Electronic Colloquium on Computational Complexity, Report No. 14 (2008)

Separating NOF communication complexity classes RP and NP

Matei David
Computer Science Department
University of Toronto
matei at cs toronto edu

Toniann Pitassi*
Computer Science Department
University of Toronto
toni at cs toronto edu

February 25, 2008

Abstract

We provide a non-explicit separation of the number-on-forehead communication complexity classes RP and NP when the number of players is up to $\delta \cdot \log n$ for any $\delta < 1$. Recent lower bounds on Set-Disjointness [10, 7] provide an explicit separation between these classes when the number of players is only up to $o(\log\log n)$.

1 Introduction

In the number-on-forehead (NOF) model of communication complexity, k players are trying to evaluate a function F defined on kn bits. The input of F is partitioned into k pieces of n bits each, call them $x_1, \ldots, x_k$, and x_i is placed, metaphorically, on the forehead of player i. Thus, each player sees $(k-1)n$ of the kn input bits. The players communicate by writing bits on a shared blackboard in order to compute F. This model was introduced by [5] and it has many applications, including circuit lower bounds [9, 11], time/space tradeoffs for Turing Machines, pseudo-random number generators for space-bounded Turing Machines [2], and proof system lower bounds [4].

In this model, a protocol is said to be "efficient" if it has complexity $(\log n)^{O(1)}$. Correspondingly, P_k^{cc}, RP_k^{cc}, BPP_k^{cc} and NP_k^{cc} are the classes of functions having efficient deterministic, one-sided-error randomized, (two-sided-error) randomized and nondeterministic protocols, respectively. The usual inclusions between these classes apply, so $\mathsf{P}_k^{cc} \subseteq \mathsf{RP}_k^{cc} \subseteq \mathsf{NP}_k^{cc}$ and $\mathsf{RP}_k^{cc} \subseteq \mathsf{BPP}_k^{cc}$. One of the most fundamental questions in NOF communication complexity is to provide separations between these classes. In [3], Beame et al. show that $\mathsf{RP}_k^{cc} \neq \mathsf{P}_k^{cc}$ for $k \leq n^{O(1)}$ players. Recently, [7, 10] show that $\mathsf{NP}_k^{cc} \not\subseteq \mathsf{BPP}_k^{cc}$ (and thus, that $\mathsf{NP}_k^{cc} \neq \mathsf{RP}_k^{cc}$) for $k \leq o(\log\log n)$ players. Our main result in this paper is the following.

Theorem 1.1 (Main Theorem). *$\mathsf{NP}_k^{cc} \not\subseteq \mathsf{BPP}_k^{cc}$ (and thus, $\mathsf{NP}_k^{cc} \neq \mathsf{RP}_k^{cc}$) for all $\delta < 1$ and all $k \leq \delta \cdot \log n$.*

Until very recently, it was far from clear how to obtain communication complexity lower bounds in the number-on-forehead model for any function that could separate nondeterministic from randomized complexity. The difficulty can be described as follows. The only method currently known for obtaining multi-party NOF lower bounds is the discrepancy method [2, 13, 8]. Lower bounds using discrepancy are obtained

* Research supported by NSERC.

ISSN 1433-8092

by showing that the function in question has small discrepancy with respect to some distribution. Unfortunately, it is not hard to see that every function with small nondeterministic complexity has high discrepancy with respect to every distribution (see, for example, Lemma 3.1 in [7].) Thus, the discrepancy method seemed doomed to failure and new techniques seemed to be required.

However, in very recent work, these difficulties were overcome to obtain a surprisingly elegant lower bound for the Set-Disjointness function [7, 10]. The idea behind their proofs as well as ours is as follows.

In a recent paper, Sherstov [15] (and implicitly also in Razborov [14]) applied the discrepancy method in a more general way for the 2-player model in order to overcome the above difficulties. The *generalized* discrepancy method was adapted to the number-on-forehead model in [7, 10] and can be described at a high level as follows. Start with some candidate function F, where F has small nondeterministic complexity, and we want to prove that F has high randomized communication complexity. Now come up with a function G and a distribution λ such that: (1) F and G are highly correlated with respect to λ; and (2) G has small discrepancy with respect to λ. It is not hard to see that if such a G can be found, then since G has small discrepancy, it requires large randomized complexity, and moreover since F and G are very correlated, this in turn implies lower bounds on the randomized complexity of F as well.

Thus, to use the generalized discrepancy method, the problem is to come up with the functions F and G. To accomplish this, we will use another wonderful idea due to Sherstov [16], and substantially generalized to apply to the number-on-forehead setting by Chattopadhyay [6]. We consider special functions of the form F^{ϕ}. This will be a function on $(k+1)n$ bits, computed by $k+1$ players. Player 0 receives an n-bit vector x. Player i, for $1 \leq i \leq k$ gets an n-bit vector y_i. The function ϕ takes as input $y_1, \ldots, y_k$ and outputs an n-bit string z, where z has exactly m 1's. We will view ϕ a selecting m bits/indices of Player 0's input, x. The function F^{ϕ} will be the OR function applied to the m bits of x as specified by $\phi(y_1, \ldots y_k)$. (In earlier terminology, the $k+1$ players will apply the OR function to Player 0's *unmasked* input.)

Note that regardless of what function ϕ is chosen, F^{ϕ} will have a small nondeterministic protocol. Player 0 simply guesses an index j that is one of the indices chosen by ϕ, and then any of the other players can easily verify whether or not x_j is 1 in that position. When ϕ is the bitwise AND function, then F^{ϕ} is the Set-Disjointness function. We will show that for almost all ϕ, the randomized communication complexity of F^{ϕ} is large as long as k is at most a constant times $\log n$. Because we will be working with a random ϕ, as a bonus, our argument is substantially simpler that the previous bounds obtained for Set-Disjointness.

2 Definitions and Notation

2.1 Communication Complexity

In the number-on-forehead (NOF) multiparty communication complexity game [5] there are k players that are trying to collaborate to compute a function $F : X_1 \times \ldots \times X_k \to \{0,1\}$ where each $X_i = \{0,1\}^n$. The kn input bits are partitioned into k sets, each of size n. For $(x_1, \ldots, x_k) \in \{0,1\}^{kn}$, and for each i, player i knows the values of all of the inputs except for x_i (which conceptually is thought of as being placed on player i's forehead).

The players exchange bits according to an agreed-upon protocol, by writing them on a public blackboard. A *protocol* specifies, for every possible blackboard contents, whether or not the communication is over, the output if over and the next player to speak if not. A protocol also specifies what each player writes as a function of the blackboard contents and of the inputs seen by that player. The *cost* of a protocol is the maximum number of bits written on the blackboard.

In a *deterministic protocol*, the blackboard is initially empty. A *randomized protocol* of cost c is simply a probability distribution over deterministic protocols of cost c, which can be viewed as a protocol in which the players have access to a *shared* random string. A *non-deterministic protocol* is one where an initial guess string appears on the blackboard at the beginning of the protocol, and the players are trying to verify that the output of the function is 1 in the usual sense: there exists a guess string where the output of the protocol is 1 if and only if the output of the function is 1.

The *deterministic communication complexity of* F, written $D_k(F)$, is the minimum cost of a deterministic protocol for F that always outputs the correct answer. For $0 \leq \varepsilon < 1/2$, let $R_{k,\varepsilon}(F)$ denote the minimum cost of a randomized protocol for F which, for every input, makes an error with probability at most ε (over the choice of the deterministic protocols). The *(two-sided-error) randomized communication complexity of* F is $R_k(F) = R_{k,1/3}(F)$. Let $R^1_{k,\varepsilon}(F)$ denote the minimum cost of a randomized protocol for F which is correct on all 0-inputs, and for every 1-input, it makes an error with probability at most ε. The *one-sided-error randomized communication complexity of* F is $R^1_k(F) = R^1_{k,1/3}(F)$. The *non-deterministic communication complexity of* F, written $N_k(F)$, is the minimum cost of a non-deterministic protocol for F. We usually drop the subscript k when the number of players is clear from the context.

Since any function F_n on kn bits can be computed using only n bits of communication, following [1], for sequences of functions $F = (F_n)_{n\in\mathbb{N}}$, protocols are considered "efficient" or "polynomial" if only polylogarithmically many bits are exchanged. Accordingly, let P^{cc}_k, RP^{cc}_k, BPP^{cc}_k and NP^{cc}_k denote the classes of function families F for which $D_k(F_n), R^1_k(F_n), R_k(F_n)$ and $N_k(F_n)$ are $(\log n)^{O(1)}$, respectively.

Even though the standard communication complexity definitions above are given for functions with range $\{0,1\}$, we find it more convenient to work with the range $\{-1,1\}$. We transform the former into the latter by mapping $0 \to 1$ (representing *false*) and $1 \to -1$ (representing *true*). Thus, for example, when the range of F is $\{-1,1\}$, in a non-deterministic protocol the players are trying to verify that the output of F is -1.

The most important method to prove lower bounds for randomized communication complexity uses the concept of discrepancy. An *i-cylinder* Γ_i in $X_1 \times \ldots \times X_k$ is a set such that for all $x_1 \in X_1, \ldots, x_k \in X_k, x'_i \in X_i$ we have $(x_1,\ldots,x_i,\ldots,x_k) \in \Gamma_i$ if and only if $(x_1,\ldots,x'_i,\ldots,x_k) \in \Gamma_i$. A *cylinder intersection* is a set of the form $\bigcap_{i=1}^k \Gamma_i$ where each Γ_i is an i-cylinder in $X_1 \times \cdots \times X_k$. For a set S, let 1_S be its characteristic function, which is 1 if the input is in S and 0 otherwise. Let λ be a distribution on the inputs of F. The *discrepancy of* F *on* Γ *under* λ is $\mathrm{disc}^{\Gamma}_{k,\lambda}(F) = |\mathbb{E}_{\bar{x}\sim\lambda}[F(\bar{x})1_\Gamma(\bar{x})]|$. The *discrepancy of* F *under* λ is $\mathrm{disc}_{k,\lambda}(F) = \max_\Gamma \mathrm{disc}^{\Gamma}_{k,\lambda}(F)$. The *standard discrepancy method* [2] connects the discrepancy of a function F with its randomized communication complexity as follows: for every distribution λ, $R_{k,\varepsilon}(F) \geq \log\left(\frac{1-2\varepsilon}{\mathrm{disc}_{k,\lambda}(F)}\right)$.

2.2 Notation

Throughout this paper, the functions whose communication complexity we are analyzing are denoted by capital letters such as F. As mentioned in the introduction, we will be restricting our attention to certain functions which are constructed from a *base* function, usually denoted by lower case f, and a *masking* function, usually denoted by ϕ. In general, m denotes the size of the input to the base function f, and the range of this function is $\{-1,1\}$. A specific base function we will work with is the OR function, which takes on the value -1 if and only if any of its input bits is 1. The masking function ϕ takes as input k strings of n bits each, usually denoted by $y_1,\dots,y_k$, and it's output is an m-element subset of $[1,n]$. We always have $m \le n$. Starting with a base function f and a masking function ϕ, we construct a function $\mathrm{Lift}(f,\phi)$ on $(k+1)n$ input bits as follows. Given n-bit inputs $x,y_1,\dots,y_k$, ϕ is evaluated on the latter k inputs to select a set of m bits in x on which we apply f. Formally, $\mathrm{Lift}(f,\phi)(x,y_1,\dots,y_k) = f(x|\phi(y_1,\dots,y_k))$, where for a set $S \subseteq [1,n]$, $x|S$ denotes the substring of x indexed by the elements in S. We are interested in the communication complexity of $\mathrm{Lift}(f,\phi)$ in the NOF model with $k+1$ players, where player 0 gets x and players 1 through k get y_1 through y_k, respectively.

2.3 Correlation, Fourier Representation and Degree

Let $f,g : \{0,1\}^m \to \mathbb{R}$. Let μ be a distribution on the set $\{0,1\}^m$. We define the *correlation between f and g under μ* to be $\mathrm{corr}_\mu(f,g) = \mathbb{E}_{x\sim\mu}[f(x)g(x)]$. Whenever we omit to mention a specific distribution when computing the correlation, an expected value or a probability, it is to be assumed that we are talking about the uniform distribution.

For $S \subseteq [1,m]$, let $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$ be the Fourier character of the set S. Let $f : \{0,1\}^m \to \mathbb{R}$ and let $f_S = \mathrm{corr}(f,\chi_S)$. Then $f(x) = \sum_{S\subseteq[1,m]} f_S\chi_S(x)$ is the Fourier representation of f. The *exact degree of f* is the size of the largest S such that f_S is non-zero. The *ε-approximate degree of f*, denoted by $\deg_\varepsilon(f)$ is the smallest d for which there exists a function g of exact degree d such that $\max_x |f(x)-g(x)| \le \varepsilon$.

2.4 Set Families

Let $\mathcal{S} = (S_1,\dots,S_z)$ be a multi-set of m-element subsets of $[1,n]$. Let the *range* of $\mathcal{S}$, denoted by $\bigcup\mathcal{S}$, be the set of indices from $[1,n]$ that appear in at least one set in $\mathcal{S}$. Let the *boundary* of $\mathcal{S}$, denoted by $\partial\mathcal{S}$, be the set of indices from $[1,n]$ that appear in exactly one set in the collection $\mathcal{S}$.

3 Statement of Results

Our main technical result is the following.

Theorem 3.1. *Let $\delta < 1$ be a constant. Let $\varepsilon = (1-\delta)/4$. Let $m = n^\varepsilon$ and let $k \le \delta\cdot\log n$. There exists a function ϕ such that $R_{k+1}(\mathrm{Lift}(\mathrm{OR},\phi)) \ge n^{\Omega(1)}$.*

Proof of Main Theorem 1.1 from Theorem 3.1. Consider the function ϕ whose existence is guaranteed by Theorem 3.1. On the one hand, the Theorem implies that $\mathrm{Lift}(\mathrm{OR},\phi) \notin \mathsf{BPP}^{cc}_{k+1}$.

On the other hand, the following is a nondeterministic protocol for $\mathrm{Lift}(\mathrm{OR},\phi)$: guess an index $i \in [1,n]$ using $\log n$ bits; player 0 (the one holding x on its forehead) locally computes $\phi(y_1,\dots,y_k)$ and communicates a 1 if i belongs to that set; player 1 communicates a 1 if $x_i = 1$. The cost of this protocol is $O(\log n)$. Easily, $\mathrm{Lift}(\mathrm{OR},\phi)(x,y_1,\dots,y_k) = -1$ iff there exists a guess i such that both players communicate a 1. Thus, $\mathrm{Lift}(\mathrm{OR},\phi) \in \mathsf{NP}^{cc}_{k+1}$. $\square$

4 Proof of Main Result

We obtain our lower bounds on the bounded-error communication complexity of $\mathrm{Lift}(\mathrm{OR},\phi)$ using an analysis that follows [7]. In their paper, Chattopadhyay and Ada analyze the Set-Disjointness function, and for that reason, their masking function ϕ must be the AND function. In our case, intuitively, we allow ϕ to be a random function. While our results no longer apply to Set-Disjointness, we still obtain a separation between BPP_k^{cc} and NP_k^{cc} because, no matter what masking function is used, $\mathrm{Lift}(\mathrm{OR},\phi)$ always has a cheap nondeterministic protocol.

At a more technical level, the results of [7] become trivial when $k \geq \log\log n$ because of the relationship between n (the size of the input to F) and m (the number of bits the base function OR gets applied to.) For their analysis to go through, they need $n = 2^{2^k} m^{O(1)}$. In our case, $n = m^{O(1)}$ is sufficient, and this allows our results to be non-trivial for $k \leq \delta \log n$ for any $\delta < 1$.

4.1 Overview of Proof

As mentioned earlier, we will start with the base function $f = \mathrm{OR}$ on m input bits, $m < n$. We lift the base function f in order to obtain the lifted function $F^\phi = \mathrm{Lift}(f,\phi)$. Recall that F^ϕ is a function on $(k+1)n$ inputs with small nondeterministic complexity, and is obtained by applying the base function (in this case the OR function) to the unmasked bits of Player 0's input, x. We want to prove that for a random ϕ, F^ϕ has high randomized communication complexity.

Paturi [12] proved that no function that is a sum of low-degree Fourier characters can well-approximate the OR function. This implies that there exists a function g (also on m bits) and a distribution μ over all m-bit inputs such that the functions g and $f = \mathrm{OR}$ are highly correlated over μ and furthermore, g is orthogonal to all small Fourier characters. This is our Lemma 4.1, and it was originally proved using duality by Sherstov [15] in the context of 2-player lower bounds for quantum communication complexity.

Now we lift the function g in order to get the function $G^\phi = \mathrm{Lift}(g,\phi)$. Define λ to be a distribution over all $(k+1)n$-bit inputs that is the natural extension of μ. Since g and $f = \mathrm{OR}$ are highly correlated over μ, it is not hard to see (using the definitions and the fact that λ is the natural extension of μ to the lifted space) that the lifted versions, F^ϕ and G^ϕ are also highly correlated over λ.

By the generalized discrepancy method (Lemma 4.2), in order to prove that the randomized complexity of F^ϕ is high, it suffices to prove that G^ϕ has small discrepancy. This final step is accomplished by Lemmas 4.4, 4.5, and 4.6, using two important properties of g and ϕ. The crucial property of g that we exploit is that it is orthogonal to the space of all small Fourier characters. This property will be used to prove Lemma 4.4. Secondly, we want ϕ to behave like a random function with respect to all sub-cubes. This second property is exploited in order to prove Lemma 4.6. We now proceed with the formal proof.

4.2 Proof of Main Theorem

The following lemma is from [15]. Intuitively it shows the following. Let f be a base function on m bits, and with the property that no function in the low-degree Fourier subspace can approximate f. (We will be interested in $f = \mathrm{OR}$.) The lemma states that this implies the existence of another function g and a distribution μ such that g is in the orthogonal subspace of low-degree Fourier characters and g wellapproximates f.

Lemma 4.1 (Orthogonality Lemma, Lemma 5.1 in [7])**.** *If $f : \{0,1\}^m \to \{-1,1\}$ is a function with δ'-approximate degree d, there exist a function $g : \{0,1\}^m \to \{-1,1\}$ and a distribution μ on $\{0,1\}^m$ such that:*

(i) $\mathrm{corr}_\mu(g, f) \geq \delta'$*; and*

(ii) for every $T \subseteq [1,m]$ with $|T| \leq d$ and every function $h : \{0,1\}^{|T|} \to \mathbb{R}$, $\mathbb{E}_{x\sim\mu}[g(x)\cdot h(x|T)] = 0$.

The next lemma is the generalized discrepancy lemma from [7]. It states that if two functions F and G are highly correlated, and if G has small discrepancy (and hence high communication complexity), then the communication complexity of F is also high.

Lemma 4.2 (Generalized Discrepancy Lemma, Lemma 3.2 in [7])**.** *Let $Z = Z_1 \times \cdots \times Z_k$. Let $F, G : Z \to \{-1,1\}$ and let λ be a distribution on Z such that* $\mathrm{corr}_\lambda(G,F) \geq \delta'$*. Then, for every $\varepsilon' < \delta'/2$,*

$$R_{k,\varepsilon'}(F) \geq \log\left(\frac{\delta' - 2\cdot\varepsilon'}{\mathrm{disc}_{k,\lambda}(G)}\right).$$

The following lemma is standard and used in every discrepancy argument. See [2, 13, 8] for details.

Lemma 4.3 (The standard BNS argument)**.** *Let $Z = X \times Y_1 \times \cdots \times Y_k$ and let $F : Z \to \{-1,1\}$. Let $\Gamma \subseteq Z$ be a cylinder intersection. We write $\bar{y}$ for $(y_1, \ldots, y_k)$. Then,*

$$\left(\mathbb{E}_{x,\bar{y}}\left[F(x,\bar{y})1_\Gamma(x,\bar{y})\right]\right)^{2^k} \leq \mathbb{E}_{\bar{y}^0,\bar{y}^1}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k} F(x, y_1^{u_1}, \ldots, y_k^{u_k})\right]\right|\right].$$

Using the above lemmas, We will now prove Theorem 3.1. By [12], $\deg_{5/6}(\mathrm{OR}) \geq c\sqrt{m}$ for some constant c. By Lemma 4.1, applied with $f = \mathrm{OR}$, there exist a function g and a distribution μ such that:

(i) $\mathrm{corr}_\mu(g, \mathrm{OR}) \geq 5/6$; and

(ii) for every $T \subseteq [1,m]$ with $T \leq c\sqrt{m}$ and every function $h : \{0,1\}^{|T|} \to \mathbb{R}$, $\mathbb{E}_{x\sim\mu}[g(x)h(x|T)] = 0$.

For every masking function ϕ, let $F^\phi = \mathrm{Lift}(\mathrm{OR}, \phi)$ and let $G^\phi = \mathrm{Lift}(g, \phi)$. As in [7], we define the distribution λ on $\{0,1\}^{(k+1)n}$ as follows. For $x \in \{0,1\}^n$ and $\bar{y} = (y_1, \ldots, y_k) \in \{0,1\}^{kn}$, let

$$\lambda(x,\bar{y}) = \frac{\mu(x|\phi(\bar{y}))}{2^{(k+1)n-m}}.$$

It can be easily verified that $\mathrm{corr}_\lambda(G^\phi, F^\phi) = \mathrm{corr}_\mu(g, \mathrm{OR}) \geq 5/6$. Thus, by Lemma 4.2,

$$R(F^\phi) \geq \log\left(\frac{5/6 - 2(1/3)}{\mathrm{disc}_\lambda(G^\phi)}\right) = \log\left(\frac{1}{\mathrm{disc}_\lambda(G^\phi)}\right) - \Theta(1).$$

Let Γ be the cylinder intersection that witnesses the discrepancy of G^ϕ under λ. Then,

$$\mathrm{disc}_\lambda(G^\phi) = \mathrm{disc}_\lambda^\Gamma(G^\phi) = \left|\mathbb{E}_{(x,\bar{y})\sim\lambda}[G^\phi(x,\bar{y})1_\Gamma(x,\bar{y})]\right| = 2^m \left|\mathbb{E}_{x,\bar{y}}[\mu(x|\phi(\bar{y}))g(x|\phi(\bar{y}))1_\Gamma(x,\bar{y})]\right|$$

where the last equality follows from the connection between λ and the uniform distribution. Finally, by Lemma 4.3, we obtain

$$\forall\phi,\left(\mathrm{disc}_{\lambda}(G^{\phi})\right)^{2^k}\le 2^{m2^k}\mathbb{E}_{\overline{y}^0,\overline{y}^1}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|\phi(y_1^{u_1},\ldots,y_k^{u_k}))g(x|\phi(y_1^{u_1},\ldots,y_k^{u_k}))\right]\right|\right].$$

It is at this point that we diverge from the analysis in [7]. Let $A=A(\overline{y}^0,\overline{y}^1)$ be the event "$\exists i$ such that $y_i^0=y_i^1$". Clearly, this event depends only on the choice of $\overline{y}^0$ and $\overline{y}^1$. By a simple union bound, $\Pr_{\overline{y}^0,\overline{y}^1}[A]\le k/2^n=2^{-n+logk}$. Furthermore, $\Pr_{\overline{y}^0,\overline{y}^1}[\overline{A}]\le 1$, and since $|\mu g|\le 1$, $\mathbb{E}_{\overline{y}^0,\overline{y}^1}[\ldots|\overline{A}]\le 1$. Thus,

$$\forall\phi,\left(\mathrm{disc}_{\lambda}(G^{\phi})\right)^{2^k}\le 2^{-n+m2^k+\log k}+2^{m2^k}\mathbb{E}_{\overline{y}^0,\overline{y}^1}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|\phi(y_1^{u_1},\ldots,y_k^{u_k}))g(x|\phi(y_1^{u_1},\ldots,y_k^{u_k}))\right]\right|\Big|\overline{A}\right].$$

For the remaining part of the analysis, we fix the choices of $\overline{y}^0$ and $\overline{y}^1$ in such a way that the event A does not occur. For $u\in\{0,1\}^k$, define $S_u=S_u(\overline{y}^0,\overline{y}^1,\phi)=\phi(y_1^{u_1},\ldots,y_k^{u_k})$. Let $\mathcal{S}=\mathcal{S}(\overline{y}^0,\overline{y}^1,\phi)$ be the multi-set $(S_u : u\in\{0,1\}^k)$. Even though the sets S_u and the multi-set $\mathcal{S}$ depend on $\overline{y}^0,\overline{y}^1$ and ϕ, we will usually omit explicitly indicating this dependence in our proofs in order to reduce the clutter. We define *the number of conflicts in* $\mathcal{S}$ to be $q(\mathcal{S})=m2^k-|\bigcup\mathcal{S}|$. Intuitively, $|\bigcup\mathcal{S}|$ measures the range of $\mathcal{S}$, while $m2^k$ is the maximum possible value for this range.

We use the following three Lemmas to complete our proof.

Lemma 4.4. *For every $\overline{y}^0,\overline{y}^1$ and ϕ, if $\overline{A}(\overline{y}^0,\overline{y}^1)$ and $q(\mathcal{S}(\overline{y}^0,\overline{y}^1,\phi))<c\cdot\sqrt{m}\cdot 2^k/2$, then*

$$\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u(\overline{y}^0,\overline{y}^1,\phi))g(x|S_u(\overline{y}^0,\overline{y}^1,\phi))\right]=0.$$

Lemma 4.5. *For every $\overline{y}^0,\overline{y}^1$ and ϕ, if $\overline{A}(\overline{y}^0,\overline{y}^1)$,*

$$\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u(\overline{y}^0,\overline{y}^1,\phi))\right]\le\frac{2^{q(\mathcal{S}(\overline{y}^0,\overline{y}^1,\phi))}}{2^{m\cdot 2^k}}.$$

Lemma 4.6. *For every $\overline{y}^0,\overline{y}^1$, if $\overline{A}(\overline{y}^0,\overline{y}^1)$, when ϕ is chosen at random,*

$$\Pr_{\phi}[q(\mathcal{S}(\overline{y}^0,\overline{y}^1,\phi))=q|\overline{A}(\overline{y}^0,\overline{y}^1)]\le\left(\frac{m\cdot 2^k}{n}\right)^q.$$

Before proving these Lemmas, we complete the proof of our main Theorem. Since the bound on $\mathrm{disc}_{\lambda}(G^{\phi})$ holds for every ϕ, we can write

$$\mathbb{E}_{\phi}\left[\left(\mathrm{disc}_{\lambda}(G^{\phi})\right)^{2^k}\right]\le 2^{-n+m2^k+\log k}+2^{m2^k}\mathbb{E}_{\overline{y}^0,\overline{y}^1,\phi}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)g(x|S_u)\right]\right|\Big|\overline{A}\right].$$

Moreover,

$$\begin{aligned}
& \mathbb{E}_{\overline{y}^0,\overline{y}^1,\phi}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)g(x|S_u)\right]\right|\Big|\overline{A}\right] \\
\leq\ & \sum_{q\geq 0}\Pr_{\phi}[q(\mathcal{S})=q|\overline{A}]\mathbb{E}_{\overline{y}^0,\overline{y}^1,\phi}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)g(x|S_u)\right]\right|\Big|\overline{A},q(\mathcal{S})=q\right] \\
\text{(by Lemma 4.4)}\quad \leq\ & \sum_{q\geq c\sqrt{m}2^k/2}\Pr_{\phi}[q(\mathcal{S})=q|\overline{A}]\mathbb{E}_{\overline{y}^0,\overline{y}^1,\phi}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)g(x|S_u)\right]\right|\Big|\overline{A},q(\mathcal{S})=q\right] \\
\text{(because } |g|=1)\quad \leq\ & \sum_{q\geq c\sqrt{m}2^k/2}\Pr_{\phi}[q(\mathcal{S})=q|\overline{A}]\mathbb{E}_{\overline{y}^0,\overline{y}^1,\phi}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)\right]\right|\Big|\overline{A},q(\mathcal{S})=q\right] \\
\text{(by Lemma 4.5)}\quad \leq\ & \sum_{q\geq c\sqrt{m}2^k/2}\Pr_{\phi}[q(\mathcal{S})=q|\overline{A}]\frac{2^q}{2^{m2^k}} \\
\text{(by Lemma 4.6)}\quad \leq\ & \sum_{q\geq c\sqrt{m}2^k/2}\left(\frac{m2^k}{n}\right)^q\frac{2^q}{2^{m2^k}} \\
=\ & \frac{1}{2^{m2^k}}\sum_{q\geq c\sqrt{m}2^k/2}\left(\frac{2m2^k}{n}\right)^q.
\end{aligned}$$

We have chosen $\varepsilon=(1-\delta)/4$, so $1-\varepsilon-\delta=3\varepsilon$. Furthermore, $m=n^\varepsilon$ and $k\leq\delta\log n$, so $m2^k/n\leq n^{-1+\varepsilon+\delta}=n^{-3\varepsilon}<1/4$ when n is large enough. Thus, $2m2^k/n<1/2$. Using $\sum_{q\geq q_0}w^q=w^{q_0}/(1-w)\leq 2w^{q_0}$ for $w<1/2$, we obtain

$$\mathbb{E}_{\overline{y}^0,\overline{y}^1,\phi}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)g(x|S_u)\right]\right|\Big|\overline{A}\right]\leq\frac{2^{1-c\sqrt{m}2^k/2}}{2^{m2^k}}.$$

Putting everything together,

$$\mathbb{E}_\phi\left[\left(\mathrm{disc}_\lambda(G^\phi)\right)^{2^k}\right]\leq 2^{-n+m2^k+\log k}+2^{m2^k}2^{-m2^k}2^{1-c\sqrt{m}2^k/2}.$$

For the exponent of the first term, note that $\log k\leq m2^k$ and $n\geq 4m2^k$, so $-n+m2^k+\log k\leq -2m2^k$. When m is large enough, $-2m2^k\leq -c\sqrt{m}2^k/4$. For the exponent of the second term, note that $1\leq c\sqrt{m}2^k/4$ when m is large enough, so $1-c\sqrt{m}2^k/2\leq -c\sqrt{m}2^k/4$. Thus, the sum of the two terms is at most $2^{1-c\sqrt{m}2^k/4}$. When m is large enough, $1\leq c\sqrt{m}2^k/8$, so

$$\mathbb{E}_\phi\left[\left(\mathrm{disc}_\lambda(G^\phi)\right)^{2^k}\right]\leq 2^{-c\sqrt{m}2^k/8}.$$

Therefore, there exists some ϕ such that $\mathrm{disc}_\lambda(G^\phi)\leq 2^{-c\sqrt{m}/8}$. For this ϕ,

$$R(F^\phi)\geq\log\left(\frac{1}{\mathrm{disc}_\lambda(G^\phi)}\right)-\Theta(1)\geq\Theta(1)\sqrt{m}=\Theta(1)n^\varepsilon\geq n^{\Omega(1)}.$$

5 Proofs of Lemmas

Proof of Lemma 4.4. We write S_u for $S_u(\bar{y}^0,\bar{y}^1,\phi)$ and $\mathcal{S}$ for $\mathcal{S}(\bar{y}^0,\bar{y}^1,\phi)$. Assume $q(\mathcal{S}) < c\sqrt{m}2^k/2$. Let $r(\mathcal{S}) = |\bigcup \mathcal{S}|$ be the size of the range of $\mathcal{S}$, and let $b(\mathcal{S}) = |\partial \mathcal{S}|$ be the size of the boundary of $\mathcal{S}$. Note that $r(\mathcal{S}) - b(\mathcal{S}) \leq q(\mathcal{S})$ because every $j \in \cup\mathcal{S} \setminus \partial\mathcal{S}$ occurs in at least 2 sets in $\mathcal{S}$, thus contributes at least 1 to $q(\mathcal{S})$. Furthermore, $r(\mathcal{S}) + q(\mathcal{S}) = m2^k$. Then, $b(\mathcal{S}) \geq r(\mathcal{S}) - q(\mathcal{S}) = m2^k - 2q(\mathcal{S}) > (m - c\sqrt{m})2^k$. There are 2^k sets in the multi-set $\mathcal{S}$ so by the pigeonhole principle, there exists v such that $|S_v \cap \partial\mathcal{S}| > m - c\sqrt{m}$. We can write

$$\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)g(x|S_u)\right] = \mathbb{E}_{x|S_v}\left[\mu(x|S_v)g(x|S_v)\mathbb{E}_{x|[1,n]\setminus S_v}\left[\prod_{u\in\{0,1\}^k,u\neq v}\mu(x|S_u)g(x|S_u)\right]\right].$$

Let $T = S_v \setminus \partial\mathcal{S}$. So $|T| \leq c\sqrt{m}$. Let $h = \mathbb{E}_{x|[1,n]\setminus S_v}\left[\prod_{u\neq v}\mu(x|S_u)g(x|S_u)\right]$. Note that h is a function that depends only on $x|T$. Then, by the property (ii) of g and μ, $\mathbb{E}_{x|S_v}[\mu(x|S_v)g(x|S_v)h(x|T)] = 0$. ☐

Proof of Lemma 4.5. We write S_u for $S_u(\bar{y}^0,\bar{y}^1,\phi)$ and $\mathcal{S}$ for $\mathcal{S}(\bar{y}^0,\bar{y}^1,\phi)$. We see that

$$\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)\right] = \mathbb{E}_{x|[1,n]\setminus\bigcup\mathcal{S}}\left[\mathbb{E}_{x|\bigcup\mathcal{S}}\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)\right]\right] = \mathbb{E}_{x|\bigcup\mathcal{S}}\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)\right].$$

Every $u \in \{0,1\}^k$ can be interpreted as an integer in the range $[0,2^k-1]$. With this in mind, for $0 \leq j \leq 2^k - 1$, let $\mathcal{S}_j$ be the sub-multi-set of $\mathcal{S}$ consisting of the sets up to and including S_j, $\mathcal{S}_j = (S_0,\ldots,S_j)$. So, $\mathcal{S} = \mathcal{S}_{2^k-1}$. Define $\mathcal{S}_{-1} = \emptyset$. For $0 \leq j \leq 2^k-1$, let $G_j = \mathbb{E}_{x|\bigcup\mathcal{S}_j}[\prod_{i=0}^{j}\mu(x|S_i)]$ and let $H_j(x|S_j\setminus\partial\mathcal{S}_j) = \mathbb{E}_{x|S_j\cap\partial\mathcal{S}_j}[\mu(x|S_j)]$. Letting $G_{-1} = 1$, observe that, for $0 \leq j \leq 2^k-1$,

$$G_j = \mathbb{E}_{x|\bigcup\mathcal{S}_{j-1}}\left[\left(\prod_{i=0}^{j-1}\mu(x|S_i)\right)H_j(x|S_j\setminus\partial\mathcal{S}_j)\right] \leq (\max(H_j))\cdot G_{j-1}.$$

To obtain a bound on $\max(H_j)$, consider an arbitrary partition of $[1,m]$ into two sets E,F. Let ν be a distribution on $[1,m]$, and let $\rho(x|E) = \mathbb{E}_{x|F}[\nu(x)]$. Then, $\rho(x|E) = \sum_{x|F}2^{-|F|}\nu(x) = 2^{-|F|}\sum_{x|F}\nu(x) \leq 2^{-|F|} = 2^{|E|-m}$, simply using the fact that ν is a probability distribution. Thus, $\max(H_j) \leq 2^{|S_j\setminus\partial\mathcal{S}_j|-m}$. Inductively,

$$\mathbb{E}_x\left[\prod_{i=0}^{2^k-1}\mu(x|S_i)\right] = G_{2^k-1} \leq \frac{2^{\sum_{j=0}^{2^k-1}|S_j\setminus\partial\mathcal{S}_j|}}{2^{m2^k}}.$$

Consider some index $z \in \bigcup\mathcal{S}$. Suppose this index appears in l sets $S_{j_1},\ldots,S_{j_l}$ from $\mathcal{S}$, with $j_1 < \cdots < j_l$. Then, this index contributes exactly $l-1$ to the expression $\sum_{j=0}^{2^k-1}|S_j\setminus\partial\mathcal{S}_j|$, once for every $j = j_2,\ldots,j_l$ (for $j = j_1$, $z \in \partial\mathcal{S}_j$ because no set before S_j contains z.) Since this holds for every index z, we see that $\sum_{j=0}^{2^k-1}|S_j\setminus\partial\mathcal{S}_j| = q(\mathcal{S})$ and therefore $\mathbb{E}_x[\prod_{u\in\{0,1\}^k}\mu(x|S_u)] \leq 2^{q(\mathcal{S})-m2^k}$. ☐

Proof of Lemma 4.6. Fix $\bar{y}^0,\bar{y}^1$ such that $\overline{A}$. The multi-set $\mathcal{S}$ is constructed from the sets $S_u = \phi(y_1^{u_1},\ldots,y_k^{u_k})$ for $u \in \{0,1\}^k$. Since A did not occur, the 2^k points where ϕ gets evaluated are distinct. Furthermore, ϕ is chosen at random, which is equivalent to choosing 2^k random m-element subsets of $[1,n]$. We can overestimate the number of conflicts in $\mathcal{S}$ as follows. Instead of choosing, for each subset, m elements from $[1,n]$ *without* replacement, suppose we chose them *with* replacement. The number of conflicts we will obtain can only be larger than in the original experiment or, equivalently, the probability of obtaining a fixed number of conflicts can only be greater in the second experiment. The maximum range of $\mathbb{S}$ is $m2^k$. Every conflict in $\mathbb{S}$ arises when we select a previously selected point from $[1,n]$. Thus, the probability of each conflict is independently at most $m2^k/n$. The probability of obtaining q conflicts is at most $(m2^k/n)^q$. □

References

[1] László Babai, P. Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *27th Annual Symposium on Foundations of Computer Science*, pages 337–347, Toronto, Ontario, October 1986. IEEE.

[2] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2):204–232, October 1992.

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

[4] P. Beame, P. Pitassi, and N. Segerlind. Lower bounds for lovasz-schrijver systems and beyond follow from multiparty communication complexity. In *Proceedings from Thirty-second ICALP*. IEEE, 2005.

[5] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, pages 94–99, Boston, MA, April 1983.

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

[7] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. In *Electronic Colloquium on Computational Complexity TR08-002*, 2008.

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

[9] Johan Håstad and M. Goldmann. On the power of small-depth threshold circuits. In *Proceedings 31st Annual Symposium on Foundations of Computer Science*, pages 610–618, St. Louis, MO, October 1990. IEEE.

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

[11] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. In *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*, pages 419–429, New Orleans, LA, May 1991.

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

[13] R. Raz. The bns-chung criterion for multiparty communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

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

[15] A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Electronic Colloquium on Computational Complexity, TR 07-100*, 2007.

[16] A. Sherstov. Separating ac^0 from depth-2 majority circuits. In *ACM FOCS*, pages 294–301, 2007.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/