Inner Product and Set Disjointness: Beyond Logarithmically Many Parties

VLADIMIR V. PODOLSKII, Steklov Mathematical Institute, Russia and Higher School of Economics, Russia
ALEXANDER A. SHERSTOV, University of California, Los Angeles, USA

A major goal in complexity theory is to understand the communication complexity of number-on-the-forehead problems $f\colon (\{0,1\}^n)^k \to \{0,1\}$ with $k \gg \log n$ parties. We study the problems of inner product and set disjointness and determine their randomized communication complexity for every $k \geq \log n$, showing in both cases that $\Theta(1 + \lceil \log n \rceil / \log \lceil 1 + k/\log n \rceil)$ bits are necessary and sufficient. In particular, these problems admit constant-cost protocols if and only if the number of parties is $k \geq n^\epsilon$ for some constant $\epsilon > 0$.

CCS Concepts: • **Theory of computation → Communication complexity**.

Additional Key Words and Phrases: communication complexity, inner product, set disjointness, number on the forehead

ACM Reference Format:
Vladimir V. Podolskii and Alexander A. Sherstov. 2020. Inner Product and Set Disjointness: Beyond Logarithmically Many Parties. *ACM Trans. Comput. Theory* X, X, Article X (December 2020), 30 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The *number-on-the-forehead* model, due to Chandra et al. [8], is the standard model of multiparty communication. The model features k collaborative players and a Boolean function $F\colon X_1 \times X_2 \times \cdots \times X_k \to \{0,1\}$ with k arguments. An input $(x_1, x_2, \ldots, x_k)$ is distributed among the k players with overlap, by giving the ith player the arguments $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$ but not x_i. This arrangement can be visualized as having the k players seated in a circle with x_i written on the ith player's forehead, whence the name of the model. The players communicate according to a protocol agreed upon in advance. The communication occurs in the form of broadcasts, with a message sent by any given player instantly reaching everyone else. The players' objective is to compute F on any given input with minimal communication. We are specifically interested in *randomized* protocols, where the players have an unbounded supply of shared random bits. The *cost* of a protocol is the total bit length of all the messages broadcast in a worst-case execution. The *ϵ-error randomized communication complexity* $R_\epsilon(F)$ is the least cost of a randomized protocol that computes F with probability of error at most ϵ on every input.

Number-on-the-forehead communication complexity is a natural subject of study in its own right, in addition to its applications to circuit complexity, pseudorandomness, and proof complexity [3,

Authors' addresses: Vladimir V. Podolskii, podolskii@mi-ras.ru, Steklov Mathematical Institute, 8 Gubkina Street, Moscow, Russia, Higher School of Economics, 11 Pokrovsky Bulvar, Moscow, Russia; Alexander A. Sherstov, sherstov@cs.ucla.edu, University of California, Los Angeles, 404 Westwood Plaza, Los Angeles, California, USA, CA 90095.

1942-3454/2020/12-ARTX $15.00
https://doi.org/10.1145/1122445.1122456

27, 15, 23, 6]. Number-on-the-forehead is the most studied model in the area because any other way of assigning arguments to players results in a less powerful formalism—provided of course that one does not assign all the arguments to some player, in which case there is never a need to communicate. The generous overlap in the players' inputs makes proving lower bounds in the number-on-the-forehead model difficult. The strongest lower bound for an explicit communication problem $F\colon (\{0,1\}^n)^k \to \{0,1\}$ is currently $\Omega(n/2^k)$, obtained by Babai et al. [3] almost thirty years ago. This lower bound becomes trivial at $k = \log n$, and it is a longstanding open problem to overcome this logarithmic barrier and prove strong lower bounds for an explicit function with $k \gg \log n$. As one would expect, the existence of such functions is straightforward to prove using a counting argument [3, 13]. In particular, it is known [13, Sec. 9.2] that for all n and k, a uniformly random function $F\colon (\{0,1\}^n)^k \to \{0,1\}$ almost surely has randomized communication complexity

$$R_{1/3}(F) \geq n - 5, \tag{1.1}$$

which essentially meets the trivial upper bound $R_{1/3}(F) \leq n + 1$.

The two most studied problems in communication complexity theory are (*generalized*) *inner product* and *set disjointness*. In the k-party versions of these problems, the inputs are subsets $S_1, S_2, \ldots, S_k \subseteq \{1, 2, \ldots, n\}$. As usual, the ith player knows $S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_k$ but not S_i. In the inner product problem, the objective is to determine whether $|\bigcap S_i|$ is odd. In set disjointness, the objective is to determine whether $\bigcap S_i = \varnothing$. In Boolean form, these two functions are given by the formulas

$$\mathrm{GIP}_{n,k}(X) = \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{k} X_{i,j},$$

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

respectively, where the input is an $n \times k$ Boolean matrix $X \in \{0,1\}^{n \times k}$ whose columns are the characteristic vectors of the input sets. In the setting of two players, the communication complexity is well-known to be $\Theta(n)$ for both inner product [11] and set disjointness [17, 22, 4]. A moment's thought reveals that the k-party communication complexity of these problems is monotonically nonincreasing in k, and determining this dependence has been the subject of extensive research in the area [3, 14, 26, 7, 19, 9, 5, 25, 24]. On the upper bounds side, Grolmusz [14] proved that k-party inner product has communication complexity $O(k\lceil n/2^k \rceil)$, which easily carries over to k-party set disjointness. The best lower bounds to date are $\Omega(n/4^k)$ for inner product, due to Babai et al. [3]; and $\Omega(n/4^k)^{1/4}$ and $\Omega(\sqrt{n}/2^k k)$ for set disjointness, due to Sherstov [25, 24].

1.1 Our results

Our work began with a basic question: how many players k does it take to compute inner product and set disjointness with *constant* communication? As discussed above, the best bounds on the communication complexity of these functions for large k prior to this paper were $\Omega(1)$ and $O(\log n)$. We close this logarithmic gap, determining the communication complexity up to a multiplicative constant for every $k \geq \log n$.

Theorem 1.1 (Main result). *For any $k \geq \log n$, inner product and set disjointness have randomized communication complexity*

$$R_{1/3}(\mathrm{GIP}_{n,k}) = \Theta\left(\frac{\log n}{\log\left\lceil 1 + \frac{k}{\log n}\right\rceil} + 1\right),$$

$$R_{1/3}(\mathrm{DISJ}_{n,k}) = \Theta\left(\frac{\log n}{\log\left\lceil 1 + \frac{k}{\log n}\right\rceil} + 1\right).$$

To our knowledge, Theorem 1.1 is the first nontrivial (i.e., superconstant) lower bound for any explicit communication problem with $k \geq \log n$ players. In particular, inner product and set disjointness have communication protocols with constant cost if and only if the number of players is $k \geq n^{\epsilon}$ for some constant $\epsilon > 0$. It is noteworthy that we prove the upper bounds in Theorem 1.1 using *simultaneous* protocols, where the players do not interact. In more detail, each player in a simultaneous protocol broadcasts all his messages at once and without regard to the messages from the other players. The output of a simultaneous protocol is fully determined by the shared randomness and the concatenation of the messages broadcast by all the players. The cost of a simultaneous protocol is defined in the usual way, as the total number of bits broadcast by all the players in a worst-case execution. Theorem 1.1 shows that as far as inner product and set disjointness are concerned, simultaneous protocols are asymptotically as powerful as general ones.

A natural next step is to construct a problem $F_{n,k} \colon (\{0,1\}^n)^k \to \{0,1\}$ whose communication complexity remains nontrivial for all k. Its *existence* follows from the lower bound (1.1) on the communication complexity of random functions. In the theorem below, we give an explicit function with communication complexity at least $c \log n$ for some absolute constant $c > 0$ and all n and k. We remind the reader that MOD_m stands for the Boolean function that evaluates to true if and only if the sum of its arguments is a multiple of m.

Theorem 1.2. *Define $F_{n,k} \colon (\{0,1\}^n)^k \to \{0,1\}$ by*

$$F_{n,k}(X) = \mathrm{MOD}_3\left(\bigoplus_{j=1}^{k} X_{1,j}, \ldots, \bigoplus_{j=1}^{k} X_{n,j}\right).$$

Then

$$R_{1/3}(F_{n,k}) \geq \frac{1}{3}\log n - \frac{1}{3}.$$

As with inner product and set disjointness, we show that the lower bound of Theorem 1.2 is asymptotically tight for all $k \geq \log n$.

1.2 Our techniques

The upper bounds in Theorem 1.1 are based on Grolmusz's deterministic protocol for multiparty inner product [14], which we are able to speed up using public randomness. The lower bounds in Theorems 1.1 and 1.2 are more subtle. First of all, it may be surprising that we are able to prove any lower bounds at all for $k \gg \log n$ players since all known techniques for explicit functions stop working at $k = \log n$ players. The key is to realize that we only need to rule out communication protocols with cost $O(\log n)$, and in any given execution of such a protocol all but $O(\log n)$ players remain silent! This makes it possible to reduce the analysis to the setting of $k \leq \log n$ players, where strong lower bounds are known. This reduction involves constructing an input distribution such that the portion of the input seen by any small set of players does not significantly help with

computing the output. Our communication lower bounds use the *discrepancy method*, which we adapt here to reflect the number of active players.

The remainder of this paper is organized as follows. Section 2 gives a thorough review of the technical preliminaries. Our results for inner product and set disjointness are presented in Sections 3 and 4, respectively. Section 5 concludes the paper with a proof of Theorem 1.2 along with a matching upper bound.

2 PRELIMINARIES

2.1 General

We use lowercase letters for vectors and strings, and uppercase letters for matrices. The empty string is denoted ε. For a bit string $x \in \{0,1\}^n$, we let $|x| = x_1 + x_2 + \cdots + x_n$ denote the Hamming weight of x. We let the bar operator $\overline{\ }$ denote either complex conjugation or set complementation, depending on the nature of the argument. For convenience, we adopt the convention that $0/0 = 0$. The notation $\log x$ refers to the logarithm of x to base 2.

We will view Boolean functions as mappings $f\colon X \to \{0,1\}$ for a finite set X, typically $X = \{0,1\}^n$. A *partial function* f on a set X is a function whose domain of definition, denoted $\operatorname{dom} f$, is a proper subset of X. For (possibly partial) Boolean functions f and g on $\{0,1\}^n$ and X, respectively, the symbol $f \circ g$ refers to the (possibly partial) Boolean function on X^n given by $(f \circ g)(x_1, x_2, \ldots, x_n) = f(g(x_1), g(x_2), \ldots, g(x_n))$. Clearly, the domain of $f \circ g$ is the set of all $(x_1, \ldots, x_n) \in (\operatorname{dom} g)^n$ for which $(g(x_1), g(x_2), \ldots, g(x_n)) \in \operatorname{dom} f$. As usual, for (possibly partial) Boolean functions f and g, the symbol $f \oplus g$ refers to the (possibly partial) Boolean function given by $(f \oplus g)(x, y) = f(x) \oplus g(y)$. Observe that in this notation, $f \oplus f$ and f are completely different functions. We abbreviate $f^{\oplus n} = f \oplus f \oplus \cdots \oplus f$ (n times). The familiar functions AND_n, OR_n, and XOR_n on the Boolean hypercube $\{0,1\}^n$ are given by $\mathrm{AND}_n(x) = \bigwedge_{i=1}^n x_i$, $\mathrm{OR}_n(x) = \bigvee_{i=1}^n x_i$, and $\mathrm{XOR}_n(x) = \bigoplus_{i=1}^n x_i$. We let $\mathrm{MOD}_3\colon \{0,1\}^* \to \{0,1\}$ be the Boolean function given by $\mathrm{MOD}_3(x) = 1 \Leftrightarrow |x| \equiv 0 \pmod 3$. Finally, we define a partial Boolean function $\widetilde{\mathrm{AND}}_n$ on $\{0,1\}^n$ as the restriction of AND_n to $\{x : |x| \geq n-1\}$. In other words,

$$\widetilde{\mathrm{AND}}_n(x) = \begin{cases} x_1 \wedge x_2 \wedge \cdots \wedge x_n & \text{if } |x| \geq n-1, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

We let $X^{n \times k}$ denote the family of $n \times k$ matrices with entries in X, the most common cases being those of real matrices ($X = \mathbb{R}$) and Boolean matrices ($X = \{0,1\}$). For a matrix $M \in \mathbb{R}^{n \times m}$ and a set $S \subseteq \{1, 2, \ldots, n\}$, we let $M|_S$ denote the submatrix of M obtained by keeping the rows with index in S. More generally, for sets $S \subseteq \{1, 2, \ldots, n\}$ and $T \subseteq \{1, 2, \ldots, m\}$, we let $M|_{S,T}$ denote the $|S| \times |T|$ submatrix of M obtained by keeping the rows with index in S and columns with index in T. We adopt the standard convention that the ordering of the rows (and columns) in a submatrix is inherited from the containing matrix.

For nonnegative integers n and k, we define

$$\binom{n}{\leq k} := \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k} = \sum_{i=0}^{\min\{k,n\}} \binom{n}{i}.$$

The following bounds are well-known [16, Proposition 1.4]:

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{\leq k} \leq \left(\frac{\mathrm{e}n}{k}\right)^k \qquad (1 \leq k \leq n). \tag{2.1}$$

2.2 Analytic preliminaries

For a finite set X, we let $\mathbb{R}^X$ denote the linear space of real functions $X \to \mathbb{R}$. This space is equipped with the usual norms and inner product:

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

The support of $\phi \in \mathbb{R}^X$ is the subset $\operatorname{supp} \phi = \{x \in X : \phi(x) \neq 0\}$. The pointwise (Hadamard) product of $\phi, \psi \in \mathbb{R}^X$ is denoted $\phi \cdot \psi \in \mathbb{R}^X$ and given by $(\phi \cdot \psi)(x) = \phi(x)\psi(x)$. The tensor product of $\phi \in \mathbb{R}^X$ and $\psi \in \mathbb{R}^Y$ is the function $\phi \otimes \psi \in \mathbb{R}^{X \times Y}$ given by $(\phi \otimes \psi)(x, y) = \phi(x)\psi(y)$. The tensor product $\phi \otimes \phi \otimes \cdots \otimes \phi$ (n times) is abbreviated $\phi^{\otimes n}$. Tensor product notation generalizes to partial functions in the natural way: if ϕ and ψ are partial real functions on X and Y, respectively, then $\phi \otimes \psi$ is a partial function on $X \times Y$ with domain $\operatorname{dom} \phi \times \operatorname{dom} \psi$ and is given by $(\phi \otimes \psi)(x, y) = \phi(x)\psi(y)$ on that domain. Similarly, $\phi^{\otimes n}$ is a partial function on X^n with domain $(\operatorname{dom} \phi)^n$.

We now recall the Fourier transform on $\{0,1\}^n$. For a subset $S \subseteq \{1, 2, \ldots, n\}$, define $\chi_S \colon \{0,1\}^n \to \{-1, +1\}$ by $\chi_S(x) = \prod_{i \in S} (-1)^{x_i}$. Then every function $\phi \colon \{0,1\}^n \to \mathbb{R}$ has a unique representation of the form $\phi = \sum_S \hat{\phi}(S)\, \chi_S$, where $\hat{\phi}(S) = \mathrm{E}_{x \in \{0,1\}^n} \phi(x)\chi_S(x)$. The reals $\hat{\phi}(S)$ are the *Fourier coefficients* of ϕ, and the mapping $\phi \mapsto \hat{\phi}$ is the *Fourier transform* of ϕ.

2.3 Probability

We view probability distributions first and foremost as real functions and use the notational shorthands above. In particular, we write $\operatorname{supp} \mu$ to refer to the support of the probability distribution μ, and $\mu \otimes \lambda$ to refer to the tensor product of the distributions μ and λ. The notation $X \sim \mu$ means that the random variable X is distributed according to μ. We let $B(n, p)$ denote the binomial distribution with n trials and success probability p.

Fact 2.1. *For any integer $n \geq 1$,*

$$\mathop{\mathrm{E}}_{s \sim B(n-1,p)} \frac{1}{\sqrt{n-s}} \leq \frac{1}{\sqrt{(1-p)n}}, \tag{2.2}$$

$$\mathop{\mathrm{E}}_{s \sim B(n-1,q)} \frac{1}{\sqrt{s+1}} \leq \frac{1}{\sqrt{qn}}, \tag{2.3}$$

$$\mathop{\mathrm{E}}_{s \sim B(n,p)} |s - pn| \leq \sqrt{p(1-p)n}. \tag{2.4}$$

Proof. For (2.2), we have

$$\begin{aligned}\left(\mathop{\mathrm{E}}_{s\sim B(n-1,p)}\frac{1}{\sqrt{n-s}}\right)^2 &\le \mathop{\mathrm{E}}_{s\sim B(n-1,p)}\frac{1}{n-s}\\ &= \sum_{s=0}^{n-1}\binom{n-1}{s}\frac{p^s(1-p)^{n-1-s}}{n-s}\\ &= \frac{1}{(1-p)n}\sum_{s=0}^{n-1}\binom{n}{s}p^s(1-p)^{n-s}\\ &= \frac{1-p^n}{(1-p)n},\end{aligned}$$

where the first step follows from the Cauchy-Schwarz inequality, and the last step uses the binomial theorem. The bound (2.3) follows from (2.2) since the distribution of $s+1$ for $B(n-1,q)$ is the same as the distribution of $n-s$ for $s\sim B(n-1,1-q)$. For (2.4),

$$\begin{aligned}\left(\mathop{\mathrm{E}}_{s\sim B(n,p)}|s-pn|\right)^2 &\le \mathop{\mathrm{E}}_{s\sim B(n,p)}[(s-pn)^2]\\ &= p(1-p)n,\end{aligned}$$

where the first step uses the Cauchy-Schwarz inequality, and the second step uses the fact that the binomial distribution $B(n,p)$ has variance $p(1-p)n$. □

2.4 Approximation by polynomials

We let $\deg p$ denote the total degree of a multivariate polynomial p. In this paper, we use the terms "degree" and "total degree" interchangeably, preferring the former for brevity. Let $\phi\colon X\to\mathbb{R}$ be given, for a finite subset $X\subset\mathbb{R}^n$. The *ϵ-approximate degree* of ϕ, denoted $\deg_\epsilon(\phi)$, is the least degree of a real polynomial p such that $\|\phi-p\|_\infty\le\epsilon$. We generalize this definition to partial functions ϕ on X by defining $\deg_\epsilon(\phi)$ as the least degree of a real polynomial p with

$$\left.\begin{aligned}&|\phi(x)-p(x)|\le\epsilon, && x\in\operatorname{dom}\phi,\\ &|p(x)|\le 1+\epsilon, && x\in X\setminus\operatorname{dom}\phi.\end{aligned}\right\} \tag{2.5}$$

For a (possibly partial) real function ϕ on a finite subset $X\subset\mathbb{R}^n$, we define $E(\phi,d)$ to be the least ϵ such that (2.5) holds for some polynomial of degree at most d. In this notation, $\deg_\epsilon(\phi)=\min\{d: E(\phi,d)\le\epsilon\}$. The canonical setting of the error parameter is $\epsilon=1/3$, which is without loss of generality since the error in a uniform approximation of a Boolean function can be reduced from any given constant in $(0,1/2)$ to any other constant in $(0,1/2)$ with only a constant-factor increase in the degree of the approximant. One of the earliest results on the approximation of Boolean functions by polynomials is the following seminal theorem due to Nisan and Szegedy [20].

Theorem 2.2 (Nisan and Szegedy).

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

2.5 Multiparty communication

An excellent introduction to communication complexity theory is the monograph by Kushilevitz and Nisan [18]. In our overview, we will limit ourselves to key definitions and notation. This paper uses the standard model of randomized multiparty communication known as the *number-on-the-forehead model* [8]. Let F be a (possibly partial) Boolean function on $X_1\times X_2\times\cdots\times X_k$, for some finite sets $X_1,X_2,\ldots,X_k$. The model features k players. A given input $(x_1,x_2,\ldots,x_k)\in X_1\times X_2\times\cdots\times X_k$ is

distributed among the players 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 players communicate according to an agreed-upon protocol by writing bits on a shared blackboard, visible to them all. They additionally have access to a shared source of random bits which they can use in deciding what messages to send. At any given point, the identity of the player who is to speak next is determined by the shared random string and the communication so far (i.e., the contents of the shared blackboard). This rules out the possibility of crosstalk. The speaker's message may depend on the arguments known to that player, in addition to the shared random string and communication so far. The players' goal is to accurately compute the value of F on any given input in the domain of F. An *ϵ-error communication 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 on any input. The *ϵ-error randomized communication complexity of F*, denoted $R_\epsilon(F)$, is the least cost of an ϵ-error randomized communication protocol for F. As usual, the standard setting of the error parameter is $\epsilon = 1/3$, which is without loss of generality since the error probability in a communication protocol can be efficiently reduced by running the protocol several times independently and outputting the majority answer.

The communication problems of interest to us are *generalized inner product* $\mathrm{GIP}_{n,k} \colon \{0,1\}^{n \times k} \to \{0,1\}$ and *set disjointness* $\mathrm{DISJ}_{n,k} \colon \{0,1\}^{n \times k} \to \{0,1\}$, given by

$$\mathrm{GIP}_{n,k}(X) = \bigoplus_{i=1}^{n} \bigwedge_{j=1}^{k} X_{i,j},$$

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

These k-party communication problems are both defined on $n \times k$ matrices, where the ith party receives as input all but the ith column of the matrix. The disjointness function evaluates to true if and only if the input matrix does not have an all-ones row, whereas the generalized inner product function evaluates to true if and only if the number of all-ones rows is odd. We also consider a partial Boolean function $\mathrm{UDISJ}_{n,k}$ on $\{0,1\}^{n \times k}$, called *unique set disjointness* and defined as the restriction of $\mathrm{DISJ}_{n,k}$ to matrices with at most one all-ones row. In other words, $\mathrm{UDISJ}_{n,k}(X)$ is undefined if X has two or more all-ones rows, and is given by $\mathrm{UDISJ}_{n,k}(X) = \mathrm{DISJ}_{n,k}(X)$ otherwise.

Let G be a (possibly partial) Boolean function on $X_1 \times X_2 \times \cdots \times X_k$, representing a k-party communication problem, and let f be a (possibly partial) Boolean function on $\{0,1\}^n$. We view the composition $f \circ G$ as a k-party communication problem on $X_1^n \times X_2^n \times \cdots \times X_k^n$. It will be helpful to keep in mind that for all positive integers m and n,

$$\mathrm{GIP}_{mn,k} = \mathrm{XOR}_m \circ \mathrm{GIP}_{n,k}, \tag{2.6}$$

$$\mathrm{DISJ}_{mn,k} = \mathrm{AND}_m \circ \mathrm{DISJ}_{n,k}, \tag{2.7}$$

$$\mathrm{UDISJ}_{mn,k} = \widetilde{\mathrm{AND}}_m \circ \mathrm{UDISJ}_{n,k}. \tag{2.8}$$

Similarly, if F_i for $i = 1, 2, \ldots, m$ is a (possibly partial) k-party communication problem on $X_{i,1} \times X_{i,2} \times \cdots \times X_{i,k}$, we view $\bigoplus_{i=1}^{m} F_i$ as a k-party communication problem on $(\prod X_{i,1}) \times (\prod X_{i,2}) \times \cdots \times (\prod X_{i,k})$.

2.6 Cylinder intersections

Let $X_1, X_2, \ldots, X_k$ be nonempty finite sets. A *cylinder intersection* on $X_1 \times X_2 \times \cdots \times X_k$ is any function $\chi : X_1 \times X_2 \times \cdots \times X_k \to \{0,1\}$ of the form

$$\chi(x_1, \ldots, x_k) = \prod_{i=1}^{k} \chi_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k), \tag{2.9}$$

where $\chi_i : X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_k \to \{0,1\}$. In other words, a cylinder intersection is the product of k Boolean functions, where the ith function does not depend on the ith coordinate but may depend arbitrarily on the other $k-1$ coordinates. For a given set $S \subseteq \{1,2,\ldots,k\}$, we further specialize this notion to *S-cylinder intersections*, defined as functions of the form

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

for some $\chi_i : X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_k \to \{0,1\}$. Finally, an *$\ell$-cylinder intersection* on $X_1 \times X_2 \times \cdots \times X_k$ is any S-cylinder intersection for a subset $S \subseteq \{1,2,\ldots,k\}$ of cardinality at most ℓ. Cylinder intersections were introduced by Babai, Nisan, and Szegedy [3] and play a fundamental role in the theory due to the following fact.

Fact 2.3. *Let $\Pi : X_1 \times X_2 \times \cdots \times X_k \to \{0,1\}$ be a deterministic k-party communication protocol with cost c. Then*

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

for some $\min\{c,k\}$-cylinder intersections $\chi_1, \ldots, \chi_{2^c}$ and some $a_1, \ldots, a_{2^c} \in \{0,1\}$.

We refer the reader to [18] for a simple proof of Fact 2.3. Recall that a randomized protocol of cost c is a probability distribution on deterministic protocols of cost c. With this in mind, one easily infers the following from Fact 2.3:

Corollary 2.4. *Let F be a (possibly partial) Boolean function on $X_1 \times X_2 \times \cdots \times X_k$. If $R_\epsilon(F) = c$, then*

$$\begin{aligned} &|F(x_1, \ldots, x_k) - \Pi(x_1, \ldots, x_k)| \leq \epsilon, && (x_1, \ldots, x_k) \in \operatorname{dom} F, \\ &|\Pi(x_1, \ldots, x_k)| \leq 1, && (x_1, \ldots, x_k) \in X_1 \times X_2 \times \cdots \times X_k, \end{aligned}$$

where $\Pi = \sum_\chi a_\chi \chi$ is a linear combination of $\min\{c,k\}$-cylinder intersections with $\sum_\chi |a_\chi| \leq 2^c$.

We follow the tradition of denoting both cylinder intersections and Fourier characters by the letter χ, subscripted as appropriate. The meaning in each case will be clear from the context.

2.7 Discrepancy

For a (possibly partial) Boolean function F on $X_1 \times X_2 \times \cdots \times X_k$, a probability distribution μ on the domain of F, and a set $S \subseteq \{1,2,\ldots,k\}$, the *$S$-discrepancy of F with respect to μ* is defined as

$$\begin{aligned} \operatorname{disc}_S(F, \mu) &= \max_\chi |\langle (-1)^F, \mu \cdot \chi \rangle| \\ &= \max_\chi \left| \sum_{x \in \operatorname{dom} F} (-1)^{F(x)} \mu(x) \chi(x) \right|, \end{aligned}$$

where the maximum is over S-cylinder intersections χ. Further maximizing over S gives the key notions of *ℓ-discrepancy* and *discrepancy*, as follows:

$$\operatorname{disc}_\ell(F,\mu) = \max_{\substack{S\subseteq\{1,2,\ldots,k\}\\ |S|\le\ell}} \operatorname{disc}_S(F,\mu),$$
$$\operatorname{disc}(F,\mu) = \max_{S\subseteq\{1,2,\ldots,k\}} \operatorname{disc}_S(F,\mu).$$

By definition,

$$\operatorname{disc}_S(F,\mu) \le \operatorname{disc}_\ell(F,\mu) \le \operatorname{disc}(F,\mu)$$

for every $\ell = 0, 1, \ldots, k$ and every set S of cardinality at most ℓ.

In light of Corollary 2.4, upper bounds on the discrepancy give lower bounds on the communication complexity. This fundamental technique is known as the *discrepancy method* [11, 3, 18]:

Theorem 2.5 (Discrepancy method). *For every (possibly partial) Boolean function F on $X_1 \times X_2 \times \cdots \times X_k$ and every probability distribution μ on the domain of F,*

$$2^{R_\epsilon(F)} \ge \frac{1-2\epsilon}{\operatorname{disc}(F,\mu)}. \tag{2.10}$$

More generally,

$$2^{R_\epsilon(F)} \ge \frac{1-2\epsilon}{\operatorname{disc}_{\min\{R_\epsilon(F),k\}}(F,\mu)}. \tag{2.11}$$

A proof of (2.10) can be found in [25]; that same proof carries over to discrepancy with respect to any given family χ of functions and in particular establishes (2.11) as well.

A useful property of discrepancy is its convexity in the second argument, as formalized by the following proposition.

Proposition 2.6 (Convexity of discrepancy). *Let F be a (possibly partial) Boolean function on $X_1 \times X_2 \times \cdots \times X_k$, and let μ and λ be probability distributions on the domain of F. Then for every $S \subseteq \{1, 2, \ldots, k\}$ and $0 \le p \le 1$,*

$$\operatorname{disc}_S(F, p\mu + (1-p)\lambda) \le p \operatorname{disc}_S(F,\mu) + (1-p)\operatorname{disc}_S(F,\lambda),$$

and likewise for disc_ℓ and disc.

By induction, Proposition 2.6 immediately generalizes to any finite convex combination of probability distributions. It is this more general form that we will invoke in our applications.

Proof of Proposition 2.6. Immediate from the following inequality for any cylinder intersection χ:

$$|\langle(-1)^F, (p\mu + (1-p)\lambda)\cdot\chi\rangle| \le p\cdot|\langle(-1)^F, \mu\cdot\chi\rangle| + (1-p)\cdot|\langle(-1)^F, \lambda\cdot\chi\rangle|,$$

where for partial functions F the inner products are restricted to the domain of F. □

It is clear that discrepancy is a continuous function of the input distribution. The following proposition quantifies this continuity.

Proposition 2.7 (Continuity of discrepancy). *For any k-party communication problem $F\colon X_1 \times X_2 \times \cdots \times X_k \to \{0,1\}$ and any probability distributions μ and $\tilde{\mu}$ on the domain of F,*

$$\operatorname{disc}(F,\mu) \le \operatorname{disc}(F,\tilde{\mu}) + \|\mu - \tilde{\mu}\|_1.$$

More generally, for any $S \subseteq \{1,2,\ldots,k\}$, any (possibly partial) k-party communication problems $F_1, F_2, \ldots, F_m$, and any probability distributions $\mu_1, \mu_2, \ldots, \mu_m$ and $\tilde{\mu}_1, \tilde{\mu}_2, \ldots, \tilde{\mu}_m$ on the corresponding domains,

$$\operatorname{disc}_S \left(\bigoplus_{i=1}^m F_i, \bigotimes_{i=1}^m \mu_i \right) \leq \sum_{A \subseteq \{1,2,\ldots,m\}} \operatorname{disc}_S \left(\bigoplus_{i \in A} F_i, \bigotimes_{i \in A} \tilde{\mu}_i \right) \prod_{i \notin A} \|\mu_i - \tilde{\mu}_i\|_1$$

and likewise for disc_ℓ *and* disc.

Proof. It suffices to prove the claim for disc_S. Fix a set $S \subseteq \{1,2,\ldots,k\}$ and an S-cylinder intersection χ with

$$\operatorname{disc}_S \left(\bigoplus_{i=1}^m F_i, \bigotimes_{i=1}^m \mu_i \right) = \left| \left\langle \bigotimes_{i=1}^m (-1)^{F_i}, \chi \cdot \bigotimes_{i=1}^m \mu_i \right\rangle \right|,$$

where as usual the inner product on the right-hand side is restricted to $\prod \operatorname{dom} F_i$. Then

$$\begin{aligned} \operatorname{disc}_S \left(\bigoplus_{i=1}^m F_i, \bigotimes_{i=1}^m \mu_i \right) &= \left| \left\langle \bigotimes_{i=1}^m (-1)^{F_i}, \chi \cdot \bigotimes_{i=1}^m (\tilde{\mu}_i + (\mu_i - \tilde{\mu}_i)) \right\rangle \right| \\ &= \left| \sum_{A \subseteq \{1,2,\ldots,m\}} \left\langle \bigotimes_{i=1}^m (-1)^{F_i}, \chi \cdot \Lambda_A \right\rangle \right|, \end{aligned} \tag{2.12}$$

where Λ_A is given by $\Lambda_A(x_1, x_2 \ldots, x_m) = \prod_{i \in A} \tilde{\mu}_i(x_i) \cdot \prod_{i \notin A} (\mu_i(x_i) - \tilde{\mu}_i(x_i))$. Continuing,

$$\begin{aligned} \left| \left\langle \bigotimes_{i=1}^m (-1)^{F_i}, \chi \cdot \Lambda_A \right\rangle \right| &= \left| \sum_{x_1,\ldots,x_m} \chi(x) \prod_{i \in A} (-1)^{F_i(x_i)} \tilde{\mu}_i(x_i) \cdot \prod_{i \notin A} (-1)^{F_i(x_i)} (\mu_i(x_i) - \tilde{\mu}_i(x_i)) \right| \\ &\leq \sum_{x_i : i \notin A} \left| \sum_{x_i : i \in A} \chi(x) \prod_{i \in A} (-1)^{F_i(x_i)} \tilde{\mu}_i(x_i) \right| \prod_{i \notin A} |\mu_i(x_i) - \tilde{\mu}_i(x_i)| \\ &\leq \sum_{x_i : i \notin A} \operatorname{disc}_S \left(\bigoplus_{i \in A} F_i, \bigotimes_{i \in A} \tilde{\mu}_i \right) \prod_{i \notin A} |\mu_i(x_i) - \tilde{\mu}_i(x_i)| \\ &= \operatorname{disc}_S \left(\bigoplus_{i \in A} F_i, \bigotimes_{i \in A} \tilde{\mu}_i \right) \prod_{i \notin A} \|\mu_i - \tilde{\mu}_i\|_1, \end{aligned} \tag{2.13}$$

where the next to last step is legitimate because for any fixing of x_i for $i \notin A$, the function χ continues to be an S-cylinder intersection with respect to the remaining coordinates. In view of (2.12) and (2.13), the proof is complete. □

3 INNER PRODUCT

In this section, we determine the communication complexity of the inner product problem $\mathrm{GIP}_{n,k}$ for $k \geq \log n$ players. Our proofs build on the classic lower and upper bounds for this problem for $k \leq \log n$, due to Babai et al. [3] and Grolmusz [14], respectively.

3.1 Lower bound

For the lower bound, we use the generalization of the discrepancy method given by Theorem 2.5. We will work with the following input distribution.

Definition 3.1. Let $\upsilon_{n,k,\ell}$ denote the probability distribution on $n \times k$ Boolean matrices whereby each row is chosen independently and uniformly at random from the set $\{u \in \{0,1\}^k : |u| \geq k - \ell\}$.

In particular, $\upsilon_{n,k,k}$ is the uniform probability distribution on $\{0,1\}^{n\times k}$. In this special case, a strong upper bound on the discrepancy of generalized inner product was obtained in the seminal work of Babai, Nisan, and Szegedy [3].

Theorem 3.2 (Babai, Nisan, and Szegedy). *For any positive integers n and k,*

$$\operatorname{disc}(\mathrm{GIP}_{n,k}, \upsilon_{n,k,k}) \leq \left(1 - \frac{1}{4^{k-1}}\right)^n.$$

We generalize this discrepancy bound to $\upsilon_{n,k,\ell}$ for any ℓ.

Theorem 3.3. *For any positive integers n, k, ℓ with $\ell \leq k$,*

$$\operatorname{disc}_\ell(\mathrm{GIP}_{n,k}, \upsilon_{n,k,\ell}) \leq \left(1 - \frac{1}{2^{\ell-2}\binom{k}{\leq \ell}}\right)^n.$$

For $\ell = k$, this discrepancy bound is identical to that of Theorem 3.2. In the setting $\ell \ll k$ of interest to us, however, the new bound is substantially stronger.

Proof of Theorem 3.3. Since the communication problem $\mathrm{GIP}_{n,k}$ and the probability distribution $\upsilon_{n,k,\ell}$ are both symmetric with respect to the k players, we have

$$\operatorname{disc}_\ell(\mathrm{GIP}_{n,k}, \upsilon_{n,k,\ell}) = \operatorname{disc}_{\{1,2,\ldots,\ell\}}(\mathrm{GIP}_{n,k}, \upsilon_{n,k,\ell}). \tag{3.1}$$

For a given set $S \subseteq \{1,2,\ldots,n\}$ and a probability distribution σ on matrices $X \in \{0,1\}^{n\times k}$, let $\sigma|_S$ stand for the probability distribution induced by σ after conditioning on the event that $(X_{i,\ell+1}, X_{i,\ell+2}, \ldots, X_{i,k}) = (1,1,\ldots,1)$ if and only if $i \in S$. Observe that $\upsilon_{n,k,\ell}|_S$ is a probability distribution on matrices $X \in \{0,1\}^{n\times k}$ whereby $X|_S$ and $X|_{\overline{S}}$ are distributed independently such that

$$X|_{S,\{1,2,\ldots,\ell\}} \sim \upsilon_{|S|,\ell,\ell}, \tag{3.2}$$

$$X|_{S,\{\ell+1,\ell+2,\ldots,k\}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}, \tag{3.3}$$

and $X|_{\overline{S}}$ does not have an all-ones row. In particular, the rows in $X|_{\overline{S}}$ do not affect the value of the function. The conditional probability distribution of $X|_S$ given any value of $X|_{\overline{S}}$ is always (3.2)–(3.3). Viewing $\upsilon_{n,k,\ell}|_S$ as the convex combination of these conditional probability distributions, corresponding to every possible value of $X|_{\overline{S}}$, we conclude by Proposition 2.6 that

$$\operatorname{disc}_{\{1,2,\ldots,\ell\}}(\mathrm{GIP}_{n,k}, \upsilon_{n,k,\ell}|_S) \leq \operatorname{disc}(\mathrm{GIP}_{|S|,\ell}, \upsilon_{|S|,\ell,\ell}). \tag{3.4}$$

We will now express $\upsilon_{n,k,\ell}$ as a convex combination of probability distributions $\upsilon_{n,k,\ell}|_S$ and use the convexity of discrepancy to complete the proof. Specifically, let $p = 2^\ell / \binom{k}{\leq \ell}$. Then

$$\begin{aligned}
\operatorname{disc}_{\{1,2,\ldots,\ell\}}(\mathrm{GIP}_{n,k}, \upsilon_{n,k,\ell}) &= \operatorname{disc}_{\{1,2,\ldots,\ell\}}\left(\mathrm{GIP}_{n,k}, \mathop{\mathrm{E}}_{s\sim B(n,p)} \mathop{\mathrm{E}}_{\substack{S\subseteq\{1,2,\ldots,n\}\\|S|=s}} \upsilon_{n,k,\ell}|_S\right) \\
&\qquad \text{by definition of } \upsilon_{n,k,\ell} \\
&\leq \mathop{\mathrm{E}}_{s\sim B(n,p)} \mathop{\mathrm{E}}_{\substack{S\subseteq\{1,2,\ldots,n\}\\|S|=s}} \operatorname{disc}_{\{1,2,\ldots,\ell\}}(\mathrm{GIP}_{n,k}, \upsilon_{n,k,\ell}|_S) \\
&\qquad \text{by Proposition 2.6} \\
&\leq \mathop{\mathrm{E}}_{s\sim B(n,p)} \operatorname{disc}(\mathrm{GIP}_{s,\ell}, \upsilon_{s,\ell,\ell}) && \text{by (3.4)} \\
&\leq \mathop{\mathrm{E}}_{s\sim B(n,p)} \left(1 - \frac{1}{4^{\ell-1}}\right)^s && \text{by Theorem 3.2} \\
&= \sum_{s=0}^{n} \binom{n}{s} \left(1 - \frac{1}{4^{\ell-1}}\right)^s p^s (1-p)^{n-s} \\
&= \left(1 - \frac{p}{4^{\ell-1}}\right)^n \\
&= \left(1 - \frac{1}{2^{\ell-2}\binom{k}{\leq \ell}}\right)^n.
\end{aligned}$$

In light of (3.1), the proof is complete. □

As a corollary to the new bound on the discrepancy of generalized inner product, we obtain our claimed communication lower bound for this function.

Theorem 3.4. *Abbreviate* $R = R_{1/3}(\mathrm{GIP}_{n,k})$. *Then*

$$\binom{k}{\leq R}^2 R \geq \Omega(n). \tag{3.5}$$

In particular,

$$R_{1/3}(\mathrm{GIP}_{n,k}) = \Omega\left(\frac{\log n}{\log\left\lceil 1 + \frac{k}{\log n}\right\rceil} + 1\right). \tag{3.6}$$

Proof. We have

$$\begin{aligned}
2^R &\geq \frac{1}{3\operatorname{disc}_{\min\{R,k\}}(\mathrm{GIP}_{n,k}, \upsilon_{n,k,\min\{R,k\}})} && \text{by Theorem 2.5} \\
&\geq \frac{1}{3}\exp\left(\frac{n}{2^{\min\{R,k\}}\binom{k}{\leq \min\{R,k\}}}\right) && \text{by Theorem 3.3} \\
&\geq \frac{1}{3}\exp\left(\frac{n}{\binom{k}{\leq R}^2}\right),
\end{aligned}$$

settling (3.5). Now, recall from (2.1) that

$$\binom{k}{\leq R} \leq \mathrm{e}^R \left\lceil \frac{k}{R} \right\rceil^R .$$

Substituting this estimate in (3.5) gives $\mathrm{e}^{2R}\lceil k/R \rceil^{2R} R \geq \Omega(n)$, whence (3.6). □

3.2 Upper bound

We now prove a matching upper bound on the communication complexity of inner product for $k \geq \log n$ players. Our proof is based on Grolmusz's well-known *deterministic* protocol [14] for this function, which we are able to speed up using shared randomness. In addition to lower communication cost, the protocol below has the advantage of being simultaneous, which was not the case in previous work [14, 2].

Theorem 3.5. *For any $k \geq \log n$ and any constant $\epsilon > 0$,*

$$R_\epsilon(\mathrm{GIP}_{n,k}) = O\left(\frac{\log n}{\log\left\lceil 1 + \frac{k}{\log n} \right\rceil} + 1 \right).$$

Moreover, this upper bound is achieved by a simultaneous protocol.

Proof. We first consider the case

$$k \geq \log 3n. \tag{3.7}$$

Recall that the generalized inner product problem, $\mathrm{GIP}_{n,k}$, is the k-party problem of determining whether a given $n \times k$ Boolean matrix X contains an odd number of all-ones rows, where the ith party ($1 \leq i \leq k$) sees all the columns of X except for the ith column. Let ℓ denote the smallest natural number, $0 \leq \ell \leq k$, with the property that

$$\binom{k}{\leq \ell} \geq 3n. \tag{3.8}$$

Such ℓ exists by (3.7). Moreover, in view of the lower bound in (2.1), it is straightforward to verify that

$$\ell = O\left(\frac{\log n}{\log\left\lceil 1 + \frac{k}{\log n} \right\rceil} + 1 \right). \tag{3.9}$$

As the first step of the protocol, the players use their shared randomness to pick a uniformly random row vector $y \in \{0,1\}^k$ with Hamming weight at least $k - \ell$. The defining property (3.8) of ℓ ensures that with probability $2/3$ or higher, y is distinct from every row of the input matrix X. We will prove that conditioned on this event, the protocol is guaranteed to output the correct answer. We emphasize that this first step requires no communication. Indeed, it is our only departure from Grolmusz's protocol [14], where the corresponding vector was computed deterministically and counted toward the communication cost.

The rest of the analysis is identical to [14]. Specifically, by renumbering if necessary the players and the columns of X, we may assume that

$$y = (\underbrace{0, 0, \ldots, 0}_{j}, 1, 1, \ldots, 1)$$

for some $j \leq \ell$. Let n_i denote the number of rows of X of the form

$$(\underbrace{0, 0, \ldots, 0}_{i}, 1, 1, \ldots, 1).$$

In this notation, $n_j = 0$ by the assumption on y, and the objective of the protocol is to compute the quantity $n_0 \bmod 2$. For $i = 1, 2, \ldots, j$, the protocol has the ith party broadcast the sum $(n_{i-1} + n_i) \bmod 2$, which is known to him because he sees all but the ith coordinate of every row of X. These j broadcasts are sufficient to compute the answer since

$$\begin{aligned} n_0 &\equiv (n_0 + n_1) + (n_1 + n_2) + \cdots + (n_{j-1} + n_j) + n_j \pmod 2 \\ &\equiv (n_0 + n_1) + (n_1 + n_2) + \cdots + (n_{j-1} + n_j) \pmod 2. \end{aligned}$$

Observe that the described protocol is simultaneous, with communication cost bounded by (3.9). Its error probability can be reduced from $1/3$ to any constant $\epsilon > 0$ by running several copies of the protocol in parallel and using the majority answer.

We handle the case $k \in [\log n, \log 3n)$ in a manner analogous to [14], using the composed structure (2.6) of generalized inner product. Specifically, the players partition the input matrix horizontally into submatrices with at most $n/3$ rows each and run the above protocol on the resulting submatrices with a small constant error parameter, simultaneously and in parallel. The protocol output is the XOR of these answers. □

4 SET DISJOINTNESS

We now turn to the set disjointness problem $\mathrm{DISJ}_{n,k}$, proving matching lower and upper bounds on its communication complexity for $k \geq \log n$ players. Analogous to the previous section, our lower bound is a reduction to the case $k \leq \log n$ followed by an appeal to a known lower bound for that setting [25]. The treatment here is more technical than for inner product.

4.1 ℓ-discrepancy

For positive integers n and k, let $\mu_{n,k}$ denote the uniform probability distribution on matrices $X \in \{0,1\}^{n \times k}$ such that $X_{i,1} = X_{i,2} = \cdots = X_{i,k-1} = 1$ for precisely one row i. The following result [25, Theorem 4.2] bounds the multiparty discrepancy of the XOR of m independent instances of the set disjointness problem, each distributed according to the probability distribution just defined.

Theorem 4.1 (Sherstov). *For any integers $n_1, n_2, \ldots, n_m$,*

$$\operatorname{disc}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^{m} \mu_{n_i,k}\right) \leq \frac{(2^{k-1}-1)^m}{\sqrt{n_1 n_2 \cdots n_m}}.$$

For the purposes of this paper, we will slightly adapt Theorem 4.1 to obtain a discrepancy bound under a more symmetric distribution. Specifically, let

$$\sigma_{n,k} = \frac{1}{2}\sigma^0_{n,k} + \frac{1}{2}\sigma^1_{n,k}, \tag{4.1}$$

where $\sigma^0_{n,k}$ is the uniform probability distribution on $n \times k$ Boolean matrices without an all-ones row, and $\sigma^1_{n,k}$ is the uniform probability distribution on $n \times k$ Boolean matrices with precisely one all-ones row.

Theorem 4.2. *For any integers $n_1, n_2, \ldots, n_m$,*

$$\operatorname{disc}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^{m} \sigma_{n_i,k}\right) \leq \left(\frac{(\sqrt{2^k-1}+1)\sqrt{2^k-2}}{2}\right)^m \frac{1}{\sqrt{n_1 n_2 \cdots n_m}}.$$

Proof. Instead of analyzing the discrepancy with respect to the tensor product $\bigotimes \sigma_{n_i,k}$, we will define a different family of probability distributions $\tilde{\sigma}_{n_i,k}$ and bound the discrepancy with respect to $\bigotimes \tilde{\sigma}_{n_i,k}$ using Theorem 4.1. We will then prove that $\sigma_{n_i,k}$ and $\tilde{\sigma}_{n_i,k}$ are close in statistical distance for each i, and appeal to Proposition 2.7 to complete the proof.

Abbreviate $p = 1/(2^k - 1)$. For a given set $S \subseteq \{1, 2, \ldots, n\}$ and a probability distribution σ on matrices $X \in \{0,1\}^{n\times k}$, let $\sigma|_S$ stand for the probability distribution induced by σ after conditioning on the event that $(X_{i,1}, X_{i,2}, \ldots, X_{i,k}) = (1, 1, \ldots, 1, 0)$ if and only if $i \in S$. Observe that for a fixed set $S \subsetneq \{1, 2, \ldots, n\}$, the convex combination

$$\frac{\sigma^1_{n,k}|_S}{2} + \mathop{\mathrm{E}}_{\substack{S' \supset S \\ |S'|=|S|+1}} \frac{\sigma^0_{n,k}|_{S'}}{2} \tag{4.2}$$

is a probability distribution on matrices $X \in \{0,1\}^{n\times k}$ whereby

$$X|_S = \begin{bmatrix} 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}$$

and

$$X|_{\overline{S}} \sim \mu_{n-|S|,k}.$$

In other words, the rows with indices in S are fixed to non-1^k values and can therefore be disregarded from the point of view of set disjointness, whereas the remaining rows have joint probability distribution $\mu_{n-|S|,k}$. It follows that for any sets $S_1, S_2, \ldots, S_m$ with $S_i \subsetneq \{1, 2, \ldots, n_i\}$,

$$\begin{aligned}\operatorname{disc}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^{m} \left(\frac{\sigma^1_{n_i,k}|_{S_i}}{2} + \mathop{\mathrm{E}}_{\substack{S_i' \supset S_i \\ |S_i'|=|S_i|+1}} \frac{\sigma^0_{n_i,k}|_{S_i'}}{2}\right)\right) &= \operatorname{disc}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i-|S_i|,k}, \bigotimes_{i=1}^{m} \mu_{n_i-|S_i|,k}\right) \\ &\leq \prod_{i=1}^{m} \frac{2^{k-1}-1}{\sqrt{n_i - |S_i|}} \\ &= \prod_{i=1}^{m} \frac{1-p}{2p\sqrt{n_i - |S_i|}},\end{aligned}$$

where the second step uses Theorem 4.1. Proposition 2.6 now implies that for any integers $s_1, s_2, \ldots, s_m$ with $0 \leq s_i < n_i$,

$$\operatorname{disc}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^{m} \left(\mathop{\mathrm{E}}_{|S_i|=s_i} \frac{\sigma^1_{n_i,k}|_{S_i}}{2} + \mathop{\mathrm{E}}_{|S_i|=s_i+1} \frac{\sigma^0_{n_i,k}|_{S_i}}{2}\right)\right) \leq \prod_{i=1}^{m} \frac{1-p}{2p\sqrt{n_i - s_i}}. \tag{4.3}$$

We now define an approximation to the probability distribution $\sigma_{n,k}$, namely,

$$\tilde{\sigma}_{n,k} = \mathop{\mathrm{E}}_{s \sim B(n-1,p)} \left[\mathop{\mathrm{E}}_{|S|=s} \frac{\sigma^1_{n,k}|_S}{2} + \mathop{\mathrm{E}}_{|S|=s+1} \frac{\sigma^0_{n,k}|_S}{2} \right]. \tag{4.4}$$

For random integers $s_1, s_2, \ldots, s_m$ distributed independently according to $s_i \sim B(n_i - 1, p)$,

$$\begin{aligned}
&\operatorname{disc}\left(\bigoplus_{i=1}^m \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^m \tilde{\sigma}_{n_i,k} \right) \\
&\quad \le \mathop{\mathrm{E}}_{s_1,\ldots,s_m} \operatorname{disc}\left(\bigoplus_{i=1}^m \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^m \left(\mathop{\mathrm{E}}_{|S_i|=s_i} \frac{\sigma^1_{n_i,k}|_{S_i}}{2} + \mathop{\mathrm{E}}_{|S_i|=s_i+1} \frac{\sigma^0_{n_i,k}|_{S_i}}{2} \right) \right) \\
&\qquad\qquad \text{by Proposition 2.6} \\
&\quad \le \mathop{\mathrm{E}}_{s_1,\ldots,s_m} \prod_{i=1}^m \frac{1-p}{2p\sqrt{n_i - s_i}} \qquad \text{by (4.3)} \\
&\quad = \prod_{i=1}^m \mathop{\mathrm{E}}_{s_i} \frac{1-p}{2p\sqrt{n_i - s_i}} \qquad \text{by independence} \\
&\quad \le \prod_{i=1}^m \frac{\sqrt{1-p}}{2p\sqrt{n_i}} \qquad \text{by Fact 2.1.}
\end{aligned}$$

Of course, this calculation shows more generally that

$$\operatorname{disc}\left(\bigoplus_{i \in A} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i \in A} \tilde{\sigma}_{n_i,k} \right) \le \prod_{i \in A} \frac{\sqrt{1-p}}{2p\sqrt{n_i}} \tag{4.5}$$

for any set $A \subseteq \{1, 2, \ldots, m\}$. This completes the first part of the program.

We proceed to bound the statistical distance between $\sigma_{n,k}$ and $\tilde{\sigma}_{n,k}$. To start with,

$$\begin{aligned}
\sigma^0_{n,k} &= \sum_{s=0}^n \binom{n}{s} p^s (1-p)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^0_{n,k}|_S, \\
\sigma^1_{n,k} &= \sum_{s=0}^{n-1} \binom{n-1}{s} p^s (1-p)^{n-1-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^1_{n,k}|_S, \\
\tilde{\sigma}_{n,k} &= \frac{1}{2} \sum_{s=1}^n \binom{n-1}{s-1} p^{s-1} (1-p)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^0_{n,k}|_S + \frac{1}{2} \sum_{s=0}^{n-1} \binom{n-1}{s} p^s (1-p)^{n-1-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^1_{n,k}|_S,
\end{aligned}$$

where the first two equations hold by definition, and the third is a restatement of (4.4). Then

$$\begin{aligned}
\|\sigma_{n,k} - \tilde{\sigma}_{n,k}\|_1 &= \left\| \frac{\sigma^0_{n,k} + \sigma^1_{n,k}}{2} - \tilde{\sigma}_{n,k} \right\|_1 \\
&= \frac{1}{2} \left\| \sum_{s=0}^{n} \binom{n}{s} p^s (1-p)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^0_{n,k}|_S - \sum_{s=1}^{n} \binom{n-1}{s-1} p^{s-1} (1-p)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^0_{n,k}|_S \right\|_1 \\
&= \frac{1}{2} \sum_{s=1}^{n} \left| \binom{n}{s} p^s (1-p)^{n-s} - \binom{n-1}{s-1} p^{s-1} (1-p)^{n-s} \right| + \frac{(1-p)^n}{2} \\
&= \frac{1}{2pn} \sum_{s=0}^{n} \binom{n}{s} p^s (1-p)^{n-s} |s - pn| \\
&= \frac{1}{2pn} \mathop{\mathrm{E}}_{s \sim B(n,p)} |s - pn| \\
&\leq \frac{1}{2} \sqrt{\frac{1-p}{pn}},
\end{aligned} \tag{4.6}$$

where the first and last steps use (4.1) and Fact 2.1, respectively. This completes the second part of the proof program.

It remains to put together the above ingredients and appeal to the continuity of discrepancy:

$$\begin{aligned}
\operatorname{disc}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^{m} \sigma_{n_i,k} \right) &\leq \sum_{A \subseteq \{1,2,\ldots,m\}} \operatorname{disc}\left(\bigoplus_{i \in A} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i \in A} \tilde{\sigma}_{n_i,k} \right) \prod_{i \notin A} \|\sigma_{n_i,k} - \tilde{\sigma}_{n_i,k}\|_1 \\
&\leq \sum_{A \subseteq \{1,2,\ldots,m\}} \prod_{i \in A} \frac{\sqrt{1-p}}{2p\sqrt{n_i}} \cdot \prod_{i \notin A} \frac{\sqrt{1-p}}{2\sqrt{pn_i}} \\
&= \prod_{i=1}^{m} \left(\frac{\sqrt{1-p}}{2p\sqrt{n_i}} + \frac{\sqrt{1-p}}{2\sqrt{pn_i}} \right) \\
&= \prod_{i=1}^{m} \frac{(\sqrt{2^k - 1} + 1)\sqrt{2^k - 2}}{2\sqrt{n_i}},
\end{aligned}$$

where the first inequality uses Proposition 2.7, and the second inequality follows from the upper bounds in (4.5) and (4.6). □

We now refine the previous theorem for cylinder intersections that contain significantly fewer cylinders than there are players. Formally, define $\sigma^0_{n,k,\ell}$ to be the probability distribution on $n \times k$ Boolean matrices whereby the rows are chosen independently and uniformly at random from the set $\{u \in \{0,1\}^k : k - \ell \leq |u| \leq k - 1\}$. Define $\sigma^1_{n,k,\ell}$ to be the probability distribution on $n \times k$ Boolean matrices whereby a randomly chosen row is set to 1^k and the remaining rows are chosen independently and uniformly at random from the set $\{u \in \{0,1\}^k : k - \ell \leq |u| \leq k - 1\}$. We will analyze the ℓ-discrepancy of set disjointness with respect to the probability distribution

$$\sigma_{n,k,\ell} = \frac{1}{2}\sigma^0_{n,k,\ell} + \frac{1}{2}\sigma^1_{n,k,\ell}. \tag{4.7}$$

This setup is indeed a generalization of the case $\ell = k$ dealt with above. Specifically,

$$\begin{aligned}\sigma^0_{n,k,k} &= \sigma^0_{n,k},\\ \sigma^1_{n,k,k} &= \sigma^1_{n,k},\\ \sigma_{n,k,k} &= \sigma_{n,k}.\end{aligned}$$

THEOREM 4.3. *For any integers $n_1, n_2, \ldots, n_m \geq 1$ and $k \geq \ell \geq 1$,*

$$\operatorname{disc}_\ell \left(\bigoplus_{i=1}^m \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^m \sigma_{n_i,k,\ell} \right) \leq (2^\ell - 1)^{m/2} \left(\binom{k}{1} + \binom{k}{2} + \cdots + \binom{k}{\ell} \right)^{m/2} \frac{1}{\sqrt{n_1 n_2 \cdots n_m}}.$$

In the setting $\ell \ll k$ of interest to us, the new bound is considerably stronger than the bound of Theorem 4.2. The proof is essentially a reprise of Theorem 4.2. Indeed, we could have combined the two theorems for a more economical presentation. Treating them separately, as we do in this paper, has the advantage of simplifying the notation and illustrating the proof idea in a simpler setting first.

PROOF OF THEOREM 4.3. We will closely follow the proof of the previous theorem. Specifically, instead of analyzing the discrepancy with respect to the tensor product $\bigotimes \sigma_{n_i,k,\ell}$, we will define a different family of probability distributions $\tilde{\sigma}_{n_i,k,\ell}$ and bound the discrepancy with respect to $\bigotimes \tilde{\sigma}_{n_i,k,\ell}$. We will then prove that $\sigma_{n_i,k,\ell}$ and $\tilde{\sigma}_{n_i,k,\ell}$ are close in statistical distance for each i, and appeal to Proposition 2.7 to complete the proof.

Since the communication problem $\mathrm{DISJ}_{n,k}$ and the probability distribution $\sigma_{n,k,\ell}$ are both symmetric with respect to the k players, we have

$$\operatorname{disc}_\ell \left(\bigoplus_{i=1}^m \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^m \sigma_{n_i,k,\ell} \right) = \operatorname{disc}_{\{1,2,\ldots,\ell\}} \left(\bigoplus_{i=1}^m \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^m \sigma_{n_i,k,\ell} \right). \tag{4.8}$$

For a given set $S \subseteq \{1, 2, \ldots, n\}$ and a probability distribution σ on matrices $X \in \{0,1\}^{n \times k}$, let $\sigma|_S$ stand for the probability distribution induced by σ after conditioning on the event that $(X_{i,\ell+1}, X_{i,\ell+2}, \ldots, X_{i,k}) = (1, 1, \ldots, 1)$ if and only if $i \in S$. Observe that for a fixed nonempty set $S \subseteq \{1, 2, \ldots, n\}$, the convex combination

$$\frac{\sigma^1_{n,k,\ell}|_S}{2} + \frac{\sigma^0_{n,k,\ell}|_S}{2} \tag{4.9}$$

is a probability distribution on matrices $X \in \{0,1\}^{n \times k}$ whereby $X|_S$ and $X|_{\overline{S}}$ are distributed independently such that

$$X|_{S,\{1,2,\ldots,\ell\}} \sim \sigma_{|S|,\ell}, \tag{4.10}$$

$$X|_{S,\{\ell+1,\ell+2,\ldots,k\}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}, \tag{4.11}$$

and $X|_{\overline{S}}$ does not have an all-ones row. In particular, the rows in $X|_{\overline{S}}$ do not affect the value of the function. Since the conditional probability distribution of $X|_S$ given $X|_{\overline{S}}$ is always (4.10)–(4.11), we

arrive at the following conclusion: for any nonempty sets $S_1, S_2, \ldots, S_m$ with $S_i \subseteq \{1, 2, \ldots, n_i\}$,

$$\operatorname{disc}_{\{1,2,\ldots,\ell\}}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^{m}\left(\frac{\sigma^1_{n_i,k,\ell}|_{S_i}}{2} + \frac{\sigma^0_{n_i,k,\ell}|_{S_i}}{2}\right)\right) \leq \operatorname{disc}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{|S_i|,\ell}, \bigotimes_{i=1}^{m} \sigma_{|S_i|,\ell}\right) \leq \prod_{i=1}^{m} \frac{(\sqrt{2^\ell - 1} + 1)\sqrt{2^\ell - 2}}{2\sqrt{|S_i|}},$$

where the second step holds by Theorem 4.2. Proposition 2.6 now implies that for any integers $s_1, s_2, \ldots, s_m$ with $1 \leq s_i \leq n_i$,

$$\operatorname{disc}_{\{1,2,\ldots,\ell\}}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^{m}\left(\mathop{\mathrm{E}}_{|S_i|=s_i} \frac{\sigma^1_{n_i,k,\ell}|_{S_i}}{2} + \mathop{\mathrm{E}}_{|S_i|=s_i} \frac{\sigma^0_{n_i,k,\ell}|_{S_i}}{2}\right)\right) \leq \prod_{i=1}^{m} \frac{(\sqrt{2^\ell - 1} + 1)\sqrt{2^\ell - 2}}{2\sqrt{s_i}}. \tag{4.12}$$

We now define the promised approximation $\tilde{\sigma}_{n,k,\ell}$ to the probability distribution $\sigma_{n,k,\ell}$, namely,

$$\tilde{\sigma}_{n,k,\ell} = \mathop{\mathrm{E}}_{s\sim B(n-1,q)} \mathop{\mathrm{E}}_{|S|=s+1} \frac{\sigma^0_{n,k,\ell}|_S + \sigma^1_{n,k,\ell}|_S}{2} \tag{4.13}$$

where

$$q = \frac{2^\ell - 1}{\binom{k}{1} + \binom{k}{2} + \cdots + \binom{k}{\ell}}.$$

Then for random integers $s_1, s_2, \ldots, s_m$ distributed independently according to $s_i \sim B(n_i - 1, q)$, we have

$$\begin{aligned}
&\operatorname{disc}_{\{1,2,\ldots,\ell\}}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^{m} \tilde{\sigma}_{n_i,k,\ell}\right) \\
&\quad\leq \mathop{\mathrm{E}}_{s_1,\ldots,s_m} \operatorname{disc}_{\{1,2,\ldots,\ell\}}\left(\bigoplus_{i=1}^{m} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^{m}\left(\mathop{\mathrm{E}}_{|S_i|=s_i+1} \frac{\sigma^0_{n_i,k,\ell}|_{S_i} + \sigma^1_{n_i,k,\ell}|_{S_i}}{2}\right)\right) && \text{by Proposition 2.6} \\
&\quad\leq \mathop{\mathrm{E}}_{s_1,\ldots,s_m} \prod_{i=1}^{m} \frac{(\sqrt{2^\ell - 1} + 1)\sqrt{2^\ell - 2}}{2\sqrt{s_i + 1}} && \text{by (4.12)} \\
&\quad= \prod_{i=1}^{m} \mathop{\mathrm{E}}_{s_i} \frac{(\sqrt{2^\ell - 1} + 1)\sqrt{2^\ell - 2}}{2\sqrt{s_i + 1}} && \text{by independence} \\
&\quad\leq \prod_{i=1}^{m} \frac{(\sqrt{2^\ell - 1} + 1)\sqrt{2^\ell - 2}}{2\sqrt{qn_i}} && \text{by Fact 2.1.}
\end{aligned}$$

Of course, this calculation shows more generally that

$$\operatorname{disc}_{\{1,2,\ldots,\ell\}}\left(\bigoplus_{i\in A} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i\in A} \tilde{\sigma}_{n_i,k,\ell}\right) \leq \prod_{i\in A} \frac{(\sqrt{2^\ell - 1} + 1)\sqrt{2^\ell - 2}}{2\sqrt{qn_i}} \tag{4.14}$$

for any set $A \subseteq \{1, 2, \ldots, m\}$. This completes the first part of the program.

In this second part of the proof, we bound the statistical distance between $\sigma_{n,k,\ell}$ and $\tilde{\sigma}_{n,k,\ell}$. We have

$$\begin{aligned}
\sigma^0_{n,k,\ell} &= \sum_{s=0}^{n} \binom{n}{s} q^s (1-q)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^0_{n,k,\ell}|_S, \\
\sigma^1_{n,k,\ell} &= \sum_{s=1}^{n} \binom{n-1}{s-1} q^{s-1} (1-q)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^1_{n,k,\ell}|_S, \\
\tilde{\sigma}_{n,k,\ell} &= \frac{1}{2} \sum_{s=1}^{n} \binom{n-1}{s-1} q^{s-1} (1-q)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^0_{n,k,\ell}|_S + \frac{1}{2} \sum_{s=1}^{n} \binom{n-1}{s-1} q^{s-1} (1-q)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^1_{n,k,\ell}|_S,
\end{aligned}$$

where the first two equations hold by definition, and the third is a restatement of (4.13). Then

$$\begin{aligned}
\|\sigma_{n,k,\ell} - \tilde{\sigma}_{n,k,\ell}\|_1 &= \left\| \frac{\sigma^1_{n,k,\ell} + \sigma^0_{n,k,\ell}}{2} - \tilde{\sigma}_{n,k,\ell} \right\|_1 \\
&= \frac{1}{2} \left\| \sum_{s=0}^{n} \binom{n}{s} q^s (1-q)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^0_{n,k,\ell}|_S \right. \\
&\qquad \left. - \sum_{s=1}^{n} \binom{n-1}{s-1} q^{s-1} (1-q)^{n-s} \mathop{\mathrm{E}}_{|S|=s} \sigma^0_{n,k,\ell}|_S \right\|_1 \\
&= \frac{1}{2} \sum_{s=1}^{n} \left| \binom{n}{s} q^s (1-q)^{n-s} - \binom{n-1}{s-1} q^{s-1} (1-q)^{n-s} \right| + \frac{(1-q)^n}{2} \\
&= \frac{1}{2qn} \sum_{s=0}^{n} \binom{n}{s} q^s (1-q)^{n-s} |s - qn| \\
&= \frac{1}{2qn} \mathop{\mathrm{E}}_{s \sim B(n,q)} |s - qn| \\
&\leq \frac{1}{2} \sqrt{\frac{1-q}{qn}},
\end{aligned} \tag{4.15}$$

where the first and last steps use (4.7) and Fact 2.1, respectively. This completes the second part of the proof program.

It remains to put together the above ingredients and appeal to the continuity of discrepancy:

$$\begin{aligned}
&\operatorname{disc}_\ell\left(\bigoplus_{i=1}^m \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^m \sigma_{n_i,k,\ell}\right) \\
&\quad= \operatorname{disc}_{\{1,2,\ldots,\ell\}}\left(\bigoplus_{i=1}^m \mathrm{DISJ}_{n_i,k}, \bigotimes_{i=1}^m \sigma_{n_i,k,\ell}\right) \\
&\quad\le \sum_{A\subseteq\{1,2,\ldots,m\}} \operatorname{disc}_{\{1,2,\ldots,\ell\}}\left(\bigoplus_{i\in A} \mathrm{DISJ}_{n_i,k}, \bigotimes_{i\in A} \tilde{\sigma}_{n_i,k,\ell}\right) \prod_{i\notin A} \|\sigma_{n_i,k,\ell} - \tilde{\sigma}_{n_i,k,\ell}\|_1 \\
&\quad\le \sum_{A\subseteq\{1,2,\ldots,m\}} \prod_{i\in A} \frac{(\sqrt{2^\ell-1}+1)\sqrt{2^\ell-2}}{2\sqrt{qn_i}} \cdot \prod_{i\notin A} \frac{\sqrt{1-q}}{2\sqrt{qn_i}} \\
&\quad= \prod_{i=1}^m \frac{(\sqrt{2^\ell-1}+1)\sqrt{2^\ell-2}+\sqrt{1-q}}{2\sqrt{qn_i}} \\
&\quad\le \prod_{i=1}^m \frac{(2^\ell-1)}{\sqrt{qn_i}},
\end{aligned}$$

where the first step is valid by (4.8), the next step uses Proposition 2.7, and the third step follows from the upper bounds in (4.14) and (4.15). □

4.2 Lower bound

We are now in a position to prove the claimed lower bound on the communication complexity of set disjointness. Following [25], we find it helpful to work in the more general setting of composed communication problems $f \circ G$, where f is any Boolean function with high approximate degree and G is an instance of set disjointness on a small number of variables. This approach is motivated by the composed structure (2.7)–(2.8) of the set disjointness problem.

The following communication lower bound was obtained in [25, Theorem 5.1].

Theorem 4.4 (Sherstov). *Let f be a (possibly partial) Boolean function on $\{0,1\}^n$. Consider the k-party communication problem $F = f \circ \mathrm{UDISJ}_{r,k}$. Then for $\epsilon, \delta \ge 0$,*

$$2^{R_\epsilon(F)} \ge (\delta-\epsilon)\left(\frac{\deg_\delta(f)\sqrt{r}}{2^k \mathrm{e} n}\right)^{\deg_\delta(f)}.$$

Using the new discrepancy upper bound in this paper, we are able to obtain the following improvement:

Theorem 4.5. *Let f be a (possibly partial) Boolean function on $\{0,1\}^n$. Consider the k-party communication problem $F = f \circ \mathrm{UDISJ}_{r,k}$. Then for $\epsilon, \delta \ge 0$,*

$$2^{R_\epsilon(F)} \ge (\delta-\epsilon)\left(\frac{\deg_\delta(f)\sqrt{r}}{2\binom{k}{\le R_\epsilon(F)}\mathrm{e} n}\right)^{\deg_\delta(f)}.$$

The lower bound of Theorem 4.5 is always at least as strong as that of Theorem 4.4, up to a small constant factor in the denominator. The improvement becomes significant in the setting $k \gg R_\epsilon(F)$ of interest to us, where the number of players far exceeds the communication requirements of the problem.

Proof of Theorem 4.5. The proof is virtually identical to that in [25], the only difference being the use of the improved discrepancy upper bound in this paper (Theorem 4.3). The idea behind the proof is to show that any low-cost communication protocol for F can be converted into a low-degree polynomial approximating f in the infinity norm. Details follow.

Put $\ell = \min\{R_\epsilon(F), k\}$ and recall the probability distribution

$$\sigma_{r,k,\ell} = \frac{1}{2}\sigma^0_{r,k,\ell} + \frac{1}{2}\sigma^1_{r,k,\ell}, \tag{4.16}$$

where $\sigma^0_{r,k,\ell}$ and $\sigma^1_{r,k,\ell}$ (both defined in Section 4.1) are probability distributions on $\mathrm{UDISJ}^{-1}_{r,k}(1)$ and $\mathrm{UDISJ}^{-1}_{r,k}(0)$, respectively. Consider the following averaging operator L, which linearly sends real functions χ on $(\{0,1\}^{r\times k})^n$ to real functions on $\{0,1\}^n$ according to

$$(L\chi)(z) = \mathop{\mathrm{E}}_{X_1\sim\sigma^{z_1}_{r,k,\ell}} \cdots \mathop{\mathrm{E}}_{X_n\sim\sigma^{z_n}_{r,k,\ell}} \chi(X_1,\ldots,X_n) \qquad (z\in\{0,1\}^n).$$

When χ is an ℓ-cylinder intersection,

$$\begin{aligned}
|\widehat{L\chi}(S)| &= \left|\mathop{\mathrm{E}}_{z\in\{0,1\}^n} \mathop{\mathrm{E}}_{X_1\sim\sigma^{z_1}_{r,k,\ell}} \cdots \mathop{\mathrm{E}}_{X_n\sim\sigma^{z_n}_{r,k,\ell}} \left[\chi(X_1,\ldots,X_n)\prod_{i\in S}(-1)^{z_i}\right]\right| \\
&= \left|\mathop{\mathrm{E}}_{X_1,\ldots,X_n\sim\sigma_{r,k,\ell}} \left[\chi(X_1,\ldots,X_n)\prod_{i\in S}(-1)^{1-\mathrm{UDISJ}_{r,k}(X_i)}\right]\right| \\
&= \left|\mathop{\mathrm{E}}_{X_1,\ldots,X_n\sim\sigma_{r,k,\ell}} \left[\chi(X_1,\ldots,X_n)\prod_{i\in S}(-1)^{\mathrm{UDISJ}_{r,k}(X_i)}\right]\right| \\
&\le \mathrm{disc}_\ell\left(\mathrm{DISJ}^{\oplus|S|}_{r,k}, (\sigma_{r,k,\ell})^{\otimes|S|}\right) \\
&\le \left(\frac{2^\ell-1}{r}\cdot\left(\binom{k}{1}+\binom{k}{2}+\cdots+\binom{k}{\ell}\right)\right)^{|S|/2} \\
&\le \left(\frac{1}{\sqrt{r}}\cdot\binom{k}{\le\ell}\right)^{|S|},
\end{aligned} \tag{4.17}$$

where the second and fifth steps use (4.16) and Theorem 4.3, respectively.

Now, fix a randomized communication protocol for F with error ϵ and cost $R_\epsilon(F)$. Approximate F as in Corollary 2.4 by a linear combination of ℓ-cylinder intersections $\Pi = \sum_\chi a_\chi\chi$, where $\sum_\chi |a_\chi| \le 2^{R_\epsilon(F)}$. We claim that $L\Pi$ is approximable by a low-degree polynomial. Indeed, let d be a positive integer to be chosen later. Discarding the Fourier coefficients of $L\Pi$ of order d and higher

gives

$$\begin{aligned} E(L\Pi, d-1) &\le \min\left\{1,\ \sum_{\chi} |a_\chi| \sum_{|S|\ge d} |\widehat{L\chi}(S)|\right\} \\ &\le \min\left\{1,\ 2^{R_\epsilon(F)} \sum_{i=d}^{n} \binom{n}{i} \left(\frac{1}{\sqrt{r}} \cdot \binom{k}{\le \ell}\right)^i\right\} \\ &\le \min\left\{1,\ 2^{R_\epsilon(F)} \sum_{i=d}^{n} \left(\frac{\mathrm{e}n}{d\sqrt{r}} \cdot \binom{k}{\le \ell}\right)^i\right\} \\ &\le 2^{R_\epsilon(F)} \left(\frac{2\mathrm{e}n}{d\sqrt{r}} \cdot \binom{k}{\le \ell}\right)^d, \end{aligned} \tag{4.18}$$

where the second and third steps use (4.17) and (2.1), respectively. On the other hand, recall from Corollary 2.4 that Π approximates F in the sense that $\|\Pi\|_\infty \le 1$ and $|F - \Pi| \le \epsilon$ on the domain of F. It follows that $\|L\Pi\|_\infty \le 1$ and $|f - L\Pi| \le \epsilon$ on the domain of f, whence

$$E(f, d-1) \le \epsilon + E(L\Pi, d-1).$$

Substituting the estimate from (4.18),

$$E(f, d-1) \le \epsilon + 2^{R_\epsilon(F)} \left(\frac{2\mathrm{e}n}{d\sqrt{r}} \cdot \binom{k}{\le \ell}\right)^d.$$

For $d = \deg_\delta(f)$, the left-hand side of this inequality by definition exceeds δ, completing the proof. □

We now specialize the previous theorem to set disjointness.

Theorem 4.6. *Abbreviate* $R = R_{1/4}(\mathrm{UDISJ}_{n,k})$. *Then*

$$\binom{k}{\le R}^2 R^4 \ge \Omega(n). \tag{4.19}$$

In particular,

$$R_{1/3}(\mathrm{DISJ}_{n,k}) \ge R_{1/3}(\mathrm{UDISJ}_{n,k}) = \Omega\left(\frac{\log n}{\log\left\lceil 1 + \frac{k}{\log n}\right\rceil} + 1\right). \tag{4.20}$$

Proof. Observe that $R_{1/4}(\mathrm{UDISJ}_{n,k})$ is monotonically nondecreasing in n. Now for all $1 \le r \le n$, we have $R \ge R_{1/4}(\widetilde{\mathrm{AND}}_{\lfloor n/r \rfloor} \circ \mathrm{UDISJ}_{r,k})$ by (2.8) and therefore

$$2^R \ge \left(\frac{1}{3} - \frac{1}{4}\right) \left(\frac{\deg_{1/3}(\widetilde{\mathrm{AND}}_{\lfloor n/r \rfloor})\sqrt{r}}{2\mathrm{e}\lfloor n/r \rfloor \binom{k}{\le R}}\right)^{\deg_{1/3}(\widetilde{\mathrm{AND}}_{\lfloor n/r \rfloor})}$$

by Theorem 4.5. This in turn simplifies to

$$2^R \ge \frac{1}{12} \left(\frac{cr}{\sqrt{n}\binom{k}{\le R}}\right)^{c\sqrt{n/r}} \tag{4.21}$$

for some absolute constant $c > 0$, by Theorem 2.2. There are two cases to examine. If $\binom{k}{\le R} \ge \frac{c}{2}\sqrt{n}$, then (4.19) is trivially true. Otherwise, letting $r = \lceil \frac{2}{c}\sqrt{n}\binom{k}{\le R} \rceil$ in (4.21) forces (4.19).

It remains to explain how the newly obtained relation (4.19) implies the communication lower bounds in the theorem statement. By (2.1),

$$\binom{k}{\leq R} \leq \mathrm{e}^R \left\lceil \frac{k}{R} \right\rceil^R.$$

Substituting this estimate in (4.19) gives $\mathrm{e}^{2R}\lceil k/R\rceil^{2R}R^4 \geq \Omega(n)$, whence (4.20). □

Remark 4.7. Theorem 4.5 is sufficiently general to imply our lower bound for generalized inner product as well (Theorem 3.4). However, considering the effort required to prove Theorem 4.5 itself, the treatment of $\mathrm{GIP}_{n,k}$ in Section 3.1 appears to be more direct.

4.3 Upper bound

To prove a matching upper bound on the communication complexity of set disjointness for $k \geq \log n$ players, we reduce this problem to inner product and appeal to our previously established Theorem 3.5.

Theorem 4.8. *For any $k \geq \log n$ and any constant $\epsilon > 0$,*

$$R_\epsilon(\mathrm{DISJ}_{n,k}) = O\left(\frac{\log n}{\log\left\lceil 1 + \frac{k}{\log n}\right\rceil} + 1\right).$$

Moreover, this upper bound is achieved by a simultaneous protocol.

Proof. Recall that for any string $y \in \{0,1\}^n$,

$$\mathop{\mathrm{P}}_{S\subseteq\{1,2,\ldots,n\}}\left[\bigoplus_{i\in S} y_i = 0\right] = \begin{cases} 1 & \text{if } y = 0^n, \\ 1/2 & \text{otherwise.}\end{cases}$$

This gives the following well-known relation between generalized inner product and set disjointness:

$$\mathop{\mathrm{P}}_{S\subseteq\{1,2,\ldots,n\}}\left[\mathrm{GIP}_{|S|,k}(X|_S) = 0\right] = \begin{cases} 1 & \text{if } \mathrm{DISJ}_{n,k}(X) = 1, \\ 1/2 & \text{otherwise.}\end{cases}$$

Thus, the players can solve an instance $X \in \{0,1\}^{n\times k}$ of set disjointness by estimating

$$\mathop{\mathrm{P}}_{S}[\mathrm{GIP}_{|S|,k}(X|_S) = 0].$$

This can be done by running the protocol of Theorem 3.5 with error parameter $1/4$ a constant number of times, simultaneously and in parallel, each time on a uniformly random subset of the rows of X. □

5 COMMUNICATION BOUNDS INDEPENDENT OF k

In this final section, we study the communication problem $F_{n,k} \colon (\{0,1\}^n)^k \to \{0,1\}$ given by

$$F_{n,k}(X) = \mathrm{MOD}_3\left(\bigoplus_{j=1}^k X_{1,j}, \ldots, \bigoplus_{j=1}^k X_{n,j}\right). \tag{5.1}$$

Ada et al. [1] proved that this function has randomized communication complexity $R_{1/3}(F_{n,k}) = \Omega(n/4^k)$. Here, we derive the incomparable lower bound $R_{1/3}(F_{n,k}) \geq \frac{1}{3}\log n - \frac{1}{3}$ for all n and k, which shows that $F_{n,k}$ requires nontrivial communication regardless of the number of players k. We also prove that for every $k \geq \log n$, our lower bound is tight up to a multiplicative constant.

5.1 Lower bound

In what follows, we use the shorthand $e(t) = \exp(2\pi \mathrm{i} t)$, where i is the imaginary unit. Our proof requires the following correlation bound for cylinder intersections, which is implied by the more general work of Ada et al. [1].

Lemma 5.1 (cf. Ada et al.). *For every cylinder intersection $\chi\colon (\{0,1\}^n)^k \to \{0,1\}$,*

$$\left| \mathop{\mathrm{E}}_{X \in \{0,1\}^{n \times k}} e\left(\frac{1}{3} \sum_{i=1}^{n} \bigoplus_{j=1}^{k} X_{i,j} \right) \chi(X) \right| < \exp\left(-\frac{n}{4^k}\right).$$

For the reader's convenience, a short and self-contained proof of Lemma 5.1 as stated above is available in Appendix A. We now sharpen this result for ℓ-cylinder intersections with $\ell \le k$.

Lemma 5.2. *For every ℓ-cylinder intersection $\chi\colon (\{0,1\}^n)^k \to \{0,1\}$,*

$$\left| \mathop{\mathrm{E}}_{X \in \{0,1\}^{n \times k}} e\left(\frac{1}{3} \sum_{i=1}^{n} \bigoplus_{j=1}^{k} X_{i,j} \right) \chi(X) \right| < \exp\left(-\frac{n}{4^\ell}\right). \tag{5.2}$$

Proof. By symmetry, we may assume that χ is a $\{1, 2, \ldots, \ell\}$-cylinder intersection. In what follows, we let X stand for a uniformly random matrix in $\{0,1\}^{n \times k}$. Define auxiliary random variables X', X'' by

$$X' = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,\ell-1} & X_{1,\ell} \oplus \cdots \oplus X_{1,k} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,\ell-1} & X_{2,\ell} \oplus \cdots \oplus X_{2,k} \\ \vdots & \vdots & & \vdots & \vdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,\ell-1} & X_{n,\ell} \oplus \cdots \oplus X_{n,k} \end{bmatrix},$$

$$X'' = \begin{bmatrix} X_{1,\ell+1} & X_{1,\ell+2} & \cdots & X_{1,k} \\ X_{2,\ell+1} & X_{2,\ell+2} & \cdots & X_{2,k} \\ \vdots & \vdots & & \vdots \\ X_{n,\ell+1} & X_{n,\ell+2} & \cdots & X_{n,k} \end{bmatrix}.$$

In this notation,

$$\mathrm{E}\left[e\left(\frac{1}{3} \sum_{i=1}^{n} \bigoplus_{j=1}^{k} X_{i,j} \right) \chi(X) \,\middle|\, X'' \right] = \mathrm{E}\left[e\left(\frac{1}{3} \sum_{i=1}^{n} \bigoplus_{j=1}^{\ell} X'_{i,j} \right) \chi(X) \,\middle|\, X'' \right]. \tag{5.3}$$

Setting X'' to any given value makes $\chi(X)$ a cylinder intersection in terms of X'. Moreover, the conditional probability distribution of X' given X'' is uniform on $\{0,1\}^{n \times \ell}$. In view of Fact 5.1, we conclude that the expectation on the right-hand side of (5.3) is smaller in absolute value than $\exp(-n/4^\ell)$. Averaging over X'' now gives (5.2). □

We are now in a position to prove our claimed lower bound on the communication complexity of $F_{n,k}$.

Theorem 5.3. *Let $F_{n,k}\colon (\{0,1\}^n)^k \to \{0,1\}$ be given by (5.1). Then*

$$\mathrm{disc}_\ell(F_{n,k}, \nu) \le 2 \exp\left(-\frac{n}{4^\ell}\right), \tag{5.4}$$

where ν is the probability distribution under which the weight of any point of $F_{n,k}^{-1}(1)$ is double the weight of any point of $F_{n,k}^{-1}(0)$. In particular,

$$R_{1/3}(F_{n,k}) \ge \frac{1}{3} \log n - \frac{1}{3}. \tag{5.5}$$

Proof. Observe that

$$e\left(\frac{t}{3}\right) + \overline{e\left(\frac{t}{3}\right)} = \begin{cases} 2 & \text{if } t \equiv 0 \pmod 3, \\ -1 & \text{if } t \equiv \pm 1 \pmod 3, \end{cases}$$

where the bar denotes complex conjugation. As a result,

$$e\left(\frac{1}{3}\sum_{i=1}^{n}\bigoplus_{j=1}^{k} X_{i,j}\right) + \overline{e\left(\frac{1}{3}\sum_{i=1}^{n}\bigoplus_{j=1}^{k} X_{i,j}\right)} = c \cdot 2^{nk}(-1)^{F_{n,k}(X)+1}\nu(X)$$

for some normalizing factor $c \geq 1$ and all X. We conclude that for every ℓ-cylinder intersection χ,

$$\begin{aligned} \left|\mathop{\mathrm{E}}_{X\sim\nu}(-1)^{F_{n,k}(X)}\chi(X)\right| &= \frac{1}{c}\left|\mathop{\mathrm{E}}_{X\in\{0,1\}^{n\times k}}\left[e\left(\frac{1}{3}\sum_{i=1}^{n}\bigoplus_{j=1}^{k} X_{i,j}\right)\chi(X) + \overline{e\left(\frac{1}{3}\sum_{i=1}^{n}\bigoplus_{j=1}^{k} X_{i,j}\right)\chi(X)}\right]\right| \\ &\leq 2\left|\mathop{\mathrm{E}}_{X\in\{0,1\}^{n\times k}}\left[e\left(\frac{1}{3}\sum_{i=1}^{n}\bigoplus_{j=1}^{k} X_{i,j}\right)\chi(X)\right]\right| \\ &\leq 2\exp\left(-\frac{n}{4^{\ell}}\right), \end{aligned}$$

where the final step uses Lemma 5.2. This establishes (5.4). Now Theorem 2.5 yields

$$2^{R_{1/3}(F_{n,k})} \geq \frac{1}{6}\exp\left(\frac{n}{4^{R_{1/3}(F_{n,k})}}\right),$$

which implies (5.5) by elementary calculus. □

5.2 Upper bound

We now show that our lower bound on the communication complexity of $F_{n,k}$ is tight up to a constant factor for all $k \geq \log n$. The idea is to alter $F_{n,k}$ in a random fashion on a small portion of the domain so as to make it representable by a polynomial of degree $k-1$, which the players can then directly evaluate. This technique was previously used in [1, 10].

Theorem 5.4. *Let $F_{n,k} : (\{0,1\}^n)^k \to \{0,1\}$ be given by (5.1). Then*

$$R_{1/3}(F_{n,k}) \leq c\log n + c$$

for some absolute constant $c > 0$ and all $n \geq 1$ and $k \geq \log n$. Moreover, this upper bound is achieved by a simultaneous protocol.

Proof. For $u \in \{0,1\}^k$, consider the multivariate polynomial $p_u \in \mathbb{Z}_3[x_1, x_2, \ldots, x_k]$ given by

$$p_u(x_1, x_2, \ldots, x_k) = \prod_{i=1}^{k}(x_i + 1) - \prod_{i=1}^{k}(x_i + u_i - 1) - 1.$$

Then

$$\deg p_u \leq k - 1, \tag{5.6}$$

$$p_u(x_1, x_2, \ldots, x_k) = x_1 \oplus x_2 \oplus \cdots \oplus x_k, \qquad x \in \{0,1\}^k \setminus \{u\}. \tag{5.7}$$

Our protocol for $F_{n,k}$ is as follows. On input $X \in \{0,1\}^{n\times k}$, the players pick a random point $u \in \{0,1\}^k$ using shared randomness, and consider the following polynomial $F_{n,k,u} \in \mathbb{Z}_3[X_{1,1}, \ldots, X_{n,k}]$:

$$F_{n,k,u}(X) = \sum_{i=1}^{n} p_u(X_{i,1}, X_{i,2}, \ldots, X_{i,k}).$$

It follows from (5.6) that $\deg F_{n,k,u} \le k-1$, which makes it possible to partition the monomials of $F_{n,k,u}$ among the players in some predetermined fashion and have each player report the sum of the monomials assigned to him. This simultaneous protocol has cost $k\lceil \log 3 \rceil = 2k$. With probability at least $1 - n2^{-k}$ over shared randomness, we have $u \neq (X_{i,1}, X_{i,2}, \ldots, X_{i,k})$ for $i = 1, 2, \ldots, n$, in which case (5.7) implies that $F_{n,k}(X) \equiv 1 - F_{n,k,u}(X)^2 \pmod 3$. In particular, the players' broadcasts uniquely identify $F_{n,k}(X)$ with probability at least $1 - n2^{-k}$ on any given input X. Thus,

$$R_{n2^{-k}}(F_{n,k}) \le 2k. \tag{5.8}$$

To complete the proof of the theorem, we consider three cases depending on the value of k.

- Equation (5.8) directly implies that $R_{1/3}(F_{n,\lceil \log 3n \rceil}) \le 2\lceil \log 3n \rceil$, which settles the case $k = \lceil \log 3n \rceil$. As usual, the error probability can be reduced from $1/3$ to any positive constant by running the protocol simultaneously multiple times and outputting the majority answer. In particular, $R_\epsilon(F_{n,\lceil \log 3n \rceil}) = O(\log n)$ for any fixed $\epsilon > 0$.
- For $k > \lceil \log 3n \rceil$, observe that

$$F_{n,k}(X) = \mathrm{MOD}_3\left(X_{1,1} \oplus \cdots \oplus X_{1,\lceil \log 3n \rceil - 1} \oplus \left(\bigoplus_{j=\lceil \log 3n \rceil}^{k} X_{1,j}\right), \ldots \right.$$
$$\left. X_{n,1} \oplus \cdots \oplus X_{n,\lceil \log 3n \rceil - 1} \oplus \left(\bigoplus_{j=\lceil \log 3n \rceil}^{k} X_{n,j}\right)\right)$$

  and in particular $R_\epsilon(F_{n,k}) \le R_\epsilon(F_{\lceil \log 3n \rceil,k})$ for all ϵ. As a result, we conclude by the first case that $R_\epsilon(F_{n,k}) \le O(\log n)$ for any fixed $\epsilon > 0$ and $k > \lceil \log 3n \rceil$.
- For $\log n \le k < \lceil \log 3n \rceil$, the players partition the input matrix X horizontally into submatrices with at most $n/3$ rows each and run the protocol from the previous cases with the error parameter set to a small constant. Then with probability at least $2/3$, the players' broadcasts uniquely determine $\sum_{i=1}^{n} \bigoplus_{j=1}^{k} X_{i,j}$ modulo 3 and thereby reveal $F_{n,k}(X)$. □

ACKNOWLEDGMENTS

The work of Vladimir Podolskii was supported by the HSE University Basic Research Program. The work of Alexander Sherstov was supported by NSF CAREER award CCF-1149018 and an Alfred P. Sloan Foundation Research Fellowship.

REFERENCES

[1] Anil Ada, Arkadev Chattopadhyay, Omar Fawzi, and Phuong Nguyen. 2015. The NOF Multiparty Communication Complexity of Composed Functions. *Computational Complexity* 24, 3 (2015), 645–694. https://doi.org/10.1007/s00037-013-0078-4

[2] László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. 2003. Communication Complexity of Simultaneous Messages. *SIAM J. Comput.* 33, 1 (2003), 137–166. https://doi.org/10.1137/S0097539700375944

[3] László Babai, Noam Nisan, and Mario Szegedy. 1992. Multiparty Protocols, Pseudorandom Generators for Logspace, and Time-Space Trade-Offs. *J. Comput. Syst. Sci.* 45, 2 (1992), 204–232. https://doi.org/10.1016/0022-0000(92)90047-M

[4] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. 2004. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.* 68, 4 (2004), 702–732. https://doi.org/10.1016/j.jcss.2003.11.006

[5] Paul Beame and Trinh Huynh. 2012. Multiparty Communication Complexity and Threshold Circuit Size of AC^0. *SIAM J. Comput.* 41, 3 (2012), 484–518. https://doi.org/10.1137/100792779

[6] Paul Beame, Toniann Pitassi, and Nathan Segerlind. 2007. Lower Bounds for Lovász-Schrijver Systems and Beyond Follow from Multiparty Communication Complexity. *SIAM J. Comput.* 37, 3 (2007), 845–869. https://doi.org/10.1137/060654645

[7] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. 2006. A Strong Direct Product Theorem for Corruption and the Multiparty Communication Complexity of Disjointness. *Computational Complexity* 15, 4 (2006), 391–432. https://doi.org/10.1007/s00037-007-0220-2
[8] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. 1983. Multi-Party Protocols. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing* (STOC). 94–99. https://doi.org/10.1145/800061.808737
[9] Arkadev Chattopadhyay and Anil Ada. 2008. Multiparty Communication Complexity of Disjointness. In *Electronic Colloquium on Computational Complexity (ECCC)*. Report TR08-002.
[10] Arkadev Chattopadhyay and Michael E. Saks. 2014. The Power of Super-logarithmic Number of Players. In *Proceedings of the Eighteenth International Workshop on Randomization and Computation* (RANDOM). 596–603. https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2014.596
[11] Benny Chor and Oded Goldreich. 1988. Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity. *SIAM J. Comput.* 17, 2 (1988), 230–261. https://doi.org/10.1137/0217015
[12] Fan R. K. Chung and Prasad Tetali. 1993. Communication Complexity and Quasi Randomness. *SIAM J. Discrete Math.* 6, 1 (1993), 110–123. https://doi.org/10.1137/0406009
[13] Jeffrey Stephen Ford. 2006. *Lower Bound Methods for Multiparty Communication Complexity.* Ph.D. Dissertation. The University of Texas at Austin.
[14] Vince Grolmusz. 1994. The BNS Lower Bound for Multi-Party Protocols in Nearly Optimal. *Inf. Comput.* 112, 1 (1994), 51–54. https://doi.org/10.1006/inco.1994.1051
[15] Johan Håstad and Mikael Goldmann. 1991. On the Power of Small-Depth Threshold Circuits. *Computational Complexity* 1 (1991), 113–129. https://doi.org/10.1007/BF01272517
[16] Stasys Jukna. 2001. *Extremal Combinatorics with Applications in Computer Science.* Springer-Verlag, Berlin. https://doi.org/10.1007/978-3-662-04650-0
[17] Bala Kalyanasundaram and Georg Schnitger. 1992. The Probabilistic Communication Complexity of Set Intersection. *SIAM J. Discrete Math.* 5, 4 (1992), 545–557. https://doi.org/10.1137/0405044
[18] Eyal Kushilevitz and Noam Nisan. 1997. *Communication complexity.* Cambridge University Press.
[19] Troy Lee and Adi Shraibman. 2009. Disjointness is Hard in the Multiparty Number-on-the-Forehead Model. *Computational Complexity* 18, 2 (2009), 309–336. https://doi.org/10.1007/s00037-009-0276-2
[20] Noam Nisan and Mario Szegedy. 1994. On the degree of Boolean functions as real polynomials. *Computational Complexity* 4 (1994), 301–313. https://doi.org/10.1007/BF01263419
[21] Ran Raz. 2000. The BNS-Chung criterion for multi-party communication complexity. *Comput. Complex.* 9, 2 (2000), 113–122. https://doi.org/10.1007/PL00001602
[22] Alexander A. Razborov. 1992. On the distributional complexity of disjointness. *Theor. Comput. Sci.* 106, 2 (1992), 385–390. https://doi.org/10.1016/0304-3975(92)90260-M
[23] Alexander A. Razborov and Avi Wigderson. 1993. $n^{\Omega(\log n)}$ Lower Bounds on the Size of Depth-3 Threshold Circuits with AND Gates at the Bottom. *Inf. Process. Lett.* 45, 6 (1993), 303–307. https://doi.org/10.1016/0020-0190(93)90041-7
[24] Alexander A. Sherstov. 2014. Communication lower bounds using directional derivatives. *J. ACM* 61, 6 (2014), 1–71. https://doi.org/10.1145/2629334 Preliminary version in *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing* (STOC), 2013.
[25] Alexander A. Sherstov. 2016. The multiparty communication complexity of set disjointness. *SIAM J. Comput.* 45, 4 (2016), 1450–1489. https://doi.org/10.1137/120891587 Preliminary version in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 2009.
[26] Pascal Tesson. 2003. *Computational complexity questions related to finite monoids and semigroups.* Ph.D. Dissertation. McGill University.
[27] Andrew Chi-Chih Yao. 1990. On ACC and Threshold Circuits. In *Proceedings of the Thirty-First Annual IEEE Symposium on Foundations of Computer Science* (FOCS). 619–627. https://doi.org/10.1109/FSCS.1990.89583

A A CORRELATION BOUND

The purpose of this appendix is to provide a short proof of Lemma 5.1. The only technical prerequisite for the proof is the following fact about cylinder intersections, which is implicit in the work of Babai et al. [3] and is derived explicitly in the followup papers by Chung and Tetali [12] and Raz [21].

Fact A.1 (Chung and Tetali; Raz). *Let $U_1, U_2, \ldots, U_k$ be finite sets. Then for any function $\phi\colon U_1 \times U_2 \times \cdots \times U_k \to \mathbb{C}$ and any cylinder intersection $\chi\colon U_1 \times U_2 \times \cdots \times U_k \to \{0,1\}$,*

$$\left|\mathop{\mathrm{E}}_{u_1,\ldots,u_k} \phi(u_1,\ldots,u_k)\chi(u_1,\ldots,u_k)\right|^{2^k} \leq \mathop{\mathrm{E}}_{\substack{u_1^0,\ldots,u_k^0\\ u_1^1,\ldots,u_k^1}} \left[\prod_{\substack{z\in\{0,1\}^k\\ |z| \text{ even}}} \phi(u_1^{z_1},\ldots,u_k^{z_k}) \cdot \prod_{\substack{z\in\{0,1\}^k\\ |z| \text{ odd}}} \overline{\phi(u_1^{z_1},\ldots,u_k^{z_k})}\right],$$

where $u_i, u_i^0, u_i^1 \in U_i$ for each i.

We are now in a position to prove Lemma 5.1, which we restate below for the reader's convenience. Recall that we use the shorthand $e(t) = \exp(2\pi \mathrm{i} t)$, where i is the imaginary unit.

Lemma (Ada et al.). *For every cylinder intersection $\chi\colon (\{0,1\}^n)^k \to \{0,1\}$,*

$$\left|\mathop{\mathrm{E}}_{X\in\{0,1\}^{n\times k}} e\left(\frac{1}{3}\sum_{i=1}^{n}\bigoplus_{j=1}^{k} X_{i,j}\right)\chi(X)\right| < \exp\left(-\frac{n}{4^k}\right).$$

Proof. We mostly follow the analysis of Ada et al. [1], who proved a more general correlation bound. As one would expect, focusing on a special case as we do here allows for a shorter and simpler presentation. We have:

$$\begin{aligned}
\left|\mathop{\mathrm{E}}_{X\in\{0,1\}^{n\times k}} e\left(\frac{1}{3}\sum_{i=1}^{n}\bigoplus_{j=1}^{k} X_{i,j}\right)\chi(X)\right|^{2^k} &\leq \mathop{\mathrm{E}}_{X^0,X^1\in\{0,1\}^{n\times k}} \prod_{z\in\{0,1\}^k} e\left(\frac{(-1)^{|z|}}{3}\sum_{i=1}^{n}\bigoplus_{j=1}^{k} X_{i,j}^{z_j}\right)\\
&= \mathop{\mathrm{E}}_{X^0,X^1\in\{0,1\}^{n\times k}} \prod_{i=1}^{n} e\left(\frac{1}{3}\sum_{z\in\{0,1\}^k}(-1)^{|z|}\bigoplus_{j=1}^{k} X_{i,j}^{z_j}\right)\\
&= \left(\mathop{\mathrm{E}}_{x^0,x^1\in\{0,1\}^k} e\left(\frac{1}{3}\sum_{z\in\{0,1\}^k}(-1)^{|z|}\bigoplus_{j=1}^{k} x_j^{z_j}\right)\right)^n\\
&= \left(\mathop{\mathrm{E}}_{x^0,x^1\in\{0,1\}^k} e\left(\frac{1}{3}\sum_{z\in\{0,1\}^k}(-1)^{|z|}\left(\prod_{j=1}^{k}(x_j^{z_j}+1)-1\right)\right)\right)^n\\
&= \left(\mathop{\mathrm{E}}_{x^0,x^1\in\{0,1\}^k} e\left(\frac{1}{3}\prod_{j=1}^{k}(x_j^0 - x_j^1)\right)\right)^n, \qquad \text{(A.1)}
\end{aligned}$$

where the first step follows from Fact A.1, and the fourth step uses the fact that

$$\bigoplus_{j=1}^{k} x_j^{z_j} \equiv \prod_{j=1}^{k}(x_j^{z_j}+1) - 1 \pmod 3.$$

A routine calculation reveals that

$$e\left(\frac{1}{3}\prod_{j=1}^{k}(x_j^0 - x_j^1)\right) = \begin{cases} 1 & \text{if } x_j^0 = x_j^1 \text{ for some } j,\\ -\frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i}\prod_{j=1}^{k}(x_j^0 - x_j^1) & \text{otherwise,}\end{cases}$$

where i is the imaginary unit. Making this substitution in (A.1) yields

$$\begin{aligned}\left|\mathop{\mathrm{E}}_{X\in\{0,1\}^{n\times k}} e\left(\frac{1}{3}\sum_{i=1}^{n}\bigoplus_{j=1}^{k} X_{i,j}\right)\chi(X)\right|^{2^k} &\leq \left(\left(1-\frac{1}{2^k}\right)\cdot 1-\frac{1}{2^k}\cdot\frac{1}{2}\right)^n \\ &\leq \left(1-\frac{1}{2^k}\right)^n \\ &< \exp\left(-\frac{n}{2^k}\right).\end{aligned}$$

□