Multiparty Communication Complexity of Disjointness

Arkadev Chattopadhyay and Anil Ada*

School of Computer Science
McGill University, Montreal, Canada
achatt3,aada@cs.mcgill.ca

Abstract

We obtain a lower bound of $n^{\Omega(1)}$ on the k-party randomized communication complexity of the Disjointness function in the 'Number on the Forehead' model of multiparty communication when k is a constant. For $k = o(\log\log n)$, the bounds remain super-polylogarithmic i.e. $(\log n)^{\omega(1)}$. The previous best lower bound for three players until recently was $\Omega(\log n)$.

Our bound separates the communication complexity classes NP_k^{CC} and BPP_k^{CC} for $k = o(\log\log n)$. Furthermore, by the results of Beame, Pitassi and Segerlind [4], our bound implies proof size lower bounds for tree-like, degree $k-1$ threshold systems and superpolynomial size lower bounds for Lovász-Schrijver proofs.

To obtain our result, we further develop the "Generalized Discrepancy Method" recently suggested by Sherstov [16]. The other main components of the proof are the "Approximation/Orthogonality Principle" that also appears in [16] and techniques to estimate discrepancy under non-uniform distribution developed by Chattopadhyay [8].

A similar bound for Disjointness has been recently and independently obtained by Lee and Shraibman.

1 Introduction

Chandra, Furst and Lipton [7] introduced the 'Number on the Forehead' model of multiparty communication as an extension of Yao's [20] two party

*authors are supported by research grants of Prof. D. Thérien. The first author thanks M. David and T. Pitassi for several discussions.

communication model. This model, besides being interesting in its own right, has found numerous connections with circuit complexity, proof complexity, branching programs, pseudo-random generators and other areas of theoretical computer science.

Both proving upper and lower bounds for this model remain a very challenging task as it is known that the overlap of information accessible to players provides significant power to it. In fact, proving a super-polylogarithmic lower bound on the communication needed by poly-logarithmic number of players for computing a function f in the restricted setting of simultaneous deterministic communication, is enough to show that f is not in ACC^0, a class for which no strong bounds are known. Although several efforts [2, 9, 14, 10] have been made, this goal currently remains out of reach as no superlogarithmic lower bounds exist for even $\log n$ players.

More modestly, one would like to be able to determine the communication complexity of simple functions for at least constant number of players. However, despite intensive research (see for example [5, 6, 19, 18]) the best known lower bounds on the communication complexity of simple functions like Disjointness and Pointer Jumping was $\Omega(\log n)$ even for three players. The root cause of this problem is that there was essentially only one method that was the backbone of almost all strong lower bounds. This method is known as the discrepancy method and was introduced in the seminal work of Babai, Nisan and Szegedy [2]. It is however known that for functions like Disjointness this method at best yields $\Omega(\log n)$ lower bounds.

In this work, we further develop a technique called the Generalized Discrepancy Method as suggested in the recent work of Sherstov [16] and implicitly by Razborov [15] that in principle applies to functions that have a large discrepancy. Combining the tools of Chattopadhyay [8] for estimating discrepancy under certain non-uniform distribution over inputs and the beautiful Approximation/Orthogonality principle discovered in [16], we are able to apply the Generalized Discrepancy Method to yield $n^{\Omega(1)}$ lower bounds on the k-party communication complexity of Disjointness in the bounded error randomized model (with public coin tosses) as long as k is a constant.

Our result has interesting consequences for communication complexity classes and proof complexity. It provides the first example of an explicit function that has small non-deterministic communication complexity, but exponentially high randomized complexity. In the language of complexity classes, this separates BPP_k^{CC} and NP_k^{CC} for $k = o(\log \log n)$. In fact, the separation is exponential when k is any constant. Although such a separation was already known from the work of [3], before our work no explicit function was known to separate these classes. By the work of Beame, Pittasi and Szegerlind [4], our lower bounds on the k-party complexity of Disjointness implies strong lower bounds on the proof size for a family of proof systems known as tree-like, degree $k-1$ threshold systems. Proving lower bounds for these systems was a major open problem in propositional proof complexity.

1.1 Our Main Result

Let $y^1, \ldots, y^{k-1}$ be $k-1$ n-bit binary strings. Define the $k-1 \times n$ boolean matrix A obtained by placing y^i in the ith row of A. For $x \in \{0,1\}^n$, let $x \Leftarrow y^1, \ldots, y^{k-1}$ be the n-bit string $x_{i_1} x_{i_2} \ldots x_{i_t} 0^{n-t}$, where $i_1, \ldots, i_t$ are the indices of the all-one columns of A.

Let $g : \{0,1\}^n \to \{-1,1\}$ be a *base* function. We define $G_k^g : (\{0,1\}^n)^k \to \{-1,1\}$ by $G_k^g(x, y^1, \ldots, y^{k-1}) := g(x \Leftarrow y^1, \ldots, y^{k-1})$. Observe that G_k^{PARITY} is the Generalized Inner Product function and G_k^{NOR} is the Disjointness function. Our main result shows how to use the high approximation degree of a base function to generate a function with high randomized communication complexity.

Let $R_k^\epsilon(f)$ denote the randomized k-party communication complexity of f with advantage ϵ. Then,

Theorem 1.1. *Let $f : \{0,1\}^m \to \{-1,1\}$ have δ-approximate degree d. Let $n \geq \left(2^{2^k}(k-1)e\right)^{k-1} m^k$, and $f' : \{0,1\}^n \to \{-1,1\}$ be such that $f(z) = f'(z0^{n-m})$. Then*

$$R_k^\epsilon(G_k^{f'}) \geq \frac{d}{2^{k-1}} + \log(\delta + 2\epsilon - 1).$$

As a corollary we show that

$$R_k^\epsilon(\text{DISJ}_k) = \Omega\left(\frac{n^{1/2k}}{2^{2^k} e(k-1)2^{k-1}}\right)$$

for a constant ϵ. In brief, this follows from the following facts. Let NOR_n denote the NOR function for inputs of length n. Then $f' = \text{NOR}_n$ and $f = \text{NOR}_m$ satisfy $f(z) = f'(z0^{n-m})$ and by a result of Paturi [13], we know that the 1/3-approximate degree of NOR_m is $\Theta(\sqrt{m})$.

A similar bound for the Disjointness function has been recently and independently obtained by Lee and Shraibman [12].

1.2 Proof Overview

To obtain our results we use three main ingredients, the first of which is the Generalized Discrepancy Method. The classical discrepancy method states that if a function has low discrepancy, then it has high randomized communication complexity. In the generalized discrepancy method this idea is extended as follows. If a function g correlates well with f and has low discrepancy, then f has high randomized communication complexity.

The second ingredient is the "Approximation/Orthogonality Principle" of Sherstov [16]. It states that given a function f with high approximation degree, we can find g and a distribution μ such that g is orthogonal to every low degree polynomial under μ.

The third ingredient, called the Orthogonality-Discrepancy Lemma, is derived from the work of Chattopadhyay [8]. This takes a function that is orthogonal with low degree polynomials and constructs a new masked function that has low discrepancy.

We can then summarize our strategy as follows. We start with a function $f : \{0,1\}^n \to \{-1,1\}$ with high approximation degree. By the Approximation/Orthogonality Principle, we obtain g that highly correlates with f and is orthogonal with low degree polynomials. From f and g we construct new masked functions F_k^f and F_k^g, similar to the construction of G_k^f. Since g is orthogonal to low degree polynomials, by the Orthogonality-Discrepancy Lemma we deduce that F_k^g has low discrepancy under an appropriate distribution. Under this distribution F_k^g and F_k^f are highly correlated and therefore applying the Generalized Discrepancy Method, we conclude that F_k^f has high randomized communication complexity. This implies, by the construction of F_k^f, that the randomized communication complexity of G_k^f is high. See Figure 1 for an outline.

2 Preliminaries

2.1 Multiparty Communication Model

In the multiparty communication model introduced by [7], k players $P_1, \ldots, P_k$ wish to collaborate to compute a function $f : \{0,1\}^n \to \{-1,1\}$. The n input bits are partitioned into k sets $X_1, \ldots, X_k \subseteq [n]$ and each participant P_i knows the values of all the input bits *except* the ones of X_i. This game is often referred to as the "Number/Input on the forehead" model since it is convenient to picture that player i has the bits of X_i written on its forehead, available to everyone but itself. Players exchange bits, according to

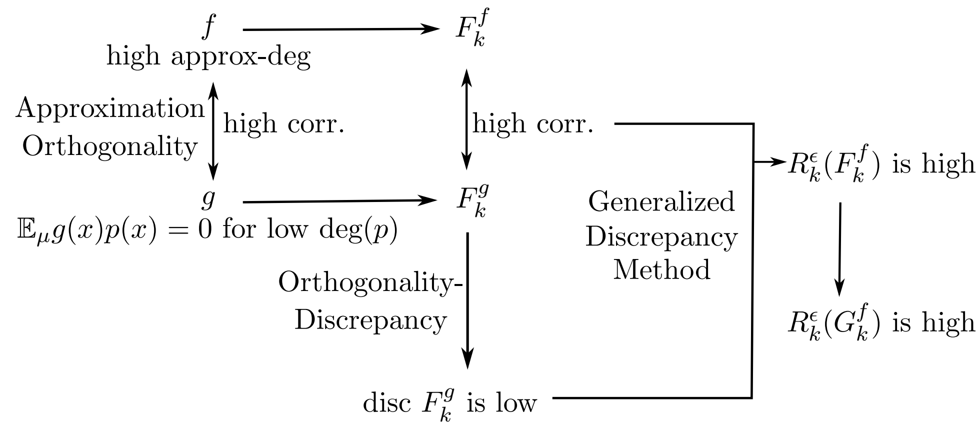

Figure 1: Proof outline

an agreed upon protocol, by writing them on a public blackboard. The protocol specifies whose turn it is to speak, and what the player broadcasts as a function of the communication history and the input the player has access to. The protocol's output is a function of what is on the blackboard after the protocol's termination. We denote by $D_k(f)$ the deterministic k-party communication complexity of f, i.e. the number of bits exchanged in the *best* deterministic protocol for f on the worst case input.

By allowing the players to access a public random string and the protocol to err, one defines the randomized communication complexity of a function. We say that a protocol computes f with ϵ advantage if the probability that $\mathcal{P}$ and f agree is at least $1/2+\epsilon$ for all inputs. We denote by $R_k^\epsilon(f)$ the cost of the best protocol that computes f with advantage ϵ. One further introduces non-determinism in protocols by allowing 'God' to help the players by furnishing a proof string. As is usual with non-determinism in other models, a correct non-deterministic protocol $\mathcal{P}$ for f has the following property: on every input x at which $f(x) = -1$, $\mathcal{P}(x,y) = -1$ for some proof string y and whenever $f(x) = 1$, $\mathcal{P}(x,y) = 1$ for all proof strings y. The length of the proof string y is now included in the cost of $\mathcal{P}$ on an input and $N_k(f)$ denotes the cost of the best non-deterministic protocol for f on the worst input.

Communication complexity classes were introduced for two players in [1] in which "efficient" protocol was defined to have cost no more than *polylog*(n). This idea naturally extends to the multiparty model giving rise to the following classes: $\mathrm{P}_k^{CC} := \{f \mid D_k(f) = \mathrm{polylog}(n)\}$, $\mathrm{BPP}_k^{CC} :=$

$\{f \mid R_k^{1/3}(f) = \text{polylog}(n)\}$ and $\text{NP}_k^{CC} := \{f \mid N_k(f) = \text{polylog}(n)\}$. Determining the relationship among these classes is an interesting research theme within the broader area of understanding the relative power of determinism, non-determinism and randomness in computation. While Beame et.al. [3] show that $\text{BPP}_k^{CC} \neq \text{NP}_k^{CC}$, no explicit function was known that separated these classes.

2.2 Cylinder Intersections and Discrepancy

The key combinatorial object that arises in the study of multiparty communication is a *cylinder-intersection.* A k-cylinder in the ith dimension is a subset S of $Y_1 \times \cdots \times Y_k$ with the property that membership in S is independent of the ith coordinate. A set S is called a cylinder-intersection if $S = \cap_{i=1}^{k} S_i$, where S_i is a cylinder in the ith dimension. One can represent a k-cylinder in the ith dimension by its characteristic function $\phi^i : (\{0,1\}^n)^k \rightarrow \{0,1\}$. Here $\phi^i(y_1, ..., y_k)$ does not depend on y_i. A cylinder intersection is represented as the product

$$\phi(y_1, ..., y_k) = \phi^1(y_1, ..., y_k)...\phi^k(y_1, ..., y_k).$$

It is well known that a protocol that computes f with cost c partitions the input space of f into at most 2^c monochromatic cylinder intersections.

An important measure, defined for a function $f : Y_1 \times ... \times Y_k \rightarrow \{-1, 1\}$, is its *discrepancy.* With respect to any probability distribution μ over $Y_1 \times \cdots \times Y_k$ and cylinder intersection ϕ, define

$$\begin{aligned} \text{disc}_{k,\mu}^{\phi}(f) &= \Big| \Pr_{\mu} \left[f(y_1, \ldots, y_k) = 1 \wedge \phi(y_1, \ldots, y_k) = 1 \right] \\ &\quad - \Pr_{\mu} \left[f(y_1, \ldots, y_k) = -1 \wedge \phi(y_1, \ldots, y_k) = 1 \right] \Big|. \end{aligned}$$

Since f is -1/1 valued, it is not hard to verify that equivalently:

$$\text{disc}_{k,\mu}^{\phi}(f) = \Big| \mathbb{E}_{y_1, \ldots, y_k \sim \mu} f(y_1, \ldots, y_k) \phi(y_1, \ldots, y_k) \Big|. \tag{1}$$

The discrepancy of f w.r.t. μ, denoted by $\text{disc}_{k,\mu}(f)$ is $\max_{\phi} \text{disc}_{k,\mu}^{\phi}(f)$. For removing notational clutter, we often drop μ from the subscript when the distribution is clear from the context. We now state the discrepancy method which connects the discrepancy and the randomized communication complexity of a function.

Theorem 2.1 (see [2, 11]). *Let $0 < \epsilon \leq 1/2$ be any real and $k \geq 2$ be any integer. For every function $f : Y_1 \times ... \times Y_k \rightarrow \{1, -1\}$ and distribution μ on inputs from $Y_1 \times \cdots \times Y_k$,*

$$R_k^{\epsilon}(f) \geq \log\left(\frac{2\epsilon}{disc_{k,\mu}(f)}\right). \tag{2}$$

2.3 Fourier Expansion

We consider the vector space of functions from $\{0,1\}^n \rightarrow \mathbb{R}$. Equip this space with the standard inner product $\langle f, g \rangle$

$$\langle f, g \rangle = \mathbb{E}_{x \sim \mathcal{U}} f(x) g(x) \tag{3}$$

For each $S \subseteq [n]$, define $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. Then it is easy to verify that the set of functions $\{\chi_S | S \subseteq [n]\}$ forms an orthonormal basis for this inner product space, and so every f can be expanded in terms of its *Fourier coefficients*

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x) \tag{4}$$

where $\hat{f}(S)$ is defined as $\langle f, \chi_S \rangle$. This expansion is unique and the *exact degree* of f is defined to be the largest d such that there exists $S \subseteq [n]$ with $|S| = d$ and $\hat{f}(S) \neq 0$.

2.4 Approximation Degree

A natural question is the following. How large degree is needed if we want to simply approximate f well? Define the *ϵ-approximate degree of f*, denoted by $\deg_\epsilon(f)$ to be the smallest integer d for which there exists a function ϕ of exact degree d such that

$$\max_{x \in \{0,1\}^n} \left| f(x) - \phi(x) \right| \leq \epsilon$$

For any $D : \{0, 1, \ldots, n\} \rightarrow \{1, -1\}$, define

$$\ell_0(D) \in \{0, 1, \ldots, \lfloor n/2 \rfloor\}$$

$$\ell_1(D) \in \{0, 1, \ldots, \lceil n/2 \rceil\}$$

such that D is constant over the interval $[\ell_0(D), n - \ell_1(D)]$ and $\ell_0(D)$ and $\ell_1(D)$ are the smallest possible values for which this happens.

Paturi's theorem provides bounds on the approximate degree of symmetric functions.

Theorem 2.2 (Paturi[13]). *Let $f : \{0,1\}^n \to \{1,-1\}$ be any symmetric function induced from the predicate $D : \{0,\ldots,n\} \to \{1,-1\}$. Then,*

$$deg_{1/3}(f) = \Theta\big(\sqrt{n(\ell_0(D) + \ell_1(D))}\big) \tag{5}$$

In particular, the 1/3-approximate degree of NOR is $\Theta(\sqrt{n})$.

3 The Generalized Discrepancy Method

Babai, Nisan and Szegedy [2] estimated the discrepancy of functions like GIP_k w.r.t k-wise cylinder intersections and the uniform distribution. These estimates resulted in the first strong lower bounds in the k-party model via Theorem 2.1. Unfortunately, the applicability of Theorem 2.1 is limited to those functions that have small discrepancy. Disjointess is a classical example of a function that does not have small discrepancy.

Lemma 3.1 (Folklore). *Under every distribution μ over the inputs,*

$$disc_{k,\mu}(DISJ_k) = \Omega(1/n).$$

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

Thus, it is impossible to obtain better than $\Omega(\log n)$ bounds on the communication complexity of Disjointness by a direct application of the discrepancy method. In fact, the above method shows that Theorem 2.1 fails to give better than polylogarithmic lower bound for every function that is in NP_k^{CC} or co-NP_k^{CC}.

For dealing with such functions we need to generalize the discrepancy method. Simply put, the generalized discrepancy method states that a low discrepancy of a function g that highly correlates with f implies a high randomized communication complexity for f. Formally we define correlation to be $\mathrm{Corr}_\mu(f,g) = |\mathbb{E}_{x\sim\mu} f(x)g(x)|$.

Lemma 3.2 (Generalized Discrepancy Method). *Denote $X = Y_1 \times \ldots \times Y_k$. Let $f : X \to \{-1,1\}$ and $g : X \to \{-1,1\}$ be such that under some distribution μ we have $Corr_\mu(f,g) \geq \delta$. Then*

$$R_k^\epsilon(f) \geq \log\left(\frac{\delta + 2\epsilon - 1}{disc_{k,\mu}(g)}\right) \tag{6}$$

Proof. Let $\mathcal{P}$ be a k-party randomized protocol that computes f with advantage ϵ and cost c. Then for every distribution μ over the inputs, we can derive a deterministic k-player protocol $\mathcal{P}'$ for f that errs only on at most $1/2-\epsilon$ fraction of the inputs (w.r.t. μ) and has cost c. Take μ to be a distribution satisfying the correlation inequality. We know $\mathcal{P}'$ partitions the input space into at most 2^c monochromatic (w.r.t. $\mathcal{P}'$) cylinder intersections. Let $\mathcal{C}$ denote this set of cylinder intersections. Then,

$$\begin{aligned}
\delta &\leq \left|\mathbb{E}_{x\sim\mu} f(x)g(x)\right| \\
&= \Big|\sum_x f(x)g(x)\mu(x)\Big| \\
&\leq \Big|\sum_x \mathcal{P}'(x)g(x)\mu(x)\Big| + \Big|\sum_x (f(x)-\mathcal{P}'(x))g(x)\mu(x)\Big|
\end{aligned}$$

Since $\mathcal{P}'$ is a constant over every cylinder intersection S in $\mathcal{C}$, we have

$$\begin{aligned}
\delta &\leq \sum_{S\in\mathcal{C}} \Big|\sum_{x\in S} \mathcal{P}'(x)g(x)\mu(x)\Big| + \sum_x \big|g(x)\big|\big|f(x)-\mathcal{P}'(x)\big|\mu(x) \\
&\leq \sum_{S\in\mathcal{C}} \Big|\sum_{x\in S} g(x)\mu(x)\Big| + \sum_x \big|f(x)-\mathcal{P}'(x)\big|\mu(x) \\
&\leq 2^c \mathrm{disc}_{k,\mu}(g) + 2(1/2-\epsilon).
\end{aligned}$$

This gives us immediately (6). ☐

Observe that when $f = g$, i.e. $\mathrm{Corr}_\mu(f,g) = 1$, we get the classical discrepancy method (Theorem 2.1).

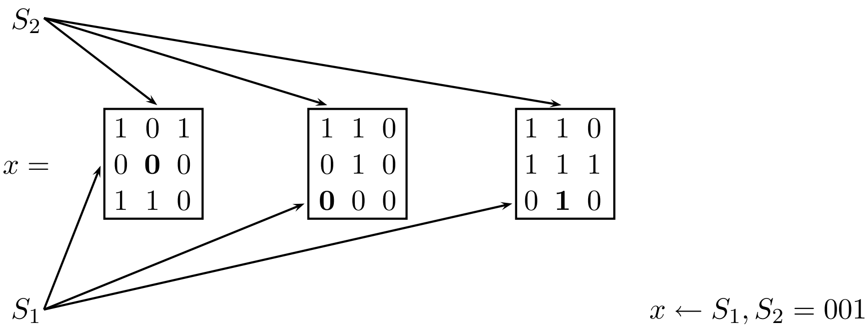

Figure 2: Illustration of the masking scheme $x \leftarrow S_1, S_2$. The parameters are $\ell = 3, m = 3, n = 27$.

4 Generating Functions With Low Discrepancy

4.1 Masking Schemes

We have already defined one masking scheme through the notation $x \Leftarrow y_1, \ldots, y_k$. This allowed us to define G_k^g for a base function g. Well-known functions such as GIP_k and DISJ_k are respresentable in this notation by G_k^{PARITY} and G_k^{NOR} respectively. Now we define a second masking scheme which plays a crucial role in lowerbounding the communication complexity of G_k^g.

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

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

where $x[i]$ refers to the ith block of x. See Figure 2 for an illustration of this masking scheme.

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

Lemma 4.1. *If $f : \{0,1\}^m \to \{-1,1\}$ and $f' : \{0,1\}^n \to \{-1,1\}$ have the property that $f(z) = f'(z0^{n-m})$ (here $n = \ell^{k-1}m$ as described in the* *construction of F_k^f), then*

$$R_k^\epsilon(F_k^f) \leq R_k^\epsilon(G_k^{f'}). \tag{7}$$

Proof Sketch. Observe that there are functions $\Gamma_i : [\ell]^m \to \{0,1\}^n$ such that $F_k^f(x, S^1, \ldots, S^{k-1}) = G_k^{f'}(x, \Gamma_1(S^1), \ldots, \Gamma_{k-1}(S^{k-1}))$ for all $x, S^1, \ldots, S^{k-1}$. Therefore the players can privately convert their inputs and apply the protocol for $G_k^{f'}$. $\square$

Note that the proof shows (7) holds not just for randomized but any model of communication.

4.2 Orthogonality and Discrepancy

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

Lemma 4.2 (Orthogonality-Discrepancy Lemma). *Let $f : \{-1,1\}^m \to \{-1,1\}$ be any (μ, d)-orthogonal function for some distribution μ on $\{-1,1\}^m$ and some integer $d > 0$. Derive the probability distribution λ on $\{-1,1\}^n \times \left([\ell]^m\right)^{k-1}$ from μ as follows: $\lambda(x, S^1, \ldots, S^{k-1}) = \frac{\mu(x \leftarrow S^1,\ldots,S^{k-1})}{\ell^{m(k-1)}2^{n-m}}$. Then,*

$$\left(disc_{k,\lambda}(F_k^f)\right)^{2^{k-1}} \leq \sum_{j=d}^{(k-1)m} \binom{(k-1)m}{j} \left(\frac{2^{2^{k-1}-1}}{\ell - 1}\right)^j \tag{8}$$

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

$$disc_{k,\lambda}(F_k^f) \leq \frac{1}{2^{d/2^{k-1}}}. \tag{9}$$

Remark. The Lemma above appears very similar to the Multiparty Degree-Discrepancy Lemma in [8] that is an extension of the two party Degree-Discrepancy Lemma of [17]. There, the magic property on the base function is high voting degree. It is worth noting that (μ, d)-orthogonality of f is equivalent to voting degree of f being at least d. Indeed the proof of the above Lemma is almost identical to the proof of the Degree-Discrepancy Lemma save for the minor details of the difference between our masking scheme and the one used in [8].

Proof of Lemma 4.2. The starting point is to write the expression for discrepancy w.r.t. an arbitrary cylinder intersection ϕ,

$$\begin{aligned}\operatorname{disc}_k^{\phi}(F_k^f) = \\ \Big| \sum_{x,S^1,\ldots,S^{k-1}} \quad & F_k^f(x,S^1,\ldots,S^{k-1})\phi(x,S^1,\ldots,S^k) \\ & \cdot\lambda(x,S^1,\ldots,S^{k-1})\Big| \end{aligned} \tag{10}$$

This changes to the more convenient expected value notation as follows:

$$\begin{aligned}\operatorname{disc}_k^{\phi}(F_k^f) = \quad & 2^m \Big| \mathbb{E}_{x,S^1,\ldots,S^{k-1}} F_k^f(x,S^1,\ldots,S^{k-1}) \\ & \times \phi(x,S^1,\ldots,S^{k-1})\mu\big(x \leftarrow S^1,\ldots,S^{k-1}\big)\Big| \end{aligned} \tag{11}$$

where, $(x,S^1,\ldots,S^{k-1})$ is now uniformly distributed over $\{0,1\}^{\ell^{k-1}m} \times \big([\ell]^m\big)^{k-1}$. Then, we use the trick of repeatedly combining triangle inequality with Cauchy-Schwarz exactly as done in Chattopadhyay[8] to obtain the following:

$$\begin{aligned}&(\operatorname{disc}_k^{\phi}(F_k^f))^{2^{k-1}} \leq \\ &2^{2^{k-1}m}\mathbb{E}_{S_0^1,S_1^1,\ldots,S_0^{k-1},S_1^{k-1}} H_k^f\big(S_0^1,S_1^1,\ldots,S_0^{k-1},S_1^{k-1}\big)\end{aligned} \tag{12}$$

where,

$$\begin{aligned}&H_k^f\big(S_0^1,S_1^1,\ldots,S_0^{k-1},S_1^{k-1}\big) \\ &= \Big|\mathbb{E}_{x\in\{0,1\}^{t^{k-1}m}} \prod_{u\in\{0,1\}^{k-1}} \Big(F_k^f(x,S_{u_1}^1,\ldots,S_{u_{k-1}}^{k-1}) \\ &\qquad \times \mu(x \leftarrow S_{u_1}^1,\ldots,S_{u_{k-1}}^{k-1})\Big)\Big|\end{aligned} \tag{13}$$

We look at a fixed S_0^i, S_1^i, for $i = 1,\ldots,k-1$. Let $r_i = \big|S_0^i \cap S_1^i\big|$ and $r = \sum_i r_i$ for $1 \leq i \leq 2^{k-1}$. We now make two claims that are analogous to Claim 15 and Claim 16 respectively in [8].

Claim 4.3.

$$H_k^f\big(S_0^1,S_1^1,\ldots,S_0^{k-1},S_1^{k-1}\big) \leq \frac{2^{(2^{k-1}-1)r}}{2^{2^{k-1}m}} \tag{14}$$

Claim 4.4. *Let $r < d$. Then,*

$$H_k^f\left(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\right) = 0 \tag{15}$$

We prove these Claims in the next section. Claim 4.3 simply follows from the fact that μ is a probability distribution and f is 1/-1 valued while Claim 4.4 uses the (μ, d) orthogonality of f. We now continue with the proof of the Orthogonality-Discrepancy Lemma assuming these claims. Applying them, we obtain

$$\begin{aligned}&\left(\mathrm{disc}_k^{\phi}(F_k)\right)^{2^{k-1}} \\ &\le \sum_{j=d}^{(k-1)m} 2^{(2^{k-1}-1)j} \sum_{j_1+\cdots+j_{k-1}=j} \Pr\left[r_1 = j_1 \wedge \cdots \wedge r_{k-1} = j_{k-1}\right]\end{aligned} \tag{16}$$

Substituting the value of the probability, we further obtain:

$$\begin{aligned}&\left(\mathrm{disc}_k^{\phi}(F_k)\right)^{2^{k-1}} \\ &\le \sum_{j=d}^{(k-1)m} 2^{(2^{k-1}-1)j} \sum_{j_1+\cdots+j_{k-1}=j} \binom{m}{j_1} \cdots \binom{m}{j_{k-1}} \frac{(\ell-1)^{m-j_1} \cdots (\ell-1)^{m-j_{k-1}}}{\ell^{(k-1)m}}\end{aligned} \tag{17}$$

The following simple combinatorial identity is well known:

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

Plugging this identity into (17) immediately yields (8) of the Orthogonality-Discrepancy Lemma. Recalling $\binom{(k-1)m}{j} \le \left(\frac{e(k-1)m}{j}\right)^j$, and choosing $\ell - 1 \ge 2^{2^k}(k-1)em$, we get (9).

□

4.3 Proofs of Claims

We identify the set of all assignments to boolean variables in $X = \{x_1, \ldots, x_n\}$ with the n-ary boolean cube $\{0,1\}^n$. For any $u \in \{0,1\}^{k-1}$, let Z_u represent the set of m variables indexed jointly by $S_{u_1}^1, \ldots, S_{u_{k-1}}^{k-1}$. There is precisely one variable chosen from each block of X. Denote by $Z_i[\alpha]$ the unique variable in Z_i that is in the αth block of X, for each $1 \le \alpha \le m$. Let $Z = \cup_u Z_u$. We abuse notation for the sake of clarity and use Z_u in the context of expected value calculations to also mean a uniformly chosen random assignment to the variables in the set Z_u.

Proof of Claim 4.4.

$$\begin{aligned} &H_k^f(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}) \\ &= \left| \mathbb{E}_{Z_{0^{k-1}}} f(Z_{0^{k-1}})\mu(Z_0)\ \mathbb{E}_{X-Z_{0^{k-1}}} \prod_{\substack{u\in\{0,1\}^{k-1} \\ u\neq 0}} f(Z_u)\mu(Z_u) \right| \end{aligned} \tag{18}$$

Observe that for any block α and any $u \neq 0^{k-1}$, $Z_u[\alpha] = Z_{0^{k-1}}[\alpha]$ iff for each i such that $u_i = 1$, $S_0^i[\alpha] = S_1^i[\alpha]$. Recall that r_i is the number of indices α such that $S_0^i[\alpha] = S_1^i[\alpha]$. Therefore, there are at most $r = \sum_{i=1}^{k-1} r_i$ many indices α such that $Z_u[\alpha] = Z_{0^{k-1}}[\alpha]$ for some $u \neq 0^{k-1}$. This means the inner expectation in (18) is a function that depends on at most r variables. Since f is orthogonal under μ with every polynomials of degree less than d and $r < d$, we get the desired result. □

Proof of Claim 4.3. Observe that since F_k^f is 1/-1 valued, we get the following:

$$\begin{aligned} H_k^f(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}) &\le \mathbb{E}_x \prod_{u\in\{0,1\}^{k-1}} \mu(x \leftarrow S_{u_1}^1, \ldots, S_{u_{k-1}}^{k-1}) \\ &= \mathbb{E}_{X-Z}\ \mathbb{E}_Z \prod_{u\in\{0,1\}^{k-1}} \mu(Z_u) \\ &= \mathbb{E}_{X-Z} \frac{1}{2^{|Z|}} \sum_{Z\in\{0,1\}^{|Z|}} \prod_{u\in\{0,1\}^{k-1}} \mu(Z_u) \quad (19) \\ &\le \mathbb{E}_{X-Z} \frac{1}{2^{|Z|}} \sum_{\substack{y^1,\ldots,y^{k-1} \\ \in\{0,1\}^m}} \prod_{i=1}^{k-1} \mu(y^i) \quad (20) \end{aligned}$$

where the last inequality holds because every product in the inner sum of (19) appears in the inner sum of (20). Using the fact that μ is a probability distribution, we get:

$$\text{RHS of (20)} = \mathbb{E}_{X-Z}\, \frac{1}{2^{|Z|}} \prod_{i=1}^{k-1} \sum_{y^i \in \{0,1\}^m} \mu(y^i)$$
$$= \mathbb{E}_{X-Z}\, \frac{1}{2^{|Z|}}$$
$$= \frac{1}{2^{|Z|}}.$$

We now find a lower bound on $|Z|$. Let t_u denote the Hamming weight of the string u and $\{j_1, \ldots, j_{t_u}\}$ denote the set of indices in $[k-1]$ at which u has a 1. Define

$$Y_u = \left\{ Z_u[\alpha] \mid S_1^{j_s}[\alpha] \neq S_0^{j_s}[\alpha];\ 1 \leq s \leq t_i;\ 1 \leq \alpha \leq m \right\} \tag{21}$$

The following follow from the above definition.

- $|Y_{0^{k-1}}| = m$ and $|Y_u| \geq m - \sum_{1 \leq s \leq t_i} r_{j_s} \geq m - r$ for all $u \neq 0^{k-1}$.
- $Y_u \cap Y_v = \emptyset$, for $u \neq v$. This follows from the following argument: wlog assume there is an index β where u has a one but v has a zero. Consider any block α such that $Z_u[\alpha]$ is in Y_u. It must be true that $S_1^{\beta}[\alpha] \neq S_0^{\beta}[\alpha]$. This means that $Z_u[\alpha] \neq Z_v[\alpha]$. Therefore $Z_u[\alpha]$ is not in Y_v and we are done.
- $Y := \cup_{u \in \{0,1\}^{k-1}} Y_u = Z$. This is because if $Z_u[\alpha]$ is not in Y_u then there are indices $j_1, \ldots, j_s$ where u contains a one and $S_0^{j_i}[\alpha] = S_1^{j_i}[\alpha]$. Let v be the string that contains a zero at positions $j_1, \ldots, j_s$ and at other positions, corresponds to u. Then by definition, $Z_u[\alpha] = Z_v[\alpha] \in Y_v$.

Thus, $|Z| = |Y| = \sum_u |Y_u| \geq m + \sum_{u \neq 0}(m - r) = 2^{k-1}m - (2^{k-1} - 1)r$ and the result follows. □

5 The Main Result

Before proving the main result, we borrow from Sherstov [16] a beautiful duality between approximability and orthogonality. The intuition is that if a function is at a large distance from the linear space spanned by the characters of degree less than d, then its projection on the dual space spanned by characters of degree at least d is large. More precisely,

Lemma 5.1. *Let $f : \{-1,1\}^m \to \mathbb{R}$ be given with $deg_\delta(f) = d \geq 1$. Then there exists $g : \{-1,1\}^m \to \{-1,1\}$ and a distribution μ on $\{-1,1\}^m$ such that g is (μ, d)-orthogonal and $Corr_\mu(f, g) > \delta$.*

We do not prove this Lemma but the interested reader can read its short proof in [16] which is based on an application of linear programming duality.

Theorem 5.2 (Main Theorem). *Let $f : \{0,1\}^m \to \{-1,1\}$ have δ-approximate degree d. Let $n \geq \left(2^{2^k}(k-1)e\right)^{k-1} m^k$, and $f' : \{0,1\}^n \to \{-1,1\}$ be such that $f(z) = f'(z0^{n-m})$. Then*

$$R_k^\epsilon(G_k^{f'}) \geq \frac{d}{2^{k-1}} + \log(\delta + 2\epsilon - 1). \tag{22}$$

Proof. Applying Lemma 5.1 we obtain a function g and a distribution μ such that $\mathrm{Corr}_\mu(f, g) > \delta$ and $\mathbb{E}_{x\sim\mu} g(x)\chi_S(x) = 0$ for $|S| < d$. These g and μ satisfy the conditions of Lemma 4.2, therefore we have

$$\mathrm{disc}_{k,\lambda}\left(F_k^g\right) \leq \frac{1}{2^{d/2^{k-1}}} \tag{23}$$

where λ is obtained from μ as stated in Lemma 4.2 and $\ell \geq 2^{2^k}(k-1)em$. Since $n = \ell^{k-1}m$, (23) holds for $n \geq \left(2^{2^k}(k-1)e\right)^{k-1} m^k$.

It can be easily verified that $\mathrm{Corr}_\lambda(F_k^f, F_k^g) = \mathrm{Corr}_\mu(f, g) > \delta$. Thus, by plugging the value of $\mathrm{disc}_{k,\lambda}\left(F_k^g\right)$ in (6) of the generalized discrepancy method we get

$$R_k^\epsilon(F_k^f) \geq \frac{d}{2^{k-1}} + \log(\delta + 2\epsilon - 1).$$

The desired result is obtained by applying Lemma 4.1. □

5.1 Disjointness Separates $\mathbf{BPP}_k^{CC}$ and $\mathbf{NP}_k^{CC}$

As a corollary to the main theorem, we obtain the following lower bound for the Disjointness function.

Corollary 5.3.

$$R_k^\epsilon(DISJ_k) = \Omega\left(\frac{n^{\frac{1}{2k}}}{2^{2^k} e(k-1)2^{k-1}}\right)$$

as long as $\epsilon < 1/6$.

Proof. Let $f = \mathrm{NOR}_m$ and $f' = \mathrm{NOR}_n$. We know $\deg_{1/3}(\mathrm{NOR}_m) = \Theta(\sqrt{m})$ by Theorem 2.2. Setting $n = \left(2^{2^k}(k-1)e\right)^{k-1} m^k$, and writing (22) in terms of n gives the result. □

Observe that we get the same bound for the function G_k^{OR}. It is not difficult to see that there is a $O(\log n)$ bit non-deterministic protocol for G_k^{OR} and therefore this function separates the communication complexity classes BPP_k^{CC} and NP_k^{CC} for all $k = o(\log \log n)$.

5.2 Other Symmetric Functions

Theorem 5.2 does not immediately provide strong bounds on the communication complexity of G_k^f for every symmetric f. For instance, if f is the MAJORITY function then one has to work a little more to derive strong lower bounds.

In this section, using the main result and Paturi's Theorem (Theorem 2.2), we obtain a lower bound on the communication complexity of G_k^f for each symmetric f. Let $f : \{0,1\}^n \to \{1,-1\}$ be the symmetric function induced from a predicate $D : \{0,1,\ldots,n\} \to \{1,-1\}$. We denote by G_k^D the function G_k^f. For $t \in \{0,1,\ldots,n-1\}$, define $D_t : \{0,1,\ldots,n-t\} \to \{1,-1\}$ by $D_t(i) = D(i+t)$. Observe that the communication complexity of G_k^D is at least the communication complexity of $G_k^{D_t}$.

Corollary 5.4. *Let $D : \{0,1,\ldots,n\}$ be any predicate and let $\ell_0 = \ell_0(D)$, $\ell_1 = \ell_1(D)$. Define $T : \mathbb{N} \to \mathbb{N}$ by*

$$T(n) = \left(\frac{n}{(2^{2^k}(k-1)e)^{k-1}}\right)^{\frac{1}{k}}$$

Then,

$$R_k^\epsilon(G_k^D) = \Omega\left(\Psi(\ell_0) + \frac{T(\ell_1)}{2^{k-1}}\right)$$

where

$$\Psi(\ell_0) = \min\{\Omega\big(\frac{\sqrt{T(n)\ell_0}}{2^{k-1}}\big), \Omega\big(\frac{T(n-\ell_0)}{2^{k-1}}\big)\}.$$

Proof. There are three cases to consider.
<u>Case 1:</u> Suppose $\ell_0 \leq T(n)/2$. Let $D' : \{0,1,\ldots,T(n)\} \to \{1,-1\}$ be such that for any $z \in \{0,1\}^{T(n)}$, we have $D(|z|) = D'(|z|)$. By Theorem 5.2, the complexity of G_k^D is $\Omega(d/2^{k-1})$ where $d = \deg_{1/3}(D')$. By Paturi's Theorem, $deg_{1/3}(D') \geq \sqrt{T(n)\ell_0(D')} = \sqrt{T(n)\ell_0}$ and so

$$R_k^\epsilon(G_k^D) = \Omega\big(\frac{\sqrt{T(n)\ell_0}}{2^{k-1}}\big)$$

Case 2: Suppose $T(n)/2 < \ell_0 \leq n/2$. We find a lower bound on the communication complexity of G^{D_t} where $t = \ell_0 - T(n-\ell_0)/2$. Let $D'_t : \{0, 1, \ldots, T(n-\ell_0)\} \to \{1, -1\}$ be such that $D'_t(|z|) = D_t(|z|)$. So by Theorem 5.2, the complexity of $G_k^{D_t}$ is $\Omega(d/2^{k-1})$ where d is the approximation degree of D'_t. We know

$$\begin{aligned} D'_t(T(n-\ell_0)/2) &= D_t(T(n-\ell_0)/2) \\ &= D(T(n-\ell_0)/2 + \ell_0 - T(n-\ell_0)/2) \\ &= D(\ell_0) \\ &\neq D(\ell_0 - 1) \\ &= D'_t(T(n-\ell_0)/2 - 1). \end{aligned}$$

Thus by Paturi's Theorem, $\deg_{1/3}(D'_t) \geq \sqrt{T(n-\ell_0)^2/2}$. This implies

$$R_k^\epsilon(G_k^D) = \Omega\big(\frac{T(n-\ell_0)}{2^{k-1}}\big).$$

Case 3: Suppose $\ell_0 = 0$ and $\ell_1 \neq 0$. The argument is similar to the one for Case 2. Consider D_t where $t = n - \ell_1 - T(\ell_1)/2$. Let $D'_t : \{0, 1, \ldots, T(\ell_1)\} \to \{1, -1\}$ be such that $D'_t(|z|) = D_t(|z|)$. As in case 2, one sees that $D'_t(T(\ell_1)/2) \neq D'_t(T(\ell_1)/2 + 1)$, so $deg_{1/3}(D'_t) \geq \sqrt{T(\ell_1)^2/2}$. Therefore,

$$R_k^\epsilon(G_k^D) = \Omega\big(\frac{T(\ell_1)}{2^{k-1}}\big).$$

Combining these three cases, we get the desired result. □

References

[1] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *FOCS*, pages 337–347, 1986.

[2] 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.

[3] P. Beame, M. David, T. Pitassi, and P. Woelfel. Separating deterministic from non-deterministic NOF multiparty communication complexity. In *ICALP*, pages 134–145, 2007.

[4] P. Beame, T. Pitassi, and N. Segerlind. Lower bounds for lovasz–schrijver systems and beyond follow from multiparty communication complexity. *SIAM Journal on Computing*, 37(3):845–869, 2007.

[5] 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.

[6] A. Chakrabarti. Lower bounds for multi-player pointer jumping. In *IEEE Conf. Computational Complexity*, pages 33–45, 2007.

[7] A. Chandra, M. Furst, and R. Lipton. Multi-party protocols. In *STOC*, pages 94–99, 1983.

[8] A. Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *FOCS*, 2007.

[9] F. Chung and P. Tetali. Communication complexity and quasi-randomness. *SIAM J. Discrete Math.*, 6(1):110–123, 1993.

[10] J. Ford and A. Gál. Hadamard tensors and lower bounds on multiparty communication complexity. In *ICALP*, pages 1163–1175, 2005.

[11] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[12] T. Lee and A. Shraibman. Disjointness is hard in the multi-party number in the forehead model. In *Electronic Colloquium on Computational Complexity*, number TR08-003, 2008.

[13] R. Paturi. On the degree of polynomials that approximate symmetric boolean functions. In *STOC*, pages 468–474, 1992.

[14] R. Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[15] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003.

[16] A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Electronic Colloquium on Computational Complexity*, number TR07-100. 2007.

[17] A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *STOC*, pages 294–301, 2007.

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

[19] E. Viola and A. Wigderson. One-way multi-party communication lower bound for pointer jumping with applications. In *FOCS*, pages 427–437, 2007.

[20] A. C.-C. Yao. Some complexity questions related to distributive computing. In *STOC*, pages 209–213, 1979.