Improved Separations between Nondeterministic and Randomized Multiparty Communication

Matei David[*] Toniann Pitassi[*] Emanuele Viola[†]

September 10, 2008

Abstract

We exhibit an explicit function $f : \{0,1\}^n \to \{0,1\}$ that can be computed by a nondeterministic number-on-forehead protocol communicating $O(\log n)$ bits, but that requires $n^{\Omega(1)}$ bits of communication for randomized number-on-forehead protocols with $k = \delta \cdot \log n$ players, for any fixed $\delta < 1$. Recent breakthrough results for the Set-Disjointness function (Sherstov, STOC '08; Lee Shraibman, CCC '08; Chattopadhyay Ada, ECCC '08) imply such a separation but only when the number of players is $k < \log\log n$.

We also show that for any $k = A \log\log n$ the above function f is computable by a small circuit whose depth is constant whenever A is a (possibly large) constant. Recent results again give such functions but only when the number of players is $k < \log\log n$.

[*] Computer Science Department, University of Toronto. {matei,toni}@cs.toronto.edu. Supported by NSERC.

[†] Computer Science Department, Columbia University. viola@cs.columbia.edu. Supported by grants NSF award CCF-0347282 and NSF award CCF-0523664.

1 Introduction

Number-on-forehead communication protocols are a fascinating model of computation where k collaborating players are trying to evaluate a function $f : (\{0,1\}^n)^k \rightarrow \{0,1\}$. The players are all-powerful, but the input to f is partitioned into k pieces of n bits each, $x_1, \ldots, x_k \in \{0,1\}^n$, and x_i is placed, metaphorically, on the forehead of player i. Thus, each player only sees $(k-1)n$ of the $k \cdot n$ input bits. In order to compute f, the players communicate by writing bits on a shared blackboard, and the complexity of the protocol is the number of bits that are communicated (i.e., written on the board). This model was introduced in [CFL83] and has found applications in a surprising variety of areas, including circuit complexity [HG91, NW93], pseudorandomness [BNS92], and proof complexity [BPS07].

In this model, a protocol is said to be *efficient* if it has complexity $\log^{O(1)} n$. Correspondingly, P_k^{cc}, RP_k^{cc}, BPP_k^{cc} and NP_k^{cc} are the number-on-forehead communication complexity analogs of the standard complexity classes [BFS86], see also [KN97]. For example, RP_k^{cc} is the class of functions having efficient one-sided-error randomized communication protocols. One of the most fundamental questions in NOF communication complexity, and the main question addressed in this paper, is to separate these classes. In [BDPW07], Beame et al. give an exponential separation between randomized and deterministic protocols for $k \leq n^{O(1)}$ players (in particular, $\mathsf{RP}_k^{cc} \neq \mathsf{P}_k^{cc}$ for $k \leq n^{O(1)}$). The breakthrough work by Sherstov [She07, She08a] sparked a flurry of exciting results in communication complexity [Cha07, LS08, CA08] which gave an exponential separation between nondeterministic and randomized protocols for $k < \log\log n$ players (in particular, $\mathsf{NP}_k^{cc} \not\subseteq \mathsf{BPP}_k^{cc}$ for $k < \log\log n$). Our main result is to improve the latter separation to larger values of k.

Theorem 1.1 (Main Theorem; $\mathsf{NP}_k^{cc} \not\subseteq \mathsf{BPP}_k^{cc}$ for $k = \delta \log n$ players)**.** *For every fixed $\delta < 1$, sufficiently large n and $k = \delta \cdot \log n$, there is an explicit function $f : (\{0,1\}^n)^k \rightarrow \{0,1\}$ such that: f can be computed by k-player nondeterministic protocols communicating $O(\log n)$ bits, but f cannot be computed by k-player randomized protocols communicating $n^{o(1)}$ bits.*

We note that the number of players $k = \delta \cdot \log n$ in the above Theorem 1.1 is state-of-the-art: it is a major open problem in number-on-forehead communication complexity to determine if every explicit function on n bits can be computed by $k = \log_2 n$ players communicating $O(\log n)$ bits. We also note that Theorem 1.1 in particular implies an exponential separation between nondeterministic and deterministic protocols (hence, $\mathsf{NP}_k^{cc} \not\subseteq \mathsf{P}_k^{cc}$ for $k = \delta \log n$ players). Similar separations follow from [BDPW07], but only for non-explicit functions.

We also address the challenge of exhibiting functions computable by small (unbounded fan-in) constant-depth circuits that require high communication for k-player protocols, which is relevant to separating various circuit classes (see, e.g., [HG91, RW93]). Previous results [Cha07, LS08, CA08] give such functions for $k < \log\log n$. We offer a slight improvement and achieve $k = A \log\log n$ for any (possibly large) constant A, where the depth of the circuit computing the function depends on A.

Theorem 1.2 (Constant-depth circuits require high communication for $k = A \log\log n$ players)**.** *For every constant $A > 1$ there is a constant B such that for sufficiently large n and $k := A \log\log n$ there is a function $f : (\{0,1\}^n)^k \rightarrow \{0,1\}$ which satisfies the following: f can be computed by circuits of* *size n^B and depth B, but f cannot be computed by k-player randomized protocols communicating $n^{o(1)}$ bits.*

1.1 Techniques

In this section we discuss the technical challenges presented by our theorems and how we have overcome them, building on previous work. An exposition of previous works and of some of the ideas in this paper also appears in the survey by Sherstov [She08b]. For concreteness, in our discussion we focus on separating nondeterministic from deterministic (as opposed to randomized) protocols, a goal which involves all the main difficulties.

Until very recently, it was far from clear how to obtain communication lower bounds in the number-on-forehead model for any explicit function f with efficient nondeterministic protocols. The difficulty can be described as follows. The standard method for obtaining number-on-forehead lower bounds is what can be called the "correlation method" [BNS92, CT93, Raz00, VW07].[1] This method goes by showing that f has *exponentially small* ($2^{-n^{\Omega(1)}}$) correlation with efficient (deterministic) protocols, and this immediately implies that f does not have efficient protocols (the correlation is w.r.t. some probability distribution which in general is not uniform). The drawback of this method is that, although for the conclusion that f does not have efficient protocols it is clearly enough to show that the correlation of f with such protocols is strictly less than one, the method actually proves the stronger exponentially small correlation bound. This is problematic in our setting because it is not hard to see that every function that has an efficient nondeterministic protocol also has *noticeable* ($\geq 2^{-\log^{O(1)} n}$) correlation with an efficient (deterministic) protocol, and thus this method does not seem useful for separating nondeterministic from deterministic protocols.

In recent work, these difficulties were overcome to obtain a surprising lower bound for a function with an efficient nondeterministic protocol: the Set-Disjointness function [LS08, CA08]. The starting point is the work by Sherstov [She08a] who applies the correlation method in a more general way for the 2-player model in order to overcome the above difficulties. This *generalized* correlation method is then adapted to handle more players $(k \gg 2)$ in [LS08, CA08]. The high-level idea of the method is as follows. Suppose that we want to prove that some specific function f does not have efficient protocols. The idea is to come up with another function f' and a distribution λ such that: (1) f and f' have constant correlation, say f and f' disagree on at most $1/10$ mass of the inputs with respect to λ, and (2) f' has exponentially small ($2^{-n^{\Omega(1)}}$) correlation with efficient protocols with respect to λ. The combination of (1) and (2) easily implies that f also has correlation at most $1/10 + 2^{-n^{\Omega(1)}} < 1$ with efficient protocols, which gives the desired lower bound for f. This method is useful because for f' we can use the correlation method, and on the other hand the correlation of f with efficient protocols is *not* shown to be exponentially small, only bounded away from 1 by a constant. Thus it is conceivable that f has efficient nondeterministic protocols, and in fact this is the case in [LS08, CA08] and in this work.

Although a framework similar to the above is already proposed in previous papers, e.g. [Raz87, Raz03], it is the work by Sherstov [She08a] that finds a way to successfully apply it to functions

[1] This method is sometimes called the "discrepancy method." We believe that lower bound proofs are easier to understand when presented in terms of correlation rather than discrepancy, cf. [VW07].

f with efficient nondeterministic protocols. For this, [She08a] uses two main ideas, generalized to apply to the number-on-forehead setting in [Cha07, LS08, CA08]. The first is to consider a special class of functions $f := \text{Lift}(\text{OR}, \phi)$ with efficient nondeterministic protocols. These are obtained by combining the "base" function OR on m bits with a "selection" function ϕ as described next. It is convenient to think of $f = \text{Lift}(\text{OR}, \phi)$ as a function on $(k+1)n$ bits distributed among $k+1$ players as follows: Player 0 receives an n-bit vector x, while Player i, for $1 \leq i \leq k$, gets an n-bit vector y_i. The selection function ϕ takes as input $y_1, \ldots, y_k$ and outputs an m-bit subset of $\{1, \ldots, n\}$. We view ϕ as selecting m bits of Player 0's input x, denoted $x|\phi(y_1, \ldots, y_k)$. $\text{Lift}(\text{OR}, \phi)$ outputs the value of OR on those m bits of x:

$$\text{Lift}(\text{OR}, \phi)(x, y_1, \ldots, y_k) := \text{OR}(x|\phi(y_1, \ldots, y_k)).$$

The second idea is to apply to such a function $f := \text{Lift}(\text{OR}, \phi)$ a certain orthogonality principle to produce a function f' that satisfies the points (1) and (2) above. The structure of $f = \text{Lift}(\text{OR}, \phi)(x, y_1, \ldots, y_k)$ is crucially exploited to argue that f' satisfies (2), and it is here that previous works require $k < \log \log n$ [Cha07, LS08, CA08].

So far we have rephrased previous arguments. We now discuss the main new ideas in this paper.

Ideas for the proof of Theorem 1.1. To prove Theorem 1.1 we start by noting that regardless of what function ϕ is chosen, $\text{Lift}(\text{OR}, \phi)$ has an efficient nondeterministic protocol: Player 0 simply guesses an index j that is one of the indices chosen by ϕ (she can do so because she knows the input to ϕ) and then any of the other players can easily verify whether or not x_j is 1 in that position. In previous work [LS08, CA08], ϕ is the bitwise AND function, and this makes $\text{Lift}(\text{OR}, \phi)$ the Set-Disjointness function. By contrast, *in this work we choose the function ϕ uniformly at random* and we argue that, for almost all ϕ, $\text{Lift}(\text{OR}, \phi)$ does not have efficient randomized protocols, whenever k is at most $\delta \log n$ for a fixed $\delta < 1$.

The above argument gives a *non-explicit* separation, due to the random choice of ϕ. To make it explicit, we derandomize the choice of ϕ. Specifically, we note that the above argument goes through as long as ϕ is 2^k-wise independent, i.e. as long as ϕ comes from a distribution such that for every 2^k fixed inputs $\vec{y}^1, \ldots, \vec{y}^{2^k} \in (\{0,1\}^n)^k$ the values $\phi(\vec{y}^1), \ldots, \phi(\vec{y}^{2^k})$ are uniform and independent (over the choice of ϕ). Known constructions of such distributions [ABI86, CG89] only require about $n \cdot 2^k = n^{O(1)}$ random bits, which can be given as part of the input. Two things should perhaps be stressed. The first is that giving a description of ϕ as part of the input does not affect the lower bound in the previous paragraph which turns out to hold even against protocols that depend on ϕ. The second is that, actually, using 2^k-wise independence seems to add the constraint $k < 1/2(\log n)$; to achieve $k = \delta \log n$ for every $\delta < 1$ we use a distribution on ϕ that is *almost* 2^k-wise independent [NN93].

Ideas for the proof of Theorem 1.2. To prove Theorem 1.2 we show how to implement the function given by Theorem 1.1 by small constant-depth circuits when k is $A \log \log n$ for a fixed, possibly large, constant A. In light of the above discussion, this only requires computing a 2^k-wise independent function by small constant-depth circuits, a problem which is studied in [GV04, HV06]. Specifically, dividing up ϕ in blocks it turns out that it is enough to compute 2^k-wise independent functions $g : \{0,1\}^t \to \{0,1\}^t$ where t is also about 2^k. When $k = A \log\log n$, g is a $(2^k = \log^A n)$-wise independent function on $\log^A n$ bits, and [HV06] shows how to compute it with circuits of size n^B and depth B where B depends on A only – and this dependence of B on A is tight even for almost 2-wise independence. This gives Theorem 1.2. Finally, we note that [HV06] gives explicit (a.k.a. uniform) circuits, and that we are not aware of an alternative to [HV06] even for non-explicit circuits.

Subsequent Work. Subsequent to our work, [BHN08] extend our main results (Theorem 1.1 and Theorem 1.2) by proving the separation in Theorem 1.1 under the stronger requirement that the function f is computable by explicit (unbounded fan-in) circuits of depth 5.

Organization. The organization of the paper is as follows. In Section 2 we give necessary definitions and background. We present the proof of our main result Theorem 1.1 in two stages. First, in Section 3 we present a non-explicit separation obtained by selecting ϕ at random. Then, in Section 4 we derandomize the choice of ϕ in order to give an explicit separation and prove Theorem 1.1. Finally, in Section 5 we prove our results about constant-depth circuits, Theorem 1.2.

2 Preliminaries

Correlation. Let $f, g : X \to \mathbb{R}$ be two functions, and let μ be a distribution on X. We define the *correlation between f and g under μ* to be $\mathrm{Cor}_\mu(f,g) := \mathbb{E}_{x\sim\mu}[f(x)g(x)]$. Let $\mathcal{G}$ be a class of functions $g : X \to \mathbb{R}$ (e.g. efficient communication protocols). We define the *correlation between f and $\mathcal{G}$ under μ* to be $\mathrm{Cor}_\mu(f,\mathcal{G}) := \max_{g\in\mathcal{G}} \mathrm{Cor}_\mu(f,g)$. Note that, whenever $\mathcal{G}$ is closed under complements, which will always be the case in this paper, this correlation is non-negative. 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 referring to the uniform distribution, which we denote by $\mathcal{U}$.

Communication Complexity. In the number-on-forehead (NOF) multiparty communication complexity model [CFL83], k players are trying to collaborate to compute a function $f : X_1 \times \ldots \times X_k \to \{-1,1\}$. For each i, player i knows the values of all of the inputs $(x_1,\ldots,x_k) \in X_1 \times \ldots \times X_k$ 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 what each player writes as a function of the blackboard content and the inputs seen by that player, and whether the protocol is over, in which case the last bit written is taken as the output of the protocol. 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* is a distribution on deterministic protocols such that for every input a protocol selected at random from the distribution errs with probability at most 1/3. In a *nondeterministic protocol*, an initial guess string is written on the blackboard at the beginning of the protocol (and counted towards communication) and the players are trying to verify that the output of the function is -1 (representing *true*) 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 *communication complexity* of a function f under one of the above types of protocols is the minimum cost of a protocol of that type computing f. In line with [BFS86], a k-player protocol computing $f:(\{0,1\}^n)^k \to \{-1,1\}$ is considered to be *efficient* if its cost is at most poly-logarithmic, $\log^{O(1)} n$. Equipped with the notion of efficiency, one has the NOF communication complexity classes BPP_k^{cc} and NP_k^{cc} that are analogues of the corresponding complexity classes.

Definition 2.1. *We denote by $\Pi^{k,c}$ the class of all deterministic k-player NOF communication protocols of cost at most c.*

The following immediate fact allows us to derive lower bounds on the randomized communication complexity of f from upper bounds on the correlation between f and the class $\Pi^{k,c}$ [KN97, Theorem 3.20].

Fact 2.2. *If there exists a distribution μ such that $\mathrm{Cor}_\mu(f,\Pi^{k,c}) \leq 1/3$ then every randomized protocol (with error $1/3$) for f must communicate at least c bits.*

In order to obtain upper bounds on the correlation between f and the class $\Pi^{k,c}$, we use the following result, which is also standard. Historically, it was first proved by Babai, Nisan and Szegedy [BNS92] using the notion of *discrepancy* of a function. It has since been rewritten in many ways [CT93, Raz00, FG05, VW07]. The formulation we use appears in [VW07], except that in [VW07] one also takes two copies of x; it is easy to modify the proof in [VW07] to obtain the following lemma.

Lemma 2.3 (The standard BNS argument)**.** *Let $f: X\times Y_1\times\cdots\times Y_k \to \mathbb{R}$. Then,*

$$\mathrm{Cor}_{\mathcal{U}}(f,\Pi^{k+1,c})^{2^k} \leq 2^{c\cdot 2^k}\cdot \mathbb{E}_{\substack{(y_1^0,\ldots,y_k^0)\in Y_1\times\cdots\times Y_k\\(y_1^1,\ldots,y_k^1)\in Y_1\times\cdots\times Y_k}}\left[\left|\mathbb{E}_{x\in X}\left[\prod_{u\in\{0,1\}^k} f(x,y_1^{u_1},\ldots,y_k^{u_k})\right]\right|\right].$$

We later write $\bar{y}$ for $(y_1,\ldots,y_k)$.

Degree. The *ε-approximate degree of f* is the smallest d for which there exists a multivariate real-valued polynomial g of degree d such that $\max_x |f(x)-g(x)| \leq \varepsilon$. We will use the following result of Nisan and Szegedy; see [Pat92] for a result that applies to more functions.

Lemma 2.4 ([NS94])**.** *There exists a universal constant $\gamma > 0$ such that the $(5/6)$-approximate degree of the* OR *function on m bits is at least $\gamma\cdot\sqrt{m}$.*

The following key result shows that if a function f has ε-approximate degree d then there is another function g and a distribution μ such that g is orthogonal to degree-d polynomials and g has correlation ε with f. Sherstov [She08a] gives references in the mathematics literature and points out a short proof by duality.

Lemma 2.5 (Orthogonality Lemma)**.** *If $f : \{0,1\}^m \to \{-1,1\}$ is a function with ε-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) $\mathrm{Cor}_\mu(g,f) \geq \varepsilon$*; and*

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

where $x|T$ denotes the m bits of x indexed by T.

3 Non-explicit Separation

In this section we prove a *non-explicit* separation between nondeterministic and randomized protocols. As mentioned in the introduction, we restrict our attention to analyzing the communication complexity of certain functions constructed from a *base* function $f : \{0,1\}^m \to \{-1,1\}$, and a *selection* function ϕ. The 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.

We now give the definition of the function we prove the lower bound for, and then the statement of the lower bound.

Definition 3.1 (Lift)**.** *Let ϕ be a function that takes as input k strings $y_1,\ldots,y_k$ and outputs an m-element subset of $[n]$. Let f be a function on m bits. We construct a* lifted function $\mathrm{Lift}(f,\phi)$ *as follows. On input $(x \in \{0,1\}^n, y_1,\ldots,y_k)$,* $\mathrm{Lift}(f,\phi)$ *evaluates ϕ on the latter k inputs to select a set of m bits in x and returns the value of f on those m bits. Formally,*

$$\mathrm{Lift}(f,\phi)(x,y_1,\ldots,y_k) := f(x|\phi(y_1,\ldots,y_k)),$$

where for a set $S = \{i_1,\ldots,i_m\} \subseteq [n]$, $x|S$ denotes the substring $x_{i_1}\cdots x_{i_m}$ of x indexed by the elements in S, where $i_1 < i_2 < \ldots < i_m$.

The inputs to $\mathrm{Lift}(f,\phi)$ *are partitioned among $k+1$ players as follows: Player 0 is given x and, for all $1 \leq i \leq k$, Player i is given y_i.*

The following is the main theorem proved in this section.

Theorem 3.2. *For every $\delta < 1$ there are constants $\varepsilon, \alpha > 0$ such that for sufficiently large n, for $k = \delta \cdot \log n$, and for $m = n^\varepsilon$, the following holds. There is a distribution λ such that if we choose a random selection function $\phi : (\{0,1\}^n)^k \to \binom{[n]}{m}$, we have:*

$$\mathbb{E}_\phi[\mathrm{Cor}_\lambda(\mathrm{Lift}(\mathrm{OR},\phi), \Pi^{k+1,n^\alpha})] \leq 1/3.$$

3.1 Overview of the Proof

We obtain our lower bound on the randomized communication complexity of $\mathrm{Lift}(\mathrm{OR},\phi)$ using an analysis that follows [CA08]. In their paper, Chattopadhyay and Ada analyze the Set-Disjointness function, and for that reason, their selection function ϕ must be the AND function. In our case, we allow ϕ to be a random function. While our results no longer apply to Set-Disjointness, we still obtain a separation between randomized and nondeterministic communication (BPP_k^{cc} and NP_k^{cc}) because, no matter what selection function is used, $\mathrm{Lift}(\mathrm{OR},\phi)$ always has an efficient nondeterministic protocol.

At a more technical level, the results of [CA08] require $k < \log\log n$ because of the relationship between n (the size of player 0's input) and m (the number of bits the base function OR gets applied to.) For their analysis to go through, they need $n > 2^{2^k} \cdot m^{O(1)}$. In our case, $n = 2^k \cdot m^{O(1)}$ is sufficient, and this allows our results to be non-trivial for $k \leq \delta \log n$ for any $\delta < 1$.

As mentioned earlier, we will start with the base function OR on m input bits, $m = n^{\varepsilon} \ll n$. We lift the base function OR in order to obtain the lifted function $\mathrm{Lift}(\mathrm{OR},\phi)$. Recall that $\mathrm{Lift}(\mathrm{OR},\phi)$ 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 selected bits of Player 0's input, x. We want to prove that for a random ϕ, $\mathrm{Lift}(\mathrm{OR},\phi)$ has high randomized communication complexity.

We start with a result of Nisan and Szegedy [NS94] who prove a lower bound on the approximate degree of the OR function. By Lemma 2.5 this implies that there exists a function g (also on m bits) and a distribution μ such that the functions g and OR are highly correlated over μ and, furthermore, g is orthogonal to low-degree polynomials. Now we lift the function g in order to get the function $\mathrm{Lift}(g,\phi)$, and we define λ to be a distribution over all $(k+1)n$-bit inputs that chooses the y_i's uniformly at random and x also uniformly at random except on the bits indexed by $\phi(y_1,\ldots,y_k)$ which are selected according to μ. Since g and OR are highly correlated with respect to μ, it is not hard to see that the lifted functions $\mathrm{Lift}(f,\phi)$ and $\mathrm{Lift}(g,\phi)$are also highly correlated with respect to λ. Therefore, to prove that $\mathrm{Lift}(f,\phi)$ has low correlation with c-bit protocols it suffices to prove that $\mathrm{Lift}(g,\phi)$ has. To prove this, we use the correlation method. This involves bounding the average value of $\mathrm{Lift}(g,\phi)$ on certain k-dimensional cubes (cf. Lemma 2.3). For this, we need to analyze the distribution of the 2^k sets that arise from evaluating ϕ on the 2^k points of the cube. Specifically, we are interested in how much these 2^k sets are "spread out," as measured by the size of their union. If the sets are not spread out, we use in Lemma 3.4 the fact that g is orthogonal to low-degree polynomials to bound the average value of $\mathrm{Lift}(g,\phi)$ on the cubes. This step is similar to [She07, Cha07, LS08, CA08]. The main novelty in our analysis is that since we choose ϕ at random, we can prove good upper bounds (Lemma 3.6) on the probability that the sets are spread out.

3.2 Proof of Theorem 3.2

Let $m := n^{\varepsilon}$ for a small $\varepsilon > 0$ to be determined later. Combining Lemma 2.4 and 2.5, we see that there exists a function g and a distribution μ such that:

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

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

Define the distribution λ on $\{0,1\}^{(k+1)n}$ as follows. For $x, y_1, \ldots, y_k \in \{0,1\}^n$, let

$$\lambda(x,y_1,\ldots,y_k) := \frac{\mu(x|\phi(y_1,\ldots,y_k))}{2^{(k+1)n-m}},$$

in words we select $y_1,\dots,y_k$ uniformly at random and then we select the bits of x indexed by $\phi(y_1,\dots,y_k)$ according to μ and the others uniformly.

It can be easily verified that $\mathrm{Cor}_\lambda(\mathrm{Lift}(g,\phi),\mathrm{Lift}(\mathrm{OR},\phi)) = \mathrm{Cor}_\mu(g,\mathrm{OR}) \geq 5/6$. Consequently, for every ϕ and c,

$$\begin{aligned}\mathrm{Cor}_\lambda(\mathrm{Lift}(\mathrm{OR},\phi),\Pi^c) &\leq \mathrm{Cor}_\lambda(\mathrm{Lift}(g,\phi),\Pi^c) + 2\cdot \Pr_\lambda[\mathrm{Lift}(\mathrm{OR},\phi) \neq \mathrm{Lift}(g,\phi)] \\ &\leq \mathrm{Cor}_\lambda(\mathrm{Lift}(g,\phi),\Pi^c) + 1/6, \qquad (1)\end{aligned}$$

where in the last inequality we use that $\mathrm{Cor}_\lambda(\mathrm{Lift}(\mathrm{OR},\phi),\mathrm{Lift}(g,\phi)) = E_\lambda[\mathrm{Lift}(\mathrm{OR},\phi)\cdot\mathrm{Lift}(g,\phi)] \geq 5/6$. Therefore, we only have to upper bound $\mathrm{Cor}_\lambda(\mathrm{Lift}(g,\phi),\Pi^c)$, and this is addressed next. We have, by the definition of λ and then Lemma 2.3:

$$\begin{aligned}\mathrm{Cor}_\lambda(\mathrm{Lift}(g,\phi),\Pi^c)^{2^k} &= 2^{m\cdot 2^k}\mathrm{Cor}_{\mathcal{U}}(\mu(x|\phi(y_1,\dots,y_k))g(x|\phi(y_1,\dots,y_k),\Pi^c)^{2^k} \\ &\leq 2^{(c+m)2^k}\mathbb{E}_{\bar{y}^0,\bar{y}^1}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|\phi(y_1^{u_1},\dots,y_k^{u_k}))g(x|\phi(y_1^{u_1},\dots,y_k^{u_k}))\right]\right|\right], \qquad (2)\end{aligned}$$

for every ϕ.

Our analysis makes extensive use of the following notation.

Definition 3.3. *Let $\mathcal{S} = (S_1,\dots,S_z)$ be a multiset of m-element subsets of $[n]$. Let the* range of $\mathcal{S}$, *denoted by $\bigcup\mathcal{S}$, be the set of indices from $[n]$ 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 $[n]$ that appear in exactly one set in the collection $\mathcal{S}$.*

For $u \in \{0,1\}^k$, define $S_u = S_u(\bar{y}^0,\bar{y}^1,\phi) = \phi(y_1^{u_1},\dots,y_k^{u_k})$. Let $\mathcal{S} = \mathcal{S}(\bar{y}^0,\bar{y}^1,\phi)$ be the multiset $(S_u : u \in \{0,1\}^k)$. We define the number of conflicts in $\mathcal{S}$ *to be $q(\mathcal{S}) := m\cdot 2^k - |\bigcup\mathcal{S}|$.*

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. The first Lemma 3.4 deals with the case where the multiset $\mathcal{S}$ has few conflicts. In this case, we argue that one of the sets $S_u \in \mathcal{S}$ has a very small intersection with the rest of the other sets, which allows us to apply Property (ii) of g and μ to obtain the stated bound. A variant of Lemma 3.4 appears in [CA08].

Lemma 3.4. *For every $\bar{y}^0,\bar{y}^1$ and ϕ, if $q(\mathcal{S}(\bar{y}^0,\bar{y}^1,\phi)) < \gamma\cdot\sqrt{m}\cdot 2^k/2$, then*

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

Lemma 3.5 gives a bound in terms of the number of conflicts in $\mathcal{S}$ which only uses the fact that μ is a probability distribution. A slightly weaker version of this lemma appeared originally in [CA08]. Independently of our work, Chattopadhyay and Ada have subsequently also derived the stronger statement we give below.

Lemma 3.5. *For every $\bar{y}^0, \bar{y}^1$ and ϕ:*

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

Lemma 3.6 is the key place where we exploit the fact that ϕ is chosen at random to obtain an upper bound on the probability of having a given number of conflicts in $\mathcal{S}$.

Lemma 3.6. *For every $q > 0$ and uniformly chosen $\bar{y}^0, \bar{y}^1, \phi$:*

$$\Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S}(\bar{y}^0,\bar{y}^1,\phi)) = q] \le \left(\frac{m^3\cdot 2^{2k}}{q\cdot n}\right)^q.$$

Before proving these Lemmas, we complete the proof of our main theorem. We have the following derivation. For a uniformly chosen ϕ:

$$\begin{aligned}
&\mathbb{E}_\phi\left[\mathrm{Cor}_\lambda(\mathrm{Lift}(g,\phi),\Pi^c)\right]^{2^k} \le \mathbb{E}_\phi\left[\mathrm{Cor}_\lambda(\mathrm{Lift}(g,\phi),\Pi^c)^{2^k}\right]\\
&\le\; 2^{(c+m)2^k}\cdot\mathbb{E}_{\bar{y}^0,\bar{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|\right] \qquad \text{(by Equation (2))}\\
&=\; 2^{(c+m)2^k}\cdot\sum_{q\ge 0}\Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S})=q]\cdot\mathbb{E}_{\bar{y}^0,\bar{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|\,\middle|\,q(\mathcal{S})=q\right]\\
&\le\; 2^{(c+m)2^k}\cdot\sum_{q\ge\gamma\sqrt{m}2^k/2}\Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S})=q]\cdot\mathbb{E}_{\bar{y}^0,\bar{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|\,\middle|\,q(\mathcal{S})=q\right]\\
&\qquad\text{(by Lemma 3.4)}\\
&\le\; 2^{(c+m)2^k}\cdot\sum_{q\ge\gamma\sqrt{m}2^k/2}\Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S})=q]\cdot\mathbb{E}_{\bar{y}^0,\bar{y}^1,\phi}\left[\left|\mathbb{E}_x\left[\prod_{u\in\{0,1\}^k}\mu(x|S_u)\right]\right|\,\middle|\,q(\mathcal{S})=q\right]\\
&\qquad\text{(because } |g|=1\text{)}\\
&\le\; 2^{(c+m)2^k}\cdot\sum_{q\ge\gamma\sqrt{m}2^k/2}\Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S})=q]\cdot\frac{2^q}{2^{m2^k}} = 2^{c\cdot 2^k}\cdot\sum_{q\ge\gamma\sqrt{m}2^k/2}\Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S})=q]\cdot 2^q\\
&\qquad\text{(by Lemma 3.5)}\\
&\le\; 2^{c\cdot 2^k}\cdot\sum_{q\ge\gamma\sqrt{m}2^k/2}\left(\frac{m^3\cdot 2^{2k}}{q\cdot n}\right)^q\cdot 2^q = 2^{c\cdot 2^k}\cdot\sum_{q\ge\gamma\sqrt{m}2^k/2}\left(\frac{2\cdot m^3\cdot 2^{2k}}{q\cdot n}\right)^q\\
&\qquad\text{(by Lemma 3.6)}\\
&\le\; 2^{c\cdot 2^k}\cdot\sum_{q\ge\gamma\sqrt{m}2^k/2}\left(\frac{1}{2}\right)^q \le 2^{c\cdot 2^k+1-\gamma\sqrt{m}2^k/2} \le 2^{2^k(c-n^{\Omega(1)})}\\
&\qquad\text{(using } q\ge\gamma\sqrt{m}2^k/2,\ k=\delta\log n \text{ where } \delta<1\text{, and taking } m=n^\varepsilon \text{ for a sufficiently small } \varepsilon\text{)}
\end{aligned}$$

Therefore, when c is a sufficiently small power of n we have that $\mathbb{E}_{\phi}[\mathrm{Cor}_{\lambda}(\mathrm{Lift}(g,\phi),\Pi^c)] \leq 1/6$. Combining this with Equation (1) we obtain:

$$\mathbb{E}_{\phi}[\mathrm{Cor}_{\lambda}(\mathrm{Lift}(\mathrm{OR},\phi),\Pi^c)] \leq 1/6+1/6 = 1/3.$$

It is left to prove the lemmas. For this, the reader may want to recall Definition 3.3.

Proof of Lemma 3.4. We write S_u for $S_u(\bar{y}^0,\bar{y}^1,\phi)$ and $\mathcal{S}$ for $\mathcal{S}(\bar{y}^0,\bar{y}^1,\phi)$. 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, $\sum_{u\in\{0,1\}^k} |S_u \cap \partial\mathcal{S}| = b(\mathcal{S}) \geq r(\mathcal{S}) - q(\mathcal{S}) = m2^k - 2q(\mathcal{S}) > (m-\gamma\sqrt{m})2^k$. By the pigeonhole principle, there exists v such that $|S_v \cap \partial\mathcal{S}| > m - \gamma\sqrt{m}$. We can write

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

Let $T = S_v \setminus \partial\mathcal{S}$. So $|T| \leq \gamma\sqrt{m}$. Let $h = \mathbb{E}_{x|[n]\setminus S_v}\left[\prod_{u\neq v}\mu(x|S_u)g(x|S_u)\right]$. 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 3.5. 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

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

as each $\mu(x|S_u)$ only depends on the bits of x in $\bigcup\mathcal{S}$. For $0 \leq j \leq 2^k - 1$, let $\mathcal{S}_j$ be the submultiset of $\mathcal{S}$ consisting of the sets up to and including S_j, $\mathcal{S}_j = (S_0,\ldots,S_j)$. We have $\mathcal{S} = \mathcal{S}_{2^k-1}$ and 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 $H_j(x|S_j \setminus \partial\mathcal{S}_j) := \mathbb{E}_{x|S_j\cap\partial\mathcal{S}_j}[\mu(x|S_j)]$, which note is a quantity that depends on the bits of x in $S_j \setminus \partial\mathcal{S}_j$, i.e. on $x|(S_j \setminus \partial\mathcal{S}_j)$. Letting $G_{-1} := 1$, observe that, for $0 \leq j \leq 2^k - 1$,

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

To obtain a bound on $\max(H_j)$, consider an arbitrary partition of $[m]$ into two sets E,F. Let ν be a distribution on $[m]$, 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, $\max_{x|(S_j\setminus\partial\mathcal{S}_j)}(H_j) \leq 2^{|S_j\setminus\partial\mathcal{S}_j|-m}$. 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 sketch of Lemma 3.6. The multiset $\mathcal{S}$ is given by the sets $S_u = \phi(y_1^{u_1}, \ldots, y_k^{u_k})$ for $u \in \{0,1\}^k$. The probability over the choice of the y's that for some i, $y_i^0 = y_i^1$, is at most $k/2^n$. When this event does not occur, the 2^k points at which ϕ gets evaluated are all distinct. Since ϕ is chosen at random, the 2^k outputs of ϕ are 2^k uniformly and independently random m-element subsets of $[n]$. We now upper bound the probability of having q conflicts in this case.

We write Q for $q(\mathcal{S})$. Let $\mathcal{S}_i = (S_1, \ldots, S_i)$ and let $\mathcal{S}_0 = \emptyset$. Let Q_i be the number of conflicts obtained while picking S_i, after having picked $\mathcal{S}_{i-1}$, and let R_i be the range of $\mathcal{S}_i$. Formally, $Q_i = |S_i \cap (\cup \mathcal{S}_{i-1})|$ and $R_i = |\cup \mathcal{S}_i|$. It is easy to see that $Q = \sum_{i=1}^{2^k} Q_i$. Then,

$$\Pr[Q = q] = \sum_{q_1 + \cdots + q_{2^k} = q} \Pr[\forall i, Q_i = q_i] = \sum_{q_1 + \cdots + q_{2^k} = q} \prod_i \Pr[Q_i = q_i | \forall j < i, Q_j = q_j].$$

By the nature of the experiment, the probability of obtaining q_i conflicts while picking S_i depends only on the range of the sets picked before, thus $\Pr[Q_i = q_i | \forall j < i, Q_j = q_j] = \Pr[Q_i = q_i | R_{i-1} = (i-1)m - \sum_{j<i} q_j]$. Let $C(q,r)$ denote the probability that, when picking an m-element subset of $[n]$ we obtain exactly q conflicts, conditioned on the fact that the range of elements picked so far is exactly r. By standard calculations, one can show that, as long as $2^k m^3 \leq n$ (which holds for sufficiently small $m = n^{\varepsilon}$), $C(q,r) \leq \binom{m2^k}{q} (4m/n)^q$. Plugging this into the expression above, $\Pr[Q = q] \leq \left(4em^2 2^{2k}/qn\right)^q$.

Taking into account the probability that the 2^k strings $y_1^{u_1}, \ldots, y_k^{u_k}$ are all distinct, we obtain

$$\Pr_{\bar{y}^0, \bar{y}^1, \phi}[q(\mathcal{S}) = q] \leq \frac{k}{2^n} + \left(\frac{4 \cdot e \cdot m^2 \cdot 2^{2k}}{q \cdot n}\right)^q \leq \left(\frac{m^3 \cdot 2^{2k}}{q \cdot n}\right)^q,$$

where the last inequality is a loose bound which is sufficient for our purposes. The bound holds because we can assume that $q \leq m \cdot 2^k$ (otherwise the probability is 0) and note that $m \cdot 2^k = n^{1-\Omega(1)}$, for a sufficiently small $m = n^{\varepsilon}$, and therefore the second summand in the left-hand side of the inequality above is greater than the first. □

4 Explicit Separation

In this section we prove our main Theorem 1.1. We proceed as follows. First, we prove a derandomized version of Theorem 3.2 from the previous section. This derandomized version is such that the distribution on ϕ can be generated using only n random bits r. Then, we observe how including the random bits r as part of the input gives an explicit function for the separation, thus proving Theorem 1.1. As we mentioned in the introduction, the idea is that the only property of the distribution over ϕ that the previous construction was using is that such a distribution is 2^k-wise independent. That is, the evaluations of ϕ at any 2^k points, fixed and distinct, are jointly uniformly distributed, over the choice of ϕ (cf. the proof of Lemma 3.6). The most straightforward way to obtain explicit constructions from our previous results is thus to replace a random ϕ with a 2^k-wise independent distribution, and then include a description of ϕ as part of the input. However, this raises some technicalities, one being that the range of our ϕ was a size-m subset of $[n]$, and it is not immediate how to give constructions with such a range. We find it more simple to use a slightly different block-wise approach as we describe next.

We think of our universe of n bits as divided in $m := n^{\varepsilon}$ *blocks* of $b := n^{1-\varepsilon}$ bits each, where as before ε is a sufficiently small constant. We consider functions $\phi(y_1,\ldots,y_k)$ whose output is a subset of $[n]$ that contains exactly one bit per block. That is, $\phi(y_1,\ldots,y_k) \in [b]^m$. The building block of our distribution is the following result about almost t-wise independent functions. We say that two distributions X and Y on the same support are *ε-close in statistical distance* if for every event E we have $|\Pr[E(X)] - \Pr[E(Y)]| \le \varepsilon$.

Lemma 4.1 (almost t-wise independence; [NN93])**.** *There is a universal constant $a > 0$ such that for every t, b (where b is a power of 2) there is a polynomial-time computable map*

$$h : \{0,1\}^t \times \{0,1\}^{a\cdot t\cdot\log b} \longrightarrow [b]$$

such that for every t distinct $x_1,\ldots,x_t \in \{0,1\}^t$, the distribution $(h(x_1;r),\ldots,h(x_t;r)) \in [b]^t$, over the choice of $r \in \{0,1\}^{a\cdot t\cdot\log b}$, is $(1/b)^t$-close in statistical distance to the uniform distribution over $[b]^t$.

Proof. Naor and Naor [NN93, Section 4] give an explicit construction of N random variables over $\{0,1\}$ such that any k of them are δ-close to uniform (over $\{0,1\}^k$) and the construction uses $O(\log N + k + \log(1/\delta))$ random bits.[2] We identify $[b]$ with $\{0,1\}^{\log b}$ and use their construction for $N := 2^t \cdot \log b, k := t\cdot\log b$, and $\delta := (1/b)^t$. We consider the N random variables as divided up in 2^t blocks of $\log b$ bits each. On input $x \in \{0,1\}^t$, our function h will output the $\log b$ random variables from the x-th block, which, again, we are going to identify with an element in $[b]$. Since we set $k = t\cdot\log b$, and for distinct $x_1,\ldots,x_t$ the distribution of $(h(x_1;r),\ldots,h(x_t;r))$ is the distribution of $t\cdot\log b$ distinct random variables in $\{0,1\}$, we have by the result in [NN93] mentioned above that $(h(x_1;r),\ldots,h(x_t;r))$ is $(\delta = (1/b)^t)$-close to the uniform distribution on $[b]^t$. To conclude, we only need to verify the amount of randomness required. Indeed, as we mentioned above, the construction in [NN93] uses $O(\log N + k + \log(1/\delta))$ random bits, which by our choice of parameters is $O(t + \log\log b + t\cdot\log b + t\cdot\log b) = O(t\cdot\log b)$. □

We now define our derandomized distribution on ϕ. This is the concatenation of m of the above functions using independent random bits, a function per block. Specifically, for each of the m blocks of b bits, we are going to use the above function h where $t := k\cdot 2^k\cdot(1+\log b)$. Jumping ahead, the large input length t is also chosen so that the probability (over the choice of the y's) that we do not obtain 2^k distinct inputs drops down exponentially with 2^k, which is needed in the analysis. On input $y_1,\ldots,y_k$ and randomness r, we break up each y_i in m blocks and also r in m

[2] They in fact achieve in [NN93, Lemma 4.2] a doubly-logarithmic dependence on N, but this improvement, which arises from combining the above bound with a construction from [CG89, ABI86], is irrelevant to this work.

blocks. The value of ϕ in the j-th block depends only on the j-th blocks of the y_i's and on the j-th block of r.

Definition 4.2 (Derandomized distribution on ϕ, given parameters n, $m = n^{\varepsilon}$, $b = n^{1-\varepsilon}$, $k = \delta \cdot \log n$; and a universal constant from Lemma 4.1)**.** *Let $l := 2^k \cdot (1 + \log b)$, $t := l \cdot k$. Let*

$$\phi : \{0,1\}^{m \cdot t} \times \{0,1\}^{m \cdot a \cdot t \cdot \log b} \rightarrow [b]^m$$

be defined as follows. On input $(y_1, \ldots, y_k) \in \{0,1\}^{m \cdot t}$ and randomness $r \in \{0,1\}^{m \cdot a \cdot t \cdot \log b}$, think of each $y_i \in \{0,1\}^{m \cdot l}$ as divided in m blocks of l bits each, i.e. $(y_i = (y_i)_1 \circ \cdots \circ (y_i)_m)$, and r as divided in m blocks of $a \cdot t \cdot \log b$ bits each, i.e. $(r = r_1 \circ \cdots \circ r_m)$. The j-th output of ϕ in $[b]$ is then

$$\phi(y_1, \ldots, y_k; r)_j := h(\underbrace{(y_1)_j, \ldots, (y_k)_j}_{l \cdot k = t \text{ bits}}; \underbrace{r_j}_{a \cdot t \cdot \log b \text{ bits}}) \in [b].$$

The distribution on ϕ *is the distribution obtained by selecting a uniform $r \in \{0,1\}^{m \cdot a \cdot t \cdot \log b}$ and then considering the map*

$$y_1, \ldots, y_k \rightarrow \phi(y_1, \ldots, y_k; r) \in [b]^m.$$

Note that, in the above definition, the input length of each y_i is $m \cdot l$ which up to polylogarithmic factors is $n^{\varepsilon} \cdot 2^k = n^{1-\Omega(1)}$, for a sufficiently small ε depending on δ.

Theorem 4.3. *For every $\delta < 1$ there are constants $\varepsilon, \alpha > 0$ such that for sufficiently large n, $k := \delta \cdot \log n$, and $m = n^{\varepsilon}$, the following holds.*

There is a distribution λ such that if $\phi : \{0,1\}^{m \cdot t} \rightarrow [b]^m$ is distributed according to Definition 4.2 we have:

$$\mathbb{E}_{\phi}[\mathrm{Cor}_{\lambda}(\mathrm{Lift}(\mathrm{OR}, \phi), \Pi^{k+1, n^{\alpha}})] \leq 1/3.$$

Proof. The proof follows very closely that of Theorem 3.2. A minor difference is that now the y_i's are over $m \cdot l$ bits as opposed to n in Theorem 3.2, but the definition of the distribution λ in Theorem 3.2 immediately translates to the new setting – λ just selects the y_i's at random. The only other place where the proofs differ is in Lemma 3.6, which is where the properties of ϕ are used. Thus we only need to verify the following Lemma. □

Lemma 4.4. *For every $q > 0$ and ϕ distributed as in Definition (4.2):*

$$\Pr_{\bar{y}^0, \bar{y}^1, \phi}[q(\mathcal{S}(\bar{y}^0, \bar{y}^1, \phi)) = q] \leq \left(\frac{m^2 \cdot 2^{2k}}{q \cdot b}\right)^q = \left(\frac{m^3 \cdot 2^{2k}}{q \cdot n}\right)^q.$$

Proof. For the multiset $\mathcal{S} = \mathcal{S}(\bar{y}^0, \bar{y}^1, \phi)$ define the *number of conflicts in the j-th block*, denoted $q(\mathcal{S})_j$, as 2^k minus the number of distinct elements in the j-th block – thus $q(\mathcal{S}) = \sum_j q(\mathcal{S})_j$. If $q(\mathcal{S}) = q$ then there must exist $q_1, \ldots, q_m$ summing up to q such that for every j, $q(\mathcal{S})_j = q_j$. As by construction the distribution $(q(\mathcal{S})_1, \ldots, q(\mathcal{S})_m)$ (over the choice of the y's and ϕ) is a product distribution, we have:

$$\Pr_{\bar{y}^0, \bar{y}^1, \phi}[q(\mathcal{S}) = q] = \sum_{\substack{q_1, \ldots, q_m: \\ \sum_j q_j = q}} \prod_{j \leq m} \Pr_{\bar{y}^0, \bar{y}^1, \phi}[q(\mathcal{S})_j = q_j]. \tag{3}$$

We now bound $\Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S})_j = q_j]$ for any fixed j. Thus we are interested in the size of

$$\bigcup_{u\in\{0,1\}^k} \{\phi(y_1^{u_1},\ldots,y_k^{u_k};r)_j\} \subseteq [b].$$

By construction, this depends only on the j-th blocks (of $l = 2^k(1+\log b)$ bits) of the y's and on the j-th block of r. Specifically,

$$\bigcup_{u\in\{0,1\}^k} \{\phi(y_1^{u_1},\ldots,y_k^{u_k};r)_j\} = \bigcup_{u\in\{0,1\}^k} \{h((y_1^{u_1})_j,\ldots,(y_k^{u_k})_j;r_j)\} \subseteq [b].$$

The probability over the choice of the y's that the 2^k strings (given by the 2^k choices of $u \in \{0,1\}^k$)

$$((y_1^{u_1})_j,\ldots,(y_k^{u_k})_j) \in \{0,1\}^t$$

are not all distinct is at most, by a union bound, $k/2^l = 2^{\log k - 2^k(\log b+1)} \leq (1/b)^{2^k}$. When this happens, the 2^k elements

$$X_u := h((y_1^{u_1})_j,\ldots,(y_k^{u_k})_j;r_j) \in [b]$$

(given by the 2^k choices of $u \in \{0,1\}^k$) are by Lemma 4.1 $(1/b)^t$-close to being uniform and independent in $[b]$ (over the choice of r), where recall $t \geq 2^k$. If the X_u's were exactly uniform and independent over $[b]$ then it is not hard to see that the probability (over r) that $q(\mathcal{S})_j = q_j$ would be at most $\binom{2^k}{q_j}(2^k/b)^{q_j}$, a bound which can be obtained by noting that if $q(\mathcal{S})_j = q_j$ then there must exist q_j distinct $i \in \{0,1\}^k$ such that $X_i \in \{X_1,\ldots,X_{i-1}\}$. Since the X_u's are $((1/b)^t \leq (1/b)^{2^k})$-close to being uniform and independent, the probability (over r) that $q(\mathcal{S})_j = q_j$ is at most $(1/b)^{2^k} + \binom{2^k}{q_j}(2^k/b)^{q_j}$. Overall,

$$\Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S})_j = q_j] \leq (1/b)^{2^k} + (1/b)^{2^k} + \binom{2^k}{q_j}(2^k/b)^{q_j} \leq \binom{2^k}{q_j}(3\cdot 2^k/b)^{q_j},$$

where the last inequality holds when $q_j > 0$ – which is the case to which we are going to restrict – also using the fact that $q_j \leq 2^k$ – otherwise the probability is 0.

Therefore, combining the above bound with Equation (3) we obtain

$$\begin{aligned}
\Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S}) = q] &\leq \sum_{\substack{q_1,\ldots,q_m:\\ \sum_j q_j = q}} \prod_{j\leq m} \Pr_{\bar{y}^0,\bar{y}^1,\phi}[q(\mathcal{S})_j = q_j] \\
&\leq \sum_{\substack{q_1,\ldots,q_m:\\ \sum_j q_j = q}} \prod_{j\leq m: 0<q_j\leq 2^k} \binom{2^k}{q_j}(3\cdot 2^k/b)^{q_j} \\
&= (3\cdot 2^k/b)^q \sum_{\substack{q_1,\ldots,q_m:\\ \sum_j q_j = q}} \prod_{j\leq m: 0<q_j\leq 2^k} \binom{2^k}{q_j} \\
&= (3\cdot 2^k/b)^q \binom{m\cdot 2^k}{q} \leq \left(\frac{3\cdot 2^k}{b}\cdot\frac{e\cdot m\cdot 2^k}{q}\right)^q \leq \left(\frac{m^2\cdot 2^{2\cdot k}}{b\cdot q}\right)^q. \qquad \square
\end{aligned}$$

We can now prove the main theorem of this work.

Theorem 1.1 (Main Theorem; $\mathsf{NP}_k^{cc} \not\subset \mathsf{BPP}_k^{cc}$ for $k = \delta \log n$ players)**.** (Restated.) *For every fixed $\delta < 1$, sufficiently large n and $k = \delta \cdot \log n$, there is an explicit function $f : (\{0,1\}^n)^k \to \{0,1\}$ such that: f can be computed by k-player nondeterministic protocols communicating $O(\log n)$ bits, but f cannot be computed by k-player randomized protocols communicating $n^{o(1)}$ bits.*

Proof. Let $f(x,(y_1,r),y_2,\ldots,y_k) := \mathrm{OR}(x|\phi(y_1,\ldots,y_k;r))$, where ϕ is as in Definition 4.2. We partition an input $(x,(y_1,r),y_2,\ldots,y_k)$ as follows: Player 0 gets x, Player 1 gets the pair (y_1,r), where r is to be thought of as selecting which ϕ to use, and player $i > 1$ gets y_i. Let p be the distribution obtained by choosing r uniformly at random, and independently $(x,y_1,\ldots,y_k)$ according to the distribution λ in Theorem 4.3.

It is not hard to see that f has a nondeterministic protocol communicating $O(\log n)$ bits: We can guess a bit position i and then the player that sees $(y_1,r),y_2,\ldots,y_k$ can verify that the position i belongs to $\phi(y_1,\ldots,y_k;r)$, and finally another player can verify that $x_i = 1$.

To see the second item observe that:

$$\begin{aligned}\mathrm{Cor}_p(f,\Pi^{k+1,n^\alpha}) &= \max_{\pi \in \Pi^{k+1,n^\alpha}} \mathbb{E}_r[\mathbb{E}_{(x,\bar{y})\sim\lambda}[\mathrm{OR}(x|\phi(\bar{y};r))\cdot\pi(x,\bar{y},r)]] \\ &\le \mathbb{E}_r[\max_{\pi \in \Pi^{k+1,n^\alpha}} \mathbb{E}_{(x,\bar{y})\sim\lambda}[\mathrm{OR}(x|\phi(\bar{y};r))\cdot\pi(x,\bar{y},r)]] \le 1/3,\end{aligned}$$

where the last inequality follows by Theorem 4.3. Again, the claim about randomized communication follows by standard techniques, cf. Fact 2.2.

To conclude, we need to verify that we can afford to give r as part of the input without affecting the bounds. Specifically, we need to verify that $|(y_1,r)| \le n$. Indeed, $|(y_1,r)| \le m \cdot l + O(m \cdot t \cdot \log b) = m \cdot 2^k(1+\log b) + O(m \cdot 2^k(1+\log b)k \cdot \log b)$ which is less than n when $k = \delta \log n$ for a fixed $\delta < 1$, $m = n^\varepsilon$ for a sufficiently small ε, and n is sufficiently large (recall $b \cdot m = n$, and in particular $b \le n$.) □

As is apparent from the proofs, and similarly to previous works [She08b], our lower bound Theorems 3.2 and 4.3 hold more generally for any function of the form $\mathrm{Lift}(f,\phi)$ for an arbitrary base function f. The communication bound is then expressed in terms of the approximate degree of f. In our paper, we focused on $f = \mathrm{OR}$ for concreteness. However, also note that the choice of $f = \mathrm{OR}$ is essential in Theorem 1.1 in order for $\mathrm{Lift}(f,\phi)$ to have a cheap nondeterministic protocol.

4.1 Communication bounds for constant-depth circuits

In this section we point out how Theorem 4.3 from the previous section gives us some new communication bounds for functions computable by constant-depth circuits. Specifically, the next theorem, which was also stated in the introduction, gives communication bounds for up to $k = A \cdot \log\log n$ players for functions computable by constant-depth circuits (whose parameters depend on A), whereas previous results [Cha07, LS08, CA08] require $k < \log\log n$.

Theorem 1.2 (Constant-depth circuits require high communication for $k = A\log\log n$ players)**.** (Restated.) *For every constant $A > 1$ there is a constant B such that for sufficiently large n and $k := A\log\log n$ there is a function $f : (\{0,1\}^n)^k \to \{0,1\}$ which satisfies the following: f can be computed by circuits of size n^B and depth B, but f cannot be computed by k-player randomized protocols communicating $n^{o(1)}$ bits.*

Proof. Use the function from the proof of Theorem 1.1. This only requires computing $(2^k = \log^A n)$-wise independent functions on $\log^{O(A)} n$ bits. (As mentioned before, although Theorem 4.3 uses the notion of *almost* t-wise independence, for small values of k, such as those of interest in the current proof, we can afford to use *exact* t-wise independence, i.e. set the distance from uniform distribution to 0). Such functions can be computed by circuits of size n^B and depth B, for a constant B that depends on A only. To see this, one can use the standard constructions based on arithmetic over finite fields [CG89, ABI86] and then the results from [HV06, Corollary 6]. Equivalently, "scale down" [HV06, Theorem 14] as described in [HV06, Section 3]. □

It is not clear to us how to prove a similar result for $k = \omega(\log\log n)$. This is because our approach would require computing almost $(2^k = \log^{\omega(1)} n)$-wise independent functions on $\log^{\omega(1)} n$ bits by $n^{O(1)}$-size circuits of constant depth, which cannot be done (even for almost 2-wise independence). The fact that this cannot be done follows from the results in [MNT90] or known results on the noise sensitivity of constant-depth circuits [LMN93, Bop97].

We point out that Theorem 1.2 can be strengthened to give a function that has correlation $2^{-n^{\Omega(1)}}$ with protocols communicating $n^{o(1)}$ bits. This can be achieved using the Minsky-Papert function instead of OR (cf. [She07, Cha07]).

Finally, Troy Lee (personal communication, May 2008) has pointed out to us that the analogous of our Theorem 1.2 for *deterministic* protocols can be easily obtained from the known lower bound for generalized inner product (GIP) [BNS92]. This is because it is not hard to see that for every constant c there is a circuit of depth $B = B(c)$ and size n^B that has correlation at least $\exp(-n/\log^c n)$ with GIP – just compute the parity in GIP by brute-force on blocks of size $\log^c n$ – but on the other hand low-communication k-party protocols have correlation at most $\exp(-\Omega(n/4^k))$ with GIP [BNS92]. However, this idea does not seem to give a bound for randomized protocols or a correlation bound, whereas our results do.

Acknowledgements. We thank Sasha Sherstov and Troy Lee for helpful comments on the write-up. Matei David and Toniann Pitassi gratefully acknowledge Arkadev Chattopadhyay and Anil Ada for several very insightful conversations. Emanuele Viola is especially grateful to Troy Lee for many stimulating conversations on communication complexity.

References

[ABI86] N. Alon, L. Babai, and A. Itai. A fast and simple randomized algorithm for the maximal independent set problem. *Journal of Algorithms*, 7:567–583, 1986. 3, 12, 16

[BDPW07] Paul Beame, Matei David, Toniann Pitassi, and Philipp Woelfel. Separating deterministic from nondet. nof multiparty communication complexity. In *ICALP*, 2007. 1

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *FOCS*, pages 337–347. IEEE, 1986. 1, 5

[BHN08] Paul Beame and Dang-Trinh Huynh-Ngoc. Multiparty communication complexity and threshold size of AC_0, 2008. Manuscript. Earlier version ECCC Technical Report TR08-061. 4

[BNS92] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. 1, 2, 5, 16

[Bop97] Ravi B. Boppana. The average sensitivity of bounded-depth circuits. *Inform. Process. Lett.*, 63(5):257–261, 1997. 16

[BPS07] Paul Beame, Toniann Pitassi, and Nathan Segerlind. Lower bounds for lovász–schrijver systems and beyond follow from multiparty communication complexity. *SIAM J. Comput.*, 37(3):845–869, 2007. 1

[CA08] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *ECCC*, Technical Report TR08-002, 2008. 1, 2, 3, 6, 7, 8, 15

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *STOC*, 1983. 1, 4

[CG89] Benny Chor and Oded Goldreich. On the power of two-point based sampling. *Journal of Complexity*, 5(1):96–106, March 1989. 3, 12, 16

[Cha07] Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *FOCS*, pages 449–458. IEEE, October 2007. 1, 3, 7, 15, 16

[CT93] Fan R. K. Chung and Prasad Tetali. Communication complexity and quasi randomness. *SIAM J. Discrete Math.*, 6(1):110–123, 1993. 2, 5

[FG05] Jeff Ford and Anna Gál. Hadamard tensors and lower bounds on multiparty communication complexity. In *ICALP*, pages 1163–1175, 2005. 5

[GV04] Dan Gutfreund and Emanuele Viola. Fooling parity tests with parity gates. In *RANDOM*, LNCS, Volume 3122, pages 381–392. Springer-Verlag, 2004. 4

[HG91] Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Comput. Complexity*, 1(2):113–129, 1991. 1

[HV06] Alexander Healy and Emanuele Viola. Constant-depth circuits for arithmetic in finite fields of characteristic two. In *STACS*, LNCS, Volume 3884, pages 672–683, 2006. 4, 16

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, Cambridge, 1997. 1, 5

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *J. Assoc. Comput. Mach.*, 40(3):607–620, 1993. 16

[LS08] Troy Lee and Adi Shraibman. Disjointness is hard in the multi-party number on the forehead model. In *CCC*. IEEE, 2008. 1, 2, 3, 7, 15

[MNT90] Yishay Mansour, Noam Nisan, and Prasoon Tiwari. The computational complexity of universal hashing. In *STOC*, pages 235–243. ACM Press, 1990. 16

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993. 3, 12

[NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994. 5, 7

[NW93] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM J. Comput.*, 22(1):211–219, 1993. 1

[Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *STOC*, pages 468–474. ACM, 1992. 5

[Raz87] Alexander A. Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Mat. Zametki*, 41(4):598–607, 623, 1987. 2

[Raz00] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Comput. Complexity*, 9(2):113–122, 2000. 2, 5

[Raz03] Alexander Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003. 2

[RW93] Alexander Razborov and Avi Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. *Inform. Process. Lett.*, 45(6):303–307, 1993. 1

[She07] Alexander Sherstov. Separating AC^0 from depth-2 majority circuits. In *STOC'07: Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 2007. 1, 7, 16

[She08a] Alexander Sherstov. The pattern matrix method for lower bounds on quantum communication. In *STOC*, 2008. 1, 2, 3, 5

[She08b] Alexander A. Sherstov. Communication lower bounds using dual polynomials. *Electronic Colloquium on Computational Complexity*, Technical Report TR08-057, 2008. 2, 15

[VW07] Emanuele Viola and Avi Wigderson. Norms, xor lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *CCC*. IEEE, 2007. To appear in the journal *Theory of Computing*. 2, 5