DIMACS Technical Report 2000-36
December 2000

Interaction in Quantum Communication and the Complexity of Set Disjointness

by

Ashwin Nayak [1]
DIMACS Center
Rutgers, P.O. Box 1179
Piscataway, NJ 08855
nayak@dimacs.rutgers.edu

Amnon Ta-Shma [2]
Department of Computer Science
Tel-Aviv University
Israel 69978
amnon@post.tau.ac.il

David Zuckerman [3]
Department of Computer Science
University of Texas
Austin, TX 78712
diz@cs.utexas.edu

[1]Supported by a joint DIMACS-AT&T Post-Doctoral Fellowship and NSF grants STC 91-19999, CCR 99-06105 and EIA 00-80234. Part of this work was completed while the author was at University of California, Berkeley, and was supported by a JSEP grant and NSF grant CCR 9800024.
[2]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.
[3]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.

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 [2, 22]. 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 [11], 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 bi-partite 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 [2]. 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 [22], 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 [2, 22] also share the feature that they require minimal *interaction* between the communicating players. For example, in the protocol of [2], 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 problems such as checking set disjointness (DISJ) seem to require much more interaction: Buhrman, Cleve and Wigderson [4] 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 $\sqrt{N}$), 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 [2, 22]?

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 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 5.4 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 [19, 7, 18, 20]. In fact, the problem we study in this paper is the subproblem of Pointer Jumping singled out in [16] (See Section 5.1 for a formal definition of the subproblem). Our analysis follows the same intuition as that behind the result of [16] (also explained in [13]), but relies on entirely new ideas from quantum information theory. The resulting lower bound is optimal (for a constant number of rounds).

In the context of quantum communication, it was observed by Buhrman and de Wolf [5] (based on a lower bound of Nayak [17]) 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 [17] 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 [11], which independently raised the question of the role of rounds in quantum communication). Our result subsumes all these.

In [11], Klauck states a simple reduction from Pointer Jumping in a bounded number of rounds to DISJ.[1] Our result thus implies new lower bounds for DISJ.

Corollary 1.2 *For any constant k, any k-message quantum protocol for Set Disjointness has communication complexity $\Omega(N^{1/k})$.*

The problem of determining the quantum communication complexity of DISJ has inspired much research in the last few years and yet the best known lower bound remains $\Omega(\log n)$ [2, 5]. Our result provides new insight into the complexity of the problem.

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 a new primitive in quantum encoding, as suggested by the following theorem.

Theorem 1.3 (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,*

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

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.

We also use another primitive derived from the work of Lo and Chau [14] and Mayers [15] which combines results of Jozsa [10], and Fuchs and van de Graaf [8]. 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 1.4 (Local transition theorem) *(based on [14, 15, 10, 8]) 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*

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

[1]The reduction is described in a recent paper [12].

These primitives may be of significance in other applications as well, especially in a cryptographic context.

Subsequent to our work, Klauck [12] proved a lower bound for the bounded round quantum communication complexity of Pointer Jumping. The lower bound is linear for a constant number of message exchanges. Klauck's result is based on ideas similar to ours, and involves a bound on "informational distance", a notion that generalizes average encoding, and the local transition theorem (Theorem 1.4 above). It implies a round hierarchy and a lower bound for DISJ similar to ours, but with a worse dependence on the number of rounds.

2 Preliminaries

In this section we briefly describe the communication model we study. Our lower bound results rely heavily on quantum information theory. The necessary background is provided in Appendix A, along with the associated notation.

2.1 The communication complexity model

In the quantum communication complexity model [24], 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 [3]) 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 $1-\epsilon$. $Q_\epsilon(f)$ denotes the complexity of the best quantum 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

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

$Q^{c,P}_{\mu,\epsilon}(f)$ 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, $Q^{c,P}_{\mu,\epsilon}(f) \leq Q^{c,P}_{\epsilon}(f)$.*

3 The average encoding theorem

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 the *pairwise* distinguishability of the encoded states. 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. Denote $\Delta = \frac{1}{2^{2m}} \sum_{x_1,x_2 \in \{0,1\}^m} \| \sigma_{x_1} - \sigma_{x_2} \|_{\mathrm{t}}$. Then $I(X:Q) \geq 1 - H(\frac{1}{2} + \frac{\Delta}{4})$.*

Proof: We start with the special case of $m = 1$. By Theorem A.3, 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, [8]), 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 set of $2^m/2$ disjoint pairs (x_{2i-1}, x_{2i}) which together cover $\{0,1\}^m$ such that*

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

Proof: The expectation of the LHS over a random pairing is $\frac{2^m}{2^m-1}\Delta$; hence there is a pairing that achieves this 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 $\Delta_i = \left\| \sigma_{x_{2i-1}} - \sigma_{x_{2i}} \right\|_{\mathrm{t}}$. 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 the entropy function, $S(Q) \geq \frac{2}{2^m}\sum_i S(Q \,|\, X \in Z_i)$, 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 1.3.

Proof of Theorem 1.3: Let $\Delta' = \frac{1}{2^m}\sum_{x_1} \| \sigma_{x_1} - \sigma \|_{\mathrm{t}}$. We have:

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

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

4 Local transition between bipartite states

Lo and Chau [14] and Mayers [15] proved:

Theorem 4.1 (Lo and Chau; Mayers) *Suppose $|\phi_1\rangle$ and $|\phi_2\rangle$ are two pure states in the Hilbert space $\mathcal{H} \otimes \mathcal{K}$, such that $\mathrm{Tr}_{\mathcal{K}} |\phi_2\rangle\langle\phi_2| = \mathrm{Tr}_{\mathcal{K}} |\phi_1\rangle\langle\phi_1|$, i.e., the reduced density matrix of $|\phi_2\rangle$ restriced to $\mathcal{H}$ is the same as the reduced density matrix of $|\phi_1\rangle$ restriced to $\mathcal{H}$. Then, there is a local unitary transformation U on $\mathcal{K}$ such that $I \otimes U\,|\phi_2\rangle = |\phi_1\rangle$.*

The theorem follows by examining the Schmidt decomposition [21] of the two states.

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 1.4. Lo and Chau [14] and Mayers [15] considered this case as well. Theorem 1.4 formalizes their intuition by using the newer results of [8] stated in Theorem A.6.

Proof of Theorem 1.4: By Theorem A.5, there exists a purification $|\phi_2'\rangle \in \mathcal{H}\otimes\mathcal{K}$ of ρ_2 such that $|\langle\phi_1 \,|\, \phi_2'\rangle|^2 = F(\rho_1, \rho_2)$. Since $|\phi_2\rangle$ and $|\phi_2'\rangle$ have the same reduced density matrix in $\mathcal{H}$, by Theorem 4.1, there is a (local) unitary transformation U on $\mathcal{K}$ such that $I \otimes U \,|\phi_2\rangle = |\phi_2'\rangle$. Moreover, by Fact A.4 we have

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

By Theorem A.6, $\sqrt{F(\rho_1, \rho_2)} \;\geq\; 1 - \frac{1}{2} \, \| \, \rho_1 - \rho_2 \, \|_{\mathrm{t}}$, so

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

This, when combined with the earlier bound on the trace distance between $|\phi_1\rangle, |\phi_2'\rangle$ gives us the required result. ■

5 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 5.4 discusses our results in the case where the number of messages grows as a function of the input size.

5.1 The communication problem and its complexity

In this section, we give the main components of the proof of Theorem 1.1. We define a sequence of problems $S_0, S_1, \ldots, S_k, \ldots$ 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 5.1 *For all constant $k \geq 1$, $0 \leq \epsilon < \frac{1}{2}$,* $\;Q_{U,\epsilon}^k(S_{k+1}) \;\geq\; \Omega\left(N^{1/(k+1)}\right)$.

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

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 5.1 now follows directly.

Proof of Theorem 5.1: 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 5.2 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 same bound from Theorem 5.2. ■

We prove Theorem 5.2 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 [17]. 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 5.3 *For any $0 \leq \epsilon \leq 1$,* $\quad Q^{1,A}_{U,\epsilon}(S_1) \;\geq\; (1 - H(\epsilon))n.$

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 5.4 *For all $k \geq 2$, $0 \leq \epsilon < \frac{1}{2}$,* $\quad Q^{k-1,\bar{P}_{k-1}}_{U,\epsilon'}(S_{k-1}) \;\leq\; \ell + \log n$, *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 5.2 above.

Proof of Theorem 5.2: We prove the theorem by induction on k. The case $k = 1$ is handled by Lemma 5.3. 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 5.4 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. ■

5.2 The key lemmas

We now prove the average case hardness of the index function.

Proof of Lemma 5.3: 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 (1) and (2) in Appendix 2, and the concavity of the binary entropy function,

$$I(X_1 \cdots X_n : Q) \;\; \geq \;\; \sum_i I(X_i : Q) \;\; \geq \;\; \sum_i (1 - H(\epsilon_i)) \;\; \geq \;\; n(1 - H(\epsilon)).$$

The second inequality follows from the fact that Bob has a measurement that predicts X_i with error ϵ_i and Fact A.1 (Fano's inequality). On the other hand, $I(X : Q)$ is bounded above by the number of qubits in the message. This poves the lemma. ■

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 [16, 13]. However, we use entirely new techniques from quantum information theory, as developed in Section 3 and 4 and also get better bounds in the process.

Proof of Lemma 5.4: 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 distribtution 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 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 5.5 *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 5.6 *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 $Q^{k-1,A}_{U,\epsilon'}(S_{k-1}) \le \ell + \log n$ as claimed. $\blacksquare$

5.3 Proof of Lemmas 5.5 and 5.6

Proof of Lemma 5.5: 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 chosen randomly.

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 $U_{a=j}$ and $y_j = z$, and 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 $t_z = \|\, \rho_M - \rho_M(z)\,\|_{\mathrm{t}}$. By Theorem 1.3 we know that $E_z t_z \le 4\sqrt{\mu_j}$.

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 1.4 there is a local unitary transformation T_z acting on register B alone, such that

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

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

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

where the second step follows from Jensen's inequality. $\blacksquare$

Proof of Lemma 5.6: 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 $i \ne j$; and Bob initializes each register Y_i for $i \ne j$ 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

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

We claim that

Claim 5.7 $\sum_i \mu_i \leq \ell_1$, *where* ℓ_1 *is the length of the message* M.

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

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 5.5, 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

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

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 4.1, 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 5.7: 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 $\mathcal{X}_{k-1}^n \times \mathcal{Y}_{k-1}^n$. So we prove the claim for the latter.

For any i, $I(Y_i : Y_1\cdots Y_{i-1}Y_{i+1}\cdots Y_n) = 0$. Therefore by Properties (1) and (2) (cf. Appendix A) we have

$$I(M:Y_1\cdots Y_n) \;\geq\; \sum_{i=1}^{n} I(MY_1\cdots Y_{i-1}:Y_i) \;\geq\; \sum_{i=1}^{n} I(M:Y_i) \;=\; \sum_i \mu_i$$

As the first message M contains only ℓ_1 qubits, we have $\sum_i \mu_i \leq I(M:Y_1\cdots Y_n) \leq \ell_1$. ■

5.4 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 5.8 *For all $k = k(n) \geq 1$, and constant $\epsilon < \frac{1}{2}$,* $Q^{k,\bar{P}_k}_{U,\epsilon}(S_k) \;\geq\; \Omega\left(\frac{n}{k^4} - k\log n\right)$

This theorem follows immediately from Lemmas 5.3 and 5.4 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.

Acknowledgements

We thank Jaikumar Radhakrishnan and Venkatesh Srinivasan for their input on the classical communication complexity of the pointer jumping and the subproblem we study in this paper, Hartmut Klauck for telling us the reduction from Pointer Jumping to Set Disjointness, and Dorit Aharonov and Pranab Sen for helpful feedback on the paper.

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*, pages 20–30, New York, May 23–26, 1998. ACM Press.

[2] 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*, pages 342–351, Los Alamitos, CA, November 8–11, 1998. IEEE Computer Society Press.

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

[4] 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*, New York, NY, May 23–26, 1998. ACM Press.

[5] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. Technical report, LANL CS archive, http://www.lanl.gov/abs/cs/9910010, 1999.

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

[7] Pavol Duris, Zvi Galil, and Georg Schnitger. Lower bounds on communication complexity. *Information and Computation*, 73(1):1–22, April 1987.

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

[9] A. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973. English translation in *Problems of Information Transmission*, volume 9, 1973, pages 177–183.

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

[11] Hartmut 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*, Portland, OR, May 21–23, 2000. ACM Press.

[12] Hartmut Klauck. On rounds in quantum communication. Technical report, quant-ph/0004100, 2000.

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

[14] 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.

[15] D. Mayers. Unconditionally secure quantum bit commitment is impossible. *Phys. Rev. Lett.*, 78:3414–3417, 1997.

[16] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on the Theory of Computing*, pages 103–111, Las Vegas, Nevada, May 29–June 1, 1995.

[17] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, pages 369–376, New York, NY, October 17–19, 1999. IEEE Computer Society Press.

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

[19] Christos H. Papadimitriou and Michael Sipser. Communication complexity. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, pages 196–200, San Francisco, California, May 5–7, 1982.

[20] Stephen J. Ponzio, Jaikumar 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*, pages 602–611, New York, May 1–4 1999. ACM Press.

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

[22] Ran Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pages 358–367, Atlanta, GA, May 1–4, 1999. ACM Press.

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

[24] Andrew Chi-Chih Yao. Quantum circuit complexity. In *Proceedings of the 34th Annual Symposium on Foundations of Computer Science*, pages 352–361, Palo Alto, CA, November 3–5, 1993. IEEE Computer Society Press.

A Information theory background

A.1 Classical entropy and mutual information

The *Shannon entropy* $S(X)$ of a classical random variable X quantifies the amount of randomness in it. If X takes values x in some finite set with probability p_x, its Shannon entropy is defined as $S(X) = -\sum_x p_x \log p_x$. The *mutual information* $I(X:Y)$ of a pair of random variables X, Y is defined by $I(X:Y) = S(X) + S(Y) - S(XY)$. It is a measure of how correlated the two random variables are.

The following are some basic facts about the mutual information function that we use in the paper. For other equivalent definitions and properties of these concepts, we refer the reader to a standard text (such as [6]) on information theory.

For any random variables X, Y, Z,

$$\begin{aligned} I(X:YZ) &= I(X:Y) + I(XY:Z) - I(Y:Z) \qquad (1) \\ I(X:YZ) &\geq I(X:Y). \qquad (2) \end{aligned}$$

Fano's inequality states that if Y can predict another random variable X with an advantage, then X and Y have large mutual information. We use it only in the following simple form.

Fact A.1 (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 $I(X:Y) \geq 1 - H(p)$, where $H(\cdot)$ is the binary entropy function.*

Finally, we need a bound on the deviation of the binary entropy function $H(p)$ from 1 as p deviates from 1/2.

Fact A.2 *For $\delta \in [-\frac{1}{2}, \frac{1}{2}]$, we have $H(\frac{1}{2} + \delta) \leq 1 - \delta^2$.*

Proof: From the definition of the binary entropy function, we have

$$H(\frac{1}{2} + \delta) = 1 - \frac{1}{2}[(1+2\delta)\log(1+2\delta) + (1-2\delta)\log(1-2\delta)].$$

Using the expansion $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$ for $|x| < 1$, and simplifying, we get

$$\begin{aligned} H(\frac{1}{2} + \delta) &= 1 - (\log \mathrm{e}) \left[\left(2 - \frac{2^2}{2 \cdot 2}\right)\delta^2 + \left(\frac{2^3}{3} - \frac{2^4}{2 \cdot 4}\right)\delta^4 + \left(\frac{2^5}{5} - \frac{2^6}{2 \cdot 6}\right)\delta^6 + \cdots \right] \\ &\leq 1 - \delta^2, \end{aligned}$$

which is the claimed bound. ■

A.2 The density matrix and the trace norm

The quantum mechanical analogue of a random variable is a probability distribution over superpositions, also called a *mixed state*. Consider the mixed state $X = \{p_i, |\phi_i\rangle\}$, where the superposition $|\phi_i\rangle$ is drawn with probability p_i. The *density matrix* of the mixed state X is $\rho_X = \sum_i p_i |\phi_i\rangle\langle\phi_i|$. The following properties of density matrices are immediate from the definition: every density matrix ρ is Hermitian, i.e., $\rho = \rho^\dagger$, has unit trace, i.e., $\mathrm{Tr}(\rho) = \sum_i \rho(i,i) = 1$, and is positive semi-definite, i.e., $\langle\psi|\,\rho\,|\psi\rangle \geq 0$ for all $|\psi\rangle$. Thus, every density matrix is *unitarily diagonalizable* and has non-negative real eigenvalues that sum up to 1.

Mixed states often arise when we consider pure states that are shared by two parties, say Alice and Bob. Consider a pure state $|\phi\rangle \in \mathcal{H} \otimes \mathcal{K}$ where the qubits of $\mathcal{H}$ are held by Alice and those of $\mathcal{K}$ by Bob. The qubits with Alice then behave as if they were in a mixed state obtained by measuring Bob's qubits (in an arbitrary basis). The corresponding density matrix ρ for Alice is given by the *partial trace* operation $\rho = \mathrm{Tr}_{\mathcal{K}} |\phi\rangle\langle\phi| = \sum_{h,h'} |h\rangle\langle h'| \sum_k \langle h,k\,|\,\phi\rangle\,\langle\phi\,|\,h',k\rangle$ where h, h' vary over an orthonormal basis for $\mathcal{H}$ and k over an orthonormal basis for $\mathcal{K}$.

Given a quantum system in a mixed state with density matrix ρ and a (general) measurement $\mathcal{O}$ on it, let $\rho^{\mathcal{O}}$ denote the classical distribution on the possible results that we get by measuring ρ according to $\mathcal{O}$. Suppose that it is some classical distribution $p_1, \ldots, p_k$ where we get result i with probability p_i. Given two different mixed states, we can ask how well one can distinguish between the two mixtures, or equivalently, how different the distributions resulting from a measurement may be. To quantify this, we consider the ℓ_1 metric: if $p = (p_1, \ldots, p_k)$ and $q = (q_1, \ldots, q_k)$ are two probability distributions over $\{1, \ldots, k\}$, then the ℓ_1 distance between them is $\|\, p - q \,\|_1 = \sum_i |p_i - q_i|$. A fundamental theorem about distinguishing density matrices (see [1]) tells us:

Theorem A.3 *Let ρ_1, ρ_2 be two density matrices on the same space $\mathcal{H}$. Then for any (general) measurement $\mathcal{O}$*

$$\left\|\, \rho_1^{\mathcal{O}} - \rho_2^{\mathcal{O}} \,\right\|_1 \;\leq\; \mathrm{Tr}\sqrt{A^\dagger A},$$

where $A = \rho_1 - \rho_2$. Furthermore, the bound is tight, and the orthogonal measurement $\mathcal{O}$ that projects a state on the eigenvectors of $\rho_1 - \rho_2$ achieves this bound.

Theorem A.3 shows that the density matrix captures all the accessible information that a quantum state contains. If two different mixtures have the same density matrix (which is indeed possible) then even though they are two distinct distributions, they are physically, and thus from a computational point of view, indistinguishable. As the behavior of a mixed state is completely characterized by its density matrix we often identify a mixed state with its density matrix.

The quantity $\mathrm{Tr}\sqrt{A^\dagger A}$ is of independent interest. (Note that this is compact notation for the sum of the (magnitudes of the) singular values of A.) If we define $\|\, A \,\|_{\mathrm{t}} = \mathrm{Tr}\sqrt{A^\dagger A}$ then $\|\cdot\|_{\mathrm{t}}$ defines a norm (the *trace norm*), and has some additional properties such as $\|\, A \otimes B \,\|_{\mathrm{t}} = \|\, A \,\|_{\mathrm{t}} \cdot \|\, B \,\|_{\mathrm{t}}$, $\|\, A \,\|_{\mathrm{t}} = 1$ for any density matrix A and $\|\, AB \,\|_{\mathrm{t}}, \|\, BA \,\|_{\mathrm{t}} \leq \|\, A \,\|_{\mathrm{t}} \cdot \|\, B \,\|_{\mathrm{t}}$. (See [1] for more details.) We single out the following fact for later use.

Fact A.4 *If $|\phi_1\rangle, |\phi_2\rangle$ are two pure states, and ρ_i is the density matrix of $|\phi_i\rangle$, then*

$$\| \rho_1 - \rho_2 \|_{\mathrm{t}} \;=\; 2\sqrt{1 - |\langle \phi_1 | \phi_2 \rangle|^2}.$$

A.3 The fidelity measure

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

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

where the supremum is taken over all purifications $|\phi_i\rangle$ of ρ_i in the same Hilbert space [10]. We note a few properties of this measure: $0 \leq F(\rho_1, \rho_2) \leq 1$, $F(\rho_1, \rho_2) = 1 \iff \rho_1 = \rho_2$ and if $\rho_1 = |\phi_1\rangle\langle\phi_1|$, then we have $F(\rho_1, \rho_2) = \langle\phi_1| \,\rho_2\, |\phi_1\rangle$. Jozsa [10] proved that the optimum is always achieved when finite dimensional density matrices are considered.

Theorem A.5 (Jozsa) *Let ρ_1, ρ_2 be any two mixed states with support in a finite dimensional Hilbert space $\mathcal{H}$, $\mathcal{K}$ a Hilbert space of dimension at least $\dim(\mathcal{H})$, and $|\phi_1\rangle$ any purification of ρ_1 in $\mathcal{H} \otimes \mathcal{K}$. Then there exists a purification $|\phi_2\rangle \in \mathcal{H} \otimes \mathcal{K}$ of ρ_2 such that $\left|\langle \phi_1 \,|\, \phi_2 \rangle\right|^2 = F(\rho_1, \rho_2)$.*

Jozsa [10] also gave a simple proof (again for the finite dimensional case) of the following remarkable equivalence first established by Uhlmann [23].

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

Using this equivalence, Fuchs and van de Graaf [8] show that the fidelity and the trace measures of distance between density matrices are closely related. They prove:

Theorem A.6 (Fuchs, van de Graaf) *For any two mixed states ρ_1, ρ_2,*

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

A.4 Von Neumann entropy and quantum mutual information

As mentioned earlier, the eigenvalues of a density matrix are all real, non-negative and sum up to one. Thus, they induce a probability distribution on the corresponding eigenvectors. Since the eigenvectors are all orthogonal, this is essentially a *classical* distribution. Every mixed state with the same density matrix is physically equivalent to such a canonical classical distribution. It is thus natural to define the entropy of a mixed state as the *Shannon entropy* of this distribution. Formally, the *von Neumann entropy* $S(\rho)$ of a density matrix ρ is defined as $S(\rho) = -\sum_i \lambda_i \log \lambda_i$, where $\{\lambda_i\}$ is the multi-set of all the eigenvalues of ρ. More compactly, $S(\rho) = -\mathrm{Tr}\, \rho \log \rho$.

Not all properties of classical Shannon entropy carry over to the quantum case. For example it is quite possible that $S(XY) < S(X)$ as can be seen by considering the pure state $\frac{1}{\sqrt{2}}(|0\rangle_X |0\rangle_Y + |1\rangle_X |1\rangle_Y)$. Nonetheless, some of the classical properties do carry over, e.g., $S(X) \geq 0$, $S(X)$ is concave and $S(XY) \leq S(X) + S(Y)$. A property of interest to us is the following, which also generalizes a classical assertion.

Fact A.7 *Suppose a quantum system A is in mixed state $\{p_i, |i\rangle\}$, where $\{|i\rangle\}$ are orthogonal, and σ_i are density matrices for another system B, then $S(\sum_i p_i \, |i\rangle\langle i| \otimes \sigma_i) \;=\; H(A) + \sum_i p_i \, S(\sigma_i)$.*

The density matrix corresponding to a mixed state with superpositions drawn from a Hilbert space $\mathcal{H}$ is said to have *support* in $\mathcal{H}$. A density matrix with support in a Hilbert space of dimension d, has d eigenvalues, hence the entropy of any such distribution is at most $\log d$. A pure-state has zero entropy. Measuring a pure-state may result in a non-trivial mixture and positive entropy. In general, orthogonal measurements increase the entropy. For a comprehensive introduction to this concept and its properties see, for instance, [21].

We define the "mutual information" $I(X : Y)$ of two disjoint systems X, Y in analogy with classical mutual information: $I(X : Y) = S(X) + S(Y) - S(XY)$, where XY is density matrix of the system that includes the qubits of both systems. Again, not all properties of classical mutual information carry over to the quantum case. For example, it is not true in general that $I(X : Y) \leq S(X)$. Nonetheless, some of the intuition we have about mutual information still applies. Equation (1) still holds, as follows immediately from the definition. Equation (2) also continues to be true, but its proof is much more involved. It is in fact equivalent to the *strong sub-additivity property* of von Neumann entropy. An important consequence of this property is that local measurements can only decrease the amount of mutual information. A special case of this is the classic Holevo theorem [9] from quantum information theory, which bounds the amount of information we can extract from a quantum encoding of classical bits.

Theorem A.8 (Holevo) *Let $x \mapsto \sigma_x$ be any quantum encoding of bit strings into density matrices. let X be a random variable with a distribution given by $\mathrm{Prob}(X = x) = p_x$, let Q be the quantum encoding of X according to this map, and let $\sigma = \sum_x p_x \sigma_x$. If Y is any random variable obtained by performing a measurement on the encoding, then*

$$I(X : Y) \;\leq\; I(X : Q) \;=\; S(\sigma) - \sum_x p_x S(\sigma_x).$$

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 with respect to a classical event.