Electronic Colloquium on Computational Complexity, Report No. 72 (2016)

Separations in communication complexity using cheat sheets and information complexity

Anurag Anshu* Aleksandrs Belovs† Shalev Ben-David‡ Mika Göös§
Rahul Jain¶ Robin Kothari‖ Troy Lee** Miklos Santha††

Abstract

While exponential separations are known between quantum and randomized communication complexity for partial functions (Raz, STOC 1999), the best known separation between these measures for a total function is quadratic, witnessed by the disjointness function. We give the first super-quadratic separation between quantum and randomized communication complexity for a total function, giving an example exhibiting a power 2.5 gap. We also present a nearly optimal quadratic separation between randomized communication complexity and the logarithm of the partition number, improving upon the previous best power 1.5 separation due to Göös, Jayram, Pitassi, and Watson.

Our results are the communication analogues of separations in query complexity proved using the recent cheat sheet framework of Aaronson, Ben-David, and Kothari (STOC 2016). Our main technical results are randomized communication and information complexity lower bounds for a family of functions, called lookup functions, that generalize and port the cheat sheet framework to communication complexity.

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

†CWI, Amsterdam, The Netherlands. `stiboh@gmail.com`

‡Massachusetts Institute of Technology. `shalev@mit.edu`

§University of Toronto. `mgoos@cs.toronto.edu`

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

‖Massachusetts Institute of Technology. `rkothari@mit.edu`

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

††IRIF, Université Paris Diderot, CNRS, 75205 Paris, France; and Centre for Quantum Technologies, National University of Singapore, Singapore. `miklos.santha@liafa.univ-paris-diderot.fr`

ISSN 1433-8092

1 Introduction

Understanding the power of different computational resources is one of the primary aims of complexity theory. Communication complexity provides an ideal setting to study these questions, as it is a nontrivial model for which we are still able to show interesting lower bounds. Moreover, lower bounds in communication complexity have applications to many other areas of complexity theory, for example yielding lower bounds for circuits, data structures, streaming algorithms, property testing and linear and semi-definite programs.

In communication complexity, two players Alice and Bob are given inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively, and their task is to compute a known function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ while minimizing the number of bits communicated between them. We call such a function a communication function. The players only need to be correct on inputs (x, y) for which $F(x, y) \in \{0, 1\}$. The function is called total if $F(x, y) \in \{0, 1\}$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, and otherwise is called partial.

A major question in communication complexity is what advantage players who exchange quantum messages can achieve over their classical counterparts. We will use $\mathrm{R}(F)$ and $\mathrm{Q}(F)$ to denote bounded-error (say 1/3) public-coin randomized and bounded-error quantum communication complexities of F, respectively. We also use $\mathrm{D}(F)$ for the deterministic communication complexity of F. Note the easy relationships $\mathrm{D}(F) \geq \mathrm{R}(F) \geq \mathrm{Q}(F)$.

There are examples of *partial* functions F for which $\mathrm{Q}(F)$ is exponentially smaller than $\mathrm{R}(F)$ [Raz99]. For total functions, however, it is an open question if $\mathrm{Q}(F)$ and $\mathrm{R}(F)$ are always polynomially related. On the other hand, the largest separation between these measures is quadratic, witnessed by the disjointness function which satisfies $\mathrm{R}(\textsc{Disj}_n) = \Omega(n)$ [KS92, Raz92] and $\mathrm{Q}(\textsc{Disj}_n) = O(\sqrt{n})$ [BCW98, AA03]. Our first result gives the first super-quadratic separation between $\mathrm{Q}(F)$ and $\mathrm{R}(F)$ for a total function.

Theorem 1. *There exists a total function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ with $\mathrm{R}(F) = \widetilde{\Omega}(\mathrm{Q}(F)^{2.5})$.*

In fact, we establish a power 2.5 separation between $\mathrm{Q}(F)$ and information complexity [BJKS04], a lower bound technique for randomized communication complexity (defined in Section 2).

Another interesting question in communication complexity is the power of different lower bound techniques. After years of work on randomized communication complexity lower bounds, there are essentially two lower bound techniques that stand at the top of the heap, the aforementioned information complexity [BJKS04] and the partition bound [JK10]. Both of these techniques are known to dominate many other techniques in the literature, such as the smooth rectangle bound, corruption bound, discrepancy, etc., but the relationship between them is not yet known. For deterministic protocols, a bound even more powerful than the partition bound, is the logarithm of the partition number. The partition number, denoted $\chi(F)$, is the smallest number of F-monochromatic rectangles in a partition of $\mathcal{X} \times \mathcal{Y}$ (see Section 2 for more precise definitions). We use the notation $\mathrm{UN}(F) = \log \chi(F)$, where UN stands for unambiguous nondeterministic communication complexity.

Showing separations between $\mathrm{R}(F)$ and $\mathrm{UN}(F)$ is very difficult because there are few techniques available to lower bound $\mathrm{R}(F)$ that do not also lower bound $\mathrm{UN}(F)$. Indeed, until recently only a factor 2 separation was known even between $\mathrm{D}(F)$ and $\mathrm{UN}(F)$, shown by Kushilevitz, Linial, and Ostrovsky [KLO99]. This changed with the breakthrough work of Göös, Pitassi, and Watson [GPW15], who exhibited a total function F with $\mathrm{D}(F) = \widetilde{\Omega}(\mathrm{UN}(F)^{1.5})$. Ambainis, Kokainis and Kothari [AKK16] improved this by constructing a total function F with $\mathrm{D}(F) \geq \mathrm{UN}(F)^{2-o(1)}$. This separation is nearly optimal as Aho, Ullman, and Yannakakis [AUY83] showed $\mathrm{D}(F) = O(\mathrm{UN}(F)^2)$ for all total F.

Göös, Jayram, Pitassi, and Watson [GJPW15] improved the original [GPW15] separation in a different direction, constructing a total F for which $\mathrm{R}(F) = \widetilde{\Omega}(\mathrm{UN}(F)^{1.5})$. In this paper, we achieve a nearly optimal separation between these measures.

Theorem 2. *There exists a total function* $F\colon \mathcal{X}\times\mathcal{Y}\to\{0,1\}$ *with* $\mathrm{R}(F)\geq \mathrm{UN}(F)^{2-o(1)}$.

In particular, this means the partition bound can be quadratically smaller than $\mathrm{R}(F)$, since the partition bound is at most $\mathrm{UN}(F)$.

1.1 Comparison with prior work

The model of query complexity provides insight into communication complexity and is usually easier to understand. Many theorems in query complexity have analogous results in communication complexity. There is also a more precise connection between these models, which we now explain. For a function $f\colon\{0,1\}^n\to\{0,1\}$, let $\mathrm{D}^{\mathrm{dt}}(f)$ be the deterministic query complexity of f, the minimum number of queries an algorithm needs to the bits of the input x to compute $f(x)$, in the worst case. Similarly, let $\mathrm{R}^{\mathrm{dt}}(f)$, $\mathrm{Q}^{\mathrm{dt}}(f)$, and $\mathrm{UN}^{\mathrm{dt}}(f)$ denote the randomized, quantum and unambiguous nondeterministic query complexities of f.

Any function f can be turned into a communication problem by composing it with a communication "gadget" $G\colon\mathcal{X}\times\mathcal{Y}\to\{0,1\}$. On input $((x_1,\ldots,x_n),(y_1,\ldots,y_n))$ the function $f\circ G$ evaluates to $f(G(x_1,y_1),\ldots,G(x_n,y_n))$. It is straightforward to see that $\mathrm{D}(f\circ G)\leq \mathrm{D}^{\mathrm{dt}}(f)\,\mathrm{D}(G)$, and analogous results hold for $\mathrm{UN}(f\circ G)$, $\mathrm{R}(f\circ G)$, and $\mathrm{Q}(f\circ G)$ (with extra logarithmic factors).

The reverse direction, that is, lower bounding the communication complexity of $f\circ G$ in terms of the query complexity of f is not always true, but can hold for specific functions G. Such results are called "lifting" theorems and are highly nontrivial. Göös, Pitassi, and Watson [GPW15], building on work of Raz and McKenzie [RM99], show a general lifting theorem for deterministic query complexity: for a specific $G\colon\{0,1\}^{20\log n}\times\{0,1\}^{n^{20}}\to\{0,1\}$, with $\mathrm{D}(G)=O(\log n)$, it holds that $\mathrm{D}(f\circ G)=\Omega(\mathrm{D}^{\mathrm{dt}}(f)\log n)$, for any $f\colon\{0,1\}^n\to\{0,1\}$.

This allowed them to achieve their separation between D and UN by first showing the analogous result in the query world, i.e., exhibiting a function $f\colon\{0,1\}^n\to\{0,1\}$ with $\mathrm{D}^{\mathrm{dt}}(f)=\widetilde{\Omega}(\mathrm{UN}^{\mathrm{dt}}(f)^{1.5})$, and then using the lifting theorem to achieve the same separation for a communication problem. The work of Ambainis, Kokainis, and Kothari [AKK16] followed the same plan and obtained their communication complexity separation by improving the query complexity separation of [GPW15] to $\mathrm{D}^{\mathrm{dt}}(f)\geq \mathrm{UN}^{\mathrm{dt}}(f)^{2-o(1)}$.

For separations against randomized communication complexity, as in our case, the situation is different. Analogs of our results have been shown in query complexity. Aaronson, Ben-David, and Kothari [ABK16] defined a transformation of a Boolean function, which they called the "cheat sheet technique." This transformation takes a function f and returns a cheat sheet function, f_{CS}, whose randomized query complexity is at least that of f. They used this method to give a total function f with $\mathrm{R}^{\mathrm{dt}}(f)=\widetilde{\Omega}(\mathrm{Q}^{\mathrm{dt}}(f)^{2.5})$. The cheat sheet technique is also used in [AKK16] to show the query analog of our Theorem 2, giving an f with $\mathrm{R}^{\mathrm{dt}}(f)\geq \mathrm{UN}^{\mathrm{dt}}(f)^{2-o(1)}$. These results, however, do not immediately imply similar results for communication complexity as no general theorem is known to lift randomized query lower bounds to randomized communication lower bounds. Such a theorem could hold and is an interesting open problem.

The most similar result to ours is that of Göös, Jayram, Pitassi, and Watson [GJPW15] who show $\mathrm{R}(F)=\widetilde{\Omega}(\mathrm{UN}(F)^{1.5})$. While the query analogue $\mathrm{R}^{\mathrm{dt}}(f)=\widetilde{\Omega}(\mathrm{UN}^{\mathrm{dt}}(f)^{1.5})$ was not hard to show, the communication separation required developing new communication complexity techniques. We similarly work directly in the setting of communication complexity, as described next.

1.2 Techniques

While a lifting theorem is not known for randomized query complexity, a lifting theorem is known for a stronger model known as *approximate conical junta degree*, denoted $\deg^+_{1/3}(f)$ (formally defined in Section 4.2). This is a query measure that satisfies $\deg^+_{1/3}(f) \leq \mathrm{R}(f)$ and has a known lifting theorem [GLM+15] (see Theorem 7). The first idea to obtain our theorems would be to show (say) that $\deg^+_{1/10}(\overline{f_{\mathrm{CS}}}) = \widetilde{\Omega}(\deg^+_{1/3}(f))$[1] and to use this lifting theorem. We were not able to show such a theorem, however, in part because $\deg^+_{\varepsilon}(f)$ does not behave well with respect to the error parameter ε.

Instead we work directly in the setting of communication complexity. We show randomized communication lower bounds for a broad family of communication functions called lookup functions. For intuition about a lookup function, consider first the query setting and the familiar address function $\textsc{Addr}\colon \{0,1\}^{c+2^c} \to \{0,1\}$. Think of the input as divided into two parts, $\mathbf{x} = (x_1, \ldots, x_c) \in \{0,1\}^c$ and the data $\mathbf{u} = (u_0, \ldots, u_{2^c-1}) \in \{0,1\}^{2^c}$. The bit string $\mathbf{x}$ is interpreted as an integer $\ell \in \{0, \ldots, 2^c - 1\}$ and the output of $\textsc{Addr}(\mathbf{x}, \mathbf{u})$ is u_ℓ.

A natural generalization of this problem is to instead have a function[2] $f\colon \{0,1\}^n \to \{0,1\}$ and functions $g_j\colon \{0,1\}^{cn} \times \{0,1\}^m \to \{0,1\}$ for $j \in \{0, \ldots, 2^c - 1\}$. Now the input consists of $\mathbf{x} = (x_1, \ldots, x_c)$ where each $x_i \in \{0,1\}^n$, and $\mathbf{u} = (u_0, \ldots, u_{2^c-1})$ where each $u_j \in \{0,1\}^m$. An address $\ell \in \{0, \ldots, 2^c - 1\}$ is defined by the string $(f(x_1), \ldots, f(x_c))$, and the output of the function is $g_\ell(\mathbf{x}, u_\ell)$. Call such a function a $(f, \{g_0, \ldots, g_{2^c-1}\})$-lookup function. The cheat sheet framework of [ABK16] naturally fits into this framework: the cheat sheet function f_{CS} of f is a lookup function where $g_\ell(x_1, \ldots, x_c, u_\ell) = 1$ if and only if u_ℓ provides certificates that $f(x_i) = \ell_i$ for each $i \in [c]$.

This idea also extends to communication complexity where one can define a $(F, \mathcal{G})$-lookup function in the same way, with F a communication function and $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$ a family of communication functions. Our main technical theorem (Theorem 5) states that, under mild conditions on the family $\mathcal{G}$, the randomized communication complexity of the $(F, \mathcal{G})$-lookup function is at least that of F. To prove the separation of Theorem 1, we take the function $f = \textsc{Simon}_n \circ \textsc{Or}_n \circ \textsc{And}_n$ and let F be f composed with the inner product communication gadget. We define the family of functions $\mathcal{G}$ in a similar fashion as in the cheat sheet framework. We show a randomized communication lower bound on F using the approximate conical junta degree and the lifting theorem of [GLM+15].

Moving on to our second result (Theorem 2), we find that just having a lower bound on the randomized communication complexity of a $(F, \mathcal{G})$-lookup function is not enough to obtain the separation. The query analogue of Theorem 2 [AKK16] relies on repeatedly composing a function with $\textsc{And}_n$ (or $\textsc{Or}_n$), which raises its randomized query complexity by $\Omega(n)$. More precisely, it relies on the fact that $\mathrm{R}^{\mathrm{dt}}(\textsc{And}_n \circ f) = \Omega(n\,\mathrm{R}^{\mathrm{dt}}(f))$. However, the analogous communication complexity claim, $\mathrm{R}(\textsc{And}_n \circ F) = \Omega(n\,\mathrm{R}(F))$, is false. For a silly example, if F itself is $\textsc{And}_n$ (under some bipartition of input bits), then $\mathrm{R}(\textsc{And}_n \circ F) \leq \mathrm{D}(\textsc{And}_{n^2}) = O(1)$. Another example is if $F\colon \{0,1\} \times \{0,1\} \to \{0,1\}$ is the equality function on 1 bit, then $\mathrm{R}(\textsc{And}_n \circ F) = O(1)$, since this is the equality function on n bits.

To circumvent this issue, we use information complexity instead of randomized communication complexity. Let $\mathrm{IC}(F)$ denote the information complexity of a function F (defined in Section 2). Information complexity, or more precisely one-sided information complexity, satisfies a composition theorem for the $\textsc{And}_n$ function (Fact 37). While one-sided information complexity upper bounds

[1] We negate the function f_{CS} because the obvious statement $\deg^+_{1/10}(f_{\mathrm{CS}}) = \widetilde{\Omega}(\deg^+_{1/3}(f))$ is false in general.

[2] For simplicity we restrict to total functions here. The full definition (Definition 19) also allows for partial functions.

can be converted to information complexity upper bounds (Fact 38), the conversion also requires upper bounding the communication complexity of the protocol. This makes the argument delicate and requires simultaneously keeping track of the information complexity and communication complexity throughout the argument. Informally, we show the following theorem.

Theorem 3 (informal)**.** *For any function F, and any family of functions $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$ let $F_\mathcal{G}$ be the $(F, \mathcal{G})$-lookup function. Provided $\mathcal{G}$ satisfies certain mild technical conditions,* $\mathrm{R}(F_\mathcal{G}) = \widetilde{\Omega}(\mathrm{R}(F))$ *and* $\mathrm{IC}(F_\mathcal{G}) = \widetilde{\Omega}(\mathrm{IC}(F))$.

We prove this formally as Theorem 5 in Section 3. This is the most technical part of the paper, requiring all the preliminary facts and notation set up in Section 2.1 and Section 2.2. The proof relies on an information theoretic argument that establishes that a correct protocol for $F_\mathcal{G}$ already has enough information to compute one copy of the base function F.

2 Preliminaries and notation

In this paper we denote query complexity (or decision tree complexity) measures using the superscript dt. For example, the deterministic, bounded-error randomized, and bounded-error quantum query complexities of a function f are denoted $\mathrm{D}^{\mathrm{dt}}(f)$, $\mathrm{R}^{\mathrm{dt}}(f)$, and $\mathrm{Q}^{\mathrm{dt}}(f)$ respectively. We refer the reader to the survey by Buhrman and de Wolf [BdW02] for formal definitions of these measures.

A function $f\colon \{0,1\}^n \to \{0,1,*\}$ is said to be a total function if $f(x) \in \{0,1\}$ for all $x \in \{0,1\}^n$ and is said to be partial otherwise. We define $\mathrm{dom}(f) \coloneqq \{x : f(x) \neq *\}$ to be the set of valid inputs to f. An algorithm computing f is allowed to output an arbitrary value for inputs outside $\mathrm{dom}(f)$. $\textsc{And}_n$ and $\textsc{Or}_n$ denote the $\textsc{And}$ and $\textsc{Or}$ functions on n bits, defined as $\textsc{And}_n(x_1, \ldots, x_n) \coloneqq \bigwedge_{i=1}^n x_i$ and $\textsc{Or}_n(x_1, \ldots, x_n) \coloneqq \bigvee_{i=1}^n x_i$. In general, f_n denotes an n-bit function.

In communication complexity, we wish to compute a function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1,*\}$ for some finite sets $\mathcal{X}$ and $\mathcal{Y}$, where the inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ are given to two players Alice and Bob, while minimizing the communication between the two. As in query complexity, F is total if $F(x,y) \in \{0,1\}$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$ and is partial otherwise. We define $\mathrm{dom}(F) \coloneqq \{(x,y) : F(x,y) \neq *\}$. As before a correct protocol may behave arbitrarily on inputs outside $\mathrm{dom}(F)$. Formal definitions of the measures studied here can be found in the textbook by Kushilevitz and Nisan [KN06].

For a function $f\colon \{0,1\}^n \to \{0,1,*\}$ we let f^c denote the function $f^c\colon \{0,1\}^{nc} \to \{0,1,*\}^c$ where $f^c(x_1, \ldots, x_c) = (f(x_1), \ldots, f(x_c))$. Note that $\mathrm{dom}(f^c) = \mathrm{dom}(f)^c$. For a communication function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ we let $F^c\colon \mathcal{X}^c \times \mathcal{Y}^c \to \{0,1\}^c$ be $F^c((x_1, \ldots, x_c), (y_1, \ldots, y_c)) = (F(x_1, y_1), \ldots, F(x_c, y_c))$.

We use $\mathrm{D}(F)$ to denote the deterministic communication complexity of F, the minimum number of bits exchanged in a deterministic communication protocol that correctly computes $F(x,y)$ for all inputs in $\mathrm{dom}(F)$. Public-coin randomized and quantum (without entanglement) communication complexities, denoted $\mathrm{R}(F)$ and $\mathrm{Q}(F)$, are defined similarly except the protocol may now err with probability at most $1/3$ on any input and may use random coins or quantum messages respectively. We use $\mathrm{N}(F)$ and $\mathrm{UN}(F)$ to denote the nondeterministic (or certificate) complexity of F and the unambiguous nondeterministic complexity of F respectively. $\mathrm{UN}(F)$ equals $\log \chi(F)$, where $\chi(F)$ is the partition number of F, the least number of monochromatic rectangles in a partition (or disjoint cover) of $\mathcal{X} \times \mathcal{Y}$. We now define these measures formally.

Given a partial function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1,*\}$ and $b \in \{0,1\}$, a b-monochromatic rectangle is a set $A \times B$ with $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$ such that all inputs in $A \times B$ evaluate to b or $*$ on F. A b-cover of F is a set of b-monochromatic rectangles that cover all the b-inputs (i.e., inputs that evaluate to b on F) of F. If the rectangles form a partition of the b-inputs, we say that the cover is unambiguous. Given a b-cover of F, a b-certificate for input (x, y) is the label of a rectangle containing (x, y) in the b-cover. The b-cover number $\mathrm{C}_b(F)$ is the size of the smallest b-cover, and we set $\mathrm{N}_b(F) := \lceil \log \mathrm{C}_b(F) \rceil$. The nondeterministic complexity of F is $\mathrm{N}(F) := \max\{\mathrm{N}_0(F), \mathrm{N}_1(F)\}$. The quantities $\mathrm{UN}_b(F)$ and the unambiguous non-deterministic complexity $\mathrm{UN}(F)$ are defined analogously from partitions.

It is useful to interpret a b-certificate for $(x, y) \in \mathrm{dom}(F)$ as a message that an all-powerful prover can send to the players to convince each of them that $F(x, y) = b$. In this interpretation, $\mathrm{N}_b(F)$ is the minimum over prover strategies of the maximum length of a message taken over all inputs. Similarly, $\mathrm{UN}_b(F)$ is the maximum length of a message when, in addition, for every input in $\mathrm{dom}(F)$, there is exactly one certificate the prover can send.

We also use $\mathrm{IC}(F)$ to denote the information complexity of F, defined formally in Section 2.2. Informally, the information complexity of a function F is the minimum amount of information about their inputs that the players have to reveal to each other to compute F. $\mathrm{IC}(F)$ is a lower bound on randomized communication complexity, because the number of bits communicated in a protocol is certainly an upper bound on the information gained by any player, since 1 bit of communication can at most have 1 bit of information.

In Section 2.1 and Section 2.2 we cover some preliminaries needed to prove Theorem 5.

2.1 Information theory

We now introduce some definitions and facts from information theory. Please refer to the textbook by Cover and Thomas [CT06] for an excellent introduction to information theory.

For a finite set S, we say $P\colon S \to \mathbb{R}^+$ is a probability distribution over S if $\sum_{s \in S} P(s) = 1$. For correlated random variables $XYZ \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, we use the same symbol represent the random variable and its distribution. If μ is a distribution over $\mathcal{X}$, we say $X \sim \mu$ to represent that X is distributed according to μ and $X \sim Y$ to represent that X and Y are similarly distributed. We use Y^x as shorthand for $(Y \mid (X = x))$. We define the joint random variable $X \otimes Y \in \mathcal{X} \times \mathcal{Y}$ as

$$\Pr(X \otimes Y = (x, y)) = \Pr(X = x) \cdot \Pr(Y = y).$$

We call X and Y independent random variables if $XY \sim X \otimes Y$.

A basic fact about random variables is Markov's inequality. We'll often make use of one particular corollary of the inequality, which we state here for convenience.

Fact 1 (Markov's Inequality)**.** *If Z is a random variable over $\mathbb{R}^+$, then for any $c \geq 1$,*

$$\Pr(Z \geq c\mathbb{E}[Z]) \leq \frac{1}{c}.$$

In particular, if f is a function mapping the domains of X and Y to $\mathbb{R}^+$, then

$$\Pr_{x \leftarrow X}(\mathbb{E}_{y \leftarrow Y^x}[f(x, y)] > \alpha) < \beta \quad \Rightarrow \quad \Pr_{(x,y) \leftarrow XY}(f(x, y) > 100\alpha) < \beta + 0.01.$$

Proof. To see that the first equation holds, note that if the elements $z \in \mathcal{Z}$ that are larger than $c\mathbb{E}[Z]$ have probability mass more than $1/c$, then they contribute more than $\mathbb{E}[Z]$ to the expectation of Z; but since the domain of Z is non-negative, this implies the expectation of Z is larger than $\mathbb{E}[Z]$, which is a contradiction.

Now suppose that $\Pr_{x \leftarrow X}(\mathbb{E}_{y \leftarrow Y^x}[f(x, y)] > \alpha) < \beta$. We classify the elements $x \in \mathcal{X}$ into two types: the "bad" ones, which satisfy $\mathbb{E}_{y \leftarrow Y^x}[f(x, y)] > \alpha$, and the "good" ones, which satisfy $\mathbb{E}_{y \leftarrow Y^x}[f(x, y)] \leq \alpha$. Note that the probability that an x sampled from X is bad is less than β. For good x, we have $\Pr_{y \leftarrow Y^x}(f(x, y) > 100\alpha) \leq 0.01$ by Markov's inequality above (using the fact that $f(x, y)$ is non-negative and $\mathbb{E}_{y \leftarrow Y^x}[f(x, y)] \leq \alpha$). Since the probability of a bad x is less than β and for good x the equation $f(x, y) \leq 100\alpha$ only fails with probability 0.01 (over choice of $y \leftarrow Y^x$), we conclude

$$\Pr_{(x,y) \leftarrow XY}(f(x, y) > 100\alpha) < \beta + 0.01$$

as desired. $\square$

2.1.1 Distance measures

We now define the main distance measures we use and some properties of these measures.

Definition 2 (Distance measures)**.** Let P and Q be probability distributions over S. We define the following distance measures between distributions.

$$\text{Total variation distance:} \quad \Delta(P, Q) \coloneqq \max_{T \subseteq S} \sum_{s \in T} \big(P(s) - Q(s)\big) = \frac{1}{2} \sum_{s \in S} |P(s) - Q(s)|.$$

$$\text{Hellinger distance:} \quad \mathrm{h}(P, Q) \coloneqq \frac{1}{\sqrt{2}} \sqrt{\sum_{s \in S} \left(\sqrt{P(s)} - \sqrt{Q(s)}\right)^2}.$$

Note that this definition can be extended to arbitrary functions $P \colon S \to \mathbb{R}^+$ and $Q \colon S \to \mathbb{R}^+$. However, when P and Q are probability distributions these measures are between 0 and 1. These extremes are achieved when $P = Q$ and when P and Q have disjoint support, respectively. Conveniently, these measures are closely related and are interchangeable up to a quadratic factor.

Fact 3 (Relation between Δ and h)**.** *Let P and Q be probability distributions. Then*

$$\frac{1}{\sqrt{2}} \Delta(P, Q) \leq \mathrm{h}(P, Q) \leq \sqrt{\Delta(P, Q)}.$$

Proof. This follows from [Das11, Theorem 15.2, p. 515]. In this reference, the quantity $\sqrt{2} \cdot \mathrm{h}(P, Q)$ is used for Hellinger distance. $\square$

In this paper, we only use Hellinger distance when we invoke Fact 16 (Pythagorean property), a key step in the proof of Theorem 5. Hence we do not require any further properties of this measure.

On the other hand, total variation distance satisfies several useful properties that we use in our arguments. We review some of its basic properties below.

Fact 4 (Facts about Δ)**.** *Let P, P', Q, Q', and R be probability distributions and let $XY \in \mathcal{X} \times \mathcal{Y}$ and $X'Y'$ in $\mathcal{X} \times \mathcal{Y}$ be correlated random variables. Then we have the following facts.*

Fact 4.A (Triangle inequality)**.** $\Delta(P, Q) \leq \Delta(P, R) + \Delta(R, Q)$.

Fact 4.B (Product distributions)**.** $\Delta(P \otimes Q, P' \otimes Q') \leq \Delta(P, P') + \Delta(Q, Q')$. *Additionally, if $Q = Q'$ then $\Delta(P \otimes Q, P' \otimes Q') = \Delta(P, P')$.*

Fact 4.C (Monotonicity)**.** $\Delta(XY, X'Y') \geq \Delta(X, X')$.

Fact 4.D (Partial measurement)**.** *If $X \sim X'$, then $\Delta(XY, X'Y') = \mathbb{E}_{x \leftarrow X}[\Delta(Y^x, Y'^x)]$.*

Proof. These facts are proved as follows.

A. Let P, Q, and R be distributions over $\mathcal{X}$. Then for any $x \in \mathcal{X}$ we have $|P(x) - Q(x)| = |P(x) - R(x) + R(x) - Q(x)| \leq |P(x) - R(x)| + |R(x) - Q(x)|$. Summing over all $x \in \mathcal{X}$ yields the inequality.

B. Let P and P' be distributions over $\mathcal{X}$; Q and Q' be distributions over $\mathcal{Y}$. Then

$$\begin{aligned}\Delta(P\otimes Q, P'\otimes Q') &= \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}|P(x)Q(y)-P'(x)Q'(y)| \\ &\le \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}|P(x)Q(y)-P(x)Q'(y)|+|P(x)Q'(y)-P'(x)Q'(y)| \\ &= \frac{1}{2}\sum_{y\in\mathcal{Y}}|Q(y)-Q'(y)|+\frac{1}{2}\sum_{x\in\mathcal{X}}|P(x)-P'(x)| \quad = \Delta(P,P')+\Delta(Q,Q').\end{aligned}$$

When $Q = Q'$, the desired result follows immediately from the first line by factoring out $Q(y)$.

C. Let the distribution of XY be $P(x,y)$ and that of $X'Y'$ be $Q(x,y)$. Let marginals on $\mathcal{X}$ be $P(x) \coloneqq \sum_y P(x,y)$ and $Q(x) \coloneqq \sum_y Q(x,y)$. Since $\Delta(X,X') = \sum_x |P(x)-Q(x)|$, we have

$$\sum_x \big|P(x)-Q(x)\big| = \sum_x \Big|\sum_y \big(P(x,y)-Q(x,y)\big)\Big| \le \sum_{xy}\big|P(x,y)-Q(x,y)\big| = \Delta(XY,X'Y').$$

D. Let the distribution of XY be $P(x,y)$ and that of $X'Y'$ be $Q(x,y)$. Let marginals on $\mathcal{X}$ be $P(x) \coloneqq \sum_y P(x,y)$ and $Q(x) \coloneqq \sum_y Q(x,y)$. Furthermore, let $P(y|x) \coloneqq P(x,y)/P(x)$ and $Q(y|x) \coloneqq Q(x,y)/Q(x)$ be the distributions of Y^x and Y'^x respectively. By assumption, we have $P(x) = Q(x)$. Then we can rewrite $\Delta(XY,X'Y') = \frac{1}{2}\sum_{xy}|P(x,y)-Q(x,y)|$ as

$$\frac{1}{2}\sum_{xy}\big|P(x,y)-Q(x,y)\big| = \frac{1}{2}\sum_x P(x)\sum_y\big|P(y|x)-Q(y|x)\big| = \mathbb{E}_{x\leftarrow X}[\Delta(Y^x,Y'^x)]. \qquad \square$$

2.1.2 Markov chains

We now define the concept of a Markov chain. We use Markov chains in our analysis because of Fact 15 (Independence) introduced in Section 2.2.

Definition 5 (Markov chain)**.** We say XYZ is a Markov chain (denoted $X \leftrightarrow Y \leftrightarrow Z$) if

$$\Pr(XYZ = (x,y,z)) = \Pr(X=x)\cdot\Pr(Y=y \mid X=x)\cdot\Pr(Z=z \mid Y=y).$$

Equivalently, XYZ is a Markov chain if for every y we have $(XZ)^y \sim X^y \otimes Z^y$.

The equivalence of the two definitions is shown in [CT06, eq. (2.118), p. 34]. We now present two facts about Markov chains.

Fact 6. *If $X_1X_2YZ_1Z_2$ are random variables and $(X_1X_2) \leftrightarrow Y \leftrightarrow (Z_1Z_2)$, then $X_1 \leftrightarrow Y \leftrightarrow Z_1$.*

Proof. Assuming for all y, $X_1^yX_2^yZ_1^yZ_2^y \sim X_1^yX_2^y \otimes Z_1^yZ_2^y$, we have

$$\begin{aligned}\Pr(X_1^yZ_1^y = (x_1,z_1)) &= \sum_{x_2,z_2}\Pr(X_1^yX_2^yZ_1^yZ_2^y = (x_1,x_2,z_1,z_2)) \\ &= \sum_{x_2,z_2}\Pr(X_1^yX_2^y = (x_1,x_2))\cdot\Pr(Z_1^yZ_2^y = (z_1,z_2)) \\ &= \Pr(X_1^y = x_1)\cdot\Pr(Z_1^y = z_1).\end{aligned}$$

Thus $(X_1Z_1)^y \sim X_1^y \otimes Z_1^y$. $\square$

Fact 7. *Let $X \leftrightarrow Y \leftrightarrow Z$ be a Markov chain. Then*

$$\Delta(XYZ, X \otimes Y \otimes Z) \leq \Delta(XY, X \otimes Y) + \Delta(YZ, Y \otimes Z).$$

Proof. This follows from the following inequalities.

$$\begin{aligned}
\Delta(XYZ, X \otimes Y \otimes Z) &= \mathbb{E}_{y \leftarrow Y} \Delta(X^y \otimes Z^y, X \otimes Z) && \text{(Fact 4.D: Partial measurement)} \\
&\leq \mathbb{E}_{y \leftarrow Y} [\Delta(X^y \otimes Z^y, X \otimes Z^y) + \Delta(X \otimes Z^y, X \otimes Z)] && \text{(Triangle inequality)} \\
&= \mathbb{E}_{y \leftarrow Y} [\Delta(X^y, X) + \Delta(Z^y, Z)] && \text{(Fact 4.B: Product distributions)} \\
&= \Delta(XY, X \otimes Y) + \Delta(YZ, Y \otimes Z). && \text{(Fact 4.D: Partial measurement)} \square
\end{aligned}$$

2.1.3 Mutual information

We now define mutual information and conditional mutual information.

Definition 8 (Mutual information)**.** Let $XYZ \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be correlated random variables. We define the following measures, where $\log(\cdot)$ denotes the base 2 logarithm.

$$\text{Mutual information:} \quad \mathbb{I}(X : Y) := \sum_{xy} \Pr(XY = (x, y)) \log \left(\frac{\Pr(XY = (x, y))}{\Pr(X = x)\Pr(Y = y)} \right).$$

$$\text{Conditional mutual information:} \quad \mathbb{I}(X : Y \mid Z) := \mathbb{E}_{z \leftarrow Z} \mathbb{I}(X : Y \mid Z = z) = \mathbb{E}_{z \leftarrow Z} \mathbb{I}(X^z : Y^z).$$

Mutual information satisfies the following basic properties.

Fact 9 (Facts about $\mathbb{I}$)**.** *Let $XYZ \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be correlated random variables. Then we have the following facts.*

Fact 9.A (Chain rule)**.** $\mathbb{I}(X : YZ) = \mathbb{I}(X : Z) + \mathbb{I}(X : Y \mid Z) = \mathbb{I}(X : Y) + \mathbb{I}(X : Z \mid Y)$.

Fact 9.B (Nonnegativity)**.** $\mathbb{I}(X : Y) \geq 0$ *and* $\mathbb{I}(X : Y \mid Z) \geq 0$.

Fact 9.C (Monotonicity)**.** $\mathbb{I}(X : YZ) \geq \mathbb{I}(X : Y)$.

Fact 9.D (Bar hopping)**.** $\mathbb{I}(X : YZ) \geq \mathbb{I}(X : Y \mid Z)$, *where equality holds if* $\mathbb{I}(X : Z) = 0$.

Fact 9.E (Independence)**.** *If Y and Z are independent, then* $\mathbb{I}(X : YZ) \geq \mathbb{I}(X : Z) + \mathbb{I}(X : Y)$.

Fact 9.F (Data processing)**.** *If $X \leftrightarrow Y \leftrightarrow Z$ is a Markov chain, then* $\mathbb{I}(X : Y) \geq \mathbb{I}(X : Z)$.

Proof. These facts are proved as follows.

A. See [CT06, Theorem 2.5.2, p. 24].

B. See [CT06, eq. (2.90), p. 28] and [CT06, eq. (2.92), p. 29].

C. Follows from Fact 9.A (Chain rule) and Fact 9.B (Nonnegativity).

D. From Fact 9.A (Chain rule) and Fact 9.B (Nonnegativity), it follows that $\mathbb{I}(X : YZ) = \mathbb{I}(X : Y \mid Z) + \mathbb{I}(X : Z) \geq \mathbb{I}(X : Y \mid Z)$.

E. Using Fact 9.A (Chain rule), we have $\mathbb{I}(X : Z \mid Y) = \mathbb{I}(Z : X \mid Y) = \mathbb{I}(Z : XY) - \mathbb{I}(Z : Y)$. Since Y and Z are independent, $\mathbb{I}(Z : Y) = 0$, and hence we get

$$\mathbb{I}(X : Z \mid Y) = \mathbb{I}(Z : XY) \geq \mathbb{I}(Z : X) = \mathbb{I}(X : Z). \qquad \text{(Fact 9.C: Monotonicity)}$$

Then using Fact 9.A gives $\mathbb{I}(X : YZ) = \mathbb{I}(X : Y) + \mathbb{I}(X : Z \mid Y) \geq \mathbb{I}(X : Y) + \mathbb{I}(X : Z)$.

F. See [CT06, Theorem 2.8.1, p. 34]. □

We now present a way to relate mutual information and total variation distance.

Fact 10 (Relation between $\mathbb{I}$ and Δ)**.** *Let $XY \in \mathcal{X} \times \mathcal{Y}$ be correlated random variables. Then*

$$\mathbb{I}(X:Y) \geq \Delta^2(XY, X \otimes Y) \qquad \textit{and} \qquad \mathbb{I}(X:Y) \geq \mathbb{E}_{x \leftarrow X} \Delta^2(Y^x, Y).$$

Proof. To prove this we will require a distance measure called relative entropy (or Kullback–Leibler divergence). For any probability distributions P and Q over S, we define

$$\mathbb{D}(P\|Q) \coloneqq \sum_{s \in S} P(s) \log \frac{P(s)}{Q(s)}.$$

We can now express $\mathbb{I}(X:Y)$ in terms of relative entropy as follows:

$$\mathbb{I}(X:Y) = \mathbb{D}(XY\|X \otimes Y) = \mathbb{E}_{x \leftarrow X} \mathbb{D}(Y^x\|Y). \tag{1}$$

The first equality follows straightforwardly from definitions, as shown in [CT06, eq. 2.29, p. 20]. For the second equality, we proceed as follows:

$$\mathbb{D}(XY\|X \otimes Y) = \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)}\right) = \sum_{x} p(x) \sum_{y} p(y|x) \log \left(\frac{p(y|x)}{p(y)}\right) = \mathbb{E}_{x \leftarrow X} \mathbb{D}(Y^x\|Y).$$

We then use Pinsker's inequality [CT06, Lemma 11.6.1, p. 370], which states

$$\mathbb{D}(P\|Q) \geq \frac{2}{\ln 2} \Delta^2(P, Q) \geq \Delta^2(P, Q).$$

Combining (1) with Pinsker's inequality completes the proof. □

In general, it is impossible to relate $\mathbb{I}$ and Δ in the reverse direction. Indeed, mutual information is unbounded, whereas variation distance is always at most 1. However, in the case where one of the variables has binary outcomes, we have the following fact.

Fact 11 ($\mathbb{I}$ vs. Δ for binary random variables)**.** *Let AB be correlated random variables with $A \in \{0,1\}$. Let $p \coloneqq \Pr(A=0)$, $B^0 := (B \mid A=0)$, and $B^1 := (B \mid A=1)$. Then*

$$\mathbb{I}(A:B) \leq 2 \log e \;\; \Delta(pB^0, (1-p)B^1).$$

Proof. For every $s \in S$, we define

$$B(s) \coloneqq pB^0(s) + (1-p)B^1(s) \qquad \text{and} \qquad D(s) \coloneqq pB^0(s) - (1-p)B^1(s),$$

which gives

$$B^0(s) = \frac{B(s)+D(s)}{2p} \qquad \text{and} \qquad B^1(s) = \frac{B(s)-D(s)}{2(1-p)}.$$

Recall that although $\Delta(P,Q)$ is a distance measure for probability distributions, it is well defined when P and Q are unnormalized. In particular, $\Delta(pB^0, (1-p)B^1) = \frac{1}{2}\sum_s |D(s)|$. We can now upper bound $\mathbb{I}(A:B)$ as follows.

$$\mathbb{I}(A:B) = \sum_{a \in \{0,1\}} \sum_{s \in S} \Pr(A=a) B^a(s) \log \left(\frac{B^a(s)}{B(s)}\right)$$

$$
\begin{aligned}
&= \sum_{s\in S}\left(pB^0(s)\log\left(\frac{B(s)+D(s)}{2pB(s)}\right)+(1-p)B^1(s)\log\left(\frac{B(s)-D(s)}{2(1-p)B(s)}\right)\right)\\
&= \mathbb{H}(p)-1+\sum_{s\in S}\left(pB^0(s)\log\left(1+\frac{D(s)}{B(s)}\right)+(1-p)B^1(s)\log\left(1-\frac{D(s)}{B(s)}\right)\right),
\end{aligned}
$$

where $\mathbb{H}(p) \coloneqq -p\log p-(1-p)\log(1-p)$ is the binary entropy function. Since $\mathbb{H}(p)\leq 1$, we have

$$
\mathbb{I}(A:B)\leq(\log e)\sum_{s\in S}\left(pB^0(s)\frac{D(s)}{B(s)}-(1-p)B^1(s)\frac{D(s)}{B(s)}\right)=(\log e)\sum_{s\in S}\frac{D(s)^2}{B(s)},
$$

using $\log(1+x)\leq x\log e$ (for all real x). Since $B^0(s)\geq 0$ and $B^1(s)\geq 0$ for all s, we have $|D(s)|\leq B(s)$. Hence

$$
\mathbb{I}(A:B)\leq(\log e)\sum_{s\in S}\frac{D(s)^2}{B(s)}=(\log e)\sum_{s}\frac{|D(s)|^2}{B(s)}\leq(\log e)\sum_{s\in S}|D(s)|=2\log e\;\;\Delta(pB^0,(1-p)B^1).
$$

□

Note that this inequality is tight up to constants. To see this, for any $\delta\in[0,1]$, consider the distributions $B^0=(1-\delta,0,\delta)$ and $B^1=(1-\delta,\delta,0)$. If $p=1/2$, then $I(A:B)=\delta$ and $\Delta(pB^0,(1-p)B^1)=\delta/2$.

Our next fact gives us a way to use high mutual information between two variables to get a good prediction of one variable using a sample from the other.

Fact 12 (Information $\Rightarrow$ prediction)**.** *Let AB be correlated random variables with $A\in\{0,1\}$. The probability of predicting A by a measurement on B is at least*

$$
\frac{1}{2}+\frac{\mathbb{I}(A:B)}{3}.
$$

Proof. Let $p=\Pr(A=0)$ and define a measurement M corresponding to output 1 as follows: $M(s)=0$ for all $s\in S$ such that $pB^0(s)\geq(1-p)B^1(s)$ and $M(s)=1$ otherwise. We view M as a vector, and let $\mathbb{1}$ represents the all-1 vector. Then the success probability of this measurement is

$$
\begin{aligned}
p\langle\mathbb{1}-M,B^0\rangle+(1-p)\langle M,B^1\rangle &= p\langle\mathbb{1},B^0\rangle+\langle M,(1-p)B^1-pB^0\rangle\\
&= p+\sum_{s\in S\,:\,(1-p)B^1(s)-pB^0(s)>0}(1-p)B^1(s)-pB^0(s)\\
&= p+\frac{1}{2}\sum_{s\in S}|(1-p)B^1(s)-pB^0(s)|+(1-p)B^1(s)-pB^0(s)\\
&= \frac{1}{2}+\frac{1}{2}\sum_{s\in S}|(1-p)B^1(s)-pB^0(s)|\\
&= \frac{1}{2}+\Delta(pB^0,(1-p)B^1)
\end{aligned}
$$

From Fact 11, we know that $\Delta(pB^0,(1-p)B^1)\geq\mathbb{I}(A:B)/(2\log e)\geq\mathbb{I}(A:B)/3$. □

2.2 Communication complexity

Let $F\colon\mathcal{X}\times\mathcal{Y}\to\{0,1,*\}$ be a partial function, with $\mathrm{dom}(F)\coloneqq\{(x,y)\in\mathcal{X}\times\mathcal{Y}: F(x,y)\neq *\}$, and let $\varepsilon\in(0,1/2)$. In a (randomized) communication protocol for computing F, Alice gets input

$x \in \mathcal{X}$, Bob gets input $y \in \mathcal{Y}$. Alice and Bob may use private and public coins. They exchange messages and at the end of the protocol, they output $O(x, y)$. We assume $O(x, y)$ is contained in the messages exchanged by Alice and Bob. We let the random variable Π represent the *transcript* of the protocol, that is the messages exchanged and the public-coins used in the protocol Π. Let μ be a distribution over $\mathrm{dom}(F)$ and let $XY \sim \mu$. We define the following quantities.

$$\begin{aligned}
\text{Worst-case error:} \quad & \mathrm{err}(\Pi) \coloneqq \max_{(x,y)\in\mathrm{dom}(F)} \{\Pr[O(x,y) \neq F(x,y)]\}.\\
\text{Distributional error:} \quad & \mathrm{err}^{\mu}(\Pi) \coloneqq \mathbb{E}_{(x,y)\leftarrow XY} \Pr[O(x,y) \neq F(x,y)].\\
\text{Distributional IC:} \quad & \mathrm{IC}^{\mu}(\Pi) \coloneqq \mathbb{I}(X : \Pi \mid Y) + \mathbb{I}(Y : \Pi \mid X).\\
\text{Max. distributional IC:} \quad & \mathrm{IC}(\Pi) \coloneqq \max_{\mu \text{ on } \mathrm{dom}(F)} \mathrm{IC}^{\mu}(\Pi).\\
\text{IC of } F\text{:} \quad & \mathrm{IC}_{\varepsilon}(F) \coloneqq \inf_{\Pi:\mathrm{err}(\Pi)\leq\varepsilon} \mathrm{IC}(\Pi) = \inf_{\Pi:\mathrm{err}(\Pi)\leq\varepsilon} \max_{\mu \text{ on } \mathrm{dom}(F)} \mathrm{IC}^{\mu}(\Pi).\\
\text{Randomized CC:} \quad & \mathrm{CC}(\Pi) \coloneqq \text{ max. number of bits exchanged in } \Pi \text{ (over inputs and coins)}.\\
\text{Randomized CC of } F\text{:} \quad & \mathrm{R}_{\varepsilon}(F) \coloneqq \min_{\Pi:\mathrm{err}(\Pi)\leq\varepsilon} \mathrm{CC}(\Pi).
\end{aligned}$$

Note that since one bit of communication can hold at most one bit of information, for any protocol Π and distribution μ we have $\mathrm{IC}^{\mu}(\Pi) \leq \mathrm{CC}(\Pi)$. Consequently, we have $\mathrm{IC}_{\varepsilon}(F) \leq \mathrm{R}_{\varepsilon}(F)$. When ε is unspecified, we assume $\varepsilon = 1/3$. Hence $\mathrm{IC}(F) \coloneqq \mathrm{IC}_{1/3}(F)$, $\mathrm{R}(F) \coloneqq R_{1/3}(F)$, and $\mathrm{IC}(F) \leq \mathrm{R}(F)$. We now present some facts that relate these measures.

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 IC and CC.

Fact 13 (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*

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

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 communication and information complexities only increase by a factor of $O\left(\frac{\log(1/\delta)}{(1/2-\varepsilon)^2}\right)$.

A useful tool in proving lower bounds on randomized communication complexity is Yao's minimax principle [Yao77], which says that the worst-case randomized communication complexity of F is the same as the maximum distributional communication complexity over distributions μ on $\mathrm{dom}(F)$. In particular, this means there always exists a hard distribution μ over which any protocol needs as much communication as in the worst case. More precisely, it states that

$$\mathrm{R}_{\varepsilon}(F) = \max_{\mu \text{ on } \mathrm{dom}(F)} \min_{\Pi:\mathrm{err}^{\mu}(\Pi)\leq\varepsilon} \mathrm{CC}(\Pi).$$

Similar to Yao's minimax principle for randomized communication complexity, we have a (slightly weaker) minimax principle for information complexity due to Braverman [Bra12].

Fact 14 (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 an information bound $I \geq 0$. Suppose $\mathcal{P}$ is a family of protocols such that for every distribution μ on $\mathrm{dom}(F)$ there exists a protocol $\Pi \in \mathcal{P}$ such that*

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

Then for any $\alpha \in (0,1)$ there exists a protocol Π' such that

$$\mathrm{err}(\Pi') \leq \varepsilon/\alpha \quad \text{and} \quad \mathrm{IC}(\Pi') \leq I/(1-\alpha).$$

Moreover, Π' is a probability distribution over protocols in $\mathcal{P}$, and hence $\mathrm{CC}(\Pi') \leq \max_{\Pi \in \mathcal{P}} \mathrm{CC}(\Pi)$.

Our next fact is the observation that if Alice's and Bob's inputs are drawn independently from each other, conditioning on the transcript at any stage of the protocol keeps the input distributions independent of each other.

Fact 15 (Independence)**.** *Let Π be a communication protocol on input $X \otimes Y$. Then $X \leftrightarrow \Pi \leftrightarrow Y$ forms a Markov chain, or equivalently, for each π in the support of Π, we have*

$$(XY)^\pi \sim X^\pi \otimes Y^\pi.$$

Proof. Follows easily by induction on the number of message exchanges in protocol Π. $\square$

The next property of communication protocols formalizes the intuitive idea that if Alice and Bob had essentially the same transcript for input pairs (x,y) and (x',y'), then if we fix Bob's input to either y or y', the transcripts obtained for the two different inputs to Alice are nearly the same. This was shown by Bar-Yossef, Jayram, Kumar, and Sivakumar [BJKS04].

Fact 16 (Pythagorean property)**.** *Let (x,y) and (x',y') be two inputs to a protocol Π. Then*

$$\mathrm{h}^2(\Pi(x,y),\Pi(x',y)) + \mathrm{h}^2(\Pi(x,y'),\Pi(x',y')) \leq 2 \cdot \mathrm{h}^2(\Pi(x,y),\Pi(x',y')).$$

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 17. *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 $XY \sim \mu$ and define the random variable $F := F(X,Y)$. Let Π be a communication protocol with input (X,Y) to Alice and Bob respectively. There exists a communication protocol Π' for F, with input (X,Y) to Alice and Bob respectively, such that*

$$\mathrm{IC}^\mu(\Pi') \leq \mathrm{IC}^\mu(\Pi) + 1, \quad \mathrm{CC}(\Pi') = \mathrm{CC}(\Pi) + 1, \quad \textit{and} \quad \mathrm{err}^\mu(\Pi') < \frac{1}{2} - \frac{\mathbb{I}(F : \Pi \mid X)}{3}.$$

Proof. In Π', Alice and Bob run the protocol Π and at the end Alice makes a prediction for F based on the transcript and her input, essentially applying Fact 12 (Information $\Rightarrow$ prediction) to the random variables F^x and Π^x. Alice then sends her prediction, a single additional bit, to Bob. Clearly,

$$\mathrm{IC}^\mu(\Pi') \leq \mathrm{IC}^\mu(\Pi) + 1 \quad \text{and} \quad \mathrm{CC}(\Pi') = \mathrm{CC}(\Pi) + 1.$$

For every input x for Alice, her prediction is successful (assuming Bob's input is sampled from Y^x) with probability at least $1/2 + \mathbb{I}(F^x : \Pi^x)/3$ by Fact 12. Hence the overall success probability of Π' is at least

$$\mathbb{E}_{x \leftarrow X}\left[\frac{1}{2} + \frac{\mathbb{I}(F^x : \Pi^x)}{3}\right] = \frac{1}{2} + \frac{\mathbb{E}_{x \leftarrow X}[\mathbb{I}(F^x : \Pi^x)]}{3} = \frac{1}{2} + \frac{\mathbb{I}(F : \Pi | X)}{3}. \qquad \square$$

The following claim is used in the proof of our main Theorem 5 to handle the easy case of a biased input distribution.

Claim 18. *Let $F \colon \mathcal{X} \times \mathcal{Y} \to \{0,1,*\}$ be a partial function and let μ be a distribution over $\mathrm{dom}(F)$. Let $\varepsilon \in (0,1/2)$ and $c \geq 1$ be a positive integer. For $i \in [c]$, let $X_iY_i \sim \mu$ be i.i.d. and define $L_i := F(X_i,Y_i)$. Define $X := (X_1,\ldots,X_c)$, $Y := (Y_1,\ldots,Y_c)$, and $L := (L_1,\ldots,L_c)$. Then either* *(a) There exists a protocol* Π *for* F *such that* $\mathrm{CC}(\Pi) = 1$, $\mathrm{IC}^{\mu}(\Pi) \leq 1$, *and* $\mathrm{err}^{\mu}(\Pi) \leq \frac{1}{2} - \varepsilon$, *or*

(b) $\Delta(XL, X \otimes W^{\otimes c}) \leq c\varepsilon$, *where* W *is the uniform distribution over* $\{0,1\}$.

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

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

Assume otherwise $\mathbb{E}_{x_1 \leftarrow X_1}[|\frac{1}{2} - q^{x_1}|] < \varepsilon$. Let W be the uniform distribution on $\{0,1\}$. This implies

$$\Delta(XL, X \otimes W^{\otimes c}) \leq c \cdot \Delta(X_1 L_1, X_1 \otimes W) = c \cdot \mathbb{E}_{x_1 \leftarrow X_1}[|\frac{1}{2} - q^{x_1}|] < c\varepsilon,$$

where the first inequality follows from Fact 4.B (Product distributions). $\square$

3 Lookup functions in communication complexity

We now describe the class of functions we will use for our separations, $(F, \mathcal{G})$-lookup functions. This class of communication functions and our applications of them are inspired by the cheat sheet functions defined in query complexity in [ABK16].

A $(F, \mathcal{G})$-lookup function, denoted $F_{\mathcal{G}}$, is defined by a (partial) communication function $F \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ and a family $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$ of communication functions, where each $G_i \colon (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$. 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 $(u_0, \ldots, u_{2^c-1}) \in \{0,1\}^{m2^c}$ and likewise Bob receives input $\mathbf{y} = (y_1, \ldots, y_c) \in \mathcal{Y}^c$ and $(v_0, \ldots, v_{2^c-1}) \in \{0,1\}^{m2^c}$. 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 G_i the players should evaluate. If $\ell \in \{0,1\}^c$, i.e., all $(x_i, y_i) \in \mathrm{dom}(F)$, then the goal of the players is to output $G_\ell(\mathbf{x}, u_\ell, \mathbf{y}, v_\ell)$; otherwise, if some $(x_i, y_i) \notin \mathrm{dom}(F)$, then the goal is to output $G_0(\mathbf{x}, u_0, \mathbf{y}, v_0)$.

The formal definition follows.

Definition 19 ($(F, \mathcal{G})$-lookup function)**.** Let $F \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ be a (partial) communication function and $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$ a family of communication functions, where each $G_i \colon (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$. A $(F, \mathcal{G})$-lookup function, denoted $F_{\mathcal{G}}$, is a total communication function $F_{\mathcal{G}} \colon (\mathcal{X}^c \times \{0,1\}^{m2^c}) \times \mathcal{Y}^c \times \{0,1\}^{m2^c}$ defined as follows. Let $\mathbf{x} = (x_1, \ldots, x_c) \in \mathcal{X}^c, \mathbf{y} = (y_1, \ldots, y_c) \in \mathcal{Y}^c, \mathbf{u} = (u_0, \ldots, u_{2^c-1}) \in \{0,1\}^{m2^c}, \mathbf{v} = (v_0, \ldots, v_{2^c-1}) \in \{0,1\}^{m2^c}$. Then

$$F_{\mathcal{G}}(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v}) = \begin{cases} G_\ell(\mathbf{x}, u_\ell, \mathbf{y}, v_\ell) & \text{if } \ell = F^c(\mathbf{x}, \mathbf{y}) \in \{0,1\}^c \\ G_0(\mathbf{x}, u_0, \mathbf{y}, v_0) & \text{otherwise.} \end{cases}$$

As lookup functions form quite a general class of functions, we will need to impose additional constraints on the family of functions $\mathcal{G}$ in order to show interesting theorems about them. To show *upper bounds* on the communication complexity of lookup functions (Theorem 4), we need a *consistency* condition. This says that whenever some $(x_i, y_i) \notin \mathrm{dom}(F)$, the output of the G_j functions can depend only on $\mathbf{x}, \mathbf{y}$ and not on u, v or j.

Definition 20 (Consistency outside F)**.** Let $F \colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ be a (partial) communication function and $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$ 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 *consistent* outside F if for all $i \in \{0, \ldots, 2^c - 1\}, u, v, u', v' \in \{0,1\}^m$ and $\mathbf{x} = (x_1, \ldots, x_c) \in \mathcal{X}^c, \mathbf{y} = (y_1, \ldots, y_c) \in \mathcal{Y}^c$ with $\ell = F^c(\mathbf{x}, \mathbf{y}) \notin \{0,1\}^c$ we have $G_0(\mathbf{x}, u, \mathbf{y}, v) = G_i(\mathbf{x}, u', \mathbf{y}, v')$.

In order to show lower bounds on the communication complexity of $F_{\mathcal{G}}$ (Theorem 5) we add two additional constraints on the family $\mathcal{G}$.

Definition 21 (Nontrivial XOR family)**.** Let $\mathcal{G} = \{G_0, \dots, G_{2^c-1}\}$ 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.

1. (Nontriviality) For all $\mathbf{x} = (x_1, \dots, x_c) \in \mathcal{X}^c$ and $\mathbf{y} = (y_1, \dots, y_c) \in \mathcal{Y}^c$, if we have $\ell = F^c(\mathbf{x}, \mathbf{y}) \in \{0,1\}^c$ then for every $i \in \{0, \dots, 2^c - 1\}$ there exists $u, v, u', v' \in \{0,1\}^m$ such that $G_i(\mathbf{x}, u, \mathbf{y}, v) \neq G_i(\mathbf{x}, u', \mathbf{y}, v')$.

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

3.1 Upper bound

We now show a general upper bound on the quantum communication complexity of a $(F, \mathcal{G})$ lookup function, when $\mathcal{G}$ is consistent outside F. A similar result holds for randomized communication complexity, but we will not need this.

Theorem 4. *Let $F \colon \mathcal{X} \times \mathcal{Y} \to \{0,1,*\}$ be a (partial) function and $\mathcal{G} = \{G_0, \dots, G_{2^c-1}\}$ a family of communication functions, where each $G_i \colon (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$. If $\mathcal{G}$ is consistent outside F (Definition 20) then*

$$\mathrm{Q}(F_{\mathcal{G}}) = O(\mathrm{Q}(F) \cdot c \log c) + \max_{i \in [2^c]} O(\mathrm{Q}(G_i))$$

where $F_{\mathcal{G}}$ is the $(F, \mathcal{G})$-lookup function.

Proof. Consider an input where Alice holds $\mathbf{x} = (x_1, \dots, x_c) \in \mathcal{X}^c$ and $\mathbf{u} = (u_0, \dots, u_{2^c-1}) \in \{0,1\}^{m2^c}$ and Bob holds $\mathbf{y} = (y_1, \dots, y_c) \in \mathcal{Y}^c$ and $\mathbf{v} = (v_0, \dots, v_{2^c-1}) \in \{0,1\}^{m2^c}$. For each $i = 1, \dots, c$, Alice and Bob run an optimal protocol for F on input (x_i, y_i) $O(\log c)$ many times and let ℓ_i be the resulting majority vote. Letting $\ell = (\ell_1, \dots, \ell_c)$, they then run an optimal protocol for G_ℓ on input $\mathbf{x}, u_\ell, \mathbf{y}, v_\ell$ a constant number of times and output the majority result.

The complexity of this protocol is clearly at most $O(\mathrm{Q}(F) \cdot c \log c) + \max_i O(\mathrm{Q}(G_i))$. We now argue correctness. First suppose that each $(x_i, y_i) \in \mathrm{dom}(F)$ for $i = 1, \dots, c$. In this case, the protocol for F computes $F(x_i, y_i)$ with error at most $1/3$. Thus by running this protocol $O(\log c)$ many times and taking a majority vote $\ell = (F(x_1, y_1), \dots, F(x_c, y_c))$ with error probability at most (say) $1/6$. Similarly by running the protocol for G_ℓ a constant number of times the error probability can be reduced to $1/6$ and thus the players' output equals $G_\ell(\mathbf{x}, u_\ell, \mathbf{y}, v_\ell)$ with error probability at most $1/3$.

If some $(x_i, y_i) \notin \mathrm{dom}(F)$ then by the consistency condition $G_1(\mathbf{x}, u_1, \mathbf{y}, v_1) = G_\ell(\mathbf{x}, u_\ell, \mathbf{y}, v_\ell)$. Thus in this case the players' also output the correct answer with error probability at most $1/3$. □

3.2 Lower bound

The next theorem is the key result of our work. It gives a lower bound on the randomized communication complexity and information complexity of any $(F, \mathcal{G})$-lookup function $F_{\mathcal{G}}$, when $\mathcal{G}$ is a nontrivial XOR family, in terms of the same quantities for F. Recall that the value of $F_{\mathcal{G}}(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v})$ is equal to $G_\ell(\mathbf{x}, u_\ell, \mathbf{y}, v_\ell)$, where $\ell = F^c(\mathbf{x}, \mathbf{y})$. Intuitively, if $\mathcal{G}$ is a nontrivial family, then to evaluate $G_\ell(\mathbf{x}, u_\ell, \mathbf{y}, v_\ell)$ the players must at least know the relevant input u_ℓ, v_ℓ. This in turn requires knowing ℓ, which can only be figured out by evaluating F.

Since the argument is long, we separate out several claims that will be proven afterwards. The overall structure of the argument is explained in the main proof, and displayed visually in Figure 1.

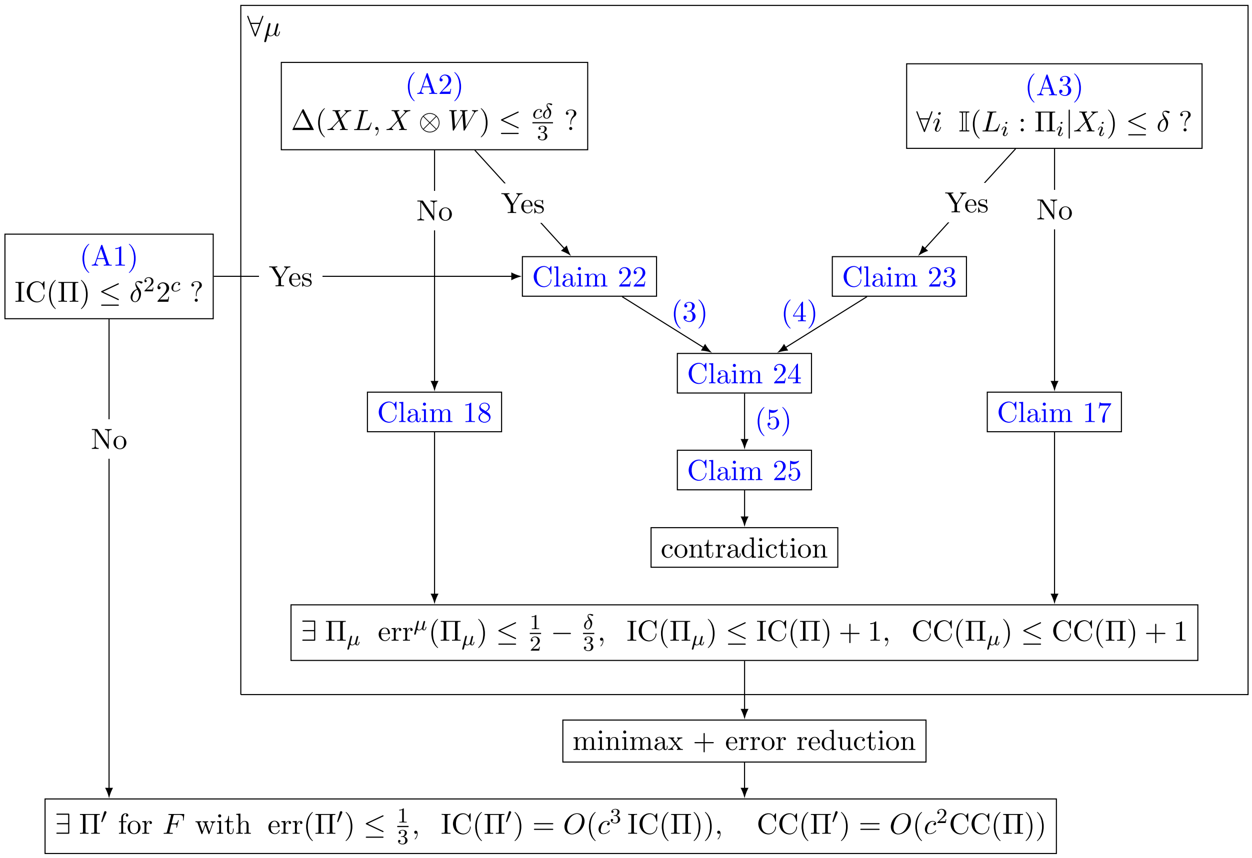

Figure 1: The structure of the proof of Theorem 5. Note that Claim 22 and Claim 24 only follow if both of their incoming arcs hold.

In Theorem 5, we are given a bounded-error protocol Π for $F_{\mathcal{G}}$, and our goal is to construct a bounded-error protocol Π' for F such that its communication complexity and information complexity do not increase by more than a polynomial in c compared to the protocol Π.

As depicted in Figure 1, if (A1) fails to hold, then we are done. Otherwise, we assume μ is an arbitrary distribution over $\mathrm{dom}(F)$, and check if (A2) or (A3) hold. We show that it is not possible for both to hold, since that leads to a contradiction. If either (A2) or (A3) fail to hold, then we have a protocol Π_μ that does well for the distrubition μ. Finally we apply a minimax argument, which converts protocols that work well against a known distribution into a protocol that works on all inputs, and obtain the desired protocol Π'.

3.2.1 Main result

Theorem 5. *Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ be a (partial) function and let $c \geq \log \mathrm{R}(F)$. Let $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$ be a nontrivial family of XOR functions (Definition 21) where each $G_i\colon (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$, and let $F_{\mathcal{G}}$ be the $(F, \mathcal{G})$-lookup function. For any $1/3$-error protocol Π for $F_{\mathcal{G}}$, there exists a $1/3$-error protocol Π' for F such that*

$$\mathrm{IC}(\Pi') \leq O(c^3\,\mathrm{IC}(\Pi)) \quad \textit{and} \quad \mathrm{CC}(\Pi') \leq O(c^2\mathrm{CC}(\Pi)).$$

In particular, $\mathrm{R}(F_{\mathcal{G}}) = \Omega(\mathrm{R}(F)/c^2)$ *and* $\mathrm{IC}(F_{\mathcal{G}}) = \Omega(\mathrm{IC}(F)/c^3)$.

Proof. In this proof, for convenience we define $\delta = \frac{1}{10^{22}c}$ (we are not trying to optimize the constants).

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

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

If this does not hold then $\mathrm{IC}(\Pi) \geq \delta^2 2^c = \Omega(\mathrm{R}(F)/c^2)$. By choosing the protocol whose communication complexity is $\mathrm{R}(F)$, we obtain a protocol Π' for F with $\mathrm{IC}(\Pi') \leq \mathrm{CC}(\Pi') = \mathrm{R}(F) = O(c^2\,\mathrm{IC}(\Pi))$ 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 14 (Minimax principle) 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:

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

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 (2).

Construct a distribution for $F_{\mathcal{G}}$. 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:

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

where for all $i \in [c]$, X_iY_i is distributed according to μ and is independent of all other random variables and for $j \in \{0, \ldots, 2^c - 1\}$, U_jV_j are uniformly distributed in $\{0,1\}^{2m}$ and independent of all other random variables.

For $i \in [c]$, we define $L_i \coloneqq F(X_i, Y_i)$. We also define $X \coloneqq (X_1, \ldots, X_c)$, $Y \coloneqq (Y_1, \ldots, Y_c)$, $L \coloneqq (L_1, \ldots, L_c)$, $U \coloneqq (U_0, \ldots, U_{2^c-1})$, and $V \coloneqq (V_0, \ldots, V_{2^c-1})$. Lastly, for $i \in [c]$ we define $X_{-i} \coloneqq X_1 \ldots X_{i-1}X_{i+1} \ldots X_c$ and $X_{<i} \coloneqq X_1 \ldots X_{i-1}$. Similar definitions hold for L and Y.

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 18 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 their public coins to sample $c-1$ inputs $X_{-i}Y_{-i}$ according to $\mu^{\otimes c}$ and inputs U and V uniformly at random. Note that Alice and Bob now have inputs XU and YV distributed according to T. The random variable corresponding to their transcript, which includes the messages exchanges and the public coins, is $\Pi_i = (\Pi, X_{-i}, U, Y_{-i}, V)$. We then claim that

$$\forall i \in [c], \quad \mathrm{CC}(\Pi_i) = \mathrm{CC}(\Pi) \quad \text{and} \quad \mathrm{IC}^{\mu}(\Pi_i) \leq \mathrm{IC}(\Pi).$$

It is obvious that $\mathrm{CC}(\Pi_i) = \mathrm{CC}(\Pi)$ because the bits transmitted in Π_i are the same as those in Π. The second part uses the following chain of inequalities, which hold for any $i \in [c]$.

$$\begin{aligned}
\mathrm{IC}(\Pi) \geq \mathrm{IC}^{T}(\Pi) &= \mathbb{I}(XU : \Pi \mid YV) + \mathbb{I}(YV : \Pi \mid XU) && \text{(definition)} \\
&\geq \mathbb{I}(X_i : \Pi \mid Y_i X_{-i} U Y_{-i} V) + \mathbb{I}(Y_i : \Pi \mid X_i X_{-i} U Y_{-i} V) && \text{(Fact 9.D: Bar hopping)} \\
&= \mathbb{I}(X_i : \Pi X_{-i} U Y_{-i} V \mid Y_i) + \mathbb{I}(Y_i : \Pi X_{-i} U Y_{-i} V \mid X_i) && \text{(Fact 9.D: Bar hopping)} \\
&= \mathrm{IC}^{\mu}(\Pi_i). && \text{(definition)}
\end{aligned}$$

The second equality used the fact that $\mathbb{I}(X_i : X_{-i} U Y_{-i} V \mid Y_i) = \mathbb{I}(Y_i : X_{-i} U Y_{-i} V \mid X_i) = 0$.

Rule out informative protocols Π_i. We then check if any of the protocols Π_i that we just constructed have a lot of information about the output L_i. If this happens then Π_i can solve F on μ and will yield the desired protocol Π_μ. More precisely, we check if the following condition holds.

$$\forall i \in [c] \quad \mathbb{I}(L_i : \Pi_i \mid X_i) \leq \delta. \tag{A3}$$

If it does not hold, then we apply Claim 17, which gives us the desired protocol Π_μ satisfying (2). Hence we may assume that (A3) holds for the rest of the proof.

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 22 to get the following from (A1) and (A2).

$$\Pr_{(x,\ell)\leftarrow XL}(\Delta((\Pi U_\ell)^x, \Pi^x \otimes U_\ell) > \sqrt{\delta}) < 0.01. \tag{3}$$

Intuitively this claim asserts that for a typical x and ℓ, the transcript Π^x has very little information about the correct cell U_ℓ, which is quantified by saying their joint distribution is close to being a product distribution. This would be false without assuming (A1) because if there was no upper bound on the information contained in Π, then the protocol could simply communicate all of U, in which case it 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.

We then apply Claim 23 to get the following from (A3).

$$\Pr_{(x,\ell)\leftarrow XL}(\Delta((\Pi U_\ell)^{x,\ell}, (\Pi U_\ell)^x) > 100\sqrt{c\delta}) \leq 0.01 \tag{4}$$

Intuitively, this claim asserts that for a typical x, the transcript (and even the contents of U_ℓ, Alice's part of the contents of the correct cheat sheet cell) does not change much upon further conditioning on ℓ. This is just one way of saying that Alice (who knows x and U) does not learn much about ℓ from the transcript Π. The assumption (A3) was necessary, because without it, it would be possible for Π to provide a lot of information about L (conditioned on X).

We then apply Claim 24, which uses (3) and (4) to obtain the following:

$$\Pr_{(x,y,\ell,u_\ell,v_\ell)\leftarrow XYLU_LV_L}\left(\Delta(\Pi^{x,y,\ell,u_\ell,v_\ell}, \Pi^{x,y,\ell}) > 6\cdot 10^6 \cdot \sqrt{c\delta}\right) < 0.09. \tag{5}$$

This equation is a key result. It says that conditioning on a typical (x, y, ℓ), the message transcript does not change much on further conditioning on a typical (u_ℓ, v_ℓ). Finally, we use Claim 25 to obtain a contradiction from (5).

Minimax argument. Note that in all branches where we did not reach a contradiction, we constructed a protocol satisfying (2). Hence we constructed, for any μ over $\mathrm{dom}(F)$, a protocol Π_μ that satisfies (2). From here it is easy to complete the proof. First we use Fact 14 (Minimax principle) with the choice $\alpha = 1 - \frac{\delta}{6}$ and $\varepsilon = \frac{1}{2} - \frac{\delta}{3}$ to get a protocol $\widetilde{\Pi}$ for F such that

$$\mathrm{IC}(\widetilde{\Pi}) \leq O\left(\frac{1}{\delta}\,\mathrm{IC}(\Pi)\right), \qquad \mathrm{CC}(\widetilde{\Pi}) \leq \mathrm{CC}(\Pi) + 1 \qquad \text{and} \qquad \mathrm{err}(\widetilde{\Pi}) \leq 1/2 - \delta/6.$$

Finally, using Fact 13 (Error reduction), we get a protocol Π' for F such that

$$\mathrm{IC}(\Pi') \leq O\left(\frac{1}{\delta^3}\,\mathrm{IC}(\Pi)\right), \qquad \mathrm{CC}(\Pi') \leq O\left(\frac{1}{\delta^2}\mathrm{CC}(\Pi)\right) \qquad \text{and} \qquad \mathrm{err}(\Pi') \leq 1/3.$$

This completes the proof since $1/\delta = O(c)$. □

This completes the proof of the theorem, except the claims we did not prove, Claim 22, Claim 23, Claim 24, and Claim 25. We now prove these claims.

3.2.2 Proofs of claims

Claim 22. *Assume the following conditions hold.*

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

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

Then we have

$$\Pr_{(x,\ell)\leftarrow XL}(\Delta((\Pi U_\ell)^x, \Pi^x \otimes U_\ell) > \sqrt{\delta}) < 0.01. \tag{3}$$

Proof. Using (A1), we have

$$\begin{aligned}
\delta^2 2^c > \mathrm{IC}(\Pi) > \mathrm{IC}^T(\Pi) &= \mathbb{I}(UX : \Pi \mid YV) + \mathbb{I}(YV : \Pi \mid XU) && \text{(definition)}\\
&\geq \mathbb{I}(UX : \Pi \mid YV) && \text{(Fact 9.B: Nonnegativity)}\\
&\geq \mathbb{I}(U : \Pi \mid XYV) && \text{(Fact 9.D: Bar hopping)}\\
&= \mathbb{I}(U : \Pi YV \mid X) && \text{(Fact 9.D: Bar hopping)}\\
&\geq \mathbb{I}(U : \Pi \mid X) && \text{(Fact 9.C: Monotonicity)}\\
&= \mathbb{E}_{x\leftarrow X}\mathbb{I}(U : \Pi \mid X = x) && \text{(Definition 8: Mutual information)}\\
&= \mathbb{E}_{x\leftarrow X}\mathbb{I}(U_1^x \cdots U_{2^c}^x : \Pi^x) && \text{(notation)}\\
&\geq \mathbb{E}_{x\leftarrow X} \textstyle\sum_{\ell=1}^{2^c} \mathbb{I}(U_\ell^x : \Pi^x) && \text{(Fact 9.E: Independence)}\\
&= 2^c\, \mathbb{E}_{x\leftarrow X}\mathbb{E}_{\ell\leftarrow W}\mathbb{I}(U_\ell^x : \Pi^x). && \text{(W is the uniform distribution)}\\
\Rightarrow \quad \delta^2 &> \mathbb{E}_{(x,\ell)\leftarrow X\otimes W}\mathbb{I}(U_\ell^x : \Pi^x).\\
\Rightarrow \quad \delta &> \Pr_{(x,\ell)\leftarrow X\otimes W}(\mathbb{I}(U_\ell^x : \Pi^x) > \delta). && \text{(Fact 1: Markov's Inequality)}
\end{aligned}$$

We now want to replace the distribution $X \otimes W$ with XL on the right hand side. Since $\Delta(XL, X \otimes W) \leq c\delta/3$ from (A2), changing the distribution from $X \otimes W$ to XL only changes the probability of any event by $2c\delta/3 \leq c\delta$. Therefore

$$\begin{aligned}
0.01 > c\delta + \delta > &\Pr_{(x,\ell)\leftarrow XL}(\mathbb{I}(U_\ell^x : \Pi^x) > \delta)\\
&\geq \Pr_{(x,\ell)\leftarrow XL}(\Delta^2((\Pi U_\ell)^x, \Pi^x \otimes U_\ell^x) > \delta) && \text{(Fact 10: Relation between $\mathbb{I}$ and Δ)}\\
&= \Pr_{(x,\ell)\leftarrow XL}(\Delta^2((\Pi U_\ell)^x, \Pi^x \otimes U_\ell) > \delta) && \text{(U_ℓ is independent of X)}\\
&= \Pr_{(x,\ell)\leftarrow XL}(\Delta((\Pi U_\ell)^x, \Pi^x \otimes U_\ell) > \sqrt{\delta}). && \square
\end{aligned}$$

Claim 23. *Assume the following condition holds.*

$$\forall i \in [c] \quad \mathbb{I}(L_i : \Pi_i \mid X_i) \le \delta \tag{A3}$$

Then we have

$$\Pr_{(x,\ell)\leftarrow XL}(\Delta((\Pi U_\ell)^{x,\ell}, (\Pi U_\ell)^x) > 100\sqrt{c\delta}) \le 0.01. \tag{4}$$

Proof. We first show that (ΠU) together carries low information about L even conditioned on X. More precisely we show that

$$c\delta \ge \mathbb{I}(L : \Pi U \mid X). \tag{6}$$

This follows from the following chain on inequalities starting with (A3).

$$\begin{aligned}
\delta &\ge \mathbb{I}(L_i : \Pi_i \mid X_i) \\
&= \mathbb{I}(L_i : \Pi X_{-i} U Y_{-i} V \mid X_i) && \text{(definition of } \Pi_i) \\
&= \mathbb{I}(L_i : \Pi X_{-i} U Y_{-i} V X_{<i} Y_{<i} \mid X_i) && (X_{<i}Y_{<i} \text{ contained in } X_{-i}Y_{-i}) \\
&\ge \mathbb{I}(L_i : X_{<i} Y_{<i} \Pi U \mid X) && \text{(Fact 9.D: Bar hopping and Fact 9.C: Monotonicity)} \\
&\ge \mathbb{I}(L_i : L_{<i} \Pi U \mid X) && \text{(Fact 9.F: Data processing)} \\
&= \mathbb{I}(L_i : \Pi U \mid L_{<i} X). && \text{(Fact 9.D: Bar hopping)}
\end{aligned}$$

By summing this inequality over i, we get

$$\begin{aligned}
c\delta &\ge \textstyle\sum_{i=1}^{c} \mathbb{I}(L_i : \Pi U \mid L_{<i} X) \\
&= \mathbb{I}(L : \Pi U \mid X). && \text{(Fact 9.A: Chain rule)}
\end{aligned}$$

This is (6), which we set out to show. Using this inequality, we have

$$\begin{aligned}
& c\delta \ge \mathbb{I}(L : \Pi U \mid X) \\
& \quad = \mathbb{E}_{x\leftarrow X} \mathbb{I}(L : \Pi U \mid X = x) && \text{(Definition 8: Mutual information)} \\
& \quad = \mathbb{E}_{x\leftarrow X} \mathbb{I}(L^x : (\Pi U)^x) && \text{(notation)} \\
& \quad \ge \mathbb{E}_{(x,\ell)\leftarrow XL} \Delta^2((\Pi U)^{x,\ell}, (\Pi U)^x) && \text{(Fact 10: Relation between } \mathbb{I} \text{ and } \Delta) \\
& \quad \ge \mathbb{E}_{(x,\ell)\leftarrow XL} \Delta^2((\Pi U_\ell)^{x,\ell}, (\Pi U_\ell)^x). && \text{(Fact 4.C: Monotonicity)} \\
\Rightarrow \quad & \sqrt{c\delta} \ge \mathbb{E}_{(x,\ell)\leftarrow XL} \Delta((\Pi U_\ell)^{x,\ell}, (\Pi U_\ell)^x). && \text{(convexity of square)} \\
\Rightarrow \quad & 0.01 \ge \Pr_{(x,\ell)\leftarrow XL}(\Delta((\Pi U_\ell)^{x,\ell}, (\Pi U_\ell)^x) > 100\sqrt{c\delta}) && \text{(Fact 1: Markov's Inequality)} \quad (4)
\end{aligned}$$

This completes the proof. □

Claim 24. *Assume the following conditions hold.*

$$\Pr_{(x,\ell)\leftarrow XL}(\Delta((\Pi U_\ell)^x, \Pi^x \otimes U_\ell) > \sqrt{\delta}) < 0.01. \tag{3}$$

$$\Pr_{(x,\ell)\leftarrow XL}(\Delta((\Pi U_\ell)^{x,\ell}, (\Pi U_\ell)^x) > 100\sqrt{c\delta}) \le 0.01. \tag{4}$$

Then we have

$$\Pr_{(x,y,\ell,u_\ell,v_\ell)\leftarrow XYLU_LV_L}\left(\Delta(\Pi^{x,y,\ell,u_\ell,v_\ell}, \Pi^{x,y,\ell}) > 6 \cdot 10^6 \cdot \sqrt{c\delta}\right) < 0.09. \tag{5}$$

Proof. First, using (4) we can show

$$
\begin{aligned}
0.01 &\geq \Pr_{(x,\ell)\leftarrow XL}(\Delta((\Pi U_\ell)^{x,\ell}, (\Pi U_\ell)^x) > 100\sqrt{c\delta}) && (4)\\
&\geq \Pr_{(x,\ell)\leftarrow XL}(\Delta(\Pi^{x,\ell}, \Pi^x) > 100\sqrt{c\delta}) && \text{(Fact 4.C: Monotonicity)}\\
&= \Pr_{(x,\ell)\leftarrow XL}(\Delta(\Pi^{x,\ell}\otimes U_\ell, \Pi^x\otimes U_\ell) > 100\sqrt{c\delta}). && \text{(Fact 4.B: Product distributions)} \quad (7)
\end{aligned}
$$

Using (3), (4), and (7), the union bound and Fact 4.A (Triangle inequality) we get

$$
\begin{aligned}
0.03 &> \Pr_{(x,\ell)\leftarrow XL}\left(\Delta((\Pi U_\ell)^{x,\ell}, \Pi^{x,\ell}\otimes U_\ell) > 300\sqrt{c\delta}\right)\\
&= \Pr_{(x,\ell)\leftarrow XL}\left(\mathbb{E}_{\pi\leftarrow\Pi^{x,\ell}}(\Delta(U_\ell^{x,\ell,\pi}, U_\ell)) > 300\sqrt{c\delta}\right) && \text{(Fact 4.D: Partial measurement)}\\
&= \Pr_{(x,\ell)\leftarrow XL}\left(\mathbb{E}_{\pi\leftarrow\Pi^{x,\ell}}(\Delta(U_\ell^{x,\ell,\pi}\otimes Y^{x,\ell,\pi}, U_\ell\otimes Y^{x,\ell,\pi})) > 300\sqrt{c\delta}\right) && \text{(Fact 4.B)}\\
&= \Pr_{(x,\ell)\leftarrow XL}\left(\mathbb{E}_{\pi\leftarrow\Pi^{x,\ell}}(\Delta(U_\ell^{x,\ell,\pi} Y^{x,\ell,\pi}, U_\ell\otimes Y^{x,\ell,\pi})) > 300\sqrt{c\delta}\right) && \text{(Fact 15: Independence)}\\
&= \Pr_{(x,\ell)\leftarrow XL}\left(\Delta(U_\ell^{x,\ell} Y^{x,\ell}\Pi^{x,\ell}, U_\ell\otimes Y^{x,\ell}\Pi^{x,\ell}) > 300\sqrt{c\delta}\right) && \text{(Fact 4.D: Partial measurement)}\\
&= \Pr_{(x,\ell)\leftarrow XL}\left(\mathbb{E}_{y\leftarrow Y^{x,\ell}}(\Delta(U_\ell^{x,y,\ell}\Pi^{x,y,\ell}, U_\ell\otimes \Pi^{x,y,\ell})) > 300\sqrt{c\delta}\right) && \text{(Fact 4.D)}
\end{aligned}
$$

where the third equality follows since for all (x,ℓ), the variables $(U_\ell\Pi Y)^{x,\ell}$ form a Markov chain. To see this, fix x and ℓ, and consider giving Alice the input x together with an input distributed from $U^{x,\ell}$. Also, give Bob an input generated from $(YV)^{x,\ell}$. Since U is uniform and independent of everything else, Alice's input is independent of Bob's. Fact 15 (Independence) then implies that $U^{x,\ell}\leftrightarrow\Pi^{x,\ell}\leftrightarrow(YV)^{x,\ell}$ is a Markov chain. Then Fact 6 allows us to conclude $U_\ell^{x,\ell}\leftrightarrow\Pi^{x,\ell}\leftrightarrow Y^{x,\ell}$.

Next, using Fact 1 (Markov's Inequality), we get

$$
0.04 > \Pr_{(x,y,\ell)\leftarrow XYL}\left(\Delta((U_\ell\Pi)^{x,y,\ell}, U_\ell\otimes(\Pi^{x,y,\ell})) > 30000\sqrt{c\delta}\right). \tag{8}
$$

By symmetry between Alice and Bob, we get

$$
0.04 > \Pr_{(x,y,\ell)\leftarrow XYL}\left(\Delta((V_\ell\Pi)^{x,y,\ell}, V_\ell\otimes(\Pi^{x,y,\ell})) > 30000\sqrt{c\delta}\right). \tag{9}
$$

Using Eqs. (8) and (9) and the union bound we get

$$
\begin{aligned}
0.08 &> \Pr_{(x,y,\ell)\leftarrow XYL}\left(\Delta((U_\ell\Pi)^{x,y,\ell}, U_\ell\otimes(\Pi^{x,y,\ell})) + \Delta((V_\ell\Pi)^{x,y,\ell}, V_\ell\otimes(\Pi^{x,y,\ell})) > 60000\sqrt{c\delta}\right)\\
&\geq \Pr_{(x,y,\ell)\leftarrow XYL}\left(\Delta((U_\ell\Pi V_\ell)^{x,y,\ell}, U_\ell\otimes(\Pi^{x,y,\ell})\otimes V_\ell) > 60000\sqrt{c\delta}\right), && \text{(Fact 7)}
\end{aligned}
$$

where the last inequality used the fact that $(U_\ell\Pi V_\ell)^{x,y,\ell}$ is a Markov chain, which follows from a similar argument to before. Using Fact 4.D (Partial measurement) and Fact 1 (Markov's Inequality), we then get

$$
0.09 > \Pr_{(x,y,\ell,u_\ell,v_\ell)\leftarrow XYLU_LV_L}\left(\Delta(\Pi^{x,y,\ell,u_\ell,v_\ell}, \Pi^{x,y,\ell}) > 6\cdot 10^6\cdot\sqrt{c\delta}\right). \qquad \square
$$

Claim 25. *If we assume*

$$
\Pr_{(x,y,\ell,u_\ell,v_\ell)\leftarrow XYLU_LV_L}\left(\Delta(\Pi^{x,y,\ell,u_\ell,v_\ell}, \Pi^{x,y,\ell}) > 6\cdot 10^6\cdot\sqrt{c\delta}\right) < 0.09 \tag{5}
$$

then we have a contradiction.

Proof. Eq. (5) implies that there exists (x, y, ℓ) such that

$$0.09 > \Pr_{(u_\ell, v_\ell) \leftarrow U_\ell V_\ell} \left(\Delta(\Pi^{x,y,\ell,u_\ell,v_\ell}, \Pi^{x,y,\ell}) > 6 \cdot 10^6 \cdot \sqrt{c\delta} \right).$$

Recall that G_ℓ only depends on the XOR of the strings u_ℓ and v_ℓ. We assume without loss of generality that the number of strings $s \in \{0,1\}^m$ such that $G_\ell(x, u_\ell, y, v_\ell) = 1$ when $u_\ell \oplus v_\ell = s$ is at least the number of strings s for which $G_\ell(x, u_\ell, y, v_\ell) = 0$ when $u_\ell \oplus v_\ell = s$. A symmetric argument holds otherwise. This implies that there exists a string s such that $G_\ell(x, u_\ell, y, v_\ell) = 1$ whenever $u_\ell \oplus v_\ell = s$ and

$$0.18 > \Pr_{u_\ell \leftarrow U_\ell} \left(\Delta(\Pi^{x,y,\ell,u_\ell,u_\ell \oplus s}, \Pi^{x,y,\ell}) > 6 \cdot 10^6 \cdot \sqrt{c\delta} \right).$$

Fix any $t \in \{0,1\}^m$ such that $G_\ell(x, u_\ell, y, v_\ell) = 0$ whenever $u_\ell \oplus v_\ell = t$. The inequality above implies that there exists a string u_ℓ such that

$$\begin{aligned} &6 \cdot 10^6 \cdot \sqrt{c\delta} \geq \Delta(\Pi^{x,y,\ell,u_\ell,u_\ell \oplus s}, \Pi^{x,y,\ell}) \\ \text{and} \quad &6 \cdot 10^6 \cdot \sqrt{c\delta} \geq \Delta(\Pi^{x,y,\ell,u_\ell \oplus t \oplus s,u_\ell \oplus t}, \Pi^{x,y,\ell}). \end{aligned}$$

Using Fact 4.A (Triangle inequality) we get

$$\begin{aligned} 0.001 \geq 12 \cdot 10^6 \cdot \sqrt{c\delta} &\geq \Delta(\Pi^{x,y,\ell,u_\ell,u_\ell \oplus s}, \Pi^{x,y,\ell,u_\ell \oplus s \oplus t,u_\ell \oplus t}) \\ &\geq \mathrm{h}^2(\Pi^{x,y,\ell,u_\ell,u_\ell \oplus s}, \Pi^{x,y,\ell,u_\ell \oplus s \oplus t,u_\ell \oplus t}) && \text{(Fact 3: Relation between } \Delta \text{ and h)} \\ &= \mathrm{h}^2((\Pi^{x,y,\ell})^{u_\ell,u_\ell \oplus s}, (\Pi^{x,y,\ell})^{u_\ell \oplus s \oplus t,u_\ell \oplus t}) && \text{(notation)} \\ &\geq \frac{1}{2}\mathrm{h}^2((\Pi^{x,y,\ell})^{u_\ell,u_\ell \oplus s}, (\Pi^{x,y,\ell})^{u_\ell \oplus s \oplus t,u_\ell \oplus s}) && \text{(Fact 16: Pythagorean property)} \\ &\geq \frac{1}{4}\Delta^2(\Pi^{x,y,\ell,u_\ell,u_\ell \oplus s}, \Pi^{x,y,\ell,u_\ell \oplus s \oplus t,u_\ell \oplus s}) && \text{(Fact 3: Relation between } \Delta \text{ and h)} \\ \Rightarrow 0.1 > \sqrt{0.004} &> \Delta(\Pi^{x,y,\ell,u_\ell,u_\ell \oplus s}, \Pi^{x,y,\ell,u_\ell \oplus s \oplus t,u_\ell \oplus s}). \end{aligned}$$

This implies that the worst case error of protocol Π is at least $0.5 - 0.1 > 1/3$. This is a contradiction because Π was assumed to have error less than $1/3$. ☐

4 Quantum vs. randomized communication

The goal of this section is to prove Theorem 1, which we restate for convenience:

Theorem 1. *There exists a total function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ with* $\mathrm{R}(F) = \widetilde{\Omega}(\mathrm{Q}(F)^{2.5})$.

Overview. We start in the world of query complexity. Let $\textsc{Tr}_{n^2} = \textsc{Or}_n \circ \textsc{And}_n$ denote the Tribes function on n^2 input bits. We study a partial function on n^3 bits:

$$\textsc{Str} \coloneqq \textsc{Simon}_n \circ \textsc{Tr}_{n^2} = \textsc{Simon}_n \circ \textsc{Or}_n \circ \textsc{And}_n. \tag{10}$$

Here $\textsc{Simon}_n$ is a certain property testing version of Simon's problem [Sim97] introduced in [BFNR08] (defined in Section 4.1 below) which witnesses a large gap between its randomized $\mathrm{R}^{\mathrm{dt}}(\textsc{Simon}_n) = \Omega(\sqrt{n})$ and quantum $\mathrm{Q}^{\mathrm{dt}}(\textsc{Simon}_n) = O(\log n \log\log n)$ query complexities. As shown in [ABK16, §3], the cheat sheet version of $\textsc{Str}$ witnesses an $\widetilde{O}(n)$-vs-$\widetilde{\Omega}(n^{2.5})$ separation between quantum and randomized query complexities. (Actually, they use $\textsc{Forrelation}$ [AA15] in place of $\textsc{Simon}$, but we find it more convenient to work with $\textsc{Simon}$.)

We follow a similar approach to the query case and first "lift" $\textsc{Str}$ to a partial two-party function $F = \textsc{Str} \circ \mathrm{IP}_b$ by composing it with IP_b, the two-party inner-product function on $b = \Theta(\log n)$ bits per party. Our final function achieving the desired separation will be a $(F, \mathcal{G})$-lookup function $F_\mathcal{G}$ where $\mathcal{G}$ forms a consistent family of nontrivial XOR functions, described in Section 4.4. By Theorem 5, to show a lower bound on the randomized communication complexity of $F_\mathcal{G}$, it suffices to show a randomized communication lower bound on $F = \textsc{Str} \circ \mathrm{IP}_b$. We show the following.

Theorem 6. *Let $b \geq c \log n$ for a sufficiently large constant c. Then $\mathrm{R}(\textsc{Str} \circ \mathrm{IP}_b) \geq \widetilde{\Omega}(n^{2.5})$.*

The basic idea of the proof of Theorem 6 is to use the query-to-communication lifting theorem of [GLM+15], which requires us to show a lower bound on the approximate *conical junta degree* of $\textsc{Str}$ (see Section 4.2 for definitions). For this, we would like to show that each of $\textsc{Simon}_n$, $\textsc{Or}_n$, $\textsc{And}_n$ individually have large junta degree and then invoke a *composition theorem* for conical junta degree [GJ16]. Because of certain technical conditions in the composition theorem, we will actually need to show a lower bound on the functions $\textsc{Simon}_n$, $\textsc{Or}_n$, $\textsc{And}_n$ in a slightly stronger model, as discussed in Section 4.3. This will prove Theorem 6.

The other half of Theorem 1 is a quantum upper bound on the communication complexity of $F_\mathcal{G}$, for a particular family of functions $\mathcal{G}$. We need that the family $\mathcal{G}$ is consistent outside F, and that each function $G_i \in G$ has $\mathrm{Q}(G_i) = \widetilde{O}(n)$. We do this in a way very analogous to the cheat sheet framework: each function $G_i(\mathbf{x}, u, \mathbf{y}, v)$ evaluates to 1 if and only if $u \oplus v$ verifies that $(x_i, y_i) \in \mathrm{dom}(F)$ for all $i \in [c]$. The players check this using a distributed version of Grover search. The formal definition of $F_\mathcal{G}$ and the upper bound on its quantum communication complexity appear in Section 4.4.

4.1 Simon's problem

The partial function $\textsc{Simon}_n \colon \{0,1\}^n \to \{0,1,*\}$ is defined as follows. (For convenience, we actually use the negation of the function defined in [BFNR08].) We interpret the input $z \in \{0,1\}^n$ as a function $z \colon \mathbb{Z}_2^d \to \{0,1\}$ where $d = \log n$ (we tacitly assume that n is a power of 2, which can be achieved by padding). Call a function z *periodic* if there is some nonzero $y \in \mathbb{Z}_2^d$ such that $z(x+y) = z(x)$ for all $x \in \mathbb{Z}_2^d$. Furthermore, z is *far from periodic* if the Hamming distance between z and every periodic function is at least $n/8$. Then

$$\textsc{Simon}_n(z) \coloneqq \begin{cases} 1 & \text{if } z \text{ is far from periodic,} \\ 0 & \text{if } z \text{ is periodic,} \\ * & \text{otherwise.} \end{cases}$$

The key properties of this function, proved in [BFNR08, §4], are:

- *Quantum query complexity:* $\mathrm{Q}^{\mathrm{dt}}(\textsc{Simon}_n) \leq O(\log n \log\log n)$.
- *Randomized query complexity:* $\mathrm{R}^{\mathrm{dt}}(\textsc{Simon}_n) \geq \Omega(\sqrt{n})$.

Moreover, it is important for us that the randomized lower bound is robust: it is witnessed by a pair of distributions (D_1, D_0) where D_i is supported on $(\textsc{Simon}_n)^{-1}(i)$ such that any small-width conjunction that accepts under D_1 also accepts under D_0 with comparable probability. We formalize this property in the following lemma; for completeness, we present its proof (which is implicit in [BFNR08, §4]). One subtlety is that the property is *one-sided* in that the statement becomes false if we switch the roles of D_1 and D_0.

Lemma 26 ([BFNR08])**.** *Let $\alpha := \sqrt{n}/2$. There exists a pair of distributions (D_1, D_0) where D_i is supported on $\textsc{Simon}_n^{-1}(i)$ such that for every conjunction C with $|C|$ literals,*

$$\Pr_{z \leftarrow D_0}[C(z) = 1] \geq (1 - |C|/\alpha) \cdot \Pr_{z \leftarrow D_1}[C(z) = 1]. \tag{11}$$

Proof. Assume $1 \leq |C| \leq \alpha$ for otherwise the claim is trivial. Define U and D_1 as the uniform distributions on $\{0,1\}^n$ and $\textsc{Simon}_n^{-1}(1)$, respectively. Define a distribution D_0 on periodic functions z as follows: choose a nonzero period $y \in \mathbb{Z}_2^d$ uniformly at random, and for every edge of the matching $(x, x+y)$ in $\mathbb{Z}_2^d$ uniformly choose $b \in \{0,1\}$ and set $b = z(x) = z(x+y)$.

Intuitively, C can distinguish between $z \leftarrow D_0$ and a uniformly random string only if C queries two input vectors whose difference is the hidden period y that was used to generate z. Indeed, let $\mathcal{S} \subseteq \mathbb{Z}_2^d$, $|\mathcal{S}| \leq \binom{|C|}{2}$, be the set of vectors of the form $x + x'$ where C queries both $z(x)$ and $z(x')$. Then, conditioned on the event "$y \notin \mathcal{S}$", the bits C reads from z are uniformly random. Hence

$$\begin{aligned}
\Pr_{z \leftarrow D_0}[C(z) = 1] &\geq \Pr_{z \leftarrow D_0}[y \notin \mathcal{S} \wedge C(z) = 1] \\
&= \Pr_{z \leftarrow D_0}[y \notin \mathcal{S}] \cdot \Pr_{z \sim D_0}[C(z) = 1 \mid y \notin \mathcal{S}] \\
&\geq \left(1 - \tbinom{|C|}{2}/(n-1)\right) \cdot 2^{-|C|} \\
&\geq \left(1 - |C|/\sqrt{n}\right) \cdot 2^{-|C|},
\end{aligned} \tag{12}$$

where the last inequality holds because $|C| \leq \alpha$.

Since there are at most $n2^{n/2}$ periodic functions, there are at most $n2^{n/2} \cdot 2^{nH(1/8)} \leq 2^{2n/3}$ functions at Hamming distance $\leq n/8$ from periodic functions (here H is the binary entropy function). Hence the total variation distance between U and D_1 is tiny: $\Delta(U, D_1) \leq 2^{-\Omega(n)}$. Thus

$$\Pr_{z \leftarrow D_1}[C(z) = 1] \leq 2^{-|C|} + 2^{-\Omega(n)} \leq (1 + 2^{-\Omega(n)}) \cdot 2^{-|C|}. \tag{13}$$

Putting (12) and (13) together we get

$$\begin{aligned}
\Pr_{z \leftarrow D_0}[C(z) = 1] &\geq \left(1 - |C|/\sqrt{n}\right)(1 - 2^{-\Omega(n)}) \cdot \Pr_{z \leftarrow D_1}[C(z) = 1] \\
&\geq \left(1 - 2|C|/\sqrt{n}\right) \cdot \Pr_{z \leftarrow D_1}[C(z) = 1].
\end{aligned}$$

□

4.2 Query-to-communication

To prove the randomized lower bound $\mathrm{R}(\textsc{Str} \circ \mathrm{IP}_b) \geq \widetilde{\Omega}(n^{2.5})$ in Theorem 6 we apply the main result of [GLM+15]: a simulation of randomized communication protocols by *conical juntas*—a nonnegative analogue of multivariate polynomials.

Conical juntas. A *conical junta* h is a nonnegative linear combination of conjunctions; more precisely, $h = \sum_C w_C C$ where $w_C \geq 0$ and the sum ranges over all conjunctions $C\colon \{0,1\}^n \to \{0,1\}$ of literals (input bits or their negations). For a conjunction C we let $|C|$ denote its width, i.e., the number of literals in C. The *conical junta degree* of h, denoted $\deg^+(h)$, is the maximum width of a conjunction C with $w_C > 0$. Any conical junta h naturally computes a nonnegative function $h\colon \{0,1\}^n \to \mathbb{R}_{\geq 0}$. For a partial boolean function $f\colon \{0,1\}^n \to \{0,1,*\}$ we say that h *ε-approximates* f if and only if $|f(x) - h(x)| \leq \varepsilon$ for all inputs $x \in \mathrm{dom}(f)$. The *ε-approximate conical junta degree of* f, denoted $\deg^+_\varepsilon(f)$, is defined as the minimum degree of a conical junta h that ε-approximates f. (Note: $\deg^+_\varepsilon(f)$ is also known as the query complexity analogue of the one-sided smooth rectangle bound [JK10].)

Lifting theorem. The following theorem is a corollary of [GLM$^+$15, Theorem 31] (originally, the theorem was stated for constant $\varepsilon > 0$, but the theorem holds more generally for $\varepsilon = 2^{-\Theta(b)}$; note also that instead of $\deg^+_\varepsilon(f)$, that paper uses the notation $\mathrm{WAPP}^{\mathrm{dt}}_\varepsilon(f)$). Here $\mathrm{IP}_b\colon \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ is the two-party inner-product function defined by $\mathrm{IP}_b(x,y) \coloneqq \langle x,y\rangle \bmod 2$.

Theorem 7 (Lifting theorem [GLM$^+$15])**.** *For any $\varepsilon > 0$ define $b \coloneqq \Theta(\log(n/\varepsilon))$ (with a large enough implicit constant). For every partial $f\colon \{0,1\}^n \to \{0,1,*\}$ we have*

$$\mathrm{R}_\varepsilon(f \circ \mathrm{IP}_b) \geq \Omega(\deg^+_\varepsilon(f) \cdot b).$$

Hence, to prove Theorem 6, it remains to show for some not-too-small $\varepsilon \geq 1/\operatorname{poly}(n)$ that

$$\deg^+_\varepsilon(\textsc{Str}) \geq \Omega(n^{2.5}). \tag{14}$$

This will be done in the next section. (Recall from Fact 13 that the error parameter for the measure R_ε can be reduced from any such ε to a constant at the cost of some $O(\log n)$ factor loss.)

4.3 Query lower bound

To prove (14) we use a composition theorem for conical junta degree due to [GJ16]. Ideally, we would like to conclude the $\Omega(n^{2.5})$ lower bound from the facts that $\textsc{Simon}_n$, $\textsc{Or}_n$, $\textsc{And}_n$ (and some of their negations) have approximate conical junta degrees $\Omega(\sqrt{n})$, $\Omega(n)$, $\Omega(n)$, respectively. These facts are indeed implicit in existing literature; for example:

- The result of [BFNR08] recorded in Lemma 26 implies $\deg^+_{1/3}(\textsc{Simon}_n) \geq \Omega(\sqrt{n})$.
- Klauck [Kla10] has proved even a communication analogue of $\deg^+_\varepsilon(\textsc{Or}_n) \geq \Omega(n)$. (For an exposition of the query version, see, e.g., [GJPW15, §4.1].)
- Jain and Klauck [JK10, §3.3] proved that $\deg^+_{1/16}(\textsc{Tr}_{n^2}) \geq \Omega(n^2)$.

Unfortunately, the composition theorem from [GJ16] assumes some regularity conditions from the *dual certificates* witnessing these lower bounds. (In fact, without regularity assumptions, a composition theorem for a related "average conical junta degree" measure is known to fail! See [GJ16, §3] for a discussion.) It is unsurprising that suitable dual certificates of a special form can be found (e.g., inspired by the above results). We now go through these straightforward verifications.

Composition theorem. We recall the necessary definitions from [GJ16] in order to state the composition theorem precisely. The theorem was originally phrased for total functions, but the result holds more generally for partial functions f provided the dual certificates are supported on the domain of f. The following definitions make these provisions.

A function $\Psi\colon \{0,1\}^n \to \mathbb{R}$ is *balanced* if $\sum_x \Psi(x) = 0$. Write $X_{\geq 0} \coloneqq \max\{0, X\}$ for short. A *two-sided (α,β)-certificate* for a partial function $f\colon \{0,1\}^n \to \{0,1,*\}$ consists of four balanced functions $\{\Psi_v, \hat{\Psi}_v\}_{v=0,1}$ satisfying the following:

- *Special form:* There exist distributions D_1 over $f^{-1}(1)$ and D_0 over $f^{-1}(0)$ such that $\Psi_1 = \alpha \cdot (D_1 - D_0)$ and $\Psi_0 = -\Psi_1$. Moreover, $\hat{\Psi}_v$ is supported on $f^{-1}(v)$.
- *Bounded 1-norm:* For each $v \in \{0,1\}$ we have $\|\hat{\Psi}_v\|_1 \leq \beta$.
- *Feasibility:* For all conjunctions C and $v \in \{0,1\}$,

$$\langle \Psi_v, C\rangle_{\geq 0} + \langle \hat{\Psi}_v, C\rangle \leq |C|\langle D_v, C\rangle. \tag{15}$$

We also define a *one-sided* (α, β)*-certificate* for f as a pair of balanced functions $\{\Psi_1, \hat{\Psi}_1\}$ that satisfies the above conditions but only for $v = 1$.

Theorem 8 (Composition theorem [GJ16]). *Suppose f admits a two-sided (resp. one-sided) (α_1, β_1)-certificate, and g admits a two-sided (α_2, β_2)-certificate. Then $f \circ g$ admits a two-sided (resp. one-sided) $(\alpha_1\alpha_2, \beta_1 + n\beta_2)$-certificate.*

Lemma 27 (Degree lower bounds from certificates [GJ16]). *Suppose f admits a one-sided (α, β)-certificate. Then $\deg^+_\varepsilon(f) \geq \Omega(\alpha)$ provided $\varepsilon < 1/4$ and $\varepsilon\beta \leq \alpha/4$.*

Certificates for Simon**.** For $\textsc{Simon}_n$, a one-sided $(\alpha, 0)$-certificate, $\alpha \coloneqq \sqrt{n}/2$, is given by

$$\Psi_1 \coloneqq \alpha \cdot (D_1 - D_0), \qquad\qquad \hat{\Psi}_1 \coloneqq 0,$$

where (D_1, D_0) are from Lemma 26. Note that (11) can be rephrased as $\langle D_0, C\rangle \geq (1-|C|/\alpha)\langle D_1, C\rangle$ since $\langle D_v, C\rangle = \Pr_{z\sim D_v}[D(z) = 1]$. The feasibility condition (15) follows:

$$\begin{aligned}
\langle \Psi_1, C\rangle_{\geq 0} + \langle \hat{\Psi}_1, C\rangle &= \langle \Psi_1, C\rangle \\
&= \alpha\langle D_1, C\rangle - \alpha\langle D_0, C\rangle \\
&\leq \alpha\langle D_1, C\rangle - \alpha(1 - |C|/\alpha)\langle D_1, C\rangle \\
&= |C|\langle D_1, C\rangle.
\end{aligned}$$

Certificates for OR. For $\textsc{Or}_n$, a two-sided $(n/2, n)$-certificate is given by

$$\begin{aligned}
\Psi_1 &\coloneqq n/2 \cdot (D_1 - D_0), &\qquad \hat{\Psi}_1 &\coloneqq n/2 \cdot (D_1 - D_2),\\
\Psi_0 &\coloneqq n/2 \cdot (D_0 - D_1), &\qquad \hat{\Psi}_0 &\coloneqq 0,
\end{aligned}$$

where D_i is the uniform distribution on inputs of Hamming weight i. To check the feasibility conditions (15), we split into cases depending on how many positive literals C contains. For notation, let C_j be a conjunction of width $w \coloneqq |C|$ having j positive literals (and thus $w - j$ negative literals). We have the following table of acceptance probabilities:

j	$\langle D_0, C_j\rangle$	$\langle D_1, C_j\rangle$	$\langle D_2, C_j\rangle$
0	1	$(n-w)/n$	$\binom{n-w}{2}/\binom{n}{2}$
1	0	$1/n$	$(n-w)/\binom{n}{2}$
2	0	0	$1/\binom{n}{2}$
≥ 3	0	0	0

For $v = 1$, it suffices to consider $j \in \{0, 1\}$ since any C_j with $j > 1$ will have $\langle D_1, C_j\rangle = 0$ and hence $\langle \Psi_1, C_j\rangle, \langle \hat{\Psi}_1, C_j\rangle \leq 0$. For $v = 0$, it suffices to consider $j = 0$ since any C_j with $j > 0$ will have $\langle D_0, C_j\rangle = 0$ and hence $\langle \Psi_0, C_j\rangle \leq 0$. Here we go:

$$\begin{aligned}
\langle \Psi_1, C_0\rangle_{\geq 0} + \langle \hat{\Psi}_1, C_0\rangle &= 0 + n/2 \cdot \langle D_1 - D_2, C_0\rangle \\
&= n/2 \cdot \left(\tfrac{n-w}{n} - \tbinom{n-w}{2}/\tbinom{n}{2}\right) = n/2 \cdot \left(\tfrac{n-w}{n}(1 - \tfrac{n-w-1}{n-1})\right) \\
&= n/2 \cdot \left(\tfrac{n-w}{n} \cdot \tfrac{w}{n-1}\right) = 1/2 \cdot (n-w) \cdot \tfrac{w}{n-1} \\
&\leq (n-w) \cdot \tfrac{w}{n} = w\langle D_1, C_0\rangle,
\end{aligned}$$

$$
\begin{aligned}
\langle \Psi_1, C_1\rangle_{\geq 0} + \langle \hat{\Psi}_1, C_1\rangle &= n/2\cdot\langle D_1, C_1\rangle + n/2\cdot\langle D_1 - D_2, C_1\rangle \\
&= n\cdot\langle D_1, C_1\rangle - n/2\cdot\langle D_2, C_1\rangle \\
&= 1 - n/2\cdot(n-w)/\tbinom{n}{2} = 1 - \tfrac{n-w}{n-1} = \tfrac{w-1}{n-1} \\
&\leq w/n = w\langle D_1, C_1\rangle,
\end{aligned}
$$

$$
\begin{aligned}
\langle \Psi_0, C_0\rangle_{\geq 0} + \langle \hat{\Psi}_0, C_0\rangle &= \langle \Psi_0, C_0\rangle = n/2\cdot\langle D_0 - D_1, C_0\rangle \\
&= n/2\cdot(1-(n-w)/n) = w/2 \\
&\leq w = w\langle D_0, C_0\rangle.
\end{aligned}
$$

Proof of Theorem 6. For $\textsc{And}_n$, a two-sided $(n/2, n)$-certificate can be obtained similarly to the above (since $\textsc{And}_n$ and $\textsc{Or}_n$ are duals of one another). We can now use Theorem 8 twice to compose $(\Omega(\sqrt{n}), 0)$-, $(n/2, n)$-, $(n/2, n)$-certificates for $\textsc{Simon}_n$, $\textsc{Or}_n$, $\textsc{And}_n$, respectively, to end up with a one-sided $(\Omega(n^{2.5}), O(n^3))$-certificate for $\textsc{Str}$. Lemma 27 now yields (14) for $\varepsilon \coloneqq \Theta(1/\sqrt{n})$. Hence we can apply Theorem 7 to complete the proof of Theorem 6. ☐

4.4 Separation

We now explicitly define the $(F, \mathcal{G})$-lookup function we will use for our quantum vs. randomized communication complexity separation. Let $F = \textsc{Str} \circ \mathrm{IP}_b$ as defined in Equation 10, for $b = \Theta(\log n)$. Let $c = 10 \log n$. The definition of the family of functions $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$, closely resembles the construction of cheat sheet functions. The most difficult property to achieve is to make $\mathcal{G}$ consistent outside F. We do this by defining $G_i(\mathbf{x}, u, \mathbf{y}, v)$ to be 1 if and only if $u \oplus v$ certifies that each (x_i, y_i) is in the domain of F (all functions G_i will be the same). This condition naturally enforces consistency outside F. We further require that $u \oplus v$ certifies this in a very specific fashion. This is done so that the players can check $u \oplus v$ has the required properties efficiently using a distributed version of Grover's search algorithm.

We first define a helper function which will be like G_i but just works to certify that a single copy (x_j, y_j) of the input is in $\mathrm{dom}(F)$. Let

$$
P\colon \left(\{0,1\}^{bn^3} \times \{0,1\}^{n(n\log n+1)}\right) \times \left(\{0,1\}^{bn^3} \times \{0,1\}^{n(n\log n+1)}\right) \to \{0,1\}.
$$

This function will be defined such that $P(x, u, y, v) = 1$ if and only if $(x, y) \in \mathrm{dom}(F)$ is witnessed by $u \oplus v$ in a specific fashion, described next. Decompose $x \in \{0,1\}^{bn^3}$ as $x = (x_{i,j,k})_{i,j,k\in[n]}$ where each $x_{ijk} \in \{0,1\}^b$, and similarly for y. Let $z_{ijk} = \mathrm{IP}_b(x_{ijk}, y_{ijk})$ for $i, j, k \in [n]$, and $z_i = \textsc{Or}_n \circ \textsc{And}_n(z_{i11}, \ldots, z_{inn})$ for $i \in [n]$. Now (x, y) will be in the domain of F if and only if $(z_1, \ldots, z_n)$ is in the domain of $\textsc{Simon}_n$.

If the players know $(z_1, \ldots, z_n)$ then they can easily verify if it is in $\mathrm{dom}(\textsc{Simon}_n)$. Let $w = u \oplus v$ and decompose this as $w = (q, \mathrm{C})$, where $q \in \{0,1\}^n$ and $\mathrm{C} = (C_1, \ldots, C_n)$ with each $C_i \in [n]^n$. Intuitively, q can be thought of as the purported value of $(z_1, \ldots, z_n)$, and C_i as a "certificate" that $q_i = z_i$. The function evaluates to 1 if these certificates check out.

Formally, $P(x, u, y, v) = 1$ if and only if

1. $q \in \mathrm{dom}(\textsc{Simon}_n)$
2. for all $i \in [n]$: if $q_i = 1$ then $C_i = (j, 0, \ldots, 0)$ and $z_{ijk} = 1$ for all $k \in [n]$, and if $q_i = 0$ then $C_i = (t_1, \ldots, t_n)$ and $z_{ijt_j} = 0$ for all $j \in [n]$.

Note that (2) ensures that if $P(x, u, y, v)$ accepts then $z_i = q_i$ for all $i \in [n]$.

Finally, we can define G_i for $i \in \{0, \ldots, 2^c - 1\}$: $G_i(\mathbf{x}, u_1, \ldots, u_c, \mathbf{y}, v_1, \ldots, v_c) = 1$ if and only if $P((x_j, u_j), (y_j, v_j)) = 1$ for all $j \in [c]$.

Claim 28. *The family of functions $\mathcal{G}$ defined above is consistent outside of F and is a nontrivial XOR function.*

Proof. Each G_i is an XOR function by definition. Also, if $F^c(\mathbf{x},\mathbf{y}) \notin \{0,1\}^c$ because (say) $(x_j, y_j) \notin \mathrm{dom}(F)$, then $P((x_j,u),(y_j,v))$ will always evaluate to 0 no matter what u, v. This is because $P((x_j,u),(y_j,v))$ can only evaluate to 1 if $u \oplus v = (q, \mathrm{C})$ where C certifies that $z_i = q_i$ for all $i \in [n]$ as in item (2) above. If this holds, then P will reject when $q = (z_1,\ldots,z_n) \notin \mathrm{dom}(F)$. This means $\mathcal{G}$ is consistent outside F.

Finally, let $(\mathbf{x},\mathbf{y}) \in \mathrm{dom}(F^c)$. Then there will exist u, v such that $u \oplus v$ provides correct certificates of this, and u', v' providing incorrect certificates. Thus each G_i is nontrivial. $\square$

To complete the separation, we need some known results on the behavior of quantum query and communication complexity under composition.

Fact 29 (Composition of quantum query complexity [Rei11])**.** *Let $f : \{0,1\}^n \to \{0,1,*\}$ and $g : \{0,1\}^m \to \{0,1\}$. Then $\mathrm{Q}^{\mathrm{dt}}(f \circ g^n) = O(\mathrm{Q}^{\mathrm{dt}}(f)\,\mathrm{Q}^{\mathrm{dt}}(g))$.*

Fact 30 (Composition with query function [BCW98])**.** *Let $f\colon \{0,1\}^n \to \{0,1,*\}$ be a (partial) function. For $i \in [n]$, let $F_i\colon \mathcal{X} \times \mathcal{Y} \to \{0,1,*\}$ be a communication problem. Then, $\mathrm{Q}(f \circ (F_1,\ldots,F_n)) = O(\mathrm{Q}^{\mathrm{dt}}(f) \log \mathrm{Q}^{\mathrm{dt}}(f) \cdot \max_i \mathrm{Q}(F_i) \log n)$.*

We can now finish the separation.

Theorem 9. *Let $F = \textsc{Str} \circ \mathrm{IP}_b$ be defined as in (10) for $b = \Theta(\log n)$, $\mathcal{G}$ be the family of functions defined above, and $F_{\mathcal{G}}$ be the $(F,\mathcal{G})$-lookup function. Then $F_{\mathcal{G}}$ is a total function satisfying*

$$\mathrm{Q}(F_{\mathcal{G}}) = \widetilde{O}(bn) = \widetilde{O}(n) \qquad \text{and} \qquad \mathrm{R}(F_{\mathcal{G}}) = \widetilde{\Omega}(n^{2.5}).$$

Proof. We start with the randomized lower bound. As $c = 10 \log n \geq \mathrm{R}(F)$ we can apply Theorem 5 to obtain $\mathrm{R}(F_{\mathcal{G}}) = \widetilde{\Omega}(\mathrm{R}(F)) = \widetilde{\Omega}(n^{2.5})$ by Theorem 6.

Now we turn to the quantum upper bound. By Theorem 4 it suffices to show $\mathrm{Q}(F) = \widetilde{O}(bn)$ and $\max_s \mathrm{Q}(G_s) = \widetilde{O}(bn)$. As $\mathrm{Q}^{\mathrm{dt}}(\textsc{Simon}_n) = O(\log n \log\log n)$ and $\mathrm{Q}^{\mathrm{dt}}(\textsc{Or}_n \circ \textsc{And}_n) = O(n)$, by the composition theorem Fact 29 $\mathrm{Q}(\textsc{Str}) = \widetilde{O}(n)$. Thus $\mathrm{Q}(F) = \widetilde{O}(bn)$ by Fact 30, as $\mathrm{Q}(\mathrm{IP}_b) \leq b$.

We now turn to show $\max_s \mathrm{Q}(G_s) = \widetilde{O}(bn)$. Fix s and let the input to G_s be $(\mathbf{x},\mathbf{u},\mathbf{y},\mathbf{v})$. For each $\ell \in [c]$ the players do the following procedure to evaluate $P(x_\ell, u_\ell, y_\ell, v_\ell)$. For ease of notation, fix ℓ and let $x = x_\ell, y = y_\ell, u = u_\ell, v = v_\ell$. As above, let $x = (x_{i,j,k})_{i,j,k \in [n]}$ where each $x_{ijk} \in \{0,1\}^b$ and similarly for y, $z_{ijk} = \mathrm{IP}_b(x_{ijk}, y_{ijk})$ for $i,j,k \in [n]$, and $z_i = \textsc{Or}_n \circ \textsc{And}_n(z_{i11},\ldots,z_{inn})$ for $i \in [n]$. Also let $w = u \oplus v$ and $w = (q, \mathrm{C})$ where $\mathrm{C} = (C_1,\ldots,C_n)$ and each $C_i \in [n]^n$. We will further decompose $C_i = (C_{i1},\ldots,C_{in})$.

Alice and Bob first exchange n bits to learn q. If $q \notin \mathrm{dom}(\textsc{Simon}_n)$ they reject. Otherwise, they proceed to check property (2) above, that C_i certifies that $q_i = z_i$ for all $i \in [n]$. They view this as a search problem on n^2 items $g_{i,t} \in \{0,1\}$ for $i,t \in [n]$. If $q_i = 1$ then $g_{i,t} = 1$ if and only if $z_{itC_{it}} = 1$. If $q_i = 0$ then $g_{i,t} = 1$ if and only if $z_{itC_{it}} = 0$. Then (x,u,y,v) satisfy property (2) in the definition of P if and only if $g_{i,t} = 1$ for all $i,t \in [n]$. Each $g_{i,t}$ can be evaluated using $O(b + \log n)$ bits of communication. Hence, using Grover search and Fact 30, it takes $\widetilde{O}(bn)$ qubits of quantum communication to verify that all $g_{i,t} = 1$. $\square$

5 Partitions vs. randomized communication

In this section, we prove Theorem 2, which we restate for convenience:

Theorem 2. *There exists a total function $F\colon \mathcal{X}\times\mathcal{Y}\to\{0,1\}$ with $\mathrm{R}(F)\geq \mathrm{UN}(F)^{2-o(1)}$.*

The proof closely follows the analogous result obtained for query complexity in [AKK16] using the cheat sheet framework. For a total communication function F, we will define a special case of $(F,\mathcal{G})$-lookup functions that are a communication analog of cheat sheets in query complexity.

Definition 31 (Cheat sheets for total functions)**.** Let $F\colon \mathcal{X}\times\mathcal{Y}\to\{0,1\}$ be a total function. Fix a cover $\mathcal{R}=\{R_0,\dots,R_{2^{\mathrm{N}(F)}-1}\}$ of $\mathcal{X}\times\mathcal{Y}$ by rectangles monochromatic for F. Let $N=\min\{\log|\mathcal{X}|,\log|\mathcal{Y}|\}$ and $c=10\log N$. Define a function

$$G:(\mathcal{X}^c\times\{0,1\}^{c\mathrm{N}(F)})\times(\mathcal{Y}^c\times\{0,1\}^{c\mathrm{N}(F)})\to\{0,1\}$$

where $G(x_1,\dots,x_c,a_1,\dots,a_c,y_1,\dots,y_c,b_1,\dots,b_c)=1$ if and only if $(x_i,y_i)\in R_{a_i\oplus b_i}$ for all $i=1,\dots,c$. The *cheat sheet* function F_{CS} of F is the $(F,\{G_0,\dots,G_{2^c-1}\})$ lookup function where $G_i=G$ for all i. In other words, $F_{\mathrm{CS}}(x_1,\dots,x_c,u_0,\dots,u_{2^c-1},y_1,\dots,y_c,v_0,\dots,v_{2^c-1})$ evaluates to $G(x_1,\dots,x_c,u_\ell,y_1,\dots,y_c,v_\ell)$, where $\ell=(F(x_1,y_1),\dots,F(x_c,y_c))$.

Remark 32. Note that F_{CS} is in particular a $(F,\mathcal{G})$-lookup function where $\mathcal{G}$ is a nontrivial $\textsc{Xor}$ family (Definition 21), thus Theorem 5 applies. Further letting $\mathcal{X}'\times\mathcal{Y}'$ be the domain of F_{CS}, note that $N'=\min\{\log|\mathcal{X}'|,\log|\mathcal{Y}'|\}=O(cN+c\cdot 2^c\mathrm{N}(F))=O(N^{12})$.

Recall that the function $\textsc{Tr}_{n^2}$ on n^2 input bits is the composition $\textsc{Or}_n\circ\textsc{And}_n$. The separating function of Theorem 2 is constructed by starting with disjointness on n variables and alternately taking the cheat sheet function of it and composing $\textsc{Tr}_{n^2}$ with it. Repeating this process gives a function with a larger and larger gap between R and UN, converging to a quadratic gap between these measures.

To prove this result, we first need to understand how the composition operations affect R and UN. We start with UN, for which we wish to prove an upper bound.

Lemma 33 (AND/OR composition)**.** *For any communication function F, the following bounds hold:*

- $\mathrm{N}_0(\textsc{And}_n\circ F)\leq \mathrm{N}_0(F)+\log n$
- $\mathrm{N}_1(\textsc{And}_n\circ F)\leq n\,\mathrm{N}_1(F)$
- $\mathrm{UN}_0(\textsc{And}_n\circ F)\leq \mathrm{UN}_0(F)+(n-1)\,\mathrm{UN}_1(F)$
- $\mathrm{UN}_1(\textsc{And}_n\circ F)\leq n\,\mathrm{UN}_1(F)$
- $\mathrm{N}_0(\textsc{Or}_n\circ F)\leq n\,\mathrm{N}_0(F)$
- $\mathrm{N}_1(\textsc{Or}_n\circ F)\leq \mathrm{N}_1(F)+\log n$
- $\mathrm{UN}_0(\textsc{Or}_n\circ F)\leq n\,\mathrm{UN}_0(F)$
- $\mathrm{UN}_1(\textsc{Or}_n\circ F)\leq (n-1)\,\mathrm{UN}_0(F)+\mathrm{UN}_1(F)$

Proof. We prove the statements for the functions of the form $\textsc{And}_n\circ F$. The proofs for the functions $\textsc{Or}_n\circ F$ are immediate by duality. A 0-certificate for $\textsc{And}_n\circ F$ on input $((x_1,y_1),\dots,(x_n,y_n))$ can be the index i such that $F(x_i,y_i)=0$, and 0-certificate for (x_i,y_i) on F. A 1-certificate for $\textsc{And}_n\circ F$ can be 1-certificates for each (x_i,y_i) on F, for $i=1,\dots,n$. For an unambiguous 0-certificate we can choose an unambiguous 0-certificate for (x_i,y_i) on F for the least i such that $F(x_i,y_i)=0$, and unambiguous 1-certificates for (x_j,y_j) on F for all $j=1,\dots,i-1$. For an unambiguous 1-certificate we can choose an unambiguous 1-certificate for each (x_i,y_i) on F, for $i=1,\dots,n$. □

We have the following corollary.

Corollary 10 (Tribes composition)**.** *Let $\textsc{Tr}_{n^2}=\textsc{Or}_n\circ\textsc{And}_n$. For any function F, we have:*

- $\mathrm{N}(\textsc{Tr}_{n^2}\circ F)=O(n\,\mathrm{N}(F)+n\log n)$
- $\mathrm{UN}(\textsc{Tr}_{n^2}\circ F)\leq n\,\mathrm{UN}_0(F)+n^2\,\mathrm{UN}_1(F)$

We now analyze the properties of N and UN under the cheat sheet operation.

Lemma 34 (Nondeterministic complexity of cheat sheet functions)**.** *Let F_{CS} be the cheat-sheet version of a total function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ where $N = \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}$. Then*

$$\mathrm{N}(F_{\text{CS}}) = O(\mathrm{N}(F)\log N), \qquad \mathrm{UN}_1(F_{\text{CS}}) = O(\mathrm{N}(F)\log N), \qquad \mathrm{UN}_0(F_{\text{CS}}) = O(\mathrm{UN}(F)\log N).$$

Proof. We first upper bound $\mathrm{N}_1(F_{\text{CS}})$ by giving a protocol. Let $\mathbf{x} = (x_1, \ldots, x_c), \mathbf{y} = (y_1, \ldots, y_c)$ and consider an input $(\mathbf{x}, u_0, \ldots, u_{2^c-1}, \mathbf{y}, v_0, \ldots, v_{2^c-1})$ to F_{CS}. The prover provides a proof of the form (ℓ, a, b) where $\ell \in \{0, \cdots, 2^c - 1\}, a, b \in \{0,1\}^{cN(F)}$. Note that the length of the proof is $O(cN(F)) = O(N(F)\log N)$. The players accept if and only if $u_\ell = a, v_\ell = b$, and $a \oplus b$ provides certificates that $F(x_i, y_i) = \ell_i$ for all $i = 1, \ldots, c$. If F_{CS} evaluates to 1 on this input, a valid proof always exists by giving $\ell = F^c(\mathbf{x}, \mathbf{y})$ and $a = u_\ell, b = v_\ell$. On the other hand if F_{CS} evaluates to 0 on this input, then by definition of the cheat sheet function for any message (ℓ, a, b) it cannot be that a, b agree with u_ℓ, v_ℓ and that $a \oplus b$ certifies that $F^c(\mathbf{x}, \mathbf{y}) = \ell$.

This protocol is in fact unambiguous. Say that F_{CS} evaluates to 1 on the input $(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v})$ and let $\ell = F^c(\mathbf{x}, \mathbf{y})$. A valid proof is given by (ℓ, u_ℓ, v_ℓ). Consider another proof (ℓ', a, b). First, if $\ell' \neq \ell$, then $a \oplus b$ cannot certify that $F^c(\mathbf{x}, \mathbf{y}) = \ell'$, as $F^c(\mathbf{x}, \mathbf{y}) = \ell$. Now if $\ell' = \ell$, then the players will only accept if $a = u_\ell$ and $b = v_\ell$. Thus there is a unique accepting proof.

We now turn to bound the N_0 complexity. Fix a cover $C_1, \ldots, C_{2^{N(F)}}$ of F by monochromatic rectangles. In this case the prover provides a message of the form $(\ell, i_1, \ldots, i_c, a, b)$, where $\ell \in \{0, \ldots, 2^c - 1\}, i_j \in \{0,1\}^{N(F)}, a, b \in \{0,1\}^{cN(F)}$. Thus the length of the proof is $O(cN(F)) = O(N \log N)$. Alice and Bob accept if and only if

1. $(x_j, y_j) \in C_{i_j}$ for all $j = 1, \ldots, c$.
2. C_{i_j} is ℓ_j-monochromatic on F for $j = 1, \ldots, c$,
3. $u_\ell = a, v_\ell = b$ and $a \oplus b$ does not provide valid certificates that $F^c(\mathbf{x}, \mathbf{y}) = \ell$.

If $F_{\text{CS}}(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v}) = 0$ then there is a valid proof by giving $\ell = F^c(\mathbf{x}, \mathbf{y})$, providing valid certificates for these values, and giving u_ℓ, v_ℓ. On the other hand, if $F_{\text{CS}}(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v}) = 1$, then if the steps 1,2 of the verification pass then it must be the case that a, b do not agree with u_ℓ, v_ℓ, as in this case $u_\ell \oplus v_\ell$ do provide valid certificates.

To upper bound the UN_0 complexity, the protocol is exactly the same except now a partition $R_1, \ldots, R_{\chi(F)}$ of rectangles monochromatic for F is used instead of a cover. In this case, there is a unique choice of witnesses $(i_1, \ldots, i_c)$ to certify the correct value $F^c(\mathbf{x}, \mathbf{y}) = \ell$. The second part (a, b) of a valid proof is also uniquely specified as it must agree with the part of the input (u_ℓ, v_ℓ). $\square$

Corollary 11. *For any total function $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ with $N = \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}$, we have*

- $\mathrm{UN}(\mathrm{TR}_{n^2} \circ F_{\text{CS}}) = O(n\,\mathrm{UN}(F)\log N + n^2\,\mathrm{N}(F)\log N)$
- $\mathrm{N}(\mathrm{TR}_{n^2} \circ F_{\text{CS}}) = O(n\,\mathrm{N}(F)\log N)$.

We put these together to get an upper bound on UN for the iterated function. Let $F_0 = \mathrm{DISJ}_n$ and $F_{i+1} \coloneqq \mathrm{TR}_{n^2} \circ (F_i)_{\text{CS}}$ for all $i \geq 0$. The function F_k for appropriately chosen k will provide the near-quadratic separation.

Claim 35. *There is a constant a such that for any $k \geq 0$, we have*

- $\mathrm{UN}(F_k) = O(n^{k+2} a^k k^k \log^k n)$
- $\mathrm{N}(F_k) = O(n^{k+1} a^k k^k \log^k n)$.

When k is constant, these simplify to $\widetilde{O}(n^{k+2})$ and $\widetilde{O}(n^{k+1})$, respectively.

Proof. This follows from Corollary 11 by induction on k. In the base case, we have $\mathrm{N}(\textsc{Disj}_n) = O(\mathrm{UN}(\textsc{Disj}_n)) = O(n)$. The induction step follows immediately from Corollary 11. The only subtlety is the size of N, which increases polynomially with each iteration, which means $\log N = O(k \log n)$. This gives the $a^k k^k \log^k n$ factor. $\square$

Next, we prove a lower bound on $\mathrm{R}(F_k)$. To do this, we need to get a handle on the behavior of R when the function is composed with $\textsc{And}_n$ and $\textsc{Or}_n$. We use the following definition and fact.

Definition 36. Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ be a (partial) function and let $\varepsilon \in (0, 1/2)$. For any protocol Π and any $b \in \{0, 1\}$,

$$\mathrm{IC}^b(\Pi) \coloneqq \max_{\mu \text{ on } F^{-1}(b)} \mathrm{IC}^\mu(\Pi).$$

The following claim shows a composition result for one-sided information complexity. A result similar in spirit for the $\textsc{Or}_n \circ \textsc{And}$ function was shown by [BJKS04].

Claim 37 (Composition)**.** *Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, *\}$ be a (partial) function, and let $\varepsilon \in (0, 1/2)$ be a constant. For any protocol Π for $\textsc{Or}_n \circ F$ with worst case error at most ε, there is a protocol Π' for F with worst error at most ε such that*

$$\mathrm{IC}^0(\Pi') = O(\mathrm{IC}^0(\Pi)/n) \quad \textit{and} \quad \mathrm{CC}(\Pi') = O(\mathrm{CC}(\Pi)).$$

Similarly, if Π is a protocol for $\textsc{And}_n \circ F$ with worst case error at most ε, there is a protocol Π' for F with worst case error at most ε, such that

$$\mathrm{IC}^1(\Pi') = O(\mathrm{IC}^1(\Pi)/n) \quad \textit{and} \quad \mathrm{CC}(\Pi') = O(\mathrm{CC}(\Pi)).$$

Proof. We show the result for $\textsc{Or}_n \circ F$. The result for $\textsc{And}_n \circ F$ follows similarly. Let μ be a distribution on $F^{-1}(0)$. We will exhibit a protocol Π^1 for F with worst case error at most ε, such that

$$\mathrm{IC}^\mu(\Pi^1) = O(\mathrm{IC}^0(\Pi)/n) \quad \text{and} \quad \mathrm{CC}(\Pi^1) = O(\mathrm{CC}(\Pi)).$$

The desired result then follows from Fact 14 (Minimax principle) and Fact 13 (Error reduction). Let us define random variables:

1. $XY = (X_1Y_1, \ldots, X_nY_n)$ where each $(X_iY_i) \sim \mu$ and i.i.d.
2. $D = (D_1, \ldots, D_n)$ where each D_i is uniformly distributed in $\{A, B\}$ and i.i.d.
3. $U = (U_1, \ldots, U_n)$ where for each i, $U_i = X_i$ if $D_i = A$ and $U_i = Y_i$ if $D_i = B$.

Using Fact 9.E (Independence) we have,

$$I(XY : \Pi \mid DU) \geq \sum_{i=1}^{n} I(X_iY_i : \Pi \mid DU).$$

This implies there exists $j \in [n]$ such that

$$\begin{aligned}
\frac{1}{n} I(XY : \Pi \mid DU) \geq\ & I(X_jY_j : \Pi \mid DU) \\
=\ & I(X_jY_j : \Pi \mid DjUjD_{-j}U_{-j}) \\
=\ & \frac{1}{2}\left(I(X_j : \Pi \mid Y_jD_{-j}U_{-j}) + I(X_j : \Pi \mid Y_jD_{-j}U_{-j})\right)
\end{aligned}$$

$$= \frac{1}{2}\left(I(X_j : \Pi D_{-j}U_{-j} \mid Y_j) + I(X_j : \Pi D_{-j}U_{-j} \mid Y_j)\right). \quad \text{(Fact 9.D: Bar hopping)}$$

Define protocol Π^1 as follows. Alice and Bob insert their inputs at the j-th coordinate and generate $(D_{-j}U_{-j})$ using public-coins. They go ahead and simulate Π afterwards. From above we have

$$\frac{1}{n} I(XY : \Pi \mid DU) \geq \frac{1}{2}\,\mathrm{IC}^{\mu}(\Pi^1). \tag{16}$$

It is clear that $\mathrm{CC}(\Pi^1) \leq \mathrm{CC}(\Pi)$ and the worst case error of Π^1 is upper bounded by the worst case error of Π. Consider,

$$\begin{aligned}
\mathrm{IC}^0(\Pi) &\geq \mathrm{IC}^{XY}(\Pi) \\
&= I(X : \Pi \mid Y) + I(Y : \Pi \mid X) \\
&= I(X : \Pi \mid Y) + I(DU : \Pi \mid XY) + I(Y : \Pi \mid X) + I(DU : \Pi \mid XY) && (DU \leftrightarrow XY \leftrightarrow \Pi) \\
&= I(XDU : \Pi \mid Y) + I(YDU : \Pi \mid X) && \text{(Fact 9.A: Chain rule)} \\
&\geq I(X : \Pi \mid YDU) + I(Y : \Pi \mid XDU) && \text{(Fact 9.D: Bar hopping)} \\
&= I(X : \Pi \mid YDU) + I(X : Y \mid DU) + I(Y : \Pi \mid XDU) && (X \leftrightarrow DU \leftrightarrow Y) \\
&= I(X : \Pi Y \mid DU) + I(Y : \Pi \mid XDU) && \text{(Fact 9.A: Chain rule)} \\
&\geq I(X : \Pi \mid DU) + I(Y : \Pi \mid XDU) && \text{(Fact 9.C: Monotonicity)} \\
&= I(XY : \Pi \mid DU). && \text{(Fact 9.A: Chain rule)}
\end{aligned}$$

This along with Eq. (16) shows the desired. □

To be able to use this, we need a way of converting between IC^0, IC^1, and IC. The following fact was shown by [GJPW15, Corollary 18] using the "information odometer" of Braverman and Weinstein [BW15] (the upper bound on $\mathrm{CC}(\Pi')$ was not stated explicitly in [GJPW15], but it traces back to [BW15, Theorem 3], which was used in [GJPW15]).

Fact 38. *Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function. Let $1/2 > \delta > \varepsilon > 0$ and $b \in \{0,1\}$. Then for any protocol Π with $\mathrm{err}(\Pi) < \varepsilon$, there is a protocol Π' with $\mathrm{err}(\Pi') < \delta$ such that*

$$\mathrm{IC}(\Pi') = O(\mathrm{IC}^b(\Pi) + \log \mathrm{CC}(\Pi)) \quad \textit{and} \quad \mathrm{CC}(\Pi') = O(\mathrm{CC}(\Pi) \log \mathrm{CC}(\Pi)).$$

Theorem 12. *There is a constant b such that for every $k \leq n^{1/10}$, we have*

$$\mathrm{R}(F_k) = \Omega\left(\frac{n^{2k+1}}{b^k k^{3k} \log^{3k} n}\right).$$

Proof. Consider the protocol Π for F_k with error at most $1/3$ such that $\mathrm{CC}(\Pi) = \mathrm{R}(F_k)$, and hence $\mathrm{IC}(\Pi) = O(\mathrm{R}(F_k))$. Recall that $F_k = \textsc{Tr}_{n^2} \circ (F_{k-1})_{\mathrm{CS}} = \textsc{Or}_n \circ \textsc{And}_n \circ (F_{k-1})_{\mathrm{CS}}$. Using Claim 37, we get a protocol Π' for $\textsc{And}_n \circ (F_{k-1})_{\mathrm{CS}}$ with $\mathrm{err}(\Pi') \leq 1/3$, $\mathrm{IC}^0(\Pi') = O(\mathrm{IC}^0(\Pi)/n) = O(\mathrm{IC}(\Pi)/n) = O(\mathrm{R}(F_k)/n)$, and $\mathrm{CC}(\Pi') = O(\mathrm{CC}(\Pi))$. Using Fact 38, we get a protocol Π'' for $\textsc{And}_n \circ (F_{k-1})_{\mathrm{CS}}$ with $\mathrm{err}(\Pi'') \leq 2/5$, $\mathrm{IC}(\Pi'') = O(\mathrm{IC}^0(\Pi') + \log \mathrm{CC}(\Pi')) = O(\mathrm{R}(F_k)/n + \log \mathrm{R}(F_k))$, and $\mathrm{CC}(\Pi'') = O(\mathrm{CC}(\Pi') \log \mathrm{CC}(\Pi')) = O(\mathrm{R}(F_k) \log \mathrm{R}(F_k))$. Using Fact 13 (Error reduction), we get a protocol Π''' for $\textsc{And}_n \circ (F_{k-1})_{\mathrm{CS}}$ with $\mathrm{err}(\Pi''') \leq 1/3$, $\mathrm{IC}(\Pi''') = O(\mathrm{R}(F_k)/n + \log \mathrm{R}(F_k))$, and $\mathrm{CC}(\Pi''') = O(\mathrm{R}(F_k) \log \mathrm{R}(F_k))$.

We can repeat this process to strip away the $\textsc{And}_n$; that is, we use Claim 37, Fact 38, and Fact 13 (Error reduction) to get a protocol Π'''' for $(F_{k-1})_{\mathrm{CS}}$ with $\mathrm{err}(\Pi'''') \leq 1/3$, $\mathrm{IC}(\Pi'''') = O(\mathrm{R}(F_k)/n^2 + \log \mathrm{R}(F_k))$, and $\mathrm{CC}(\Pi'''') = O(\mathrm{R}(F_k) \log^2 \mathrm{R}(F_k))$. Then Theorem 5 gives a protocol

Π'''''' for F_{k-1} with $\mathrm{err}(\Pi'''''') \le 1/3$, $\mathrm{IC}(\Pi'''''') = O((\mathrm{R}(F_k)\log^3 N)/n^2 + \log \mathrm{R}(F_k) \cdot \log^3 N)$, and $\mathrm{CC}(\Pi'''''') = O(\mathrm{R}(F_k)\log^2 \mathrm{R}(F_k)\log^2 N)$, where N is the input size of F_{k-1}. Here $N = n^{O(k)}$, so $\log N = O(k\log n)$ and $\log \mathrm{R}(F_k) = O(k\log n)$, and hence $\mathrm{IC}(\Pi'''''') = O((\mathrm{R}(F_k)k^3\log^3 n)/n^2 + k^4\log^4 n)$ and $\mathrm{CC}(\Pi'''''') = O(\mathrm{R}(F_k)k^4\log^4 n)$.

We now repeat this k times to get a protocol Ψ for $F_0 = \textsc{Disj}_n$. Then we have $\mathrm{CC}(\Psi) = O(b^k\,\mathrm{R}(F_k)k^{4k}\log^{4k} n)$ for some constant b, and the communication complexity of every intermediate protocol in the construction is also at most $O(b^k\,\mathrm{R}(F_k)k^{4k}\log^{4k} n)$. To calculate $\mathrm{IC}(\Psi)$, note that each iteration divides IC by n^2, adds a $\log \mathrm{CC}$ term, and multiplies by $k^3\log^3 n$. Thus we get, for some constant b,

$$\mathrm{IC}(\Psi) = O\left((\mathrm{R}(F_k)b^k k^{3k}\log^{3k} n)/n^{2k} + k^3\log^3 n \cdot \log \mathrm{CC}(\Psi)\sum_{i=0}^{k-1}\left(\frac{k^3\log^3 n}{n^2}\right)^i\right).$$

Since $k = O(n^{1/10})$, the sum is $O(1)$, so we get

$$\begin{aligned}\mathrm{IC}(\Psi) &= O((\mathrm{R}(F_k)b^k k^{3k}\log^{3k} n)/n^{2k} + k^3\log^3 n \cdot \log \mathrm{CC}(\Psi))\\ &= O((\mathrm{R}(F_k)b^k k^{3k}\log^{3k} n)/n^{2k} + k^3\log^3 n \cdot (\log \mathrm{R}(F_k) + k\log k + k\log\log n))\\ &= O((\mathrm{R}(F_k)b^k k^{3k}\log^{3k} n)/n^{2k} + k^3\log^3 n \cdot \log \mathrm{R}(F_k) + n^{8/10}). \qquad (\text{since } k = O(n^{1/10}))\end{aligned}$$

Now, since $\mathrm{IC}(\textsc{Disj}_n) = \Omega(n)$, we get either $\mathrm{R}(F_k) = 2^{\Omega(n/k^3\log^3 n)} = \Omega(2^{\sqrt{n}})$ or

$$\mathrm{R}(F_k) = \Omega\left(\frac{n^{2k+1}}{b^k k^{3k}\log^{3k} n}\right).$$

Because $k = O(n^{1/10})$, the value of $2^{\sqrt{n}}$ is even larger than the desired lower bound, so the desired result follows. □

Finally, we get prove the near-quadratic separation.

Theorem 2. *There exists a total function $F\colon \mathcal{X}\times\mathcal{Y} \to \{0,1\}$ with $\mathrm{R}(F) \ge \mathrm{UN}(F)^{2-o(1)}$.*

Proof. We take $F = F_k$ with k some slowly growing function of n. In particular, let $k = \sqrt{\frac{\log n}{\log\log n}}$. This gives $\mathrm{R}(F_k) \ge \frac{n^{2k+1}}{2^{O(\sqrt{\log n\log\log n})}}$ and $\mathrm{UN}(F_k) \le n^{k+2}2^{O(\sqrt{\log n\log\log n})}$, so $\log \mathrm{UN}(F_k) = \log^{3/2} n/\log\log^{1/2} n + O(\sqrt{\log n\log\log n})$ and

$$\begin{aligned}\log \mathrm{R}(F_k) &= 2\log^{3/2} n/\log\log^{1/2} n - O(\sqrt{\log n\log\log n})\\ &= 2\log \mathrm{UN}(F_k) - O(\log^{2/3}\mathrm{UN}(F_k)\log\log^{4/3}\mathrm{UN}(F_k)).\end{aligned}$$

Thus

$$\mathrm{R}(F_k) \ge \mathrm{UN}(F_k)^{2-O(\alpha(\mathrm{UN}(F_k)))}$$

where $\alpha(x) = \frac{\log\log^{4/3} x}{\log^{1/3} x} = o(1)$. □

Acknowledgements

Part of this work was performed when the authors met during the workshop "Semidefinite and Matrix Methods for Optimization and Communication" hosted at the Institute for Mathematical Sciences, Singapore. We thank them for their hospitality.

This work is partially supported by ARO grant number W911NF-12-1-0486, by the Singapore Ministry of Education and the National Research Foundation, also through NRF RF Award No. NRF-NRFF2013-13, and the Tier 3 Grant "Random numbers from quantum processes" MOE2012-T3-1-009. This research is also partially supported by the European Commission IST STREP project Quantum Algorithms (QALGO) 600700 and by the French ANR Blanc program under contract ANR-12-BS02- 005 (RDAM project). M.G. is partially supported by the Simons Award for Graduate Students in TCS.

R.J would like to thank Ankit Garg for helpful discussions.

References

[AA03] Scott Aaronson and Andris Ambainis. Quantum search of spatial regions. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS 2003)*, pages 200–209, 2003. doi:10.1109/SFCS.2003.1238194. [p. 1]

[AA15] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *Proceedings of the 47th ACM Symposium on Theory of Computing (STOC 2015)*, pages 307–316, 2015. doi:10.1145/2746539.2746547. [p. 21]

[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)*, 2016. arXiv:1511.01937. [pp. 2, 3, 13, 21]

[AKK16] Andris Ambainis, Martins Kokainis, and Robin Kothari. Nearly optimal separations between communication (or query) complexity and partitions. *To appear in Proceedings of CCC 2016*, 2016. Preprints: arXiv:1512.00661 and arXiv:1512.01210. [pp. 1, 2, 3, 28]

[AUY83] Alfred V. Aho, Jeffrey D. Ullman, and Mihalis Yannakakis. On notions of information transfer in VLSI circuits. In *Proceedings of the 15th ACM Symposium on Theory of Computing (STOC 1983)*, pages 133–139, 1983. doi:10.1145/800061.808742. [p. 1]

[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. [pp. 1, 27]

[BdW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21 – 43, 2002. doi:10.1016/S0304-3975(01)00144-X. [p. 4]

[BFNR08] Harry Buhrman, Lance Fortnow, Ilan Newman, and Hein Röhrig. Quantum property testing. *SIAM Journal on Computing*, 37(5):1387–1400, 2008. doi:10.1137/S0097539704442416. [pp. 21, 22, 23, 24]

[BJKS04] Ziv Bar-Yossef, T.S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004. doi:10.1016/j.jcss.2003.11.006. [pp. 1, 12, 30]

[Bra12] Mark Braverman. Interactive information complexity. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, STOC '12, pages 505–524, 2012. doi:10.1145/2213977.2214025. [p. 11]

[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. 31]

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006. [pp. 5, 7, 8, 9]

[Das11] Anirban DasGupta. *Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics*. Springer Texts in Statistics. Springer-Verlag New York, 2011. doi:10.1007/978-1-4419-9634-3. [p. 6]

[GJ16] Mika Göös and T.S. Jayram. A composition theorem for conical juntas. In *Proceedings of the 31st Annual Conference on Computational Complexity*, CCC '16, 2016. [pp. 22, 24, 25]

[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. [pp. 1, 2, 24, 31]

[GLM+15] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 257–266, 2015. doi:10.1145/2746539.2746596. [pp. 3, 22, 23, 24]

[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. [pp. 1, 2]

[JK10] Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity*, CCC '10, pages 247–258, 2010. doi:10.1109/CCC.2010.31. [pp. 1, 23, 24]

[Kla10] Hartmut Klauck. A strong direct product theorem for disjointness. In *Proceedings of the 42nd Symposium on Theory of Computing (STOC)*, pages 77–86, 2010. doi:10.1145/1806689.1806702. [p. 24]

[KLO99] Eyal Kushilevitz, Nathan Linial, and Rafail Ostrovsky. The linear-array conjecture in communication complexity is false. *Combinatorica*, 19(2):241–254, 1999. doi:10.1007/s004930050054. [p. 1]

[KN06] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 2006. URL: http://books.google.ca/books?id=dHH7rdhKwzsC. [p. 4]

[KS92] Bala Kalyanasundaram and Georg Schintger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992. doi:10.1137/0405044. [p. 1]

[Raz92] Alexander Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992. doi:10.1016/0304-3975(92)90260-M. [p. 1]

[Raz99] Ran Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, STOC '99, pages 358–367, 1999. doi:10.1145/301250.301343. [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. 27]

[RM99] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Combinatorica*, 19(3):403–435, 1999. doi:10.1109/SFCS.1997.646112. [p. 2]

[Sim97] Daniel R. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997. doi:10.1137/S0097539796298637. [p. 21]

[Yao77] Andrew Yao. Probabilistic computations: Toward a unified measure of complexity. In *Proceedings of the 18th IEEE Symposium on Foundations of Computer Science (FOCS 1977)*, pages 222–227, 1977. doi:10.1109/SFCS.1977.24. [p. 11]

ECCC ISSN 1433-8092

http://eccc.hpi-web.de