Bounded-Error Quantum State Identification and Exponential Separations in Communication Complexity

Dmitry Gavinsky [*] Julia Kempe [†] Oded Regev [‡] Ronald de Wolf [§]

ABSTRACT

We consider the problem of bounded-error quantum state identification: given either state α_0 or state α_1, we are required to output '0', '1' or '?' ("don't know"), such that conditioned on outputting '0' or '1', our guess is correct with high probability. The goal is to maximize the probability of not outputting '?'. We prove a direct product theorem: if we're given two such problems, with optimal probabilities a and b, respectively, and the states in the first problem are pure, then the optimal probability for the joint bounded-error state identification problem is $O(ab)$. Our proof is based on semidefinite programming duality and may be of wider interest.

Using this result, we present two exponential separations in the simultaneous message passing model of communication complexity. First, we describe a relation that can be computed with $O(\log n)$ classical bits of communication in the presence of shared randomness, but needs $\Omega(n^{1/3})$ communication if the parties don't share randomness, even if communication is quantum. This shows the optimality of Yao's recent exponential simulation of shared-randomness protocols by quantum protocols without shared randomness. Second, we describe a relation that can be computed with $O(\log n)$ classical bits of communication in the presence of shared entanglement, but needs $\Omega((n/\log n)^{1/3})$ communication if the parties share randomness but no entanglement, even if communication is quantum. This is the first example in communication complexity where entanglement buys you much more than quantum communication does.

Categories and Subject Descriptors

E.4 [**Coding and information theory**]: Formal models of communication; F.1.2 [**Computation by Abstract Devices**]: Modes of Computation; F.1.3 [**Computation by Abstract Devices**]: Complexity Measures and Classes—*Relations among complexity measures*

General Terms

Algorithms, Theory

Keywords

quantum computing, state identification, communication complexity, randomness, entanglement

1. INTRODUCTION

1.1 Bounded-error quantum state identification

Suppose we are given one of two mixed quantum states, α_0 or α_1, each with probability 1/2. We know what α_0 and α_1 are. Our goal is to identify which one we are given. It is well known that we can output the correct answer (0 or 1) with probability $1/2 + \|\alpha_0 - \alpha_1\|_{tr}/2$, where $\|\cdot\|_{tr}$ is the trace norm (the sum of the singular values, divided by 2). This is optimal. In particular, if α_0 and α_1 are very close in trace norm, the best measurement will do little better than a fair coin flip. In some situations, however, we cannot afford to output the wrong answer with such high probability, and would rather settle for a measurement that sometimes claims ignorance, but that is usually correct in the case where it does give an output.

To illustrate this, suppose the states involved are the following pure states:

$$|\alpha_0\rangle = \sqrt{a}|0\rangle + \sqrt{1-a}|2\rangle$$
$$|\alpha_1\rangle = \sqrt{a}|1\rangle + \sqrt{1-a}|2\rangle$$

If we cannot afford to make a mistake at all, it is clear what measurement we should apply: measure in the computational basis, and if the outcome is 0 the state must have been

[*]University of Calgary.

[†]CNRS & LRI, Univ. de Paris-Sud, Orsay. Supported in part by ACI Sécurité Informatique SI/03 511 and ACI-Cryptologie CR/02 20040 grants of the French Research Ministry, the EU fifth framework project RESQ, IST-2001-37559, and the EU sixth framework project QAP. Supported by ARO grant DAAD19-03-1-0082 while visiting MSRI.

[‡]Department of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Supported by an Alon Fellowship, by the Binational Science Foundation, by the Israel Science Foundation, and by the EU sixth framework project QAP.

[§]CWI, Amsterdam. Supported by a Veni grant from the Netherlands Organization for Scientific Research (NWO) and by the EU fifth framework project RESQ, IST-2001-37559, and the EU sixth framework project QAP.

STOC'06, May21–23, 2006, Seattle, Washington, USA.

α_0; if the outcome is 1 the state must have been α_1; if the outcome is 2 we claim ignorance. Note that the probability of getting an answer (0 or 1) for the identification problem is now only a. We have thus increased our confidence in the answer, at the expense of decreasing the probability of getting an answer at all. Now consider a slightly more "fudged" example, for some small ε:

$$|\alpha_0\rangle = \sqrt{(1-\varepsilon)a}|0\rangle + \sqrt{\varepsilon a}|1\rangle + \sqrt{1-a}|2\rangle$$
$$|\alpha_1\rangle = \sqrt{\varepsilon a}|0\rangle + \sqrt{(1-\varepsilon)a}|1\rangle + \sqrt{1-a}|2\rangle$$

If we apply the same procedure as before, we have now a small probability of error: on both states our measurement outputs a guess (0 or 1) with probability a, and *if* we output a guess, then that guess is wrong with probability only ε. If ε is sufficiently small, this may still be acceptable for many applications.

More generally, let A be some classical random variable, and B be another random variable whose range includes the special symbol '?'. We call B an (a,ε)-*predictor* for A if $\Pr[B \neq ?] \geq a$ and $\Pr[A = B \mid B \neq ?] \geq 1-\varepsilon$. For example, the above measurement applied to state α_X where X is a random bit, gives us an (a,ε)-predictor for X if we interpret output 2 as '?'. Motivated by the above examples—and by our applications in later sections—we define the bounded-error state identification problem:

> Given a register containing α_X, with X a uniformly random bit, and an $\varepsilon > 0$, what is the maximal a for which there exists a quantum measurement on the register whose outcome is an (a,ε)-predictor for X?

We use $D_\varepsilon(\alpha_0,\alpha_1)$ to denote the maximal value a. We stress again that the error probability is a *conditional* probability, conditioned on actually outputting a guess for the bit (0 or 1). Unlike the straightforward distinguishing problem, where the optimal success probability is determined by the trace distance $\|\alpha_0-\alpha_1\|_{tr}$, we do not know of any simple metric on density matrices that determines the value $D_\varepsilon(\alpha_0,\alpha_1)$. However, as was also noted by Eldar [11], one can easily express quantities like this as the optimal value of a semidefinite program, as we will do in Section 3.

Now suppose we are given another identification problem in a second register, quantum state β_Y for a random bit Y, and suppose $b = D_\varepsilon(\beta_0,\beta_1)$ is the largest value for which we can obtain a (b,ε)-predictor for Y. We now want to determine the optimal probability with which we can identify (again with error at most ε or something related) *both* states simultaneously. That is, what is the maximal probability $p = D_\varepsilon(\alpha_0\otimes\beta_0, \alpha_0\otimes\beta_1, \alpha_1\otimes\beta_0, \alpha_1\otimes\beta_1)$ such that a joint measurement on $\alpha_X\otimes\beta_Y$ gives us a (p,ε)-predictor for XY? Since the two registers are completely independent, it seems there is nothing much better we can do except applying the optimal measurement for both registers separately.[1] Thus our intuition suggests that $p \leq ab$, or at least $p \leq O(ab)$. This problem has a flavor similar to "direct product theorems" in computational complexity theory, where one is usually interested in $k \geq 2$ independent instances of some computational problem, and the aim is to show that the overall success probability of some algorithm for the k-fold problem is close to the product of the k individual success probabilities. Another problem with a similar flavor is the notoriously hard quantum information theory issue of multiplicativity of norms of superoperators under tensor product [17].

Proving our intuition actually turned out to be quite a hard problem, and here we briefly mention some reasons why. First, classically the intuition turns out to be true, but the optimal 2-register strategy is *not* the product of two separate optimal 1-register strategies (see the longer version of this paper [14] for more details). Second, one would expect the same intuition to hold in the case where one wants to obtain the parity $X \oplus Y$ instead of the two bits X and Y separately. Yet in that case, we know an example where $p \approx a, b$ and not $p \leq O(ab)$; more details may be found in [14]. In an earlier preprint [13] we were only able to prove it for $\varepsilon = 0$, which was then used by us in [13] and [12] to obtain various zero-error separations in communication complexity. The present paper supersedes all of these unpublished results and gives in Section 3 the first proof of the $p \leq O(ab)$ bound for the case where at least one of the two sides is pure (i.e., α_0 and α_1 are both pure, or β_0 and β_1 are both pure). More precisely, we show

$$D_{\varepsilon/2}(\alpha_0\otimes\beta_0, \alpha_0\otimes\beta_1, \alpha_1\otimes\beta_0, \alpha_1\otimes\beta_1) \leq O(D_\varepsilon(\alpha_0,\alpha_1)\cdot D_\varepsilon(\beta_0,\beta_1)). \quad (1)$$

Our proof relies heavily on a semidefinite programming formulation for the quantities involved and on an analysis of their duals. Note that because of the $\varepsilon/2$ on the left hand side, this bound is slightly weaker than what we have promised; as we explain in the longer version of this paper [14], this modification is (somewhat surprisingly) necessary. This is a third reason why the bounded-error state identification problem is quite subtle.

1.2 Exponential separations in communication complexity

Apart from being an interesting information theoretic problem in its own right, the bounded-error state identification problem and our direct product theorem have interesting applications. We give two new exponential separations, both in the *simultaneous message passing* (SMP) model of communication complexity. The area of communication complexity deals with the amount of communication required for solving computational problems with distributed input. This area is interesting for its own sake, but also has many applications to lower bounds on circuit size, data structures, etc. The SMP model involves three parties: Alice, Bob, and a referee. Alice gets input x, Bob gets input y. They each send one message to the referee, to enable him to compute something depending on both x and y, such as a Boolean function or some relational property. The *cost* or *complexity* of a communication protocol is the length of the total communication for a worst-case input, and the complexity of a problem is the cost of the best protocol that solves our problem with small error probability.

The SMP model is arguably the weakest setting of communication complexity that is still interesting. Even this simple setting is not well understood. In the case of deterministic protocols, the optimal communication is determined by the number of distinct rows (and columns) in the communication matrix, which is a simple property. However, as soon as we add randomization to the model things become much more complicated. For one, we can choose to

[1] This actually gives slightly worse error $2\varepsilon - \varepsilon^2$ for the prediction of XY, so potentially it could be that $p \ll ab$.

either add *shared* (a.k.a. public) or *private* randomness. In other communication models this difference affects the optimal communication by at most an additive $O(\log n)$ [23], but in the SMP model the difference can be huge. For example, the equality function for n-bit strings requires about $\sqrt{n}$ bits of communication if the parties have only private randomness [1, 24, 2], but only constant communication with shared randomness! No simple characterization of SMP communication complexity with either private or shared randomness is known.[2]

The situation becomes more complicated still when we throw in *quantum* communication. Buhrman et al. [8] exhibited a quantum protocol for the equality function with $O(\log n)$ qubits of communication. This is exponentially better than classical private-randomness protocols, but slightly worse than shared-randomness protocols. Roughly speaking, their quantum fingerprinting technique may be viewed as replacing the shared randomness by a quantum superposition.

1.2.1 Shared randomness beats quantum communication

The fingerprinting idea of [8] was generalized by Yao [29], who showed that every classical shared-randomness protocol with c-bit messages for a Boolean function can be simulated by a quantum fingerprinting protocol that uses $O(2^{4c} \log n)$ qubits of communication. This has since been improved to $O(2^{2c} \log n)$ qubits [13, 15]. In particular, every $O(1)$-bit shared-randomness protocol can be simulated by an $O(\log n)$-qubit quantum protocol. Again, quantum superposition replaces shared randomness in this construction.

This raises the question whether something similar always holds in the SMP model: can every classical shared-randomness protocol be efficiently simulated by some protocol that sends qubits but shares neither randomness nor entanglement? Since the appearance of Yao's paper, quite a number of people have tried to address this. Our first separation, presented in Section 4, gives a negative answer to this question. Suppose Alice receives inputs $x, s \in \{0,1\}^n$ with the property that s has Hamming weight $n/2$ and Bob receives input $y \in \{0,1\}^n$. The referee should output, with probability at least $1-\varepsilon$, a triple (i, x_i, y_i) for an i satisfying $s_i = 1$. We prove that protocols where Alice and Bob share randomness can solve this task with $O(\log n)$ classical bits of communication, while every bounded-error quantum protocol without shared randomness needs $\Omega(n^{1/3})$ qubits of communication. This shows for the first time that the resource of shared randomness cannot be efficiently traded for quantum communication. The quantum lower bound relies crucially on our direct product theorem for bounded-error state identification.

Yao's exponential simulation can be made to work for relations as well, and our quantum lower bound shows that it is essentially optimal, since the required quantum communication is exponentially larger than the classical shared-randomness complexity for our relational problem. We expect a similar gap to hold for (promise) Boolean functions as well. Our separation complements a separation in the other direction: Bar-Yossef et al. [3] exhibited a relation where quantum SMP protocols are exponentially *more* efficient than classical SMP protocols even with shared randomness (also in their case it is open whether there is a similar gap for a Boolean function). Accordingly, the quantum SMP model is incomparable with the classical shared-randomness SMP model.

[2]Kremer et al. [19] claimed a characterization of shared-randomness complexity as the largest of the two one-way complexities, but Bar-Yossef et al. [4, Section 4] exhibited a function where their characterization fails.

1.2.2 Shared entanglement beats quantum communication with shared randomness

The second application of our state identification result is again in the SMP model. While the previous application separated classical protocols with shared randomness from quantum protocols without shared randomness, this one separates classical protocols with *entanglement* (EPR-pairs, 2-qubit states $\frac{1}{2}(|00\rangle + |11\rangle)$) from quantum protocols with shared randomness.

The additional power that prior entanglement gives is one of the most fundamental questions in quantum communication complexity. This additional power is not well understood. We basically know two ways in which entanglement can help: it can be used for teleportation (where one EPR-pair and two classical bits of communication replace one qubit of communication) and it can be used for shared randomness (if Alice and Bob each measure their side of their shared EPR-pair in the computational basis, they get the same random bit). Neither saves very much communication, and it has in fact been conjectured for the standard two-party one-round and many-round protocols that the model of classical communication with entanglement [9] and the model of quantum communication without entanglement [28] are essentially equivalent.

Our second separation shows that the situation is very different in the SMP model: the qubit-communication model cannot efficiently simulate the entanglement model. In Section 5 we exhibit a relational problem, inspired by the problem of Bar-Yossef et al. mentioned above, that can be solved with $\log n$ EPR-pairs shared between Alice and Bob and $O(\log n)$ classical bits of communication. In contrast, if only shared randomness is available instead of entanglement, every bounded-error SMP protocol needs $\Omega((n/\log n)^{1/3})$ quantum bits of communication. Again, our direct product theorem is crucial for proving the quantum lower bound. This is the first example of a communication problem where entanglement is much more useful than quantum communication.

2. PRELIMINARIES

2.1 Quantum computing

The essentials needed for this paper are quantum states and their measurement. First, an m-qubit *pure state* is a superposition $|\phi\rangle = \sum_{z\in\{0,1\}^m} \alpha_z |z\rangle$ over all classical m-bit states. The α_z's are complex numbers called *amplitudes*, and $\sum_z |\alpha_z|^2 = 1$. Hence a pure state $|\phi\rangle$ is a unit vector in $\mathbb{C}^{2^m}$. Its complex conjugate (a row vector with entries conjugated) is denoted $\langle\phi|$. The inner product between $|\phi\rangle$ and $|\psi\rangle = \sum_z \beta_z |z\rangle$ is the dot product $\langle\phi| \cdot |\psi\rangle = \langle\phi|\psi\rangle = \sum_z \alpha_z^* \beta_z$. The *norm* of a vector v is $\|v\| = \sqrt{\langle v|v\rangle}$. Second, a *mixed state* $\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$ corresponds to a probability distribution over pure states, where $|\phi_i\rangle$ is given with probability p_i. A k-outcome *positive operator-valued measure* (POVM) is given by k positive semidefinite operators

$E_1, \ldots, E_k$ with the property that $\sum_{i=1}^{k} E_i = I$. When this POVM is applied to a mixed state ρ, the probability of the i-th outcome is given by the trace $\mathrm{Tr}[E_i\rho]$. See Nielsen and Chuang [25] for more details.

2.2 Communication complexity

We now give a somewhat informal description of the simultaneous message passing model discussed in our two applications. For a more formal description, we refer to Kushilevitz and Nisan [20] for classical communication complexity and to the surveys [18, 6, 27] for the quantum variant. In the simultaneous message passing model, Alice receives input x, Bob receives input y, they each send a message to a referee who should then output either $f(x, y)$ in the case of a functional problem, or an element from some set $R(x, y)$ in the case of a relational problem. We use $R_\varepsilon^{\|}(P)$, $R_\varepsilon^{\|,pub}(P)$, $R_\varepsilon^{\|,ent}(P)$ to denote, respectively, the optimal communication complexity of classical protocols that solve problem P with worst-case error probability ε, using, respectively, private randomness, shared randomness between Alice and Bob, and shared entanglement between Alice and Bob (EPR pairs). The number of shared coin flips or shared EPR-pairs is unlimited and does not count towards the communication cost of the protocol. We use $Q_\varepsilon^{\|}(P)$, $Q_\varepsilon^{\|,pub}(P)$, $Q_\varepsilon^{\|,ent}(P)$ for the variety that allows quantum communication.

2.3 The random access code argument

Here we will describe a slight extension of a quantum information theory argument due to Nayak [22] that we will apply several times in our communication lower bounds. We call this the "random access code argument". We assume familiarity with classical information theory [10] and quantum information theory [25].

LEMMA 1. *["Random Access Code Argument"] Let $X = X_1 \ldots X_n$ be a classical random variable of n uniformly distributed bits. Suppose for each instantiation $X = x$ we have a quantum state M_x of q qubits. Suppose also that for each $i \in [n]$ of our choice we can apply a quantum measurement to M_X whose outcome is a $(\lambda_i, \varepsilon_i)$-predictor for X_i. Then*

$$\sum_{i=1}^{n} \lambda_i (1 - H(\varepsilon_i)) \leq q.$$

Before giving the proof, notice the following special case: if we can predict each X_i with bias η_i (i.e., we have a $(1, 1/2 - \eta_i)$-predictor), then the above bound becomes

$$\sum_{i=1}^{n} (1 - H(1/2 - \eta_i)) \leq q.$$

Since $1 - H(1/2 - \eta_i) = \Theta(\eta_i^2)$, the left hand side is essentially the sum of squares of the η_i.

PROOF. First, let Y be a classical random variable corresponding to a uniformly distributed bit. Let B be another random variable that is a (λ, ε)-predictor of Y. Using $H(Y \mid B, B \neq ?) \leq H(\varepsilon)$ and $\Pr[B \neq ?] \geq \lambda$, we can upper bound the entropy of Y given B:

$$\begin{aligned} H(Y \mid B) &= \Pr[B = ?] \cdot H(Y \mid B, B = ?) \\ &\quad + \Pr[B \neq ?] \cdot H(Y \mid B, B \neq ?) \\ &\leq (1 - \Pr[B \neq ?]) \cdot 1 + \Pr[B \neq ?] \cdot H(\varepsilon) \\ &\leq 1 - \lambda(1 - H(\varepsilon)), \end{aligned}$$

and hence lower bound the mutual information between Y and B:

$$I(Y : B) = H(Y) - H(Y \mid B) \geq \lambda(1 - H(\varepsilon)).$$

Now let B_i be the outcome of the measurement corresponding to i applied to M_X. We have

$$S(X_i : M_X) \geq I(X_i : B_i) \geq \lambda_i(1 - H(\varepsilon_i))$$

by Holevo's theorem [16] (the left hand side is equal to the Holevo χ-quantity).

Using [25, Theorem 11.8.5] we have

$$\begin{aligned} S(X : M_X) &= S(X) + S(M_X) - S(X, M_X) \\ &= S(M_X) - \frac{1}{2^n} \sum_{x \in \{0,1\}^n} S(M_x) \leq S(M_X) \leq q. \end{aligned}$$

Abbreviating $X_{1:i-1} = X_1 \ldots X_{i-1}$, a chain rule for mutual information gives

$$S(X : M_X) = \sum_{i=1}^{n} S(X_i : M_X \mid X_{1:i-1}).$$

Using strong subadditivity and the fact $S(X_i \mid X_{1:i-1}) = S(X_i)$ we get

$$\begin{aligned} S(X_i : M_X \mid X_{1:i-1}) &= S(X_i \mid X_{1:i-1}) - S(X_i \mid M_X X_{1:i-1}) \\ &\geq S(X_i) - S(X_i \mid M_X) = S(X_i : M_X). \end{aligned}$$

Combining our inequalities gives the bound on q. □

3. BOUNDED-ERROR QUANTUM STATE IDENTIFICATION: DIRECT PRODUCT

In this section we prove our main results about the 2-register quantum state identification problem. We use the powerful technique of semidefinite programming duality. For details on semidefinite programming, see e.g. [21, 26]. Recall that in the first register we are given a quantum state α_X, with X a random bit, and the optimal probability with which we can get an ε-predictor for X is a. In the second register we're given β_Y, with Y a random bit, and the optimal probability with which we can get an ε-predictor for Y is b. We now want to know the optimal probability p with which a joint measurement on both registers can obtain an ε-predictor for XY. We will actually prove two bounds. First, for the case where α_0, α_1 are pure and β_0, β_1 are unrestricted, our Theorem 1 implies

$$\begin{aligned} &D_{\varepsilon/2}(\alpha_0 \otimes \beta_0, \alpha_0 \otimes \beta_1, \alpha_1 \otimes \beta_0, \alpha_1 \otimes \beta_1) \\ &\qquad \leq O(D_\varepsilon(\alpha_0, \alpha_1) \cdot D_\varepsilon(\beta_0, \beta_1)). \end{aligned} \tag{2}$$

Second, if we allow all of $\alpha_0, \alpha_1, \beta_0, \beta_1$ to be mixed states then our Corollary 1 gives

$$\begin{aligned} &D_{\varepsilon/2}(\alpha_0 \otimes \beta_0, \alpha_0 \otimes \beta_1, \alpha_1 \otimes \beta_0, \alpha_1 \otimes \beta_1) \\ &\qquad \leq O(\|\alpha_0 - \alpha_1\|_{tr} \cdot D_\varepsilon(\beta_0, \beta_1)). \end{aligned}$$

The second bound will follow from the first by purifying the mixed states α_0 and α_1.

Let us first characterize $D_\varepsilon(\alpha_0, \alpha_1)$. Recall that any measurement whose outcome is an (a, ε)-predictor outputs the correct answer with probability at least $1 - \varepsilon$ *conditioned* on outputting a guess (0 or 1, but not ?). Denote the three

measurement operators by E_0, E_1, $E_?$. Then we require

$$\varepsilon \geq \Pr[\text{wrong guess} \mid \text{guess}] = \frac{\Pr[\text{wrong guess}]}{\Pr[\text{guess}]} = \frac{\frac{1}{2}\mathrm{Tr}[E_0\alpha_1] + \frac{1}{2}\mathrm{Tr}[E_1\alpha_0]}{\mathrm{Tr}\left[(E_0+E_1)\alpha\right]}, \tag{3}$$

where $\alpha = \frac{1}{2}(\alpha_0+\alpha_1)$ is the average state. To our knowledge there is no simple expression for $D_\varepsilon(\alpha_0,\alpha_1)$ in terms of α_0 and α_1. However, one can easily express it as a solution to a semidefinite program (SDP). For fixed density matrices α_0, α_1 and fixed $\varepsilon \in [0, 1/2)$, the optimal value $a = D_\varepsilon(\alpha_0,\alpha_1)$ is given by the SDP:

$$\begin{array}{ll} \max & \mathrm{Tr}[(E_0+E_1)\alpha] \\ \text{s.t.} & 0 \preceq E_0, E_1, \\ & E_0 + E_1 \preceq I, \\ & \frac{1}{2}\mathrm{Tr}[E_0\alpha_1] + \frac{1}{2}\mathrm{Tr}[E_1\alpha_0] \leq \varepsilon \mathrm{Tr}[(E_0+E_1)\alpha]. \end{array} \tag{4}$$

The first two constraints state that the operators E_0, E_1 together with a third operator $E_? = I - E_0 - E_1$ form a valid quantum measurement. The last constraint bounds the conditional error probability, as in Eq. (3). An analogous SDP can be written for $b = D_\varepsilon(\beta_0,\beta_1)$.

Similarly we can write the primal SDP that optimizes $p = D_\varepsilon(\alpha_0\otimes\beta_0, \alpha_0\otimes\beta_1, \alpha_1\otimes\beta_0, \alpha_1\otimes\beta_1)$:

$$\begin{array}{ll} \max & \mathrm{Tr}[(E_{00}+E_{01}+E_{10}+E_{11})\,\alpha\otimes\beta] \\ \text{s.t.} & 0 \preceq E_{00}, E_{01}, E_{10}, E_{11}, \\ & E_{00}+E_{01}+E_{10}+E_{11} \preceq I, \\ & \frac{1}{4}\mathrm{Tr}\left[(E_{01}+E_{10}+E_{11})\,\alpha_0\otimes\beta_0 + \right. \\ & \qquad (E_{00}+E_{10}+E_{11})\,\alpha_0\otimes\beta_1 + \\ & \qquad (E_{00}+E_{01}+E_{11})\,\alpha_1\otimes\beta_0 + \\ & \qquad \left.(E_{00}+E_{01}+E_{10})\,\alpha_1\otimes\beta_1\right] \\ & \qquad \leq \varepsilon\mathrm{Tr}[(E_{00}+E_{01}+E_{10}+E_{11})\,\alpha\otimes\beta]. \end{array} \tag{5}$$

Here $\alpha\otimes\beta = \frac{1}{4}(\alpha_0\otimes\beta_0 + \alpha_0\otimes\beta_1 + \alpha_1\otimes\beta_0 + \alpha_1\otimes\beta_1)$ is the average state.

THEOREM 1. *Let $0 \leq \varepsilon < \frac{1}{2}$ and α_0, α_1, β_0, β_1 be density matrices, where α_0, α_1 correspond to pure states $|\alpha_0\rangle, |\alpha_1\rangle$. Let $b = D_\varepsilon(\beta_0,\beta_1)$ and $p = D_{\varepsilon/2}(\alpha_0\otimes\beta_0, \alpha_0\otimes\beta_1, \alpha_1\otimes\beta_0, \alpha_1\otimes\beta_1)$. Then*

$$p \leq 16(1-|\langle\alpha_0|\alpha_1\rangle|^2)\cdot b.$$

Since α_0 and α_1 are pure, $a = D_\varepsilon(\alpha_0,\alpha_1) \geq D_0(\alpha_0,\alpha_1) \geq \frac{1}{2}(1-|\langle\alpha_0|\alpha_1\rangle|^2)$, where the last inequality follows by considering the projective measurement on $|\alpha_0\rangle$ and $|\alpha_0^\perp\rangle$. Hence this theorem implies Eq. (2).

PROOF. The idea behind our proof is the following. Both b and p are the solution of an SDP and so any feasible solution of the corresponding dual SDP yields an upper bound to b resp. p. We will show that a feasible solution with value $d_b \geq b$ for the dual for b can be used to construct a feasible solution with value $16(1-|\langle\alpha_0|\alpha_1\rangle|^2)\cdot d_b$ for the dual for p. This value then upper bounds p. The dual SDP for b is strictly feasible in our case, which means that we can make d_b as close to b as we want. This implies the theorem.

Let $\delta := \sqrt{1-|\langle\alpha_0|\alpha_1\rangle|^2}$. Then we want to show $p \leq 16\delta^2 b$. The dual SDP for b is

$$\begin{array}{ll} \min & \mathrm{Tr}[X_b] \\ \text{s.t.} & X_b \succeq 0, z_b \geq 0, \\ & X_b \succeq \frac{1}{2}\left((1+\varepsilon z_b)\beta_0 + (1-(1-\varepsilon)z_b)\beta_1\right) =: X_1, \\ & X_b \succeq \frac{1}{2}\left((1+\varepsilon z_b)\beta_1 + (1-(1-\varepsilon)z_b)\beta_0\right) =: X_2. \end{array} \tag{6}$$

This SDP is strictly feasible, for example, $z_b = \frac{1}{2}, X_b = 2I$ is a strictly feasible solution. Hence by strong duality its optimal value is exactly b.

The dual SDP for p is

$$\begin{array}{ll} \min & \mathrm{Tr}[X] \\ \text{s.t.} & X \succeq 0, z \geq 0, \\ & X \succeq \frac{1}{4}\left(\left\{(1+\frac{\varepsilon}{2}z)\alpha_0 + (1-(1-\frac{\varepsilon}{2})z)\alpha_1\right\}\otimes\beta_0 + \right. \\ & \qquad \left.(1-(1-\frac{\varepsilon}{2})z)(\alpha_0+\alpha_1)\otimes\beta_1\right) =: X_1', \\ & X \succeq \frac{1}{4}\left(\left\{(1+\frac{\varepsilon}{2}z)\alpha_0 + (1-(1-\frac{\varepsilon}{2})z)\alpha_1\right\}\otimes\beta_1 + \right. \\ & \qquad \left.(1-(1-\frac{\varepsilon}{2})z)(\alpha_0+\alpha_1)\otimes\beta_0\right) =: X_2', \\ & X \succeq \frac{1}{4}\left(\left\{(1+\frac{\varepsilon}{2}z)\alpha_1 + (1-(1-\frac{\varepsilon}{2})z)\alpha_0\right\}\otimes\beta_0 + \right. \\ & \qquad \left.(1-(1-\frac{\varepsilon}{2})z)(\alpha_0+\alpha_1)\otimes\beta_1\right) =: X_3', \\ & X \succeq \frac{1}{4}\left(\left\{(1+\frac{\varepsilon}{2}z)\alpha_1 + (1-(1-\frac{\varepsilon}{2})z)\alpha_0\right\}\otimes\beta_1 + \right. \\ & \qquad \left.(1-(1-\frac{\varepsilon}{2})z)(\alpha_0+\alpha_1)\otimes\beta_0\right) =: X_4'. \end{array} \tag{7}$$

For what follows we need to define the positive part of a Hermitian matrix. Any Hermitian matrix A can be written uniquely as $A = A^+ - A^-$, where A^+, A^- are positive semidefinite ($A^+, A^- \succeq 0$) and have orthogonal support. Then define $\mathrm{Pos}(A) = A^+$. We need the following simple properties:

CLAIM 1. *1. If $A \preceq B$ then $A \preceq Pos(B)$.*

2. If $A \succeq 0$ then $Pos(A\otimes B) = A\otimes Pos(B)$.

3. If $A \preceq B$ then $Tr[Pos(A)] \leq Tr[Pos(B)]$.

NB: it is *not* true that $A \preceq B$ implies $\mathrm{Pos}(A) \preceq \mathrm{Pos}(B)$.

PROOF. The first part follows from $B \preceq \mathrm{Pos}(B)$. The second part can be seen by diagonalizing the matrices (note that the non-zero eigenvalues of $\mathrm{Pos}(B)$ are exactly the positive eigenvalues of B). The third part can be seen for instance by using majorization (see e.g. [5]). If $A \preceq B$, then the vector of eigenvalues of A is submajorized by the vector of eigenvalues of B ([5], Eq. (II.16), Ky Fan Maximum Principle). This means that if we order the eigenvalues of A (resp. B) as $\lambda_1 \geq \lambda_2 \geq \ldots$ (resp. $\mu_1 \geq \mu_2 \geq \ldots$) then for all $k \geq 1$, $\sum_{i=1}^k \lambda_i \leq \sum_{i=1}^k \mu_i$. Together with the fact that the trace of $\mathrm{Pos}(A)$ is the sum of the positive eigenvalues of A, the property follows. □

We also need the following technical claim, which we will prove afterwards:

CLAIM 2. *Let $0 \leq \varepsilon < 1/2$ and $\sigma_0, \sigma_1, \rho_0, \rho_1$ be density matrices, where ρ_0 and ρ_1 are 2-dimensional of rank 1 (i.e., pure states). Denote by $\rho_1^\perp = I - \rho_1$ the rank 1 density matrix whose support is orthogonal to that of ρ_1. Then for all $z_b \geq 0$ there exists $z = z(\varepsilon, z_b) \geq 0$ such that*

$$\begin{aligned} 4\delta^2\rho_1^\perp\otimes\frac{1}{2}\left\{(1+\varepsilon z_b)\sigma_0 + (1-(1-\varepsilon)z_b)\sigma_1\right\} \\ \succeq \frac{1}{4}\left(\left\{(1+\frac{\varepsilon}{2}z)\rho_0 + (1-(1-\frac{\varepsilon}{2})z)\rho_1\right\}\otimes\sigma_0 \right. \\ \left. +(1-(1-\frac{\varepsilon}{2})z)(\rho_0+\rho_1)\otimes\sigma_1\right). \end{aligned}$$

Fix a dual solution (X_b, z_b) for (6). Our goal is to find a feasible solution (X, z) to (7) such that $\mathrm{Tr}[X] \leq 16\delta^2\mathrm{Tr}[X_b]$. Since $|\alpha_0\rangle$ and $|\alpha_1\rangle$ are pure states, we can assume without loss of generality that they are in a two dimensional space,

and therefore we can apply Claim 2 with $\rho_0 = \alpha_0$, $\rho_1 = \alpha_1$, $\sigma_0 = \beta_0$ and $\sigma_1 = \beta_1$. Let

$$Y_1 = 4\delta^2\alpha_1^\perp \otimes \frac{1}{2}\left\{(1+\varepsilon z_b)\beta_0 + (1-(1-\varepsilon)z_b)\beta_1\right\} = 4\delta^2\alpha_1^\perp \otimes X_1.$$

Claim 2 gives $z = z(\varepsilon, z_b)$ such that $Y_1 \succeq X_1'$ (see (7) for the definition of X_1'). Since $\alpha_1^\perp \succeq 0$ we can use Claim 1.2:

$$\mathrm{Pos}(Y_1) = 4\delta^2\alpha_1^\perp \otimes \mathrm{Pos}\frac{1}{2}\left\{(1+\varepsilon z_b)\beta_0 + (1-(1-\varepsilon)z_b)\beta_1\right\} = 4\delta^2\alpha_1^\perp \otimes \mathrm{Pos}(X_1).$$

Because $\alpha_1^\perp \succeq 0$, $\mathrm{Tr}[\mathrm{Pos}(Y_1)] = 4\delta^2\mathrm{Tr}[\mathrm{Pos}(X_1)]$. Moreover, $X_1 \preceq X_b$ by definition (see (6)) and $X_b = \mathrm{Pos}(X_b)$, hence $\mathrm{Tr}[\mathrm{Pos}(Y_1)] \leq 4\delta^2\mathrm{Tr}[\mathrm{Pos}(X_b)] = 4\delta^2\mathrm{Tr}[X_b]$ (using Claim 1.3).

However, $\mathrm{Pos}(Y_1)$ is not a solution of the dual SDP in (7) because it need not satisfy the last three inequalities. We construct three more matrices Y_2, Y_3 and Y_4 such that $Y_i \succeq X_i'$ for *the same* z as before. For this we apply Claim 2 three more times (for $Y_2 = 4\delta^2\alpha_1^\perp \otimes X_2$ with $(\rho_0,\rho_1,\sigma_0,\sigma_1) = (\alpha_0,\alpha_1,\beta_1,\beta_0)$, for $Y_3 = 4\delta^2\alpha_0^\perp \otimes X_1$ with $(\rho_0,\rho_1,\sigma_0,\sigma_1) = (\alpha_1,\alpha_0,\beta_0,\beta_1)$ and for $Y_4 = 4\delta^2\alpha_0^\perp \otimes X_2$ with $(\rho_0,\rho_1,\sigma_0,\sigma_1) = (\alpha_1,\alpha_0,\beta_1,\beta_0)$). Because z depends only on z_b and ε, which are the same in all four applications, we obtain each time the same z. Now define $X = \sum_{i=1}^4 \mathrm{Pos}(Y_i)$. Clearly (X, z) is a feasible solution to the SDP (7) since $X \succeq 0$ by definition and $X \succeq \mathrm{Pos}(Y_i) \succeq X_i'$ for $i = 1 \ldots 4$ (using Claim 1.1). But $\mathrm{Tr}[X] = \sum_{i=1}^4 \mathrm{Tr}[\mathrm{Pos}(Y_i)] \leq 16\delta^2\mathrm{Tr}[X_b]$. As $\mathrm{Tr}[X]$ is an upper bound on p, and $\mathrm{Tr}[X_b]$ can be made arbitrarily close to b, this implies the theorem.

Proof of Claim 2. Because σ_0 and σ_1 are positive semidefinite, it suffices to find a $z \geq 0$ for which

$$4\delta^2\rho_1^\perp\frac{1}{2}(1+\varepsilon z_b) \succeq \frac{1}{4}\left\{(1+\frac{\varepsilon}{2}z)\rho_0 + (1-(1-\frac{\varepsilon}{2})z)\rho_1\right\} \quad (8)$$

and

$$4\delta^2\rho_1^\perp\frac{1}{2}(1-(1-\varepsilon)z_b) \succeq \frac{1}{4}(1-(1-\frac{\varepsilon}{2})z)(\rho_0+\rho_1) \quad (9)$$

are true.

Let $|\rho_0\rangle, |\rho_1\rangle$ and $|\rho_1^\perp\rangle$ be pure states whose density matrices are ρ_0, ρ_1 and $\rho_1^\perp$. We choose their global phase such that $|\rho_0\rangle = \sqrt{1-\delta^2}|\rho_1\rangle + \delta|\rho_1^\perp\rangle$. Then, in the basis given by $|\rho_1\rangle$, $|\rho_1^\perp\rangle$, Eqs. (8) and (9) become

$$\begin{pmatrix} z(1-\varepsilon+\delta^2\frac{\varepsilon}{2})+\delta^2-2 & -\delta\sqrt{1-\delta^2}(1+\frac{\varepsilon}{2}z) \\ -\delta\sqrt{1-\delta^2}(1+\frac{\varepsilon}{2}z) & \delta^2(7+8\varepsilon z_b - \frac{\varepsilon}{2}z)\end{pmatrix} \succeq 0 \quad (10)$$

and

$$\begin{pmatrix} ((1\text{-}\frac{\varepsilon}{2})z-1)(2\text{-}\delta^2) & \delta\sqrt{1-\delta^2}((1\text{-}\frac{\varepsilon}{2})z-1) \\ \delta\sqrt{1-\delta^2}((1\text{-}\frac{\varepsilon}{2})z-1) & \delta^2(7-8(1\text{-}\varepsilon)z_b+(1\text{-}\frac{\varepsilon}{2})z)\end{pmatrix} \succeq 0 \quad (11)$$

To show that a 2×2 Hermitian matrix is positive semidefinite it suffices to show that both its determinant and at least one of its diagonal entries are positive. We choose

$$z = 16\frac{1-\varepsilon}{1-\varepsilon/2}z_b + \frac{4}{1-\varepsilon}.$$

Since $z \geq 4$, the upper diagonal entries of the matrices in Eqs. (10) and (11) are positive. Moreover, if $\delta = 0$ these matrices are trivially positive. If $\delta > 0$ then we can cancel $\delta^2 > 0$ from both terms that appear in their determinants. Hence, for Eqs. (10) and (11) to be true it suffices to show

$$\big(z(1-\varepsilon)-2\big)(7+8\varepsilon z_b - \frac{\varepsilon}{2}z) - (1+\frac{\varepsilon}{2}z)^2 > 0 \quad (12)$$

and

$$(2-\delta^2)((1-\frac{\varepsilon}{2})z-1)(7-8(1-\varepsilon)z_b+(1-\frac{\varepsilon}{2})z) - (1-\delta^2)((1-\frac{\varepsilon}{2})z-1)^2 > 0. \quad (13)$$

To derive Eq. (12) we have replaced the term $z(1-\varepsilon+\delta^2\frac{\varepsilon}{2})+\delta^2-2$ by the smaller positive term $z(1-\varepsilon)-2$, which is allowed because this equation is only true if $7+8\varepsilon z_b-\frac{\varepsilon}{2}z > 0$. Using $(2-\delta^2)/(1-\delta^2) \geq 2$ and $(1-\frac{\varepsilon}{2})z-1 > 0$, Eq. (13) is implied by

$$2(7-8(1-\varepsilon)z_b+(1-\frac{\varepsilon}{2})z) > (1-\frac{\varepsilon}{2})z-1$$

which is equivalent to

$$z > 16z_b\frac{1-\varepsilon}{1-\frac{\varepsilon}{2}} - \frac{15}{1-\frac{\varepsilon}{2}}.$$

This inequality is true for our choice of z. It remains to show that our z satisfies Eq. (12). Substituting for z we see that the quadratic term in z_b cancels and we obtain

$$\Big(17-\frac{4}{(1-\varepsilon)^2}\Big) + 16z_b\Big(\frac{7}{1-\frac{\varepsilon}{2}} - 17\varepsilon\Big) > 0.$$

This linear inequality is satisfied (for $z_b \geq 0$) because both its constant coefficient and the coefficient of z_b are positive for $0 \leq \varepsilon < \frac{1}{2}$. □

Using this result, we also obtain a second, "asymmetric" direct product theorem when α_0, α_1 and β_0, β_1 are all mixed states:

Corollary 1. *Let $0 \leq \varepsilon < \frac{1}{2}$ and $\alpha_0, \alpha_1, \beta_0, \beta_1$ be density matrices. Let $a = \|\alpha_0 - \alpha_1\|_{tr}$, $b = D_\varepsilon(\beta_0, \beta_1)$, and $p = D_{\varepsilon/2}(\alpha_0\otimes\beta_0, \alpha_0\otimes\beta_1, \alpha_1\otimes\beta_0, \alpha_1\otimes\beta_1)$. Then $p \leq 32\,a\cdot b$.*

Proof. The idea is to work with purifications of α_0 and α_1. By Uhlmann's theorem [25, p.410] there exist purifications $|\tilde\alpha_0\rangle$ and $|\tilde\alpha_1\rangle$ that preserve the fidelity, i.e., $F(\alpha_0,\alpha_1) = F(|\tilde\alpha_0\rangle, |\tilde\alpha_1\rangle) = |\langle\tilde\alpha_0|\tilde\alpha_1\rangle|$. Using known properties of the fidelity [25, Section 9.2.3], we have

$$F(\alpha_0,\alpha_1) \geq 1 - \|\alpha_0-\alpha_1\|_{tr} = 1-a.$$

Hence $1-|\langle\tilde\alpha_0|\tilde\alpha_1\rangle|^2 \leq 2a$. If $\tilde\alpha_i = |\tilde\alpha_i\rangle\langle\tilde\alpha_i|$, then

$$\begin{aligned} p &= D_{\varepsilon/2}(\alpha_0\otimes\beta_0, \alpha_0\otimes\beta_1, \alpha_1\otimes\beta_0, \alpha_1\otimes\beta_1) \\ &\leq D_{\varepsilon/2}(\tilde\alpha_0\otimes\beta_0, \tilde\alpha_0\otimes\beta_1, \tilde\alpha_1\otimes\beta_0, \tilde\alpha_1\otimes\beta_1)\end{aligned}$$

because one can obtain α_0, α_1 by tracing out the purification degrees of freedom of $\tilde\alpha_0, \tilde\alpha_1$. Theorem 1 now gives $p \leq 16(1-|\langle\tilde\alpha_0|\tilde\alpha_1\rangle|^2)\cdot b \leq 32\,a\cdot b$. □

4. SHARED RANDOMNESS CAN BE EXPONENTIALLY STRONGER THAN QUANTUM COMMUNICATION

4.1 The problem

In this section we analyze the following communication problem P_1 in the SMP model:

Alice's input: strings $x, s \in \{0,1\}^n$, with Hamming weight $|s| = n/2$
Bob's input: a string $y \in \{0,1\}^n$
Goal: the referee should output (i, x_i, y_i) for some i such that $s_i = 1$

We allow the referee a small constant error probability $\varepsilon < 1/8$. In the next two subsections we show that this problem is easy if we have classical communication and shared randomness, and hard if we have quantum communication without shared randomness:[3]

THEOREM 2. *For the relational problem P_1 defined above we have*

$$R_\varepsilon^{\|,pub}(P_1) \leq O(\log n) \text{ and } Q_\varepsilon^{\|}(P_1) \geq \Omega(n^{1/3}).$$

4.2 Upper bound with classical communication and shared randomness

Shared randomness gives the parties enough coordination to easily solve this problem. Alice and Bob just send (i, x_i, s_i) and (i, y_i), respectively, to the referee for $\log(1/\varepsilon)$ public random i's. With probability $1-\varepsilon$, $s_i = 1$ for at least one of those i's and the referee outputs the corresponding (i, x_i, y_i). With probability ε he doesn't see an i for which $s_i = 1$, in which case he outputs something random. Hence $R_\varepsilon^{\|,pub}(P) \leq O(\log n \log(1/\varepsilon))$.

4.3 Lower bound for quantum communication with private randomness

Consider some quantum protocol that solves our problem with error probability $\varepsilon < 1/8$, and where the messages that Alice and Bob send to the referee are at most q qubits long. Our goal is to show $q \geq \Omega(n^{1/3})$.

First consider the mixed state message β_y that Bob sends given input y. For $i \in [n]$, let

$$\beta_{i0} = \frac{1}{2^{n-1}} \sum_{y: y_i = 0} \beta_y$$

be the uniform mixture of all β_y with $y_i = 0$ and define β_{i1} similarly. Let $b_i = D_{4\varepsilon}(\beta_{i0}, \beta_{i1})$. Then by the random access code argument (Lemma 1) we have

$$\sum_{i=1}^n b_i(1 - H(4\varepsilon)) \leq q.$$

By Markov's inequality, there is a set S of $n/2$ i's such that $b_i \leq 2q/n(1 - H(4\varepsilon)) \leq O(q/n)$ for all $i \in S$. We fix Alice's input s to be the n-bit string with support corresponding to S.

[3]We believe the problem remains hard if we drop Alice's input x, but our proof doesn't seem to work in that case.

We now analyze Alice's message. Let α_x be the mixed state she sends given input x and our fixed s. Define α_{i0} as the uniform mixture of all α_x with $x_i = 0$, similarly define α_{i1}, and $a_i = \|\alpha_{i0} - \alpha_{i1}\|_{tr}$. The optimal probability with which we can distinguish α_{i0} from α_{i1} is $\frac{1}{2} + \frac{a_i}{2}$. The random access code argument gives

$$\sum_{i=1}^n a_i^2 \leq O(q).$$

Now we look at the protocol. Let $X = X_1 \ldots X_n$ and $Y = Y_1 \ldots Y_n$ be uniformly distributed random variables giving Alice's first and Bob's only input, and I, B_1, B_2 be the random variables describing the referee's output. We call an index $i \in [n]$ *good*, if the protocol is correct with high probability when it outputs $(i, *, *)$:

$$i \text{ is good iff } i \in S \text{ and}$$
$$\Pr[B_1 = X_i, B_2 = Y_i \mid I = i] \geq 1 - 2\varepsilon.$$

The index is called *bad* otherwise. Define $p_i = \Pr[I = i]$ to be the probability that the referee outputs something of the form $(i, *, *)$. Because the protocol is correct with probability at least $1 - \varepsilon$, a Markov argument shows that the good indices have most of the probability:

$$1 - \varepsilon \leq \sum_{\text{good } i} p_i + \sum_{\text{bad } i} (1 - 2\varepsilon) p_i = 1 - 2\varepsilon + 2\varepsilon \sum_{\text{good } i} p_i,$$

hence

$$\frac{1}{2} \leq \sum_{\text{good } i} p_i.$$

Note that for each good i we can use the protocol to get a $(p_i, 2\varepsilon)$-predictor for $X_i Y_i$: just run the protocol and return '?' if the protocol's output is not of the form $(i, *, *)$, and otherwise return the last two bits of the protocol's output. Therefore Corollary 1 implies $p_i \leq O(a_i b_i)$. Also, $b_i \leq O(q/n)$ for all good i, hence

$$\frac{1}{2} \leq \sum_{\text{good } i} p_i \leq \sum_{\text{good } i} O(a_i b_i) \leq O\left(\frac{q}{n} \sum_{i=1}^n a_i\right)$$
$$\leq O\left(\frac{q}{n} \sqrt{n \sum_{i=1}^n a_i^2}\right) \leq O\left(\frac{q^{3/2}}{n^{1/2}}\right),$$

where we applied Cauchy-Schwarz in the fourth step. This implies $q \geq \Omega(n^{1/3})$.

Remark: The best no-shared-randomness protocol we know for P_1 communicates $O(\sqrt{n})$ bits. The idea is to arrange the n-bit inputs in a $\sqrt{n} \times \sqrt{n}$ matrix. Alice picks a random row index in $[\sqrt{n}]$, and then sends that index and the indexed row of x and of s to the referee. Bob picks a random column index in $[\sqrt{n}]$, and then sends that index and the indexed column of y to the referee. The row and the column intersect in exactly one (uniformly random) point $i \in [n]$. With probability $1/2$, $s_i = 1$ and we are done. Repeating this a few times in parallel reduces the error probability to a small constant. A matching lower bound would follow from the general direct product theorem $p \leq O(ab)$, for the case of the 2-register identification problem where both sides are allowed to be mixed.

5. SHARED ENTANGLEMENT CAN BE EXPONENTIALLY STRONGER THAN QUANTUM COMMUNICATION WITH SHARED RANDOMNESS

5.1 The problem

For n a power of 2, consider the following relational problem P_2, inspired by a one-way communication problem due to Bar-Yossef et al. [3]:

Alice's input: a perfect matching $M \subset \binom{[n]}{2}$ and a string $x \in \{0,1\}^{n/2}$ containing a bit x_e for each $e \in M$
Bob's input: a string $y \in \{0,1\}^n$
Goal: the referee should output $(i, j, x_{(i,j)}, y_i \oplus y_j)$ for some edge $(i,j) \in M$

We show that this problem is easy if we have classical communication and prior entanglement, and hard if we have quantum communication without entanglement:

THEOREM 3. *For the relational problem P_2 defined above we have*

$$R_\varepsilon^{\|,ent}(P_2) \leq O(\log n) \text{ and } Q_\varepsilon^{\|,pub}(P_2) \geq \Omega((n/\log n)^{1/3}).$$

5.2 Upper bound with classical communication and entanglement

The following protocol solves the problem with success probability 1, using $O(\log n)$ classical bits of communication and $\log n$ EPR-pairs shared between Alice and Bob. It is a modification of an unpublished protocol due to Harry Buhrman [7], which is in turn based on a one-way protocol from [3]. The starting state of Alice and Bob is

$$\frac{1}{\sqrt{n}} \sum_{i \in \{0,1\}^{\log n}} |i\rangle|i\rangle.$$

Bob adds his bits as phases:

$$\frac{1}{\sqrt{n}} \sum_i |i\rangle(-1)^{y_i}|i\rangle.$$

Alice measures with the $n/2$ projectors $E_{ij} = |i\rangle\langle i| + |j\rangle\langle j|$ induced by the $n/2$ pairs $(i,j) \in M$. This gives her a random $(i,j) \in M$ and the resulting joint state is

$$\frac{1}{\sqrt{2}} \left(|i\rangle(-1)^{y_i}|i\rangle + |j\rangle(-1)^{y_j}|j\rangle\right).$$

Now both players apply a Hadamard transform to each of the $\log n$ qubits of their part of the state, which becomes (ignoring normalization)

$$\sum_{k,\ell} \left((-1)^{y_i + (k+\ell)\cdot i} + (-1)^{y_j + (k+\ell)\cdot j}\right) |k\rangle|\ell\rangle.$$

Note that $|k\rangle|\ell\rangle$ has non-zero amplitude iff $y_i + (k+\ell)\cdot i = y_j + (k+\ell)\cdot j \mod 2$, equivalently

$$(k+\ell)\cdot(i+j) = y_i \oplus y_j.$$

Alice and Bob both measure their part of the state in the computational basis, obtaining some k and ℓ, respectively, satisfying the above equality. Alice sends i, j, k, and $x_{(i,j)}$ to the referee, Bob sends ℓ; a total of $O(\log n)$ bits of communication. The referee calculates $y_i \oplus y_j$ from i, j, k, ℓ and outputs $(i, j, x_{(i,j)}, y_i \oplus y_j)$.

5.3 Lower bound for quantum communication without entanglement

We make use of some ideas from the classical lower bound of Bar-Yossef et al. [3]. For $k \in \{0, \ldots, n/2-1\}$, let M_k denote the matching $\{(i, (i+k-1 \mod n/2)+n/2+1\}_{i=1}^{n/2}$. For example, $M_1 = \{(1, n/2+2), (2, n/2+3), (3, n/2+4), \ldots, (n/2-1, n), (n/2, n/2+1)\}$. We will prove our lower bound for the special case where Alice's matching is one of the M_k.[4] Consider a quantum protocol where Alice and Bob share randomness but no entanglement, each communicates at most q qubits to the referee, and they solve problem P_2 with error probability $\varepsilon < 1/16$ for each input. Our goal is to show $q \geq \Omega((n/\log n)^{1/3})$.

We consider the following input distribution. Let K be a uniformly random number between 0 and $n/2-1$, M_K be Alice's first input, and $X \in \{0,1\}^{n/2}$ and $Y \in \{0,1\}^n$ be uniformly distributed random variables for Alice's second and Bob's only input. Since the protocol has error at most ε for all inputs, we can (and will) fix a value for the shared randomness such that the resulting protocol has average error at most ε under the above input distribution.

Let α_{kx} be Alice's message on input M_k, x. For edge $e = (i,j) \in M_k$, define α_{ke0} as the uniform mixture of all α_{kx} with $x_e = 0$, similarly define α_{ke1}, and $a_{ke} = \|\alpha_{ke0} - \alpha_{ke1}\|_{tr}$. The optimal probability with which we can distinguish α_{ke0} from α_{ke1} is $1/2 + a_{ke}/2$. Hence for every k, the random access code argument gives

$$\sum_{e \in M_k} a_{ke}^2 \leq O(q).$$

Let β_y be Bob's message on input y. For any $e = (i,j)$ (not necessarily part of any matching), define β_{e0} as the uniform mixture over all β_y with $y_i \oplus y_j = 0$ and similarly define β_{e1}. Let $b_e = D_{8\varepsilon}(\beta_{e0}, \beta_{e1})$. We now prove two claims upper bounding sums of these b_e.

CLAIM 3. *For any forest (i.e., acyclic graph) F on $[n]$ we have* $\sum_{e \in F} b_e \leq O(q)$.

PROOF. Denote by $|F|$ the number of edges in F. For every $e = (i,j) \in F$ we can obtain a $(b_e, 8\varepsilon)$-predictor for the bit $Y_i \oplus Y_j$ given the q-qubit state β_Y. Intuitively, since F is a forest, these $|F|$ bits are independent and therefore represent $|F|$ bits of information. To make this formal, define for each $w \in \{0,1\}^{|F|}$ the set

$$T_w = \{y \in \{0,1\}^n \mid \forall e = (i,j) \in F, y_i \oplus y_j = w_e\}.$$

Since F is a forest, $\{T_w\}_{w \in \{0,1\}^{|F|}}$ is a partition of $\{0,1\}^n$ into $2^{|F|}$ sets of size $2^{n-|F|}$ each.

For any bit string $w \in \{0,1\}^{|F|}$ we define ξ_w as the uniform mixture of β_y over all $y \in T_w$. For each $e \in F$, define

[4] Note that if we restrict attention to so few matchings, then Alice can communicate her matching to the referee in $\log n$ classical bits. Hence her second input x, which seems somewhat redundant at first sight, is actually crucial for our lower bound. Without it, there would be a cheap qubit protocol (Bob just sends the uniform superposition with his n bits as phases).

ξ_{e0} as the uniform mixture of ξ_w over all w with $w_e = 0$ and similarly define ξ_{e1}. Then, it is easy to see that $\xi_{e0} = \beta_{e0}$ and $\xi_{e1} = \beta_{e1}$. Hence, $D_{8\varepsilon}(\xi_{e0}, \xi_{e1}) = b_e$ and by applying the random access code argument to the encoding of w as the q-qubit state ξ_w, we get $\sum_{e \in F} b_e(1 - H(8\varepsilon)) \leq q$. □

CLAIM 4. $\sum_{k=0}^{n/2-1} \sum_{e \in M_k} b_e^2 \leq O(q^2 \log n)$.

PROOF. By construction all our M_k's are disjoint, hence the set $M = \cup_k M_k$ contains each edge in the above sum exactly once. Making some bijection between edges in M and numbers $\ell \in [|M|]$, we order the b_e in non-increasing order as

$$b_1 \geq b_2 \geq \cdots \geq b_{|M|}.$$

Now consider the graph consisting of the first ℓ edges in this ordering. This graph must contain at least $\sqrt{2\ell}$ non-isolated vertices, since v vertices give only $\binom{v}{2} \leq v^2/2$ distinct edges. Let F be a forest consisting of a spanning tree for each connected component of this graph. This F has at least $\sqrt{2\ell}/2 = \sqrt{\ell/2}$ edges, and for each of those edges e we have $b_e \geq b_\ell$. Now we use Claim 3:

$$\sqrt{\frac{\ell}{2}} \cdot b_\ell \leq \sum_{e \in F} b_e \leq O(q).$$

Hence for all $\ell \leq |M|$ we have $b_\ell \leq O(q/\sqrt{\ell})$. Summing over all ℓ gives

$$\sum_{e \in M} b_e^2 = \sum_{\ell=1}^{|M|} b_\ell^2 \leq \sum_{\ell=1}^{n^2/4} O(q^2/\ell) \leq O(q^2 \log n). \quad \square$$

Since the protocol has average error at most ε, by Markov's inequality there is a set $\mathcal{M}$ of at least $n/4$ of our matchings M_k such that the protocol has error at most 2ε for that M_k and uniformly random X and Y. Since $\mathcal{M}$ contains at least $n/4$ elements, Claim 4 implies there is a matching $M_k \in \mathcal{M}$ such that

$$\sum_{e \in M_k} b_e^2 \leq O\left(\frac{q^2 \log n}{n}\right).$$

We now fix this matching on Alice's side. Let I, J, B_1, B_2 be the random variables giving the referee's output. Suppose we run the protocol with M_k, and uniformly random x and y as input. We call an edge (i, j) *good*, if the protocol is correct with high probability when it outputs $(i, j, *, *)$:

$$e = (i, j) \text{ is good iff } e \in M_k \text{ and}$$
$$\Pr[B_1 = X_e, B_2 = Y_i \oplus Y_j \mid I = i, J = j] \geq 1 - 4\varepsilon.$$

The edge is called *bad* otherwise. Let $p_e = \Pr[I = i, J = j]$ be the probability that the protocol outputs edge e. Since $M_k \in \mathcal{M}$, the success probability (averaged over x and y) is at least $1 - 2\varepsilon$, so by a Markov argument, the good edges must have most of the probability:

$$1 - 2\varepsilon \leq \sum_{\text{good } e} p_e + \sum_{\text{bad } e} p_e(1 - 4\varepsilon) = 1 - 4\varepsilon + 4\varepsilon \sum_{\text{good } e} p_e,$$

hence

$$\frac{1}{2} \leq \sum_{\text{good } e} p_e.$$

For every good edge e, we can construct a $(p_e, 4\varepsilon)$-predictor for $(X_e, Y_i \oplus Y_j)$. Hence, by Corollary 1, $p_e \leq O(a_{ke} b_e)$. Using Cauchy-Schwarz:

$$\begin{aligned} \frac{1}{2} &\leq \sum_{\text{good } e} p_e \leq \sum_{\text{good } e} O(a_{ke} b_e) \\ &\leq O\left(\sqrt{\sum_{\text{good } e} a_{ke}^2 \cdot \sum_{\text{good } e} b_e^2}\right) \leq O\left(\sqrt{\frac{q^3 \log n}{n}}\right). \end{aligned}$$

This implies the promised bound $q \geq \Omega((n/\log n)^{1/3})$.

Remark: Our bound is tight up to $\log n$ factors. To see this, we briefly sketch a protocol which uses $O(n^{1/3} \log n)$ qubits of communication: Alice and Bob use their shared randomness to fix a subset $S \subset [n]$ of size $n^{2/3}$. With high probability the number of edges from M contained in $S \times S$ is roughly $n^{1/3}$. For each of the edges $(i, j) \in M \cap S \times S$, Alice sends $(i, j, x_{(i,j)})$ to the referee, which is $O(n^{1/3} \log n)$ bits of communication. Bob prepares $n^{1/3}$ copies of the state

$$\frac{1}{\sqrt{|S|}} \sum_{i \in S} (-1)^{y_i} |i\rangle \tag{14}$$

and sends them to the referee, giving a total of $O(n^{1/3} \log n)$ qubits of communication. On each of the copies, the referee measures with the projectors $E_{ij} = |i\rangle\langle i| + |j\rangle\langle j|$ induced by the edges in S that Alice has sent, completed by $E_{garbage} = I - \sum E_{ij}$. Given the state in Eq. (14), the probability to not measure "garbage" is roughly $n^{-1/3}$. This means that with some constant probability the referee will measure one of the edges E_{ij} on one of the states Bob sent. This state then collapses to $\frac{1}{\sqrt{2}}((-1)^{y_i}|i\rangle + (-1)^{y_j}|j\rangle)$, and a measurement in the basis $|i\rangle \pm |j\rangle$ gives $y_i \oplus y_j$.

6. CONCLUSION AND FUTURE WORK

We studied the bounded-error quantum state identification problem and proved a direct product theorem for two independent instances of this problem (one involving pure states) using SDP duality. We applied our direct product theorem to obtain two exponential separations in the simultaneous message passing model of communication complexity. These two separations nicely complement each other: the first shows that shared randomness is much more powerful than private randomness, the second shows that prior entanglement is much more powerful than shared randomness. Moreover, both separations are shown in the strongest possible sense: the stronger model is restricted to classical communication while the weaker model is allowed quantum communication.

We identify some interesting open questions. First, for the bounded-error quantum state identification problem, prove the direct product theorem $p \leq O(ab)$ in the general case where both sides have mixed states instead of one side pure and one side mixed. That result would lift, for instance, our quantum communication lower bound for the problem P_1 to the optimal $\Omega(\sqrt{n})$. Second, show similar communication complexity separations for decision problems (Boolean functions, possibly with a promise on the input) instead of for relational problems. Finally, we hope our direct product theorem will be useful for other applications as well.

Acknowledgments

We thank Harry Buhrman for permission to include his protocol, which we modified to the one of Section 5.2. DG is grateful to Richard Cleve for helpful discussions.

7. REFERENCES

[1] A. Ambainis. Communication complexity in a 3-computer model. *Algorithmica*, 16(3):298–301,1996.

[2] L. Babai and P. G. Kimmel. Randomized simultaneous messages: Solution of a problem of Yao in communication complexity. In *Proceedings of 12th IEEE Conference on Computational Complexity*, p. 239–246, 1997.

[3] Z. Bar-Yossef, T. S. Jayram, and I. Kerenidis. Exponential separation of quantum and classical one-way communication complexity. In *Proceedings of 36th ACM STOC*, p. 128–137, 2004.

[4] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *Proceedings of 17th IEEE Conference on Computational Complexity*, p. 93–102, 2002.

[5] R. Bhatia. *Matrix Analysis*. Number 169 in Graduate Texts in Mathematics. Springer, 1997.

[6] H. Buhrman. Quantum computing and communication complexity. *EATCS Bulletin*, 70:131–141, Feb. 2000.

[7] H. Buhrman. Personal communication, Nov. 2003.

[8] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16), Sep. 26, 2001.

[9] R. Cleve and H. Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201–1204, 1997.

[10] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[11] Y. C. Eldar. A semidefinite programming approach to optimal unambiguous discrimination of quantum states. *IEEE Transactions on Information Theory*, 49:446–456, 2003.

[12] D. Gavinsky. A note on shared randomness and shared entanglement in communication. quant-ph/0505088, 12 May 2005.

[13] D. Gavinsky, J. Kempe, and R. de Wolf. Quantum communication cannot simulate a public coin. quant-ph/0411051, 8 Nov 2004.

[14] D. Gavinsky, J. Kempe, O. Regev, and R. de Wolf. Bounded-error quantum state identification and exponential separations in communication complexity. quant-ph/0511013, 2 Nov 2005.

[15] A. Golinsky and P. Sen. A note on the power of quantum fingerprinting. quant-ph/0510091, Dec. 2003.

[16] A. S. 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*, 9:177–183, 1973.

[17] C. King and M-B. Ruskai. Comments on multiplicativity of maximal p-norms when $p = 2$. In O. Hirota, editor, *Quantum Information, Statistics, Probability (Festschrift for A. Holevo)*. Rinton Press, 2004.

[18] H. Klauck. Quantum communication complexity. In *Proceedings Workshop on Boolean Functions and Applications at 27th ICALP*, p.241–252, 2000.

[19] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999.

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

[21] L. Lovász. Semidefinite programs and combinatorial optimization. At http://research.microsoft.com/users/lovasz/notes.htm, 2000.

[22] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of 40th IEEE FOCS*, p. 369–376, 1999.

[23] I. Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67–71, 1991.

[24] I. Newman and M. Szegedy. Public vs. private coin flips in one round communication games. In *Proceedings of 28th ACM STOC*, p. 561–570, 1996.

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

[26] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1996.

[27] R. de Wolf. Quantum communication and complexity. *Theoretical Computer Science*, 287(1):337–353, 2002.

[28] A. C-C. Yao. Quantum circuit complexity. In *Proceedings of 34th IEEE FOCS*, p. 352–360, 1993.

[29] A. C-C. Yao. On the power of quantum fingerprinting. In *Proceedings of 35th ACM STOC*, p. 77–81, 2003.