Interaction in Quantum Communication and the Complexity of Set Disjointness

Hartmut Klauck [*]
CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands
klauck@cwi.nl

Ashwin Nayak [†]
Computer Science Dept.
Caltech, MC 256-80
Pasadena, CA 91125
nayak@cs.caltech.edu

Amnon Ta-Shma [‡]
Dept. of Computer Science
Tel-Aviv University
Israel 69978
amnon@post.tau.ac.il

David Zuckerman [§]
Dept. of Computer Science
University of Texas
Austin, TX 78712
diz@cs.utexas.edu

ABSTRACT

One of the most intriguing facts about communication using quantum states is that these states cannot be used to transmit more classical bits than the number of qubits used, yet in some scenarios there are ways of conveying information with exponentially fewer qubits than possible classically [3, 26]. Moreover, these methods have a very simple structure—they involve only few message exchanges between the communicating parties.

We consider the question as to 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, 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, an $\Omega(N^{1/k})$ lower bound for k message protocols for Set Disjointness for constant k.

Our result builds on two primitives, *local transitions in bipartite states* (based on previous work) and *average encoding* which may be of significance in other contexts as well.

1. 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 [3]. They show that it is possible to deal a random set of $\sqrt{N}$ cards each from a set of N by the exchange of $O(\log N)$ quantum bits between two players. Another example is given by Raz [26], 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.

The protocols presented by [3, 26] also share the feature that they require minimal *interaction* between the communicating players. For example, in the protocol of [3], one player prepares a set of qubits in a certain state and sends half of them across as the message, after which both players measure their qubits to obtain the result. In contrast, efficient quantum protocols for computing total functions such as checking Set Disjointness (DISJ) seem to require much more interaction: Buhrman, Cleve, and Wigderson [6] give an $O(\sqrt{N}\log N)$ qubit protocol for DISJ that has $O(\sqrt{N})$ message exchanges. This represents quadratic savings in communication cost, but also an *unbounded* increase in the number of messages exchanged (from one message to $\sqrt{N}$),

[*] Supported by the EU 5th framework program QAIP IST-1999-11234 and by NWO grant 612.055.001.

[†] Supported by Charles Lee Powell Foundation, NSF grant CCR 98761722, and Institute for Quantum Information. This work was done while the author was at University of California, Berkeley, supported by a JSEP grant and NSF grant CCR 9800024, and later at DIMACS Center and AT&T Labs, supported by NSF grants STC 91-19999, CCR 99-06105 and EIA 00-80234.

[‡] Most of this work was done while the author was at the University of California at Berkeley, and supported in part by a David and Lucile Packard Fellowship for Science and Engineering and NSF NYI Grant CCR-9457799.

[§] This work was done while the author was on leave at the University of California at Berkeley. Supported in part by a David and Lucile Packard Fellowship for Science and Engineering, NSF Grant CCR-9912428, NSF NYI Grant CCR-9457799, and an Alfred P. Sloan Research Fellowship.

STOC'01, July 6–8, 2001, Hersonissos, Crete, Greece.
Copyright 2001 ACM 1-58113-349-9/01/0007 ...$5.00.

as compared to classical protocols. Are there simpler protocols for DISJ with similar efficiency? Can we exploit the features of quantum communication and always reduce interaction while maintaining the same communication cost? In other words, do all efficient quantum protocols have the simple structure shared by those of [3, 26]?

In this paper, we study the effect of interaction on the quantum communication complexity of problems. We show that for any constant k, allowing even *one* more message may lead to an exponential decrease in the communication complexity of a problem, thus answering the above question in the negative. More formally,

Theorem 1.1. *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 4.5 and implies a round hierarchy even when the number of messages k grows as a function of input size.

The role of interaction in classical communication is well-studied, especially in the context of the Pointer Jumping function [23, 10, 22, 13, 24]. In fact, the problem S_k in Theorem 1.1 is the subproblem of Pointer Jumping singled out in [19] (see Section 4.1 for a formal definition of S_k). Our analysis follows the same intuition as that behind the result of [19] (also explained in [16]), 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 4.6.

Theorem 1.2. *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.*

We also show an improved upper bound on the classical complexity of Pointer Jumping, further closing the gap between upper and lower bounds.

The input length for the Pointer Jumping function f_k is $N = 2n \log n$, independent of k unlike in the subproblem S_k, where the input length is exponential in k. The function f_k is thus usually more appropriate to study the effect of rounds on communication when k grows rapidly as a function of the input length. The lower bound of Theorem 1.2 however decays doubly exponentially in k, and leads to separation results for $k = O(\log \log N)$. We believe it is possible to improve this dependence on k, but leave it as an open problem.

In the context of quantum communication, it was observed by Buhrman and de Wolf [7] (based on a lower bound of Nayak [20]) that any one message quantum protocol for DISJ has linear communication complexity. Thus, allowing more interaction leads to a quadratic improvement in communication cost. The lower bound of [20] immediately implies a much stronger separation: it shows that the two message complexity of a problem may be exponentially smaller than its one message complexity (see also [14], which independently raised the question of the role of rounds in quantum communication). Our results subsume all these previous results. In [14] it is also shown that any Las Vegas quantum protocol achieving a non-constant speedup against deterministic one-way communication for a total function uses more than one round of communication.

We describe a simple reduction from Pointer Jumping in a bounded number of rounds to DISJ. Our results above thus imply new lower bounds for the problem.

Corollary 1.3. *For any constant k, the communication complexity of any k-message quantum protocol for Set Disjointness is $\Omega(N^{1/k})$.*

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 is $\Omega(\log n)$ [3, 7]. Our result provides new insight into the complexity of the problem.

A model of quantum communication complexity that has also been studied in the literature is that of communication with prior entanglement (see, e.g., [8, 7]). In this model, the communicating parties may hold an arbitrary input-independent entangled state in the beginning. One can use superdense coding [4] 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 superlogarithmic 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. More specifically, we present the following new primitive in quantum encoding.

Theorem 1.4 (Average encoding theorem). *Let $x \mapsto \sigma_x$ be a quantum encoding mapping m bit strings $x \in \{0,1\}^m$ into mixed states σ_x. Let X be distributed uniformly over $\{0,1\}^m$, let Q be the encoding of X according to this map, and let $\sigma = \frac{1}{2^m}\sum_x \sigma_x$. Then,*

$$\frac{1}{2^m}\sum_x \| \sigma - \sigma_x \|_{\mathrm{t}} \quad \leq \quad 4\sqrt{I(Q:X)}.$$

In other words, if an encoding Q is only weakly correlated to a random variable X, then the "average encoding" σ (corresponding to a random string) is on average a good approximation of any encoded state. Thus, in certain situations, we may dispense with the encoding altogether, and use the single state σ instead.

Actually we are able to give a more general theorem which implies the average encoding theorem. Let $S(\rho||\sigma)$ denote the relative von Neumann entropy between density matrices ρ and σ.

Theorem 1.5. *For all density matrices ρ, σ:*

$$S(\rho||\sigma) \quad \geq \quad \frac{1}{2\ln 2}\| \rho - \sigma \|_{\mathrm{t}}^2 .$$

Since $I(Q:X) = S(\sigma_{QX} \| \sigma_Q \otimes \sigma_X)$ we get the average encoding theorem as a special case. This more general theorem seems to be of independent interest. A classical version of the theorem can be found in, e.g., [9].

We also use another primitive derived from the work of Lo and Chau [17] and Mayers [18] which combines results of Jozsa [12], and Fuchs and van de Graaf [11]. Consider two bi-partite pure states such that one party sharing the states cannot locally distinguish between the two states with significant probability. Then the other party can locally transform any of the states to a state that is close to the other.

Theorem 1.6 (Local transition theorem). *(based on [17, 18, 12, 11]) Let ρ_1, ρ_2 be two mixed states with support in a Hilbert space $\mathcal{H}$, $\mathcal{K}$ any Hilbert space of dimension at least $\dim(\mathcal{H})$, and $|\phi_i\rangle$ any purifications of ρ_i in $\mathcal{H} \otimes \mathcal{K}$. Then, there is a local unitary transformation U on $\mathcal{K}$ that maps $|\phi_2\rangle$ to $|\phi_2'\rangle = I \otimes U\, |\phi_2\rangle$ such that*

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

These primitives may be of significance in other applications as well, especially in a cryptographic context. In fact, the idea of local transitions has very recently been used by Ambainis [2] to prove lower bounds for bias in quantum coin-flipping protocols.

2. PRELIMINARIES

In this section we first describe the communication model we study. Our lower bound results rely heavily on quantum information theory. The necessary background is provided in Section 2.2, along with the associated notation. See also [25] for a thorough introduction into the field.

2.1 The communication complexity model

In the quantum communication complexity model [28], Alice and Bob hold qubits. When the game starts Alice holds a superposition $|x\rangle$ and Bob holds $|y\rangle$ (representing the input to the two players), and so the initial joint state is simply $|x\rangle \otimes |y\rangle$. 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. At the end of the protocol, one player makes a measurement and declares that as the result of the protocol.

In general, each player may also (partially) measure her qubits during her turn. However, we assume (by invoking the principle of safe storage [5]) that all such measurements are postponed to the end. We also assume that the two players do not modify the qubits holding the input superposition during the protocol. Neither of these affects the aspect of communication we focus on in this paper.

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 $P \in \{\text{Alice, Bob}\}$, $Q_\epsilon^{c,P}(f)$ 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

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

$Q_{\mu,\epsilon}^{c,P}(f)$ 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. The following is immediate.

Fact 2.1. *For any distribution μ, number of messages c and player P, $Q_{\mu,\epsilon}^{c,P}(f) \leq Q_\epsilon^{c,P}(f)$.*

2.2 Information theory background

Measures of distinguishability

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 $\rho_X = \sum_i p_i \, |\phi_i\rangle\langle\phi_i|$.

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

$$\| \, |\phi_1\rangle\langle\phi_1| - |\phi_2\rangle\langle\phi_2| \, \|_{\mathrm{t}} \quad = \quad 2\sqrt{1 - |\langle\phi_1 \, | \, \phi_2\rangle|^2}. \qquad (1)$$

A fundamental theorem about distinguishing mixed states is the following.

Theorem 2.2. *Let ρ_1, ρ_2 be two density matrices on the same space $\mathcal{H}$. Then for any measurement $\mathcal{O}$,*

$$\left\| \, \rho_1^{\mathcal{O}} - \rho_2^{\mathcal{O}} \, \right\|_1 \quad \leq \quad \| \, \rho_1 - \rho_2 \, \|_{\mathrm{t}} \, ,$$

where $\rho^{\mathcal{O}}$ denotes the classical distribution on outcomes resulting from the measurement of ρ, and $\| \cdot \|_1$ is the ℓ_1 norm.

See [1] for more details.

A useful alternative to the trace metric as a measure of closeness of density matrices is *fidelity*, which is defined in terms of the pure states that can give rise to those density matrices. A *purification* of a mixed state ρ with support in a Hilbert space $\mathcal{H}$ 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

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

where the supremum is taken over all purifications $|\phi_i\rangle$ of ρ_i in the same Hilbert space [12].

Jozsa [12] gave a simple proof, for the finite dimensional case, of the following remarkable equivalence first established by Uhlmann [27].

Theorem 2.3 (Jozsa). *For any two density matrices ρ_1, ρ_2 on the same finite dimensional space $\mathcal{H}$,*

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

Using this equivalence, Fuchs and van de Graaf [11] relate fidelity to the trace distance.

THEOREM 2.4 (FUCHS, VAN DE GRAAF). *For any two mixed states ρ_1, ρ_2,*

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

Entropy and mutual information

The *Shannon entropy* $S(X)$ of a classical random variable X and *mutual information* $I(X : Y)$ of a pair of random variables X, Y are defined as usual (see, e.g., [9]). $H(\cdot)$ denotes the binary entropy function. We use a simple form of Fano's inequality.

FACT 2.5 (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 $I(X : Y) \geq 1 - H(p)$.*

We also need the following bound.

FACT 2.6. $H(\frac{1}{2} + \delta) \;\; \leq \;\; 1 - \delta^2$, *for $\delta \in [-\frac{1}{2}, \frac{1}{2}]$.*

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. We also consider *relative* von Neumann entropy of two density matrices, defined by $S(\rho || \sigma) = \mathrm{Tr}\, \rho \log \rho - \mathrm{Tr}\, \rho \log \sigma$. For properties of these two functions see [25].

We define the "mutual information" $I(X : Y)$ of two disjoint quantum systems X, Y as $I(X : Y) = S(X) + S(Y) - S(XY)$, where XY is density matrix of the system that includes the qubits of both systems. Then

$$\begin{aligned} I(X : YZ) &= I(X : Y) + I(XY : Z) - I(Y : Z) && (2) \\ I(X : YZ) &\geq I(X : Y). && (3) \end{aligned}$$

Equation (3) is in fact equivalent to the *strong sub-additivity property* of von Neumann entropy.

In analogy with classical conditional entropy, we define *conditional* von Neumann entropy $S(Y|X) = \sum_x p_x S(\sigma_x)$, when X is a classical random variable and Y is a quantum encoding of it given by $x \mapsto \sigma_x$. Thus, for example, $I(X : Y) = S(Y) - S(Y|X)$. We similarly define conditional mutual information for quantum states.

3. THE TECHNICAL THEOREMS

3.1 Average encoding

The average encoding theorem asserts that if a quantum encoding has little correlation with the encoded classical information then the encoded states are essentially indistinguishable. In particular, they are all "close" to the *average* encoding. This theorem formalizes a very intuitive idea and might seem to be immediate from Holevo's theorem. However, there is a subtle difference: in Holevo's theorem one is interested in a *single* measurement that *simultaneously* distinguishes all the states, whereas in our case we are interested in their *pairwise* distinguishability. We first prove:

THEOREM 3.1. *Let $x \mapsto \sigma_x$ be a quantum encoding mapping m bit strings $x \in \{0,1\}^m$ into mixed states σ_x. Let X be distributed uniformly over $\{0,1\}^m$ and let Q be the encoding of X according to this map. Then $I(X : Q) \geq 1 - H(\frac{1}{2} + \frac{\Delta}{4})$, where $\Delta = \frac{1}{2^{2m}} \sum_{x_1, x_2 \in \{0,1\}^m} \| \sigma_{x_1} - \sigma_{x_2} \|_{\mathrm{t}}$.*

PROOF. We start with the special case of $m = 1$. It is known [1] that there is a measurement $\mathcal{O}$ on Q that realizes the trace norm distance $t = \| \sigma_0 - \sigma_1 \|_{\mathrm{t}}$ between σ_0 and σ_1. Using Bayes' strategy (see, for example, [11]), the resulting distributions can be identified with probability $\frac{1}{2} + \frac{t}{4}$. Let Y denote the classical random variable holding the result of this entire procedure. We have $\mathrm{Prob}(Y = X) = \frac{1}{2} + \frac{t}{4}$. Thus, by Fano's Inequality, $I(X : Y) \geq 1 - H(\frac{1}{2} + \frac{t}{4})$. We complete the proof for $m = 1$ by noticing that $\Delta = \frac{t}{2}$, and that measurements can only reduce mutual information, so $I(X : Q) \geq I(X : Y)$.

To prove the theorem for general m we reduce it to the $m = 1$ case. We do this by partitioning the set of strings into pairs with "easily" distinguishable encoding.

LEMMA 3.2. *There is a partition of $\{0,1\}^m$ into a set of $2^m/2$ disjoint pairs (x_{2i-1}, x_{2i}) such that*

$$\frac{2}{2^m} \sum_i \| \sigma_{x_{2i-1}} - \sigma_{x_{2i}} \|_{\mathrm{t}} \;\; \geq \;\; \Delta.$$

PROOF. The expectation of the LHS over a random pairing is $\frac{2^m}{2^m - 1} \Delta$, so there is a pairing that achieves average distance Δ. □

We now fix this pairing. Let Z_i denote the set of elements in the i'th pair, i.e., $Z_i = \{x_{2i-1}, x_{2i}\}$ and $\Delta_i = \| \sigma_{x_{2i-1}} - \sigma_{x_{2i}} \|_{\mathrm{t}}$. We know that $\frac{2}{2^m} \sum \Delta_i \geq \Delta$. Let us also denote $f(\delta) = 1 - H(\frac{1}{2} + \frac{\delta}{4})$. From the base case $m = 1$, we know that for any $i = 1, \ldots, 2^m/2$, $I(X : Q \,|\, X \in Z_i) \geq f(\Delta_i)$. Thus we get:

$$S(Q \,|\, X \in Z_i) - \frac{1}{2}[S(\sigma_{x_{2i}}) + S(\sigma_{x_{2i+1}})] \;\; \geq \;\; f(\Delta_i).$$

Averaging all the $2^m/2$ equations yields:

$$\frac{2}{2^m} \sum_i S(Q \,|\, X \in Z_i) - \frac{1}{2^m} \sum_x S(\sigma_x) \;\; \geq \;\; \frac{2}{2^m} \sum_i f(\Delta_i)$$

By the concavity of entropy, $S(Q) \geq \frac{2}{2^m} \sum_i S(Q \,|\, X \in Z_i)$, and by definition $\frac{1}{2^m} \sum_x S(\sigma_x) = S(Q|X)$. Therefore,

$$I(X : Q) \;\; = \;\; S(Q) - S(Q|X) \;\; \geq \;\; \frac{2}{2^m} \sum_i f(\Delta_i).$$

Since f is convex, $\frac{2}{2^m} \sum_i f(\Delta_i) \geq f(\frac{2}{2^m} \sum_i \Delta_i)$. Also, $f(\delta)$ is monotone increasing for $0 \leq \delta \leq 2$, so $f(\frac{2}{2^m} \sum_i \Delta_i) \geq f(\Delta)$. Together this yields $I(X : Q) \geq f(\Delta)$, as required.

Now, we can easily deduce Theorem 1.4.

PROOF OF THEOREM 1.4. Let $\Delta' = \frac{1}{2^m} \sum_{x_1} \| \sigma_{x_1} - \sigma \|_{\mathrm{t}}$. We have:

$$\begin{aligned} \Delta' &= \frac{1}{2^m} \sum_{x_1} \left\| \frac{1}{2^m} \sum_{x_2} (\sigma_{x_1} - \sigma_{x_2}) \right\|_{\mathrm{t}} \\ &\leq \frac{1}{2^{2m}} \sum_{x_1, x_2} \| \sigma_{x_1} - \sigma_{x_2} \|_{\mathrm{t}} \;\; = \;\; \Delta. \end{aligned}$$

By Theorem 3.1, $I(X : Q) \geq 1 - H(\frac{1}{2} + \frac{\Delta}{4})$, and by Fact 2.6 we have $1 - H(\frac{1}{2} + \frac{\Delta}{4}) \geq 1 - (1 - (\frac{\Delta}{4})^2) = \frac{\Delta^2}{16}$. Thus, $\Delta' \leq \Delta \leq 4\sqrt{I(X : Q)}$. □

As stated in the introduction we have a more general theorem that also implies the average encoding theorem. We now describe its proof.

PROOF OF THEOREM 1.5. We show below that it suffices to prove the theorem for 2×2 density matrices, and omit the necessary computations for the 2×2 case from this extended abstract. For details see [15].

All eigenvalues of $\rho-\sigma$ are real, since the matrix is is Hermitian. Let S be the multiset of all nonnegative eigenvalues of $\rho-\sigma$ and R the multiset of all its negative eigenvalues. Now if the dimension of the space $\mathcal{H}_S$ spanned by the eigenvectors corresponding to S has dimension k and the space $\mathcal{H}_R$ spanned by the eigenvectors corresponding to R has dimension $n-k$, we increase (if required) the size of the underlying Hilbert space so that both spaces have the same dimension $n' = \max\{k, n-k\}$. The density matrices have zero entries at the corresponding positions. Now we view the density matrices as density matrices over a product space $\mathcal{H}_2 \otimes \mathcal{H}_{n'}$, where the $\mathcal{H}_2$ space "indicates" the space $\mathcal{H}_S$ or $\mathcal{H}_R$.

We trace out the space $\mathcal{H}_{n'}$ in $\rho, \sigma, \rho-\sigma$, to obtain the 2×2 matrices $\widetilde{\rho}, \widetilde{\sigma}, \widetilde{\rho-\sigma}$. Note that the matrix $\widetilde{\rho-\sigma}$ is diagonalized and contains the sum of all nonnegative eigenvalues, and the sum of all negative eigenvalues on its diagonal. Furthermore $\widetilde{\rho-\sigma} = \widetilde{\rho}-\widetilde{\sigma}$.

Due to Lindblad-Uhlmann monotonicity of the relative von Neumann entropy (see [25]) we get $S(\rho||\sigma) \geq S(\widetilde{\rho}||\widetilde{\sigma})$. Thus, it suffices to bound the latter by

$$\frac{1}{2\ln 2}\,\|\,\widetilde{\rho}-\widetilde{\sigma}\,\|_{\mathrm{t}}^2 \quad = \quad \frac{1}{2\ln 2}\left\|\,\widetilde{\rho-\sigma}\,\right\|_{\mathrm{t}}^2,$$

and then conclude the theorem, since the trace norm of $\widetilde{\rho-\sigma}$ is the sum of absolute values of its eigenvalues, which is the sum of absolute values of eigenvalues of $\rho-\sigma$ by construction, i.e., $||\widetilde{\rho}-\widetilde{\sigma}||_{\mathrm{t}} = ||\rho-\sigma||_{\mathrm{t}}$. □

3.2 Local transition between bipartite states

The idea of local transitions has been used by Lo and Chau [17] and Mayers [18] in showing the impossibility of ideal coin-tossing and bit-commitment. They show that if two bi-partite states are indistinguishable by one party sharing the states, then the other party can locally transform one state to the other. This follows directly from a result due to Jozsa [12], a part of which was stated as Theorem 2.3:

THEOREM 3.3 (JOZSA). *Suppose $|\phi_1\rangle, |\phi_2\rangle \in \mathcal{H}\otimes\mathcal{K}$ 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 $F(\rho_1,\rho_2) = |\langle\phi_1|\,(I\otimes U)\,|\phi_2\rangle|^2$.*

Thus, if $\rho_1 = \rho_2$, the transformation U may be chosen so that $(I\otimes U)\,|\phi_2\rangle = |\phi_1\rangle$. A natural generalization of this is to the case where the reduced density matrices are close to each other but not quite the same, which is what appears in Theorem 1.6. Lo and Chau [17] and Mayers [18] considered this case as well. Theorem 1.6 formalizes their intuition by using the newer results of [11] stated in Theorem 2.4.

PROOF OF THEOREM 1.6. By Theorem 3.3, 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: $F(\rho_1,\rho_2) = |\langle\phi_1\,|\,\phi_2'\rangle|^2$.

Moreover, by fact (1) we have

$$\begin{aligned}&\big\|\,|\phi_1\rangle\langle\phi_1| - |\phi_2'\rangle\langle\phi_2'|\,\big\|_{\mathrm{t}}\\ &\quad=\; 2\sqrt{1-|\langle\phi_1\,|\,\phi_2'\rangle|^2} \;=\; 2\sqrt{1-F(\rho_1,\rho_2)}.\end{aligned}$$

By Theorem 2.4, $\sqrt{F(\rho_1,\rho_2)} \;\geq\; 1-\frac{1}{2}\,\|\,\rho_1-\rho_2\,\|_{\mathrm{t}}$, so

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

Combining these gives us the required result. □

4. 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. We first define a communication problem and state our results formally (giving an overview of the proof), and then give the details of the proofs. For the most part, we will concentrate on communication in a constant number of rounds. Section 4.4 describes the application to the disjointness problem. Section 4.5 discusses our results in the case where the number of messages grows as a function of the input size. Section 4.6 analyzes the quantum communication complexity of the Pointer Jumping function.

4.1 The communication problem and its complexity

In this section, we give the main components of the proof of Theorem 1.1. We define 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 $S_{k-1} : \mathcal{X}_{k-1}\times\mathcal{Y}_{k-1} \rightarrow \{0,1\}$. 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 ath 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 require small probability of error only on average. (The 'U' below stands for the uniform distribution over the inputs.)

THEOREM 4.1. *For all constant $k\geq 1$, $0\leq\epsilon<\frac{1}{2}$,*

$$Q^k_{U,\epsilon}(S_{k+1}) \quad\geq\quad \Omega\left(N^{1/(k+1)}\right).$$

In fact, 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.

Theorem 4.2. *For all constant $k \geq 1$, $0 \leq \epsilon < \frac{1}{2}$, $Q^{k,\bar{P}_k}_{U,\epsilon}(S_k) \geq \Omega(n) = \Omega\left(N^{1/k}\right)$.*

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

Theorem 4.1 now follows directly.

Proof of Theorem 4.1. 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 $Q^{k,P_{k+1}}_{U,\epsilon}(S_k) \leq Q^{k,P_{k+1}}_{U,\epsilon}(S_{k+1})$. But $P_{k+1} = \bar{P}_k$, so the bound of Theorem 4.2 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: $Q^{k+1,\bar{P}_{k+1}}_{U,\epsilon}(S_{k+1}) \leq Q^{k,\bar{P}_{k+1}}_{U,\epsilon}(S_{k+1})$. So we again get the bound from Theorem 4.2. □

We prove Theorem 4.2 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 in [20]. 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 4.3. *For any $0 \leq \epsilon \leq 1$, $Q^{1,A}_{U,\epsilon}(S_1) \geq (1-H(\epsilon))n$.*

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 almost the same total length, again with the wrong player starting, at the cost of a slight increase in the average probability of error.

Lemma 4.4. *For all $k \geq 2$, $0 \leq \epsilon < \frac{1}{2}$, $Q^{k-1,\bar{P}_{k-1}}_{U,\epsilon'}(S_{k-1}) \leq \ell + \log n$, where $\ell = Q^{k,\bar{P}_k}_{U,\epsilon}(S_k)$, and $\epsilon' = \epsilon + 4(\ell/n)^{1/4}$.*

We defer the proof of this lemma to a later section, but show how it implies Theorem 4.2 above.

Proof of Theorem 4.2. We prove the theorem by induction on k. The case $k = 1$ is handled by Lemma 4.3. Suppose the theorem holds for $k-1$. We prove by contradiction that it holds for k as well.

If $Q^{k,\bar{P}_k}_{U,\epsilon}(S_k) = o(n)$, then by Lemma 4.4 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 the same communication complexity $o(n)$. This contradicts the induction hypothesis. □

In the case of communication with prior entanglement, the lower bound in Lemma 4.3 decreases by a factor of two. Lemma 4.4, however, may be strengthened so that we get a slightly better lower bound in Theorem 4.8. The details are omitted.

4.2 The key lemmas

We now prove average case hardness of the index function.

Proof of Lemma 4.3. 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 Q the message qubits sent by Alice. From Properties (2) and (3) in Section 2, and the concavity of binary entropy,

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

The second inequality follows from the fact that Bob has a measurement that predicts X_i with error ϵ_i and Fact 2.5 (Fano's inequality). On the other hand, $I(X:Q)$ is bounded above by the number of qubits in the message. □

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 [19, 16]. However, we use entirely new techniques from quantum information theory, as developed in Sections 3.1 and 3.2 and also get better bounds.

Proof of Lemma 4.4. 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 ℓ message qubits, error ϵ, and k messages starting with Bob. We would like to concentrate on inputs where a is fixed to a particular value in $[n]$ (which is also known to Bob). This would give rise to an instance of S_{k-1} that is also solved by $\mathcal{P}$, but with k messages. An easy argument shows that if it is much smaller than n qubits long, the first message M carries almost no information about y_a for a which is picked at random. We would like to argue that it is therefore not relevant for solving S_{k-1}. However, the correctness of the protocol relies on the message, so we try to reconstruct the message with *Alice* starting the protocol instead. We give the details below.

We first derive a protocol $\mathcal{P}'$ which has low error on an input for S_k generated as below (we call the resulting distribution $U_{a=j}$): $x_1, \ldots, x_n$ are chosen uniformly at random from $\mathcal{X}_{k-1}$, a is set to j, y_j is chosen uniformly at random from $\mathcal{Y}_{k-1}$, and for all $i \neq j$, register Y_i is initialized to the state $\sum_{z \in \mathcal{Y}_{k-1}} |z\rangle$ (normalized).

Let ϵ_j denote the error of $\mathcal{P}$ with respect to the distribution $U_{a=j}$. Note that $\frac{1}{n}\sum_i \epsilon_i \leq \epsilon$, since having the Y_i in a uniform superposition over all possible inputs has the same effect on the result of the protocol as having it randomly distributed over the inputs (recall that we require that the input registers are not changed during a quantum protocol). Let μ_j be the mutual information $I(M:Y_j)$ in the protocol $\mathcal{P}$ when run on the mixed state $U_{a=j}$ with y_j being chosen randomly.

Lemma 4.5. *There is a protocol $\mathcal{P}'$ which solves S_k with respect to the distribution $U_{a=j}$ with error $\delta_j = \epsilon_j + 4\mu_j^{1/4}$ error, ℓ message qubits and k rounds starting with Bob, such that $I(M:Y_j) = 0$.*

The protocol $\mathcal{P}'$ is obtained by slightly modifying the first message in protocol $\mathcal{P}$ so that it is *completely* independent of Y_j. This only affects the average probability of error. Intuitively this means that Alice does not need to get that message at all, or equivalently that she can recreate it herself. This gives a protocol for solving $S_{k-1}(x_j, y_j)$ with $k-1$ messages and with Alice starting.

Lemma 4.6. *There is a protocol $\mathcal{P}''$ that solves S_{k-1} with respect to U with ϵ' error, $\ell + \log n$ message qubits and $k-1$ messages starting with Alice.*

Together we get $Q^{k-1,A}_{U,\epsilon'}(S_{k-1}) \leq \ell + \log n$. □

4.3 Proof of lemmas 4.5 and 4.6

Proof of Lemma 4.5. First consider the case when Y_j is fixed to some z, but the rest of the inputs are as in $U_{a=j}$. In protocol $\mathcal{P}$ Bob applies a unitary transformation V on his qubits and computes $|\phi(z)\rangle = V\,|\bar{0}, Y_1, \ldots, Y_n\rangle$ in register M (for the message) and B (for Bob's ancilla and input). In $\mathcal{P}'$ the message computation is slightly different. Instead of computing $|\phi(z)\rangle$, Bob computes $|\phi'\rangle = V\,|\bar{0}, Y_1, \ldots, Y_{j-1}\rangle\,|\psi\rangle\,|Y_{j+1}, \ldots, Y_n\rangle$, where $|\psi\rangle$ is the uniform superposition over $\mathcal{Y}_{k-1}$. Clearly, in $\mathcal{P}'$ the state $|\phi'\rangle$ and hence the message M does not depend on $y_j = z$, hence $I(M : Y_j) = 0$ when Y_j is uniformly random.

Let us denote by $\rho_M(z)$ the reduced density matrix of the message register M in $\mathcal{P}$ when the input is drawn according to $U_{a=j}$ but $y_j = z$, let the corresponding density matrix for $\mathcal{P}'$ be ρ_M. Clearly, $\rho_M = \frac{1}{|\mathcal{Y}_{k-1}|}\sum_{z\in\mathcal{Y}_{k-1}} \rho_M(z)$. Let $t_z = \|\,\rho_M - \rho_M(z)\,\|_{\mathrm{t}}$, Theorem 1.4 implies $E_z t_z \le 4\sqrt{\mu_j}$.

Protocol $\mathcal{P}'$ generates the pure state $|\phi'\rangle$, while the desired pure state is $|\phi(z)\rangle$. Bob, who knows $y_j = z$ knows both $|\phi(z)\rangle$ and $|\phi'\rangle$. By Theorem 1.6 there is a local unitary transformation T_z acting on register B alone, such that

$$\big\|\, |T_z\phi'\rangle\langle T_z\phi'| - |\phi(z)\rangle\langle\phi(z)| \,\big\|_{\mathrm{t}} \quad\le\quad 2\sqrt{t_z}.$$

The next step in protocol $\mathcal{P}'$ is that Bob applies the transformation T_z to his register B. After that, protocol $\mathcal{P}'$ proceeds exactly as in $\mathcal{P}$. Therefore, for a given z, the probability that $\mathcal{P}$ and $\mathcal{P}'$ disagree on the result is at most $2\sqrt{t_z}$, and the error probability of $\mathcal{P}'$ on $U_{a=j}$ is at most

$$\delta_j \quad=\quad \epsilon_j + 2E_z\sqrt{t_z} \quad\le\quad \epsilon_j + 2\sqrt{E_z t_z} \quad\le\quad \epsilon_j + 4\mu_j^{1/4},$$

where the second step follows from Jensen's inequality. □

Proof of Lemma 4.6. Protocol $\mathcal{P}''$ solves an instance of S_{k-1}. Alice is given an input $\hat{x} \in_R \mathcal{X}_{k-1}$ and Bob is given an input $\hat{y} \in_R \mathcal{Y}_{k-1}$. The protocol proceeds as follows. Alice and Bob first reduce the problem to an S_k instance taken from the distribution $U_{a=j}$ for a random j. To do that, Alice picks $j \in [n]$ at random, sets $a = j$ and sends it to Bob; Alice sets $x_j = \hat{x}$ and Bob sets $y_j = \hat{y}$; Alice picks $x_i \in_R \mathcal{X}_{k-1}$ for $i \neq j$; and Bob initializes each register Y_i for $i \neq j$ with $\sum_{z\in\mathcal{Y}_{k-1}} |z\rangle$ (normalized).

Notice that if Alice and Bob run the protocol $\mathcal{P}'$ over this input, then they get the answer $S_{k-1}(x, y)$ with probability of error at most $\epsilon' = \frac{1}{n}\sum_{i=1}^{n}\delta_i$, which by Lemma 4.5 is bounded by

$$\frac{1}{n}\sum_{i=1}^{n}\epsilon_i + 4\frac{1}{n}\sum_{i=1}^{n}\mu_i^{1/4} \;\le\; \epsilon + 4\left[\frac{1}{n}\sum_{i=1}^{n}\mu_i\right]^{\frac{1}{4}}.$$

We claim that

Claim 4.7. *$\sum_i \mu_i \le \ell_1$, where ℓ_1 is the length of the message M.*

Hence $\epsilon' \;\le\; \epsilon + 4(\ell/n)^{1/4}$.

Alice and Bob do not run the protocol $\mathcal{P}'$ itself, but a modification of it in which Alice sends the first message instead of Bob, thus reducing the number of rounds to $k-1$.

Let ρ_M be the reduced density matrix of register M holding the first message that Bob sends to Alice in $\mathcal{P}'$, for the input given above. By Lemma 4.5, we know that ρ_M does not depend on $y_j = \hat{y}$. So ρ_M is known *in advance* to Alice. Alice starts the protocol $\mathcal{P}''$ by purifying ρ_M. More specifically, let $\{|e_i\rangle\}$ be an eigenvector basis for ρ_M with real and positive eigenvalues λ_i. Alice constructs the superposition $\sum_i \sqrt{\lambda_i}\,|e_i, i\rangle_{MB}$ over two registers M (containing the eigenvectors) and B (containing the label i), and sends register B to Bob. She also sends the index a (chosen as above). The state of the system after this message in $\mathcal{P}''$ is

$$|\xi\rangle \quad=\quad |x_1, \ldots, x_n\rangle_A \otimes \sum_i \sqrt{\lambda_i}\,|e_i\rangle_M\,|i\rangle_B$$

whereas in $\mathcal{P}'$ it is $|\chi(\hat{y})\rangle = |x_1, \ldots, x_n\rangle_A \otimes |T_{\hat{y}}\phi'\rangle_{MB}$.

The reduced density matrix of $|\xi\rangle$ restricted to registers AM is the same as the reduced density matrix of $|\chi(\hat{y})\rangle$ restricted to registers AM. By Theorem 3.3, Bob has a *local* unitary transformation $V_{\hat{y}}$ (operating on his register B) that transforms $|\xi\rangle$ to $|\chi(\hat{y})\rangle$. Bob applies $V_{\hat{y}}$, and Alice and Bob then simulate the rest of the protocol $\mathcal{P}'$. From this stage on, the runs of the protocols $\mathcal{P}'$ and $\mathcal{P}''$ are identical have the same communication complexity and success probability. □

Proof of Claim 4.7. Note that μ_j is the same as the mutual information $I(M : Y_j)$ when $\mathcal{P}$ is run on the uniform distribution on $\mathcal{X}_{k-1}^n \times \mathcal{Y}_{k-1}^n$. In the latter case, Properties (2) and (3) imply the claim, as in the proof of Lemma 4.3. □

4.4 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 the $Q^1_\epsilon(\mathrm{DISJ}) \ge (1 - H(\epsilon))n$ [14, 7]. Furthermore $Q^{O(\sqrt{n})}_{1/3}(\mathrm{DISJ}) = O(\sqrt{n}\log n)$ by an application of Grover search [6]. We now prove a lower bound by reduction.

Proof of Corollary 1.3. 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 4.1 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 ath 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 leads 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 Theorem 4.2, 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 $Q^k_{1/3}(\text{DISJ}) = \Omega(N^{1/k})$ for $k = O(1)$. □

4.5 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 size). We may state the result as follows.

THEOREM 4.8. *For all $k = k(n) \geq 1$, and constant $\epsilon < \frac{1}{2}$, $Q^{k,\bar{P}_k}_{U,\epsilon}(S_k) \quad \geq \quad \Omega\left(\frac{n}{k^4} - k \log n\right)$.*

This theorem follows immediately from Lemmas 4.3 and 4.4 by keeping careful track of the probability of error and the communication cost as a protocol for S_k is reduced to a protocol for S_1.

The above theorem implies a gap in communication complexity between k and $k+1$ message protocols for k up to $\Theta((n/\log n)^{1/5}) = \Theta(\log N/\log\log N)$, and also lower bounds for DISJ for such k.

4.6 The pointer jumping function

The Pointer Jumping function is considered in most results showing a round-hierarchy for classical communication complexity [10, 22, 24, 13]. This problem is a particularly natural candidate for such results.

DEFINITION 4.1. *Let V_A and V_B be disjoint sets of n vertices each.*

Let $F_A = \{f_A | f_A : V_A \to V_B\}$, and $F_B = \{f_B | f_B : V_B \to V_A\}$.

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

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

Then $g_k : F_A \times F_B \to (V_A \cup V_B)$ is defined by $g_k(f_A, f_B) = f^{(k+1)}_{f_A,f_B}(v_1)$, where $v_1 \in V_A$ is fixed. The function $f_k : F_A \times F_B \to \{0,1\}$ is the XOR of all bits in the binary code of the output of g_k.

Nisan and Wigderson proved in [22] that f_k has a randomized k round communication complexity of $\Omega(n/k^2 - k\log n)$ if Bob starts communicating and a deterministic k round communication complexity of $k \log n$ if Alice starts. The lower bound can also be improved to $\Omega(n/k + k)$, see [14]. The advantage of the lower bounds for Pointer Jumping compared to the bounds for the problem investigated in Section 4.1 is that they are linear in the number n of vertices (for constant k). The input length of Pointer Jumping is $2n \log n$. With techniques similar to the ones in this section we can also 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, which is better than the above lower bounds for small constant values of k.

Nisan and Wigderson describe a randomized protocol for computing g_k with communication $O((n/k)\log n + k \log n)$ in the situation where Bob starts and k rounds are allowed [22]. Ponzio *et al.* show that the deterministic communication complexity of f_k is $O(n)$ then, assuming $k = O(1)$ [24]. First we give a new upper bound which combines ideas from [22] and [24]. Its proof is given in [15].

THEOREM 4.9. $R^{k,B}_\epsilon(g_k) \leq O(\frac{n}{k\epsilon} \cdot (\log^{(k/2)} n + \log k) + k \log n)$.

If $k \geq 2\log^*(n)$ then $R^{k,B}(g_k) \leq O((\frac{n}{k} + k)\log k)$.

Previous lower bounds for Pointer Jumping [22, 24] take the following approach. They consider the complexity of deterministic protocols with error under the uniform distribution. Then they show that at a random leaf of the protocol tree with high probability the entropy of v_{k+2} is still large, where $v_1, v_2, \ldots, v_{k+2}$ denote the vertices of the relevant path in the input graph.

One is tempted to think that a "simpler" approach is possible, and that the information between the messages of all rounds and v_{k+2} is small. Or more generally that the information between the first t messages and v_{t+1} is small. But this is not true, as the protocol for the Pointer Jumping function g_k resp. f_k given in the proof of Theorem 4.9 shows.

So in our lower bound we replace the usual notion of information by another quantity called informational distance, which is based on distinguishability.

Due to Theorem 1.5, the following holds for a bipartite state ρ_{AB}:

$$I(A:B) = S(\rho_{AB} || \rho_A \otimes \rho_B) \geq \frac{1}{2\ln 2} ||\rho_{AB} - \rho_A \otimes \rho_B||^2_1.$$

Thus the measurable distance between the tensor product state and the "real" bipartite state can be bounded in terms of the information. We will call the value $D(A:B) = ||\rho_{AB} - \rho_A \otimes \rho_B||_{\mathrm{t}}$ the *informational distance*. The next lemma collects a few properties of informational distance.

LEMMA 4.10. *For all states ρ_{ABC} the following holds:*

1. *$D(A:B) = D(B:A)$.*
2. *$D(AB:C) \geq D(A:C)$.*
3. *$0 \leq D(A:B) \leq 2$.*
4. *$D(A:B) \geq ||F(\rho_{AB}) - F(\rho_A \otimes \rho_B)||_{\mathrm{t}}$ for all completely positive and trace-preserving superoperators F.*
5. *$D(A:B) \leq \sqrt{2I(A:B)}$.*

Now we state our lower bound for Pointer Jumping. Note that the lower bound is linear in n for constant k and leads to Theorem 1.2.

THEOREM 4.11. $Q^{k,B}_{1/3}(f_k) \geq n/2^{2^{O(k)}} - k \log n$.

PROOF. We consider some quantum protocol for f_k with error 1/3, k rounds, Bob starting.

At any time in the protocol Alice has access to qubits containing her input, some "work" qubits and some of the qubits used in messages so far, the same holds for Bob. We require the protocol to satisfy some properties. First we require that in round t the vertex $v_t = f^{(t-1)}(v_1)$ is communicated by a classical message and stored by the receiving player. This increases the communication by an additive $k \log n$ term. Furthermore we demand the protocol be of the form described in Section 2.

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. The superposition on inputs is measured after the protocol has finished.

The density matrix of the global state of the protocol is $\rho_{M_{A,t}M_{B,t}F_AF_B}$. Here 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.

We demand that before round t the t th vertex of the path is measured. This vertex is stored in some qubits V_t. V_1 has the fixed value v_1. Before some later round t the global state is a probabilistic mixture over the possibilities to fix the first $t-1$ vertices of the path. For each pure state in the mixture the first $t-1$ vertices are fixed and V_t is either $F_A(v_{t-1})$ or $F_B(v_{t-1})$ and may be measured in the standard basis. Note that the fixed vertices are included in previous messages. The measurements do not affect the correctness of the protocol.

We assume that the communication complexity of the protocol is now δn and prove a lower bound $\delta \geq 2^{-2^{O(k)}}$.

The general strategy of the proof is an induction over the rounds. We show that $D(M_{A,t+1}F_A : F_B(V_{t+1})) \leq 4\sqrt{D(M_{B,t}F_B : F_A(V_t))} + \sqrt{4\delta}$ (and the same with A and B exchanged). Actually this is a slight abuse of terminology, since V_{t+1} is not fixed but determined by $M_{A,t+1}$ and F_A, and so for different messages different pointers are considered. However, at the time, when we consider $F_B(V_{t+1})$ we have that V_{t+1} is a classical random variable, whose value is fixed in $M_{A,t+1}$.

Bob sends the first message. Then obviously $I(M_{B,1}F_B : V_2) = 0$, because Bob has seen no message yet, and V_2 is determined by F_A. This implies $D(M_{B,1}F_B : V_2) = 0$. The invariant of the induction will be that $D(M_{A,t}F_A : V_{t+1})$ resp. $D(M_{B,t}F_B : V_{t+1})$ is small.

First we consider the information Alice has on Bob's input (for a proof see [15]).

Lemma 4.12. *$I(M_{A,t}F_A : F_B) \leq 2\delta n$ at all times t.*

Now consider the situation that F_B is uniformly random instead of being in the Hadamard superposition. Then $\sum_{i=1}^{n} \frac{1}{n} I(M_{A,t}F_A : F_B(i)) \leq 2\delta$, because the $F_B(i)$ are mutually independent. The value of $I(M_{A,t}F_A : F_B(i))$ stays the same, if all $F_B(j)$ for $j \neq i$ are in superposition and $F_B(i)$ is random, instead of all of F_B being random.

By Lemma 4.10 we get:

$$\sum_{i=1}^{n} \frac{1}{n} D(M_{A,t}F_A : F_B(i)) \leq \sqrt{4\delta}, \tag{4}$$

where (4) holds at all times in the protocol, if we consider the situation that $F_B(i)$ is random instead of being in superposition.

We use the induction hypothesis that $D(M_{A/B,t}F_{A/B} : V_{t+1}) \leq \gamma_t$ and let $\gamma_t = 4\sqrt{\gamma_{t-1}} + \sqrt{4\delta}$ and $\gamma_1 = 0$. Then $\gamma_{t+1} \leq 3^t \delta^{1/2^t}$ for all $t \geq 0$.

W.l.o.g. let Alice be the speaker in round $t+1$. Before that 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. Any reduced state containing at least all qubits of one player is block diagonal with respect to the possible values of the vertices $V_1, \ldots, V_t$, since they are either in a player's input or received messages.

We may assume by induction that $D(M_{B,t}F_B : V_{t+1}) \leq \gamma_t$ and consequently $D(M_{B,t+1}F_B : V_{t+1}) \leq \gamma_t$, because in round t Bob has received no qubits. Let $M_A = M_{A,t+1}$ and $M_B = M_{B,t+1}$. Consider some fixed path p which is a value of $V_1, \ldots, V_t$. Let v be some value of V_{t+1}. For any p, v let $\rho^p_{M_AM_BF_AF_B}$ denote the state with the path $V_1, \ldots, V_t$ fixed to p and $\rho^{p,v}_{M_AM_BF_AF_B}$ denote the state with $V_1, \ldots, V_{t+1}$ fixed to p, v.

We know that $\rho_{M_BF_BV_{t+1}}$ and $\rho_{M_BF_B} \otimes \rho_{V_{t+1}}$ are close in the trace distance: $||\rho_{M_BF_BV_{t+1}} - \rho_{M_BF_B} \otimes \rho_{V_{t+1}}||_{\mathrm{t}} = E_pE_v||\rho^{p,v}_{M_BF_B} - \rho^p_{M_BF_B}||_{\mathrm{t}} \leq \gamma_t$.

Then $E_{p,v}\gamma_{p,v} \leq \gamma_t$ for

$$\gamma_{p,v} = ||\rho^{p,v}_{M_BF_B} - \rho^p_{M_BF_B}||_{\mathrm{t}}. \tag{5}$$

Furthermore (4) implies

$$||\rho^p_{M_AF_AF_B(i)} - \rho^p_{M_AF_A} \otimes \rho_{F_B(i)}||_{\mathrm{t}} \quad = \quad \beta_{p,i} \tag{6}$$

with $E_{p,i}\beta_{p,i} \leq \sqrt{4\delta}$, where i is uniformly distributed. Here $F_B(i)$ is assumed to be uniformly random (i.e., measured). Equation (6) also holds if V_{t+1} is not yet measured.

Now we are interested in the value $D(M_AF_A : F_B(V_{t+1}))$

$$= \quad E_pE_v||\rho^{p,v}_{M_AF_AF_B(v)} - \rho^{p,v}_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}}$$

We now employ Theorem 1.6. $\rho^{p,v}_{M_AM_BF_AF_BR}$ is a purification of $\rho^{p,v}_{M_BF_B}$ and $\rho^p_{M_AM_BF_AF_BR}$ a purification of $\rho^p_{M_BF_B}$ for all p, v, where R is some additional space used to purify the random V_{t+1} in $\rho^p_{M_AM_BF_AF_B}$ and left blank in the former.

Due to the Theorem 1.6 there is a unitary transformation U such that the application of U to F_AM_AR changes $\rho^p_{M_AM_BF_AF_BR}$ to some state $\sigma^p_{M_AM_BF_AF_BR}$ that has by (5) distance $2\sqrt{\gamma_{p,v}}$ from $\rho^{p,v}_{M_AM_BF_AF_BR}$.

Then since tracing out cannot increase the distance:

$$||\rho^{p,v}_{M_AF_AF_B(v)} - \sigma^p_{M_AF_AF_B(v)}||_{\mathrm{t}} \leq 2\sqrt{\gamma_{p,v}}. \tag{7}$$

This also holds, if $F_B(v)$ is measured. Since we have to show the induction step only for a state with uniformly random V_{t+2}, we consider $F_B(v)$ as uniformly random from now on. Since U is unitary and acts on M_AF_AR only for all p, v:

$$\begin{aligned}
||\sigma^p_{M_AF_AF_B(v)} \quad &- \quad \sigma^p_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&\leq \quad ||\sigma^p_{M_AF_ARF_B(v)} - \sigma^p_{M_AF_AR} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&= \quad ||\rho^p_{M_AF_ARF_B(v)} - \rho^p_{M_AF_AR} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&\overset{*}{=} \quad ||\rho^p_{M_AF_AF_B(v)} - \rho^p_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&\overset{(6)}{=} \quad \beta_{p,v}.
\end{aligned} \tag{8}$$

Here (*) holds for the states, when $F_B(v)$ is uniformly random, but V_{t+1} is not yet measured. For all p, v:

$$\begin{aligned}
&||\rho^{p,v}_{M_AF_AF_B(v)} - \rho^{p,v}_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&\quad \leq \quad ||\rho^{p,v}_{M_AF_AF_B(v)} - \sigma^p_{M_AF_AF_B(v)}||_{\mathrm{t}} \\
&\qquad\quad + \quad ||\sigma^p_{M_AF_AF_B(v)} - \rho^{p,v}_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&\quad \overset{(7)}{\leq} \quad 2\sqrt{\gamma_{p,v}} + ||\sigma^p_{M_AF_AF_B(v)} - \rho^{p,v}_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&\quad \leq \quad 2\sqrt{\gamma_{p,v}} + ||\sigma^p_{M_AF_AF_B(v)} - \sigma^p_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&\qquad\quad + \quad ||\sigma^p_{M_AF_A} \otimes \rho_{F_B(v)} - \rho^{p,v}_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&\quad \overset{(7)}{\leq} \quad 4\sqrt{\gamma_{p,v}} + ||\sigma^p_{M_AF_AF_B(v)} - \sigma^p_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\
&\quad \overset{(8)}{\leq} \quad 4\sqrt{\gamma_{p,v}} + \beta_{p,v}.
\end{aligned}$$

Thus, $D(M_AF_A : F_B(V_{t+1}))$

$$\begin{aligned} &= E_{p,v}||\rho^{p,v}_{M_AF_AF_B(v)} - \rho^{p,v}_{M_AF_A} \otimes \rho_{F_B(v)}||_{\mathrm{t}} \\ &\leq E_{p,v}[4\sqrt{\gamma_{p,v}} + \beta_{p,v}] \leq 4\sqrt{E_{p,v}\gamma_{p,v}} + \sqrt{4\delta} \leq \gamma_{t+1}. \end{aligned}$$

After round round k one player, say Alice, announces the result which is supposed to be the parity of v_{k+2} and included in $M_{A,k+1}$. But $D(M_{A,k+1} : V_{k+2}) \leq \gamma_{k+1} \leq 3^k\delta^{1/2^k}$. It is not hard to see that for error 1/3:

$$\gamma_{k+1} \geq D(M_{A,k+1} : V_{k+2}) \geq D(M_{A,k+1} : \bigoplus V_{k+2}) \geq 1/3.$$

Thus $3^k\delta^{1/2^k} \geq 1/3$ and $\delta \geq 2^{-2^{O(k)}}$. □

Acknowledgements

We thank Jaikumar Radhakrishnan and Venkatesh Srinivasan for their input on the classical communication complexity of Pointer Jumping and the subproblem S_k, and Dorit Aharonov and Pranab Sen for helpful feedback.

5. REFERENCES

[1] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pp. 20–30, 1998.

[2] A. Ambainis. A new protocol and lower bounds for quantum coin flipping. In these proceedings.

[3] A. Ambainis, L. J. Schulman, A. Ta-Shma, U. Vazirani, and A. Wigderson. The quantum communication complexity of sampling. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, pp. 342–351, 1998.

[4] C.H. Bennett and S.J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69:2881–2884, 1992.

[5] E. Bernstein and U. Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, October 1997.

[6] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, 1998.

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

[8] R. Cleve, W. van Dam, M. Nielsen and A. Tapp. Quantum entanglement and the communication complexity of the inner product function. In *Proceedings of the 1st International Conference on Quantum Computing and Quantum Communication*, vol. 1509 of *Lecture Notes in Computer Science*, pp. 61–74. Springer-Verlag, 1998.

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

[10] P. Duris, Z. Galil, and G. Schnitger. Lower bounds on communication complexity. *Information and Computation*, 73(1):1–22, April 1987.

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

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

[13] H. Klauck. Lower bounds for computation with limited nondeterminism. In *Proceedings of the 13th Annual IEEE Conference on Computational Complexity*, 1998.

[14] H. Klauck. On quantum and probabilistic communication: Las vegas and one-way protocols. In *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing*, pp. 644–651, 2000.

[15] H. Klauck. On rounds in quantum communication. LANL Preprint archive, quant-ph/0004100, 2000.

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

[17] H. Lo and H. Chau. Why quantum bit commitment and ideal quantum coin tossing are impossible. *Physica D*, 120:177–187, 1998. See also quant-ph/9711065.

[18] D. Mayers. Unconditionally secure quantum bit commitment is impossible. *Physical Review Letters*, 78:3414–3417, 1997.

[19] P.B. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on the Theory of Computing*, pp. 103–111, 1995.

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

[21] A. Nayak, A. Ta-Shma, and D. Zuckerman. Interaction in quantum communication and the complexity of set disjointness. Technical report TR 2000-36, DIMACS Center, Rutgers University. Earlier version quant-ph/0005106, 2000.

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

[23] C.H. Papadimitriou and M. Sipser. Communication complexity. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, pp. 196–200, 1982.

[24] S.J. Ponzio, J. Radhakrishnan, and S. Venkatesh. The communication complexity of pointer chasing, applications of entropy and sampling. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pp. 602–611, 1999.

[25] J. Preskill. Lecture notes. http://www.theory.caltech.edu/people/preskill/ph229/.

[26] R. Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pp. 358–367, 1999.

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

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