Electronic Colloquium on Computational Complexity, Report No. 145 (2011)

THE MULTIPARTY COMMUNICATION COMPLEXITY OF SET DISJOINTNESS

ALEXANDER A. SHERSTOV*

ABSTRACT. We study the set disjointness problem in the number-on-the-forehead model.

(i) We prove that k-party set disjointness has randomized and nondeterministic communication complexity $\Omega(n/4^k)^{1/4}$ and Merlin-Arthur complexity $\Omega(n/4^k)^{1/8}$. These bounds are close to tight. Previous lower bounds (2007-2008) for $k \geqslant 3$ parties were weaker than $n^{1/(k+1)}/2^{k^2}$ in all three models.

(ii) We prove that solving ℓ instances of set disjointness requires $\ell \cdot \Omega(n/4^k)^{1/4}$ bits of communication, even to achieve correctness probability exponentially close to $1/2$. This gives the first direct-product result for multiparty set disjointness, solving an open problem due to Beame, Pitassi, Segerlind, and Wigderson (2005).

(iii) We construct a read-once $\{\wedge, \vee\}$-circuit of depth 3 with exponentially small discrepancy for up to $k \approx \frac{1}{2}\log n$ parties. This result is optimal with respect to depth and solves an open problem due to Beame and Huynh-Ngoc (FOCS '09), who gave a depth-6 construction. Applications to circuit complexity are given.

The proof technique of this paper departs significantly from previous work and is of independent interest.

1. INTRODUCTION

In a seminal paper thirty years ago, Yao [50] introduced the two-party model of communication complexity. In this model, two parties seek to evaluate a function $f(x, y)$ with minimal communication, where the first party knows only x and the second party only y. To capture communication among three or more parties, one considers a function f with several arguments that are somehow distributed among the parties, possibly with overlap. For a model to be meaningful, no party should know *all* the arguments (making communication necessary to evaluate f), and every argument should be known to *some* party (making communication sufficient). The *number-on-the-forehead* model of multiparty communication, due to Chandra, Furst, and Lipton [16], is the most powerful model that obeys the two principles. This model features k parties and a function $f(x_1, x_2, \ldots, x_k)$ with k arguments. The ith party knows all the arguments except for x_i—one can think of x_i as written on the ith party's forehead, hence the name of the model. Communication occurs in broadcast, a bit sent by any given party instantly reaching everyone else. The main research question is whether f has low *communication complexity,* i.e., can be computed by a protocol in which the number of bits communicated is small on every input. We will primarily be interested in *randomized* protocols, which are allowed to err with a small constant probability, as well as *nondeterministic* and *Merlin-Arthur* protocols. The multiparty model is a natural computational model in its own right and has additionally found a variety of applications, including streaming algorithms, circuit complexity, pseudorandomness, and proof complexity [6, 51, 25, 41, 12].

The multiparty model draws its richness from the overlap in the parties' inputs, which makes it challenging to prove lower bounds. For this reason, several fundamental questions in this model remain open despite much research. One such unresolved question is the

* Computer Science Department, UCLA, Los Angeles, California 90095. ✉ sherstov@cs.ucla.edu.

ISSN 1433-8092

communication complexity of *set disjointness*, arguably the most studied problem in the area [3, 27, 4, 7, 40, 47, 31, 13, 49, 42, 46, 35, 19, 11, 14, 30]. In the k-party version of set disjointness, the inputs are sets $S_1, S_2, \ldots, S_k \subseteq \{1, 2, \ldots, n\}$, and the ith party knows all the inputs except for S_i. The goal is to determine whether the intersection $\bigcap_{i=1}^{k} S_i$ is empty. One also studies a promise version of this problem called *unique set disjointness*, in which the input sets $S_1, S_2, \ldots, S_k$ either have an empty intersection or intersect in a unique element, i.e., either $|\bigcap S_i| = 0$ or $|\bigcap S_i| = 1$. It is common to represent set disjointness in function form as

$$\mathrm{DISJ}_{n,k}(x_1, x_2, \ldots, x_k) = \bigwedge_{j=1}^{n} \bigvee_{i=1}^{k} \overline{x}_{ij},$$

where the bit strings $x_1, x_2, \ldots, x_k \in \{0,1\}^n$ are the characteristic vectors of the k sets. Unique set disjointness $\mathrm{UDISJ}_{n,k}$ is represented by an identical formula, with the understanding that the strings $x_1, x_2, \ldots, x_k$ are legal inputs if and only if their bitwise conjunction $x_1 \wedge x_2 \wedge \cdots \wedge x_k$ has at most one nonzero bit. In communication complexity, set disjointness plays a role closely similar to the role of satisfiability in computational complexity. Outside of communication complexity the study of set disjointness is motivated by a number of applications, which we will discuss shortly in the context of our results.

Previous work. In the model with two parties, the communication complexity of set disjointness is thoroughly understood. One of the earliest results in the area is a tight lower bound of $n + 1$ bits for deterministic protocols solving set disjointness. For randomized protocols, a lower bound of $\Omega(\sqrt{n})$ was obtained by Babai, Frankl, and Simon [3] and strengthened to a tight $\Omega(n)$ by Kalyanasundaram and Schnitger [27]. Simpler proofs of the linear lower bound were discovered by Razborov [39] and Bar-Yossef et al. [7]. All three proofs [27, 39, 7] of the linear lower bound apply to unique set disjointness. Finally, Razborov [40] obtained a tight lower bound of $\Omega(\sqrt{n})$ on the bounded-error quantum communication complexity of set disjointness and unique set disjointness, with a simpler proof discovered several years later in [42]. Already in the two-party model, set disjointness has been a driving force for various technical innovations, including ideas from combinatorics, Kolmogorov complexity, information theory, matrix analysis, and Fourier analysis.

Progress on the communication complexity of set disjointness for $k \geqslant 3$ parties is summarized in Table A. In a surprising result, Grolmusz [24] proved an upper bound of $O(\log^2 n + k^2 n/2^k)$ on the deterministic communication complexity of this problem. Proving a strong lower bound, even for $k = 3$, turned out to be difficult. Tesson [47] and Beame et al. [13] obtained a lower bound of $\Omega\big(\frac{1}{k}\log n\big)$ for randomized protocols. Four years later, Lee and Shraibman [35] and Chattopadhyay and Ada [19] gave an improved result. These authors generalized the two-party method of [43, 42] to $k \geqslant 3$ parties and thereby obtained a lower bound of $\Omega\big(n^{1/(k+1)}/2^{2^k}\big)$ on the randomized communication complexity of set disjointness. The only subsequent work of which we are aware is due to Beame and Huynh-Ngoc [11], who proved a lower bound of $2^{\Omega(\sqrt{\log n}/\sqrt{k})}/2^k$ on the randomized communication complexity. This improves on the previous bound for k sufficiently large. For more than three years now, set disjointness has seen no progress. In what follows, we state the new results of this paper on set disjointness and related questions.

Randomized communication complexity of set disjointness. To summarize Table A, lower bounds on the k-party communication complexity of set disjointness prior to this paper, both deterministic and randomized, were weaker than $n^{1/(k+1)}/2^{k^2}$. In particular,

no polynomial lower bounds were known for $k = \omega(1)$ parties. Our first result is the following theorem, where R_ϵ denotes randomized communication complexity with error probability ϵ.

THEOREM 1.1. *Set disjointness and unique set disjointness have randomized communication complexity*

$$R_{1/3}(\mathrm{DISJ}_{n,k}) \geqslant R_{1/3}(\mathrm{UDISJ}_{n,k}) = \Omega\Big(\frac{n}{4^k}\Big)^{1/4}.$$

Theorem 1.1 comes close to matching Grolmusz's longstanding upper bound and shows in particular that the randomized communication complexity of set disjointness remains polynomial for up to $k \approx \frac{1}{2}\log n$ parties. This is representative of the state of the art in multiparty communication complexity in general: no explicit function $F\colon (\{0,1\}^n)^k \to \{0,1\}$ is currently known with nontrivial communication complexity for $k \geqslant \log n$ parties. Theorem 1.1 subsumes all previous multiparty lower bounds, with a strict improvement starting at $k = 4$. Finally, several restrictions of the number-on-the-forehead model have been considered [4, 47, 13, 49, 14, 30], including simultaneous message passing, one-way protocols, and certain intermediate models. The strongest communication lower bound [47, 13] in these restricted models was $\Omega(n^{1/(k-1)}/k)$, which is already weaker than Theorem 1.1 starting at $k = 6$.

XOR lemmas and direct product theorems. A natural question to ask of any computational model is how the resources needed to solve ℓ instances of a problem scale with ℓ. Suppose that solving a single instance of a given decision problem, with probability of correctness $2/3$, requires R units of a computational resource such as time, memory, communication, or queries. How many units of the resource are needed to solve ℓ independent instances of the problem? Common sense suggests that the answer should be $\Omega(\ell R)$. After all, having less than $\epsilon \ell R$ units overall, for a small constant $\epsilon > 0$, leaves less than ϵR units per instance, intuitively forcing the algorithm to guess random answers for many of the instances and resulting in overall correctness probability $2^{-\Omega(\ell)}$. Such a statement is called a

Bound	**Reference**
$O(\log^2 n + k^2 n/2^k)$	Grolmusz [24]
$\Omega\left(\frac{1}{k}\log n\right)$	Tesson [47]
	Beame, Pitassi, Segerlind, and Wigderson [13]
$\Omega\big(n^{1/(k+1)}/2^{2^k}\big)$	Lee and Shraibman [35]
	Chattopadhyay and Ada [19]
$2^{\Omega(\sqrt{\log n}/\sqrt{k})}/2^k$	Beame and Huynh-Ngoc [11]
$\Omega(n/4^k)^{1/4}$	This paper

Table A: Bounds for the randomized number-on-the-forehead communication complexity of set disjointness for $k \geqslant 3$ parties.

strong direct product theorem. A related notion is an *XOR lemma*, which asserts that computing the XOR of the answers to the ℓ problem instances requires $\Omega(\ell R)$ resources, even to achieve correctness probability $\frac{1}{2} + 2^{-\Omega(\ell)}$. While intuitively satisfying, XOR lemmas and strong direct product theorems are hard to prove and sometimes simply not true.

In communication complexity, the direct-product question has been studied for over twenty years. We refer the reader to [30, 45] for an up-to-date overview of the literature, focusing here exclusively on set disjointness. The direct-product question for *two-party* set disjointness has been resolved completely and definitively [31, 13, 14, 26, 30, 45], including classical one-way protocols [26], classical two-way protocols [13, 30], quantum one-way protocols [14], and quantum two-way protocols [31, 45]. Starting at $k = 3$, however, we are not aware of direct-product results of any kind for set disjointness. In fact, obtaining such a result was posed as an open problem by Beame et al. [13, p. 426]. We prove a direct-product result for up to $k \approx \frac{1}{2}\log n$ parties.

THEOREM 1.2. *The following tasks require $\ell \cdot \Omega(n/4^k)^{1/4}$ bits of communication each:*

(i) *computing the XOR of ℓ independent instances of unique set disjointness* $\mathrm{UDISJ}_{n,k}$ *with correctness probability $\frac{1}{2} + 2^{-\Omega(\ell)}$;*

(ii) *solving with probability $2^{-\epsilon\ell}$ at least $(1-\epsilon)\ell$ among ℓ instances of unique set disjointness* $\mathrm{UDISJ}_{n,k}$*, where $\epsilon > 0$ is a small enough constant.*

Clearly, this result also holds for set disjointness, a problem harder than $\mathrm{UDISJ}_{n,k}$. Theorem 1.2 generalizes Theorem 1.1, showing that $\Omega(n/4^k)^{1/4}$ is in fact a lower bound on the per-instance cost of set disjointness. Note that by (ii), this lower bound remains valid even if the protocol only needs to solve a $1-\epsilon$ fraction of the given ℓ instances, rather than all ℓ instances. Results of this type are known as *threshold direct product theorems*.

Nondeterministic and Merlin-Arthur communication. Nondeterministic communication is a natural counterpart to determinism and randomization. Analogous to computational complexity, a nondeterministic protocol starts with a guess string, whose length counts toward the protocol's communication cost, and proceeds deterministically thenceforth. A nondeterministic protocol for a given communication problem F is required to output the correct answer for *all* guess strings when $F = 0$, and for *some* guess string when $F = 1$. Observe that the complement of set disjointness has a highly efficient nondeterministic protocol. Indeed, it suffices to guess an element $i \in \{1, 2, \ldots, n\}$ and verify with two bits of communication that $i \in S_1 \cap S_2 \cap \cdots \cap S_k$. We show that set disjointness, unlike its complement, has high nondeterministic complexity.

THEOREM 1.3. *Set disjointness has nondeterministic communication complexity*

$$N(\mathrm{DISJ}_{n,k}) = \Omega\Big(\frac{n}{4^k}\Big)^{1/4}.$$

The best previous lower bound [23] on the nondeterministic complexity of set disjointness was $n^{\Omega(1/k)}/2^{2^k}$.

We further consider *Merlin-Arthur* protocols [2, 5], a communication model that combines the power of randomization and nondeterminism. As before, a Merlin-Arthur protocol for a given problem F starts with a guess string, whose length counts toward the communication cost. From then on, the parties run an ordinary randomized protocol. The

randomized phase in a Merlin-Arthur protocol must produce the correct answer with probability $2/3$ for *all* guess strings when $F = 0$ and for *some* guess string when $F = 1$. We prove that set disjointness has high Merlin-Arthur complexity, denoted *MA*:

THEOREM 1.4. *Set disjointness has Merlin-Arthur communication complexity*

$$MA(\mathrm{DISJ}_{n,k}) = \Omega\Big(\frac{n}{4^k}\Big)^{1/8}.$$

Theorem 1.4 can be viewed as a generalization of Theorems 1.1 and 1.3 on randomized and nondeterministic communication, respectively. These lower bounds are close to optimal, in view of Grolmusz's deterministic upper bound. As with nondeterminism, the best previous lower bound [23] on the Merlin-Arthur complexity of set disjointness was $n^{\Omega(1/k)}/2^{2^k}$.

Theorems 1.1, 1.3, and 1.4 shed new light on communication complexity classes, defined in the seminal work of Babai, Frankl, and Simon [3]. An infinite family $\{F_n\}_{n=1}^{\infty}$, where each $F_n\colon (\{0,1\}^n)^k \to \{0,1\}$ is a k-party number-on-the-forehead communication problem, is considered to be efficiently solvable by a given class of protocols if F_n has communication complexity at most $\log^c n$, for a large enough constant $c > 1$ and all $n > c$. This convention allows one to define BPP_k, NP_k, coNP_k, and MA_k as the classes of families with efficient randomized, nondeterministic, co-nondeterministic, and Merlin-Arthur protocols, respectively. In recent years, the relationships among these classes have been almost fully determined [9, 35, 19, 22, 11, 10, 23]. It particular, for $k \leqslant \Theta(\log n)$, it is known [10, 23] that coNP_k is not contained in BPP_k, NP_k, or even MA_k. As a corollary to Theorem 1.4, we show that coNP_k can be separated from all these classes by a particularly simple function, set disjointness.

COROLLARY. *For $k \leqslant (\frac{1}{2} - \epsilon)\log n$, where $\epsilon > 0$ is any constant,*

$$\mathrm{DISJ}_{n,k} \in \mathsf{coNP}_k \setminus \mathsf{BPP}_k,$$
$$\mathrm{DISJ}_{n,k} \in \mathsf{coNP}_k \setminus \mathsf{NP}_k,$$
$$\mathrm{DISJ}_{n,k} \in \mathsf{coNP}_k \setminus \mathsf{MA}_k.$$

Prior to this paper, the separation $\mathrm{DISJ}_{n,k} \in \mathsf{coNP}_k \setminus \mathsf{BPP}_k$ was known to hold for up to $k \leqslant \Theta(\log^{1/3} n)$ parties [11], with a much weaker lower bound on randomized communication complexity (see Table A). The other two separations were known to hold for up to $k \leqslant \Theta(\log\log n)$ parties [23], again with a much weaker lower bound on nondeterministic and Merlin-Arthur communication complexity.

Discrepancy and circuit complexity. Theorem 1.1 rules out an efficient protocol that solves set disjointness with correctness probability $\frac{2}{3}$. However, for any number of parties k, set disjointness has a simple and efficient protocol with nonnegligible correctness probability, $\frac{1}{2} + n^{-\Theta(1)}$. In fact, such a protocol exists not just for set disjointness but any function computable by a polynomial-size $\{\wedge, \vee, \neg\}$-circuit of depth 2, regardless of how the bits are assigned to the parties. We show that this phenomenon is special to depth 2, by constructing a read-once $\{\wedge, \vee\}$-circuit of depth 3 whose communication complexity remains high even for correctness probability exponentially close to $\frac{1}{2}$.

THEOREM 1.5. *There is a k-party communication problem $H_{n,k}\colon (\{0,1\}^n)^k \to \{0,1\}$, given by an explicit read-once $\{\wedge, \vee\}$-formula of depth 3, such that solving $H_{n,k}$ with correctness probability $\frac{1}{2} + \exp(-\Omega(n/4^k)^{1/7})$ requires communication $\Omega(n/4^k)^{1/7}$.*

To use a technical term, Theorem 1.5 shows that depth-3 circuits have exponentially small *discrepancy* for up to $k \approx \frac{1}{2}\log n$ parties, i.e., exponentially small correlation with low-cost communication protocols. As we mentioned in the previous paragraph, Theorem 1.5 is optimal with respect to circuit depth. It is also qualitatively optimal with respect to the number of parties k: by the results in [1, 25], every polynomial-size $\{\wedge, \vee, \neg\}$-circuit of constant depth admits a $\log^c n$-party protocol with communication $\log^c n$ and correctness probability $\frac{1}{2} + 2^{-\log^c n}$, where $c > 1$ is a suitably large constant. Theorem 1.5 solves an open problem posed by Beame and Huynh-Ngoc [11], who constructed a similarly hard depth-6 formula and asked whether the depth can be reduced. The communication lower bound in Theorem 1.5 is stronger than in [11], where a lower bound of $\Omega(n/2^{31k})^{1/29}$ bits is derived for correctness probability $\frac{1}{2} + \exp(-\Omega(n/2^{31k})^{1/29})$.

Theorem 1.5 has applications to circuit complexity, which we now pause to explain. Circuits of majority gates are a biologically inspired computational model whose study spans several decades and several disciplines. Research has shown that majority circuits of depth 3 already are surprisingly powerful. In particular, Allender [1] proved that depth-3 majority circuits of quasipolynomial size can simulate all of AC^0, the class of $\{\wedge, \vee, \neg\}$-circuits of constant depth and polynomial size. Yao [51] further proved that depth-3 majority circuits of quasipolynomial size can simulate all of ACC, the class of $\{\wedge, \vee, \neg, \bmod m\}$-circuits of constant depth and polynomial size for an arbitrary but fixed modulus m. For several years, it was open whether these simulations are optimal. Håstad and Goldmann [25] showed that Yao's simulation of ACC is optimal with respect to *circuit depth*, by exhibiting a function in ACC whose simulation by a depth-2 majority circuit requires exponential size. The analogous question for AC^0 remained open [32]. It was solved several years ago in [42, 43], where an AC^0 function was constructed whose simulation by depth-2 majority circuits requires exponential size. The simulations of Allender [1] and Yao [51] were thus shown to be optimal with respect to circuit depth.

Another natural parameter to study is the *fan-in* of a circuit's bottom gates. The simulations of Allender [1] and Yao [51] had bottom fan-in $\log^{O(1)} n$. The paper of Håstad and Goldmann [25] showed that this fan-in is not far from optimal, in that simulating ACC by a depth-3 majority circuit with bottom fan-in $\approx \frac{1}{2}\log n$ requires exponentially many gates. The analogous question for AC^0 was considered by Chattopadhyay [17], who generalized the method of [42, 43] to show that depth-3 majority circuits with constant bottom fan-in require exponentially many gates to simulate AC^0. More recently, Beame and Huynh-Ngoc [11] proved an analogous result for bottom fan-in $\approx \frac{1}{31}\log n$. It was thus shown that the simulations of Allender [1] and Yao [51] are close to optimal in bottom fan-in.

The lower bounds surveyed in the previous paragraphs [25, 42, 43, 17, 11] apply not only to majority circuits but all circuits of type MAJ $\circ$ SYMM $\circ$ ANY (with a majority gate at the top, arbitrary symmetric gates at the middle level, and arbitrary gates at the bottom). This line of research is summarized in Table B, with quantitative detail. Theorem 1.5 in this paper implies the following new lower bound.

THEOREM 1.6. *Let* $H_{n,k} \colon (\{0,1\}^n)^k \to \{0,1\}$ *be the depth-*3 *read-once* $\{\wedge, \vee\}$*-formula constructed in Theorem* 1.5. *Then every circuit of type* MAJ $\circ$ SYMM $\circ$ ANY *with bottom fan-in at most* k *that computes* $H_{n,k+1}$ *has size*

$$\exp\left(\frac{1}{k} \cdot \Omega\Big(\frac{n}{4^k}\Big)^{1/7}\right).$$

As Table B shows, Theorem 1.6 improves on previous AC^0 constructions [17, 11] with respect to all parameters: the function is simpler than those considered previously, whereas the circuit lower bound is stronger and applies to MAJ $\circ$ SYMM $\circ$ ANY circuits with larger fan-in. In particular, the construction in Theorem 1.6 has optimal depth because $\{\wedge, \vee\}$-circuits of depth 2 are clearly computable by MAJ $\circ$ SYMM circuits of the same size.

Using the method of random restrictions, Razborov and Wigderson [41] discovered a way to convert lower bounds for MAJ $\circ$ SYMM $\circ$ ANY circuits with restricted fan-in into lower bounds for MAJ $\circ$ SYMM $\circ$ AND without any fan-in restrictions. Using that technique, we obtain the following consequence of Theorem 1.6.

THEOREM 1.7. *Every circuit of type* MAJ $\circ$ SYMM $\circ$ AND *that computes the function*

$$x \mapsto \bigvee_{i=1}^{n} \bigwedge_{j=1}^{n} \bigvee_{k=1}^{\log n} \bigoplus_{\ell=1}^{\log^2 n} x_{i,j,k,\ell}$$

has size $n^{\Omega(\log\log n)}$.

Again, Theorem 1.7 improves on previous work [11], where the same lower bound was derived for a more complicated, depth-8 AC^0 function.

Additional results and generalizations. Theorem 1.1 on the randomized communication complexity of set disjointness and Theorem 1.2 on the direct product property are proved here in greater generality. Specifically, our results apply to any k-party communication problem of the form $F = f(\mathrm{UDISJ}_{r,k}, \mathrm{UDISJ}_{r,k}, \ldots, \mathrm{UDISJ}_{r,k})$, i.e., an arbitrary Boolean function f composed componentwise with independent instances of the k-party set disjointness problem on a small number of variables r. We bound the ϵ-error randomized complexity of F from below in terms of the *ϵ-approximate degree* of f, defined as the

Function	**Circuit lower bound**	**Reference**
$\bigoplus_{j=1}^{n} \bigwedge_{i=1}^{k+1} x_{ij}$	$\exp\left(\frac{1}{k} \cdot \Omega\left(\frac{n}{4^k}\right)\right)$	Håstad and Goldmann [25]
read-once depth-3 $\{\wedge, \vee\}$-formula	$\exp\left(\Omega\left(n^{1/3}\right)\right)$ for $k = 1$	Sherstov [42, 43]
depth-3 $\{\wedge, \vee\}$-formula	$\exp\left(\Omega\left(n^{\frac{1}{6k2^k}}\right)\right)$	Chattopadhyay [17]
read-once depth-6 $\{\wedge, \vee\}$-formula	$\exp\left(\frac{1}{k} \cdot \Omega\left(\frac{n}{2^{31k}}\right)^{1/29}\right)$	Beame and Huynh-Ngoc [11]
read-once depth-3 $\{\wedge, \vee\}$-formula	$\exp\left(\frac{1}{k} \cdot \Omega\left(\frac{n}{4^k}\right)^{1/7}\right)$	This paper

Table B: Lower bounds for computing functions in ACC and AC^0 by circuits of type MAJ $\circ$ SYMM $\circ$ ANY with bottom fan-in k. All functions are on $n(k+1)$ bits.

least degree of a real polynomial that approximates f within ϵ pointwise. The approximate degree is a thoroughly studied quantity, with tight estimates known for various ϵ and various functions of interest to us. By taking $\epsilon = 1/3$, we derive lower bounds for bounded-error communication, including the lower bound for set disjointness ($f = \mathrm{AND}$). Letting $\epsilon \nearrow 1/2$, we obtain lower bounds for protocols with error exponentially close to random guessing, including the discrepancy result for constant-depth circuits.

In the setting of bounded-error communication, we are further able to give a near-optimal lower bound on the k-party communication complexity of every composition of the form $f(\mathrm{OR}_k \vee \mathrm{AND}_k, \ldots, \mathrm{OR}_k \vee \mathrm{AND}_k)$, where f is an arbitrary Boolean function. The same holds for XOR lemmas and direct-product theorems. Finally, in this introduction and throughout the paper, we focus on randomized communication in the context of *classical* multiparty protocols; however, by the results of [34, 15], our randomized lower bounds carry over in full to the quantum model. We defer formal statements of the quantum multiparty lower bounds and background on quantum multiparty protocols to the final version of this paper.

Previous analyses. In a precise technical sense, our approach to set disjointness is the opposite of previous multiparty analyses [17, 35, 19, 21, 22, 11]. In the overview that follows, we describe the limitations of previous analyses and how this paper overcomes them. Let $F\colon (\{0,1\}^n)^k \to \{0,1\}$ be a given k-party communication problem. A fundamental fact [6] in communication complexity is that a cost-c deterministic protocol for F gives a representation $F = \chi_1 + \chi_2 + \cdots + \chi_{2^c}$, where the χ_i are highly structured Boolean functions called *cylinder intersections*. This fact immediately generalizes to randomized communication since a cost-c randomized protocol is a probability distribution on cost-c deterministic protocols. Specifically, a cost-c randomized protocol for F gives a representation $F \approx \sum_\chi a_\chi \chi$, where the sum is over cylinder intersections and $\sum |a_\chi| \leqslant 2^c$. How these representations arise is immaterial in this discussion; what matters is that proofs of high communication complexity typically work by bounding the *correlations* of a relevant function with cylinder intersections. The simplest such technique is the *discrepancy method* [20, 6, 33], whereby one carefully chooses a probability distribution μ on the domain $(\{0,1\}^n)^k$ and argues that F has small correlation under μ with all cylinder intersections. This property of F is referred to as *small discrepancy* with respect to μ, hence the name of the technique. A more powerful technique is the *generalized discrepancy method* [28, 40], whereby one constructs a real function Ψ such that Ψ is highly correlated with F but almost uncorrelated with cylinder intersections.

Even in the two-party setting, it is difficult to construct the right μ or Ψ and analyze the associated correlations. To illustrate, it was an open problem until recently whether AC^0 circuits have small two-party discrepancy with respect to some distribution. This problem was solved in considerable generality four years ago in [43, 42]. The technique developed in that work, called the *pattern matrix method*, automates the choice of μ and Ψ as well as the subsequent analysis of correlations for a class of communication problems. The communication problems F to which the pattern matrix method applies are of the following form. Let $f\colon \{0,1\}^m \to \{0,1\}$ be a given function, fixed once and for all. In the two-party model, the first party receives a bit string $x \in \{0,1\}^n$ where $n \gg m$, the second party receives a subset $S \subset \{1,2,\ldots,n\}$ of cardinality m, and their goal is to compute $F(x,S) = f(x|_S)$. In other words, the answer only depends on a few of the input bits $x_1, x_2, \ldots, x_n$; the first party knows all the bits but does not know which ones are relevant, whereas the second party knows which bits are relevant but does not know their values. For

any f, the pattern matrix method uses a well-studied approximation-theoretic property—namely, the impossibility of approximating f in the infinity norm by a real polynomial of given degree d—to construct the desired μ and Ψ. This results in a lower bound of $\Omega(d)$ on the randomized communication complexity.

Originally formulated in [43, 42] for the two-party model, the pattern matrix method has been adapted to three or more parties by several authors [17, 35, 19, 21, 22, 11], resulting among other things in improved multiparty lower bounds for set disjointness. Analogous to the two-party setting, one starts with a function $f\colon \{0,1\}^m \to \{0,1\}$. In the case of k parties, the inputs to the communication problem are Boolean strings $x, y_1, y_2, \ldots, y_{k-1}$ and the goal is to compute $F(x, y_1, y_2, \ldots, y_{k-1}) = f\big(x|_{S(y_1,y_2,\ldots,y_{k-1})}\big)$, where the *selector* $S(y_1, y_2, \ldots, y_{k-1})$ is some mapping into cardinality-m subsets. In other words, in the multiparty setting, the bit strings $y_1, y_2, \ldots, y_{k-1}$ jointly determine to which bits f is to be applied. What fundamentally differentiates the various multiparty extensions of the pattern matrix method [17, 35, 19, 21, 22, 11] is the definition of the selector. The simpler the selector, the more widely applicable the communication lower bounds—and the harder they are to prove. Arguably the simplest meaningful selector is a small CNF or DNF formula. This is the selector used in the original two-party pattern matrix method [43, 42] as well as its first multiparty adaptations [17, 35, 19]. This selector is desirable in that it embeds nicely in the disjointness function and thus directly gives communication lower bounds for this problem. While the simple selector works well for $k = 2$ parties, the situation changes qualitatively at $k = 3$, and the multiparty lower bounds degrade rapidly with k. In the case of set disjointness, one obtains a lower bound of $\Omega(\sqrt{n})$ for $k = 2$ parties [42] and $\Omega(n^{1/(k+1)}/2^{2^k})$ for $k \geqslant 3$ parties [17, 35]. At the other extreme, one can use the most complicated selector possible, namely, a random mapping. This is the approach taken in [21]. The communication lower bounds for the random selector remain very strong up to $k \approx \log n$ parties, which is excellent. The problem here, of course, is that the random selector cannot be computed by any constant-depth circuit, let alone a small CNF formula, and thus the communication lower bounds do not apply to set disjointness. Finally, selectors of intermediate complexity were considered in [22, 11], using bounded independence and XOR gates. These ideas were fruitful, giving an explicit separation of NP_k and BPP_k in communication [22] and strong multiparty lower bounds for constant-depth circuits [11]. However, a strong lower bound for set disjointness has remained off-limits—the above approaches have not yielded a bound better than $\Omega(n^{1/(k+1)}/2^{k^2})$ for the problem.

At a technical level, previous approaches to multiparty set disjointness face the following fundamental difficulty. Recall that the goal in multiparty lower bounds is to bound correlations of relevant functions with cylinder intersections. For the past twenty-two years [6], the maximum correlation of a given real function $\Psi(x, y_1, \ldots, y_{k-1})$ with a cylinder intersection is bounded in terms of the expected product of various subfunctions of Ψ. Concretely, one bounds the correlation in terms of the quantity

$$\Delta(\Psi) = \left(\mathop{\mathbf{E}}_{y_1', y_1''} \mathop{\mathbf{E}}_{y_2', y_2''} \cdots \mathop{\mathbf{E}}_{y_{k-1}', y_{k-1}''} \left| \mathop{\mathbf{E}}_{x} \left[\prod_{z \in \{',''\}^{k-1}} \Psi(x, y_1^{z_1}, y_2^{z_2}, \ldots, y_{k-1}^{z_{k-1}}) \right] \right| \right)^{1/2^{k-1}}.$$

In the pattern matrix method, the relevant function is of the form $\Psi(x, y_1, y_2, \ldots, y_{k-1}) = \psi(x|_{S(y_1,\ldots,y_{k-1})})$, where the low-order Fourier coefficients of ψ are zero. This Fourier-theoretic fact is what allows one to bound $\Delta(\Psi)$. Specifically, previous papers—starting with the original two-party work [43, 42]—argue that the 2^{k-1}-fold product will likely have 0 for the constant Fourier coefficient and thus zero expectation. To handle the unlikely complementary event, one needs to additionally control the growth of the 2^{k-1}-fold product. As the number of parties k grows, this argument requires an increasingly complex selector.

Our proof. Our proof *reverses* the steps in the above argument: we first apply the Fourier-theoretic property and then the correlation bound. In more detail, we first write ψ in terms of its Fourier expansion $\psi = \sum_A \hat{\psi}(A)\chi_A$, where χ_A denotes a character of the Fourier transform. We then observe that by linearity, the correlation of Ψ with a cylinder intersection is bounded by $\sum_A |\hat{\psi}(A)|\Delta(\chi_A)$. From then on, we work with the quantities $\Delta(\chi_A)$, whereas previous multiparty papers work directly with $\Delta(\psi)$. Intuitively, the switch from an arbitrary real function ψ to χ_A is motivated by the convenient multiplicative structure of the characters and their global boundedness.

This leaves us with the challenge of proving a strong upper bound on $\Delta(\chi_A)$ with a selector as weak as a small CNF formula. Prior to this paper, it was unclear whether it could be done at all, let alone how to do it. Indeed, the "reverse" argument has been widely known to researchers since 2007; it was introduced as a refinement [42] of the two-party pattern matrix method and as part of another duality-based technique [46]. Previous attempts to use the reverse argument in multiparty communication were unsuccessful. In particular, it was shown in [18, p. 189] that its direct application gives a lower bound worse than $(n/k^k)^{1/(2+k2^k)}$ on the $(k+1)$-party communication complexity of set disjointness, which is substantially weaker than the lower bound [35, 19] obtained by following the steps of the pattern matrix method in the original order.

We are nevertheless able to prove a strong, essentially exact bound on $\Delta(\chi_A)$ for a selector which is computable by a small CNF formula and thus efficiently embeds in the disjointness function. This part of the proof exploits *metric* properties of distributions induced by set disjointness on the Boolean cube (such as conditional independence), in contrast to the *Fourier-theoretic* content of previous work. Specifically, we use conditioning to make appropriate variables independent and thereby simulate XOR-like behavior with an OR gate. This simulation is of course only approximate, and the bounding of error terms is done via a different careful conditioning. The argument proceeds by induction on the number of parties, the base case admitting a first-principles solution.

Once $\Delta(\chi_A)$ has been bounded, we are in a strong position to prove Theorem 1.1 and its generalization to arbitrary compositions and arbitrary error rates. To be more precise, we give two alternate proofs of this result. One proof is based, like previous work [43, 42, 46, 17, 35, 19, 21, 22, 11], on the dual view of the problem. The other proof is quite different and works with the primal view, explicitly converting a low-cost protocol for a given communication problem into a low-degree approximant for the given Boolean function. In addition to being more intuitive, the primal approach allows us to prove the direct product theorems and XOR lemmas for set disjointness (Theorem 1.2), by reducing them to a corresponding direct product theorem and XOR lemma for polynomial approximation and appealing to known results in that setting [45].

Once Theorem 1.1 has been established, we are able to use it almost like a black box to obtain Theorems 1.3 and 1.4 on nondeterministic and Merlin-Arthur communication.

Specifically, we are able to apply an earlier argument for these models, due to Gavinsky and the author [23], using the new randomized lower bound of Theorem 1.1 in place of earlier bounds.

2. PRELIMINARIES

From now on we will view Boolean functions as mappings $f: X \to \{-1, +1\}$ for some finite set X, where -1 and $+1$ correspond to "true" and "false," respectively. A *partial function* f on a set X is a function whose domain of definition, denoted $\operatorname{dom} f$, is a proper subset of X. For emphasis, we will sometimes refer to functions with $\operatorname{dom} f = X$ as *total.* We use lowercase letters (x, y, u, v) for vectors and Boolean strings, and uppercase letters (A, B, X, Y) for real and Boolean matrices. The empty string is denoted ε. The complement of a set S is denoted $\overline{S}$.

For a bit string $x \in \{0,1\}^n$, we let $|x| = x_1 + x_2 + \cdots + x_n$. The componentwise conjunction of $x, y \in \{0,1\}^n$ is denoted $x \wedge y = (x_1 \wedge y_1, \ldots, x_n \wedge y_n)$. In particular, $|x \wedge y|$ refers to the number of components in which x and y both have a 1. The bitwise negation of a string $x \in \{0,1\}^n$ is denoted $\overline{x} = (1 - x_1, \ldots, 1 - x_n)$. For a string $x = (x_1, \ldots, x_n)$ and a set $S \subseteq \{1, 2, \ldots, n\}$, we adopt the shorthand $x|_S = (x_{i_1}, x_{i_2}, \ldots, x_{i_{|S|}})$, where $i_1 < i_2 < \cdots < i_{|S|}$ are the elements of S. For convenience, we adopt the convention that $0/0 = 0$. The symbol $\doteq$ means "equal by definition." The indicator function of a logical condition C is given by

$$\mathbf{I}[C] = \begin{cases} 1 & \text{if } C \text{ holds,} \\ 0 & \text{otherwise.} \end{cases}$$

The set membership sign $\in$, when used in the subscript of an expectation operator, means that the expectation is taken over a uniformly random member of the indicated set. The uniform distribution on $\{0,1\}^n$ is denoted $\mathscr{U}_n$. The notation $\log x$ refers to the logarithm of x to base 2. For a real function ϕ on a finite set X, the *support* of ϕ is the subset $\operatorname{supp} \phi = \{x \in X : \phi(x) \neq 0\}$. For probability distributions μ and λ on finite sets X and Y, respectively, the symbol $\mu \times \lambda$ refers to the probability distribution on $X \times Y$ given by $(\mu \times \lambda)(x, y) = \mu(x)\lambda(y)$. The total degree of a multivariate real polynomial p is denoted $\deg p$. For (possibly partial) Boolean functions f and g on $\{-1, +1\}^n$ and X, respectively, the symbol $f \circ g$ refers to the (possibly partial) Boolean function on X^n given by $(f \circ g)(x_1, x_2, \ldots, x_n) = f(g(x_1), g(x_2), \ldots, g(x_n))$. Clearly, the domain of $f \circ g$ is the set of all $(x_1, x_2, \ldots, x_n) \in (\operatorname{dom} g)^n$ for which $(g(x_1), g(x_2), \ldots, g(x_n)) \in \operatorname{dom} f$.

The symbol $\{0,1\}^{n \times k}$ denotes the family of $n \times k$ matrices with entries 0, 1. The notation $(\{0,1\}^n)^k$ refers to the set of vector sequences $(x_1, x_2, \ldots, x_k)$, where each $x_i \in \{0,1\}^n$. Throughout this paper, we identify the sets $\{0,1\}^{n \times k}$ and $(\{0,1\}^n)^k$. This means that $(x_1, x_2, \ldots, x_k)$ can be viewed both as a sequence of vectors in $\{0,1\}^n$ and as a matrix of size $n \times k$ with columns $x_1, x_2, \ldots, x_k$. Taking this convention a step further, we view $(X_1, X_2, \ldots, X_r)$ as an element of $(\{0,1\}^n)^{k_1 + \cdots + k_r} \equiv \{0,1\}^{n \times (k_1 + \cdots + k_r)}$ whenever $X_i \in \{0,1\}^{n \times k_i}$ $(i = 1, 2, \ldots, r)$.

For $X \in \{0,1\}^{n \times k}$, the *disjointness predicate* $D(X) \in \{-1, +1\}$ is defined as $D(X) = -1$ if and only if each row of X has a 0 entry. By the convention of the previous paragraph, this also gives meaning to the symbols $D(X_1, X_2, \ldots, X_r)$ and $D(x_1, x_2, \ldots, x_k)$, where

$X_i \in \{0,1\}^{n \times k_i}$ and $x_i \in \{0,1\}^n$. For example,

$$D(x_1, x_2, \ldots, x_k) = \begin{cases} -1 & \text{if } x_1 \wedge x_2 \wedge \cdots \wedge x_k = 0^n, \\ 1 & \text{otherwise.} \end{cases}$$

By convention,

$$D(\varepsilon) = -1.$$

For a Boolean matrix $X = [X_{i,j}] \in \{0,1\}^{n \times k}$ and a string $y \in \{0,1\}^n$, we let $X|_y$ denote the submatrix of X obtained by keeping only those rows i for which $y_i = 1$. More formally,

$$X|_y = \begin{bmatrix} X_{i_1,1} & X_{i_1,2} & \ldots & X_{i_1,k} \\ X_{i_2,1} & X_{i_2,2} & \ldots & X_{i_2,k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{i_{|y|},1} & X_{i_{|y|},2} & \ldots & X_{i_{|y|},k}, \end{bmatrix}$$

where $i_1 < i_2 < \cdots < i_{|y|}$ are the indices with $y_{i_1} = y_{i_2} = \cdots = y_{i_{|y|}} = 1$. In particular, $X|_{0^n} = \varepsilon$. It is useful to keep in mind that

$$D(X, y) \equiv D(X|_y).$$

The familiar functions AND_n, OR_n, and PARITY_n on the Boolean hypercube $\{-1,+1\}^n$ are given by $\mathrm{AND}_n(x) = \bigwedge_{i=1}^n x_i$, $\mathrm{OR}_n(x) = \bigvee_{i=1}^n x_i$, and $\mathrm{PARITY}_n(x) = \bigoplus_{i=1}^n x_i$. We also define a partial Boolean function $\widetilde{\mathrm{AND}}_n$ on $\{-1,+1\}^n$ as the restriction of AND_n to the set $\{x : |\{i : x_i = -1\}| \geqslant n-1\}$. In other words,

$$\widetilde{\mathrm{AND}}_n(x) = \begin{cases} \mathrm{AND}_n(x) & \text{if } |\{i : x_i = -1\}| \geqslant n-1, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Analogously, we define a partial Boolean function $\widetilde{\mathrm{OR}}_n$ on $\{-1,+1\}^n$ as the restriction of OR_n to the set $\{x : |\{i : x_i = -1\}| \leqslant 1\}$.

Norms and products. For a finite set X, the linear space of real functions on X is denoted $\mathbb{R}^X$. This space is equipped with the usual norms and inner product:

$$\begin{aligned} \|\phi\|_\infty &= \max_{x \in X} |\phi(x)| && (\phi \in \mathbb{R}^X), \\ \|\phi\|_1 &= \sum_{x \in X} |\phi(x)| && (\phi \in \mathbb{R}^X), \\ \langle \phi, \psi \rangle &= \sum_{x \in X} \phi(x)\psi(x) && (\phi, \psi \in \mathbb{R}^X). \end{aligned}$$

The tensor product of $\phi \in \mathbb{R}^X$ and $\psi \in \mathbb{R}^Y$ is the function $\phi \otimes \psi \in \mathbb{R}^{X \times Y}$ given by $(\phi \otimes \psi)(x, y) = \phi(x)\psi(y)$. The tensor product $\phi \otimes \phi \otimes \cdots \otimes \phi$ (n times) is abbreviated $\phi^{\otimes n}$. When specialized to real matrices, the tensor product is the usual Kronecker product. The pointwise (Hadamard) product of $\phi, \psi \in \mathbb{R}^X$ is denoted $\phi \circ \psi \in \mathbb{R}^X$ and given by $(\phi \circ \psi)(x) = \phi(x)\psi(x)$. Note that as functions, $\phi \circ \psi$ is a restriction of $\phi \otimes \psi$. Tensor product notation generalizes to partial functions in the natural way: if ϕ and ψ are partial real functions on X and Y, respectively, then $\phi \otimes \psi$ is a partial function on $X \times Y$ with domain $\operatorname{dom}\phi \times \operatorname{dom}\psi$ and is given by $(\phi \otimes \psi)(x, y) = \phi(x)\psi(y)$ on that domain. Similarly, $\phi^{\otimes n} = \phi \otimes \phi \otimes \cdots \otimes \phi$ (n times) is a partial function on X^n with domain $(\operatorname{dom}\phi)^n$.

The spectral norm of a real matrix A is given by $\|A\| = \max_{x \neq 0} \|Ax\|_2/\|x\|_2$, where $\|\cdot\|_2$ stands for the Euclidean norm on vectors. The spectral norm is multiplicative with respect to tensor product: $\|A \otimes B\| = \|A\|\,\|B\|$.

Fourier transform. Consider the real vector space of functions $\{-1,+1\}^n \to \mathbb{R}$. For $S \subseteq \{1,2,\ldots,n\}$, define $\chi_S\colon \{-1,+1\}^n \to \{-1,+1\}$ by $\chi_S(x) = \prod_{i \in S} x_i$. Then every function $\phi\colon \{-1,+1\}^n \to \mathbb{R}$ has a unique representation of the form $\phi = \sum_S \hat{\phi}(S)\,\chi_S$, where $\hat{\phi}(S) = 2^{-n} \sum_{x \in \{-1,+1\}^n} \phi(x)\chi_S(x)$. The reals $\hat{\phi}(S)$ are called the *Fourier coefficients of* ϕ. The following fact is immediate from the definition of $\hat{\phi}(S)$:

PROPOSITION 2.1. *For all* $\phi\colon \{-1,+1\}^n \to \mathbb{R}$,

$$\max_{S \subseteq \{1,2,\ldots,n\}} |\hat{\phi}(S)| \leqslant 2^{-n}\|\phi\|_1.$$

Approximation by polynomials. Let $\phi\colon X \to \mathbb{R}$ be given, for a finite subset $X \subset \mathbb{R}^n$. The *ϵ-approximate degree* of ϕ, denoted $\deg_\epsilon(\phi)$, is the least degree of a real polynomial p such that $\|\phi - p\|_\infty \leqslant \epsilon$. We generalize this definition to partial functions ϕ on X by letting $\deg_\epsilon(\phi)$ be the least degree of a real polynomial p with

$$\text{(2.1)} \quad \begin{cases} |\phi(x) - p(x)| \leqslant \epsilon, & x \in \operatorname{dom}\phi, \\ |p(x)| \leqslant 1 + \epsilon, & x \in X \setminus \operatorname{dom}\phi. \end{cases}$$

For a (possibly partial) real function ϕ on a finite subset $X \subset \mathbb{R}^n$, we define $E(\phi, d)$ to be the least ϵ such that (2.1) holds for some polynomial of degree at most d. In this notation, $\deg_\epsilon(\phi) = \min\{d : E(\phi,d) \leqslant \epsilon\}$. We will need the following dual characterization of the approximate degree.

FACT 2.2. *Let* ϕ *be a (possibly partial) real function on* $\{-1,+1\}^n$. *Then* $\deg_\epsilon(\phi) > d$ *if and only if there exists* $\psi\colon \{-1,+1\}^n \to \mathbb{R}$ *such that*

$$\sum_{x \in \operatorname{dom}\phi} \phi(x)\psi(x) - \sum_{x \notin \operatorname{dom}\phi} |\psi(x)| - \epsilon\|\psi\|_1 > 0,$$

and $\hat{\psi}(S) = 0$ *for* $|S| \leqslant d$.

Fact 2.2 follows from linear programming duality; see [45, 42] for details.

A related notion is the *threshold degree* $\deg_\pm(f)$, defined for a (possibly partial) Boolean function f as the limit $\deg_\pm(f) = \lim_{\epsilon \searrow 0} \deg_{1-\epsilon}(f)$. Equivalently, $\deg_\pm(f)$ is the least degree of a real polynomial p with $f(x) = \operatorname{sgn} p(x)$ for $x \in \operatorname{dom} f$. We recall two well-known results on the polynomial approximation of Boolean functions, the first due to Minsky and Papert [37] and the second due to Nisan and Szegedy [38].

THEOREM 2.3 (Minsky and Papert). *The function* $\mathrm{MP}_n(x) = \bigvee_{i=1}^n \bigwedge_{j=1}^{4n^2}$ *obeys*

$$\deg_\pm(\mathrm{MP}_n) = n.$$

THEOREM 2.4 (Nisan and Szegedy). *The functions* AND_n *and* $\widetilde{\mathrm{AND}}_n$ *obey*

$$\deg_{1/3}(\mathrm{AND}_n) \geqslant \deg_{1/3}(\widetilde{\mathrm{AND}}_n) = \Theta(\sqrt{n}).$$

Multiparty communication. An excellent reference on communication complexity is the monograph by Kushilevitz and Nisan [33]. In this overview, we will limit ourselves to key definitions and notation. The simplest model of communication in this work is the two-party randomized model. Consider a (possibly partial) Boolean function F on $X \times Y$, where X and Y are finite sets. Alice receives an input $x \in X$, Bob receives $y \in Y$, and their objective is to compute $F(x, y)$ with high accuracy whenever $(x, y) \in \operatorname{dom} F$. To this end, Alice and Bob share a communication channel and have an unlimited supply of shared random bits. Alice and Bob's protocol is said to have *error* ϵ if on every input $(x, y) \in \operatorname{dom} F$, the computed output differs from the correct answer $F(x, y)$ with probability no greater than ϵ. The *cost* of a given protocol is the maximum number of bits exchanged on any input. The *ϵ-error randomized communication complexity* of F, denoted $R_\epsilon(F)$, is the least cost of an ϵ-error protocol for F. The canonical quantity to study is $R_{1/3}(F)$, where the choice of $1/3$ is largely arbitrary since the error probability of a protocol can be decreased from $1/3$ to any other positive constant at the expense of increasing the communication cost by a constant factor.

A generalization of two-party communication is the *multiparty number-on-the-forehead* model, due to Chandra, Furst, and Lipton [16]. Here one considers a (possibly partial) Boolean function F on $X_1 \times X_2 \times \cdots \times X_k$, for some finite sets $X_1, X_2, \ldots, X_k$. There are k parties. A given input $(x_1, x_2, \ldots, x_k) \in X_1 \times X_2 \times \cdots \times X_k$ is distributed among the parties by placing x_i on the forehead of party i (for $i = 1, 2, \ldots, k$). In other words, party i knows $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$ but not x_i. The parties communicate by writing bits on a shared blackboard, visible to all. They additionally have access to a shared source of random bits. Their goal is to devise a communication protocol that will allow them to accurately predict the value of F everywhere on the domain of F. As before, an ϵ-error protocol for F is one which, on every input $(x_1, x_2, \ldots, x_k) \in \operatorname{dom} F$, produces the correct answer $F(x_1, x_2, \ldots, x_k)$ with probability at least $1 - \epsilon$. The *cost* of a communication protocol is the total number of bits written to the blackboard on the worst-case input. Analogous to the two-party case, the randomized communication complexity $R_\epsilon(F)$ is the least cost of an ϵ-error communication protocol for F in this model.

Let G be a (possibly partial) Boolean function on $X_1 \times X_2 \times \cdots \times X_k$, representing a k-party communication problem, and let f be a (possibly partial) Boolean function on $\{-1, +1\}^n$. We view the composition $f \circ G$ as a k-party communication problem on $X_1^n \times X_2^n \times \cdots \times X_k^n$. The primary problem of interest to us is *set disjointness* $\mathrm{DISJ}_{n,k} \colon (\{0,1\}^n)^k \to \{-1, +1\}$, given by $\mathrm{DISJ}_{n,k}(x_1, x_2, \ldots, x_k) = D(x_1, x_2, \ldots, x_k)$. We will also consider the partial Boolean function $\mathrm{UDISJ}_{n,k}$ on $(\{0,1\}^n)^k$ given by: $\mathrm{UDISJ}_{n,k}(x_1, x_2, \ldots, x_n) = D(x_1, x_2, \ldots, x_n)$ when $|x_1 \wedge x_2 \wedge \cdots \wedge x_k| \leqslant 1$, and undefined otherwise. The k-party communication problem that corresponds to $\mathrm{UDISJ}_{n,k}$ is known as *unique set disjointness*. In other words, unique set disjointness is a promise version of set disjointness where the input matrix $(x_1, x_2, \ldots, x_k) \in \{0,1\}^{n \times k}$ is guaranteed to have at most one row consisting entirely of ones. It will be helpful to keep in mind that for all positive integers r, s, one has $\mathrm{DISJ}_{rs,k} = \mathrm{AND}_r \circ \mathrm{DISJ}_{s,k}$ and analogously $\mathrm{UDISJ}_{rs,k} = \widetilde{\mathrm{AND}}_r \circ \mathrm{UDISJ}_{s,k}$.

A k-dimensional *cylinder intersection* is a function $\chi \colon X_1 \times X_2 \times \cdots \times X_k \to \{0, 1\}$ of the form

$$\chi(x_1, \ldots, x_k) = \prod_{i=1}^{k} \chi_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k),$$

where $\chi_i: X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_k \to \{0,1\}$. In other words, a k-dimensional cylinder intersection is the product of k functions with range $\{0,1\}$, where the ith function does not depend on the ith coordinate but may depend arbitrarily on the other $k-1$ coordinates. Cylinder intersections were introduced by Babai, Nisan, and Szegedy [6] and play a fundamental role in the theory due to the following fact.

FACT 2.5. *Let $\Pi: X_1 \times X_2 \times \cdots \times X_k \to \{-1,+1\}$ be a deterministic k-party communication protocol with cost c. Then*

$$\Pi = \sum_{i=1}^{2^c} a_i \chi_i$$

for some cylinder intersections $\chi_1, \ldots, \chi_{2^c}$ with pairwise disjoint support and $a_1, \ldots, a_{2^c} \in \{-1,+1\}$.

Recall that a randomized protocol with cost c is a probability distribution on deterministic protocols of cost c. Therefore, Fact 2.5 immediately implies the following two results on randomized communication complexity.

COROLLARY 2.6. *Let F be a (possibly partial) Boolean function on $X_1 \times X_2 \times \cdots \times X_k$. If $R_\epsilon(F) = c$, then*

$$|F(x_1,\ldots,x_k) - \Pi(x_1,\ldots,x_k)| \leqslant \frac{\epsilon}{1-\epsilon}, \qquad (x_1,\ldots,x_k) \in \operatorname{dom} F,$$

$$|\Pi(x_1,\ldots,x_k)| \leqslant \frac{1}{1-\epsilon}, \qquad (x_1,\ldots,x_k) \in X_1 \times \cdots \times X_k,$$

where $\Pi = \sum_\chi a_\chi \chi$ is a linear combination of cylinder intersections with $\sum_\chi |a_\chi| \leqslant 2^c/(1-\epsilon)$.

COROLLARY 2.7. *Let Π be a randomized k-party protocol with domain $X_1 \times X_2 \times \cdots \times X_k$. If Π has communication cost c bits, then*

$$\mathbf{P}[\Pi(x_1, x_2, \ldots, x_k) = -1] \equiv \sum_\chi a_\chi \chi(x_1, x_2, \ldots, x_k)$$

on $X_1 \times X_2 \times \cdots \times X_k$, where the sum is over cylinder intersections and $\sum_\chi |a_\chi| \leqslant 2^c$.

Discrepancy and generalized discrepancy. For a communication problem $F: X_1 \times X_2 \times \cdots \times X_k \to \{-1,+1\}$ and a probability distribution P on $X_1 \times X_2 \times \cdots \times X_k$, the *discrepancy* of F with respect to P is defined as

$$\operatorname{disc}_P(F) = \max_\chi |\langle F \circ P, \chi \rangle|,$$

where the maximum is over cylinder intersections. We generalize this definition to partial functions as follows: for a partial Boolean function F on $X_1 \times X_2 \times \cdots \times X_k$ and a probability distribution P on $X_1 \times X_2 \times \cdots \times X_k$,

$$\operatorname{disc}_P(F) = \sum_{x \notin \operatorname{dom} F} P(x) + \max_\chi \left| \sum_{x \in \operatorname{dom} F} F(x) P(x) \chi(x) \right|,$$

where the maximum is again over cylinder intersections; this agrees with the previous definition if $\operatorname{dom} F = X_1 \times X_2 \times \cdots \times X_k$. The least discrepancy over all distributions is denoted $\operatorname{disc}(F) = \min_P \operatorname{disc}_P(F)$. Estimating the discrepancy is difficult and represents

a central obstacle in multiparty communication complexity. In two-party communication, the following method is frequently useful.

PROPOSITION 2.8 (see Kushilevitz and Nisan [33]). *Fix a function $F\colon X \times Y \to \{-1,+1\}$ and probability distribution P on $X \times Y$. Define $\Phi \doteq [F(x,y)P(x,y)]_{x\in X, y\in Y}$. Then*

$$\operatorname{disc}_P(F) \leqslant \|\Phi\| \sqrt{|X|\,|Y|}.$$

As Fact 2.5 suggests, upper bounds on the discrepancy give lower bounds on communication complexity. This technique is known as the *discrepancy method*. The original treatment of the discrepancy method [20, 6, 33] was specialized to total Boolean functions. In the theorem that follows, we extend the method to partial functions.

THEOREM 2.9 (Discrepancy method). *Let F be a (possibly partial) Boolean function on $X_1 \times X_2 \times \cdots \times X_k$. Then*

$$2^{R_\epsilon(F)} \geqslant \frac{1-2\epsilon}{\operatorname{disc}(F)}.$$

Proof. (Based on [33, pp. 36–38].) Fix a probability distribution P such that $\operatorname{disc}(F) = \operatorname{disc}_P(F)$. The distribution P' induced by P on $\operatorname{dom} F$ satisfies $\operatorname{disc}_{P'}(F) \leqslant \operatorname{disc}_P(F)$. Thus, we may assume that $\operatorname{supp} P \subseteq \operatorname{dom} F$ to start with.

Now, suppose that F has a communication protocol with error ϵ and cost c. Approximate F uniformly by $\Pi = \sum_\chi a_\chi \chi$ as in Corollary 2.6. Then

$$\begin{aligned}\sum_{\operatorname{dom} F} (F(x) - \Pi(x))F(x)P(x) &\leqslant \left(\max_{\operatorname{dom} F} |F(x) - \Pi(x)|\right) \sum_{\operatorname{dom} F} P(x) \\ &\leqslant \frac{\epsilon}{1-\epsilon}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\sum_{\operatorname{dom} F} (F(x) - \Pi(x))F(x)P(x) &= \sum_{\operatorname{dom} F} P(x) - \sum_{\operatorname{dom} F} \Pi(x)F(x)P(x) \\ &\geqslant 1 - \sum_\chi |a_\chi| \left| \sum_{\operatorname{dom} F} \chi(x)F(x)P(x) \right| \\ &\geqslant 1 - \frac{2^c}{1-\epsilon} \operatorname{disc}_P(F).\end{aligned}$$

The claimed lower bound on 2^c follows. □

A more general technique, originally applied by Klauck [28] in the two-party quantum model and subsequently adapted to many other settings [40, 36, 42, 35, 19], is the *generalized discrepancy method*. Again, previous treatments focused on total Boolean functions. In what follows, we derive a version of the method that applies to partial functions as well.

THEOREM 2.10 (Generalized discrepancy method). *Let F be a (possibly partial) Boolean function on $X_1 \times X_2 \times \cdots \times X_k$. Then for every nonzero $\Psi\colon X_1 \times X_2 \times \cdots \times X_k \to \mathbb{R}$,*

$$2^{R_\epsilon(F)} \geqslant \frac{1-\epsilon}{\max_\chi |\langle \chi, \Psi \rangle|} \left\{ \sum_{x \in \operatorname{dom} F} F(x)\Psi(x) - \sum_{x \notin \operatorname{dom} F} |\Psi(x)| - \frac{\epsilon}{1-\epsilon} \|\Psi\|_1 \right\},$$

where the maximum is over cylinder intersections χ.

Proof. (Based on [40, 42, 35, 19]). Suppose that F has a communication protocol with error ϵ and cost c. Extend F to a total function $G\colon X_1 \times \cdots \times X_k \to \mathbb{R}$ by letting $G = 0$ outside dom F. By Corollary 2.6, there is a linear combination of cylinder intersections $\Pi = \sum_\chi a_\chi \chi$ such that $\sum_\chi |a_\chi| \leqslant 2^c/(1-\epsilon)$ and Π approximates F in the sense that $\|\Pi\|_\infty \leqslant 1/(1-\epsilon)$ and $|F - \Pi| \leqslant \epsilon/(1-\epsilon)$ on the domain of F. It follows that

$$\langle G - \Pi, \Psi\rangle \leqslant \frac{\epsilon}{1-\epsilon} \sum_{x \in \operatorname{dom} F} |\Psi(x)| + \frac{1}{1-\epsilon} \sum_{x \notin \operatorname{dom} F} |\Psi(x)|.$$

However,

$$\begin{aligned}\langle G - \Pi, \Psi\rangle &\geqslant \sum_{x \in \operatorname{dom} F} F(x)\Psi(x) - \sum_\chi |a_\chi|\, |\langle \Psi, \chi\rangle| \\ &\geqslant \sum_{x \in \operatorname{dom} F} F(x)\Psi(x) - \frac{2^c}{1-\epsilon} \max_\chi |\langle \Psi, \chi\rangle|.\end{aligned}$$

Comparing these two estimates of $\langle G - \Pi, \Psi\rangle$ gives the claimed lower bound on 2^c. □

3. Preparatory Work

For positive integers n, k, we let $\mu_{n,k}$ denote the uniform probability distribution on those matrices in $\{0,1\}^{n\times k}$ that have exactly one row composed of all ones. Thus $\mu_{n,k}$ is supported on $n(2^k - 1)^{n-1}$ matrices, each occurring with the same probability. For a Boolean matrix $Y = (y_1, \ldots, y_{k-1}) \in \{0,1\}^{n\times(k-1)}$, we consider the marginal probability distribution

$$\begin{aligned}\mu_{n,k}(Y) &\doteq \sum_{u \in \{0,1\}^n} \mu_{n,k}(Y, u) \\ &= \frac{|y_1 \wedge \cdots \wedge y_{k-1}|\, 2^{n - |y_1 \wedge \cdots \wedge y_{k-1}|}}{n(2^k - 1)^{n-1}}, \qquad (3.1)\end{aligned}$$

and the conditional probability

$$\mu_{n,k}(u \mid Y) \doteq \frac{\mu_{n,k}(Y, u)}{\mu_{n,k}(Y)}. \qquad (3.2)$$

Note that the argument to $\mu_{n,k}$ is a matrix of size either $n \times k$ or $n \times (k-1)$, depending on the meaning intended. Finally, we let $\lambda_{n,k}$ be the probability distribution on $\{0,1\}^{n\times(k-1)} \times \{0,1\}^n \times \{0,1\}^n$ given by

$$\lambda_{n,k}(Y, u, v) \doteq \mu_{n,k}(Y)\, \mu_{n,k}(u \mid Y)\, \mu_{n,k}(v \mid Y).$$

In other words, $\lambda_{n,k}$ corresponds to an experiment whereby one first chooses Y according to the marginal distribution (3.1) and then, given Y, chooses u and v independently according to the conditional distribution (3.2). The remainder of this section is devoted to establishing various metric properties of $\lambda_{n,k}$. The four lemmas that follow, Lemmas 3.1–3.4, are independent and can be read in any order. We alert the reader that we will refer to the distributions $\mu_{n,k}$ and $\lambda_{n,k}$ in later sections as well, without restating the definitions just given.

LEMMA 3.1 (On decomposition). *For each (Y,u,v) in the support of $\lambda_{n,k}$ and each $x \in \{0,1\}^n$,*

$$D(x,Y,u)D(x,Y,v) = \begin{cases} D(x,Y,u\wedge\overline{v})D(x,Y,\overline{u}\wedge v) & \text{if } D(Y,u,v)=-1,\\ 1 & \text{otherwise.}\end{cases}$$

Proof. By definition of $\lambda_{n,k}$, there is exactly one coordinate, call it i, where u and the columns of Y all have a 1. Analogously, there is exactly one coordinate, call it j, where v and the columns of Y all have a 1. When $D(Y,u,v)=1$, it follows that $i=j$ and hence $D(x,Y,u)D(x,Y,v) = (-1)^{x_i+x_j} = 1$. When $D(Y,u,v)=-1$, we have

$$\begin{aligned} D(x,Y,u) &= D(x,Y,u,v)\wedge D(x,Y,u,\overline{v}) = D(x,Y,u,\overline{v}),\\ D(x,Y,v) &= D(x,Y,u,v)\wedge D(x,Y,\overline{u},v) = D(x,Y,\overline{u},v),\end{aligned}$$

whence $D(x,Y,u)D(x,Y,v) = D(x,Y,u,\overline{v})D(x,Y,\overline{u},v)$ as claimed. □

LEMMA 3.2 (On conditional independence). *Let $(Y,u,v)\sim\lambda_{n,k}$. Conditioned on fixed values of u, v, $Y|_{u\wedge v}$, and $Y|_{\overline{u}\wedge\overline{v}}$ with $D(Y|_{u\wedge v})=-1$, the remaining parts $Y|_{u\wedge\overline{v}}$ and $Y|_{\overline{u}\wedge v}$ are independent and distributed according to $\mu_{|u\wedge\overline{v}|,k-1}$ and $\mu_{|\overline{u}\wedge v|,k-1}$, respectively.*

Proof. Put $Y=(y_1,\dots,y_{k-1})$. By (3.1) and (3.2), the support of $\lambda_{n,k}$ consists of tuples (Y,u,v) with

$$|y_1\wedge\cdots\wedge y_{k-1}\wedge u| = |y_1\wedge\cdots\wedge y_{k-1}\wedge v| = 1, \tag{3.3}$$

each such tuple with probability

$$\lambda_{n,k}(Y,u,v) = \frac{2^{|y_1\wedge\cdots\wedge y_{k-1}|}}{n(2^k-1)^{n-1}2^n\,|y_1\wedge\cdots\wedge y_{k-1}|}. \tag{3.4}$$

Now assign values to u, v, $Y|_{u\wedge v}$, and $Y|_{\overline{u}\wedge\overline{v}}$ such that $D(Y|_{u\wedge v})=-1$. We are to determine the conditional distribution of the remaining variables $Y|_{u\wedge\overline{v}}$ and $Y|_{\overline{u}\wedge v}$. It follows from (3.3) that each of these two matrices will have exactly one row made up entirely of ones. But by (3.4), any such assignment to $Y|_{u\wedge\overline{v}}$ and $Y|_{\overline{u}\wedge v}$ carries the same probability. Hence, $Y|_{u\wedge\overline{v}}$ and $Y|_{\overline{u}\wedge v}$ are independent and have the claimed distributions. □

LEMMA 3.3 (On expected intersection size). *For $\lambda_{n,k}$ defined above,*

$$\mathop{\mathbf{E}}_{\lambda_{n,k}}\left[\frac{\mathbf{I}[D(Y,u,v)=-1]}{\sqrt{|u\wedge\overline{v}|\,|\overline{u}\wedge v|}}\right] \leqslant \frac{4}{n}\cdot\frac{2^k-1}{2^k-2}.$$

Proof. By convexity,

$$\frac{1}{\sqrt{|u\wedge\overline{v}|\,|\overline{u}\wedge v|}} \leqslant \frac{1}{2|u\wedge\overline{v}|} + \frac{1}{2|\overline{u}\wedge v|}.$$

It is clear by symmetry that the strings $u\wedge\overline{v}$ and $\overline{u}\wedge v$ have identical distributions, leading to

$$\mathop{\mathbf{E}}_{\lambda_{n,k}}\left[\frac{\mathbf{I}[D(Y,u,v)=-1]}{\sqrt{|u\wedge\overline{v}|\,|\overline{u}\wedge v|}}\right] \leqslant \mathop{\mathbf{E}}_{\lambda_{n,k}}\left[\frac{\mathbf{I}[D(Y,u,v)=-1]}{|u\wedge\overline{v}|}\right]. \tag{3.5}$$

Let $y_1,\dots,y_{k-1}$ denote the columns of Y. Recall that $|y_1\wedge\cdots\wedge y_{k-1}\wedge u| = 1$ on the support of $\lambda_{n,k}$, whence by symmetry the right member of (3.5) is unchanged after

conditioning on $y_1 \wedge \cdots \wedge y_{k-1} \wedge u = 0^{n-1}1$. But then the first $n-1$ bits of $u \wedge \overline{v}$ are distributed independently, each taking on 1 with probability $p \doteq \frac{2^{k-1}-1}{2^k-1} \cdot \frac{1}{2}$, whereas the nth bit of $u \wedge \overline{v}$ takes on 1 whenever $D(Y,u,v) = -1$. These facts bound the right member of (3.5) from above by

$$\sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \frac{1}{i+1} = \frac{1-(1-p)^n}{pn} \leqslant \frac{4}{n} \cdot \frac{2^k-1}{2^k-2}. \qquad \square$$

LEMMA 3.4 (On the probability of disjointness). *For $\lambda_{n,k}$ defined above,*

$$\mathop{\mathbf{P}}_{\lambda_{n,k}} [D(Y,u,v) = 1] \leqslant \frac{2^k-1}{n}. \tag{3.6}$$

Proof. Conditioned on $Y = (y_1, \ldots, y_{k-1})$, the probability that $y_1 \wedge \cdots \wedge y_{k-1} \wedge u \wedge v$ is not the zero vector is exactly $1/|y_1 \wedge \cdots \wedge y_{k-1}|$. Thus, the left member of (3.6) equals

$$\begin{aligned}
&\sum_{y_1,\ldots,y_{k-1}} \frac{\mu_{n,k}(y_1,\ldots,y_{k-1})}{|y_1 \wedge \cdots \wedge y_{k-1}|} \\
&\qquad \leqslant \frac{2^{nk}}{n(2^k-1)^{n-1}} \mathop{\mathbf{E}}_{y_1,\ldots,y_{k-1} \in \{0,1\}^n} \left[\frac{1}{2^{|y_1 \wedge \cdots \wedge y_{k-1}|}} \right] \\
&\qquad = \frac{2^{nk}}{n(2^k-1)^{n-1}} \left(1 - \frac{1}{2^k}\right)^n \\
&\qquad = \frac{2^k-1}{n},
\end{aligned}$$

where the inequality holds by (3.1). $\square$

4. A DISCREPANCY RESULT

The goal of this section is to analyze the k-party discrepancy of $(\mathrm{UDISJ}_{n,k})^{\otimes m}$, the XOR of m independent copies of the unique disjointness problem. In actuality, we will derive a somewhat more general result. For positive integers $n_1, n_2, \ldots, n_m$, define

$$\Gamma_k(n_1,n_2,\ldots,n_m) \doteq \max_{\chi} \left| \mathop{\mathbf{E}}_{(x^1,W^1),\ldots,(x^m,W^m)} \left[\chi \cdot \prod_{i=1}^m D(x^i, W^i) \right] \right|,$$

where $(x^i, W^i) \sim \mathscr{U}_{n_i} \times \mu_{n_i,k}$ independently for each i, and the maximum is taken over all $(k+1)$-dimensional cylinder intersections $\chi\colon (\{0,1\}^{n_1+n_2+\cdots+n_m})^{k+1} \to \{0,1\}$. Our objective is to bound Γ_k from above. The proof will use induction on k, the base case corresponding to the following proposition.

PROPOSITION 4.1. *For all positive integers $n_1, n_2, \ldots, n_m$,*

$$\Gamma_1(n_1,n_2,\ldots,n_m) \leqslant \frac{1}{\sqrt{n_1 n_2 \cdots n_m}}.$$

Proof. Define $M_i = [(-1)^{x_j}]_{j,x}$, where the indices range as follows: $j = 1, 2, \ldots, n_i$ and $x \in \{0,1\}^{n_i}$. Then $\|M_i\| = \sqrt{2^{n_i}}$, whence by Proposition 2.8

$$\Gamma_1(n_1, n_2, \ldots, n_m) \leqslant \left\| \bigotimes_{i=1}^{m} \frac{M_i}{n_i 2^{n_i}} \right\| \left(\prod_{i=1}^{m} n_i 2^{n_i} \right)^{1/2} = \frac{1}{\sqrt{n_1 n_2 \cdots n_m}}. \qquad \square$$

We now proceed to bound Γ_k for all k. We will use the general technique of Babai, Nisan, and Szegedy [6] to pass from $(k+1)$-party discrepancy to k-party discrepancy. Unlike the functions studied in [6], however, set disjointness does not have a multiplicative structure. This complicates the passage from $(k+1)$-party problems to k-party subproblems. To overcome this difficulty, we use the metric properties of the distribution $\lambda_{n,k}$ established in the previous section. More concretely, we use conditional independence among variables to *simulate* a multiplicative structure.

THEOREM 4.2. *For all positive integers $n_1, n_2, \ldots, n_m$, and k,*

$$\Gamma_k(n_1, n_2, \ldots, n_m) \leqslant \frac{(2^k - 1)^m}{\sqrt{n_1 n_2 \cdots n_m}}.$$

Proof. We adopt the following notational shorthand: if X, Y are random variables with some joint distribution and $\Phi(X, Y)$ is a real function, we abbreviate $\mathbf{E}_X \mathbf{E}_Y[\Phi(X,Y)] \doteq \mathbf{E}_X \mathbf{E}_Y[\Phi(X,Y) \mid X]$. In other words, the inner expectation is always with respect to the conditional probability distribution induced by the outer variable.

The proof will proceed by induction on k, the base case $k = 1$ having already been established in Proposition 4.1. For the inductive step, fix $k \geqslant 2$ and consider a cylinder intersection $\chi\colon (\{0,1\}^{n_1+\cdots+n_m})^{k+1} \to \{0,1\}$ for which Γ_k is achieved:

$$\Gamma_k(n_1, n_2, \ldots, n_m) = \left| \mathop{\mathbf{E}}_{(x^1,Y^1,u^1),\ldots,(x^m,Y^m,u^m)} \left[\chi \cdot \prod_{i=1}^{m} D(x^i, Y^i, u^i) \right] \right|, \tag{4.1}$$

where $(x^i, (Y^i, u^i)) \sim \mathscr{U}_{n_i} \times \mu_{n_i,k}$ independently for each i, and the symbol χ is shorthand for $\chi(x^1, \ldots, x^m, Y^1, \ldots, Y^m, u^1, \ldots, u^m)$. Recall that one has the representation

$$\begin{aligned} &\chi(x^1, \ldots, x^m, Y^1, \ldots, Y^m, u^1, \ldots, u^m) \\ &\qquad = \chi_{u^1,\ldots,u^m}(x^1, \ldots, x^m, Y^1, \ldots, Y^m)\, \xi(x^1, \ldots, x^m, Y^1, \ldots, Y^m), \end{aligned} \tag{4.2}$$

where $\chi_{u^1,\ldots,u^m}\colon (\{0,1\}^{n_1+\cdots+n_m})^k \to \{0,1\}$ is a k-dimensional cylinder intersection for each $(u^1, \ldots, u^m)$, and ξ is some function into $\{0,1\}$. Rearranging the right member of (4.1) gives

$$\begin{aligned} \Gamma_k(n_1, \ldots, n_m) &= \left| \mathop{\mathbf{E}}_{(x^1,Y^1),\ldots,(x^m,Y^m)} \mathop{\mathbf{E}}_{u^1,\ldots,u^m} \left[\chi \cdot \prod_{i=1}^{m} D(x^i, Y^i, u^i) \right] \right| \\ &\leqslant \mathop{\mathbf{E}}_{(x^1,Y^1),\ldots,(x^m,Y^m)} \left| \mathop{\mathbf{E}}_{u^1,\ldots,u^m} \left[\chi_{u^1,\ldots,u^m} \cdot \prod_{i=1}^{m} D(x^i, Y^i, u^i) \right] \right|. \end{aligned}$$

We now apply the technique of Babai, Nisan, and Szegedy [6]. Squaring both sides and using the Cauchy-Schwarz inequality, one arrives at

$$\begin{aligned}\Gamma_k(n_1,\ldots,n_m)^2 &\leqslant \mathop{\mathbf{E}}_{(x^1,Y^1),\ldots,(x^m,Y^m)}\left[\mathop{\mathbf{E}}_{u^1,\ldots,u^m}\left[\chi_{u^1,\ldots,u^m}\cdot\prod_{i=1}^m D(x^i,Y^i,u^i)\right]^2\right]\\ &= \mathbf{E}\left[\chi'\cdot\prod_{i=1}^m D(x^i,Y^i,u^i)D(x^i,Y^i,v^i)\right], \qquad (4.3)\end{aligned}$$

where $\chi' \doteq \chi_{u^1,\ldots,u^m}\cdot\chi_{v^1,\ldots,v^m}$ and the expectation in (4.3) is taken with respect to $(x^i,(Y^i,u^i,v^i)) \sim \mathscr{U}_{n_i}\times\lambda_{n_i,k}$, independently for each i. It is clear from (4.2) that with $u^1,\ldots,u^m$ and $v^1,\ldots,v^m$ fixed, χ' is a cylinder intersection on $(\{0,1\}^{n_1+\cdots+n_m})^k$.

We now need to analyze (4.3). It is here that similarities with Babai, Nisan, and Szegedy [6] end, and we must exploit properties specific to set disjointness. To restate Lemma 3.1,

$$\begin{aligned}&D(x^i,Y^i,u^i)D(x^i,Y^i,v^i) \qquad (4.4)\\ &\quad= \mathbf{I}\left[D\left(Y^i|_{u^i\wedge v^i}\right)=1\right]\\ &\quad+\mathbf{I}\left[D\left(Y^i|_{u^i\wedge v^i}\right)=-1\right]D\left((x^i,Y^i)|_{u^i\wedge\overline{v}^i}\right)D\left((x^i,Y^i)|_{\overline{u}^i\wedge v^i}\right)\end{aligned}$$

on the support of $\lambda_{n_i,k}$. To bound (4.3), we will take advantage of conditioning. Specifically, using (4.4) and conditioning on $u_i, v_i, Y^i|_{u^i\wedge v^i}$, and $Y^i|_{\overline{u}^i\wedge\overline{v}^i}$ for each i, we arrive at the following expectation over the remaining variables x^i, $Y^i|_{u^i\wedge\overline{v}^i}$, and $Y^i|_{\overline{u}^i\wedge v^i}$:

$$\begin{aligned}&\mathbf{E}\left[\chi'\cdot\prod_{i=1}^m D(x^i,Y^i,u^i)D(x^i,Y^i,v^i)\right] \qquad (4.5)\\ &\quad=\sum_{z\in\{-1,+1\}^m}\mathbf{E}\left[\chi'\cdot\prod_{i:z_i=-1}D\left((x^i,Y^i)|_{u^i\wedge\overline{v}^i}\right)D\left((x^i,Y^i)|_{\overline{u}^i\wedge v^i}\right)\right]\\ &\qquad\times\prod_{i=1}^m\mathbf{I}\left[D\left(Y^i|_{u^i\wedge v^i}\right)=z_i\right].\end{aligned}$$

The expectations in the right member of (4.5) admit direct analysis. By Lemma 3.2, conditioning on any fixed value of $u^i, v^i, Y^i|_{u^i\wedge v^i}, Y^i|_{\overline{u}^i\wedge\overline{v}^i}$ with $D(Y^i|_{u^i\wedge v^i})=-1$ makes the remaining variables $Y^i|_{u^i\wedge\overline{v}^i}$ and $Y^i|_{\overline{u}^i\wedge v^i}$ independent and distributed according to $\mu_{|u^i\wedge\overline{v}^i|,k-1}$ and $\mu_{|\overline{u}^i\wedge v^i|,k-1}$, respectively. Since χ' is a cylinder intersection for fixed $u^1,\ldots,u^m$ and $v^1,\ldots,v^m$, the inductive hypothesis applies to the right member of (4.5), bounding it in absolute value by

$$\sum_{z\in\{-1,+1\}^m}\prod_{i:z_i=-1}\frac{(2^{k-1}-1)^2\,\mathbf{I}\left[D\left(Y^i|_{u^i\wedge v^i}\right)=-1\right]}{\sqrt{|u^i\wedge\overline{v}^i|\,|\overline{u}^i\wedge v^i|}}\cdot\prod_{i:z_i=1}\mathbf{I}\left[D\left(Y^i|_{u^i\wedge v^i}\right)=1\right].$$

Passing to expectations, one concludes that $\Gamma_k(n_1, n_2, \ldots, n_m)^2$ is bounded from above by

$$\sum_{z \in \{-1,+1\}^m} \prod_{i:z_i=-1} \mathop{\mathbf{E}}_{\lambda_{n_i,k}} \left[\frac{(2^{k-1}-1)^2 \, \mathbf{I}\left[D\left(Y^i|_{u^i \wedge v^i}\right) = -1\right]}{\sqrt{|u^i \wedge \overline{v}^i| \, |\overline{u}^i \wedge v^i|}} \right] \times \prod_{i:z_i=1} \mathop{\mathbf{P}}_{\lambda_{n_i,k}} \left[D\left(Y^i|_{u^i \wedge v^i}\right) = 1\right]$$

$$= \prod_{i=1}^{m} \left((2^{k-1}-1)^2 \mathop{\mathbf{E}}_{\lambda_{n_i,k}} \left[\frac{\mathbf{I}\left[D\left(Y^i|_{u^i \wedge v^i}\right) = -1\right]}{\sqrt{|u^i \wedge \overline{v}^i| \, |\overline{u}^i \wedge v^i|}} \right] + \mathop{\mathbf{P}}_{\lambda_{n_i,k}} \left[D\left(Y^i|_{u^i \wedge v^i}\right) = 1\right] \right).$$

The probabilities and expectations in the final expression are given by Lemmas 3.3 and 3.4, leading to $\Gamma_k(n_1, n_2, \ldots, n_m)^2 \leqslant (2^k - 1)^{2m}/(n_1 n_2 \cdots n_m)$ and thereby completing the inductive proof. □

Notes. We are only interested in $\Gamma_k(n_1, n_2, \ldots, n_m)$ for $n_1 = n_2 = \cdots = n_m$. However, the above inductive proof requires consideration of the more general quantity.

The base case, given by Proposition 4.1, could have been handled by a first-principles argument analogous to Theorem 4.2. However, we find the above treatment more concise and modular.

5. Randomized Communication

Combining the technical work of the previous sections with additional ideas, we will now derive a general lower bound on randomized communication complexity for composed functions (Section **5.1**). We will specifically be interested in compositions of the form $f \circ \mathrm{UDISJ}_{r,k}$. In Sections **5.2** and **5.3**, we will apply our findings to the bounded-error and small-bias communication complexity of AC^0 circuits, including set disjointness itself.

5.1. Master theorem. We start with the main technical result of the section. We present two different proofs for it, one based on the primal view of the problem and the other, on the dual view. The primal proof appears more intuitive to this author, whereas the dual proof is more versatile. Each of the proofs will be used in later sections to obtain additional results.

Theorem 5.1. *Let f be a (possibly partial) Boolean function on $\{-1,+1\}^n$. Consider the k-party communication problem $F = f \circ \mathrm{UDISJ}_{r,k}$. Then for $\epsilon, \delta \geqslant 0$,*

$$2^{R_\epsilon(F)} \geqslant (\delta - \epsilon(1+\delta)) \left(\frac{\deg_\delta(f)\sqrt{r}}{2^k e n} \right)^{\deg_\delta(f)}. \tag{5.1}$$

The idea of the primal proof is to convert a communication protocol for F into a low-degree polynomial approximating f in the infinity norm. The dual proof proceeds in the reverse direction and manipulates explicit witness objects, in the sense of Fact 2.2 and Theorem 2.10. More specifically, the dual proof converts a witness of f's inapproximability

by polynomials to a witness of F's high communication complexity. The primal point of view is original to this paper, whereas the dual approach is due to [42, 46].

Primal proof of Theorem 5.1. Define $\mu = \mathscr{U}_r \times \mu_{r,k-1}$, a probability distribution on the domain of $\mathrm{UDISJ}_{r,k}$. Let μ_{-1} and μ_{+1} stand for the probability distributions induced by μ on $\mathrm{UDISJ}_{r,k}^{-1}(-1)$ and $\mathrm{UDISJ}_{r,k}^{-1}(+1)$, respectively. Consider the following averaging operator M, which linearly sends real functions χ on $(\{0,1\}^{r\times k})^n$ to real functions on $\{-1,+1\}^n$:

$$(M\chi)(z) \doteq \mathop{\mathbf{E}}_{X_1\sim\mu_{z_1}} \cdots \mathop{\mathbf{E}}_{X_n\sim\mu_{z_n}} [\chi(X_1,\ldots,X_n)].$$

When χ is a k-dimensional cylinder intersection,

$$\begin{aligned}
|\widehat{M\chi}(S)| &= \left| \mathop{\mathbf{E}}_{z\in\{-1,+1\}^n} \mathop{\mathbf{E}}_{X_1\sim\mu_{z_1}} \cdots \mathop{\mathbf{E}}_{X_n\sim\mu_{z_n}} \left[\chi(X_1,\ldots,X_n) \prod_{i\in S} z_i \right] \right| \\
&= \left| \mathop{\mathbf{E}}_{X_1,\ldots,X_n\sim\mu} \left[\chi(X_1,\ldots,X_n) \prod_{i\in S} \mathrm{UDISJ}_{r,k}(X_i) \right] \right| \\
&\leqslant \Gamma_{k-1}(\overbrace{r,r,\ldots,r}^{|S|}) \\
&\leqslant \left(\frac{2^{k-1}}{\sqrt{r}}\right)^{|S|}, \qquad (5.2)
\end{aligned}$$

where the second equality uses the fact that $\mu = (\mu_{-1}+\mu_{+1})/2$, and the final step follows by Theorem 4.2.

Fix a randomized protocol for F with error ϵ and cost $c \doteq R_\epsilon(F)$. Approximate F as in Corollary 2.6 by a linear combination of cylinder intersections $\Pi = \sum_\chi a_\chi \chi$, where $\sum_\chi |a_\chi| \leqslant 2^c/(1-\epsilon)$. We claim that $M\Pi$ is approximable by a low-degree polynomial. Indeed, let d be a positive integer to be chosen later. Discarding the Fourier coefficients of $M\Pi$ of order d and higher gives

$$\begin{aligned}
E(M\Pi, d-1) &\leqslant \min\left\{ \frac{1}{1-\epsilon}, \sum_\chi |a_\chi| \sum_{|S|\geqslant d} |\widehat{M\chi}(S)| \right\} \\
&\leqslant \min\left\{ \frac{1}{1-\epsilon}, \frac{2^c}{1-\epsilon} \sum_{i=d}^{n} \binom{n}{i} \left(\frac{2^{k-1}}{\sqrt{r}}\right)^i \right\} \\
&\leqslant \frac{2^c}{1-\epsilon} \left(\frac{2^k e n}{d\sqrt{r}}\right)^d, \qquad (5.3)
\end{aligned}$$

where the second step uses (5.2). On the other hand, recall from Corollary 2.6 that Π approximates F in the sense that $\|\Pi\|_\infty \leqslant 1/(1-\epsilon)$ and $|F-\Pi| \leqslant \epsilon/(1-\epsilon)$ on the domain of F. It follows that $\|M\Pi\|_\infty \leqslant 1/(1-\epsilon)$ and $|f - M\Pi| \leqslant \epsilon/(1-\epsilon)$ on the domain of f, whence

$$E(f, d-1) \leqslant \frac{\epsilon}{1-\epsilon} + E(M\Pi, d-1).$$

Substituting the estimate from (5.3),

$$E(f, d-1) \leqslant \frac{\epsilon}{1-\epsilon} + \frac{2^c}{1-\epsilon}\left(\frac{2^k e n}{d\sqrt{r}}\right)^d. \qquad (5.4)$$

For $d = \deg_\delta(f)$, the left member of (5.4) must exceed δ, forcing the claimed lower bound on 2^c. □

We now present an alternate proof, which combines the argument in [42, 46] with the discrepancy result in this paper.

Dual proof of Theorem 5.1. As before, consider the distribution $\mu = \mathscr{U}_r \times \mu_{r,k-1}$ on the domain of $\mathrm{UDISJ}_{r,k}$. For $d \doteq \deg_\delta(f)$, Fact 2.2 provides $\psi\colon \{-1,+1\}^n \to \mathbb{R}$ with

$$\sum_{z \in \operatorname{dom} f} f(z)\psi(z) - \sum_{z \notin \operatorname{dom} f} |\psi(z)| > \delta, \tag{5.5}$$

$$\|\psi\|_1 = 1, \tag{5.6}$$

$$\hat{\psi}(S) = 0, \qquad |S| < d. \tag{5.7}$$

Define $\Psi\colon (\{0,1\}^{r\times k})^n \to \mathbb{R}$ by

$$\Psi(X_1,\ldots,X_n) = 2^n \psi\left(\mathrm{DISJ}_{r,k}(X_1),\ldots,\mathrm{DISJ}_{r,k}(X_n)\right) \prod_{i=1}^n \mu(X_i).$$

Since μ places equal weight on $\mathrm{UDISJ}_{r,k}^{-1}(-1)$ and $\mathrm{UDISJ}_{r,k}^{-1}(+1)$, we have

$$\|\Psi\|_1 = 2^n \mathop{\mathbf{E}}_{z \in \{-1,+1\}^n} [|\psi(z)|] = 1 \tag{5.8}$$

and analogously

$$\begin{aligned} \sum_{\operatorname{dom} F} F(X_1,\ldots,X_n)\Psi(X_1,\ldots,X_n) - \sum_{\overline{\operatorname{dom} F}} |\Psi(X_1,\ldots,X_n)| &= \sum_{z \in \operatorname{dom} f} f(z)\psi(z) - \sum_{z \notin \operatorname{dom} f} |\psi(z)| \\ &> \delta, \end{aligned} \tag{5.9}$$

where the final step in the two derivations uses (5.5) and (5.6). It remains to bound the inner product of Ψ with a k-dimensional cylinder intersection χ. By (5.7),

$$\begin{aligned} |\langle \Psi, \chi \rangle| &\leqslant 2^n \sum_{|S| \geqslant d} |\hat{\psi}(S)| \left| \mathop{\mathbf{E}}_{X_1,\ldots,X_n \sim \mu} \left[\chi(X_1,\ldots,X_n) \prod_{i \in S} \mathrm{DISJ}_{r,k}(X_i) \right] \right| \\ &\leqslant 2^n \sum_{|S| \geqslant d} |\hat{\psi}(S)| \, \Gamma_{k-1}(\overbrace{r, r, \ldots, r}^{|S|}) \\ &\leqslant \sum_{|S| \geqslant d} \left(\frac{2^{k-1}}{\sqrt{r}} \right)^{|S|}, \end{aligned}$$

where the final step uses Proposition 2.1 and Theorem 4.2. Combining this with the trivial bound $|\langle \Psi, \chi \rangle| \leqslant \|\Psi\|_1 \|\chi\|_\infty = 1$,

$$|\langle \Psi, \chi \rangle| \leqslant \min\left\{ 1, \sum_{i=d}^n \binom{n}{i} \left(\frac{2^{k-1}}{\sqrt{r}} \right)^i \right\} \leqslant \left(\frac{2^k e n}{d\sqrt{r}} \right)^d. \tag{5.10}$$

By (5.8)–(5.10) and Theorem 2.10, the proof is complete. □

5.2. Bounded-error communication. The general theorem that we have just proved allows one to obtain lower bounds on bounded-error communication in terms of the $1/3$-approximate degree, as follows.

THEOREM 5.2. *Let f be a (possibly partial) Boolean function on $\{-1,+1\}^n$. Let $d = \deg_{1/3}(f)$. Then*

$$R_{1/3}\left(f \circ \mathrm{UDISJ}_{4^{k+2}\lceil \frac{n}{d}\rceil^2,k}\right) \geqslant \Omega(d).$$

Proof. Let $\epsilon = 1/5,\ \delta = 1/3,\ r = 4^{k+2}\lceil n/d\rceil^2$ in Theorem 5.1. □

In particular, we obtain the following lower bound on the randomized communication complexity of set disjointness.

THEOREM 1.1 (restated). *The k-party set disjointness problem obeys*

$$R_{1/3}(\mathrm{DISJ}_{n,k}) \geqslant R_{1/3}(\mathrm{UDISJ}_{n,k}) = \Omega\Big(\frac{n}{4^k}\Big)^{1/4}.$$

Proof. Recall that $\mathrm{UDISJ}_{nr,k} = \widetilde{\mathrm{AND}}_n \circ \mathrm{UDISJ}_{r,k}$ for all integers n, r. Theorem 2.4 shows that $\deg_{1/3}(\widetilde{\mathrm{AND}}_n) > \delta\sqrt{n}$ for some constant $\delta > 0$. Thus, taking $f = \widetilde{\mathrm{AND}}_n$ and $d = \delta\sqrt{n}$ in Theorem 5.2 gives $R_{1/3}(\mathrm{UDISJ}_{4^{k+2}n\lceil\sqrt{n}/\delta\rceil^2,k}) = \Omega(\sqrt{n})$, which is equivalent to the claimed bound. □

Theorem 5.2 gives a general lower bound on bounded-error communication complexity for compositions $f \circ G$, where G is a gadget on a relatively large number of variables. We will now derive an alternate lower bound, in which the gadget G is essentially as simple as possible and in particular depends on only $2k$ variables.

We recall some combinatorial complexity measures. Let $f\colon \{-1,+1\}^n \to \{-1,+1\}$ be given. For a string $x \in \{-1,+1\}^n$ and a subset $S \subseteq \{1,2,\ldots,n\}$, let x^S stand for the string obtained from x by negating the bit positions in S, i.e.,

$$(x^S)_i = \begin{cases} -x_i & \text{if } i \in S, \\ x_i & \text{otherwise.} \end{cases}$$

The *block sensitivity of f*, denoted $\mathrm{bs}(f)$, is the maximum number of nonempty, pairwise disjoint subsets $S_1, S_2, S_3, \ldots \subseteq \{1,2,\ldots,n\}$ such that $f(x) \neq f(x^{S_1}) = f(x^{S_2}) = f(x^{S_3}) = \cdots$ for some string $x \in \{-1,+1\}^n$. The *sensitivity of f*, denoted $\mathrm{s}(f)$, is defined analogously with the additional requirement that $S_1, S_2, S_3, \ldots$ contain exactly one element each. In other words, the sensitivity of f is the maximum of $|\{i : f(x) \neq f(x_1,\ldots,x_{i-1},-x_i,x_{i+1},\ldots,x_n)\}|$ over all strings $x \in \{-1,+1\}^n$. The *decision tree complexity of f*, denoted $\mathrm{dt}(f)$, is the minimum depth of a decision tree for f. Surprisingly, the quantities $\mathrm{bs}(f)$, $\mathrm{dt}(f)$, $\deg_{1/3}(f)$ are polynomially related for every total Boolean function [38]. Finally, a *contraction* of a Boolean function $f\colon \{-1,+1\}^n \to \{-1,+1\}$ is any function $g\colon \{-1,+1\}^n \to \{-1,+1\}$ such that $g(x) \equiv f(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$ for some indices $i_1, i_2, \ldots, i_n \in \{1,2,\ldots,n\}$. Informally, a contraction is the result of replacing groups of variables by a single variable (and possibly permuting the variables).

We are now in a position to prove the alternate lower bound on communication for composed functions. We will study compositions of the form $f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$, where $\mathrm{OR}_k \vee \mathrm{AND}_k$ refers to the function $(x_1,\ldots,x_{2k}) \mapsto x_1 \vee \cdots \vee x_k \vee (x_{k+1} \wedge \cdots \wedge x_{2k})$. It is clear that any such composition has a deterministic k-party communication protocol

with cost $3\,\mathrm{dt}(f)$. In what follows, we prove that this trivial upper bound is close to tight, even for randomized protocols.

THEOREM 5.3. *Let* $f\colon\{-1,+1\}^n \to \{-1,+1\}$ *be given,* $\deg_{1/3}(f) = d$. *Then*

$$\begin{aligned} R_{1/3}(f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)) &\geqslant \Omega\left(\frac{\mathrm{bs}(f)}{4^k}\right)^{1/4} \\ &\geqslant \Omega\left(\frac{\mathrm{dt}(f)^{1/3}}{4^k}\right)^{1/4} \\ &\geqslant \Omega\left(\frac{d^{1/3}}{4^k}\right)^{1/4}. \end{aligned}$$

By symmetry, the theorem also holds with $\mathrm{OR}_k \wedge \mathrm{AND}_k$ in place of $\mathrm{OR}_k \vee \mathrm{AND}_k$. Finally, note that the above result implies a lower bound of $\Omega(n/4^k)^{1/4}$ on the randomized communication complexity of set disjointness. However, this does not give a *new* proof of Theorem 1.1 because Theorem 1.1 is actually used in a crucial way to prove Theorem 5.3.

Proof of Theorem 5.3. It is clear that $d \leqslant \mathrm{dt}(f)$, and it is known [8, p. 791] that $\mathrm{dt}(f) \leqslant \mathrm{bs}(f)^3$. Thus, it suffices to prove the lower bound in terms of $\mathrm{bs}(f)$. We will actually prove the following stronger result: for some fixed $z \in \{-1,+1\}^n$,

$$R_{1/3}(f_z \circ \mathrm{AND}_k) \geqslant \Omega\left(\frac{\mathrm{bs}(f)}{4^k}\right)^{1/4},$$

where $f_z\colon\{-1,+1\}^n \to \{-1,+1\}$ is defined by $f_z(x) = f(z \oplus x)$.

Choose z such that $f(z) \neq f(z^{S_1}) = f(z^{S_2}) = \cdots = f(z^{S_{\mathrm{bs}(f)}})$ for some nonempty, pairwise disjoint subsets $S_1, S_2, \ldots, S_{\mathrm{bs}(f)} \subseteq \{1,2,\ldots,n\}$. This means that $f_z(\mathbf{1}) \neq f_z(\mathbf{1}^{S_1}) = f_z(\mathbf{1}^{S_2}) = \cdots = f_z(\mathbf{1}^{S_{\mathrm{bs}(f)}})$, where $\mathbf{1} = (1,1,\ldots,1)$. But then there is a contraction g of f_z such that

$$g(1,1,\ldots,1) \neq g(\underbrace{1,\ldots,1,-1}_{i},1,\ldots,1)$$

for $i = 1,2,\ldots,\mathrm{bs}(f)$. Indeed, such a contraction can be obtained from f_z by replacing the variables in each block S_i by a single variable, and suitably permuting the resulting variable set. In other words, $\widetilde{\mathrm{OR}}_{\mathrm{bs}(f)}$ is a subfunction of g. Therefore, $\widetilde{\mathrm{OR}}_{\mathrm{bs}(f)} \circ \mathrm{AND}_k = \neg\mathrm{UDISJ}_{\mathrm{bs}(f),k}$ is a subfunction of $g \circ \mathrm{AND}_k$, and $R_{1/3}(g \circ \mathrm{AND}_k) = \Omega(\mathrm{bs}(f)/4^k)^{1/4}$ by Theorem 1.1. The same lower bound holds for $f_z \circ \mathrm{AND}_k$ since passage to a contraction cannot increase communication complexity. □

It is tempting to go further and try to bound $R_{1/3}(f \circ \mathrm{AND}_k)$ from below in terms of the approximate degree of f. Unfortunately, the gap between $R_{1/3}(f \circ \mathrm{AND}_k)$ and $\deg_{1/3}(f)$ can be as large as $\Theta(1)$ versus $\Theta(\sqrt{n})$; simply take $f = \mathrm{AND}_n$. As it turns out, the right approach is to consider the *maximum* of the communication complexities of $f \circ \mathrm{AND}_k$ and $f \circ \mathrm{OR}_k$.

THEOREM 5.4. *Let $f:\{-1,+1\}^n \to \{-1,+1\}$ be given,* $\deg_{1/3}(f) = d$. *Then*

$$\max\{R_{1/3}(f \circ \mathrm{OR}_k),\ R_{1/3}(f \circ \mathrm{AND}_k)\} \geqslant \Omega\left(\frac{\mathrm{bs}(f)^{1/4}}{2^k}\right)^{1/2} \geqslant \Omega\left(\frac{\mathrm{dt}(f)^{1/12}}{2^k}\right)^{1/2} \geqslant \Omega\left(\frac{d^{1/12}}{2^k}\right)^{1/2}. \tag{5.11}$$

Proof. As in the previous proof, it suffices to prove the first inequality, and moreover we may replace f in the left member of (5.11) by a contraction of f. The argument is closely analogous to the one in [44] for two-party communication. Specifically, by [44, Lem. 3.3], there is a contraction g of f such that $\mathrm{s}(g) \geqslant \alpha\sqrt{\mathrm{bs}(f)}$ for some absolute constant $\alpha > 0$. So, fix a subset $S \subseteq \{1,2,\ldots,n\}$ of size $|S| \geqslant \alpha\sqrt{\mathrm{bs}(f)}$ and a string $z \in \{-1,+1\}^n$ such that $g(z) \neq g(z_1,\ldots,z_{i-1},-z_i,z_{i+1},\ldots,z_n)$ for all $i \in S$. We consider two cases.

(i) If $z|_S$ has more "-1" entries than "$+1$" entries, then $\widetilde{\mathrm{AND}}_{|S|/2}$ is a subfunction of g. As a result, $\mathrm{UDISJ}_{|S|/2,k}$ is a subfunction of $g \circ \mathrm{OR}_k$ (up to negations of the input variables), and the proof is complete in view of Theorem 1.1.

(ii) If $z|_S$ has no more "-1" entries than "$+1$" entries, then $\widetilde{\mathrm{OR}}_{|S|/2}$ is a subfunction of g. As a result, $\neg\mathrm{UDISJ}_{|S|/2,k}$ is a subfunction of $g \circ \mathrm{AND}_k$, and the proof is again complete in view of Theorem 1.1. □

5.3. Small-bias communication and discrepancy. The counterpart to bounded-error communication is small-bias communication, when the protocol is only required to produce the correct output with probability vanishingly close to $1/2$. Theorem 5.1 gives communication lower bounds in this setting as well, in terms of the approximate degree with an appropriate error parameter.

THEOREM 5.5. *Let f be a (possibly partial) Boolean function on $\{-1,+1\}^n$. Then*

$$R_{\frac{1}{2}-\frac{\epsilon}{2}}\left(f \circ \mathrm{UDISJ}_{4^{k+3}\lceil n/\deg_{1-\gamma}(f)\rceil^2,k}\right) \geqslant \deg_{1-\gamma}(f) - \log\frac{1}{\epsilon-\gamma},$$
$$R_{\frac{1}{2}-\frac{\epsilon}{2}}\left(f \circ \mathrm{UDISJ}_{4^{k+3}\lceil n/\deg_{\pm}(f)\rceil^2,k}\right) \geqslant \deg_{\pm}(f) - \log\frac{1}{\epsilon}.$$

Proof. The first lower bound follows by taking $\delta = 1-\gamma$ and $r = 4^{k+3}\lceil n/\deg_{1-\gamma}(f)\rceil^2$ in Theorem 5.1. The second lower bound follows from the first by letting $\gamma \searrow 0$. □

Finally, Theorem 5.1 allows one to directly prove upper bounds on discrepancy, a complexity measure of interest in its own right. We have:

THEOREM 5.6. *Let f be a (possibly partial) Boolean function on $\{-1,+1\}^n$. Then for every $\gamma > 0$,*

$$\mathrm{disc}(f \circ \mathrm{UDISJ}_{r,k}) < \left(\frac{2^k e n}{\deg_{1-\gamma}(f)\sqrt{r}}\right)^{\deg_{1-\gamma}(f)} + \gamma.$$

In particular,

$$\operatorname{disc}(f \circ \mathrm{UDISJ}_{r,k}) \leqslant \left(\frac{2^k e n}{\deg_\pm(f)\sqrt{r}}\right)^{\deg_\pm(f)}.$$

Proof. The second bound follows from the first by letting $\gamma \searrow 0$. To prove the first bound, take $\delta = 1-\gamma$, $d = \deg_\delta(f)$, and define $\Psi\colon (\{0,1\}^{r\times k})^n \to \mathbb{R}$ as in the dual proof of Theorem 5.1. Then (5.8) shows that $\Psi = H \circ P$, where H is a sign tensor and P a probability distribution. Letting $F = f \circ \mathrm{UDISJ}_{r,k}$, we can restate (5.9) as

$$\sum_{\operatorname{dom} F} F(x)H(x)P(x) - P(\overline{\operatorname{dom} F}) > 1-\gamma. \tag{5.12}$$

For every cylinder intersection χ,

$$\begin{aligned}
&\left|\sum_{\operatorname{dom} F} F(x)P(x)\chi(x)\right| \\
&\qquad = \left|\langle H\circ P, \chi\rangle + \sum_{\operatorname{dom} F}(F(x)-H(x))P(x)\chi(x) - \sum_{\overline{\operatorname{dom} F}} H(x)P(x)\chi(x)\right| \\
&\qquad \leqslant \operatorname{disc}_P(H) + \sum_{\operatorname{dom} F}|F(x)-H(x)|P(x) + P(\overline{\operatorname{dom} F}) \\
&\qquad = \operatorname{disc}_P(H) + P(\operatorname{dom} F) - \sum_{\operatorname{dom} F} F(x)H(x)P(x) + P(\overline{\operatorname{dom} F}) \\
&\qquad < \operatorname{disc}_P(H) + P(\operatorname{dom} F) - 1 + \gamma,
\end{aligned} \tag{5.13}$$

where the last step uses (5.12). Maximizing over all cylinder intersections χ,

$$\begin{aligned}
\operatorname{disc}_P(F) &= \max_\chi \left|\sum_{\operatorname{dom} F} F(x)P(x)\chi(x)\right| + P(\overline{\operatorname{dom} F}) \\
&< \operatorname{disc}_P(H) + \gamma \\
&\leqslant \left(\frac{2^k e n}{d\sqrt{r}}\right)^d + \gamma,
\end{aligned}$$

where the second step uses (5.13) and the third uses (5.10). ∎

As an application of the above results on small-bias computation, we exhibit a particularly hard multiparty communication problem $F \in \mathsf{AC}^0$. This k-party communication problem is given by an $\{\wedge,\vee\}$-circuit of size kn and depth 3 and has exponentially small discrepancy: $\operatorname{disc}(F) \leqslant \exp(-\Omega(n/4^k)^{1/7})$. In particular, the communication complexity of F remains high even to achieve an exponentially small advantage over random guessing. A more detailed statement follows.

THEOREM 5.7. *Consider the k-party communication problem $F_{n,k}\colon \left(\{0,1\}^{4^k n^7}\right)^k \to \{-1,+1\}$ given by*

$$F_{n,k}(x) = \bigvee_{i=1}^{n} \bigwedge_{j=1}^{4^k n^6} (x_{i,j,1} \vee x_{i,j,2} \vee \cdots \vee x_{i,j,k}).$$

Then

$$\operatorname{disc}(F_{4n,k}) \leqslant 2^{-n},$$
$$R_{\frac{1}{2}-\frac{\gamma}{2}}(F_{4n,k}) \geqslant n - \log\frac{1}{\gamma} \qquad (\gamma > 0).$$

This construction achieves optimal circuit depth because AC^0 circuits of depth less than 3 have multiparty discrepancy $1/n^{O(1)}$, regardless of how the bits are assigned to the parties. The previous best construction, due to Beame and Huynh-Ngoc [11], was a depth-6 circuit of size kn with discrepancy $\exp(-\Omega(n/2^{31k})^{1/29})$.

Proof of Theorem 5.7. Let MP_n be given by Theorem 2.3, so that $\deg_{\pm}(\mathrm{MP}_n) = n$. Since $\mathrm{MP}_n \circ \mathrm{DISJ}_{4^{k+5}n^4,k}$ is a subfunction of $F_{4n,k}$ (up to negations of the input variables), Theorem 5.6 yields the discrepancy bound. The communication lower bound follows by Theorem 2.9. □

Theorem 5.7 settles Theorem 1.5 from the introduction.

6. XOR Lemmas and Direct Product Theorems

In Section 5, we proved that $\Omega(n/4^k)^{1/4}$ bits of communication are required to solve the set disjointness problem with probability of correctness $2/3$. In this section, we consider the task of simultaneously solving ℓ instances of set disjointness and prove that $\ell \cdot \Omega(n/4^k)^{1/4}$ bits of communication are necessary to even achieve advantage $2^{-\Omega(\ell)}$ over random guessing. We prove an analogous result for computing the XOR of ℓ instances. The theorems in this section hold in somewhat greater generality, applying to compositions $f \circ G$ where f is an arbitrary function and G is an instance of set disjointness on a suitable number of variables. Our proof works by reducing these communication statements to analogous statements about polynomial approximation. We then appeal to known direct product theorems and XOR lemmas for the latter setting, which were recently obtained in [45].

6.1. XOR lemmas. We start by proving the XOR lemma for set disjointness, which happens to admit a more intuitive and direct analysis than the corresponding direct product theorem. We recall an analogous XOR lemma for polynomial approximation [45, Cor. 5.2].

Theorem 6.1 (Sherstov). *Let f be a (possibly partial) Boolean function on $\{-1,+1\}^n$. Then for some absolute constant $c > 0$ and every ℓ,*

$$\deg_{1-2^{-\ell-1}}(\underbrace{f \otimes \cdots \otimes f}_{\ell}) \geqslant c\ell \deg_{1/3}(f).$$

Using the small-bias version of the master theorem (Theorem 5.5), we are able to immediately translate this result to communication.

Theorem 6.2 (XOR lemma). *Let f be a (possibly partial) Boolean function on $\{-1,+1\}^n$. Define $d = \deg_{1/3}(f)$. Then for some absolute constant $C > 1$, the k-party communication problem*

$$F = f \circ \mathrm{UDISJ}_{4^k\lceil \frac{Cn}{d} \rceil^2,k}$$

obeys

$$R_{\frac{1}{2}-(\frac{1}{2})^{\ell+1}}(\underbrace{F \otimes \cdots \otimes F}_{\ell}) \geqslant \ell \cdot \Omega(d).$$

Proof. Define $g = f^{\otimes \ell}$. Theorem 6.1 provides an absolute constant $c > 0$ such that $\deg_{1-1/2^{\ell+1}}(g) \geqslant c\ell d$. Letting $C = 8/c$, Theorem 5.5 implies that the composition $g \circ \mathrm{UDISJ}_{4^k \lceil \frac{Cn}{d} \rceil^2, k} = F^{\otimes \ell}$ obeys $R_{1/2-1/2^{\ell+1}}(F^{\otimes \ell}) = \ell \cdot \Omega(d)$. □

The desired XOR lemma for set disjointness, stated as Theorem 1.2(i) in the introduction, now falls out as a corollary.

COROLLARY 6.3. *For every* ℓ,

$$R_{\frac{1}{2}-(\frac{1}{2})^{\ell+1}}(\underbrace{\mathrm{UDISJ}_{n,k} \otimes \cdots \otimes \mathrm{UDISJ}_{n,k}}_{\ell}) \geqslant \ell \cdot \Omega\Big(\frac{n}{4^k}\Big)^{1/4}.$$

Proof. Theorem 2.4 shows that $\deg_{1/3}(\widetilde{\mathrm{AND}}_n) \geqslant \epsilon\sqrt{n}$ for a constant $\epsilon > 0$. Thus, letting $f = \widetilde{\mathrm{AND}}_n$ and $d = \epsilon\sqrt{n}$ in Theorem 6.2 gives $R_{1/2-1/2^{\ell+1}}\big(\mathrm{UDISJ}_{4^k n \lceil C\sqrt{n}/\epsilon \rceil^2, k}{}^{\otimes \ell}\big) \geqslant \ell \cdot \Omega(\sqrt{n})$, which is equivalent to the claimed bound. □

Using the argument of Theorem 5.3, we are now able to give an XOR lemma for arbitrary compositions of the form $f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$. For this, we will use the combinatorial complexity measures $\mathrm{bs}(f)$ and $\mathrm{dt}(f)$, defined in Section 5.

THEOREM 6.4. *Let* $f\colon \{-1,+1\}^n \to \{-1,+1\}$ *be given. Put* $F = f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$. *Then for every* ℓ,

$$\begin{aligned} R_{\frac{1}{2}-(\frac{1}{2})^{\ell+1}}(\overbrace{F \otimes F \otimes \cdots \otimes F}^{\ell}) &\geqslant \ell \cdot \Omega\left(\frac{\mathrm{bs}(f)}{4^k}\right)^{1/4} \\ &\geqslant \ell \cdot \Omega\left(\frac{\mathrm{dt}(f)^{1/3}}{4^k}\right)^{1/4} \\ &\geqslant \ell \cdot \Omega\left(\frac{\deg_{1/3}(f)^{1/3}}{4^k}\right)^{1/4}. \end{aligned}$$

Recall from Section 5 that F has a *deterministic* protocol with cost $3\,\mathrm{dt}(f)$, and thus $F^{\otimes \ell}$ has a deterministic protocol with cost $3\ell\,\mathrm{dt}(f)$. In other words, Theorem 6.4 is reasonably close to tight.

Proof of Theorem 6.4. The argument is essentially identical to that of Theorem 5.3. As argued there, any communication protocol for $f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$ also solves $\mathrm{UDISJ}_{\mathrm{bs}(f),k}$, so that the first inequality is immediate from Corollary 6.3. The other two inequalities follow from general relationships among $\mathrm{bs}(f)$, $\mathrm{dt}(f)$, and $\deg_{1/3}(f)$; see the proof of Theorem 5.3. □

6.2. Direct product theorems. Let $F\colon X_1 \times X_2 \times \cdots \times X_k \to \{-1,+1\}$ be a given k-party communication problem. We are interested here in the communication complexity of simultaneously solving ℓ instances of F. More formally, the communication protocol now

receives ℓ inputs $x^1, \ldots, x^\ell \in X_1 \times X_2 \times \cdots \times X_k$ and outputs a string $\{-1,+1\}^\ell$, representing a guess at $(F(x^1), \ldots, F(x^\ell))$. As usual, an ϵ-error protocol is one whose output differs from the correct answer with probability no greater than ϵ, on any given input. The least cost of such a protocol for solving ℓ instances of F is denoted $R_\epsilon(F, F, \ldots, F)$, where the number of instances will always be specified with an underbrace.

It is also meaningful to consider communication protocols that solve all but m of the ℓ instances (m for "mistake"), where the ratio m/ℓ is a small constant. In other words, given ℓ input instances $x^1, \ldots, x^\ell$, the protocol is required to output, with probability at least $1-\epsilon$, a vector $z \in \{-1,+1\}^\ell$ such that $z_i = F(x^i)$ for at least $\ell - m$ indices i. We let

$$R_{\epsilon,m}(\underbrace{F, F, \ldots, F}_{\ell})$$

stand for the least cost of such a protocol. When referring to this formalism, we will write that a protocol "solves with probability $1-\epsilon$ at least $\ell - m$ of the ℓ instances." This setting corresponds to *threshold direct product theorems*, as opposed to the more restricted notion of *strong direct product theorems* for which $m = 0$. All of our results belong to the former category. The following definition from [45] analytically formalizes the simultaneous solution of ℓ instances.

DEFINITION 6.5 (Approximants). Let f be a (possibly partial) Boolean function on a finite set X. A (σ, m, ℓ)-*approximant* for f is any system $\{\phi_z\}$ of functions $\phi_z \colon X^\ell \to \mathbb{R}$, $z \in \{-1,+1\}^\ell$, such that for all $x^1, \ldots, x^\ell \in X$,

$$\sum_{z \in \{-1,+1\}^\ell} |\phi_z(x^1, \ldots, x^\ell)| \leqslant 1, \qquad x^1, \ldots, x^\ell \in X,$$

$$\sum_{\substack{z \in \{-1,+1\}^\ell \\ |\{i : z_i = -1\}| \leqslant m}} \phi_{(z_1 f(x^1), \ldots, z_\ell f(x^\ell))}(x^1, \ldots, x^\ell) \geqslant \sigma, \qquad x^1, \ldots, x^\ell \in \operatorname{dom} f.$$

The following result [45, Cor. 5.7] on polynomial approximation is naturally regarded as a threshold direct product theorem in that model of computation.

THEOREM 6.6 (Sherstov). *Let f be a (possibly partial) Boolean function on $\{-1,+1\}^n$. Let $\beta > 0$ be a small enough absolute constant. Then every $(2^{-\beta\ell}, \beta\ell, \ell)$-approximant $\{\phi_z\}$ for f obeys*

$$\max_{z \in \{-1,+1\}^\ell} \{\deg \phi_z\} \geqslant \beta \ell \deg_{1/3}(f).$$

Using the technique of Theorem 5.1, we are able to translate this result to multiparty communication complexity.

THEOREM 6.7 (Direct product theorem). *Let $\alpha > 0$ be a sufficiently small absolute constant. Let f be a (possibly partial) Boolean function on $\{-1,+1\}^n$ with approximate degree $d = \deg_{1/3}(f)$. Then the k-party communication problem*

$$F = f \circ \mathrm{UDISJ}_{4^k \lceil \frac{n}{\alpha d} \rceil^2, k}$$

obeys

$$R_{1-2^{-\alpha\ell}, \alpha\ell}(\underbrace{F, F, \ldots, F}_{\ell}) \geqslant \alpha \ell d.$$

Proof. Our proof strategy will be to convert a low-cost communication protocol for solving ℓ instances of F into a low-degree approximant for f, in the sense of Definition 6.5. Such an approximant would be in contradiction to Theorem 6.6, thus ruling out the assumed low-cost protocol. The critical part of the proof is the passage from communication protocols to polynomials, to which end we will mimic the primal proof of Theorem 5.1.

Abbreviate $r = 4^k \left\lceil \frac{n}{\alpha d} \right\rceil^2$ and let $\mu = \mathscr{U}_r \times \mu_{r,k-1}$, a probability distribution on the domain of $\mathrm{UDISJ}_{r,k}$. Let μ_{-1} and μ_{+1} stand for the probability distributions induced by μ on $\mathrm{UDISJ}_{r,k}^{-1}(-1)$ and $\mathrm{UDISJ}_{r,k}^{-1}(+1)$, respectively. Consider the following averaging operator M, which linearly sends real functions ϕ on $(\{0,1\}^{r\times k})^{\ell n}$ to real functions on $\{-1,+1\}^{\ell n}$:

$$(M\phi)(z) \doteq \mathop{\mathbf{E}}_{X_{1,1}\sim\mu_{z_{1,1}}} \cdots \mathop{\mathbf{E}}_{X_{\ell,n}\sim\mu_{z_{\ell,n}}} \big[\phi(X_{1,1},\ldots,X_{\ell,n})\big].$$

Now fix a cost-c randomized protocol Π which solves, with probability $2^{-\alpha\ell}$, at least $(1-\alpha)\ell$ from among ℓ instances of F. We will take $\alpha = \alpha(\beta) > 0$ small enough, where $\beta > 0$ is the constant from Theorem 6.6. Starting with the assumption that $c < \alpha\ell d$, we will arrive at a contradiction.

For $z \in \{-1,+1\}^\ell$, consider the protocol Π_z with Boolean output which on input from $(\{0,1\}^{r\times k})^{\ell n}$ runs Π and outputs -1 if and only if Π outputs z. Let $\phi_z\colon (\{0,1\}^{r\times k})^{\ell n} \to [0,1]$ be the acceptance probability function for Π_z. Then $\phi_z = \sum a_\chi \chi$ by Corollary 2.7, where the sum is over k-dimensional cylinder intersections and $\sum |a_\chi| \leqslant 2^c$. By the argument in the primal proof of Theorem 5.1, for every positive integer D,

$$E(M\phi_z, D-1) \leqslant 2^c \left(\frac{2^k \mathrm{e}\ell n}{D\sqrt{r}}\right)^D. \tag{6.1}$$

Observe that $\{\phi_z\}$ is a $(2^{-\alpha\ell}, \alpha\ell, \ell)$-approximant for F, and analogously $\{M\phi_z\}$ is a $(2^{-\alpha\ell}, \alpha\ell, \ell)$-approximant for f. By (6.1), each $M\phi_z$ can in turn be approximated by a polynomial of degree less than $\beta\ell d$ to within $2^{\alpha\ell d}(\alpha\mathrm{e}/\beta)^{\beta\ell d}$. Taking $\alpha = \alpha(\beta) > 0$ small enough, we arrive at a $(2^{-\beta\ell}, \beta\ell, \ell)$-approximant for f of degree less than $\beta\ell d$, in contradiction to Theorem 6.6. Hence, $c \geqslant \alpha\ell d$. □

As a corollary, we obtain a direct product result for set disjointness, originally stated as Theorem 1.2(ii) in the introduction.

COROLLARY 6.8. *For some absolute constant $\alpha > 0$ and every ℓ,*

$$R_{1-2^{-\alpha\ell},\alpha\ell}\underbrace{(\mathrm{UDISJ}_{n,k},\ldots,\mathrm{UDISJ}_{n,k})}_{\ell} \geqslant \ell \cdot \Omega\Big(\frac{n}{4^k}\Big)^{1/4}.$$

Proof. Theorem 2.4 shows that $\deg_{1/3}(\widetilde{\mathrm{AND}}_n) \geqslant \epsilon\sqrt{n}$ for a constant $\epsilon > 0$. As a result, taking $f = \widetilde{\mathrm{AND}}_n$ and $d = \epsilon\sqrt{n}$ in Theorem 6.7 gives

$$R_{1-2^{-\alpha\ell},\alpha\ell}\underbrace{(\ldots,\mathrm{UDISJ}_{4^k n\lceil\sqrt{n}/\alpha\epsilon\rceil^2,k},\ldots)}_{\ell} = \ell\cdot\Omega(\sqrt{n}),$$

which is equivalent to the claimed bound. □

Again, the above corollary readily generalizes to arbitrary compositions of the form $F = f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$.

THEOREM 6.9. *Let $f:\{-1,+1\}^n \to \{-1,+1\}$ be given. Put $F = f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$. Then for some absolute constant $\alpha > 0$ and every ℓ,*

$$R_{1-2^{-\alpha\ell},\alpha\ell}(\overbrace{F, F, \ldots, F}^{\ell}) \geqslant \ell \cdot \Omega\left(\frac{\mathrm{bs}(f)}{4^k}\right)^{1/4} \geqslant \ell \cdot \Omega\left(\frac{\mathrm{dt}(f)^{1/3}}{4^k}\right)^{1/4} \geqslant \ell \cdot \Omega\left(\frac{\deg_{1/3}(f)^{1/3}}{4^k}\right)^{1/4}.$$

Proof. Identical to Theorem 6.4, with Corollary 6.8 invoked in place of Corollary 6.3. □

We have focused here on XOR lemmas and direct product theorems for ℓ instances of the *same* communication problem. The results and proofs above generalize easily to ℓ distinct communication problems, by invoking, in place of Theorems 6.1 and 6.6, correspondingly more general results from [45] on polynomial approximation.

7. NONDETERMINISTIC AND MERLIN-ARTHUR COMMUNICATION

In this section, we study the communication complexity of set disjointness in the nondeterministic and Merlin-Arthur multiparty models. We will see that the lower bound of Theorem 1.1 carries over. We will reinterpret our findings in terms of communication complexity classes.

7.1. Definitions. We start by describing the *nondeterministic* k-party model of communication, which is similar in some ways and different in others from the randomized model. As in the randomized model, one considers a function $F: X_1 \times X_2 \times \cdots \times X_k \to \{-1,+1\}$ for some finite sets $X_1, X_2, \ldots, X_k$. An input $(x_1, x_2, \ldots, x_k) \in X_1 \times X_2 \times \cdots \times X_k$ is distributed among the k parties as before, giving the ith party all the arguments except x_i. Beyond this setup, nondeterministic computation proceeds as follows. At the start of the protocol, c_1 bits appear on the shared blackboard. Given the values of those bits, the parties execute an agreed-upon deterministic protocol with communication cost at most c_2. A nondeterministic protocol for F is required to output the correct answer for *at least one* nondeterministic choice of the c_1 bits when $F(x_1, x_2, \ldots, x_k) = -1$ and for *all* possible choices when $F(x_1, x_2, \ldots, x_k) = +1$. The *cost* of a nondeterministic protocol is defined as $c_1 + c_2$. The *nondeterministic communication complexity* of F, denoted $N(F)$, is the least cost of a nondeterministic protocol for F. The *co-nondeterministic communication complexity* of F is the quantity $N(-F)$.

The *Merlin-Arthur* model [2, 5] combines the power of randomization and nondeterminism. Similar to the nondeterministic model, the protocol starts with a nondeterministic guess of c_1 bits, followed by c_2 bits of communication. However, the communication can be randomized, and the requirement is that the error probability be at most ϵ for *at least one* nondeterministic guess when $F(x_1, x_2, \ldots, x_k) = -1$ and for *all* possible nondeterministic guesses when $F(x_1, x_2, \ldots, x_k) = +1$. The *cost* of a Merlin-Arthur protocol is defined as $c_1 + c_2$. The *ϵ-error Merlin-Arthur communication complexity* of F, denoted $MA_\epsilon(F)$, is the least cost of an ϵ-error Merlin-Arthur protocol for F. Clearly, $MA_\epsilon(F) \leqslant \min\{N(F), R_\epsilon(F)\}$ for every F.

7.2. Communication lower bounds. To analyze the nondeterministic and Merlin-Arthur complexity of set disjointness, we start with a criterion for high communication complexity in these models. It was derived recently by Gavinsky and the author [23, Thm. 4.1] and builds on earlier work by Klauck [28, 29], including the generalized discrepancy method.

THEOREM 7.1 (Gavinsky and Sherstov). *Let $F\colon X \to \{-1,+1\}$ be a given k-party communication problem, where $X = X_1 \times X_2 \times \cdots \times X_k$. Fix a function $H\colon X \to \{-1,+1\}$ and a probability distribution P on X. Put*

$$\begin{aligned}\alpha &= P(F^{-1}(-1) \cap H^{-1}(-1)),\\ \beta &= P(F^{-1}(-1) \cap H^{-1}(+1)),\\ Q &= \log \frac{\alpha}{\beta + \operatorname{disc}_P(H)}.\end{aligned}$$

Then

$$\begin{aligned}&N(F) \geqslant Q,\\ &MA_{1/3}(F) \geqslant \min\left\{\Omega(\sqrt{Q}),\ \Omega\left(\frac{Q}{\log(2/\alpha)}\right)\right\}.\end{aligned}$$

A key technical ingredient in [23] is the following property of the AND function, which we will use in a similar way in this paper.

THEOREM 7.2 (Gavinsky and Sherstov). *There is a function $\psi\colon \{-1,+1\}^n \to \mathbb{R}$ with*

$$\begin{aligned}&\langle \psi, \mathrm{AND}_n\rangle > \frac{1}{3},\\ &\|\psi\|_1 = 1,\\ &\hat{\psi}(S) = 0, \qquad\qquad |S| < \deg_{1/3}(\mathrm{AND}_n),\\ &\psi(-1,\ldots,-1) < -\frac{1}{6}. \qquad (7.1)\end{aligned}$$

Proof (Gavinsky and Sherstov). The first three properties of ψ are guaranteed by Fact 2.2. To establish the remaining property, note that $\langle \psi, 1\rangle = 0$ because $\hat{\psi}(\varnothing) = 0$. Thus,

$$\begin{aligned}-2\psi(-1,\ldots,-1) &= \sum_{z\in\{-1,+1\}^n} \psi(z)\{\mathrm{AND}_n(z) - 1\}\\ &= \sum_{z\in\{-1,+1\}^n} \psi(z)\mathrm{AND}_n(z)\\ &> \frac{1}{3}.\end{aligned}$$

□

Two years ago, Gavinsky and the author [23] obtained a lower bound of $n^{\Omega(1/k)}/2^{2^k}$ on the nondeterministic and Merlin-Arthur communication complexity of set disjointness. The main result of this section, which we are about to establish, is an improved lower bound of $\Omega(n/4^k)^{1/4}$ for nondeterministic and $\Omega(n/4^k)^{1/8}$ for Merlin-Arthur protocols. Our proof closely follows the proof in [23], i.e., we use Theorem 7.2 to construct H and P for Theorem 7.1. The main difference resides in the discrepancy calculation, for which we turn to the master theorem in this paper on randomized communication complexity.

THEOREM 7.3 (restatement of Theorems 1.3 and 1.4). *The set disjointness problem obeys*

$$N(\mathrm{DISJ}_{n,k}) \geqslant \Omega\Big(\frac{n}{4^k}\Big)^{1/4},$$
$$MA_{1/3}(\mathrm{DISJ}_{n,k}) \geqslant \Omega\Big(\frac{n}{4^k}\Big)^{1/8}.$$

Proof. Let r be a parameter to be set later. Put $f = \mathrm{AND}_n$, $d = \deg_{1/3}(\mathrm{AND}_n)$, and fix $\psi\colon \{-1,+1\}^n \to \mathbb{R}$ as in Theorem 7.2. Let $F = f \circ \mathrm{DISJ}_{r,k}$ and define $\Psi\colon (\{0,1\}^{r\times k})^n \to \mathbb{R}$ as in the dual proof of Theorem 5.1, viz.,

$$\Psi(X_1,\dots,X_n) = 2^n \psi\left(\mathrm{DISJ}_{r,k}(X_1),\dots,\mathrm{DISJ}_{r,k}(X_n)\right) \prod_{i=1}^n \mu(X_i),$$

where $\mu = \mathscr{U}_r \times \mu_{r,k-1}$ as before. Then (5.8) shows that $\Psi = H \circ P$ for some sign tensor H and probability distribution P. In particular, (5.10) asserts that

$$\mathrm{disc}_P(H) \leqslant \left(\frac{2^k e n}{d\sqrt{r}}\right)^d. \tag{7.2}$$

By (7.1), we have $\psi(z) < 0$ whenever $f(z) = -1$, so that

$$P(F^{-1}(-1) \cap H^{-1}(+1)) = 0. \tag{7.3}$$

Also,

$$P(F^{-1}(-1) \cap H^{-1}(-1)) = P(F^{-1}(-1)) = |\psi(-1,\dots,-1)| > \frac{1}{6}, \tag{7.4}$$

where the first step uses (7.3), the second step uses the fact that μ places equal weight on the sets $\mathrm{DISJ}_{r,k}^{-1}(-1)$ and $\mathrm{DISJ}_{r,k}^{-1}(+1)$, and the final inequality uses (7.1). By (7.2)–(7.4) and Theorem 7.1,

$$N(F) = \Omega\left(d \log\left\{\frac{d\sqrt{r}}{2^k e n}\right\}\right), \qquad MA_{1/3}(F) = \Omega\left(d \log\left\{\frac{d\sqrt{r}}{2^k e n}\right\}\right)^{1/2}.$$

Recall now from Theorem 2.4 that $d > c\sqrt{n}$ for some constant $c > 0$. As a result, setting $r = 4^{k+2} n \lceil 1/c \rceil^2$ gives $N(F) = \Omega(\sqrt{n})$ and $MA_{1/3}(F) = \Omega(n^{1/4})$. It remains to note that $F = \mathrm{DISJ}_{4^{k+2}n^2\lceil 1/c\rceil^2,k}$. □

7.3. Applications to communication classes. Babai, Frankl, and Simon [3] defined analogues of computational complexity classes in communication. We will only mention a few of them, namely, those corresponding to efficient randomized, nondeterministic, and Merlin-Arthur protocols. For a given number of parties $k = k(n)$, fix a family $\{F_n\}_{n=1}^\infty$ of k-party communication problems, where $F_n\colon (\{0,1\}^n)^k \to \{-1,+1\}$. The family $\{F_n\}$ is said to belong to the communication class BPP_k if and only if $R_{1/3}(F_n) \leqslant \log^c n$ for some constant $c > 1$ and all $n > c$. Analogously, the family $\{F_n\}$ is said to belong to NP_k and MA_k if and only if the communication complexity of F_n in the nondeterministic and Merlin-Arthur models, respectively, is at most $\log^c n$ for some constant $c > 1$ and all $n > c$. The derived classes coNP_k and coMA_k have the usual definition, e.g., $\{F_n\} \in \mathsf{coNP}_k$ if and only if $\{-F_n\} \in \mathsf{NP}_k$.

A corollary to Theorem 7.3 is that set disjointness separates coNP_k from NP_k, BPP_k, and even MA_k for $k < (\frac{1}{2} - \epsilon)\log n$, where $\epsilon > 0$ is any constant.

THEOREM 7.4. *For $k \leqslant \left(\frac{1}{2} - \epsilon\right)\log n$, where $\epsilon > 0$ is any constant,*

$$\begin{aligned}
&\mathrm{DISJ}_{n,k} \in \mathsf{coNP}_k \setminus \mathsf{NP}_k,\\
&\mathrm{DISJ}_{n,k} \in \mathsf{coNP}_k \setminus \mathsf{BPP}_k,\\
&\mathrm{DISJ}_{n,k} \in \mathsf{coNP}_k \setminus \mathsf{MA}_k.
\end{aligned}$$

Proof. It suffices to prove the final statement, since MA_k contains NP_k and BPP_k.

Theorem 7.3 shows that $\mathrm{DISJ}_{n,k} \notin \mathsf{MA}_k$. On the other hand, it is well-known that $N(-\mathrm{DISJ}_{n,k}) \leqslant \lceil \log n \rceil + 2$. Specifically, the parties choose $i \in \{1, 2, \ldots, n\}$ nondeterministically and compute $x_{1,i} \wedge \cdots \wedge x_{k,i}$ with two bits of communication. As a result, $\mathrm{DISJ}_{n,k} \in \mathsf{coNP}_k$. □

8. APPLICATIONS TO CIRCUIT COMPLEXITY

We will now apply our results on small-bias communication to circuit complexity. We start with a well-known connection between multiparty communication and circuits, due to Håstad and Goldmann [25].

PROPOSITION 8.1 (Håstad and Goldmann). *Let f be a Boolean function computable by a* MAJ∘SYMM∘ANY *circuit, where the top gate has fan-in m, the middle gates have fan-in at most s, and the bottom gates have fan-in at most $k-1$. Then the k-party communication complexity of f obeys*

$$R_{\frac{1}{2} - \frac{1}{2(m+1)}}(f) \leqslant k\lceil \log(s+1) \rceil,$$

regardless of how the bits are assigned to the parties.

Proof (Håstad and Goldmann). The parties pick a random gate G at the middle level, evaluate it deterministically using $k\lceil \log(s+1) \rceil$ bits of communication, and output the answer. The deterministic computation is possible because every input to G can be computed by some party without communication, which makes it possible to partition the bottom gates among the parties and have each party report the sum of those inputs to G assigned to him. Since G is symmetric, the sum of its inputs uniquely determines its output. □

We arrive at the first result of this section, a lower bound on the size of MAJ ∘ SYMM ∘ ANY circuits with small bottom fan-in computing a depth-3 formula.

THEOREM 8.2. *Let $F_{n,k} \colon \{0,1\}^{4^k n^7 k} \to \{-1,+1\}$ be the depth-3 read-once $\{\wedge, \vee\}$-formula defined in Theorem* 5.7. *Then any circuit of type* MAJ ∘ SYMM ∘ ANY *with bottom fan-in at most $k-1$ computing $F_{n,k}$ has size $2^{\Omega(n/k)}$.*

Proof. We interpret $F_{n,k}$ as the k-party communication problem defined in Theorem 5.7. Let C be a circuit of type MAJ ∘ SYMM ∘ ANY that computes $F_{n,k}$, where the bottom fan-in of C is at most $k-1$. If C has size s, then the fan-in of the gates at the top and middle levels is bounded by s, which in view of Proposition 8.1 gives

$$R_{\frac{1}{2} - \frac{1}{2(s+1)}}(F_{n,k}) \leqslant k\lceil \log(s+1) \rceil.$$

By Theorem 5.7, this leads to $s \geqslant \exp(\Omega(n/k))$. □

Theorem 8.2 establishes Theorem 1.6 from the introduction.

Consider now a different computational model, that of MAJ ∘ SYMM ∘ AND circuits without any fan-in restrictions. We will prove a superpolynomial lower bound in this model as well. We will use a well-known argument due to Razborov and Wigderson [41] which reduces the task of proving lower bounds for MAJ ∘ SYMM ∘ AND circuits to proving lower bounds for MAJ ∘ SYMM ∘ ANY circuits with small bottom fan-in. This argument has already been used by several authors [48, 11] in the context of proving lower bounds for MAJ ∘ SYMM ∘ AND circuits computing AC^0 functions.

THEOREM 8.3 (Razborov and Wigderson). *Let $\delta > 0$ be a sufficiently small absolute constant, $f\colon\{0,1\}^N \to \{-1,+1\}$ a given function. For $\epsilon \in (0,\delta)$, define*

$$F_\epsilon = f \circ \mathrm{PARITY}_{\frac{1}{\epsilon}\ln^2 N}.$$

If F_ϵ is computable by a MAJ ∘ SYMM ∘ AND *circuit C of size at most $N^{\epsilon^2 \ln\ln N}$, then f is computable by a* MAJ ∘ SYMM ∘ AND *circuit of the same size with bottom fan-in at most $\epsilon \ln N$.*

For completeness, we include the short proof of this theorem. In what follows, we let G_ρ denote the result of applying a random restriction ρ to a gate or function G.

Proof of Theorem 8.3 (adapted from [41, 48, 11]). Let ρ be a random restriction that leaves each variable unset independently with probability $p \doteq 2\epsilon/\ln N$, and otherwise sets it to 0 or 1 with equal probability. For a conjunction K, let $|K|$ denote the number of literals in K. We claim that for every conjunction K,

$$\mathbf{P}[|K_\rho| \geqslant \epsilon \ln N] \leqslant N^{-\Theta(\epsilon \ln\ln N)}. \tag{8.1}$$

Indeed, for $|K| \leqslant \ln N \ln\ln N$,

$$\mathbf{P}[|K_\rho| \geqslant \epsilon \ln N] \leqslant \binom{|K|}{\epsilon \ln N} p^{\epsilon \ln N} \leqslant N^{-\Theta(\epsilon \ln\ln N)},$$

whereas for $|K| > \ln N \ln\ln N$

$$\mathbf{P}[|K_\rho| \geqslant \epsilon \ln N] \leqslant \mathbf{P}[K_\rho \not\equiv 1] = \left(\frac{1+p}{2}\right)^{|K|} \leqslant N^{-\Theta(\ln\ln N)}.$$

Applying (8.1) with a union bound across the bottom gates of C, we find that with probability $1 - o(1)$ the bottom fan-in of C_ρ is at most $\epsilon \ln N$. Furthermore, the probability that ρ does not turn any parity gate in F_ϵ into a constant is at least $1 - N(1-p)^{\frac{1}{\epsilon}\ln^2 N} = 1 - o(1)$. In particular, there is a random restriction ρ such that on the one hand, C_ρ has bottom fan-in at most $\epsilon \ln N$, and on the other hand f is a subfunction of C_ρ. □

We are now in a position to prove the promised lower bound.

THEOREM 1.7 (restated). *Every* MAJ ∘ SYMM ∘ AND *circuit that computes*

$$H_n(x) = \bigvee_{i=1}^{n} \bigwedge_{j=1}^{n} \bigvee_{k=1}^{\log n} \bigoplus_{\ell=1}^{\log^2 n} x_{i,j,k,\ell}$$

has size $n^{\Omega(\log\log n)}$.

Proof. Without loss of generality, we may assume that n is a power of 2. Let $F_{n,k}$ be the depth-3 read-once $\{\wedge,\vee\}$-formula constructed in Theorem 5.7. By Theorem 8.2, every MAJ ∘ SYMM ∘ ANY circuit with bottom fan-in $\log n - 1$ that computes $F_{n,\log n}$ has size

$2^{\Omega(n/\log n)}$. As a result, Theorem 8.3 gives a lower bound of $n^{\Omega(\log\log n)}$ on the size of any MAJ $\circ$ SYMM $\circ$ AND circuit computing the composition $H'_n = F_{n,\log n} \circ \mathrm{PARITY}_{c\log^2 n}$, where $c > 1$ is a sufficiently large constant. It remains to note that H'_n is a subfunction of H_{nC} for a large enough constant $C = C(c) > 1$. □

REFERENCES

[1] E. Allender. A note on the power of threshold circuits. In *Proceedings of the Thirtieth Annual IEEE Symposium on Foundations of Computer Science*, pages 580–584, 1989.

[2] L. Babai. Trading group theory for randomness. In *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, pages 421–429, 1985.

[3] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of the Twenty-Seventh Annual IEEE Symposium on Foundations of Computer Science*, pages 337–347, 1986.

[4] L. Babai, T. P. Hayes, and P. G. Kimmel. The cost of the missing bit: Communication complexity with help. *Combinatorica*, 21(4):455–488, 2001.

[5] L. Babai and S. Moran. Arthur-Merlin games: A randomized proof system, and a hierarchy of complexity classes. *J. Comput. Syst. Sci.*, 36(2):254–276, 1988.

[6] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. Syst. Sci.*, 45(2):204–232, 1992.

[7] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[8] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[9] P. Beame, M. David, T. Pitassi, and P. Woelfel. Separating deterministic from nondeterministic NOF multiparty communication complexity. In *Proc. of the 34th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 134–145, 2007.

[10] P. Beame, M. David, T. Pitassi, and P. Woelfel. Separating deterministic from randomized multiparty communication complexity. *Theory of Computing*, 6(1):201–225, 2010.

[11] P. Beame and D.-T. Huynh-Ngoc. Multiparty communication complexity and threshold circuit complexity of AC^0. In *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science*, pages 53–62, 2009. Preliminary version in ECCC Report TR08-082, September 2008.

[12] P. Beame, T. Pitassi, and N. Segerlind. Lower bounds for Lovász-Schrijver systems and beyond follow from multiparty communication complexity. *SIAM J. Comput.*, 37(3):845–869, 2007.

[13] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Computational Complexity*, 15(4):391–432, 2006.

[14] A. Ben-Aroya, O. Regev, and R. de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs. In *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science*, pages 477–486, 2008.

[15] J. Briet, H. Buhrman, T. Lee, and T. Vidick. Multiplayer XOR games and quantum communication complexity with clique-wise entanglement. Manuscript at `http://arxiv.org/abs/0911.4007`, 2009.

[16] A. K. Chandra, M. L. Furst, and R. J. Lipton. Multi-party protocols. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, pages 94–99, 1983.

[17] A. Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proceedings of the Forty-Eighth Annual IEEE Symposium on Foundations of Computer Science*, pages 449–458, 2007.

[18] A. Chattopadhyay. *Circuits, Communication, and Polynomials*. PhD thesis, McGill University, 2008.

[19] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. In *Electronic Colloquium on Computational Complexity (ECCC)*, January 2008. Report TR08-002.

[20] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988.

[21] M. David and T. Pitassi. Separating NOF communication complexity classes RP and NP. In *Electronic Colloquium on Computational Complexity (ECCC)*, February 2008. Report TR08-014.

[22] M. David, T. Pitassi, and E. Viola. Improved separations between nondeterministic and randomized multiparty communication. *ACM Transactions on Computation Theory (TOCT)*, 1(2), 2009.

[23] D. Gavinsky and A. A. Sherstov. A separation of NP and coNP in multiparty communication complexity. *Theory of Computing*, 6(10):227–245, 2010.

[24] V. Grolmusz. The BNS lower bound for multi-party protocols in nearly optimal. *Inf. Comput.*, 112(1):51–54, 1994.

[25] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1:113–129, 1991.

[26] R. Jain, H. Klauck, and A. Nayak. Direct product theorems for classical communication complexity via subdistribution bounds. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 599–608, 2008.

[27] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[28] H. Klauck. Lower bounds for quantum communication complexity. In *Proceedings of the Forty-Second Annual IEEE Symposium on Foundations of Computer Science*, pages 288–297, 2001.

[29] H. Klauck. Rectangle size bounds and threshold covers in communication complexity. In *Proceedings of the Eighteenth Annual IEEE Conference on Computational Complexity*, pages 118–134, 2003.

[30] H. Klauck. A strong direct product theorem for disjointness. In *Proceedings of the Forty-Second Annual ACM Symposium on Theory of Computing*, pages 77–86, 2010.

[31] H. Klauck, R. Špalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM J. Comput.*, 36(5):1472–1493, 2007.

[32] M. Krause and P. Pudlák. On the computational power of depth-2 circuits with threshold and modulo gates. *Theor. Comput. Sci.*, 174(1–2):137–156, 1997.

[33] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, 1997.

[34] T. Lee, G. Schechtman, and A. Shraibman. Lower bounds on quantum multiparty communication complexity. In *Proceedings of the Twenty-Fourth Annual IEEE Conference on Computational Complexity*, pages 254–262, 2009.

[35] T. Lee and A. Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009. Preliminary version at `http://arxiv.org/abs/0712.4279v1`, December 2007.

[36] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Struct. Algorithms*, 34(3):368–394, 2009.

[37] M. L. Minsky and S. A. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, Mass., 1969.

[38] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[39] A. A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[40] A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Sciences, Mathematics*, 67:145–159, 2002.

[41] A. A. Razborov and A. Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. *Inf. Process. Lett.*, 45(6):303–307, 1993.

[42] A. A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 85–94, 2008. Preliminary version in ECCC Report TR07-100, September 2007.

[43] A. A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. Preliminary version in Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing, 2007.

[44] A. A. Sherstov. On quantum-classical equivalence for composed communication problems. *Quantum Information & Computation*, 10(5-6):435–455, 2010.

[45] A. A. Sherstov. Strong direct product theorems for quantum communication and query complexity. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, pages 41–50, 2011. Full version available as ECCC Report TR11-040.

[46] Y. Shi and Y. Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5–6):444–460, 2009.

[47] P. Tesson. *Computational complexity questions related to finite monoids and semigroups*. PhD thesis, McGill University, 2003.

[48] E. Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM J. Comput.*, 36(5):1387–1403, 2007.

[49] E. Viola and A. Wigderson. One-way multiparty communication lower bound for pointer jumping with applications. *Combinatorica*, 29(6):719–743, 2009.

[50] A. C.-C. Yao. Some complexity questions related to distributive computing. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.

[51] A. C.-C. Yao. On ACC and threshold circuits. In *Proceedings of the Thirty-First Annual IEEE Symposium on Foundations of Computer Science*, pages 619–627, 1990.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de