Proof: (Of Theorem III.5) In the setting of the Average encoding theorem we have a random variable that is distributed over $\{0,1\}^m$, and a quantum encoding $x \mapsto \rho_x$ mapping m bit strings $x \in \{0,1\}^m$ into mixed states with density matrices ρ_x. Let X be the register holding the input x and Q be the register holding the encoding. Let us also define the average encoding $\bar{\rho} = \sum_x p_x \rho_x$.

Then, by Theorem III.1,

$$I(Q:X) \;\; = \;\; S(\rho_{QX} \| \rho_Q \otimes \rho_X) \;\; \geq \;\; \frac{1}{2 \ln 2} \left\| \rho_{QX} - \rho_Q \otimes \rho_X \right\|_{\mathrm{t}}^2$$

The density matrix ρ_X of the X register alone is diagonal and contains the values p_x on the diagonal, the density matrix ρ_Q of the Q register alone is $\bar{\rho}$, and the density matrix $\rho_Q \otimes \rho_X$ is block diagonal and the x'th block is of the form $p_x \bar{\rho}$. Also, the density matrix ρ_{QX} of the whole system is block diagonal, with $p_x \rho_x$ in the x'th block. Thus, $\left\| \rho_{QX} - \rho_Q \otimes \rho_X \right\|_{\mathrm{t}} = \sum_x p_x \left\| \rho_x - \bar{\rho} \right\|_{\mathrm{t}}$, and so $\mathbf{E}_x \left\| \rho_x - \bar{\rho} \right\|_{\mathrm{t}} \leq \sqrt{2 \ln 2} \sqrt{I(Q:X)}$.

The second statement follows analogously using Theorem III.2. ∎

IV. The Communication Complexity Model

In the quantum communication complexity model [35], two parties Alice and Bob hold qubits. When the game starts Alice holds a classical input x and Bob holds y, and so

the initial joint state is simply $|x\rangle \otimes |y\rangle$. Furthermore each player has an arbitrarily large supply of private qubits in some fixed basis state. The two parties then play in turns. Suppose it is Alice's turn to play. Alice can do an arbitrary unitary transformation on her qubits and then send one or more qubits to Bob. Sending qubits does not change the overall superposition, but rather changes the ownership of the qubits, allowing Bob to apply his next unitary transformation on the newly received qubits. Alice may also (partially) measure her qubits during her turn. At the end of the protocol, one player makes a measurement and declares the result of the protocol. In a classical probabilistic protocol the players may only exchange classical messages.

In both the classical and quantum settings we can also define a public coin model. In the classical public coin model the players are also allowed to access a shared source of random bits without any communication cost. The classical public and private coin models are strongly related [36]. Similarly, in the quantum public coin model Alice and Bob initially share an arbitrary number of quantum bits which are in some pure state that is independent of the inputs. This is better known as communication with prior entanglement [15], [12].

The complexity of a quantum (or classical) protocol is the number of qubits (respectively, bits) exchanged between the two players. We say a protocol *computes* a function $f : \mathcal{X} \times \mathcal{Y} \mapsto \{0, 1\}$ with $\epsilon \geq 0$ error if, for any input $x \in \mathcal{X}, y \in \mathcal{Y}$, the probability that the two players compute $f(x, y)$ is at least $1 - \epsilon$. $Q_\epsilon(f)$ (resp. $R_\epsilon(f)$) denotes the complexity of the best quantum (resp. probabilistic) protocol that computes f with at most ϵ error. For a player $P \in \{$Alice, Bob$\}$, $Q_\epsilon^{c,P}(f)$ denotes the complexity of the best quantum protocol that computes f with at most ϵ error with only c messages (called rounds in the literature), where the first message is sent by P. If the name of the player is omitted from the superscript, either player is allowed to start the protocol. We say a protocol $\mathcal{P}$ *computes* f with ϵ error with respect to a distribution μ on $\mathcal{X} \times \mathcal{Y}$, if

$$\text{Prob}_{(x,y)\in\mu,\mathcal{P}}(\mathcal{P}(x, y) = f(x, y)) \geq 1 - \epsilon.$$

$Q_{\mu,\epsilon}^{c,P}(f)$ is the complexity of computing f with at most ϵ error with respect to μ, with only c messages where the first message is sent by player P. We will use the notation $\tilde{Q}$ (rather than Q^*, as in the literature) for communication complexity in the public coin

model. In all the above definitions, we may replace μ with U when μ is the uniform distribution over the inputs.

The following is immediate.

Fact IV.1: For any distribution μ, number of messages c and player P, $\tilde{Q}_{\mu,\epsilon}^{c,P}(f) \leq Q_{\mu,\epsilon}^{c,P}(f) \leq Q_{\epsilon}^{c,P}(f)$.

We put two constraints on protocols in the above definitions:

• We assume that the two players do not modify the qubits holding the classical input during the protocol. This does not affect the aspect of communication we focus on in this paper.

• We demand that the length of the i'th message sent in a protocol is known in advance. This restriction is also implicit in Yao's definition of quantum communication complexity using interacting quantum circuits [35].

To illustrate this, think of a public coin classical protocol in which Alice first looks at a public coin and if the coin is "head" sends in the first round a message of c qubits and in the second round a message of 1 qubit, otherwise she sends one qubit in the first round and c qubits in the second. In such a protocol the number of message bits sent in the first round is not known in advance, and so such a protocol is not allowed in our model.

A k round protocol with communication complexity c in the more general model, in which the restriction above is absent, can be simulated in our model losing a factor of k in the communication complexity. To show this one invokes the principle of safe storage. The principle says that instead of a mixed state depending on measurement results, we may have a superposition over the measurement results and the messages. Note that in such a superposition there may be messages of different lengths (augmented by some blanks). In the worst case, the length of a single message is now c, so the overall communication cost is at most kc, and the number of rounds used is always the worst case number of rounds. In the example above we get a $2c$ communication complexity.

V. The Role Of Interaction In Quantum Communication

In this section, we prove that allowing more interaction between two players in a quantum communication game can substantially reduce the amount of communication required. In Section V-A we define a communication problem and formally state our results (giving

an overview of the proof), then in Section V-B we give the details of the proofs. For the most part, we will concentrate on communication in a constant number of rounds. Section V-C describes the application to the disjointness problem. Section V-D discusses our results in the case where the number of messages grows as a function of the input size.

A. The Communication Problem And Its Complexity

We define a sequence of problems $S_1, S_2, \ldots, S_k, \ldots$ by induction. The problem S_1 is the index function, i.e., Alice has an n-bit string $x \in \mathcal{X}_1 = \{0,1\}^n$, Bob has an index $i \in \mathcal{Y}_1 = [n]$ and the desired output is $S_1(x,i) = x_i$. Suppose we have already defined the function $S_{k-1} : \mathcal{X}_{k-1} \times \mathcal{Y}_{k-1} \to \{0,1\}$. In the problem S_k, Alice has as input her part of n independent instances of S_{k-1}, i.e., $x \in \mathcal{X}_{k-1}^n$, Bob has his share of n independent instances of S_{k-1}, i.e., $y \in \mathcal{Y}_{k-1}^n$, and in addition, there is an extra input $a \in [n]$ which is given to Alice if k is even and to Bob if k is odd. The output we seek is the solution to the a'th instance of S_{k-1}. In other words, $S_k(x_1, \ldots, x_n, a, y_1, \ldots, y_n) = S_{k-1}(x_a, y_a)$.

Note that the size of the input to the problem S_k is $N = \Theta(n^k)$. If we allow k message exchanges for solving the problem, it can be solved by exchanging $\Theta(\log N) = \Theta(k \log n)$ bits: for $k = 1$, Bob sends Alice the index i and Alice then knows the answer; for $k > 1$, the player with the index a sends it to the other player and then they recursively solve for $S_{k-1}(x_a, y_a)$. However, we show that if we allow one less message, then no quantum protocol can compute S_k as efficiently. In fact, no quantum protocol can compute the function as efficiently even if we allow error, and only require small probability of error on average.

Theorem V.1: For all constant $k \geq 1$ and $0 \leq \epsilon < \frac{1}{2}$ we have

$$\tilde{Q}_{U,\epsilon}^k(S_{k+1}) = \Omega\left(N^{1/(k+1)}\right).$$

To prove this theorem we prove a stronger intermediate claim. Let P_1 be Bob, and for $k \geq 2$, let P_k denote the player that holds the index a in an instance of S_k (a indicates which of the n instances of S_{k-1} to solve). Let $\bar{P}_k$ denote the other player. We refer to $\bar{P}_k$ as the "wrong" player to start a protocol for S_k. The stronger claim is that any k message protocol for S_k in which the wrong player starts is exponentially inefficient as compared to the $\log N$ protocol described above.

Lemma V.2: For all constant $k \geq 1$ and $0 \leq \epsilon < \frac{1}{2}$ we have $\tilde{Q}_{U,\epsilon}^{k,\bar{P}_k}(S_k) = \Omega(n) = \Omega\left(N^{1/k}\right).$

Indeed, there is a classical k-message, $O(n)$-bit protocol in which the wrong player starts, so our lower bound is optimal.

Theorem V.1 now follows directly.

Proof: (Of Theorem V.1): It is enough to show the lower bound for the two cases when the protocol starts either with P_{k+1} or with the other player.

Let P_{k+1} be the player to start. Note that if we set a to a fixed value, say 1, then we get an instance of S_k. So $\tilde{Q}_{U,\epsilon}^{k,P_{k+1}}(S_k) \leq \tilde{Q}_{U,\epsilon}^{k,P_{k+1}}(S_{k+1})$. But $P_{k+1} = \bar{P}_k$, so the bound of Lemma V.2 applies.

Let player $\bar{P}_{k+1}$ be the one to start. Then, observe that if we allow one more message (i.e., $k+1$ messages in all), the complexity of the problem only decreases: $\tilde{Q}_{U,\epsilon}^{k+1,\bar{P}_{k+1}}(S_{k+1}) \leq \tilde{Q}_{U,\epsilon}^{k,\bar{P}_{k+1}}(S_{k+1})$. So we again get the bound from Lemma V.2. ∎

We prove Lemma V.2 by induction. First, we show that the index function is hard to solve with one message if the wrong player starts. This essentially follows from the lower bound for random access codes [13], [14]. The only difference is that we seek a lower bound for a protocol that has low error probability *on average* rather than in the worst case, so we need a refinement of the original argument. We give this in the next section.

Lemma V.3: For any $0 \leq \epsilon \leq 1$ we have $\tilde{Q}_{U,\epsilon}^{1,A}(S_1) \geq \frac{1}{2}(1 - H(\epsilon))n.$

Next, we show that if we can solve S_k with k messages with the wrong player starting, then we can also solve S_{k-1} with only $k-1$ messages of smaller total length, again with the wrong player starting, at the cost of a slight increase in the average probability of error.

Lemma V.4: For $k \geq 2$ and $0 \leq \epsilon < \frac{1}{2}$, let $\mathcal{P}$ be any protocol that solves S_k with respect to the uniform distribution U with error ϵ, and k messages starting with $\bar{P}_k$. Let the communication complexity of $\mathcal{P}$ be $\ell = \ell_1 + \bar{\ell}$ with ℓ_1 being the length of the first message sent. Then, $\tilde{Q}_{U,\epsilon'}^{k-1,\bar{P}_{k-1}}(S_{k-1}) \leq \bar{\ell}$, where $\epsilon' = \epsilon + 2(\ell_1/n)^{1/2}$.

We defer the proof of this lemma to a later section, but show how it implies Lemma V.2 above.

Proof: (Of Lemma V.2): We prove the lemma by induction on k. The case $k = 1$ is

handled by Lemma V.3. Suppose the statement holds for $k-1$. We prove by contradiction that it holds for k as well. If $\ell = \tilde{Q}_{U,\epsilon}^{k,\bar{P}_k}(S_k) = o(n)$, then by Lemma V.4 there is a $k-1$ message protocol for S_{k-1} with the wrong player starting, with error $\epsilon' = \epsilon + o(1) < \frac{1}{2}$, and with communication complexity at most $\ell = o(n)$. This contradicts the induction hypothesis. ∎

B. The Key Lemmas

We now prove average case hardness of the index function.

Proof: (Of Lemma V.3): Consider any protocol for S_1 with Alice sending the first (and only) message. Let ϵ_i be the probability of error when the input to Alice is uniformly random but the input to Bob is i. Note that $\epsilon = \sum_i \epsilon_i/n$. Let X denote the random variable containing Alice's input, and let M_B denote the qubits held by Bob after he has received Alice's message, including his part of the shared entangled state. From Properties (1) and (2) of mutual information in Section II-C, and the concavity of binary entropy,

$$I(X : M_B) \geq \sum_i I(X_i : M_B) \geq \sum_i (1 - H(\epsilon_i)) \geq n(1 - H(\epsilon)).$$

The second inequality follows from the fact that Bob has a measurement that predicts X_i with error ϵ_i and Fact II.9 (Fano's inequality). On the other hand, $I(X : M_B)$ is bounded above by twice the number of qubits in the message [15, Theorem 2]. The lemma follows.
∎

Note that for public-coin randomized protocols we do not have the factor of $\frac{1}{2}$, and obtain a lower bound of $n(1 - H(\epsilon))$.

Next, we show how an efficient protocol for S_k gives rise to an efficient protocol for S_{k-1}. The intuition behind the argument is the same as in proofs for classical communication [10], [36]. However, we use entirely new techniques from quantum information theory, as developed in Section III and also get better bounds.

Proof: (Of Lemma V.4): For concreteness, we assume that k is even, so that $\bar{P}_k$ is Bob. Let $\mathcal{P}$ be a protocol that solves S_k with respect to the uniform distribution U with error ϵ, k messages starting with Bob. Let the communication complexity of $\mathcal{P}$ be $\ell = \ell_1 + \bar{\ell}$ with ℓ_1 being the length of the first message sent.

Given the protocol $\mathcal{P}$, we devise a protocol $\mathcal{P}'$ for solving S_{k-1} with respect to the uniform distribution, but with *Alice* starting, and with only $k-1$ messages. The intuition behind the protocol $\mathcal{P}'$ is the following. It first tries to recreate, from some shared prior entanglement, the state after the first message in the run of $\mathcal{P}$ on a specially chosen S_k instance, and then simulates the remaining $k-1$ rounds of communication of the protocol $\mathcal{P}$ on the recreated state. The instance of S_k is such that the solution to that instance coincides with the solution to the given S_{k-1} instance. We thus get a protocol for S_{k-1} with the desired properties. The details follow.

We start by describing the joint pure state that Alice and Bob share in $\mathcal{P}'$ prior to being given the inputs to the problem S_{k-1}. Consider the protocol $\mathcal{P}$ computing S_k. Let M_A, M_B be the private qubits (or "registers") held by Alice and Bob respectively. Let $Y = Y_1 Y_2 \cdots Y_n$ denote the register containing the input to Bob. Consider the state $|\chi\rangle$ of the registers $M_A M_B Y$, after Bob sends the first message in $\mathcal{P}$, when Y is initialized to a uniform superposition over $\mathcal{Y}_k = \mathcal{Y}_{k-1}^n$. The prior entanglement that Alice and Bob share in $\mathcal{P}'$ is then defined as

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} |j\rangle_A |\chi\rangle_{AB} |j\rangle_B \,,$$

where the qubits M_A in $|\chi\rangle$ are given to Alice and M_B, Y to Bob. It simplifies the description of the protocol if Alice and Bob measure the first and the last register, respectively, of the shared state to get a common random index $j \in [n]$. Since these registers will not be modified during the course of the protocol, the behavior of $\mathcal{P}'$ is not affected by this measurement.

We are ready to describe the steps of the protocol $\mathcal{P}'$. Given the inputs x, y to S_{k-1},

1. Alice, who gets the input x, initializes a register X to $|\phi\rangle^{\otimes(j-1)} |x\rangle |\phi\rangle^{\otimes(n-j)} |j\rangle$, where $|\phi\rangle$ is the uniform superposition over $\mathcal{X}_{k-1}$.

Note that the state of the registers $X M_A M_B Y$ is now exactly as after the first message in a run of the protocol $\mathcal{P}$ on an input for S_k where $a = j$, all input registers X_i but for X_j are in uniform superposition over $\mathcal{X}_{k-1}$, $X_j = x$, and all Y_i are in uniform superposition over $\mathcal{Y}_{k-1}$.

2. Bob, who gets the input y, applies a unitary transformation $V_{j,y}$ (to be defined below)

to the registers M_BY. This step is intended to bring the state of the registers M_AM_BY close to $|\chi(y)\rangle$, the state after the first message in a run of the protocol $\mathcal{P}$ on an input for S_k with $X, Y_1, Y_2, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_n$ as above, except that register Y_j is set to y rather than the uniform superposition over $\mathcal{Y}_{k-1}$. Note that on an input as in $|\chi(y)\rangle$, the result of a protocol for S_k is expected to be the same as $S_{k-1}(x, y)$.

3. Alice and Bob now simulate the protocol $\mathcal{P}$ from the second message onwards starting with the registers XM_AM_BY, and declare the result of that procedure as the output of the protocol $\mathcal{P}'$.

The transformation $V_{j,y}$ is defined as follows. Consider the state $|\chi(j, y)\rangle$ of the registers M_AM_BY (analogous to $|\chi\rangle$) obtained by running $\mathcal{P}$ till the first message is sent, when the register Y is initialized to $|\psi\rangle^{\otimes(j-1)} |y\rangle |\psi\rangle^{\otimes(n-j)}$, where $|\psi\rangle$ is the uniform superposition over $\mathcal{Y}_{k-1}$. Let $\rho = \text{Tr}_{M_BY} |\chi\rangle\langle\chi|$, and $\rho_{j,y} = \text{Tr}_{M_BY} |\chi(j, y)\rangle\langle\chi(j, y)|$ be the restriction of the two states to Alice. The transformation $V_{j,y}$ is defined as the local unitary operator on M_BY, given by Theorem II.7, that achieves the fidelity between ρ and $\rho_{j,y}$. This completes the description of $\mathcal{P}'$.

Observe that $\mathcal{P}'$ has $k - 1$ messages starting with Alice, and has complexity $\bar{\ell}$. We now analyze its probability of error, under a uniform distribution on inputs.

Bob's part of the input to S_k in $|\chi\rangle$ and $|\chi(j, y)\rangle$ differ only in the register Y_j: in the first state, this is uniform over $\mathcal{Y}_{k-1}$, whereas in the second state, this is set to y. Thus, the state $|\chi\rangle$ when restricted to Alice is the *average encoding*, over all $y \in \mathcal{Y}_{k-1}$, of the state $|\chi(j, y)\rangle$ restricted to her:

$$\rho \;=\; \frac{1}{|\mathcal{Y}_{k-1}|} \sum_{z \in \mathcal{Y}_{k-1}} \rho_{j,z}.$$

The Average encoding theorem tells us that ρ and $\rho_{j,y}$ are close to each other on average, provided the mutual information $\mu_j = I(Y_j : M_A)$ between Alice's state and Y_j in a run of $\mathcal{P}$ on the uniform distribution on all inputs is small:

$$\frac{1}{|\mathcal{Y}_{k-1}|} \sum_z h^2\left(\rho, \rho_{j,z}\right) \;\leq\; \left(\frac{\ln 2}{2}\right) \mu_j. \tag{3}$$

As in the proof of Lemma V.3, it is not hard to see that if the length ℓ_1 of the first message M is small relative to n, then for a random j, this mutual information is small.

Claim V.5: $\sum_i \mu_i \leq 2\ell_1$. Thus, $\mathbf{E}_j \, \mu_j \leq 2\ell_1/n$.

By Lemma II.8, the transformation $V_{j,y}$ maps $|\chi\rangle$ to a state close to $|\chi(j,y)\rangle$, and by Lemma II.6

$$
\begin{aligned}
&\| \, |V_{j,y}\chi\rangle\langle V_{j,y}\chi| - |\chi(j,y)\rangle\langle\chi(j,y)| \, \|_{\mathrm{t}} \\
&\leq \;\; 2\sqrt{2} \;\; h\left(|V_{j,y}\chi\rangle\langle V_{j,y}\chi| \, , \, |\chi(j,y)\rangle\langle\chi(j,y)|\right) \\
&= \;\; 2\sqrt{2} \;\; h\left(\rho, \rho_{j,y}\right).
\end{aligned}
\tag{4}
$$

For a random $y \in \mathcal{Y}_{k-1}$, and a random $j \in [n]$, then, the average error in approximating the state $|\chi(j,y)\rangle$ is

$$
\begin{aligned}
\mathbf{E}_{j,y} & \| \, |V_{j,y}\chi\rangle\langle V_{j,y}\chi| - |\chi(j,y)\rangle\langle\chi(j,y)| \, \|_{\mathrm{t}} \\
&\leq \;\; 2\sqrt{2} \;\; \mathbf{E}_{j,y} \, h\left(\rho, \rho_{j,y}\right) && \text{From equation (4)} \\
&\leq \;\; 2\sqrt{2} \;\; \mathbf{E}_j \left[\mathbf{E}_y \, h^2\left(\rho, \rho_{j,y}\right)\right]^{1/2} && \text{By Jensen's inequality} \\
&\leq \;\; 2\sqrt{2} \;\; \mathbf{E}_j \left(\frac{\ln 2}{2} \mu_j\right)^{1/2} && \text{From equation (3)} \\
&\leq \;\; 2\sqrt{\ln 2} \;\; \left[\mathbf{E}_j \mu_j\right]^{1/2} && \text{By Jensen's inequality} \\
&\leq \;\; 3\left(\ell_1/n\right)^{1/2}. && \text{From Claim V.5}
\end{aligned}
$$

Running the protocol $\mathcal{P}$ on the input described in step 2 of $\mathcal{P}'$ finds $S_{k-1}(x,y)$ with probability of error at most ϵ on average when x, y are chosen at random. Thus, running the protocol $\mathcal{P}$ on the state resulting from step 2 of the protocol $\mathcal{P}'$ gives us the answer to $S_{k-1}(x,y)$ with average probability of error only slightly higher than ϵ:

$$
\epsilon' \;=\; \epsilon + \frac{1}{2} \, \mathbf{E}_{j,y} \, \| \, |V_{j,y}\chi\rangle\langle V_{j,y}\chi| - |\chi(j,y)\rangle\langle\chi(j,y)| \, \|_{\mathrm{t}} \;\; \leq \;\; \epsilon + 2\left(\ell_1/n\right)^{1/2},
$$

as claimed. ∎

For classical randomized protocols, it is possible to simplify the reduction of S_{k-1} to S_k described above: This is accomplished as follows. Recall that Alice and Bob share public random coins. They use this to sample a (common) message m from the distribution over classical messages in the first round of the protocol $\mathcal{P}$ for S_k, where the inputs are chosen uniformly at random. They also pick a common random index $j \in [n]$. Alice now picks X_i, $i \neq j$ uniformly at random from $\mathcal{X}_{k-1}$, and sets $X_j = x$, and $a = j$. Bob picks $Y_1, \ldots, Y_n$

from the uniform distribution over $\mathcal{Y}_{k-1}^{j-1} \times \{y\} \times \mathcal{Y}_{k-1}^{n-j}$, conditioned on the first message in the protocol $\mathcal{P}$ on such a random input being equal to m. The distance between the joint state so constructed and the joint state in the original protocol differs (in ℓ_1-distance) by at most the distance between Alice's marginal distributions. Alice and Bob now simulate the protocol $\mathcal{P}$ from the second message onwards on the input X, Y. A straightforward analysis using the Average encoding theorem shows that the initial state (consisting of the message and the inputs) constructed above differs from the corresponding state in the protocol $\mathcal{P}$ by only $(2\ell_1/n)^{1/2}$. This simpler argument was noted in Ref. [37] and independently in Ref. [38].

C. The Disjointness Problem

We now investigate the bounded round complexity of the disjointness problem. Here Alice and Bob each receive the incidence vector of a subset of a size n universe. They reject iff the sets are disjoint. It is known [39], [12] that $Q_\epsilon^1(\text{DISJ}) \geq (1 - H(\epsilon))n$ and $\tilde{Q}_\epsilon^1(\text{DISJ}) \geq (1 - H(\epsilon))n/2$. Furthermore $Q_{1/3}^{O(\sqrt{n})}(\text{DISJ}) = O(\sqrt{n} \log n)$ by an application of Grover search [1]. This upper bound was later improved [20] to $O(\sqrt{n})$, although the number of rounds remained $O(\sqrt{n})$. We now prove a lower bound by reduction.

Proof: (Of Corollary I.3): Suppose we are given a k round quantum protocol for the disjointness problem having error $1/3$ and using c qubits. W.l.o.g. we can assume Bob starts the communication, because the problem is symmetrical, and that k is even. We reduce the communication problem S_k from Section V-A to DISJ.

We visualize an instance of S_k as defining a subtree of the n-ary tree with $k+1$ levels and the edges at alternate levels known to Alice and Bob, respectively. The leaves of the tree are labelled by Boolean values known to Alice (since k is even). The only edge at the root connects it to the a'th child, where $a \in [n]$ is the input that specifies which instance of S_{k-1} is to be solved. The subtrees at the second level are defined recursively according to the n instances of S_{k-1}.

There are at most n^k possible paths of length k that could start at the root vertex. With each such path we associate an element in the universe for the disjointness problem. Given the edges originating from each of their levels, Alice and Bob construct an instance of DISJ on a universe of size $N = n^k$. Alice checks for each possible path of length k

whether the path is consistent with her input and whether the paths lead to a leaf which corresponds to the bit 1. In this case she takes the corresponding element of the universe into her subset. Bob similarly constructs his subset. Now, if the two subsets intersect, then the (unique) element in the intersection witnesses a length k path leading to 1-leaf. If the subsets do not intersect, then the length k path from the root leads to a 0-leaf.

We thus obtain a k round protocol for S_k in which Bob starts. By Lemma V.2, the communication c is $\Omega(n)$ for any constant k. Since the input length for the constructed instance of DISJ is $N = n^k$, we get $\tilde{Q}^k_{1/3}(\text{DISJ}) = \Omega(N^{1/k})$ for $k = O(1)$. ■

D. Beyond A Constant Number Of Messages

So far, we have discussed the complexity of solving S_k in the context of protocols with a constant number of messages. In fact, we may derive a meaningful lower bound even when k grows as a function of the parameter n (hence as a function of $N = n^k$, the input length). We may state the result as follows.

Theorem V.6: For all $k = k(n) \geq 1$ and constant $\epsilon < \frac{1}{2}$ we have $\tilde{Q}^{k,\bar{P}_k}_{U,\epsilon}(S_k) = \Omega\left(\frac{n}{k} + k\right)$.

Proof: Let $\ell = \tilde{Q}^{k,\bar{P}_k}_{U,\epsilon}(S_k)$. Then, there is a protocol that achieves this communication complexity with $\ell_1, \ell_2, \ldots, \ell_k$ qubits of communication in the k rounds, respectively. By repeated application of Lemma V.4 there is a quantum protocol that solves S_1 with one message, the wrong player starting, ℓ_k communication qubits and error

$$
\begin{aligned}
\epsilon_1 \quad &= \quad \epsilon + 2 \sum_{i=1}^{k-1} \left(\frac{\ell_i}{n}\right)^{1/2} \\
&\leq \quad \epsilon + 2 \left(\frac{k \sum_{i<k} \ell_i}{n}\right)^{1/2} \qquad \text{By Jensen's inequality} \\
&\leq \quad \epsilon + 2 \left(\frac{k\ell}{n}\right)^{1/2}
\end{aligned}
$$

For a constant $\delta \in (\epsilon, \frac{1}{2})$, if $\ell \leq (\frac{\delta-\epsilon}{2})^2 \frac{n}{k}$ then $\epsilon_1 \leq \delta$ and by Lemma V.3 we have $\ell \geq \ell_k \geq \frac{1-H(\delta)}{2}n$. This implies that $k \leq (\frac{\delta-\epsilon}{2})^2 \cdot 2 \cdot \frac{1}{1-H(\delta)}$. For some δ close enough to ϵ we get $k < 1$. A contradiction. This proves that $\ell \geq \Omega(\frac{n}{k})$. Also, every k round protocol has at least k communication qubits and so $\ell \geq k$. ■

Note that this lower bound of $\Omega(n/k + k)$ also applies to classical randomized protocols.

The above theorem implies a gap in communication complexity between k and $k+1$ message protocols for k up to $\Theta((n/\log n)^{1/2}) = \Theta(\log N/\log\log N)$, and also lower bounds for DISJ for such k.

VI. The Pointer Jumping Function

The pointer jumping function is considered in most results showing a round-hierarchy for classical communication complexity [6], [7], [9], [8]. This problem is a particularly natural candidate for such results.

Definition VI.1 (Pointer Jumping) Let V_A and V_B be disjoint sets of n vertices each. Let $\mathcal{F}_A = \{f_A | f_A : V_A \to V_B\}$, and $\mathcal{F}_B = \{f_B | f_B : V_B \to V_A\}$, and

$$f(v) = f_{f_A, f_B}(v) = \begin{cases} f_A(v) & \text{if } v \in V_A, \\ f_B(v) & \text{if } v \in V_B. \end{cases}$$

Define $f^{(0)}(v) = v$ and $f^{(k)}(v) = f(f^{(k-1)}(v))$.

Then $g_k : \mathcal{F}_A \times \mathcal{F}_B \to (V_A \cup V_B)$ is defined by $g_k(f_A, f_B) = f_{f_A, f_B}^{(k+1)}(v_1)$, where $v_1 \in V_A$ is fixed. The *pointer jumping function* $f_k : \mathcal{F}_A \times \mathcal{F}_B \to \{0,1\}$ is the XOR of all the bits in the output of g_k.

In the corresponding communication problem, Alice is given a function $f_A \in \mathcal{F}_A$, and Bob a function $f_B \in \mathcal{F}_B$, and they are required to compute $f_k(f_A, f_B)$.

A. Previous Work

If Alice starts, f_k has a deterministic k round communication complexity of $k \log n$. If Bob starts, Nisan and Wigderson [7] proved that f_k has a randomized k round communication complexity of $\Omega(\frac{n}{k^2} - k \log n)$. The lower bound can also be improved to $\Omega(\frac{n}{k} + k)$, see Klauck [39]. With techniques similar to the ones in this section it is also possible to show a lower bound of $\frac{(1-2\epsilon)^2 n}{2k^2} - k \log n$ for the randomized k round complexity of f_k when Bob starts. We omit the details.

The lower bounds are not far from the known upper bound. Nisan and Wigderson [7] describe a randomized protocol for computing g_k with complexity $O(\frac{n}{k} \log n + k \log n)$ in the situation where Bob starts and k rounds are allowed. Ponzio *et al.* [9] show that when $k = O(1)$, the deterministic communication complexity of f_k is $O(n)$.

B. A New Upper Bound

We first give a new classical upper bound which combines ideas from Nisan and Wigderson [7] and Ponzio *et al.* [9]. For $n \geq 1$, define $\log^{(1)}(n) = \log n$ and for $k > 1$, define

$$\log^{(k)}(n) = \log(\max\{\log^{(k-1)}(n), 1\}).$$

Furthermore let $\log^*(n) = \min\{k : \log^{(k)}(n) \leq 1\}$.

Theorem VI.1: $R_\epsilon^{k,B}(g_k) \leq O(k \log n + \frac{n}{k} \cdot \log \frac{1}{\epsilon} \cdot (\log^{(\lceil k/2 \rceil)}(n) + \log k))$.

Proof: The claim is trivial for $k = 1$.

For greater k Bob starts and we have the following protocol. At the first round Bob guesses (with public random bits) a set S_0 of δn random vertices from V_B, we specify δ later. For each chosen vertex v Bob communicates the first ℓ_0 bits of $f_B(v)$, we specify ℓ_0 later. Note that the names of the chosen vertices are accessible to Alice without communication, by reading the public random bits. The protocol then proceeds in two stages.

- Denote $v_t = f^{(t-1)}(v_1)$. For each round $i = 1, \ldots, k$ the active player sends v_i. I.e., at the first round Bob sends nothing (as v_1 is known), at the second round Alice sends $v_2 = f(v_1)$, then Bob sends $f(v_2)$ and so on. Also, at each round i Alice checks whether $v_i \in S_0$. Let t be the first round in which this happens. If $t > \frac{k}{2}$ the two players abort the protocol.

- The rounds $t, t+1, \ldots, k$ take a special form. Let us start with round t. Alice knows $v_t \in S_0$ and therefore knows the first ℓ_0 bits of $f_B(v_t)$. Alice defines a set S_1 that contains all elements of V_A with that prefix. I.e., $|S_1| \leq \frac{n}{2^{\ell_0}}$ and $v_{t+1} = f(v_t) \in S_1$. For each $v \in S_1$ Alice sends the first ℓ_1 bits of $f_A(v)$. In general, in the $(t+i)$'th round the active player knows ℓ_i bits of $f(v_{t+i})$. The active player then defines a set S_{i+1} that contains all the elements of his side with that prefix. I.e., $|S_{i+1}| \leq \frac{n}{2^{\ell_i}}$ and $v_{t+i+1} = f(v_{t+i}) \in S_{i+1}$. For each $v \in S_{i+1}$ the active player sends the first ℓ_{i+1} bits of $f(v)$.

We now specify the parameters. First we choose $\delta = \frac{4}{k} \ln \frac{1}{\epsilon}$. W.l.o.g. we can assume the vertices $v_2, v_4, \ldots$ are all distinct, or Alice can easily save two rounds and the players finish on time. For any choice of $\frac{k}{4}$ distinct vertices $v_2, \ldots, v_{k/2}$ the probability, over the choice of S_0, that during the first $\frac{k}{2}$ rounds Alice will not visit S_0 is at most $(1 - \frac{k}{4n})^{\delta n} \leq e^{-\frac{\delta k}{4}} \leq \epsilon$. So assume indeed that $t \leq \frac{k}{2}$.

We now chose $\ell_i = \log^{(\lceil k/2 \rceil - i)} n + 3 \log k$. It follows that for some $i < \frac{k}{2}$ we have $\ell_i \geq \log n$ and $|S_i| = 1$ and the active player who holds v_{t+i} also knows $f(v_{t+i})$, so he can save two rounds and the computation ends on time.

We now count the number of communication bits. We need $k \log n$ bits for communicating v_i, $i = 1, \ldots, k$. Also, we need $\sum_{i=0}^{\lceil k/2 \rceil} |S_i| \ell_i$ bits for communicating the first ℓ_i bits of each element in S_i. Notice, however, that $\ell_i \leq \frac{2^{\ell_{i-1}}}{k^2}$ and so:

$$
\begin{aligned}
\sum_{i=0}^{\lceil k/2 \rceil} |S_i| \ell_i &\leq n[\delta \ell_0 + \sum_{i=1}^{\lceil k/2 \rceil} \frac{\ell_i}{2^{\ell_{i-1}}}] \leq n[\delta \ell_0 + \frac{1}{k^2} \sum_{i=1}^{\lceil k/2 \rceil} 1] \\
&= O(\frac{n}{k} \cdot \log \frac{1}{\epsilon} \cdot (\log^{(\lceil k/2 \rceil)} n + \log k))
\end{aligned}
$$

which completes the proof. ∎

Corollary VI.2: If $k \geq 2 \log^*(n)$ then $R_{1/3}^{k,B}(g_k) \leq O((\frac{n}{k} + k) \log k)$.

C. A Lower Bound On The Quantum Communication Complexity

In this section we prove a lower bound on the quantum communication complexity of the pointer jumping function f_k, for the situation that k rounds are allowed and Bob sends the first message. The proof uses the same ingredients as the proof of the lower bound for the function S_k in Theorem V.1, namely the Average Encoding Theorem and the Local Transition Lemma. We will consider a quantity d_t capturing the information the active player has in round t on vertex $t + 1$ of the path. This quantity will be the informational distance between the active player's qubits and vertex $t + 1$. Our goal will be to bound d_t in terms of d_{t-1} (which is the information gain so far) plus a term related to the average information on pointers in the other player's input (which is low as long as the number of qubits sent is small). This leads to a recursion imposing a lower bound on the communication complexity, since in the end the protocol must have reasonably large information to produce the output, and in the beginning the corresponding information d_0 is 0.

Let Alice be active in the $(t+1)$'th round. The informational distance d_{t+1} measures the distance between the state of, say, Alice's qubits together with the next vertex $F_B(V_{t+1})$ of the path, and the tensor product of the states of Alice's qubits and $F_B(V_{t+1})$. In the product state Alice has no information about $F_B(V_{t+1})$, so if the two states are close

Alice's powers to say something about the vertex are very limited. We will use the triangle inequality to bound d_{t+1} by the sum of three intermediate distances. In the first step we move from the state given by the protocol to a state in which the $(t + 1)$'th vertex is replaced by a uniformly random vertex, independent of previous communications. The penalty we have to pay for that is proportional to d_t which is a bound on the amount of information Bob gained on V_{t+1}. We use the local transition lemma to conceal Bob's ability to detect such a replacement. Once the $(t + 1)$'th vertex is random, we deal with the average information a player (Bob) can get on a random pointer in the other player's input, and this term is small when the number of communicated qubits is small. The last step is similar to the first and reverses the first one's effect, i.e., replaces the "randomized" $(t + 1)$-th vertex by its real value again. We arrive at the desired product state.

Theorem VI.3: $\tilde{Q}_{1/8}^{k,B}(f_k) \geq \frac{n}{2^{O(k)}} - k \log n.$

Note that the lower bound is linear in n for constant k and leads to Theorem I.2. It implies a separation between the k and $k + 1$ round complexity of Pointer Jumping for k upto $\Theta(\log n) = \Theta(\log N)$, where $N = n \log n$ is the input size.

Proof: (of Theorem VI.3) Fix a quantum protocol for f_k with probability of error $\frac{1}{8}$, k rounds, and with Bob starting. Usually a protocol gets some classical f_A and f_B as inputs, but we will investigate what happens if the protocol is started on a superposition over all inputs, in which all inputs have the same amplitude, i.e., on

$$\sum_{f_A \in \mathcal{F}_A, f_B \in \mathcal{F}_B} \frac{1}{n^n} |f_A\rangle |f_B\rangle .$$

Note that $|\mathcal{F}_A| = |\mathcal{F}_B| = n^n$. The superposition over all inputs is measured after the protocol has finished, so that a uniformly random input and the result of the protocol on that input are produced.

We also require that before round t the active player computes and measures the vertex $v_t = f^{(t-1)}(v_1)$, and includes it in the message that is sent to the other player, who stores it in some qubits V_t. Thus, at the first round Bob sends v_0 (which is known in advance) to Alice, at the second round Alice sends $v_2 = F_A(v_1)$ to Bob and so on. This increases the communication by an additive $k \log n$ term. Notice that F_A, F_B are in a uniform superposition over all possible inputs, and so if we don't measure F_A and F_B the register

V_i is also in a uniform superposition for every $i > 1$. The density matrix of the global state of the protocol before the communication of round t is $\rho_{M_{A,t}M_{B,t}F_A F_B}$, where F_A, F_B are the qubits holding the inputs of Alice and Bob and $M_{A,t}$ resp. $M_{B,t}$ are the other qubits in the possession of Alice and Bob before the communication of round t. The state of the latter two systems of qubits may be entangled. In the beginning these qubits are independent of the input. We also denote $\tilde{\rho}_{M_{A,t}M_{B,t}F_A F_B}$ the density matrix of the system in the case where we do not measure any of the V_i.

Let us denote $d_t = D^2(M_{B,t}F_B : F_A(V_t))$ when t is odd, where the register $F_A(V_t)$ has been measured. Notice that at this stage V_t is measured and $F_A(V_t)$ is a subregister of F_A. The quantity d_t is a measure of Bob's information on the value $F_A(v)$ Alice is going to compute. We similarly let $d_t = D^2(M_{A,t}F_A : F_B(V_t))$ when t is even, where the register $F_B(V_t)$ has been measured.

We assume that the communication complexity of the protocol is δn and prove a lower bound $\delta \geq 2^{-O(k)}$. The general strategy of the proof is induction over the rounds, to successively bound $d_1, d_2, \ldots, d_{k+1}$. Bob sends the first message. As Bob has seen no message yet, we have that $I(M_{B,1}F_B : F_A(V_1)) = 0$, and hence $d_1 = 0$. We show that

Lemma VI.4: $d_{t+1} \leq 8d_t + 4\delta$.

We see that $d_{t+1} \leq 9^t\delta$ for all $t \geq 0$. After round k one player, say Alice, announces the result which is supposed to be the parity of $F_B(V_{k+1})$ and included in $M_{A,k+1}$. On the one hand $d_{k+1} = D^2(M_{A,k+1} : F_B(V_{k+1})) \leq 9^k\delta$. On the other hand, by Lemma III.4(2) $D^2(M_{A,k+1} : \bigoplus F_B(V_{k+1})) \geq 1/8 - 1/16 = 1/16$. Together, $\frac{1}{16} \leq 9^k\delta$, so $\delta \geq 2^{-O(k)}$. ■

We now turn to proving Lemma VI.4.

Proof: (Of Lemma VI.4): W.l.o.g. let Alice be active in round $t+1$. Let $M_A = M_{A,t+1}$ and $M_B = M_{B,t+1}$. Before the $t+1$ round $V_{t+1} = F_A(V_t)$ is measured. The resulting state is a probabilistic ensemble over the possibilities to fix $V_1, \ldots, V_{t+1}$, which are then classically distributed. Alice's reduced state is block diagonal with respect to the possible values of the vertices $V_1, \ldots, V_{t+1}$. For any value v of V_{t+1} let $\rho^v_{M_A M_B F_A F_B} = \rho^{V_{t+1}=v}_{M_A M_B F_A F_B}$ denote the pure state with vertex V_{t+1} fixed to v. Our first goal is to bound the amount of information Bob has at this stage about Alice's value V_{t+1}. We define:

$$\gamma_v \stackrel{\text{def}}{=} h^2 \left(\rho^v_{M_B F_B}, \rho_{M_B F_B} \right).$$

I.e., we look at Bob's view before the $t+1$ message, and in particular before Alice sends V_{t+1} to him, and we let γ_v measure how much Bob's view when $V_{t+1} = v$ differs from Bob's average view. We show that these two are typically close to each other, namely:

Lemma VI.5: $\mathbf{E}_v \gamma_v \le d_t$.

Loosely speaking this says that Bob does not know more than d_t units of information about F_A.

The next step is to replace the actual state $\rho^v_{M_A M_B F_A F_B}$ where $V_{t+1} = v$ with the average case $\rho_{M_A M_B F_A F_B R}$ where nothing is known about V_{t+1}. As we saw, typically, Bob can not distinguish between the actual encoding and the average one, so this should not matter much to Bob. We let $\rho^v_{M_A M_B F_A F_B R}$ be a purification of $\rho^v_{M_A M_B F_A F_B}$ where R is some additional space used to purify the random path $V_1, \ldots, V_t$. I.e., $\rho^v_{M_A M_B F_A F_B R}$ reflects a purification of Bob's view, when $V_{t+1} = v$. We let $\rho_{M_A M_B F_A F_B R}$ be a purification of $\rho_{M_A M_B F_A F_B}$ where R is some additional space used to purify the random path $V_1, \ldots, V_{t+1}$. Now, due to Lemma II.8 there is a local unitary transformation U_v acting only on $F_A M_A R$ such that $\sigma^v_{M_A M_B F_A F_B R} \stackrel{\text{def}}{=} U_v \rho_{M_A M_B F_A F_B R} U_v^\dagger$, and $\rho^v_{M_A M_B F_A F_B R}$ are close to each other. $\sigma^v_{M_A M_B F_A F_B R}$ reflects a purification of Bob's average view with Alice locally adding $V_{t+1} = v$ to it . Notice that in $\sigma^v_{M_A M_B F_A F_B R}$, v is arbitrary and in particular can be different than V_{t+1}. By Lemma II.8 for all vertices $v \in V_B$,

$$
\begin{aligned}
h^2 \left(\rho^v_{M_A F_A}, \sigma^v_{M_A F_A} \right) &\le h^2 \left(\rho^v_{M_A F_A F_B(v)}, \sigma^v_{M_A F_A F_B(v)} \right) \\
&\le h^2 \left(\rho^v_{M_A M_B F_A F_B R}, \sigma^v_{M_A M_B F_A F_B R} \right) \\
&= h^2 \left(\rho^v_{M_B F_B}, \rho_{M_B F_B} \right) = \gamma_v, \quad (5)
\end{aligned}
$$

We are interested in the value

$$
d_{t+1} = D^2(M_A F_A : F_B(V_{t+1})) = \mathbf{E}_v \; h^2 \left(\rho^v_{M_A F_A F_B(v)}, \rho^v_{M_A F_A} \otimes \rho_{F_B(v)} \right),
$$

where $F_B(v)$ is measured and the expectation is over the uniform distribution on vertices v. We now study this expression under the average case scenario, i.e., we look at $h^2 \left(\sigma^v_{M_A F_A F_B(v)}, \sigma^v_{M_A F_A} \otimes \rho_{F_B(v)} \right)$. We prove:

Lemma VI.6: For all vertices $v \in V_B$,

$$h^2 \left(\sigma^v_{M_A F_A F_B(v)}, \sigma^v_{M_A F_A} \otimes \rho_{F_B(v)} \right) \leq \tilde{d}_{t+1}(v)$$

where,

$$\tilde{d}_{t+1}(v) \stackrel{\text{def}}{=} h^2 \left(\tilde{\rho}_{M_A F_A F_B(v)}, \tilde{\rho}_{M_A F_A} \otimes \rho_{F_B(v)} \right), \tag{6}$$

where $F_B(v)$ is assumed to have been measured. Recall that in $\tilde{\rho}$ we let $V_1, \ldots, V_{t+1}$ go unmeasured and that v is an arbitrary value not necessarily equal to V_{t+1}. We then prove:

Lemma VI.7: $\mathbf{E}_v \; \tilde{d}_{t+1}(v) \leq 2\delta$.

Assuming the above lemma, we see that for all v:

$$
\begin{aligned}
& h \left(\rho^v_{M_A F_A F_B(v)}, \rho^v_{M_A F_A} \otimes \rho_{F_B(v)} \right) \\
\leq \;& h \left(\rho^v_{M_A F_A F_B(v)}, \sigma^v_{M_A F_A F_B(v)} \right) \\
& + h \left(\sigma^v_{M_A F_A F_B(v)}, \sigma^v_{M_A F_A} \otimes \rho_{F_B(v)} \right) \\
& + h \left(\sigma^v_{M_A F_A} \otimes \rho_{F_B(v)}, \rho^v_{M_A F_A} \otimes \rho_{F_B(v)} \right) \\
\leq \;& 2\sqrt{\gamma_v} + h \left(\sigma^v_{M_A F_A F_B(v)}, \sigma^v_{M_A F_A} \otimes \rho_{F_B(v)} \right) && \text{From equation (5)} \\
\leq \;& 2\sqrt{\gamma_v} + \sqrt{\tilde{d}_{t+1}(v)} && \text{From Lemma (VI.6).}
\end{aligned}
$$

Squaring both sides,

$$
\begin{aligned}
h^2 \left(\rho^v_{M_A F_A F_B(v)}, \rho^v_{M_A F_A} \otimes \rho_{F_B(v)} \right) &\leq \left(2\sqrt{\gamma_v} + \sqrt{\tilde{d}_{t+1}(v)} \right)^2 \\
&\leq 8\gamma_v + 2\tilde{d}_{t+1}(v). \tag{7}
\end{aligned}
$$

I.e., we paid an $8\gamma_v$ penalty, and we switched to the scenario where Bob has no information about V_{t+1}. Now,

$$
\begin{aligned}
D^2(M_A F_A : F_B(V_{t+1})) &= \mathbf{E}_v \; h^2 \left(\rho^v_{M_A F_A F_B(v)}, \rho^v_{M_A F_A} \otimes \rho_{F_B(v)} \right) \\
&\leq \mathbf{E}_v \left[8\gamma_v + 2\tilde{d}_{t+1}(v) \right] && \text{By equation (7)} \\
&\leq 8d_t + 4\delta && \text{By Lemma VI.7.}
\end{aligned}
$$

This completes the proof of Lemma VI.4. ∎

We finish the proof of Theorem VI.3 by proving the remaining Lemmas.

Proof: (Of Lemma VI.5): By definition $\mathbf{E}_v \gamma_v$ is $\mathbf{E}_u h^2 \left(\rho_{M_B F_B F_A(u)}^{V_t=u}, \rho_{M_B F_B}^{V_t=u} \otimes \rho_{F_A(u)} \right) = D^2(M_B F_B : F_A(V_t))$. Now, $D^2(M_{B,t+1} F_B : F_A(V_t)) \leq D^2(M_{B,t} F_B : F_A(V_t)) = d_t$ because Bob sends the t'th message, and this only decreases the informational distance. ■

Proof: (Of Lemma VI.6):

$$
\begin{aligned}
& h^2 \left(\sigma_{M_A F_A F_B(v)}^v, \sigma_{M_A F_A}^v \otimes \rho_{F_B(v)} \right) \\
\leq\; & h^2 \left(\sigma_{M_A F_A R F_B(v)}^v, \sigma_{M_A F_A R}^v \otimes \rho_{F_B(v)} \right) \\
=\; & h^2 \left(\rho_{M_A F_A R F_B(v)}, \rho_{M_A F_A R} \otimes \rho_{F_B(v)} \right) \qquad \text{By unitarity} \\
=\; & h^2 \left(\tilde{\rho}_{M_A F_A F_B(v)}, \tilde{\rho}_{M_A F_A} \otimes \rho_{F_B(v)} \right) \qquad\qquad (8) \\
=\; & \tilde{d}_{t+1}(v) \qquad\qquad\qquad\qquad\qquad \text{By definition (6).}
\end{aligned}
$$

For equation (8), notice that R holds the path $V_1, \ldots, V_{t+1}$, which is determined by $M_A F_A$. We can apply a unitary transformation that "erases" this. We then get a pure state that is ρ with $V_1, \ldots, V_{t+1}$ unmeasured, i.e., what we called $\tilde{\rho}$ ■

Proof: (Of Lemma VI.7): We first bound the information Alice has on Bob's input. For all t, $I(M_{A,t} F_A : F_B)$ is bounded above by twice the number of qubits in the messages so far due to Lemma II.10, assuming that F_B is measured, i.e., $I(M_{A,t} F_A : F_B) \leq 2\delta n$. Thus considering the situation that F_B is distributed uniformly instead of being in the uniform superposition we get $\mathbf{E}_v I(M_A F_A : F_B(v)) \leq 2\delta$ (where v is uniformly random), using Equation (1) and that the $F_B(v)$ are mutually independent. Now,

$$
\begin{aligned}
\mathbf{E}_v \tilde{d}_{t+1}(v) &= \mathbf{E}_v h^2 \left(\tilde{\rho}_{M_A F_A F_B(v)}, \tilde{\rho}_{M_A F_A} \otimes \rho_{F_B(v)} \right) \\
&= \mathbf{E}_v D^2(M_A F_A : F_B(v)),
\end{aligned}
$$

where $F_A M_A M_B F_B$ are as in the protocol without measurements. Also $I(M_A F_A : F_B(v))$ is invariant if $F_B(i)$ is in superposition or measured for $i \neq v$. So,

$$
\begin{aligned}
\mathbf{E}_v D^2(M_A F_A : F_B(v)) &\leq \mathbf{E}_v I(M_A F_A : F_B(v)) \qquad \text{By Lemma III.3} \\
&= 2\delta.
\end{aligned}
$$

■

Acknowledgements

References

[1] H. Buhrman, R. Cleve, and A. Wigderson, "Quantum vs. classical communication and computation," in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, 1998, pp. 63–68.

[2] A. Ambainis, L. J. Schulman, A. Ta-Shma, U. Vazirani, and A. Wigderson, "The quantum communication complexity of sampling," *SIAM Journal on Computing*, vol. 32, no. 6, pp. 1570–1585, 2003.

[3] R. Raz, "Exponential separation of quantum and classical communication complexity," in *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, 1999, pp. 358–367.

[4] A. Kitaev and J. Watrous, "Parallelization, amplification, and exponential time simulation of quantum interactive proof systems," in *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, 2000, pp. 608–617.

[5] C.H. Papadimitriou and M. Sipser, "Communication complexity," in *Proceedings of the 14th Annual ACM Symposium on Theory of Computing*, 1982, pp. 196–200.

[6] P. Duris, Z. Galil, and G. Schnitger, "Lower bounds on communication complexity," *Information and Computation*, vol. 73(1), pp. 1–22, 1987.

[7] N. Nisan and A. Wigderson, "Rounds in communication complexity revisited," *SIAM Journal on Computing*, vol. 22(1), pp. 211–219, 1993.

[8] H. Klauck, "Lower bounds for computation with limited nondeterminism," in *Proceedings of the 13th Annual IEEE Conference on Computational Complexity*, 1998, pp. 141–153.

[9] S.J. Ponzio, J. Radhakrishnan, and S. Venkatesh, "The communication complexity of pointer chasing, applications of entropy and sampling," in *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, 1999, pp. 602–611.

[10] P.B. Miltersen, N. Nisan, S. Safra, and A. Wigderson, "On data structures and asymmetric communication complexity," in *Proceedings of the 27th Annual ACM Symposium on Theory of Computing*, 1995, pp. 103–111.

[11] H. Klauck, A. Nayak, A. Ta-Shma, and D. Zuckerman, "Interaction in quantum communication and the complexity of set disjointness," in *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, 2001, pp. 124–133.

[12] H. Buhrman and R. de Wolf, "Communication complexity lower bounds by polynomials," in *Proceedings of the 16th Annual IEEE Conference on Computational Complexity*, 2001.

[13] A. Nayak, "Optimal lower bounds for quantum automata and random access codes," in *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, 1999, pp. 369–376.

[14] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma, and Umesh Vazirani, "Dense quantum coding and quantum finite automata," *Journal of the ACM*, vol. 49, no. 4, pp. 1–16, July 2002.

[15] R. Cleve, W. van Dam, M. Nielsen, and A. Tapp, "Quantum entanglement and the communication complexity of the inner product function," in *QCQS: NASA International Conference on Quantum Computing and Quantum Communications, QCQS*. 1998, LNCS.

[16] C.H. Bennett and S.J. Wiesner, "Communication via one- and two-particle operators on einstein- podolsky-rosen states," *Physical review letters*, vol. 69, pp. 2881–2884, 1992.

[17] J. Preskill, "Lecture notes.," http://www.theory.caltech.edu/people/preskill/ph229/., 1998.

[18] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.

[19] A.A. Razborov, "Quantum communication complexity of symmetric predicates," *Izvestiya of the Russian Academy of Science, Mathematics*, vol. 67, pp. 145–159, 2003, see also quant-ph/0204025.

[20] Scott Aaronson and Andris Ambainis, "Quantum search of spatial regions," *Theory of Computing*, vol. 1, pp. 47–79, 2005.

[21] R. Jain, J. Radhakrishnan, and P. Sen, "A lower bound for bounded round quantum communication complexity of set disjointness," in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003, pp. 220–229, see also quant-ph/0303138.

[22] R. Jain, J. Radhakrishnan, and P. Sen, "The quantum communication complexity of the pointer chasing problem: the bit version," in *Proceedings of the 22nd Conference on Foundations of Software Technology and Theoretical Computer Science*, 2002, pp. 218–229.

[23] R. Jain, J. Radhakrishnan, and P. Sen, "Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states," in *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, 2002, pp. 429–438.

[24] S. Venkatesh and P. Sen, "Lower bounds in the quantum cell probe model," in *Proceedings of the 28th International Colloquium on Automata, Languages, and Programming*, 2001, pp. 358–369, see also cs.CC/030903.

[25] H. Klauck, "Quantum and approximate privacy," *Theory of Computing Systems*, vol. 37, pp. 221–246, 2004.

[26] Amit Chakrabarti and Oded Regev, "An optimal randomised cell probe lower bound for approximate nearest neighbour searching," in *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, 2004.

[27] R. Jozsa, "Fidelity for mixed quantum states.," *Journal of Modern Optics*, vol. 41(12), pp. 2315–2323, 1994.

[28] A. Uhlmann, "The 'transition probability' in the state space of a ∗-algebra.," *Reports on Mathematical Physics*, vol. 9, pp. 273–279, 1976.

[29] C.A. Fuchs and J. van de Graaf, "Cryptographic distinguishability measures for quantum-mechanical states.," *IEEE Transactions on Information Theory*, vol. 45(4), pp. 1216–1227, 1999.

[30] H. Lo and H. Chau, "Why quantum bit commitment and ideal quantum coin tossing are impossible.," *Physica D*, vol. 120, pp. 177–187, 1998, see also quant-ph/9711065.

[31] D. Mayers, "Unconditionally secure quantum bit commitment is impossible," *Physical review letters*, vol. 78, pp. 3414–3417, 1997.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications. John Wiley & Sons, New York, NY, USA, 1991.

[33] Masanori Ohya and Denes Petz, *Quantum Entropy and its Use*, Texts and Monographs in Physics. Springer-Verlag, Heidelberg, 1993, Second edition, 2004.

[34] D. Dacunha-Castelle, "Vitesse de convergence pour certains problemes statistiques," in *Ecole d'Ete de Probabilites de Saint-Flour VII-1977, Lecture Notes in Mathematics 678*, 1978, pp. 1–172.

[35] A.C.-C. Yao, "Quantum circuit complexity," in *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, 1993, pp. 352–361.

[36] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, Cambridge, 1997.

[37] Elitza Maneva, "Interactive communication on noisy channels," B.S. Thesis, California Institute of Technology, Pasadena, CA, USA, 2001.

[38] Pranab Sen, "Lower bounds for predecessor searching in the cell probe model," in *Proceedings of the 18th Annual IEEE Conference on Computational Complexity*, 2003, pp. 73–83.

[39] H. Klauck, "On quantum and probabilistic communication: Las vegas and one-way protocols," in *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, 2000, pp. 644–651.