Electronic Colloquium on Computational Complexity, Report No. 95 (2016)

Small Error Versus Unbounded Error Protocols in the NOF Model

Arkadev Chattopadhyay*[1] and Nikhil S.Mande†[1]

[1] *School of Technology and Computer Science, TIFR, Mumbai*

Abstract

We show that a simple function has small unbounded error communication complexity in the k-party number-on-forehead (NOF) model but every probabilistic protocol that solves it with sub-exponential advantage over random guessing has cost essentially $\Omega\left(\frac{\sqrt{n}}{4^k}\right)$ bits. Such a separation was first shown for $k = 2$ independently by Buhrman et al. [8] and Sherstov [24]. A very recent work of Sherstov [27] can be combined with an earlier work of Beigel [7] to yield such a separation for up to $k = \Theta(\log\log n)$ players. To the best of our knowledge, our result provides the first such separation that works all the way up to $k \leq \delta \log n$ players, where $\delta < 1$ is a constant. Additionally, our communication lower bounds for k-party probabilistic protocols for a function that has efficient unbounded error protocols are quantitatively stronger than previous bounds. In particular, for any constant k, our bound on communication for such a function is $\Omega(n^{1/2})$ in contrast to the best known previous bound of $\Omega(n^{1/3})$.

This has the following consequence for boolean Threshold circuits: let THR and MAJ denote respectively the classes of linear threshold functions that have unbounded weights and polynomially bounded weights. Further, let PAR_k (ANY_k) denote the class of functions that are parities of k bits (any k-junta). For every $2 \leq k \leq \delta \log n$, we show that there exists a function in linear size $\text{THR} \circ \text{PAR}_k$ that needs $2^{n^{\Omega(1)}}$ size to be computed by every circuit in the class $\text{MAJ} \circ \text{SYM} \circ \text{ANY}_{k-1}$, where SYM represents the class of all symmetric functions. Applying a result of Goldmann et al. [15] to the above, similar lower bounds on the size of circuits of the form $\text{MAJ} \circ \text{THR} \circ \text{ANY}_{k-1}$ for computing the function follow.

The main technical ingredient of our result is to show that a composed function of the form $f \circ \text{PAR}$ has exponentially small discrepancy while f has sign degree just 1.

*arkadev.c@tifr.res.in

†nikhil.mande@tifr.res.in

ISSN 1433-8092

1 Introduction

Chandra, Furst and Lipton [9] introduced the "number-on-forehead" (NOF) model of multi-party communication, over thirty years ago, to obtain lower bounds on the size of branching programs. In this model, there are k players each having an input that is metaphorically held on their foreheads. Every forehead is visible to a player except her own. The two features that make this model much more subtle than its classical two-party counterpart, are the mutual overlap of information and the fact that as k grows, each player misses less information. Indeed, starting with the surprising work of Grolmusz [16], several work (see for example [3, 1, 13]) have shown that there are very counter-intuitive protocols especially when k is larger than $\log n$. This makes proving multi-party lower bounds on the cost of protocols quite challenging. However, researchers have been well motivated to take on this challenge due to many well known applications of such lower bounds in diverse areas like circuit complexity, proof complexity, and pseudo-random generators. More recently new applications have emerged in areas like data-structures [21] and distributed computing [14].

In a seminal work, Babai, Frankl and Simon [2] introduced communication complexity classes for the 2-party model. Corresponding to polynomial time being the notion of efficient computation on a Turing machine, [2] argues that poly-log of the input length of communication is a natural notion of efficient protocols. Armed with this notion, most computational complexity classes have their analogues in communication complexity. This also extends easily to the NOF model and gives rise to complexity classes $\mathrm{P}_k^{\mathrm{cc}}, \mathrm{BPP}_k^{\mathrm{cc}}, \mathrm{NP}_k^{\mathrm{cc}}, \mathrm{PP}_k^{\mathrm{cc}}$ etc. While it is very hard to separate Turing machine complexity classes, many separation in the communication world is known when $k = 2$. For instance Equality function easily separates $\mathrm{P}_2^{\mathrm{cc}}$ from $\mathrm{BPP}_2^{\mathrm{cc}}$. Set-Disjointness famously separates $\mathrm{BPP}_2^{\mathrm{cc}}$ from $\mathrm{PP}_2^{\mathrm{cc}}$. However, for $k \geq 3$ things become much more delicate. While for $k \geq 3$ Beame et al. [5] separated $\mathrm{P}_k^{\mathrm{cc}}$ from $\mathrm{BPP}_k^{\mathrm{cc}}$ not too long ago, it is still outstanding to find an explicit function witnessing this separation for even $k = 3$. A very recent line of work [20, 12, 11, 26, 27, 22] showed that Set-Disjointness also separates $\mathrm{BPP}_k^{\mathrm{cc}}$ and $\mathrm{PP}_k^{\mathrm{cc}}$ for $k \leq \delta \cdot \log n$ for some constant $\delta < 1$.

In this paper, we consider the class $\mathrm{PP}_k^{\mathrm{cc}}$. Babai et al. realized that the Turing machine complexity class PP has two different natural versions in the communication world. Let ϵ be the advantage of a probabilistic protocol over random guessing. Then, one way to measure cost of the protocol is to sum up the total number of bits communicated in the worst case with $\log \frac{1}{\epsilon}$. Functions that admit k-party probabilistic protocols of poly-logarithmic cost in this model form the class $\mathrm{PP}_k^{\mathrm{cc}}$. The other model is unrestricted: it does not penalize by adding the $\log \frac{1}{\epsilon}$ term to the cost, i.e. the cost is just the total number of bits communicated in the worst case. Protocols in this model are allowed to use only *private* random coins (see Section 2.1) and must, on each input, have non-zero advantage over random guessing. Functions that have efficient k-party protocol in this model form the class $\mathrm{UPP}_k^{\mathrm{cc}}$. It is not difficult to see $\mathrm{PP}_k^{\mathrm{cc}} \subseteq \mathrm{UPP}_k^{\mathrm{cc}}$. The fact this inclusion is strict for $k = 2$ was relatively recently shown independently by Buhrman, Vereshchagin and de Wolf [8] and by Sherstov [24]. Although not explicitly stated in the literature, a very recent work of Sherstov [27] can be combined with an earlier work of Beigel [7] to yield such a separation for up to $k = \Theta(\log\log n)$ players[1]. Our main theorem in this work addresses this problem for $k = \omega(\log\log n)$. We show that $\mathrm{PP}_k^{\mathrm{cc}}$ is strictly contained in $\mathrm{UPP}_k^{\mathrm{cc}}$ for $2 \leq k \leq \delta \cdot \log n$, for some constant $\delta < 1$. More precisely, we extend the function defined by Goldmann, Håstad and Razborov [15] as follows:

Definition 1. *Let*

$$P(x, y_1, \ldots, y_k) \equiv \sum_{i=0}^{n-1} \sum_{j=0}^{n4^k-1} 2^i y_{1j} \ldots y_{kj}(x_{i,2j} + x_{i,2j+1})$$

[1] See Section 1.1 to get more details of the relation between our and earlier work.

where $x \in \{\pm 1\}^{2n^2 4^k}$, $y_i \in \{\pm 1\}^{n4^k}$ *for each* i.
Then, $\mathrm{GHR}_k^N(x, y_1, \ldots, y_k) \equiv sgn(2P(x, y_1, \ldots, y_k) + 1)$, *where* $N = 2n^2 4^k$.

Note that the function GHR_k^N is a $k+1$-partite function for which in a $k+1$-party communication game the inputs are assigned to players in the following natural way: inputs $x, y_1, \ldots, y_k$ are held on foreheads of Player 1, Player 2, ..., Player $k+1$ respectively. Our main theorem is given below.

Theorem 1 (Main Theorem)**.** *Let* Π *be any* $k+1$*-party probabilistic public-coin protocol solving the* GHR_k^N *function with advantage* $\epsilon > 0$. *Then,*

$$\mathrm{Cost}(\Pi) + \log(1/\epsilon) \geq \Omega\left(\frac{\sqrt{N}}{4^k} - \log N - k\right).$$

Observe that Theorem 1 lower bounds precisely the cost of a $\mathrm{PP}_{k+1}^{\mathrm{cc}}$ protocol computing GHR_k^N. On the other hand, note that GHR_k^N is a composition of a linear threshold function with N parities of arity $k+1$. A well known simple fact (refer to Section 3 for a proof) says that every such function has a $\mathrm{UPP}_{k+1}^{\mathrm{cc}}$ protocol of cost $O(\log N)$. This immediately yields the following separation result:

Corollary 1. *For all* $1 \leq k \leq \delta \cdot \log n$, *the* GHR_k^N *function is not in* $\mathrm{PP}_{k+1}^{\mathrm{cc}}$ *but is in the class* $\mathrm{UPP}_{k+1}^{\mathrm{cc}}$, *where* $\delta > 0$ *is some constant.*

An additional motivation for our work comes from the study of constant-depth circuits with Threshold gates. There are two types of Threshold gates that have been considered in the literature. The first one is with unbounded weights and the class of such gates is denoted by THR. The second is with polynomially bounded weight, called Majority gates. We denote the class of such gates by MAJ. Goldmann et al. [15] showed that although THR is strictly contained in $\mathrm{MAJ} \circ \mathrm{MAJ}$, a simple function computable by linear size $\mathrm{THR} \circ \mathrm{PAR}_2$ needs exponential size to be computed by $\mathrm{MAJ} \circ \mathrm{SYM}$ circuits, where SYM denotes gates computing arbitrary symmetric functions. We strengthen their result to depth-three circuits as follows:

Theorem 2. *For each* $k \geq 2$, *the function* $\mathrm{GHR}_k^N \in \mathrm{THR} \circ \mathrm{PAR}_{k+1}$ *needs size* $2^{\Omega\left(\frac{\sqrt{N}}{4^k} - \frac{\log N}{k}\right)}$ *to be computed by depth-three circuits of the form* $\mathrm{MAJ} \circ \mathrm{SYM} \circ \mathrm{ANY}_k$.

Let us remark that Theorem 2 continues to yield non-trivial bounds as long as $k < \delta \log n$, for a certain constant $\delta > 0$. It is also worth noting that a result of [15] immediately yields, from the above theorem, the following interesting result:

Corollary 2. *There exists a function that is computed very efficiently by* $\mathrm{THR} \circ \mathrm{PAR}_{k+1}$ *circuits but requires* $2^{\Omega\left(\frac{\sqrt{N}}{4^k} - \frac{\log N}{k}\right)}$ *size to be computed by depth-three circuits of the form* $\mathrm{MAJ} \circ \mathrm{THR} \circ \mathrm{ANY}_k$.

While earlier communication bounds (using the work of Sherstov [27] and Beigel [7]) would also yield a separation between depth-two and three circuits, Corollary 2 significantly improves over previous results in the following ways: first the lower bound on the size of depth-3 circuits implied by earlier work is essentially $2^{\Omega\left(\frac{N^{1/3}}{4^k}\right)}$ which is weaker than ours. Second, the bottom fan-in allowed for depth-3 circuits for previous work would be at most $O(\log \log n)$. Even for the regime of constant k, the separation between the bottom fan-in of the depth-2 and depth-3 circuits in previous work is larger than the necessary difference of one that we obtain. These differences with previous work is discussed in more detail in the next section.

1.1 Previous Work

The standard way to lower bound the cost of PP_k^{cc} protocols for computing a function f is to establish upper bounds on the discrepancy over cylinder intersections of f under an appropriately chosen input distribution. Doing this for $k \geq 3$ is delicate and essentially the only known method is due to Babai, Nisan and Szegedy [4]. This general prescription has been refined in Raz [23] and then successfully applied for various functions [10, 6, 26, 27, 1]. However, trying to directly use these ideas faces the following problem: First, almost all of these works bound the discrepancy of a composed function of the form $h \circ g$, where g is some nicely behaved function and h crucially has high sign degree. However, the fact that h has high sign degree seems to make the composed function difficult for UPP_k^{cc} protocols. In particular, when $k = 2$ and h is symmetric, Sherstov [25] proves that such functions have high *sign rank* and consequently are hard for even UPP_2^{cc} protocols. This gives rise to a natural challenge of proving multi-party discrepancy bounds when h has low sign degree.

However, some results are known when h has low sign degree. In a recent work, Theorem 5.7 of Sherstov [27] upper bounds the discrepancy of a composed function $f \circ g$ in terms of the ϵ-approximate degree of f and a quantity called the repeated discrepancy of g. Beigel showed in [7] that a polynomial $p : \{0,1\}^n \to \{-1,1\}$ with integer coefficients of degree d which sign represents the ODD-MAX-BIT function (which is a linear threshold function) must have weight $2^{\Omega(n/d^2)}$. A closer inspection of the proof reveals that any real polynomial of degree d which approximates ODD-MAX-BIT upto error $1-\epsilon$ must have $\epsilon \leq 2^{-\Omega(n/d^2)}$. Consider a function F that is a composition of ODD-MAX-BIT with a hard inner k-party function g with block size m. Plugging these values into Sherstov's theorem generates a discrepancy upper bound for F of roughly $\big(c \cdot G(k,m)\big)^{-n^{1/3}}$ for some constant δ, where $G(k,m)$ is the repeated discrepancy of g that depends on k and m and c is a large constant. There are various choices for g like IP or unique disjointness that will then yield for F discrepancy upper bounds of the form $2^{-n^{1/3}}$ for values of k upto $\delta \cdot \log(n)$. The main problem with this approach is that the UPP_k^{cc} upper bound seems to break down when $k = \omega(\log\log n)$ for the following reason: to get these bounds it seems that the block size of the inner function g increases exponentially with k. Hence, when $k = \omega(\log\log n)$ we have $m = (\log n)^{\omega(1)}$. A naive unbounded error protocol for a composed function, where the outer function is a linear threshold function is as in the proof of Corollary 1. In this protocol, roughly, we (randomly) pick an instance of inner function g with an advantage that can be exponentially small due to the large weights of an outer function ODD-MAX-BIT. This forces the protocol to solve the inner function correctly with probability almost 1. The hardness of g makes the communication cost now almost $\Omega(m)$ which as we saw for $k = \omega(\log\log n)$ is super-poly-logarithmic. This renders the UPP_k^{cc} protocol inefficient and it is not clear therefore how to use this approach to separate PP_k^{cc} from UPP_k^{cc} for such large values of k.

The constraint of the hardness of the inner function also weakens the circuit lower bounds that arise as compared to those that arise from our results. The natural implementation of F is by a depth-3 circuit. When one tries to implement it as a depth-2 circuit, one can obtain polynomially sized $\mathrm{THR} \circ \mathrm{PAR}_{\alpha(k,m)}$, where again α blows up exponentially in k. In contrast, our target function is expressed naturally by $\mathrm{THR} \circ \mathrm{PAR}_{k+1}$ circuits which we show cannot be computed by small sized $\mathrm{MAJ} \circ \mathrm{THR} \circ \mathrm{ANY}_k$ circuits (as stated in Corollary 2). This allows us to have bottom fan-in of $\Omega(\log n)$ in our depth-3 circuits whereas previous results would allow fan-in $O(\log\log n)$.

Finally, our proof is self-contained and simpler than the much more involved work of Sherstov [27]. Of course, the arguments in [27] are more general and thus the two approaches seem incomparable.

1.2 Our Proof Technique and Organization

We work with the GHR function which is easily seen to be the composition of the *universal* threshold function [18] and Parity. The universal threshold function derives its name from the fact that by setting some of its bits appropriately one can induce any arbitrary threshold function. In that sense, it is the hardest function of sign degree 1. To estimate the discrepancy of GHR_k^N, we extend ideas from [15] who estimated this in the setting of two players. The basic intuition can be seen after observing that for a given setting of $y_1, \dots, y_k$ the GHR_k^N function essentially depends on the sign of a plus-minus combination of A_j's for $0 \le j \le n4^k - 1$, where

$$A_j \equiv \frac{1}{2}\sum_{i=0}^{n-1} 2^i \big(x_{i,2j} + x_{i,2j+1}\big).$$

The relevant sign of each A_j depends on the parity of the bits $y_{1,j}, \dots, y_{k,j}$. Further, the set of bits in x that each A_j depends on is disjoint from the set of bits that $A_{j'}$ depends on. We sample x such that each A_j are i.i.d. binomial distributions centered at 0 with range $[-2^n + 1, 2^n - 1]$. Let this distribution be μ_x. We sample each y_i uniformly at random. We want to ensure that the GHR_k^N function, under this distribution, behaves in a way that leaves the players with little clue about the outcome unless the relevant sign to be associated with each A_j is determined. To do this, as done in [15], one is forced to sample in a slightly more involved way: first sample y's uniformly at random. Then sample x according to μ_x, conditioned on the fact that $P = \sum_{j=0}^{n4^k-1} A_j y_{1,j} \cdots y_{k,j}$ is very close to its mean compared to its standard deviation (which is as high as $2^{\Omega(n)}$). Note that the median of each A_j is about $2^{n/2}$, which gives us plenty of room to exploit. This turns out to be the hard distribution but to establish this requires technical work.

Organization: Section 2 develops the basic notions and lemmas. Section 3 establishes our main technical result, Theorem 3, which upper bounds the k-wise discrepancy of the GHR function. Using this, we prove Theorem 1 and Corollary 1. Section 4 derives the circuit consequences of Theorem 2 and Corollary 2. Finally, Section 5 concludes with some open problems.

2 Preliminaries

2.1 The NOF model

In the k-party model of Chandra et al.[9], k players with unlimited computational power wish to compute a function $f : X_1 \times \cdots \times X_k \to \{-1, 1\}$ on some input $x = (x_1, \dots, x_k)$. For the purpose of this paper, we consider inputs of the form $X_i \in \{-1, 1\}^{n_i}$. On input x, player i is given $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$, which is why it is figuratively said that x_i is on the ith player's forehead. Players communicate by writing on a blackboard, so every player sees every message. We denote by $D_k(f)$ the deterministic k-party communication complexity of f, namely the number of bits exchanged in the best deterministic protocol for f on the worst case input.

A probabilistic protocol Π with access to public (private) randomness computes f with advantage ϵ if the probability that Π and f agree is at least $1/2 + \epsilon$ for all inputs. The cost of Π is the maximum number of bits it communicates over it's internal random choices in the worst case. $R_\epsilon(f)$ is the cost of the best such protocol. Let us define two other notions.

$$\mathrm{PP}_k(f) \equiv \min_\epsilon \left[R_\epsilon^{pub}(f) + \log\left(\frac{1}{\epsilon}\right)\right], \qquad \mathrm{UPP}_k(f) \equiv \min_\epsilon \left[R_\epsilon^{priv}(f)\right] \tag{1}$$

Note that privateness of the random coins is essential. It is a simple exercise to show that every function can be computed using 2 bits if allowed public coins. Define $\mathrm{PP}_k^{\mathrm{cc}} = \{f :$

$\mathrm{PP}_k(f) = \mathrm{polylog}(n)\}$ and $\mathrm{UPP}_k^{\mathrm{cc}} = \{f : \mathrm{UPP}_k(f) = \mathrm{polylog}(n)\}$, where n is the maximum length of an input to a player.

2.2 Cylinder intersections, discrepancy and the cube norm

Let $f : X_1 \times \cdots \times X_k \to \{-1, 1\}$. A subset $S_i \subseteq X_1 \times \cdots \times X_k$ is a cylinder in the ith direction if membership in S does not depend on the ith coordinate. A subset S is called a cylinder intersection if it can be represented as the intersection of k cylinders, $S = \cap_{i=1}^k S_i$, where S_i is a cylinder in the ith direction.

Definition 2. *Let μ be a distribution on $X_1 \times \cdots \times X_k$. The discrepancy of f according to μ, $Disc_\mu^k(f)$ is*

$$\max_S \left| \Pr_\mu[f(x_1, \ldots, x_k) = 1 \wedge (x_1, \ldots, x_k) \in S] - \Pr_\mu[f(x_1, \ldots, x_k) = -1 \wedge (x_1, \ldots, x_k) \in S] \right|$$

where the maximum is taken over all cylinder intersections S.

The k in $Disc_\mu^k$ denotes the dimension of the underlying cylinder intersections. We will drop this superscript when it is clear from the context what k is. Let $Disc(f) = \min_\mu Disc_\mu^k(f)$.
The discrepancy method is a powerful tool that lower bounds the randomized communication complexity in terms of the discrepancy. The following lemma can be found for example in [19].

Lemma 1. *$R_\epsilon(f) \geq \log(2\epsilon / Disc(f))$.*

We now recall a useful technique that upper bounds the discrepancy of a function under a product distribution. It is a standard lemma and can be found in [11] and [23] for example.

Lemma 2. *Let $f : X \times Y_1 \times \cdots \times Y_k \to \{-1, 1\}$, and μ any product distribution. Then,*

$$(Disc_\mu^{k+1}(f))^{2^k} \leq \mathbb{E}_{y_1^0, y_1^1, \ldots, y_k^0, y_k^1} \left[\left| \mathbb{E}_x \Pi_{a \in \{0,1\}^k} f(x, y_1^{a_1}, \ldots, y_k^{a_k}) \right| \right]$$

2.3 The binomial distribution

Definition 3. *Let $B(N)$ denote the distribution obtained as the sum of $2N$ independent Bernoulli variables, each of which take values $1/2, -1/2$ with probability $1/2$ each.*

A few important things to observe are that $B(N)$ takes only integral values, it is centered and symmetric around 0, so $B(N)$ is identically distributed to $-B(N)$. Its range is $[-N, N]$. Let us denote $\Pr[B(N) = 0]$ by p_0. It is a well known fact that $p_0 = \frac{\binom{2N}{N}}{4^N} = \Theta\left(\frac{1}{N^{1/2}}\right)$. The following lemma tells us that the probability of a binomial distribution taking any value close to its mean is significantly high.

Lemma 3. *Let W be a binomial random variable distributed according to $B(N)$. Let p_0 denote $\Pr[W = 0]$. Then for all $j \in [-N, N]$,*

$$p_0 - O\left(\frac{j^2}{N^{3/2}}\right) \leq \Pr[W = j] \leq p_0$$

Proof. Note that for $j \geq N/2$, the bound to be proved is trivial. Thus we assume $j < N/2$.

$$\begin{aligned}
\Pr[W = j-1] - \Pr[W = j] &= \left[\binom{2N}{N+j-1} - \binom{2N}{N+j}\right] \cdot \frac{1}{2^{2N}} \\
&= \left[\frac{(2N)!}{(N+j-1)!(N-j+1)!} - \frac{(2N)!}{(N+j)!(N-j)!}\right] \cdot \frac{1}{2^{2N}} \\
&= \frac{(2N)!}{(N+j-1)!(N-j)!} \cdot \frac{2j-1}{(N-j+1)(N+j)} \cdot \frac{1}{2^{2N}} \\
&= \binom{2N}{N+j} \cdot \frac{2j-1}{N-j+1} \cdot \frac{1}{2^{2N}} \\
&\leq \binom{2N}{N} \cdot \frac{1}{2^{2N}} \cdot \frac{2j}{N-j}
\end{aligned}$$

since the middle binomial coefficient is the maximum. Thus, we have $\forall i, |i| \leq j$,

$$\Pr[W = i-1] - \Pr[W = i] \leq \binom{2N}{N} \frac{2j}{N-j} \cdot \frac{1}{2^{2N}}$$

Since $\frac{\binom{2N}{N}}{4^N} = \Theta\left(\frac{1}{N^{1/2}}\right)$,

$$\begin{aligned}
\Pr[W = 0] - \Pr[W = j] &\leq \sum_{i=1}^{j} |\Pr[W = i-1] - \Pr[W = i]| \leq \frac{2j^2}{N-j} \cdot O\left(\frac{1}{N^{1/2}}\right) \\
&\leq \frac{2 \cdot 2j^2}{N} \cdot O\left(\frac{1}{N^{1/2}}\right) \qquad \text{Since } j \leq N/2 \\
&\leq O\left(\frac{j^2}{N^{3/2}}\right)
\end{aligned}$$

□

3 A discrepancy upper bound for the multiparty GHR function

In this section, we prove essentially an $O\left(2^{-\sqrt{N}/4^k}\right)$ upper bound on the discrepancy of the GHR_k^N function where the first player gets N input bits. This gives us an inverse exponential upper bound on the discrepancy if $k < \epsilon \log(N)$ for any constant ϵ. Goldmann et al. [15] showed that when $k = 2$, if there is a low cost one-way protocol for GHR_2^N, then it must have low advantage. Sherstov [24] noted that the same proof technique can be adapted to prove an upper bound on the discrepancy on GHR_2^N. We generalize this for higher k. In particular, we show

Theorem 3. *For any $k \geq 1$,*

$$\mathit{Disc}(\mathrm{GHR}_k^N) = O\left(\frac{(8e)^k N^{1/4}}{2^{\sqrt{N}/4^k} \cdot 2^{k/2}}\right),$$

where GHR_k^N is defined as in Definition 1, and N is the maximum number of bits a player gets (in this case the first player).

Proof of Theorem 1. It follows directly from Theorem 3 and Lemma 1. □

Proof of Corollary 1. From Theorem 1, it follows that for all $1 \leq k \leq \delta \cdot \log n$, the GHR_k^N function is not in PP_{k+1}^{cc} where $\delta > 0$ is some constant. Let us see an easy unbounded error protocol for GHR_k^N. Note that all the weights of the top threshold are positive. One player chooses and announces a bottom layer Parity gate with probability proportional to it's corresponding weight. The cost of announcing this is $O(\log(N))$. The probability of success equals $\sum w_i^+/w$, where w_i^+'s are the weights of the gates which agree with the output. Since $\sum w_i^+ > \sum w_i^-$ (the gates which disagree with the output), the probability of success is strictly greater than 1/2. □

Recall that $N = 2n^2 4^k$. The proof technique of Theorem 3 is inspired from that of Goldmann et al. [15].

Proof of Theorem 3. Let $A_j = \frac{1}{2}\sum_{i=0}^{n-1} 2^i(x_{i,2j} + x_{i,2j+1})$. It is easy to see that A_j can take any integer value in $[-2^n+1, 2^n-1]$. Let μ_x be a distribution on the x's that make the A_j's independent and binomially distributed according to $B(2^n-1)$ as defined in Definition 3. Such a distribution exists because each A_j depends on a disjoint set of variables. For each $i \in \{1, \ldots, k\}$, let μ_i be the uniform distribution on the y_i. We choose a tuple $(x, y_1, \ldots, y_k)$ by first picking $y_i \sim \mu_i$ independently for each i, and then picking $x \sim \mu_x$ under the condition that $|P(x, y_1, \ldots, y_k)| = 2^k$. Let us call this distribution μ.

We will now show an upper bound on the discrepancy of GHR_k^N under the distribution μ. Let S denote the characteristic function (0-1 valued) of a cylinder intersection. By Definition 2, the discrepancy of GHR_k^N according to μ is

$$Disc_\mu(\mathrm{GHR}_k^N) = \max_S \left|\mathbb{E}_\mu\left[\mathrm{GHR}_k^N(x, y_1, \ldots, y_k) S(x, y_1, \ldots, y_k)\right]\right| \tag{2}$$

The following lemma will enable us to switch to working with a product distribution on the inputs, for which we have nice techniques for proving lower bounds.

Lemma 4. *For $\mu_x, \mu_1, \ldots, \mu_k$ as defined above,*

$$\Pr_{\mu_x, \mu_1, \ldots, \mu_k}[|P(x, y_1, \ldots, y_k)| = 2^k] \geq \Omega\left(\frac{1}{\sqrt{n}2^{(n+2k)/2}}\right)$$

Proof. We will show that for any fixed $y_1, \ldots, y_k$, if we sample x according to μ_x, then $P(x, y_1, \ldots, y_k)/2 = \sum_{j=0}^{n4^k-1} A_j y_{1j} \cdots y_{kj}$ is distributed according to $B(n4^k(2^n-1))$. First note that no matter what the values of $y_1, \ldots, y_k$, $A_j y_{1j} \cdots y_{kj}$ is always distributed according to $B(2^n-1)$. This is because this distribution is symmetric around 0. Next, observe that the sum of binomials is a binomial. This shows that $\sum_{j=0}^{n4^k-1} A_j y_{1j} \cdots y_{kj}$ is distributed according to $B(n4^k(2^n-1))$.

Hence by plugging in $N = n4^k(2^n-1)$ and $j = 2^k$ in Lemma 3,

$$\begin{aligned}\Pr_{\mu_x, \mu_1, \ldots, \mu_k}[|P(x, y_1, \ldots, y_k)| = 2^k] &\geq \Theta\left(\frac{1}{(n4^k(2^n-1))^{1/2}}\right) - O\left(\frac{4^k}{(n4^k(2^n-1))^{3/2}}\right) \\ &= \Omega\left(\frac{1}{\sqrt{n}2^{(n+2k)/2}}\right)\end{aligned}$$

We could discard the second term since it equals $O\left(\frac{1}{(4^k)^{1/2}\cdot(n(2^n-1)^{3/2})}\right)$, and is dominated by the first term. □

Define a function q by

$$q(x, y_1, \ldots, y_k) = \begin{cases} P(x, y_1, \ldots, y_k)/2^k & \text{if } |P(x, y_1, \ldots y_k)| = 2^k \\ 0 & \text{otherwise} \end{cases}$$

This means that if $(x, y_1, \ldots, y_k)$ is chosen according to the distribution $\mu_x \times \mu_1 \times \cdots \times \mu_k$, then $q(x, y_1, \ldots, y_k) = \mathrm{GHR}_k^N(x, y_1, \ldots, y_k)$ on the domain of μ, and 0 otherwise. Using Lemma 4,

$$Disc_\mu(\mathrm{GHR}_k^N) \leq |\mathbb{E}_{\mu_x,\mu_1,\ldots,\mu_k}[q(x, y_1, \ldots, y_k)S(x, y_1, \ldots, y_k)]| \cdot O\left(\sqrt{n}2^{\frac{n+2k}{2}}\right) \tag{3}$$

This can be seen by expanding the expectation in the above equation and in Equation 2 and comparing term by term. We can then use the definition of conditional probability to obtain the above inequality. It suffices to show

$$|\mathbb{E}_{\mu_x,\mu_1,\ldots,\mu_k}[q(x, y_1, \ldots, y_k)S(x, y_1, \ldots, y_k)]| \leq O\left(2^{-\frac{n+2k}{2}-\epsilon}\right) \tag{4}$$

for some constant $\epsilon > 0$ to give us an inverse exponential discrepancy. For notational convenience, use $\mathbb{E}_x$ when we mean $\mathbb{E}_{x\sim\mu_x}$ from now on. Now that we have a product distribution, we can use Lemma 2,

$$\begin{aligned} &|\mathbb{E}_{\mu_x,\mu_1,\ldots,\mu_k}[q(x, y_1, \ldots, y_k)S(x, y_1, \ldots, y_k)]| \\ &\qquad \leq \left(\mathbb{E}_{y_1^0,y_1^1,\ldots,y_k^0,y_k^1}\left|\mathbb{E}_x\left[\prod_{a_1,\ldots,a_k\in\{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k})\right]\right|\right)^{1/2^k} \end{aligned} \tag{5}$$

We will now upper bound the RHS of the above equation by splitting the outer expectation into two terms, the first of which has low probability. We will require certain properties of Hadamard matrices to upper bound the second term. Let $\alpha \in \{0,1\}^k$. Define 2^k subsets of indices as $I_\alpha = \{j \in [n4^k] : \forall i \in [k],\ (y_i^0)_j = -1^{\alpha_i} \cdot (y_i^1)_j\}$. Since our distribution on y_i^0, y_i^1's are uniform and independent, the probability of each I_α being empty is equal. An easy counting gives us the probability of I_α being empty as $\left(\frac{2^k-1}{2^k}\right)^{n4^k}$. By a union bound, the probability that any one of them is empty is at most $2^k \cdot \left(\frac{2^k-1}{2^k}\right)^{n4^k}$. We have the following.

$$\left(\mathbb{E}_{y_1^0,y_1^1,\ldots,y_k^0,y_k^1}\left|\mathbb{E}_x\left[\prod_{a_1,\ldots,a_k\in\{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k})\right]\right|\right)^{1/2^k} \leq \left(2^k\left(1 - \frac{1}{2^k}\right)^{n4^k} + Z\right)^{1/2^k}$$

where $Z = \mathbb{E}_{y_1^0,y_1^1,\ldots,y_k^0,y_k^1:\forall\alpha,I_\alpha\neq\emptyset}\left|\mathbb{E}_x \prod_{a_1,\ldots,a_k\in\{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k})\right|$

Claim 1. *For all $y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$ such that I_α is non-empty for each $\alpha \in \{0,1\}^k$, we have*

$$\left|\mathbb{E}_x\left[\prod_{a_1,\ldots,a_k\in\{0,1\}} q(x, y_1^{a_1}, \ldots, y_k^{a_k})\right]\right| \leq O\left(2^{k\log(e)2^k} \cdot 2^{2^k}\frac{1}{(2^{n/2})^{2^k-1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)$$

Let us assume the claim to be true for now. We have from Equation 3 that

$$
\begin{aligned}
Disc_\mu(\mathrm{GHR}_k^N) &\le |\mathbb{E}_{\mu_x,\mu_1,\ldots,\mu_k}[q(x,y_1,\ldots,y_k)S(x,y_1,\ldots,y_k)]|\; O\left(\sqrt{n}2^{\frac{n+2k}{2}}\right)\\
&\le \left(2^k\left(1-\frac{1}{2^k}\right)^{n4^k} + O\left(2^{k\log(e)2^k}\cdot 2^{2^k}\frac{1}{(2^{n/2})^{2^k-1}}\cdot\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)\right)^{1/2^k}\\
&\cdot O\left(\sqrt{n}2^{\frac{n+2k}{2}}\right)\\
&\le \left[2^{k/2^k}\left(1-\frac{1}{2^k}\right)^{n2^k} + O\left(\frac{(4e)^k}{(2^{\frac{n}{2}})^{1-\frac{1}{2^k}}\cdot 2^{\frac{3n}{2}\cdot\frac{1}{2^k}}}\right)\right] O\left(\sqrt{n}2^{\frac{n+2k}{2}}\right)\\
&\le O\left(\left(e^{-1/2^k}\right)^{n2^k}\cdot 2^{n/2+k} + \frac{(8e)^k\sqrt{n}}{2^{(\frac{3n}{2}-\frac{n}{2})\cdot\frac{1}{2^k}}}\right)\\
&\qquad\qquad\qquad\qquad\text{Using the fact that } \left(1-\tfrac{1}{\beta}\right) < e^{-1/\beta}\\
&= O\left(e^{-n}\cdot 2^{n/2+k} + \frac{(8e)^k\sqrt{n}}{2^{n/2^k}}\right) = O\left(\frac{(8e)^k\sqrt{n}}{2^{n/2^k}}\right) \qquad \text{Assuming } k < n/2
\end{aligned}
$$

which proves Theorem 3. Assuming $k \le \epsilon\log(n)$ for any constant $\epsilon < 1$ gives us an inverse exponential upper bound on the discrepancy. □

Now it only remains to prove Claim 1.

3.1 Proof of Claim 1

Recall that we need to show the following. For all $y_1^0,\ldots,y_k^0,y_1^1,\ldots,y_k^1$ such that I_α is non-empty for each α, we want

$$
\left|\mathbb{E}_x\left[\prod_{a_1,\ldots,a_k\in\{0,1\}} q(x,y_1^{a_1},\ldots,y_k^{a_k})\right]\right| \le O\left(2^{k\log(e)2^k}\cdot 2^{2^k}\frac{1}{(2^{n/2})^{2^k-1}}\cdot\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)
$$

Fix any such $y_1^0,\ldots,y_k^0,y_1^1,\ldots,y_k^1$. Note that the LHS of the above equation is

$$
\left|\Pr_x\left[\prod_{a_1,\ldots,a_k\in\{0,1\}} q(x,y_1^{a_1},\ldots,y_k^{a_k})=1\right] - \Pr_x\left[\prod_{a_1,\ldots,a_k\in\{0,1\}} q(x,y_1^{a_1},\ldots,y_k^{a_k})=-1\right]\right|
$$

For convenience, for all $a\in\{0,1\}^k$ let us denote $P(x,y_1^{a_1},\ldots,y_k^{a_k})$ by P_a and let S_a denote $P_a/2$. By the definition of q, we have

$$
\begin{aligned}
\left|\mathbb{E}_x\left[\prod_{a\in\{0,1\}^k} q(x,y_1^{a_1},\ldots,y_k^{a_k})\right]\right| &= \left|\Pr\left[\prod_{a\in\{0,1\}^k}\frac{P_a}{2^k}=1\right] - \Pr\left[\prod_{a\in\{0,1\}^k}\frac{P_a}{2^k}=-1\right]\right|\\
&= \left|\Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}\right] - \Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}\right]\right| \qquad (6)
\end{aligned}
$$

Let $W_\alpha = \sum_{j\in I_\alpha} A_j(y_1^0)_j\ldots(y_k^0)_j$. It will be useful to note here that W_α only takes integral values. We will use this fact crucially later. Let $\mathbf{P_k}$ denote the $2^k\times 1$ matrix whose rows are indexed by $a=(a_1,\ldots,a_k)\in\{0,1\}^k$, and the ath row of $\mathbf{P_k}$ is $P(x,y_1^{a_1},\ldots,y_k^{a_k})$. Similarly define matrices $\mathbf{S_k}$ and $\mathbf{W_k}$.

Claim 2. *The following holds true for all k*

$$\mathbf{P_k} = 2\mathbf{S_k} = 2\mathbf{H_k} \cdot \mathbf{W_k}$$

where $\mathbf{H_k}$ is a $2^k \times 2^k$ Hadamard matrix defined as $\mathbf{H_k} = \begin{bmatrix} \mathbf{H_{k-1}} & \mathbf{H_{k-1}} \\ \mathbf{H_{k-1}} & -\mathbf{H_{k-1}} \end{bmatrix}$ and $\mathbf{H_0} = [1]$.

Let us first state a well known property of $\mathbf{H_k} = \begin{bmatrix} \mathbf{H_{k-1}} & \mathbf{H_{k-1}} \\ \mathbf{H_{k-1}} & -\mathbf{H_{k-1}} \end{bmatrix}$ where $\mathbf{H_0} = [1]$.

Fact 1. *Let $\mathbf{H_k}$ be as defined above. Then, $(\mathbf{H_k})_{ij} = (-1)^{\langle i,j \rangle}$ for all $i, j \in \{0,1\}^k$.*

In other words, $\mathbf{H_k}$ is the communication matrix of the inner product (modulo 2) function. Let us now prove Claim 2.

Proof. Let $a \in \{0,1\}^k$. $P_a = 2\sum_{j=1}^{n4^k} A_j (y_1^{a_1})_j \cdots (y_k^{a_k})_j$ and $W_\alpha = \sum_{j \in I_\alpha} A_j (y_1^0)_j \cdots (y_k^0)_j$. Say $j \in I_\alpha$ where $\alpha \in \{0,1\}^k$. Note that $(y_i^{a_i})_j = -1 \cdot (y_i^0)_j$ iff $a_i = 1, \alpha_i = 1$. Hence $(y_1^{a_1})_j \cdots (y_k^{a_k})_j = (-1)^{\left(\sum_i a_i \cdot \alpha_i\right)} (y_1^0)_j \cdots (y_k^0)_j = (-1)^{\langle a, \alpha \rangle} (y_1^0)_j \cdots (y_k^0)_j$.

$$\begin{aligned} P_a = 2\sum_{j=1}^{n4^k} A_j (y_1^{a_1})_j \cdots (y_k^{a_k})_j &= 2\left(\sum_{\alpha \in \{0,1\}^k} \sum_{j \in I_\alpha} (-1)^{\langle a, \alpha \rangle} A_j (y_1^0)_j \cdots (y_k^0)_j \right) \\ &= 2\left(\sum_{\alpha \in \{0,1\}^k} (-1)^{\langle a, \alpha \rangle} W_\alpha \right) \\ &= 2(\mathbf{H_k})_a \cdot \mathbf{W_k} \end{aligned}$$

where $(\mathbf{H_k})_a$ denotes the ath row of $\mathbf{H_k}$. Thus, $\mathbf{P_k} = 2\mathbf{S_k} = 2\mathbf{H_k} \cdot \mathbf{W_k}$. □

3.1.1 On integral solutions to Hadamard constraints

In this subsection, we will prove that the number of integral solutions to $\mathbf{W_k}$ such that $\prod_{a \in \{0,1\}^k} S_a = 2^{(k-1)2^k}$ is equal to the number of integral solutions to $\mathbf{W_k}$ such that $\prod_{a \in \{0,1\}^k} S_a = -2^{(k-1)2^k}$. Moreover, we show that the total number of integral solutions is small, and the values of $|W_a|$ are not too large in any integral solutions. Recall from Equation 6 that for all $y_1^0, \ldots, y_k^0, y_1^1, \ldots, y_k^1$ such that I_α is non-empty for each α, we have

$$\begin{aligned} &\left| \mathbb{E}_x \left[\prod_{a \in \{0,1\}^k} q(x, y_1^{a_1}, \ldots, y_k^{a_k}) \right] \right| \\ &\qquad = \left| \Pr_x \left[\prod_{a \in \{0,1\}^k} S_a = 2^{(k-1)2^k} \right] - \Pr_x \left[\prod_{a \in \{0,1\}^k} S_a = -2^{(k-1)2^k} \right] \right| \end{aligned}$$

This allows us to pair the "positive" and "negative" solutions, and higher order terms in the difference of probabilities $\left| \Pr_x \left[\prod_{a \in \{0,1\}^k} S_a = 2^{(k-1)2^k} \right] - \Pr_x \left[\prod_{a \in \{0,1\}^k} S_a = -2^{(k-1)2^k} \right] \right|$ cancel out. We will require the following well known property of Hadamard matrices.

Fact 2. *Let $\mathbf{H}$ be an $N \times N$ Hadamard matrix. Then, $\mathbf{H}$ is invertible, and $\mathbf{H}^{-1} = \frac{1}{N}\mathbf{H}$.*

Claim 3. *The number of integral solutions to $\mathbf{W_k}$ such that $\prod_{a \in \{0,1\}^k} S_a = +2^{(k-1)2^k}$ equals the number of integral solutions such that $\prod_{a \in \{0,1\}^k} S_a = -2^{(k-1)2^k}$.*

Proof. The constraints we have are $\mathbf{H_k} \cdot \mathbf{W_k} = \mathbf{S_k}$. Since W_a is integral for all a, and $\mathbf{H_k}$ is a ± 1 matrix, this implies that S_a's are integral as well. Thus, using Fact 2 we get $\frac{1}{2^k}\mathbf{H_k} \cdot \mathbf{S_k} = \mathbf{W_k}$, or $\mathbf{H_k} \cdot \frac{\mathbf{S_k}}{2^k} = \mathbf{W_k}$. Let us consider two cases, one where $\forall a \in \{0,1\}^k, \left|\frac{S_a}{2^k}\right| = 1/2$, and another where there exists an a such that $\left|\frac{S_a}{2^k}\right| \neq 1/2$.

- Let us assume $\forall a, \left|\frac{S_a}{2^k}\right| = 1/2$. We show something slightly stronger, namely that every setting of each $\frac{S_a}{2^k}$ to $\pm 1/2$ gives us an integer solution to the W_a's. Since $\mathbf{H_k}$ is a ± 1 matrix of even dimension, the parity of the number of appearances of $+1/2$ equals the parity of number of appearances of $-1/2$ in the sum $(\mathbf{H_k})_R \cdot \frac{\mathbf{S_k}}{2^k}$, where $(\mathbf{H_k})_R$ is the Rth row of $\mathbf{H_k}$. This holds for every row R. Thus, W_R is always an integer. This means the number of positive solutions equals the number of negative solutions in this case.

- The absolute value of S_a must equal a power of 2 for each a since the product of them is a power of 2. If there exists an S_a whose value is not $\pm 2^{k-1}$, then there must exist an S_b (consider the last such one) which is a multiple of 2^k since $\prod_{a\in\{0,1\}^k} S_a = \pm 2^{(k-1)2^k}$. Since $S_b/2^k$ is an integer, and we had an integral solution to $\mathbf{W_k}$, flipping the sign of S_b can change the value of any W_c to $W_c \pm 2 \cdot S_j/2^k$, which remains an integer. This is a bijection between positive and negative solutions.

□

The following lemmas just require $\mathbf{H_k}$ to be the $2^k \times 2^k$ Hadamard matrix as defined in Claim 2, S_a's to be integer valued such that $\prod_{a\in\{0,1\}^k} S_a = \pm 2^{(k-1)2^k}$, and $\mathbf{H_k} \cdot \mathbf{W_k} = \mathbf{S_k}$.

Lemma 5. *The number of integral solutions to* $\mathbf{W_k}$ *is at most* $2^{k\log(e)2^k}$.

We will require the following standard fact about binomial coefficients.

Fact 3. *For all n and for all $k \in [n]$,* $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{n \cdot e}{k}\right)^k$.

Proof of Lemma 5. Suppose $\prod_{a\in\{0,1\}^k} S_a = \pm 2^{(k-1)2^k}$. This means we have to distribute $(k-1)2^k$ powers of 2 among $2^k S_a$'s (which are all integers). This equals the number of non-negative integer solutions to $m_1 + \cdots + m_{2^k} = (k-1)2^k$, which equals $\binom{k2^k-1}{(k-1)2^k}$. This is at most $\binom{k2^k}{(k-1)2^k}$, which is at most $\left(\frac{k2^k \cdot e}{(k-1)2^k}\right)^{(k-1)2^k}$ by Fact 3. Now we will use the fact that $1 + x \leq e^x$ and conclude that this is bounded above by e^{k2^k}, which equals $2^{k\log(e)2^k}$. Each of these can give at most 1 integral solution to the W_a's because the system of constraints is linearly independent. □

We now state an upper bound on the value of $|W_a|$ in every integral solution.

Lemma 6. *For all $a \in \{0,1\}^k, |W_a| \leq 2^{(k+1)2^k}$.*

Proof. First note that for each a, $|W_a| \leq \sum_{a\in\{0,1\}^k} \frac{|S_a|}{2^k}$ since $\mathbf{H_k} \cdot \mathbf{S_k} = \mathbf{W_k}$. We show that $\sum_{a\in\{0,1\}^k} |S_a|$ is at most 2^{k2^k}. Suppose not. By a simple averaging argument, there must be an b such that $|S_b| > \frac{2^{k2^k}}{2^k}$, which is $2^{k(2^k-1)}$, which is at least $2^{(k-1)2^k}$ if $k \geq 1$. But this is not possible since $\prod_{a\in\{0,1\}^k} S_a = \pm 2^{(k-1)2^k}$ and S_i's are integers. □

3.1.2 Using properties of the binomial distribution

Recall from Equation 6 that for all $y_1^0, \dots, y_k^0, y_1^1, \dots, y_k^1$ such that I_α is non-empty for each α, we want to upper bound $\left|\Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}\right] - \Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}\right]\right|$. Recall that we defined $W_\alpha = \sum_{j\in I_\alpha} A_j(y_1^0)_j \dots (y_k^0)_j$. For any $\alpha \in \{0,1\}^k$, note that W_α is always distributed according to $B(c_\alpha(2^n-1))$, where $c_\alpha = |I_\alpha| \neq 0$. We can prove this in a manner similar to that in the proof of Lemma 4. In Claim 3, we showed that the number of integral solutions to $\mathbf{W_k}$ such that $\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}$ equals the number of integral solutions such that $\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}$. Note that if the solution to $\mathbf{W_k}$ is not integral, then it has probability 0 since for each a, W_a takes only integral values. Let us call a solution to $\mathbf{W_k}$ to be positive if the corresponding value of $\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}$, and negative if the value is $\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}$. Arbitrarily pair up the positive and negative solutions. We will bound the difference of probabilities of each pair.

$$\begin{aligned}&\left|\Pr_x\left[\textstyle\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}\right] - \Pr_x\left[\textstyle\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}\right]\right| \\ &\leq \sum_{w,w'} \left|\Pr_x[\mathbf{W_k} = w] - \Pr_x[\mathbf{W_k} = w']\right|\end{aligned}$$

where $w = (w_a)_{a\in\{0,1\}^k}, w' = (w'_a)_{a\in\{0,1\}^k}$ are positive and negative solutions respectively to $\mathbf{W_k}$ such that $\prod_{a\in\{0,1\}^k} S_a = \pm 2^{(k-1)2^k}$. The term $\Pr_x[\mathbf{W_k} = w]$ equals $\Pr_x[\bigwedge_{a\in\{0,1\}^k} W_a = w_a]$. In Lemma 6 we showed that for each α, the absolute value of W_α in any integral solution can be at most $2^{(k+1)2^k}$. Each W_α is distributed according to $B(c_\alpha(2^n-1)), c_\alpha > 0$, since $|I_\alpha| > 0$.

Now for a particular positive solution w, negative solution w' and any $y_1^0, \dots, y_k^0, y_1^1, \dots, y_k^1$ such that I_α is non-empty for each α,

$$\left|\Pr_x[\mathbf{W_k} = w] - \Pr_x[\mathbf{W_k} = w']\right| = \left|\Pr[\bigwedge_{a\in\{0,1\}^k} W_a = w_a] - \Pr[\bigwedge_{a\in\{0,1\}^k} W_a = w'_a]\right|$$

By plugging in $N = c_\alpha(2^n - 1)$ and $j = 2^{(k+1)2^k}$ in Lemma 3, we obtain $p_0 \geq \Pr_x[W_\alpha = w_\alpha] \geq p_0 - O\left(\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)$, where $p_0 = \Pr[W_\alpha = 0] = O\left(\frac{1}{2^{n/2}}\right)$. For convenience in calculations, let us say $\Pr_x[W_\alpha = w_\alpha] \in \left(p_0 \pm O\left(\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)\right)$. Recall that W_α's are independent of each other since they depend on disjoint variables. Thus,

$$\begin{aligned}&\left|\Pr[\textstyle\bigwedge_{a\in\{0,1\}^k} W_a = w_a] - \Pr[\bigwedge_{a\in\{0,1\}^k} W_a = w'_a]\right| \\ &\leq \left|\left(p_0 \pm O\left(\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)\right)^{2^k} - \left(p_0 \pm O\left(\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\right)\right)^{2^k}\right| \leq \frac{2^{2^k}}{(2^{n/2})^{2^k-1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}\end{aligned}$$

The last inequality holds because the highest order term after binomially expanding both terms is $\frac{1}{2^{n2^k/2}}$, which cancel each other. Note that the sum of the binomial coefficients is 2^{2^k}, and each term after the first is at most $\frac{1}{(2^{n/2})^{2^k-1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}$. Thus, the sum of the remaining terms can be bounded above by $2^{2^k}\frac{1}{(2^{n/2})^{2^k-1}} \cdot \frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}$. By Claim 5, the number of solutions (and hence number of pairs) is at most $2^{k\log(e)2^k}$. Thus,

$$\left|\Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = 2^{(k-1)2^k}\right] - \Pr_x\left[\prod_{a\in\{0,1\}^k} S_a = -2^{(k-1)2^k}\right]\right|$$
$$\le \sum_{w,w'}\left|\Pr_x[\mathbf{W_k}=w]-\Pr_x[\mathbf{W_k}=w']\right| \le 2^{k\log(e)2^k}\cdot 2^{2^k}\frac{1}{(2^{n/2})^{2^k-1}}\cdot\frac{2^{(k+1)2^{k+1}}}{2^{3n/2}}$$

which proves Claim 1. Using Equation 3, this proves Theorem 3.

4 Circuit Lower Bounds

In this section, we will show how we obtain depth-3 circuit lower bounds on the class $\mathrm{MAJ}\circ\mathrm{THR}\circ\mathrm{ANY}_k$ for functions that are in $\mathrm{THR}\circ\mathrm{PAR}_{k+1}$. First let us state the results that were known prior to this work.

Lemma 7 (Folklore). *Any function $f\in\mathrm{SYM}\circ\mathrm{ANY}_k$ of size s has a deterministic simultaneous $(k+1)$ player protocol of cost $O(k\log(s))$ for any partitioning of the input bits.*

Proof. Since each of the bottom layer gates have fan-in at most k, there must exist a player who sees all the inputs to it. The protocol decides beforehand which gate 'belongs' to which player. All players simultaneously broadcast their contribution to the top SYM gate using at most $\log(s)$ bits each. □

A consequence of this is an upper bound for randomized protocols for depth-3 circuits, which may be found in [10] for example and is stated below without proof.

Lemma 8 (Folklore). *Given any function $f\in\mathrm{MAJ}\circ\mathrm{SYM}\circ\mathrm{ANY}_k$ of size s, and any partition of the input bits, there exists a randomized protocol computing f with advantage $\Omega(1/s)$ and cost $O(k\log(s))$.*

Let us now prove Theorem 2.

Proof. Suppose GHR_k^N could be computed by $\mathrm{MAJ}\circ\mathrm{SYM}\circ\mathrm{ANY}_k$ circuits of size $s=2^{o(\sqrt{N}/4^k)}$. Using the protocol mentioned in Lemma 8, the cost of the protocol is $O(k\log(s))$ and advantage $\Omega(1/s)$. Using Theorem 1, $O(k\log(s)+\log(s))\ge\Omega\left(\frac{\sqrt{N}}{4^k}-\log(N)-k\right)$, which gives $\log(s)\ge\Omega\left(\frac{\sqrt{N}}{4^k}-\frac{\log(N)}{k}-k\right)$ Thus, $s\ge 2^{\Omega\left(\frac{\sqrt{N}}{4^k}-\frac{\log(N)}{k}-1\right)}\ge 2^{\Omega\left(\frac{\sqrt{N}}{4^k}-\frac{\log(N)}{k}\right)}$ □

By definition, $\mathrm{MAJ}\circ\mathrm{MAJ}\subseteq\mathrm{MAJ}\circ\mathrm{SYM}$. Also, Goldmann et al. [15] (Theorem 26) showed that $\mathrm{MAJ}\circ\mathrm{THR}$ circuits can be simulated by $\mathrm{MAJ}\circ\mathrm{MAJ}$ circuits with a polynomial blowup. More precisely, a $\mathrm{MAJ}\circ\mathrm{THR}$ circuit of size s can be simulated by a $\mathrm{MAJ}\circ\mathrm{MAJ}$ circuit of size $s^\alpha\cdot n^\beta$ for some constants α,β. Hence, Corollary 2 follows by a similar proof as that of Lemma 8.

5 Conclusion

We have shown that GHR_k^N needs essentially $\Omega(\sqrt{N}/4^k)$ cost to be solved in the $\mathrm{PP}_{k+1}^{\mathrm{cc}}$ model. Since it follows almost from the definition of GHR_k^N that it has $O(\log N)$ cost $\mathrm{UPP}_{k+1}^{\mathrm{cc}}$ protocols, this provides the first separation of $\mathrm{PP}_k^{\mathrm{cc}}$ from $\mathrm{UPP}_k^{\mathrm{cc}}$ for the NOF model when $O(\log\log N)\le k\le\delta\cdot\log N$ for some constant $\delta>0$. In general, current techniques do not allow us to go beyond $\log N$ number of players to prove lower bounds for the cost of even deterministic protocols. This remains one of the most interesting problems in NOF complexity. However, let us remark that for many of the functions used in the literature (see for example [16, 3, 1, 13]), there are surprisingly efficient protocols when $k > \log N$. Moreover these protocols are typically deterministic and either simultaneous or barely interactive. On the other hand, we do not immediately see an efficient randomized interactive protocol for GHR_k^N at $k > \log N$. This raises the following question: Is GHR_k^N a hard function for even $k > \log N$?

Another question that may be within reach to answer is the following: our work shows that the $\mathrm{PP}_k^{\mathrm{cc}}$ complexity of GHR_k^N is $\Omega(\sqrt{N})$ for any constant k. Is there a function that has $\Omega(N)$ cost in $\mathrm{PP}_k^{\mathrm{cc}}$ but has efficient $\mathrm{UPP}_k^{\mathrm{cc}}$ protocols? This is open even for the two-player case.

Finally, proving super-logarithmic lower bounds for $\mathrm{UPP}_k^{\mathrm{cc}}$ protocols for any explicit function remains a very interesting challenge for even $k = 3$. Hansen and Podolskii [17] have shown that meeting this challenge is enough to yield super-polynomial lower bounds for $\mathrm{THR} \circ \mathrm{THR}$ circuits.

Acknowledgements

We are grateful to an anonymous reviewer for pointing out to us that the results of Sherstov [27] and Beigel [7] can be combined to get a separation between $\mathrm{PP}_k^{\mathrm{cc}}$ and $\mathrm{UPP}_k^{\mathrm{cc}}$ for k at most $O(\log \log n)$. We would also like to thank Kristoffer Hansen for directing our attention to the result of Goldmann, Håstad and Razborov [15].

References

[1] Anil Ada, Arkadev Chattopadhyay, Omar Fawzi, and Phuong Nguyen. The NOF multiparty communication complexity of composed functions. *Computational Complexity*, 24(3):645–694, 2015.

[2] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*, pages 337–347, 1986.

[3] László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication complexity of simultaneous messages. *SIAM J. Comput.*, 33(1):137–166, 2003.

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

[5] Paul Beame, Matei David, Toniann Pitassi, and Philipp Woelfel. Separating deterministic from nondeterministic NOF multiparty communication complexity. In *Automata, Languages and Programming, 34th International Colloquium, ICALP 2007, Wroclaw, Poland, July 9-13, 2007, Proceedings*, pages 134–145, 2007.

[6] Paul Beame and Dang-Trinh Huynh-Ngoc. Multiparty communication complexity and threshold circuit size of acˆ0. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, October 25-27, 2009, Atlanta, Georgia, USA*, pages 53–62, 2009.

[7] Richard Beigel. Perceptrons, pp, and the polynomial hierarchy. *Computational Complexity*, 4:339–349, 1994.

[8] Harry Buhrman, Nikolay Vereshchagin, and Ronald de Wolf. On computation and communication with small bias. In *Proceedings of the Twenty-Second Annual IEEE*

Conference on Computational Complexity, CCC '07, pages 24–32, Washington, DC, USA, 2007. IEEE Computer Society.

[9] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 25-27 April, 1983, Boston, Massachusetts, USA*, pages 94–99, 1983.

[10] Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*, pages 449–458, 2007.

[11] Arkadev Chattopadhyay. *Circuits, Communication and Polynomials.* PhD thesis, McGill University, 2009.

[12] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *CoRR*, abs/0801.3624, 2008.

[13] Arkadev Chattopadhyay and Michael E. Saks. The power of super-logarithmic number of players. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014, September 4-6, 2014, Barcelona, Spain*, pages 596–603, 2014.

[14] Andrew Drucker, Fabian Kuhn, and Rotem Oshman. On the power of the congested clique model. In *ACM Symposium on Principles of Distributed Computing, PODC '14, Paris, France, July 15-18, 2014*, pages 367–376, 2014.

[15] Mikael Goldmann, Johan Håstad, and Alexander A. Razborov. Majority gates VS. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

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

[17] Kristoffer Arnsfelt Hansen and Vladimir V. Podolskii. Polynomial threshold functions and boolean threshold circuits. *Inf. Comput.*, 240:56–73, 2015.

[18] Johan Håstad. On the size of weights for threshold gates. *SIAM J. Discrete Math.*, 7(3):484–492, 1994.

[19] Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, 1997.

[20] Troy Lee and Adi Shraibman. Disjointness is hard in the multi-party number-on-the-forehead model. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, CCC 2008, 23-26 June 2008, College Park, Maryland, USA*, pages 81–91, 2008.

[21] Mihai Patrascu. Towards polynomial lower bounds for dynamic problems. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 603–610, 2010.

[22] Anup Rao and Amir Yehudayoff. Simplified lower bounds on the multiparty communication complexity of disjointness. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, pages 88–101, 2015.

[23] Ran Raz. The bns-chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[24] Alexander A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

[25] Alexander A. Sherstov. The unbounded-error communication complexity of symmetric functions. *Combinatorica*, 31(5):583–614, 2011.

[26] Alexander A. Sherstov. The multiparty communication complexity of set disjointness. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 525–548, 2012.

[27] Alexander A. Sherstov. Communication lower bounds using directional derivatives. *J. ACM*, 61(6):34:1–34:71, 2014.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de