Electronic Colloquium on Computational Complexity, Report No. 5 (2013)

COMMUNICATION LOWER BOUNDS USING DIRECTIONAL DERIVATIVES

ALEXANDER A. SHERSTOV

ABSTRACT. We prove that the set disjointness problem has randomized communication complexity $\Omega(\sqrt{n}/2^k k)$ in the number-on-the-forehead model with k parties, a quadratic improvement on the previous bound $\Omega(\sqrt{n}/2^k)^{1/2}$. Our result remains valid for quantum protocols, where it is essentially tight. Proving it was an open problem since 1997, even in restricted settings such as one-way classical protocols with $k = 4$ parties. We obtain other near-optimal results for multiparty set disjointness, including an XOR lemma, a direct product theorem, and lower bounds for nondeterministic and Merlin-Arthur protocols. The proof contributes a novel technique for lower bounds on multiparty communication, based on directional derivatives of communication protocols over the reals.

CONTENTS

* Computer Science Department, UCLA, Los Angeles, California 90095. ✉ sherstov@cs.ucla.edu
Supported by NSF CAREER award CCF-1149018.

ISSN 1433-8092

1. Introduction

Set disjointness is by far the most studied problem in communication complexity theory [4, 31, 44, 5, 8, 45, 51, 34, 11, 54, 47, 50, 38, 20, 10, 12, 33]. The simplest version of the problem features two parties, Alice and Bob. Alice receives as input a subset $S \subseteq \{1, 2, \ldots, n\}$, Bob receives a subset $T \subseteq \{1, 2, \ldots, n\}$, and their goal is to determine with minimal communication whether the subsets are disjoint. One also studies a promise version of this problem called *unique set disjointness*, in which the intersection $S \cap T$ is either empty or contains a single element. The communication complexity of two-party 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 [4] and strengthened to a tight $\Omega(n)$ by Kalyanasundaram and Schnitger [31]. Simpler proofs of the linear lower bound were discovered by Razborov [44] and Bar-Yossef et al. [8]. All three proofs [31, 44, 8] of the linear lower bound apply to unique set disjointness. Finally, Razborov [45] 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 [47]. Already in the two-party setting, the study of set disjointness contributed to communication complexity theory a variety of techniques, including ideas from combinatorics, Kolmogorov complexity, information theory, matrix analysis, and Fourier analysis.

We study the complexity of set disjointness in the model with three or more parties. We use the *number-on-the-forehead* model of multiparty communication, due to Chandra, Furst, and Lipton [18]. This model features k parties and a function $f(x_1, x_2, \ldots, x_k)$ with k arguments. Communication occurs in broadcast, a bit sent by any given party instantly reaching everyone else. The input $(x_1, x_2, \ldots, x_k)$ is distributed among the parties by giving the ith party the arguments $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$ but not x_i. One can think of x_i as written on the ith party's forehead, hence the name of the model. The number-on-the-forehead model is the main model in the area because any other way of assigning arguments to parties results in a less powerful model (provided of course that one does not assign all the arguments to some party, in which case there is never a need to communicate).

In the k-party version of set disjointness, the inputs are $S_1, S_2, \ldots, S_k \subseteq \{1, 2, \ldots, n\}$, and the ith party knows all the inputs except for S_i. The goal is to determine whether the sets have empty intersection: $S_1 \cap S_2 \cap \cdots \cap S_k = \varnothing$. For unique set disjointness, the parties additionally know that the intersection $S_1 \cap S_2 \cap \cdots \cap S_k$ is either empty or contains a unique element. It is common to represent the input to set disjointness by a $k \times n$ Boolean matrix $X = [x_{ij}]$, whose rows correspond to the characteristic vectors of the input sets. In this notation, set disjointness is given by simple CNF formula:

$$\mathrm{DISJ}_{k,n}(X) = \bigwedge_{j=1}^{n} \bigvee_{i=1}^{k} \overline{x_{ij}}. \tag{1.1}$$

Unique set disjointness $\mathrm{UDISJ}_{k,n}$ is given by an identical formula, with the understanding that the input matrix X contains at most one column consisting entirely of ones.

Progress on the communication complexity of set disjointness for $k \geqslant 3$ parties is summarized in Table 1. In a surprising result, Grolmusz [27] 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 [51] and Beame et al. [11] obtained a lower bound of $\Omega\big(\frac{1}{k} \log n\big)$ for randomized protocols. Four years later, Lee and Shraibman [38] and Chattopadhyay and Ada [20] gave an improved result. These authors generalized the two-party method of [46, 47] to $k \geqslant 3$ parties and thereby obtained a lower bound of $\Omega(n/2^{2^k k})^{1/(k+1)}$ on the randomized communication complexity of set disjointness. Their lower bound was strengthened by Beame and Huynh-Ngoc [10] to $(n^{\Omega(\sqrt{k/\log n})}/2^{k^2})^{1/(k+1)}$, which is an improvement for k large enough. All lower bounds listed up to this point are weaker than $\Omega(n/2^{k^3})^{1/(k+1)}$, which means that they become subpolynomial as soon as the number of parties k starts to grow. Three years later, a lower bound of $\Omega(n/4^k)^{1/4}$ was obtained in [48] on the randomized communication complexity of set disjointness, which remains polynomial for up to $k \approx \frac{1}{2}\log n$ and comes close to matching Grolmusz's upper bound.

The $\Omega(n/4^k)^{1/4}$ lower bound is not an accidental numerical value. It represents what we call the *triangle inequality barrier* in multiparty communication complexity, described in detail at the end of the introduction. We are able to break this barrier and obtain a quadratically stronger lower bound. In the theorem that follows, R_ϵ denotes ϵ-error randomized communication complexity.

THEOREM 1.1 (Main result). *Set disjointness and unique set disjointness have randomized communication complexity*

$$R_{1/3}(\mathrm{DISJ}_{k,n}) \geqslant R_{1/3}(\mathrm{UDISJ}_{k,n}) = \Omega\left(\frac{\sqrt{n}}{2^k k}\right).$$

Bound	**Reference**
$O\left(\log^2 n + \frac{k^2 n}{2^k}\right)$	Grolmusz [27]
$\Omega\left(\frac{\log n}{k}\right)$	Tesson [51], Beame, Pitassi, Segerlind, and Wigderson [11]
$\Omega\left(\frac{n}{2^{2^k}k}\right)^{\frac{1}{k+1}}$	Lee and Shraibman [38], Chattopadhyay and Ada [20]
$\left(\frac{n^{\Omega(\sqrt{k/\log n})}}{2^{k^2}}\right)^{\frac{1}{k+1}}$	Beame and Huynh-Ngoc [10]
$\Omega\left(\frac{n}{4^k}\right)^{1/4}$	Sherstov [48]
$\Omega\left(\frac{\sqrt{n}}{2^k k}\right)$	This paper

TABLE 1. Communication complexity of k-party set disjointness.

Two remarks are in order. Over the years, the lack of progress on set disjointness prompted researchers to consider restricted multiparty protocols, such as *one-way* protocols where the parties $1, 2, \ldots, k$ speak in that order and the last party announces the answer. An even more restricted form of communication is a *simultaneous* protocol, in which the parties simultaneously and independently send a message to a referee who then announces the answer. In 1997, Wigderson proved a lower bound of $\Omega(\sqrt{n})$ for solving set disjointness by a simultaneous protocol with $k = 3$ parties (unpublished by Wigderson, the proof appeared in [5]). Since then, several papers [5, 51, 11, 54, 12, 33] have examined the multiparty complexity of set disjointness for simultaneous, one-way, and other restricted kinds of protocols. The strongest communication lower bound [51, 11] obtained in that line of research was $\Omega(n/k^k)^{1/(k-1)}$. To summarize, prior to our work it was an open problem to generalize Wigderson's 1997 lower bound even to $k = 4$ parties, communicating one-way or simultaneously.

Second, by the results of [37, 14], all communication lower bounds in this paper generalize to the *quantum* model. In particular, Theorem 1.1 implies a lower bound of $\sqrt{n}/2^{k+o(k)}$ on the bounded-error quantum communication complexity of set disjointness. This lower bound essentially matches the well-known quantum protocol for set disjointness due to Buhrman, Cleve, and Wigderson [16], with cost $\lceil \sqrt{n/2^k} \rceil \log^{O(1)} n$. For the reader's convenience, we provide a sketch of the protocol in Remark 5.4. Thus, our results essentially settle the bounded-error quantum communication complexity of set disjointness:

$$\frac{\sqrt{n}}{2^{k+o(k)}} \leqslant Q_{1/3}(\mathrm{UDISJ}_{k,n}) \leqslant Q_{1/3}(\mathrm{DISJ}_{k,n}) \leqslant \left\lceil \frac{\sqrt{n}}{2^{k/2}} \right\rceil \log^{O(1)} n.$$

Our technique allows us to obtain several additional results, discussed next.

XOR lemmas and direct product theorems. In a seminal paper, Yao [55] asked whether computation admits economies of scale. More concretely, 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). Common sense suggests that solving ℓ independent instances of the problem requires $\Omega(\ell R)$ units of the resource. 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 $2^{-\Theta(\ell)}$. Such a statement is called a *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 $\frac{1}{2} + 2^{-\Theta(\ell)}$. XOR lemmas and direct product theorems are motivated by basic intellectual curiosity as well as a number of applications, including separations of circuit classes, improvement of soundness in proof systems, inapproximability results for optimization problems, and time-space trade-offs.

In communication complexity, the direct product question has been studied for over twenty years. We refer the reader to [33, 49] 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 [34, 11, 12, 30, 33, 49], including classical one-way protocols [30], classical two-way protocols [11, 33], quantum one-way protocols [12], and quantum two-way protocols [34, 49]. Proving any kind of direct product result for three or more parties remained an open problem until the paper [48] a year ago. In [48], a communication lower bound of $\ell \cdot \Omega(n/4^k)^{1/4}$ was proved for the following tasks: computing the XOR of ℓ instances of set disjointness with probability of correctness $\frac{1}{2} + 2^{-\Theta(\ell)}$; solving ℓ instances of set disjointness simultaneously with probability of correctness at least $2^{-\Theta(\ell)}$. We obtain an improved result:

THEOREM 1.2. *Let $\epsilon > 0$ be a sufficiently small absolute constant. The following tasks require $\ell \cdot \Omega(\sqrt{n}/2^k k)$ bits of communication each:*

(i) *computing the XOR of ℓ instances of* $\mathrm{UDISJ}_{k,n}$ *with probability at least $\frac{1}{2} + 2^{-\epsilon\ell}$;*
(ii) *solving with probability $2^{-\epsilon\ell}$ at least $(1-\epsilon)\ell$ among ℓ instances of* $\mathrm{UDISJ}_{k,n}$.

Theorem 1.2 generalizes Theorem 1.1, showing that $\Omega(\sqrt{n}/2^k k)$ is in fact a lower bound on the per-instance cost of set disjointness. The communication lower bound in Theorem 1.2 is quadratically stronger than in previous work [48]. Clearly, Theorem 1.2 also holds for set disjointness, a problem harder than $\mathrm{UDISJ}_{k,n}$. Finally, this theorem generalizes to quantum protocols, where it is essentially tight.

Nondeterministic and Merlin-Arthur communication. Nondeterministic communication is defined in complete analogy with 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 presented with a negative instance of F, and for *some* guess string when presented with a positive instance. We further consider *Merlin-Arthur* protocols [3, 6], 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 at least $2/3$ for *all* guess strings when presented with a negative instance of F, and for *some* guess string when presented with a positive instance.

Nondeterministic and Merlin-Arthur protocols have been extensively studied for $k = 2$ parties but are much less understood for $k \geqslant 3$. To illustrate, it was only four years ago that the first nontrivial lower bound, $n^{\Omega(1/k)}/2^{2^k}$, was obtained [23] on the multiparty communication complexity of set disjointness in these models. That lower bound was improved in [48] to $\Omega(n/4^k)^{1/4}$ for nondeterministic protocols and $\Omega(n/4^k)^{1/8}$ for Merlin-Arthur protocols, both of which are tight up to a polynomial. In this paper, we obtain quadratically stronger lower bounds in both models.

THEOREM 1.3. *Set disjointness has nondeterministic and Merlin-Arthur complexity*

$$N(\mathrm{DISJ}_{k,n}) = \Omega\left(\frac{\sqrt{n}}{2^k k}\right),$$

$$MA(\mathrm{DISJ}_{k,n}) = \Omega\left(\frac{\sqrt{n}}{2^k k}\right)^{1/2}.$$

Set disjointness should be contrasted in this regard with its complement $\neg\mathrm{DISJ}_{k,n}$, whose nondeterministic complexity is at most $\log n + O(1)$. Indeed, it suffices to guess an element $i \in \{1, 2, \ldots, n\}$ and verify with two bits of communication that $i \in S_1 \cap S_2 \cap \cdots \cap S_k$.

Small-bias communication and discrepancy. Much of the work in communication complexity revolves around the notion of *discrepancy*. Roughly speaking, the discrepancy of a function F is the maximum correlation of F with a constant-cost communication protocol. One of the many uses of discrepancy is proving lower bounds for *small-bias protocols*, which are randomized protocols with probability of correctness vanishingly close to the trivial value $1/2$. Quantitatively speaking, any function with discrepancy γ requires $\log \frac{1}{\sqrt{\gamma}}$ bits of communication to achieve correctness probability $\frac{1}{2} + \frac{1}{2}\sqrt{\gamma}$. The converse also holds, up to minor numerical adjustments. Thus, the study of discrepancy is the study of small-bias communication.

In a famous result, Babai, Nisan, and Szegedy [7] proved that the *generalized inner product* function $\bigoplus_{j=1}^{n} \bigwedge_{i=1}^{k} x_{ij}$ has exponentially small discrepancy, $\exp(-\Omega(n/4^k))$. Thus, generalized inner product does not admit an efficient protocol with any nonnegligible advantage over random guessing, much less a bounded-error protocol. The proof in [7] crucially exploits the XOR function, and until several years ago it was unknown whether *any* constant-depth $\{\wedge, \vee, \neg\}$-circuit of polynomial size has small discrepancy. The most natural candidate, set disjointness, is of no use here: while its bounded-error communication complexity is high, its discrepancy turns out to be $\Theta(1/n)$. The question was finally resolved for $k = 2$ parties in [17, 46, 47], with a bound of $\exp(-\Omega(n^{1/3}))$ on the discrepancy of an $\{\wedge, \vee\}$-formula of depth 3 and size n. Since then, a series of papers have studied the question for $k \geqslant 3$ parties. Table 2 gives a quantitative summary of this line of research. The best multiparty bound prior to this paper was $\exp(-\Omega(n/4^k)^{1/7})$, obtained in [48] for an $\{\wedge, \vee\}$-formula of depth 3 and size nk. We prove the following stronger result.

THEOREM 1.4. *There is an explicit k-party communication problem $H_{k,n}$, given by an $\{\wedge, \vee\}$-formula of depth 3 and size nk, with discrepancy*

$$\operatorname{disc}(H_{k,n}) = \exp\left\{-\Omega\left(\frac{n}{4^k k^2}\right)^{1/3}\right\}.$$

Depth	Discrepancy	Reference
3	$\exp\{-\Omega(n^{1/3})\}, \quad k = 2$	Buhrman, Vereshchagin, and de Wolf [17] Sherstov [46, 47]
3	$\exp\left\{-\Omega\left(\frac{n}{4^k}\right)^{1/(6k2^k)}\right\}$	Chattopadhyay [19]
6	$\exp\left\{-\Omega\left(\frac{n}{2^{31k}}\right)^{1/29}\right\}$	Beame and Huynh-Ngoc [10]
3	$\exp\left\{-\Omega\left(\frac{n}{4^k}\right)^{1/7}\right\}$	Sherstov [48]
3	$\exp\left\{-\Omega\left(\frac{n}{4^k k^2}\right)^{1/3}\right\}$	This paper

TABLE 2. Multiparty discrepancy of constant-depth $\{\wedge, \vee\}$-circuits of size nk.

In particular,

$$R_{\frac{1}{2}-\exp\left\{-\Omega\left(\frac{n}{4^k k^2}\right)^{1/3}\right\}}(H_{k,n}) = \Omega\left(\frac{n}{4^k k^2}\right)^{1/3}.$$

Theorem 1.4 is satisfying in that it matches the state of the art for two-party communication, i.e., even in the setting of two parties no bound is known better than the multiparty bound of Theorem 1.4. This theorem is qualitatively optimal with respect to the number of parties k: by the results in [2, 29], every polynomial-size $\{\wedge, \vee, \neg\}$-circuit of constant depth has discrepancy at least $2^{-\log^c n}$ for $k \geqslant \log^c n$ parties, where $c > 1$ is a constant. Theorem 1.4 is also optimal with respect to circuit depth because polynomial-size DNF and CNF formulas have discrepancy at least $1/n^{O(1)}$, regardless of the number of parties k. In Section **6.4**, we give applications of Theorem 1.4 to circuit complexity.

The triangle inequality barrier. To properly set the stage for our approach, we first outline the proof of the previous best lower bound for set disjointness [48] and explain its fundamental limitations. Let G be a k-party communication problem, with domain $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$. In what follows, we refer to G as a *gadget*. The paper [48] studies the communication complexity of composed functions of the form $F = f(G, G, \ldots, G)$, where $f : \{0,1\}^n \to \{0,1\}$. Thus, the new communication problem F has domain $\mathscr{X}^n = \mathscr{X}_1^n \times \mathscr{X}_2^n \times \cdots \times \mathscr{X}_k^n$. The motivation for studying composed communication problems is obvious from (1.1), which shows that $\mathrm{DISJ}_{k,nm} = \mathrm{AND}_n(\mathrm{DISJ}_{k,m}, \ldots, \mathrm{DISJ}_{k,m})$.

Assume henceforth that f does not approximate in the infinity norm to a low-degree real polynomial, an assumption that holds for $f = \mathrm{AND}_n$ and any other function of the smallest practical significance. Now, consider a linear operator L that maps real functions $\Pi : \mathscr{X}^n \to \mathbb{R}$ to real functions $L\Pi : \{0,1\}^n \to \mathbb{R}$ in the following natural way: the value $(L\Pi)(x_1, x_2, \ldots, x_n)$ is obtained by averaging Π one way or another on the set $G^{-1}(x_1) \times G^{-1}(x_2) \times \cdots \times G^{-1}(x_n)$. The operator L is defined the way it is in order to ensure that $f = LF$. The approach of [48] is to prove that if $\Pi : \mathscr{X}^n \to [0,1]$ is the acceptance probability of any low-cost ϵ-error randomized protocol, then $L\Pi$ approximates in the infinity norm to a low-degree real polynomial $\tilde{f}$. This immediately rules out an efficient protocol for F, for its existence would force

$$|f - \tilde{f}| = |LF - \tilde{f}| \approx |LF - L\Pi| = |L(\underbrace{F - \Pi}_{\leqslant \epsilon})| \leqslant \epsilon,$$

in contradiction to the inapproximability of f by low-degree polynomials.

The difficult part of the above program is proving that $L\Pi$ approximates to a low-degree polynomial. The paper [48] does so constructively, by arguing that the Fourier spectrum of $L\Pi$ resides almost entirely on low-order characters:

$$|\widehat{L\Pi}(S)| < 2^r \cdot 2^{-|S|} \binom{n}{|S|}^{-1}, \qquad S \subseteq \{1, 2, \ldots, n\}, \tag{1.2}$$

where r is the cost of the communication protocol. In particular, an approximating polynomial for $L\Pi$ results from truncating the Fourier spectrum at degree $r + O(1)$. The type of concentration in (1.2) is quite extreme, and it requires a fairly complicated gadget G. A key technical result in [48] is that the set disjointness problem on $O(4^k n^2/|S|^2)$ variables is an acceptable choice for G. A fundamental limitation of this approach is that the size of the gadget G must grow with n, for the obvious reason that the number of Fourier coefficients of $L\Pi$ grows with n and we apply a triangle inequality across them. We refer to this obstacle as the *triangle inequality barrier.*

Our proof. For our main result, we must make do with a gadget G whose size is independent of n. This requires breaking the triangle inequality barrier, i.e., finding a way to approximate protocols by low-degree polynomials without summing discarded Fourier coefficients term by term. In the setting of $k = 2$ parties, the triangle inequality barrier was successfully overcome in 2007 using matrix analysis [47]. For multiparty communication, the problem remained wide open prior to this paper because matrix-analytic tools do not apply to $k \geqslant 3$.

Our solution involves two steps. First, we derive a criterion for the approximability of *any* given function $\phi\colon \{0,1\}^n \to \mathbb{R}$ by low-degree polynomials. Specifically, recall that the directional derivative of ϕ in the direction $S \subseteq \{1, 2, \ldots, n\}$ at the point $x \in \{0,1\}^n$ is given by $(\partial\phi/\partial S)(x) = \frac{1}{2}\phi(x) - \frac{1}{2}\phi(x \oplus \mathbf{1}_S)$, where $\mathbf{1}_S$ denotes the characteristic vector of S. Directional derivatives of higher order are obtained by differentiating repeatedly. We prove that the error in an infinity-norm approximation of ϕ by a degree-d polynomial does not exceed

$$K^{d+1}\Delta(\phi, d+1) + K^{d+2}\Delta(\phi, d+2) + K^{d+3}\Delta(\phi, d+3) + \cdots, \tag{1.3}$$

where $K > 2$ is an absolute constant and $\Delta(\phi, i)$ is the maximum magnitude of a degree-i directional derivative of ϕ with respect to any pairwise disjoint sets $S_1, S_2, \ldots, S_i$. The crucial point is that the dimension n of the ambient hypercube never figures in (1.3). Thus, we are able to prove that ϕ approximates to a low-degree polynomial by taking a *global* view of the analytic structure of ϕ, without assuming any bound on the sum of the absolute values of ϕ's high-order Fourier coefficients. This makes it possible to approximate a large class of functions ϕ that were off limits to previous techniques, including communication protocols. The author finds this result to be of general interest in Boolean function analysis, independent of its use in this paper to prove communication lower bounds.

To apply the above criterion to multiparty communication, we must bound the directional derivatives of $L\Pi$ for every Π derived from a low-cost communication protocol. This is equivalent to bounding the *repeated discrepancy* of the gadget G, a new communication complexity measure that we introduce. The standard notion of discrepancy, reviewed above, involves fixing a probability distribution μ on the domain of G and challenging a constant-cost communication protocol to solve an instance X of G chosen at random according to μ. In computing the repeated discrepancy of G, one presents the communication protocol with *infinitely many* instances $X_1, X_2, X_3, \ldots$ of the given communication problem G, each chosen independently from μ conditioned on $G(X_1) = G(X_2) = G(X_3) = \cdots$. Thus, the instances are either all positive or all negative, and the protocol's challenge is to tell which is the case. It is considerably harder to bound the repeated discrepancy than the usual discrepancy because the additional instances $X_2, X_3, \ldots$ generally reveal new information about the truth status of X_1. In fact, it is not clear a priori whether there is *any* distribution μ under which set disjointness has repeated discrepancy less than the maximum possible value 1, let alone $o(1)$ as our application requires. By a detailed probabilistic analysis, we are able to prove the desired $o(1)$ bound for a suitable distribution μ.

Once we have overcome the above two challenges, we obtain an efficient way to transform communication protocols into approximating polynomials. This passage then allows us to expeditiously prove all the results of this paper, stated as Theorems 1.1–1.4 above.

2. PRELIMINARIES

There are two common ways to encode the Boolean values "true" and "false," the classic encoding using $1, 0$ and a more recent one using $-1, +1$. The former is more convenient in combinatorial applications, whereas the latter is more natural and economical when working with analytic tools such as the Fourier transform or matrix analysis. In this paper, we will use both encodings depending on context. To exclude any possibility of confusion, we reserve the term *Boolean predicate* in the remainder of the paper for mappings of the form $\mathscr{X} \to \{0,1\}$, and the term *Boolean function* for mappings $\mathscr{X} \to \{-1,+1\}$. As a notational aid to distinguish predicates from functions, we always typeset the former with an asterisk, as in PARITY* and AND*, reserving unstarred symbols such as PARITY and AND for the corresponding Boolean functions. More generally, to every Boolean function f we associate the corresponding Boolean predicate $f^* = (1-f)/2$. A *partial function* f on $\mathscr{X}$ is a function whose domain of definition, denoted $\operatorname{dom} f$, is a nonempty proper subset of $\mathscr{X}$. For emphasis, we will sometimes refer to functions with $\operatorname{dom} f = \mathscr{X}$ as *total.* For (possibly partial) Boolean functions f and g on $\{0,1\}^n$ and $\mathscr{X}$, respectively, we let $f \circ g$ denote the componentwise composition of f with g, i.e., the (possibly partial) Boolean function on $\mathscr{X}^n$ given by $(f \circ g)(x_1, x_2, \dots, x_n) = f(g^*(x_1), g^*(x_2), \dots, g^*(x_n))$. Clearly, the domain of $f \circ g$ is the set of all $(x_1, x_2, \dots, x_n) \in (\operatorname{dom} g)^n$ for which $(g^*(x_1), g^*(x_2), \dots, g^*(x_n)) \in \operatorname{dom} f$.

We let ε denote the empty string, which is the only element of the zero-dimensional hypercube $\{0,1\}^0$. For a bit string $x \in \{0,1\}^n$, we let $|x| = x_1 + x_2 + \dots + x_n$ denote the Hamming weight of x. The *kth level* of the Boolean hypercube $\{0,1\}^n$ is the subset $\{x \in \{0,1\}^n : |x| = k\}$. The componentwise conjunction and componentwise XOR of $x, y \in \{0,1\}^n$ are denoted $x \wedge y = (x_1 \wedge y_1, \dots, x_n \wedge y_n)$ and $x \oplus y = (x_1 \oplus y_1, \dots, x_n \oplus y_n)$. 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 $\overline{x} = (x_1 \oplus 1, \dots, x_n \oplus 1)$. The notation $\log x$ refers to the logarithm of x to base 2. For a subset $S \subseteq \{1, 2, \dots, n\}$, its characteristic vector $\mathbf{1}_S$ is given by

$$(\mathbf{1}_S)_i = \begin{cases} 1 & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

For $i = 1, 2, \dots, n$, we define $e_i = \mathbf{1}_{\{i\}}$. In other words, e_i is the vector with 1 in the ith component and zeroes everywhere else. We identify $\{0,1\}^n$ with the n-dimensional vector space $\mathrm{GF}(2)^n$, with addition corresponding to componentwise XOR. This makes available standard vector space notation, e.g., $ax \oplus by = (\dots, (a_i x_i) \oplus (b_i y_i), \dots)$ for $a, b \in \{0,1\}$ and strings $x, y \in \{0,1\}^n$. A more complicated instance of this notation that we will use many times is $w \oplus z_1 \mathbf{1}_{S_1} \oplus z_2 \mathbf{1}_{S_2} \oplus \dots \oplus z_d \mathbf{1}_{S_d}$, where $z_1, z_2, \dots, z_d \in \{0,1\}$, $w \in \{0,1\}^n$, and $S_1, S_2, \dots, S_d \subseteq \{1, 2, \dots, n\}$.

The *parity* of a Boolean string $x \in \{0,1\}^n$, denoted PARITY$^*(x) \in \{0,1\}$, is defined as usual by PARITY$^*(x) = \bigoplus_{i=1}^n x_i$. We adopt the convention that

$$\binom{n}{-1} = \binom{n}{-2} = \binom{n}{-3} = \dots = 0$$

for every positive integer n. For positive integers n, m, k, one has

$$\sum_{i=0}^{k} \binom{n}{i} \binom{m}{k-i} = \binom{n+m}{k}, \tag{2.1}$$

a combinatorial identity known as *Vandermonde's convolution.* The total degree of a multivariate real polynomial p is denoted $\deg p$. The Kronecker delta is given by

$$\delta_{x,y} = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise,} \end{cases}$$

where x, y are elements of some set. We let $\mathbb{Z}^+ = \{1, 2, 3, \ldots, \}$ and $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ denote positive integers and natural numbers, respectively. We adopt the convention that the linear span of the empty set is the zero vector: $\operatorname{span} \varnothing = \{0\}$. The symmetric group of order n is denoted S_n. For a string $x \in \{0,1\}^n$ and a permutation $\sigma \in S_n$, we define $\sigma x = (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$. A function $f\colon \{0,1\}^n \to \mathbb{R}$ is called *symmetric* if $f(x) = f(\sigma x)$ for all x and all $\sigma \in S_n$. Equivalently, f is symmetric if and only if it is determined uniquely by the Hamming weight $|x|$ of the input.

The familiar functions $\mathrm{AND}_n, \mathrm{OR}_n\colon \{0,1\}^n \to \{-1,+1\}$ are given by $\mathrm{AND}_n(x) = \bigwedge_{i=1}^n x_i$ and $\mathrm{OR}_n(x) = \bigvee_{i=1}^n x_i$. We also define a partial Boolean function $\widetilde{\mathrm{AND}}_n$ on $\{0,1\}^n$ as the restriction of AND_n to the set $\{x : |x| \geqslant n-1\}$. In other words,

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

Analogously, we define a partial Boolean function $\widetilde{\mathrm{OR}}_n$ on $\{0,1\}^n$ as the restriction of OR_n to the set $\{x : |x| \leqslant 1\}$.

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

$$\|f\|_\infty = \max_{x \in \mathscr{X}} |f(x)| \qquad (f \in \mathbb{R}^{\mathscr{X}}), \tag{2.2}$$

$$\|f\|_1 = \sum_{x \in \mathscr{X}} |f(x)| \qquad (f \in \mathbb{R}^{\mathscr{X}}), \tag{2.3}$$

$$\langle f, g \rangle = \sum_{x \in \mathscr{X}} f(x)g(x) \qquad (f, g \in \mathbb{R}^{\mathscr{X}}). \tag{2.4}$$

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

The *support* of a function $f\colon \mathscr{X} \to \mathbb{R}$ is defined as the set $\operatorname{supp} f = \{x \in \mathscr{X} : f(x) \neq 0\}$. For a real number λ and subsets $F, G \subseteq \mathbb{R}^{\mathscr{X}}$, we use the standard notation $\lambda F = \{\lambda f : f \in F\}$ and $F + G = \{f + g : f \in F, g \in G\}$. Clearly, λF and $F + G$ are convex whenever F and G are convex. More generally, we adopt the shorthand $\lambda_1 F_1 + \lambda_2 F_2 + \cdots + \lambda_k F_k = \{\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_k f_k : f_1 \in F_1, f_2 \in F_2, \ldots, f_k \in F_k\}$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are reals and $F_1, F_2, \ldots, F_k \subseteq \mathbb{R}^{\mathscr{X}}$. A *conical combination* of $f_1, f_2, \ldots, f_k \in \mathbb{R}^{\mathscr{X}}$ is any function of the form $\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_k f_k$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are nonnegative. A *convex combination* of $f_1, f_2, \ldots, f_k \in \mathbb{R}^{\mathscr{X}}$ is any function of the form $\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_k f_k$, where $\lambda_1, \lambda_2, \dots, \lambda_k$ are nonnegative and additionally sum to 1. The *convex hull* of $F \subseteq \mathbb{R}^{\mathscr{X}}$, denoted conv F, is the set of all convex combinations of functions in F.

2.2. Matrices. For a set $\mathscr{X}$ such as $\mathscr{X} = \{0, 1\}$ or $\mathscr{X} = \mathbb{R}$, the symbol $\mathscr{X}^{n\times m}$ denotes the family of $n \times m$ matrices with entries in $\mathscr{X}$. The symbol $\mathscr{X}^{n\times *}$ denotes the family of matrices that have n rows and entries in $\mathscr{X}$, and analogously $\mathscr{X}^{*\times m}$ denotes matrices with m columns and entries in $\mathscr{X}$. The notation (2.2)–(2.4) applies to any real matrices: $\|A\|_\infty = \max |A_{i,j}|$, $\|A\|_1 = \sum_{i,j} |A_{i,j}|$, and $\langle A, B\rangle = \sum_{i,j} A_{i,j} B_{i,j}$. For a matrix $A = [A_{i,j}]$ of size $n \times m$ and a permutation $\sigma \in S_m$, we let $\sigma A = [A_{i,\sigma(j)}]_{i,j}$ denote the result of permuting the columns of A according to σ. The notation $A =_{\circlearrowleft} B$ means that the matrices A, B are the same up to a permutation of columns, i.e., $A = \sigma B$ for some permutation σ. A *submatrix of* A is a matrix obtained from A by discarding zero or more rows and zero or more columns, keeping unchanged the relative ordering of the remaining rows and columns. For a Boolean matrix $A \in \{0, 1\}^{n\times m}$ and a string $x \in \{0, 1\}^m$, we let $A|_x$ denote the submatrix of A obtained by removing those columns i for which $x_i = 0$:

$$A|_x = \begin{bmatrix} A_{1,i_1} & A_{1,i_2} & \cdots & A_{1,i_{|x|}} \\ A_{2,i_1} & A_{2,i_2} & \cdots & A_{2,i_{|x|}} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,i_1} & A_{n,i_2} & \cdots & A_{n,i_{|x|}} \end{bmatrix},$$

where $i_1 < i_2 < \cdots < i_{|x|}$ are the distinct indices such that $x_{i_1} = x_{i_2} = \cdots = x_{i_{|x|}} = 1$. By convention, $A|_{0^m} = \varepsilon$. The notation $B \sqsubseteq A$ means that

$$B = \begin{bmatrix} A_{i_1,j_1} & A_{i_1,j_2} & \cdots & A_{i_1,j_m} \\ A_{i_2,j_1} & A_{i_2,j_2} & \cdots & A_{i_2,j_m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{i_n,j_1} & A_{i_n,j_2} & \cdots & A_{i_n,j_m} \end{bmatrix}$$

for some row indices $i_1 < i_2 < \cdots < i_n$ and some distinct column indices $j_1, j_2, \dots, j_m$, where $n \times m$ is the dimension of B. In other words, $B \sqsubseteq A$ means that B is a submatrix of A, up to a permutation of columns.

We use lowercase letters (a, b, u, v, w, x, y, z) for row vectors and Boolean strings, and uppercase letters (A, B, M, X, Y) for real and Boolean matrices. The convention of using lowercase letters for *row vectors* is somewhat unusual, and for that reason we emphasize it. We identify Boolean strings with corresponding row vectors, e.g., the string 00111 is used interchangeably with the row vector $\begin{bmatrix} 0 & 0 & 1 & 1 & 1 \end{bmatrix}$. Similarly, $111\dots1$ refers to an all-ones row, and 0^m1^m refers to the row vector whose $2m$ components are m zeroes followed by m ones. On occasion, we will use bracket notation to emphasize that the string should be interpreted as a row vector, e.g., $[0^m1^m]$. We use standard matrix-theoretic notation to typeset block matrices, e.g.,

$$\begin{bmatrix} A^{00} & A^{01} & A^{10} & A^{11} \end{bmatrix}, \begin{bmatrix} A \\ 111\dots1 \end{bmatrix}, \begin{bmatrix} B \\ b \\ b' \end{bmatrix}.$$

Here the first matrix is composed of four blocks, the second matrix is obtained by adding an all-ones row to A, and the third matrix is obtained by adding the row vectors b and b' to B. When warranted, we will use vertical and horizontal lines as in (4.13) to emphasize block structure.

The *set disjointness function* DISJ on Boolean matrices X is defined by

$$\mathrm{DISJ}(X) = \begin{cases} +1 & \text{if } X \text{ contains an all-ones column,} \\ -1 & \text{otherwise.} \end{cases}$$

In particular, $\mathrm{DISJ}^{-1}(+1)$ is the family of all Boolean matrices with an all-ones column. By convention, $\mathrm{DISJ}(\varepsilon) = -1$. Note that

$$\mathrm{DISJ}\begin{bmatrix} X \\ x \end{bmatrix} = \mathrm{DISJ}(X|_x)$$

for any matrix $X \in \{0,1\}^{n\times m}$ and any row vector $x \in \{0,1\}^m$. We let $\mathrm{DISJ}_{k,n}\colon \{0,1\}^{k\times n} \to \{-1,+1\}$ be the restriction of DISJ to matrices of size $k \times n$. In Boolean notation,

$$\mathrm{DISJ}_{k,n}(X) = \bigwedge_{j=1}^{n} \bigvee_{i=1}^{k} \overline{X_{i,j}}. \tag{2.5}$$

The partial function $\mathrm{UDISJ}_{k,n}$ on $\{0,1\}^{k\times n}$, called *unique set disjointness*, is defined as the restriction of $\mathrm{DISJ}_{k,n}$ to $k \times n$ Boolean matrices with at most one column consisting entirely of ones. In other words,

$$\mathrm{UDISJ}_{k,n}(X) = \begin{cases} \mathrm{DISJ}_{k,n}(X) & \text{if } |x_1 \wedge x_2 \wedge \cdots \wedge x_k| \leqslant 1, \\ \text{undefined} & \text{otherwise,} \end{cases} \tag{2.6}$$

where $x_1, x_2, \ldots, x_k$ are the rows of X. As usual, $\mathrm{DISJ}^*_{k,n}$ and $\mathrm{UDISJ}^*_{k,n}$ denote the corresponding Boolean predicates, given by $\mathrm{DISJ}^*_{k,n} = (1-\mathrm{DISJ}_{k,n})/2$ and $\mathrm{UDISJ}^*_{k,n} = (1-\mathrm{UDISJ}_{k,n})/2$.

2.3. Probability. We view probability distributions first and foremost as real functions. This makes available various notational devices introduced above. In particular, for probability distributions μ and λ, the symbol $\operatorname{supp}\mu$ denotes the support of μ, and $\mu\otimes\lambda$ denotes the probability distribution given by $(\mu\otimes\lambda)(x,y) = \mu(x)\lambda(y)$. We define $\mu\times\lambda = \mu\otimes\lambda$, the former notation being more standard for probability distributions. The *Hellinger distance* between probability distributions μ and λ on a finite set $\mathscr{X}$ is given by

$$H(\mu,\lambda) = \left(\frac{1}{2}\sum_{x\in\mathscr{X}} (\sqrt{\mu(x)} - \sqrt{\lambda(x)})^2\right)^{1/2} = \left(1 - \sum_{x\in\mathscr{X}} \sqrt{\mu(x)\lambda(x)}\right)^{1/2}.$$

The term "distance" is used here in the proper sense, i.e., H is a metric. The *statistical distance* between μ and λ is defined to be $\frac{1}{2}\|\mu-\lambda\|_1$. The Hellinger distance between two random variables taking values in the same finite set $\mathscr{X}$ is defined to be the Hellinger distance between their respective probability distributions. Analogously, one defines the statistical distance between two random variables. The following classical fact [36, 42] gives basic properties of Hellinger distance and relates it to statistical distance.

FACT 2.1. *For any probability distributions $\mu, \mu_1, \mu_2, \ldots, \mu_n$ and $\lambda, \lambda_1, \lambda_2, \ldots, \lambda_n$,*

(i) $0 \leqslant H(\mu,\lambda) \leqslant 1$,
(ii) $2H(\mu,\lambda)^2 \leqslant \|\mu-\lambda\|_1 \leqslant 2\sqrt{2}H(\mu,\lambda)$,
(iii) $H(\mu_1\otimes\cdots\otimes\mu_n, \lambda_1\otimes\cdots\otimes\lambda_n) \leqslant \sqrt{H(\mu_1,\lambda_1)^2 + \cdots + H(\mu_n,\lambda_n)^2}$.

Parts (ii) and (iii) give a useful technique for bounding the statistical distance between product distributions and have seen several uses in the complexity literature; cf. [43, 9]. For the reader's convenience, we include a proof of this classical result.

Proof. Part (i) is immediate from the defining equations for Hellinger distance. For (ii), we have

$$\begin{aligned}
2H(\mu,\lambda)^2 &= \sum_{x\in\mathscr{X}} (\sqrt{\mu(x)} - \sqrt{\lambda(x)})^2 \\
&\leqslant \sum_{x\in\mathscr{X}} |\sqrt{\mu(x)} - \sqrt{\lambda(x)}|(\sqrt{\mu(x)} + \sqrt{\lambda(x)}) \\
&= \|\mu - \lambda\|_1,
\end{aligned}$$

and in the reverse direction

$$\begin{aligned}
\|\mu - \lambda\|_1 &= \sum_{x\in\mathscr{X}} |\sqrt{\mu(x)} - \sqrt{\lambda(x)}|(\sqrt{\mu(x)} + \sqrt{\lambda(x)}) \\
&\leqslant \left(\sum_{x\in\mathscr{X}} (\sqrt{\mu(x)} - \sqrt{\lambda(x)})^2\right)^{1/2} \left(\sum_{x\in\mathscr{X}} (\sqrt{\mu(x)} + \sqrt{\lambda(x)})^2\right)^{1/2} \\
&= \sqrt{2}H(\mu,\lambda)\left(\sum_{x\in\mathscr{X}} (\sqrt{\mu(x)} + \sqrt{\lambda(x)})^2\right)^{1/2} \\
&= 2H(\mu,\lambda)\left(1 + \sum_{x\in\mathscr{X}} \sqrt{\mu(x)\lambda(x)}\right)^{1/2} \\
&= 2H(\mu,\lambda)\sqrt{2 - H(\mu,\lambda)^2} \\
&\leqslant 2\sqrt{2}H(\mu,\lambda).
\end{aligned}$$

For (iii), let $\mathscr{X}_i$ denote the domain of μ_i and λ_i. Then

$$\begin{aligned}
&H(\mu_1 \otimes \cdots \otimes \mu_n, \lambda_1 \otimes \cdots \otimes \lambda_n)^2 \\
&\qquad = 1 - \sum_{x_1\in\mathscr{X}_1} \cdots \sum_{x_n\in\mathscr{X}_n} \sqrt{\mu_1(x_1)\cdots\mu_n(x_n)\lambda_1(x_1)\cdots\lambda_n(x_n)} \\
&\qquad = 1 - \prod_{i=1}^{n}\left(\sum_{x_i\in\mathscr{X}_i} \sqrt{\mu_i(x_i)\lambda_i(x_i)}\right) \\
&\qquad = 1 - \prod_{i=1}^{n}(1 - H(\mu_i,\lambda_i)^2) \\
&\qquad \leqslant \sum_{i=1}^{n} H(\mu_i,\lambda_i)^2,
\end{aligned}$$

where the final step uses (i). ∎

The set membership symbol $\in$, when used in the subscript of an expectation operator, means that the expectation is taken over a uniformly random element of the indicated set.

2.4. Fourier transform. Consider the real vector space of functions $\{0,1\}^n \to \mathbb{R}$. For $S \subseteq \{1,2,\dots,n\}$, define $\chi_S\colon \{0,1\}^n \to \{-1,+1\}$ by $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$. Then every function $f\colon \{0,1\}^n \to \mathbb{R}$ has a unique representation of the form

$$f = \sum_{S\subseteq\{1,2,\dots,n\}} \hat{f}(S)\,\chi_S,$$

where $\hat{f}(S) = 2^{-n}\sum_{x\in\{0,1\}^n} f(x)\chi_S(x)$. The reals $\hat{f}(S)$ are called the *Fourier coefficients of f*. Formally, the Fourier transform is the linear transformation $f \mapsto \hat{f}$, where $\hat{f}$ is viewed as a function on the power set of $\{1,2,\dots,n\}$. This makes available the shorthands

$$\|\hat{f}\|_1 = \sum_{S\subseteq\{1,2,\dots,n\}} |\hat{f}(S)|, \qquad\qquad \|\hat{f}\|_\infty = \max_{S\subseteq\{1,2,\dots,n\}} |\hat{f}(S)|.$$

PROPOSITION 2.2. *For all functions* $f,g\colon \{0,1\}^n \to \mathbb{R}$,

(i) $\|\hat{f}\|_\infty \leqslant 2^{-n}\|f\|_1$,
(ii) $\|\hat{f}\|_1 \leqslant \|f\|_1$,
(iii) $\|\widehat{f+g}\|_1 \leqslant \|\hat{f}\|_1 + \|\hat{g}\|_1$,
(iv) $\|\widehat{f\cdot g}\|_1 \leqslant \|\hat{f}\|_1\|\hat{g}\|_1$.

Proof. Item (i) is immediate by definition, and (ii) follows directly from (i). Item (iii) is trivial. The submultiplicativity (iv) can be verified as follows:

$$\begin{aligned}
\|\widehat{f\cdot g}\|_1 &= \sum_{S\subseteq\{1,2,\dots,n\}} |\widehat{f\cdot g}(S)| \\
&= \sum_{S\subseteq\{1,2,\dots,n\}} \left| \sum_{T\subseteq\{1,2,\dots,n\}} \hat{f}(T)\hat{g}(S\oplus T)\right| \\
&\leqslant \sum_{S\subseteq\{1,2,\dots,n\}} \sum_{T\subseteq\{1,2,\dots,n\}} |\hat{f}(T)|\,|\hat{g}(S\oplus T)| \\
&= \|\hat{f}\|_1\|\hat{g}\|_1,
\end{aligned}$$

where $S\oplus T = (S\cap\overline{T})\cup(\overline{S}\cap T)$ denotes the symmetric difference of sets. ∎

The *convolution* of $f,g\colon \{0,1\}^n \to \mathbb{R}$ is the function $f * g\colon \{0,1\}^n \to \mathbb{R}$ given by

$$(f*g)(x) = \sum_{y\in\{0,1\}^n} f(y)g(x\oplus y).$$

Some papers define convolution using an additional normalizing factor of 2^{-n}, but the above definition is more classical and better serves our needs. The Fourier spectrum of the convolution is given by

$$\widehat{f*g}(S) = 2^n\hat{f}(S)\hat{g}(S), \qquad\qquad S\subseteq\{1,2,\dots,n\}.$$

In particular, convolution is a symmetric operation: $f*g = g*f$. It also follows that convolving f with the function $2^{-n}\sum_{|S|\geqslant d}\chi_S$ is tantamount to discarding the Fourier coefficients of f of order less than d:

$$\left(2^{-n}\sum_{|S|\geqslant d}\chi_S\right) * f = \sum_{|S|\geqslant d}\hat{f}(S)\chi_S. \tag{2.7}$$

For any given $f\colon\{0,1\}^n \to \mathbb{R}$, it is straightforward to verify the existence and uniqueness of a multilinear real polynomial $\tilde{f}\colon\mathbb{R}^n \to \mathbb{R}$ such that $f \equiv \tilde{f}$ on $\{0,1\}^n$. Following standard practice, we will identify f with its multilinear extension $\tilde{f}$ to $\mathbb{R}^n$. In particular, we define $\deg f = \deg \tilde{f}$. The polynomial $\tilde{f}$ can be read off from the Fourier expansion of f, with the useful consequence that $\deg f = \max\{|S| : \hat{f}(S) \neq 0\}$.

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

$$\left.\begin{aligned} &|f(x) - p(x)| \leqslant \epsilon, && x \in \operatorname{dom} f,\\ &|p(x)| \leqslant 1 + \epsilon, && x \in \mathscr{X} \setminus \operatorname{dom} f. \end{aligned}\right\} \tag{2.8}$$

For a (possibly partial) real function f on a finite subset $\mathscr{X} \subset \mathbb{R}^n$, we define $E(f,d)$ to be the least ϵ such that (2.8) holds for some polynomial of degree at most d. In this notation, $\deg_\epsilon(f) = \min\{d : E(f,d) \leqslant \epsilon\}$. When f is a total function, $E(f,d)$ is simply the least error to which f can be approximated by a real polynomial of degree no greater than d. We will need the following dual characterization of the approximate degree.

FACT 2.3. *Let f be a (possibly partial) real function on $\{0,1\}^n$. Then $\deg_\epsilon(f) > d$ if and only if there exists $\psi\colon\{0,1\}^n \to \mathbb{R}$ such that*

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

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

Fact 2.3 follows from linear programming duality; see [47, 49] for details. A related notion is that of *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 [40] and the second due to Nisan and Szegedy [41].

THEOREM 2.4 (Minsky and Papert). *The function $\mathrm{MP}_n(x) = \bigvee_{i=1}^n \bigwedge_{j=1}^{4n^2} x_{ij}$ obeys*

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

THEOREM 2.5 (Nisan and Szegedy). *The functions AND_n and $\widetilde{\mathrm{AND}}_n$ obey*

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

2.6. Multiparty communication. An excellent reference on communication complexity is the monograph by Kushilevitz and Nisan [35]. In this overview, we will limit ourselves to key definitions and notation. The main model of communication of interest to us is the randomized multiparty number-on-the-forehead model, due to Chandra, Furst, and Lipton [18]. Here one considers a (possibly partial) Boolean function F on $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$, for some finite sets $\mathscr{X}_1, \mathscr{X}_2, \ldots, \mathscr{X}_k$. There are k parties. A given input $(x_1, x_2, \ldots, x_k) \in \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$ 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 also 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. 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 in the worst case. The *ϵ-error randomized communication complexity* of F, denoted $R_\epsilon(F)$, is the least cost of an ϵ-error communication protocol for F in this model. 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.

The *nondeterministic* model is similar in some ways and different in others from the randomized model. As in the randomized model, one considers a (possibly partial) Boolean function F on $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$, for some finite sets $\mathscr{X}_1, \mathscr{X}_2, \ldots, \mathscr{X}_k$. An input $(x_1, x_2, \ldots, x_k) \in \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$ 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 $F(x_1, x_2, \ldots, x_k) = +1$. As usual, the protocol is allowed to behave arbitrarily on inputs outside the domain of F. 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 *Merlin-Arthur* model [3, 6] 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 now 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.

In much of this paper, the input to a k-party communication problem will be an ordered sequence of matrices $X_1, X_2, \ldots, X_n \in \{0,1\}^{k,*}$, with the understanding that the ith party sees rows $1, \ldots, i-1, i+1, \ldots, k$ of every matrix. The main communication problem of interest to us is the *k-party set disjointness problem* $\mathrm{DISJ}_{k,n}$, defined in (2.5). In words, the goal in the set disjointness problem is to determine whether a given $k \times n$ Boolean matrix contains an all-ones column, where the ith party sees the entire matrix except for the ith row. We will also consider the k-party communication problem $\mathrm{UDISJ}_{k,n}$ called *unique set disjointness*, given by (2.6). Observe that $\mathrm{UDISJ}_{k,n}$ is a promise version of set disjointness, the promise being that the input matrix has at most one column consisting entirely of ones.

A common operation in this paper is that of composing functions to obtain communication problems. Specifically, let G be a (possibly partial) Boolean function on $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$, representing a k-party communication problem, and let f be a (possibly partial) Boolean function on $\{0,1\}^n$. We view the composition $f \circ G$ as a k-party communication problem on $\mathscr{X}_1^n \times \mathscr{X}_2^n \times \cdots \times \mathscr{X}_k^n$. With these conventions, one has

$$\begin{aligned} \mathrm{DISJ}_{k,rs} &= \mathrm{AND}_r \circ \mathrm{DISJ}_{k,s}, \\ \mathrm{UDISJ}_{k,rs} &= \widetilde{\mathrm{AND}}_r \circ \mathrm{UDISJ}_{k,s} \end{aligned}$$

for all positive integers r, s.

2.7. Discrepancy and generalized discrepancy. A k-dimensional *cylinder intersection* is a function $\chi\colon \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k \to \{0,1\}$ 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 $\chi_i\colon \mathscr{X}_1 \times \cdots \times \mathscr{X}_{i-1} \times \mathscr{X}_{i+1} \times \cdots \times \mathscr{X}_k \to \{0,1\}$. 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. In particular, a one-dimensional cylinder intersection is one of the two constant functions $0, 1$. Cylinder intersections were introduced by Babai, Nisan, and Szegedy [7] and play a fundamental role in the theory due to the following fact.

FACT 2.6. *Let $\Pi\colon \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k \to \{-1,+1\}$ be a deterministic k-party communication protocol with cost r. Then*

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

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

Since a randomized protocol with cost r is a probability distribution on deterministic protocols of cost r, Fact 2.6 implies the following two results on randomized communication complexity.

COROLLARY 2.7. *Let F be a (possibly partial) Boolean function on $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$. If $R_\epsilon(F) = r$, then*

$$\begin{aligned} |F(X) - \Pi(X)| &\leqslant \frac{\epsilon}{1-\epsilon}, && X \in \operatorname{dom} F, \\ |\Pi(X)| &\leqslant \frac{1}{1-\epsilon}, && X \in \mathscr{X}_1 \times \cdots \times \mathscr{X}_k, \end{aligned}$$

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

COROLLARY 2.8. *Let Π be a randomized k-party protocol with domain $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$. If Π has communication cost r bits, then*

$$\mathbf{P}[\Pi(X) = -1] \equiv \sum_\chi a_\chi \chi(X), \qquad X \in \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k,$$

where the sum is over cylinder intersections and $\sum_\chi |a_\chi| \leqslant 2^r$.

For a (possibly partial) Boolean function F on $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$ and a probability distribution P on $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$, the *discrepancy of F with respect to P* is given by

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

where the maximum is over cylinder intersections. The least discrepancy over all distributions is denoted $\operatorname{disc}(F) = \min_P \operatorname{disc}_P(F)$. As Fact 2.6 suggests, upper bounds on the discrepancy give lower bounds on communication complexity. This technique is known as the *discrepancy method* [21, 7, 35]:

THEOREM 2.9 (Discrepancy method). *Let F be a (possibly partial) Boolean function on $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$. Then*

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

A more general technique, originally applied by Klauck [32] in the two-party quantum model and subsequently adapted to many other settings [45, 39, 47, 38, 20], is the *generalized discrepancy method*.

THEOREM 2.10 (Generalized discrepancy method). *Let F be a (possibly partial) Boolean function on $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$. Then for every nonzero $\Psi\colon \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k \to \mathbb{R}$,*

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

where the maximum is over cylinder intersections χ.

Complete proofs of Theorems 2.9 and 2.10 can be found in [48, Theorems 2.9, 2.10]. The ideas of the generalized discrepancy method have been adapted to nondeterministic and Merlin-Arthur communication. The following result [23, Theorem 4.1] gives a criterion for high communication complexity in these models.

THEOREM 2.11 (Gavinsky and Sherstov). *Let F be a (possibly partial) k-party communication problem on $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$. Fix a function $H\colon \mathscr{X} \to \{-1,+1\}$ and a probability distribution P on* dom F. *Put*

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

Then

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

Theorem 2.11 was stated in [23] for total functions F, but the proof in that paper applies to partial functions as well.

3. Directional derivatives and approximation

Directional derivatives are meaningful for any function on the Boolean hypercube with values in a ring R. The directional derivative of $f\colon\{0,1\}^n \to R$ in the direction $S \subseteq \{1,2,\ldots,n\}$ is usually defined as the function $(\partial f/\partial S)(x) = f(x) - f(x \oplus \mathbf{1}_S)$. Directional derivatives of higher order are obtained by differentiating more than once. As a special case, partial derivatives are given by $(\partial f/\partial\{i\})(x) = f(x) - f(x \oplus e_i)$. Directional derivatives have been studied mostly for the field $R = \mathbb{F}_2$, motivated by applications to circuit complexity and cryptography [53, 1, 25, 26, 52, 22]. In particular, the uniformity norm U^d of Gowers [25, 26] is defined in terms of the expected value of a randomly chosen order-d directional derivative for $R = \mathbb{F}_2$. To a lesser extent, directional derivatives have been studied for R a finite field [24] and the field of reals [13]. In this work, derivatives serve the purpose of determining how well a given function $f\colon\{0,1\}^n \to \mathbb{R}$ can be approximated by a polynomial $p \in \mathbb{R}[x_1, x_2, \ldots, x_n]$ of given degree d. Consequently, we work with the field $R = \mathbb{R}$.

3.1. Definition and basic properties. Let d be a positive integer. For a given function $f\colon\{0,1\}^n \to \mathbb{R}$ and sets $S_1, S_2, \ldots, S_d \subseteq \{1,2,\ldots,n\}$, we define the *directional derivative of f with respect to $S_1, S_2, \ldots, S_d$* to be the function $\partial^d f/\partial S_1\,\partial S_2 \cdots \partial S_d\colon\{0,1\}^n \to \mathbb{R}$ given by

$$\frac{\partial^d f}{\partial S_1\,\partial S_2\,\cdots\,\partial S_d}(x) = \mathop{\mathbf{E}}_{z\in\{0,1\}^d}\left[(-1)^{|z|} f\left(x \oplus \bigoplus_{i=1}^{d} z_i \mathbf{1}_{S_i}\right)\right]. \tag{3.1}$$

The *order* of the directional derivative is the number of sets involved. Thus, (3.1) is a directional derivative of order d. We collect basic properties of directional derivatives in the following proposition.

Proposition 3.1 (Folklore). *Let $f\colon\{0,1\}^n \to \mathbb{R}$ be a given function, $S_1, S_2, \ldots, S_d \subseteq \{1,2,\ldots,n\}$ given sets, and $\sigma\colon\{1,2,\ldots,d\} \to \{1,2,\ldots,d\}$ a permutation. Then*

(i) *$\partial^d/\partial S_1\partial S_2\cdots\partial S_d$ is a linear transformation of $\mathbb{R}^{\{0,1\}^n}$ into itself;*
(ii) *$\partial^d f/\partial S_1\partial S_2\cdots\partial S_d \equiv \partial(\partial^{d-1} f/\partial S_1\partial S_2\cdots\partial S_{d-1})/\partial S_d$;*
(iii) *$\partial^d f/\partial S_1\partial S_2\cdots\partial S_d \equiv \partial^d f/\partial S_{\sigma(1)}\partial S_{\sigma(2)}\cdots\partial S_{\sigma(d)}$;*
(iv) *$\|\partial^d f/\partial S_1\partial S_2\cdots\partial S_d\|_\infty \leqslant \|f\|_\infty$;*
(v) *$\partial^d f/\partial S_1\partial S_2\cdots\partial S_d \equiv 0$ whenever $S_i = \varnothing$ for some i;*
(vi) *$\partial^d f/\partial S_1\partial S_2\cdots\partial S_d \equiv 0$ whenever $S_1, S_2, \ldots, S_d$ are pairwise disjoint and $\deg f \leqslant d-1$.*

Proof. Items (i)–(iv) follow immediately from the definition. Since $\partial f/\partial\varnothing \equiv 0$ for any function f, item (v) follows directly from (ii) and (iii). To prove (vi), we may assume by (i) that $f = \chi_T$ with $|T| \leqslant d-1$. For such f, observe that $\partial f/\partial S_i \equiv 0$ whenever $T \cap S_i = \varnothing$. Since $|T| \leqslant d-1$ and $S_1, S_2, \ldots, S_d$ are pairwise disjoint, we have $T \cap S_i = \varnothing$ for some i, thus forcing $\partial f/\partial S_i \equiv 0$. That $\partial^d f/\partial S_1\,\partial S_2\,\cdots\,\partial S_d \equiv 0$ now follows from (ii) and (iii). ∎

Item (vi) in Proposition 3.1 provides intuition for why directional derivatives might be relevant in characterizing the least error in an approximation of f by a real polynomial of given degree. This intuition will be borne out at the end of Section 3. The disjointness assumption in Proposition 3.1(vi) cannot be removed, even when $\mathbf{1}_{S_1}, \mathbf{1}_{S_2}, \ldots, \mathbf{1}_{S_d}$ are linearly independent as vectors in $\mathbb{F}_2^n$. For example, $\partial^2 x_1/\partial\{1,2\}\,\partial\{1,3\} = x_1 - \frac{1}{2} \not\equiv 0$.

We now define the key complexity measure in our study.

DEFINITION 3.2. Let $f\colon\{0,1\}^n \to \mathbb{R}$ be a given function. For $d = 1, 2, \ldots, n$, define

$$\Delta(f, d) = \max_{S_1,\ldots,S_d} \left\| \frac{\partial^d f}{\partial S_1\, \partial S_2\, \cdots\, \partial S_d} \right\|_\infty,$$

where the maximum is over nonempty pairwise disjoint sets $S_1, S_2, \ldots, S_d \subseteq \{1, 2, \ldots, n\}$. Define $\Delta(f, n+1) = \Delta(f, n+2) = \cdots = 0$.

It is helpful to think of $\Delta(f, d)$ as a measure of smoothness. Our ultimate goal is to understand how this complexity measure relates to the approximation of f by polynomials. As a first step in that direction, we have:

THEOREM 3.3. *For all functions $f\colon\{0,1\}^n \to \mathbb{R}$ and all $d = 1, 2, \ldots, n$,*

$$E(f, d-1) \geqslant \Delta(f, d).$$

Furthermore,

$$E(\mathrm{AND}_n^*, d-1) \geqslant 2^{d(1-O(\frac{d}{n}))-1} \Delta(\mathrm{AND}_n^*, d).$$

Proof. Write $f = p + \xi$, where p is a polynomial of degree at most $d-1$ and $\|\xi\|_\infty \leqslant E(f, d-1)$. Then

$$\begin{aligned} \Delta(f, d) &\leqslant \Delta(p, d) + \Delta(\xi, d) \\ &= \Delta(\xi, d) && \text{by Proposition 3.1(vi)} \\ &\leqslant E(f, d-1) && \text{by Proposition 3.1(iv).} \end{aligned}$$

To prove the second part, note that $\Delta(\mathrm{AND}_n^*, d) = 2^{-d}$ since AND_n^* is supported on exactly one point and takes on 1 at that point. At the same time, Buhrman et al. [15] show that $E(\mathrm{AND}_n^*, d-1) \geqslant 2^{-1-\Theta(d^2/n)}$. ∎

Thus, $\Delta(f, d)$ is always a lower bound on the least error in an approximation of f by a polynomial of degree less than d, and the gap between the two quantities can be considerable. Our challenge is to prove a partial converse to this result. Specifically, we will be able to show that

$$\begin{aligned} E(f, d-1) \leqslant K^d \Delta(f, d) + K^{d+1} \Delta(f, d+1) + \cdots \\ + K^{d+i} \Delta(f, d+i) + \cdots, \end{aligned} \tag{3.2}$$

where $K > 2$ is an absolute constant.

3.2. Elementary dual functions. The proof of (3.2) requires considerable preparatory work. Basic building blocks in it are the linear functionals to which partial derivatives correspond. We start with a formal definition of these fundamental objects.

DEFINITION 3.4 (Elementary dual function). For a string $w \in \{0,1\}^n$ and nonempty pairwise disjoint subsets $S_1, \ldots, S_d \subseteq \{1, 2, \ldots, n\}$, let $\psi_{w,S_1,\ldots,S_d}\colon\{0,1\}^n \to \mathbb{R}$ be the function that has support

$$\operatorname{supp} \psi_{w,S_1,\ldots,S_d} = \left\{ w \oplus \bigoplus_{i=1}^{d} z_i \mathbf{1}_{S_i} : z \in \{0,1\}^d \right\}$$

and is defined on that support by

$$\psi_{w,S_1,\dots,S_d}\left(w \oplus \bigoplus_{i=1}^{d} z_i \mathbf{1}_{S_i}\right) = \frac{(-1)^{|z|}}{2^d}, \qquad z \in \{0,1\}^d.$$

An *elementary dual function of degree* d is any of the functions $\psi_{w,S_1,\dots,S_d}$, where $w \in \{0,1\}^n$ and $S_1,\dots,S_d \subseteq \{1,2,\dots,n\}$ are nonempty pairwise disjoint sets.

An elementary dual function can be written in several ways using the above notation. For example, $\psi_{w,S_1,\dots,S_d} \equiv \psi_{w,S_{\sigma(1)},\dots,S_{\sigma(d)}}$ for any permutation σ on $\{1,2,\dots,d\}$. One also has $\psi_{w,S,T} \equiv \psi_{w\oplus \mathbf{1}_S \oplus \mathbf{1}_T,S,T}$ and more generally $\psi_{w\oplus z_1\mathbf{1}_{S_1}\oplus\cdots\oplus z_d\mathbf{1}_{S_d},S_1,\dots,S_d} = (-1)^{|z|}\psi_{w,S_1,\dots,S_d}$. We now establish key properties of elementary dual functions, motivating the term itself and relating it to directional derivatives.

THEOREM 3.5 (On elementary dual functions).

(i) *For every* $f\colon\{0,1\}^n \to \mathbb{R}$, *one has* $\langle f, \psi_{w,S_1,\dots,S_d}\rangle = (\partial^d f/\partial S_1 \partial S_2 \cdots \partial S_d)(w)$.

(ii) *The negation of a degree-d elementary dual function is a degree-d elementary dual function.*

(iii) *If p is a polynomial of degree less than d, then* $\langle \psi_{w,S_1,\dots,S_d}, p\rangle = 0$. *Equivalently,* $\psi_{w,S_1,\dots,S_d} \in \operatorname{span}\{\chi_S : |S| \geqslant d\}$.

(iv) *Every χ_S with $|S| \geqslant d$ is the sum of 2^n elementary dual functions of degree d. In particular, every function in* $\operatorname{span}\{\chi_S : |S| \geqslant d\}$ *is a linear combination of degree-d elementary dual functions.*

(v) *For every function $f\colon\{0,1\}^n \to \mathbb{R}$ and $d = 1,2,\dots,n$, one has* $\Delta(f,d) = \max\langle f, \psi_{w,S_1,\dots,S_d}\rangle = \max|\langle f, \psi_{w,S_1,\dots,S_d}\rangle|$, *where the maximum is taken over degree-d elementary dual functions $\psi_{w,S_1,\dots,S_d}$.*

Proof. Item (i) is immediate from the definitions, and (ii) follows from $-\psi_{w,S_1,\dots,S_d} = \psi_{w\oplus\mathbf{1}_{S_1},S_1,\dots,S_d}$. Item (iii) follows from (i) and Proposition 3.1(vi).

For (iv), it suffices by symmetry to consider $\chi_{\{1,2,\dots,D\}}$ for $D = d, d+1, \dots, n$. For every $u \in \{0,1\}^{n-d}$,

$$\psi_{0^d u,\{1\},\dots,\{d\}}(x) = \begin{cases} 2^{-d}\chi_{\{1,\dots,d\}}(x) & \text{if } (x_{d+1},\dots,x_n) = u, \\ 0 & \text{otherwise.}\end{cases}$$

Therefore,

$$\chi_{\{1,\dots,D\}}(x) = 2^d \sum_{u\in\{0,1\}^{n-d}} (-1)^{u_1+\cdots+u_{D-d}} \psi_{0^d u,\{1\},\dots,\{d\}}(x).$$

By (ii), each of the functions in the final summation is a degree-d elementary dual function, so that $\chi_{\{1,\dots,D\}}$ is indeed the sum of 2^n elementary dual functions of degree d.

Finally, (v) is immediate from (i) and (ii). ∎

DEFINITION. For $d = 1,\dots,n$, define $\Psi_{n,d} \subseteq \mathbb{R}^{\{0,1\}^n}$ to be the convex hull of degree-d elementary dual functions, $\Psi_{n,d} = \operatorname{conv}\{\chi_{w,S_1,\dots,S_d}\}$. Define $\Psi_{n,n+1}, \Psi_{n,n+2}, \dots \subseteq \mathbb{R}^{\{0,1\}^n}$ by $\Psi_{n,n+1} = \Psi_{n,n+2} = \cdots = \{0\}$.

By Theorem 3.5(ii), the convex sets $\Psi_{n,1}, \Psi_{n,2}, \dots, \Psi_{n,n}$ are all closed under negation and hence contain 0. As a result, we have $c\Psi_{n,d} \subseteq C\Psi_{n,d}$ for all $C \geqslant c \geqslant 0$. We will use this fact without mention throughout this section, including Lemmas 3.6, 3.10, and 3.11 and Theorems 3.7 and 3.12. The next lemma establishes useful analytic properties of $\Psi_{n,d}$.

LEMMA 3.6. *Let* $f, \psi\colon \{0,1\}^n \to \mathbb{R}$ *be given,* $d \in \{1, 2, \ldots, n\}$. *Then*

(i) $\psi \in 2^n \|\hat{\psi}\|_1 \Psi_{n,d} \subseteq 2^n \|\psi\|_1 \Psi_{n,d}$ *whenever* $\psi \in \operatorname{span}\{\chi_S : |S| \geqslant d\}$;
(ii) $(f * \psi_{w,S_1,\ldots,S_d})(x) = (\partial^d f/\partial S_1 \partial S_2 \cdots \partial S_d)(x \oplus w)$;
(iii) $\|f * \psi\|_\infty \leqslant \Delta(f,d)$ *whenever* $\psi \in \Psi_{n,d}$.

Proof. (i) Recall from Theorem 3.5(iv) that $\chi_S \in 2^n \Psi_{n,d}$ for every subset $S \subseteq \{1, 2, \ldots, n\}$ with $|S| \geqslant d$. Therefore, $\psi \in 2^n \|\hat{\psi}\|_1 \Psi_{n,d}$ by convexity. The containment $2^n \|\hat{\psi}\|_1 \Psi_{n,d} \subseteq 2^n \|\psi\|_1 \Psi_{n,d}$ is immediate from Proposition 2.2(ii).

(ii) Writing out the convolution explicitly,

$$\begin{aligned}(f * \psi_{w,S_1,\ldots,S_d})(x) &= \sum_{y \in \{0,1\}^n} f(y) \psi_{w,S_1,\ldots,S_d}(x \oplus y) \\ &= \langle f, \psi_{x \oplus w, S_1, \ldots, S_d} \rangle \\ &= \left(\frac{\partial^d f}{\partial S_1 \partial S_2 \cdots \partial S_d} \right)(x \oplus w),\end{aligned}$$

where the final step uses Theorem 3.5(i).

(iii) It is a direct consequence of (ii) that $\|f * \psi_{w,S_1,\ldots,S_d}\|_\infty \leqslant \Delta(f,d)$ for every elementary dual function $\psi_{w,S_1,\ldots,S_d}$. By convexity, (iii) follows. ∎

Recall that our goal is to establish a partial converse to Theorem 3.3, i.e., prove that functions with small derivatives can be approximated well by low-degree polynomials. To help the reader build some intuition for the proof, we illustrate our technique in a particularly simple setting. Specifically, we give a short proof that $E(f, d-1) \leqslant 2^n \Delta(f,d)$. We actually prove something stronger, namely, that every f can be approximated pointwise within $2^n \Delta(f,d)$ by its truncated Fourier polynomial $\sum_{|S| \leqslant d-1} \hat{f}(S) \chi_S$. We do so by expressing the discarded part of the Fourier spectrum,

$$\sum_{|S| \geqslant d} \hat{f}(S) \chi_S(x), \tag{3.3}$$

as a linear combination of order-d directional derivatives of f at appropriate points, where the absolute values of the coefficients in the linear combination sum to at most 2^n. Since the magnitude of an order-d derivative of f cannot exceed $\Delta(f,d)$, we arrive at the desired upper bound on the approximation error.

THEOREM 3.7. *For all functions* $f\colon \{0,1\}^n \to \mathbb{R}$ *and all* $d = 1, 2, \ldots, n$,

$$E(f, d-1) \leqslant \left\| \sum_{|S| \geqslant d} \hat{f}(S) \chi_S \right\|_\infty \leqslant 2^n \Delta(f,d).$$

Proof. Define $\psi\colon \{0,1\}^n \to \mathbb{R}$ by $\psi(x) = 2^{-n} \sum_{|S| \geqslant d} \chi_S$. Then by Lemma 3.6(i),

$$\psi \in 2^n \Psi_{n,d}. \tag{3.4}$$

As a result,

$$\left\| \sum_{|S|\geqslant d} \hat{f}(S)\chi_S \right\|_\infty = \|f * \psi\|_\infty \qquad \text{by (2.7)}$$
$$\leqslant 2^n \max_{\psi' \in \Psi_{n,d}} \|f * \psi'\|_\infty \qquad \text{by (3.4)}$$
$$\leqslant 2^n \Delta(f,d) \qquad \text{by Lemma 3.6(iii).} \qquad \square$$

Theorem 3.7 serves an illustrative purpose and is of little interest by itself. To obtain the actual result that we want, (3.2), we need to consider directional derivatives of *all* orders starting at d. Specifically, we express the discarded portion of the Fourier spectrum, (3.3), as a linear combination of directional derivatives of f of orders $i = d, d+1, \ldots, n$, where the sum of the absolute values of the coefficients of order-i derivatives is K^i for some absolute constant $K > 2$. Since an order-i derivative cannot exceed $\Delta(f,i)$, the desired bound (3.2) follows.

To find the kind of linear combination described in the previous paragraph, we will express the function $\psi = 2^{-n}\sum_{|S|\geqslant d}\chi_S$ as a linear combination of elementary dual functions of orders $d, d+1, \ldots, n$ with small coefficients. That is a nontrivial task, and it will take up the next few pages. Once we have obtained the needed representation for ψ, we will be able to prove (3.2) easily using a convolution argument, cf. Theorem 3.7.

3.3. Symmetric extensions. Consider the operation of extending a symmetric function $g\colon \{0,1\}^m \to \mathbb{R}$ to a larger domain $\{0,1\}^n$, illustrated schematically in Figure 3.1. The extended function is again symmetric, supported on $m+1$ equispaced levels of the hypercube, and normalized such that the sum on each of these levels is the same as for g. Here, we relate the metric and Fourier-theoretic properties of the original function to those of its extension.

LEMMA 3.8. *Let n, m, Δ be positive integers, $m\Delta \leqslant n$. Let $g\colon \{0,1\}^m \to \mathbb{R}$ be a given symmetric function. Consider the symmetric function $G\colon \{0,1\}^n \to \mathbb{R}$ given by*

$$G(x) = \begin{cases} \binom{n}{|x|}^{-1}\binom{m}{|x|/\Delta} g(1^{|x|/\Delta}00\ldots0) & \text{if } |x| = 0, \Delta, 2\Delta, \ldots, m\Delta, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

Then:

(i) *the Fourier coefficients of G are given by*

$$\hat{G}(S) = 2^{-n}\sum_{i=0}^{m}\binom{m}{i} g(1^i 0^{m-i}) \mathop{\mathbf{E}}_{x\in\{0,1\}^n,\,|x|=i\Delta}[\chi_S(x)];$$

(ii) $G \in \operatorname{span}\{\chi_S : |S| \geqslant d\}$ *if and only if* $g \in \operatorname{span}\{\chi_S : |S| \geqslant d\}$;

(iii) $G \in \Psi_{n,d}$ *whenever* $g \in \Psi_{m,d}$.

Proof. (i) By the symmetry of g and G,

$$\hat{G}(S) = 2^{-n}\sum_{i=0}^{n}\binom{n}{i} G(1^i 0^{n-i}) \mathop{\mathbf{E}}_{|x|=i}[\chi_S(x)]$$
$$= 2^{-n}\sum_{i=0}^{m}\binom{m}{i} g(1^i 0^{m-i}) \mathop{\mathbf{E}}_{|x|=i\Delta}[\chi_S(x)].$$

FIGURE 3.1. Extending a symmetric function from $\{0,1\}^m$ to $\{0,1\}^n$.

(ii) Since g is symmetric, $g \notin \operatorname{span}\{\chi_S : |S| \geqslant d\}$ if and only if

$$\sum_{x\in\{0,1\}^m} g(x)p(x_1+\cdots+x_m) \neq 0$$

for some univariate polynomial p of degree less than d. Analogously, $G \notin \operatorname{span}\{\chi_S : |S| \geqslant d\}$ if and only if

$$\sum_{x\in\{0,1\}^n} G(x)q(x_1+\cdots+x_n) \neq 0$$

for some univariate polynomial q of degree less than d. Finally, the definition of G ensures

$$\sum_{x\in\{0,1\}^m} g(x)p(x_1+\cdots+x_m) = \sum_{x\in\{0,1\}^n} G(x)p\left(\frac{x_1+\cdots+x_n}{\Delta}\right)$$

for every polynomial p, regardless of degree.

(iii) For nonempty pairwise disjoint subsets $T_1, T_2, \ldots, T_m \subseteq \{1,2,\ldots,n\}$, define $L_{T_1,\ldots,T_m}$ to be the linear transformation that sends a function $\phi\colon \{0,1\}^m \to \mathbb{R}$ into the function $L_{T_1,\ldots,T_m}\phi\colon \{0,1\}^n \to \mathbb{R}$ such that

$$(L_{T_1,\ldots,T_m}\phi)(u_1\mathbf{1}_{T_1} \oplus \cdots \oplus u_m\mathbf{1}_{T_m}) = \phi(u), \qquad u \in \{0,1\}^m,$$

and $(L_{T_1,\ldots,T_m}\phi)(x) = 0$ whenever $x \neq u_1\mathbf{1}_{T_1} \oplus \cdots \oplus u_m\mathbf{1}_{T_m}$ for any u. We claim that

$$G = \mathop{\mathbf{E}}_{T_1,\ldots,T_m}[L_{T_1,\ldots,T_m} g], \tag{3.6}$$

where the expectation is over pairwise disjoint subsets $T_1, T_2, \ldots, T_m \subseteq \{1,2,\ldots,n\}$ of cardinality Δ each. Indeed, the right-hand side of (3.6) is a function $\{0,1\}^n \to \mathbb{R}$ that is symmetric, sums to $\binom{m}{i}g(1^i0^{m-i})$ on inputs of Hamming weight $i\Delta$ ($i = 0,1,2,\ldots,m$), and vanishes on all other inputs. There is only one function that has these three properties, namely, the function G in the statement of the lemma.

In view of (3.6) it suffices to show that under $L_{T_1,\ldots,T_m}$, the image of an elementary dual function $\{0,1\}^m \to \mathbb{R}$ is an elementary dual function $\{0,1\}^n \to \mathbb{R}$ of the same degree. By definition, the elementary dual function $\psi_{w,S_1,\ldots,S_d}\colon \{0,1\}^m \to \mathbb{R}$ satisfies

$$\psi_{w,S_1,\ldots,S_d}(w \oplus z_1\mathbf{1}_{S_1} \oplus \cdots \oplus z_d\mathbf{1}_{S_d}) = \frac{(-1)^{|z|}}{2^d}, \qquad z \in \{0,1\}^d,$$

and vanishes on the remaining $2^m - 2^d$ points of $\{0,1\}^m$. Thus, $L_{T_1,\ldots,T_m}\psi_{w,S_1,\ldots,S_d}$ obeys

$$(L_{T_1,\ldots,T_m}\psi_{w,S_1,\ldots,S_d})\left(\bigoplus_{i=1}^{m} w_i \mathbf{1}_{T_i} \oplus \bigoplus_{i=1}^{d} z_i \mathbf{1}_{R_i}\right) = \frac{(-1)^{|z|}}{2^d}, \qquad z \in \{0,1\}^d,$$

and vanishes on the remaining $2^n - 2^d$ points of $\{0,1\}^n$, where $R_1, \ldots, R_d \subseteq \{1,2,\ldots,n\}$ are the nonempty pairwise disjoint sets $R_i = \bigcup_{j \in S_i} T_j$. Therefore, $L_{T_1,\ldots,T_m}\psi_{w,S_1,\ldots,S_d}$ is a degree-d elementary dual function. ∎

The next lemma takes as given the Fourier coefficients of the extended symmetric function and solves for the values of the original symmetric function.

LEMMA 3.9. *Let $F\colon \{0,1\}^n \to \mathbb{R}$ be a symmetric function and $m \in \{1,2,\ldots,n\}$. Then there exist reals $g_0, g_1, \ldots, g_m$ such that*

$$\sum_{i=0}^{m} g_i \mathop{\mathbf{E}}_{\substack{x \in \{0,1\}^n \\ |x| = i\lfloor n/m \rfloor}} [\chi_S(x)] = \hat{F}(S) \qquad (|S| \leqslant m), \tag{3.7}$$

$$\sum_{i=0}^{m} |g_i| \leqslant (8^m - 1)\|\hat{F}\|_\infty. \tag{3.8}$$

Proof. Abbreviate $\Delta = \lfloor n/m \rfloor$, so that $m\Delta \leqslant n \leqslant 2m\Delta$. The expectation in (3.7) depends only on the cardinality of S. As a result, it suffices to prove the lemma for $S = \varnothing, \{1\}, \{1,2\}, \{1,2,3\}, \ldots, \{1,2,\ldots,m\}$. To that end, consider the matrix

$$A = \left[\mathop{\mathbf{E}}_{|x| = i\Delta}[\chi_{\{1,2,\ldots,j\}}(x)]\right]_{j,i},$$

where $i, j = 0, 1, 2, \ldots, m$. Then the sought reals $g_0, g_1, \ldots, g_m$ are given by

$$\begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix} = A^{-1} \begin{bmatrix} \hat{F}(\varnothing) \\ \hat{F}(\{1\}) \\ \hat{F}(\{1,2\}) \\ \vdots \\ \hat{F}(\{1,2,\ldots,m\}) \end{bmatrix}$$

whenever A is nonsingular. Consequently, the proof will be complete once we show that the inverse of A exists and obeys

$$\|A^{-1}\|_1 \leqslant 8^m - 1. \tag{3.9}$$

We will calculate A^{-1} explicitly. Consider polynomials $p_0, p_1, \ldots, p_m\colon \{0,1\}^n \to \mathbb{R}$, each of degree m, given by

$$p_j(x) = \frac{(-1)^{m-j}}{m!\Delta^m}\binom{m}{j}\prod_{\substack{i=0 \\ i \neq j}}^{m}(|x| - i\Delta), \qquad j = 0, 1, \ldots, m.$$

Then

$$p_j(x) = \begin{cases} 1 & \text{if } |x| = j\Delta, \\ 0 & \text{if } |x| \in \{0, \Delta, 2\Delta, \ldots, m\Delta\} \setminus \{j\Delta\}. \end{cases}$$

It follows that

$$\begin{aligned}
\delta_{i,j} &= \mathop{\mathbf{E}}_{|x|=i\Delta}[p_j(x)] \\
&= \sum_{k=0}^{m} \hat{p}_j(\{1,2,\ldots,k\}) \mathop{\mathbf{E}}_{|x|=i\Delta}\left[\sum_{\substack{S\subseteq\{1,2,\ldots,n\}\\|S|=k}} \chi_S(x)\right] \\
&= \sum_{k=0}^{m} \hat{p}_j(\{1,2,\ldots,k\}) \binom{n}{k} \mathop{\mathbf{E}}_{|x|=i\Delta}[\chi_{\{1,\ldots,k\}}(x)] \\
&= \sum_{k=0}^{m} \hat{p}_j(\{1,2,\ldots,k\}) \binom{n}{k} A_{k,i} \qquad\qquad (i,j=0,1,\ldots,m),
\end{aligned}$$

where the second and third steps use the symmetry of p_j and the symmetry of the expectation operator, respectively. This gives the explicit form

$$A^{-1} = \left[\binom{n}{k}\hat{p}_j(\{1,2,\ldots,k\})\right]_{j,k},$$

showing in particular that A is nonsingular. It remains to prove (3.9). Applying Proposition 2.2(iii)–(iv),

$$\begin{aligned}
\|\hat{p}_j\|_1 &\leqslant \frac{1}{m!\Delta^m}\binom{m}{j}\prod_{\substack{i=0\\i\neq j}}^{m} \|\widehat{|x|-i\Delta}\|_1 \\
&= \frac{1}{m!\Delta^m}\binom{m}{j}\prod_{\substack{i=0\\i\neq j}}^{m}\left(\frac{n}{2}+\left|\frac{n}{2}-i\Delta\right|\right) \\
&\leqslant \frac{1}{m!\Delta^m}\binom{m}{j}\prod_{\substack{i=0\\i\neq j}}^{m}\left(\frac{2m\Delta}{2}+\left|\frac{2m\Delta}{2}-i\Delta\right|\right) \qquad \text{since } n\leqslant 2m\Delta \\
&= \frac{m}{2m-j}\binom{2m}{m}\binom{m}{j}.
\end{aligned}$$

Hence,

$$\|A^{-1}\|_1 = \sum_{j=0}^{m}\|\hat{p}_j\|_1 \leqslant \binom{2m}{m}\sum_{j=0}^{m}\binom{m}{j} = 2^m\binom{2m}{m} \leqslant 8^m - 1. \qquad \square$$

3.4. Bounding the global error. At this point, we have all the tools at our disposal to express $\psi = 2^{-n}\sum_{|S|\geqslant d}\chi_S$ as a linear combination of elementary dual functions of orders $d, d+1, \ldots, n$ with small coefficients. We do so by means of an iterative process that can be visualized as "chasing the bulge," to borrow the metaphor from linear algebra. Originally, the Fourier spectrum of ψ is supported on characters of degree d or higher. In the ith iteration, the smallest degree of a nonzero Fourier coefficient grows by a factor of c, and the magnitude of the nonzero Fourier coefficients grows by a factor of at most $8^{c^i d}$. In this way, each iteration pushes the Fourier spectrum further back at the expense of a controlled increase in the magnitude of the remaining coefficients, which results in

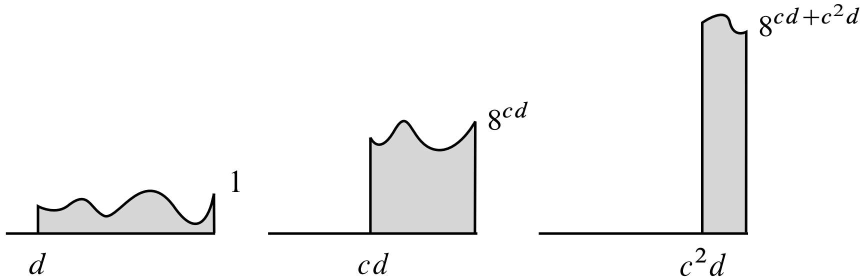

FIGURE 3.2. Chasing the bulge.

a growing "bulge" of Fourier mass on characters of high degree. This process is shown schematically in Figure 3.2. The next lemma corresponds to a single iteration.

LEMMA 3.10. *Let D be a given integer, $1 \leqslant D \leqslant n$. Let $F\colon \{0,1\}^n \to \mathbb{R}$ be a symmetric function with $F \in \operatorname{span}\{\chi_S : |S| \geqslant D\}$. Then for every integer $m \geqslant D$, there is a symmetric function $G\colon \{0,1\}^n \to \mathbb{R}$ such that*

$$G \in 2^n\, 16^m \|\hat{F}\|_\infty \Psi_{n,D}, \tag{3.10}$$

$$F - G \in \operatorname{span}\{\chi_S : |S| \geqslant m+1\}, \tag{3.11}$$

$$\|\widehat{F-G}\|_\infty \leqslant 8^m \|\hat{F}\|_\infty. \tag{3.12}$$

Proof. When $m > n$, Lemma 3.6(i) shows that $F \in 2^n \|\hat{F}\|_1 \Psi_{n,D} \subseteq 2^n\, 2^n \|\hat{F}\|_\infty \Psi_{n,D} \subseteq 2^n\, 16^m \|\hat{F}\|_\infty \Psi_{n,D}$. As a result, the lemma holds in that case with $G = F$.

In the remainder of the proof, we treat the complementary case $m \leqslant n$. Define $\Delta = \lfloor n/m \rfloor \geqslant 1$. By Lemma 3.9, there exist reals $g_0, g_1, \ldots, g_m$ that obey (3.7) and (3.8). Let $g\colon \{0,1\}^m \to \mathbb{R}$ be the symmetric function given by $g(x) = 2^n \binom{m}{|x|}^{-1} g_{|x|}$. Then (3.7) and (3.8) can be restated as

$$\hat{F}(S) = 2^{-n} \sum_{i=0}^{m} \binom{m}{i} g(1^i 0^{m-i}) \mathop{\mathbf{E}}_{\substack{x\in\{0,1\}^n \\ |x| = i\Delta}} [\chi_S(x)] \qquad (|S| \leqslant m), \tag{3.13}$$

$$\|g\|_1 \leqslant 2^n (8^m - 1) \|\hat{F}\|_\infty. \tag{3.14}$$

Now define $G\colon \{0,1\}^n \to \mathbb{R}$ by (3.5). Then Lemma 3.8(i) gives

$$\hat{F}(S) = \hat{G}(S), \qquad |S| \leqslant m. \tag{3.15}$$

Since the Fourier spectrum of F is supported on characters of order D or higher (where $D \leqslant m$), we conclude that $G \in \operatorname{span}\{\chi_S : |S| \geqslant D\}$. This results in the following chain of implications:

$$\begin{aligned}
&g \in \operatorname{span}\{\chi_S : |S| \geqslant D\} && \text{by Lemma 3.8(ii)},\\
&g \in 2^m \|g\|_1 \Psi_{m,D} && \text{by Lemma 3.6(i)},\\
&g \in 2^n\, 16^m \|\hat{F}\|_\infty \Psi_{m,D} && \text{by (3.14)},\\
&G \in 2^n\, 16^m \|\hat{F}\|_\infty \Psi_{n,D} && \text{by Lemma 3.8(iii)}.
\end{aligned} \tag{3.16}$$

Finally,

$$\begin{aligned}
\|\widehat{F-G}\|_\infty &\leqslant \|\hat{F}\|_\infty + \|\hat{G}\|_\infty \\
&\leqslant \|\hat{F}\|_\infty + 2^{-n}\|g\|_1 && \text{by Lemma 3.8(i)} \\
&\leqslant 8^m \|\hat{F}\|_\infty && \text{by (3.14).} \qquad (3.17)
\end{aligned}$$

Now (3.10)–(3.12) follow from (3.16), (3.15), and (3.17), respectively. ∎

By iteratively applying the previous lemma, we obtain the desired representation for $2^{-n}\sum_{|S|\geqslant d}\chi_S$.

LEMMA 3.11. *Let $F\colon\{0,1\}^n \to \mathbb{R}$ be a symmetric function with $F \in \operatorname{span}\{\chi_S : |S| \geqslant d\}$, where d is an integer with $1 \leqslant d \leqslant n$. Then for every real $c > 1$,*

$$F \in 2^n \|\hat{F}\|_\infty \sum_{i=0}^{\infty} \left(2^{\frac{4c^2-c}{c-1}}\right)^{c^i d} \Psi_{n,\lceil c^i d\rceil}. \tag{3.18}$$

Proof. We will construct symmetric functions $F_1, F_2, \ldots, F_i, \ldots\colon\{0,1\}^n \to \mathbb{R}$, where

$$F_i \in 2^n \|\hat{F}\|_\infty \left(2^{\frac{4c^2-c}{c-1}}\right)^{c^{i-1} d} \Psi_{n,\lceil c^{i-1} d\rceil}, \tag{3.19}$$

$$F - F_1 - F_2 - \cdots - F_i \in \operatorname{span}\{\chi_S : |S| \geqslant \lceil c^i d\rceil\}, \tag{3.20}$$

$$\|\widehat{F - F_1 - F_2 - \cdots - F_i}\|_\infty \leqslant 8^{\frac{c^{i+1}d - cd}{c-1}} \|\hat{F}\|_\infty. \tag{3.21}$$

Before carrying out the construction, let us finish the proof assuming the existence of such a sequence. Since $c^i d > n$ for all i sufficiently large, (3.19) implies that only finitely many functions in the sequence $\{F_i\}_{i=1}^\infty$ are nonzero. The series $\sum_{i=1}^\infty F_i$ is therefore well-defined, and (3.20) gives $F = \sum_{i=1}^\infty F_i$. Property (3.19) now settles (3.18).

We will construct $F_1, F_2, \ldots, F_i, \ldots$ using induction to ensure properties (3.19)–(3.21). The base case $i = 0$ is immediate from the assumed membership $F \in \operatorname{span}\{\chi_S : |S| \geqslant d\}$. For the inductive step, fix $i \geqslant 1$ and assume that the symmetric functions $F_1, F_2, \ldots, F_{i-1}$ have already been constructed. Then by the inductive hypothesis,

$$F - F_1 - \cdots - F_{i-1} \in \operatorname{span}\{\chi_S : |S| \geqslant \lceil c^{i-1} d\rceil\},$$

$$\|\widehat{F - F_1 - \cdots - F_{i-1}}\|_\infty \leqslant 8^{\frac{c^i d - cd}{c-1}} \|\hat{F}\|_\infty. \tag{3.22}$$

There are two cases to consider. In the degenerate case when $\lceil c^{i-1}d\rceil = \lceil c^i d\rceil$, one obtains (3.19)–(3.21) trivially by letting $F_i = 0$. In the complementary case when $\lceil c^{i-1}d\rceil < \lceil c^i d\rceil$, we have $\lceil c^{i-1}d\rceil \leqslant \lfloor c^i d\rfloor$. As a result, Lemma 3.10 is applicable with parameters $D = \lceil c^{i-1}d\rceil$ and $m = \lfloor c^i d\rfloor$ to the symmetric function $F - F_1 - F_2 - \cdots - F_{i-1}$ and yields a symmetric function F_i such that

$$F_i \in 2^n\, 16^{\lfloor c^i d\rfloor} \|\widehat{F - F_1 - \cdots - F_{i-1}}\|_\infty \Psi_{n,\lceil c^{i-1}d\rceil},$$

$$(F - F_1 - \cdots - F_{i-1}) - F_i \in \operatorname{span}\{\chi_S : |S| \geqslant \lfloor c^i d\rfloor + 1\},$$

$$\|\widehat{(F - F_1 - \cdots - F_{i-1}) - F_i}\|_\infty \leqslant 8^{\lfloor c^i d\rfloor} \|\widehat{F - F_1 - \cdots - F_{i-1}}\|_\infty.$$

These three properties establish (3.19)–(3.21) in view of (3.22). ∎

We have reached the main result of Section 3, stated earlier as (3.2).

THEOREM 3.12. *Let $c > 1$ be a given real number. Then for every $d = 1, 2, \dots, n$ and every function $f : \{0,1\}^n \to \mathbb{R}$,*

$$E(f, d-1) \leqslant \left\| \sum_{|S| \geqslant d} \hat{f}(S) \chi_S \right\|_\infty$$

$$\leqslant \sum_{i=0}^{\infty} \left(2^{\frac{4c^2 - c}{c-1}} \right)^{c^i d} \Delta(f, \lceil c^i d \rceil). \tag{3.23}$$

In particular,

$$E(f, d-1) \leqslant \sum_{i=d}^{n} 5^{6i} \Delta(f, i), \tag{3.24}$$

$$E(f, d-1) \leqslant \sum_{i=0}^{\lfloor \log \frac{n}{d} \rfloor} \left(2^{14}\right)^{2^i d} \Delta(f, 2^i d). \tag{3.25}$$

Proof. The function $c \mapsto 2^{(4c^2 - c)/(c-1)}$ attains its minimum on $(1, \infty)$ at the point $c = 1 + \sqrt{3/4} = 1.8660\dots$. Substituting this value in (3.23) and noting that $\lceil (1 + \sqrt{3/4})^i d \rceil < \lceil (1 + \sqrt{3/4})^{i+1} d \rceil$ gives (3.24). For the alternate bound (3.25), let $c = 2$ in (3.23).

It remains to prove (3.23). Abbreviate $K = 2^{(4c^2 - c)/(c-1)}$ and define $\psi : \{0,1\}^n \to \mathbb{R}$ by $\psi(x) = 2^{-n} \sum_{|S| \geqslant d} \chi_S$. Then by Lemma 3.11,

$$\psi \in \sum_{i=0}^{\infty} K^{c^i d} \Psi_{n, \lceil c^i d \rceil}. \tag{3.26}$$

As a result,

$$\begin{aligned} \left\| \sum_{|S| \geqslant d} \hat{f}(S) \chi_S \right\|_\infty &= \| f * \psi \|_\infty && \text{by (2.7)} \\ &\leqslant \sum_{i=0}^{\infty} K^{c^i d} \max_{\psi' \in \Psi_{n, \lceil c^i d \rceil}} \| f * \psi' \|_\infty && \text{by (3.26)} \\ &\leqslant \sum_{i=0}^{\infty} K^{c^i d} \Delta(f, \lceil c^i d \rceil). && \text{by Lemma 3.6(iii).} \end{aligned}$$

∎

4. REPEATED DISCREPANCY OF SET DISJOINTNESS

Let G be a multiparty communication problem, such as set disjointness. The classic notion of discrepancy, reviewed in Section 2, involves fixing a probability distribution π on the domain of G and challenging a communication protocol to solve an instance X of G chosen at random according to π. If some low-cost protocol solves this task with nonnegligible accuracy, one says that G has high discrepancy with respect to π. In this paper, we introduce a rather different notion which we call *repeated discrepancy*. Here, one presents the communication protocol with arbitrarily many instances $X_1, X_2, X_3, \dots$ of the given communication problem G, each chosen independently from π conditioned on $G(X_1) = G(X_2) = G(X_3) = \cdots$. Thus, the instances are either all positive or all negative, and the protocol's challenge is to tell which is the case. The formal definition given next is somewhat more subtle, but the intuition is exactly the same.

DEFINITION 4.1. Let G be a (possibly partial) k-party communication problem on $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$ and π a probability distribution on the domain of G. The *repeated discrepancy of G with respect to π* is

$$\operatorname{rdisc}_\pi(G) = \sup_{n,r\in\mathbb{Z}^+} \max_{\chi} \left| \mathop{\mathbf{E}}_{\dots,X_{i,j},\dots} \left[\chi(\dots, X_{i,j}, \dots) \prod_{i=1}^{n} G(X_{i,1}) \right] \right|^{1/n},$$

where the maximum is over k-dimensional cylinder intersections χ on $\mathscr{X}^{nr} = \mathscr{X}_1^{nr} \times \mathscr{X}_2^{nr} \times \cdots \times \mathscr{X}_k^{nr}$, and the arguments $X_{i,j}$ $(i = 1,2,\dots,n,\ j = 1,2,\dots,r)$ are chosen independently according to π conditioned on $G(X_{i,1}) = G(X_{i,2}) = \cdots = G(X_{i,r})$ for each i.

We focus on probability distributions π that are *balanced* on the domain of G, meaning that negative and positive instances carry equal weight: $\pi(G^{-1}(-1)) = \pi(G^{-1}(+1))$. We define

$$\operatorname{rdisc}(G) = \inf_{\pi} \operatorname{rdisc}_\pi(G),$$

where the infimum is over all probability distributions on the domain of G that are balanced. Our motivation for studying repeated discrepancy comes from the approximation theoretic contribution of this paper, Theorem 3.12. Using it, we will now prove that repeated discrepancy gives a highly efficient way to approximate multiparty protocols by polynomials.

THEOREM 4.2. *Let G be a (possibly partial) k-party communication problem on $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$. For an integer $n \geqslant 1$ and a balanced probability distribution π on* $\operatorname{dom} G$, *consider the linear operator $L_{\pi,n} \colon \mathbb{R}^{\mathscr{X}^n} \to \mathbb{R}^{\{0,1\}^n}$ given by*

$$(L_{\pi,n}\chi)(x) = \mathop{\mathbf{E}}_{X_1\sim\pi_{x_1}} \cdots \mathop{\mathbf{E}}_{X_n\sim\pi_{x_n}} \chi(X_1,\dots,X_n), \qquad x \in \{0,1\}^n,$$

where π_0 and π_1 are the probability distributions induced by π on $G^{-1}(+1)$ and $G^{-1}(-1)$, respectively. Then for some absolute constant $c > 0$ and every k-dimensional cylinder intersection χ on $\mathscr{X}^n = \mathscr{X}_1^n \times \mathscr{X}_2^n \times \cdots \times \mathscr{X}_k^n$,

$$E(L_{\pi,n}\chi, d-1) \leqslant (c \operatorname{rdisc}_\pi(G))^d, \qquad d = 1,2,\dots,n.$$

Proof. Put $\Delta_d = \max_\chi \Delta(L_{\pi,n}\chi, d)$, where the maximum is over k-dimensional cylinder intersections. In light of (3.24), it suffices to prove that

$$\Delta_d \leqslant (2 \operatorname{rdisc}_\pi(G))^d, \qquad d = 1,2,\dots,n. \tag{4.1}$$

Fix $w \in \{0,1\}^n$ and pairwise disjoint sets $S_1, S_2, \dots, S_d \subseteq \{1,2,\dots,n\}$ such that

$$\Delta_d = \max_\chi \left| \frac{\partial^d\ (L_{\pi,n}\chi)}{\partial S_1\ \partial S_2\ \cdots\ \partial S_d}(w) \right|, \tag{4.2}$$

where the maximum is over k-dimensional cylinder intersections. Then by the definition of directional derivative,

$$\Delta_d = \max_{\chi} \left| \mathop{\mathbf{E}}_{z\in\{0,1\}^d} \mathop{\mathbf{E}}_{X_1,X_2,\ldots,X_n} \chi(X_1, X_2, \ldots, X_n)\,(-1)^{|z|} \right|, \tag{4.3}$$

where

$$X_i \sim \begin{cases} \pi_{w_i\oplus z_1} & \text{if } i \in S_1, \\ \pi_{w_i\oplus z_2} & \text{if } i \in S_2, \\ \vdots & \vdots \\ \pi_{w_i\oplus z_d} & \text{if } i \in S_d, \\ \pi_{w_i} & \text{otherwise.} \end{cases}$$

In other words, the cylinder intersection χ receives zero or more arguments distributed independently according to π_{z_1}, zero or more arguments distributed independently according to $\pi_{\overline{z_1}}$, zero or more arguments distributed independently according to π_{z_2}, and so on, for a total of n arguments. To simplify the remainder of the proof, we will manipulate the input to χ as follows.

(i) We will discard any arguments X_i whose probability distribution does not depend on z, simply by fixing them so as to maximize the expectation in (4.3) with respect to the remaining arguments. This simplification is legal because after one or more arguments X_i are fixed, χ continues to be a cylinder intersection with respect to the remaining arguments.

(ii) We will provide the cylinder intersection with *additional* arguments drawn independently from each of the probability distributions $\pi_{z_1}, \pi_{\overline{z_1}}, \ldots, \pi_{z_d}, \pi_{\overline{z_d}}$, so that there are exactly n arguments per distribution. This simplification is legal because the cylinder intersection can always choose to ignore the newly provided arguments.

Applying these two simplifications, we arrive at

$$\Delta_d \leqslant \max_{\chi} \left| \mathop{\mathbf{E}}_{z\in\{0,1\}^d} \mathop{\mathbf{E}}_{\substack{X_{1,1},\ldots,X_{1,n}\sim\pi_{z_1} \\ Y_{1,1},\ldots,Y_{1,n}\sim\pi_{\overline{z_1}}}} \cdots \mathop{\mathbf{E}}_{\substack{X_{d,1},\ldots,X_{d,n}\sim\pi_{z_d} \\ Y_{d,1},\ldots,Y_{d,n}\sim\pi_{\overline{z_d}}}} \chi(\ldots, X_{i,1}, \ldots, X_{i,n}, Y_{i,1}, \ldots, Y_{i,n}, \ldots)\,(-1)^{|z|} \right|. \tag{4.4}$$

It remains to eliminate $\pi_{\overline{z_1}}, \ldots, \pi_{\overline{z_d}}$. Rewriting (4.4) in tensor notation,

$$\begin{aligned}
\Delta_d &\leqslant 2^{-d} \max_{\chi} \left| \sum_{z\in\{0,1\}^d} (-1)^{|z|} \left\langle \chi, \bigotimes_{i=1}^{d} (\pi_{z_i}^{\otimes n} \otimes \pi_{\overline{z_i}}^{\otimes n}) \right\rangle \right| \\
&= 2^{-d} \max_{\chi} \left| \left\langle \chi, \bigotimes_{i=1}^{d} (\pi_0^{\otimes n} \otimes \pi_1^{\otimes n} - \pi_1^{\otimes n} \otimes \pi_0^{\otimes n}) \right\rangle \right| \\
&= 2^{-d} \max_{\chi} \left| \left\langle \chi, \bigotimes_{i=1}^{d} (\pi_0^{\otimes n} \otimes \pi_0^{\otimes n} - \pi_1^{\otimes n} \otimes \pi_0^{\otimes n} \right.\right. \\
&\qquad\qquad \left.\left. - \pi_0^{\otimes n} \otimes \pi_0^{\otimes n} + \pi_0^{\otimes n} \otimes \pi_1^{\otimes n}) \right\rangle \right| \\
&= 2^{-d} \max_{\chi} \left| \sum_{y\in\{0,1\}^d} (-1)^{|y|} \sum_{z\in\{0,1\}^d} (-1)^{|z|} \left\langle \chi, \bigotimes_{i=1}^{d} (\pi_{z_i\wedge\overline{y_i}}^{\otimes n} \otimes \pi_{z_i\wedge y_i}^{\otimes n}) \right\rangle \right| \\
&\leqslant \max_{y\in\{0,1\}^d} \max_{\chi} \left| \sum_{z\in\{0,1\}^d} (-1)^{|z|} \left\langle \chi, \bigotimes_{i=1}^{d} (\pi_{z_i\wedge\overline{y_i}}^{\otimes n} \otimes \pi_{z_i\wedge y_i}^{\otimes n}) \right\rangle \right|. \qquad (4.5)
\end{aligned}$$

For every $y \in \{0,1\}^d$, the probability distribution $\bigotimes_{i=1}^{d} (\pi_{z_i\wedge\overline{y_i}}^{\otimes n} \otimes \pi_{z_i\wedge y_i}^{\otimes n})$ is the same as $(\pi_{z_1}^{\otimes n} \otimes \cdots \otimes \pi_{z_d}^{\otimes n}) \otimes (\pi_0^{\otimes n} \otimes \cdots \otimes \pi_0^{\otimes n})$, up to a permutation of the coordinates. The inner maximum in (4.5) is therefore the same for all y, namely,

$$2^d \max_{\chi} \left| \mathop{\mathbf{E}}_{z\in\{0,1\}^d} \mathop{\mathbf{E}}_{X_{1,1},\ldots,X_{1,n}\sim\pi_{z_1}} \cdots \mathop{\mathbf{E}}_{X_{d,1},\ldots,X_{d,n}\sim\pi_{z_d}} \mathop{\mathbf{E}}_{Y_{1,1},\ldots,Y_{d,n}\sim\pi_0} \chi(\ldots, X_{i,j}, Y_{i,j}, \ldots)\,(-1)^{|z|} \right|.$$

The variables $Y_{i,j}$ can be discarded, as argued in (i) at the beginning of this proof. This leaves us with

$$\Delta_d \leqslant 2^d \max_{\chi} \left| \mathop{\mathbf{E}}_{z\in\{0,1\}^d} \mathop{\mathbf{E}}_{X_{1,1},\ldots,X_{1,n}\sim\pi_{z_1}} \cdots \mathop{\mathbf{E}}_{X_{d,1},\ldots,X_{d,n}\sim\pi_{z_d}} \chi(\ldots, X_{i,j}, \ldots) \prod_{i=1}^{d} G(X_{i,1}) \right|.$$

Since π is balanced, (4.1) follows immediately. ∎

Theorem 1.1 gives a highly efficient way to transform communication protocols for composed problems $f \circ G$ into approximating polynomials for f, as long as the base communication problem G has repeated discrepancy smaller than a certain absolute constant. This result is centrally relevant to the set disjointness problem in light of its composed structure: $\mathrm{DISJ}_{k,rs} = \mathrm{AND}_r \circ \mathrm{DISJ}_{k,s}$ for any integers r, s. In the remainder of this section, we will obtain a near-tight upper bound on the repeated discrepancy of set disjointness, which we will later use to prove the main result of this paper.

4.1. Key distributions and definitions. Let T_k be the $k \times 2^{k-1}$ matrix whose columns are the 2^{k-1} distinct columns of the same parity as the all-ones vector 1^k. Let F_k be the $k \times 2^{k-1}$ matrix whose columns are the 2^{k-1} distinct columns of the same parity as the vector 01^{k-1}. Thus, the columns of T_k and F_k form a partition of $\{0,1\}^k$. We use T_k and F_k to encode true and false instances of set disjointness, respectively, hence the choice of notation. Let H_k be the $k \times 2^k$ matrix whose columns are the 2^k distinct vectors in $\{0,1\}^k$, and let H'_k be the $k \times (2^k - 1)$ matrix whose columns are the $2^k - 1$ distinct vectors in $\{0,1\}^k \setminus \{1^k\}$. The choice of letter for H_k and H'_k is a reference to the hypercube. For definitiveness one may assume that the columns of T_k, F_k, H_k, H'_k are ordered lexicographically, although the choice of ordering is immaterial for our purposes. For an integer $m \geqslant 1$, we define shorthands

$$H_{k,m} = [\,\underbrace{H_k \quad H_k \quad \dots \quad H_k}_{m}\,], \qquad\qquad H'_{k,m} = [\,\underbrace{H'_k \quad H'_k \quad \dots \quad H'_k}_{m}\,].$$

For a Boolean matrix A, we define

$$\overline{A} = A \oplus \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

When A is a Boolean matrix of dimension 1×1, this notation is consistent with our earlier shorthand $\overline{a} = a \oplus 1$ for $a \in \{0,1\}$. Observe that for any matrices $A, A_1, A_2, \dots, A_n$,

$$\overline{\overline{A}} = A, \tag{4.6}$$

$$[\,\overline{A_1} \quad \overline{A_2} \quad \cdots \quad \overline{A_n}\,] = \overline{[\,A_1 \quad A_2 \quad \cdots \quad A_n\,]}. \tag{4.7}$$

Moreover,

$$\overline{H_{k,m}} =_{\circlearrowright} H_{k,m}, \tag{4.8}$$

$$\overline{T_k} =_{\circlearrowright} F_k, \tag{4.9}$$

$$\overline{F_k} =_{\circlearrowright} T_k. \tag{4.10}$$

In this section, we will encounter a variety of probability distributions on matrix sequences. Describing them formulaically, using probability mass functions, is both tedious and unenlightening from the standpoint of proof. Instead, we will define each probability distribution algorithmically, by giving a procedure for generating a random element. We refer to such a specification as an *algorithmic description* of the given distribution. We will often use the following shorthand: for fixed matrices $A_1, A_2, \dots, A_t$, the notation

$$(A_1^{\circlearrowright}, A_2^{\circlearrowright}, \dots, A_t^{\circlearrowright}) \tag{4.11}$$

stands for a random tuple of matrices obtained from $(A_1, A_2, \dots, A_t)$ by permuting the columns in each of the t matrices independently and uniformly at random. In other words, (4.11) refers to a random tuple $(\sigma_1 A_1, \sigma_2 A_2, \dots, \sigma_t A_t)$, where $\sigma_1, \sigma_2, \dots, \sigma_t$ are column permutations chosen independently and uniformly at random. We will also use (4.11) to refer to the resulting probability distribution on matrix tuples, which will enable us to use shorthands like $B \sim A^{\circlearrowright}$ and $(B_1, B_2, \dots, B_t) \sim (A_1^{\circlearrowright}, A_2^{\circlearrowright}, \dots, A_t^{\circlearrowright})$. As an important special case, the $\circlearrowright$ notation applies to row vectors, which are matrices with a single row. On occasion, it will be necessary to apply the $\circlearrowright$ notation to a submatrix rather than the

Algorithm 1: Alternate algorithmic description of $\mu_{k,m}$ for $k \geqslant 2$

(i) Choose $f \in \{0,1\}$ uniformly at random.

(ii) Choose 2^k row vectors $a_{0^{k-1}}, b_{0^{k-1}}, \ldots, a_{1^{k-1}}, b_{1^{k-1}}$ independently according to

$$\begin{aligned}
&a_{1^{k-1}} \sim [0^{2m} \quad f]^{\circlearrowright}, && \\
&a_z \sim [0^{2m}1^{2m} \quad f \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}}]^{\circlearrowright}, && z \neq 1^{k-1}, \\
&b_z \sim [0^m 1^m \quad f \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}}]^{\circlearrowright}, && z \in \{0,1\}^{k-1}.
\end{aligned}$$

(iii) Define A_z, B_z for $z \in \{0,1\}^{k-1}$ by

$$A_z = \left[\begin{array}{cccc} z_1 & z_1 & \cdots & z_1 \\ z_2 & z_2 & \cdots & z_2 \\ \vdots & \vdots & & \vdots \\ z_{k-1} & z_{k-1} & \cdots & z_{k-1} \\ \hline & & a_z & \end{array}\right], \qquad B_z = \left[\begin{array}{cccc} z_1 & z_1 & \cdots & z_1 \\ z_2 & z_2 & \cdots & z_2 \\ \vdots & \vdots & & \vdots \\ z_{k-1} & z_{k-1} & \cdots & z_{k-1} \\ \hline & & b_z & \end{array}\right].$$

(iv) Output $([A_{0^{k-1}} \quad \cdots \quad A_{1^{k-1}}]^{\circlearrowright}, [B_{0^{k-1}} \quad \cdots \quad B_{1^{k-1}}]^{\circlearrowright})$.

entire matrix. For example, $[\ [0^m 1^m]^{\circlearrowright} \quad 0 \quad 0 \quad 1]$ refers to a random row vector whose last three components are $0,0,1$ and the first $2m$ components are a uniformly random permutation of the row vector $0^m 1^m$.

We say that a probability distribution μ on matrix sequences $(A_1, A_2, \ldots, A_t)$ is *invariant under column permutations* if $\mu(A_1, A_2, \ldots, A_t) = \mu(\sigma_1 A_1, \sigma_2 A_2, \ldots, \sigma_t A_t)$ for every choice of column permutations $\sigma_1, \sigma_2, \ldots, \sigma_t$. Most of the randomized procedures in this section involve choosing $(A_1, A_2, \ldots, A_t)$ by some process and outputting $(A_1^{\circlearrowright}, A_2^{\circlearrowright}, \ldots, A_t^{\circlearrowright})$, so that the resulting probability distribution on matrix sequences is invariant under column permutations.

We now define the main probability distribution of interest to us, which we call $\mu_{k,m}$. Nearly all of the work in this section is devoted to understanding various metric properties of $\mu_{k,m}$ and of probability distributions derived from it.

DEFINITION. For positive integers k, m, let $\mu_{k,m}$ be the probability distribution whose algorithmic description is as follows: choose $M \in \{T_k, F_k\}$ uniformly at random and output $([M \quad H'_{k,2m}]^{\circlearrowright}, [M \quad H_{k,m}]^{\circlearrowright})$.

We will need to establish an alternate procedure for sampling from $\mu_{k,m}$, whereby one first chooses rows $1, 2, \ldots, k-1$ and then the remaining row according to the conditional probability distribution. Such a procedure is given by Algorithm 1.

PROPOSITION 4.3. *Algorithm* 1 *is a valid algorithmic description of* $\mu_{k,m}$ *for* $k \geqslant 2$*.*

Proof. By inspection, the output distribution of Algorithm 1 has the following properties: (i) it is invariant under column permutations; (ii) with probability $1/2$, the output is a matrix pair (A, B) with $A =_{\circlearrowright} [T_k \quad H'_{k,2m}]$ and $B =_{\circlearrowright} [T_k \quad H_{k,m}]$; (iii) with probability $1/2$, the output is a matrix pair (A, B) with $A =_{\circlearrowright} [F_k \quad H'_{k,2m}]$ and $B =_{\circlearrowright} [F_k \quad H_{k,m}]$. There is only one probability distribution with these three properties, namely, $\mu_{k,m}$. $\square$

Algorithm 2: Alternate algorithmic description of $\lambda_{k,m}$

(i) Choose $f, f' \in \{0,1\}$ uniformly at random.

(ii) Choose 2^{k+1} row vectors a_z, a'_z, b_z, b'_z, for $z \in \{0,1\}^{k-1}$, independently according to

$$\begin{aligned}
a_{1^{k-1}} &\sim [0^{2m} \quad f]^{\circlearrowright}, && \\
a'_{1^{k-1}} &\sim [0^{2m} \quad f']^{\circlearrowright}, && \\
a_z &\sim [0^{2m}1^{2m} \quad f \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}}]^{\circlearrowright}, && z \neq 1^{k-1}, \\
a'_z &\sim [0^{2m}1^{2m} \quad f' \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}}]^{\circlearrowright}, && z \neq 1^{k-1}, \\
b_z &\sim [0^{m}1^{m} \quad f \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}}]^{\circlearrowright}, && z \in \{0,1\}^{k-1}, \\
b'_z &\sim [0^{m}1^{m} \quad f' \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}}]^{\circlearrowright}, && z \in \{0,1\}^{k-1}.
\end{aligned}$$

(iii) Define A_z, B_z for $z \in \{0,1\}^{k-1}$ by

$$A_z = \left[\begin{array}{cccc} z_1 & z_1 & \cdots & z_1 \\ z_2 & z_2 & \cdots & z_2 \\ \vdots & \vdots & & \vdots \\ z_{k-1} & z_{k-1} & \cdots & z_{k-1} \\ \hline \multicolumn{4}{c}{a_z} \\ \multicolumn{4}{c}{a'_z} \end{array}\right], \quad B_z = \left[\begin{array}{cccc} z_1 & z_1 & \cdots & z_1 \\ z_2 & z_2 & \cdots & z_2 \\ \vdots & \vdots & & \vdots \\ z_{k-1} & z_{k-1} & \cdots & z_{k-1} \\ \hline \multicolumn{4}{c}{b_z} \\ \multicolumn{4}{c}{b'_z} \end{array}\right]. \tag{4.12}$$

(iv) Output $([A_{0^{k-1}} \quad \cdots \quad A_{1^{k-1}}]^{\circlearrowright}, [B_{0^{k-1}} \quad \cdots \quad B_{1^{k-1}}]^{\circlearrowright})$.

We now define a key probability distribution $\lambda_{k,m}$ derived from $\mu_{k,m}$.

DEFINITION. For integers $k \geqslant 2$ and $m \geqslant 1$, define $\lambda_{k,m}$ to be the probability distribution with the following algorithmic description:

(i) pick a matrix pair $(A, B) \in \{0,1\}^{k-1,*} \times \{0,1\}^{k-1,*}$ according to the marginal distribution of $\mu_{k,m}$ on the first $k-1$ rows;

(ii) consider the probability distribution induced by $\mu_{k,m}$ on matrix pairs of the form $\left(\begin{bmatrix} A \\ * \end{bmatrix}, \begin{bmatrix} B \\ * \end{bmatrix}\right)$, and choose $\left(\begin{bmatrix} A \\ a \end{bmatrix}, \begin{bmatrix} B \\ b \end{bmatrix}\right), \left(\begin{bmatrix} A \\ a' \end{bmatrix}, \begin{bmatrix} B \\ b' \end{bmatrix}\right)$ independently according to that distribution;

(iii) output $\left(\begin{bmatrix} A \\ a \\ a' \end{bmatrix}, \begin{bmatrix} B \\ b \\ b' \end{bmatrix}\right)$.

By symmetry of the columns, $\lambda_{k,m}$ is invariant under column permutations. To reason effectively about $\lambda_{k,m}$, we need a more explicit algorithmic description.

PROPOSITION 4.4. *Algorithm 2 is a valid algorithmic description of $\lambda_{k,m}$.*

Proof. Immediate from the description of $\mu_{k,m}$ given by Algorithm 1. ∎

In analyzing the repeated discrepancy of set disjointness, we will need to argue that the last two rows of a matrix pair drawn according to $\lambda_{k,m}$ do not reveal too much information

Algorithm 3: Definition of $\nu^i_{k,m}$ $(i = 0, 1)$

(i) Choose $f, f' \in \{0,1\}$ uniformly at random.

(ii) Choose 2^{k+1} row vectors a_z, a'_z, b_z, b'_z, for $z \in \{0,1\}^{k-1}$, independently according to

$$\begin{aligned}
&a_{1^{k-1}} = [0^{2m-1} \quad f \quad 0], \\
&a'_{1^{k-1}} = [0^{2m-1} \quad 0 \quad f'], \\
&a_z \sim [\,[0^{2m-1}1^{2m-1}]^{\circlearrowright} \quad 1 \quad f \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}} \quad 0], && z \neq 1^{k-1}, \\
&a'_z \sim [\,[0^{2m-1}1^{2m-1}]^{\circlearrowright} \quad 1 \quad 0 \quad f' \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}}], && z \neq 1^{k-1}, \\
&b_z \sim [\,[0^{m-1}1^{m-1}]^{\circlearrowright} \quad \overline{i} \quad f \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}} \quad i], && z \in \{0,1\}^{k-1}, \\
&b'_z \sim [\,[0^{m-1}1^{m-1}]^{\circlearrowright} \quad \overline{i} \quad i \quad f' \oplus \overline{z_1} \oplus \cdots \oplus \overline{z_{k-1}}], && z \in \{0,1\}^{k-1}.
\end{aligned}$$

(iii) Define A_z, B_z for $z \in \{0,1\}^{k-1}$ by (4.12).

(iv) Output $([A_{0^{k-1}} \quad \cdots \quad A_{1^{k-1}}]^{\circlearrowright}, [B_{0^{k-1}} \quad \cdots \quad B_{1^{k-1}}]^{\circlearrowright})$.

about the remaining rows. We will do so by showing that $\lambda_{k,m}$ is close in statistical distance to certain probability distributions $\nu^0_{k,m}, \nu^1_{k,m}$ in which no information is revealed.

DEFINITION. For integers $k \geqslant 2$ and $m \geqslant 1$, define $\nu^0_{k,m}$ and $\nu^1_{k,m}$ to be the probability distributions whose algorithmic descriptions are given by Algorithm 3.

Comparing Algorithms 2 and 3, we see that the new distributions $\nu^0_{k,m}$ and $\nu^1_{k,m}$ differ from $\lambda_{k,m}$ exclusively in step (ii) of the algorithmic description. An alternate, global view of $\nu^0_{k,m}$ and $\nu^1_{k,m}$ is given by the following proposition.

PROPOSITION 4.5. *Algorithm* 4 *is a valid algorithmic description of* $\nu^0_{k,m}$ *and* $\nu^1_{k,m}$.

Proof. Algorithm 4 is obtained from Algorithm 3 by reordering the columns prior to the application of the $\circlearrowright$ operator. Specifically, in the notation of Algorithm 3, the last two columns of $A_{1^{k-1}}$ and the last three columns of each A_z $(z \neq 1^{k-1})$ are moved up front and listed before any of the remaining columns; likewise, the last three columns of each B_z $(z \in \{0,1\}^{k-1})$ are moved up front and listed before any of the remaining columns. The subsequent application of the $\circlearrowright$ operator in both algorithms ensures that the output distributions are the same. ∎

Closely related to $\nu^0_{k,m}$ and $\nu^1_{k,m}$ are the distributions $\nu^0_{k,P_1,\ldots,P_8}$ and $\nu^1_{k,P_1,\ldots,P_8}$, defined next.

DEFINITION 4.6. Let $P_1, \ldots, P_8 \in \{0,1\}^{k-1,*}$. The probability distribution $\nu^0_{k,P_1,\ldots,P_8}$ on matrix pairs is given by choosing $M_1, M_2 \in \{T_{k-1}, F_{k-1}\}$ uniformly at random and

Algorithm 4: Alternate algorithmic description of $\nu^i_{k,m}$ $(i = 0, 1)$

(i) Choose 2^{k+1} row vectors a_z, a'_z, b_z, b'_z, for $z \in \{0,1\}^{k-1}$, independently according to

$$a_{1^{k-1}} = a'_{1^{k-1}} = 0^{2m-1},$$
$$a_z, a'_z \sim [0^{2m-1}1^{2m-1}]^{\circlearrowleft}, \qquad z \neq 1^{k-1},$$
$$b_z, b'_z \sim [0^{m-1}1^{m-1}]^{\circlearrowleft}, \qquad z \in \{0,1\}^{k-1}.$$

(ii) Define A_z, B_z for $z \in \{0,1\}^{k-1}$ by (4.12).

(iii) Choose $M_1, M_2 \in \{T_{k-1}, F_{k-1}\}$ uniformly at random and output the matrix pair

$$\left[\begin{array}{c|c|c|c|c|ccc} H'_{k-1} & M_1 & \overline{M_1} & M_2 & \overline{M_2} & & & \\ \cline{1-5} 11\ldots1 & 11\ldots1 & 00\ldots0 & 00\ldots0 & 00\ldots0 & A_{0^{k-1}} & \cdots & A_{1^{k-1}} \\ 11\ldots1 & 00\ldots0 & 00\ldots0 & 11\ldots1 & 00\ldots0 & & & \end{array}\right]^{\circlearrowleft},$$

$$\left[\begin{array}{c|c|c|c|c|ccc} H_{k-1} & M_1 & \overline{M_1} & M_2 & \overline{M_2} & & & \\ \cline{1-5} \bar{i}\,\bar{i}\ldots\bar{i} & 11\ldots1 & 00\ldots0 & i\,i\ldots i & i\,i\ldots i & B_{0^{k-1}} & \cdots & B_{1^{k-1}} \\ \bar{i}\,\bar{i}\ldots\bar{i} & i\,i\ldots i & i\,i\ldots i & 11\ldots1 & 00\ldots0 & & & \end{array}\right]^{\circlearrowleft}.$$

---

outputting the pair

$$\left[\begin{array}{cc|cc|ccc|c} M_1 & P_1 & M_2 & P_2 & \overline{M_1} & \overline{M_2} & P_3 & P_4 \\ \hline \multicolumn{2}{c|}{111\ldots1} & \multicolumn{2}{c|}{000\ldots0} & \multicolumn{3}{c|}{00000000\ldots0} & 11\ldots1 \\ \multicolumn{2}{c|}{000\ldots0} & \multicolumn{2}{c|}{111\ldots1} & \multicolumn{3}{c|}{00000000\ldots0} & 11\ldots1 \end{array}\right]^{\circlearrowleft},$$

$$\left[\begin{array}{cc|cc|ccc|c} M_1 & P_5 & M_2 & P_6 & \overline{M_1} & \overline{M_2} & P_7 & P_8 \\ \hline \multicolumn{2}{c|}{111\ldots1} & \multicolumn{2}{c|}{000\ldots0} & \multicolumn{3}{c|}{00000000\ldots0} & 11\ldots1 \\ \multicolumn{2}{c|}{000\ldots0} & \multicolumn{2}{c|}{111\ldots1} & \multicolumn{3}{c|}{00000000\ldots0} & 11\ldots1 \end{array}\right]^{\circlearrowleft}. \tag{4.13}$$

The probability distribution $\nu^1_{k,P_1,\ldots,P_8}$ on matrix pairs is given by choosing $M_1, M_2 \in \{T_{k-1}, F_{k-1}\}$ uniformly at random and outputting the pair

$$\left[\begin{array}{cc|cc|ccc|c} M_1 & P_1 & M_2 & P_2 & \overline{M_1} & \overline{M_2} & P_3 & P_4 \\ \hline \multicolumn{2}{c|}{111\ldots1} & \multicolumn{2}{c|}{000\ldots0} & \multicolumn{3}{c|}{00000000\ldots0} & 11\ldots1 \\ \multicolumn{2}{c|}{000\ldots0} & \multicolumn{2}{c|}{111\ldots1} & \multicolumn{3}{c|}{00000000\ldots0} & 11\ldots1 \end{array}\right]^{\circlearrowleft},$$

$$\left[\begin{array}{cc|cc|c|ccc} \overline{M_2} & P_5 & \overline{M_1} & P_6 & P_7 & M_1 & M_2 & P_8 \\ \hline \multicolumn{2}{c|}{111\ldots1} & \multicolumn{2}{c|}{000\ldots0} & 00\ldots0 & \multicolumn{3}{c}{11111111\ldots1} \\ \multicolumn{2}{c|}{000\ldots0} & \multicolumn{2}{c|}{111\ldots1} & 00\ldots0 & \multicolumn{3}{c}{11111111\ldots1} \end{array}\right]^{\circlearrowleft}.$$

It is not hard to see, as we will soon, that $\nu^0_{k,m}$ is a convex combination of probability distributions $\nu^0_{k,P_1,\ldots,P_8}$, and analogously for $\nu^1_{k,m}$. This will enable us to replace $\nu^0_{k,m}$ and $\nu^1_{k,m}$ in our arguments by particularly simple and highly structured distributions.

DEFINITION. A matrix pair (A, B) is (k, m, α)-*good* if

$$\left[\begin{array}{c|c|c} H'_{k-1,2m'} & H'_{k-1,2m'} & H_{k-1,2m'} \\ \hline 11\ldots1 & 00\ldots0 & 00\ldots0 \\ 00\ldots0 & 11\ldots1 & 00\ldots0 \end{array}\right] \sqsubseteq A$$

and $H_{k+1,m'} \sqsubseteq B$, where $m' = \lceil \frac{1-\alpha}{2} \cdot m \rceil$. A matrix pair (A, B) is (k, m, α)-*bad* if it is not (k, m, α)-good.

It will be necessary to control the quantitative contribution of bad matrix pairs in the analysis of set disjointness. In the definition that follows, we give a special name to probability distributions $\nu^i_{k,P_1,\dots,P_8}$ supported on good matrix pairs.

DEFINITION. Let $\mathscr{G}^0_{k,m,\alpha}$ denote the set of all probability distributions $\nu^0_{k,P_1,\dots,P_8}$ that are supported on (k, m, α)-good matrix pairs. (The letter $\mathscr{G}$ is a reference to good pairs.) Analogously, let $\mathscr{G}^1_{k,m,\alpha}$ denote the set of all probability distributions $\nu^1_{k,P_1,\dots,P_8}$ that are supported on (k, m, α)-good matrix pairs.

The following proposition gives a convenient characterization of probability distributions in $\mathscr{G}^0_{k,m,\alpha}$ and $\mathscr{G}^1_{k,m,\alpha}$.

PROPOSITION 4.7. *Let $k \geqslant 2$ and $m \geqslant 1$ be integers, $m' = \lceil \frac{1-\alpha}{2} \cdot m \rceil$. Fix matrices $P_1, \dots, P_8 \in \{0,1\}^{k-1,*}$ and $i \in \{0,1\}$. Then $\nu^i_{k,P_1,\dots,P_8} \in \mathscr{G}^i_{k,m,\alpha}$ if and only if the following three conditions hold:*

(i) $H'_{k-1,2m'} \sqsubseteq P_1, P_2$,
(ii) $H_{k-1,2m'} \sqsubseteq P_3$,
(iii) $H_{k-1,m'} \sqsubseteq P_5, P_6, P_7, P_8$.

Proof. Immediate from the definitions of $\nu^i_{k,P_1,\dots,P_8}$ and (k, m, α)-good matrix pairs. ∎

4.2. Technical lemmas. We now establish key properties of the probability distributions introduced so far. Our main result here, Theorem 4.12, will be an approximate representation of $\lambda_{k,m}$ out of the convex hulls of $\mathscr{G}^0_{k,m,\alpha}$ and $\mathscr{G}^1_{k,m,\alpha}$, with careful control of the error term. We start with an auxiliary lemma which we will use to show the proximity of $\lambda_{k,m}$, $\nu^0_{k,m}$, and $\nu^1_{k,m}$ in statistical distance.

LEMMA 4.8. *For an integer $m \geqslant 1$, consider the probability distributions $\alpha_{m,1}, \alpha_{m,2}, \beta_m$ on $\{1, 2, \dots, m+2\}$ given by*

$$\alpha_{m,j}(i) = \binom{m}{i-j}^2 \binom{2m}{m}^{-1}, \qquad j = 1, 2,$$

$$\beta_m(i) = \binom{m+2}{i}\binom{m+1}{i-1}\binom{2m+3}{m+1}^{-1}.$$

Then there is an absolute constant $c > 0$ such that

$$H(\alpha_{m,j}, \beta_m) \leqslant \frac{c}{\sqrt{m}}, \qquad j = 1, 2.$$

That the functions $\alpha_{m,1}, \alpha_{m,2}, \beta_m$ are probability distributions follows from Vandermonde's convolution, (2.1).

Proof of Lemma 4.8. For $j = 1, 2$, elementary arithmetic gives

$$1-\frac{c}{m}-\frac{c|i-\frac{m}{2}|}{m} \leqslant \frac{\alpha_{m,j}(i)}{\beta_m(i)} \leqslant 1+\frac{c}{m}+\frac{c|i-\frac{m}{2}|}{m} \qquad (i = 1, 2, \ldots, m+2)$$

for some absolute constant $c > 0$, so that $|1 - \sqrt{\alpha_{m,j}(i)/\beta_m(i)}| \leqslant \frac{c}{m}(1 + |i - \frac{m}{2}|)$. As a result,

$$\begin{aligned} 2H(\alpha_{m,j}, \beta_m)^2 &= \mathop{\mathbf{E}}_{i\sim\beta_m}\left[\left(1-\sqrt{\frac{\alpha_{m,j}(i)}{\beta_m(i)}}\right)^2\right] \\ &\leqslant \frac{c^2}{m^2}\left\{1+2\mathop{\mathbf{E}}_{\beta_m}\left|i-\frac{m}{2}\right| + \mathop{\mathbf{E}}_{\beta_m}\left[\left(i-\frac{m}{2}\right)^2\right]\right\} \\ &\leqslant \frac{c^2}{m^2}\left\{1+2\sqrt{\mathop{\mathbf{E}}_{\beta_m}\left[\left(i-\frac{m}{2}\right)^2\right]} + \mathop{\mathbf{E}}_{\beta_m}\left[\left(i-\frac{m}{2}\right)^2\right]\right\}, \end{aligned} \tag{4.14}$$

where we used the fact that $\mathbf{E}\,X \leqslant \sqrt{\mathbf{E}[X^2]}$ for a real random variable X. Furthermore,

$$\begin{aligned} \mathop{\mathbf{E}}_{\beta_m}[i] &= \binom{2m+3}{m+1}^{-1}\sum_{i=1}^{m+2} i\binom{m+2}{i}\binom{m+1}{i-1} \\ &= \binom{2m+3}{m+1}^{-1}(m+2)\sum_{i=1}^{m+2}\binom{m+1}{i-1}^2 \\ &= \binom{2m+3}{m+1}^{-1}(m+2)\binom{2m+2}{m+1} \\ &= \frac{(m+2)^2}{2m+3} \end{aligned}$$

and

$$\begin{aligned} \mathop{\mathbf{E}}_{\beta_m}[i(i-1)] &= \binom{2m+3}{m+1}^{-1}\sum_{i=1}^{m+2} i(i-1)\binom{m+2}{i}\binom{m+1}{i-1} \\ &= \binom{2m+3}{m+1}^{-1}(m+1)(m+2)\sum_{i=2}^{m+2}\binom{m}{i-2}\binom{m+1}{i-1} \\ &= \binom{2m+3}{m+1}^{-1}(m+1)(m+2)\sum_{i=0}^{m}\binom{m}{i}\binom{m+1}{m-i} \\ &= \binom{2m+3}{m+1}^{-1}(m+1)(m+2)\binom{2m+1}{m} \\ &= \frac{(m+1)(m+2)^2}{2(2m+3)}, \end{aligned}$$

whence

$$\mathop{\mathbf{E}}_{\beta_m}\left[\left(i-\frac{m}{2}\right)^2\right] = \frac{m^2}{4} - (m-1)\mathop{\mathbf{E}}_{\beta_m}[i] + \mathop{\mathbf{E}}_{\beta_m}[i(i-1)] = O(m).$$

In view of (4.14), the proof is complete. ∎

A fairly direct consequence of the previous lemma is that the probability distributions $\lambda_{k,m}$, $\nu^0_{k,m}$, and $\nu^1_{k,m}$ are within $O(2^k/\sqrt{m})$ of each other in statistical distance. In what follows, we prove a superior upper bound of $O(\sqrt{2^k/m})$, which is tight. The analysis exploits the multiplicative property of Hellinger distance.

LEMMA 4.9. *There is a constant $c > 0$ such that for all integers $k \geqslant 2$ and $m \geqslant 1$,*

$$\|\lambda_{k,m} - \nu^i_{k,m}\|_1 \leqslant \sqrt{\frac{c2^k}{m}}, \qquad i = 0, 1.$$

Proof. Throughout the proof, the term "algorithmic description" will refer to Algorithm 2 in the case of $\lambda_{k,m}$ and Algorithm 3 in the case of $\nu^0_{k,m}$ and $\nu^1_{k,m}$. As we have noted earlier, the algorithmic descriptions of these three distributions are identical except for step (ii). In particular, observe that

$$\lambda_{k,m} = \frac{1}{4} \sum_{f,f' \in \{0,1\}} \lambda^{f,f'}_{k,m},$$
$$\nu^i_{k,m} = \frac{1}{4} \sum_{f,f' \in \{0,1\}} \nu^{i,f,f'}_{k,m}, \qquad i = 0, 1,$$

where $\lambda^{f,f'}_{k,m}, \nu^{0,f,f'}_{k,m}, \nu^{1,f,f'}_{k,m}$ are the distributions that result from $\lambda_{k,m}, \nu^0_{k,m}, \nu^1_{k,m}$, respectively, when one conditions on the choice of f, f' in step (i) of the algorithmic description. Therefore,

$$\begin{aligned} \|\lambda_{k,m} - \nu^i_{k,m}\|_1 &\leqslant \max_{f,f'} \left\| \lambda^{f,f'}_{k,m} - \nu^{i,f,f'}_{k,m} \right\|_1 \\ &\leqslant 2\sqrt{2} \max_{f,f'} H\left(\lambda^{f,f'}_{k,m}, \nu^{i,f,f'}_{k,m}\right), \qquad i = 0, 1, \end{aligned} \tag{4.15}$$

where the second step uses Fact 2.1.

In the remainder of the proof, we consider f, f' fixed. Define the *column histogram* of a matrix $X \in \{0,1\}^{k+1,*}$ to be the vector of 2^{k+1} natural numbers indicating how many times each string in $\{0,1\}^{k+1}$ occurs as a column of X. If D_1 and D_2 are two probability distributions on $\{0,1\}^{k+1,*}$ that are invariant under column permutations, then the Hellinger distance between D_1 and D_2 is obviously the same as the Hellinger distance between the column histograms of matrices drawn from D_1 versus D_2. An analogous statement holds for probability distributions D_1, D_2 on matrix pairs. As a result, we need only consider the column histograms of matrix pairs drawn from $\lambda^{f,f'}_{k,m}, \nu^{0,f,f'}_{k,m}, \nu^{1,f,f'}_{k,m}$. Furthermore, for every matrix pair

$$(A, B) \in \operatorname{supp} \lambda^{f,f'}_{k,m} \cup \operatorname{supp} \nu^{0,f,f'}_{k,m} \cup \operatorname{supp} \nu^{1,f,f'}_{k,m},$$

the column histograms of A and B are uniquely determined by the number of occurrences of

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{k-1} \\ f \oplus z_1 \oplus z_2 \oplus \cdots \oplus z_{k-1} \\ f' \oplus z_1 \oplus z_2 \oplus \cdots \oplus z_{k-1} \end{bmatrix} \tag{4.16}$$

as a column of A and B, respectively, for each $z \in \{0,1\}^{k-1}$. Thus, we need 2^{k-1} numbers per matrix, rather than 2^{k+1}, to describe the column histograms of A and B.

With this in mind, for $(A, B) \sim \lambda_{k,m}^{f,f'}$, define $a_{\lambda,z}$ and $b_{\lambda,z}$ (where $z \in \{0,1\}^{k-1}$) to be the number of occurrences of (4.16) as a column in A and B, respectively. Analogously, for $(A, B) \sim \nu_{k,m}^{i,f,f'}$, define $a_{\nu^i,z}$ and $b_{\nu^i,z}$ ($z \in \{0,1\}^{k-1}$) to be the number of occurrences of (4.16) as a column in A and B, respectively. By the preceding discussion, the Hellinger distance between $\lambda_{k,m}^{f,f'}$ and $\nu_{k,m}^{i,f,f'}$ is the same as the Hellinger distance between $(\dots, a_{\lambda,z}, b_{\lambda,z}, \dots)$ and $(\dots, a_{\nu^i,z}, b_{\nu^i,z}, \dots)$, viewed as random variables in $\mathbb{N}^{2^k}$. By step (ii) of the algorithmic description, the random variables

$$a_{\lambda,0^{k-1}}, b_{\lambda,0^{k-1}}, \dots, a_{\lambda,1^{k-1}}, b_{\lambda,1^{k-1}}$$

are independent. Similarly, for each $i = 0, 1$, the random variables

$$a_{\nu^i,0^{k-1}}, b_{\nu^i,0^{k-1}}, \dots, a_{\nu^i,1^{k-1}}, b_{\nu^i,1^{k-1}}$$

are independent. Therefore,

$$\begin{aligned} H\left(\lambda_{k,m}^{f,f'}, \nu_{k,m}^{i,f,f'}\right) &= H((\dots, a_{\lambda,z}, b_{\lambda,z}, \dots), (\dots, a_{\nu^i,z}, b_{\nu^i,z}, \dots)) \\ &\leqslant \sqrt{\sum_{z\in\{0,1\}^{k-1}} H(a_{\lambda,z}, a_{\nu^i,z})^2 + \sum_{z\in\{0,1\}^{k-1}} H(b_{\lambda,z}, b_{\nu^i,z})^2}, \end{aligned} \tag{4.17}$$

where the second step uses Fact 2.1. The probability distributions of these random variables are easily calculated from step (ii) of the algorithmic description. From first principles,

$$\begin{aligned} H(a_{\lambda,1^{k-1}}, a_{\nu^i,1^{k-1}}) &\leqslant \sqrt{\frac{1}{2}\left(1 - \sqrt{1 - \frac{1}{2m+1}}\right)^2 + \frac{1}{2}\left(0 - \sqrt{\frac{1}{2m+1}}\right)^2} \\ &= O\left(\frac{1}{\sqrt{m}}\right). \end{aligned} \tag{4.18}$$

In the notation of Lemma 4.8, the remaining variables are governed by

$$\left.\begin{array}{l} a_{\lambda,z} \sim \beta_{2m-1}, \\ a_{\nu^i,z} \sim \alpha_{2m-1,1} \text{ or } a_{\nu^i,z} \sim \alpha_{2m-1,2} \end{array}\right\} \qquad z \neq 1^{k-1},$$

$$\left.\begin{array}{l} b_{\lambda,z} \sim \beta_{m-1}, \\ b_{\nu^i,z} \sim \alpha_{m-1,1} \text{ or } b_{\nu^i,z} \sim \alpha_{m-1,2} \end{array}\right\} \qquad z \in \{0,1\}^{k-1},$$

where the precise distribution of $a_{\nu^i,z}$ and $b_{\nu^i,z}$ depends on f, f'. By Lemma 4.8,

$$H(a_{\lambda,z}, a_{\nu^i,z}) \leqslant \frac{c'}{\sqrt{m}}, \qquad z \neq 1^{k-1}, \tag{4.19}$$

$$H(b_{\lambda,z}, b_{\nu^i,z}) \leqslant \frac{c'}{\sqrt{m}}, \qquad z \in \{0,1\}^{k-1}, \tag{4.20}$$

for an absolute constant $c' > 0$. By (4.15) and (4.17)–(4.20), the proof is complete. ∎

Our next result shows that $\lambda_{k,m}$ is supported almost entirely on good matrix pairs.

LEMMA 4.10. *For $0 < \alpha < 1$, the probability distribution $\lambda_{k,m}$ places at most $2^{-c\alpha^2 m+k}$ probability mass on (k, m, α)-bad matrix pairs, where $c > 0$ is an absolute constant.*

Proof. Define $m' = \lceil (1-\alpha)m/2 \rceil$. Throughout the proof, we will refer to the description of $\lambda_{k,m}$ given by Algorithm 2. We may assume that $m \geqslant 2$, in which case $2m - 1 \geqslant 2m'$ and the matrix $A_{1^{k-1}}$ in the algorithm is guaranteed to have at least $2m'$ occurrences of the column

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \end{bmatrix}. \tag{4.21}$$

As a result, the output of the algorithm is (k, m, α)-good provided that the four vectors

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{k-1} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{k-1} \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{k-1} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{k-1} \\ 1 \\ 1 \end{bmatrix} \tag{4.22}$$

each occur at least $2m'$ times as a column of A_z (for $z \in \{0,1\}^{k-1}$, $z \neq 1^{k-1}$) and at least m' times as a column of B_z (for $z \in \{0,1\}^{k-1}$). Let E_{A_z} and E_{B_z} be the events that A_z and B_z, respectively, enjoy this property. Then

$$\begin{aligned} \mathbf{P}[\neg E_{A_z}] &\leqslant \binom{4m+1}{2m}^{-1} \left\{ \sum_{i=0}^{2m'-1} \binom{2m}{i}\binom{2m+1}{i+1} \right. \\ &\qquad\qquad \left. + \sum_{i=2m-2m'+1}^{2m} \binom{2m}{i}\binom{2m+1}{i+1} \right\} \\ &\leqslant \binom{4m+1}{2m}^{-1} \binom{2m+1}{m} \left\{ \sum_{i=0}^{2m'-1} \binom{2m}{i} + \sum_{i=2m-2m'+1}^{2m} \binom{2m}{i} \right\} \\ &\leqslant 2^{-\Omega(\alpha^2 m)}, \end{aligned}$$

where the final step uses Stirling's approximation and the Chernoff bound. Similarly,

$$\begin{aligned} \mathbf{P}[\neg E_{B_z}] &= \binom{2m+1}{m}^{-1} \left\{ \sum_{i=0}^{m'-1} \binom{m}{i}\binom{m+1}{i+1} + \sum_{i=m-m'+1}^{m} \binom{m}{i}\binom{m+1}{i+1} \right\} \\ &\leqslant 2^{-\Omega(\alpha^2 m)}. \end{aligned}$$

Applying a union bound over all z, we find that a (k, m, α)-bad matrix pair is generated with probability no greater than $2^{-c\alpha^2 m + k}$ for some constant $c > 0$. ∎

We now prove an analogous result for the probability distributions $\nu^0_{k,m}$ and $\nu^1_{k,m}$, showing along the way that $\nu^i_{k,m}$ can be accurately approximated by a convex combination of probability distributions in $\mathscr{G}^i_{k,m,\alpha}$.

LEMMA 4.11. *For $0 < \alpha < 1$ and any integers $k \geqslant 2$ and $m \geqslant 1$, one has*

$$\nu^i_{k,m} = \nu^{i,\mathrm{good}}_{k,m} + \nu^{i,\mathrm{bad}}_{k,m} \qquad (i = 0, 1), \tag{4.23}$$

where:

(i) $\nu^{i,\mathrm{good}}_{k,m}$ is a conical combination of probability distributions $\nu^i_{k,P_1,\dots,P_8} \in \mathscr{G}^i_{k,m,\alpha}$ such that P_1, P_2, P_4 do not contain an all-ones column,

(ii) $\|\nu^{i,\mathrm{good}}_{k,m}\|_1 \leqslant 1$,

(iii) $\|\nu^{i,\mathrm{bad}}_{k,m}\|_1 \leqslant 2^{-c\alpha^2 m+k}$ for an absolute constant $c > 0$.

Proof. Fix $i \in \{0, 1\}$ for the remainder of the proof and consider the description of $\nu^i_{k,m}$ given by Algorithm 4. Conditioned on the choice of matrices A_z, B_z in steps (i)–(ii) of the algorithm, the output is distributed according to $\nu^i_{k,P_1,\dots,P_8}$ for some $P_1, \dots, P_8$ such that P_1, P_2, P_4 do not contain an all-ones column. This gives the representation (4.23), where $\nu^{i,\mathrm{good}}_{k,m}$ and $\nu^{i,\mathrm{bad}}_{k,m}$ are conical combinations of probability distributions $\nu^i_{k,P_1,\dots,P_8} \in \mathscr{G}^i_{k,m,\alpha}$ and $\nu^i_{k,P_1,\dots,P_8} \notin \mathscr{G}^i_{k,m,\alpha}$, respectively, for which P_1, P_2, P_4 do not contain an all-ones column.

It remains to prove (iii). Define $m' = \lceil (1-\alpha)m/2 \rceil$. We may assume that $m \geqslant 2$, in which case $2m - 1 \geqslant 2m'$ and the vector (4.21) is guaranteed to occur at least $2m'$ times as a column of $A_{1^{k-1}}$ in Algorithm 4. We infer that, conditioned on steps (i)–(ii) of the algorithm, the output is (k, m, α)-good whenever the four vectors (4.22) each occur at least $2m'$ times as a column of A_z (for $z \in \{0,1\}^{k-1}, z \neq 1^{k-1}$) and at least m' times as a column of B_z (for $z \in \{0,1\}^{k-1}$). The $2^k - 1$ matrices A_z, B_z simultaneously enjoy this property with probability at least $1 - 2^{-c\alpha^2 m+k}$ for an absolute constant $c > 0$, by a calculation analogous to that in Lemma 4.10. It follows that

$$\|\nu^{i,\mathrm{bad}}_{k,m}\|_1 = 1 - \|\nu^{i,\mathrm{good}}_{k,m}\|_1 \leqslant 2^{-c\alpha^2 m+k}. \qquad \square$$

We have reached the main result of this subsection, which states that $\lambda_{k,m}$ can be accurately approximated by a convex combination of probability distributions in $\mathscr{G}^0_{k,m,\alpha}$ or $\mathscr{G}^1_{k,m,\alpha}$, with the statistical distance supported almost entirely on good matrix pairs.

THEOREM 4.12. *Let $c > 0$ be a sufficiently small absolute constant. For every $\alpha \in (0, 1)$, the probability distribution $\lambda_{k,m}$ can be expressed as*

$$\lambda_{k,m} = \lambda^i_1 + \lambda^i_2 + \lambda^i_3 \qquad (i = 0, 1), \tag{4.24}$$

where:

(i) *λ^i_1 is a conical combination of probability distributions $\nu^i_{k,P_1,\dots,P_8} \in \mathscr{G}^i_{k,m,\alpha}$ such that P_1, P_2, P_4 do not contain an all-ones column, and moreover $\|\lambda^i_1\|_1 \leqslant 1$;*

(ii) *λ^i_2 is a real function such that $\|\lambda^i_2\|_1 \leqslant \sqrt{2^k/(cm)} + 2^{-c\alpha^2 m+k}$, with support on (k, m, α)-good matrix pairs;*

(iii) *λ^i_3 is a real function with $\|\lambda^i_3\|_1 \leqslant 2^{-c\alpha^2 m+k}$.*

Proof. Decompose

$$\nu^i_{k,m} = \nu^{i,\mathrm{good}}_{k,m} + \nu^{i,\mathrm{bad}}_{k,m}$$

as in Lemma 4.11, so that

$$\|\nu_{k,m}^{i,\mathrm{bad}}\|_1 \leqslant 2^{-c'\alpha^2 m+k} \tag{4.25}$$

for some absolute constant $c' > 0$. Analogously, write

$$\lambda_{k,m} = \lambda_{k,m}^{\mathrm{good}} + \lambda_{k,m}^{\mathrm{bad}},$$

where $\lambda_{k,m}^{\mathrm{good}}$ and $\lambda_{k,m}^{\mathrm{bad}}$ are nonnegative functions supported on (k,m,α)-good and (k,m,α)-bad matrix pairs, respectively. Then

$$\|\lambda_{k,m}^{\mathrm{bad}}\|_1 \leqslant 2^{-c'\alpha^2 m+k} \tag{4.26}$$

by Lemma 4.10. Letting

$$\begin{aligned}
\lambda_1^i &= \nu_{k,m}^{i,\mathrm{good}},\\
\lambda_2^i &= \lambda_{k,m}^{\mathrm{good}} - \nu_{k,m}^{i,\mathrm{good}},\\
\lambda_3^i &= \lambda_{k,m}^{\mathrm{bad}},
\end{aligned}$$

we immediately have (4.24). Furthermore,

$$\begin{aligned}
\|\lambda_2^i\|_1 &= \|(\lambda_{k,m} - \lambda_{k,m}^{\mathrm{bad}}) - (\nu_{k,m}^i - \nu_{k,m}^{i,\mathrm{bad}})\|_1\\
&\leqslant \|\lambda_{k,m} - \nu_{k,m}^i\|_1 + \|\lambda_{k,m}^{\mathrm{bad}}\|_1 + \|\nu_{k,m}^{i,\mathrm{bad}}\|_1\\
&\leqslant \sqrt{\frac{2^k}{c''m}} + 2\cdot 2^{-c'\alpha^2 m+k}
\end{aligned} \tag{4.27}$$

for an absolute constant $c'' > 0$, where the final step uses (4.25), (4.26), and Lemma 4.9. Now items (i)–(iii) follow from Lemma 4.11(i)–(ii), (4.27), and (4.26), respectively, by taking $c = c(c', c'') > 0$ small enough. ∎

We close this subsection with a few basic observations regarding k-party protocols. On several occasions in this manuscript, we will need to argue that a communication problem does not become easier from the standpoint of communication complexity if we manipulate the protocol's input in a particular way. The input will always come in the form of a matrix sequence $(X_1, X_2, \ldots, X_n)$, and manipulations that we will encounter include discarding one or more of the arguments, reordering the arguments, applying a uniformly random column permutation to one of the arguments, adding a fixed matrix to one of the arguments, and so on. Rather than treat these instances individually as they arise, we find it more economical to address them all at once.

DEFINITION 4.13. Let $(X_1, X_2, \ldots, X_n)$ be a random variable with range $\{0,1\}^{k\times m_1} \times \{0,1\}^{k\times m_2} \times \cdots \times \{0,1\}^{k\times m_n}$. The following random variables are said to be *derivable from* $(X_1, X_2, \ldots, X_n)$ *in one step without communication:*

(i) $(X_2, \ldots, X_n)$;
(ii) $(X_1, \ldots, X_n, X_1)$;
(iii) $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$, where $\sigma \in S_n$ is a fixed permutation;
(iv) $(\sigma_1 X_1, \ldots, \sigma_n X_n)$, where $\sigma_1, \ldots, \sigma_n$ are fixed column permutations;
(v) $(X_1, \ldots, X_n, \sigma X_1)$, where σ is a uniformly random column permutation, independent of any other variables;
(vi) $(X_1, \ldots, X_n, A)$, where A is a fixed Boolean matrix;
(vii) $([X_1 \quad A_1], \ldots, [X_n \quad A_n])$, where $A_1, \ldots, A_n$ are fixed Boolean matrices;
(viii) $(X_1 \oplus A_1, \ldots, X_n \oplus A_n)$, where $A_1, \ldots, A_n$ are fixed Boolean matrices;

(ix) $(X_1, \dots, X_n, \sigma[X_1 \quad A])$, where A is a fixed Boolean matrix and σ is a uniformly random column permutation, independent of any other variables.

A random variable $(Y_1, \dots, Y_r)$ is said to be *derivable from* $(X_1, \dots, X_n)$ *with no communication*, denoted $(X_1, \dots, X_n) \rightsquigarrow (Y_1, \dots, Y_r)$, if there exists a finite sequence of random variables starting with $(X_1, \dots, X_n)$ and ending with $(Y_1, \dots, Y_r)$, where every random variable in the sequence is derivable in one step with no communication from the one immediately preceding it.

If $(Y_1, \dots, Y_r)$ is a random variable derivable from $(X_1, \dots, X_n)$ with no communication, then the former is the result of deterministic or randomized processing of the latter. In particular, $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_r)$ will in general be highly correlated. The following proposition shows that there is no advantage to providing a communication protocol with $(Y_1, \dots, Y_r)$ instead of $(X_1, \dots, X_n)$.

PROPOSITION 4.14. *Consider random variables*

$$\begin{aligned} X &= (X_1, \dots, X_n) \in \{0,1\}^{k \times m_1} \times \cdots \times \{0,1\}^{k \times m_n}, \\ X' &= (X'_1, \dots, X'_{n'}) \in \{0,1\}^{k \times m'_1} \times \cdots \times \{0,1\}^{k \times m'_{n'}}, \\ X'' &= (X''_1, \dots, X''_{n''}) \in \{0,1\}^{k \times m''_1} \times \cdots \times \{0,1\}^{k \times m''_{n''}}, \end{aligned}$$

where $X \rightsquigarrow X' \rightsquigarrow X''$. Then for every real function f,

$$\max_{\chi} \left| \mathbf{E}\, \chi(X'') f(X) \right| \leqslant \max_{\chi} \left| \mathbf{E}\, \chi(X') f(X) \right|, \tag{4.28}$$

where the maximum is over k-dimensional cylinder intersections χ.

Proof. By induction, we may assume that X'' is derivable from X' in one step with no communication. In other words, it suffices to consider cases (i)–(ix) in Definition 4.13. In what follows, we let γ denote the right-hand side of (4.28).

Cases (i)–(iv) are trivial because as a function family, cylinder intersections are closed under the operations of removing, duplicating, and reordering columns of the input matrix. For (v), we have

$$\max_{\chi} \left| \mathop{\mathbf{E}}_{\sigma} \mathop{\mathbf{E}}_{X,X'} \chi(X'_1, \dots, X'_{n'}, \sigma X'_1) f(X) \right| \leqslant \mathop{\mathbf{E}}_{\sigma} \max_{\chi} \left| \mathop{\mathbf{E}}_{X,X'} \chi(X'_1, \dots, X'_{n'}, \sigma X'_1) f(X) \right|.$$

The final expression is bounded by γ, by a combination of (ii) followed by (iv). For (vi),

$$\max_{\chi} \left| \mathop{\mathbf{E}}_{X,X'} \chi(X'_1, \dots, X'_{n'}, A) f(X) \right| \leqslant \max_{\chi} \left| \mathop{\mathbf{E}}_{X,X'} \chi(X'_1, \dots, X'_{n'}) f(X) \right|$$

because with A fixed, χ is a cylinder intersection with respect to the remaining arguments $X'_1, \dots, X'_{n'}$. The proof for (vii) is analogous. Case (viii) is immediate because as a function family, cylinder intersections are closed under the operation of adding a fixed matrix to the input matrix. Finally, (ix) is a combination of (ii), (vii), and (v), in that order. $\square$

4.3. Discrepancy analysis. Building on the work in the previous two subsections, we will now prove the desired upper bound on the repeated discrepancy of set disjointness. We start by defining the probability distribution with respect to which we will bound the discrepancy.

DEFINITION. For positive integers k, m, let $\pi_{k,m}$ be the probability distribution whose algorithmic description is as follows: choose $M \in \{T_k, F_k\}$ uniformly at random and output $[M \quad H'_{k,m}]^{\circlearrowright}$.

In words, we are interested in the probability distribution whereby true and false instances of set disjointness are generated by randomly permuting the columns of $[F_k \quad H'_{k,m}]$ and $[T_k \quad H'_{k,m}]$, respectively. For our purposes, a vital property of this probability distribution is the equivalence of the following tasks from the standpoint of communication complexity:

(i) for X drawn according to $\pi_{k,m}$, determine $\mathrm{DISJ}(X)$;
(ii) for $X_1, X_2, \ldots, X_i, \ldots$ drawn independently according to $\pi_{k,m}$ conditioned on $\mathrm{DISJ}(X_1) = \mathrm{DISJ}(X_2) = \cdots = \mathrm{DISJ}(X_i) = \cdots$, determine $\mathrm{DISJ}(X_1)$.

Thus, it does not help to have access to additional instances of set disjointness with the same truth status as the given instance. This is a *very* unusual property for a probability distribution to have, and in particular the probability distribution used in the previous best lower bound for set disjointness [48] fails badly in this regard.

This property of $\pi_{k,m}$ comes at a cost: the columns of $X \sim \pi_{k,m}$ are highly interdependent, and the inductive analysis of the discrepancy is considerably more involved than in [48]. As a matter of fact, $\pi_{k,m}$ is not directly usable in an inductive argument because it does not lead to a decomposition into subproblems with like distributions. (To be more precise, forcing an inductive argument with $\pi_{k,m}$ would result in a much weaker bound on the repeated discrepancy of set disjointness than what we prove.) Instead, we will need to analyze the discrepancy of set disjointness under a distribution more exotic than $\pi_{k,m}$, which provides the communication protocol with additional information.

A description of this exotic distribution is as follows. We will analyze the XOR of several independent instances of set disjointness, rather than a single instance. Fix a nonnegative integer d and subsets $Z_1, Z_2, \ldots, Z_n \subseteq \{0,1\}^d$. Given matrix pairs $(A_{t,z}, B_{t,z})$, where $t = 1, 2, \ldots, n$ and $z \in Z_t$, the symbol

$$\mathrm{enc}(\ldots, A_{t,z}, B_{t,z}, \ldots)$$

shall denote the following ordered list of matrices:

(i) the matrices $A_{t,z}$, listed in lexicographic order by (t, z);
(ii) the matrices $[B_{t,z} \quad B_{t,z'}]$ for all t and all $z, z' \in Z_t$ such that $|z \oplus z'| = 1$, listed in lexicographic order by (t, z, z');
(iii) the matrices $[\overline{B_{t,z}} \quad B_{t,z'}]$ for all t and all $z, z' \in Z_t$ such that $|z \oplus z'| = 1$, listed in lexicographic order by (t, z, z').

The abbreviation enc stands for "encoding" and highlights the fact that the communication protocol does not have direct access to the matrix pairs $A_{t,z}, B_{t,z}$. In particular, for $d = 0$ the matrices $B_{t,z}$ do not appear on the list $\mathrm{enc}(\ldots, A_{t,z}, B_{t,z}, \ldots)$ at all. The symbol

$$\sigma\,\mathrm{enc}(\ldots, A_{t,z}, B_{t,z}, \ldots) \tag{4.29}$$

refers to the result of permuting the columns for each of the matrices in the ordered list $\mathrm{enc}(\ldots, A_{t,z}, B_{t,z}, \ldots)$ according to σ, where $\sigma = (\ldots, \sigma_{t,z}, \ldots, \sigma_{t,z,z'}, \ldots, \sigma'_{t,z,z'}, \ldots)$ is an ordered list of column permutations, one for each of the matrices on the matrix list. In our analysis, σ will always be chosen uniformly at random, so that (4.29) is simply the result of permuting the columns for each of the matrices on the list independently and uniformly at random. With these notations in place, we define

$$\begin{aligned}
&\Gamma(k,m,d,Z_1,\dots,Z_n)\\
&\quad=\max_{\chi}\left|\mathop{\mathbf{E}}_{\dots,(A_{t,z},B_{t,z}),\dots}\ \mathop{\mathbf{E}}_{\sigma}\ \chi(\sigma\,\mathrm{enc}(\dots,A_{t,z},B_{t,z},\dots))\prod_{t=1}^{n}\prod_{z\in Z_t}\mathrm{DISJ}(A_{t,z})\right|,
\end{aligned}$$

where: the maximum is over k-dimensional cylinder intersections χ; the first expectation is over the matrix pairs $(A_{t,z},B_{t,z})$ distributed independently according to $\mu_{k,m}$; and the second expectation is over column permutations chosen independently and uniformly at random for each matrix on the list $\mathrm{enc}(\dots,A_{t,z},B_{t,z},\dots)$. This completes the description. Note the conceptual duality: from the point of view of the communication protocol, its input is an array of bits distributed according to some probability distribution, namely, the one to which we referred as "exotic" above; from the point of view of analysis, the input is an encoding of smaller and simpler objects, namely, the matrix pairs $(A_{t,z},B_{t,z})$.

For nonnegative integers $\ell_1,\ell_2,\dots,\ell_n$, we let

$$\Gamma(k,m,d,\ell_1,\dots,\ell_n)=\max_{|Z_1|=\ell_1}\cdots\max_{|Z_n|=\ell_n}\Gamma(k,m,d,Z_1,\dots,Z_n),$$

where the maximum is over all possible subsets $Z_1,Z_2,\dots,Z_n\subseteq\{0,1\}^d$ of cardinalities $\ell_1,\ell_2,\dots,\ell_n$, respectively. Observe that $\Gamma(k,m,d,\ell_1,\dots,\ell_n)$ is only defined for $\ell_1,\dots,\ell_n\in\{0,1,2,3,\dots,2^d\}$. The only setting of interest to us is $d=0$ and $\ell_1=\ell_2=\cdots=\ell_n=1$, in which case $\mathrm{enc}(\dots,A_{t,z},B_{t,z},\dots)=(\dots,A_{t,z},\dots)$ and

$$\begin{aligned}
&\Gamma(k,m,0,\overbrace{1,\dots,1}^{n})\\
&\qquad\qquad=\max_{\chi}\left|\mathop{\mathbf{E}}_{X_1,\dots,X_n\sim\pi_{k,2m}}\ \chi(X_1,\dots,X_n)\prod_{i=1}^{n}\mathrm{DISJ}(X_i)\right|. \qquad (4.30)
\end{aligned}$$

However, the inductive analysis below requires consideration of $\Gamma(k,m,d,\ell_1,\dots,\ell_n)$ for all possible parameters. We start by deriving a recurrence relation for Γ.

LEMMA 4.15. *Let $c>0$ be the absolute constant from Theorem 4.12. Then for $0<\alpha<1$ and $k\geqslant 2$, the quantity $\Gamma(k,m,d,\ell_1,\dots,\ell_n)^2$ does not exceed*

$$\begin{aligned}
&\sum_{i_1,j_1=0}^{\ell_1}\cdots\sum_{i_n,j_n=0}^{\ell_n}\left\{\prod_{t=1}^{n}\binom{\ell_t}{i_t}\binom{\ell_t-i_t}{j_t}\left(\sqrt{\frac{2^k}{cm}}+\frac{2^k}{2^{c\alpha^2 m}}\right)^{i_t}\left(\frac{2^k}{2^{c\alpha^2 m}}\right)^{j_t}\right\}\\
&\quad\times\max_{\substack{\ell_1'\geqslant 2\max\{0,\ell_1-i_1-(d+1)j_1\}\\ \vdots\\ \ell_n'\geqslant 2\max\{0,\ell_n-i_n-(d+1)j_n\}}}\Gamma\left(k-1,\left\lceil\frac{(1-\alpha)m}{2}\right\rceil,d+1,\ell_1',\dots,\ell_n'\right).
\end{aligned}$$

Moreover,

$$\Gamma(1,m,d,\ell_1,\dots,\ell_n)=\begin{cases}0 & \textit{if } \ell_1+\cdots+\ell_n>0,\\ 1 & \textit{otherwise.}\end{cases}$$

Proof. The claim regarding $\Gamma(1, m, d, \ell_1, \ldots, \ell_n)$ is obvious because the probability distribution $\mu_{k,m}$ places equal weight on the positive and negative instances of set disjointness. In what follows, we prove the recurrence relation.

Abbreviate $\Gamma = \Gamma(k, m, d, \ell_1, \ldots, \ell_n)$. Let $Z_1, \ldots, Z_n \subseteq \{0,1\}^d$ be subsets of cardinalities $\ell_1, \ldots, \ell_n$, respectively, such that $\Gamma(k, m, d, Z_1, \ldots, Z_n) = \Gamma$. Let χ be a k-dimensional cylinder intersection for which

$$\Gamma = \Bigg| \mathop{\mathbf{E}}_{\ldots, \left[\begin{smallmatrix} A_{t,z} \\ a_{t,z} \end{smallmatrix}\right], \left[\begin{smallmatrix} B_{t,z} \\ b_{t,z} \end{smallmatrix}\right], \ldots} \; \mathop{\mathbf{E}}_{\sigma} \; \chi\left(\sigma \operatorname{enc}\left(\ldots, \begin{bmatrix} A_{t,z} \\ a_{t,z} \end{bmatrix}, \begin{bmatrix} B_{t,z} \\ b_{t,z} \end{bmatrix}, \ldots\right)\right) \times \prod_{t=1}^{n} \prod_{z \in Z_t} \operatorname{DISJ} \begin{bmatrix} A_{t,z} \\ a_{t,z} \end{bmatrix} \Bigg|,$$

where the inner expectation is over the independent permutation of the columns for each of the matrices on the encoded list, and the outer expectation is over matrix pairs

$$\left(\begin{bmatrix} A_{t,z} \\ a_{t,z} \end{bmatrix}, \begin{bmatrix} B_{t,z} \\ b_{t,z} \end{bmatrix} \right), \qquad t = 1, 2, \ldots, n, \qquad z \in Z_t,$$

each drawn independently according to $\mu_{k,m}$ (as usual, $a_{t,z}$ and $b_{t,z}$ denote row vectors). The starting point in the proof is a reduction to $(k-1)$-dimensional cylinder intersections using the Cauchy-Schwarz inequality, a technique due to Babai, Nisan, and Szegedy [7]. Rearranging,

$$\Gamma \leqslant \mathop{\mathbf{E}}_{\sigma} \mathop{\mathbf{E}}_{\ldots, A_{t,z}, B_{t,z}, \ldots} \Bigg| \mathop{\mathbf{E}}_{\ldots, a_{t,z}, b_{t,z}, \ldots} \chi\left(\sigma \operatorname{enc}\left(\ldots, \begin{bmatrix} A_{t,z} \\ a_{t,z} \end{bmatrix}, \begin{bmatrix} B_{t,z} \\ b_{t,z} \end{bmatrix}, \ldots\right)\right) \times \prod_{t=1}^{n} \prod_{z \in Z_t} \operatorname{DISJ} \begin{bmatrix} A_{t,z} \\ a_{t,z} \end{bmatrix} \Bigg|, \tag{4.31}$$

where the second expectation is over the marginal probability distribution on the pairs $(A_{t,z}, B_{t,z})$, and the third expectation is over the conditional probability distribution on the pairs $(a_{t,z}, b_{t,z})$ for fixed $(A_{t,z}, B_{t,z})$. Recall that χ is the pointwise product of two functions $\chi = \phi \cdot \chi'$, where ϕ depends only on the first $k-1$ rows and has range $\{0,1\}$, and χ' is a $(k-1)$-dimensional cylinder intersection with respect to the first $k-1$ rows for any fixed value of the kth row. Since the innermost expectation in (4.31) is over $(\ldots, a_{t,z}, b_{t,z}, \ldots)$ for fixed $(\ldots, A_{t,z}, B_{t,z}, \ldots)$, the function ϕ can be taken outside the innermost expectation and absorbed into the absolute value operator:

$$\Gamma \leqslant \mathop{\mathbf{E}}_{\sigma} \mathop{\mathbf{E}}_{\ldots, A_{t,z}, B_{t,z}, \ldots} \Bigg| \mathop{\mathbf{E}}_{\ldots, a_{t,z}, b_{t,z}, \ldots} \chi'\left(\sigma \operatorname{enc}\left(\ldots, \begin{bmatrix} A_{t,z} \\ a_{t,z} \end{bmatrix}, \begin{bmatrix} B_{t,z} \\ b_{t,z} \end{bmatrix}, \ldots\right)\right) \times \prod_{t=1}^{n} \prod_{z \in Z_t} \operatorname{DISJ} \begin{bmatrix} A_{t,z} \\ a_{t,z} \end{bmatrix} \Bigg|.$$

Squaring both sides and applying the Cauchy-Schwarz inequality,

$$\begin{aligned}\Gamma^2 \leqslant \mathop{\mathbf{E}}_{\sigma} \mathop{\mathbf{E}}_{\dots,A_{t,z},B_{t,z},\dots} \Bigg[\Bigg\{ \mathop{\mathbf{E}}_{\dots,a_{t,z},b_{t,z},\dots} \chi'\left(\sigma \operatorname{enc}\left(\dots,\begin{bmatrix}A_{t,z}\\a_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b_{t,z}\end{bmatrix},\dots\right)\right) \\ \times \prod_{t=1}^{n}\prod_{z\in Z_t} \operatorname{DISJ}\begin{bmatrix}A_{t,z}\\a_{t,z}\end{bmatrix}\Bigg\}^2\Bigg] \\ = \mathop{\mathbf{E}}_{\dots,\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix},\dots} \mathop{\mathbf{E}}_{\sigma} \chi'\left(\sigma \operatorname{enc}\left(\dots,\begin{bmatrix}A_{t,z}\\a_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b_{t,z}\end{bmatrix},\dots\right)\right)\times \\ \chi'\left(\sigma \operatorname{enc}\left(\dots,\begin{bmatrix}A_{t,z}\\a'_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b'_{t,z}\end{bmatrix},\dots\right)\right)\prod_{t=1}^{n}\prod_{z\in Z_t}\operatorname{DISJ}\begin{bmatrix}A_{t,z}\\a_{t,z}\end{bmatrix}\operatorname{DISJ}\begin{bmatrix}A_{t,z}\\a'_{t,z}\end{bmatrix},\end{aligned}$$

where the outer expectation is over matrix pairs drawn according to $\lambda_{k,m}$. Since the product of two cylinder intersections is a cylinder intersection, we arrive at

$$\begin{aligned}\Gamma^2 \leqslant \mathop{\mathbf{E}}_{\dots,\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix},\dots} \mathop{\mathbf{E}}_{\sigma} \chi''\left(\sigma \operatorname{enc}\left(\dots,\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix},\dots\right)\right) \\ \times\prod_{t=1}^{n}\prod_{z\in Z_t}\operatorname{DISJ}\begin{bmatrix}A_{t,z}\\a_{t,z}\end{bmatrix}\operatorname{DISJ}\begin{bmatrix}A_{t,z}\\a'_{t,z}\end{bmatrix}, \qquad (4.32)\end{aligned}$$

where χ'' is a $(k-1)$-dimensional cylinder intersection with respect to the first $k-1$ rows for any fixed value of the kth and $(k+1)$st rows. This completes the promised reduction to the $(k-1)$-dimensional case.

Theorem 4.12 states that

$$\lambda_{k,m} = \lambda_1^i + \lambda_2^i + \lambda_3^i \qquad (i = 0, 1), \qquad (4.33)$$

where: λ_1^i is a conical combination of probability distributions $\nu^i_{k,P_1,\dots,P_8} \in \mathscr{G}^i_{k,m,\alpha}$ for which P_1, P_2, P_4 do not contain an all-ones column; λ_2^i is a real function supported on (k, m, α)-good matrix pairs; and furthermore

$$\|\lambda_1^i\|_1 \leqslant 1 \qquad (i = 0, 1), \qquad (4.34)$$

$$\|\lambda_2^i\|_1 \leqslant \sqrt{\frac{2^k}{cm}} + \frac{2^k}{2^{c\alpha^2 m}} \qquad (i = 0, 1), \qquad (4.35)$$

$$\|\lambda_3^i\|_1 \leqslant \frac{2^k}{2^{c\alpha^2 m}} \qquad (i = 0, 1). \qquad (4.36)$$

Define

$$\begin{aligned}\Phi\left(\dots,\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix},\dots\right) = \mathop{\mathbf{E}}_{\sigma} \chi''\left(\sigma \operatorname{enc}\left(\dots,\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix},\dots\right)\right) \\ \times\prod_{t=1}^{n}\prod_{z\in Z_t}\operatorname{DISJ}\begin{bmatrix}A_{t,z}\\a_{t,z}\end{bmatrix}\operatorname{DISJ}\begin{bmatrix}A_{t,z}\\a'_{t,z}\end{bmatrix}.\end{aligned}$$

CLAIM 4.16. *Fix functions* $\iota_t \colon Z_t \to \{1,2,3\}$ $(t = 1,2,\ldots,n)$. *Define* $i_t = |\iota_t^{-1}(2)|$ *and* $j_t = |\iota_t^{-1}(3)|$. *Then*

$$\left\langle \Phi, \bigotimes_{t=1}^{n} \bigotimes_{z \in Z_t} \lambda_{\iota_t(z)}^{\mathrm{PARITY}^*(z)} \right\rangle \leqslant \left\{ \prod_{t=1}^{n} \left(\sqrt{\frac{2^k}{cm}} + \frac{2^k}{2^{c\alpha^2 m}} \right)^{i_t} \left(\frac{2^k}{2^{c\alpha^2 m}} \right)^{j_t} \right\}$$
$$\times \max_{\substack{\ell_1' \geqslant 2\max\{0,\ell_1 - i_1 - (d+1)j_1\} \\ \vdots \\ \ell_n' \geqslant 2\max\{0,\ell_n - i_n - (d+1)j_n\}}} \Gamma\left(k-1, \left\lceil \frac{(1-\alpha)m}{2} \right\rceil, d+1, \ell_1', \ldots, \ell_n'\right).$$

Before settling the claim, we will finish the proof of the lemma:

$$\begin{aligned}
\Gamma^2 &\leqslant \left\langle \Phi, \bigotimes_{t=1}^{n} \bigotimes_{z \in Z_t} \lambda_{k,m} \right\rangle && \text{by (4.32)} \\
&= \left\langle \Phi, \bigotimes_{t=1}^{n} \bigotimes_{z \in Z_t} \left(\lambda_1^{\mathrm{PARITY}^*(z)} + \lambda_2^{\mathrm{PARITY}^*(z)} + \lambda_3^{\mathrm{PARITY}^*(z)} \right) \right\rangle && \text{by (4.33)} \\
&= \sum_{\iota_1, \iota_2, \ldots, \iota_n} \left\langle \Phi, \bigotimes_{t=1}^{n} \bigotimes_{z \in Z_t} \lambda_{\iota_t(z)}^{\mathrm{PARITY}^*(z)} \right\rangle,
\end{aligned}$$

where the sum is over all possible functions $\iota_1, \iota_2, \ldots, \iota_n$ with domains $Z_1, Z_2, \ldots, Z_n$, respectively, and range $\{1,2,3\}$. Using the bound of Claim 4.16 for the inner products in the final expression, one immediately arrives at the recurrence in the statement of the lemma. ∎

Proof of Claim 4.16. For $t = 1,2,\ldots,n$, define Y_t to be the collection of all $z \in \iota_t^{-1}(1)$ for which $\{z' \in Z_t : |z \oplus z'| = 1\} \cap \iota_t^{-1}(3) = \varnothing$. This set $Y_t \subseteq Z_t$ has the following intuitive interpretation. View Z_t as an undirected graph in which two vertices $z, z' \in Z_t$ are connected by an edge if and only if they are neighbors in the ambient hypercube, i.e., $|z \oplus z'| = 1$. We will refer to the vertices in $\iota_t^{-1}(1), \iota_t^{-1}(2)$, and $\iota_t^{-1}(3)$ as *good, neutral,* and *bad,* respectively. In this terminology, Y_t is simply the set of all good vertices that do not have a bad neighbor. Since the degree of every vertex in the graph is at most d, we obtain

$$\begin{aligned}
|Y_t| &\geqslant |\iota_t^{-1}(1)| - d|\iota_t^{-1}(3)| \\
&= (|Z_t| - |\iota_t^{-1}(2)| - |\iota_t^{-1}(3)|) - d|\iota_t^{-1}(3)| \\
&= \ell_t - i_t - (d+1)j_t, \qquad t = 1,2,\ldots,n. \qquad (4.37)
\end{aligned}$$

Now, consider the quantity

$$\gamma = \max \left| \mathop{\mathbf{E}}_{\dots,\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix},\dots} \mathop{\mathbf{E}}_{\sigma} \chi''\left(\sigma \operatorname{enc}\left(\dots, \begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix}, \begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix}, \dots\right)\right) \right.$$
$$\left. \times \prod_{t=1}^{n} \prod_{z\in Z_t} \mathrm{DISJ}\begin{bmatrix}A_{t,z}\\a_{t,z}\end{bmatrix} \mathrm{DISJ}\begin{bmatrix}A_{t,z}\\a'_{t,z}\end{bmatrix} \right|, \quad (4.38)$$

where: the maximum is over all matrix pairs

$$\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix}, \begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix}, \qquad t = 1,2,\dots,n, \quad z \in (\iota_t^{-1}(1) \setminus Y_t) \cup \iota_t^{-1}(2) \quad (4.39)$$

that are (k,m,α)-good and over all possible matrix pairs

$$\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix}, \begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix}, \qquad t = 1,2,\dots,n, \quad z \in \iota_t^{-1}(3), \quad (4.40)$$

and the outer expectation is over the remaining matrix pairs

$$\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix}, \begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix}, \qquad t = 1,2,\dots,n, \quad z \in Y_t \quad (4.41)$$

which are distributed independently, each according to some distribution $\nu_{k,P_1,\dots,P_8}^{\mathrm{PARITY}^*(z)} \in \mathscr{G}_{k,m,\alpha}^{\mathrm{PARITY}^*(z)}$ such that P_1, P_2, P_4 do not contain an all-ones column. Since $\lambda_2^{\mathrm{PARITY}^*(z)}$ is supported on (k,m,α)-good matrix pairs and since $\lambda_1^{\mathrm{PARITY}^*(z)}$ is a conical combination of probability distributions $\nu_{k,P_1,\dots,P_8}^{\mathrm{PARITY}^*(z)} \in \mathscr{G}_{k,m,\alpha}^{\mathrm{PARITY}^*(z)}$ such that P_1, P_2, P_4 do not contain an all-ones column, it follows by convexity that

$$\begin{aligned}\left\langle \Phi, \bigotimes_{t=1}^{n} \bigotimes_{z\in Z_t} \lambda_{\iota_t(z)}^{\mathrm{PARITY}^*(z)} \right\rangle &\leqslant \gamma \prod_{t=1}^{n} \prod_{z\in Z_t} \left\| \lambda_{\iota_t(z)}^{\mathrm{PARITY}^*(z)} \right\|_1 \\ &\leqslant \gamma \prod_{t=1}^{n} \left(\sqrt{\frac{2^k}{cm}} + \frac{2^k}{2^{c\alpha^2 m}} \right)^{i_t} \left(\frac{2^k}{2^{c\alpha^2 m}} \right)^{j_t},\end{aligned}$$

where the second step uses the estimates (4.34)–(4.36). As a result, the proof will be complete once we show that

$$\gamma \leqslant \max_{\substack{\ell'_1 \geqslant 2\max\{0,\ell_1 - i_1 - (d+1)j_1\} \\ \vdots \\ \ell'_n \geqslant 2\max\{0,\ell_n - i_n - (d+1)j_n\}}} \Gamma(k-1, m', d+1, \ell'_1, \dots, \ell'_n), \quad (4.42)$$

where $m' = \lceil (1-\alpha)m/2 \rceil$. In the remainder of the proof, we will fix an assignment to the matrix pairs (4.39) and (4.40) for which the maximum is achieved in (4.38). The argument involves three steps: splitting the input to χ'' into tuples of smaller matrices, determining the individual probability distribution of each tuple, and recombining the results to characterize the joint probability distribution of the input to χ''.

Step I: partitioning into submatrices. Think of every matrix M on the encoded matrix list in (4.38) as partitioned into four submatrices $M^{00}, M^{01}, M^{10}, M^{11}$ of the form

$$\begin{bmatrix} & * & & \\ \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \begin{bmatrix} & * & & \\ \\ 0 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix}, \begin{bmatrix} & * & & \\ \\ 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \begin{bmatrix} & * & & \\ \\ 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \end{bmatrix},$$

respectively, with the relative ordering of columns in each submatrix inherited from the original matrix M. A uniformly random column permutation of M can be realized as

$$\upsilon\left[\sigma^{00}M^{00} \quad \sigma^{01}M^{01} \quad \sigma^{10}M^{10} \quad \sigma^{11}M^{11}\right],$$

where $\sigma^{00},\ldots,\sigma^{11}$ are uniformly random column permutations of the four submatrices and υ is a uniformly random column permutation of the entire matrix. We will reveal υ completely to the cylinder intersection (this corresponds to allowing the cylinder intersection to depend on υ) but keep $\sigma^{00},\ldots,\sigma^{11}$ secret.

In more detail, define

$$\begin{aligned} A^{00}_{t,z} &= A_{t,z}|_{\overline{a_{t,z}}\wedge\overline{a'_{t,z}}}, & B^{00}_{t,z} &= B_{t,z}|_{\overline{b_{t,z}}\wedge\overline{b'_{t,z}}}, \\ A^{01}_{t,z} &= A_{t,z}|_{\overline{a_{t,z}}\wedge a'_{t,z}}, & B^{01}_{t,z} &= B_{t,z}|_{\overline{b_{t,z}}\wedge b'_{t,z}}, \\ A^{10}_{t,z} &= A_{t,z}|_{a_{t,z}\wedge\overline{a'_{t,z}}}, & B^{10}_{t,z} &= B_{t,z}|_{b_{t,z}\wedge\overline{b'_{t,z}}}, \\ A^{11}_{t,z} &= A_{t,z}|_{a_{t,z}\wedge a'_{t,z}}, & B^{11}_{t,z} &= B_{t,z}|_{b_{t,z}\wedge b'_{t,z}}, \end{aligned}$$

where $t=1,2,\ldots,n$ and $z\in Z_t$. Then by the argument of the previous paragraph,

$$\begin{aligned}\gamma \leqslant \Bigg| \mathop{\mathbf{E}}_{\ldots,\begin{bmatrix}A_{t,z}\\a_{t,z}\\a'_{t,z}\end{bmatrix},\begin{bmatrix}B_{t,z}\\b_{t,z}\\b'_{t,z}\end{bmatrix},\ldots} \; \mathop{\mathbf{E}}_{\upsilon} \; \mathop{\mathbf{E}}_{\sigma^{00},\ldots,\sigma^{11}} \; \prod_{t=1}^{n}\prod_{z\in Z_t} \mathrm{DISJ}\begin{bmatrix}A_{t,z}\\a_{t,z}\end{bmatrix}\mathrm{DISJ}\begin{bmatrix}A_{t,z}\\a'_{t,z}\end{bmatrix}\times \\ \times\chi''_{\upsilon}(\sigma^{00}\,\mathrm{enc}(\ldots,A^{00}_{t,z},B^{00}_{t,z},\ldots), \\ \sigma^{01}\,\mathrm{enc}(\ldots,A^{01}_{t,z},B^{01}_{t,z},\ldots), \\ \sigma^{10}\,\mathrm{enc}(\ldots,A^{10}_{t,z},B^{10}_{t,z},\ldots), \\ \sigma^{11}\,\mathrm{enc}(\ldots,A^{11}_{t,z},B^{11}_{t,z},\ldots))\Bigg|,\end{aligned}$$

where $\sigma^{00},\sigma^{01},\sigma^{10},\sigma^{11}$ are permutation lists chosen independently and uniformly at random, υ is a joint column permutation from an appropriate probability distribution, and each χ''_{υ} is a $(k-1)$-dimensional cylinder intersection. Note that the final property crucially uses the fact that χ'' is a $(k-1)$-dimensional cylinder intersection for any fixed value of the bottom two rows. Taking the expectation with respect to υ outside the absolute value operator, we conclude that there is some $(k-1)$-dimensional cylinder intersection χ''' such that

$$\gamma \leqslant \Bigg| \mathop{\mathbf{E}}_{\ldots,\begin{bmatrix} A_{t,z} \\ a_{t,z} \\ a'_{t,z} \end{bmatrix},\begin{bmatrix} B_{t,z} \\ b_{t,z} \\ b'_{t,z} \end{bmatrix},\ldots} \;\; \mathop{\mathbf{E}}_{\sigma^{00},\sigma^{01},\sigma^{10},\sigma^{11}} \;\; \prod_{t=1}^{n} \prod_{z \in Z_t} \mathrm{DISJ}\begin{bmatrix} A_{t,z} \\ a_{t,z} \end{bmatrix} \mathrm{DISJ}\begin{bmatrix} A_{t,z} \\ a'_{t,z} \end{bmatrix} \times$$

$$\begin{aligned} \times \chi'''(\sigma^{00}\,&\mathrm{enc}(\ldots, A^{00}_{t,z}, B^{00}_{t,z}, \ldots), \\ \sigma^{01}\,&\mathrm{enc}(\ldots, A^{01}_{t,z}, B^{01}_{t,z}, \ldots), \\ \sigma^{10}\,&\mathrm{enc}(\ldots, A^{10}_{t,z}, B^{10}_{t,z}, \ldots), \\ \sigma^{11}\,&\mathrm{enc}(\ldots, A^{11}_{t,z}, B^{11}_{t,z}, \ldots))\Bigg|. \end{aligned} \tag{4.43}$$

Step II: distribution of the induced matrix sequences. We will now take a closer look at the matrix sequence $(A^{00}_{t,z}, A^{01}_{t,z}, A^{10}_{t,z}, A^{11}_{t,z}, B^{00}_{t,z}, B^{01}_{t,z}, B^{10}_{t,z}, B^{11}_{t,z})$ and characterize its distribution depending on t, z. In what follows, the symbol $*$ denotes a fixed Boolean matrix, and the symbol $\star$ denotes a fixed Boolean matrix without an all-ones column. We will use $*$ and $\star$ to designate matrices whose entries are immaterial to the proof. It is important to remember that $*$ and $\star$ are semantic shorthands rather than variables, i.e., every occurrence of $*$ and $\star$ may refer to a different matrix.

(a) *Sequences with* $t = 1, 2, \ldots, n,\ z \in (\iota_t^{-1}(1) \setminus Y_t) \cup \iota_t^{-1}(2)$. For such t, z, the matrices $(A_{t,z}, B_{t,z})$ are fixed to some (k, m, α)-good matrix pairs, which by definition forces

$$\begin{aligned} A^{00}_{t,z} &=_{\circlearrowright} \begin{bmatrix} * & H_{k-1,2m'} \end{bmatrix}, & B^{00}_{t,z} &=_{\circlearrowright} \begin{bmatrix} * & H_{k-1,m'} \end{bmatrix}, \\ A^{01}_{t,z} &=_{\circlearrowright} \begin{bmatrix} * & H'_{k-1,2m'} \end{bmatrix}, & B^{01}_{t,z} &=_{\circlearrowright} \begin{bmatrix} * & H_{k-1,m'} \end{bmatrix}, \\ A^{10}_{t,z} &=_{\circlearrowright} \begin{bmatrix} * & H'_{k-1,2m'} \end{bmatrix}, & B^{10}_{t,z} &=_{\circlearrowright} \begin{bmatrix} * & H_{k-1,m'} \end{bmatrix}, \\ A^{11}_{t,z} &=_{\circlearrowright} \begin{bmatrix} * \end{bmatrix}, & B^{11}_{t,z} &=_{\circlearrowright} \begin{bmatrix} * & H_{k-1,m'} \end{bmatrix}. \end{aligned}$$

(b) *Sequences with* $t = 1, 2, \ldots, n,\ z \in \iota^{-1}(3)$. Each such sequence is fixed to some unknown tuple of matrices over which we have no control:

$$\begin{aligned} A^{00}_{t,z} &= \begin{bmatrix} * \end{bmatrix}, & B^{00}_{t,z} &= \begin{bmatrix} * \end{bmatrix}, \\ A^{01}_{t,z} &= \begin{bmatrix} * \end{bmatrix}, & B^{01}_{t,z} &= \begin{bmatrix} * \end{bmatrix}, \\ A^{10}_{t,z} &= \begin{bmatrix} * \end{bmatrix}, & B^{10}_{t,z} &= \begin{bmatrix} * \end{bmatrix}, \\ A^{11}_{t,z} &= \begin{bmatrix} * \end{bmatrix}, & B^{11}_{t,z} &= \begin{bmatrix} * \end{bmatrix}. \end{aligned}$$

(c) *Sequences with* $t = 1, 2, \ldots, n,\ z \in Y_t$. Each such sequence is distributed independently of the others. The exact distribution of a given sequence depends on the parity of z and is given by the following table, where $M_{t,z0}, M_{t,z1}$ refer to independent random variables distributed uniformly in $\{T_{k-1}, F_{k-1}\}$.

$$\begin{array}{lll}
 & \textbf{Distribution for } |z| \textbf{ even} & \textbf{Distribution for } |z| \textbf{ odd} \\
A^{00}_{t,z} & \begin{bmatrix} * & H_{k-1,2m'} & \overline{M_{t,z0}} & \overline{M_{t,z1}} \end{bmatrix}^{\circlearrowleft} & \begin{bmatrix} * & H_{k-1,2m'} & \overline{M_{t,z0}} & \overline{M_{t,z1}} \end{bmatrix}^{\circlearrowleft} \\
A^{01}_{t,z} & \begin{bmatrix} \star & H'_{k-1,2m'} & M_{t,z0} \end{bmatrix}^{\circlearrowleft} & \begin{bmatrix} \star & H'_{k-1,2m'} & M_{t,z1} \end{bmatrix}^{\circlearrowleft} \\
A^{10}_{t,z} & \begin{bmatrix} \star & H'_{k-1,2m'} & M_{t,z1} \end{bmatrix}^{\circlearrowleft} & \begin{bmatrix} \star & H'_{k-1,2m'} & M_{t,z0} \end{bmatrix}^{\circlearrowleft} \\
A^{11}_{t,z} & \begin{bmatrix} \star \end{bmatrix}^{\circlearrowleft} & \begin{bmatrix} \star \end{bmatrix}^{\circlearrowleft} \\
B^{00}_{t,z} & \begin{bmatrix} * & H_{k-1,m'} & \overline{M_{t,z0}} & \overline{M_{t,z1}} \end{bmatrix}^{\circlearrowleft} & \begin{bmatrix} * & H_{k-1,m'} \end{bmatrix}^{\circlearrowleft} \\
B^{01}_{t,z} & \begin{bmatrix} * & H_{k-1,m'} & M_{t,z0} \end{bmatrix}^{\circlearrowleft} & \begin{bmatrix} * & H_{k-1,m'} & \overline{M_{t,z0}} \end{bmatrix}^{\circlearrowleft} \\
B^{10}_{t,z} & \begin{bmatrix} * & H_{k-1,m'} & M_{t,z1} \end{bmatrix}^{\circlearrowleft} & \begin{bmatrix} * & H_{k-1,m'} & \overline{M_{t,z1}} \end{bmatrix}^{\circlearrowleft} \\
B^{11}_{t,z} & \begin{bmatrix} * & H_{k-1,m'} \end{bmatrix}^{\circlearrowleft} & \begin{bmatrix} * & H_{k-1,m'} & M_{t,z0} & M_{t,z1} \end{bmatrix}^{\circlearrowleft}
\end{array}$$

To verify, recall that each matrix pair in (4.41) is distributed independently according to $\nu^{\mathrm{PARITY}^*(z)}_{k,P_1,\ldots,P_8} \in \mathscr{G}^{\mathrm{PARITY}^*(z)}_{k,m,\alpha}$ for some $P_1,\ldots,P_8$, where P_1, P_2, P_4 do not contain an all-ones column. The stated description is now immediate by letting

$$(M_1, M_2) = \begin{cases} (M_{t,z1}, M_{t,z0}) & \text{if } |z| \text{ is even,} \\ (M_{t,z0}, M_{t,z1}) & \text{if } |z| \text{ is odd} \end{cases}$$

in Definition 4.6 and recalling that $P_1,\ldots,P_8$ have submatrix structure given by Proposition 4.7.

An important consequence of the newly obtained characterization is that

$$\begin{aligned}
\mathrm{DISJ}\begin{bmatrix} A_{t,z} \\ a_{t,z} \end{bmatrix} \mathrm{DISJ}\begin{bmatrix} A_{t,z} \\ a'_{t,z} \end{bmatrix} &= \mathrm{DISJ}\begin{bmatrix} A^{10}_{t,z} & A^{11}_{t,z} \end{bmatrix} \mathrm{DISJ}\begin{bmatrix} A^{01}_{t,z} & A^{11}_{t,z} \end{bmatrix} \\
&= \mathrm{DISJ}(M_{t,z0})\,\mathrm{DISJ}(M_{t,z1})
\end{aligned}$$

for all $z \in Y_t$. Since for $z \notin Y_t$ the values $A_{t,z}, a_{t,z}, a'_{t,z}$ are fixed, (4.43) simplifies to

$$\begin{aligned}
\gamma \leqslant \Bigg| \mathop{\mathbf{E}}_{\ldots,\begin{bmatrix} A_{t,z} \\ a_{t,z} \\ a'_{t,z} \end{bmatrix},\begin{bmatrix} B_{t,z} \\ b_{t,z} \\ b'_{t,z} \end{bmatrix},\ldots} \; \mathop{\mathbf{E}}_{\sigma^{00},\sigma^{01},\sigma^{10},\sigma^{11}} \; \prod_{t=1}^{n} \prod_{z \in Y_t} \mathrm{DISJ}(M_{t,z0})\mathrm{DISJ}(M_{t,z1}) \times \\
\times \chi'''(\sigma^{00}\,\mathrm{enc}(\ldots, A^{00}_{t,z}, B^{00}_{t,z}, \ldots), \\
\sigma^{01}\,\mathrm{enc}(\ldots, A^{01}_{t,z}, B^{01}_{t,z}, \ldots), \\
\sigma^{10}\,\mathrm{enc}(\ldots, A^{10}_{t,z}, B^{10}_{t,z}, \ldots), \\
\sigma^{11}\,\mathrm{enc}(\ldots, A^{11}_{t,z}, B^{11}_{t,z}, \ldots)) \Bigg|.
\end{aligned} \tag{4.44}$$

Step III: recombining. Having examined the new submatrices, we are now in a position to fully characterize the probability distribution of the input to χ''' in (4.44). To start with, χ''' receives as input the matrices $A^{00\circlearrowleft}_{t,z}, A^{01\circlearrowleft}_{t,z}, A^{10\circlearrowleft}_{t,z}, A^{11\circlearrowleft}_{t,z}$. If $z \in Y_t$, then by Step II (c) each of them is distributed according to one of the distributions

$$\begin{bmatrix} * & H_{k-1,2m'} & \overline{M_{t,z0}} & \overline{M_{t,z1}} \end{bmatrix}^{\circlearrowleft}, \tag{4.45}$$

$$\begin{bmatrix} * & H'_{k-1,2m'} & M_{t,z0} \end{bmatrix}^{\circlearrowleft}, \tag{4.46}$$

$$\begin{bmatrix} * & H'_{k-1,2m'} & M_{t,z1} \end{bmatrix}^{\circlearrowleft}, \tag{4.47}$$

$$\begin{bmatrix} * \end{bmatrix}^{\circlearrowleft}. \tag{4.48}$$

If $z \notin Y_t$, then each of the matrices in question is distributed according to (4.48). The only other input to χ''' is

$$[B^{\varepsilon_1\varepsilon_2}_{t,z} \quad B^{\varepsilon_1\varepsilon_2}_{t,z'}]^{\circlearrowleft}, \quad [\overline{B^{\varepsilon_1\varepsilon_2}_{t,z}} \quad B^{\varepsilon_1\varepsilon_2}_{t,z'}]^{\circlearrowleft}, \tag{4.49}$$

where $\varepsilon_1, \varepsilon_2 \in \{0, 1\}$ and the strings $z, z' \in Z_t$ satisfy $|z \oplus z'| = 1$. If $z, z' \in Y_t$, then Step II (c) reveals that each of the matrices in (4.49) is distributed according to one of the probability distributions

$$\begin{bmatrix} * & H_{k-1,2m'} & M_{t,w} & M_{t,w'} \end{bmatrix}^{\circlearrowleft}, \tag{4.50}$$

$$\begin{bmatrix} * & H_{k-1,2m'} & \overline{M_{t,w}} & M_{t,w'} \end{bmatrix}^{\circlearrowleft}, \tag{4.51}$$

$$\begin{bmatrix} * & H_{k-1,2m'} & \overline{M_{t,w}} & \overline{M_{t,w'}} \end{bmatrix}^{\circlearrowleft}, \tag{4.52}$$

where $w, w' \in Y_t \times \{0, 1\}$ are some Boolean strings with $|w \oplus w'| = 1$. If $z \notin Y_t$ and $z' \notin Y_t$, then each of the matrices in (4.49) is distributed according to (4.48). In the remaining case when $z \in Y_t$ and $z' \notin Y_t$, we have by definition of Y_t that $z' \in (\iota_t^{-1}(1) \setminus Y_t) \cup \iota_t^{-1}(2)$, and therefore by Step II (a)(c) each of the matrices in (4.49) is distributed according to one of the probability distributions

$$\begin{bmatrix} * \end{bmatrix}^{\circlearrowleft}, \tag{4.53}$$

$$\begin{bmatrix} * & H_{k-1,2m'} & M_{t,z0} & M_{t,z1} \end{bmatrix}^{\circlearrowleft}, \tag{4.54}$$

$$\begin{bmatrix} * & H_{k-1,2m'} & \overline{M_{t,z0}} & \overline{M_{t,z1}} \end{bmatrix}^{\circlearrowleft}, \tag{4.55}$$

$$\begin{bmatrix} * & H_{k-1,2m'} & M_{t,z0} \end{bmatrix}^{\circlearrowleft}, \tag{4.56}$$

$$\begin{bmatrix} * & H_{k-1,2m'} & \overline{M_{t,z0}} \end{bmatrix}^{\circlearrowleft}, \tag{4.57}$$

$$\begin{bmatrix} * & H_{k-1,2m'} & M_{t,z1} \end{bmatrix}^{\circlearrowleft}, \tag{4.58}$$

$$\begin{bmatrix} * & H_{k-1,2m'} & \overline{M_{t,z1}} \end{bmatrix}^{\circlearrowleft}. \tag{4.59}$$

In the terminology of Definition 4.13, each of the random variables (4.45)–(4.48), (4.50)–(4.59) is derivable with no communication from

$$\begin{bmatrix} M_{t,w} & H'_{k-1,2m'} \end{bmatrix}^{\circlearrowleft},$$

$$\begin{bmatrix} M_{t,w} & H_{k-1,m'} & \overline{M_{t,w'} \quad H_{k-1,m'}} \end{bmatrix}^{\circlearrowleft},$$

$$\begin{bmatrix} M_{t,w} & H_{k-1,m'} & M_{t,w'} & H_{k-1,m'} \end{bmatrix}^{\circlearrowleft},$$

where $t = 1, 2, \ldots, n$ and w, w' range over all strings in $Y_t \times \{0, 1\}$ at Hamming distance 1. This follows easily from (4.6)–(4.10). As a result, the input to χ''' in (4.44) is derivable with no communication from $\sigma\,\mathrm{enc}(\ldots, [M_{t,w} \quad H'_{k-1,2m'}], [M_{t,w} \quad H_{k-1,m'}], \ldots)$, where $t = 1, 2, \ldots, n$, $w \in Y_t \times \{0,1\}$, and σ is chosen uniformly at random. Then by Proposition 4.14,

$$\gamma \leqslant \max_{\chi} \left| \mathop{\mathbf{E}}_{\ldots, M_{t,w}, \ldots} \mathop{\mathbf{E}}_{\sigma} \prod_{t=1}^{n} \prod_{w \in Y_t \times \{0,1\}} \mathrm{DISJ}(M_{t,w}) \right.$$
$$\left. \times\, \chi(\sigma\, \mathrm{enc}(\ldots, [M_{t,w} \quad H'_{k-1,2m'}], [M_{t,w} \quad H_{k-1,m'}], \ldots)) \right|,$$

where the maximum is over $(k-1)$-dimensional cylinder intersections χ. The right-hand side is by definition $\Gamma(k-1, m', d+1, Y_1 \times \{0,1\}, \ldots, Y_n \times \{0,1\})$. Recalling the lower bound (4.37) on the size of $Y_1, \ldots, Y_n$, we arrive at the desired inequality (4.42). ∎

This completes the proof of Lemma 4.15. To solve the newly obtained recurrence for Γ, we prove a technical result.

LEMMA 4.17. *Fix reals $p_1, p_2, \ldots > 0$ and $q_1, q_2, \ldots > 0$. Let $A\colon \mathbb{Z}^+ \times \mathbb{N}^{n+1} \to [0,1]$ be any function that satisfies*

$$A(1, d, \ell_1, \ell_2, \ldots, \ell_n) = \begin{cases} 0 & \textit{if } \ell_1 + \ell_2 + \cdots + \ell_n > 0, \\ 1 & \textit{otherwise,} \end{cases}$$

and for $k \geqslant 2$,

$$A(k, d, \ell_1, \ell_2, \ldots, \ell_n)^2 \leqslant \sum_{i_1, j_1 = 0}^{\ell_1} \cdots \sum_{i_n, j_n = 0}^{\ell_n} \left\{ \prod_{t=1}^{n} \binom{\ell_t}{i_t} \binom{\ell_t - i_t}{j_t} p_k^{i_t} q_k^{j_t} \right\}$$
$$\times \max_{\substack{\ell'_1 \geqslant 2 \max\{0, \ell_1 - i_1 - (d+1) j_1\} \\ \vdots \\ \ell'_n \geqslant 2 \max\{0, \ell_n - i_n - (d+1) j_n\}}} A(k-1, d+1, \ell'_1, \ldots, \ell'_n).$$

Then

$$A(k, d, \ell_1, \ell_2, \ldots, \ell_n) \leqslant \left(\sum_{i=1}^{k} p_i + 8 \sum_{i=1}^{k} q_i^{1/(d+k-i+1)} \right)^{\frac{\ell_1 + \ell_2 + \cdots + \ell_n}{2}}. \tag{4.60}$$

Proof. The proof is by induction on k. In the base case $k = 1$, the bound (4.60) follows immediately from the definition of $A(1, d, \ell_1, \ell_2, \ldots, \ell_n)$. For the inductive step, fix $k \geqslant 2$ and define

$$a = \sum_{i=1}^{k-1} p_i + 8 \sum_{i=1}^{k-1} q_i^{1/(d+k-i+1)}.$$

We may assume that $a \leqslant 1$ since (4.60) is trivial otherwise. Then from the inductive hypothesis,

$$\begin{aligned}
&A(k, d, \ell_1, \ell_2, \dots, \ell_n)^2 \\
&\quad \leqslant \sum_{i_1, j_1 = 0}^{\ell_1} \cdots \sum_{i_n, j_n = 0}^{\ell_n} \left\{ \prod_{t=1}^{n} \binom{\ell_t}{i_t} \binom{\ell_t - i_t}{j_t} p_k^{i_t} q_k^{j_t} \right\} a^{\sum_{t=1}^{n} \max\{0, \ell_t - i_t - (d+1) j_t\}} \\
&\quad = \prod_{t=1}^{n} \left\{ \sum_{i,j=0}^{\ell_t} \binom{\ell_t}{i} \binom{\ell_t - i}{j} p_k^i q_k^j a^{\max\{0, \ell_t - i - (d+1) j\}} \right\}. \qquad (4.61)
\end{aligned}$$

Claim. *For any integers $\ell \geqslant 0$ and $D \geqslant 1$ and a real number $0 < q \leqslant 1$,*

$$\sum_{j=0}^{\ell} \binom{\ell}{j} q^j a^{\max\{0, \ell - Dj\}} \leqslant (a + 8q^{1/D})^{\ell}.$$

Proof:

$$\begin{aligned}
&\sum_{j=0}^{\ell} \binom{\ell}{j} q^j a^{\max\{0, \ell - Dj\}} \\
&\quad = \sum_{j \geqslant \lfloor \ell/D \rfloor + 1} \binom{\ell}{j} q^j + a^{\ell - D\lfloor \ell/D \rfloor} \sum_{j=0}^{\lfloor \ell/D \rfloor} \binom{\ell}{j} q^j (a^D)^{\lfloor \ell/D \rfloor - j} \\
&\quad \leqslant 2^{\ell} q^{\ell/D} + a^{\ell - D\lfloor \ell/D \rfloor} \sum_{j=0}^{\lfloor \ell/D \rfloor} \binom{\lfloor \ell/D \rfloor}{j} (2\mathrm{e}Dq)^j (a^D)^{\lfloor \ell/D \rfloor - j} \\
&\quad = 2^{\ell} q^{\ell/D} + a^{\ell - D\lfloor \ell/D \rfloor} (a^D + 2\mathrm{e}Dq)^{\lfloor \ell/D \rfloor} \\
&\quad \leqslant 2^{\ell} q^{\ell/D} + a^{\ell - D\lfloor \ell/D \rfloor} (a + (2\mathrm{e}Dq)^{1/D})^{D\lfloor \ell/D \rfloor} \\
&\quad \leqslant 2^{\ell} q^{\ell/D} + (a + (2\mathrm{e}Dq)^{1/D})^{\ell} \\
&\quad \leqslant (2q^{1/D} + a + (2\mathrm{e}Dq)^{1/D})^{\ell} \\
&\quad \leqslant (a + 8q^{1/D})^{\ell}.
\end{aligned}$$

∎

We may assume that $q_k \leqslant 1$ since (4.60) is trivial otherwise. Invoking the above claim with $\ell = \ell_t - i,\ q = q_k,\ D = d + 1$, we have from (4.61) that

$$\begin{aligned}
A(k, d, \ell_1, \ell_2, \dots, \ell_n)^2 &\leqslant \prod_{t=1}^{n} \left\{ \sum_{i=0}^{\ell_t} \binom{\ell_t}{i} p_k^i \left(a + 8q_k^{1/(d+1)} \right)^{\ell_t - i} \right\} \\
&= \prod_{t=1}^{n} \left(a + p_k + 8q_k^{1/(d+1)} \right)^{\ell_t},
\end{aligned}$$

completing the inductive step. ∎

Using the previous two lemmas, we will now obtain a closed-form upper bound on Γ.

THEOREM 4.18. *There exists an absolute constant $C > 1$ such that*

$$\Gamma(k, m, d, \ell_1, \ldots, \ell_n) \leqslant \left(C\sqrt{\frac{k^2 2^k}{m}} + C\exp\left\{-\frac{m}{C2^k(d+k)}\right\}\right)^{(\ell_1+\cdots+\ell_n)/2}.$$

Proof. It follows from Proposition 4.14 that Γ is monotonically decreasing in the second argument, a fact that we will use several times without further mention. Let m be an arbitrary positive integer. Set $\epsilon = 3/4$ and define

$$m_k = \left\lceil \frac{2^k m}{(1-\epsilon)(1-\epsilon^2)\cdots(1-\epsilon^k)} \right\rceil, \qquad k = 1, 2, 3, \ldots,$$

$$p_k = \sqrt{\frac{2^k}{cm_k}} + \frac{2^k}{2^{c\epsilon^{2k} m_k}}, \qquad k = 1, 2, 3, \ldots,$$

$$q_k = \frac{2^k}{2^{c\epsilon^{2k} m_k}}, \qquad k = 1, 2, 3, \ldots,$$

where $c > 0$ is the absolute constant from Theorem 4.12. Consider the real function $A : \mathbb{Z}^+ \times \mathbb{N}^{n+1} \to [0, 1]$ given by

$$A(k, d, \ell_1, \ldots, \ell_n) = \begin{cases} \Gamma(k, m_k, d, \ell_1, \ldots, \ell_n) & \text{if } \ell_1, \ldots, \ell_n \in \{0, 1, \ldots, 2^d\}, \\ 0 & \text{otherwise.} \end{cases}$$

Taking $\alpha = \epsilon^k$ in Lemma 4.15 shows that $A(k, d, \ell_1, \ldots, \ell_n)$ obeys the recurrence in Lemma 4.17. In particular, on the domain of Γ one has

$$\begin{aligned} \Gamma(k, m_k, d, \ell_1, \ldots, \ell_n) &= A(k, d, \ell_1, \ldots, \ell_n) \\ &\leqslant \left(\sum_{i=1}^{k} p_i + 8\sum_{i=1}^{k} q_i^{1/(d+k-i+1)}\right)^{(\ell_1+\cdots+\ell_n)/2} \end{aligned} \tag{4.62}$$

by Lemma 4.17.

One easily verifies that $p_i \leqslant (cm)^{-1/2} + 2^{-cm(9/8)^i + i}$ and $q_i \leqslant 2^{-cm(9/8)^i+i}$. Substituting these estimates in (4.62) gives

$$\Gamma(k, m_k, d, \ell_1, \ldots, \ell_n) \leqslant \left(\frac{k}{\sqrt{cm}} + c'\exp\left\{-\frac{c''m}{d+k}\right\}\right)^{(\ell_1+\cdots+\ell_n)/2}$$

for some absolute constants $c', c'' > 0$. Since $m_k = \Theta(2^k m)$, the proof is complete. ∎

COROLLARY 4.19. *For every n and every k-dimensional cylinder intersection χ,*

$$\left| \mathop{\mathbf{E}}_{X_1,\ldots,X_n \sim \pi_{k,m}} \left[\chi(X_1, \ldots, X_n) \prod_{i=1}^{n} \mathrm{DISJ}(X_i) \right] \right| \leqslant \left(\frac{ck^2 2^k}{m}\right)^{n/4}, \tag{4.63}$$

where $c > 0$ is an absolute constant.

Proof. By Proposition 4.14, the left-hand side of (4.63) cannot decrease if we replace $\pi_{k,m}$ with $\pi_{k,m-1}$. As a result, we may assume that m is even (if not, replace $\pi_{k,m}$ with $\pi_{k,m-1}$ in what follows). As we have already pointed out in (4.30), in this case the left-hand side of (4.63) does not exceed

$$\Gamma\left(k, \frac{m}{2}, 0, \underbrace{1, 1, \ldots, 1}_{n}\right).$$

The claimed bound is now immediate from Theorem 4.18. □

We have reached the main result of this section, an upper bound on the repeated discrepancy of set disjointness.

THEOREM 4.20. *For some absolute constant $c > 0$ and all positive integers k, m,*

$$\operatorname{rdisc}(\mathrm{UDISJ}_{k,m}) \leqslant \left(\frac{ck2^k}{\sqrt{m}}\right)^{1/2}.$$

Proof. We will prove the equivalent bound

$$\operatorname{rdisc}(\mathrm{UDISJ}_{k,M}) \leqslant \left(\frac{ck^2 2^k}{m}\right)^{1/4}, \tag{4.64}$$

where $c > 0$ is an absolute constant and $M = m(2^k - 1) + 2^{k-1}$. We will work with the probability distribution $\pi_{k,m}$, which is balanced on the domain of $\mathrm{UDISJ}_{k,M}$. By the definition of repeated discrepancy,

$$\operatorname{rdisc}_{\pi_{k,m}}(\mathrm{UDISJ}_{k,M}) = \sup_{n,r\in\mathbb{Z}^+} \max_{\chi} \left| \mathop{\mathbf{E}}_{\ldots,X_{i,j},\ldots} \chi(\ldots, X_{i,j}, \ldots) \prod_{i=1}^{n} \mathrm{DISJ}(X_{i,1}) \right|^{1/n}, \tag{4.65}$$

where $X_{i,j}$ $(i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, r)$ are chosen independently according to $\pi_{k,m}$ conditioned on $\mathrm{DISJ}(X_{i,1}) = \mathrm{DISJ}(X_{i,2}) = \cdots = \mathrm{DISJ}(X_{i,r})$ for all i. Recall that $\pi_{k,m}$ is a convex combination of $[T_k \quad H'_{k,m}]^{\circlearrowright}$ and $[F_k \quad H'_{k,m}]^{\circlearrowright}$. In particular,

$$X_{i,2}, X_{i,3}, \ldots, X_{i,r} \sim X_{i,1}^{\circlearrowright}$$

for each i. This means that the input to χ in (4.65) is derivable with no communication from $(X_{1,1}, X_{2,1}, \ldots, X_{n,1})$. As a result, Proposition 4.14 implies that

$$\operatorname{rdisc}_{\pi_{k,m}}(\mathrm{UDISJ}_{k,M}) \leqslant \sup_{n\in\mathbb{Z}^+} \max_{\chi} \left| \mathop{\mathbf{E}}_{X_{1,1},X_{2,1},\ldots,X_{n,1}\sim\pi_{k,m}} \chi(X_{1,1}, X_{2,1}, \ldots, X_{n,1}) \prod_{i=1}^{n} \mathrm{DISJ}(X_{i,1}) \right|^{1/n}.$$

The claimed upper bound (4.64) is now immediate by Corollary 4.19. □

5. RANDOMIZED COMMUNICATION

In the remainder of the paper, we will derive lower bounds for multiparty communication using the reduction to polynomials given by Theorems 4.2 and 4.20. The proofs of these applications are similar to those in [48], the main difference being the use of the newly obtained passage from protocols to polynomials in place of the less efficient reduction in [48]. We start with randomized communication, which covers protocols with small constant error as well as those with vanishing advantage over random guessing.

5.1. A master theorem. We will derive all our results on randomized communication from a single "master" theorem, which we are about to prove. Following [48], we present two proofs for it, one based on the primal view of the problem and the other, on the dual view. The idea of the primal proof is to convert a communication protocol for $f \circ \mathrm{UDISJ}_{k,m}$ into a low-degree polynomial approximating f in the infinity norm. The dual proof proceeds in the opposite direction and manipulates explicit witness objects, in the sense of Fact 2.3 and Theorem 2.10. The primal proof is probably more intuitive, 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 $\{0,1\}^n$. For every (possibly partial) k-party communication problem G and all $\epsilon, \delta \geqslant 0$,*

$$R_\epsilon(f \circ G) \geqslant \deg_\delta(f) \log\left(\frac{1}{c\,\mathrm{rdisc}(G)}\right) - \log\frac{1}{\delta - 2\epsilon}, \tag{5.1}$$

where $c > 0$ is an absolute constant. In particular,

$$R_\epsilon(f \circ \mathrm{UDISJ}_{k,m}) \geqslant \frac{\deg_\delta(f)}{2} \log\left(\frac{\sqrt{m}}{c2^k k}\right) - \log\frac{1}{\delta - 2\epsilon} \tag{5.2}$$

for some absolute constant $c > 0$.

Primal proof of Theorem 5.1. Abbreviate $F = f \circ G$. Let π be any balanced probability distribution on the domain of G and define the linear operator $L_{\pi,n}$ as in Theorem 4.2, so that $L_{\pi,n}F = f$ on the domain of f. Corollary 2.7 gives an approximation to F by a linear combination of cylinder intersections $\Pi = \sum_\chi a_\chi \chi$ with $\sum_\chi |a_\chi| \leqslant 2^{R_\epsilon(F)}/(1-\epsilon)$, 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 $\|L_{\pi,n}\Pi\|_\infty \leqslant 1/(1-\epsilon)$ and $|f - L_{\pi,n}\Pi| = |L_{\pi,n}(F - \Pi)| \leqslant \epsilon/(1-\epsilon)$ on the domain of f, whence

$$E(f, d-1) \leqslant \frac{\epsilon}{1-\epsilon} + E(L_{\pi,n}\Pi, d-1)$$

for any positive integer d. By Theorem 4.2,

$$E(L_{\pi,n}\Pi, d-1) \leqslant \sum_\chi |a_\chi| E(L_{\pi,n}\chi, d-1) \leqslant \frac{2^{R_\epsilon(F)}}{1-\epsilon}(c\,\mathrm{rdisc}_\pi(G))^d$$

for some absolute constant $c > 0$, whence

$$E(f, d-1) \leqslant \frac{\epsilon}{1-\epsilon} + \frac{2^{R_\epsilon(F)}}{1-\epsilon}(c\,\mathrm{rdisc}_\pi(G))^d.$$

For $d = \deg_\delta(f)$, the left-hand side of this inequality must exceed δ, forcing (5.1). The other lower bound (5.2) now follows immediately by Theorem 4.20. ∎

We now present an alternate proof, which is based directly on the generalized discrepancy method.

Dual proof of Theorem 5.1. Again, it suffices to prove (5.1). We closely follow the proof in [48] except at the end. Let $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$ be the input space of G. Let π be an arbitrary balanced probability distribution on the domain of G, and define $d = \deg_\delta(f)$. By Fact 2.3, there exists a function $\psi\colon \{0,1\}^n \to \mathbb{R}$ with

$$\sum_{x \in \operatorname{dom} f} f(x)\psi(x) - \sum_{x \notin \operatorname{dom} f} |\psi(x)| > \delta, \tag{5.3}$$

$$\|\psi\|_1 = 1, \tag{5.4}$$

$$\hat{\psi}(S) = 0, \qquad |S| < d. \tag{5.5}$$

Define $\Psi\colon \mathscr{X}^n \to \mathbb{R}$ by

$$\Psi(X_1, \dots, X_n) = 2^n \psi(G^*(X_1), \dots, G^*(X_n)) \prod_{i=1}^n \pi(X_i)$$

and let $F = f \circ G$. Since π is balanced on the domain of G,

$$\|\Psi\|_1 = 2^n \mathop{\mathbf{E}}_{x \in \{0,1\}^n} [|\psi(x)|] = 1 \tag{5.6}$$

and analogously

$$\begin{aligned} \sum_{\operatorname{dom} F} F(X_1, \dots, X_n)\Psi(X_1, \dots, X_n) &- \sum_{\overline{\operatorname{dom} F}} |\Psi(X_1, \dots, X_n)| \\ &= \sum_{x \in \operatorname{dom} f} f(x)\psi(x) - \sum_{x \notin \operatorname{dom} f} |\psi(x)| \\ &> \delta, \end{aligned} \tag{5.7}$$

where the final step in the two derivations uses (5.3) and (5.4). It remains to bound the inner product of Ψ with a k-dimensional cylinder intersection χ. We have

$$\begin{aligned} \langle \Psi, \chi \rangle &= 2^n \mathop{\mathbf{E}}_{X_1, \dots, X_n \sim \pi} \left[\psi(G^*(X_1), \dots, G^*(X_n)) \chi(X_1, \dots, X_n) \right] \\ &= \sum_{x \in \{0,1\}^n} \psi(x) \mathop{\mathbf{E}}_{X_1 \sim \pi_{x_1}} \cdots \mathop{\mathbf{E}}_{X_n \sim \pi_{x_n}} \chi(X_1, \dots, X_n) \\ &= \langle \psi, L_{\pi,n} \chi \rangle, \end{aligned}$$

where π_0 and π_1 are the probability distributions induced by π on $G^{-1}(+1)$ and $G^{-1}(-1)$, respectively, and $L_{\pi,n}$ is as defined in Theorem 4.2. Continuing,

$$\begin{aligned} |\langle \Psi, \chi \rangle| &\leqslant \|\psi\|_1 E(L_{\pi,n}\chi, d-1) && \text{by (5.5)} \\ &\leqslant (c \operatorname{rdisc}_\pi(G))^d && \text{by (5.4) and Theorem 4.2,} \end{aligned} \tag{5.8}$$

where $c > 0$ is an absolute constant. Now (5.1) is immediate by (5.6)–(5.8) and the generalized discrepancy method (Theorem 2.10). ∎

5.2. Bounded-error communication. Specializing the master theorem to bounded-error communication gives the following lower bound for composed communication problems in terms the $1/3$-approximate degree.

THEOREM 5.2. *There exists an absolute constant $c > 0$ such that for every* (*possibly partial*) *Boolean function f on* $\{0,1\}^n$,

$$R_{1/3}(f \circ \mathrm{UDISJ}_{k, c4^k k^2}) \geqslant \deg_{1/3}(f).$$

Proof. Take $\epsilon = 1/7$, $\delta = 1/3$, and $m = c'4^k k^2$ in the lower bound (5.2) of Theorem 5.1, where $c' > 0$ is a sufficiently large integer constant. ∎

As a consequence we obtain the main result of this paper, stated in the introduction as Theorem 1.1.

COROLLARY 5.3.

$$R_{1/3}(\mathrm{DISJ}_{k,n}) \geqslant R_{1/3}(\mathrm{UDISJ}_{k,n}) = \Omega\left(\frac{\sqrt{n}}{2^k k}\right).$$

Proof. Recall that $\mathrm{UDISJ}_{k,nm} = \widetilde{\mathrm{AND}}_n \circ \mathrm{UDISJ}_{k,m}$ for all integers n, m. Theorem 2.5 shows that $\deg_{1/3}(\widetilde{\mathrm{AND}}_n) = \Omega(\sqrt{n})$. Thus, taking $f = \widetilde{\mathrm{AND}}_n$ in Theorem 5.2 gives $R_{1/3}(\mathrm{UDISJ}_{k,c4^k k^2 n}) = \Omega(\sqrt{n})$ for some absolute constant $c > 0$, which is equivalent to the claimed bound. ∎

REMARK 5.4. As shown by the dual proof of Theorem 5.1, we obtain the $\Omega(\sqrt{n}/2^k k)$ lower bound for set disjointness using the generalized discrepancy method. By the results of [37, 14], the generalized discrepancy method applies to *quantum* multiparty protocols as well. In particular, Corollary 5.3 in this paper gives a lower bound of $\Omega(\sqrt{n}/2^k k) - O(k^4)$ on the bounded-error k-party quantum communication complexity of set disjointness. This lower bound nearly matches the well-known upper bound of $\lceil \sqrt{n/2^k} \rceil \log^{O(1)} n$ due to Buhrman, Cleve, and Wigderson [16]. For the reader's convenience, we include a sketch of the protocol. Let G be any k-party communication problem and $f\colon \{0,1\}^n \to \{-1,+1\}$ a given function. An elegant simulation in [16] shows that $f \circ G$ has bounded-error quantum communication complexity $O(Q_{1/3}(f)D(G)k^2 \log n)$, where $Q_{1/3}(f)$ and $D(G)$ are the bounded-error quantum query complexity of f and the deterministic classical communication complexity of G, respectively. Letting $\mathrm{DISJ}_{k,n} = \mathrm{AND}_{n/2^k} \circ \mathrm{DISJ}_{k,2^k}$, we have $Q_{1/3}(\mathrm{AND}_{n/2^k}) = O(\sqrt{n/2^k})$ by Grover's search algorithm [28] and $D(\mathrm{DISJ}_{k,2^k}) = O(k^2)$ by Grolmusz's result [27]. Therefore, set disjointness has bounded-error quantum communication complexity at most $\lceil \sqrt{n/2^k} \rceil \log^{O(1)} n$.

Theorem 5.2 gives a lower bound on bounded-error communication complexity for compositions $f \circ G$, where G is a gadget whose size grows exponentially with the number of parties. Following [48], we will derive an alternate lower bound, in which the gadget G is essentially as simple as possible and in particular depends on only $2k$ variables. The resulting lower bound will be in terms of the approximate degree as well as two combinatorial complexity measures, defined next. The *block sensitivity* of a Boolean function $f\colon \{0,1\}^n \to \{-1,+1\}$, 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 $f(x) \neq f(x \oplus \mathbf{1}_{S_1}) = f(x \oplus \mathbf{1}_{S_2}) = f(x \oplus \mathbf{1}_{S_3}) = \cdots$ for some string $x \in \{0,1\}^n$. The *decision tree complexity* of f, denoted $\mathrm{dt}(f)$, is the minimum depth of a decision tree for f. We have:

THEOREM 5.5. *For every* $f\colon \{0,1\}^n \to \{-1,+1\}$,

$$R_{1/3}(f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)) \geqslant \Omega\left(\frac{\sqrt{\mathrm{bs}(f)}}{2^k k}\right) \geqslant \Omega\left(\frac{\mathrm{dt}(f)^{1/6}}{2^k k}\right) \geqslant \Omega\left(\frac{\deg_{1/3}(f)^{1/6}}{2^k k}\right)$$

and

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

Here OR_k and AND_k refer to the k-party communication problems $x \mapsto \bigvee_{i=1}^k x_i$ and $x \mapsto \bigwedge_{i=1}^k x_i$, where the ith party sees all the bits except for x_i. Analogously, $\mathrm{OR}_k \vee \mathrm{AND}_k$ refers to the k-party communication problem $x \mapsto x_1 \vee \cdots \vee x_k \vee (x_{k+1} \wedge \cdots \wedge x_{2k})$ in which the ith party sees all the bits except for x_i and x_{k+i}. It is clear that the composed communication problems $f \circ \mathrm{OR}_k$, $f \circ \mathrm{AND}_k$, and $f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$ each have a deterministic k-party communication protocol with cost $3\,\mathrm{dt}(f)$. The above theorem shows that this upper bound is reasonably close to tight, even for randomized protocols. Note that it is impossible to go beyond Theorem 5.5 and bound $R_{1/3}(f \circ \mathrm{AND}_k)$ from below in terms of the approximate degree of f: taking $f = \mathrm{AND}_n$ shows that the gap between $R_{1/3}(f \circ \mathrm{AND}_k)$ and $\deg_{1/3}(f)$ can be as large as $\Theta(1)$ versus $\Theta(\sqrt{n})$. Theorem 5.5 is a quadratic improvement on the lower bounds in [48].

Proof of Theorem 5.5. Identical to the proofs of Theorems 5.3 and 5.4 in [48], with Corollary 5.3 used instead of the earlier lower bound for set disjointness in [48]. ∎

5.3. Small-bias communication and discrepancy. We now specialize Theorem 5.1 to the setting of small-bias communication, where the protocol is only required to produce the correct output with probability vanishingly close to $1/2$.

THEOREM 5.6. *Let f be a* (*possibly partial*) *Boolean function on* $\{0,1\}^n$. *For every* (*possibly partial*) *k-party communication problem G and all $\epsilon, \gamma \geqslant 0$,*

$$R_{\frac{1}{2}-\frac{\epsilon}{2}}(f \circ G) \geqslant \deg_{1-\gamma}(f) \log\left(\frac{1}{c\,\mathrm{rdisc}(G)}\right) - \log\frac{1}{\epsilon-\gamma}, \tag{5.9}$$

$$R_{\frac{1}{2}-\frac{\epsilon}{2}}(f \circ G) \geqslant \deg_{\pm}(f) \log\left(\frac{1}{c\,\mathrm{rdisc}(G)}\right) - \log\frac{1}{\epsilon}, \tag{5.10}$$

where $c > 0$ is an absolute constant. In particular,

$$R_{\frac{1}{2}-\frac{\epsilon}{2}}(f \circ \mathrm{UDISJ}_{k,c4^k k^2}) \geqslant \deg_{1-\gamma}(f) - \log\frac{1}{\epsilon-\gamma}, \tag{5.11}$$

$$R_{\frac{1}{2}-\frac{\epsilon}{2}}(f \circ \mathrm{UDISJ}_{k,c4^k k^2}) \geqslant \deg_{\pm}(f) - \log\frac{1}{\epsilon} \tag{5.12}$$

for an absolute constant $c > 0$.

Proof. One obtains (5.9) by taking $\delta = 1-\gamma$ in (5.1). Letting $\gamma \searrow 0$ in (5.9) gives (5.10). The remaining two lower bounds are now immediate in view of Theorem 4.20. ∎

The method of Theorem 5.1 allows one to directly prove upper bounds on discrepancy, a complexity measure of interest in its own right.

THEOREM 5.7. *For every (possibly partial) Boolean function f on $\{0,1\}^n$, every (possibly partial) k-party communication problem G, and every $\gamma > 0$, one has*

$$\operatorname{disc}(f \circ G) \leqslant (c \operatorname{rdisc}(G))^{\deg_{1-\gamma}(f)} + \gamma, \tag{5.13}$$

$$\operatorname{disc}(f \circ G) \leqslant (c \operatorname{rdisc}(G))^{\deg_{\pm}(f)}, \tag{5.14}$$

where $c > 0$ is an absolute constant. In particular,

$$\operatorname{disc}(f \circ \mathrm{UDISJ}_{k,m}) \leqslant \left(\frac{c2^k k}{\sqrt{m}}\right)^{\deg_{1-\gamma}(f)/2} + \gamma, \tag{5.15}$$

$$\operatorname{disc}(f \circ \mathrm{UDISJ}_{k,m}) \leqslant \left(\frac{c2^k k}{\sqrt{m}}\right)^{\deg_{\pm}(f)/2} \tag{5.16}$$

for an absolute constant $c > 0$.

Proof. The proof is virtually identical to that in [48], with the difference that we use Theorems 4.2 and 4.20 in place of the earlier passage from protocols to polynomials. For the reader's convenience, we include a complete proof.

Let $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$ be the input space of G, and let π be an arbitrary balanced probability distribution on the domain of G. Take $\delta = 1-\gamma$, $d = \deg_\delta(f)$, and define $\Psi\colon \mathscr{X}^n \to \mathbb{R}$ as in the dual proof of Theorem 5.1. Then (5.6) shows that Ψ is the pointwise product $\Psi = H \cdot P$, where H is a sign tensor and P a probability distribution. Abbreviating $F = f \circ G$, we can restate (5.7) and (5.8) as

$$\sum_{\operatorname{dom} F} F(X)H(X)P(X) - P(\overline{\operatorname{dom} F}) > 1-\gamma, \tag{5.17}$$

$$\operatorname{disc}_P(H) \leqslant (c \operatorname{rdisc}_\pi(G))^d, \tag{5.18}$$

respectively, where $c > 0$ is an absolute constant. For every cylinder intersection χ,

$$\begin{aligned}
&\left|\sum_{\operatorname{dom} F} F(X)P(X)\chi(X)\right| \\
&\quad= \left|\langle H\cdot P, \chi\rangle + \sum_{\operatorname{dom} F} (F(X)-H(X))P(X)\chi(X) - \sum_{\overline{\operatorname{dom} F}} H(X)P(X)\chi(X)\right| \\
&\quad\leqslant \operatorname{disc}_P(H) + \sum_{\operatorname{dom} F} |F(X)-H(X)|P(X) + P(\overline{\operatorname{dom} F}) \\
&\quad= \operatorname{disc}_P(H) + P(\operatorname{dom} F) - \sum_{\operatorname{dom} F} F(X)H(X)P(X) + P(\overline{\operatorname{dom} F}) \\
&\quad< \operatorname{disc}_P(H) + P(\operatorname{dom} F) - 1 + \gamma,
\end{aligned} \tag{5.19}$$

where the last step uses (5.17). Therefore,

$$\begin{aligned}
\operatorname{disc}_P(f \circ G) &= \max_\chi \left|\sum_{\operatorname{dom} F} F(X)P(X)\chi(X)\right| + P(\overline{\operatorname{dom} F}) \\
&< \operatorname{disc}_P(H) + \gamma \\
&\leqslant (c \operatorname{rdisc}_\pi(G))^d + \gamma,
\end{aligned}$$

where the second step uses (5.19) and the third uses (5.18). This completes the proof of (5.13). Letting $\gamma \searrow 0$, one arrives at (5.14). The remaining two lower bounds (5.15) and (5.16) are now immediate by Theorem 4.20. ∎

COROLLARY 5.8. *Consider the Boolean function*

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

viewed as a k-party communication problem in which the rth party $(r = 1, 2, \ldots, k)$ is missing the bits $x_{i,j,r}$ for all i, j. Then

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

Proof. Let MP_n be given by Theorem 2.4, so that $\deg_{\pm}(\mathrm{MP}_n) = n$. Let $c > 0$ be the constant from (5.16). Since $\mathrm{MP}_n \circ \mathrm{DISJ}_{k,c^2 4^{k+1} k^2}$ is a subfunction of $F_{k,\lceil 4c\rceil n}(\overline{x})$, Theorem 5.7 yields the discrepancy bound. The communication lower bound follows by Theorem 2.9. ∎

Corollary 5.8 gives a hard k-party communication problem computable by an AC^0 circuit family of depth 3. This depth is optimal because AC^0 circuits of smaller depth have multiparty discrepancy $1/n^{O(1)}$, regardless of how the bits are assigned to the parties. Quantitatively, the corollary gives an upper bound of $\exp(-\Omega(n/4^k k^2)^{1/3})$ on the discrepancy of a size-nk circuit family in AC^0, considerably improving on the previous best bound of $\exp(-\Omega(n/4^k)^{1/7})$ in [48], itself preceded by $\exp(-\Omega(n/2^{31k})^{1/29})$ in [10].

Corollary 5.8 settles Theorem 1.4 from the introduction.

6. ADDITIONAL APPLICATIONS

We conclude this paper with several additional results in communication complexity. In what follows, we give improved XOR lemmas and direct product theorems for composed communication problems as well as a quadratically stronger lower bound on the nondeterministic and Merlin-Arthur complexity of set disjointness. Lastly, we give applications of our work to circuit complexity.

6.1. XOR lemmas. In Section 5, we proved an $\Omega(\sqrt{n}/2^k k)$ communication lower bound for solving the set disjointness problem $\mathrm{DISJ}_{k,n}$ with probability of correctness $2/3$. In this section, we consider the communication problem $\mathrm{DISJ}_{k,n}{}^{\otimes\ell}$. As one would expect, we show that its randomized communication complexity is $\ell\cdot\Omega(\sqrt{n}/2^k k)$. More interestingly, we show that the same lower bound remains valid even for probability of correctness $\frac{1}{2} - 2^{-\Omega(\ell)}$, a statement known as an *XOR lemma.* We prove an analogous result for the unique set disjointness problem and more generally for composed problems $f \circ G$, where G has small repeated discrepancy. Our proofs are nearly identical to those in [48], the main difference being that we use Theorems 4.2 and 4.20 in place of the earlier and less efficient passage from protocols to polynomials.

We first recall an XOR lemma for polynomial approximation, proved in [49, Cor. 5.2].

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

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

Using the small-bias version of the master theorem (Theorem 5.6), we are able to immediately translate this result to communication.

THEOREM 6.2. *For every (possibly partial) Boolean function f on $\{0,1\}^n$ and every (possibly partial) k-party communication problem G,*

$$R_{\frac{1}{2}-(\frac{1}{2})^{\ell+1}}((f \circ G)^{\otimes \ell}) \geqslant c\ell \deg_{1/3}(f) \cdot \log \frac{c}{\operatorname{rdisc}(G)}, \tag{6.1}$$

where $c > 0$ is an absolute constant. In particular,

$$R_{\frac{1}{2}-(\frac{1}{2})^{\ell+1}}((f \circ \mathrm{UDISJ}_{k,c4^k k^2})^{\otimes \ell}) \geqslant \ell \deg_{1/3}(f) \tag{6.2}$$

for an absolute constant $c > 0$.

Proof. Theorem 6.1 provides an absolute constant $c_1 > 0$ such that $\deg_{1-2^{-\ell-1}}(f^{\otimes \ell}) \geqslant c_1 \ell \deg_{1/3}(f)$. Applying Theorem 5.6 to $f^{\otimes \ell} \circ G = (f \circ G)^{\otimes \ell}$ with parameters $\epsilon = 2^{-\ell}$ and $\gamma = 2^{-\ell-1}$, one arrives at

$$R_{1/2-1/2^{\ell+1}}((f \circ G)^{\otimes \ell}) \geqslant c_1 \ell \deg_{1/3}(f) \cdot \log \left(\frac{1}{c_2 \operatorname{rdisc}(G)} \right) - \ell - 1$$

for some absolute constant $c_2 > 0$. This conclusion is logically equivalent to (6.1). In view of Theorem 4.20, the other lower bound (6.2) is immediate from (6.1). ∎

COROLLARY 6.3.

$$R_{\frac{1}{2}-(\frac{1}{2})^{\ell+1}}({\mathrm{UDISJ}_{k,n}}^{\otimes \ell}) \geqslant \ell \cdot \Omega\left(\frac{\sqrt{n}}{2^k k} \right).$$

Proof. Theorem 2.5 shows that $\deg_{1/3}(\widetilde{\mathrm{AND}}_n) \geqslant \Omega(\sqrt{n})$. Thus, letting $f = \widetilde{\mathrm{AND}}_n$ in (6.2) gives $R_{1/2-1/2^{\ell+1}}({\mathrm{UDISJ}_{k,c4^k k^2 n}}^{\otimes \ell}) \geqslant \ell \cdot \Omega(\sqrt{n})$ for a constant $c > 0$, which is equivalent to the claimed bound. ∎

The above corollary settles Theorem 1.2(i) from the introduction. It is a quadratic improvement on the previous best XOR lemma for multiparty set disjointness [48]. As a consequence, we obtain stronger XOR lemmas for arbitrary compositions of the form $f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$, improving quadratically on the work in [48].

THEOREM 6.4. *Let $f\colon \{0,1\}^n \to \{-1,+1\}$ be given. Then the k-party communication problem $F = f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$ obeys*

$$R_{\frac{1}{2}-(\frac{1}{2})^{\ell+1}}(F^{\otimes \ell}) \geqslant \ell \cdot \Omega\left(\frac{\sqrt{\operatorname{bs}(f)}}{2^k k} \right) \geqslant \ell \cdot \Omega\left(\frac{\operatorname{dt}(f)^{1/6}}{2^k k} \right)$$
$$\geqslant \ell \cdot \Omega\left(\frac{\deg_{1/3}(f)^{1/6}}{2^k k} \right).$$

Proof. The argument is identical to that in [48, Theorem 5.3]. As argued there, any communication protocol for $f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$ also solves $\mathrm{UDISJ}_{k,\mathrm{bs}(f)}$, so that the first inequality is immediate from the newly obtained XOR lemma for unique set disjointness. The other two inequalities follow from general relationships among $\mathrm{bs}(f)$, $\mathrm{dt}(f)$, and $\deg_{1/3}(f)$; see [48, Theorem 5.3]. ∎

6.2. Direct product theorems. Given a (possibly partial) k-party communication problem F on $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_k$, consider the task of simultaneously solving ℓ instances of F. More formally, the communication protocol now receives ℓ inputs $X_1, \dots, X_\ell \in \mathscr{X}$ and outputs a string $\{-1,+1\}^\ell$, representing a guess at $(F(X_1), \dots, F(X_\ell))$. An ϵ-error protocol is one whose output differs from the correct answer with probability no greater than ϵ on any given input $X_1, \dots, X_\ell \in \operatorname{dom} F$. We let $R_\epsilon(F, F, \dots, F)$ denote the least cost of such a protocol for solving ℓ instances of F, where the number of instances will always be specified with an underbrace.

It is also meaningful to consider communication protocols that solve almost all ℓ instances. In other words, the protocol receives instances $X_1, \dots, X_\ell$ and 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

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

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." The parameter m, for "mistake," should be thought of as a small constant fraction of ℓ. This regime 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 [49] analytically formalizes the simultaneous solution of ℓ instances.

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

$$\sum_{z \in \{-1,+1\}^\ell} |\phi_z(x_1, \dots, x_\ell)| \leqslant 1, \qquad x_1, \dots, x_\ell \in \mathscr{X},$$

$$\sum_{\substack{z \in \{-1,+1\}^\ell \\ |\{i : z_i = -1\}| \leqslant m}} \phi_{(z_1 f(x_1), \dots, z_\ell f(x_\ell))}(x_1, \dots, x_\ell) \geqslant \sigma, \qquad x_1, \dots, x_\ell \in \operatorname{dom} f.$$

The following result [49, Corollary 5.7] on polynomial approximation can be thought of as a threshold direct product theorem in that model of computation.

THEOREM 6.6 (Sherstov). *There exists an absolute constant $\alpha > 0$ such that for every (possibly partial) Boolean function f on $\{0,1\}^n$ and every $(2^{-\alpha\ell}, \alpha\ell, \ell)$-approximant $\{\phi_z\}$ for f,*

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

We will now translate this result to multiparty communication complexity. Our proof is closely analogous to that in [48, Theorem 6.7], the main difference being our use of

Theorems 4.2 and 4.20 in place of the earlier and less efficient passage from protocols to polynomials.

THEOREM 6.7. *There is an absolute constant* $0 < c < 1$ *such that for every* (*possibly partial*) *Boolean function* f *on* $\{0,1\}^n$ *and every* (*possibly partial*) k*-party communication problem* G,

$$R_{1-2^{-c\ell},c\ell}\,\underbrace{(f\circ G,\ldots,f\circ G)}_{\ell} \geqslant c\ell\deg_{1/3}(f)\cdot\log\frac{c}{\operatorname{rdisc}(G)}. \tag{6.3}$$

In particular,

$$R_{1-2^{-c\ell},c\ell}\left(\underbrace{f\circ \mathrm{UDISJ}_{k,\lceil\frac{4^k k^2}{c}\rceil}\ldots,f\circ \mathrm{UDISJ}_{k,\lceil\frac{4^k k^2}{c}\rceil}}_{\ell}\right)\geqslant \ell\deg_{1/3}(f)$$

for some absolute constant $0 < c < 1$.

Proof. Let $\mathscr{X} = \mathscr{X}_1\times\mathscr{X}_2\times\cdots\times\mathscr{X}_k$ be the input space of G. Let $\alpha > 0$ be the absolute constant from Theorem 6.6, and let $c \in (0,\alpha)$ be a sufficiently small absolute constant to be named later. Consider any randomized protocol Π which solves with probability $2^{-c\ell}$ at least $(1-c)\ell$ from among ℓ instances of $f\circ G$, and let r denote the cost of this protocol. For $z \in \{-1,+1\}^\ell$, let Π_z denote the protocol with Boolean output which on input from $(\mathscr{X}^n)^\ell$ runs Π and outputs -1 if and only if Π outputs z. Let $\phi_z\colon(\mathscr{X}^n)^\ell\to[0,1]$ be the acceptance probability function for Π_z. Then $\phi_z = \sum a_\chi\chi$ by Corollary 2.8, where the sum is over k-dimensional cylinder intersections and $\sum|a_\chi|\leqslant 2^r$.

Now let π be any balanced probability distribution on the domain of G and define the linear operator $L_{\pi,\ell n}\colon\mathbb{R}^{(\mathscr{X}^n)^\ell}\to\mathbb{R}^{(\{0,1\}^n)^\ell}$ as in Theorem 4.2. By Theorem 4.2 and linearity,

$$E(L_{\pi,\ell n}\,\phi_z, D-1)\leqslant 2^r\left(\frac{\operatorname{rdisc}_\pi(G)}{c'}\right)^D$$

for every z and every positive integer D, where $c' > 0$ is an absolute constant. Abbreviate $d = \deg_{1/3}(f)$ in what follows. Letting $D = \lceil\alpha\ell d\rceil$, we arrive at

$$E(L_{\pi,\ell n}\,\phi_z,\lceil\alpha\ell d\rceil-1)\leqslant 2^r\left(\frac{\operatorname{rdisc}_\pi(G)}{c'}\right)^{\lceil\alpha\ell d\rceil} \tag{6.4}$$

for every z. On the other hand, we claim that

$$E(L_{\pi,\ell n}\,\phi_z,\lceil\alpha\ell d\rceil-1)\geqslant\frac{2^{-c\ell}-2^{-\alpha\ell}}{2^\ell(1+2^{-\alpha\ell})} \tag{6.5}$$

for at least one value of z. To see this, observe that $\{\phi_z\}$ is a $(2^{-c\ell},\alpha\ell,\ell)$-approximant for $f\circ G$, and analogously $\{L_{\pi,\ell n}\,\phi_z\}$ is a $(2^{-c\ell},\alpha\ell,\ell)$-approximant for f. As a result, if every function $L_{\pi,\ell n}\,\phi_z$ can be approximated within ϵ by a polynomial of degree less than $\alpha\ell d$, one obtains a $((2^{-c\ell}-2^\ell\epsilon)/(1+2^\ell\epsilon),\alpha\ell,\ell)$-approximant for f with degree less than $\alpha\ell d$. The inequality (6.5) now follows from Theorem 6.6, which states that f does not admit a $(2^{-\alpha\ell},\alpha\ell,\ell)$-approximant of degree less than $\alpha\ell d$.

Comparing (6.4) and (6.5) yields for claimed lower bound (6.3) on r, provided that $c = c(c',\alpha) > 0$ is small enough. The other lower bound in the theorem statement follows from (6.3) by Theorem 4.20. ∎

Theorem 6.7 readily generalizes to compositions of the form $f \circ (\mathrm{OR}_k \vee \mathrm{AND}_k)$, as illustrated above for XOR lemmas.

COROLLARY 6.8. *For some absolute constant* $0 < c < 1$ *and every* ℓ,

$$R_{1-2^{-c\ell},c\ell}(\underbrace{\mathrm{UDISJ}_{k,n},\ldots,\mathrm{UDISJ}_{k,n}}_{\ell}) \geqslant \ell\cdot\Omega\left(\frac{\sqrt{n}}{2^k k}\right).$$

Proof. Theorem 2.5 shows that $\deg_{1/3}(\widetilde{\mathrm{AND}}_n) \geqslant \Omega(\sqrt{n})$. As a result, Theorem 6.7 for $f = \widetilde{\mathrm{AND}}_n$ gives

$$R_{1-2^{-c\ell},c\ell}\left(\underbrace{\mathrm{UDISJ}_{k,n\left\lceil\frac{4^k k^2}{c}\right\rceil},\ldots,\mathrm{UDISJ}_{k,n\left\lceil\frac{4^k k^2}{c}\right\rceil}}_{\ell}\right) = \ell\cdot\Omega(\sqrt{n}),$$

which is equivalent to the claimed bound. ∎

This settles Theorem 1.2(ii) from the introduction.

6.3. Nondeterministic and Merlin-Arthur communication. We now turn to the nondeterministic and Merlin-Arthur communication complexity of set disjointness. The best lower bounds [48] prior to this paper were $\Omega(n/4^k)^{1/4}$ for nondeterministic protocols and $\Omega(n/4^k)^{1/8}$ for Merlin-Arthur protocols, both of which are tight up to a polynomial. In what follows, we prove quadratically stronger lower bounds in both models. The proof in this paper is nearly identical to those in [23, 48], the only difference being the passage from communication protocols to polynomials. We use Theorems 4.2 and 4.20 for this purpose, in place of the less efficient passage in previous works.

THEOREM 6.9. *There exists an absolute constant* $c > 0$ *such that for every* (*possibly partial*) k*-party communication problem* G,

$$N(\mathrm{AND}_n \circ G) \geqslant \Omega\left(\sqrt{n}\log\frac{1}{c\,\mathrm{rdisc}(G)}\right), \tag{6.6}$$

$$MA_{1/3}(\mathrm{AND}_n \circ G) \geqslant \Omega\left(\sqrt{n}\log\frac{1}{c\,\mathrm{rdisc}(G)}\right)^{1/2}. \tag{6.7}$$

In particular,

$$N(\mathrm{DISJ}_{k,n}) \geqslant \Omega\left(\frac{\sqrt{n}}{2^k k}\right),$$

$$MA_{1/3}(\mathrm{DISJ}_{k,n}) \geqslant \Omega\left(\frac{\sqrt{n}}{2^k k}\right)^{1/2}.$$

Proof. Define $f = \mathrm{AND}_n$, $F = f \circ G$, and $d = \deg_{1/3}(\mathrm{AND}_n)$. As shown in [23] and [48, Theorem 7.2], there exists a function $\psi\colon\{0,1\}^n \to \mathbb{R}$ that obeys (5.4), (5.5), and

$$\psi(1,1,\ldots,1) < -\frac{1}{6}. \tag{6.8}$$

Now fix an arbitrary balanced probability distribution π on the domain of G and define

$$\Psi(X_1, \dots, X_n) = 2^n \psi(G^*(X_1), \dots, G^*(X_n)) \prod_{i=1}^{n} \pi(X_i),$$

as in the dual proof of Theorem 5.1. Then (5.6) shows that Ψ is the pointwise product $\Psi = H \cdot P$ for some sign tensor H and probability distribution P. In particular, (5.8) asserts that

$$\operatorname{disc}_P(H) \leqslant (c \operatorname{rdisc}_\pi(G))^d \tag{6.9}$$

for an absolute constant $c > 0$. By (6.8), we have $\psi(x) < 0$ whenever $f(x) = -1$, so that

$$P(F^{-1}(-1) \cap H^{-1}(+1)) = 0. \tag{6.10}$$

Also,

$$P(F^{-1}(-1) \cap H^{-1}(-1)) = P(F^{-1}(-1)) = |\psi(1, 1, \dots, 1)| > \frac{1}{6}, \tag{6.11}$$

where the first step uses (6.10), the second step uses the fact that π is balanced on the domain of G, and the final inequality uses (6.8). By Theorem 2.5,

$$d = \Omega(\sqrt{n}). \tag{6.12}$$

Now (6.6) and (6.7) are immediate from (6.9)–(6.12) and Theorem 2.11.

Taking $G = \mathrm{DISJ}_{k,c'4^k k^2}$ in (6.6) for a sufficiently large integer constant $c' \geqslant 1$ gives

$$N(\mathrm{DISJ}_{k,c'4^k k^2 n}) \geqslant \Omega\left(\sqrt{n} \log \frac{1}{c \operatorname{rdisc}(\mathrm{DISJ}_{k,c'4^k k^2})}\right) \geqslant \Omega(\sqrt{n}),$$

where the second inequality uses Theorem 4.20. Analogously $MA_{1/3}(\mathrm{DISJ}_{k,c'4^k k^2 n}) \geqslant \Omega(n^{1/4})$. These lower bounds on the nondeterministic and Merlin-Arthur complexity of set disjointness are equivalent to those in the theorem statement. ∎

This settles Theorem 1.3 from the introduction.

6.4. Circuit complexity. 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 [2] 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. Allender's result prompted a study of the computational limitations of depth-2 majority circuits and more generally of depth-3 majority circuits with restricted bottom fan-in. Most of the results in this line of work exploit the following reduction to multiparty communication complexity, where the shorthand $\mathrm{MAJ} \circ \mathrm{SYMM} \circ \mathrm{ANY}$ refers to the family of circuits with a majority gate at the top, arbitrary symmetric gates at the middle level, and arbitrary gates at the bottom.

PROPOSITION 6.10 (Håstad and Goldmann). *Let f be a Boolean function computable by a $\mathrm{MAJ} \circ \mathrm{SYMM} \circ \mathrm{ANY}$ 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 number-on-the-forehead 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.

Depth	Circuit lower bound	Reference
3	$\exp\{\Omega(n^{1/3})\}, \quad k=2$	Buhrman, Vereshchagin, and de Wolf [17] Sherstov [46, 47]
3	$\exp\left\{\Omega\left(\frac{n}{4^k}\right)^{1/(6k2^k)}\right\}$	Chattopadhyay [19]
6	$\exp\left\{\frac{1}{k}\cdot\Omega\left(\frac{n}{2^{31k}}\right)^{1/29}\right\}$	Beame and Huynh-Ngoc [10]
3	$\exp\left\{\frac{1}{k}\cdot\Omega\left(\frac{n}{4^k}\right)^{1/7}\right\}$	Sherstov [48]
3	$\exp\left\{\frac{1}{k}\cdot\Omega\left(\frac{n}{4^k k^2}\right)^{1/3}\right\}$	This paper

TABLE 3. Lower bounds for computing functions in AC^0 by circuits of type MAJ $\circ$ SYMM $\circ$ ANY with bottom fan-in $k-1$. All functions are on nk bits.

Using Håstad and Goldmann's observation, a series of papers [17, 46, 47, 19, 10, 48] have studied the circuit complexity of AC^0 functions, culminating in a proof [48] that MAJ $\circ$ SYMM $\circ$ ANY circuits with bottom fan-in $(\frac{1}{2}-\epsilon)\log n$ require exponential size to simulate AC^0 functions, for any $\epsilon>0$. This circuit lower bound comes close to matching Allender's simulation of AC^0 by quasipolynomial-size depth-3 majority circuits, where the bottom fan-in is $\log^{O(1)} n$. Table 3 gives a quantitative summary of this line of research. We are able to contribute the following sharper lower bound.

THEOREM 6.11. *There is an (explicitly given) read-once* $\{\wedge,\vee\}$*-formula* $H_{k,n}\colon\{0,1\}^{nk}\to\{-1,+1\}$ *of depth* 3 *such that any circuit of type* MAJ $\circ$ SYMM $\circ$ ANY *with bottom fan-in at most* $k-1$ *computing* $H_{k,n}$ *has size*

$$\exp\left\{\frac{1}{k}\cdot\Omega\left(\frac{n}{4^k k^2}\right)^{1/3}\right\}.$$

Proof. Define

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

We interpret $F_{k,n}$ as the k-party communication problem in Corollary 5.8. Let C be a circuit of type MAJ $\circ$ SYMM $\circ$ ANY that computes $F_{k,n}$, where the bottom fan-in of C is at most $k-1$. Let s denote the size of C. The proof will be complete once we show that $s\geqslant 2^{\Omega(n/k)}$.

Since C has size s, the fan-in of the gates at the top and middle levels is bounded by s, which in view of Proposition 6.10 gives

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

By Corollary 5.8, this leads to the desired lower bound: $s \geqslant 2^{\Omega(n/k)}$. ∎

REFERENCES

[1] A. N. Alekseichuk. On complexity of computation of partial derivatives of Boolean functions realized by Zhegalkin polynomials. *Cybernetics and Systems Analysis*, 37(5):648–653, 2001.

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

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

[4] 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* (FOCS), pages 337–347, 1986.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] B. Barak, M. Hardt, I. Haviv, A. Rao, O. Regev, and D. Steurer. Rounding parallel repetitions of unique games. In *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 374–383, 2008.

[10] 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* (FOCS), pages 53–62, 2009.

[11] 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.

[12] 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* (FOCS), pages 477–486, 2008.

[13] E. Boros and P. L. Hammer. Pseudo-Boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155–225, 2002.

[14] 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.

[15] H. Buhrman, R. Cleve, R. de Wolf, and C. Zalka. Bounds for small-error and zero-error quantum algorithms. In *Proceedings of the Fortieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 358–368, 1999.

[16] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing* (STOC), pages 63–68, 1998.

[17] H. Buhrman, N. K. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity* (CCC), pages 24–32, 2007.

[18] 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* (STOC), pages 94–99, 1983.

[19] 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* (FOCS), pages 449–458, 2007.

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

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

[22] Á. M. del Rey, G. R. Sánchez, and A. de la Villa Cuenca. On the Boolean partial derivatives and their composition. *Appl. Math. Lett.*, 25(4):739–744, 2012.

[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] P. Gopalan, A. Shpilka, and S. Lovett. The complexity of Boolean functions in different characteristics. *Computational Complexity*, 19(2):235–263, 2010.

[25] B. Green. Finite field models in additive combinatorics. *Surveys in Combinatorics, London Math. Soc. Lecture Notes*, 327:1–27, 2005.

[26] B. Green and T. Tao. An inverse theorem for the Gowers U^3-norm, with applications. *Proc. Edinburgh Math. Soc.*, 51(1):73–153, 2008.

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

[28] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* (STOC), pages 212–219, 1996.

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

[30] 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* (STOC), pages 599–608, 2008.

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

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

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

[34] 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.

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

[36] L. Le Cam and G. L. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer, 2nd edition, 2000.

[37] 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* (CCC), pages 254–262, 2009.

[38] T. Lee and A. Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

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

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

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

[42] D. Pollard. *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, 2001.

[43] R. Raz. A counterexample to strong parallel repetition. *SIAM J. Comput.*, 40(3):771–777, 2011.

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

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

[46] 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* (STOC), 2007.

[47] A. A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. Preliminary version in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing* (STOC), 2008.

[48] A. A. Sherstov. The multiparty communication complexity of set disjointness. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), pages 525–544, 2012.

[49] A. A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM J. Comput.*, 41(5):1122–1165, 2012. Preliminary version in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (STOC), 2011.

[50] Y. Shi and Y. Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5–6):444–460, 2009.
[51] P. Tesson. *Computational complexity questions related to finite monoids and semigroups*. PhD thesis, McGill University, 2003.
[52] P. S. Thomas W. Cusick. *Cryptographic Boolean Functions and Applications*. Academic Press, 2009.
[53] G. Y. Vichniac. Boolean derivatives on cellular automata. *Physica D*, 45:63–74, 1990.
[54] E. Viola and A. Wigderson. One-way multiparty communication lower bound for pointer jumping with applications. *Combinatorica*, 29(6):719–743, 2009.
[55] A. C.-C. Yao. Theory and applications of trapdoor functions. In *Proceedings of the Twenty-Third Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 80–91, 1982.

ECCC ISSN 1433-8092
http://eccc.hpi-web.de