SIGACT News Complexity Theory Column 67

Lane A. Hemaspaandra
Dept. of Computer Science, University of Rochester
Rochester, NY 14627, USA

Introduction to Complexity Theory Column 67

Warmest thanks to Arkadev and Toniann for telling the fascinating story of set disjointness. Upcoming guest articles in this column include Marius Zimand on Kolmogorov complexity extraction, Madhu Sudan on invariance in property testing, and John Watrous (topic TBA).

As this issue arrives in mailboxes, a new academic year will be starting. I wish each of you a year of exciting advances.

Guest Column: The Story of Set Disjointness

Arkadev Chattopadhyay and Toniann Pitassi[1]

1 Introduction

The satisfiability problem has emerged as the queen of the complexity zoo. She is the quintessential NP-complete hard-to-find but easy-to-recognize search problem in computer science. There are hundreds if not thousands of problems that are now known to be equivalent to SAT, and our rich theory of complexity classes is centered around its queen.

In the world of communication complexity, the set disjointness problem has similarly emerged as the quintessential hard-to-find but easy-to-recognize problem. There is an impressive collection of problems in many diverse areas whose hardness boils down to the hardness of the set disjointness problem in some model of communication complexity. Moreover, we will argue that proving lower bounds for the set disjointness function in a particular communication

[1]Computer Science Department, University of Toronto, Toronto, ON M5S1A4, CANADA. arkadev@cs.toronto.edu, toni@cs.torontno.edu. Supported by NSERC.

model (as opposed to proving lower bounds for some other function such as the inner product function), has often required important new insights and/or ideas.

In this article, we will first define the set disjointness function and its relatives. We will present several lower bounds for the set disjointness function in various communication models (deterministic, randomized, 2-party, multiparty, etc.), where our main goal will be to explain/expose the important lower bound techniques that have been developed to date in communication complexity. We will mention a handful of diverse applications, all of which require a lower bound for set disjointness.

2 Definitions

Two party communication complexity was first introduced in the seminal paper by Yao [57], and it has since been shown to have many diverse applications in complexity theory. (See [33] for an excellent exposition of the basic theory of communication complexity including applications. Two excellent sources for more advanced material are [35, 36].) The "number-on-forehead" model (NOF), first introduced by Chandra, Furst, and Lipton [15], generalizes the 2-party model. In this model, the input is partitioned into k parts, so that player i can see all parts except for the i^{th} part (since it is "written on his forehead").

Lower bounds for multiparty complexity in the number-on-forehead model are connected to major open problems in complexity theory: it has been established that $(\log n)^{\omega(1)}$ communication complexity lower bounds in the NOF model for any explicit function with polylogarithmically many players would imply explicit lower bounds for ACC^0 [11, 26]. The best lower bound obtained so far is $\Omega(n/2^k)$, which breaks down when the number of players is logarithmic [5, 19, 49, 23]. Lower bounds in the NOF model have many other important applications as well, including: constructions of pseudorandom generators for space bounded computation, universal traversal sequences, time-space tradeoffs [5], circuit complexity bounds [26, 44, 39], and proof complexity [9]. (Note: another generalization of the two-party model is the number-in-hand model. While this model is also quite interesting, with applications for streaming, we will focus on the number-on-forehead model in this paper.)

In the NOF multiparty communication complexity model of computation [15] there are k players, numbered 1 to k, that are trying to collaborate to compute a function $f_{k,n} : X_1 \times \ldots \times X_k \to \{0,1\}$ where each $X_i = \{0,1\}^n$. In general, we allow k to be a function of n. The kn input bits are partitioned into k sets, each of size n. For $(x_1, \ldots, x_k) \in \{0,1\}^{kn}$, and for each i, player i knows the values of all of the inputs except for x_i (which conceptually is thought of as being placed on player i's forehead).

The players exchange bits according to an agreed-upon protocol, by writing them on a public blackboard. A *protocol* specifies, for every possible blackboard contents, whether or not the communication is over, the output if over and the next player to speak if not. A protocol also specifies what each player writes as a function of the blackboard contents and of the inputs seen by that player. The *cost* of a protocol is the maximum number of bits written on the blackboard.

In a *deterministic protocol*, the blackboard is initially empty. A *public-coin randomized protocol* of cost c is simply a probability distribution over deterministic protocols of cost c, which can be viewed as a protocol in which the players have access to a shared random string. A *private-coin randomized protocol* is a protocol in which each player has access to a private random string. A *nondeterministic protocol* is a randomized private coin protocol with 1-sided error (only false negatives) and an error probability less than 1.

The *deterministic communication complexity of* $f_{k,n}$, written $D_k(f_{k,n})$, is the minimum cost of a deterministic protocol for $f_{k,n}$ that always outputs the correct answer. If $\mu : X_1 \times X_2 \times \ldots X_k \to [0,1]$ is a probability distribution and $\epsilon \geq 0$, the ϵ-error complexity of f for distribution μ, which we denote by $D_k^{\epsilon,\mu}(f_{k,n})$, is the minimum number of bits communicated

in any deterministic protocol π that computes f and errs on at most an ϵ fraction of the inputs with respect to μ. Note that $D_k^{\epsilon,\mu}(f_{k,n}) \leq D_k(f_{k,n}) \leq n+1$ for any f, μ, ϵ.

For $0 \leq \epsilon < 1/2$, let $R_{k,\epsilon}^{\text{pub}}(f_{k,n})$ denote the minimum cost of a public-coin randomized protocol for $f_{k,n}$ which, for every input, makes an error with probability at most ϵ (over the choice of the deterministic protocols). The *public-coin randomized communication complexity of* $f_{k,n}$ is $R_k^{\text{pub}}(f_{k,n}) = R_{k,1/3}^{\text{pub}}(f_{k,n})$. Let $R_{k,\epsilon}(f_{k,n})$ denote the minimum cost of a private-coin randomized protocol for $f_{k,n}$ which, for every input, makes an error with probability at most ϵ (over the choice of the private random strings). The *private-coin randomized communication complexity of* $f_{k,n}$ is $R_k(f_{k,n}) = R_{k,1/3}(f_{k,n})$. For both public-coin and private-coin complexities we add a superscript 1 if we require that the protocol makes error only on 1-inputs (i.e., false-negatives), and superscript 0 if we require that the protocol makes error only on 0-inputs (i.e., false-positives). For example, $R_{k,\epsilon}^{0,\text{pub}}(f_{k,n})$ is the minimum cost of a k-player public-coin protocol for $f_{k,n}$ which is always correct on 1-inputs and makes error at most ϵ on 0-inputs.

The standard way to prove lower bound on the randomized complexity of a function f is to prove lower bounds for the distributional complexity of f under a conveniently selected probability distribution over inputs. This works due to the following characterization by Yao.

Theorem 1 *(Yao's min-max principle)*

$$R_{k,\epsilon}(f_{k,n}) = max_\mu\{D_k^{\epsilon,\mu}(f_{k,n})\}.$$

Thus proving randomized lower bounds is equivalent to finding a distribution μ over the inputs such that any efficient protocol has large error with respect to μ. A *product distribution* on $X_1 \times X_2 \ldots \times X_k$ is a distribution of the form $\mu_{X_1} \times \mu_{X_2} \times \ldots \times \mu_{X_k}$ where μ_{X_i} is a distribution over X_i.

The *nondeterministic communication complexity of* $f_{k,n}$, written $N_k(f_{k,n})$, is the minimum cost of a nondeterministic protocol for $f_{k,n}$.

For a function $k = k(n)$, for a function family $f = (f_{k(n),n})_{n\in\mathbb{N}}$, and for any complexity measure C defined above, we write $C_k(f)$ for the function $(C_k(f))(n) = C_{k(n)}(f_{k(n),n})$.

Since any function $f_{k,n}$ can be computed using only $n+1$ bits of communication, following [4], for sequences of functions $f = (f_{k,n})_{n\in\mathbb{N}}$, communication protocols are considered "efficient" or "polynomial" if only polylogarithmically many bits are exchanged. Accordingly, let P_k^{cc} denote the class of function families f for which $D_k(f)$ is $(\log n)^{O(1)}$, let NP_k^{cc} denote the class of function families f with nondeterministic complexity $(\log n)^{O(1)}$, and let RP_k^{cc} denote the class of function families f for which $R_k^1(f_n)$ is $(\log n)^{O(1)}$. The classes BPP_k^{cc}, coRP_k^{cc} and coNP_k^{cc} can be defined similarly to their computational complexity counterparts.

Multiparty communication complexity lower bounds are proven by analyzing properties of functions on *cylinder intersections*.

Definition 2 An *i-cylinder* C_i in $X_1 \times \ldots \times X_k$ is a set such that for all $x_1 \in X_1, \ldots, x_k \in X_k, x_i' \in X_i$ we have $(x_1, \ldots, x_i, \ldots, x_k) \in C_i$ if and only if $(x_1, \ldots, x_i', \ldots, x_k) \in C_i$. A *cylinder intersection* is a set of the form $\bigcap_{i=1}^k C_i$ where each C_i is an i-cylinder in $X_1 \times \cdots \times X_k$.

Note that when $k = 2$, cylinder intersections become combinatorial rectangles; that is, $F \times G$, where $F \subseteq X_1$ and $G \subseteq X_2$. Any deterministic two-party b-bit communication protocol partitions $X_1 \times X_2$ into 2^b disjoint, monochromatic combinatorial rectangles; An ϵ-error two-party b-bit protocol over distribution μ partitions $X_1 \times X_2$ into 2^b disjoint combinatorial rectangles where each rectangle is nearly monochromatic. For simplicity of notation, we will suppress the subscript k in the case of two-party communication complexity.

Definition 3 (Set Disjointness) The *set disjointness* function family is $\text{DISJ} = (\text{DISJ}_{k,n})_{n\in\mathbb{N}}$, with $\text{DISJ}_{k,n} = 1$ if and only if there exists some $i \in [n]$ such that $x_{1,i} = \ldots\ x_{k,i} = 1$.

If we view the vectors x_i as sets, the set disjointness function is 1 if and only if the intersection of the sets is nonempty. It is easy to see that this problem lies in NP_k^{cc} for all k. For $k = 2$

it is complete; moreover it is complete in the number-in-hand model for every k. However for larger k it is unlikely to be NP_k^{cc}-complete. (NP_k^{cc}-completeness would imply that the number-on-forehead and number-in-hand models are equivalent.)

For data structure applications, it is often very useful to look at lopsided versions of the set disjointness problem (for two players), where Alice receives a small subset of the universe, and Bob receives an arbitrary subset of the universe.

Definition 4 (Lopsided Set Disjointness) The two-player (N, B) lopsided set disjointness function is as follows. The input is S, T where S is a set $S \subseteq [N \cdot B]$, $|S| = N$, and $T \subset [N \cdot B]$.

3 Lower Bounds for Set Disjointness

3.1 Two Party Deterministic Lower Bounds

We start our story with a very simple deterministic lower bound for two-party set disjointness using the "fooling set method."

Lemma 5 *Any two party deterministic protocol for solving DISJ requires $n+1$ bits of communication.*

Proof: Consider the 2^n different input pairs (x, y) such that $x + y$ (the bitwise sum of x and y) is the all 1 vector. Hence, for each such pair, $\mathrm{DISJ}(x, y) = 0$. We will show that no two of these pairs can be contained in the same rectangle. Suppose (x, y) and (x', y') are in the same rectangle, R. Then (x, y') and (x', y) are also contained in R. But this is not possible: since $x \neq x'$, there exists an index i such that $x_i \neq x'_i$. Suppose that $x_i = 0$ and $x'_i = 1$. Then $x'_i = y_i = 1$ and therefore $DISJ(x', y) = 1$, whereas $DISJ(x, y) = DISJ(x', y') = 0$. Similarly, if $x_i = 1$ and $x'_i = 0$, then $DISJ(x, y') = 1$, yielding a contradiction. These shows that we need at least 2^n rectangles to cover these input pairs. We need at least one other rectangle to cover points at which DISJ outputs 1. ∎

3.2 Two Party Randomized Lower Bounds for Product Distributions

Babai, Frankl, and Simon [4] established the first strong *randomized* lower bound for Disjointness.

Theorem 6 (Babai, Frankl, and Simon) *Consider the following product distribution μ on inputs: sample S, T independently and uniformly at random from the set of all subsets of $[n]$ that have cardinality $\lfloor \sqrt{n} \rfloor$. Then, $D^{\epsilon}_{\mu}(DISJ) = \Omega(\sqrt{n})$, for $\epsilon < 1/100$.*

Proof: Our distribution μ is a product distribution, where x is uniformly chosen from X, and y is uniformly chosen from Y, and where both X and Y consist of all vectors with exactly $\sqrt{n}$ 1's. We will prove the lower bound by showing that there are no large nearly 0-monochromatic rectangles. Specifically we will prove that for any rectangle $F \times G$, where at most an ϵ fraction of the pairs in $F \times G$ are intersecting, either $|F|$ or $|G|$ is small (less than $|X| \cdot 2^{-c\sqrt{n}+1}$), implying that the total number of rectangles is $exp(\sqrt{n})$, and thus $D^{\epsilon}_{\mu}(DISJ)$ is $\Omega(\sqrt{n})$.

We will say that F is large if $|F| \geq |X| \cdot 2^{-c\sqrt{n}+1}$; otherwise F is small. If F is small, then we are done. Otherwise, if $|F|$ is large, then there must be a large subset of $|F|$, where the union of these sets spans nearly all of $[n]$. But if $F \times G$ is nearly 0-monochromatic, this means that any subset $y \in G$ must avoid nearly all of $[n]$, and hence $|G|$ must be small. We proceed now to the details.

It suffices to show that if F is large, then G must be small. First, focus on $F_1 \subset F$, where F_1 are those vectors x that intersect with at most a 2ϵ fraction of the y's in G. Since $F \times G$ is nearly 0-monochromatic, $|F_1| \geq |F|/2 = |X| \cdot 2^{-c\sqrt{n}}$.

Since $|F_1|$ is still large, we claim that there exists $\sqrt{n}/3$ vectors $x_1, \ldots, x_k$ such that each x_l contains $\sqrt{n}/2$ new points relative to $x_1 \ldots x_{l-1}$. This can be proven by induction. Let z be the union of the sets x_i, $i < l$. We infer $|z| < l\sqrt{n} < n/3$. The number of $x \in X$ satisfying $|x \cap z| > \sqrt{n}/2$ is less than

$$n\binom{n/3}{\sqrt{n}/2}\binom{2n/3}{\sqrt{n}/2} < \binom{n}{\sqrt{n}}2^{-c\sqrt{n}}.$$

Therefore, $|x_l \cap z| < \sqrt{n}/2$ for some $x_l \in F$.

Take these $\sqrt{n}/3$ vectors where the ith one has $\sqrt{n}/2$ new elements (not in the first $i-1$ sets. Now we have a set of x's in F_1 whose union is of size at least $n/3$, and such that each of them intersects with only a few y's in G. But this means that G must be small: By a simple averaging argument, given any k elements in F_1, at most $|G|/2$ of the y's in G are good in that they intersect more than $4\epsilon k$ of the x_i. There are $\binom{k}{4\epsilon k}$ ways to select the $(4\epsilon k)$ of the x_i which a good $y \in G$ is allowed to intersect. Then the union of the remaining x_i's have size at least $n/9$ which must be avoided, so we get

$$|G| < 2\binom{k}{4\epsilon k}\binom{8n/9}{\sqrt{n}} < |Y|2^{-c\sqrt{n}}.$$

∎

It is worth noting that the above lower bound is essentially tight as Babai et al. show that $D^{\epsilon,\mu}(\text{DISJ}) = O(\sqrt{n}\log n)$ for every product distribution μ and every constant $\epsilon < 1$.

3.3 Two Party Randomized Lower Bounds For a Non-Product Distribution

We next show that by considering nonproduct distributions, one can improve the lower bound of Babai et al. to linear lower bounds. This was first achieved by Kalyanasundaram and Schnitger [27]. Razborov [51] provides a simplified argument for this. It is worthwhile to note that Raz [48] mentions that some of the ideas for proving his famous and difficult parallel repetition theorem are based on Razborov's technique.

Theorem 7 *Let $n = 4\ell - 1$ and consider the following stochastic process: first choose a random partition $P \equiv \{P_S, P_T, \{i\}\}$ of $[n]$, where $P_S, P_T \subset [n]$ with $|P_S| = |P_T| = 2\ell - 1$ and $i \in [n]$. Then, S (T) is a random subset of $P_S \cup \{i\}$ ($P_T \cup \{i\}$) with cardinality ℓ.*

If μ is the probability distribution on $S \times T$ corresponding to the above random process, then $D^{\epsilon}_{\mu}(DISJ) = \Omega(n)$, where ϵ is a small constant.

Before we begin the formal proof, let us note that the distribution μ is supported on two sets of inputs: the set of inputs, denoted by A, where S and T are disjoint and the set B where they intersect. Every set pair $(S,T) \in B$, is *barely* intersecting, i.e., $|S \cap T| = 1$. Hence, intuitively, it should be hard to distinguish A from B.

Proof: First note that $\mu(A)$ is large. This is because of the following: for *each* partition P, we generate a pair in B iff $i \in S$ and $i \in T$. Each of these happens with probability $1/2$. Hence, $\mu(B) = 1/4$. Thus, $\mu(A) = 3/4$. The argument establishes that every large rectangle R is corrupted w.r.t. A, i.e., almost a constant fraction of the probability mass of R rests on points in B. Formally,

Lemma 8 *There exists constants $\alpha, \delta > 0$ such that for every combinatorial rectangle $R = C \times D$,*

$$\mu(R \cap B) \geq \alpha\mu(R \cap A) - 2^{-\delta n}.$$

Roughly speaking, the above is established by analyzing the contribution of each partition P to R. In order to do so, we define $\mathrm{Row}(P) = \Pr\big[S \in C \,|\, P\big]$ and $\mathrm{Col}(P) = \Pr\big[T \in D \,|\, P\big]$. Further, let $\mathrm{Row}_0(P) = \Pr\big[S \in C \,|\, P, i \notin S\big]$ and $\mathrm{Row}_1(P) = \Pr\big[S \in C \,|\, P, i \in S\big]$. Likewise one defines $\mathrm{Col}_0(P)$ and $\mathrm{Col}_1(P)$. Then the following is simple to verify:

$$\begin{aligned} \mathrm{Row}(P) &= \frac{1}{2}\big(\mathrm{Row}_0(P) + \mathrm{Row}_1(P)\big), \\ \mathrm{Col}(P) &= \frac{1}{2}\big(\mathrm{Col}_0(P) + \mathrm{Col}_1(P)\big). \end{aligned} \tag{1}$$

Intuitively, the above defined quantities measure the contribution of each partition towards rectangle R. The right notion of contribution emerges from the following:

Fact 9

$$\begin{aligned} \mu\big(B \cap R\big) &= \frac{1}{4}\mathbb{E}_P\Big[\mathit{Row}_0\big(P\big)\,\mathit{Col}_0\big(P\big)\Big], \\ \mu\big(A \cap R\big) &= \frac{3}{4}\mathbb{E}_P\Big[\mathit{Row}_1\big(P\big)\,\mathit{Col}_1\big(P\big)\Big]. \end{aligned}$$

Proof:

$$\mu\big(A \cap R\big) = \mu\big(A\big)\mu\big(R|A\big)$$

Recalling $\mu(A) = 3/4$, we get $\mu\big(A \cap R\big) = \frac{3}{4}\mu(R|A)$. Now note that by symmetry,

$$\sum_P \Pr[P] \Pr\big[S = x | P, i \notin S\big] \Pr[T = y | P, i \notin T\big]$$

is just another way of writing the uniform distribution on A. Hence,

$$\mu\big(R|A\big) = \sum_P \Pr[P] \Pr\big[S \in R | P, i \notin S\big] \Pr[T \in R | P, i \notin T\big].$$

Thus, combining things, and plugging in the definition of $\mathit{Row}_0(P)$ and $\mathit{Col}_0(P)$, we are done for proving the claim w.r.t. $\mu(A \cap R)$. We leave the argument for $\mu(B \cap R)$ to the reader as it is very similar and slightly simpler.

∎

Having formulated how we are going to track the contribution of each partition P towards R, let us state when P contributes in a good way (with, of course, the aim of proving our corruption bound). We say P is S-bad if $\mathrm{Row}_1(P) < \mathrm{Row}_0(P)/3 - 2^{-\delta n}$. Similarly, P is T-bad if $\mathrm{Col}_1(P) < \mathrm{Col}_0(P)/3 - 2^{-\delta n}$. P is bad if it is S-bad or T-bad, otherwise it is good. Indeed, it is clear why good partitions help us in establishing the sought corruption bound. The next lemma shows that there are not many bad partitions.

Lemma 10 *(1) For every value of P_T, $\Pr_P\Big[P \text{ is } S-\mathit{bad} \,|\, P_T\Big] \leq \frac{1}{5}$. (2) Symmetrically, for every value of P_S, $\Pr_P\Big[P \text{ is } T-\mathit{bad} \,|\, P_S\Big] \leq \frac{1}{5}$.*

We defer the proof of this until later. Let us point out that we are not quite done. All we have established at this point is that by far, most partitions are good. It is still possible that the contribution of the bad partitions is significantly more than the good partitions. The next lemma rules this out. Let $\mathrm{Bad}_S(P)$ ($\mathrm{Bad}_T(P)$) be an indicator random variable for the event that P is S-bad (T-bad).

Lemma 11

$$\mathbb{E}_P\Big[Row_0(P)\,Col_0(P)\,Bad(P)\Big] \;\le\; \frac{4}{5}\mathbb{E}_P\Big[Row_0(P)\,Col_0(P)\Big].$$

Proof: We establish that $\mathbb{E}_P\big[\mathrm{Row}_0(P)\mathrm{Col}_0(P)\mathrm{Bad}_S(P)|P_T\big] \le \frac{2}{5}\mathbb{E}_P\big[\mathrm{Row}_0(P)\mathrm{Col}_0(P)|P_T\big]$ and symmetrically, $\mathbb{E}_P\big[\mathrm{Row}_0(P)\mathrm{Col}_0(P)\mathrm{Bad}_T(P)|P_S\big] \le \frac{2}{5}\mathbb{E}_P\big[\mathrm{Row}_0(P)\mathrm{Col}_0(P)|P_S\big]$. Clearly, adding the two inequalities yields our desired result.

We state some useful and easily verifiable facts:

Fact 12 *Col_0 and Row are just functions of P_T.*

Fact 13 $Row(P) = \frac{1}{2}\big(Row_0(P) + Row_1(P)\big)$.

Fact 14 $Row(P_T) = \mathbb{E}_P\Big[Row_0(P)|P_T\Big]$.

We apply these observations as below:

$$\begin{aligned}
\mathbb{E}_P\Big[\mathrm{Row}_0(P)\mathrm{Col}_0(P)\mathrm{Bad}_S(P)|P_T\Big] &=^{\text{Fact 12}} \mathrm{Col}_0(P_T)\mathbb{E}_P\Big[\mathrm{Row}_0(P)\mathrm{Bad}_S(P)|P_T\Big] \\
&\le^{\text{Fact 13}} \mathrm{Col}_0(P_T)\mathbb{E}_P\Big[2\mathrm{Row}(P)\mathrm{Bad}_S(P)|P_T\Big] \\
&=^{\text{Fact 12}} 2\mathrm{Col}_0(P_T)\mathrm{Row}(P_T)\mathbb{E}_P\Big[\mathrm{Bad}_S(P)|P_T\Big] \\
&\le^{\text{Lemma 10}} \frac{2}{5}\mathrm{Col}_0(P_T)\mathrm{Row}(P_T) \\
&=^{\text{Fact 14}} \frac{2}{5}\mathrm{Col}_0(P_T)\mathbb{E}_P\Big[\mathrm{Row}_0(P)|P_T\Big] \\
&=^{\text{Fact 12}} \frac{2}{5}\mathbb{E}_P\Big[\mathrm{Row}_0(P)\mathrm{Col}_0(P)|P_T\Big]
\end{aligned}$$

∎

We finally show below how knowing that the contribution of the bad partitions is not large allows us to establish the corruption bound.

$$\begin{aligned}
\mu(B\cap R) &=^{\text{Fact 9}} \frac{1}{4}\mathbb{E}_P\Big[\mathrm{Row}_1(P)\mathrm{Col}_1(P)\Big] \\
&\ge \frac{1}{4}\mathbb{E}_P\Big[\mathrm{Row}_1(P)\mathrm{Col}_1(P)\big(1-\mathrm{Bad}(P)\big)\Big] \\
&\ge \frac{1}{4}\mathbb{E}_P\Big[\Big(\mathrm{Row}_0(P)-2^{-\delta n}\Big)\Big(\mathrm{Col}_0(P)-2^{-\delta n}\Big)\big(1-\mathrm{Bad}(P)\big)\Big] \\
&> \frac{1}{4}\frac{1}{9}\mathbb{E}_P\Big[\mathrm{Row}_0(P)\mathrm{Col}_0(P)\big(1-\mathrm{Bad}(P)\big)\Big] - 2^{-\delta n} \\
&\ge^{\text{Lemma 11}} \frac{1}{4}\frac{1}{9}\frac{1}{5}\mathbb{E}_P\Big[\mathrm{Row}_0(P)\mathrm{Col}_0(P)\Big] - 2^{-\delta n} \\
&=^{\text{Fact 9}} \frac{1}{4}\frac{1}{9}\frac{1}{5}\frac{4}{3}\mu(A\cap R) - 2^{-\delta n}
\end{aligned}$$

Setting $\alpha = \frac{1}{4}\frac{1}{9}\frac{1}{5}\frac{4}{3}$ finishes the argument.

All that remains now is to prove Lemma 10. The intuition is simple. Consider partitions of $[n]$ such that P_T is a fixed set. Wlog assume that this fixed set is $\{1, \ldots, 2\ell-1\}$. Then, the range set of S has size $\binom{2\ell}{\ell}$ and the conditional distribution of S is just the uniform distribution on its range. If the set C of rows in the rectangle R is equal to the range of S (i.e., $\mathrm{Row}(P_T) = 1$), then clearly for every $i \geq 2\ell$, $\Pr[S \in C | i \in S] = \Pr[S \in C | i \notin S]$. It is natural to expect that the two probabilities will be close to each other for most i if C is a large subset of the range of S. This is what we formally show below via a simple entropy argument.

For any i, let $C_S = \{x \in C | x \subseteq [n] - P_T\}$. Then, $C_i = \{x \in C_S | i \in x\}$ and $C_{-i} = \{x \in C_S | i \notin x\}$. Further, let $\binom{2\ell}{\ell}_{-i}$ (and $\binom{2\ell}{\ell}_{i}$) denote the set of those ℓ-subsets of $[n] - P_T$ that do not contain i (and contain i). Note that $|\binom{2\ell}{\ell}_{i}| = |\binom{2\ell}{\ell}_{-i}| = \frac{1}{2}\binom{2\ell}{\ell}$. Thus,

$$\mathrm{Row}_0(P_T, i) = \frac{|C_{-i}|}{|\binom{2\ell}{\ell}_{-i}|} = 2\frac{|C_{-i}|}{|C_S|}\frac{|C_S|}{\binom{2\ell}{\ell}}$$

and

$$\mathrm{Row}_1(P_T, i) = \frac{|C_i|}{|\binom{2\ell}{\ell}_{i}|} = 2\frac{|C_i|}{|C_S|}\frac{|C_S|}{\binom{2\ell}{\ell}}$$

Hence if selecting i makes the partition P S-bad, then $\frac{|C_i|}{|C_S|} < \frac{1}{3}\frac{|C_{-i}|}{|C_S|}$. In other words, if we select a set x uniformly at random from C_S then with probability less than a $1/4$, i is in x. We show that this cannot be true for more than a fifth of the 2ℓ indices i, if the size of C_S is large, i.e., $|C_S| \geq 2^{-\delta n}\binom{2\ell}{\ell}$.

Assume the contrary, i.e., at least a fifth of the indices are bad. Wlog, let $[n] - P_T = \{1, \ldots, 2\ell\}$. Consider 2ℓ random boolean variables, $s_1, \ldots, s_{2\ell}$, one for each of the 2ℓ elements of $[n] - P_T$. We pick a subset at random from C_S. Variable s_i takes value 1 iff i is an element of the random set. By basic properties of entropy,

$$H(s_1, \ldots, s_{2\ell}) \leq H(s_1) + \cdots + H(s_{2\ell}).$$

By our assumption, for at least a fifth of i's, $H(s_i) < H(1/4)$. So,

$$H((s_1, \ldots, s_{2\ell}) < 2\ell(\frac{1}{5}H(1/4) + \frac{4}{5}).$$

However, by our assumption on the size of C_S and the definition of entropy,

$$H(s_1, \ldots, s_{2\ell}) = \log(|C_S|) \geq \log\left(\binom{2\ell}{\ell}2^{-\delta n}\right) = \Theta(2\ell - \delta n).$$

Noting that $H(1/4)$ is a constant strictly less than 1, we observe that the above gives a contradiction by choosing a sufficiently small constant δ. This finishes the entire argument. ∎

3.4 Two Party Linear Bounds using Information Complexity

A communication protocol aims to reveal the minimum information about inputs held by players that still allows them to compute the target function. Thus, one may hope to prove lower bounds on the communication complexity of a target function by showing that every protocol is forced to reveal large information about the inputs. Although this sounds like a mere change of language, this point of view could exploit powerful, yet fairly intuitive, tools from information theory. Indeed, Bar-Yossef et al. [6] reproved an $\Omega(n)$ lower bound on the communication complexity of set-disjointness using a very elegant information theoretic argument. They formalize the idea mentioned above through the notion of an *information cost* of a protocol. This concept was

introduced in the earlier work of Chakrabarti et al. [14] and is implicit in earlier works [2, 52]. We sketch the argument of Bar-Yossef et al. below, starting with basic definitions.

Let Ω be a finite set and let P be a probability distribution over Ω. The entropy of a random variable X distributed according to P is

$$H(X) = \sum_{x \in \Omega} P(x) \log \left(\frac{1}{P(x)} \right).$$

Entropy quantifies the amount of uncertainty in a distribution. The conditional entropy $H(X|Y)$ is equal to $E_y[H(X|Y=y)]$, where $H(X|Y=y) = \sum_x P(x|y) \log \frac{1}{P(x|y)}$.
Finally the joint entropy of X, Y is $H(X,Y) = H(X) + H(Y|X)$.

For two random variables Z and Π, the mutual information is defined as

$$I(Z;\Pi) = H(Z) - H(Z|\Pi) = H(\Pi) - H(\Pi|Z).$$

Intuitively, this is the average amount of uncertainty about X given that we know Π, or symmetrically[2] the amount of uncertainty about Π given that we know X.

In communication complexity, we will be studying the mutual information between (X,Y) and Π, where (X,Y) is the distribution over inputs and Π is the communication transcript generated by a protocol.

Consider a random variable $\Pi((X,Y), R_A, R_B)$ over transcripts. This variable depends on the distribution of inputs (X,Y) and Alice's random bits R_A and Bob's random bits R_B. For the discussion in this section, it will be convenient to assume that each player uses its own random coin-tosses. Define the information cost of the protocol Π as $I((X,Y);\Pi)$, i.e., it is the information learned about the inputs by an eavesdropper who only has access to the transcript. The ϵ-error information cost of a function, $IC_\epsilon(f)$, is the minimum information cost over all randomized protocols for f that err with probability at most ϵ. The first thing to note is that clearly the information complexity of a function is at most its communication complexity as the information revealed by a protocol can be at most the total number of bits communicated in a transcript.

The following fact represents a simple but useful direct sum property of mutual information.

Fact 15 *If $Z = (Z_1, \ldots, Z_n)$ are mutually independent then*

$$I(Z;\Pi) \geq I(Z_1;\Pi) + \ldots + I(Z_n;\Pi).$$

We define a distribution on inputs where $(X_1,Y_1), \ldots, (X_n,Y_n)$ are mutually independent. In this case, using Fact 15,

$$I((X,Y);\Pi) \geq I((X_1,Y_1);\Pi) + \ldots + I((X_n,Y_n);\Pi).$$

The distribution μ we will use on (X_i,Y_i) is $P(0,0) = 1/2$, $P(1,0) = 1/4$, $P(0,1) = 1/4$. Although μ is not a product distribution, we sample by viewing it as a mixture of product distributions. First, choose D_i to be 0 or 1, each with probability 1/2. Once D_i is chosen, then we have a product distribution on (X_i,Y_i) as follows: if $D_i = 0$, then set $X_i = 0$ and otherwise select X_i to be 0 or 1 each with probability 1/2; if $D_i = 1$, then set $Y_i = 0$ and otherwise select Y_i to be 0 or 1 each with probability 1/2 again. Let this distribution on $(X_i,Y_i), D_i$ be denoted by ν. Given such a mixture of product distributions ν, the conditional information cost of a protocol Π is defined as $I((X,Y);\Pi\,|D)$. The ϵ-error conditional information complexity of a function f, denoted by $CIC_\epsilon(f)$, is then the minimal conditional information cost over all ϵ-error protocols for f.

We next outline the proof of the following useful fact.

[2] the fact that mutual information is a symmetric quantity requires an argument.

Claim: Let Π be any ϵ-error protocol for computing Disjointness and let AND denote the conjunction of two bits. Then,

$$I((X_i, Y_i); \Pi \mid D) \geq CIC_\epsilon(AND).$$

Proof: Let $D_{-i} = D_1, \ldots, D_{i-1}, D_{i+1}, \ldots, D_n$. Then, by definition, $I((X_i, Y_i); \Pi \mid D) = \mathbb{E}_d\big[I((X_i, Y_i); \Pi \,|\, D_i, D_{-i} = d)\big]$. We derive a protocol Π_d for AND from Π for every d that errs with probability at most ϵ.

Given two bits, U to Alice and V to Bob, Alice generates $X_{-i} = X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ and Bob generates $Y_{-i} = Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n$ conditioned on $D_{-i} = d$. They do this by themselves as any pair X_j and Y_j are independent given D_j for any $j \neq i$. They, then embed $X_i = U$ and $Y_i = V$ and run Π on X, Y. The key thing to note is that $AND(U, V) = DISJ(X, Y)$.

The argument gets completed by simply verifying that the joint distribution of (U, V, D, Π_d) is identical to that of $(X_i, Y_i, D_i, \Pi(X, Y))$ conditioned on $D_{-i} = d$. ∎

Combining Fact 15 and Claim 3.4, one concludes the following:

Theorem 16 $IC_\epsilon\big(DISJ\big) \geq n \cdot CIC_\epsilon\big(AND\big)$.

Thus, it suffices to prove an $\Omega(1)$ lower bound on the conditional information complexity of the AND function on two bits. In the remaining part of this section, we prove such a lower bound for the mixture ν of product distributions.

Let Γ be any ϵ-error protocol for AND. We want to lower bound $I((U, V); \Gamma \mid D) = \frac{1}{2} I((U, V); \Gamma \mid D = 0) + \frac{1}{2} I((U, V); \Gamma \mid D = 1)$. As these are symmetrical, we will focus on $I((U, V); \Gamma \mid D = 1) = I(U; \Gamma(U, 0) \mid D = 1)$. It is intuitive to expect that the information above is related to some appropriately defined notion of distance between the two distributions $\Gamma(1, 0)$ and $\Gamma(0, 0)$. Bar-Yossef et al. discovered that the *Hellinger* distance is a convenient measure for that. Let $\Omega = \{\omega_1, \ldots, \omega_t\}$ be a discrete domain. Any probability distribution μ over Ω can naturally be viewed as a unit vector Ψ_μ (with the euclidean norm) in $\mathbb{R}^\Omega$ whose jth co-ordinate is simply $\sqrt{\mu(\omega_j)}$. With this transformation, the Hellinger distance between two vectors Ψ_1 and Ψ_2, denoted by $h(\Psi_1, \Psi_2)$, is just $1/\sqrt{2}$ of the Euclidean distance between Ψ_1 and Ψ_2, i.e., $h(\Psi_1, \Psi_2) = \frac{1}{\sqrt{2}} ||\Psi_1 - \Psi_2||$. This immediately implies that Hellinger satisfies the triangle inequality. We need three following key properties of Hellinger distance:

- **Hellinger distance and Information:** Let $u, v \in \{0, 1\}^2$ be two inputs to AND, and $U \in_R \{u, v\}$. As before, let $\Psi(u)$ be the unit vector formed by the entrywise square root of $\Gamma(u)$.
$$I(U; \Gamma) \geq \frac{1}{2} ||\Psi(u) - \Psi(v)||^2.$$
- **Soundness:** If $AND(u) \neq AND(v)$ and Γ is a protocol with error at most ϵ, then we expect that the two distributions on transcripts $\Gamma(u)$ and $\Gamma(v)$ are far apart. Indeed, one can easily show,
$$\frac{1}{2} ||\Psi(u) - \Psi(v)||^2 \geq 1 - 2\sqrt{\epsilon}.$$
- **Cut and Paste:** Let $u = (x, y)$, $v = (x', y')$, $u' = (x, y')$ and $v' = (x', y)$. Then, the rectangular nature of deterministic 2-party protocols is captured by the following simple fact: if such a protocol generates the same transcript for input instances u and v, then this transcript is also generated for instances u' and v'. This property manifests itself on transcript distributions for randomized protocols in the following natural form:
$$||\Psi(u) - \Psi(v)|| = ||\Psi(u') - \Psi(v')||.$$

The Theorem follows from an application of Cauchy-Schwartz plus the above three properties:

$$\begin{aligned}
I((U,V);\Gamma \mid D) &= \frac{1}{2}||\Psi(0,0)-\Psi(1,0)||^2 + ||\Psi(0,0)-\Psi(0,1)||^2 \\
&\geq^{\text{Cauchy-Schwartz}} \frac{1}{4}(||\Psi(0,0)-\Psi(1,0)|| + ||\Psi(0,0)-\Psi(0,1)||)^2 \\
&\geq^{\text{Triangle Inequality}} \frac{1}{4}||\Psi(1,0)-\Psi(0,1)||^2 \\
&=^{\text{Cut-Paste}} \frac{1}{4}||\Psi(0,0)-\Psi(1,1)||^2 \\
&\geq^{\text{Soundness}} \frac{1}{2}(1-2\sqrt{\epsilon}).
\end{aligned}$$

This immediately yields the desired linear bound.

Theorem 17 *The randomized complexity of Disjointness is at least* $\frac{1}{2}\big(1-2\sqrt{\epsilon}\big)n$.

3.5 NOF Lower Bounds and the Generalized Discrepancy Method

Recall that we described three different techniques to prove strong lower bounds on the randomized 2-party communication complexity of Disjointness. It is not known if any of these techniques can be extended to three or more players in the NOF model. In fact, until recently, the best known lower bound for Disjointness in the k-player model was $\Omega(\log n)$, due to Tesson [56]. This was significantly improved to $n^{\Omega(1)}$ for any constant k in the independent works of Lee and Shraibman [34] and Chattopadhyay and Ada [18]. Both build upon the recent breakthrough work of Sherstov [54]. In this section, we give an overview of these developments.

The main difficulty one faces in the k-player case is that k-wise cylinder intersections are frustratingly difficult to analyze. Recall that a k-player protocol partitions the input space into such cylinder intersections. When $k = 2$, these intersections correspond to rectangles, a much simpler object to understand. For instance, we presented a relatively simple argument showing that at least $2^n + 1$ rectangles are needed to partition the input space into monochromatic rectangles w.r.t. Disjointness. No such simple argument is known for proving the same for 3-wise cylinder intersections.

One successful way of handling cylinder intersections was introduced in the seminal work of Babai, Nisan, and Szegedy [5] that proved the first strong lower bounds in the NOF model. They employ an analytical trick achieved by using the Cauchy-Schwartz inequality very elegantly that we briefly review later in this section. However, the only known way to use this trick is in computing cylinder intersection discrepancy of functions. Using this technique, [5] were able to prove exponentially small upper bounds on the discrepancy of some functions like the Generalized Inner Product. Later, [49, 19, 23] have shown the applicability of this technique to a wider class of functions.

Consequently, this technique could not be made to work for functions that have polynomially high discrepancy. Disjointness is a cardinal example of such a function. In order to describe how these recent works overcame the problem, let us quickly review some basic notions.

For a distribution μ over $\{0,1\}^{X_1\times\cdots\times X_k}$ the discrepancy of function f over a cylinder intersection C, denoted by $\text{disc}_\mu^C(f)$, is given by $\left|\sum_{(x_1,\ldots,x_k)\in C} f(x_1,\ldots,x_k)\mu(x_1,\ldots,x_k)\right|$. Here, wlog, we have assumed f to be 1/-1 valued. Thus, a small upper bound on the discrepancy implies that all cylinder intersections that have significant probability mass under μ are far from being monochromatic, i.e., their mass is almost equally distributed between the set of points where f evaluates to 1 and the set of points where it evaluates to -1. It is not hard to verify that $D_k^{\epsilon,\mu}(f) \geq \log\left(\frac{2\epsilon}{\text{disc}_{\mu,k}(f)}\right)$. Thus, proving an exponentially small upper bound on the discrepancy yields strong lower bounds on the communication complexity.

3.5.1 Dealing With High Discrepancy

We first observe that Disjointness has high discrepancy with respect to every distribution.

Lemma 18 (Folklore) *Under every distribution μ over the inputs, $disc_{k,\mu}(DISJ_k) \geq \frac{1}{2n} - \frac{1}{2n^2}$.*

Proof: Let X^+ and X^- be the set of disjoint and nondisjoint inputs respectively. The first thing to observe is that if $|\mu(X^+) - \mu(X^-)| \geq (1/n)$, then we are done immediately by considering the discrepancy over the intersection corresponding to the entire set of inputs. Hence, we may assume $|\mu(X^+) - \mu(X^-)| < (1/n)$. Thus, $\mu(X^-) \geq 1/2 - (1/2n)$. However, X^- can be covered by the following n *monochromatic* cylinder intersections: let C_i be the set of inputs in which the ith column is an all-one column. Then $X^- = \cup_{i=1}^n C_i$. By averaging, there exists an i such that $\mu(C_i) \geq 1/2n - (1/2n^2)$. Taking the discrepancy of this C_i, we are done. ∎

It is therefore impossible to obtain better than $\Omega(\log n)$ bounds on the communication complexity of Disjointness by a direct application of the discrepancy method. In fact, the above argument shows that this method fails to give better than polylogarithmic lower bounds for any function that is in NP_k^{cc} or co-NP_k^{cc}.

Fortunately, there is a simple generalization of the Discrepancy Method that is effective for dealing with several functions that have large discrepancy. The origins of this idea can be found in the work of Klauck [29][3]. Klauck considered, in the setting of two players, functions of the form $f(x, y) = g(x \wedge y)$ where the $\wedge$ operation is naturally applied bitwise to the bits of x and y. He observed that if g correlates well with a parity function on some large subset S of $\{1, \ldots, n\}$ under the uniform distribution[4], then f correlates well with the inner-product function of the columns indexed by elements of S, denoted by IP_S, under a simple product distribution μ. The ingenuity in Klauck's argument is that he shows IP_S having small discrepancy under μ implies that f has large distributional complexity under μ. This, as he correctly adds, follows despite the possibility that f itself has large discrepancy. Indeed, Klauck proves that IP has very small rectangular discrepancy under μ. Klauck goes on to show that this "generalized form of the discrepancy method" can be used to obtain a lower bound of $\Omega(n/\log n)$ on the quantum (and hence classical randomized) communication complexity of $\mathrm{MAJ}(x \wedge y)$ despite the fact that it has large discrepancy.

The main idea in Klauck's work was abstracted by Sherstov [54] in following terms: A function f may have high discrepancy and still correlate well under some distribution μ with a function h that has small discrepancy under μ. Exhibiting such a h, yields lower bounds on the bounded-error communication complexity of f. We re-express it in a form that appears in [18] and follows straightforwardly from basic definitions: for functions f, g having range $\{1, -1\}$, and a distribution μ over their common domain, define their correlation, denoted by $\mathrm{Corr}_\mu(f, g)$, to be $\mathbb{E}_{x \sim \mu}\big[f(x)g(x)\big]$. Then,

Lemma 19 (Generalized Discrepancy Method) *Denote $X = Y_1 \times ... \times Y_k$. Let $f : X \to \{-1, 1\}$ and $g : X \to \{-1, 1\}$ be such that under some distribution μ we have $Corr_\mu(f, g) \geq \delta$. Then*

$$D_k^{\epsilon,\mu}(f) \geq \log\left(\frac{\delta + 2\epsilon - 1}{disc_{k,\mu}(g)}\right). \tag{2}$$

3.5.2 Dual Polynomials

The main challenge in applying the Generalized Discrepancy Method to a given function like Disjointness is the following: how do we come up with a function g and distribution μ such that

[3]The full version of Klauck's work appears in [30].

[4]In other words, g has a large high-order Fourier coefficient, i.e., $\hat{f}(S)$ is large.

g correlates well with Disjointness under μ and g has small k-wise discrepancy under μ. This was achieved in the independent works of Sherstov [54] and Shi and Zhu [55] by a clever use of dual polynomials. In order to describe that, we have to recall some notions from the theory of polynomial representations of boolean functions.

We view the boolean cube as $\mathcal{B}^n \equiv \{1,-1\}^n$. Then the space of functions from the cube to reals is a vector space of dimension 2^n. A convenient basis for this is the Fourier basis of all parities or, equivalently, multilinear monomials $\mathcal{M} = \{\chi_S = \prod_{i \in S} x_i \,|\, S \subseteq [n]\}$. Thus, every boolean function $f : \{1,-1\}^n \to \{1,-1\}$ is uniquely represented by a real linear combination of monomials from $\mathcal{M}$, i.e., a polynomial with real coefficients. The *exact degree* of f is the degree of this polynomial. It is well-known that the degree of the functions OR, AND and Parity is n.

One could naturally relax the notion of representation as follows: a polynomial P that is always within δ of the function f is a δ-approximation of f, i.e., $|f(x) - P(x)| \leq \delta$ for each $x \in \{1,-1\}^n$ and $\delta \geq 0$. The δ-approximation degree of f, denoted by $\deg_\delta(f)$, is the minimal degree such that there exists a polynomial of that degree which is a δ-approximation of f. It follows that $\deg(f) \leq \deg_\delta(f)$ for any $\delta < 1$. The following result, due to Nisan and Szegedy, shows that even this relaxed notion does not decrease the degree for AND substantially.

Theorem 20 (Nisan and Szegedy [43]) *Let f be either the AND or the OR function. Then, $deg_{1/3}(f) = \Theta(\sqrt{n})$.*

One way to interpret approximation degree is the following: if f has approximation degree d, then it is at a large distance from the linear space spanned by monomials of degree less than d. It is natural to expect that the projection of f on the dual space spanned by characters of degree at least d is large. This intuition works and is formalized below:

Lemma 21 (Approximation-Orthogonality [54, 55]) *Let $f : \{-1,1\}^n \to \mathbb{R}$ be given with $deg_\delta(f) = d \geq 1$. Then there exists $g : \{-1,1\}^n \to \{-1,1\}$ and a distribution μ on $\{-1,1\}^n$ such that $Corr_\mu(f,g) > \delta$ and $Corr_\mu(f,\chi_S) = 0$ for all $|S| < d$.*

This Approximation-Orthogonality Principle is a classical result in functional analysis. There are two known ways, closely related to each other, to exploit it for applying the Generalized Discrepancy Method to the Disjointness function. One is due to Sherstov [54] and the other due to Shi and Zhu [55]. Both of them were originally used to obtain lower bounds for Disjointness in the 2-player quantum communication model. Sherstov's strategy, that he called the pattern matrix method, yields tight lower bounds of $\Omega(\sqrt{n})$ that was first obtained by Razborov [50] using a different and more involved argument. The strategy of Shi and Zhu, called the block composition method, yields less than tight, but still $n^{\Omega(1)}$ lower bound. The pattern matrix strategy was extended to the multiparty NOF model by Lee and Shraibman [34] and Chattopadhyay and Ada [18] independently. Chattopadhyay [17] extended the block-composition technique to the multiparty setting. Both extensions yield $n^{\Omega(1)}$ lower bounds on the k-party communication complexity if k is a constant, that significantly improves previous bounds. Due to limitations of space, we describe the pattern matrix extension since it yields stronger bounds that remain interesting and nontrivial up to k slightly less than $\log \log n$.

3.5.3 Pattern Tensors

We lay out the general strategy for proving lower bounds on the k-party communication complexity of a function G: we start with an appropriate function $f : \{0,1\}^n \to \{-1,1\}$ with high approximation degree. By the Approximation-Orthogonality Principle, we obtain g that highly correlates with f and is orthogonal with low degree polynomials under a distribution μ. From f and g we construct new *masked functions* F_k^f and F_k^g with the property that F_k^f is a promised or restricted version of the target function G, i.e., the set of inputs of F_k^f is a subset

of the set of inputs of G and on this subset F_k^f behaves exactly like G. Using the property that g is orthogonal to low degree polynomials (under distribution μ), we deduce that F_k^g has low discrepancy under an appropriate distribution λ that is naturally constructed out of μ. This calculation of discrepancy uses ideas from the work of Babai, Nisan, and Szegedy [5]. Under distribution λ, F_k^g and F_k^f remain highly correlated and therefore applying the Generalized Discrepancy Method, we conclude that F_k^f has high randomized communication complexity. This completes the argument as this implies that even a restricted form of G, represented as F_k^f, has large communication complexity.

It is worthwhile to note that this strategy involves taking a function f that is mildly hard (is hard for polynomials to approximate pointwise) and generating a function F_k^f that is much harder (hard for k-party protocols to approximate). Such hardness amplification is a recurring theme in different areas of complexity. In particular, there is a compelling similarity with the much earlier work of Krause and Pudlák [32] in which the authors did the following amplification: they showed that if f is hard in the sense that it cannot be sign represented by low degree polynomials, then F_2^f is harder in the sense that it cannot be sign represented by polynomials with few monomials. Krause and Pudlák give the construction only for $k = 2$ and for this k it coincides with the pattern matrix construction that we present here. However, the hardness for f assumed in Krause and Pudlák (high sign-degree) is stronger than the assumption of high approximation degree considered here. The conclusion about the hardness of F_k^f are incomparable in the two cases.

Our description below is taken from [18]. Let $S^1, \ldots, S^{k-1} \in [\ell]^m$ for some positive ℓ and m. Let $x \in \{0,1\}^n$ where $n = \ell^{k-1}m$. Here it is convenient to think of x to be divided into m equal blocks where each block is a $k-1$-dimensional array with each dimension having size ℓ. Each S^i is a vector of length m with each co-ordinate being an element from $\{1, \ldots, \ell\}$. The $k-1$ vectors $S^1, \ldots, S^{k-1}$ jointly unmask m bits of x, denoted by $x \leftarrow S^1, \ldots, S^{k-1}$, precisely one from each block of x, i.e.,

$$x_1\big[S^1[1], S^2[1], ..., S^{k-1}[1]\big], \ldots, x_m\big[S^1[m], S^2[m], \ldots, S^{k-1}[m]\big].$$

where x_i refers to the ith block of x.

For a given base function $f : \{0,1\}^m \to \{-1,1\}$, we define $F_k^f : \{0,1\}^n \times ([\ell]^m)^{k-1} \to \{-1,1\}$ as $F_k^f(x, S^1, \ldots, S^{k-1}) = f(x \leftarrow S^1, \ldots, S^{k-1})$.

Now we prove that if the base function f has a certain nice property, then the masked function F_k^f has small discrepancy. To describe the nice property, let us define the following: for a distribution μ on the inputs, f is (μ, d)-orthogonal if $\mathbb{E}_{x\sim\mu} f(x)\chi_S(x) = 0$, for all $|S| < d$. Then,

Lemma 22 (Orthogonality-Discrepancy Lemma) *Let $f : \{-1,1\}^m \to \{-1,1\}$ be any (μ, d)-orthogonal function for some distribution μ on $\{-1,1\}^m$ and some integer $d > 0$. Derive the probability distribution λ on $\{-1,1\}^n \times \big([\ell]^m\big)^{k-1}$ from μ as follows: $\lambda(x, S^1, \ldots, S^{k-1}) = \frac{\mu(x\leftarrow S^1,\ldots,S^{k-1})}{\ell^{m(k-1)}2^{n-m}}$. Then,*

$$\left(disc_{k,\lambda}\big(F_k^f\big)\right)^{2^{k-1}} \leq \sum_{j=d}^{(k-1)m} \binom{(k-1)m}{j} \left(\frac{2^{2^{k-1}-1}}{\ell - 1}\right)^j \tag{3}$$

Hence, for $\ell - 1 \geq \frac{2^{2^k}(k-1)em}{d}$ and $d > 2$,

$$disc_{k,\lambda}\big(F_k^f\big) \leq \frac{1}{2^{d/2^{k-1}}}. \tag{4}$$

Remark The Lemma above appears very similar to the Multiparty Degree-Discrepancy Lemma in [16] that is an extension of the two party Degree-Discrepancy Theorem of [53]. There, the magic property on the base function is high voting degree. It is worth noting that (μ, d)-orthogonality of f is equivalent to voting degree of f being at least d. Indeed the proof of the above Lemma is almost identical to the proof of the Degree-Discrepancy Lemma save for the minor details of the difference between the two masking schemes.

Proof: We briefly outline the main steps involved here. The missing details can be easily filled in from [18]. The starting point is to write the expression for discrepancy w.r.t. an arbitrary cylinder intersection ϕ,

$$\mathrm{disc}_k^{\phi}(F_k^f) = 2^m \left| \mathbb{E}_{x,S^1,\ldots,S^{k-1}} F_k^f(x, S^1, \ldots, S^{k-1}) \times \ \phi(x, S^1, \ldots, S^{k-1}) \mu\big(x \leftarrow S^1, \ldots, S^{k-1}\big) \right| \tag{5}$$

where, $(x, S^1, \ldots, S^{k-1})$ is uniformly distributed over $\{0,1\}^{\ell^{k-1}m} \times \big([\ell]^m\big)^{k-1}$.

Applying repeatedly Cauchy-Schwartz inequality along with triangle inequality (very similar to in [16, 49]), one rewrites

$$\left(\mathrm{disc}_k^{\phi}(F_k^f)\right)^{2^{k-1}} \leq 2^{2^{k-1}m} \mathbb{E}_{S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}} H_k^f\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) \tag{6}$$

where,

$$\begin{aligned} &H_k^f\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) \\ &= \left| \mathbb{E}_{x \in \{0,1\}^{\ell^{k-1}m}} \prod_{u \in \{0,1\}^{k-1}} \left(F_k^f(x, S_{u_1}^1, \ldots, S_{u_{k-1}}^{k-1}) \mu(x \leftarrow S_{u_1}^1, \ldots, S_{u_{k-1}}^{k-1}) \right) \right| \end{aligned} \tag{7}$$

We look at a fixed S_0^i, S_1^i, for $i = 1, \ldots, k-1$. Let $r_i = \left|S_0^i \cap S_1^i\right|$ and $r = \sum_i r_i$ for $1 \leq i \leq 2^{k-1}$. We now make two claims:

Claim:

$$H_k^f\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) \leq \frac{2^{(2^{k-1}-1)r}}{2^{2^{k-1}m}}. \tag{8}$$

Claim: Let $r < d$. Then,

$$H_k^f\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) = 0. \tag{9}$$

We leave the proof of these claims with the following remarks. Claim 3.5.3 simply follows from the fact that μ is a probability distribution and f is 1/-1 valued while Claim 3.5.3 uses the (μ, d)-orthogonality of f. We now continue with the proof of the Orthogonality-Discrepancy Lemma assuming these claims. Applying them, we obtain

$$\begin{aligned} &\left(\mathrm{disc}_k^{\phi}(F_k^f)\right)^{2^{k-1}} \\ &\leq \sum_{j=d}^{(k-1)m} 2^{(2^{k-1}-1)j} \sum_{j_1+\cdots+j_{k-1}=j} \Pr\left[r_1 = j_1 \wedge \cdots \wedge r_{k-1} = j_{k-1}\right]. \end{aligned} \tag{10}$$

Substituting the value of the probability, we further obtain:

$$\begin{aligned} &\left(\mathrm{disc}_k^{\phi}(F_k^f)\right)^{2^{k-1}} \\ &\leq \sum_{j=d}^{(k-1)m} 2^{(2^{k-1}-1)j} \sum_{j_1+\cdots+j_{k-1}=j} \binom{m}{j_1} \cdots \binom{m}{j_{k-1}} \frac{(\ell-1)^{m-j_1} \cdots (\ell-1)^{m-j_{k-1}}}{\ell^{(k-1)m}}. \end{aligned} \tag{11}$$

The following simple combinatorial identity is well known:

$$\sum_{j_1+\cdots+j_{k-1}=j} \binom{m}{j_1} \cdots \binom{m}{j_{k-1}} = \binom{(k-1)m}{j}.$$

Plugging this identity into (11) immediately yields (3) of the Orthogonality-Discrepancy Lemma. Recalling $\binom{(k-1)m}{j} \leq \left(\frac{e(k-1)m}{j}\right)^j$, and choosing $\ell - 1 \geq 2^{2^k}(k-1)em/d$, we get (4). ∎

We can now combine things to obtain the main theorem about the k-party communication complexity of Disjointness below.

Theorem 23 ([34, 18])

$$R_{k,\epsilon}\big(DISJ_{k,n}\big) = \Omega\left(\frac{n^{\frac{1}{k+1}}}{2^{2^k}(k-1)2^{k-1}}\right)$$

for any constant $\epsilon > 0$.

Proof: Let $f = \mathrm{OR}_m$ on m variables. The theorem of Nisan and Szegedy gives $\deg_{1/3}(\mathrm{OR}_m) = \Theta(\sqrt{m}) = d$. Consider F_k^f. It is simple to verify that this is a restricted version of k-wise Disjointness, for $n = \ell^{k-1}m$. Applying Approximation-Orthogonality principle, we obtain a g and a distribution μ such that g is (μ, d)-orthogonal for $d = \Theta(\sqrt{m})$ and $\mathrm{Corr}(\mathrm{OR}, g) \geq 1/3 = \delta$. Orthogonality-Discrepancy Lemma prescribes us to set $\ell = 2^{2^k}(k-1)em/d$. This implies $n = \left(\frac{2^{2^k}(k-1)e}{\deg_{1/3}(\mathrm{OR}_m)}\right)^{k-1} m^k$. For, these setting of parameters, we conclude

$$\mathrm{disc}_{k,\lambda}\big(F_k^g\big) \leq \frac{1}{2^{d/2^{k-1}}}.$$

Noting that $\mathrm{Corr}_\lambda\big(F_k^{\mathrm{OR}}, F_k^g\big) = \mathrm{Corr}_\mu\big(\mathrm{OR}, g\big) \geq 1/3$ and expressing all parameters in terms of n, we apply the Generalized Discrepancy Method to get the desired result for any constant $\epsilon > 1/6$. The bound can be made to work for every constant ϵ by a standard boosting argument. ∎

Remarks: There has been a flurry of recent research combining the Generalized Discrepancy Method and the Approximation/Orthogonality principle. First, David, Pitassi and Viola [21] observed that one can consider more general masking schemes than the pattern matrix technique. In particular, using a random scheme they proved lower bounds for a function that is in NP_k^{cc} but has no efficient k-party randomized protocols, for $k = O(\log n)$. This yielded the first explicit separation of NP_k^{cc} and BPP_k^{cc} for $k > \log\log n$. Later, Beame and Huynh-Ngoc [7] showed, using a more involved notion than approximate degree, that there are functions in depth-5 AC^0 that have no efficient k-party randomized protocols for $k = \Theta(\log n)$. As a consequence, they obtain a lower bound of $\Omega\big(2^{\sqrt{(\log n/k)-k}}\big)$ for the k-party complexity of Disjointness. This yields interesting lower bounds up to $k = \Theta(\log^{1/3} n)$ as opposed to the bound of Theorem 23 that peters out at $k = \log\log n$. On the other hand, observe that for $k = o(\log\log n)$, the bounds of Theorem 23 are much better.

4 Applications

There are many beautiful applications of the set disjointness lower bounds to many diverse areas of computer science. Here we highlight some applications in the areas of: streaming, data structures, circuit complexity, proof complexity, game theory and quantum computation.

4.1 Streaming

Let $S \in [n]^m$ be a length m stream, where each item in the stream is an item in $[n]$, and let $f(S)$ be some function of S. In a streaming model, the algorithm sees S one symbol at a time and the goal is to compute f using as little memory as possible. We desire low-space approximations to $f(S)$: an ϵ-approximation is an algorithm that computes a value that is within an ϵ factor of the correct answer (with high probability).

The seminal paper by Alon, Matias, and Szegedy [3] proves lower bounds for a class of important statistical functions called frequency moments. Let $M_i = |\{j \in [m] \mid S_j = i\}|$. The k^{th} frequency moment, F_k is equal to $\sum_{i=1}^{n} M_i^k$. Thus F_0 equals the number of distinct elements in the stream, F_1 equals the length of the stream, and F_∞ equals the number of occurrences of the most frequent item in the stream.

Our first simple application shows that computing F_∞ in the streaming model implies an efficient two-party communication protocol for set disjointness. Suppose that A is a streaming algorithm for F_∞ using C bits of memory. Given an input (x, y) to DISJ, Alice converts x into a stream $a_x = \{i \mid x_i = 1\}$ and similarly Bob converts y into a stream $b_y = \{i \mid y_i = 1\}$. Alice simulates A on a_x and then sends C bits of information representing the state of A after processing a_x; Bob then continues the simulation on b_x to compute $F_\infty(a_x b_x)$. Clearly if x and y are disjoint, then $F_\infty(a_x b_x) = 1$, and if x and y are not disjoint, then $F_\infty(a_x b_x) = 2$. Thus the lower bounds for DISJ imply $\Omega(n)$ space bounds for streaming algorithms computing F_∞.

The reduction can be generalized to obtain lower bounds for other frequency moments as well. In the *number-in-hand* model, there are p players; each player has a private input x_i, $|x_i| = n$. The players communicate via a shared blackboard in order to compute a function $f(x_1, \ldots, x_p)$. The promise version of disjointness, $UDISJ(x_1, \ldots, x_p)$, is equal to 1 if the intersection size is 1; is equal to 0 if they are pairwise disjoint, and otherwise the output can be anything. [13, 24] prove that the randomized p-player communication complexity of UDISJ is $\Omega(n/p)$. By a reduction to UDISJ it follows that any streaming algorithm for computing F_k requires space $\Omega(n^{1-2/k}$. Fix k and let $p = n^{1/k}$. As before, let A be a C-space streaming algorithm for computing F_k. Let $x_1, \ldots, x_p$ be an input to UDISJ and let the total number of one's in all of the strings be n. On input $(x_1, \ldots, x_p)$ to UDISJ, player i converts his/her input x_i into a stream $a_i = \{j \mid$ the j^{th} bit of x_i equals 1.$\}$. On the stream $S = a_1, \ldots, a_p$, player i will simulate the computation of A on the a_i portion of the stream, and then communicate the state (C bits) to the next player, to obtain an $Cn^{1/k}$ bit protocol. This solves UDISJ since $UDISJ(x_1, \ldots, x_p) = 0$ implies $F_k(S) = n$, and $UDISJ(x_1, \ldots, x_p) = 1$ implies $F_k(S) \geq n - p + n = 2n - p$.

4.2 Data Structures

The cell probe model was introduced over thirty years ago by Yao [58]. In this model, memory is modelled by an array of cells, each having w bits. The data is allowed to occupy S consecutive cells, called the space. Certain queries and update operations are to be supported. The runtime of an operation is the number of cell probes (reads and writes executed). There are two distinct types of problems in the cell probe literature: Dynamic problems are characterized by a tradeoff between the update time and the query time, whereas for static problems, the tradeoff is between the space S used, and the runtime.

The connection to communication complexity and asymmetric set disjointness was made explicit by [38]. (See [37] for somewhat dated but an excellent survey of the area.) Consider a communication game where Alice holds a query and Bob holds a database. A cell probe algorithm implies a communication protocol for computing the query on the database: Alice sends $\log S$ bits (an address), and Bob replies with w bits (value of that memory location). In the asymmetric version of set disjointness, called Lopsided Set Disjointness (LSD) Alice has a subset $S \subseteq [N \cdot B]$ of size N, and Bob has a subset $T \subseteq [N \cdot B]$. They want to determine whether

their sets intersect. In a beautiful paper, Patrascu [46] proved that for any $\delta > 0$, any bounded error protocol for LSD requires that either Alice sends $\delta N \log B$ bits, or that Bob sends at least $NB^{1-O(\delta)}$ bits, and from this lower bound he obtains lower bounds for a wide variety of both static and dynamic problems in the cell probe model. For example, lower bounds for lopsided set disjointness imply the first lower bound for reachability oracles, as well as cell probe lower bounds for high-dimensional problems where the goal is to show large space bounds.

Another version of set disjointness was very recently introduced in [47]. The 3-party set disjointness problem is defined as follows. Alice has $i \in [k]$ on her forehead; Bob has sets $S_1, \ldots, S_k$, $S_i \subseteq [n]$ on his forehead; and Charlie has a set $T \subseteq [n]$ on his forehead. In Stage 1, Alice communicates nM bits of information privately to Bob. Then in Stage 2, Bob and Charlie communicate back and forth, sending M bits in total, and at the end, they announce whether or not S_i intersects T. Proving that $M \geq n^\epsilon$ for the 3-party set disjointness problem implies polynomial lower bounds for many dynamic data structure problems, resolving a major open problem [47].

4.3 Circuit Complexity

There are important applications of communication complexity lower bounds to complexity theory, most notably the connection between k-player NOF lower bounds ($k > \log n$) for any explicit function and the circuit class ACC. Here we give two explicit applications of set disjointness lower bounds to circuit complexity.

Our first result, due to Nisan and Wigderson [44], relates communication complexity lower bounds for a disjointness function to a circuit lower bound. Let H be a family of 2-universal hash functions, so $h \in H$ is a 2-universal hash function mapping $\{0,1\}^n$ to itself, where $|h| = O(n)$. The function F associated with H is to compute the vector $h(y)$ on input h and y. One such family H is a succinct description of a family of n-by-n matrices M, in which case we think of the function as matrix multiplication where the allowable matrices are succinctly represented (by n bits rather than by n^2 bits). The obvious circuit for carrying out this computation has size n^2.

The 3-player version of this function is as follows. Alice has $j \in [n]$ on her forehead. Bob has some $h \in H$ on his forehead, where $|h| = n$, and Charlie has a string y, $|y| = n$, on his forehead. They want to compute whether the j^{th} bit of $h(y)$ is greater than 0. Notice that if we are working over the integers, then this problem is a 3-player version of set disjointness.

Theorem 24 *[44] If F can be computed by a circuit of fan-in 2, size $O(n)$ and depth $O(\log n)$, then the simultaneous communication complexity of the 3-player version of F is $O(n/\log\log n)$.*

The proof uses Valiant's graph theoretic lemma which states that given any depth $O(\log n)$, size $O(n)$ circuit C with n inputs and n outputs, there exists at most $O(n/\log\log n)$ wires in C whose removal leaves a circuit with the property that each output gate j depends only on a small set S_j of inputs, where $|S_j| \leq \sqrt{n}$. Using this lemma, the protocol is as follows. Alice, who sees h and y (the entire input to C), sends the values of the $O(n/\log\log n)$ wires; Bob and Charlie, who both see j and half of the input, send the values of the inputs in S_j. With all of this information, the answer can be computed by a referee. Thus, a lower bound of $\omega(n/\log\log n)$ on the simultaneous communication complexity of F is enough to prove a superlinear lower bound on the number of wires for a log-depth circuit to compute F. Proving such a lower bound for an explicit function is a major open problem in circuit complexity.

Another application of set disjointness lower bounds are algebraic oracle separations among complexity classes [1]. For example, using set disjointness lower bounds, Aaronson and Wigderson prove that resolving many important complexity separations, such as separating NP from P and separating NP from BPP, must require nonalgebrizing techniques.

4.4 Proof Complexity

The central problem in proof complexity is to establish superpolynomial lower bounds on the proof length required to refute hard unsatisfiable formulas in natural proof systems, such as Resolution, Cutting Planes, or Frege systems. One starts with an unsatisfiable formula, typically in 3CNF form. The goal is to apply rules from a standard axiomatic propositional proof system in order to derive the identically false formula "0". Proving superpolynomial lower bounds for all proof systems is equivalent to proving $NP \neq coNP$. While this goal seems beyond reach, a more reasonable goal is to prove lower bounds for specific proof systems. Most proof systems can be classified in terms of their representational strength. For example, Resolution proofs only allow one to derive clauses from previous clauses; thus the representational strength of Resolution is depth-1 AC_0. Cutting Planes proofs allow one to derive linear inequalities from previously derived linear inequalities, and thus has the representational strength of depth-1 TC_0. Similarly, $Th(k)$ proofs are defined, and their representational strength corresponds to degree k polynomial inequalities. $Th(k)$ proofs are quite powerful and include as special cases not only Resolution, but also a wide variety of matrix cut systems such as all of the Lovasz-Schrijver systems, low rank Sherali-Adams and Lasserre systems. (For more details on propositional proof complexity, see [10].)

The following theorem gives lower bounds for $Th(k)$ proofs via lower bounds for NOF set disjointness.

Theorem 25 *[9] For every constant k, there exists a family of CNF tautologies T_n, where for each n, T_n has n variables and has size polynomial in n, such that if T_n has polynomial-size tree-like $Th(k)$ proofs, then there exists an efficient (polylogarithmic many bits) probabilistic protocol for set disjointness in the $(k+1)$-player NOF model.*

While the proof of the above theorem is quite complicated, the high level argument is as follows. For any unsatisfiable 3CNF formula f, the search problem associated with f, S_f, takes as input a truth assignment α to the variables of f, and outputs a clause of f that is violated by α. In order to prove lower bounds for $Th(k)$ refutations via communication complexity lower bounds, the goal is to find a specific family of hard unsatisfiable formulas and prove that any small tree-like $Th(k)$ refutation of f_n implies an efficient $(k+1)$-party NOF protocol for S_{f_n}, the search problem associated with f_n. Notice that for *any* unsatisfiable 3CNF f, S_f has an efficient nondeterministic protocol in any communication complexity model, as the players can simply guess and check the clause that is violated. Thus, proof complexity lower bounds obtained via this method require a function that is hard to solve in a randomized or deterministic model, but that is easy to solve nondeterministically. In [9], a carefully concocted family of formulas T_n are constructed, and it is shown that small tree-like $Th(k+1)$ refutations for T_n imply small k-player NOF protocols for set disjointness.

In a recent paper [8], the concept of hardness escalation is introduced, and a much more general lower bound for $Th(k)$ proofs was obtained. In hardness amplification, one begins with a boolean function f, and constructs an *amplified* function g (based on f) such that: if f cannot be computed with respect to some circuit or complexity class C, then g cannot be computed with respect to another circuit or complexity class C', where C' is stronger than C. An example of hardness escalation is the dual polynomial method presented earlier. The same idea can also be applied in proof complexity: now one begins with an unsatisfiable CNF formula f, and from f, we construct an amplified unsatisfiable CNF formula $Lift(f)$, such that if f requires superpolynomial size refutations in proof system P, then $Lift(f)$ requires superpolynomial size refutations with respect to P', where P' is a more powerful proof system than P. The main theorem in [8] is the following hardness escalation theorem for proof complexity.

Theorem 26 *Let f_n be any unsatisfiable family of formulae requiring superpolynomial size tree*

Resolution proofs. Then, an amplified function $Lift_{f_n}(k)$ requires superpolynomial size tree-like $Th(k)$ proofs.

The proof of the above theorem applies the dual polynomial method used to obtain NOF lower bounds for set disjointness (described earlier), but adapted to the proof complexity context.

4.5 Game Theory

There are two primary applications of lower bounds for set disjointness to game theory. The first are lower bounds for combinatorial auctions and mechanism design, and the second family of results are lower bounds for finding pure Nash equilibrium. In these applications, understanding the communication complexity of certain game theoretic tasks is our main goal. This is unlike applications from previous sections, where communication complexity entered as a useful tool for attaining other goals.

Combinatorial Auctions and Mechanism Design. A classical problem in algorithmic game theory is a combinatorial auction. There are some number m of goods that need to be allocated to n agents. Each agent has a valuation function v_i that specifies for each subset S of the goods, the bidders desire or value $v_i(S)$ of obtaining S. The goal is to find a partition of the goods, $S_1, \ldots, S_n$ that maximizes social welfare, $\sum_i v_i(S_i)$. In [42], it is shown that the communication complexity of determining an optimal allocation is nearly maximal, via lower bounds for 2-player set disjointness. The result was generalized in [40] to show that even approximation algorithms require large communication, via set disjointness lower bounds in the k-player NIH model.

Algorithm mechanism design attempts to design protocols in distributed settings where the different agents have their own selfish goals, and thus they may attempt to optimize their own goals rather than to follow the prescribed protocol. A central goal in this area is to develop incentive-compatible mechanisms. These are protocols/algorithms together with payment schemes that motivate the agents to truthfully follow the protocol.

The famous VCG mechanism achieves incentive compatibility for combinatorial auctions. However, the problem of finding an optimal partition of the goods is NP-hard, and thus the VCG mechanism applied to this problem is too inefficient to be useful. On the other hand, there are polynomial-time approximation schemes for solving the purely computational problem of finding a good partition of the goods, but the VCG payment rule no longer leads to incentive compatibility when we move to approximation algorithms. This leads us to a central question in mechanism design, first posed by Dobzinski and Nisan [22]: "To what extent do the strategic requirements degrade the quality of the solution beyond the degradation implied by the purely computational constraints?" Specifically for combinatorial auctions, the question is whether or not it is possible to design an auction protocol that is both incentive compatible and computationally efficient, and still achieves a good approximation ratio.

In [22], Dobzinski and Nisan raised this question, and studied a subclass of algorithms for combinatorial auctions in the communication complexity setting. (In this setting, one measures the amount of communication between the agents.) They prove a strong lower bound on the approximation factor that can be achieved by incentive compatible VCG-based mechanisms for combinatorial auctions via a two part case analysis where one case reduces to lower bounds for set disjointness.

Recently, Papadimitriou, Shapira, and Singer [45] gave an unconditional negative answer to Dobzinski and Nisan's question for a different problem, the combinatorial public projects problem (CPPP), again in the communication complexity setting. Specifically they prove that the communication complexity of any constant-factor incentive-compatible algorithm for CPPP is nearly linear, via a reduction to set disjointness.

Pure Nash Equilibrium. Another central problem in game theory concerns the complexity of reaching an equilibrium in a game. Computational complexity is concerned with the time required to reach equilibrium; communication complexity is concerned with how much information must be exchanged between the players in order to reach equilibrium (assuming honest players).

In [20], Conitzer and Sandholm proved that any deterministic or randomized protocol for computing a pure Nash equilibrium in a 2-person game requires nearly maximal communication complexity. In 2007, Hart and Mansour [25] proved a similar result for multi player games. All of these results are obtained via lower bounds for set disjointness. We give the basic idea behind these reductions, following Nisan's presentation [41] For two players, Alice and Bob each hold their utility function $u_i : S_1 \times \ldots \times S_n \to R$, where the S_i's are the strategy sets, each of size m. Thus each player's input consists of m^n real numbers, which we will assume are in some small integer range. We want to show how a protocol for determining the existence of a pure Nash equilibrium for games with n players each having m strategies can be used to solve DISJ on strings of length $N = \Omega(m^n)$.

Here we will give the idea for $n = 2$, where we will solve set disjointness on strings of length $N = (m-2)^2$. We will view both Alice and Bob's inputs x and y as a square matrices, $x_{i,j}$, $i, j \leq (m-2)$. Alice will interpret x as a utility matrix u_A, and similarly Bob will interpret y as a utility matrix u_B. (Entry (i, j) of Alice's utility matrix describes the payoff for Alice when Alice plays strategy i and Bob plays strategy j. Similarly, entry (i, j) of Bob's utility matrix describes Bob's payoff in this situation. A Nash equilibrium is a pair of strategies (i^*, j^*) such that Alice's strategy i^* is optimal given that Bob plays strategy j^* and similarly Bob's strategy j^* is optimal given that Alice plays i^*.) Any cell where $x_i = y_i = 1$ will be a Nash equilibrium since both players get the highest utility possible in this matrix. We just need to make sure that no other cell is a Nash equilibrium. One easy way to do this is to add two extra rows and two extra columns to u_A, as follows.

$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,4}$	0	0
$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,4}$	0	0
$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$	0	0
$x_{4,1}$	$x_{4,2}$	$x_{4,3}$	$x_{4,4}$	0	0
1	1	1	1	1	0
1	1	1	1	0	1

The matrix u_B is defined similarly, with $y_{i,j}$ replacing $x_{i,j}$, and 0 interchanged with 1. With the addition of these rows and columns, each player's best reply always obtains a value of 1, and thus only an entry where both u_A and u_B are 1 will be a Nash equilibrium.

This idea can be generalized for any n by adding $n-2$ dummy players (in addition to Alice and Bob) that have utilities that are identically 0. Thus a pure Nash equilibrium will be determined solely by Alice and Bob. In this more general setting, we will interpret Alice (and Bob's) input as n-dimensional matrices of size m^n, and thus after adding the extra rows and columns, we will answer set disjointness on vectors of length $N = (m-2)^2 m^{n-2}$.

4.6 Quantum Computation

As discussed earlier, any probabilistic protocol for set disjointness requires $\Omega(n)$ bits. However for quantum communication, Buhrman, Cleve, and Wigderson gave an $O(\sqrt{n}\log n)$ qubit algorithm for set disjointness, based on Grover's quantum search algorithm [12]. There is a nearly matching $\Omega(\sqrt{n})$ lower bound due to Razborov [50]. Thus the quantum communication complexity lower bound for set disjointness shows that quantum communication is of limited help for solving this problem: it does not generate exponential savings in communication as compared to the classical randomized model in this case. This question in general is still open.

There are several applications of the quantum set disjointness lower bound in quantum computation. One application is the result by Aaronson and Wigderson [1] showing that any polynomial-time quantum algorithm for solving an NP-complete problem must be nonalgebrizing. A second application is a strong direct product theorem for quantum communication complexity [31], which makes essential use of the quantum communication complexity lower bound for disjointness.

5 Discussion

There are many other notable results for set disjointness that are beyond the scope of this survey. For example, Klauck [28] has recently proven a strong direct product theorem for set disjointness, showing that if we want to compute k independent instances of set disjointness using less than k times the resources needed for one instance, then the overall success probability will be exponentially small in k.

There are many important open problems still to be solved. We mention just a few here. First, is there a complete problem for nondeterministic NOF communication for $k \geq 3$? Secondly, the current lower bounds for multiparty complexity of set disjointness are not known to be tight. In particular, for constant k, the bounds are of the form $n^{1/k+1}$, whereas the best known upper bound is $O(n)$. The current best bounds of Beame and Huynh-Ngoc stop giving anything nontrivial beyond $k = \Theta(\log^{1/3} n)$. The best upper bounds are of the form $O(n/4^k + k \log n)$. This upper bound is essentially due to a very general protocol by Grolmusz that in particular, also works for the Generalized Inner Product (GIP) function. However, for GIP we know that these bounds are tight. Are they tight for Disjointess? Thirdly, the information complexity lower bound for two players has not been generalized to the NOF setting. An important open problem is to prove lower bounds via the information complexity approach in the multiparty NOF setting.

Acknowledgements

We wish to thank the following people for very helpful comments: Scott Aaronson, Joshua Brody, Amit Chakrabarti, Jeff Edmonds, Faith Ellen, Brendan Lucier, Noam Nisan, Michel Schapira, and Avi Wigderson.

References

[1] S. Aaronson and A. Wigderson. Algebrization: A New Barrier in Complexity Theory. In *Proceedings STOC*, 2008.

[2] F. Ablayev. Lower Bounds for One-way Probabilistic Communication Complexity and their Application to Space Complexity. *Theoretical Computer Science*, 157(2):139–159, 1996.

[3] N. Alon, Y. Matias, and M. Szegedy. The Space Complexity of Approximating the Frequency Moments. In *Proceedings STOC*, 1996.

[4] L. Babai, P. Frankl, and J. Simon. Complexity Classes in Communication Complexity Theory. In *Proceedings FOCS*, pages 337–347, 1986.

[5] L. Babai, N. Nisan, and M. Szegedy. Multiparty Protocols, Pseudorandom Generators for Logspace, and Time-Space Trade-offs. *JCSS*, 45(2):204–232, 1992.

[6] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An Information Statistics Approach to Data Stream and Communication Complexity. *JCSS*, 68:702–732, 2004.

[7] P. Beame and D. Huynh-Ngoc. Multiparty Communication Complexity and Threshold Circuit Complexity of AC^0. In *IEEE 50th Annual Symposium on Foundations of Computer Science (FOCS)*, 2009.

[8] P. Beame, T. Ngoc, and T. Pitassi. Hardness amplification in proof complexity. In *Proceedings STOC*, 2010.

[9] P. Beame, T. Pitassi, and N. Segerlind. Lower Bounds for Lovasz Schrijver from Multiparty Communication Complexity. In *SIAM J. Computing*, 2007.

[10] P. W. Beame and T. Pitassi. Propositional Proof Complexity: Past, Present, Future. *EATCS Bulletin*, 1998.

[11] R. Beigel and J. Tarui. On ACC. In *IEEE FOCS*, pages 783–792, 1991.

[12] R. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. Classical Communication and Computation. In *30th ACM STOC*, pages 63–86, 1998.

[13] A. Chakrabarti, S. Khot, and X. Sun. Near-optimal Lower Bounds on the Multiparty Communication Complexity of Set Disjointness. In *IEEE CCC*, 2008.

[14] A. Chakrabarti, Y. Shi, A. Wirth, and A. C. C. Yao. Informational Complexity and the Direct Sum Problem for Simultaneous Message Complexity. In *42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 270–278, 2001.

[15] A. K. Chandra, M. L. Furst, and R. J. Lipton. Multi-party Protocols. In *Proc. STOC*, pages 94–99, 1983.

[16] A. Chattopadhyay. Discrepancy and the Power of Bottom Fan-in in Depth-three Circuits. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 449–458, 2007.

[17] A. Chattopadhyay. *Circuits, Communication and Polynomials*. PhD thesis, Computer Science, McGill University, November 2008.

[18] A. Chattopadhyay and A. Ada. Multiparty Communication Complexity of Disjointness. In *Electronic Colloquium on Computational Complexity TR08-002*, 2008.

[19] F. Chung and P. Tetali. Communication Complexity and Quasi randomness. *SIAM J. Discrete Math*, 6(1):110–125, 1993.

[20] V. Conitzer and T. Sandholm. Communication Complexity as a Lower Bound for Learning in Games. In *Proceedings 24th ICML*, 2004.

[21] M. David, T. Pitassi, and E. Viola. Separating RP from NP for NOF Communication Complexity. In *Proc. RANDOM*, 2008.

[22] S. Dobzinski and N. Nisan. Limitations of VCG-Based Mechanisms. In *Proc. ACM STOC*, pages 338–344, 2007.

[23] J. Ford and A. Gal. Hadamard Tensors and Lower Bounds on Multiparty Communication Complexity. In *Complexity of Boolean Functions*, 2006.

[24] A. Gronemeier. Asymptotically Optimal Lower Bounds on the NIH Multiparty Information Complexity of the AND Function and Disjointness. In *Proceedings 26th STACS*, pages 505–516, 2009.

[25] S. Hart and Y. Mansour. The Communication Complexity of Uncoupled Nash Equilibrium. In *Proc. STOC*, 2007.

[26] J. Håstad and M. Goldmann. On the Power of Small-Depth Threshold Circuits. In *Proceedings 31st Annual Symposium on Foundations of Computer Science*, pages 610–618, St. Louis, MO, Oct. 1990. IEEE.

[27] B. Kalyanasundaram and G. Schnitger. The Probabilistic Communication Complexity of Set Intersection. In *Proceedings CCC*, pages 41–49, 1987.

[28] H. Klauck. A Strong Direct Product Theorem for Set Disjointness. In *Proceedings STOC 2010*, pages 77–86.

[29] H. Klauck. Lower bounds for quantum communication complexity. In *42nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 288–297, 2001.

[30] H. Klauck. Lower bounds for quantum communication complexity. *SIAM J. Computing*, 37(1):20–46, 2007.

[31] H. Klauck, R. Spalek, and R. de Wolf. Quantum and Classical Strong Direct Product Theorems and Optimal Time-Space Tradeoffs. In *http://arxiv.org/abs/quant-ph/0402123*, 2004.

[32] M. Krause and P. Pudlák. On the Computational Power of Depth 2 Circuits with Threshold and Modulo Gates. In *26th Annual Symposium on Theory of Computing (STOC)*, pages 48–57. ACM, 1994.

[33] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, Cambridge [England] ; New York, 1997.

[34] T. Lee and A. Shraibman. Disjointness is Hard in the Multi-party Number-on-the-Forehead Model. In *23rd Annual IEEE Conference on Computational Complexity*, 2008.

[35] T. Lee and A. Shraibman. Lower Bounds in Communication Complexity. In *Manuscript*, 2010.

[36] S. Lokam. Complexity Lower Bounds using Linear Algebra. *Foundations and Trends in Theoretical Computer Science*, 4(1):1–155, 2008.

[37] P. Miltersen. Cell Probe Complexity - A Survey. In *Advances in Data Structures*, 1999.

[38] P. B. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. *Journal of Computer and System Sciences*, 57(1):37–49, 1998.

[39] N. Nisan. The Communication Complexity of Threshold Gates. In *Proc. of "Combinatorics, Paul Erdos is Eighty"*, pages 301–315, 1993.

[40] N. Nisan. The Communication Complexity of Approximate Set Packing and Covering. In *Proceedings ICALP*, 2002.

[41] N. Nisan. Communication Complexity of Reaching Equilibrium. In *http://agtb.wordpress.com/2009/08/18/communication-complexity-of-reaching-equilibrium/*, 2009.

[42] N. Nisan and I. Segal. The Communication Requirements of Efficient Allocations and Supporting Lindahl Prices. *Journal of Economic Theory*, 2006.

[43] N. Nisan and M. Szegedy. On the Degree of Boolean Functions as Real Polynomials. *Computational Complexity*, 4:301–313, 1994.

[44] N. Nisan and A. Wigderson. Rounds in Communication Complexity Revisited. In *Proc. STOC*, pages 419–429, 1991.

[45] C. Papadimitriou, M. Schapira, and Y. Singer. On the Hardness of Being Truthful. In *Proc. IEEE FOCS*, 2008.

[46] M. Patrascu. Unifying the Landscape of Cell Probe Lower Bounds. In *IEEE FOCS*, 2008.

[47] M. Patrascu. Toward Polynomial Lower Bounds for Dynamic Problems. In *ACM STOC*, 2010.

[48] R. Raz. A Parallel Repetition Theorem. *SIAM J. Computing*, 27(3):763–803, 1998.

[49] R. Raz. The BNS-Chung Criterion for Multiparty Communication Complexity. *Computational Complexity*, 9(2):113–122, 2000.

[50] A. Razborov. Quantum Communication Complexity of Classical Predicates. 67(1):145–159, 2003.

[51] A. A. Razborov. On the Distributional Complexity of Disjointness. In *Proc. ICALP*, pages 249–253, 1990.

[52] M. Saks and X. Sun. Space Lower Bounds for Distance Approximation in the Data Stream Model. In *34th Annual Symposium on Theory of Computing (STOC)*, pages 360–369.

[53] A. Sherstov. Separating AC^0 from Depth-2 Majority Circuits. In *39th ACM Symposium on Theory of Computing (STOC)*, pages 294–301, 2007.

[54] A. Sherstov. The Pattern Matrix Method for Lower Bounds on Quantum Communication. In *40th Annual Symposium on Theory of Computing (STOC)*, pages 85–94, 2008.

[55] Y. Shi and Y. Zhu. The Quantum Communication Complexity of Block-Composed Functions. 2007.

[56] P. Tesson. *Computational Complexity Questions Relating to Finite Monoids and Semigroups*. PhD thesis, McGill University, 2003.

[57] A. C. Yao. Some Complexity Questions Related to Distributive Computing. In *Proc. ACM STOC*, pages 209–213, 1979.

[58] A. C. Yao. Should Tables be Sorted? *Journal of the ACM*, 28:615–628, July 1981.