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

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

December 22, 2006

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 are 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.

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. Combined with an earlier separation in the other direction due to Bar-Yossef et al., this shows that the quantum SMP model is incomparable with the classical shared-randomness SMP model. 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 of a situation where entanglement buys you much more than quantum communication.

*University of Calgary. Supported in part by Canada's NSERC.

†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 and the EU fifth framework project RESQ, IST-2001-37559, and by the European Commission under the Integrated Project Qubit Applications (QAP) funded by the IST directorate as Contract Number 015848. Hospitality of the MSRI, Berkeley, where part of this work was done, is gratefully acknowledged.

‡Department of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Supported by an Alon Fellowship, the Binational Science Foundation, the Israel Science Foundation, and the European Commission under the Integrated Project Qubit Applications (QAP) funded by the IST directorate as Contract Number 015848.

§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 by the European Commission under the Integrated Project Qubit Applications (QAP) funded by the IST directorate as Contract Number 015848.

1 Introduction

1.1 Bounded-error quantum state identification

Suppose we are given one of two known mixed quantum states, α_0 or α_1, each with probability 1/2. Our goal is to identify which one we are given. It is well known that the optimal probability of outputting the correct answer (0 or 1) is $1/2 + \|\alpha_0 - \alpha_1\|_{tr}/2$, where $\|\cdot\|_{tr}$ is the trace norm (the sum of the singular values, divided by 2). In particular, if α_0 and α_1 are very close in trace norm, then even 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 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:

Definition 1.1 *Let A be a 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 of 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 [13], one can easily express quantities like this as the optimal value of a semidefinite program, as we do in Section 3.2.

Now, suppose we are given another identification problem in a second register. This register contains a 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 $p = O(ab)$. 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 [16].

Proving our intuition actually turned out to be quite a hard problem, and we indicate some reasons why in Section 3.1. In Section 3 we prove the $p = O(ab)$ 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

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

Notice that because of the $\varepsilon/2$ on the left hand side, this bound is slightly weaker than what we have promised; as we indicate in Section 3.1, this modification is (somewhat surprisingly) necessary. Our proof relies heavily on a semidefinite programming formulation for the quantities involved and on an analysis of their duals. The case when *both* sides can be mixed states is still open, but fortunately the special case where at least one of the two sides is pure suffices for our applications.

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 *simultaneous message passing* (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 *communication cost* of a protocol is the length of the total communication for a worst-case input, and the *communication complexity* of a problem is the cost of the best protocol that solves the 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

[1] This actually gives slightly worse conditional error probability $1 - (1 - \varepsilon)^2 = 2\varepsilon - \varepsilon^2$ for the prediction of XY, so it could even be that to achieve error as low as ε we need a p that is strictly smaller than ab.

optimal communication for Alice is determined by the number of distinct rows in the communication matrix (and for Bob by the number of distinct columns), 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 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 $O(\log n)$ [22], 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, 23, 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. [10] 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 classical shared-randomness protocols. Roughly speaking, their technique, known as "quantum fingerprinting", may be viewed as replacing the shared randomness by a quantum superposition. Later, Bar-Yossef et al. [3] showed an exponential separation even between quantum protocols and classical protocols with *shared* randomness. Their separation was for a relational problem.

1.2.1 Shared randomness beats quantum communication

The fingerprinting idea of [10] was generalized by Yao [28], 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 [14, 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 holds in general: can *every* classical shared-randomness SMP protocol be efficiently simulated by some SMP protocol that sends qubits but shares neither randomness nor entanglement? Our first result, 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 quantum protocol without shared randomness needs $\Omega(n^{1/3})$ qubits of communication. The quantum lower bound relies crucially on our direct product theorem for bounded-error state identification. This shows for the first time that the resource of shared randomness cannot be efficiently traded for quantum communication.

It is not hard to see that Yao's exponential simulation can be made to work for relations as well. Our quantum lower bound shows that this 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

[2]Kremer et al. [18] 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.

with shared randomness. Accordingly, the quantum SMP model is incomparable with the classical shared-randomness SMP model.

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 of the form $\frac{1}{2}(|00\rangle + |11\rangle)$) from quantum protocols with shared randomness.

The additional power that prior entanglement gives is one of the 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 [11] and the model of quantum communication without entanglement [27] are essentially equivalent.[3]

Our second separation, given in Section 5, shows that the situation is very different in the simultaneous message passing model. 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 states and measurements

The essentials needed for this paper are quantum states, distances, and measurements. 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 the 2^m-dimensional Hilbert space $\mathbb{C}^{2^m}$. 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.

The *trace distance* between two mixed states ρ and σ is defined as $1/2$ times the sum of the singular values of the matrix $\rho - \sigma$, and denoted by $\|\rho - \sigma\|_{tr}$. A basic fact is that the optimal probability with which we can distinguish ρ from σ, equals $1/2 + \|\rho - \sigma\|_{tr}/2$. The *fidelity* of ρ and σ is $F(\rho, \sigma) = \mathrm{Tr}[\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}]$, where "Tr[$\cdot$]" is the trace function (sum of diagonal entries).

[3] Note that classical communication complexity with entanglement and quantum communication complexity with entanglement are equivalent up to a factor of 2 for one-round and many-round protocols, thanks to teleportation.

The trace distance is close to 0 iff the fidelity is close to 1; we refer to Chapter 9 of Nielsen and Chuang [24] for more details.

A pure state in some bipartite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ is called *entangled* if it cannot be written as a product state $|\phi\rangle_A \otimes |\psi\rangle_B$. The 2-qubit *EPR-pair*

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

is the canonical example of an entangled pure state.

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 $\mathrm{Tr}[E_i \rho]$. Again, see Nielsen and Chuang [24] for more details.

2.2 Communication complexity

We now give a somewhat informal description of the simultaneous message passing model used in our two applications. For a more detailed description, we refer to Kushilevitz and Nisan [19] for classical communication complexity and to the surveys [17, 8, 26] 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 the optimal communication cost 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 variant that allows quantum communication.

2.3 The random access code argument

Here we describe a slight extension of a quantum information theoretic argument due to Ashwin Nayak [21] that we will apply several times in our communication complexity lower bounds. We call this the "random access code argument".

We quickly introduce the basic information theory needed. For more details, we refer to [12] for classical information theory and to [24] for quantum information theory. We start with classical information theory. If A is a random variable with probability distribution $p_1, \ldots, p_m$ ($p_i \geq 0$, $\sum_i p_i = 1$), then its entropy is defined by

$$H(A) = H(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i.$$

This always lies between 0 and $\log m$. If $m = 2$, we abbreviate $H(p, 1-p)$ to $H(p)$. If random variables A and B have some joint (possibly correlated) distribution, then the *conditional* entropy of A given B is

$$H(A \mid B) = H(A, B) - H(B) = \mathrm{E}_b[H(A \mid B = b)],$$

where $H(A,B)$ is the entropy of the joint distribution, "$A \mid B = b$" is the distribution of A conditioned on the event "$B = b$", and the expectation is taken according to the marginal distribution of B. The *mutual information* between A and B is

$$I(A:B) = H(A) + H(B) - H(A,B) = H(A) - H(A \mid B).$$

This quantity is always nonnegative. The following claim lower bounds the mutual information between a uniform bit and its predictor.

Claim 2.1 *If A is a uniform random bit and B is another random variable that is a (λ, ε)-predictor of A for some $\lambda \geq 0$ and $\varepsilon \geq 1/2$, then $I(A:B) \geq \lambda(1 - H(\varepsilon))$.*

Proof: By definition,

$$I(A:B) = H(A) - H(A \mid B).$$

The first term is 1 by our assumption on A. For the second term, let $(p_0, p_1, p_?)$ be the marginal distribution of B, let q_0 be the probability that $A = 1$ conditioned on $B = 0$ and similarly let q_1 be the probability that $A = 0$ conditioned on $B = 1$. Our assumption on B is then that $p_0 + p_1 \geq \lambda$ and $(p_0 q_0 + p_1 q_1)/(p_0 + p_1) \leq \varepsilon$. Using the definition of conditional entropy, we have

$$\begin{aligned} H(A \mid B) &\leq p_? \cdot 1 + p_0 H(q_0) + p_1 H(q_1) \\ &\leq p_? + (p_0 + p_1) H\left(\frac{p_0 q_0 + p_1 q_1}{p_0 + p_1}\right) \\ &\leq 1 - (p_0 + p_1)(1 - H(\varepsilon)) \end{aligned}$$

where the second inequality follows from the concavity of $H(\cdot)$. The claim now follows from $p_0 + p_1 \geq \lambda$. ■

We now describe some basics of *quantum* information theory. If ρ is an m-dimensional mixed state with eigenvalues $\lambda_1, \ldots, \lambda_m$, then these form a probability distribution, and we define the *von Neumann entropy* of ρ as

$$S(\rho) = H(\lambda_1, \ldots, \lambda_m).$$

If

$$\rho_{AB} = \sum_{i,i',j,j'} \alpha_{ii'jj'} |i\rangle\langle i'| \otimes |j\rangle\langle j'|$$

is a bipartite mixed state in the tensor space $A \otimes B$, where i, i' each range over an orthonormal basis for A and j, j' each range over an orthonormal basis for B, then we define the state of the A-register via the partial trace:

$$\rho_A = \sum_{i,i',j,j'} \alpha_{ii'jj'} |i\rangle\langle i'| \otimes \mathrm{Tr}(|j\rangle\langle j'|) = \sum_{i,i'} \left(\sum_j \alpha_{ii'jj} \right) |i\rangle\langle i'|.$$

We often write $S(A)$ for $S(\rho_A)$ and $S(A,B)$ for $S(\rho_{AB})$. Equipped with these definitions we can define the *conditional von Neumann entropy*, the *quantum mutual information*, and the *quantum conditional mutual information* by

$$\begin{aligned} &S(A \mid B) = S(A,B) - S(B), \\ &S(A:B) = S(A) + S(B) - S(A,B) = S(A) - S(A \mid B), \text{ and} \\ &S(A:B \mid C) = S(A,C) + S(B,C) - S(A,B,C) - S(C) = S(A \mid C) - S(A \mid B,C). \end{aligned}$$

The *strong subadditivity* of von Neumann entropy [24, Theorem 11.15.1] says that the quantum conditional mutual information is always nonnegative, or equivalently, that removing a conditional cannot increase entropy, i.e.,

$$S(A \mid B, C) \leq S(A \mid C).$$

Finally, we need to describe the Holevo bound. Consider a classically correlated mixed state ρ_{XM}, i.e., a state of the form

$$\rho_{XM} = \sum_{x \in \{0,1\}^n} p_x |x\rangle\langle x| \otimes M_x,$$

where M_x is a mixed state on the second register that depends on x. For such states, the quantum mutual information becomes

$$S(X : M) = S(M) - S(M \mid X) = S(M) - \sum_{x \in \{0,1\}^n} p_x S(M_x). \tag{2}$$

Assume we apply a quantum measurement to the second register, and obtain some classical random variable Y as outcome. Then the Holevo bound [24, Theorem 12.1] implies that

$$I(X : Y) \leq S(X : M). \tag{3}$$

In other words, it says that the (classical) information on X we can obtain by a measurement of M is at most $S(X : M)$. Now we are in a position to state and prove the random access code argument.

Lemma 2.2 *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 $i \in [n]$ there exists a quantum measurement of M_X whose outcome is a $(\lambda_i, \varepsilon_i)$-predictor for X_i with some $\varepsilon_i \leq 1/2$ and $\lambda_i \geq 0$. 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: We will analyze the classically correlated state

$$\rho_{XM} = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} |x\rangle\langle x| \otimes M_x.$$

By (2), $S(X : M) \leq S(M) \leq q$ and hence it suffices to lower bound the quantum mutual information $S(X : M)$. By definition, we have the chain rule

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

where we abbreviate $X_{1:i-1} = X_1 \dots X_{i-1}$. Moreover, by strong subadditivity and the fact that $S(X_i \mid X_{1:i-1}) = S(X_i)$ we get

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

It therefore suffices to show that for each i, $S(X_i : M) \geq \lambda_i(1 - H(\varepsilon_i))$. For this, let B_i be the outcome of the measurement corresponding to i applied to M. By (3) we have $S(X_i : M) \geq I(X_i : B_i)$. Finally, we complete the proof by noting that Claim 2.1 implies that $I(X_i : B_i) \geq \lambda_i(1-H(\varepsilon_i))$. ■

2.4 Semidefinite programs

A semidefinite program (SDP) is a particular type of convex optimization problem. Although SDPs can be formulated in many equivalent forms, they all amount to the maximization of a linear function over the intersection of the cone of positive semidefinite matrices with some affine subspace. As all convex optimization problems, semidefinite programs enjoy a powerful duality theory that allows us to bound the optimal value of a program by the value achieved by any feasible solution to another program, known as the *dual program*. Moreover, under very mild conditions, the optimal values of the two programs coincide; this property is known as *strong duality*. We now describe these notions in some detail. We use the conic programming approach of Ben-Tal and Nemirovski [5] since it emphasizes the elegant symmetry between the primal and dual programs. For proofs and definitions of some of the terms below, see [5, Lecture 2]. For other references on the topic, see, e.g., [20, 25, 7].

We consider the space of Hermitian $n \times n$ matrices equipped with the Frobenius inner product given by $\langle A, B\rangle := \mathrm{Tr}[AB]$. This space is a real vector space of dimension n^2 (one for each diagonal entry, and two for each off-diagonal pair, corresponding to their real and imaginary parts). The set of all positive semidefinite matrices forms a *cone* in this space known as the (Hermitian) positive semidefinite cone.

A *(complex) semidefinite program* (SDP) is a maximization program of a linear function over the intersection of the positive semidefinite cone with some affine subspace.[4] In more detail, any linear subspace $\mathcal{L}$ in the space of Hermitian matrices, and Hermitian matrices B, D define the semidefinite program

$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}[BX] \\ \text{subject to} & X \succeq 0 \\ & X \in \mathcal{L} + D \end{array} \tag{4}$$

where the inequality in the second line means that X must be positive semidefinite. The set of *feasible* solutions is simply the set of X that satisfy the two conditions above. If this set is non-empty we say that the SDP is *feasible*. Moreover, if a feasible solution X satisfies $X \succ 0$, i.e., is positive definite, then we say that it is a *strictly feasible* solution and the SDP is said to be strictly feasible.

The method of Lagrange multipliers allows us to associate a *dual problem* to any optimization problem. In the case of SDPs, the Lagrange dual problem turns out to also be an SDP. Indeed, it

[4] Most of the literature deals with semidefinite programs over (real) symmetric matrices. Since both are conic programs, exactly the same theory applies.

is not difficult to show that the Lagrange dual of the "primal" SDP (4) is

$$\begin{array}{ll} \text{minimize} & \mathrm{Tr}[BD] + \mathrm{Tr}[DY] \\ \text{subject to} & Y \succeq 0 \\ & Y \in \mathcal{L}^{\perp} - B \end{array} \tag{5}$$

where $\mathcal{L}^{\perp}$ is the orthogonal subspace to $\mathcal{L}$. Notice that $\mathrm{Tr}[BD]$ is a constant term. Also note that the dual of (5) is exactly the primal SDP (4).

The basic property of a Lagrange dual is that any feasible solution to it gives a bound on the optimal value of the primal program. This is readily seen in our case: let Y be any feasible solution to (5). Then for any feasible solution X to (4),

$$\mathrm{Tr}[BD] + \mathrm{Tr}[DY] - \mathrm{Tr}[BX] = \mathrm{Tr}[XY] \geq 0$$

where we used that $X - D$ is orthogonal to $Y + B$ and that the inner product of any two positive semidefinite matrices is nonnegative. Hence the optimal value of (5) upper bounds the optimal value of (4). This property is known as *weak duality.*

In addition to weak duality, SDPs (and conic programs in general) usually satisfy that the optimal value of the dual program is *equal* to that of the primal program. This property is known as *strong duality.* One sufficient condition for strong duality to hold is simply that the dual (or primal) program is strictly feasible. This condition, known as Slater's condition, guarantees that we avoid those pathological cases where strong duality does not hold.

2.5 On the positive part of a matrix

We now prove a basic claim regarding the *positive part* of a Hermitian matrix which will be used in Section 3. 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. We define $\mathrm{Pos}(A) = A^+$. By definition, the trace of $\mathrm{Pos}(A)$ is given by the sum of the positive eigenvalues of A. Equivalently,

$$\mathrm{Tr}[\mathrm{Pos}(A)] = \max_k \sum_{i=1}^{k} \lambda_i \tag{6}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots$ are the eigenvalues of A.

Claim 2.3 *For any Hermitian matrices A, B the following holds.*

1. *If $A \preceq B$ then $A \preceq \mathrm{Pos}(B)$.*
2. *If $A \succeq 0$ then $\mathrm{Pos}(A \otimes B) = A \otimes \mathrm{Pos}(B)$.*
3. *If $A \preceq B$ then $\mathrm{Tr}[\mathrm{Pos}(A)] \leq \mathrm{Tr}[\mathrm{Pos}(B)]$.*

Note that it is *not* true that if $A \preceq B$ then $\mathrm{Pos}(A) \preceq \mathrm{Pos}(B)$.

Proof: The first part follows from $B \preceq \mathrm{Pos}(B)$. The second part can be seen by diagonalizing the matrices (note that the nonzero 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., [6]). If $A \preceq B$, then the vector of eigenvalues of A is submajorized by the vector of eigenvalues of B ([6], Eq. (II.16)). This means that if we order the eigenvalues of A (resp. B) as $\lambda_1 \geq \lambda_2 \geq \cdots$ (resp. $\mu_1 \geq \mu_2 \geq \cdots$) then for all $k \geq 1$, $\sum_{i=1}^{k} \lambda_i \leq \sum_{i=1}^{k} \mu_i$. The third part now follows from (6). ■

3 Bounded-error quantum state identification: Direct product

3.1 Why this is a delicate problem

We briefly recall the 2-register state identification problem from the introduction. In the first register we are given a quantum state α_X, with X a random bit, and a is the largest value for which we can get an (a, ε)-predictor for X. In the second register we are given β_Y, with Y a random bit, and b is the largest value for which we can get a (b, ε)-predictor for Y. We now want to know the optimal probability p such that there is a (p, ε)-predictor for XY. As mentioned in the introduction, intuition suggests that $p = O(ab)$. Before proceeding to prove a slightly weaker form of this statement (namely the special case where α_0 and α_1 are pure), we will pause to sketch two variants of the problem where the same intuition is provably *false*, even for pure states! This points to the subtleness of the state identification problem: seemingly small changes to the setup change everything.

First, suppose that instead of a (p, ε)-predictor for XY we want a (p, ε)-predictor for the parity $X \oplus Y$ of the two bits. This might be slightly easier than getting both bits separately, but intuition still suggests that because both registers are independent, the best we can do is predict both registers separately and output their parity if both measurements gave an answer. So we expect $p = O(ab)$. However, this intuition is *false*. Consider the following counterexample, with δ very small:

$$
\begin{aligned}
|\alpha_0\rangle &= |\beta_0\rangle = |0\rangle \\
|\alpha_1\rangle &= |\beta_1\rangle = \sqrt{1-\delta^2}|0\rangle + \delta|1\rangle
\end{aligned}
$$

It is not hard to convince oneself[5] that for any fixed $\varepsilon < 1/2$, the optimal a and b are $\Theta(\delta^2)$, so our intuition suggests $p = O(ab) = O(\delta^4)$ for the parity problem. However, if we apply the measurement with operator E_0 that projects onto the state $\frac{1}{\sqrt{2+\delta^2}}(\delta|00\rangle - |01\rangle - |10\rangle)$, $E_1 = 0$, and $E_? = I - E_0$, then on the parity-0 inputs $\alpha_0 \otimes \beta_0$ and $\alpha_1 \otimes \beta_1$ the measurement gives outcome 0 with probability roughly δ^2, while on the parity-1 inputs it gives the (incorrect) outcome 0 with probability only about δ^6, which is much smaller than $\varepsilon\delta^2$. Thus, both for 0 and for 1-inputs the probability to output the incorrect answer is at most ε (conditioned on actually outputting an answer), while the probability p to actually output an answer is of the same order as a and b instead of their product.

In our second example, we return to the original setting where we want to obtain a predictor for XY (not their parity). We consider the case where in the left hand side of Eq. (1) from the introduction we replace $\varepsilon/2$ with a slightly larger error parameter. Surprisingly, we show that in this case the bound $p = O(ab)$ is *false*. Choose ε to be, say, 0.49, and replace $\varepsilon/2$ in the left hand side of (1) with something slightly larger, say, 0.251.[6] To construct this example, we use the same states as in the previous example. For our choice of ε, we still have $a, b = \Theta(\delta^2)$. Now consider the measurement where operator E_{00} projects onto the state $\frac{1}{\sqrt{8/9+\delta^2}}(\delta|00\rangle - \frac{2}{3}|01\rangle - \frac{2}{3}|10\rangle)$, $E_{01} = E_{10} = E_{11} = 0$, and $E_? = I - E_{00}$. Then on the state $\alpha_0 \otimes \beta_0$ we get outcome 00 with probability roughly $9\delta^2/8$, while on each of the other three states this probability is roughly $\delta^2/8$. Conditioned on outputting an answer, our error probability is roughly $(3/8)/(9/8 + 3/8) = 1/4$, so we obtain a $(p, 0.251)$-predictor for XY, with $p \approx \delta^2/8$. We see that again, contrary to our intuition, p is of the same order as a and b.

[5] A rigorous proof can be obtained from the SDP formulation of this problem.

[6] With some effort, this example can be generalized to other values of ε.

Finally, to get a better feel for this problem and for why it is non-trivial, let us consider the classical case. This is the special case of the problem in which all states involved are classical probability distributions. In other words the density matrices α_0, α_1 are diagonal in the same basis and similarly for β_0, β_1.[7] In this case, one can give a characterization of the optimal measurement. Let α_0 (resp., α_1) correspond to some probability distribution on n elements with probabilities $p_1, \ldots, p_n$ (resp., $q_1, \ldots, q_n$). Assume without loss of generality that the n elements are sorted by non-increasing order of $\max\{p_i/(p_i+q_i), q_i/(p_i+q_i)\}$. For any $k \geq 1$, consider the measurement that maps the outcome i for $1 \leq i \leq k$ to either 0 if $p_i > q_i$ or 1 otherwise, and maps any outcome $i > k$ to '?'. This means that for each $i \leq k$ we output the guess (α_0 or α_1) that is more likely, conditioned on i. Note that $\max\{p_i/(p_i+q_i), q_i/(p_i+q_i)\}$ represents the probability that our guess is correct, given i. Then, for any error parameter ε, one can show that the best measurement is obtained by taking k as large as possible while still keeping the error probability of the resulting measurement below ε.[8]

Now assume we have probability distributions $\alpha_0, \alpha_1, \beta_0, \beta_1$ (equivalently, diagonal matrices) and we want to predict XY based on a sample from $\alpha_X \otimes \beta_Y$ (the tensor can be described classically as one sample from α_X together with one independent sample from β_Y). The optimal measurement in the two-register case can be obtained by a straightforward generalization of the measurement we have described in the single register case. As mentioned in the introduction, one might expect the optimal measurement to use the first register to predict X and the second register to predict Y separately, i.e., to be a tensor product measurement. It is perhaps somewhat surprising that this is *not* true in general, as can be seen using some simple examples. The intuitive reason for this is that if a sample (i, j) from $\alpha_X \otimes \beta_Y$ is such that i gives a very strong indication of (say) α_0, then we might be willing to predict the state $\alpha_0 \otimes \beta_0$ even if j gives only a weak indication of β_0.

Nevertheless, the direct product theorem of Eq. (1) does hold in the classical case, even when we replace $\varepsilon/2$ with ε. One proof of this is based on a similar approach to the one we will take in the quantum case: first, formulate the problem in terms of linear programs (which are very similar to the semidefinite programs that arise in the quantum case) and then bound the dual solution of the joint system. Bounding the dual solution is the most demanding step technically, and amounts to solving some inequalities on real numbers. In the general quantum case, this step involves some (rather involved) matrix inequalities that seem quite difficult to solve. In the special case that we consider below, these matrix inequalities turn out to have a sufficiently nice form that can be analyzed.

3.2 Proof of the direct product theorem

In this section we prove our main results about the 2-register quantum state identification problem by using the powerful technique of semidefinite programming duality (see Section 2.4). The main theorem is the following.

Theorem 3.1 *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*

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

[7] This is related to optimal detector design, see, e.g. [7], Section 7.3.

[8] To be precise, we should also allow non-integer k in the sense that when the outcome is $\lfloor k \rfloor + 1$, one should output either 0 or 1 (depending on whether $p_{\lceil k \rceil} > q_{\lceil k \rceil}$) with probability $k - \lfloor k \rfloor$ and '?' otherwise.

Before presenting the proof, let us mention two consequences of this theorem. First, notice that since α_0 and α_1 are pure,

$$D_\varepsilon(\alpha_0, \alpha_1) \geq D_0(\alpha_0, \alpha_1) \geq \frac{1}{2}(1 - |\langle\alpha_0|\alpha_1\rangle|^2),$$

where the last inequality follows by considering the projective measurement on $|\alpha_0\rangle$ and $|\alpha_0^\perp\rangle$. Hence Theorem 3.1 implies that whenever α_0, α_1 are pure,

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

Second, by using purifications, we can derive a useful bound even in the case where α_0 and α_1 are not necessarily pure states,

$$D_{\varepsilon/2}(\alpha_0 \otimes \beta_0, \alpha_0 \otimes \beta_1, \alpha_1 \otimes \beta_0, \alpha_1 \otimes \beta_1) = O(\|\alpha_0 - \alpha_1\|_{tr} \cdot D_\varepsilon(\beta_0, \beta_1)).$$

This "asymmetric" bound is given in the following corollary.

Corollary 3.2 *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 [24, 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 [24, Section 9.2.3], we have

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

This implies $1 - |\langle\tilde{\alpha}_0|\tilde{\alpha}_1\rangle|^2 \leq 2a$. With slight abuse of notation, let $\tilde{\alpha}_i = |\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i|$. Then,

$$p = D_{\varepsilon/2}(\alpha_0 \otimes \beta_0, \alpha_0 \otimes \beta_1, \alpha_1 \otimes \beta_0, \alpha_1 \otimes \beta_1) \leq D_{\varepsilon/2}(\tilde{\alpha}_0 \otimes \beta_0, \tilde{\alpha}_0 \otimes \beta_1, \tilde{\alpha}_1 \otimes \beta_0, \tilde{\alpha}_1 \otimes \beta_1)$$

because one can obtain α_0, α_1 by tracing out the purification degrees of freedom of $\tilde{\alpha}_0, \tilde{\alpha}_1$. Theorem 3.1 now gives $p \leq 16(1 - |\langle\tilde{\alpha}_0|\tilde{\alpha}_1\rangle|^2) \cdot b \leq 32\, a \cdot b$. ■

In the rest of this section we prove Theorem 3.1.

Proofof Theorem 3.1: We first write $b = D_\varepsilon(\beta_0, \beta_1)$ as the optimal value of an SDP. Recall that any measurement whose outcome is a (b, ε)-predictor outputs the correct answer with probability at least $1 - \varepsilon$ *conditioned* on outputting a guess (0 or 1, but not ?). Let E_0, E_1, $E_? \succeq 0$ be the three measurement operators with $E_0 + E_1 + E_? = I$. Then we require

$$\varepsilon \geq \Pr[\text{wrong guess} \mid \text{guess}] = \frac{\Pr[\text{wrong guess}]}{\Pr[\text{guess}]} = \frac{\frac{1}{2}\mathrm{Tr}[E_0\beta_1] + \frac{1}{2}\mathrm{Tr}[E_1\beta_0]}{\mathrm{Tr}\left[(E_0 + E_1)\beta\right]}, \tag{7}$$

where $\beta = \frac{1}{2}(\beta_0 + \beta_1)$ is the average state. To our knowledge there is no simple expression for $D_\varepsilon(\beta_0, \beta_1)$ in terms of β_0 and β_1. However, one can easily express it as the optimal value of an SDP. Indeed, for fixed density matrices β_0, β_1 and fixed $\varepsilon \in [0, 1/2)$, $b = D_\varepsilon(\beta_0, \beta_1)$ is given by the following SDP.

$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}[(E_0 + E_1)\beta] \\ \text{subject to} & E_0, E_1, E_? \succeq 0, \\ & E_0 + E_1 + E_? = I, \\ & \frac{1}{2}\mathrm{Tr}[E_0\beta_1] + \frac{1}{2}\mathrm{Tr}[E_1\beta_0] \leq \varepsilon \mathrm{Tr}[(E_0 + E_1)\beta]. \end{array}$$

The first two constraints state that $E_0, E_1, E_?$ form a valid quantum measurement. The last constraint bounds the conditional error probability, as in Eq. (7). By introducing an auxiliary real variable e, we obtain the equivalent form

$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}[(E_0 + E_1)\beta] \\ \text{subject to} & E_0, E_1, E_? \succeq 0, \ e \geq 0, \\ & E_0 + E_1 + E_? = I, \\ & e - \varepsilon \mathrm{Tr}[(E_0 + E_1)\beta] + \frac{1}{2}\mathrm{Tr}[E_0\beta_1] + \frac{1}{2}\mathrm{Tr}[E_1\beta_0] = 0. \end{array}$$

We can write this SDP in the form (4) by optimizing over Hermitian matrices X of dimension $d + d + d + 1$ where d is the dimension of β_0 and β_1. Namely, let B, D be square matrices of dimension $d + d + d + 1$ given by

$$B = \begin{pmatrix} \beta & & & \\ & \beta & & \\ & & 0 & \\ & & & 0 \end{pmatrix} \quad D = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & I & \\ & & & 0 \end{pmatrix},$$

and take $\mathcal{L}$ to be the space of Hermitian matrices whose first three diagonal blocks sum to zero and are orthogonal to

$$\begin{pmatrix} \frac{1}{2}\beta_1 - \varepsilon\beta & & & \\ & \frac{1}{2}\beta_0 - \varepsilon\beta & & \\ & & 0 & \\ & & & 1 \end{pmatrix}. \tag{8}$$

Then we obtain a program of the form (4) where the blocks on the diagonal of X correspond to E_0, E_1, $E_?$, and e. Notice that we did not restrict $\mathcal{L}$ to matrices whose elements outside the diagonal blocks are zero, as this does not change the optimal value of the program. This follows from the fact that the matrix obtained from any $X \succeq 0$ by replacing all its elements outside the diagonal blocks with zero is still positive semidefinite (see, e.g., [6], Eq. (II.39)).

The dual SDP is now given by (5). The space $\mathcal{L}^\perp$ is spanned by the matrix in (8) and the space of matrices whose first three diagonal blocks are identical. In other words, the space $\mathcal{L}^\perp$ consists of all matrices of the form

$$\begin{pmatrix} X_b + (\frac{1}{2}\beta_1 - \varepsilon\beta)z_b & & & \\ & X_b + (\frac{1}{2}\beta_0 - \varepsilon\beta)z_b & & \\ & & X_b & \\ & & & z_b \end{pmatrix}.$$

for some Hermitian matrix X_b and real z_b. A straightforward calculation now shows that the dual SDP can be written as

$$\begin{array}{ll} \text{minimize} & \mathrm{Tr}[X_b] \\ \text{subject to} & X_b \succeq 0, z_b \geq 0, \\ & X_b \succeq \frac{1}{2}\left((1 + \varepsilon z_b)\beta_0 + (1 - (1 - \varepsilon)z_b)\beta_1\right) =: X_1, \\ & X_b \succeq \frac{1}{2}\left((1 + \varepsilon z_b)\beta_1 + (1 - (1 - \varepsilon)z_b)\beta_0\right) =: X_2. \end{array} \tag{9}$$

This SDP is strictly feasible as can be seen by taking, say, $X_b = 2I$, $z_b = 1$. This implies that its optimal value is also b and in particular, that there exist feasible solutions (X_b, z_b) whose value $\mathrm{Tr}[X_b]$ is as close to b as we wish.

A similar calculation allows us to derive the SDP formulation of $p = D_{\varepsilon/2}(\alpha_0\otimes\beta_0, \alpha_0\otimes\beta_1, \alpha_1\otimes\beta_0, \alpha_1\otimes\beta_1)$. The primal SDP is the following.

$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}[(E_{00}+E_{01}+E_{10}+E_{11})\,\alpha\otimes\beta] \\ \text{subject to} & E_{00},E_{01},E_{10},E_{11},E_? \succeq 0, \\ & E_{00}+E_{01}+E_{10}+E_{11}+E_? = I, \\ & \frac{1}{4}\mathrm{Tr}\left[(E_{01}+E_{10}+E_{11})\,\alpha_0\otimes\beta_0 + (E_{00}+E_{10}+E_{11})\,\alpha_0\otimes\beta_1 + \right. \\ & \qquad \left.(E_{00}+E_{01}+E_{11})\,\alpha_1\otimes\beta_0 + (E_{00}+E_{01}+E_{10})\,\alpha_1\otimes\beta_1\right] \\ & \qquad \le \frac{\varepsilon}{2}\mathrm{Tr}[(E_{00}+E_{01}+E_{10}+E_{11})\,\alpha\otimes\beta]. \end{array}$$

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. The dual SDP is easily shown to be the following.

$$\begin{array}{ll} \text{minimize} & \mathrm{Tr}[X] \\ \text{subject to} & X\succeq 0, z\ge 0, \\ \multicolumn{2}{l}{X \succeq \frac{1}{4}\left(\left\{(1+\frac{\varepsilon}{2}z)\alpha_0 + (1-(1-\frac{\varepsilon}{2})z)\alpha_1\right\}\otimes\beta_0 + (1-(1-\frac{\varepsilon}{2})z)(\alpha_0+\alpha_1)\otimes\beta_1\right) =: X_1',} \\ \multicolumn{2}{l}{X \succeq \frac{1}{4}\left(\left\{(1+\frac{\varepsilon}{2}z)\alpha_0 + (1-(1-\frac{\varepsilon}{2})z)\alpha_1\right\}\otimes\beta_1 + (1-(1-\frac{\varepsilon}{2})z)(\alpha_0+\alpha_1)\otimes\beta_0\right) =: X_2',} \\ \multicolumn{2}{l}{X \succeq \frac{1}{4}\left(\left\{(1+\frac{\varepsilon}{2}z)\alpha_1 + (1-(1-\frac{\varepsilon}{2})z)\alpha_0\right\}\otimes\beta_0 + (1-(1-\frac{\varepsilon}{2})z)(\alpha_0+\alpha_1)\otimes\beta_1\right) =: X_3',} \\ \multicolumn{2}{l}{X \succeq \frac{1}{4}\left(\left\{(1+\frac{\varepsilon}{2}z)\alpha_1 + (1-(1-\frac{\varepsilon}{2})z)\alpha_0\right\}\otimes\beta_1 + (1-(1-\frac{\varepsilon}{2})z)(\alpha_0+\alpha_1)\otimes\beta_0\right) =: X_4'.} \end{array} \tag{10}$$

Let $\delta := \sqrt{1-|\langle\alpha_0|\alpha_1\rangle|^2}$. Below we will show that any feasible solution (X_b, z_b) with value $\mathrm{Tr}[X_b]\ge b$ to (9) can be used to construct a feasible solution (X,z) to (10) whose value $\mathrm{Tr}[X]$ satisfies

$$\mathrm{Tr}[X] \le 16\delta^2\mathrm{Tr}[X_b]. \tag{11}$$

By weak duality, this implies that $p \le 16\delta^2\mathrm{Tr}[X_b]$. Moreover, as observed before, the SDP (9) is strictly feasible. Therefore, by strong duality, we can make $\mathrm{Tr}[X_b]$ as close to b as we wish and the theorem follows from (11).

It remains to show how to obtain (X,z) from (X_b,z_b) as above. So fix some feasible solution (X_b,z_b) to (9). We need the following technical claim, which we prove afterwards.

Claim 3.3 *Let $0\le\varepsilon<1/2$ and $z_b\ge 0$. There exists $z = z(\varepsilon, z_b)$ with the following property: for all density matrices σ_0, σ_1, ρ_0, and ρ_1 where ρ_0 and ρ_1 are 2-dimensional with rank 1 we have*

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

where $\rho_1^\perp = I-\rho_1$ and $\delta = \sqrt{1-\mathrm{Tr}[\rho_0\rho_1]}$.

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 3.3 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 3.3 gives a $z=z(\varepsilon,z_b)$ such that $Y_1\succeq X_1'$ (see (10) for the definition of X_1'). Since $\alpha_1^\perp\succeq 0$, Item 2 in Claim 2.3 implies that

$$\mathrm{Pos}(Y_1) = 4\delta^2\alpha_1^\perp\otimes\mathrm{Pos}(X_1).$$

Because $\mathrm{Tr}[\alpha_1^\perp] = 1$, we have $\mathrm{Tr}[\mathrm{Pos}(Y_1)] = 4\delta^2 \mathrm{Tr}[\mathrm{Pos}(X_1)]$. Moreover, $X_1 \preceq X_b$ by definition (see (9)) and $X_b = \mathrm{Pos}(X_b)$, hence $\mathrm{Tr}[\mathrm{Pos}(Y_1)] \le 4\delta^2 \mathrm{Tr}[\mathrm{Pos}(X_b)] = 4\delta^2 \mathrm{Tr}[X_b]$ (using Item 3 in Claim 2.3).

However, $\mathrm{Pos}(Y_1)$ is not a feasible solution to (10) because it need not satisfy the last three inequalities. We therefore 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 3.3 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 (10) since $X \succeq 0$ by definition and $X \succeq \mathrm{Pos}(Y_i) \succeq X'_i$ for $i = 1, \ldots, 4$ (using Item 1 of Claim 2.3). Since $\mathrm{Tr}[X] = \sum_{i=1}^4 \mathrm{Tr}[\mathrm{Pos}(Y_i)] \le 16\delta^2 \mathrm{Tr}[X_b]$, Eq. (11) follows and we are done. ■

It remains to prove the claim.

Proofof Claim 3.3: Because σ_0 and σ_1 are positive semidefinite, it suffices to find a $z \ge 0$ for which both

$$4\delta^2 \rho_1^\perp \frac{1}{2}(1 + \varepsilon z_b) \succeq \frac{1}{4}\left\{\left(1 + \frac{\varepsilon}{2} z\right)\rho_0 + \left(1 - \left(1 - \frac{\varepsilon}{2}\right) z\right)\rho_1\right\} \tag{12}$$

and

$$4\delta^2 \rho_1^\perp \frac{1}{2}(1 - (1-\varepsilon) z_b) \succeq \frac{1}{4}\left(1 - \left(1 - \frac{\varepsilon}{2}\right) z\right)(\rho_0 + \rho_1) \tag{13}$$

are true.

Let $|\rho_0\rangle, |\rho_1\rangle$ and $|\rho_1^\perp\rangle$ be pure states corresponding to the density matrices ρ_0, ρ_1 and $\rho_1^\perp$. We can assume without loss of generality that $|\rho_0\rangle = \sqrt{1-\delta^2}|\rho_1\rangle + \delta|\rho_1^\perp\rangle$ (by choosing an appropriate global phase). Then, in the basis given by $|\rho_1\rangle$, $|\rho_1^\perp\rangle$, Eqs. (12) and (13) become

$$\begin{pmatrix} -(1+\frac{\varepsilon}{2}z)(1-\delta^2) - (1-(1-\frac{\varepsilon}{2})z) & -\delta\sqrt{1-\delta^2}(1+\frac{\varepsilon}{2}z) \\ -\delta\sqrt{1-\delta^2}(1+\frac{\varepsilon}{2}z) & 8\delta^2(1+\varepsilon z_b) - \delta^2(1+\frac{\varepsilon}{2}z) \end{pmatrix} = \begin{pmatrix} z(1-\varepsilon+\delta^2\frac{\varepsilon}{2}) + \delta^2 - 2 & -\delta\sqrt{1-\delta^2}(1+\frac{\varepsilon}{2}z) \\ -\delta\sqrt{1-\delta^2}(1+\frac{\varepsilon}{2}z) & \delta^2(7+8\varepsilon z_b - \frac{\varepsilon}{2}z) \end{pmatrix} \succeq 0 \tag{14}$$

and

$$\begin{pmatrix} -(1-(1-\frac{\varepsilon}{2})z)(2-\delta^2) & -\delta\sqrt{1-\delta^2}(1-(1-\frac{\varepsilon}{2})z) \\ -\delta\sqrt{1-\delta^2}(1-(1-\frac{\varepsilon}{2})z) & 8\delta^2(1-(1-\varepsilon)z_b) - \delta^2(1-(1-\frac{\varepsilon}{2})z) \end{pmatrix} = \begin{pmatrix} ((1-\frac{\varepsilon}{2})z-1)(2-\delta^2) & \delta\sqrt{1-\delta^2}((1-\frac{\varepsilon}{2})z-1) \\ \delta\sqrt{1-\delta^2}((1-\frac{\varepsilon}{2})z-1) & \delta^2(7-8(1-\varepsilon)z_b + (1-\frac{\varepsilon}{2})z) \end{pmatrix} \succeq 0. \tag{15}$$

We choose

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

To show that a 2×2 Hermitian matrix is positive semidefinite it suffices to show that its determinant is nonnegative and at least one of its diagonal entries is positive. Since $z \ge 4$, the top-left entries of the matrices in Eqs. (14) and (15) are positive. When $\delta = 0$, the two matrices are clearly positive semidefinite, so we assume that $0 < \delta \le 1$. By canceling $\delta^2 > 0$ from both terms that appear in the determinants and some algebraic manipulations, we find that for Eqs. (14) and (15) to be true, it suffices to show

$$\Big(z(1-\varepsilon)-2\Big)\left(7+8\varepsilon z_b-\frac{\varepsilon}{2}z\right)-\left(1+\frac{\varepsilon}{2}z\right)^2\geq 0 \tag{16}$$

and

$$(2-\delta^2)\left(\left(1-\frac{\varepsilon}{2}\right)z-1\right)\left(7-8(1-\varepsilon)z_b+\left(1-\frac{\varepsilon}{2}\right)z\right)-(1-\delta^2)\left(\left(1-\frac{\varepsilon}{2}\right)z-1\right)^2\geq 0. \tag{17}$$

To see how (16) implies (14), notice that it implies $7+8\varepsilon z_b-\frac{\varepsilon}{2}z\geq 0$ and hence we can replace the term $z(1-\varepsilon)-2$ by the larger term $z(1-\varepsilon+\delta^2\frac{\varepsilon}{2})+\delta^2-2$.

Using $(2-\delta^2)\geq 2(1-\delta^2)$ and $(1-\frac{\varepsilon}{2})z-1>0$, Eq. (17) is implied by

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

which is equivalent to

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

This inequality is true for our choice of z.

It remains to show that our z satisfies Eq. (16). Substituting for z we see that the quadratic term in z_b cancels and we obtain

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

This inequality holds for any $z_b\geq 0$ because both its constant coefficient and the coefficient of z_b are positive for $0\leq\varepsilon<\frac{1}{2}$. ■

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 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 some 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. More precisely, we will prove:

Theorem 4.1 *For the relational problem P_1 defined above we have*

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

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 does not see an i for which $s_i = 1$, in which case his output is arbitrary. Hence $R_\varepsilon^{\|,pub}(P) = O(\log n \log(1/\varepsilon))$.

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 $q = \Omega(n^{1/3})$.

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 2.2) 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

$$b_i \leq \frac{2q}{n(1-H(4\varepsilon))} = O(q/n)$$

for all $i \in S$. We now fix Alice's input s to be the n-bit string with support corresponding to S.

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

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

Now we look at the protocol's behavior. 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 $i \in S$ *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 must together 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

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

Notice that for each good i we can use the protocol to get a $(p_i, 2\varepsilon)$-predictor for X_iY_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 3.2 implies $p_i = O(a_ib_i)$. Also, $b_i = O(q/n)$ for all good i so we can bound

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

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

Remark The best private-randomness protocol for P_1 that we could come up with communicates $O(\sqrt{n})$ bits. The idea, inspired by Ambainis [1], 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 $p = O(ab)$, for the case of the 2-register identification problem where both sides are allowed to be mixed.

5 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 edge $e \in M$
> **Bob's input:** a string $y \in \{0,1\}^n$
> **Goal:** referee should output $(i, j, x_{(i,j)}, y_i \oplus y_j)$ for some $(i, j) \in M$

Below 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 5.1 *For the relational problem P_2 defined above we have*

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

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 [9], 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 of Alice and Bob 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 nonzero 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 $(i, j, x_{(i,j)}, y_i \oplus y_j)$ as required.

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. 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 $q = \Omega((n/\log n)^{1/3})$.

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 (Lemma 2.2) gives

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

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

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 $w \in \{0,1\}^{|F|}$ (with positions indexed by the $e \in F$) the set of inputs y whose F-parities coincide with the string w:

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

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 $2^{n-|F|}$.

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 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 5.3 $\displaystyle\sum_{k=0}^{n/2-1} \sum_{e \in M_k} b_e^2 = O(q^2 \log n)$.

Proof: By construction all our M_k's are edge-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 can use Claim 5.2:

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

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

Summing over all ℓ gives $\sum_{e \in M} b_e^2 = \sum_{\ell=1}^{|M|} b_\ell^2 \leq \sum_{\ell=1}^{n^2/4} O(q^2/\ell) = O(q^2 \log n)$. ■

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 each of those M_k and uniformly random X and Y. Since $\mathcal{M}$ contains at least $n/4$ matchings as elements, Claim 5.3 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 $(i,j) \in M_k$ *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 3.2, $p_e = O(a_{ke} b_e)$. Using Cauchy-Schwarz:

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

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

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

$$\frac{1}{\sqrt{|S|}} \sum_{i \in S} (-1)^{y_i} |i\rangle \tag{18}$$

and sends them to the referee. This gives 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. (18), 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 problems left open by our work. First, for the bounded-error quantum state identification problem, prove the direct product theorem $p = O(ab)$ 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. Third, it is quite possible that the situation with entanglement is analogous to the situation with shared randomness: shared randomness sometimes gives savings in the SMP model (compared to private randomness) in the SMP model, but it gives at most logarithmic additive savings in the more general models of one-way or two-communication [22]. We showed here that entanglement can give exponential savings (compared to shared randomness) in the SMP model. To complete the analogy, we would need to show that entanglement doesn't help much in the models of one-way and two-way communication. 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 eventually modified to the protocol of Section 5.2. DG is grateful to Richard Cleve for helpful discussions. Thanks to the two anonymous SICOMP referees for suggestions to improve the presentation of the paper.

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 the 12th IEEE Conference on Computational Complexity*, pages 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*, pages 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*, pages 93–102, 2002.

[5] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[6] R. Bhatia. *Matrix Analysis*. Number 169 in Graduate Texts in Mathematics. Springer-Verlag, New York, 1997.

[7] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

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

[9] H. Buhrman, 2003. Personal communication.

[10] H. Buhrman, R. Cleve, J. Watrous, and R. d. Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16), 2001. quant-ph/0102001.

[11] R. Cleve and H. Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201–1204, 1997. quant-ph/9704026.

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

[13] Y. C. Eldar. A semidefinite programming approach to optimal unambiguous discrimination of quantum states. *IEEE Transactions on Information Theory*, 49:446–456, 2003. quant-ph/0206093.

[14] D. Gavinsky, J. Kempe, and R. d. Wolf. Strengths and weaknesses of quantum fingerprinting. In *Proceedings of 21st IEEE Conference on Computational Complexity*, pages 288–295, 2006. quant-ph/0603173.

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

[16] 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. quant-ph/0401026.

[17] H. Klauck. Quantum communication complexity. In *Proceedings of Workshop on Boolean Functions and Applications at 27th ICALP*, pages 241–252, 2000. quant-ph/0005032.

[18] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999. Earlier version in STOC'95. Correction at `http://www.eng.tau.ac.il/~danar/Public/KNR-fix.ps`.

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

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

[21] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of 40th IEEE FOCS*, pages 369–376, 1999. quant-ph/9904093.

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

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

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

[25] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38(1):49–95, 1996.

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

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

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