2016 IEEE 57th Annual Symposium on Foundations of Computer Science

Separations in communication complexity using cheat sheets and information complexity

[Extended abstract]*

Anurag Anshu[a], Aleksandrs Belovs[b], Shalev Ben-David[c], Mika Göös[d], Rahul Jain[a,e,f], Robin Kothari[c], Troy Lee[a,f,g], and Miklos Santha[a,h††]
[a] *CQT, National University of Singapore*
[b] *CWI, Amsterdam*
[c] *CSAIL, Massachusetts Institute of Technology*
[d] *SEAS, Harvard University*
[e] *Dept. of CS, National University of Singapore*
[f] *MajuLab, UMI 3654, Singapore*
[g] *SPMS, Nanyang Technological University*
[h] *IRIF, Université Paris Diderot, CNRS*

***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 further present a 1.5 power separation between exact quantum and randomized communication complexity, improving on the previous ≈ 1.15 separation by Ambainis (STOC 2013). Finally, we 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.

I. 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 semidefinite 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 and $\mathrm{Q}_E(F)$ for the exact quantum communication complexities of F, respectively. Note the easy relationships $\mathrm{D}(F) \geq \mathrm{R}(F) \geq \mathrm{Q}(F)$ and $\mathrm{D}(F) \geq \mathrm{Q}_E(F) \geq \mathrm{Q}(F)$.

There are examples of *partial* functions F for which $\mathrm{Q}(F)$ is exponentially smaller than $\mathrm{R}(F)$ [2]. 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 $\mathrm{R}(\textsc{Disj}_n) = \Omega(n)$ [3, 4] and $\mathrm{Q}(\textsc{Disj}_n) = O(\sqrt{n})$ [5, 6]. 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 [7], a well-known lower bound technique for randomized communication complexity.

We also give a 1.5 power separation between randomized communication complexity and *exact* quantum communication complexity. This improves the previous best separation

* All proofs appear in the full version of this work [1].

0272-5428/16 $31.00 © 2016 IEEE
DOI 10.1109/FOCS.2016.66

of ≈ 1.15 due to Ambainis [8].

Theorem 2. *There exists a total function* $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ *with* $\mathrm{R}(F) = \widetilde{\Omega}(\mathrm{Q}_E(F)^{1.5})$.

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 [7] and the partition bound [9]. 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 II 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 [10]. This changed with the breakthrough work of Göös, Pitassi, and Watson [11], who exhibited a total function F with $\mathrm{D}(F) = \widetilde{\Omega}(\mathrm{UN}(F)^{1.5})$. Ambainis, Kokainis and Kothari [12] 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 [13] showed $\mathrm{D}(F) = O(\mathrm{UN}(F)^2)$ for all total F.

Göös, Jayram, Pitassi, and Watson [14] improved the original [11] 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 3. *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 $\mathrm{UN}(F)$.

A. 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 [11], building on work of Raz and McKenzie [15], 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 [12] followed the same plan and obtained their communication complexity separation by improving the query complexity separation of [11] 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 [16] 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 [12] to show the query analog of our Theorem 3, 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 [14] 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.

B. 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 the full version [1]). This is a query measure that satisfies $\deg^+_{1/3}(f) \leq \mathrm{R}(f)$ and has a known lifting theorem [17]. The first idea to obtain our theorems would be to show (say) that $\deg^+_{1/10}(\neg f_{\mathrm{CS}}) = \widetilde{\Omega}(\deg^+_{1/3}(f))$[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 [16] 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 6) 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 [17]. The separation of Theorem 2 follows a very similar plan, starting instead with the query function $h = \textsc{Pr-Or}_n \circ \textsc{And}_m$ for $m = \Theta(\sqrt{n})$, where $\textsc{Pr-Or}_n$ is a promise version of the $\textsc{Or}_n$ function restricted to inputs of Hamming weight 0 or 1.

Moving on to our third result (Theorem **??**), 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 **??** [12] 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. Information complexity, or more precisely one-sided information complexity, is known [7] to satisfy a composition theorem for the $\textsc{And}_n$ function. While one-sided information complexity upper bounds can be converted to information complexity upper bounds [14], 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 4 (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 state this more formally as Theorem 6 in Section III. This is the main technical result of this work; 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.

II. Preliminaries

In this paper we denote query complexity (or decision tree complexity) measures using the superscript dt. For example, the deterministic, bounded-error randomized, exact quantum, and bounded-error quantum query complexities of a function f are denoted $\mathrm{D}^{\mathrm{dt}}(f)$, $\mathrm{R}^{\mathrm{dt}}(f)$, $\mathrm{Q}^{\mathrm{dt}}_E(f)$, and $\mathrm{Q}^{\mathrm{dt}}(f)$ respectively. We refer the reader to the survey by Buhrman and de Wolf [18] 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

[1] We negate the function f_{CS} because the 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 1) also allows for partial functions.

f is allowed to output an arbitrary value for inputs outside $\mathrm{dom}(f)$. $\textsc{And}_n$ and $\textsc{Or}_n$ denote the AND and 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 [19].

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. Exact quantum communication complexity, denoted $\mathrm{Q}_E(F)$, is defined similarly, except it must output the correct answer with certainty.

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 $\mathrm{N}_b(F) \coloneqq \lceil\log \mathrm{C}_b(F)\rceil$. The nondeterministic complexity of F is $\mathrm{N}(F) \coloneqq \max\{\mathrm{N}_0(F),\mathrm{N}_1(F)\}$. 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 usual information complexity of F (see the full version [1] for formal definitions). 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.

III. 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 [16].

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 1 ($(F,\mathcal{G})$-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 **??**), 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 2 (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 6) we add two additional constraints on the family $\mathcal{G}$.

Definition 3 (Nontrivial XOR family)**.** Let $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$ a family of communication functions, where each $G_i\colon (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$. We say that $\mathcal{G}$ is a nontrivial XOR family if the following conditions hold.

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

A. 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 5. *Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1,*\}$ be a (partial) function and $\mathcal{G} = \{G_0, \ldots, G_{2^c-1}\}$ a family of communication functions, where each $G_i\colon (\mathcal{X}^c \times \{0,1\}^m) \times (\mathcal{Y}^c \times \{0,1\}^m) \to \{0,1\}$. If $\mathcal{G}$ is consistent outside F (Definition 2) then*

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

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

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

Proof: We first give the proof for the bounded-error quantum communication complexity.

Consider an input where Alice holds $\mathbf{x} = (x_1, \ldots, x_c) \in \mathcal{X}^c$ and $\mathbf{u} = (u_0, \ldots, u_{2^c-1}) \in \{0,1\}^{m2^c}$ and Bob holds $\mathbf{y} = (y_1, \ldots, y_c) \in \mathcal{Y}^c$ and $\mathbf{v} = (v_0, \ldots, v_{2^c-1}) \in \{0,1\}^{m2^c}$. For each $i = 1, \ldots, c$, 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 $\ell = (\ell_1, \ldots, \ell_c)$, 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 $i = 1, \ldots, c$. 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 $\ell = (F(x_1, y_1), \ldots, F(x_c, y_c))$ 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 $1/3$.

The proof for the exact quantum communication complexity follows similarly. In this case, Alice and Bob run an exact quantum protocol for F on each input (x_i, y_i) to obtain $\ell = (\ell_1, \ldots, \ell_c)$, and then run an exact quantum protocol to evaluate G_ℓ on input $\mathbf{x}, u_\ell, \mathbf{y}, v_\ell$.

If each $(x_i, y_i) \in \mathrm{dom}(F)$ for $i = 1, \ldots, c$ then $\ell = F(x_1, y_1), \ldots, F(x_c, y_c)$ and the output will be correct. Otherwise, the output will also be correct as $\mathcal{G}$ is consistent outside of F. ∎

B. 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.

Theorem 6. *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 3) 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*

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

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

The proof is given in the full version [1].

IV. Proofs of Theorem 1 and Theorem 2

We give a high level overview of the proofs of Theorem 1 and Theorem 2. These proofs can be found in the full version [1].

In both cases, we begin in the world of query complexity. The starting point of Theorem 1 is the partial function

$$\textsc{Str} \coloneqq \textsc{Simon}_n \circ \textsc{Or}_n \circ \textsc{And}_n. \tag{1}$$

Here $\textsc{Simon}_n$ is a certain property testing version of Simon's problem [20] introduced in [21] 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 [16, §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 Forrelation [22] in place of Simon, but we find it more convenient to work with 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.

By Theorem 6, to show a lower bound on the randomized communication complexity of $F_{\mathcal{G}}$, it suffices to show a randomized communication lower bound on $F = \textsc{Str} \circ \mathrm{IP}_b$. To do this, we use the query-to-communication lifting theorem of [17], which requires us to show a lower bound on the approximate *conical junta degree* of $\textsc{Str}$. 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 [23]. 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, giving dual certificates for these functions of a special form. This will prove Theorem 7.

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.

For the separation between randomized and exact quantum communication complexity, we begin in the setting of query complexity with the partial function

$$\textsc{Ptr}_{n,m} \coloneqq \textsc{Pr-Or}_n \circ \textsc{And}_m, \tag{2}$$

where we eventually choose $m = \Theta(\sqrt{n})$ and $\textsc{Pr-Or}_n$ is a promise version of the $\textsc{Or}_n$ function

$$\textsc{Pr-Or}_n(x) = \begin{cases} 0 & \text{if } |x| = 0 \\ 1 & \text{if } |x| = 1 \\ * & \text{otherwise} \end{cases}.$$

The exact quantum query complexity of $\textsc{Ptr}$ is $O(\sqrt{n}m)$, while its randomized query complexity is $\Omega(nm)$. As shown in [16, §6.4], taking $m = \Theta(\sqrt{n})$, the cheat sheet version of $\textsc{Ptr}$ is a total function that witnesses an $\widetilde{O}(n)$ versus $\Omega(n^{3/2})$ separation between randomized and exact quantum query complexities.

We again lift $\textsc{Ptr}$ to a partial two-party function $H \coloneqq \textsc{Ptr} \circ \mathrm{IP}_b$ by composing it with IP_b. The final function for the separation of Theorem 2 will be a $(H,\mathcal{T})$-lookup function for a particular family of XOR functions $\mathcal{T}$ that consistent outside of H and defined in a similar fashion to the family $\mathcal{G}$ described above.

The main randomized lower bounds we prove for the separations of Theorem 1 and Theorem 2 are the following.

Theorem 7. *Let* $m \le n$ *and let* $b \ge t \log n$ *for a sufficiently large constant* t*. Then*

$$\mathrm{R}(\textsc{Str} \circ \mathrm{IP}_b) = \widetilde{\Omega}(n^{2.5}) \textit{ and } \mathrm{R}(\textsc{Ptr}_{n,m} \circ \mathrm{IP}_b) = \Omega(nm).$$

The plan for both of these lower bounds is similar, as outlined in Figure IV.

A. Randomized vs. bounded-error quantum

Given the randomized lower bounds of Theorem 7, we now finish the proof of Theorem 1. The proof of Theorem 2 proceeds similarly and is deferred to the full version [1].

We first need two preliminary results.

Fact 4 (Composition of quantum query complexity [24])**.** *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 5 (Composition with query function [5])**.** *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* $\mathrm{Q}(f \circ (F_1,\dots,F_n)) = O(\mathrm{Q}^{\mathrm{dt}}(f) \log \mathrm{Q}^{\mathrm{dt}}(f) \cdot \max_i \mathrm{Q}(F_i) \log n)$.

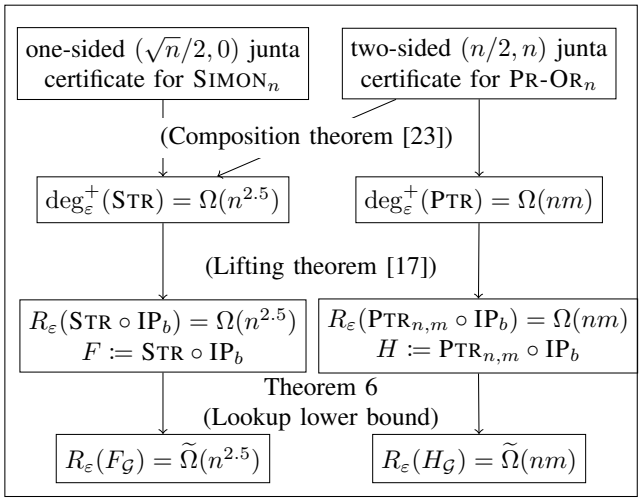

Figure 1. Overview of the randomized communication complexity lower bounds for Theorem 1 and Theorem 2.

B. Proof of Theorem 1

Let $F = \textsc{Str} \circ \text{IP}_b$ as defined in Equation 1, 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 $\text{dom}(F)$. Let $A = \{0,1\}^{bn^3} \times \{0,1\}^{n(n \log n + 1)}$. Then $P : A \times A \to \{0,1\}$. This function will be defined such that $P(x, u, y, v) = 1$ if and only if $(x, y) \in \text{dom}(F)$ 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 $z_{ijk} = \text{IP}_b(x_{ijk}, y_{ijk})$ 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 $\text{dom}(\textsc{Simon}_n)$. Let $w = u \oplus v$ and decompose this as $w = (q, \text{C})$, where $q \in \{0,1\}^n$ and $\text{C} = (C_1, \ldots, C_n)$ 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) $q \in \text{dom}(\textsc{Simon}_n)$
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 6. *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) $(x_j, y_j) \notin \text{dom}(F)$, 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 $u \oplus v = (q, \text{C})$ 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 $q = (z_1, \ldots, z_n) \notin \text{dom}(F)$. This means $\mathcal{G}$ is consistent outside F.

Finally, let $(\mathbf{x}, \mathbf{y}) \in \text{dom}(F^c)$. 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. ∎

We can now finish the separation.

Theorem 8. *Let $F = \textsc{Str} \circ \text{IP}_b$ be defined as in Equation 1 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*

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

Proof: We start with the randomized lower bound. As $c = 10 \log n \geq \text{R}(F)$ we can apply Theorem 6 to obtain $\text{R}(F_\mathcal{G}) = \widetilde{\Omega}(\text{R}(F)) = \widetilde{\Omega}(n^{2.5})$ by Theorem 7.

Now we turn to the quantum upper bound. By Theorem 5 it suffices to show $\text{Q}(F) = \widetilde{O}(bn)$ and $\max_s \text{Q}(G_s) = \widetilde{O}(bn)$. As $\text{Q}^{\text{dt}}(\textsc{Simon}_n) = O(\log n \log\log n)$ and $\text{Q}^{\text{dt}}(\textsc{Or}_n \circ \textsc{And}_n) = O(n)$, by the composition theorem Theorem 4 $\text{Q}(\textsc{Str}) = \widetilde{O}(n)$. Thus $\text{Q}(F) = \widetilde{O}(bn)$ by Fact 5, as $\text{Q}(\text{IP}_b) \leq b$.

We now turn to show $\max_s \text{Q}(G_s) = \widetilde{O}(bn)$. 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, $z_{ijk} = \text{IP}_b(x_{ijk}, y_{ijk})$ 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 $w = (q, \text{C})$ where $\text{C} = (C_1, \ldots, C_n)$ 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 $q \notin \text{dom}(\textsc{Simon}_n)$ 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 5, it takes $\widetilde{O}(bn)$ qubits of quantum communication to verify that all $g_{i,t} = 1$. ■

V. Partitions vs. randomized communication

In this section, we give an overview of Theorem 3, which we restate for convenience:

Theorem 3. *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 [12] 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 7 (Cheat sheets for total functions)**.** Let $F\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a total function. Fix a cover $\mathcal{R} = \{R_0,\dots,R_{2^{N(F)}-1}\}$ 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

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

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 8. Note that F_{CS} is in particular a $(F,\mathcal{G})$-lookup function where $\mathcal{G}$ is a nontrivial XOR family (Definition 3), thus Theorem 6 applies. Further letting $\mathcal{X}' \times \mathcal{Y}'$ be the domain of F_{CS}, note that $N' = \min\{\log|\mathcal{X}'|,\log|\mathcal{Y}'|\} = O(cN + c\cdot 2^c N(F)) = O(N^{12})$.

Recall that the function TR_{n^2} on n^2 input bits is the composition $\mathrm{OR}_n \circ \mathrm{AND}_n$. The separating function of Theorem 3 is constructed by starting with disjointness on n variables and alternately taking the cheat sheet function of it and composing 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 9 (AND/OR composition)**.** *For any communication function F, the following bounds hold:*

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

Proof: We prove the statements for the functions of the form $\mathrm{AND}_n \circ F$. The proofs for the functions $\mathrm{OR}_n \circ F$ are immediate by duality. A 0-certificate for $\mathrm{AND}_n \circ F$ 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 $\mathrm{AND}_n \circ F$ 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 9 (Tribes composition)**.** *Let $\mathrm{TR}_{n^2} = \mathrm{OR}_n \circ \mathrm{AND}_n$. For any function F, we have:*

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

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

Lemma 10 (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*

$$\begin{aligned}\mathrm{N}(F_{\mathrm{CS}}) &= O(\mathrm{N}(F)\log N)\\ \mathrm{UN}_1(F_{\mathrm{CS}}) &= O(\mathrm{N}(F)\log N)\\ \mathrm{UN}_0(F_{\mathrm{CS}}) &= O(\mathrm{UN}(F)\log N).\end{aligned}$$

Proof: We first upper bound $\mathrm{N}_1(F_{\mathrm{CS}})$ by giving a protocol. Let $\mathbf{x} = (x_1,\dots,x_c), \mathbf{y} = (y_1,\dots,y_c)$ and consider an input $(\mathbf{x},u_0,\dots,u_{2^c-1},\mathbf{y},v_0,\dots,v_{2^c-1})$ 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 $i = 1,\dots,c$. 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 $F_{\mathrm{CS}}(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v}) = 0$ 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 $F_{\mathrm{CS}}(\mathbf{x}, \mathbf{u}, \mathbf{y}, \mathbf{v}) = 1$, 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). ∎

Corollary 10. *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*

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

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

Claim 11. *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 10 by induction on k. In the base case, we have $\mathrm{N}(\mathrm{DISJ}_n) = O(\mathrm{UN}(\mathrm{DISJ}_n)) = O(n)$. The induction step follows immediately from Corollary 10. 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. ∎

To prove Theorem 3, the remaining task is to lower bound $\mathrm{R}(F_k)$. We show the following theorem.

Theorem 11. *There is a constant b such that for every $k \leq n^{1/10}$, we have*

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

The proof of Theorem 11 is given in the full version [1]. Finally, we get prove the near-quadratic separation.

Theorem 3. *There exists a total function $F \colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ with $\mathrm{R}(F) \geq \mathrm{UN}(F)^{2-o(1)}$.*

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

$$\mathrm{R}(F_k) \geq \frac{n^{2k+1}}{2^{O(\sqrt{\log n \log\log n})}}$$

and

$$\mathrm{UN}(F_k) \leq n^{k+2} 2^{O(\sqrt{\log n \log\log n})},$$

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

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

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

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their comments. R.J. would like to thank Ankit Garg for helpful discussions.

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.

REFERENCES

[1] A. Anshu, A. Belovs, S. Ben-David, M. Göös, R. Jain, R. Kothari, T. Lee, and M. Santha, "Separations in communication complexity using cheat sheets and information complexity," Electronic Colloquium on Computational Complexity (ECCC), Tech. Rep. TR16-072, 2016, Full version. [Online]. Available: http://eccc.hpi-web.de/report/2016/072/

[2] R. Raz, "Exponential separation of quantum and classical communication complexity," in *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, ser. STOC '99, 1999, pp. 358–367.

[3] B. Kalyanasundaram and G. Schintger, "The probabilistic communication complexity of set intersection," *SIAM Journal on Discrete Mathematics*, vol. 5, no. 4, pp. 545–557, 1992.

[4] A. Razborov, "On the distributional complexity of disjointness," *Theoretical Computer Science*, vol. 106, no. 2, pp. 385–390, 1992.

[5] H. Buhrman, R. Cleve, and A. Wigderson, "Quantum vs. classical communication and computation," in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, ser. STOC '98, 1998, pp. 63–68.

[6] S. Aaronson and A. Ambainis, "Quantum search of spatial regions," in *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS 2003)*, 2003, pp. 200–209.

[7] Z. Bar-Yossef, T. Jayram, R. Kumar, and D. Sivakumar, "An information statistics approach to data stream and communication complexity," *Journal of Computer and System Sciences*, vol. 68, no. 4, pp. 702–732, 2004.

[8] A. Ambainis, "Superlinear advantage for exact quantum algorithms," in *Proceedings of the 45th ACM Symposium on Theory of Computing (STOC 2013)*, 2013, pp. 891–200.

[9] R. Jain and H. Klauck, "The partition bound for classical communication complexity and query complexity," in *Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity*, ser. CCC '10, 2010, pp. 247–258.

[10] E. Kushilevitz, N. Linial, and R. Ostrovsky, "The linear-array conjecture in communication complexity is false," *Combinatorica*, vol. 19, no. 2, pp. 241–254, 1999.

[11] M. Göös, T. Pitassi, and T. Watson, "Deterministic communication vs. partition number," in *Proceedings of the 56th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2015, pp. 1077–1088.

[12] A. Ambainis, M. Kokainis, and R. Kothari, "Nearly Optimal Separations Between Communication (or Query) Complexity and Partitions," in *31st Conference on Computational Complexity (CCC 2016)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), vol. 50, 2016, pp. 4:1–4:14.

[13] A. V. Aho, J. D. Ullman, and M. Yannakakis, "On notions of information transfer in VLSI circuits," in *Proceedings of the 15th ACM Symposium on Theory of Computing (STOC 1983)*, 1983, pp. 133–139.

[14] M. Göös, T. Jayram, T. Pitassi, and T. Watson, "Randomized communication vs. partition number," *Electronic Colloquium on Computational Complexity (ECCC) http://eccc.hpi-web.de/report/2015/169/TR15-169*, 2015.

[15] R. Raz and P. McKenzie, "Separation of the monotone NC hierarchy," *Combinatorica*, vol. 19, no. 3, pp. 403–435, 1999.

[16] S. Aaronson, S. Ben-David, and R. Kothari, "Separations in query complexity using cheat sheets," in *Proceedings of the 48th ACM Symposium on Theory of Computing (STOC 2016)*, 2016.

[17] M. Göös, S. Lovett, R. Meka, T. Watson, and D. Zuckerman, "Rectangles are nonnegative juntas," in *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing*, ser. STOC '15, 2015, pp. 257–266.

[18] H. Buhrman and R. de Wolf, "Complexity measures and decision tree complexity: a survey," *Theoretical Computer Science*, vol. 288, no. 1, pp. 21 – 43, 2002.

[19] E. Kushilevitz and N. Nisan, *Communication Complexity*. Cambridge University Press, 2006. [Online]. Available: http://books.google.ca/books?id=dHH7rdhKwzsC

[20] D. R. Simon, "On the power of quantum computation," *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1474–1483, 1997.

[21] H. Buhrman, L. Fortnow, I. Newman, and H. Röhrig, "Quantum property testing," *SIAM Journal on Computing*, vol. 37, no. 5, pp. 1387–1400, 2008.

[22] S. Aaronson and A. Ambainis, "Forrelation: A problem that optimally separates quantum from classical computing," in *Proceedings of the 47th ACM Symposium on Theory of Computing (STOC 2015)*, 2015, pp. 307–316.

[23] M. Göös and T. S. Jayram, "A Composition Theorem for Conical Juntas," in *31st Conference on Computational Complexity (CCC 2016)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), vol. 50, 2016, pp. 5:1–5:16.

[24] B. W. Reichardt, "Reflections for quantum query algorithms," in *Proceedings of the 22nd ACM-SIAM Symposium on Discrete Algorithms (SODA 2011)*, ser. SODA '11, 2011, pp. 560–569. [Online]. Available: http://dl.acm.org/citation.cfm?id=2133036.2133080