THEORY OF COMPUTING, Volume 6 (2010), pp. 227–245
www.theoryofcomputing.org

A Separation of NP and coNP in Multiparty Communication Complexity

Dmitry Gavinsky Alexander A. Sherstov

Received: January 4, 2010; published: November 16, 2010.

Abstract: We prove that $\mathsf{coNP} \not\subseteq \mathsf{MA}$ in the number-on-forehead model of multiparty communication complexity for up to $k = (1-\varepsilon)\log n$ players, where $\varepsilon > 0$ is any constant. Specifically, we construct an explicit function $F : (\{0,1\}^n)^k \to \{0,1\}$ with co-nondeterministic complexity $O(\log n)$ and Merlin-Arthur complexity $n^{\Omega(1)}$. The problem was open for $k \geqslant 3$. As a corollary, we obtain an explicit separation of NP and coNP for up to $k = (1-\varepsilon)\log n$ players, complementing an independent result by Beame et al. (2010) who separate these classes nonconstructively for up to $k = 2^{(1-\varepsilon)n}$ players.

ACM Classification: F.1.3, F.2.3

AMS Classification: 68Q17, 68Q15

Key words and phrases: multiparty communication complexity, nondeterminism, Merlin-Arthur computations, separations and lower bounds

1 Introduction

The number-on-forehead model of multiparty communication complexity, introduced by Chandra, Furst, and Lipton [5], features k communicating players whose goal is to compute a given distributed function. More precisely, one considers a Boolean function $F : (\{0,1\}^n)^k \to \{-1,+1\}$ whose arguments $x_1, \dots, x_k \in \{0,1\}^n$ are placed on the foreheads of players 1 through k, respectively. Thus, player i sees all the arguments except for x_i. The players communicate by writing bits on a shared blackboard, visible to all. Their goal is to compute $F(x_1, \dots, x_k)$ with minimum communication. The multiparty model has found a variety of applications, including circuit complexity, pseudorandomness, and proof complexity [2, 30, 13, 25, 4]. This model draws its richness from the overlap in the players' inputs, which makes it challenging to prove lower bounds for explicit functions. Several fundamental questions in the multiparty model remain open despite much research.

 DOI: 10.4086/toc.2010.v006a010

1.1 Previous work and our results

In a seminal paper, Babai, Frankl, and Simon [1] defined analogues of computational complexity classes in communication and initiated their systematic study. In particular, the k-party number-on-forehead model gives rise to the complexity classes NP_k^{cc}, coNP_k^{cc}, BPP_k^{cc}, and MA_k^{cc}, corresponding to communication problems $F : (\{0,1\}^n)^k \to \{-1,+1\}$ with efficient nondeterministic, co-nondeterministic, randomized, and Merlin-Arthur protocols, respectively. An efficient protocol is one with communication cost $\log^{O(1)} n$. Determining the exact relationships among these classes is a natural goal in complexity theory.

For example, it had been open to show that nondeterministic protocols can be more powerful than randomized, for $k \geqslant 3$ players. This problem was recently solved by Lee and Shraibman [18] and Chattopadhyay and Ada [7] for up to $k = (1-o(1))\log_2\log_2 n$ players, and later strengthened by David and Pitassi [10] to $k = (1-\varepsilon)\log_2 n$ players, where $\varepsilon > 0$ is any given constant. An explicit separation for the latter case was obtained by David, Pitassi, and Viola [11].

The contribution of this paper is to relate the power of nondeterministic, co-nondeterministic, and Merlin-Arthur protocols. For $k = 2$ players, the relations among these models are well understood: Papadimitriou and Sipser [20] showed that $\mathsf{coNP}_2^{cc} \neq \mathsf{NP}_2^{cc}$, and Klauck [16] proved that additionally $\mathsf{coNP}_2^{cc} \nsubseteq \mathsf{MA}_2^{cc}$. Starting at $k = 3$, however, it has been open even to separate NP_k^{cc} and coNP_k^{cc}. Our main result is that $\mathsf{coNP}_k^{cc} \nsubseteq \mathsf{MA}_k^{cc}$ for up to $k = (1-\varepsilon)\log_2 n$ players, where $\varepsilon > 0$ is an arbitrary constant. The separation is by an explicitly given function. In particular, our work shows that $\mathsf{NP}_k^{cc} \neq \mathsf{coNP}_k^{cc}$ and also subsumes the separation in [10, 11], since $\mathsf{NP}_k^{cc} \subseteq \mathsf{MA}_k^{cc}$ and $\mathsf{BPP}_k^{cc} \subseteq \mathsf{MA}_k^{cc}$. Let the symbols $N(F)$, $N(-F)$, and $MA(F)$ denote the nondeterministic, co-nondeterministic, and Merlin-Arthur complexity of F in the k-party number-on-forehead model.

Theorem 1.1 (Main Result). *Let $k \leqslant (1-\varepsilon)\log_2 n$, where $\varepsilon > 0$ is any given constant. Then there is an (explicitly given) function $F : (\{0,1\}^n)^k \to \{-1,+1\}$ with*

$$N(-F) = O(\log n)$$

and

$$MA(F) = n^{\Omega(1)}.$$

In particular, $\mathsf{coNP}_k^{cc} \nsubseteq \mathsf{MA}_k^{cc}$ *and* $\mathsf{NP}_k^{cc} \neq \mathsf{coNP}_k^{cc}$.

Independently of our work, Beame, David, Pitassi, and Woelfel [3] proved nonconstructively that $\mathsf{NP}_k^{cc} \neq \mathsf{coNP}_k^{cc}$ for $k \leqslant 2^{(1-\varepsilon)n}$. An advantage of Theorem 1.1 is that it gives an explicit separation and additionally applies to Merlin-Arthur complexity. Theorem 1.1 is state-of-the-art with respect to the number of players: Babai, Nisan, and Szegedy [2] obtained the first strong lower bounds for multiparty communication complexity with up to $k = (1-\varepsilon)\log_2 n$ players, and it has since been an open problem to exhibit a function with nontrivial multiparty complexity for $k \geqslant \log_2 n$.

The proof of Theorem 1.1, described below, is based on Sherstov's *pattern matrix method* [28, 27] and its multiparty generalization in [10, 11]. In the final section of this paper, we revisit several other multiparty generalizations [6, 18, 7] of the pattern matrix method. By applying our techniques in these other settings, we are able to obtain similar exponential separations for smaller k, by functions as simple as constant-depth circuits.

1.2 Previous techniques

Perhaps the best-known method for lower bounds on communication complexity, both in the number-on-forehead multiparty model and various two-party models, is the *discrepancy method*. To our knowledge, this technique was introduced by Chor and Goldreich [8] in the context of two-party communication and later generalized to multiple parties by Babai, Nisan, and Szegedy [2]; see [17, pp. 36–38] for a detailed overview. The discrepancy method consists in exhibiting a distribution P with respect to which the function F of interest has negligible discrepancy, in other words, has negligible correlation with all low-cost protocols. A more powerful technique is the *generalized discrepancy method*, introduced by Klauck [15] and Razborov [24]. This method consists in exhibiting a distribution P and a function H such that, on the one hand, the function F of interest is well-correlated with H with respect to P, but on the other hand, H has negligible discrepancy with respect to P.

In practice, considerable effort is required to find suitable P and H and to analyze the resulting discrepancies. In particular, no strong bounds were available on the discrepancy or generalized discrepancy of constant-depth circuits AC^0. The *pattern matrix method*, introduced recently in [28, 27], solves this problem for AC^0 and a large family of other matrices. More specifically, the method uses standard analytic properties of Boolean functions (such as approximate degree or threshold degree) to determine the discrepancy and generalized discrepancy of the associated communication problems.

Originally formulated in [28, 27] for the two-party model, the pattern matrix method has been adapted to the multiparty model by several authors [6, 18, 7, 10, 11]. The first adaptation of the method to the multiparty model gave improved lower bounds for the multiparty disjointness function [18, 7]. This line of work was combined in [10, 11] with probabilistic arguments to separate the classes NP_k^{cc} and BPP_k^{cc} for up to $k = (1-\varepsilon)\log_2 n$ players, by an explicit function. Further details on this body of research and on other duality-based approaches [29] can be found in the survey article [26].

1.3 Our approach

To obtain our main result, we combine the work in [10, 11] with several new ideas. First, we derive a new criterion for high nondeterministic communication complexity, inspired by the Klauck-Razborov generalized discrepancy method [15, 24]. Similar to Klauck-Razborov, we also look for a hard function H that is well-correlated with the function F of interest, but we additionally quantify the agreement of H and F on the set $F^{-1}(-1)$. This agreement ensures that $F^{-1}(-1)$ does not have a small cover by cylinder intersections, thus placing F outside NP_k^{cc}. To handle the more powerful Merlin-Arthur model, we combine this development with an earlier technique due to Klauck [16] for proving lower bounds against two-party Merlin-Arthur protocols.

In keeping with the philosophy of the pattern matrix method, we then reformulate the agreement requirement for H and F as a suitable analytic property of the underlying Boolean function f and prove this property directly, using linear programming duality. The function f in question happens to be OR.

Finally, we apply our program to the specific function F constructed in [11] for the purpose of separating NP_k^{cc} and BPP_k^{cc}. Since F has small nondeterministic complexity by design, the proof of our main result is complete once we apply our machinery to $-F$ and derive a lower bound on $MA(-F)$.

1.4 Organization

We start in Section 2 with relevant technical preliminaries and standard background on multiparty communication complexity. In Section 3, we review the original discrepancy method, the generalized discrepancy method, and the pattern matrix method. In Section 4, we derive the new criterion for high nondeterministic and Merlin-Arthur communication complexity. The proof of Theorem 1.1 comes next, in Section 5. In the final section of the paper, we explore some implications of this work in the light of other multiparty papers [6, 18, 7].

2 Preliminaries

We view Boolean functions as mappings $X \to \{-1,+1\}$, where X is a finite set such as $X = \{0,1\}^n$ or $X = \{0,1\}^n \times \{0,1\}^n$. We identify -1 and $+1$ with "true" and "false," respectively. The notation $[n]$ stands for the set $\{1,2,\ldots,n\}$. For integers N,n with $N \geqslant n$, the symbol $\binom{[N]}{n}$ denotes the family of all size-n subsets of $\{1,2,\ldots,N\}$. For a string $x \in \{-1,+1\}^N$ and a set $S \in \binom{[N]}{n}$, we define $x|_S = (x_{i_1}, x_{i_2}, \ldots, x_{i_n}) \in \{-1,+1\}^n$, where $i_1 < i_2 < \cdots < i_n$ are the elements of S. For $x \in \{0,1\}^n$, we write $|x| = x_1 + \cdots + x_n$. Throughout this manuscript, "log" refers to the logarithm to base 2. For a function $f : X \to \mathbb{R}$, where X is an arbitrary finite set, we write $\|f\|_\infty = \max_{x \in X} |f(x)|$.

We will need the following observation regarding discrete probability distributions on the hypercube, cf. [28].

Proposition 2.1. *Let $\mu(x)$ be a probability distribution on $\{0,1\}^n$. Fix $i_1,\ldots,i_n \in \{1,2,\ldots,n\}$. Then*

$$\sum_{x \in \{0,1\}^n} \mu(x_{i_1}, \ldots, x_{i_n}) \leqslant 2^{n-|\{i_1,\ldots,i_n\}|}.$$

For functions $f, g : X_1 \times \cdots \times X_k \to \mathbb{R}$ (where X_i is a finite set, $i = 1,2,\ldots,k$), we define $\langle f, g \rangle = \sum_{(x_1,\ldots,x_k)} f(x_1,\ldots,x_k) g(x_1,\ldots,x_k)$. When f and g are vectors or matrices, this is the standard definition of inner product. The *Hadamard product* of f and g is the tensor $f \circ g : X_1 \times \cdots \times X_k \to \mathbb{R}$ given by $(f \circ g)(x_1,\ldots,x_k) = f(x_1,\ldots,x_k) g(x_1,\ldots,x_k)$.

The symbol $\mathbb{R}^{m \times n}$ refers to the family of all $m \times n$ matrices with real entries. The (i,j)th entry of a matrix A is denoted by A_{ij}. In most matrices that arise in this work, the exact ordering of the columns (and rows) is irrelevant. In such cases, we describe a matrix using the notation $[F(i,j)]_{i \in I,\, j \in J}$, where I and J are some index sets.

We conclude with a review of the Fourier transform over $\mathbb{Z}_2^n$ (cf. [12] for more details). Consider the vector space of functions $\{0,1\}^n \to \mathbb{R}$. For $S \subseteq [n]$, define $\chi_S : \{0,1\}^n \to \{-1,+1\}$ by $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. Then every function $f : \{0,1\}^n \to \mathbb{R}$ has a unique representation of the form $f = \sum_{S \subseteq [n]} \hat{f}(S)\, \chi_S$, where $\hat{f}(S) = 2^{-n} \sum_{x \in \{0,1\}^n} f(x) \chi_S(x)$. The reals $\hat{f}(S)$ are called the *Fourier coefficients* of f.

Communication complexity

An excellent reference on communication complexity is the monograph by Kushilevitz and Nisan [17]. In this overview, we will limit ourselves to key definitions and notation. The simplest model of communication in this work is the two-party randomized model. Consider a function $F : X \times Y \to \{-1,+1\}$, where X and Y are finite sets. Alice receives an input $x \in X$, Bob receives $y \in Y$, and their objective is to predict $F(x,y)$ with high accuracy. To this end, Alice and Bob share a communication channel and have an unlimited supply of shared random bits. Alice and Bob's protocol is said to have error ε if on every input (x,y), the computed output differs from the correct answer $F(x,y)$ with probability no greater than ε. The *cost* of a given protocol is the maximum number of bits exchanged on any input. The *randomized* communication complexity of F, denoted $R_\varepsilon(F)$, is the least cost of an ε-error protocol for F. It is standard practice to use the shorthand $R(F) = R_{1/3}(F)$. Recall that the error probability of a protocol can be decreased from $1/3$ to any other positive constant at the expense of increasing the communication cost by a constant factor. We will use this fact in our proofs without further mention.

A generalization of two-party communication is the *multiparty number-on-forehead* model of communication. Here one considers a function $F : X_1 \times \cdots \times X_k \to \{-1,+1\}$ for some finite sets $X_1,\dots,X_k$. There are k players. A given input $(x_1,\dots,x_k) \in X_1 \times \cdots \times X_k$ is distributed among the players by placing x_i on the forehead of player i (for $i = 1,\dots,k$). In other words, player i knows $x_1,\dots,x_{i-1},x_{i+1},\dots,x_k$ but not x_i. The players communicate by writing bits on a shared blackboard, visible to all. They additionally have access to a shared source of random bits. Their goal is to devise a communication protocol that will allow them to accurately predict the value of F on every input. Analogous to the two-party case, the randomized communication complexity $R_\varepsilon(F)$ is the least cost of an ε-error communication protocol for F in this model, and $R(F) = R_{1/3}(F)$.

Another model in this paper is the number-on-forehead *nondeterministic* model. As before, one considers a function $F : X_1 \times \cdots \times X_k \to \{-1,+1\}$ for some finite sets $X_1,\dots,X_k$. An input from $X_1 \times \cdots \times X_k$ is distributed among the k players as before. At the start of the protocol, c_1 nondeterministic bits appear on the shared blackboard. Given the values of those bits, the players behave deterministically, exchanging an additional c_2 bits by writing them on the blackboard. A nondeterministic protocol for F must output the correct answer for *at least one* nondeterministic choice of the c_1 bits when $F(x_1,\dots,x_k) = -1$ and for *all* possible choices when $F(x_1,\dots,x_k) = +1$. The cost of a nondeterministic protocol is defined as $c_1 + c_2$. The *nondeterministic* communication complexity of F, denoted $N(F)$, is the least cost of a nondeterministic protocol for F. The *co-nondeterministic* communication complexity of F is the quantity $N(-F)$.

The number-on-forehead *Merlin-Arthur* model combines the power of randomized and nondeterministic models. Similar to the nondeterministic case, the protocol starts with a nondeterministic guess of c_1 bits, followed by c_2 bits of communication. However, the communication can be randomized, and the requirement is that the error probability be at most ε for *at least one* nondeterministic choice when $F(x_1,\dots,x_k) = -1$ and for *all* possible nondeterministic choices when $F(x_1,\dots,x_k) = +1$. The cost of a protocol is defined as $c_1 + c_2$. The *Merlin-Arthur* communication complexity of F, denoted $MA_\varepsilon(F)$, is the least cost of an ε-error Merlin-Arthur protocol for F. We put $MA(F) = MA_{1/3}(F)$. Clearly, $MA(F) \leqslant \min\{N(F), R(F)\}$ for every F.

Babai, Frankl, and Simon [1] defined analogues of computational complexity classes in communication. We will only study a few of these communication classes, namely, those corresponding to efficient randomized, nondeterministic, co-nondeterministic, and Merlin-Arthur protocols. For a given number of players $k = k(n)$, fix a family of k-party communication problems $F_n\colon (\{0,1\}^n)^k \to \{-1,+1\}$, $n = 1,2,3,\ldots$. The family $\{F_n\}$ is said to belong to the class BPP_k^{cc} if the randomized communication complexity of F_n is bounded by $(\log n)^c$ for some constant $c > 1$ and all $n > c$. Analogously, the family $\{F_n\}$ is said to belong to NP_k^{cc}, coNP_k^{cc}, MA_k^{cc} if the communication complexity of F_n in the nondeterministic, co-nondeterministic, and Merlin-Arthur model, respectively, is at most $(\log n)^c$ for some constant $c > 1$ and all $n > c$.

3 Generalized discrepancy and pattern matrices

A common tool for proving communication lower bounds is the *discrepancy method.* Given a function $F : X \times Y \to \{-1,+1\}$ and a distribution μ on $X \times Y$, the *discrepancy of F with respect to μ* is defined as

$$\operatorname{disc}_\mu(F) = \max_{\substack{S\subseteq X,\\ T\subseteq Y}} \left| \sum_{x\in S}\sum_{y\in T} \mu(x,y)F(x,y) \right| .$$

This definition generalizes to the multiparty case as follows. Consider a function $F : X_1 \times \cdots \times X_k \to \{-1,+1\}$ and a distribution μ on $X_1 \times \cdots \times X_k$. The *discrepancy of F with respect to μ* is defined as

$$\operatorname{disc}_\mu(F) = \max_{\chi} \left| \sum_{\substack{(x_1,\ldots,x_k)\\ \in X_1\times\cdots\times X_k}} \mu(x_1,\ldots,x_k)F(x_1,\ldots,x_k)\chi(x_1,\ldots,x_k) \right| ,$$

where the maximum ranges over functions $\chi : X_1 \times \cdots \times X_k \to \{0,1\}$ of the form

$$\chi(x_1,\ldots,x_k) = \prod_{i=1}^{k} \phi_i(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_k) \tag{3.1}$$

for some $\phi_i : X_1 \times \cdots X_{i-1} \times X_{i+1} \times \cdots X_k \to \{0,1\}$, $i = 1,2,\ldots,k$. A function χ of the form (3.1) is called a *rectangle* for $k = 2$ and a *cylinder intersection* for $k \geqslant 3$, the latter notion introduced by Babai, Nisan, and Szegedy [2]. Note that for $k = 2$, the multiparty definition of discrepancy agrees with the one given earlier for the two-party model. Put

$$\operatorname{disc}(F) = \min_{\mu} \operatorname{disc}_\mu(F) .$$

Discrepancy is difficult to analyze as defined. Typically, one uses the following estimate from the pioneering work in [2], derived by repeated applications of the Cauchy-Schwarz inequality.

Theorem 3.1 ([2, 9, 22])**.** *Fix $F : X_1 \times \cdots \times X_k \to \{-1,+1\}$ and a distribution μ on $X_1 \times \cdots \times X_k$. Put $\psi(x_1,\ldots,x_k) = F(x_1,\ldots,x_k)\mu(x_1,\ldots,x_k)$. Then*

$$\left(\frac{\operatorname{disc}_\mu(F)}{|X_1|\cdots|X_k|}\right)^{2^{k-1}} \leqslant \mathop{\mathbf{E}}_{\substack{x_1^0\in X_1\\ x_1^1\in X_1}} \cdots \mathop{\mathbf{E}}_{\substack{x_{k-1}^0\in X_{k-1}\\ x_{k-1}^1\in X_{k-1}}} \left| \mathop{\mathbf{E}}_{x_k\in X_k} \prod_{z\in\{0,1\}^{k-1}} \psi(x_1^{z_1},\ldots,x_{k-1}^{z_{k-1}},x_k) \right| .$$

To our knowledge, no alternate technique has been discovered for bounding the discrepancy of explicit multiparty functions. In the case of $k = 2$ parties, there are other ways to estimate the discrepancy, including the spectral norm of a matrix (see for example [27]).

For a function $F : X_1 \times \cdots \times X_k \to \{-1,+1\}$ and a distribution μ over $X_1 \times \cdots \times X_k$, let $D^{\mu}_{\varepsilon}(F)$ denote the least cost of a deterministic protocol for F whose probability of error with respect to μ is at most ε. This quantity is known as the *μ-distributional complexity* of F. Since a randomized protocol can be viewed as a probability distribution over deterministic protocols, we immediately have that $R_{\varepsilon}(F) \geqslant \max_{\mu} D^{\mu}_{\varepsilon}(F)$. We are now ready to state the discrepancy method, which was introduced by Chor and Goldreich [8] in the context of two-party communication and generalized to multiple parties by Babai, Nisan, and Szegedy [2].

Theorem 3.2 (Discrepancy method [8, 2]; see also [17, pp. 36–38])**.** *For every $F : X_1 \times \cdots \times X_k \to \{-1,+1\}$, every distribution μ on $X_1 \times \cdots \times X_k$, and $0 < \gamma \leqslant 1$,*

$$R_{1/2-\gamma/2}(F) \geqslant D^{\mu}_{1/2-\gamma/2}(F) \geqslant \log \frac{\gamma}{\operatorname{disc}_{\mu}(F)}\,.$$

In words, a function with small discrepancy is hard to compute to any nontrivial advantage over random guessing, let alone compute it to high accuracy.

3.1 Generalized discrepancy method

The discrepancy method is particularly strong in that it gives communication lower bounds not only for bounded-error protocols but also for protocols with error vanishingly close to $1/2$. This strength of the discrepancy method is at once a weakness. For example, the disjointness function $\text{DISJ}(x,y) = \bigvee_{i=1}^{n}(x_i \wedge y_i)$ has a randomized protocol with error $1/2 - \Omega(1/n)$ and communication $O(\log n)$. As a result, the disjointness function has high discrepancy, and no strong lower bounds can be obtained for it via the discrepancy method. Yet it is well-known that DISJ and $\neg$DISJ have communication complexity $\Theta(n)$ in the randomized model [14, 23], and $\neg$DISJ additionally has complexity $\Omega(\sqrt{n})$ in the Merlin-Arthur model [16].

The *generalized discrepancy method* is an extension of the traditional discrepancy method that avoids the difficulty just cited. This technique was first applied by Klauck [15] and reformulated in its current form by Razborov [24]. The development in [15, 24] takes place in the quantum model of communication. However, the same idea works in a variety of models, as illustrated in [27]. The version of the generalized discrepancy method for the two-party randomized model is as follows.

Theorem 3.3 ([27, §2.4])**.** *Fix a function $F : X \times Y \to \{-1,+1\}$ and $0 \leqslant \varepsilon < 1/2$. Then for all functions $H : X \times Y \to \{-1,+1\}$ and all probability distributions P on $X \times Y$,*

$$R_{\varepsilon}(F) \geqslant \log \frac{\langle F, H \circ P\rangle - 2\varepsilon}{\operatorname{disc}_P(H)}\,.$$

The usefulness of Theorem 3.3 stems from its applicability to functions that have efficient protocols with error close to random guessing, such as $1/2-\Omega(1/n)$ for the disjointness function. Note that one recovers Theorem 3.2, the ordinary discrepancy method, by setting $H=F$ in Theorem 3.3.

Proof of Theorem 3.3 (adapted from [27], pp. 88–89). Put $c=R_\varepsilon(F)$. A public-coin protocol with cost c can be thought of as a probability distribution on deterministic protocols with cost at most c. In particular, there are random variables $\underline{\chi}_1,\underline{\chi}_2,\dots,\underline{\chi}_{2^c} : X\times Y\to\{0,1\}$, each a rectangle, as well as random variables $\underline{\sigma}_1,\underline{\sigma}_2,\dots,\underline{\sigma}_{2^c}\in\{-1,+1\}$, such that

$$\left\| F-\mathbf{E}\left[\sum \underline{\sigma}_i\underline{\chi}_i\right]\right\|_\infty \leqslant 2\varepsilon\,.$$

Therefore,

$$\left\langle F-\mathbf{E}\left[\sum \underline{\sigma}_i\underline{\chi}_i\right], H\circ P\right\rangle \leqslant 2\varepsilon\,.$$

On the other hand,

$$\left\langle F-\mathbf{E}\left[\sum \underline{\sigma}_i\underline{\chi}_i\right], H\circ P\right\rangle \geqslant \langle F, H\circ P\rangle - 2^c\,\mathrm{disc}_P(H)$$

by the definition of discrepancy. The theorem follows at once from the last two inequalities. □

Theorem 3.3 extends word-for-word to the multiparty model, as follows:

Theorem 3.4 ([18, 7]). *Fix a function $F:X\to\{-1,+1\}$ and $\varepsilon\in[0,1/2)$, where $X=X_1\times\cdots\times X_k$. Then for all functions $H:X\to\{-1,+1\}$ and all probability distributions P on X,*

$$R_\varepsilon(F)\geqslant \log\frac{\langle F,H\circ P\rangle-2\varepsilon}{\mathrm{disc}_P(H)}\,.$$

Proof. Identical to the two-party case (Theorem 3.3), with the word "rectangles" replaced by "cylinder intersections." □

3.2 Pattern matrix method

To apply the generalized discrepancy method to a given Boolean function F, one needs to identify a Boolean function H which is well correlated with F under some distribution P but has low discrepancy with respect to P. The pattern matrix method [28, 27] is a systematic technique for finding such H and F. To simplify the exposition of our main results, we will now review this method and sketch its proof.

Recall that the *ε-approximate degree* of a function $f:\{0,1\}^n\to\mathbb{R}$, denoted $\deg_\varepsilon(f)$, is the least degree of a polynomial p with $\|f-p\|_\infty\leqslant\varepsilon$. A starting point in the pattern matrix method is the following dual formulation of the approximate degree.

Fact 3.5. *Fix $\varepsilon \geqslant 0$. Let $f : \{0,1\}^n \to \mathbb{R}$ be given with $d = \deg_\varepsilon(f) \geqslant 1$. Then there is a function $\psi : \{0,1\}^n \to \mathbb{R}$ such that:*

$$\begin{aligned} \hat{\psi}(S) &= 0 \qquad \textit{for } |S| < d, \\ \sum_{z \in \{0,1\}^n} |\psi(z)| &= 1, \\ \sum_{z \in \{0,1\}^n} \psi(z) f(z) &> \varepsilon\,. \end{aligned}$$

See [27] for a proof of this fact using linear programming duality. The crux of the method is the following theorem.

Theorem 3.6 ([28])**.** *Fix a function $h : \{0,1\}^n \to \{-1,+1\}$ and a probability distribution μ on $\{0,1\}^n$ such that*

$$\widehat{h \circ \mu}(S) = 0 \qquad \textit{for } |S| < d.$$

Let N be a given integer. Define

$$H = [h(x|_V)]_{x,V}\,, \qquad P = 2^{-N+n} \binom{N}{n}^{-1} [\mu(x|_V)]_{x,V}\,,$$

where the rows are indexed by $x \in \{0,1\}^N$ and columns by $V \in \binom{[N]}{n}$. Then

$$\operatorname{disc}_P(H) \leqslant \left(\frac{4\mathrm{e}n^2}{Nd}\right)^{d/2}.$$

At last, we are ready to state the (two-party) pattern matrix method.

Theorem 3.7 ([27])**.** *Let $f : \{0,1\}^n \to \{-1,+1\}$ be a given function, $d = \deg_{1/3}(f)$. Let N be a given integer. Define $F = [f(x|_V)]_{x,V}$, where the rows are indexed by $x \in \{0,1\}^N$ and columns by $V \in \binom{[N]}{n}$. If $N \geqslant 16\mathrm{e}n^2/d$, then*

$$R(F) = \Omega\left(d \log\left\{\frac{Nd}{4\mathrm{e}n^2}\right\}\right).$$

Proof (adapted from [27]). Let $\varepsilon = 1/10$. By Fact 3.5, there exists a function $h : \{0,1\}^n \to \{-1,+1\}$ and a probability distribution μ on $\{0,1\}^n$ such that

$$\widehat{h \circ \mu}(S) = 0\,, \qquad |S| < d, \tag{3.2}$$

and

$$\sum_{z \in \{0,1\}^n} f(z)\mu(z)h(z) > \frac{1}{3}\,. \tag{3.3}$$

Letting $H = [h(x|_V)]_{x,V}$ and $P = 2^{-N+n}\binom{N}{n}^{-1}[\mu(x|_V)]_{x,V}$, we obtain from (3.2) and Theorem 3.6 that

$$\mathrm{disc}_P(H) \leqslant \left(\frac{4\mathrm{e}n^2}{Nd}\right)^{d/2}. \tag{3.4}$$

At the same time, one sees from (3.3) that

$$\langle F, H \circ P\rangle > \frac{1}{3}. \tag{3.5}$$

The theorem now follows from (3.4) and (3.5) in view of the generalized discrepancy method, Theorem 3.3. □

Remark. Presented above is a weaker, combinatorial version of the pattern matrix method. The communication lower bounds in Theorems 3.6 and 3.7 were improved to optimal in [27] using matrix-analytic techniques. Unlike the combinatorial argument above, however, the matrix-analytic proof is not known to extend to the multiparty model and is not used in the follow-up multiparty papers [6, 18, 7, 10, 11] or our work.

An alternate technique based on Fact 3.5 is the *block-composition method* of Shi and Zhu [29], developed independently of the pattern matrix method. See [26, §5.3] for a comparative discussion.

4 A new criterion for nondeterministic and Merlin-Arthur complexity

In this section, we derive a new criterion for high communication complexity in the nondeterministic and Merlin-Arthur models. This criterion, inspired by the generalized discrepancy method, will allow us to obtain our main result.

Theorem 4.1. *Let $F : X \to \{-1,+1\}$ be given, where $X = X_1 \times \cdots \times X_k$. Fix a function $H : X \to \{-1,+1\}$ and a probability distribution P on X. Put*

$$\begin{aligned} \alpha &= P(F^{-1}(-1) \cap H^{-1}(-1)), \\ \beta &= P(F^{-1}(-1) \cap H^{-1}(+1)), \\ Q &= \log \frac{\alpha}{\beta + \mathrm{disc}_P(H)}. \end{aligned}$$

Then

$$N(F) \geqslant Q \tag{4.1}$$

and

$$MA(F) \geqslant \min\left\{\Omega(\sqrt{Q}),\ \Omega\left(\frac{Q}{\log\{2/\alpha\}}\right)\right\}. \tag{4.2}$$

Proof. Put $c = N(F)$. It is well-known [2, 17] that there is a cover of $F^{-1}(-1)$ by 2^c cylinder intersections, each contained in $F^{-1}(-1)$. Fix one such cover, $\chi_1, \chi_2, \ldots, \chi_{2^c} : X \to \{0,1\}$. By the definition of discrepancy,

$$\langle \sum \chi_i, -H \circ P \rangle \leqslant 2^c \operatorname{disc}_P(H) .$$

On the other hand, $\sum \chi_i$ ranges between 1 and 2^c on $F^{-1}(-1)$ and vanishes on $F^{-1}(+1)$. Therefore,

$$\langle \sum \chi_i, -H \circ P \rangle \geqslant \alpha - 2^c \beta .$$

These two inequalities force (4.1).

We now turn to the Merlin-Arthur model. Let $c = MA(F)$ and $\delta = \alpha 2^{-c-1}$. The first step is to improve the error probability of the Merlin-Arthur protocol by repetition from $1/3$ to δ. Specifically, following Klauck [16] we observe that there exist randomized protocols $F_1, \ldots, F_{2^c} : X \to \{0,1\}$, each a random variable of the coin tosses and each having communication cost $c' = O(c \log\{1/\delta\})$, such that the sum $\sum \mathbf{E}[F_i]$ ranges in $[1-\delta, 2^c]$ on $F^{-1}(-1)$ and in $[0, \delta 2^c]$ on $F^{-1}(+1)$. As a result,

$$\left\langle \sum \mathbf{E}[F_i], -H \circ P \right\rangle \geqslant \alpha(1-\delta) - \beta 2^c - (1 - \alpha - \beta)\delta 2^c] . \tag{4.3}$$

At the same time,

$$\left\langle \sum \mathbf{E}[F_i], -H \circ P \right\rangle \leqslant \sum_{i=1}^{2^c} 2^{c'} \operatorname{disc}_P(H) = 2^{c+c'} \operatorname{disc}_P(H) . \tag{4.4}$$

The bounds in (4.3) and (4.4) force (4.2). □

Since sign tensors H and $-H$ have the same discrepancy under any given distribution, we have the following alternate form of Theorem 4.1.

Corollary 4.2. *Let $F : X \to \{-1,+1\}$ be given, where $X = X_1 \times \cdots \times X_k$. Fix a function $H : X \to \{-1,+1\}$ and a probability distribution P on X. Put*

$$\alpha = P(F^{-1}(+1) \cap H^{-1}(+1)) ,$$
$$\beta = P(F^{-1}(+1) \cap H^{-1}(-1)) ,$$
$$Q = \log \frac{\alpha}{\beta + \operatorname{disc}_P(H)} .$$

Then

$$N(-F) \geqslant Q ,$$
$$MA(-F) \geqslant \min \left\{ \Omega(\sqrt{Q}),\ \Omega\left(\frac{Q}{\log\{2/\alpha\}} \right) \right\} .$$

At first glance, it is unclear how the nondeterministic bound of Theorem 4.1 and its counterpart Corollary 4.2 relate to the generalized discrepancy method. We now pause to make this relationship quite explicit. Nondeterminism and randomized computation are related in that a nondeterministic protocol with cost c for a function F gives a cost-c randomized protocol with error probability at most $1-2^{-c}$ on $F^{-1}(-1)$ and error probability 0 everywhere else. This is the setting of Theorem 4.1. The generalized discrepancy method, on the other hand, has a single error parameter ε for all inputs. To best convey this distinction between the two methods, we formulate a more general criterion yet, which allows for different errors on each input.

Theorem 4.3. *Let $F : X \to \{-1,+1\}$ be given, where $X = X_1 \times \cdots \times X_k$. Let c be the least cost of a public-coin protocol for F with error probability $E(x)$ on $x \in X$, for some $E : X \to [0,1]$. Then for all functions $H : X \to \{-1,+1\}$ and all probability distributions P on X,*

$$2^c \geqslant \frac{\langle F, H \circ P\rangle - 2\langle P, E\rangle}{\operatorname{disc}_P(H)}.$$

Proof. A public-coin protocol with cost c is a probability distribution on deterministic protocols with cost at most c. Then by hypothesis, there are random variables $\underline{\chi}_1, \underline{\chi}_2, \ldots, \underline{\chi}_{2^c} : X \to \{0,1\}$, each a cylinder intersection, and random variables $\underline{\sigma}_1, \underline{\sigma}_2, \ldots, \underline{\sigma}_{2^c} \in \{-1,+1\}$, such that

$$\left| F(x) - \mathbf{E}\left[\sum \underline{\sigma}_i \underline{\chi}_i(x)\right]\right| \leqslant 2E(x) \qquad \text{for } x \in X.$$

Therefore,

$$\left\langle F - \mathbf{E}\left[\sum \underline{\sigma}_i \underline{\chi}_i\right], H \circ P \right\rangle \leqslant 2\langle P, E\rangle.$$

On the other hand,

$$\left\langle F - \mathbf{E}\left[\sum \underline{\sigma}_i \underline{\chi}_i\right], H \circ P \right\rangle \geqslant \langle F, H \circ P\rangle - 2^c \operatorname{disc}_P(H)$$

by the definition of discrepancy. The theorem follows at once from the last two inequalities. □

5 Main result

We now prove the claimed separations of nondeterministic, co-nondeterministic, and Merlin-Arthur communication complexity. It will be easier to first obtain these separations by a probabilistic argument and only then sketch an explicit construction.

We start by deriving a suitable analytic property of the OR function.

Theorem 5.1. *There is a function $\psi : \{0,1\}^m \to \mathbb{R}$ such that:*

$$\sum_{z \in \{0,1\}^m} |\psi(z)| = 1, \tag{5.1}$$

$$\hat{\psi}(S) = 0 \qquad \textit{for } |S| \leqslant \Theta(\sqrt{m}), \tag{5.2}$$

$$\psi(0) > \frac{1}{6}. \tag{5.3}$$

Proof. Let $f : \{0,1\}^m \to \{-1,+1\}$ be given by $f(z) = 1 \Leftrightarrow z = 0^m$. It is well-known [19, 21] that $\deg_{1/3}(f) \geqslant \Omega(\sqrt{m})$. By Fact 3.5, there is a function $\psi : \{0,1\}^m \to \mathbb{R}$ that obeys (5.1), (5.2), and additionally satisfies

$$\sum_{z\in\{0,1\}^m} \psi(z)f(z) > \frac{1}{3}.$$

Finally,

$$2\psi(0) = \sum_{z\in\{0,1\}^m} \psi(z)\{f(z)+1\} = \sum_{z\in\{0,1\}^m} \psi(z)f(z) > \frac{1}{3},$$

where the second equality follows from $\hat{\psi}(\emptyset) = 0$. □

For the remainder of this section, it will be convenient to establish some additional notation following David and Pitassi [10]. Fix integers n,m with $n > m$. Let $\psi : \{0,1\}^m \to \mathbb{R}$ be a given function with $\sum_{z\in\{0,1\}^m} |\psi(z)| = 1$. Let d denote the least order of a nonzero Fourier coefficient of ψ. Fix a Boolean function $h : \{0,1\}^m \to \{-1,+1\}$ and the distribution μ on $\{0,1\}^m$ such that $\psi(z) \equiv h(z)\mu(z)$. For a mapping $\alpha : (\{0,1\}^n)^k \to \binom{[n]}{m}$, define a $(k+1)$-party communication problem $H_\alpha : (\{0,1\}^n)^{k+1} \to \{-1,+1\}$ by $H_\alpha(x,y_1,\ldots,y_k) = h(x|_{\alpha(y_1,\ldots,y_k)})$. Define a distribution P_α on $(\{0,1\}^n)^{k+1}$ by $P_\alpha(x,y_1,\ldots,y_k) = 2^{-(k+1)n+m}\mu(x|_{\alpha(y_1,\ldots,y_k)})$. The following theorem combines the pattern matrix method with a probabilistic argument.

Theorem 5.2 ([10]). *Assume that $n \geqslant 16\mathrm{e}m^2 2^k$. Then for a uniformly random choice of $\alpha : (\{0,1\}^n)^k \to \binom{[n]}{m}$,*

$$\mathop{\mathbf{E}}_{\alpha}\left[\mathrm{disc}_{P_\alpha}(H_\alpha)^{2^k}\right] \leqslant 2^{-n/2} + 2^{-d2^k+1}.$$

For completeness, we include a detailed proof of this result.

Proof (reproduced from the survey article [26], pp. 88–89). By Theorem 3.1,

$$\mathrm{disc}_{P_\alpha}(H_\alpha)^{2^k} \leqslant 2^{m2^k} \mathop{\mathbf{E}}_{Y} |\Gamma(Y)|, \tag{5.4}$$

where we put $Y = (y_1^0, y_1^1, \ldots, y_k^0, y_k^1) \in (\{0,1\}^n)^{2k}$ and

$$\Gamma(Y) = \mathop{\mathbf{E}}_{x}\left[\prod_{z\in\{0,1\}^k} \psi\left(x|_{\alpha\left(y_1^{z_1},y_2^{z_2},\ldots,y_k^{z_k}\right)}\right)\right].$$

For a fixed choice of α and Y, we will use the shorthand $S_z = \alpha(y_1^{z_1},\ldots,y_k^{z_k})$. To analyze $\Gamma(Y)$, one proves two key claims analogous to those in the two-party Theorem 3.6 (see [28, 26] for more detail).

Claim 5.3. *Assume that $\left|\bigcup_{z\in\{0,1\}^k} S_z\right| > m2^k - d2^{k-1}$. Then $\Gamma(Y) = 0$.*

Proof. If $|\bigcup S_z| > m2^k - d2^{k-1}$, then some S_z must feature more than $m-d$ elements that do not occur in $\bigcup_{u\neq z} S_u$. Since the Fourier transform of ψ is supported on characters of order d and higher, we conclude that the 2^k-fold product in the definition of $\Gamma(Y)$ has zero for its constant Fourier coefficient. This conclusion is of course equivalent to $\Gamma(Y) = 0$. $\square$

Claim 5.4. *For every Y, $|\Gamma(Y)| \leqslant 2^{-|\cup S_z|}$.*

Proof. Immediate from Proposition 2.1, by considering the distribution $\mu \times \cdots \times \mu$ on $(\{0,1\}^m)^{2^k}$. $\square$

In view of (5.4) and Claims 5.3 and 5.4, we have

$$\mathop{\mathbf{E}}_{\alpha}\left[\operatorname{disc}_{P_\alpha}(H_\alpha)^{2^k}\right] \leqslant \sum_{i=d2^{k-1}}^{m2^k-m} 2^i \mathop{\mathbf{P}}_{Y,\alpha}\left[\left|\bigcup S_z\right| = m2^k - i\right].$$

It remains to bound the probabilities in the last expression. With probability at least $1-k2^{-n}$ over the choice of Y, we have $y_j^0 \neq y_j^1$ for each $j = 1,2,\ldots,k$. Conditioning on this event, the fact that α is chosen uniformly at random means that the 2^k sets S_z are distributed independently and uniformly over $\binom{[n]}{m}$. A calculation now reveals that

$$\mathop{\mathbf{P}}_{Y,\alpha}\left[\left|\bigcup S_z\right| = m2^k - i\right] \leqslant k2^{-n} + \binom{m2^k}{i}\left(\frac{m2^k}{n}\right)^i \leqslant k2^{-n} + 8^{-i},$$

which completes the proof after summing over i. $\square$

We are ready to prove our main result. It may be helpful to contrast the proof to follow with the proof of the pattern matrix method (Theorem 3.7).

Theorem 5.5. *Let $k \leqslant (1-\varepsilon)\log n$, where $\varepsilon > 0$ is any given constant. Then there exists a function $F_\alpha : (\{0,1\}^n)^{k+1} \to \{-1,+1\}$ such that:*

$$N(F_\alpha) = O(\log n) \tag{5.5}$$

and

$$MA(-F_\alpha) = n^{\Omega(1)}. \tag{5.6}$$

In particular, $\mathsf{coNP}_k^{cc} \not\subseteq \mathsf{MA}_k^{cc}$ *and* $\mathsf{NP}_k^{cc} \neq \mathsf{coNP}_k^{cc}$*.*

Proof. Let $m = \lfloor n^\delta \rfloor$ for a sufficiently small constant $\delta = \delta(\varepsilon) > 0$. As usual, define $\mathrm{OR}_m : \{0,1\}^m \to \{-1,+1\}$ by $\mathrm{OR}_m(z) = 1 \Leftrightarrow z = 0^m$. Let $\psi : \{0,1\}^m \to \mathbb{R}$ be as guaranteed by Theorem 5.1. For a mapping $\alpha : (\{0,1\}^n)^k \to \binom{[n]}{m}$, let H_α and P_α be defined in terms of ψ as described earlier in this section. Then Theorem 5.2 shows the existence of α such that

$$\operatorname{disc}_{P_\alpha}(H_\alpha) \leqslant 2^{-\Omega(\sqrt{m})}. \tag{5.7}$$

Define $F_\alpha : (\{0,1\}^n)^{k+1} \to \{-1,+1\}$ by $F_\alpha(x,y_1,\ldots,y_k) = \text{OR}_m(x|_{\alpha(y_1,\ldots,y_k)})$. It is immediate from the properties of ψ that

$$P_\alpha(F_\alpha^{-1}(+1) \cap H_\alpha^{-1}(+1)) > \frac{1}{6}, \tag{5.8}$$

$$P_\alpha(F_\alpha^{-1}(+1) \cap H_\alpha^{-1}(-1)) = 0. \tag{5.9}$$

The sought lower bound in (5.6) now follows from (5.7)–(5.9) and Corollary 4.2.

On the other hand, as observed in [10], the function F_α has an efficient nondeterministic protocol. Namely, player 1 (who knows $y_1,\ldots,y_k$) nondeterministically selects an element $i \in \alpha(y_1,\ldots,y_k)$ and writes i on the shared blackboard. Player 2 (who knows x) then announces x_i as the output of the protocol. This yields the desired upper bound in (5.5). ☐

As promised, we will now sketch an explicit construction of the function whose existence has just been proven. For this, it suffices to invoke previous work by David, Pitassi, and Viola [11], who derandomized the choice of α in Theorem 5.2. More precisely, instead of working with a family $\{H_\alpha\}$ of functions, each given by $H_\alpha(x,y_1,\ldots,y_k) = h(x|_{\alpha(y_1,\ldots,y_k)})$, the authors of [11] posited a single function $H(\alpha,x,y_1,\ldots,y_k) = h(x|_{\alpha(y_1,\ldots,y_k)})$, where the new argument α is known to all players and ranges over a small, explicitly given subset A of all mappings $(\{0,1\}^n)^k \to \binom{[n]}{m}$. By choosing A to be pseudorandom, the authors of [11] forced the same qualitative conclusion in Theorem 5.2. This development carries over unchanged to our setting, and we obtain our main result.

Theorem 1.1 (Restated from p. 228). *Let $k \leqslant (1-\varepsilon)\log_2 n$, where $\varepsilon > 0$ is any given constant. Then there is an (explicitly given) function $F : (\{0,1\}^n)^k \to \{-1,+1\}$ with*

$$N(-F) = O(\log n)$$

and

$$MA(F) = n^{\Omega(1)}.$$

In particular, $\mathsf{coNP}_k^{cc} \not\subseteq \mathsf{MA}_k^{cc}$ *and* $\mathsf{NP}_k^{cc} \neq \mathsf{coNP}_k^{cc}$.

Proof. Identical to Theorem 5.5, with the described derandomization of α. ☐

6 A separation by the disjointness function

In this section, we revisit recent multiparty analyses of the disjointness function [6, 18, 7]. We will see that the program of the previous sections applies here essentially unchanged.

We start with some notation. Fix a function $\phi : \{0,1\}^m \to \mathbb{R}$ and an integer N with $m \mid N$. Define the (k,N,m,ϕ)-*pattern tensor* as the k-argument function $A : \{0,1\}^{m(N/m)^{k-1}} \times [N/m]^m \times \cdots \times [N/m]^m \to \mathbb{R}$ given by $A(x,V_1,\ldots,V_{k-1}) = \phi(x|_{V_1,\ldots,V_{k-1}})$, where

$$x|_{V_1,\ldots,V_{k-1}} = \left(x_{1,V_1[1],\ldots,V_{k-1}[1]},\ \ldots,\ x_{m,V_1[m],\ldots,V_{k-1}[m]}\right) \in \{0,1\}^m$$

and $V_j[i]$ denotes the ith element of the m-dimensional vector V_j. (Note that we index the string x by viewing it as a k-dimensional array of $m \times (N/m) \times \cdots \times (N/m) = m(N/m)^{k-1}$ bits.) This definition extends *pattern matrices* [28, 27] to higher dimensions. The two-party Theorem 3.6 has been adapted as follows to $k \geqslant 3$ players.

Theorem 6.1 ([6, 18, 7])**.** *Fix a function $h : \{0,1\}^m \to \{-1,+1\}$ and a probability distribution μ on $\{0,1\}^m$ such that*

$$\widehat{h \circ \mu}(S) = 0, \qquad |S| < d.$$

Let N be a given integer, $m \mid N$. Let P be the $(k,N,m,2^{-m(N/m)^{k-1}+m}(N/m)^{-m(k-1)}\mu)$-tensor. Let H be the (k,N,m,h)-pattern tensor. If $N \geqslant 4\mathrm{e}m^2(k-1)2^{2^{k-1}}/d$, then

$$\mathrm{disc}_P(F) \leqslant 2^{-d/2^{k-1}}.$$

A proof of this exact formulation is available in the survey article [26], pp. 85–86. We are now prepared to apply our techniques to the disjointness function.

Theorem 6.2. *Let N be a given integer, $m \mid N$. Let F be the (k,N,m,OR_m)-pattern tensor. If $N \geqslant 4\mathrm{e}m^2(k-1)2^{2^{k-1}}/d$, then*

$$N(-F) \geqslant \Omega\left(\frac{\sqrt{m}}{2^k}\right), \qquad MA(-F) \geqslant \Omega\left(\frac{\sqrt[4]{m}}{2^{k/2}}\right).$$

Proof. Let $\psi : \{0,1\}^m \to \mathbb{R}$ be as guaranteed by Theorem 5.1. Fix a function $h : \{0,1\}^m \to \{-1,+1\}$ and a distribution μ on $\{0,1\}^m$ such that $\psi(z) \equiv h(z)\mu(z)$. Let H be the (k,N,m,h)-pattern tensor. Let P be the $(k,N,m,2^{-m(N/m)^{k-1}+m}(N/m)^{-m(k-1)}\mu)$-pattern tensor, which is a probability distribution. Then by Theorem 6.1,

$$\mathrm{disc}_P(H) \leqslant 2^{-\Omega(\sqrt{m}/2^k)}. \tag{6.1}$$

On the other hand, it is clear from the properties of ψ that

$$P(F^{-1}(+1) \cap H^{-1}(+1)) > \frac{1}{6}, \tag{6.2}$$

$$P(F^{-1}(+1) \cap H^{-1}(-1)) = 0. \tag{6.3}$$

In view of (6.1)–(6.3) and Corollary 4.2, the proof is complete. □

The function F in Theorem 6.2 is a subfunction of the multiparty disjointness function $\mathrm{DISJ} : (\{0,1\}^n)^k \to \{-1,+1\}$, where $n = m(N/m)^{k-1}$ and

$$\mathrm{DISJ}(x_1,\ldots,x_k) = \bigvee_{j=1}^{n} \bigwedge_{i=1}^{k} x_{ij}.$$

Recall that disjointness has trivial nondeterministic complexity, $O(\log n)$. In particular, Theorem 6.2 shows that the disjointness function separates NP_k^{cc} from coNP_k^{cc} and witnesses that $\mathsf{coNP}_k^{cc} \not\subseteq \mathsf{MA}_k^{cc}$ for up to $k = \Theta(\log\log n)$ players.

References

[1] LÁSZLÓ BABAI, PÉTER FRANKL, AND JANOS SIMON: Complexity classes in communication complexity theory. In *Proc. 27th FOCS*, pp. 337–347. IEEE Comp. Soc. Press, 1986. [doi:10.1109/SFCS.1986.15] 228, 231

[2] LÁSZLÓ BABAI, NOAM NISAN, AND MÁRIÓ SZEGEDY: Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Computer and System Sciences*, 45(2):204–232, 1992. [doi:10.1016/0022-0000(92)90047-M] 227, 228, 229, 232, 233, 237

[3] PAUL BEAME, MATEI DAVID, TONIANN PITASSI, AND PHILIPP WOELFEL: Separating deterministic from randomized multiparty communication complexity. *Theory of Computing*, 6:201–225, 2010. [doi:10.4086/toc.2010.v006a009] 228

[4] 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. [doi:10.1137/060654645] 227

[5] ASHOK K. CHANDRA, MERRICK L. FURST, AND RICHARD J. LIPTON: Multi-party protocols. In *Proc. 15th STOC*, pp. 94–99. ACM Press, 1983. [doi:10.1145/800061.808737] 227

[6] ARKADEV CHATTOPADHYAY: Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proc. 48th FOCS*, pp. 449–458. IEEE Comp. Soc. Press, 2007. [doi:10.1109/FOCS.2007.30] 228, 229, 230, 236, 241, 242

[7] ARKADEV CHATTOPADHYAY AND ANIL ADA: Multiparty communication complexity of disjointness. In *Electron. Colloq. on Comput. Complexity (ECCC)*, January 2008. Report TR08-002. [ECCC:TR08-002] 228, 229, 230, 234, 236, 241, 242

[8] BENNY CHOR AND ODED GOLDREICH: Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988. [doi:10.1137/0217015] 229, 233

[9] FAN R. K. CHUNG AND PRASAD TETALI: Communication complexity and quasi randomness. *SIAM J. Discrete Math.*, 6(1):110–123, 1993. [doi:10.1137/0406009] 232

[10] MATEI DAVID AND TONIANN PITASSI: Separating NOF communication complexity classes RP and NP. In *Electron. Colloq. on Comput. Complexity (ECCC)*, February 2008. Report TR08-014. [ECCC:TR08-014] 228, 229, 236, 239, 241

[11] MATEI DAVID, TONIANN PITASSI, AND EMANUELE VIOLA: Improved separations between nondeterministic and randomized multiparty communication. *ACM Trans. Comput. Log.*, 1(2), 2009. [doi:10.1145/1595391.1595392] 228, 229, 236, 241

[12] RONALD DE WOLF: *A Brief Introduction to Fourier Analysis on the Boolean Cube*. Number 1 in Graduate Surveys. Theory of Computing Library, 2008. [doi:10.4086/toc.gs.2008.001] 230

[13] JOHAN HÅSTAD AND MIKAEL GOLDMANN: On the power of small-depth threshold circuits. *Comput. Complexity*, 1(2):113–129, 1991. [doi:10.1007/BF01272517] 227

[14] BALA KALYANASUNDARAM AND GEORG SCHNITGER: The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992. [doi:10.1137/0405044] 233

[15] HARTMUT KLAUCK: Lower bounds for quantum communication complexity. In *Proc. 42nd FOCS*, pp. 288–297. IEEE Comp. Soc. Press, 2001. [doi:10.1109/SFCS.2001.959903] 229, 233

[16] HARTMUT KLAUCK: Rectangle size bounds and threshold covers in communication complexity. In *Proc. 18th Conf. Comput. Complexity (CCC)*, pp. 118–134. IEEE Comp. Soc. Press, 2003. [doi:10.1109/CCC.2003.1214415] 228, 229, 233, 237

[17] EYAL KUSHILEVITZ AND NOAM NISAN: *Communication Complexity*. Cambridge University Press, New York, 1997. 229, 231, 233, 237

[18] TROY LEE AND ADI SHRAIBMAN: Disjointness is hard in the multi-party number-on-the-forehead model. In *Proc. 23rd Conf. Comput. Complexity (CCC)*, pp. 81–91. IEEE Comp. Soc. Press, 2008. [doi:10.1109/CCC.2008.29] 228, 229, 230, 234, 236, 241, 242

[19] NOAM NISAN AND MARIO SZEGEDY: On the degree of Boolean functions as real polynomials. *Comput. Complexity*, 4(4):301–313, 1994. [doi:10.1007/BF01263419] 239

[20] CHRISTOS H. PAPADIMITRIOU AND MICHAEL SIPSER: Communication complexity. In *Proc. 14th STOC*, pp. 196–200. ACM Press, 1982. [doi:10.1145/800070.802192] 228

[21] RAMAMOHAN PATURI: On the degree of polynomials that approximate symmetric Boolean functions. In *Proc. 24th STOC*, pp. 468–474. ACM Press, 1992. [doi:10.1145/129712.129758] 239

[22] RAN RAZ: The BNS-Chung criterion for multi-party communication complexity. *Comput. Complexity*, 9(2):113–122, 2000. [doi:10.1007/PL00001602] 232

[23] A. A. RAZBOROV: On the distributional complexity of disjointness. *Theoret. Comput. Sci.*, 106(2):385–390, 1992. [doi:10.1016/0304-3975(92)90260-M] 233

[24] A. A. RAZBOROV: Quantum communication complexity of symmetric predicates. *Izv. Math.*, 67(1):145–159, 2003. [doi:10.1070/IM2003v067n01ABEH000422] 229, 233

[25] 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. [doi:10.1016/0020-0190(93)90041-7] 227

[26] ALEXANDER A. SHERSTOV: Communication lower bounds using dual polynomials. *Bull. Eur. Assoc. Theor. Comput. Sci. (EATCS)*, 95:59–93, 2008. 229, 236, 239, 242

[27] ALEXANDER A. SHERSTOV: The pattern matrix method for lower bounds on quantum communication. In *Proc. 40th STOC*, pp. 85–94. ACM Press, 2008. [doi:10.1145/1374376.1374392] 228, 229, 233, 234, 235, 236, 242

[28] ALEXANDER A. SHERSTOV: Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. Preliminary version in 39th STOC, 2007. [doi:10.1137/08071421X] 228, 229, 230, 234, 235, 239, 242

[29] YAOYUN SHI AND YUFAN ZHU: Quantum communication complexity of block-composed functions. *Quantum Inf. Comput.*, 9(5–6):444–460, 2009. 229, 236

[30] ANDREW CHI-CHIH YAO: On ACC and threshold circuits. In *Proc. 31st FOCS*, pp. 619–627. IEEE Comp. Soc. Press, 1990. [doi:10.1109/FSCS.1990.89583] 227

AUTHORS

Dmitry Gavinsky
NEC Laboratories America Inc.
4 Independence Way, Suite 200
Princeton, NJ 08540
dmitry.gavinsky@gmail.com
http://www.nec-labs.com/~dmitry

Alexander A. Sherstov
Microsoft Research
Cambridge, MA 02142
sherstov@cs.utexas.edu
http://www.cs.utexas.edu/~sherstov

ABOUT THE AUTHORS

DMITRY GAVINSKY is a research staff member of the Quantum IT group at NEC Laboratories America. He completed his Ph. D. at the University of Calgary under the supervision of John Watrous and Richard Cleve. His research interests include quantum computing and computational complexity theory.

ALEXANDER SHERSTOV recently completed his Ph. D. at the University of Texas at Austin under the supervision of Adam Klivans. His research interests include computational complexity theory, computational learning theory, and quantum computing. He is currently a postdoctoral researcher at Microsoft Research.