A lower bound for bounded round quantum communication complexity of set disjointness

Rahul Jain* Jaikumar Radhakrishnan* Pranab Sen*

Abstract

We show lower bounds in the multi-party quantum communication complexity model. In this model, there are t parties where the ith party has input $X_i \subseteq [n]$. These parties communicate with each other by transmitting qubits to determine with high probability the value of some function F of their combined input $(X_1, X_2, \ldots, X_t)$. We consider the class of functions whose value depends only on the intersection of $X_1, X_2, \ldots, X_t$; that is, for each F in this class there is an $f_F : 2^{[n]} \to \{0, 1\}$, such that

$$F(X_1, X_2, \ldots, X_t) = f_F(X_1 \cap X_2 \cap \ldots \cap X_t).$$

We show that the t-party k-round communication complexity of F is $\Omega(s_m(f_F)/(k^2))$, where $s_m(f_F)$ stands for the 'monotone sensitivity of f_F' and is defined by

$$s_m(f_F) \stackrel{\Delta}{=} \max_{S \subseteq [n]} |\{i : f_F(S \cup \{i\}) \neq f_F(S)\}|.$$

For two-party quantum communication protocols for the *set disjointness problem*, this implies that the two parties must exchange $\Omega(n/k^2)$ qubits. An upper bound of $O(n/k)$ can be derived from the $O(\sqrt{n})$ upper bound due to Aaronson and Ambainis (see also [BCW98] and [HdW02]). For $k = 1$, our lower bound matches the $\Omega(n)$ lower bound observed by Buhrman and de Wolf [BdW01] (based on a result of Nayak [Nay99]), and for $2 \leq k \ll n^{1/4}$, improves the lower bound of $\Omega(\sqrt{n})$ shown by Razborov [Raz02]. (For protocols with no restrictions on the number of rounds, we can conclude that the two parties must exchange $\Omega(n^{1/3})$ qubits. This, however, falls short of the optimal $\Omega(\sqrt{n})$ lower bound shown by Razborov [Raz02].)

Our result is obtained by adapting to the quantum setting the elegant *information-theoretic* arguments of Bar-Yossef, Jayram, Kumar and Sivakumar [BJKS02]. Using this method we can show similar lower bounds for the $\mathcal{L}_\infty$ function considered in [BJKS02].

1 Introduction

Classical communication complexity: The communication complexity model of Yao [Yao79] provides an abstract setting for studying the communication required for computing a function whose inputs are distributed between several parties. In its most widely studied version, there are two parties, Alice and Bob with inputs $X_A, X_B \subseteq [n]$, who exchange messages based on a fixed protocol in order to determine the value of some function $F(X_A, X_B)$. The goal is to design a protocol so that the parties need to exchange as few bits as possible. This model of communication is relatively well-understood (see the book of Kushilevitz

*School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India. Email: {rahulj, jaikumar, pranab}@tcs.tifr.res.in. Rahul Jain was supported partially by the Kanwal Rekhi Career Development Scholarship.

and Nisan [NK97]) both in the deterministic and the randomized setting. In this paper, we will be interested in the randomized setting, where the parties are allowed to err with some small probability (say at most $\frac{1}{3}$). Tight lower bounds are known for several functions, in this model, for example, the equality function $X_A \overset{?}{=} X_B$ [Yao79, LS81], the set disjointness function $X_A \cap X_B \overset{?}{=} \emptyset$ [KS92, Raz92] and the inner-product function $|X_A \cap X_B|$ (mod 2) [CG88].

Quantum communication complexity: The two-party quantum communication model (see Section 2.1) was introduced by Yao [Yao93], in order to investigate if communication costs for computing functions distributively reduces significantly when the parties are allowed to exchange qubits and perform quantum operations locally. Since then, there has been a flurry of results in this model. We will be mainly interested in the bounded error version of this model, where the two parties are allowed to err with some small probability (say at most $\frac{1}{3}$). It was observed early that for the equality and the inner-product functions the quantum model does not provide any significant savings: the complexity of the equality function is still $\Theta(\log n)$ [Kre95] and the complexity of the inner-product function is still $\Theta(n)$ [Kre95, CvDNT98].

The set disjointness function: For the set disjointness function, however, quantum protocols were found to be strictly more powerful than their classical randomized counterparts. Since the communication complexity of the set disjointness function is central to the work presented in this paper, we describe its history in greater detail. In the bounded error classical setting Babai, Frankl and Simon [BFS86] showed a lower bound of $\Omega(\sqrt{n})$. This was improved to an $\Omega(n)$ lower bound by Kalyanasundaram and Schnitger [KS92]; their proof was simplified by Razborov [Raz92]. There is a straightforward protocol with $n+1$ bits of communication where Alice sends her entire input to Bob, who computes the answer and returns it to Alice. Interest in the communication complexity of several problems related to the set disjointness function has been revived recently because of their connection to showing lower bounds in the classical datastream model [AMS99, FKS02, GGI$^+$02, Ind00, GMMO00, JKS03, SS02]. One of these problem is the $\mathcal{L}_\infty$ promise problem: Alice and Bob are given inputs $X_A, X_B \in \{0, 1, \ldots, m\}^n$, with the promise that either for all $i \in [n]$, $|X_A[i] - X_B[i]| \leq 1$ or there exists an $i \in [n]$, such that $|X[i] - Y[i]| = m$; they must communicate in order to distinguish between these two types of inputs. For this problem, Saks and Sun [SS02] showed a lower bound of $\Omega(n/m^2)$ in a restricted model; their lower bound was strengthened by Bar-Yossef, Jayram, Kumar and Sivakumar [BJKS02], who obtained the same lower bound without any restrictions.

In the quantum setting, the set disjointness function was first addressed by Buhrman, Cleve and Wigderson [BCW98], who showed that there is a protocol for this problem with $O(\sqrt{n}\log n)$ bits of communication. This bound was improved to $O(\sqrt{n}c^{\log^* n})$, where c is a small constant, by Hoyer and de Wolf [HdW02], and recently to $O(\sqrt{n})$ by Aaronson and Ambainis [AA03]. By a result of Razborov [Raz02] this last bound is optimal.

Multi-party classical communication complexity: There are several ways to generalize the two-party model to the multi-party model. In this paper, we will consider the version where there are t parties $P_1, P_2, \ldots, P_t$ with respective inputs $X_1, X_2, \ldots, X_t \subseteq [n]$. In each round of communication some party sends a message to another party. The party who receives the last message can determine the desired value $F(X_1, X_2, \ldots, X_t)$ based on his current state at that point. Recently, because of its connection to the problem of computing *frequency moments* in the data stream model [AMS99], the following *promise set disjointness* problem has been studied. Here, the parties are required to distinguish between two extreme types of inputs: in the first type, $X_1, X_2, \ldots, X_t$ are pairwise disjoint; in the second type, $X_1, X_2, \ldots, X_t$

have exactly one element in common but are otherwise disjoint. For this problem, Chakrabarti, Khot and Sun [CKS03] show a lower bound of $\Omega(n/(t \log t))$, improving an earlier $\Omega(n/t^2)$ lower bounds of Bar-Yossef, Jayram, Kumar and Sivakumar [BJKS02] and an $\Omega(n/t^4)$ lower bound of Alon, Matias and Szegedy [AMS99]. A slight variant of this problem, called the approximate set disjointness problem, was considered by Nisan [Nis02]; the lower bounds mentioned above apply to Nisan's version as well. The multi-party quantum communication complexity of these problems has not been considered before this work.

1.1 Our results

The upper and lower bounds on the two-party quantum communication complexity of the set disjointness function are tight up to constant factors, if there are no restrictions imposed on the number of rounds (i.e. the number of messages) in the protocol. The best upper bound uses $O(\sqrt{n})$ rounds of communication, and from it one can derive a k-round protocol where the parties exchange a total of at most $O(n/k)$ qubits. For $k = 1$, Buhrman and de Wolf [BdW01] observed that the lower bound of $\Omega(n)$ follows from the results of Nayak [Nay99] for the index-function problem. For $k \geq 2$, Klauck, Nayak, Ta-Shma and Zuckerman [KNTZ01] showed a lower bound of $n^{1/k}$, but this is subsumed by Razborov's [Raz02] lower bound of $\Omega(\sqrt{n})$ which holds even if there is no restriction on the number of rounds. However, for small k, Razborov's lower bound is far from the best upper bound known, namely $O(n/k)$. Our first result, gives lower bounds for the two-party k-round communication complexity that comes closer to the upper bound.

Result 1 *The two-party k-round quantum communication complexity of the set disjointness function is $\Omega(n/k^2)$.*

In fact, this lower bound holds even if the protocol is only required to distinguish between disjoint sets and sets with exactly one element in common. Using easy reductions one can conclude that a similar lower bound holds for several other functions. A function F is said to be *set disjointness-like* if its value depends only on the intersection of X_A, X_B; that is, there is an $f_F : 2^{[n]} \rightarrow \{0,1\}$, such that $F(X_A, X_B) = f_F(X_A \cap X_B)$. We obtain a non-trivial lower on the communication complexity of such functions F, if the underlying function f_F has high *monotone sensitivity:* $s_m(f_F) \stackrel{\Delta}{=} \max_{S \subseteq [n]} |\{i : f_F(S \cup \{i\}) \neq f_F(S)\}|$.

Result 1': The two-party k-round quantum communication complexity of the a set disjointness-like function F is $\Omega(s_m(f_F)/(k^2))$.

For the $\mathcal{L}_\infty$ promise problem we get the following.

Result 2 *The two-party k-round quantum communication complexity of the $\mathcal{L}_\infty$ promise problem is $\Omega(n/(k^3 m^{(k+1)}))$.*

We define a model for multi-party quantum communication complexity and show the following[1].

Result 3 *The t-party k-round quantum communication complexity of the promise set disjointness problem is $\Omega(n/k^2)$. [This lower bound also holds for Nisan's approximate set disjointness problem.]*

All our lower bounds hold even if the parties start with arbitrary prior entanglement that is independent of the inputs.

[1] Our lower bound appears to contradict the $\tilde{O}(n/t)$ upper bound of [BJKS02]. This is because that upper bound is in the simultaneous message model, whereas in our definition of quantum protocols one is required to pass fixed length messages from one party to another.

1.2 Techniques used

The original lower bounds for the set disjointness problem in the classical setting are based on deep analyses of the communication matrix and can be said to be based on the *discrepancy method* [Cha00]. Razborov's recent $\Omega(\sqrt{n})$ lower bound for quantum protocols also uses the discrepancy method. The discrepancy method for quantum protocols was formulated explicitly by Kremer [Kre95] (see also Klauck [Kla01] and Yao [Yao93]), but Razborov's proof extends it substantially by developing interesting and powerful tools based on the spectral theory of matrices.

Recently, however, Bar-Yossef et al. [BJKS02] proposed an information-theoretic approach for studying set disjointness-like problems in the classical setting. Using a refinement of the notion of information of communication protocol originally defined by Chakrabarti, Shi, Wirth and Yao [CSWCCY01], they showed that a linear lower bound for the set disjointness problem follows from $\Omega(1)$ lower bound on a certain information cost of a two-party communication protocol computing the AND of just two bits! Their work provided a compelling and beautiful illustration of information-theoretic tools in the analysis of communication protocols.

We adapt their approach to the quantum setting. In order to bring out the contribution of this paper more clearly, we will now informally describe the information-theoretic argument underlying their proof and discuss how we adapt them to the quantum setting. The argument has two parts: in the first part, using a direct-sum argument for information from Bar-Yossef et al. [BJKS02], one reduces the set disjointness problem to a communication problem associated with the AND of two bits (one with Alice and one with Bob); in the second part, one shows that this problem on two bits is hard.

The information cost approach: The first part of the argument is based on the notion of information cost of communication protocols, defined (by [CSWCCY01]) to be the mutual information between the inputs (which are assumed to come from some distribution) and the transcript of the protocol. Bar-Yossef et al. [BJKS02] examine the information cost of the protocol for several distributions. Let the number of bits transmitted by the protocol be c. Then, the information cost is also bounded by c for each distribution.

At this point it will be convenient to view the inputs of Alice and Bob as elements of $\{0, 1\}$ and the set disjointness function as $\bigvee_{i=1}^{n} X_A[i] \wedge X_B[i]$. A typical distribution considered by Bar-Yossef et al. is defined as follows. For each i, independently, one party is given the input 0 and the other party is given a random bit. Using the sub-additivity property of mutual information, one concludes that the sum over i of the mutual information between the transcript and the input $X_A[i]$ is bounded by c; a similar statement holds for Bob's inputs. It is then not hard to argue using a standard averaging argument that there is an i and a product distribution D^*, for inputs $(X_A[j], X_B[j] : j \neq i)$ such that the following conditions hold:

- For all $j \neq i$, $X_A[j] \wedge X_B[j] = 0$ (with probability 1).
- If $X_A[i]$ is set to zero and $X_B[i]$ is chosen at random (and the remaining bits are chosen according to the product distribution D^*), then the mutual information between the transcript and $X_B[i]$ is at most $2c/n$; similarly, if $X_B[i]$ is set to 0 and $X_A[i]$ is chosen at random (an the remaining bits are chosen according to the product distribution D^*), then the mutual information between the transcript and $X_A[i]$ is at most $2c/n$.

From the first condition, by viewing $(X_A[j], X_B[j] : j \neq i)$ as private random bits of the two-parties, we obtain from the protocol for set disjointness a protocol that computes the AND of the two bits $X_A[i]$ and $X_B[i]$. The stage is thus set for analysing the information cost of computing the AND function: a lower bound of ϵ on this quantity translates to a lower bound of $\Omega(\epsilon n)$ on the communication complexity of the set disjointness function.

In order to implement this programme in the quantum setting, one has to define a notion of information cost for quantum protocols. It is not immediately clear how this can be done, because quantum operations are notorious for destroying the states on which they act; in particular, it is not reasonable to expect that the complete transcript of all messages is part of the final global state of the algorithm. Even if the complete transcript is available in the final global state of the algorithm, it may not contain any information about the inputs of either party. If the parties are allowed prior entanglement, then using quantum teleportation, one can implement any protocol such that the messages are classical and completely random. So, the transcript will just be a random string of length c independent of the actual inputs!

The definition of information loss for quantum protocols: We address these difficulties by considering the information carried by each message separately. As observed above messages may themselves carry no information, so we examine the information carried in the message by including the context in which it is received. For example, consider a protocol for the AND problem. Fix some distribution for the inputs of Alice and Bob. We account for the information carried in a message sent by Alice to Bob, by considering the mutual information between Alice's input and the *entire* state of Bob, including the message just received. The information loss (we use the term loss instead of cost) of the protocol (for the given distribution) is defined to be the sum of these quantities (both for Alice and Bob) taken over all rounds. With this definition, the arguments of [BJKS02] are easily carried over to the quantum setting. We can then conclude that if the information loss of computing the AND of two bits is ϵ then the communication complexity of the set disjointness function is $\Omega(n\epsilon)$.

We have arrived at the second part of the programme, that is, to show non-trivial lower bounds on the information loss of computing the AND of two bits. In the original argument of [BJKS02] this was achieved by a direct argument using certain distance measures between probability distributions. Since, we are working with a different notion of information loss, this argument does not appear to be immediately applicable in our case; so, instead of reviewing it, we will now directly describe our argument. We are given a quantum communication protocol for computing the AND function. We consider two kinds of inputs: first, Alice has 0 and Bob has a random bit; second, Bob has a 0 and Alice has a random bit. Suppose we are given that for such distributions at no stage does a receiver of a message gain more than ϵ bits of information about the input of the sender. We wish to show that if ϵ is very small, then this leads to a contradiction. Our argument can be understood at an intuitive level in the framework of round-elimination. Suppose, Alice sends the first message. We know that her message does not deliver much information about her input to Bob, that is, the combined state of Bob at the end of the first round is essentially the same when the Alice's input is 0 and when Alice's input is 1. So, Alice might as well send exactly the same message in the two cases, and incur a small error in the correctness. That is, no matter what her actual input is, Alice sends her first message assuming that her input is 0. Since we allow prior entanglement, we can eliminate this round of Alice, and obtain a protocol with one fewer round of communication. Now, it is Bob's turn. Our hypothesis says that his second message does not deliver much information about his input to Alice, when her input is 0. But the modified protocol so far has proceeded as if Alice's input is 0 (even though her actual input might be something else). We can thus eliminate Bob's first message as well. If ϵ is small, then the increase in error probability on account of this manoeuvre is also small. Proceeding in this manner we eliminate all rounds. But it is obvious that if the parties exchange no messages they cannot compute any non-trivial function unless one allows huge error probability. Since, there are at most k rounds of communication, this gives us a lower bound of the form $\epsilon \geq \epsilon(k)$. Using these ideas one can show an

$\Omega(n/k^2)$ lower bound on two-party quantum communication complexity of the set disjointness function.

There are two aspects of our proof that require further comment.

Local transition: Recall the argument used above to eliminate Alice's first message. We know that Bob's state is roughly the same even if Alice generates her message assuming that her input is 0. However, this does not immediately imply that the error probability of the protocol is not changed much. The final answer is not just a function of Bob's state but the combined state of Alice and Bob. In particular, even though the Bob's state is similar after the first round for the two inputs of Alice, his work qubits might be entangled with Alice's qubits differently in the two cases. This problem arises often in round elimination arguments and by now standard solutions exist for it by considering the *fidelity* [Joz94] between quantum states. This allows Alice to perform a *local transition* [KNTZ01] on her work qubits, in order to restore them to the correct state should she discover later that her actual input is different from what was assumed while generating her first message to Bob.

A paradox?: In our notion of information loss of quantum protocols it is important that the parties start in a *pure* global state. In fact, this notion is unsuited for classical randomized communication complexity. Consider the following classical protocol for computing the AND of two bits (a, b). Alice sends Bob a random bit r, retaining a copy of r if and only if $a = 1$. Bob sends Alice $r \oplus b$; if $a = 1$, Alice can recover b using the copy of r she has and determine $a \wedge b$. Now, clearly, the first message does not deliver any information to Bob. Furthermore, when Alice has a 0, Bob's message delivers no information about his inputs, because Alice does not retain a copy of r in this case. So, according to our definition this protocol has zero information loss for both the distributions considered above. Yet, the protocol computes the AND correctly! Interestingly, no such quantum protocol starting with a pure global state is possible.

1.3 The rest of the paper

In the next section, we give some the definition and notation used in the rest of the paper. In Section 3, we prove Result 1. Result 2 and Result 3 also follow using similar arguments, but their proofs are not included in this abstract.

2 Preliminaries

2.1 Quantum communication

We define t-party quantum communication protocols which are a natural extension of the two-party quantum communication protocols as defined by Yao [Yao93]. Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \cdots \mathcal{X}_t \rightarrow \mathcal{Z}$ be a function. There are t parties, $P_1, P_2, \cdots, P_t$, who hold qubits. When the communication protocol Π starts, P_i holds $|x_i\rangle$ where $x_i \in \mathcal{X}_i$ together with some ancilla qubits in the state $|0\rangle$. These parties may also share an input independent prior entanglement (say $|\psi\rangle$). Different parties possess different qubits of $|\psi\rangle$. The parties take turns to communicate to compute $f(x_1, x_2, \cdots, x_t)$. Suppose it is P_1's turn to communicate to P_2. P_1 can make an arbitrary unitary transformation on her qubits and then send one or more qubits to P_2. The number of qubits send is predetermined and is independent of the input x_1. Sending qubits does not change the overall superposition, but rather changes the ownership of the qubits, allowing P_2 to apply her next unitary transformation on her original qubits plus the newly received qubits. At the end of the protocol, the last recipient of a message performs a von Neumann measurement in the computational basis of some qubits in her possession (the 'answer qubits') to output an answer $\Pi(x_1, x_2, \cdots, x_t)$. We say that protocol Π computes f with δ-error in the worst case (or simply with error δ), if $\max_{x_1,x_2,\cdots,x_t} \Pr[\Pi(x_1, x_2, \cdots, x_t) \neq f(x_1, x_2, \cdots, x_t)] \leq \delta$. The communication cost of Π is the number of qubits exchanged in Π between all the parties. The k-round δ-error quantum communication complexity of f, denoted by $Q^k_\delta(f)$, is the communication cost of the best k-round δ-error quantum protocol with prior entanglement for f. When δ is omitted, we mean that $\delta = \frac{1}{3}$.

We require that the parties make a 'safe' copy of their inputs (using, for example, CNOT gates) before beginning protocol Π. This is possible without loss of generality because the inputs are in computational basis states. Thus, the input qubits of the parties are never sent as messages, their state remains unchanged throughout the execution of Π, and they are never measured i.e. some work qubits are measured to determine the result $\Pi(x_1, x_2, \cdots, x_t)$. We call such protocols *safe*, and henceforth, we will assume that all our protocols are safe.

Suppose A, B, C are three disjoint finite dimensional quantum systems having some joint density matrix ρ. Let ρ_A be the reduced density matrix of A. Then $S(A) \stackrel{\Delta}{=} S(\rho) \stackrel{\Delta}{=} -\mathrm{Tr}\, \rho \log \rho$ is the *von Neumann entropy* of A. The *mutual information* of A and B is defined as $I(A : B) \stackrel{\Delta}{=} S(A)+S(B)-S(AB)$. The *conditional mutual information* of A and B given C is defined as $I((A : B) \mid C) \stackrel{\Delta}{=} S(AC)+S(BC)-S(C)-S(ABC)$. If C is a classical random variable taking the classical value $|c\rangle$ with probability p_c, it is easy to see that $I((A : B) \mid C) = \sum_c p_c I(A^c : B^c)$, where $(AB)^c$ denotes the joint density matrix of A and B when $C = |c\rangle$. We also write $I(A : B \mid C = c)$ for $I(A^c : B^c)$.

Fact 1 (see [CvDNT98]) *Let* Alice *have a classical random variable X. Suppose* Alice *and* Bob *share a pure state on some qubits (a prior entanglement) independent of X. Initially* Bob*'s qubits have no information about X. Now let* Alice *and* Bob *run a quantum communication protocol, at the end of which* Bob*'s qubits possess m bits of information about X. Then,* Alice *has to totally send at least $m/2$ qubits to* Bob*.*

Fact 2 (Sub-additivity of information,see [KNTZ01]) *Let D be a classical random variable. Let $X_1, \ldots, X_n$ be classical random variables which are independent given D. Let M be a quantum encoding of $X \stackrel{\Delta}{=} X_1 \ldots X_n$. Then, $I((X : M) \mid D) \geq \sum_{i=1}^n I((X_i : M) \mid D)$.*

Definition 1 (Trace distance) *Let ρ, σ be density matrices in the same finite dimensional Hilbert space. The trace distance between ρ and σ is defined as follows: $\|\rho - \sigma\|_t \stackrel{\Delta}{=} \mathrm{Tr}\, \sqrt{(\rho - \sigma)^\dagger(\rho - \sigma)}$.*

Definition 2 (Fidelity) *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathcal{H}$. Their fidelity is defined as $B(\rho, \sigma) \stackrel{\Delta}{=} \sup_{\mathcal{K},|\psi\rangle,|\phi\rangle} |\langle\psi|\phi\rangle|$, where $\mathcal{K}$ ranges over all finite dimensional Hilbert spaces and $|\psi\rangle, |\phi\rangle$ range over all purifications of ρ, σ respectively in $\mathcal{H} \otimes \mathcal{K}$.*

Fact 3 (see [AKN98]) *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathcal{H}$. Let $\mathcal{F}$ be a measurement (POVM) on $\mathcal{H}$. Then, $\|\mathcal{F}\rho - \mathcal{F}\sigma\|_1 \leq \|\rho - \sigma\|_t$.*

The following lemmas are derived in the appendix.

Lemma 1 *Let ρ_1, ρ_2 be two density matrices in the same finite dimensional Hilbert space $\mathcal{H}$, $\mathcal{K}$ any Hilbert space of dimension at least the dimension of $\mathcal{H}$, 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 $|\phi'_2\rangle \stackrel{\Delta}{=} (I \otimes U)|\phi_2\rangle$ (I is the identity operator on $\mathcal{H}$) such that*

$$\big\||\phi_1\rangle\langle\phi_1| - |\phi'_2\rangle\langle\phi'_2|\big\|_t \leq 2\sqrt{1 - B(\rho_1, \rho_2)^2} \leq 2\sqrt{2(1 - B(\rho_1, \rho_2))}.$$

Lemma 2 *Suppose X and Q are disjoint quantum systems, where X is a classical random variable uniformly distributed over $\{0, 1\}$ and Q is a quantum encoding $x \to \sigma_x$ of X. Then, $1 - B(\sigma_1, \sigma_2) \leq I(X : Q)$.*

2.2 Conditional information loss

Let D, X_A and X_B be random variables taking values in some finite sets $\mathcal{D}$, $\mathcal{X}_A$ and $\mathcal{X}_B$ respectively. We say that D partitions $X = (X_A, X_B)$ if for all $d \in \mathcal{D}$, X_A and X_B are independent conditioned on the event $D = d$. Given random variables X and D, the random variable $(X, D)^n$ is obtained by taking n independent copies of (X, D). Thus, $(X, D)^n$ takes values in $(\mathcal{X}_A \times \mathcal{X}_B)^n$ which we identify with $\mathcal{X}_A^n \times \mathcal{X}_B^n$. Suppose D partitions X, and $(\mathbf{X}, \mathbf{D}) = (X, D)^n$, then it is easy to verify that $\mathbf{D}$ partitions $\mathbf{X}$.

Definition 3 (Embedding) *For $\mathbf{a} \in \mathcal{A}^n$, $j \in [n]$, and $u \in \mathcal{A}$, let $\mathsf{embed}(\mathbf{a}, j, u)$ be the element of $\mathcal{A}^n$ obtained by replacing $\mathbf{a}[j]$ by u, that is, $\mathsf{embed}(\mathbf{a}, j, u)[i] \stackrel{\Delta}{=} \mathbf{a}[i]$ for $i \neq j$, and $\mathsf{embed}(\mathbf{a}, j, u)[j] \stackrel{\Delta}{=} u$.*

Definition 4 (Collapsing input) *Suppose $F : \mathcal{X}^n \to \mathcal{Z}$. We say that $\mathbf{x} \in X^n$ collapses F to the function $h : \mathcal{X} \to \mathcal{Z}$ if for all $u \in \mathcal{X}$, $F(\mathsf{embed}(\mathbf{x}, j, u)) = h(u)$. We say that a random variable X taking values in $\mathcal{X}^n$ collapses F to h if it is collapses F to h with probability 1.*

Definition 5 (Conditional Information loss) *Let Π be a two-party k-round δ-error quantum protocol for computing $F : \mathcal{X}_A \times \mathcal{X}_B \to \mathcal{Z}$. Let Alice start the protocol and let $A^i B^i$ be the joint state of Alice and Bob just after the ith message has been received. Let $X = (X_A, X_B)$ be random variable taking values in $\mathcal{X}_A \times \mathcal{X}_B$ which is partitioned by the random variable D. Then, the* conditional information loss *of Π under (X, D) is defined by*

$$\mathsf{IL}(\Pi \mid (X, D)) \stackrel{\Delta}{=} \textstyle\sum_{i=1,\ i \text{ odd}}^{k} I(X_A : B^i \mid D) + \sum_{i=1,\ i \text{ even}}^{k} I(X_B : A^i \mid D).$$

The k-round δ-error conditional information loss of F under (X, D), denoted by $\mathsf{IL}_{k,\delta}(F \mid (X, D))$, is the minimum $\mathsf{IL}(\Pi \mid (X, D))$ taken over all k-round δ-error quantum protocols Π for F. [Note that δ bounds the error for all *inputs. In particular, this error bound applies even to inputs not in the support of X.]*

3 Lower bound for set disjointness

Lemma 3 *Let $F : \mathcal{X}_A^n \times \mathcal{X}_B^n \to \mathcal{Z}$. Let X be a random variable taking values in $\mathcal{X} \stackrel{\Delta}{=} \mathcal{X}_A \times \mathcal{X}_B$; suppose X is partitioned by a random variable D taking values in some set $\mathcal{D}$. Let $(\mathbf{X}, \mathbf{D}) = (X, D)^n$. Suppose $\mathbf{X}$ collapses F to the function $h : \mathcal{X}_A \times \mathcal{X}_B \to \mathcal{Z}$. Then, $\mathsf{IL}_{k,\delta}(h \mid (X, D)) \leq \frac{2k}{n} Q_\delta^k(F)$.*

Proof: Suppose Π is a k-round δ-error quantum protocol for F with total communication c. Let us assume that Alice starts the communication. Our goal is to show that there is a k-round δ-error protocol for h with small information loss under (X, D). While analysing Π, we will need to maintain that the combined state of Alice and Bob is pure at all times. However, we will run Π on random inputs drawn from certain product distributions. In such a situation, we will adopt the following convention. We will assume that in addition to the usual input registers IN_A, Alice has another set of registers $\widetilde{\mathsf{IN}}_A$. When we require that Alice's inputs be some random variable X_A, we in fact, start with the following state in the registers $\mathsf{IN}_A\widetilde{\mathsf{IN}}_A$: $\sum_{x \in \mathcal{X}_A} \sqrt{p_x} |x\rangle |x\rangle$, where $p_x \stackrel{\Delta}{=} \Pr[X_A = x]$. Similarly, we simulate Bob's random input X_B in registers IN_B and $\widetilde{\mathsf{IN}}_B$. Then, we run the protocol Π as before with input registers IN_A and IN_B. During this execution no quantum gates are applied to registers $\widetilde{\mathsf{IN}}_A$ and $\widetilde{\mathsf{IN}}_B$. From now on $\mathbf{X}_A$ (similarly $\mathbf{X}_B$) denotes the state of the registers IN_A, which stays constant because the protocol Π is safe. In this revised protocol Π', let A^iB^i denote the state of the entire system immediately after the ith message has been received (note that A^i includes the register $\widetilde{\mathsf{IN}}_A$ and B^1 includes $\widetilde{\mathsf{IN}}_B$). Consider the execution of Π' on input $\mathbf{X} = (\mathbf{X}_A, \mathbf{X}_B)$ conditioned on $\mathbf{D} = \mathbf{d}$; note that under this condition $\mathbf{X}_A$ and $\mathbf{X}_B$ are independent and the convention described above for simulating random inputs applies. Then, we have

$$\forall i, 1 \leq i \leq k, i \text{ odd}, \quad \textstyle\sum_{j=1}^{n} I((\mathbf{X}_A[j] : B^i) \mid \mathbf{D} = \mathbf{d}) \leq I((\mathbf{X}_A : B^i) \mid \mathbf{D} = \mathbf{d}) \leq 2c.$$

The first inequality above follows from Fact 2 because by our definition of $(\mathbf{X}, \mathbf{D})$, $(\mathbf{X}_A[j] : 1 \leq j \leq n)$ are independent random variables when conditioned on $\mathbf{D} = \mathbf{d}$; the second inequality follows from Fact 1.

Averaging over the possibilities for $\mathbf{D}$, we obtain: $\forall i, 1 \leq i \leq k, i$ odd, $\sum_{j=1}^{n} I((\mathbf{X}_A[j] : B^i) \mid \mathbf{D}) \leq 2c$. Similarly, we obtain $\forall i, 1 \leq i \leq k, i$ even, $\sum_{j=1}^{n} I((\mathbf{X}_B[j] : A^i) \mid \mathbf{D}) \leq 2c$. Summing these inequalities over all rounds i, we obtain

$$\textstyle\sum_{j=1}^{n} \left(\sum_{i=1,\, i \text{ odd}}^{k} I(\mathbf{X}_A[j] : B^i \mid \mathbf{D}) + \sum_{i=1,\, i \text{ even}}^{k} I(\mathbf{X}_B[j] : A^i \mid \mathbf{D}) \right) \leq 2ck,$$

which implies:

$$\exists j, 1 \leq j \leq n, \textstyle\sum_{i=1,\, i \text{ odd}}^{k} I((\mathbf{X}_A[j] : B^i) \mid \mathbf{D}) + \sum_{i=1,\, i \text{ even}}^{k} I((\mathbf{X}_B[j] : A^i) \mid \mathbf{D}) \leq \frac{2ck}{n}. \tag{1}$$

Fix a value of j so that the last inequality holds. For $\mathbf{d} \in \mathcal{D}^n$, let

$$I(\mathbf{d}) \triangleq \textstyle\sum_{i=1,\, i \text{ odd}}^{k} I((\mathbf{X}_A[j] : B^i) \mid \mathbf{D} = \mathbf{d}) + \sum_{i=1,\, i \text{ even}}^{k} I((\mathbf{X}_B[j] : A^i) \mid \mathbf{D} = \mathbf{d}). \tag{2}$$

Then, from (1), and the definition of conditional mutual information $\mathrm{E}_{\mathbf{D}}[I(\mathbf{D})] \leq \frac{2ck}{n}$.

We will now obtain a protocol for h by 'embedding' its input as the jth input of Π'. Using a straightforward averaging argument we first fix a value $\hat{\mathbf{d}} \in \mathcal{D}^n$ so that

$$\sum_{d \in \mathcal{D}} \Pr[D = d] I(\mathsf{embed}(\hat{\mathbf{d}}, j, d)) = \mathop{\mathrm{E}}_{D}[I(\mathsf{embed}(\hat{\mathbf{d}}, j, D)) \leq \frac{2ck}{n}. \tag{3}$$

Consider the following protocol Π_h for computing $h(u_A, u_B)$. On input $u_A \in \mathcal{X}_A$, Alice prepares her input registers as follows. In the registers $(\mathsf{IN}_A[\ell], \widetilde{\mathsf{IN}}_A[\ell] : \ell \neq j)$ Alice places the superposition $\sum_{x \in \mathcal{X}^{n-1}} \sqrt{p_x} |x\rangle|x\rangle$, where $p_x = \Pr[(\mathbf{X}_A[\ell] : \ell \neq j) = x \mid \mathbf{D} = \hat{\mathbf{d}}]$; register $\mathsf{IN}_A[j]$ is set to $|u_A\rangle$. On input $u_B \in \mathcal{X}_B$, Bob prepares his input registers in a similar fashion. Then, Alice and Bob apply the protocol Π', treating IN_A and IN_B as input registers. Note that $\widetilde{\mathsf{IN}}_A$ and $\widetilde{\mathsf{IN}}_B$ do not exist in Π_h.

We need to verify that this protocol for computing h has two properties. First, it computes h correctly with high probability. For this, we note that in this protocol, at all times, the state of the registers that were present in the original protocol Π (that is all registers except $\widetilde{\mathsf{IN}}_A$ and $\widetilde{\mathsf{IN}}_B$) is identical to their state when the original protocol Π is run with input $\mathsf{embed}(\mathbf{X}, j, (u_A, u_B))$ conditioned on the event $\mathbf{D} = \hat{\mathbf{d}}$. Since $\mathbf{X}$ collapses F to h, we conclude that Π_h computes $h(u_A, u_B)$ with probability at least $1 - \delta$.

Second, we need verify that $\mathsf{IL}(\Pi_h \mid (X, D))$ is small. We expand the LHS of (3) using the definition (2) of $I(\mathbf{d})$ and show that each term in it is at least the corresponding term in $\mathsf{IL}(\Pi_h \mid (X, D))$. For example, consider the term $I(X_A : B^i \mid (D = d))$ in the definition of $\mathsf{IL}(\Pi_h \mid (X, D))$. Note that the state (X_A, B^i) of Π_h on input X conditioned on $D = d$, is identical to the state obtained from $(\mathbf{X}_A[j], B^i)$ of Π' by omitting the register $\widetilde{\mathsf{IN}}_B[j]$, when Π' is run on input $\mathbf{X}$ conditioned on $\mathbf{D} = \mathsf{embed}(\hat{\mathbf{d}}, j, d)$. It follows from the monotonicity property of information that $I(X_A : B^i \mid (D = d))$ is at most $I(\mathbf{X}_A[j] : B^i \mid (\mathbf{D} = \mathsf{embed}(\hat{\mathbf{d}}, j, d)))$. We can then conclude (details omitted) that $\mathsf{IL}(\Pi_h, (X, D)) \leq \frac{2ck}{n}$. ■

As in [BJKS02], let D be a random variable taking values in $\{A, B\}$ uniformly. Let $\mathcal{X}_A, \mathcal{X}_B = \{0, 1\}$ and $X = (X_A, X_B)$ be a random variable taking values in $\mathcal{X}_A \times \mathcal{X}_B = \{0, 1\}^2$, whose correlation with D is described $\Pr[X = 00 \mid D = A], \Pr[X = 10 \mid D = A], \Pr[X = 00 \mid D = B], \Pr[X = 01 \mid D = B] = \frac{1}{2}$. It is clear that conditioned on $D = A$ and $D = B$, X_A and X_B are independent. Note that X^n collapses DISJ to AND . We now show a lower bound for the conditional information loss of AND under (X, D).

Lemma 4 *Let (X, D) be as above. Let $\epsilon > 0$. Then $\mathsf{IL}_{k,\delta}(\mathrm{AND} \mid (X, D)) \geq \frac{(1-2\epsilon)^2}{4k}$.*

Proof: Let Π be a k-round ϵ-error quantum protocol for AND with $\eta \stackrel{\Delta}{=} \mathsf{IL}(\Pi, (X, D)) = \mathsf{IL}_{k,\delta}(\mathrm{AND} \mid (X, D))$. Consider the situation in Π just after the ith message has been sent. Let m^i denote the qubits of the ith message. Let X_A, X_B denote the random variables corresponding to Alice's and Bob's inputs respectively in Π. Suppose $(X_A, X_B) = (x, y)$. Let $|\phi^i_{xy}\rangle$ be the global state vector of Alice's and Bob's qubits, and let A^i, B^i denote Alice's qubits and Bob's qubits respectively at this point in time. Suppose m^i is sent from Alice to Bob. Then $B^i = B^{i-1} \cup m^i$ and $A^{i-1} = A^i \cup m^i$, and the unions are over disjoint sets of qubits. Let $\alpha^i_{xy}, \beta^i_{xy}$ denote the reduced density matrices of A^i, B^i in the state $|\phi^i_{xy}\rangle$. Define $c^i_B \stackrel{\Delta}{=} I(X : B^i \mid D)$ and $c^i_A \stackrel{\Delta}{=} I(Y : A^i \mid D)$. Let $U^{A^{i-1}}$ denote the unitary transformation that Alice applies to A^{i-1} after receiving the $(i-1)$st message from Bob, in order to prepare the ith message. Then, $|\phi^i_{xy}\rangle = U^{A^{i-1}}|\phi^{i-1}_{xy}\rangle$. By Lemma 2, $1 - B(\beta^i_{00}, \beta^i_{10}) \leq I((X : B^i) \mid Y = 0) \leq 2c^i_B$ and $1 - B(\alpha^{i-1}_{00}, \alpha^{i-1}_{01}) \leq I((Y : A^{i-1}) \mid X = 0) \leq 2c^{i-1}_A$. To keep our notation concise, for state vectors $|\phi\rangle$ and $|\psi\rangle$ we write $\||\phi\rangle - |\psi\rangle\|_t$ instead of $\||\phi\rangle\langle\phi| - |\psi\rangle\langle\psi|\|_t$. By Lemma 1, there exist unitary transformations $V^{B^{i-1}}_{00\to 01}$ acting on B^{i-1} and $V^{A^i}_{00\to 10}$ acting on A^i such that

$$\left\|V^{B^{i-1}}_{00\to 01}|\phi^{i-1}_{00}\rangle - |\phi^{i-1}_{01}\rangle\right\|_t \leq 4(c^{i-1}_A)^{1/2} \;\text{ and }\; \left\|V^{A^i}_{00\to 10}|\phi^i\rangle_{00} - |\phi^i_{10}\rangle\right\|_t \leq 4(c^i_B)^{1/2}. \tag{4}$$

Define $\delta_{i-1} \stackrel{\Delta}{=} \left\|V^{B^{i-1}}_{00\to 01}|\phi^{i-1}_{10}\rangle - |\phi^{i-1}_{11}\rangle\right\|_t$. Using the unitary invariance and triangle inequality of the trace norm, the fact that unitary transformations on disjoint sets of qubits commute, and (4),

$$\begin{aligned}
\delta_i &= \left\|V^{A^i}_{00\to 10}|\phi^i_{01}\rangle - |\phi^i_{11}\rangle\right\|_t = \left\|(V^{B^{i-1}}_{00\to 01})^{-1}V^{A^i}_{00\to 10}|\phi^i_{01}\rangle - (V^{B^{i-1}}_{00\to 01})^{-1}|\phi^i_{11}\rangle)\right\|_t \\
&= \left\|V^{A^i}_{00\to 10}(V^{B^{i-1}}_{00\to 01})^{-1}|\phi^i_{01}\rangle - (V^{B^{i-1}}_{00\to 01})^{-1}|\phi^i_{11}\rangle\right\|_t \\
&= \left\|V^{A^i}_{00\to 10}(V^{B^{i-1}}_{00\to 01})^{-1}U^{A^{i-1}}|\phi^{i-1}_{01}\rangle - (V^{B^{i-1}}_{00\to 01})^{-1}U^{A^{i-1}}|\phi^{i-1}_{11}\rangle\right\|_t \\
&= \left\|V^{A^i}_{00\to 10}U^{A^{i-1}}(V^{B^{i-1}}_{00\to 01})^{-1}|\phi^{i-1}_{01}\rangle - U^{A^{i-1}}(V^{B^{i-1}}_{00\to 01})^{-1}|\phi^{i-1}_{11}\rangle\right\|_t \\
&\leq \left\|V^{A^i}_{00\to 10}U^{A^{i-1}}(V^{B^{i-1}}_{00\to 01})^{-1}|\phi^{i-1}_{01}\rangle - V^{A^i}_{00\to 10}U^{A^{i-1}}|\phi^{i-1}_{00}\rangle\right\|_t + \\
&\quad \left\|V^{A^i}_{00\to 10}U^{A^{i-1}}|\phi^{i-1}_{00}\rangle - U^{A^{i-1}}|\phi^{i-1}_{10}\rangle\right\|_t + \left\|U^{A^{i-1}}|\phi^{i-1}_{10}\rangle - U^{A^{i-1}}(V^{B^{i-1}}_{00\to 01})^{-1}|\phi^{i-1}_{11}\rangle\right\|_t \\
&= \left\|(V^{B^{i-1}}_{00\to 01})^{-1}|\phi^{i-1}_{01}\rangle - |\phi^{i-1}_{00}\rangle\right\|_t + \left\|V^{A^i}_{00\to 10}|\phi^i_{00}\rangle - |\phi^i_{10}\rangle\right\|_t + \left\||\phi^{i-1}_{10}\rangle - (V^{B^{i-1}}_{00\to 01})^{-1}|\phi^{i-1}_{11}\rangle\right\|_t \\
&= \left\||\phi^{i-1}_{01}\rangle - V^{B^{i-1}}_{00\to 01}|\phi^{i-1}_{00}\rangle\right\|_t + \left\|V^{A^i}_{00\to 10}|\phi^i_{00}\rangle - |\phi^i_{10}\rangle\right\|_t + \left\|V^{B^{i-1}}_{00\to 01}|\phi^{i-1}_{10}\rangle - |\phi^{i-1}_{11}\rangle\right\|_t \\
&\leq 4(c^{i-1}_A)^{1/2} + 4(c^i_B)^{1/2} + \delta_{i-1}.
\end{aligned}$$

It is easy to check that $\delta_0 = 0$. Hence using concavity of the fourth root function, $\delta_k \leq 4k\left(\frac{\eta}{k}\right)^{1/2}$. Now a correct k-round ϵ-error protocol for AND must have (from Fact 3 and using the fact that a local unitary transformation does not affect the density matrix of the remote system), $\delta_k \geq \left\|\beta^k_{10} - \beta^k_{11}\right\|_t \geq 2 - 4\epsilon$. Hence, $\eta \geq \frac{(1-2\epsilon)^2}{4k}$. ■

The following is now immediate from Lemma 3 and Lemma 4.

Theorem 1 *Any two-party k-round bounded error quantum protocol for the set disjointness problem needs to have communication cost at least $\Omega\left(\frac{n}{k^2}\right)$.*

Corollary 1 *Any two-party bounded error quantum protocol for the set disjointness problem needs to have communication cost at least $\Omega\left(n^{1/3}\right)$.*

Acknowledgements

Our original proof gave a lower bound of $\Omega(n/k^4)$ for set disjointness. We later improved it to $\Omega(n/k^2)$ using an inequality from [Lin91]. Hartmut Klauck independently pointed out to us that similar improvements can also be obtained using an inequality from [DCHR78]. We thank him for sharing with us his insights and pointing reference [DCHR78] to us.

References

[AA03] S. Aaronson and A. Ambainis. Quantum search of spatial regions. Manuscript at quant-ph/0303041, 2003.

[AKN98] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 20–30, 1998. Also quant-ph/9806029.

[AMS99] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 63–68, 1998. Also quant-ph/9702040.

[BdW01] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th IEEE Conference on Computational Complexity*, pages 120–130, 2001.

[BFS86] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of the 30th ACM Symposium on Theory of Computing*, pages 337–347, 1986.

[BJKS02] Z. Bar-Yossef, T. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, 2002.

[CG88] B Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal of Computing*, 17(2):230–261, 1988.

[Cha00] B. Chazelle. *The Discrepancy Method*. Cambridge University Press, 2000.

[CKS03] A. Chakrabarti, S. Khot, and X. Sun. Near-optimal lower bounds on the multiparty communication complexity of set-disjointness. In *Proceedings of the 18th Annual IEEE Conference on Computational Complexity*, 2003. To appear.

[CSWCCY01] A. Chakrabarti, Y. Shi, A. Wirth, and A. C-C Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001.

[CvDNT98] 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 NASA International Conference on Quantum Computing and Quantum Communications*, Lecture Notes in Computer Science, vol. 1509, pages 61–74. Springer-Verlag, 1998. Also quant-ph/9708019.

[DCHR78] D. Dacunha-Castelle, H. Heyer, and B. Roynette. *Ecole d'Eté de Probabilités de Saint-Flour VII*. Lecture Notes in Mathematics, vol. 678. Springer-Verlag, 1978.

[FC95] C. Fuchs and C. Caves. Mathematical techniques for quantum communication theory. *Open Systems and Information Dynamics*, 3(3):345–356, 1995. Also quant-ph/9604001.

[FKS02] J. Feigenbaum, S. Kannan, and M. Strauss. An approximate l^1-difference algorithm for massive data streams. *SIAM Journal of Computing*, 32:131–151, 2002.

[GGI$^+$02] A. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. Strauss. Fast small space algorithms for approximate histogram maintenance. In *Proceedings of the 34th Annual ACM Symposium Theory of Computing*, pages 389–398, 2002.

[GMMO00] S. Guha, N. Mishra, R. Motwani, and L. O'Callaghan. Clustering data streams. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 359–366, 2000.

[HdW02] P. Hoyer and R. de Wolf. Improved quantum communication complexity bounds for disjointness and equality. In *Symposium on Theoretical Aspects of Computer Science*, pages 299–310, 2002. Also quant-ph/0109068.

[Ind00] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computations. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 189–197, 2000.

[JKS03] T.S. Jayram, R. Kumar, and D. Sivakumar. Two applications of information complexity. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, 2003. To appear.

[Joz94] R. Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41(12):2315–2323, 1994.

[JRS02] R. Jain, J. Radhakrishnan, and P. Sen. The quantum communication complexity of the pointer chasing problem: the bit version. In *Proceedings of the 22nd Foundations of Software Technology and Theoretical Computer Science Conference*, pages 218–229, 2002.

[Kla01] H. Klauck. Lower bounds for quantum communication complexity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 288–297, 2001. Also at quant-ph/0106160.

[KNTZ01] H. Klauck, A. Nayak, A. Ta-Shma, and D. Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 124–133, 2001.

[Kre95] I. Kremer. Quantum communication. 1995. Master's thesis, Hebrew University, Jerusalem.

[KS92] B. Kalyansundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992.

[Lin91] J Lin. Divergence measures based on shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991.

[LS81] R.J. Lipton and R. Sedgewick. Lower bounds for vlsi. In *Proceedings of the 13th Annual ACM Symposium on Theory of Computing*, pages 300–307, 1981.

[Nay99] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the 40rd Annual IEEE Symposium on Foundations of Computer Science*, pages 369–376, 1999.

[NC00] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[Nis02] N. Nisan. The communication complexity of approximate set packing and covering. In *Proceedings of the 29th ICALP*, pages 868–875, 2002.

[NK97] N. Nisan and E. Kushelevitz. *Communication complexity*. Cambridge University Press, 1997.

[Raz92] A.A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[Raz02] A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya Math*, 6, 2002. In Russian. To appear. English version at quant-ph/0204025.

[SS02] M. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 360–369, 2002.

[Yao79] A. C-C. Yao. Some complexity questions related to distributed computing. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.

[Yao93] A. C-C. Yao. Quantum circuit complexity. In *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–361, 1993.

A Quantum information theory background

In this section we give some basic quantum information-theoretic definitions and facts which will be useful in stating and proving our main results. For an excellent introduction to quantum information theory, see the book by Nielsen and Chuang [NC00].

Suppose P, Q are probability distributions on the same finite sample space $[k]$. Their *total variation distance* is defined as follows: $\|P - Q\|_1 \stackrel{\Delta}{=} \sum_{i \in [k]} |P(i) - Q(i)|$. The quantum generalisation of the total variation distance of a pair of probability distributions is the *trace distance* of a pair of density matrices. Recall that a density matrix over a finite dimensional Hilbert space $\mathcal{H}$ is a unit trace, Hermitian, positive semidefinite linear operator on $\mathcal{H}$.

Definition 6 (Relative Entropy) *If ρ, σ are density matrices in the same Hilbert space, their relative entropy is defined as $S(\rho \| \sigma) \stackrel{\Delta}{=} \mathrm{Tr}\,(\rho(\log \rho - \log \sigma))$.*

Let ρ be a density matrix in a finite dimensional Hilbert space $\mathcal{H}$. Suppose $\mathcal{F}$ is a measurement (POVM) on $\mathcal{H}$. Then $\mathcal{F}\rho$ denotes the probability distribution on the (finite number of) possible outcomes of $\mathcal{F}$ got by performing the measurement $\mathcal{F}$ on the state ρ. The following fundamental facts (see [AKN98]) show that both the trace distance and relative entropy only decrease on performing a measurement.

Fact 4 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathcal{H}$. Let $\mathcal{F}$ be a measurement (POVM) on $\mathcal{H}$. Then, $\|\mathcal{F}\rho - \mathcal{F}\sigma\|_1 \leq \|\rho - \sigma\|_t$.*

Fact 5 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathcal{H}$. Let $\mathcal{F}$ be a measurement (POVM) on $\mathcal{H}$. Then, $S(\mathcal{F}\rho \| \mathcal{F}\sigma) \leq S(\rho \| \sigma)$.*

Jozsa [Joz94] gave an elementary proof for finite dimensional Hilbert spaces of the following basic and remarkable property about fidelity.

Fact 6 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathcal{H}$. Then for any finite dimensional Hilbert space $\mathcal{K}$ such that $\dim(\mathcal{K}) \geq \dim(\mathcal{H})$, there exist purifications $|\psi\rangle, |\phi\rangle$ of ρ, σ in $\mathcal{H} \otimes \mathcal{K}$, such that*

$$B(\rho, \sigma) = |\langle \psi | \phi \rangle|.$$

Also,

$$B(\rho, \sigma) = \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_t.$$

We will also need the following result about fidelity, proved by Fuchs and Caves [FC95].

Fact 7 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathcal{H}$. Then*

$$B(\rho, \sigma) = \inf_{F_1, \ldots, F_k} \sum_{i=1}^{k} \sqrt{\mathrm{Tr}\,(F_i \rho)\,\mathrm{Tr}\,(F_i \sigma)},$$

where $\{F_1, \ldots, F_k\}$ ranges over POVMs on $\mathcal{H}$. In fact, the infimum above can be attained by a complete orthogonal measurement on $\mathcal{H}$.

The following relation is known between fidelity and trace distance between two density matrices [NC00].

Fact 8 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathcal{H}$. Then*

$$2(1 - B(\rho, \sigma)) \leq \|\rho - \sigma\|_t \leq 2\sqrt{1 - B(\rho, \sigma)^2}.$$

The following information-theoretic facts follows easily from the definitions.

Fact 9 *Let X be a classical random variable and M be a quantum encoding of X. Let X take the values $1, \ldots, l$ with probabilities $p_1, \ldots, p_l$ and let $\sigma_1, \ldots, \sigma_l$ be the respective density matrices of M. Let $\sigma \overset{\Delta}{=} \sum_{j=1}^{l} p_j \sigma_j$ be the average density matrix of M. Then, $I(X : M) = \sum_{j=1}^{l} p_j S(\sigma_j \| \sigma)$.*

B Improved Average encoding and Local transition theorem

In this section, we observe that the following lemma from [DCHR78] can be used to improve the average encoding and local transition arguments of [KNTZ01]. If Lemmas 6 and 7 are used in their place, the factor k^4 in the denominator of some existing lower bounds (e.g. [KNTZ01] and [JRS02]) can be replaced by k^2.

Lemma 5 *Let ρ and σ be two density matrices such that $S(\rho\|\sigma)$ is finite. Then,*

$$B(\rho, \sigma) \geq 2^{-S(\rho\|\sigma)/2}.$$

Proof: Let M be the complete orthogonal measurement which achieves the infimum as in the Fact 7. Let P and Q be the classical distributions resulting after the measurement M is performed. From Fact 5 and concavity of the $\log$ function it follows that:

$$\begin{aligned} -(1/2)S(\rho\|\sigma) \leq -(1/2)S(P\|Q) &= \sum_i p_i \log \sqrt{q_i/p_i} \\ &\leq \log \sum_i \sqrt{q_i p_i} \\ &= \log B(P, Q) = \log B(\rho, \sigma). \end{aligned}$$

■

Corollary 2 *Let ρ and σ be two density matrices such that $S(\rho\|\sigma)$ is finite. Then,*

$$1 - B(\rho, \sigma) \leq ((\ln 2)/2) S(\rho\|\sigma).$$

Proof: If $((\ln 2)/2)S(\rho\|\sigma) \geq 1$ then the inequality is trivial since $B(\, , \,) \geq 0$. Therefore when $((\ln 2)/2)S(\rho\|\sigma) \leq 1$,

$$\begin{aligned} B(\rho, \sigma) &\geq 2^{-S(\rho\|\sigma)/2} \\ &\geq \exp^{-((\ln 2)/2)S(\rho\|\sigma)} \\ &\geq 1 - ((\ln 2)/2)S(\rho\|\sigma) \;\; (\text{since} \;\; \exp^{-x} \geq 1 - x, \;\; \text{for} \;\; 0 \leq x \leq 1) \\ \Rightarrow 1 - B(\rho, \sigma) &\leq ((\ln 2)/2) S(\rho\|\sigma). \end{aligned}$$

■

The following lemma follows immediately from the above corollary and Fact 9.

Lemma 6 (Average encoding theorem) *Suppose X, Q are two disjoint quantum systems, where X is a classical random variable which takes value x with probability p_x, and Q is a quantum encoding $x \mapsto \sigma_x$ of X. Let the density matrix of the average encoding be $\sigma \stackrel{\Delta}{=} \sum_x p_x \sigma_x$. Then,*

$$\sum_x p_x (1 - B(\rho, \rho_x)) \leq (\ln 2/2) I(X : Q).$$

The following lemma follows immediately from Fact 6 and Fact 8 and Corollary 2

Lemma 7 (Local transition theorem) *Let ρ_1, ρ_2 be two density matrices in the same finite dimensional Hilbert space $\mathcal{H}$, $\mathcal{K}$ any Hilbert space of dimension at least the dimension of $\mathcal{H}$, 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 $|\phi_2'\rangle \stackrel{\Delta}{=} (I \otimes U)|\phi_2\rangle$ (I is the identity operator on $\mathcal{H}$) such that*

$$\left\| |\phi_1\rangle\langle\phi_1| - |\phi_2'\rangle\langle\phi_2'| \right\|_t \leq 2\sqrt{1 - B(\rho_1, \rho_2)^2} \leq 2\sqrt{2(1 - B(\rho_1, \rho_2))} \leq 2\sqrt{\ln 2 (S(\rho_1 \| \rho_2))}.$$

Fact 10 ([Lin91]) *Suppose X and Q are two classical correlated random variables, where X is uniformly distributed over $\{0, 1\}$ and Q is an encoding $x \rightarrow P_x$ of X. Then,*

$$1 - B(P_1, P_2) \leq I(X : Q).$$

Following corollary is immediate from Fact 7 and monotonicity of information,

Corollary 3 *Suppose X and Q are disjoint quantum systems, where X is a classical random variable uniformly distributed over $\{0, 1\}$ and Q is a quantum encoding $x \rightarrow \sigma_x$ of X. Then, $1 - B(\sigma_1, \sigma_2) \leq I(X : Q)$.*