Separating quantum communication and approximate rank

Anurag Anshu[*]   Shalev Ben-David[†]   Ankit Garg[‡]

Rahul Jain[§]   Robin Kothari[¶]   Troy Lee[‖]

Abstract

One of the best lower bound methods for the quantum communication complexity of a function H (with or without shared entanglement) is the logarithm of the approximate rank of the communication matrix of H. This measure is essentially equivalent to the approximate γ_2 norm and generalized discrepancy, and subsumes several other lower bounds. All known lower bounds on quantum communication complexity in the general unbounded-round model can be shown via the logarithm of approximate rank, and it was an open problem to give any separation at all between quantum communication complexity and the logarithm of the approximate rank.

In this work we provide the first such separation: We exhibit a total function H with quantum communication complexity almost quadratically larger than the logarithm of its approximate rank. We construct H using the communication lookup function framework of Anshu et al. (FOCS 2016) based on the cheat sheet framework of Aaronson et al. (STOC 2016). From a starting function F, this framework defines a new function $H = F_{\mathcal{G}}$. Our main technical result is a lower bound on the quantum communication complexity of $F_{\mathcal{G}}$ in terms of the discrepancy of F, which we do via quantum information theoretic arguments. We show the upper bound on the approximate rank of $F_{\mathcal{G}}$ by relating it to the Boolean circuit size of the starting function F.

---

[*]Centre for Quantum Technologies, National University of Singapore, Singapore. `a0109169@u.nus.edu`

[†]Massachusetts Institute of Technology. `shalev@mit.edu`

[‡]Microsoft Research New England. `garga@microsoft.com`

[§]Centre for Quantum Technologies, National University of Singapore and MajuLab, UMI 3654, Singapore. `rahul@comp.nus.edu.sg`

[¶]Center for Theoretical Physics, Massachusetts Institute of Technology. `rkothari@mit.edu`

[‖]SPMS, Nanyang Technological University and Centre for Quantum Technologies and MajuLab, UMI 3654, Singapore. `troyjlee@gmail.com`

1 Introduction

Communication complexity studies how much two parties Alice and Bob need to communicate in order to compute a function when each party only has partial knowledge of the input. The model of quantum communication complexity allows the players to send quantum messages back and forth, and measures the total number of qubits that need to be exchanged in order to compute the function. Communication complexity has become a fundamental area in theoretical computer science with applications to circuit complexity, data structures, streaming algorithms, property testing, and linear and semi-definite programs. Many of these applications require showing communication complexity *lower bounds*, which raises the importance of studying lower bound techniques in communication complexity.

In this paper we study lower bounds on quantum communication complexity. For a two-party function $F : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, we denote by $\mathrm{Q}(F)$ the minimum number of qubits needed by a quantum protocol to compute F with error probability at most $1/3$.

One of the strongest lower bounds on $\mathrm{Q}(F)$ comes by viewing F as a Boolean $|\mathcal{X}| \times |\mathcal{Y}|$ matrix, known as the communication matrix, which we will also denote by F. The approximate rank of F, denoted $\mathrm{rk}_{1/3}(F)$, is the minimum rank of a matrix $\tilde{F}$ that is entrywise close to F, that is, satisfying $\ell_\infty(\tilde{F} - F) \leq 1/3$. Building on the work of Kremer [Kre95] and Yao [Yao93], Buhrman and de Wolf [BdW01] showed that $\mathrm{Q}(F) = \Omega(\log \mathrm{rk}_{1/3}(F))$. Later, it was shown that approximate rank can also be used to lower bound quantum communication complexity with shared entanglement, denoted $\mathrm{Q}^*(F)$. More precisely, $\mathrm{Q}^*(F) = \Omega(\log \mathrm{rk}_{1/3}(F)) - O(\log \log(|\mathcal{X}| \cdot |\mathcal{Y}|))$ [LS08]. As this paper studies quantum communication complexity lower bounds, we will focus on the measure $\mathrm{Q}^*(F)$, which makes our results stronger.

The logarithm of the approximate rank dominates nearly all other lower bounds on quantum communication complexity, including the discrepancy method [Kre95], the approximate trace norm [Raz03, LS09], the generalized discrepancy method [Kla07, Raz03, She11], and the approximate γ_2 norm bound [LS09].[1] In fact, to the best of our knowledge, all known lower bounds for general two-way quantum communication can be obtained using approximate rank. Besides being a powerful lower bound method, approximate rank is a robust measure posessing several desirable properties such as error reduction, direct sum and strong direct product theorems [She12], and an optimal lifting theorem [She11, SZ09].

Given our current state of knowledge, it is consistent that $\mathrm{Q}^*(F) = O(\log \mathrm{rk}_{1/3}(F))$ for every function F, that is, the logarithm of the approximate rank *characterizes* quantum communication complexity. As it is widely believed that this is not the case, this state of affairs points to the limitations of our current lower bound techniques for quantum communication complexity.

In this paper, we show the first superlinear separation between quantum communication complexity and the logarithm of the approximate rank.

Theorem 1. *There is a family of total functions* $F : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ *with* $\mathrm{Q}^*(F) = \widetilde{\Omega}\Big(\log^2 \mathrm{rk}_{1/3}(F)\Big)$.

As far as we are aware, Theorem 1 is the first superlinear separation between quantum communication complexity and the logarithm of the approximate rank even for *partial* functions, which are functions defined only on a subset of the domain $\mathcal{X} \times \mathcal{Y}$.[2]

---

[1]In fact, the generalized discrepancy method, logarithm of approximate γ_2 norm, and logarithm of approximate rank are all equivalent, up to constant mutliplicative factors and an additive logarithmic term.

[2]For partial functions, we require the approximate low-rank decomposition of the communication matrix to take values between 0 and 1 even on inputs on which the function is undefined. Without this constraint it is easy to construct large partial function separations.

One alternative to approximate rank for showing lower bounds on quantum communication complexity is the recently introduced quantum information complexity [Tou15]. This bound has been shown to dominate the logarithm of the approximate rank [BGK+15], and has nice properties like characterizing amortized quantum communication complexity. The quantum information complexity, however, is difficult to bound for an explicit function and has not yet been used to show a new lower bound in the general unbounded-round model of quantum communication complexity.

By analogy with the log rank conjecture, which postulates that $D(F) = O(\text{polylog}(\text{rk}(F)))$, where $D(F)$ is the deterministic communication complexity of F, it is natural to state an approximate log rank conjecture. The quantum version of the approximate log rank conjecture states $Q^*(F) = O(\text{polylog}(\text{rk}_{1/3}(F)))$. Our results show that the exponent of the logarithm in such a statement must be at least 2. The largest gap we currently know between $D(F)$ and $\log \text{rk}(F)$ is also quadratic [GPW15]. One could also consider a randomized version of the log rank conjecture, stating $R(F) = O(\text{polylog}(\text{rk}_{1/3}(F)))$, where $R(F)$ is the 1/3-bounded-error randomized communication complexity. This conjecture is actually known to imply the usual deterministic log rank conjecture [KMSY14]. The largest known gap between $R(F)$ and $\log \text{rk}_{1/3}(F)$ is 4th power [GJPW15].

Our separation is established using quantum information theoretic arguments to lower bound quantum communication complexity of a particular family of functions known as lookup functions, introduced in [ABB+16b]. We use Boolean circuit size to upper bound the logarithm of approximate rank of lookup functions. We now provide an overview of lookup functions and our proof techniques.

1.1 Techniques

Many questions in communication complexity have analogs in the (usually simpler) model of query complexity. The query complexity quantity that is analogous to approximate rank is the approximate polynomial degree. Using the quantum adversary lower bound, Ambainis [Amb03] gave a function f with an n versus $n^{1.32}$ separation between its approximate polynomial degree and quantum query complexity. This result is the main reason for the belief that there should also be a separation between the logarithm of approximate rank and quantum communication complexity. One way to do this would be to "lift" the quantum query lower bound for f into a quantum communication lower bound for a related communication problem by composing f with an appropriate communication gadget. While such a lifting theorem is known for the approximate polynomial degree [She11, SZ09], it remains an open question to show a lifting theorem for quantum query complexity or the quantum adversary method. The lack of an analog of the adversary lower bound in the setting of quantum communication complexity is part of the difficulty of separating the logarithm of approximate rank and quantum communication complexity.

There has recently been a great deal of progress in showing new separations between complexity measures in query complexity [GPW15, ABB+16a, ABK16]. The work in query complexity most closely related to ours is the *cheat sheet* method of Aaronson et al. [ABK16]. The cheat sheet method is a way to transform a function f into its "cheat sheet" version f_{CS} so that, for some complexity measures, f_{CS} retains the hardness of f, while other complexity measures are drastically reduced by this transformation. Among other things, Aaronson et al. [ABK16] use this method to improve Ambainis' separation and give a 4th power separation between quantum query complexity and approximate polynomial degree.

[ABB+16b] generalize the cheat sheet method to communication complexity. They are able to lift several query results of [ABK16] to communication complexity, such as an example of a total function with a super-quadratic separation between its randomized and quantum communication complexities. They do this by introducing the idea of a *lookup* function. To motivate a lookup

function, consider first a communication version of the familiar address function. Alice receives inputs $x \in \{0,1\}^c$ and $u_0, \ldots, u_{2^c-1} \in \{0,1\}$ and Bob receives $y \in \{0,1\}^c$ and $v_0, \ldots, v_{2^c-1} \in \{0,1\}$. The desired output is found by interpreting $x \oplus y$ as the binary representation of a number $\ell \in \{0, \ldots, 2^c - 1\}$ and outputting $u_\ell \oplus v_\ell$.

The $(F, \mathcal{G})$ lookup function $F_{\mathcal{G}}$ is defined by a function $F : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and a function family $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$, with $G_i : (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$. Alice receives input $\mathbf{x} = (x_1, \ldots, x_c) \in \mathcal{X}^c$ and $u_0, \ldots, u_{2^c-1} \in \{0,1\}^m$ and Bob receives inputs $\mathbf{y} = (y_1, \ldots, y_c) \in \mathcal{Y}^c$ and $v_0, \ldots, v_{2^c-1} \in \{0,1\}^m$. Now the address is determined by interpreting $(F(x_1, y_1), \ldots, F(x_c, y_c)) \in \{0,1\}^c$ as an integer $\ell \in \{0, \ldots, 2^c - 1\}$ and the goal of the players is to output $G_\ell((\mathbf{x}, u_\ell), (\mathbf{y}, v_\ell))$. Note that, in contrast to the case with the address function, in a lookup function, G_ℓ can depend on $\mathbf{x}$ and $\mathbf{y}$. This is the source of difficulty in showing lower bounds for lookup functions, and also key to their interesting properties.

Lower bound. The main result of [ABB+16b] showed that, given some mild restrictions on the family of functions $\mathcal{G}$, the randomized communication complexity of $F_{\mathcal{G}}$ is at least that of F. Our main result shows that, given mild restrictions on the function family $\mathcal{G}$, if there is a quantum protocol with q qubits of communication for $F_{\mathcal{G}}$, then there is a q qubit protocol for F with *non-negligible bias*. Because of the round-by-round nature of our quantum information theoretic argument, the success probability of the quantum protocol for F decays with the number of rounds of the quantum protocol for $F_{\mathcal{G}}$. Thus to apply this theorem, we need to start with a function F that has high quantum communication complexity even for protocols with small bias. As the discrepancy method lower bounds quantum communication complexity even with small bias, we can informally state our main theorem as follows.

Theorem 2 (Informal restatement of Corollary 7). *For any (F, G) lookup function $F_{\mathcal{G}}$, provided $\mathcal{G}$ satisfies certain mild technical conditions, $\mathrm{Q}^*(F_{\mathcal{G}}) = \Omega(\log(1/\operatorname{disc}(F)))$.*

Let us call such theorems, where we lower bound the complexity of a lookup function $F_{\mathcal{G}}$ (or a cheat sheet function f_{CS}) in terms of a measure of the original function F (or f), "cheat sheet theorems." Essentially optimal cheat sheet theorems have been shown in a number of computational models such as deterministic, randomized, and quantum query complexity [ABK16] and randomized communication complexity [ABB+16b]. Cheat sheet theorems are in spirit similar to joint computation results such as direct sum and direct product theorems [BBCR13, BRWY13, BW15, Dru12, LR13, She12, Tou15].[3] Direct sum and direct product theorems are widely applicable tools and are often an important goal by themselves. Cheat sheet theorems have become useful tools recently and for example, the cheat sheet theorems proven in [ABK16] were later used in [AKK16]. We hope that our quantum cheat sheet theorem will find further applications.

We now provide a high-level overview of the proof of our quantum cheat sheet theorem. We would like to rule out the existence of a quantum protocol Π that solves the lookup function $F_{\mathcal{G}}$ and whose communication cost is much smaller than the quantum communication complexity of F (with inverse polynomial bias, for technical reasons explained below). Since Π has small communication cost, during the course of the protocol Alice and Bob do not know the value of the index $\ell = (F(x_1, y_1), \ldots, F(x_c, y_c))$. Also since there are too many cells in the array, which has length $2^c \gg \mathrm{Q}^*(F)$, and Π has small communication cost, Alice and Bob cannot talk about too many cells of the array. We first show that these two conditions imply that Alice and Bob have

---

[3]One point of difference is that in direct sum and direct product theorems, the lower bounds on the amount of resources (query, communication, etc.) usually scale with c, the number of copies of the function F. In the cheat sheet theorem we prove (and also in prior works), the lower bounds do not scale with c. This is due to the fact that the value of c is usually small in our applications.

little information about the contents of the correct cell of the other player's array, i.e., Alice has little information about v_ℓ and Bob has little information about u_ℓ.

In the hypothesis of the theorem, we assume that G_ℓ satisfies a *nontriviality condition*: this states that $G_\ell(\mathbf{x}, \mathbf{y}, u_\ell, v_\ell)$ takes both values 0 and 1 as (u_ℓ, v_ℓ) range over all possible values. Thus the fact that Alice has little information about v_ℓ and Bob has little information about u_ℓ sounds like we have reached a contradiction already. The issue is that we do not have any control over the *bias* of G_ℓ. This situation is reminiscent of the quantum information theoretic arguments in the proof of quantum communication complexity lower bounds for the disjointness function [JRS03]. In that case, one has to argue that a quantum protocol that solves the AND function on 2 bits exchanges non-trivial amount of information even on distributions which are extremely biased towards the AND being 0. We use similar arguments (namely the *quantum cut-and-paste* argument) to obtain a contradiction for our lookup function. Quantum cut-and-paste arguments usually have a round dependence (which is provably needed for the disjointness lower bound) but which may not be needed for our lookup function. Improving our quantum cheat sheet theorem or proving that it is tight remains an excellent open question.

At a high level our proof follows the same strategy as the proof for randomized communication complexity in [ABB+16b], but the implementation of the steps of the argument is different due to the quantum nature of the protocol. A quantum communication protocol presents several challenges, such as the fact that there is no notion of a communication transcript, since it is not possible to store all the quantum messages exchanged during the protocol. Hence arguments that applied to the overall communication transcript do not work in the quantum setting. Several technical lemmas, such as the Markov chain property of classical communication protocols used in [ABB+16b], fail to hold in the quantum setting.

Upper bound. We devise a general technique for proving upper bounds on the logarithm of approximate rank of lookup functions for carefully constructed function families $\mathcal{G}$. Given a circuit $\mathcal{C}$ for F, a cell in the array tries to certify the computation of F by the circuit $\mathcal{C}$. More formally, $G_\ell(\mathbf{x}, \mathbf{y}, u_\ell, v_\ell) = 1$ iff $(F(x_1, y_1), \dots, F(x_c, y_c)) = \ell$ and $u_\ell \oplus v_\ell$ provides the values of the inputs and outputs to all the gates in $\mathcal{C}$ for each of the c different evaluations of $\mathcal{C}$ on inputs $(x_1, y_1), \dots, (x_c, y_c)$. We show that a small circuit for F implies a good upper bound on the approximate rank of the lookup function $F_\mathcal{G}$.

Theorem 3 (Informal restatement of Theorem 6)**.** *For any Boolean function F, there exists a family of functions $\mathcal{G}$ satisfying certain nontrivality conditions such that the lookup function $F_\mathcal{G}$ satisfies $\log \mathrm{rk}_{1/3}(F_\mathcal{G}) = \widetilde{O}(\sqrt{\mathrm{size}(F)})$.*

Here $\mathrm{size}(F)$ denotes the size of the smallest circuit (i.e., the one with the least number of gates) for F over some constant-sized gate set, such as the set of all 2-bit gates. The high level idea for the upper bound is the following. Suppose an all-knowing prover Merlin provided Alice and Bob the value $\ell = (F(x_1, y_1), \dots, F(x_c, y_c))$. Then they can "unambiguously" verify Merlin's answer with a small amount of quantum communication. Essentially they look at the ℓth cell of the array and try to find an inconsistency in the circuit values. This can then be done with quadratically less communication by a quantum protocol by using a distributed version of Grover's algorithm [Gro96, BCW98]. We then show that this sort of upper bound on "unambiguously certifiable quantum communication" provides an upper bound on the log of approximate rank of the lookup function $F_\mathcal{G}$. A similar upper bound was also used in the query complexity separations of [ABK16].

Putting these upper and lower bounds together, if we choose F to be the inner product function, which has exponentially small discrepancy and linear circuit size, Theorem 2 and Theorem 3 give

4

us the desired quadratic separation between quantum communication complexty and the log of approximate rank for a lookup function $F_{\mathcal{G}}$.

One intriguing aspect of Theorem 3 is that if one can prove lower bounds on $\log \mathrm{rk}_{1/3}(F_{\mathcal{G}}) \gg \sqrt{n}$ for *every* nontrivial function family $\mathcal{G}$, then one proves nontrivial circuit lower bounds for F! This theorem is similar in flavor to the theorem [LLS06, Rei11] that the square of the quantum query complexity of a function f is a lower bound on the formula size of f. It might seem hopeless to prove a lower bound on $\log \mathrm{rk}_{1/3}(F_{\mathcal{G}})$ for every nontrivial function family $\mathcal{G}$, but this is exactly what our quantum cheat sheet theorem achieves for quantum communication complexity, and what the results of [ABB$^+$16b] achieve for randomized communication complexity.

2   Preliminaries and notation

We will use X, Y, Z to denote random variables as well as their distributions. $x \leftarrow X$ will stand for x being sampled from the distribution of X. For joint random variables XY, Y^x will denote the distribution of $Y|X = x$.

We now state some classical complexity measures that will be used in this paper. We define quantum measures in more detail in Section 2.1 and Section 2.2. We first formally define approximate rank.

Definition 4 (Approximate rank). Let $\varepsilon \in [0, 1/2)$ and F be an $|\mathcal{X}| \times |\mathcal{Y}|$ matrix. The ε-approximate rank of F is defined as

$$\mathrm{rk}_\varepsilon(F) = \min_{\tilde{F}} \left\{ \mathrm{rk}(\tilde{F}) : \forall x \in \mathcal{X}, y \in \mathcal{Y}, |\tilde{F}(x,y) - F(x,y)| \leq \varepsilon \right\}$$

As discussed in the introduction, approximate rank lower bounds bounded-error quantum communication complexity with shared entanglement. It also lower bounds ε-error quantum communication [LS08]:

Fact 5. *For any two-party function* $F : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$ *and* $\varepsilon \in [0, 1/3]$, *we have* $\mathrm{Q}_\varepsilon^*(F) = \Omega(\log \mathrm{rk}_\varepsilon(F)) - O(\log \log(|\mathcal{X}| \cdot |\mathcal{Y}|))$.

Another classical lower bound measure that we use is the discrepancy of a function [KN06].

Definition 6 (Discrepancy). Let F be an $|\mathcal{X}| \times |\mathcal{Y}|$ Boolean-valued matrix and P a probability distribution over $\mathcal{X} \times Y$. The discrepancy of F with respect to P is

$$\mathrm{disc}_P(F) = \max_R \left| \sum_{(x,y) \in R} P(x,y)(-1)^{F(x,y)} \right| ,$$

where the maximum is taken with respect to all combinatorial rectangles R. The discrepancy of F, denoted $\mathrm{disc}(F)$, is defined as $\mathrm{disc}(F) = \min_P \mathrm{disc}_P(F)$, where the minimum is taken over all probability distributions P.

The discrepancy bound lower bounds not only bounded-error quantum communication complexity, but also quantum communication complexity with error exponentially close (in the discrepancy) to $1/2$. More precisely, we have the following [Kre95, LS09].

Theorem 7. *Let* $F : \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$ *be a two-party function and* $\varepsilon \in [0, 1/2)$. *Then*

$$\mathrm{Q}_\varepsilon^*(F) = \Omega \left(\log \frac{1 - 2\varepsilon}{\mathrm{disc}(F)} \right).$$

Finally we define the Boolean circuit size of a function. To do this, we first fix a gate set, say the set of all gates with 2 input bits (although we could have chosen any constant instead of 2).

Definition 8 (Circuit size). For a function $F : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$, we define size$(F)$ to be the size (i.e., number of gates) of the smallest circuit over the gates set of all 2-input Boolean gates that computes F.

Note that here the encoding of Alice's and Bob's input is important, since different input representations may yield different sized circuits, unlike in communication complexity. When we use this size measure, we only deal with functions defined on bits where the input encoding is clearly specified.

2.1 Quantum Information

We now introduce some quantum information theoretic notation. We assume the reader is familiar with standard notation in quantum computing [NC00, Wat16].

Let $\mathcal{H}$ be a finite-dimensional complex Euclidean space, i.e., $\mathbb{C}^n$ for some positive integer n with the usual complex inner product $\langle \cdot, \cdot \rangle$, which is defined as $\langle u, v \rangle = \sum_{i=1}^{n} u_i^* v_i$. We will also refer to $\mathcal{H}$ as a Hilbert space. We will usually denote vectors in $\mathcal{H}$ using braket notation, e.g., $|\psi\rangle \in \mathcal{H}$.

The ℓ_1 norm (also called the trace norm) of an operator X on $\mathcal{H}$ is $\|X\|_1 \coloneqq \mathrm{Tr}(\sqrt{X^\dagger X})$, which is also equal to (vector) ℓ_1 norm of the vector of singular values of X.

A *quantum state* (or a *density matrix* or simply a *state*) ρ is a positive semidefinite matrix on $\mathcal{H}$ with $\mathrm{Tr}(\rho) = 1$. The state ρ is said to be a *pure state* if its rank is 1, or equivalently if $\mathrm{Tr}(\rho^2) = 1$, and otherwise it is called a *mixed state*. Let $|\psi\rangle$ be a unit vector on $\mathcal{H}$, that is $\langle \psi | \psi \rangle = 1$. With some abuse of notation, we use ψ to represent the vector $|\psi\rangle$ and also the density matrix $|\psi\rangle\langle\psi|$, associated with $|\psi\rangle$. Given a quantum state ρ on $\mathcal{H}$, the *support of* ρ, denoted supp(ρ) is the subspace of $\mathcal{H}$ spanned by all eigenvectors of ρ with nonzero eigenvalues.

A *quantum register* A is associated with some Hilbert space $\mathcal{H}_A$. Define $|A| \coloneqq \log \dim(\mathcal{H}_A)$. Let $\mathcal{L}(A)$ represent the set of all linear operators on $\mathcal{H}_A$. We denote by $\mathcal{D}(A)$ the set of density matrices on the Hilbert space $\mathcal{H}_A$. We use subscripts (or superscripts according to whichever is convenient) to denote the space to which a state belongs, e.g, ρ with subscript A indicates $\rho_A \in \mathcal{H}_A$. If two registers A and B are associated with the same Hilbert space, we represent this relation by $A \equiv B$. For two registers A and B, we denote the combined register as AB, which is associated with Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. For two quantum states $\rho \in \mathcal{D}(A)$ and $\sigma \in \mathcal{D}(B)$, $\rho \otimes \sigma \in \mathcal{D}(AB)$ represents the tensor product (or Kronecker product) of ρ and σ. The identity operator on $\mathcal{H}_A$ is denoted $\mathbb{1}_A$.

Let $\rho_{AB} \in \mathcal{D}(AB)$. We define the *partial trace with respect to A* of ρ_{AB} as

$$\rho_B \coloneqq \mathrm{Tr}_A(\rho_{AB}) \coloneqq \sum_i (\langle i| \otimes \mathbb{1}_B)\rho_{AB}(|i\rangle \otimes \mathbb{1}_B),$$

where $\{|i\rangle\}_i$ is an orthonormal basis for the Hilbert space $\mathcal{H}_A$. The state $\rho_B \in \mathcal{D}(B)$ is referred to as a *reduced density matrix* or a *marginal state*. Unless otherwise stated, a missing register from subscript in a state will represent partial trace over that register. Given a $\rho_A \in \mathcal{D}(A)$, a *purification* of ρ_A is a pure state $\rho_{AB} \in \mathcal{D}(AB)$ such that $\mathrm{Tr}_B(\rho_{AB}) = \rho_A$. Any quantum state has a purification using a register B with $|B| \leq |A|$. The purification of a state, even for a fixed B, is not unique as any unitary applied on register B alone does not change ρ_A.

An important class of states that we will consider is the *classical quantum states*. They are of the form $\rho_{AB} = \sum_a \mu(a) |a\rangle\langle a|_A \otimes \rho_B^a$, where μ is a probability distribution. In this case, ρ_A can

be viewed as a probability distribution and we shall continue to use the notations that we have introduced for probability distribution, for example, $\mathbb{E}_{a \leftarrow A}$ to denote the average $\sum_a \mu(a)$.

A quantum *super-operator* (or a *quantum channel* or a *quantum operation*) $\mathcal{E} : A \to B$ is a completely positive and trace preserving (CPTP) linear map (mapping states from $\mathcal{D}(A)$ to states in $\mathcal{D}(B)$). The identity operator in Hilbert space $\mathcal{H}_A$ (and associated register A) is denoted $\mathbb{1}_A$. A *unitary* operator $\mathcal{U}_A : \mathcal{H}_A \to \mathcal{H}_A$ is such that $\mathcal{U}_A^\dagger \mathcal{U}_A = \mathcal{U}_A \mathcal{U}_A^\dagger = \mathbb{1}_A$. The set of all unitary operations on register A is denoted by $\mathcal{U}(A)$.

A 2-outcome quantum measurement is defined by a collection $\{M, \mathbb{1} - M\}$, where $0 \preceq M \preceq \mathbb{1}$ is a positive semidefinite operator, where $A \preceq B$ means $B - A$ is positive semidefinite. Given a quantum state ρ, the probability of getting outcome corresponding to M is $\text{Tr}(\rho M)$ and getting outcome corresponding to $\mathbb{1} - M$ is $1 - \text{Tr}(\rho M)$.

2.1.1 Distance measures for quantum states

We now define the distance measures we use and some properties of these measures. Before defining the distance measures, we introduce the concept of *fidelity* between two states, which is not a distance measure but a similarity measure.

Definition 9 (Fidelity). Let $\rho_A, \sigma_A \in \mathcal{D}(A)$ be quantum states. The fidelity between ρ and σ is defined as

$$\text{F}(\rho_A, \sigma_A) := \|\sqrt{\rho_A}\sqrt{\sigma_A}\|_1 .$$

For two pure states $|\psi\rangle$ and $|\phi\rangle$, we have $\text{F}(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|) = |\langle\psi|\phi\rangle|$. We now introduce the two distance measures we use.

Definition 10 (Distance measures). Let $\rho_A, \sigma_A \in \mathcal{D}(A)$ be quantum states. We define the following distance measures between these states.

$$\text{Trace distance:} \quad \Delta(\rho_A, \sigma_A) := \frac{1}{2}\|\rho_A - \sigma_A\|_1$$
$$\text{Bures metric:} \quad \text{B}(\rho_A, \sigma_A) := \sqrt{1 - \text{F}(\rho_A, \sigma_A)}.$$

Note that for any two quantum states ρ_A and σ_A, these distance measures lie in $[0, 1]$. The distance measures are 0 if and only if the states are equal, and the distance measures are 1 if and only if the states have orthogonal support, i.e., if $\rho_A \rho_B = 0$.

Conveniently, these measures are closely related.

Fact 11. *For all quantum states $\rho_A, \sigma_A \in \mathcal{D}(A)$, we have*

$$1 - \text{F}(\rho_A, \sigma_A) \leq \Delta(\rho_A, \sigma_A) \leq \sqrt{2} \cdot \text{B}(\rho_A, \sigma_A).$$

Proof. The Fuchs-van de Graaf inequalities [FvdG99, Wat16] state that

$$1 - \text{F}(\rho_A, \sigma_A) \leq \Delta(\rho_A, \sigma_A) \leq \sqrt{1 - \text{F}^2(\rho_A, \sigma_A)}.$$

Our fact follows from this and the relation $1 - \text{F}^2(\rho_A, \sigma_A) \leq 2 - 2\text{F}(\rho_A, \sigma_A)$. $\qquad\square$

A fundamental fact about quantum states is Uhlmann's theorem [Uhl76].

Fact 12 (Uhlmann's theorem). *Let $\rho_A, \sigma_A \in \mathcal{D}(A)$. Let $\rho_{AB} \in \mathcal{D}(AB)$ be a purification of ρ_A and $\sigma_{AB} \in \mathcal{D}(AB)$ be a purification of σ_A with. There exists a unitary $\mathcal{U} : \mathcal{H}_B \to \mathcal{H}_B$ such that*

$$\text{F}(|\theta\rangle\langle\theta|_{AB}, |\rho\rangle\langle\rho|_{AB}) = \text{F}(\rho_A, \sigma_A),$$

where $|\theta\rangle_{AB} = (\mathbb{1}_A \otimes \mathcal{U})|\sigma\rangle_{AB}$. Trivially, the same holds for the Bures metric B as well.

We now review some properties of the Bures metric that we use in our proofs.

Fact 13 (Facts about B). *For all quantum states $\rho_A, \rho'_A, \sigma_A, \sigma'_A \in \mathcal{D}(A)$, we have the following.*

Fact 13.A (Triangle inequality [Bur69]). *The following triangle inequality and a weak triangle inequality hold for the Bures metric and the square of the Bures metric.*

1. $\mathrm{B}(\rho_A, \sigma_A) \leq \mathrm{B}(\rho_A, \tau_A) + \mathrm{B}(\tau_A, \sigma_A)$.

2. $\mathrm{B}^2(\rho_A^1, \rho_A^{t+1}) \leq t \cdot \sum_{i=1}^{t} \mathrm{B}^2(\rho_A^i, \rho_A^{i+1})$.

Fact 13.B (Product states). $\mathrm{B}(\rho_A \otimes \sigma_A, \rho'_A \otimes \sigma'_A) \leq \mathrm{B}(\rho_A, \rho'_A) + \mathrm{B}(\sigma_A, \sigma'_A)$. *Additionally, if $\sigma_A = \sigma'_A$ then* $\mathrm{B}(\rho_A \otimes \sigma_A, \rho'_A \otimes \sigma'_A) = \mathrm{B}(\rho_A, \rho'_A)$.

Fact 13.C (Partial measurement). *For classical-quantum states $\theta_{XB}, \theta'_{XB}$ with same probability distribution on the classical part, we have*

$$\mathrm{B}^2(\theta_{XB}, \theta'_{XB}) = \mathbb{E}_{x \leftarrow X}[\mathrm{B}^2(\theta_B^x, \theta_B'^x)].$$

Proof. These facts are proved as follows.

A. Proof of part 2 follows from triangle inequality and the fact that for positive reals $a_1, a_2, \ldots a_t$,

$$\left(\sum_i a_i\right)^2 = \sum_i a_i^2 + 2\sum_{i<j} a_i \cdot a_j \leq \sum_i a_i^2 + \sum_{i<j}\left(a_i^2 + a_j^2\right) \leq t\left(\sum_i a_i^2\right).$$

B. Follows easily from the triangle inequality.

C. Let $\theta_{XB} = \sum_x p(x) |x\rangle\langle x| \otimes \theta_B^x$ and $\theta'_{XB} = \sum_x p(x) |x\rangle\langle x| \otimes \theta_B'^x$. Then

$$\begin{aligned}
\mathrm{F}(\theta_{XB}, \theta'_{XB}) &= \mathrm{Tr}\left(\sqrt{\sum_x p^2(x) |x\rangle\langle x| \otimes \sqrt{\theta_B^x}\theta_B'^x\sqrt{\theta_B^x}}\right) \\
&= \mathrm{Tr}\left(\sum_x p(x) |x\rangle\langle x| \otimes \sqrt{\sqrt{\theta_B^x}\theta_B'^x\sqrt{\theta_B^x}}\right) \\
&= \sum_x p(x)\mathrm{F}(\theta_B^x, \theta_B'^x) \\
&= \mathbb{E}_{x \leftarrow X}[\mathrm{F}(\theta_B^x, \theta_B'^x)],
\end{aligned}$$

which proves the fact. □

Finally, an important property of both these distance measures is monotonicity under quantum operations [Lin75, BCF$^+$96].

Fact 14 (Monotonicity under quantum operations). *For quantum states $\rho_A, \sigma_A \in \mathcal{D}(A)$, and a quantum operation $\mathcal{E}(\cdot) : \mathcal{L}(A) \to \mathcal{L}(B)$, it holds that*

$$\Delta(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq \Delta(\rho_A, \sigma_A) \quad \text{and} \quad \mathrm{B}(\mathcal{E}(\rho_A), \mathcal{E}(\sigma_A)) \leq \mathrm{B}(\rho_A, \sigma_A),$$

with equality if $\mathcal{E}$ is unitary. In particular, for bipartite states $\rho_{AB}, \sigma_{AB} \in \mathcal{D}(AB)$, it holds that

$$\Delta(\rho_{AB}, \sigma_{AB}) \geq \Delta(\rho_A, \sigma_A) \quad \text{and} \quad \mathrm{B}(\rho_{AB}, \sigma_{AB}) \geq \mathrm{B}(\rho_A, \sigma_A).$$

2.1.2 Mutual information and relative entropy

We start with the following fundamental information theoretic quantities. We refer the reader to the excellent sources for quantum information theory [Wil12, Wat16] for further study.

Definition 15. Let $\rho_A \in \mathcal{D}(A)$ be a quantum state and $\sigma_A \in \mathcal{D}(A)$ be another quantum state on the same space with $\mathrm{supp}(\rho_A) \subset \mathrm{supp}(\sigma_A)$. We then define the following.

$$\text{von Neumann entropy:} \quad \mathrm{S}(\rho_A) := -\mathrm{Tr}(\rho_A \log \rho_A).$$
$$\text{Relative entropy:} \quad \mathrm{S}(\rho_A \| \sigma_A) := \mathrm{Tr}(\rho_A \log \rho_A) - \mathrm{Tr}(\rho_A \log \sigma_A).$$

We now define mutual information and conditional mutual information.

Definition 16 (Mutual information). Let $\rho_{ABC} \in \mathcal{D}(ABC)$ be a quantum state. We define the following measures.

$$\text{Mutual information:} \quad \mathbb{I}(A:B)_\rho := \mathrm{S}(\rho_A) + \mathrm{S}(\rho_B) - \mathrm{S}(\rho_{AB}) = \mathrm{S}(\rho_{AB} \| \rho_A \otimes \rho_B).$$
$$\text{Conditional mutual information:} \quad \mathbb{I}(A:B \mid C)_\rho := \mathbb{I}(A:BC)_\rho - \mathbb{I}(A:C)_\rho.$$

We will need the following basic properties.

Fact 17 (Properties of S and $\mathbb{I}$). *Let $\rho_{ABC} \in \mathcal{D}(ABC)$ be a quantum state. We have the following.*

Fact 17.A (Nonnegativity).

$$\mathrm{S}(A\|B)_\rho \geq 0 \ and \ |A| \geq \mathrm{S}(A)_\rho \geq 0$$
$$\mathbb{I}(A:B)_\rho \geq 0 \ and \ \mathbb{I}(A:B \mid C)_\rho \geq 0.$$

Fact 17.B (Partial measurement). *For classical-quantum states, $\theta_{XB}, \theta'_{XB}$ with same classical distribution on register X:*
$$\mathrm{S}\left(\theta_{XB} \| \theta'_{XB}\right) = \mathbb{E}_{x \leftarrow X}[\mathrm{S}\left(\theta_B^x \| \theta_B'^x\right)].$$

Fact 17.C (Chain rule). $\mathbb{I}(A:BC)_\rho = \mathbb{I}(A:C)_\rho + \mathbb{I}(A:B \mid C)_\rho = \mathbb{I}(A:B)_\rho + \mathbb{I}(A:C \mid B)_\rho.$

Fact 17.D (Monotonicity). *For a quantum operation $\mathcal{E}(\cdot) : \mathcal{L}(A) \to \mathcal{L}(B)$, $\mathbb{I}(A:\mathcal{E}(B)) \leq I(A:B)$ with equality when $\mathcal{E}$ is unitary. In particular $\mathbb{I}(A:BC)_\rho \geq \mathbb{I}(A:B)_\rho.$*

Fact 17.E (Bar hopping). $\mathbb{I}(A:BC)_\rho \geq \mathbb{I}(A:B \mid C)_\rho$, *where equality holds if $\mathbb{I}(A:C)_\rho = 0$.*

Fact 17.F (Independence). *If $\mathbb{I}(B:C)_\rho = 0$, then $\mathbb{I}(A:BC)_\rho \geq \mathbb{I}(A:B)_\rho + \mathbb{I}(A:C)_\rho.$*

Fact 17.G (Araki-Lieb inequality). $|\mathrm{S}(\rho_{AB}) - \mathrm{S}(\rho_B)| \leq \mathrm{S}(\rho_A).$

Fact 17.H (Information bound).

$$\mathbb{I}(A:BC)_\rho \leq \mathbb{I}(A:C)_\rho + 2\mathrm{S}(\rho_B).$$

Fact 17.I (Stronger version of Pinsker's inequality). *For quantum states ρ and σ:*

$$\mathrm{S}(\rho\|\sigma) \geq 1 - \mathrm{F}(\rho,\sigma) = \mathrm{B}^2(\rho,\sigma).$$

Fact 17.J. *For classical-quantum state (register X is classical) ρ_{XAB}:*

$$\mathbb{I}(A;B|X)_\rho = \mathbb{E}_{x \leftarrow X}\mathrm{S}(\rho_{AB}^x \| \rho_A^x \otimes \rho_B^x) \geq \mathbb{E}_{x \leftarrow X}\mathrm{B}^2\left(\rho_{AB}^x, \rho_A^x \otimes \rho_B^x\right).$$
$$\mathbb{I}(X;A) = \mathrm{S}(\rho_{XA} \| \rho_X \otimes \rho_A) = \mathbb{E}_{x \leftarrow X}\mathrm{S}(\rho_A^x \| \rho_A).$$
$$\mathbb{I}(X;A) = \mathbb{I}(f(X)X;A), \ where \ f \ is \ any \ function.$$

Proof. These facts are proved as follows.

A. For nonnegativity of relative entropy, see [NC00, Theorem 11.7]. For nonnegativity of mutual information and conditional mutual information, see [Wil12, Theorem 11.6.1] and [Wil12, Theorem 11.7.1].

B. Let $\theta_{XB} = \sum_x p(x) \, |x\rangle\langle x| \otimes \theta_B^x$ and $\theta'_{XB} = \sum_x p(x) \, |x\rangle\langle x| \otimes \theta_B'^x$. Then

$$
\begin{aligned}
\mathrm{S}\!\left(\theta_{XB}\middle\|\theta'_{XB}\right) &= \sum_x \mathrm{Tr}(p(x)\,|x\rangle\langle x| \otimes \theta_B^x (\log\theta_{XB} - \log\theta'_{XB})) \\
&= \sum_x p(x)\mathrm{Tr}(\theta_B^x(\log(p(x)\theta_B^x) - \log(p(x)\theta_B'^x))) \\
&= \sum_x p(x)\mathrm{Tr}(\theta_B^x(\log\theta_B^x - \log\theta_B'^x)) \\
&= \mathbb{E}_{x\leftarrow X}\mathrm{S}\!\left(\theta_B^x\middle\|\theta_B'^x\right),
\end{aligned}
$$

which proves the fact.

C. Follows from direct calculation.

D. See [NC00] [Theorem 11.15].

E. Follows from Chain rule (Fact 17.C) and Non-negativity (Fact 17.A).

F. Consider the following relations that use chain rule:

$$
\begin{aligned}
\mathbb{I}(A:BC)_\rho &= \mathbb{I}(A:B)_\rho + \mathbb{I}(A:C\mid B)_\rho \\
&= \mathbb{I}(A:B)_\rho + \mathbb{I}(AB:C)_\rho - \mathbb{I}(B:C)_\rho \\
&\geq \mathbb{I}(A:B)_\rho + \mathbb{I}(A:C)_\rho.
\end{aligned}
$$

The last line uses $\mathbb{I}(B:C)_\rho = 0$ and monotonicity (Fact 17.D).

G. See [NC00] [Section 11.3.4].

H. Consider,

$$
\begin{aligned}
\mathbb{I}(A:BC)_\rho &= \mathbb{I}(A:C)_\rho + \mathbb{I}(CA:B)_\rho - \mathbb{I}(B:C) \\
&\leq \mathbb{I}(A:C)_\rho + \mathbb{I}(CA:B)_\rho \\
&\leq \mathbb{I}(A:C)_\rho + \mathrm{S}(B) + \mathrm{S}(CA) - \mathrm{S}(CAB) \\
&\leq \mathbb{I}(A:C)_\rho + 2\mathrm{S}(B). && \text{(Fact 17.G)}
\end{aligned}
$$

I. Using Corollary 4.2 and Proposition 4.5 in [Tom16], we find that

$$
\mathrm{S}(\rho\|\sigma) \geq -2\log \mathrm{F}(\rho,\sigma).
$$

The fact now follows since for any positive $x < 1$, $2^x > 2 \cdot x^2$.

J. For the first relation, we proceed as follows, and then use Pinsker's inequality.

$$
\begin{aligned}
\mathbb{I}(A:B\mid X)_\rho &= \mathbb{I}(A:BX)_\rho - \mathbb{I}(A:X)_\rho \\
&= \mathrm{S}(\rho_{ABX}\|\rho_A \otimes \rho_{BX}) - \mathrm{S}(\rho_{AX}\|\rho_A \otimes \rho_X)
\end{aligned}
$$

$$\begin{aligned}
&= \mathbb{E}_{x \leftarrow X}[S(\rho_{AB}^x \| \rho_A \otimes \rho_B^x) - S(\rho_A^x \| \rho_A)]\\
&= \mathbb{E}_{x \leftarrow X}[-S(\rho_{AB}^x) - \text{Tr}(\rho_A^x \log \rho_A) + S(\rho_B^x) + S(\rho_A^x) + \text{Tr}(\rho_A^x \log \rho_A)]\\
&= \mathbb{E}_{x \leftarrow X}[-S(\rho_{AB}^x) + S(\rho_B^x) + S(\rho_A^x)] = \mathbb{E}_{x \leftarrow X}[S(\rho_{AB}^x \| \rho_A^x \otimes \rho_B^x)],
\end{aligned}$$

where in third line, we have used Fact 17.B. The second relation follows by direct calculation and Fact 13.C. The third relation follows by monotonicity under the maps $|x\rangle\langle x| \to |x\rangle\langle x| \otimes |f(x)\rangle\langle f(x)|$ and partial trace. $\qquad\square$

We will need the following relation between $\mathbb{I}$ and Δ for binary classical-quantum states (see also [JN06]).

Claim 18. *Let $\rho_{AB} \in \mathcal{D}(AB)$ be a classical quantum state of the form $\rho_{AB} = p\,|0\rangle\langle 0|_A \otimes \rho_B^0 + (1 - p)\,|1\rangle\langle 1| \otimes \rho_B^1$. Then*

$$\mathbb{I}(A : B)_\rho \le 2\log(2) \cdot \Delta(p\rho_B^0, (1 - p)\rho_B^1).$$

Proof. We drop the register index from ρ_B^0, ρ_B^1. Let $\rho_{av} = p\rho^0 + (1 - p)\rho^1$. Consider

$$\begin{aligned}
\mathbb{I}(A : B)_\rho &= pS\left(\rho^0 \| \rho_{av}\right) + (1 - p)S\left(\rho^1 \| \rho_{av}\right) && \text{(Fact 17.J)}\\
&= S\left(p\rho^0 \Big\| \tfrac{1}{2}\rho_{av}\right) + S\left((1 - p)\rho^1 \Big\| \tfrac{1}{2}\rho_{av}\right) - p\log(2) - (1 - p)\log(2) + S(p)\\
&\le S\left(p\rho^0 \Big\| \tfrac{1}{2}\rho_{av}\right) + S\left((1 - p)\rho^1 \Big\| \tfrac{1}{2}\rho_{av}\right).
\end{aligned}$$

The last inequality follows from $S(p) \le \log(2)$. Now, using [Aud14, Theorem 9], which states that

$$S\left(p\rho^0 \Big\| \tfrac{1}{2}\rho_{av}\right) \le \log(2)\Delta(p\rho^0, (1 - p)\rho^1) \quad \text{and} \quad S\left((1 - p)\rho^1 \Big\| \tfrac{1}{2}\rho_{av}\right) \le \log(2)\Delta(p\rho^0, (1 - p)\rho^1),$$

the claim follows. $\qquad\square$

Our next claim gives us a way to use high mutual information between two registers in a classical quantum state to make a prediction about the classical part using measurement on the quantum part.

Claim 19 (Information $\Rightarrow$ prediction). *Let $\rho_{AB} \in \mathcal{D}(AB)$ be a classical quantum state of the form $\rho_{AB} = p\,|0\rangle\langle 0|_A \otimes \rho_B^0 + (1 - p)\,|1\rangle\langle 1| \otimes \rho_B^1$. The probability of predicting A by a measurement on B is at least*

$$\frac{1}{2} + \frac{\mathbb{I}(A : B)}{2\log 2}.$$

Proof. We drop the register label B. Let M be a projector on the support of positive eigenvectors of the state $p\rho^0 - (1 - p)\rho^1$. Let the measurement be $\{M, \mathbb{1} - M\}$ and first outcome imply 0 in register A and second outcome imply 1. Then probability of success is

$$\begin{aligned}
p\text{Tr}(\rho^0 M) + (1 - p)\text{Tr}(\rho^1(\mathbb{1} - M)) &= (1 - p) + \text{Tr}((p\rho^0 - (1 - p)\rho^1)M)\\
&= (1 - p) + \frac{1}{2}(\|p\rho^0 - (1 - p)\rho^1\|_1 + \text{Tr}(p\rho^0 - (1 - p)\rho^1))\\
&= (1 - p) + \frac{1}{2}(\|p\rho^0 - (1 - p)\rho^1\|_1 + 2p - 1)\\
&= \frac{1}{2} + \frac{1}{2}\|p\rho^0 - (1 - p)\rho^1\|_1\\
&= \frac{1}{2} + \Delta(p\rho^0, (1 - p)\rho^1).
\end{aligned}$$

From Claim 18, we know that $\Delta(p\rho^0, (1 - p)\rho^1) \ge \mathbb{I}(A : B)/(2\log 2)$. $\qquad\square$

2.2 Quantum Communication complexity

In quantum communication complexity, two players wish to compute a classical function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ for some finite sets $\mathcal{X}$ and $\mathcal{Y}$. The inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ are given to two players Alice and Bob, and the goal is to minimize the quantum communication between them required to compute the function.

While the players have classical inputs, the players are allowed to exchange quantum messages. Depending on whether or not we allow the players arbitrary shared entanglement, we get $\mathrm{Q}(F)$, bounded-error quantum communication complexity without shared enganglement and $\mathrm{Q}^*(F)$, for the same measure with shared entanglement. Obviously $\mathrm{Q}^*(F) \leq \mathrm{Q}(F)$. In this paper we will only work with $\mathrm{Q}^*(F)$, which makes our results stronger since we prove lower bounds in this work.

Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1,*\}$ be a partial function, with $\mathrm{dom}(F) := \{(x,y) \in \mathcal{X} \times \mathcal{Y} : F(x,y) \neq *\}$, and let $\varepsilon \in (0,1/2)$.

An entanglement assisted quantum communication protocol Π for this function is as follows. Alice and Bob start with a preshared entanglement. Upon receiving inputs (x,y), where Alice gets x and Bob gets y, they exchange quantum states and then Alice applies a measurement on her qubits to output 1 or 0. Let $O(x,y)$ be the random variable output by Alice in Π, given input (x,y). Let μ be a distribution over $\mathrm{dom}(F)$.

Let inputs to Alice and Bob be given in registers X and Y in the state

$$\sum_{x,y} \mu(x,y) |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \,.$$

Let these registers be purified by R_X and R_Y respectively, which are not accessible to either players. Let Alice and Bob initially hold register A_0, B_0 with shared entanglement Θ_{0,A_0B_0}. Then the initial state is

$$|\Psi_0\rangle_{XYR_XR_YA_0B_0} := \sum_{x,y} \sqrt{\mu(x,y)} |xxyy\rangle_{XR_XYR_Y} |\Theta_0\rangle_{A_0B_0}$$

Alice applies a unitary $U^1 : XA_0 \to XA_1C_1$ such that the unitary acts on A_0 conditioned on X. She sends C_1 to Bob. Let $B_1 \equiv B_0$ be a relabelling of Bob's register B_0. He applies $U^2 : YC_1B_1 \to YC_2B_2$ such that the unitary acts on C_1B_0 conditioned on Y. He sends C_2 to Alice. Players proceed in this fashion till end of the protocol. At any round r, let the registers be $A_rC_rB_r$, where C_r is the message register, A_r is Alice's register and B_r is Bob's register. If r is odd, then $B_r \equiv B_{r-1}$ and if r is even, then $A_r \equiv A_{r-1}$. Let the joint state in registers $A_rC_rB_r$ be $\Theta_{r,A_rC_rB_r}$. Then the global state at round r is

$$|\Psi_r\rangle_{XYR_XR_YA_rC_rB_r} := \sum_{x,y} \sqrt{\mu(x,y)} |xxyy\rangle_{XR_XYR_Y} |\Theta_r\rangle_{A_rC_rB_r} \,.$$

We define the following quantities.

$$\text{Worst-case error:} \quad \mathrm{err}(\Pi) := \max_{(x,y)\in\mathrm{dom}(F)} \{\Pr[O(x,y) \neq F(x,y)]\}.$$

$$\text{Distributional error:} \quad \mathrm{err}^\mu(\Pi) := \mathbb{E}_{(x,y)\leftarrow\mu}\Pr[O(x,y) \neq F(x,y)].$$

$$\text{Quantum CC of a protocol:} \quad \mathrm{QCC}(\Pi) := \sum_i |C_i|.$$

$$\text{Quantum CC of } F\colon \quad \mathrm{Q}^*_\varepsilon(F) := \min_{\Pi:\mathrm{err}(\Pi)\leq\varepsilon} \mathrm{QCC}(\Pi).$$

Our first fact justifies using $\varepsilon = 1/3$ by default since the exact constant does not matter since the success probability of a protocol can be boosted for QCC.

Fact 20 (Error reduction). *Let $0 < \delta < \varepsilon < 1/2$. Let Π be a protocol for F with $\mathrm{err}(\Pi) \leq \varepsilon$. There exists protocol Π' for F such that $\mathrm{err}(\Pi') \leq \delta$ and*

$$\mathrm{QCC}(\Pi') \leq O\left(\frac{\log(1/\delta)}{\left(\frac{1}{2} - \varepsilon\right)^2} \cdot \mathrm{QCC}(\Pi) \right).$$

This fact is proved by simply repeating the protocol sufficiently many times and taking the majority vote of the outputs. If the error ε is close to $1/2$, we can first reduce the error to a constant by using $O(\frac{1}{(1/2 - \varepsilon)^2})$ repetitions. Then $O(\log(1/\delta))$ repetitions suffice to reduce the error down to δ. Hence the quantum communication only increases by a factor of $O\left(\frac{\log(1/\delta)}{(1/2 - \varepsilon)^2} \right)$.

We have the following relation between worst-case and average-case error quantum communication complexities. It follows for example from standard application of Sion's minimax theorem [Sio58].

Fact 21 (Minimax principle). *Let $F \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ be a partial function. Fix an error parameter $\varepsilon \in (0, 1/2)$ and a quantum communication bound $q \geq 0$. Suppose $\mathcal{F}$ is a family of protocols such that for every distribution μ on $\mathrm{dom}(F)$ there exists a protocol $\Pi \in \mathcal{F}$ such that*

$$\mathrm{err}^\mu(\Pi) \leq \varepsilon \quad \text{and} \quad \mathrm{QCC}(\Pi) \leq q.$$

Then there exists a protocol Π' such that

$$\mathrm{err}(\Pi') \leq \varepsilon \quad \text{and} \quad \mathrm{QCC}(\Pi') \leq q.$$

Our next claim shows that having some information about the output of a Boolean function F allows us to predict the output of F with some probability greater than $1/2$.

Claim 22. *Let $F \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ be a partial function and μ be a distribution over $\mathrm{dom}(F)$. Let XY be registers with the state $\sum_{x,y} \mu(x, y) |x\rangle\langle x| \otimes |y\rangle\langle y|$ and define a register F that contains the value of $F(x, y)$. Let Π be a quantum communication protocol with registers X, Y input to Alice and Bob respectively and number of rounds r (which is even). There either*

- *There exists a quantum communication protocol Π' for F with r rounds, with input (X, Y) to Alice and Bob respectively, such that*

$$\mathrm{QCC}(\Pi') = \mathrm{QCC}(\Pi) + 1, \quad \text{and} \quad \mathrm{err}^\mu(\Pi') < \frac{1}{2} - \frac{\mathbb{I}(F : A_r C_r \mid X)_{\Psi_r}}{2 \log(2)}.$$

- *Or, there exists a quantum communication protocol Π' for F with r rounds, with input (X, Y) to Alice and Bob respectively, such that*

$$\mathrm{QCC}(\Pi') \leq \mathrm{QCC}(\Pi), \quad \text{and} \quad \mathrm{err}^\mu(\Pi') < \frac{1}{2} - \frac{\mathbb{I}(F : B_r C_r \mid Y)_{\Psi_r}}{2 \log(2)}.$$

Proof. We first prove the first case. In Π', Alice and Bob run the protocol Π, after which Alice proceeds as follows. Consider the state $\Psi_{r, XFA_r C_r}$ in registers $XFA_r C_r$ (note that we have added a new register F to the state Ψ_r, which can be done naturally). Let

$$\Psi_{r, XFA_r C_r} = \sum_x \mu(x) |x\rangle\langle x|_X \otimes \Psi^x_{r, FA_r C_r}$$

13

be the decomposition of Ψ_{r,XFA_rC_r}, which is possible since X is classical. Note that Ψ^x_{r,FA_rC_r} is a classical quantum state between the registers F and A_rC_r. Alice, essentially applying Claim 19 makes a prediction about the content of register F. Then she outputs the prediction. Clearly,

$$\text{QCC}(\Pi') = \text{QCC}(\Pi) + 1.$$

For every input x for Alice, her prediction is successful with probability at least $1/2 + \mathbb{I}(F : A_rC_r)_{\Psi^x_r}/2\log(2)$ by Claim 19. Hence the overall success probability of Π' is at least

$$\mathbb{E}_{x\leftarrow X}\left[\frac{1}{2} + \frac{\mathbb{I}(F : A_rC_r)_{\Psi^x_r}}{2\log(2)}\right] = \frac{1}{2} + \frac{\mathbb{I}(F : A_rC_r|X)_{\Psi_r}}{2\log(2)}. \qquad \square$$

Second case follows with same argument, but applied on Bob' side before he sends C_r to Alice. Bob then sends the outcome to Alice instead of C_r.

The following claim is used in our proof to handle the easy case of a biased input distribution.

Claim 23. *Let $F \colon \mathcal{X} \times \mathcal{Y} \to \{0,1,*\}$ be a partial function and let μ be a distribution over $\text{dom}(F)$. Let $\varepsilon \in (0,1/2)$ and $c \geq 1$ be a positive integer. For $i \in [c]$, let X_i, Y_i be registers with the state $\sum_{x,y} \mu(x,y) |x\rangle\langle x|_{X_i} \otimes |y\rangle\langle y|_{Y_i}$ and define register L_i that holds the value $F(x_i, y_i)$. Define $X := X_1 \dots X_c$, $Y := Y_1 \dots Y_c$, and $L := L_1 \dots L_c$. Let Ψ_{XYL} be the joint state in registers X, Y, L. Then either*

(a) There exists a protocol Π for F such that $\text{QCC}(\Pi) = 1$, and $\text{err}^\mu(\Pi) \leq \frac{1}{2} - \varepsilon$, or

(b) $\Delta(\Psi_{XL}, \Psi_X \otimes W_{L_1} \otimes \dots W_{L_c}) \leq c\varepsilon$, where W_{L_i} is the maximally mixed state in register L_i.

Proof. Define, $q^{x_1} := \Pr[F = 0 \mid X_1 = x_1]$. Assume $\mathbb{E}_{x_1\leftarrow X_1}\left[\left|\frac{1}{2} - q^{x_1}\right|\right] \geq \varepsilon$. Let Π be a protocol where Alice, on input x_1, outputs 0 if $q^{x_1} \geq 1/2$ and 1 otherwise. Then,

$$\text{err}^\mu(\Pi) = \frac{1}{2} - \mathbb{E}_{x_1\leftarrow X_1}\left|\frac{1}{2} - q^{x_1}\right| \leq \frac{1}{2} - \varepsilon.$$

Assume otherwise $\mathbb{E}_{x_1\leftarrow X_1}\left|\frac{1}{2} - q^{x_1}\right| < \varepsilon$. This implies

$$\Delta(\Psi_{XL}, \Psi_X \otimes W_{L_1} \otimes \dots W_{L_c}) \leq c \cdot \Delta(\Psi_{X_1L_1}, \Psi_{X_1} \otimes W_{L_1}) = c \cdot \mathbb{E}_{x_1\leftarrow X_1}\left|\frac{1}{2} - q^{x_1}\right| < c\varepsilon,$$

where the first inequality follows from Fact 13.B. $\qquad \square$

In below, let A'_r, B'_r represent Alice and Bob's registers at round r. That is, at even round r, $A'_r = A_rC_r, B'_r = B_r$ and at odd r, $A'_r = A_r, B'_r = B_rC_r$. We will need the following version of quantum-cut-and-paste lemma from [NT16] (also see [JRS03] for a similar argument, where it is used to lower bound quantum communication complexity of disjointness). This is a special case of [NT16, Lemma 7] and we have rephrased it using our notation.

Lemma 24 (Quantum cut-and-paste). *Let Π be a quantum protocol with classical inputs and consider distinct inputs u, u' for Alice and v, v' for Bob. Let $|\Psi_{0,A_0B_0}\rangle$ be the initial shared state between Alice and Bob. Also let $\left|\Psi^{u'',v''}_{k,A'_kB'_k}\right\rangle$ be the shared state after round k of the protocol when the inputs to Alice and Bob are (u'', v'') respectively. For k odd, let*

$$h_k = \text{B}\left(\Psi^{u,v}_{k,B'_k}, \Psi^{u',v}_{k,B'_k}\right)$$

14

and for even k, let

$$h_k = \mathrm{B}\left(\Psi_{k,A'_k}^{u,v}, \Psi_{k,A'_k}^{u,v'}\right).$$

Then

$$\mathrm{B}\left(\Psi_{r,A'_r}^{u',v}, \Psi_{r,A'_r}^{u',v'}\right) \leq h_r + h_{r-1} + 2\sum_{k=1}^{r-2} h_k.$$

The following lemma (see also [CvNT13]) formalizes the following intuition: In a quantum protocol with communication q, the amount of information that Bob has about Alice's input at any time point is at most $2q$ (note that the factor of 2 is necessary because of super-dense coding.).

Lemma 25. *Let Π be a quantum protocol with the inputs of Alice and Bob (X, Y) being jointly distributed. Alice has an additional input U which is independent of both (X, Y). Let μ denote the distribution of inputs so that $\mu(x, u, y) = \mu(x, y)\mu(u)$. Let the total pure state after the kth round of the protocol be*

$$|\Psi_k\rangle_{X\widetilde{X}Y\widetilde{Y}A'_k B'_k} = \sum_{x,y} \sqrt{\mu(x,y)\mu(u)}\, |xxuu\rangle_{X\widetilde{X}U\widetilde{U}}\, |yy\rangle_{Y\widetilde{Y}}\, |\Theta_k^{x,u,y}\rangle_{A'_k B'_k}.$$

Then

$$\mathbb{I}(B'_k Y\widetilde{Y} : U|X)_{\Psi_k} \leq 2q_k.$$

Here q_k is communication cost up to round k. A similar statement holds by reversing the roles of Alice and Bob.

Proof. We prove the first inequality by induction on k. The inequality holds trivially for $k = 0$. First suppose k is even, so that Bob sent the last message. Then,

$$\begin{aligned}
\mathbb{I}(B'_k Y\widetilde{Y} : U|X)_{\Psi_k} &\leq \mathbb{I}(B'_{k-1} Y\widetilde{Y} : U|X)_{\Psi_{k-1}} && \text{(Fact 17.D)} \\
&\leq 2q_{k-1} \leq 2q_k,
\end{aligned}$$

where the first inequality follows by induction step.

Now suppose k is odd, so that Alice sent the last message. By our notation, $B'_k \equiv C_k B_k$ where C_k is Alice's message. Then,

$$\begin{aligned}
\mathbb{I}(B'_k Y\widetilde{Y} : U|X)_{\Psi_k} &= \mathbb{I}(C_k B_k Y\widetilde{Y} : U|X)_{\Psi_k} \\
&\leq \mathbb{I}(B_k Y\widetilde{Y} : U|X)_{\Psi_k} + 2\mathrm{S}(C_k|X) && \text{(Fact 17.H)} \\
&= \mathbb{I}(B'_{k-1} Y\widetilde{Y} : U|X)_{\Psi_{k-1}} + 2\mathrm{S}(C_k|X) \\
&\leq 2q_{k-1} + 2|C_k| = 2q_k,
\end{aligned}$$

where last inequality follows from induction step. $\square$

3 Separation

In this section we establish the main result, a nearly quadratic separation between quantum communication complexity and the logarithm of approximate rank, which we restate below.

Theorem 1. *There is a family of total functions $F : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ with $\mathrm{Q}^*(F) = \widetilde{\Omega}\left(\log^2 \mathrm{rk}_{1/3}(F)\right)$.*

Our proof is organized as follows. In Section 3.1 we define lookup functions, which we will use to construct the function achieving the separation in Theorem 1. Then in Section 3.2 we prove Theorem 1 using results from later sections. More precisely, we prove the upper bound on our function's approximate rank using Theorem 6, proved in Section 4. We prove the lower bound using Corollary 7, which follows from Theorem 11 in Section 5. Theorem 6 and Corollary 7 provide a black-box way of using the results of Section 4 and Section 5 without delving into their proofs.

3.1 Lookup functions

We define a simpler version of lookup functions than the ones used in [ABB+16b], since we only deal with total functions in this paper. This is only for simplicity, and the lower bound shown in this paper also applies to the more general lookup functions for partial functions defined in [ABB+16b].

First, for any function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and integer $c > 0$, we can define a new function $F^c\colon \mathcal{X}^c \times \mathcal{Y}^c \to \{0,1\}^c$ as $F^c((x_1,\ldots,x_c),(y_1,\ldots,y_c)) = (F(x_1,y_1),\ldots,F(x_c,y_c))$, which takes c inputs to F and outputs the answers to all c inputs. F^c is simply the problem of computing F on c independent inputs and outputting all c answers.

An $(F,\mathcal{G})$-lookup function, denoted $F_\mathcal{G}$, is defined by a function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and a family $\mathcal{G} = \{G_0,\ldots,G_{2^c-1}\}$ of functions, where each $G_i\colon (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$. It can be viewed as a generalization of the address function. Alice receives input $\mathbf{x} = (x_1,\ldots,x_c) \in \mathcal{X}^c$ and $\mathbf{u} = (u_0,\ldots,u_{2^c-1}) \in \{0,1\}^{m2^c}$ and likewise Bob receives input $\mathbf{y} = (y_1,\ldots,y_c) \in \mathcal{Y}^c$ and $\mathbf{v} = (v_0,\ldots,v_{2^c-1}) \in \{0,1\}^{m2^c}$. We refer to the inputs $(\mathbf{x},\mathbf{y})$ as the "address part" of the input and the inputs $(\mathbf{u},\mathbf{v})$ as the "array part" of the input. We will refer to u_i and v_i as a "cell" of the array. The *address*, ℓ, is determined by the evaluation of F on $(x_1,y_1),\ldots,(x_c,y_c)$, that is $\ell = F^c(\mathbf{x},\mathbf{y}) \in \{0,1\}^c$. This address (interpreted as an integer in $\{0,\ldots,2^c-1\}$) then determines which function, out of the 2^c functions G_i, the players should evaluate and which pair of cells, out of the 2^c possible pairs (u_i,v_i), of the array are relevant to the output of the function. The goal of the players is to output $G_\ell(\mathbf{x},u_\ell,\mathbf{y},v_\ell)$. The formal definition is the following.

Definition 4 (($F,\mathcal{G}$)-lookup function for total F)**.** Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function and $\mathcal{G} = \{G_0,\ldots,G_{2^c-1}\}$ a family of functions, where each $G_i\colon (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$. An $(F,\mathcal{G})$-lookup function, denoted $F_\mathcal{G}$, is a function

$$F_\mathcal{G}\colon (\mathcal{X}^c \times \{0,1\}^{m2^c}) \times \mathcal{Y}^c \times \{0,1\}^{m2^c} \to \{0,1\}$$

defined as follows. Let $\mathbf{x} = (x_1,\ldots,x_c) \in \mathcal{X}^c$, $\mathbf{y} = (y_1,\ldots,y_c) \in \mathcal{Y}^c$, $\mathbf{u} = (u_0,\ldots,u_{2^c-1}) \in \{0,1\}^{m2^c}$, and $\mathbf{v} = (v_0,\ldots,v_{2^c-1}) \in \{0,1\}^{m2^c}$. Then

$$F_\mathcal{G}(\mathbf{x},\mathbf{u},\mathbf{y},\mathbf{v}) = G_\ell(\mathbf{x},u_\ell,\mathbf{y},v_\ell),$$

where $\ell = F^c(\mathbf{x},\mathbf{y})$.

Since we only deal with total functions F, we will not need to impose a consistency condition for instances where some input to F is outside its domain. (In [ABB+16b], this condition was called "consistency outside F.")

In order to show lower bounds on the communication complexity of $F_\mathcal{G}$ (Theorem 11) we add two constraints on the family $\mathcal{G}$ as in [ABB+16b].

Definition 5 (Nontrivial XOR family)**.** Let $\mathcal{G} = \{G_0,\ldots,G_{2^c-1}\}$ a family of communication functions, where each $G_i\colon (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$. We say that $\mathcal{G}$ is a nontrivial XOR family if the following conditions hold.

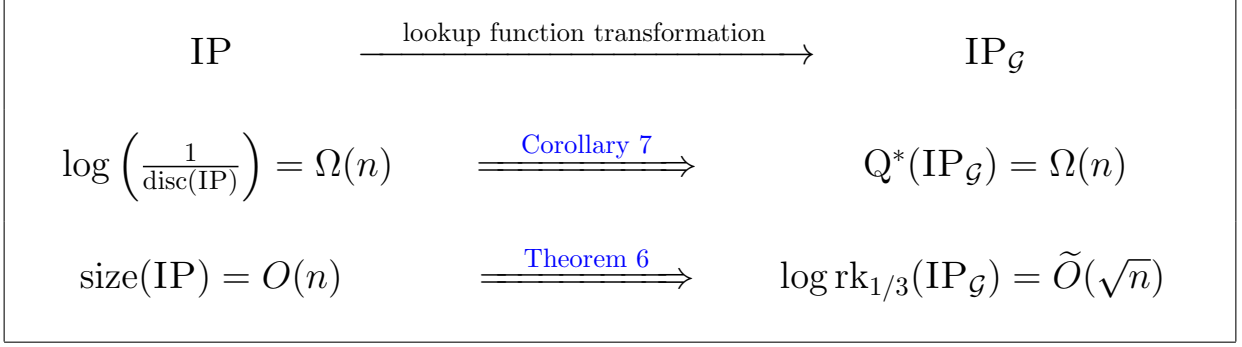Figure 1: High-level overview of our separation. Here IP : $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is the inner product function, disc is the discrepancy, and size is the circuit size.

1. (Nontriviality) For all $\mathbf{x} = (x_1, \ldots, x_c) \in \mathcal{X}^c$ and $\mathbf{y} = (y_1, \ldots, y_c) \in \mathcal{Y}^c$, if we have $\ell = F^c(\mathbf{x}, \mathbf{y}) \in \{0,1\}^c$ then there exist $u, v, u', v' \in \{0,1\}^m$ such that $G_\ell(\mathbf{x}, u, \mathbf{y}, v) \neq G_\ell(\mathbf{x}, u', \mathbf{y}, v')$.

2. (XOR function) For all $i \in \{0, \ldots, 2^c - 1\}, u, u', v, v' \in \{0,1\}^m$ and $\mathbf{x} = (x_1, \ldots, x_c) \in \mathcal{X}^c, \mathbf{y} = (y_1, \ldots, y_c) \in \mathcal{Y}^c$ if $u \oplus v = u' \oplus v'$ then $G_i(\mathbf{x}, u, \mathbf{y}, v) = G_i(\mathbf{x}, u', \mathbf{y}, v')$.

The first condition simply enforces that the content of the correct part of the array, i.e., (u_ℓ, v_ℓ), is relevant to the output of the function in the sense that there is some setting of these bits that makes the function true and another setting that makes it false.

The second condition enforces that the output of the function only depends on $u_\ell \oplus v_\ell$, and not u_ℓ and v_ℓ individually. This is just one way of combining the arrays of Alice and Bob to form one virtual array that contains 2^c cells. Other combining functions are also possible.

3.2 Separation

We can now prove the separation using results from Section 4 and Section 5. Our proof strategy is depicted in Figure 1.

The separating function is going to be a lookup function $F_\mathcal{G}$ defined by a function $F \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and a function family $\mathcal{G} = \{G_0, \ldots, G_{2^c - 1}\}$. We will choose F to be the well-known inner product function IP : $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ defined as

$$\mathrm{IP}(x, y) = \bigoplus_{i=1}^{n} (x_i \wedge y_i).$$

The communication complexity of the inner product function is well understood and is $\Theta(n)$ in all the models discussed in this paper. In fact, even $\log \text{sign-rank}(F) = \Theta(n)$ [For02], where sign-rank(F) is defined as the minimum rank of a matrix G such that $\ell_\infty(F - G) < 1/2$.

To define our function family $\mathcal{G}$, we use the following theorem proved in Section 4.

Theorem 6. *Let F be a total function with circuit size $\text{size}(F)$. Then for all $c > 0$, there exists a nontrivial family of XOR functions $\mathcal{G} = \{G_0, G_1, \ldots, G_{2^c - 1}\}$, such that*

$$\log \mathrm{rk}_{1/3}(F_\mathcal{G}) = \widetilde{O}(c^{3/2}\sqrt{\text{size}(F)}).$$

This theorem gives us a function family $\mathcal{G}$ and proves that for this family we have

$$\log \mathrm{rk}_{1/3}(\mathrm{IP}_\mathcal{G}) = \widetilde{O}(c^{3/2}\sqrt{\text{size}(\mathrm{IP})}) = \widetilde{O}(c^{3/2}\sqrt{n}), \tag{1}$$

17

where we use the fact that $\text{size}(\text{IP}) = O(n)$. This follows because IP is a parity of size n composed with an AND function on two bits, and has a circuit of size $O(n)$ consisting of a $\log n$-depth tree of fanin-2 XOR gates with fanin-2 AND gates at the bottom.

To show the lower bound, we use the following corollary of Theorem 11.

Corollary 7. *Let $F_{\mathcal{G}}$ be an $(F, \mathcal{G})$-lookup function for a function F and a nontrivial family of XOR functions $\mathcal{G} = \{G_0, G_1, \ldots, G_{2^c-1}\}$ with $c = \Theta(\log(Q^*(F)))$. Then*

$$Q^*(F_{\mathcal{G}}) = \Omega(\log(1/\text{disc}(F))).$$

Here $\text{disc}(F)$ is the discrepancy of F (Definition 6). Since $\log(1/\text{disc}(\text{IP})) = \Omega(n)$ [KN06, Example 3.19], using Theorem 11 we have

$$Q^*(\text{IP}_{\mathcal{G}}) = \Omega(\log(1/\text{disc}(\text{IP}))) = \Omega(n). \tag{2}$$

We can now choose $c = \Theta(\log n)$ to satisfy the conditions of Corollary 7. Thus (1) yields

$$\log \text{rk}_{1/3}(\text{IP}_{\mathcal{G}}) = \widetilde{O}(\sqrt{n}),$$

which together with (2) gives us $Q^*(\text{IP}_{\mathcal{G}}) = \widetilde{\Omega}(\log^2(\text{rk}_{1/3}(\text{IP}_{\mathcal{G}})))$, proving Theorem 1.

4 Upper bound on approximate rank of lookup functions

The aim of this section is to prove Theorem 6. Proving this will require some work and we will need to carefully choose our function family $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$. To do this, we first introduce the concept of an *unambiguous* lookup function.

Definition 8. Let $F_{\mathcal{G}}$ be an $(F, \mathcal{G})$-lookup function for a function $F : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ and a function family $\mathcal{G} = \{G_0, G_1, \ldots, G_{2^c-1}\}$. We say that $F_{\mathcal{G}}$ is an *unambiguous* lookup function if G_ℓ evaluating to 1 certifies that $F^c(\mathbf{x}, \mathbf{y}) = \ell$. That is, for all $\mathbf{x}, u, \mathbf{y}, v$, $G_\ell(\mathbf{x}, u, \mathbf{y}, v) = 1 \Rightarrow F^c(\mathbf{x}, \mathbf{y}) = \ell$.

Note that not all lookup functions are unambiguous even if we enforce the nontrivial XOR family condition (Definition 5), since the condition for when G_i evaluates to 1 need not even depend on $\mathbf{x}$ and $\mathbf{y}$. For example, $G_i(\mathbf{x}, u, \mathbf{y}, v)$ could simply be some nonconstant function of the string $u \oplus v$. However, the condition of unambiguity is quite natural, and the lookup functions used in prior work are unambiguous lookup functions (or can be slightly modified to be unambiguous).

The advantage of unambiguous lookup functions is that we can upper bound their approximate rank as follows.

Lemma 9. *Let $F_{\mathcal{G}}$ be an unambiguous $(F, \mathcal{G})$-lookup function. Then we have*

$$\log \text{rk}_{1/3}(F_{\mathcal{G}}) = O(c \cdot \max_i Q^*(G_i)).$$

Proof. We start by observing that the unambiguity condition implies that for any input $(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v})$, at most one of the functions $G_i(\mathbf{x}, u_i, \mathbf{y}, v_i)$ equals 1. Indeed, only $G_\ell(\mathbf{x}, u_\ell, \mathbf{y}, v_\ell)$ can potentially evaluate to 1, where $\ell = F^c(\mathbf{x}, \mathbf{y})$.

In other words, when $F_{\mathcal{G}}(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v}) = 1$ we must have $G_\ell(\mathbf{x}, u_\ell, \mathbf{y}, v_\ell) = 1$ for $\ell = F^c(x, y)$ and $G_i(\mathbf{x}, u_i, \mathbf{y}, v_i) = 0$ for all $i \neq \ell$. On the other hand, when $F_{\mathcal{G}}(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v}) = 0$ we must have $G_i(\mathbf{x}, u_i, \mathbf{y}, v_i) = 0$ for all $i \in \{0, \ldots, 2^c - 1\}$.

This means the communication matrix of $F_{\mathcal{G}}$ equals the sum of the communication matrices of G_i over all i. More precisely, we extend the definition of G_i to have it take all of $(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v})$ as

18

input in the natural way (i.e., it ignores all the other cells of the array except u_i and v_i). This observation directly yields

$$\mathrm{rk}(F_{\mathcal{G}}) \leq \sum_{i=0}^{2^c-1} \mathrm{rk}(G_i).$$

The same inequality does not immediately hold for approximate rank, because the errors in the approximation can add up. So even though $A = \sum_i B_i$, if $\tilde{B}_i$ satisfies $\ell_\infty(\tilde{B}_i - B_i) \leq 1/3$, it is not necessarily the case that $\ell_\infty(A - \sum_i \tilde{B}_i) \leq 1/3$. However, if each $\tilde{B}_i$ is an excellent approximation to B_i, then their sum will still be a good approximation to A. More precisely, it is still the case that

$$\mathrm{rk}_{1/3}(F_{\mathcal{G}}) \leq \sum_{i=0}^{2^c-1} \mathrm{rk}_\varepsilon(G_i),$$

where $\varepsilon \leq 2^{-c}/3$, since the definition of approximate rank allows error at most $1/3$. This yields

$$\mathrm{rk}(F_{\mathcal{G}}) \leq 2^c \max_i \mathrm{rk}_\varepsilon(G_i) \implies \log \mathrm{rk}_{1/3}(F_{\mathcal{G}}) \leq c + \max_i \log \mathrm{rk}_\varepsilon(G_i).$$

Since log of approximate rank lower bounds quantum communication complexity, we have that $\log \mathrm{rk}_\varepsilon(G_i) \leq Q^*_\varepsilon(G_i)$. By using standard error reduction, we have that $Q^*_\varepsilon(G_i)$ for $\varepsilon = 2^{-c}/3$ is at most $O(c\, Q^*(G_i))$. Hence $\log \mathrm{rk}_{1/3}(F_{\mathcal{G}}) = O(c \cdot \max_i Q^*(G_i))$. $\square$

To prove Theorem 6, we need a tool for taking a function F and finding a collection $\mathcal{G}$ such that $F_{\mathcal{G}}$ is an unambiguous lookup function, and $Q^*(G_i)$ is small for all $G_i \in \mathcal{G}$. The following lemma provides such a tool.

Lemma 10. *Let $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a total function with circuit size $\mathrm{size}(F)$ (i.e., F can be computed by a Boolean circuit with $\mathrm{size}(F)$ gates of constant fanin).*

Then for all $c > 0$, there exists a nontrivial family of XOR functions $\mathcal{G} = \{G_0, G_1, \ldots, G_{2^c-1}\}$, such that $F_{\mathcal{G}}$ is an unambiguous lookup function and for all $i \in \{0, \ldots, 2^c - 1\}$,

$$Q^*(G_i) = \widetilde{O}(\sqrt{c\, \mathrm{size}(F)}).$$

Proof. We need to construct functions $G_i(\mathbf{x}, u, \mathbf{y}, v)$ that lead to an unambiguous lookup function (Definition 8), that are a nontrivial XOR family (Definition 5) and have $Q^*(G_i) = \widetilde{O}(\sqrt{c\, \mathrm{size}(F)})$.

Each G_i will check that $u_i \oplus v_i$ has a very special type of certificate that proves that $F^c(x, y) = i$. If it contains such a certificate, G_i will output 1 and otherwise it will output 0. This takes care of the unambiguity condition. Since G_i only depends on $u_i \oplus v_i$, it will be an XOR family and since it only evaluates to 1 on a certificate, it will be nontrivial.

We now construct the certificate. Let $\mathrm{size}(F) = m$, which means that there is a circuit that takes in (x, y) as input and outputs $F(x, y)$ using at most m constant fanin gates. The cell $u_i \oplus v_i$ will contain c certificates, each certifying that the corresponding input to F evaluates to correct bit of i. For one instance of F, the certificate is constructed as follows. The certificate has to provide a full evaluation of the circuit of size m on (x, y) by providing the correct values for the inputs and outputs of all m gates. The final gate should, of course, evaluate the the claimed output value for F. The inputs to the first level, which are inputs belonging to either Alice or Bob, should be consistent with the true inputs that Alice and Bob hold. For a circuit of size m, a certificate of this sort has size $\widetilde{O}(m)$ (with a log factor to account for describing the labels of gates), and hence the entire certificate has size $\widetilde{O}(cm)$.

If the inputs are consistent with Alice's and Bob's input, and all the gates are evaluated correctly, then the output of the circuit will be $F(x, y)$ and the output string for all c circuits will indeed be $F^c(\mathbf{x}, \mathbf{y}) = \ell$. If this output string is consistent with i, then G_i accepts and otherwise rejects.

It is easy to see that $\mathcal{G}$ satisfies the first two properties we wanted. It remains to upper bound $\mathrm{Q}^*(G_i)$. As a warmup, note that the deterministic communication complexity of G_i is at most $\widetilde{O}(cm)$. This is because Alice and Bob can simply send all of u_i and v_i to each other, which costs $\widetilde{O}(cm)$ communication. They can then check that the their inputs are correct, the circuit evaluation is correct, and the circuits evaluate to i.

A similar algorithm, using Grover's algorithm to search for a discrepancy, yields the quantum algorithm. Alice and Bob first check that the $O(cm)$ inputs in the circuits (there are $O(m)$ inputs per F, and there are c copies of F) are consistent with their part of the input using $\widetilde{O}(\sqrt{cm})$ communication using Grover's algorithm. They can then Grover search over all cm gates to check if their inputs and outputs are consistent, which again takes $\widetilde{O}(\sqrt{cm})$ communication. The final step is to check that the output bits equal i. This takes $\widetilde{O}(\sqrt{c})$ communication using Grover search. Hence the total quantum communication complexity of G_i is $\widetilde{O}(\sqrt{cm}) = \widetilde{O}(\sqrt{c\,\mathrm{size}(F)})$. $\qquad\square$

Lemma 9 and Lemma 10 straightforwardly imply Theorem 6.

5 Lower bound on quantum communication complexity of lookup functions

In this section, we prove our main theorem, which is the following:

Theorem 11. *Let $F : \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ be a (partial) function, $c \geq 5\log(\mathrm{Q}^*_{1/3}(F))$ and $r \geq 1$ be an integer. Let $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$ be a nontrivial family of XOR functions where each $G_i : (\mathcal{X}^c \times \{0, 1\}^m) \times (\mathcal{Y}^c \times \{0, 1\}^m) \to \{0, 1\}$, and let $F_{\mathcal{G}}$ be the $(F, \mathcal{G})$-lookup function. Let $\delta = \frac{1}{10^9 cr^2}$. For any $1/3$-error r-round protocol Π for $F_{\mathcal{G}}$, there exists a $\frac{1}{2} - \frac{\delta}{3}$-error protocol Π' for F such that*

$$\mathrm{QCC}(\Pi') = O(\mathrm{QCC}(\Pi)).$$

Before proving this, we show how it implies the corollary used in Section 3, which we restate.

Corollary 7. *Let $F_{\mathcal{G}}$ be an $(F, \mathcal{G})$-lookup function for a function F and a nontrivial family of XOR functions $\mathcal{G} = \{G_0, G_1, \ldots, G_{2^c-1}\}$ with $c = \Theta(\log(\mathrm{Q}^*(F)))$. Then*

$$\mathrm{Q}^*(F_{\mathcal{G}}) = \Omega(\log(1/\mathrm{disc}(F))).$$

Proof. Let Π be a protocol for $F_{\mathcal{G}}$ with $\mathrm{QCC}(\Pi) = \mathrm{Q}^*(F)$. Then from Theorem 11, we have $\mathrm{Q}^*_\varepsilon(F) = O(\mathrm{Q}^*(F_{\mathcal{G}}))$, where $\varepsilon = \frac{1}{2} - \frac{\delta}{3}$, $\delta = \frac{1}{10^9 cr^2}$, and $r \leq \mathrm{QCC}(\Pi) = \mathrm{Q}^*(F_{\mathcal{G}}))$ is the number of rounds in Π. Now from Theorem 7, we know that $\mathrm{Q}^*_\varepsilon(F) = \Omega\big(\log\frac{1-2\varepsilon}{\mathrm{disc}(F)}\big)$. Combining these with the fact that $cr^2 = O(\mathrm{Q}^*(F_{\mathcal{G}}))$ we get

$$\mathrm{Q}^*(F_{\mathcal{G}}) = \Omega\left(\log \frac{1-2\varepsilon}{\mathrm{disc}(F)}\right) = \Omega\left(\log\left(\frac{1}{\mathrm{disc}(F)}\right) - \log(cr^2)\right) = \Omega\left(\log\left(\frac{1}{\mathrm{disc}(F)}\right) - \log \mathrm{Q}^*(F_{\mathcal{G}})\right),$$

which implies the statement to be proved. $\qquad\square$

Proof of Theorem 11. We explain here the overall structure of the argument which is also displayed visually in Figure 2.
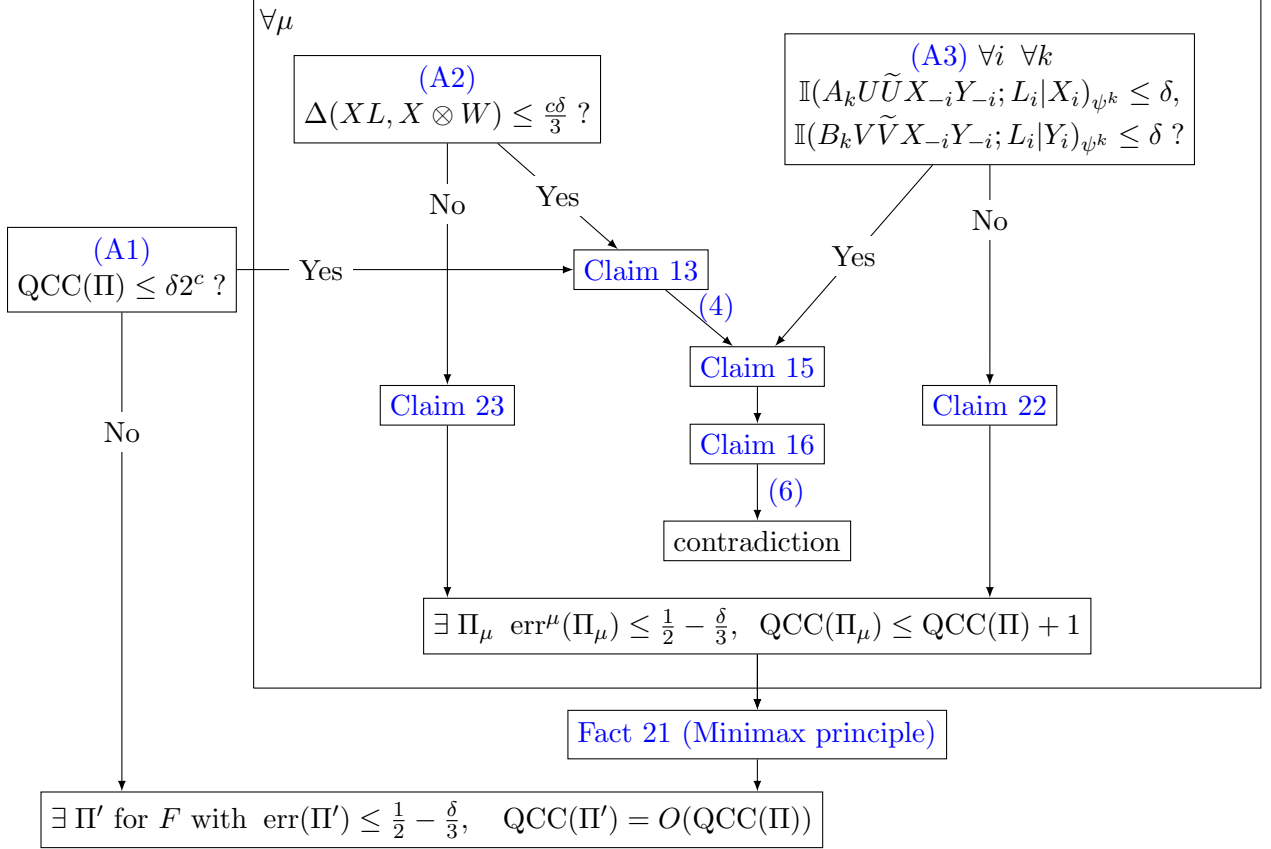
20

Figure 2: The structure of the proof of Theorem 11. Note that Claim 13 and Claim 15 only follow if both of their incoming arcs hold.

Rule out trivial protocols. We first rule out the easy case where the protocol we are given, Π, has high quantum communication cost. More precisely, we check if the following condition holds.

$$\mathrm{QCC}(\Pi) < \delta 2^c. \tag{A1}$$

If this does not hold then $\mathrm{QCC}(\Pi) \geq \delta 2^c = \Omega(\mathrm{Q}^*(F))$. By choosing the protocol whose communication complexity is $\mathrm{Q}^*(F)$, we obtain a protocol Π' for F with $\mathrm{QCC}(\Pi') = \mathrm{Q}^*(F) = O(\mathrm{QCC}(\Pi))$ and we are done. Hence for the rest of the proof we may assume (A1).

Protocols correct on a distribution. Instead of directly constructing a protocol Π' for F that is correct on all inputs with bounded error, we instead construct for every distribution μ on $\mathrm{dom}(F)$, a protocol Π_μ that does well on μ and then use Fact 21 to construct our final protocol. More precisely, for every μ over $\mathrm{dom}(F)$ we construct a protocol Π_μ for F that has the following properties:

$$\mathrm{QCC}(\Pi_\mu) = \mathrm{QCC}(\Pi) + 1 \qquad \text{and} \qquad \mathrm{err}^\mu(\Pi_\mu) < 1/2 - \delta/3. \tag{3}$$

Hence for the remainder of the proof let μ be any distribution over $\mathrm{dom}(F)$ and our aim is to construct a protocol satisfying (3).

Construct a distribution for $\mathbf{F_{\mathcal{G}}}$. Using the distribution μ on $\mathrm{dom}(F)$, we now construct a distribution over the inputs to $F_{\mathcal{G}}$. Let the random variable T be defined as follows:

$$T := (X_1, \ldots, X_c, U_0, \ldots, U_{2^c-1}, Y_1, \ldots, Y_c, V_0, \ldots, V_{2^c-1}),$$

21

where for all $i \in [c]$, $X_i Y_i$ is distributed according to μ and independent of all other random variables and for $j \in \{0, \ldots, 2^c - 1\}$, $U_j V_j$ are uniformly distributed in $\{0,1\}^{2m}$ and independent of all other variables. For $i \in [c]$, we define $L_i := F(X_i, Y_i)$. We also define $X := (X_1, \ldots, X_c)$, $Y := (Y_1, \ldots, Y_c)$, $L := (L_1, \ldots, L_c)$, $U := (U_0, \ldots, U_{2^c - 1})$ and $V := (V_1, \ldots, V_{2^c - 1})$. Lastly, for $i \in [c]$, we define $X_{-i} := X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_c$ and $X_{<i} := X_1, \ldots, X_{i-1}$. Similar definitions hold for L and Y. Let A_k, B_k be the registers of Alice and Bob after round k of protocol Π. The total pure state after round k can be written as follows:

$$|\psi_k\rangle_{X \widetilde{X} U \widetilde{U} Y \widetilde{Y} V \widetilde{V} A_k B_k} = \sum_{x,u,y,v} \sqrt{\mu_T(x,u,y,v)} \, |xx\rangle_{X \widetilde{X}} \, |uu\rangle_{U \widetilde{U}} \, |yy\rangle_{Y \widetilde{Y}} \, |vv\rangle_{V \widetilde{V}} \, \big|\psi_k^{x,u,y,v}\big\rangle_{A_k B_k}$$

Here μ_T is the distribution of the random variable T. $\widetilde{X}, \widetilde{U}, \widetilde{Y}, \widetilde{V}$ are registers that purify the classical inputs X, U, Y, V respectively.

Rule out easy distributions μ. We now show that if μ is such that the output of $F(X,Y)$ is predictable simply by looking at Alice's input X, then this distribution is easy and we can construct a protocol Π_μ that does well on this distribution since Alice can simply guess the value of $F(X,Y)$ after seeing X. More precisely, we check if the following condition holds.

$$\Delta(XL, X \otimes W) \leq c\delta/3, \tag{A2}$$

where W is the uniform distribution on $\{0,1\}^c$.

If the condition does not hold, we invoke Claim 23 with $\varepsilon = \delta/3$. Then we must be in case (a) of this claim and hence we get the desired protocol Π_μ. Therefore we can assume (A2) holds.

Construct new protocols Π_i. We now define a collection of protocols Π_i for each $i \in [c]$. Π_i is a protocol in which Alice and Bob receive inputs from $\mathrm{dom}(F)$. We construct Π_i as follows: Given the input pair (X_i, Y_i) distributed according to μ, Alice and Bob use shared entanglement $X_{-i} \widetilde{X_{-i}} Y_{-i} \widetilde{Y_{-i}}$ (Alice holds $X_{-i} \widetilde{X_{-i}}$ and Bob holds $Y_{-i} \widetilde{Y_{-i}}$), where $X_{-i} Y_{-i}$ are distributed according to $\mu^{\otimes c-1}$ and $\widetilde{X_{-i}} \widetilde{Y_{-i}}$ purify $X_{-i} Y_{-i}$ in a canonical way. They also use shared entanglement $U \widetilde{U} V \widetilde{V}$ (Alice holds $U \widetilde{U}$ and Bob holds $V \widetilde{V}$), where U and V are uniformly distributed and $\widetilde{U} \widetilde{V}$ purify UV in a canonical way. Note that Alice and Bob now have inputs XU and YV distributed according to T. They then run protocol Π. It is clear that for all $i \in [c]$, $\mathrm{QCC}(\Pi_i) = \mathrm{QCC}(\Pi)$.

Rule out informative protocols Π_i. If any of the protocols Π_i that we constructed has a lot of information about L_i, then we can use Claim 22 to design a protocol for F. Hence, we can assume that for each $1 \leq k \leq r$,

$$\mathbb{I}(A_k U \widetilde{U} X_{-i} Y_{-i}; L_i | X_i)_{\psi_k}, \mathbb{I}(B_k V \widetilde{V} X_{-i} Y_{-i}; L_i | Y_i)_{\psi_k} \leq \delta. \tag{A3}$$

Obtain a contradiction. We have already established that (A1), (A2), and (A3) must hold, otherwise we have obtained our protocol Π_μ. We will now show that if (A1), (A2), and (A3) simultaneously hold, then we obtain a contradiction. To show this, we use some claims that are proved after this theorem.

First we apply Claim 12 to get the following from (A1) and (A2).

$$\forall k \in \{1, \ldots, r\}: \quad \mathbb{E}_{x,l \leftarrow XL} \mathrm{B}^2\left(\psi_{k, B_k Y \widetilde{Y} V \widetilde{V} U_l}^x, \psi_{k, B_k Y \widetilde{Y} V \widetilde{V}}^x \otimes \psi_{U_l}\right) \leq \frac{q}{2^c} + \frac{c\delta}{3}. \tag{4}$$

Here $q = \mathrm{QCC}(\Pi)/2$. Intuitively this claim asserts that for a typical x and ℓ, Bob (conditioned on $X = x$) has very little information about the cell U_ℓ at the end of round k, which is quantified

by saying their joint state is close to being a product state. This would be false without assuming (A1) because if there was no upper bound on the communication in Π, then Alice could simply communicate all of U, in which case Bob would have a lot of information about any U_j. We need (A2) as well, since otherwise it is possible that the correct answer ℓ is easily predicted by Alice by looking at her input alone, in which case she can send over the contents of cell U_ℓ to Bob. A symmetric statement also follows with Alice and Bob interchanged.

We then apply Claim 13 to get the following from (A3).

$$\forall k \in \{1, \ldots, r\}: \quad \mathbb{E}_{x,l \leftarrow XL} \mathrm{B}^2 \left(\psi^{x,l}_{k,B_k Y \widetilde{Y} V \widetilde{V} U_l}, \psi^{x,l}_{k,B_k Y \widetilde{Y} V \widetilde{V}} \otimes \psi_{U_l} \right) \leq 3 \cdot \left(\frac{q}{2^c} + \frac{c\delta}{3} + 2c\delta \right). \quad (5)$$

Intuitively, this claim asserts that for a typical x and ℓ, Bob (conditioned on $X = x$ and $L = \ell$) has very little information about the cell U_ℓ at the end of round k, which is quantified by saying their joint state is close to being a product state. A symmetric statement also follows for Alice. Equation 5 implies the following relation, which is proved in Claim 14: $\Pr_{x,y,l,u_l,v_l \leftarrow X,Y,L,U_L,V_L}[G_l(x, y, u_l, v_l) = \alpha(x, y)] \leq 1/100$, where $\alpha(x, y)$ is either 0 or 1. We then proceed to apply Claim 15.

We then apply Claim 16, which uses (4) and (5) and Claim 15, to obtain the following. There exists, $x, y, l, \widetilde{u}_l, \widetilde{v}_l, \widetilde{\widetilde{u}}_l, \widetilde{\widetilde{v}}_l$ such that,

$$\Delta \left((\psi^{x,y,l,\widetilde{u}_l,\widetilde{v}_l}_{r,A_r U_{-l} \widetilde{U}_{-l}}, \psi^{x,y,l,\widetilde{u}_l,\widetilde{\widetilde{v}}_l}_{r,A_r U_{-l} \widetilde{U}_{-l}} \right) \leq 1000r \cdot \sqrt{\left(\frac{q}{2^c} + \frac{c\delta}{3} + 2c\delta \right)} < 0.1,$$
$$G_l(x, y, \widetilde{u}_l, \widetilde{v}_l) = 1 \text{ and } G_l(x, y, \widetilde{u}_l, \widetilde{\widetilde{v}}_l) = 0. \quad (6)$$

We assume (w.l.o.g) that Alice gives the answer in round r. From above

$$|\Pr(\text{Alice outputs 1 on } (x, y, \widetilde{u}_l, \widetilde{v}_l)) - \Pr(\text{Alice outputs 1 on } (x, y, \widetilde{u}_l, \widetilde{\widetilde{v}}_l))| < 0.1.$$

This is a contradiction since $G_l(x, y, \widetilde{u}_l, \widetilde{v}_l) = 1$ and $G_l(x, y, \widetilde{u}_l, \widetilde{\widetilde{v}}_l) = 0$ and the error of Π on any input is at most $1/3$.

Minimax argument. Note that in all branches where we did not reach a contradiction, we constructed a protocol satisfying (3). Hence we constructed, for any μ over $\mathrm{dom}(F)$, a protocol Π_μ that satisfies (3). We now use Fact 21 to complete the proof. $\qquad \square$

This completes the proof of the theorem, except for the claims Claim 12, Claim 13, Claim 14, Claim 15, and Claim 16 that we did not prove. We now prove these claims.

5.1 Proof of claims

Claim 12. *Suppose* $\mathrm{QCC}(\Pi) = 2q$ *and* $\Delta(XL, X \otimes W) \leq \delta_1$. *Then*

$$\mathbb{E}_{x,l \leftarrow XL} \mathrm{B}^2 \left(\psi^x_{k,B_k Y \widetilde{Y} V \widetilde{V} U_l}, \psi^x_{k,B_k Y \widetilde{Y} V \widetilde{V}} \otimes \psi_{U_l} \right) \leq \frac{q}{2^c} + \delta_1.$$

for all $1 \leq k \leq r$. *Here* ψ_{U_l} *is the maximally mixed state on the register* U_l *(in other words a random variable which is uniformly distributed.)*

Proof. We have

$$q \geq \mathbb{I}(B_k Y \widetilde{Y} V \widetilde{V}; U_0, \ldots, U_{2^c-1} | X)_{\psi_k} \qquad \text{(Lemma 25)}$$

$$\geq \sum_{l=0}^{2^c-1} \mathbb{I}(B_k Y \widetilde{Y} V \widetilde{V}; U_l | X)_{\psi_k} \qquad \text{(Fact 17.F)}$$

$$= 2^c \cdot \mathbb{E}_{x,l \leftarrow X \otimes W} \mathbb{I}(B_k Y \widetilde{Y} V \widetilde{V}; U_l | X = x)_{\psi_k}$$

$$\geq 2^c \cdot \mathbb{E}_{x,l \leftarrow X \otimes W} \mathrm{B}^2 \left(\psi_{k,B_k Y \widetilde{Y} V \widetilde{V} U_l}^x, \psi_{k,B_k Y \widetilde{Y} V \widetilde{V}}^x \otimes \psi_{U_l} \right) \qquad \text{(Fact 17.J)}.$$

This implies that

$$\mathbb{E}_{x,l \leftarrow X \otimes W} \mathrm{B}^2 \left(\psi_{k,B_k Y \widetilde{Y} V \widetilde{V} U_l}^x, \psi_{k,B_k Y \widetilde{Y} V \widetilde{V}}^x \otimes \psi_{U_l} \right) \leq \frac{q}{2^c}.$$

Since $\Delta(XL, X \otimes W) \leq \delta_1$ and $\mathrm{B}^2(\rho, \sigma) \leq 1$ always, this proves the claim as well. $\qquad \square$

The next claim intuitively says that, if the communication cost of Π is small, then at any point during the protocol, Bob's register has small information about the correct cheat sheet cell.

Claim 13. *Assume in addition to the assumptions of Claim 12, the following condition holds: for all $i \in [c]$, let*

$$\mathbb{I}(A_k U \widetilde{U} X_{-i} Y_{-i}; L_i | X_i)_{\psi_k} \leq \delta.$$

Then

$$\mathbb{E}_{x,l \leftarrow XL} \mathrm{B}^2 \left(\psi_{k,B_k Y \widetilde{Y} V \widetilde{V} U_l}^{x,l}, \psi_{k,B_k Y \widetilde{Y} V \widetilde{V}}^{x,l} \otimes \psi_{U_l} \right) \leq 3 \cdot \left(\frac{q}{2^c} + \delta_1 + 2c\delta \right)$$

for all $1 \leq k \leq r$.

Proof. We first prove that the register A_k carries low information about L i.e.

$$\mathbb{I}(A_k U \widetilde{U}; L | X)_{\psi^k} \leq c\delta.$$

This follows from the following chain of inequalities:

$$\begin{aligned}
\delta &\geq \mathbb{I}(A_k U \widetilde{U} X_{-i} Y_{-i}; L_i | X_i)_{\psi_k} \\
&\geq \mathbb{I}(A_k U \widetilde{U} X_{-i} L_{<i}; L_i | X_i)_{\psi_k} \qquad &\text{(Fact 17.D and Fact 17.J)} \\
&\geq \mathbb{I}(A_k U \widetilde{U}; L_i | L_{<i}, X)_{\psi_k} \qquad &\text{(Fact 17.E)}.
\end{aligned}$$

By summing the inequality over i, we get

$$\begin{aligned}
c\delta &\geq \sum_{i=1}^{c} \mathbb{I}(A_k U \widetilde{U}; L_i | L_{<i}, X)_{\psi_k} \\
&= \mathbb{I}(A_k U \widetilde{U}; L | X)_{\psi_k} \qquad &\text{(Fact 17.C)}.
\end{aligned}$$

This implies using Fact 17.J:

$$\mathbb{E}_{x,l \leftarrow XL} \mathrm{B}^2 \left(\psi_{k,A_k U \widetilde{U}}^{x,l}, \psi_{k,A_k U \widetilde{U}}^x \right) \leq c\delta. \qquad (7)$$

Now consider the following two pure states (one conditioned on x, l and the other conditioned on x):

$$\left| \psi^{x,l} \right\rangle_{k, Y \widetilde{Y} V \widetilde{V} U \widetilde{U} A_k B_k} = \sum_{y,v,u} \sqrt{\mu_T(y, v, u | X = x, L = l)} \, |uu\rangle_{U \widetilde{U}} \, |yy\rangle_{Y \widetilde{Y}} \, |vv\rangle_{V \widetilde{V}} \, |\psi^{x,u,y,v}\rangle_{k, A_k B_k}$$

and

$$|\psi^x\rangle_{k,Y\widetilde{Y}V\widetilde{V}U\widetilde{U}A_kB_k} = \sum_{y,v,u} \sqrt{\mu_T(y,v,u|X=x)}\, |uu\rangle_{U\widetilde{U}}\, |yy\rangle_{Y\widetilde{Y}}\, |vv\rangle_{V\widetilde{V}}\, |\psi^{x,u,y,v}\rangle_{k,A_kB_k}.$$

The marginals of these states on the systems $A_kU\widetilde{U}$ are close as shown above. Now by Uhlmann's theorem (Fact 12), there exists a unitary acting on the systems $B_kY\widetilde{Y}V\widetilde{V}$ (and the unitary depends on x,l) $\mathcal{U}^{x,l}_{B_kY\widetilde{Y}V\widetilde{V}}$ s.t.

$$\mathrm{B}^2\left(\mathbb{1}_{A_kU\widetilde{U}} \otimes \mathcal{U}^{x,l}_{B_kY\widetilde{Y}V\widetilde{V}}\left|\psi^{x,l}\right\rangle_{k,A_kU\widetilde{U}B_kY\widetilde{Y}V\widetilde{V}}, |\psi^x\rangle_{k,A_kU\widetilde{U}B_kY\widetilde{Y}V\widetilde{V}}\right) = \mathrm{B}^2\left(\psi^{x,l}_{k,A_kU\widetilde{U}}, \psi^x_{k,A_kU\widetilde{U}}\right).$$

The unitary $\mathcal{U}^{x,l}_{B_kY\widetilde{Y}V\widetilde{V}}$ should be intuitively thought of as implementing the operation of "forgetting L". Hence Equation (7) gives us that:

$$\mathbb{E}_{x,l\leftarrow XL}\mathrm{B}^2\left(\mathbb{1}_{A_kU\widetilde{U}} \otimes \mathcal{U}^{x,l}_{B_kY\widetilde{Y}V\widetilde{V}}\left|\psi^{x,l}\right\rangle_{k,A_kU\widetilde{U}B_kY\widetilde{Y}V\widetilde{V}}, |\psi^x\rangle_{k,A_kU\widetilde{U}B_kY\widetilde{Y}V\widetilde{V}}\right) \le c\delta. \tag{8}$$

For all (x,ℓ), define,

$$\phi^{x,\ell} = \mathbb{1}_{A_kU\widetilde{U}} \otimes \mathcal{U}^{x,l}_{B_kY\widetilde{Y}V\widetilde{V}}\left|\psi^{x,l}\right\rangle_{k,A_kU\widetilde{U}B_kY\widetilde{Y}V\widetilde{V}}.$$

Combining Equation (8) with the monotonicity of Bures metric (Fact 14), we obtain the following:

$$\mathbb{E}_{x,l\leftarrow XL}\mathrm{B}^2\left(\phi^{x,l}_{k,B_kY\widetilde{Y}V\widetilde{V}U_l}, \psi^x_{k,B_kY\widetilde{Y}V\widetilde{V}U_l}\right) \le c\delta \tag{9}$$

and

$$\mathbb{E}_{x,l\leftarrow XL}\mathrm{B}^2\left(\phi^{x,l}_{k,B_kY\widetilde{Y}V\widetilde{V}}, \psi^x_{k,B_kY\widetilde{Y}V\widetilde{V}}\right) \le c\delta. \tag{10}$$

Furthermore, combining Equation (10) with Fact 13.B, we obtain:

$$\mathbb{E}_{x,l\leftarrow XL}\mathrm{B}^2\left(\phi^{x,l}_{k,B_kY\widetilde{Y}V\widetilde{V}} \otimes \psi_{U_l}, \psi^x_{k,B_kY\widetilde{Y}V\widetilde{V}} \otimes \psi_{U_l}\right) \le c\delta. \tag{11}$$

Claim 12 gives us that:

$$\mathbb{E}_{x,l\leftarrow XL}\mathrm{B}^2\left(\psi^x_{k,B_kY\widetilde{Y}V\widetilde{V}U_l}, \psi^x_{k,B_kY\widetilde{Y}V\widetilde{V}} \otimes \psi_{U_l}\right) \le \frac{q}{2^c} + \delta_1. \tag{12}$$

Now combining Equations (9), (11) and (12) along with weak triangle inequality for square of Bures metric (Fact 13.A) and Fact 14, we obtain:

$$\begin{aligned}
&\mathbb{E}_{x,l\leftarrow XL}\mathrm{B}^2\left(\psi^{x,l}_{k,B_kY\widetilde{Y}V\widetilde{V}U_l}, \psi^{x,l}_{k,B_kY\widetilde{Y}V\widetilde{V}} \otimes \psi_{U_l}\right) \\
&= \mathbb{E}_{x,l\leftarrow XL}\mathrm{B}^2\left(\phi^{x,l}_{k,B_kY\widetilde{Y}V\widetilde{V}U_l}, \phi^{x,l}_{k,B_kY\widetilde{Y}V\widetilde{V}} \otimes \psi_{U_l}\right) \\
&\le 3\cdot\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right). \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square
\end{aligned}$$

Claim 14. *Assuming the conclusion from Claim 13, it holds that*

$$\Pr_{x,y,l,u_l,v_l\leftarrow X,Y,L,U_L,V_L}[G_l(x,y,u_l,v_l) = \alpha(x,y)] \le 1/100,$$

where $\alpha(x,y)$ is either 0 or 1 for every x,y.

25

Proof. Using monotonicity and partial measurement ([Fact 17.D](#) and [Fact 17.B](#)), we have that:

$$\mathbb{E}_{x,y,l,u_l,v_l \leftarrow XYLU_LV_L} \mathrm{B}^2\left(\psi_{r,B_r}^{x,y,l,u_l,v_l}, \psi_{r,B_r}^{x,y,l,v_l}\right) \leq 3 \cdot \left(\frac{q}{2^c} + \frac{c\delta}{3} + 2c\delta\right)$$

Let the output register be called O. Then, from our choice of parameters and monotonicity ([Fact 17.D](#)), above inequality implies

$$\mathbb{E}_{x,y,l,u_l,v_l \leftarrow XYLU_LV_L} \mathrm{B}^2\left(\psi_{r,O}^{x,y,l,u_l,v_l}, \psi_{r,O}^{x,y,l,v_l}\right) \leq 1/400 \tag{13}$$

Since protocol makes an error of at most $1/400$ (which can be assumed due to [Fact 20](#)), we have that

$$\mathbb{E}_{x,y,l,u_l,v_l \leftarrow XYLU_LV_L} \mathrm{B}^2(\psi_{r,O}^{x,y,l,u_l,v_l}, |G_l(x,y,u_l,v_l)\rangle\langle G_l(x,y,u_l,v_l)|) \leq 1/400. \tag{14}$$

On the other hand, since the look-up function is an XOR family, we find that for a fixed x,y (and hence a fixed l),

$$\mathbb{E}_{u_l \leftarrow U_L} |G_l(x,y,u_l,v_l)\rangle\langle G_l(x,y,u_l,v_l)| = \mathrm{Pr}_{u_l,v_l \leftarrow U_l,V_l|x,y,l}[G_l(x,y,u_l,v_l)=0] |0\rangle\langle 0|$$
$$+ \mathrm{Pr}_{u_l,v_l \leftarrow U_l,V_l|x,y,l}[G_l(x,y,u_l,v_l)=1] |1\rangle\langle 1|.$$

Define $p_{x,y,l}^0 = \mathrm{Pr}_{u_l,v_l \leftarrow U_l,V_l|x,y,l}[G_l(x,y,u_l,v_l)=0]$ and $p_{x,y,l}^1 = \mathrm{Pr}_{u_l,v_l \leftarrow U_l,V_l|x,y,l}[G_l(x,y,u_l,v_l)=1]$. Then above equation, along with Equation ([14](#)) implies that

$$\mathbb{E}_{x,y,l,u_l,v_l \leftarrow XYLU_lV_l} \mathrm{B}^2(\psi_{r,O}^{x,y,l,v_l}, p_{x,y,l}^0 |0\rangle\langle 0| + p_{x,y,l}^1 |1\rangle\langle 1|) \leq 1/400$$

which in conjunction with Equation [13](#) and triangle inequality gives us

$$\mathbb{E}_{x,y,l,u_l,v_l \leftarrow XYLU_lV_l} \mathrm{B}^2(|G_l(x,y,u_l,v_l)\rangle\langle G_l(x,y,u_l,v_l)|, p_{x,y,l}^0 |0\rangle\langle 0| + p_{x,y,l}^1 |1\rangle\langle 1|) \leq 1/100. \tag{15}$$

This directly implies that we cannot have both $p_{x,y,l}^0, p_{x,y,l}^1$ large. More formally, for every x,y, let $\alpha(x,y)$ be such that $p_{x,y,l}^{\alpha(x,y)} < p_{x,y,l}^{1-\alpha(x,y)}$. Then it is clear that

$$\mathrm{B}^2(|G_l(x,y,u_l,v_l)\rangle\langle G_l(x,y,u_l,v_l)|, p_{x,y,l}^0 |0\rangle\langle 0| + p_{x,y,l}^1 |1\rangle\langle 1|) > p_{x,y,l}^{\alpha(x,y)},$$

which in turn implies (when used in Equation [15](#)),

$$\mathbb{E}_{x,y,l \leftarrow XYL} p_{x,y,l}^{\alpha(x,y)} = \mathbb{E}_{x,y,l,u_l,v_l \leftarrow XYLU_lV_l} p_{x,y,l}^{\alpha(x,y)} \leq 1/100.$$

Recalling the definition of $p_{x,y,l}^{\alpha(x,y)}$, this immediately gives us

$$\mathbb{E}_{x,y,l \leftarrow XYL} \mathrm{Pr}_{u_l,v_l \leftarrow U_l,V_l|x,y,l}[G_l(x,y,u_l,v_l) = \alpha(x,y)] \leq 1/100.$$

This completes the proof. $\qquad\square$

Claim 15. *Assume that the assumptions of [Claim 12](#) and [Claim 13](#) hold. In addition,*

$$\mathbb{I}(B_k V \widetilde{V} X_{-i} Y_{-i}; L_i|Y_i)_{\psi_k} \leq \delta$$

and

$$\mathrm{Pr}_{x,y,l,u_l,v_l \leftarrow X,Y,L,U_L,V_L}[G_l(x,y,u_l,v_l) = \alpha(x,y)] \leq 1/100$$

also hold for $\alpha(x,y) \in \{0,1\}$ for every x,y. Then there exist $x,y,l = l(x,y), \widetilde{u}_l, \widetilde{v}_l, \widetilde{\widetilde{u}}_l, \widetilde{\widetilde{v}}_l$ s.t. the following conditions hold:

1. $G_l(x, y, \widetilde{u}_l, \widetilde{v}_l) = \alpha(x, y)$.

2. $G_l(x, y, \widetilde{u}_l, \widetilde{\widetilde{v}}_l) = G_l(x, y, \widetilde{\widetilde{u}}_l, \widetilde{\widetilde{v}}_l) = G_l(x, y, \widetilde{\widetilde{u}}_l, \widetilde{v}_l) = 1 - \alpha(x, y)$.

3. $\sum_{k=1}^{r} \mathrm{B}\left(\psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,u,v}, \psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,v}\right) \leq 80r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}$,
   for any choice of $(u, v) = (\widetilde{u}_l, \widetilde{v}_l), (\widetilde{\widetilde{u}}_l, \widetilde{\widetilde{v}}_l), (\widetilde{\widetilde{u}}_l, \widetilde{v}_l)$.

4. $\sum_{k=1}^{r} \mathrm{B}\left(\psi_{k, A_k U_{-l}\widetilde{U}_{-l}}^{x,y,u,v}, \psi_{k, A_k U_{-l}\widetilde{U}_{-l}}^{x,y,u}\right) \leq 80r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}$,
   for any choice of $(u, v) = (\widetilde{u}_l, \widetilde{v}_l), (\widetilde{\widetilde{u}}_l, \widetilde{v}_l), (\widetilde{\widetilde{u}}_l, \widetilde{v}_l)$.

Proof. By Claim 13, we have that for all $1 \leq k \leq r$,

$$\mathbb{E}_{x,l\leftarrow XL}\mathrm{B}^2\left(\psi_{k, B_k Y\widetilde{Y}V\widetilde{V}U_l}^{x,l}, \psi_{k, B_k Y\widetilde{Y}V\widetilde{V}}^{x,l} \otimes \psi_{U_l}\right) \leq 3 \cdot \left(\frac{q}{2^c} + \delta_1 + 2c\delta\right).$$

By monotonicity of Bures metric (Fact 14), we get that

$$\mathbb{E}_{x,l\leftarrow XL}\mathrm{B}^2\left(\psi_{k, B_k YV_{-l}\widetilde{V}_{-l}U_l V_l}^{x,l}, \psi_{k, B_k YV_{-l}\widetilde{V}_{-l}V_l}^{x,l} \otimes \psi^{U_l}\right) \leq 3 \cdot \left(\frac{q}{2^c} + \delta_1 + 2c\delta\right).$$

Note that in both the states above, the marginal state on registers $U_l V_l$ is maximally mixed. Then by the partial measurement property of the square of Bures metric, Fact 13.C, we get that

$$\mathbb{E}_{x,y,l,u_l,v_l\leftarrow XYLU_LV_L}\mathrm{B}^2\left(\psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,u_l,v_l}, \psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,v_l}\right) \leq 3 \cdot \left(\frac{q}{2^c} + \delta_1 + 2c\delta\right).$$

Convexity of square gives us that

$$\mathbb{E}_{x,y,l,u_l,v_l\leftarrow XYLU_LV_L}\mathrm{B}\left(\psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,u_l,v_l}, \psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,v_l}\right) \leq \sqrt{3} \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}. \tag{16}$$

Similarly we get that for all $1 \leq k \leq r$,

$$\mathbb{E}_{x,y,l,u_l,v_l\leftarrow XYLU_LV_L}\mathrm{B}\left(\psi_{k, A_k U_{-l}\widetilde{U}_{-l}}^{x,y,l,u_l,v_l}, \psi_{k, A_k U_{-l}\widetilde{U}_{-l}}^{x,y,l,u_l}\right) \leq \sqrt{3} \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}. \tag{17}$$

Summing Equations (16) and (17) over k, we get the following:

$$\mathbb{E}_{x,y,l,u_l,v_l\leftarrow XYLU_LV_L}\sum_{k=1}^{r} \mathrm{B}\left(\psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,u_l,v_l}, \psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,v_l}\right) \leq 2r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}.$$

and

$$\mathbb{E}_{x,y,l,u_l,v_l\leftarrow XYLU_LV_L}\sum_{k=1}^{r} \mathrm{B}\left(\psi_{k, A_k U_{-l}\widetilde{U}_{-l}}^{x,y,l,u_l,v_l}, \psi_{k, A_k U_{-l}\widetilde{U}_{-l}}^{x,y,l,u_l}\right) \leq 2r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}.$$

Now by Markov's inequality, we can find $x, y, l = l(x, y)$ s.t. the following hold:

$$\Pr_{u_l,v_l\leftarrow U_l,V_l}[G_l(x, y, u_l, v_l) = \alpha(x, y)] \leq 1/25, \tag{18}$$

$$\mathbb{E}_{u_l,v_l\leftarrow U_lV_l}\sum_{k=1}^{r} \mathrm{B}\left(\psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,u_l,v_l}, \psi_{k, B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,v_l}\right) \leq 8r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}, \tag{19}$$

$$\mathbb{E}_{u_l,v_l \leftarrow U_l V_l} \sum_{k=1}^{r} \mathrm{B}\left(\psi_{k,A_k U_{-l}\widetilde{U}_{-l}}^{x,y,l,u_l,v_l}, \psi_{k,A_k U_{-l}\widetilde{U}_{-l}}^{x,y,l,u_l}\right) \leq 8r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}. \tag{20}$$

Without loss of generality, assume that $\alpha(x,y) = 1$. Let us have the following two notations:

$$\kappa_A(u_l, v_l) := \sum_{k=1}^{r} \mathrm{B}\left(\psi_{k,A_k U_{-l}\widetilde{U}_{-l}}^{x,y,l,u_l,v_l}, \psi_{k,A_k U_{-l}\widetilde{U}_{-l}}^{x,y,l,u_l}\right),$$

$$\kappa_B(u_l, v_l) := \sum_{k=1}^{r} \mathrm{B}\left(\psi_{k,B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,u_l,v_l}, \psi_{k,B_k V_{-l}\widetilde{V}_{-l}}^{x,y,l,v_l}\right).$$

Recall that for $l = l(x,y)$, $G_l(x,y,u_l,v_l)$ is a non-trivial XOR function of the inputs u_l, v_l. So there exists a $t \in \{0,1\}^m$ s.t. $G_l(x,y,u,u \oplus t) = 1$ for all $u \in \{0,1\}^m$. Now we will choose $\widetilde{u}_l, \widetilde{\widetilde{u}}_l, \widetilde{\widetilde{v}}_l$ uniformly and independently from $\{0,1\}^m$ and set $\widetilde{v}_l = \widetilde{u}_l \oplus t$. Note that marginally, the distribution of (u,v) is uniform over $\{0,1\}^m \times \{0,1\}^m$, for any choice of $(u,v) = (\widetilde{u}_l, \widetilde{v}_l), (\widetilde{\widetilde{u}}_l, \widetilde{v}_l), (\widetilde{u}_l, \widetilde{v}_l)$. Hence for any choice of $(u,v) = (\widetilde{u}_l, \widetilde{v}_l), (\widetilde{\widetilde{u}}_l, \widetilde{v}_l), (\widetilde{u}_l, \widetilde{v}_l)$, from Equations (18), (19) and (20), we get the following:

$$\mathrm{Pr}_{\widetilde{u}_l, \widetilde{\widetilde{u}}_l, \widetilde{\widetilde{v}}_l}[G_l(x,y,u,v) = 1] \leq 1/25,$$

$$\mathbb{E}_{\widetilde{u}_l, \widetilde{\widetilde{u}}_l, \widetilde{\widetilde{v}}_l} \kappa_A(u,v) \leq 8r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)},$$

$$\mathbb{E}_{\widetilde{u}_l, \widetilde{\widetilde{u}}_l, \widetilde{\widetilde{v}}_l} \kappa_B(u,v) \leq 8r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}.$$

Now by a simple application of Markov's inequality, there exists a setting of $(\widetilde{u}_l, \widetilde{\widetilde{u}}_l, \widetilde{\widetilde{v}}_l)$ so that for any choice of $(u,v) = (\widetilde{u}_l, \widetilde{v}_l), (\widetilde{\widetilde{u}}_l, \widetilde{v}_l), (\widetilde{u}_l, \widetilde{v}_l)$,

$$G_l(x,y,u,v) = 0,$$

$$\kappa_A(u,v) \leq 80r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)},$$

$$\kappa_B(u,v) \leq 80r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}.$$

This completes the proof. Note that we chose $\widetilde{v}_l$ so that $G_l(x,y,\widetilde{u}_l,\widetilde{v}_l) = 1$. $\qquad\square$

The next claim will follow from the quantum-cut-and-paste lemma applied to Claim 13.

Claim 16. *Assume that the assumptions of Claim 12, Claim 13 and Claim 15 hold. Then for the $x,y,l,\widetilde{u}_l,\widetilde{v}_l,\widetilde{\widetilde{u}}_l,\widetilde{\widetilde{v}}_l$ in Claim 15, it holds that*

$$\Delta\left((\psi_{r,A_r U_{-l}\widetilde{U}_{-l}}^{x,y,l,\widetilde{u}_l,\widetilde{v}_l}, \psi_{r,A_r U_{-l}\widetilde{U}_{-l}}^{x,y,l,\widetilde{u}_l,\widetilde{v}_l}\right) \leq 1000r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}.$$

Proof. Let us define the following registers: $\widetilde{A}_k := A_k U_{-l}\widetilde{U}_{-l}$ and $\widetilde{B}_k := B_k V_{-l}\widetilde{V}_{-l}$. Also we will define the following:

$$\delta_{k,A} := \mathrm{B}\left(\psi_{k,\widetilde{A}_k}^{x,y,\widetilde{\widetilde{u}}_l,\widetilde{v}_l}, \psi_{k,\widetilde{A}_k}^{x,y,\widetilde{u}_l,\widetilde{v}_l}\right),$$

$$\delta_{k,B} := \mathrm{B}\left(\psi_{k,\widetilde{B}_k}^{x,y,\widetilde{u}_l,\widetilde{v}_l}, \psi_{k,\widetilde{B}_k}^{x,y,\widetilde{u}_l,\widetilde{v}_l}\right).$$

By the triangle inequality for Bures metric Fact 13.A,

$$\delta_{k,A} \le \mathrm{B}\left(\psi_{k,\widetilde{A}_k}^{x,y,\widetilde{\widetilde{u}}_l,\widetilde{v}_l}, \psi_{k,\widetilde{A}_k}^{x,y,\widetilde{u}_l}\right) + \mathrm{B}\left(\psi_{k,\widetilde{A}_k}^{x,y,\widetilde{\widetilde{u}}_l,\widetilde{\widetilde{v}}_l}, \psi_{k,\widetilde{A}_k}^{x,y,\widetilde{u}_l}\right), \tag{21}$$

$$\delta_{k,B} \le \mathrm{B}\left(\psi_{k,\widetilde{B}_k}^{x,y,\widetilde{u}_l,\widetilde{v}_l}, \psi_{k,\widetilde{B}_k}^{x,y,\widetilde{\widetilde{v}}_l}\right) + \mathrm{B}\left(\psi_{k,\widetilde{A}_k}^{x,y,\widetilde{\widetilde{u}}_l,\widetilde{\widetilde{v}}_l}, \psi_{k,\widetilde{A}_k}^{x,y,\widetilde{\widetilde{v}}_l}\right). \tag{22}$$

Combining Equations (21), (22) and Claim 15, we get the following:

$$\sum_{k=1}^{r} \delta_{k,A} \le 160r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)},$$

$$\sum_{k=1}^{r} \delta_{k,B} \le 160r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}.$$

Note that the state $\psi_{k,\widetilde{A}_k,\widetilde{B}_k}^{x,y,u,v}$ is a pure state for every k,x,y,u,v. Also for a fixed x,y, these states can be formed by a quantum protocol Π' where Alice gets the input u and Bob gets the input v (since they are originally formed by running the protocol Π and $U_{-l}\widetilde{U}_{-l}$ and $V_{-l}\widetilde{V}_{-l}$ are registers that can be owned by Alice and Bob respectively at the start of Π'). Hence we can apply Lemma 24 (by setting $u = \widetilde{\widetilde{u}}_l$, $u' = \widetilde{u}_l$, $v = \widetilde{\widetilde{v}}_l$, $v' = \widetilde{v}_l$) to conclude that

$$\mathrm{B}\left(\psi_{r,\widetilde{A}_r}^{x,y,l,\widetilde{u}_l,\widetilde{v}_l}, \psi_{r,\widetilde{A}_r}^{x,y,l,\widetilde{u}_l,\widetilde{v}_l}\right) \le 2\sum_{k=1}^{r} \left(\delta_{k,A} + \delta_{k,B}\right)$$

$$\le 640r \cdot \sqrt{\left(\frac{q}{2^c} + \delta_1 + 2c\delta\right)}.$$

Now the proof is finished by Fact 11 and monotonicity of trace distance (Fact 14). $\square$

6   Conclusion and open problems

We prove a nearly quadratic separation between the log of approximate rank and quantum communication complexity for a family of total functions, which is also the first superlinear separation between these two measures. Our separation is based on a lookup function constructed from the inner product function. To prove the lower bound on the quantum communication complexity of this lookup function, we prove a general purpose cheat sheet theorem for quantum communication complexity. We also prove a general theorem about an upper bound on log of approximate rank of lookup functions based on the circuit size of the base function. This proves the upper bound for an appropriate lookup function on inner product because the inner product function has a linear size circuit.

Several interesting open problems arise out of our work. We state some of them here:

1. Can we eliminate the round dependence in Theorem 11? Can we prove a similar result for quantum information complexity instead of quantum communication complexity, thereby separating quantum information complexity from log of approximate rank?

2. Can we separate the quantum partition bound [LLR12] from quantum communication complexity? Is the quantum partition bound a stronger lower bound measure than log of approximate rank?

3. Can we prove some sort of cheat sheet theorem for log of approximate rank? A simpler question might be to prove that for the inner product function on n bits, any lookup function contructed using a nontrivial XOR family of functions has log of approximate rank at least $\Omega(\sqrt{n})$.

Acknowledgements

References

[ABB+16a] Andris Ambainis, Kaspars Balodis, Aleksandrs Belovs, Troy Lee, Miklos Santha, and Juris Smotrovs. Separations in query complexity based on pointer functions. In *Proceedings of the 48th ACM Symposium on Theory of Computing (STOC 2016)*, 2016. doi:10.1145/2897518.2897524. [p. 2]

[ABB+16b] Anurag Anshu, Aleksandrs Belovs, Shalev Ben-David, Mika Göös, Rahul Jain, Robin Kothari, Troy Lee, and Miklos Santha. Separations in communication complexity using cheat sheets and information complexity. *Proceedings of the 57h IEEE Symposium on Foundations of Computer Science (FOCS 2016)*, 2016. arXiv preprint arXiv:1605.01142. [pp. 2, 3, 4, 5, 16, 30]

[ABK16] Scott Aaronson, Shalev Ben-David, and Robin Kothari. Separations in query complexity using cheat sheets. In *Proceedings of the 48th ACM Symposium on Theory of Computing (STOC 2016)*, pages 863–876, 2016. doi:10.1145/2897518.2897644. [pp. 2, 3, 4]

[AKK16] Andris Ambainis, Martins Kokainis, and Robin Kothari. Nearly Optimal Separations Between Communication (or Query) Complexity and Partitions. In *31st Conference on Computational Complexity (CCC 2016)*, volume 50 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 4:1–4:14, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2016.4. [p. 3]

[Amb03] Andris Ambainis. Polynomial degree vs. quantum query complexity. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS 2003)*, pages 230–239, 2003. doi:10.1109/SFCS.2003.1238197. [p. 2]

[Aud14] Koenraad M. R. Audenaert. Quantum skew divergence. *Journal of Mathematical Physics*, 55(11), 2014. doi:10.1063/1.4901039. [p. 11]

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013. doi:10.1137/100811969. [p. 3]

[BCF+96] Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schmacher. Noncommuting mixed states cannot be broadcast. *Phys. Rev. Lett.*, 76(15):2818–2821, 1996. doi:10.1103/PhysRevLett.76.2818. [p. 8]

[BCW98] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, STOC '98, pages 63–68, 1998. doi:10.1145/276698.276713. [p. 4]

[BdW01] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th IEEE Conference on Computational Complexity*, pages 120–130. IEEE, 2001. doi:10.1109/CCC.2001.933879. [p. 1]

[BGK+15] Mark Braverman, Ankit Garg, Young Kun Ko, Jieming Mao, and Dave Touchette. Near-optimal bounds on bounded-round quantum communication complexity of disjointness. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 773–791, Oct 2015. doi:10.1109/FOCS.2015.53. [p. 2]

[BRWY13] M. Braverman, A. Rao, O. Weinstein, and A. Yehudayoff. Direct products in communication complexity. In *54th Annual Symposium on Foundations of Computer Science (FOCS 2013)*, pages 746–755, Oct 2013. doi:10.1109/FOCS.2013.85. [p. 3]

[Bur69] Donald Bures. An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite ω^*-algebras. *Transactions of the American Mathematical Society*, 135:199–212, 1969. doi:10.2307/1995012. [p. 8]

[BW15] Mark Braverman and Omri Weinstein. An interactive information odometer and applications. In *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 341–350, 2015. doi:10.1145/2746539.2746548. [p. 3]

[CvNT13] Richard Cleve, Wim van Dam, Michael Nielsen, and Alain Tapp. Quantum entanglement and the communication complexity of the inner product function. *Theoretical Computer Science*, 486:11–19, 2013. doi:10.1016/j.tcs.2012.12.012. [p. 15]

[Dru12] Andrew Drucker. *The Complexity of Joint Computation*. PhD thesis, Massachusetts Institute of Technology, 2012. [p. 3]

[For02] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *Journal of Computer and System Sciences*, 65(4):612 – 625, 2002. Special Issue on Complexity 2001. doi:10.1016/S0022-0000(02)00019-3. [p. 17]

[FvdG99] Christopher A. Fuchs and Jeroen van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, May 1999. doi:10.1109/18.761271. [p. 7]

[GJPW15] Mika Göös, T.S. Jayram, Toniann Pitassi, and Thomas Watson. Randomized communication vs. partition number. *Electronic Colloquium on Computational Complexity (ECCC) TR15-169*, 2015. [p. 2]

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of the 56th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1077–1088, 2015. doi:10.1109/FOCS.2015.70. [p. 2]

[Gro96]   Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th ACM Symposium on Theory of Computing (STOC)*, pages 212–219, 1996. doi:10.1145/237814.237866. [p. 4]

[JN06]   Rahul Jain and Ashwin Nayak. Accessible versus Holevo information for a binary random variable. 2006. arXiv:quant-ph/0603278. [p. 11]

[JRS03]   Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of set disjointness. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS 2003)*, pages 220–229, Oct 2003. doi:10.1109/SFCS.2003.1238196. [pp. 4, 14]

[Kla07]   Hartmut Klauck. Lower bounds for quantum communication complexity. *SIAM Journal on Computing*, 37(1):20–46, 2007. doi:10.1137/S0097539702405620. [p. 1]

[KMSY14]   Gillat Kol, Shay Moran, Amir Shpilka, and Amir Yehudayoff. Approximate nonnegative rank is equivalent to the smooth rectangle bound. In *Automata, Languages, and Programming: 41st International Colloquium (ICALP 2014)*, pages 701–712. Springer Berlin Heidelberg, 2014. doi:10.1007/978-3-662-43948-7_58. [p. 2]

[KN06]   Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 2006. URL: http://books.google.ca/books?id=dHH7rdhKwzsC. [pp. 5, 18]

[Kre95]   Ilan Kremer. Quantum communication. Master's thesis, The Hebrew University of Jerusalem, 1995. URL: www.cs.huji.ac.il/~noam/kremer-thesis.ps. [pp. 1, 5]

[Lin75]   Göran Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40(2):147–151, 1975. doi:10.1007/BF01609396. [p. 8]

[LLR12]   Sophie Laplante, Virginie Lerays, and Jérémie Roland. Classical and quantum partition bound and detector inefficiency. In *Proceedings of the 39th International Colloquium Conference on Automata, Languages, and Programming - Volume Part I*, ICALP'12, pages 617–628, Berlin, Heidelberg, 2012. Springer-Verlag. URL: http://dx.doi.org/10.1007/978-3-642-31594-7_52, doi:10.1007/978-3-642-31594-7_52. [p. 29]

[LLS06]   Sophie Laplante, Troy Lee, and Mario Szegedy. The quantum adversary method and classical formula size lower bounds. *Computational Complexity*, 15:163–196, 2006. doi:10.1007/s00037-006-0212-7. [p. 5]

[LR13]   Troy Lee and Jérémie Roland. A strong direct product theorem for quantum query complexity. *Computational Complexity*, 22(2):429–462, 2013. doi:10.1007/s00037-013-0066-8. [p. 3]

[LS08]   Troy Lee and Adi Shraibman. An approximation algorithm for approximation rank. In *Proceedings of the 24th IEEE Conference on Computational Complexity*, pages 351–357, 2008. doi:10.1109/CCC.2009.25. [pp. 1, 5]

[LS09]   Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures & Algorithms*, 34(3):368–394, 2009. doi:10.1002/rsa.20232. [pp. 1, 5]

[NC00]   Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge Series on Information and the Natural Sciences. Cambridge University Press, 2000. [pp. 6, 10]

[NT16]   Ashwin Nayak and Dave Touchette. Augmented index and quantum streaming algorithms for DYCK(2). *arXiv preprint arXiv:1610.04937*, 2016. [p. 14]

[Raz03]   Alexander Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145, 2003. doi:10.1070/IM2003v067n01ABEH000422. [p. 1]

[Rei11]   Ben W. Reichardt. Reflections for quantum query algorithms. In *Proceedings of the 22nd ACM-SIAM Symposium on Discrete Algorithms (SODA 2011)*, SODA '11, pages 560–569, 2011. URL: http://dl.acm.org/citation.cfm?id=2133036.2133080. [p. 5]

[She11]   Alexander A. Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011. doi:10.1137/080733644. [pp. 1, 2]

[She12]   Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM Journal on Computing*, 41(5):1122–1165, 2012. doi:10.1137/110842661. [pp. 1, 3]

[Sio58]   Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 1:171176, 1958. [p. 13]

[SZ09]   Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum information and computation*, 9(5,6):444–460, 2009. arXiv:0710.0095. [pp. 1, 2]

[Tom16]   Marco Tomamichel. *Quantum Information Processing with Finite Resources: Mathematical Foundations*. SpringerBriefs in Mathematical Physics. Springer, 2016. doi:10.1007/978-3-319-21891-5. [p. 10]

[Tou15]   Dave Touchette. Quantum information complexity. In *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 317–326. ACM, 2015. doi:10.1145/2746539.2746613. [pp. 2, 3]

[Uhl76]   A. Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9:273–279, 1976. doi:10.1016/0034-4877(76)90060-4. [p. 7]

[Wat16]   John Watrous. *Theory of Quantum Information*. Unpublished, January 2016. Available at https://cs.uwaterloo.ca/~watrous/TQI/. [pp. 6, 7, 9]

[Wil12]   Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, Cambridge, 12 2012. doi:10.1017/CBO9781139525343. [pp. 9, 10]

[Yao93]   Andrew Yao. Quantum circuit complexity. In *Proceedings of the 34th IEEE Symposium on Foundations of Computer Science (FOCS 1993)*, pages 352–360, 1993. doi:10.1109/SFCS.1993.366852. [p. 1]