Discrepancy and the Power of Bottom Fan-in in Depth-three Circuits

Arkadev Chattopadhyay*
School of Computer Science
McGill University, Montreal, Canada
achatt3@cs.mcgill.ca

November 30, 2007

Abstract

We develop a new technique of proving lower bounds for the randomized communication complexity of boolean functions in the multiparty 'Number on the Forehead' model. Our method is based on the notion of voting polynomial degree of functions and extends the Degree-Discrepancy Lemma in the recent work of Sherstov [24]. Using this we prove that depth three circuits consisting of a MAJORITY gate at the output, gates computing arbitrary symmetric function at the second layer and arbitrary gates of bounded fan-in at the base layer i.e. circuits of type MAJ $\circ$ SYMM $\circ$ ANY$_{O(1)}$ cannot simulate the circuit class AC^0 in sub-exponential size. Further, even if the fan-in of the bottom ANY gates are increased to $o(\log\log n)$, such circuits cannot simulate AC^0 in quasi-polynomial size. This is in contrast to the classical result of Yao and Beigel-Tarui that shows that such circuits, having only MAJORITY gates, can simulate the class ACC^0 in quasi-polynomial size when the bottom fan-in is increased to poly-logarithmic size.

In the second part, we simplify the arguments in the breakthrough work of Bourgain [7] for obtaining exponentially small upper bounds on the correlation between the boolean function MOD$_q$ and functions represented by polynomials of small degree over $\mathbb{Z}_m$, when $m, q \geq 2$ are co-prime integers. Our calculation also shows similarity with techniques used to estimate discrepancy of functions in the multiparty communication setting. This results in a slight improvement of the estimates of [7, 14]. It is known that such estimates imply that circuits of type MAJ $\circ$ MOD$_m$ $\circ$ AND$_{\epsilon \log n}$ cannot compute the MOD$_q$ function in sub-exponential size. It remains a major open question to determine if such circuits can simulate ACC^0 in polynomial size when the bottom fan-in is increased to poly-logarithmic size.

1 Introduction

Understanding the computational power of constant depth circuits made of MAJORITY and MOD counting gates remains a very important and challenging open problem in theoretical computer science. We do not even completely understand such circuits of depth three. It is however well known that they have surprising power. A classical result of Allender [1] shows that all functions in AC^0 (circuits using AND and OR gates of constant depth and polynomial size) can be computed by quasi-polynomial sized circuits of type MAJ $\circ$ MAJ $\circ$ MAJ$_{(\log n)^{O(1)}}$ i.e. circuits of depth three having only MAJORITY gates in which the gates at the base layer are restricted to have polylog fan-in. More surprisingly, the work of Yao [26] and Beigel-Tarui [6] shows that such circuits are powerful enough to simulate the strictly bigger class ACC^0 i.e. functions

*supported by a NSERC graduate scholarship and research grants of Prof. D. Thérien. I am grateful to F. Green, A. Sherstov, M. Szegedy and P. Tesson for several useful discussions on the topics of this paper. I thank A. Sherstov for sharing an early version of his paper [24].

computable by circuits of constant depth and poly-size that use MOD_q gates in addition to AND and OR gates, for any fixed $q > 1$.

Håstad and Goldmann [16] showed that if such depth three circuits were restricted to have sub-logarithmic fan-in at the bottom layer, then they cannot simulate ACC^0 in sub-exponential size. This left open the question whether such restricted circuits, even when they have constant fan-in at the bottom, could simulate AC^0 in quasi-polynomial size. In fact until very recently, no super-polynomial lower bounds were known on the size of depth-two circuits of type $\mathrm{MAJ} \circ \mathrm{MAJ}$ for simulating AC^0. Introducing a powerful Degree-Discrepancy Lemma to analyze two party communication games, Sherstov [24] has settled the depth two question. Håstad and Goldman, on the other hand, invoked a result of Babai, Nisan and Szegedy [4] for the stronger 'Number on the Forehead' model of multiparty communication (introduced by [10]) to show their lower bound on the size of depth three circuits computing the generalized inner product function.

The 'Number on the Forehead' model is a fascinating but poorly undertstood model of communication that is under intensive research (see [20]). Obtaining superpolylogarithmic lower bounds on the number of bits needed to compute a function f by deterministic protocols for poly-logarithmic number of players is enough to show that f is not in ACC^0. Currently no such function is known. In fact, the communication complexity of simple functions like Disjointness and Pointer Jumping (see [5, 9]), is unknown even for *three* players.

In the first part of this paper, we show for every fixed $k \geq 2$, there exists a function that is computable by AC^0 circuits in depth three and almost linear size, but requires $n^{\Omega(1)}$ communication by k-players in the (public-coin) randomized two sided error model as long as the players are required to err with probability less that $1/2 - \epsilon$ and ϵ is quasi-polynomially small. Our construction is based on the notion of the *voting polynomial degree* of boolean functions. This notion has ben recently used by Sherstov [24] and in the past for obtaining circuit lower bounds (see [3, 18, 19]) and in computational learning theory (see [17]). Let f be any boolean function (called the *base* function) on inputs of length m having voting polynomial degree d. Let $k \geq 2$ be any integer. We will create a function F_k that takes as input a string x of length somewhat larger than m, and a set of bits that *mask* every bit of x except some m bits that are left unmasked. F_k essentially computes f on the unmasked bits. More precisely, define $F_k : X \times S^1 \times \cdots \times S^{k-1} \to \{0, 1\}$, where $X \in \{0,1\}^{M^{k-1}}$ and each S^j is a m-element subset of $[M]$, in the following way: $F_k(x, S^1, \ldots, S^{k-1}) = f(x_{i_1^1,\ldots,i_1^{k-1}}, \ldots, x_{i_m^1,\ldots,i_m^{k-1}})$, where $S^j = \{i_1^j, \ldots, i_m^j\}$. We partition the inputs of F_k among the k-players in the following way: Player 1's forehead is assigned X and each of the other $k-1$ foreheads receives a distinct set S^i. Let the k-party randomized communication complexity of a function f with error probabilty $1/2 - \epsilon$ (in the two-sided error model) be denoted by $R_k^\epsilon(f)$. We show the following:

Theorem 1 *Let f, defined on inputs of length m, have voting degree d. For any $k \geq 2$, define F_k using f as before on inputs of length $n = O(M^{k-1})$, where $M \geq 2^k(k-1)em^2$. Then, $R_k^\epsilon(F_k) = \Omega(d^{1/2^{k-1}} + \log \epsilon)$.*

We prove Theorem 1 by developing a new lower bound technique for the multiparty model that should be of independent interest. The main ingredient of our technique is the following extension of Sherstov's Degree-Discrepancy Lemma:

Lemma 2 (Multiparty Degree-Discrepancy Lemma) *Let $f : \{-1,1\}^m \to \{-1,1\}$ have voting polynomial degree d. Then for any $k \geq 2$, there exists a probability distribution λ such that for $M \geq m$,*

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

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

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

Here $\mathrm{disc}_{k,\lambda}(F_k)$ denotes the discrepancy of F_k over k-cylinder intersections under the input distribution λ.

By considering a simple base function that was used by [24], we show that our k-wise masked function F_{k+1} has $(n^{\Omega(1)})$ k-party complexity whenever k is a constant. On the other hand, it is simple to verify that F_{k+1} is in AC^0. It is the first example of a function in AC^0 that is hard for randomized mulitparty protocols. Let ANY represent an arbitrary gate and SYMM represent a gate that computes an arbitrary symmetric function of its inputs. An established argument of Hastad and Goldmann [16] can then be used to derive the following circuit consequence:

Corollary 3 *Circuits of type MAJ$\circ$SYMM$\circ$ANY$_k$ need size $2^{\Omega(n^{1/6k2^k})}$ to simulate depth-three AC^0. Specifically, if k is a constant ($o(\log\log n)$) then such circuits cannot simulate AC^0 in sub-exponential (quasi-polynomial) size.*

In particular, the above shows that Allender's classic construction to simulate AC^0 is reasonably close to being optimal. In fact, Allender's original construction shows that qpoly size circuits of type $\mathrm{MAJ}\circ\mathrm{MOD}_m\circ\mathrm{AND}_{(\log n)^{O(1)}}$ can simulate $\mathrm{ACC}^0[p^r]$ (i.e. circuits with MOD_{p^r} gates in addition to AND/OR gates), for every prime p that divides m and any fixed r. A long line of research (see for example [8, 12, 13, 2]) seeks to show that such depth three circuits cannot simulate ACC^0 in quasipoly size. The so called ϵ-discriminator lemma of Hajnal et al.[15] implies that obtaining an exponentially small upper bound on the correlation between a function f and and any boolean function that is represented by a polynomial of poly-logarithmic degree over $\mathbb{Z}_m$, is enough to prove that f cannot be computed in sub-exponential size by such depth three circuits. It is commonly believed that the simple function MOD_q has small correlation with such low degree polynomials over $\mathbb{Z}_m$, if m and q are co-prime.

In the second part of the paper, we simplify Bourgain's breakthrough method [7, 14] of estimating the correlation between polynomials of degree d over $\mathbb{Z}_m$ and MOD_q when $(m,q) = 1$. We argue that the notion of discrepancy, suitably modified, can be used conveniently to obtain this estimate. This approach also points out the similarities between the techniques used for estimating cylindrical discrepancy in the communication setting and the techniques used for obtaining correlation. Interestingly, our estimates for correlation are slightly better than previous estimates of [7, 14]. For the special case of $m = 2$, they match the recent bounds obtained by Viola and Wigderson[25]. It is not known if techniques of [25], based on Gower's norm, can be extended to all m.

2 Basic Notions

In the k-party 'Number on the Forehead' model of communication, k players wish to collaboratively compute a function f on n input bits. The input bits are partitioned into k sets $Y_1, \ldots, Y_k \subseteq [n]$. Each player P_i knows the value of all the input bits *except* the ones in Y_i that are written on his own forehead. In the deterministic model, players communicate (broadcast) bits according to a fixed protocol by writing them on a public blackboard. The protocol specifies whose turn it is to speak and what a player communicates is entirely determined by the communication history until that point and what the player sees written on others' forehead. The boolean output of the protocol is just a function of the communication history at its termination. The cost of a protocol is the number of bits that players communicate for the worst case input. The deterministic k-party communication complexity of f, denoted by $D_k(f)$ is the cost of the best k-party protocol for f.

In the (*public coin*) randomized model, players flip some coins and randomly select a deterministic protocol. Then they follow the deterministic protocol. Additionally, players are now allowed to err. This means that some of the protocols that players choose may not produce the correct output for all input instance. The cost of a randomized protocol is simply the maximum number of bits communicated by the players over all possible coin flips and over all possible input instances. The k-party randomized communication complexity of f with error $1/2 - \epsilon$, denoted by $R_k^\epsilon(f)$ is the cost of the best protocol $\mathcal{P}$ that computes f with error at most $1/2 - \epsilon$, i.e. $\Pr[\mathcal{P}(Y_1, \ldots, Y_k) \neq f(Y_1, \ldots, Y_k)] \leq 1/2 - \epsilon$ for all input assignments $Y_1, \ldots, Y_k$.

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 $\{-1,1\}^{Y_1 \times \cdots \times Y_k}$ with the property that membership in S is independent of the ith co-ordinate. 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. Equivalently, every cylinder-intersection can be viewed as a function $\phi : \{-1,1\}^{Y_1 \times \cdots \times Y_n} \to \{0,1\}$, such that it can be factored as $\phi = \phi^1 \times \cdots \times \phi^k$, where $\phi^i(x_1, \ldots, x_i, \ldots, x_k) = \phi^i(x_1, \ldots, x_i', \ldots, x_k)$ for all $x_1, \ldots, x_k$ and x_i'.

An important measure, defined on boolean functions, is its *discrepancy*. With respect to any probability distribution μ over $\{-1,1\}^{Y_1 \times \cdots \times Y_k}$ and cylinder intersection ϕ, define

$$\begin{aligned}
&\mathrm{disc}_{k,\mu}^{\phi}(f) = \\
&\Big| \Pr_\mu \big[f(Y_1, \ldots, Y_k) = 1 \wedge \phi(Y_1, \ldots, Y_k) = 1\big] - \\
&\Pr_\mu \big[f(Y_1, \ldots, Y_k) = -1 \wedge \phi(Y_1, \ldots, Y_k) = 1\big] \Big|
\end{aligned} \tag{3}$$

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

$$\begin{aligned}
&\mathrm{disc}_{k,\mu}^{\phi}(f) = \\
&\Big| \sum_{Y_1, \ldots, Y_k} f(Y_1, \ldots, Y_k) \phi(Y_1, \ldots, Y_k) \mu(Y_1, \ldots, Y_k) \Big|
\end{aligned} \tag{4}$$

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

Theorem 4 (see [4, 20]) *Let $0 < \epsilon < 1/2$ be any real and $k \geq 2$ be any integer. For every boolean function f and distribution μ on inputs from $Y_1 \times \cdots \times Y_k$,*

$$R_k^\epsilon(f) \geq \log\left(\frac{2\epsilon}{disc_{k,\mu}(f)}\right) \tag{5}$$

In the first part, we will assume boolean functions are defined from $\{-1,1\}^n$ into $\{-1,1\}$. For any $S \subseteq [n]$, let χ_S represent the multilinear monomial function $\chi_S(x) = \prod_{i \in S} x_i$. Consider a polynomial P over the reals i.e. $P = \sum_{S \subseteq [n]} a_S \chi_S$, where the coefficients a_S are real numbers. Then P is a voting representation of a boolean function f if $f(x) = \mathrm{sign}\big(P(x)\big)$. For example, polynomials $P_1(x) = x_1 + \cdots + x_n$ and $P_2(x) = \prod_{i=1}^n x_i$ voting represent MAJORITY and PARITY respectively. It is not hard to verify that all boolean functions can be voting represented by some polynomial. The degree of a representation is simply the degree of the polynomial P involved i.e. the largest integer $k \leq n$ such that there exists a set S of size k for which the coefficient a_S is non-zero. Thus, in our examples before, MAJORITY has a linear representation and that of PARITY was n. The voting degree of a function f, denoted by $\deg(f)$, is the minimum degree over all possible voting representations of f. [3, 22] are good sources to read about some basic properties of voting representations. We state below the key result that we need here:

Theorem 5 (see [24]) *For any boolean function $f : \{-1,1\}^n \to \{-1,1\}$, precisely one of the following holds:*

- *$deg(f) \le d$.*
- *there exists a distribution μ over $\{-1,1\}^n$, such that for all $|S| \le d$, $\mathbf{E}_{x\sim\mu} f(x)\chi_S(x) = 0$.*

In particular, this means that if $\deg(f) \ge d$, then for any function g that depends on at most $d-1$ variables, $\mathbf{E}_{x\sim\mu} f(x)g(x) = 0$.

A related measure on a pair of boolean functions g and f, called *correlation* and denoted by $\mathrm{Corr}(g,f)$, was defined by [15]. This measure can be defined w.r.t any distribution over the cube, but we will be solely interested in the uniform distribution for discussing correlation in this paper. Let $A \subseteq f^{-1}(1)$ and $B \subseteq f^{-1}(0)$ be two subsets of the cube. Then,

$$\begin{aligned}&\mathrm{Corr}_{A,B}(g,f) = \\ &\big|\Pr_x[g(x)=1|x\in A] - \Pr_x[g(x)=1|x\in B]\big|\end{aligned} \tag{6}$$

In the literature, g is said to ϵ-discriminate f, w.r.t. sets A, B if $\mathrm{Corr}_{A,B}(g,f) \ge \epsilon$. The usefulness of this measure in proving circuit lower bounds comes from the following connection made by [15]:

Lemma 6 (Discriminator Lemma) *Consider a circuit C with a MAJORITY gate at its output and s arbitrary sub-circuits, $C_1,\dots,C_s$ feeding into it. If C computes the function f, then for every $A \subseteq f^{-1}(1)$, $B \subseteq f^{-1}(0)$, there exists a sub-circuit C_i that $1/s$-discriminates f w.r.t A, B.*

3 Multiparty Degree-Discrepancy Lemma

For the sake of exposition, we will prove Lemma 2 (stated in Introduction) for the case of three players. The argument for the general case of k-players proceeds in an identical fashion and is given in the Appendix.

Let boolean function f, defined over m input bits, have voting degree d. Then, let μ be the distribution guaranteed to exist from Theorem 5 so that $\mathbf{E}_{x\sim\mu} f(x)g(x) = 0$ for any g that depends on less than d variables. The function that we form out of our 'base' function f is $F_3 : \{-1,1\}^{M^2} \times \binom{\{1,\dots,M\}}{m} \times \binom{\{1,\dots,M\}}{m} \to \{-1,1\}$, with $F_3(x, S^1, S^2) = f(x_{i_1,j_1},\dots,x_{i_m,j_m})$ where $S^1 = \{i_1,\dots,i_m\}$, $S^2 = \{j_1,\dots,j_m\}$ are each m-element subsets of $[M]$. We consider the partition in which Players 1,2 and 3 get respectively x, S^1 and S^2 written on their foreheads. The probability distibution λ that we consider on the set of inputs is derived out of μ as follows: $\lambda(x,S^1,S^2) = \frac{\mu_{S^1,S^2}(x)}{\binom{M}{m}^2 2^{M^2-m}}$, where $\mu_{S^1,S^2}(x) = \mu(x_{i_1,j_1},\dots,x_{i_m,j_m})$. It is not hard to see that the denominator in the expression of λ is just the right normalizing factor. Thus, the discrepancy of any cylinder intersection $\phi = \phi^1(x,S^1)\phi^2(x,S^2)\phi^3(S^1,S^2)$ w.r.t λ can be represented as follows (using (4)):

$$\begin{aligned}&\mathrm{disc}_3^{\phi}(F_3) = \\ &\Big|\sum_{x,S^1,S^2} F_3(x,S^1,S^2)\phi(x,S^1,S^2)\lambda(x,S^1,S^2)\Big|\end{aligned} \tag{7}$$

Using the definition of λ, we change over to the more convenient expected value notation, with (x, S^1, S^2) *uniformly distributed* over $\{-1,1\}^{M^2} \times \binom{[M]}{m}^2$:

$$\mathrm{disc}_3^{\phi}(F_3) =$$
$$2^m\left|\mathbf{E}_{x,S^1,S^2}F_3(x,S^1,S^2)\phi(x,S^1,S^2)\mu_{S^1,S^2}(x)\right| \tag{8}$$

Clearly, using the fact that ϕ^1 is 0/1 valued we get RHS of (8) $\leq 2^m\mathbf{E}_{x,S^1}Z$ where,

$$Z = \left|\mathbf{E}_{S^2}\left[F_3(x,S^1,S^2)\phi^2(x,S^2)\phi^3(S^1,S^2)\mu_{S^1,S^2}(x)\right]\right| \tag{9}$$

As in [4], we use Cauchy-Schwartz inequality i.e. $(\mathbf{E}Z)^2 \leq \mathbf{E}(Z)^2$. Recall that $(\mathbf{E}_z f(z))^2 = \mathbf{E}_{z_0,z_1}f(z_0)f(z_1)$, where z_1, z_2 are independent and identical copies of z. Noting that ϕ^2 is 0/1 valued we get:

$$(\mathrm{disc}_3^{\phi}(F_3))^2 \leq 2^{2m}\mathbf{E}_{x,S_0^2,S_1^2}\mathcal{U} \tag{10}$$

where,

$$\mathcal{U} = \left|\mathbf{E}_{S^1}\left[\prod_{\ell\in\{0,1\}} F_3(x,S^1,S_\ell^2)\mu_{S^1,S_\ell^2}(x)\phi^3(S^1,S_\ell^2)\right]\right| \tag{11}$$

and S_0^2, S_1^2 are independent and identically distributed as S^2. Using another round of Cauchy-Schwartz and very similar argument, we finally obtain:

$$(\mathrm{disc}_3^{\phi}(F_3))^4 \leq 2^{4m}\mathbf{E}_{S_0^1,S_1^1,S_0^2,S_1^2}\mathcal{V} \tag{12}$$

with,

$$\mathcal{V} = \left|\mathbf{E}_x\left[\prod_{\ell,j\in\{0,1\}} F_3(x,S_j^1,S_\ell^2)\mu_{S_j^1,S_\ell^2}(x)\right]\right| \tag{13}$$

Consider any fixed $S_0^1, S_1^1, S_0^2, S_1^2$. The following claim ties in the voting polynomial degree d of f to our argument. Let $r = \max\{|S_0^1\cap S_1^1|,\ |S_0^2\cap S_1^2|\}$. Then,

Claim 7 *If r is smaller than the voting degree d of f, the following holds:*

$$\mathbf{E}_x\left[\prod_{i,j\in\{0,1\}} F_3(x,S_i^1,S_j^2)\mu_{S_i^1,S_j^2}(x)\right] = 0 \tag{14}$$

Proof: Wlog, let us assume that $r = |S_0^1\cap S_1^1|$, $t = |S_0^2\cap S_1^2|$, with $t \leq r$. Further, again wlog we assume $S_0^1 = S_0^2 = \{1,\ldots,m\}$, $S_1^1 = \{1,\ldots,r,m+1,\ldots,2m-r\}$ and $S_1^2 = \{1,\ldots,t,m+1,\ldots,2m-t\}$. We will expand the product in the LHS of (14) in a convenient way. First note that $F_3(x,S_i^1,S_j^2)$ depends on precisely m of the variables in x for each i,j. We will call this set Z_{ij}. We will treat $Z_{00} = \{x_{1,1},\cdots,x_{m,m}\}$ separately for reasons that will become clear shortly.

$$Z_{01} = \{x_{1,1},\cdots,x_{t,t},x_{t+1,m+1},\cdots,x_{m,2m-t}\}$$
$$Y_{01} = \{x_{t+1,m+1},\cdots,x_{m,2m-t}\}$$

$$Z_{10} = \{x_{1,1}, \cdots, x_{r,r}, x_{m+1,r+1}, \cdots, x_{2m-r,m}\}$$
$$Y_{10} = \{x_{m+1,r+1}, \cdots, x_{2m-r,m}\}$$

$$\begin{aligned} Z_{11} = & \quad \{x_{1,1}, \cdots, x_{t,t}, x_{t+1,m+1}, \cdots, x_{r,m+r-t} \\ & \quad , x_{m+1,m+r-t+1}, \cdots, x_{2m-r,2m-t}\} \\ Y_{11} = & \quad \{x_{r+1,m+r-t+1}, \cdots, x_{2m-r,2m-t}\} \end{aligned}$$

Let, $g_{11} = \mathbf{E}_{Y_11} f(Z_{11})\mu(Z_{11})$, $g_{10} = \mathbf{E}_{Y_{10}} g_{11} f(Z_{10})\mu(Z_{10})$, and $g_{01} = \mathbf{E}_{Y_{01}} g_{10} f(Z_{01})\mu(Z_{01})$. Then, clearly g_{01} is just a function of the r variables $x_{1,1} \ldots, x_{r,r}$. It further gets verified easily that

$$\begin{aligned} & \text{LHS of (14)} = \\ & \mathbf{E}_{x_{1,1},\cdots,x_{m,m}} \quad \big[f(x_{1,1}, \cdots, x_{m,m})\mu(x_{1,1}, \cdots, x_{m,m}) \\ & \qquad \cdot g_{01}\big(x_{1,1}, \cdots, x_{r,r}\big)\big] \end{aligned}$$

Now invoking the property of μ from Theorem 5, we immediately see that (14) evaluates to zero. ∎

We make another claim whose simple proof, based on the fact that μ is a probability distribution, is given in the Appendix[1]:

Claim 8 *For all fixed $S_0^1, S^1, 1, S_0^2, S_1^2$ and $r = max\{|S_0^1 \cap S_1^1|, |S_0^2 \cap S_1^2|\}$,*

$$\left|\mathbf{E}_x\left[\prod_{i,j\in\{0,1\}} F_3(x, S_i^1, S_j^2)\mu_{S_i^1,S_j^2}(x)\right]\right| \leq \frac{2^{3r}}{2^{4m}} \tag{15}$$

Claim 7 and Claim 8 shows that the inner expectation in (12) can be upper bounded by a function of two numbers, namely $|S_0^i \cap S_1^i|$, for $i = 1, 2$. Using the definition for the outer expectation, we obtain:

$$\begin{aligned} & (\text{disc}_3^{\phi}(F_3))^4 \leq \\ & \sum_{j=d}^{m} 2^{3j} \sum_{j_1+j_2=j} \Pr\left[|S_0^1 \cap S_1^1| = j_1 \wedge |S_0^2 \cap S_1^2| = j_2\right] \end{aligned} \tag{16}$$

Recalling the fact that $S_0^1, S_1^1, S_0^2, S_1^2$ are being chosen independently, we have:

$$\begin{aligned} & \text{RHS of (16)} = \\ & \sum_{j=d}^{m} 2^{3j} \sum_{j_1+j_2=j} \binom{m}{j_1}\binom{m}{j_2} \frac{\binom{M-m}{m-j_1}\binom{M-m}{m-j_1}}{\binom{M}{m}^2} \end{aligned} \tag{17}$$

We recall the following fact about binomial coefficients:

Fact 9 *For every $M \geq m$,*

$$\frac{\binom{M-m}{m-k}}{\binom{M}{m}} \leq \left(\frac{m}{M}\right)^k \tag{18}$$

[1] in the Appendix, we directly prove Claim 15 that is a generalization of Claim 8 to k-players.

Using (18) with the combinatorial identity $\sum_{j_1+j_2=j}\binom{m}{j_1}\binom{m}{j_2}=\binom{2m}{j}$, (17) yields

$$\text{RHS of (17)} \leq \sum_{j=d}^{m} 2^{3j}\binom{2m}{j}\left(\frac{m}{M}\right)^{j} \tag{19}$$

Using $\binom{2m}{j} \leq (\frac{2em}{j})^j$, one sees that for $M \geq 32em^2$ and for $d > 2$, the RHS of (19) is less than $1/2^d$. Thus, $\mathrm{disc}_3^{\phi} \leq 1/2^{d/4}$, for every cylinder intersection ϕ proving the Multiparty Degree-Discrepancy Lemma for three players.

A simple combination of Theorem 4 with the Multiparty Degree-Discrepancy Lemma proves the bound on randomized communication complexity in Theorem 1.

4 Circuit consequences

Just as in [24], our base function f will be the following function, studied first in [21]: $\mathrm{MP}(x) = \vee_{i=1}^{\ell} \wedge_{j=1}^{4\ell^2} x_{i,j}$. [21] shows that the voting polynomial degree of MP, defined on $4\ell^3$ variables, is ℓ. We choose $m = 4\ell^3$ and our base function $f(x) = \mathrm{MP}(x)$. Then for each $k \geq 2$, we create our k-wise masked function F_k from MP according to the masking rules prescribed by the Multiparty Degree-Discrepancy Lemma in Section 1. We can view the domain of function F_k, for any $k \geq 2$ as $\{-1,1\}^{M^{k-1}} \times \{-1,1\}^{(k-1)m\log M}$, by naturally encoding each of the $k-1$ m-element subsets of $[M]$ in the following way: each element of a subset is encoded by binary strings having $\log M$ bits. Note that several inputs in this encoding may be illegal as in a legal input we do not allow repetitions of an element in a subset. However, the output of F_k on illegal inputs is immaterial i.e. we describe a depth-three AC^0 circuit C to compute F_k correctly on the legal inputs and we conveniently define the value of F_k on each illegal input to be the same as the output of C on that input.

Consider the decoding function $U : \{-1,1\}^{M^{k-1}} \times \{-1,1\}^{(k-1)\log M}$ that on input (x,y) interprets y to be a set of $k-1$ positive integers from $[M]$ and then outputs the bit of x corresponding to this set. It is not hard to verify that U could be computed by a depth-two AND $\circ$ OR circuit of size M^{k-1}. Now, on a legal input $F_k(x,y) = \mathrm{MP}(U(x,y_1),\ldots,U(x,y_m))$, where each y_i is the binary string of length $(k-1)\log M$ encoding the ith element of each set $S^1,\ldots,S^{k-1}$. By definition, the MP function can be computed by depth-two OR $\circ$ AND circuits of size m. This directly gives a depth-four circuit to compute F_k on legal inputs. Collapsing the two middle layers of AND gates finally yields a depth-three circuit of size mM^{k-1}. Summarizing,

Fact 10 (follows from [24]) *The function $F_k : \{-1,1\}^{M^{k-1}} \times \{-1,1\}^{(k-1)m\log M} \to \{-1,1\}$ is in depth-three AC^0.*

We recall here an established connection between randomized communication complexity of a function f and the size of depth-three circuits needed to compute f.

Fact 11 (see [16]) *If f is computed by a circuit of type MAJ$\circ$SYMM$\circ$ANY$_k$, of size s, then $R_{k+1}^{1/2-1/2s}(f) \leq k\log s$.*

Proof:[Of Corollary 3] The $k+1$-party randomized communication complexity of F_{k+1} with error $1/2-\epsilon$, by Theorem 1, is at least $d^{1/2^k} + \log\epsilon$. Here $d = \ell$, $m = 4\ell^3$, $M = 2^{k+1}kem^2$ and $n = M^k$. Combining this information, we obtain that $R_{k+1}^{\epsilon}(F_{k+1}) \geq (1/\alpha)n^{1/(6k2^k)} - \log\left(\frac{1}{2\epsilon}\right)$, where $\alpha = (4\sqrt{2^{k+1}ke})^{1/3\cdot 2^k}$. Let F_{k+1} be computed by a circuit of type MAJ $\circ$ SYMM $\circ$ ANY$_k$ with size s. Then, applying Fact 11 on the randomized complexity of F_{k+1}, we get that

$$(1/\alpha)n^{1/6k2^k} - \log\, s \leq k. \log\, s \tag{20}$$

Note that $(1/\alpha) \to 1$, with k quite rapidly. Thus, $s = 2^{\Omega(n^{1/6k2^k})}$. Corollary 3 follows quite easily from this. ∎

5 Correlation

Let P be any multi-linear polynomial of degree d over $\mathbb{Z}_m$ in n variables. For any $q \geq 2$, the boolean function MOD_q is defined to be 1 iff the sum of the input bits is non-zero nodulo q. Let L_q be the linear polynomial $x_1 + \cdots + x_n$ evaluated over $\mathbb{Z}_q$. Let $f : \{0,1\}^n \to \mathbb{Z}_q$. Consider a distribution μ such that f is almost balanced under μ i.e. $\Pr_x[f(x) = b] = 1/q + 2^{-\Omega(n)}$. For example, L_q is almost balanced under the uniform distribution for every q. Let the *mod-m polynomial discrepnacy* of f w.r.t. polynomial P and $a \in \mathbb{Z}_m$ under μ, denoted by $\mathrm{Pdisc}^{P,a}_{\mu,m}(f)$, be the following:

$$\begin{aligned}\mathrm{Pdisc}^{P,a}_{\mu,m}(f) \;=\; \max_{b\in\mathbb{Z}_m}\Bigg| &\Pr_{x\sim\mu}\big[f(x) = b \wedge P(x) = a\big] \\ &- (1/q)\Pr_{x\sim\mu}\big[P(x) = a\big]\Bigg|\end{aligned} \tag{21}$$

The Mod-m Polynomial Discrepancy of f under μ for degree d, denoted by $\mathrm{Pdisc}_{d,\mu,m}(f)$ is simply $\max\{\mathrm{Pdisc}^{P,a}_{\mu,m}(f) | \mathrm{d}$ $d, a \in \mathbb{Z}_m\}$. In this paper, for polynomial discrepancy the default distribution is uniform. Hence we will drop the subscript denoting the distribution explicitly.

Our main technical lemma, in this section, is the following :

Lemma 12 (Polynomial Discrepancy Lemma) *Let $m, q > 1$ be integers that are co-prime and $d \geq 1$. Then, there exists a constant $\beta = \beta(m,q)$, such that the following holds:*

$$Pdisc_{d,m}(L_q) \leq exp(-\frac{\beta n}{(m2^{m-1})^d}) \tag{22}$$

In words, (22) shows that $P^{-1}(a)$, for each a, looks *uniform* to a MOD_q counter i.e. every $L_q^{-1}(b)$ is almost equally represented in the set, provided the size of the set is large compared to the size of the cube. We identify the similarities between the calculation of polynomial discrepancy of the L_q function and the method used by [4] to estimate the cylindrical discrepancy for the generalized inner product function. In both estimates, the key technical ingredient is to raise the sum in question to its appropriate power.

This easily leads to an upper bound of $\exp(-\Omega(n/(m2^{m-1})^d))$ on correlation between the MOD_q function and functions represented by polynomials of degree d over $\mathbb{Z}_m$. In particular, this implies the bound of $\exp(-\Omega(n/4^d))$ for the special case of $m = 2$ that was first reported in the recent work of [25]. Let $e_m(y)$ denote $\exp(-2\pi jy/m)$, where j is the square root of -1. Recall the elementary identity for roots of unity: $\sum_{a=0}^{m-1} e_m(ay) = 1$ if y is a multiple of m and is zero otherwise. We start by estimating, using complex roots of unity, the quantity $\Pr_x[P(x) = a \wedge L_q(x) = b]$ for any polynomial P over $\mathbb{Z}_m$ and for any $a \in \mathbb{Z}_m, b \in \mathbb{Z}_q$ as follows:

$$
\begin{aligned}
&\Pr_x\big[P(x) = a \wedge L_q(x) = b\big] \;= \\
&\mathbf{E}_x\Bigg[\bigg(\frac{1}{m}\sum_{\alpha=0}^{m-1} e_m\big(\alpha(P(x)-a)\big)\bigg) \\
&\times\bigg(\frac{1}{q}\sum_{\beta=0}^{q-1} e_q\big(\beta(x_1+\cdots+x_n-b)\big)\bigg)\Bigg] \qquad (23)
\end{aligned}
$$

Expanding the sum inside the second multiplicand and treating the case of $\beta = 0$ separately, one gets

$$
\begin{aligned}
(23) \;=\; &\frac{1}{q}\mathbf{E}_x\bigg[\frac{1}{m}\sum_{\alpha=0}^{m-1} e_m\big(\alpha(P(x)-a)\big)\bigg] \\
+\; &\frac{1}{mq}\sum_{\alpha\in[m],\beta\in[q]-\{0\}} S_n^{m,q}(\alpha,\beta,P)e_m(-a\alpha)e_q(-b\beta) \qquad (24)
\end{aligned}
$$

where,

$$
S_n^{m,q}(\alpha,\beta,P) = \mathbf{E}_x\bigg[e_m(\alpha P(x))\cdot e_q\big(\beta(x_1+\cdots+x_n)\big)\bigg] \qquad (25)
$$

Observing that the first term in (24) is simply $(1/q)\Pr[P(x) = a]$ and $|e_m(-a\alpha)| = |e_q(-b\beta)| = 1$, we get :

$$
\mathrm{Pdisc}_m^{P,a}(L_q) \leq \frac{1}{mq}\sum_{\alpha\in[m],\beta\in[q]-\{0\}} |S_n^{m,q}(\alpha,\beta,P)| \qquad (26)
$$

It is simple to verify that the Polynomial Discrepancy Lemma gets established by the bound on $|S_n^{m,q}(\alpha,\beta,P)|$ provided below.

Lemma 13 *For each pair of co-prime integers $m, q > 1$ there exists a constant $\beta = \beta(q)$ such that for every polynomial P of degree $d > 0$ over $\mathbb{Z}_m$ and numbers $\alpha \in [m]$, $\beta \in [q] - \{0\}$, the following holds :*

$$
|S_n^{m,q}(\alpha,\beta,P)| \leq exp\bigg(-\frac{\beta n}{(m2^{m-1})^d}\bigg) \qquad (27)
$$

Before we begin our formal calculations, we remind the reader that a slightly weaker estimate of $|S_n^{m,q}(\alpha,\beta,P)|$ was first obtained in [7, 14]. The case when P is a linear polynomial was essentially dealt with in [8].

Observe that the quantity $S_n^{m,q}$, defined in (25), looks very similar to the sum that was obtained in Babai, Nisan and Szegedy [4] to calculate the discrepancy of GIP. There, they were interested in bounding discrepancy of GIP w.r.t k-cylinder intersections. Here, we are interested in bounding the discrepancy of L_q w.r.t to a set that is the image of a polynomial. The key idea, introduced in [4], is that squaring the sum is effective in dealing with cylinder intersections. This is something that we adapted to our proof of the Degree-Discrepancy Lemma in the previous section. Here, the analogue of the BNS trick will be to raise the sum in (25) to its mth power.

In order to further explain the intuition behind our proof of Lemma 13, we introduce some definitions and notations. Let $f : \{0,1\}^n \to \mathbb{Z}_m$ be any function. Consider any set $I \subseteq [n]$. Note that each binary vector v of length $|I|$ can be thought of as a partial assignment to the input variables of f by assigning v to the variables in I in a natural way. Let $f^{I(v)}$ be the subfunction of f on variables not indexed in I induced by the partial assignment v to variables indexed in I. For any sequence $Y = \{y_1, \ldots, y_t\}$ having t boolean vectors from $\{0,1\}^n$, let f_Y be the function defined by $f_Y(x) = f(x) + \sum_{i=1}^{t} f(x \oplus y_i)$, where the sum is taken in $\mathbb{Z}_m$. Let $I[Y] \subseteq [n]$ be the set of those indices on which every vector in Y is zero and $J[Y]$ be just the complement of $I[Y]$. Then, the following observation will be very useful in our calculation :

Observation 14 *Let P be a polynomial of degree d in n variables over Z_m for any $m > 1$. Then, for each sequence Y of $m-1$ boolean vectors in $\{0,1\}^n$, the polynomial $P_Y^{J[Y](v)}$ is a polynomial of degree $d-1$ in variables from $I[Y]$ for each vector $v \in \{0,1\}^{|J[Y]|}$.*

A point worth mentioning is that, P_Y behaves almost like a *discrete derivative* of polynomial P. In fact, if $m = 2$, then this operation coincides with the notion of discrete derivative as used in the work of [11, 23].

Proof Sketch:[of Lemma 13] We drop the superscript from $S_n^{m,q}$ to avoid clutter in the following discussion. We shall induce on the degree d of the polynomial. Our IH is that there exists a positive real constant $\mu_{d-1} < 1$ such that for all polynomials R of degree at most $d-1$ and for all $n \geq 0$ we have $|S_n(\alpha, \beta, R)| \leq 2^n \mu_{d-1}^n$. The base case of $d = 0$ is easily verified and is dealt with in earlier works on correlation. Note that μ_0 depends only on q. Our inductive step will yield a relationship between μ_{d-1} and μ_d that will also give us our desired explicit bound of (27).

As in [7, 14], we raise S_n to its mth power. Our point of departure from the earlier techniques, is to write $(S_n)^m$ in a different way.

$$\begin{aligned}(S_n)^m =&\mathbf{E}_{y^1,\ldots,y^{m-1}}\mathbf{E}_x\left[e_m\left(P(x) + \sum_{j=1}^{m-1} P(x \oplus y^j)\right)\right.\\ &\left.\times\, e_q\left(\sum_{i=1}^{n} x_i + \sum_{k=1}^{m-1}\sum_{i=1}^{n}(x_i \oplus y_i^k)\right)\right] \end{aligned} \tag{28}$$

Let Y be the sequence of length $m-1$ formed by a given set of vectors $y^1, \ldots, y^{m-1}$. We denote by u and v respectively the projection of x to $I[Y]$ and $J[Y]$. Let n_I and n_J be the cardinality of $I[Y]$ and $J[Y]$ (note that $n_I + n_J = n$) . Then, one can verify

$$\begin{aligned}&(28) = \\ &\mathbf{E}_{y^1,\ldots,y^{m-1}} \quad \mathbf{E}_v\Big[e_m\big(Q^{y^1,\ldots,y^{m-1}}(v)\big)e_q(n_J)\\ &\qquad\cdot\,\mathbf{E}_u e_m\big(P_Y^{I[Y](v)}(u)\big)e_q\big(m\sum_{i=1}^{n_I} u_i\big)\Big]\end{aligned} \tag{29}$$

where $Q^{y^1,\ldots,y^{m-1}}$ is some polynomial that is determined by $y^1, \ldots, y^{m-1}$ and polynomial P.

The key thing to note is that Observation 14 implies $P_Y^{I[Y](v)}$ to be a polynomial of degree at most $d-1$ over u for every sequence $Y = y^1, \ldots, y^{m-1}$ and every vector v. Hence, the inside sum of (29) over the variable u can be estimated using our inductive hypothesis. Note that raising to the mth power in (28) has achieved a degree reduction of the polynomial in a manner that is very reminiscent of how [4] does dimension reduction of cylinder intersections in the proof of their Lemma 2.5.

The rest of the calulation proceeds exactly as in Green et. al. [14], which again is very similar to the series of final steps in the proof of Lemma 2.5 in [4]. We repeat them in the Appendix for the sake of self-containment. ∎

Consider $A = L_q^{-1}(1)$ and $B = L_q^{-1}(0)$. Then using the estimate on the mod-m polynomial discrepancy of L_q, it gets easily verified that for every circuit C of type $\mathrm{MOD}_m \circ \mathrm{AND}_d$,

$$\mathrm{Corr}_{A,B}(C, \mathrm{MOD}_q) \leq \exp\left(-\frac{\beta n}{(m2^{m-1})^d} \right) \tag{30}$$

Combining the Discriminator Lemma (Lemma 6) with (30) leads to super-polynomial lower bounds on the fan-in of the output gate in circuits of type $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_d$ for computing MOD_q, if m, q are co-prime and $d = \epsilon \log n$ for some constant $\epsilon > 0$.

References

[1] E. Allender. A note on the power of threshold circuits. In *FOCS*, pages 580–584, 1989.

[2] N. Alon and R. Beigel. Lower bounds for approximations by low degree polynomials over $\mathbb{Z}_m$. In *IEEE Conf. on Computational Complexity*, pages 184–187, 2001.

[3] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):1–14, 1994.

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

[5] P. Beame, T. Pitassi, and N. Segerlind. Lower bounds for lovász-schrijver systems and beyond follow from multiparty communication complexity. In *ICALP*, pages 1176–1188, 2005.

[6] R. Beigel and J. Tarui. On ACC. *Computational Complexity*, 4:350–356, 1994.

[7] J. Bourgain. Estimates of certain exponential sums arising in complexity theory. *C.R.Acad.Sci.Paris*, Ser I 340(9):627–631, 2005.

[8] J. Cai, F. Green, and T. Thierauf. On the correlation of symmetric functions. *Mathematical systems theory*, 29(3):245–258, 1996.

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

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

[11] B. Green and T. Tao. An inverse theorem for the Gower's $\mathcal{U}^3$ norm. Technical report, 2005. arXiv.org:math/0503014.

[12] F. Green. Exponential sums and circuits with a single threshold gate and mod-gates. *Theory Comput. Systems*, 32:453–466, 1999.

[13] F. Green. The correlation between parity and quadratic polynomials mod 3. *J.Comp.Syst.Sci*, 69(1):28–44, 2004.

[14] F. Green, A. Roy, and H. Straubing. Bounds on an exponential sum arising in boolean circuit complexity. *C.R.Acad.Sci.Paris*, Ser I 341:279–282, 2005.

[15] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *J.Comput.Syst.Sci*, 46(2):129–154, 1993.

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

[17] A. Klivans and R. Servedio. Learning DNF in time $2^{O(n^{1/3})}$. In *STOC*, pages 258–265, 2001.

[18] M. Krause and P. Pudlák. On the computational power of depth 2 circuits with threshold and modulo gates. In *STOC*, pages 48–57, 1994.

[19] M. Krause and P. Pudlák. Computing boolean functions by polynomials and threshold circuits. *Computational Complexity*, 7(4):346–370, 1998.

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

[21] M. Minsky and S. Papert. *Perceptrons:expanded edition*. MIT Press, Cambridge, MA, USA, 1988.

[22] R. O'Donnell and R. Servedio. Extremal properties of polynomial threshold functions. In *IEEE Conf. Computational Complexity*, pages 3–12, 2003.

[23] A. Samorodnitsky. Low degree tests at large distances. In *STOC*, 2007.

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

[25] E. Viola and A. Wigderson. Norms, XOR Lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *IEEE Conf. on Computational Complexity*, pages 141–154, 2007.

[26] A. Yao. On ACC and Threshold circuits. In *FOCS*, pages 619–627, 1990.

Appendix

k-player Degree Discrepancy Lemma

The argument for 3-players naturally extends to k players in general. We define $F_k : \{-1,1\}^{M^{k-1}} \times \binom{[M]}{m}^{k-1} \to \{-1,1\}$. The partition of inputs is again the natural extension of the three player case: Player 1 gets a binary string of length M^{k-1} and each of the other $k-1$ players receives a subset of $[M]$. The distribution λ that we choose on our inputs is $\frac{\mu_{S^1,\ldots,S^{k-1}}(x)}{\binom{M}{m}^{k-1} 2^{M^{k-1}-m}}$. We sketch the argument below.

The starting point is to write the expression for discrepancy w.r.t an arbitrary cylinder intersection ϕ, generalizing (7)

$$\mathrm{disc}_k^{\phi}(F_k) = \left| \sum_{x,S^1,\ldots,S^{k-1}} F_k(x,S^1,\ldots,S^{k-1})\phi(x,S^1,\ldots,S^k) \cdot\lambda(x,S^1,\ldots,S^{k-1}) \right| \tag{31}$$

where ϕ is the intersection of k cylinders $\phi_1, \ldots, \phi_k$, and can be expressed as below:

$$\phi(x, S^1, \ldots, S^k) =$$
$$\Big(\prod_{i=1}^{k-1} \phi^i(x, S^1, \ldots, S^{k-i-1}, S^{k-i+1}, \ldots, S^{k-1})\Big)$$
$$\times\, \phi^k(S^1, \ldots, S^{k-1})$$

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

$$\mathrm{disc}_k^{\phi}(F_k) = 2^m \Big| \mathbf{E}_{x,S^1,\ldots,S^{k-1}} F_k(x, S^1, \ldots, S^{k-1}) \times \phi(x, S^1, \ldots, S^{k-1}) \mu_{S^1,\ldots,S^{k-1}}(x) \Big| \tag{32}$$

where, as before, $(x, S^1, \ldots, S^{k-1})$ is now uniformly distributed over $\{0,1\}^{M^{k-1}} \times \binom{[M]}{m}^{k-1}$. Then, we use very similar argument of combining triangle inequality with Cauchy-Schwarz as was used in the three player case for going from (8) to (12). Applying this $k-1$ times to (32), we get the following generalization of (12):

$$(\mathrm{disc}_k^{\phi}(F_k))^{2^{k-1}} \leq 2^{2^{k-1}m} \mathbf{E}_{S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}} G_k\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) \tag{33}$$

where,

$$G_k\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) = \Big| \mathbf{E}_{x \in \{0,1\}^{M^{k-1}}} \prod_{u \in \{0,1\}^{k-1}} \Big(F_k(x, S_{u_1}^1, \ldots, S_{u_{k-1}}^{k-1}) \times \mu_{S_{u_1}^1, \ldots, S_{u_{k-1}}^{k-1}}(x) \Big) \Big| \tag{34}$$

As before we look at a fixed S_0^i, S_1^i, for $i = 1, \ldots, k-1$. Let $r = \max\{|S_0^1 \cap S_1^1|, \ldots, |S_0^{k-1} \cap S_1^{k-1}|\}$. We now generalize Claim 8:

Claim 15

$$G_k\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) \leq \frac{2^{(2^{k-1}-1)r}}{2^{2^{k-1}n}} \tag{35}$$

Proof: For any boolean string u, let $u[i]$ denote its ith bit. Since F_k is -1/1 valued, we have

$$G_k\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) \leq \Big| \mathbf{E}_{x \in \{0,1\}^{M^{k-1}}} \Big[\prod_{u \in \{0,1\}^{k-1}} \mu_{S_{u[1]}^1, \ldots, S_{u[k-1]}^{k-1}}(x) \Big] \Big| \tag{36}$$

Wlog, assume $r_1 \leq r_2 \leq \cdots \leq r_{k-1} = r$. Consider any arbitrary total order on points in $\{0,1\}^{k-1}$ that implies $x < y$ if the hamming weight of x is less than that of y. Let $u_0, \ldots, u_{2^{k-1}-1}$ be the enumeration of points in the cube according to increasing order. So, $u_0 = 00\ldots0$ and $u_{2^{k-1}-1} = 11\ldots1$. Denote by t_i, the Hamming weight of u_i for $0 \leq i \leq 2^{k-1}-1$. Let the set of indices at which u_i has a 1 be $\{j_1, \ldots, j_{t_i}\}$. Let A_i be the set of size m, consisting of $k-1$-tuples in M^{k-1} indexed by the $k-1$ sets $S^1_{u_i[1]}, \ldots, S^{k-1}_{u_i[k-1]}$. For any $k-1$-tuple w, let $w[i]$ denote its ith co-ordinate. Let,

$$Y_i = \{x_w \big| w \in A_i; \forall 1 \leq \ell \leq t_i : w[j_\ell] \in S_1^{j_\ell} - S_0^{j_\ell}\} \tag{37}$$

$$Z_i = \{x_w | w \in A_i\} \tag{38}$$

Note that $|Z_i| = m$ for all i. For $i = 0$, $t_0 = 0$ and hence, $Y_0 = Z_0$. Thus $|Y_0| = m$. For $i > 0$, $|Y_i| = m - r_{j_{t_i}} \geq m - r$. Then, for $0 \leq i < 2^{k-1}-1$, define recursively

$$\begin{aligned} &H_{u_i}(Z_i - Y_i, S_0^1, S_1^1 \ldots, S_0^{k-1}, S_1^{k-1}) \\ &= \mathbf{E}_{Y_i}\Big[\mu(Z_i) H_{u_{i+1}}(Z_{i+1} - Y_{i+1}, S_0^1, \ldots, S_1^{k-1})\Big] \end{aligned} \tag{39}$$

and for $i = 2^{k-1}-1$, let

$$H_{u_i}(Z_i - Y_i, S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}) = \mathbf{E}_{Y_i}\big[\mu(Z_i)\big] \tag{40}$$

It is not hard to verify (recalling that $Z_0 = X_0$),

$$\text{RHS of (36)} = H_0(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}) \tag{41}$$

Let γ_i be the maximal value of function H_{u_i}. Then, recalling that μ is just a probability distribution, one immediately obtains that $\gamma_i \leq 2^{-|Y_i|}\gamma_{i+1}$, for $i < 2^{k-1}-1$. Since $|Y_0| = m$, $\gamma_0 \leq 2^{-m}\gamma_1$. For $1 < i < 2^{k-1}-1$, recall $|Y_i| \geq m - r$, whence $\gamma_i \leq 2^{-(m-r)}\gamma_{i+1}$. Combining all these with the fact that $\gamma_{2^{k-1}-1} \leq 2^{-(m-r)}$, we obtain $\gamma_0 \leq 2^{(2^{k-1}-1)r}/2^{2^{k-1}m}$ that proves Claim 15. ∎

Claim 7 generalizes to the following:

Claim 16 *Let $r < d$. Then,*

$$G_k\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) = 0 \tag{42}$$

Proof: Let us consider the same total order of points in $\{0,1\}^{k-1}$ as in the proof of Claim 15. Let Y_i and Z_i be as given by (37) and (38) respectively. Define $g_i = \mathbf{E}_{Y_i} f(Z_i)\mu(Z_i)g_{i+1}$, for $1 \leq i \leq 2^{k-1}-1$ and $g_i = 1$ for $i = 2^{k-1}$. Then,

$$\begin{aligned} &G_k\big(S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}\big) \\ &= \Big|\mathbf{E}_{Y_0}\big[f(Z_0)\mu(Z_0) \cdot g_1(x)\big]\Big| \end{aligned} \tag{43}$$

where $g_1(x)$ is a function of at most r variables in $Z_0 = Y_0$. Thus, recalling that $r < d$ and $\mathbf{E}_{x \sim \mu} f(x)g(x) = 0$ for any g that depends on less than d variables, we see that (43) evaluates to zero. ∎

Application of Claim 15 and Claim 16 to (34) leads to following generalization of (16):

$$\begin{aligned}&(\mathrm{disc}_k^{\phi}(F_k))^{2^{k-1}}\\ &\leq \sum_{j=d}^{m} 2^{(2^{k-1}-1)j}\\ &\quad \times \sum_{j_1+\cdots+j_{k-1}=j} \Pr\left[N^1 = j_1 \wedge \cdots \wedge N^{k-1} = j_{k-1}\right]\end{aligned} \tag{44}$$

where, $N^i = |S_0^i \cap S_1^i|$ for $1 \leq i \leq k-1$.

This further generalizes (17) to get:

$$\begin{aligned}&(\mathrm{disc}_k^{\phi}(F_k))^{2^{k-1}}\\ &\leq \sum_{j=d}^{m} 2^{(2^{k-1}-1)j}\\ &\times \sum_{j_1+\cdots+j_{k-1}=j} \binom{m}{j_1}\cdots\binom{m}{j_{k-1}} \frac{\binom{M-m}{m-j_1}\cdots\binom{M-m}{m-j_{k-1}}}{\binom{M}{m}^{k-1}}\end{aligned} \tag{45}$$

Applying simple combinatorial identities as in the last section, (45) leads to (1), proving the Multiparty Degree-Discrepancy Lemma.

Finishing the proof of Lemma 13

We continue from (29). Noting that the number of sequences Y for which $|I_Y| = k$ is exactly $\binom{n}{k}(2^{m-1} - 1)^{n-k}$ and using the triangle inequality with the binomial theorem, we get.

$$\begin{aligned}|S_n|^m &\leq \sum_{k=0}^{n} \binom{n}{k} (2^{m-1}-1)^{n-k} 2^{n-k} 2^k \mu_{d-1}^k\\ &= 2^{nm}\left(1 - \frac{1-\mu_{d-1}}{2^{m-1}}\right)^n\end{aligned} \tag{46}$$

Taking the mth root of both sides of (46), using the inequality $(1-x)^{1/m} \leq 1 - x/m$ if $0 \leq x < 1$ amd $m > 1$ after rearranging, we obtain

$$1 - \mu_d \geq \frac{1-\mu_{d-1}}{m2^{m-1}} \geq \frac{1-\mu_0}{\left(m2^{m-1}\right)^d} \tag{47}$$

Substituting $\beta = 1 - \mu_0$, one gets $\mu_d \leq \exp\left(-\frac{\beta}{(m2^{m-1})^d}\right)$. This immediately yields (27) in Lemma 13.