Electronic Colloquium on Computational Complexity, Report No. 82 (2008)

Multiparty Communication Complexity and Threshold Circuit Size of AC^0

Paul Beame*
Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350
beame@cs.washington.edu

Dang-Trinh Huynh-Ngoc †
Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350
trinh@cs.washington.edu

September 11, 2008

Abstract

We prove an $n^{\Omega(1)}/2^{O(k)}$ lower bound on the randomized k-party communication complexity of read-once depth 4 AC^0 functions in the number-on-forehead (NOF) model for up to $\Theta(\log n)$ players. These are the first non-trivial lower bounds for general NOF multiparty communication complexity for any AC^0 function for $\omega(\log\log n)$ players. For non-constant k the bounds are larger than all previous lower bounds for any AC^0 function even for simultaneous communication complexity.

Our lower bounds imply the first superpolynomial lower bounds for the simulation of AC^0 by general $\mathsf{MAJ}\circ\mathsf{SYMM}\circ\mathsf{AND}$ circuits, showing that the well-known quasipolynomial simulations of AC^0 by such circuits are qualitatively optimal, even for read-once formulas of small constant depth.

We also exhibit a read-once depth 5 formula in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for k up to $\Theta(\log n)$ and derive an $\Omega(2^{\sqrt{\log n}/\sqrt{k}})$ lower bound on the randomized k-party NOF communication complexity of set disjointness for up to $\Theta(\log^{1/3} n)$ players which is significantly larger than the $O(\log\log n)$ players allowed in the best previous lower bounds for multiparty set disjointness given by Lee and Shraibman [18] and Chattopadhyay and Ada [9] (though the bound is not as strong as those in [18, 9] for $o(\log\log n)$ players). In addition, we prove lower bounds and separations for read-once AC^0 formulas of smaller depth: a depth 4 function in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}}-\mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for $k = O(\log n/\log\log n)$ and a lower bound of $n^{\Omega(1/k)}/2^{O(k)}$ for depth 3 read-once AC^0 formulas.

We derive all these results by extending the k-party generalization in [8, 18, 9] of the pattern matrix method of Sherstov [24, 26]. Using this technique, we derive a new sufficient criterion for stronger communication complexity lower bounds based on functions having many diverse subfunctions that do not have good low-degree polynomial approximations. This criterion guarantees that such functions have orthogonalizing distributions that are "max-smooth" as opposed to the "min-smooth" orthogonalizing distributions used by Sherstov [27] and Razborov and Sherstov [23] to analyze the sign-rank of symmetric and AC^0 functions. In order to obtain our strongest results we also need to consider a broader class of selector functions than those used in the pattern matrix method.

*Research supported by NSF grants CCF-0514870 and CCF-0830626
†Research supported by a Vietnam Education Foundation Fellowship

ISSN 1433-8092

1 Introduction

The multiparty communication complexity of AC^0 in the number-on-forehead (NOF) model has been an open question since Håstad and Goldmann [15] showed that any AC^0 or ACC^0 function has polylogarithmic randomized multiparty NOF communication complexity when its input bits are divided arbitrarily among a polylogarithmic number of players. This result is based on the simulations of AC^0 circuits [1] (and indeed ACC^0 circuits [31]) by quasipolynomial-size depth-3 circuits that consist of two layers of MAJORITY gates whose inputs are polylogarithmic-size AND gates of literals. By [12] these circuits can be converted to quasipolynomial-size depth-2 $\mathsf{SYMM} \circ \mathsf{AND}$ circuits consisting of a symmetric gate whose inputs are polylogarithmic AND gates, yielding deterministic communication complexity protocols with the same complexity as the above randomized protocols for AC^0 and ACC^0. These protocols can also be expressed as simpler simultaneous NOF protocols in which the players in parallel send their information to a referee who computes the answer [2].

It is natural to ask whether these upper bounds can be improved. In the case of ACC^0, Razborov and Wigderson [22] showed that quasipolynomial size is required to simulate ACC^0 based on results of Babai, Nisan, and Szegedy [4] which showed that the Generalized Inner Product function in ACC^0 requires multiparty NOF communication complexity $\Omega(n/4^k)$ which is polynomial in n for k up to $\Theta(\log n)$ players.

However, for AC^0 functions much less has been known. For the communication complexity of the set disjointness function with k players (which is in AC^0) there are lower bounds of the form $\Omega(n^{1/(k-1)}/(k-1))$ in the simultaneous NOF [28, 6] and $n^{\Omega(1/k)}/k^{O(k)}$ in the one-way NOF model [30]. These are sub-polynomial lower bounds for all non-constant values of k and, at best, polylogarithmic when k is $\Omega(\log n/\log\log n)$. Until recently, there were no lower bounds for general multiparty NOF communication complexity of any AC^0 function. As for circuit simulations of AC^0, Sherstov [24] recently showed that AC^0 cannot be simulated by polynomial-size $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits. However, there have been no non-trivial size lower bounds for the simulation of AC^0 by $\mathsf{MAJ} \circ \mathsf{MAJ} \circ \mathsf{AND}$ or even $\mathsf{SYMM} \circ \mathsf{AND}$ circuits with $\omega(\log\log n)$ bottom fan-in. As shown by Viola [29], sufficiently strong lower bounds for AC^0 in the multiparty NOF communication model, even for sub-logarithmic numbers of players, can yield quasipolynomial circuit size lower bounds.

We indeed produce such strong lower bounds. We show that there is an explicit linear-size fixed-depth AC^0 function that requires randomized k-party NOF communication complexity of $n^{\Omega(1)}/2^{O(k)}$ even for protocols that have error exponentially close to 1/2 (only exponentially small advantage). For all non-constant numbers of players this bound is larger than all previous multiparty NOF communication complexity lower bounds for AC^0 functions, even those in the weaker simultaneous model. The bound is non-trivial for up to $\Theta(\log n)$ players and is sufficient to apply Viola's arguments to produce fixed-depth AC^0 functions that require $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuits of $n^{\Omega(\log\log n)}$ size, thus showing that quasipolynomial size is indeed necessary for the simulation of AC^0.

The function for which we derive our strongest communication complexity lower bound is computable in depth 6 AC^0. In the case of protocols with error 1/3, we exhibit a hard function computable by simple read-once depth 4 formulas. We further show that the same lower bound applies to a function having read-once depth 5 formulas that also has $O(\log^2 n)$ nondeterministic communication complexity which shows that AC^0 contains functions in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for k up to $\Theta(\log n)$. As a consequence of the lower bound for this depth 5 function, we obtain $\Omega(2^{\sqrt{\log n}/\sqrt{k}-k})$

lower bounds on the k-party NOF communication complexity of set disjointness which is non-trivial for up to $\Theta(\log^{1/3} n)$ players. The best previous lower bounds of Lee and Shraibman [18] and Chattopadhyay and Ada [9] for set disjointness do not apply for $\omega(\log\log n)$ players.

We also show somewhat weaker lower bounds of $n^{\Omega(1)}/k^{O(k)}$, which is polynomial in n for up to $k = \Theta(\log / \log\log n)$ players, for another function in depth 4 AC^0 that has $O(\log^3 n)$ nondeterministic communication complexity and yet another in depth 3 AC^0 that has $n^{\Omega(1/k)}/2^{O(k)}$ randomized k-party communication complexity for $k = \Omega(\sqrt{\log n})$ players.

Methods and Related Work Recently, Sherstov introduced the so-called pattern matrix method which is a general method to convert analytic properties of Boolean functions to yield communication lower bounds for related Boolean functions [24, 26]. In [24], the analytic property used was large threshold degree, where it was also used to derive the lower bounds for simulations of AC^0 by $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits mentioned earlier. This was extended and generalized to large approximate degree in [26] to yield a strong new method for obtaining lower bounds for two-party communication complexity even with quantum communication.

The original version of the method [24] was then generalized for $k \geq 2$ players to pattern tensors by Chattopadhyay [8] to yield the first lower bounds for the general multiparty number-on-forehead communication complexity of any AC^0 function for $k \geq 3$ and used to prove exponential lower bounds for computation of AC^0 functions by $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{ANY}$ circuits with $o(\log\log n)$ bottom level fan-in. The general version of the pattern matrix method [26] was then extended in [18, 9] to pattern tensors to yield the first lower bounds for the general multiparty number-on-forehead communication complexity of set disjointness for more than 2 players, improving a long line of research on the problem [3, 28, 6, 30, 16, 7]. The communication lower bound for k players is $\Omega(n^{\frac{1}{k+1}})/2^{2^{O(k)}}$ which yields a non-trivial separation between randomized and nondeterministic k-party models for $k \leq \log\log n - O(\log\log\log n)$. This separation between randomized and nondeterministic communication complexity was extended by David and Pitassi and David, Pitassi, and Viola to $\Omega(\log n)$ players for significantly more complex functions than disjointness based on pseudorandom generators [11]. Their construction uses a generalization of the simple pattern tensor method used in [9]. More details about the pattern matrix method and the relationship between these papers may be found in [25]. David, Pitassi, and Viola asked the question of whether one could prove a separation for $\Omega(\log n)$ players using an AC^0 function or even whether one could prove any non-trivial lower bound for $\omega(\log\log n)$ players for any AC^0 function since their functions are also only in AC^0 for $k = O(\log\log n)$, a problem which our results resolve.

The high-level idea of the k-party version of the pattern matrix method (also known as the pattern tensor method) as described in [9] is as follows. Suppose that we want to prove k-party lower bounds for a function F. The general idea is to show that F can express some $f \circ \psi^m_{k,\ell}$ (specified below) which is a function that under any projection pattern chosen by the selector function $\psi_{k,\ell}$ yields f. If f has large approximate degree, then Sherstov showed [26] that there exists another function g and a distribution μ on inputs such that, with respect to μ, g is both highly correlated with f and orthogonal to all low-degree polynomials. It follows that $f \circ \psi^m_{k,\ell}$ is also highly correlated with $g \circ \psi^m_{k,\ell}$ and, using the discrepancy method for communication complexity lower bounds, it suffices to prove a discrepancy lower bound for the latter function. Thanks to the orthogonality of g to all low degree polynomials this is possible using an iterated application of the Cauchy-Schwartz inequality as in [4, 10]. For example, the bound for set disjointness $\textsc{Disj}_{k,n}(x) = \vee_{i=1}^n \wedge_{j=1}^k x_{ji}$, which more properly should be called set intersection, corresponds to the case that $f = \textsc{Or}$ which has approximate degree $\Omega(\sqrt{n})$.

In the two party case, Sherstov [27] and Razborov and Sherstov [23] extended Sherstov's pattern matrix method to yield sign-rank lower bounds for arbitrary symmetric functions applied to the size of a set intersection and for the AC^0_3 function $\mathrm{MP} \circ \psi^m_{2,\ell}$ where $\mathrm{MP}(x) = \wedge_{i=1}^{q} \vee_{j=1}^{4q^2} x_{ij}$ is the so-called Minsky-Papert function which has threshold degree (and therefore, approximate degree) $\Omega(q)$. The key to their arguments is to show that there are orthogonalizing distributions μ for their functions that are "min-smooth" in that they assign at least some fixed positive probability to any input vector on which their functions evaluate to true. (For example, probability at least $8^{-q}2^{-m-1}$ in the case of MP.)

We prove our results by showing that any function f for which there is a diverse collection of partial assignments ρ such that each of the subfunctions $f|_\rho$ of f requires large approximate degree, there is an orthogonalizing distribution μ for f that is "max-smooth" in that the probability of subsets defined by partial assignments cannot be too much larger than under the uniform distribution. The diversity of the partial assignments is determined by a parameter α so we call the degree bound the (ϵ, α)-approximate degree. This property is somewhat delicate but we give a general technique that allows us to convert functions of large ϵ-approximate degree to functions of large (ϵ, α)-approximate degree. (The property unfortunately does not apply to $\textsc{Or}$ but we are able to derive our lower bounds for $\textsc{Disj}_{k,n}$ via reduction.)

To obtain some of our results, in addition to using the construction of max-smooth orthogonalizing distributions we also further generalize the pattern tensor method beyond the version considered in [11] to consider other classes of selector functions.

The Hard Functions For any function f defined on $\{0,1\}^m$ and any "selector" function $\psi : \{0,1\}^{ks} \to \{0,1\}$ one can define a function $f \circ \psi^m$ on $\{0,1\}^{kms}$ bits by

$$f \circ \psi^m(x_0, \ldots, x_{k-1}) = f(\psi(x_{01}, \ldots, x_{(k-1)1}), \ldots, \psi(x_{0m}, \ldots, x_{(k-1)m}))$$

where each $x_{ij} \in \{0,1\}^s$. In the k-party NOF communication problem for $f \circ \psi^m$ on $x_0, x_1, \ldots, x_{k-1} \in \{0,1\}^n$ each player i holds x_i (but can only see all x_j for $j \neq i$) and they need to compute $f \circ \psi^m(x_0, \ldots, x_{k-1})$.

The pattern tensor selector function, which we denote by $\psi_{k,\ell}$, that is used in [8, 9, 18] is as follows: Let $s = \ell^{k-1}$. The s bits of each of its k inputs $X_0, \ldots, X_{k-1}$ are viewed as being indexed in a $(k-1)$-dimensional array of side ℓ and there is the promise that for each $j \in \{1, \ldots, k-1\}$, the bits of X_j are 1 in all entries in one of the ℓ slices in the j-th dimension and are 0 otherwise. In particular this implies that there is precisely one $s' \in [s]$ such that $\wedge_{j=1}^{k-1} X_{js'}$ is 1. Given the promise, the function $\psi_{k,\ell}(X)$ can be written as $\bigvee_{s'=1}^{s} \bigwedge_{j=0}^{k-1} X_{js'}$ which is the set disjointness function $\textsc{Disj}_{k,s}$. Note that because of the promise we can alternatively view $\psi_{k,\ell}$ as $\psi'_k(X) = \bigoplus_{s'=1}^{s} \bigwedge_{j=0}^{k-1} X_{js'}$ which is the Generalized Inner Product function $GIP_{k,s}$. (We could equally well replace $\oplus$ or $\bigvee$ by any function $t : \{0,1\}^s \to \{0,1\}$ that maps the all 0's input to 0 and each input with precisely one 1 to 1.)

If we set f to be $\textsc{Or}_m$ then $f \circ \psi^m_{k,\ell}$ is the set disjointness function $\textsc{Disj}_{k,n}$ and setting f to be $\textsc{Parity}_m$ makes $f \circ GIP^m_{k,s}$ the Generalized Inner Product (GIP) function $GIP_{k,n}$.

Our strongest lower bounds use a different selector function $\psi = \textsc{Index}_{\oplus^a_{k-1}}$ which computes the bit-wise $\oplus$ of the inputs to players 1 to $k-1$ and uses the result as an index to select a bit from player 0's input.

Organization In Section 2 we give the relevant properties of correlation and the construction of orthogonalizing distributions for functions of large ϵ-approximate degree used in previous work. In Section 3 we describe a general form of the method of [26, 9, 11] based on these orthogonalizing distributions and briefly discuss its limitations.

In Sections 4 and 5 we define a new notion which we call the (ϵ, α)-approximate degree of a function and show an example of how it can be applied by considering constructions based on the pattern tensor selector function $\psi_{k,\ell}$. In Section 6 we give our general method for producing functions of large (ϵ, α)-approximate degree from functions of large ϵ-approximate degree. In particular we prove that our construction applied to the $\textsc{Or}_q$ function, which yields the function $\textsc{Tribes}_{p,q}(x) = \vee_{i=1}^{q} \wedge_{j=1}^{p} x_{i,j}$, has (ϵ, α)-approximate degree for $\epsilon = 5/6$ for suitable values of p and q. We use $f = \textsc{Tribes}_{p,q}$ in our lower bounds for 1/3-error protocols. We also prove that the construction applied to a different function given by a read-once $\mathsf{AND} \circ \mathsf{OR}$ circuits has (ϵ, α)-approximate degree for every $\epsilon < 1$. We use this function in our lower bounds for protocols having exponentially small advantage.

In Section 7, we introduce the $\textsc{Index}_{\oplus^{a}_{k-1}}$ selector function and combine it with the functions from Section 6 to produce lower bounds on k-party randomized NOF communication complexity for AC^0 functions and the depth 5 separating functions between $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}}$ and $\mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for $k = O(\log n)$. We also use these results to derive communication complexity lower bounds for set disjointess. In Section 8 we derive the size lower bounds for $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ computing AC^0 functions.

In the appendix we derive lower bounds for somewhat simpler functions constructed from other selector functions, though the bounds are not as large as those in Section 7. In Appendix A we apply the lower bound from Section 4 for constructions using the pattern tensor selector function $\psi_{k,\ell}$ to produce k-party NOF communication complexity lower bounds for depth 3 functions for $k = O(\sqrt{\log n})$. In Appendix B we analyze a selector function that is a small parity of pattern tensor selector functions and show that from it we obtain depth 4 separating functions in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for $k = O(\log n / \log\log n)$.

2 Preliminaries

2.1 Notations and Terminology

We follow the notation used in [11]. We will assume that a Boolean function on m bits is a map $f : \{0,1\}^m \to \{-1,1\}$.

Correlation Let $f, g : \{0,1\}^m \to \mathbb{R}$ be two functions, and let μ be a distribution on $\{0,1\}^m$. We define the *correlation* between f and g under μ to be $\mathrm{Cor}_\mu(f,g) := \mathbf{E}_{x\sim\mu}[f(x)g(x)]$. If $\mathcal{G}$ is a class of functions $g : \{0,1\}^m \to \mathbb{R}$, we define the correlation between f and $\mathcal{G}$ under μ to be $\mathrm{Cor}_\mu(f, G) := \max_{g\in\mathcal{G}} \mathrm{Cor}_\mu(f,g)$.

Communication complexity We denote by $D^k(f)$, $R^k_\epsilon(f)$, and $N^k(f)$ the cost of the best k-party deterministic communication protocol for f, the cost of the best k-party randomized NOF communication protocol for f with two-sided error at most ϵ, and the cost of the best k-party nondeterministic communication protocol for f, respectively. We denote by Π^c_k the class of output functions of all deterministic k-party communication protocols of cost at most c.

Fact 2.1 (cf. [17]). *If there exists a distribution μ such that $\mathrm{Cor}_\mu(f, \Pi^c_k) \leq \epsilon$ then $R^k_{1/2-\epsilon/2}(f) \geq c$.*

Because of the following property of multiparty communication complexity, for the remainder of the paper we will find it convenient to designate the input to player 0 as x and the inputs to players 1 through $k-1$ as $y_1, \ldots, y_{k-1}$.

Lemma 2.2 ([4, 10])**.** *Let $f : \{0,1\}^{m \times k} \to \mathbb{R}$ and $\mathcal{U}$ be the uniform distribution over some set $X \times Y$ where $Y = Y_1 \times \cdots \times Y_{k-1}$. Then,*

$$\mathrm{Cor}_{\mathcal{U}}(f, \Pi_k^c)^{2^{k-1}} \leq 2^{c \cdot 2^{k-1}} \cdot \mathbf{E}_{y^0, y^1 \in Y}\left[\left|\mathbf{E}_{x \in X}\left[\prod_{u \in \{0,1\}^{k-1}} f(x, y^u)\right]\right|\right]$$

where $y^u = (y_1^{u_1}, \ldots, y_{k-1}^{u_{k-1}})$ for $u \in \{0,1\}^{k-1}$.

Approximate and threshold degree Given any $0 \leq \epsilon < 1$, the ϵ-approximate degree of f, $deg_\epsilon(f)$, is the smallest d for which there exists a multivariate real-valued polynomial p of degree d such that $||f - p||_\infty = \max_x |f(x) - p(x)| \leq \epsilon$. Following [21] we have the following property of the approximate degree of OR.

Proposition 2.3. *Let $\mathrm{OR}_m : \{0,1\}^m \to \{1, -1\}$. For $0 \leq \epsilon < 1$, $deg_\epsilon(\mathrm{OR}_m) \geq \sqrt{(1-\epsilon)m/2}$.*

The threshold degree of f, $thr(f)$, is the smallest d for which there exists a multivariate real-valued polynomial p of degree d such that $f(x) = sign(p(x))$. Because the domain of f is finite, we can assume without loss of generality that $p(x) \neq 0$ for all x since we can shift p by adding the constant $\frac{1}{2} \cdot \max_{x : f(x) < 0} |f(x)|$ to p. Thus the condition on p can be replaced by $f(x)p(x) > 0$ on every input x. Hence it follows that $thr(f) = \min_{\epsilon < 1} deg_\epsilon(f)$. For this reason, we write $thr(f) = deg_{<1}(f)$.

Following [20] we have the following property of the threshold degree of (the dual of) the Minsky-Papert function.

Proposition 2.4. *Let $\mathrm{MP'}_{q,q'} : \{0,1\}^{q \cdot q'} \to \{-1, 1\}$ be defined by $\mathrm{MP'}_{q,q'}(x) := \wedge_{i=1}^{q} \vee_{j=1}^{q'} x_{ij}$. Let $m > 0$ be such that $m \leq q$ and $4m^2 \leq q'$. Then $thr(\mathrm{MP'}_{q,q'}) \geq m$.*

Define an inner product $\langle, \rangle$ on the set of functions $f : \{0,1\}^m \to \mathbb{R}$ by $\langle f, g \rangle = \mathbf{E}[f \cdot g]$. For $S \subseteq [m]$, let $\chi_S : \{0,1\}^m \to \{-1, 1\}$ be the function $\chi_S = \prod_{i \in S}(-1)^{x_i}$. The χ_S for $S \subseteq [m]$ form an orthonormal basis of this space.

Lemma 2.5 ([26])**.** *If $f : \{0,1\}^m \to \{-1, 1\}$ is a Boolean function with $deg_\epsilon(f) \geq d$ then there exists a function $g : \{0,1\}^m \to \{-1, 1\}$ and a distribution μ on $\{0,1\}^m$ such that:*

1. *$\mathrm{Cor}_\mu(g, f) > \epsilon$; and*
2. *for every $S \subseteq [m]$ with $|S| < d$ and every function $h : \{0,1\}^{|S|} \to \mathbb{R}$, $\mathbf{E}_{x \sim \mu}[g(x) \cdot h(x|S)] = 0$.*

Proof. Let Φ_d be the space of polynomials of degree less than d. By definition, $deg_\epsilon(f) \geq d$ if and only if $\min_{q \in \Phi_d} ||f - q||_\infty > \epsilon$. By duality of norms we have $\min_{q \in \Phi_d} ||f - q||_\infty = \max_{p \in \Phi_d^\perp,\ ||p||_1 = 1} \langle f, p \rangle$. Writing $\mu(x) = |p(x)|$ the condition $||p||_1 = 1$ implies that μ is a probability distribution and letting $g(x) = p(x)/\mu(x)$ for $\mu(x) \neq 0$ and $g(x) = 1$ if $\mu(x) = 0$. Then $p(x) = \mu(x)g(x)$. Therefore

$$\epsilon < \langle f, p \rangle = \mathbf{E}[f \cdot p] = \mathbf{E}[f \cdot g \cdot \mu] = \mathbf{E}_{x \sim \mu}[f(x)g(x)] = \mathrm{Cor}_\mu(f, g).$$

Moreover since $p \in \Phi_d^\perp$, we have $0 = \langle \chi_S, p \rangle = \mathbf{E}_{x \sim \mu}[\chi_S(x)g(x)]$. Now for $h : \{0,1\}^{|S|} \to \mathbb{R}$ for $|S| \leq d$, $h(x|S)$ can be expressed as a degree $|S|$ polynomial and by linearity $\mathbf{E}_{x \sim \mu}[g(x) \cdot h(x|S)] = 0$. □

We will extend this lemma in Section 4 using more general LP duality.

Circuit complexity $\mathsf{AC^0}$ is the class of functions $f : \{0,1\}^* \to \{0,1\}$ computed by polynomial size circuits (or formulas) of constant depth having $\neg$ gates and unbounded fan-in $\wedge$ and $\vee$ gates. A formula is a $\mathsf{\Sigma_1}$ formula if it is a clause and a $\mathsf{\Pi_1}$ formula if it is a term. For $i \geq 1$, a $\mathsf{\Sigma_{i+1}}$ formula is an unbounded fan-in $\vee$ of $\mathsf{\Pi_i}$ formulas and a $\mathsf{\Pi_{i+1}}$ formula is an unbounded fan-in $\wedge$ of $\mathsf{\Sigma_i}$ formulas. A circuit F is a *read-once formula* if and only if its circuit graph is a tree and its leaves are labelled by distinct variables. The output gate of F is at the top and its inputs are at the bottom of the circuit. We let AND denote the class of all unbounded fan-in $\wedge$ functions, SYMM denote the class of all symmetric functions and $\mathsf{MAJ} \subset \mathsf{SYMM}$ denote the class of all majority functions. Given a sequence of classes of functions $\mathsf{C_1}, \mathsf{C_2}, \ldots \mathsf{C_d}$, we let $\mathsf{C_1} \circ \mathsf{C_2} \circ \cdots \circ \mathsf{C_d}$ to be the class of all circuits of depth d whose inputs are given by variables and their negations and whose gates at the i-th level from the top are chosen from $\mathsf{C_i}$. Thus, for example, $\mathsf{\Pi_{i+1}} = \mathsf{AND} \circ \mathsf{\Sigma_i}$.

3 A General Form of the Correlation Method

We give a generalization of the method as described in [9, 11], which extend ideas of [24, 26] from 2-party to k-party communication complexity. We will concentrate on specific details at those points where we are extending the method.

Definition We call any function $\psi : \{0,1\}^{ks} \to \{0,1\}$ with the following property a *selector function*:

- There exist sets $D_{\psi,1}, \ldots, D_{\psi,(k-1)} \subseteq \{0,1\}^s$ such that for any $Y = (Y_1, \ldots, Y_{k-1}) \in D_\psi := D_{\psi,1} \times \cdots \times D_{\psi,(k-1)}$,

$$\Pr_{X \in \{0,1\}^s}[\psi(X,Y) = 0] = \Pr_{X \in \{0,1\}^s}[\psi(X,Y) = 1] = 1/2.$$

Example *One example of a selector function ψ is the* pattern tensor *function $\psi_{k,\ell}$ mentioned in the introduction and used in [9, 18]. In this example, $s = \ell^{k-1}$ and the s bits are viewed as being indexed in a $(k-1)$-dimensional array of side ℓ. $D_{\psi_{k,\ell},j}$ consists of the ℓ vectors $Y_j \in \{0,1\}^s$ that are 1 in all entries in one of the ℓ slices along the j-th dimension of this array and are 0 in every other entry. For $X \in \{0,1\}^s$ and $Y = (Y_1, \ldots, Y_{k-1}) \in \{0,1\}^{(k-1)s}$ such that each $Y_j \in D_{\psi_{k,\ell},j}$ we can express*

$$\psi_{k,\ell}(X,Y) = \bigvee_{s'=1}^{s} (X_{s'} \wedge \bigwedge_{j=1}^{k-1} Y_{js'}).$$

Because of the promise on the entries of Y, we know that $\bigwedge_{j=1}^{k-1} Y_{js'}$ is 1 for precisely one choice of s' so $\psi_{k,\ell}$ has the effect of selecting precisely one bit of X.

Note that, as in the example of $\psi_{k,\ell}$, any function that uses the vector of inputs for players 1 to $k-1$ to select a single bit from player 0's input as the output is a selector function. This is the form considered in [11]. Although our more general definition is not necessary for most of our results we will make use of it in section B in the appendix.

For any function $f : \{0,1\}^m \to \{1,-1\}$ and any selector function ψ we can define a new function $f \circ \psi^m$ on $\{0,1\}^{kms}$ bits by

$$\begin{aligned} f \circ \psi^m(x,y) &= f \circ \psi^m(x, y_1, \ldots, y_{k-1}) \\ &= f(\psi(x_1, y_{*1}), \ldots, \psi(x_m, y_{*m})) \end{aligned}$$

where $y_{*i} = (y_{1i}, \ldots, y_{(k-1)i})$ for $i \in [m]$. We will write $z_i = \psi(x_i, y_{*i})$ and $z = (z_1, \ldots, z_m)$ for the input to f. In the k-party NOF communication problem for $f \circ \psi^m$ on input $x, y_1, \ldots, y_{k-1} \in \{0,1\}^n$, player 0 holds x and can see all the y_i and each other player i holds y_i (but can only see x and all y_j for $j \neq i$) and they need to compute $f \circ \psi^m(x, y_1, \ldots, y_{k-1})$.

Given a Boolean function f on m bits having large 5/6-approximate degree d we want to lower bound $R^k_{1/3}(f \circ \psi^m)$.

From Lemma 2.5, we obtain another Boolean function g and a distribution μ such that:

1. $\mathrm{Cor}_\mu(g, f) \geq 5/6$; and

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

Based on μ, we define a distribution λ on $kn = kms$ bits in a straightforward way as follows: Choose $y = (y_1, \ldots, y_{k-1})$ uniformly from $D^{(m)}_\psi := D^m_{\psi,1} \times \cdots \times D^m_{\psi,(k-1)}$ and choose x uniformly subject to the constraint that z is distributed according to μ. More precisely, define

$$\lambda(x, y) := \frac{\mu(z_1, \ldots, z_m)}{2^{n-m}|D_\psi|^m}$$

where $z_i = \psi(x_i, y_{*i})$ for $y \in D^{(m)}_\psi$ and 0 otherwise. (Note that $|D^{(m)}_\psi| = |D_\psi|^m$.)

Example *In particular, when ψ is the pattern tensor $\psi_{k,\ell}$, $|D_{\psi,j}| = \ell$ for all j, $|D_\psi| = \ell^{k-1}$, and* $\lambda(x, y) = \dfrac{\mu(z_1, \ldots, z_m)}{2^{n-m}\ell^{(k-1)m}} = \dfrac{\mu(z_1, \ldots, z_m)}{2^{n-m}s^m}$.

It is straightforward that, since each $z_i = \psi(x_i, y_{*i})$ is a uniformly chosen random bit given a fixed $y_{*i} \in D_\psi$ and random x_i, we have $\mathrm{Cor}_\lambda(f \circ \psi^m, g \circ \psi^m) = \mathrm{Cor}_\mu(f, g) \geq 5/6$. Consequently, by the triangle inequality,

$$\mathrm{Cor}_\lambda(f \circ \psi^m, \Pi^c_k) \leq \mathrm{Cor}_\lambda(g \circ \psi^m, \Pi^c_k) + 1/6.$$

Therefore we only need to bound $\mathrm{Cor}_\lambda(g \circ \psi^m, \Pi^c_k)$.

Let $z_i = \psi(x_i, y_{*i})$ for $i \in [m]$. By Lemma 2.2, if we let $\mathcal{U}$ be the uniform distribution on the set of $(x, y) \in \{0,1\}^{ms} \times D^{(m)}_\psi$ we have

$$\begin{aligned}\mathrm{Cor}_\lambda(g \circ \psi^m, \Pi^c_k)^{2^{k-1}} &= 2^{m2^{k-1}} \mathrm{Cor}_{\mathcal{U}}(\mu(z_1, \ldots, z_m) g(z_1, \ldots, z_m), \Pi^c_k)^{2^{k-1}}, \\ &\leq 2^{(c+m)\cdot 2^{k-1}} \cdot \mathbf{E}_{y^0, y^1 \in D^{(m)}_\psi} H(y^0, y^1),\end{aligned}$$

where

$$H(y^0, y^1) := \Big|\mathbf{E}_x\big[\prod_{u \in \{0,1\}^{k-1}} \mu(z^u) g(z^u)\big]\Big|,$$

for $z^u = (z^u_1, \ldots, z^u_m)$ where $z^u_i = \psi(x_i, y^u_{*i})$ and $y^u = (y^{u_1}_1, \ldots, y^{u_{k-1}}_{k-1})$. (Note that $y^{0\ldots0} = y^0$ and $y^{1\ldots1} = y^1$.)

For fixed $y^0, y^1 \in D^{(m)}_\psi$ and $i \in [m]$, we call i *good for* (y^0, y^1) if the set of 2^{k-1} random variables $z^u_i = \psi(x_i, y^u_{*i})$ for $u \in \{0,1\}^{k-1}$ are mutually independent; otherwise we call i *bad for* (y^0, y^1). Let $R_\psi(y^0, y^1)$ be the set of $i \in [m]$ that are bad for (y^0, y^1) and let $r_\psi(y^0, y^1) = |R_\psi(y^0, y^1)|$.

Example *In the case that ψ is the pattern tensor $\psi_{k,\ell}$, the random variables z_i^u for $u \in \{0,1\}^{k-1}$ will be independent if and only if y_{*i}^u and y_{*i}^v select different bits of x_i for every $u \neq v$. This will be true for u and v if and only if there is some $j \in [k-1]$ such that $y_{ji}^u \neq y_{ji}^v$. However, since this must hold for every u and v, in particular those that agree everywhere except for a single bit, it is necessary and sufficient for independence that $y_{ji}^0 \neq y_{ji}^1$ for every $j \in [k-1]$. Therefore $r_{\psi_{k,\ell}}(y^0, y^1)$ is the number of $i \in [m]$ such that $y_{ji}^0 = y_{ji}^1$ for some $j \in [k-1]$.*

We rely on the following three propositions to continue the proof. The special cases of Proposition 3.1 and Proposition 3.2 when ψ is the pattern tensor $\psi_{k,\ell}$ were proved in [9]. Proposition 3.3 only concerns $\psi_{k,\ell}$. For completeness we give the proofs of Proposition 3.1 and Proposition 3.3 at the end of this section. We will prove an extension of Proposition 3.2 in Section 4 so we will not give the details here.

Proposition 3.1. *If $r = r_\psi(y^0, y^1) < d$, then $H(y^0, y^1) = 0$.*

Proposition 3.2. *If $r = r_\psi(y^0, y^1)$ then $H(y^0, y^1) \leq \dfrac{2^{(2^{k-1}-1)r}}{2^{2^{k-1}m}}$.*

Proposition 3.3. *For $r \geq d$, $\Pr_{y^0, y^1 \in D_{\psi_{k,\ell}}^{(m)}}[r_{\psi_{k,\ell}}(y^0, y^1) = r] \leq \left(\frac{e(k-1)m}{r\ell}\right)^r$.*

In [9, 18], to prove the lower bound for $\textsc{Disj}_{k,n}$, the function f is set to $\textsc{Or}_m$ (and ψ is $\psi_{k,\ell}$). By Proposition 2.3, $d = deg_{5/6}(\textsc{Or}_m) \geq \sqrt{m/12}$. Plugging the bound in Proposition 3.3 together with the bounds from Proposition 3.1 for $r < d$ and from Proposition 3.2 when $r \geq d$ into the above correlation inequality it is not hard to show that

$$\mathrm{Cor}_\lambda(g \circ \psi_{k,\ell}^m, \Pi_k^c) \leq \frac{2^c}{2^{d/2^k}},$$

for $\ell > \frac{2^{2^k} kem}{d}$. Hence for suitable $k = O(\log\log n)$ we derive a polynomial lower bound on $R_{1/3}^k(\textsc{Disj}_{k,n}) \geq c$.

The key limitation of the above technique is the required lower bound on ℓ which follows from the weakness of the upper bound in Proposition 3.2 and from the inefficiency of the selector function $\psi_{k,\ell}$.

We first address the weakness of Proposition 3.2, which is implied by how little can be assumed about the orthogonalizing distribution μ given by Lemma 2.5. In particular, the arguments in [26, 9, 18] all allow that μ may assign all of its probability mass to small subsets of points defined by partial assignments. Indeed, when the function f is $\textsc{Or}_m$, this is not far from tight. However, we will show that for other very simple functions f one can choose the orthogonalizing distribution μ so that it does not assign too much weight on such small sets of points; that is, μ is "max-smooth". To guarantee this property of μ we need to strengthen Lemma 2.5 by assuming more of f than just large approximate degree which we will do in the next section.

Later in Section B we address the inefficiency of the selector function. David, Pitassi, and Viola [11] already tackled some of this inefficiency by using 2^k-wise independent distributions which yield selector functions that are unfortunately outside of AC^0 for $k = \omega(\log\log n)$. We use our more general notion of selector functions to design efficient selector functions that are in AC^0 and produce $n^{\Omega(1)}$ lower bounds for k up to $\Theta(\log n/\log\log n)$ players.

Before moving on to detail these improvements, for completeness we give the proofs of Proposition 3.1 and Proposition 3.3.

Proof of Proposition 3.1. We have $H(y^0,y^1) = \left|\mathbf{E}_x\left[\prod_{u\in\{0,1\}^{k-1}}\mu(z^u)g(z^u)\right]\right|$. Let $\mathcal{Z} = \mathcal{Z}^{0\ldots0}\mathcal{Z}^{0\ldots1}\cdots\mathcal{Z}^{1\ldots1}$ be the joint distribution induced on $\{z^u\}_{u\in\{0,1\}^{k-1}}$ by taking x uniformly at random. By construction when x is taken uniformly at random, z^u is uniformly distributed in $\{0,1\}^m$ for any $u\in\{0,1\}^{k-1}$ so each $\mathcal{Z}^u$ is a uniform distribution. For each choice of $z^{0\ldots0}$ we will also consider the conditional distribution $\mathcal{Z}^{\neq0\ldots0}|z^{0\ldots0}$ on $\{z^u\}_{u\neq0\ldots0}$ which is derived from $\mathcal{Z}$ by conditioning on $\mathcal{Z}^{0\ldots0}=z^{0\ldots0}$. Then,

$$\begin{aligned}H(y^0,y^1) &= \left|\mathbf{E}_{\{z^u\}_{u\in\{0,1\}^{k-1}}\sim\mathcal{Z}}\left[\prod_{u\in\{0,1\}^{k-1}}\mu(z^u)g(z^u)\right]\right| \\ &= \left|\mathbf{E}_{z^{0\ldots0}}\left[\mu(z^{0\ldots0})g(z^{0\ldots0})\cdot\mathbf{E}_{\{z^u\}_{u\neq0\ldots0}\sim\mathcal{Z}^{\neq0\ldots0}|z^{0\ldots0}}\prod_{u\neq0\ldots0}\mu(z^u)g(z^u)\right]\right|.\end{aligned}$$

We now consider the conditional distribution in the inner expectation above. For any i that is good for (y^0,y^1) the set of 2^{k-1} random variables $\{z_i^u\}_{u\in\{0,1\}^{k-1}}$ are independent. Therefore for any i that is good for (y^0,y^1), conditioning of $\mathcal{Z}^{\neq0\ldots0}$ on $z^{0\ldots0}$ is equivalent to conditioning on $(z_i^{0\ldots0})_{i\in R_\psi(y^0,y^1)}$, the portions of $z^{0\ldots0}$ on those $i\in[m]$ that are bad for (y^0,y^1). Therefore

$$\begin{aligned}&\mathbf{E}_{\{z^u\}_{u\neq0\ldots0}\sim\mathcal{Z}^{\neq0\ldots0}|z^{0\ldots0}}\prod_{u\neq0\ldots0}\mu(z^u)g(z^u) \\ &= \mathbf{E}_{\{z^u\}_{u\neq0\ldots0}\sim\mathcal{Z}^{\neq0\ldots0}|(z_i^{0\ldots0})_{i\in R_\psi(y^0,y^1)}}\prod_{u\neq0\ldots0}\mu(z^u)g(z^u).\end{aligned}$$

This quantity is some function Q of $z^{0\ldots0}$ that depends on only the $r=r_\psi(y^0,y^1)$ variables $(z_i^{0\ldots0})_{i\in R_\psi(y^0,y^1)}$. Therefore

$$\begin{aligned}H(y^0,y^1) &= \left|\mathbf{E}_{z^{0\ldots0}}\left[\mu(z^{0\ldots0})g(z^{0\ldots0})Q(z^{0\ldots0})\right]\right| \\ &= 0\end{aligned}$$

by the second property of μ and g given in Lemma 2.5 since $r<d$. □

Proof of Proposition 3.3. As we have noted, $r_{\psi_{k,\ell}}(y^0,y^1)$ is the number of $i\in[m]$ such that $y^0_{ji}=y^1_{ji}$ for some $j\in[k-1]$. There are ℓ elements in $D_{\psi_{k,\ell},j}$ for each j so the probability that $y^0_{ji}=y^1_{ji}$ is $1/\ell$. Therefore the probability that $y^0_{ji}=y^1_{ji}$ for some $j\in[k-1]$ is at most $(k-1)/\ell$. By the independence of the choices for different $i\in[m]$ $\Pr_{y^0,y^1\in D^{(m)}_{\psi_{k,\ell}}}[r_{\psi_{k,\ell}}(y^0,y^1)=r]\le\binom{m}{r}(\frac{k-1}{\ell})^r\le(\frac{em(k-1)}{r\ell})^r$.
□

4 Beyond approximate degree: a new sufficient criterion for strong communication complexity bounds

In this section we introduce our notion of (ϵ,α)-approximate degree and how it implies our main technical theorem on the general correlation method. We also give an example of how it can be applied in conjunction with the pattern tensor selector function to produce larger lower bounds than those based on ϵ-approximate degree.

A $\rho \in \{0,1,*\}^m$ is called a *restriction*. For any restriction ρ, let $\mathrm{unset}(\rho) \subseteq [m]$ be the set of star positions in ρ, let $|\rho| = m - |\mathrm{unset}(\rho)|$, and let C_ρ be the set of all $x \in \{0,1\}^m$ such that for any $1 \leq i \leq m$, either $\rho_i = *$ or $\rho_i = x_i$. Hence $|C_\rho| = 2^{m-|\rho|}$. Given a restriction $\rho \in \{0,1,*\}^m$ and a function f on $\{0,1\}^m$, we define $f|_\rho$ on $\{0,1\}^{m-|\rho|}$ in the natural way. We also define $R^r_m \subset \{0,1,*\}^m$ to be the set of all restrictions with $|\mathrm{unset}(\rho)| = r$.

The approximate degree of a function f says how hard it is to approximate f. In this paper, we need a stronger notion which requires that many widely distributed restrictions of f also require large approximate degree.

Definition Given $0 < \epsilon, \alpha \leq 1$ and $d > 0$, let $\Pi = \Pi_{d,\epsilon}(f) \subseteq \{0,1,*\}^m$ be the set of restrictions such that for any $\pi \in \Pi$, $deg_\epsilon(f|_\pi) \geq d$. We say that f the (ϵ,α)-approximate degree of f is d, denoted as $deg_{\epsilon,\alpha}(f) = d$, if restrictions in (a subset of) Π are spread out "evenly".

Formally, the (ϵ,α)-approximate degree of f is the minimum integer d such that there is a distribution ν on $\Pi = \Pi_{d,\epsilon}(f)$ such that for any $\rho \in \{0,1,*\}^m$ with $|\rho| \geq d$,

$$\Pr_{\pi \sim \nu}[C_\rho \cap C_\pi \neq \emptyset] \leq 2^{|\rho|^\alpha - |\rho|}.$$

The distribution ν is the *witness* for the (ϵ,α)-approximate degree of f. Note that $deg_\epsilon(f) = deg_{\epsilon,1}(f)$.

Definition We define the $(<\epsilon,\alpha)$-approximate degree of f to be the minimum over all $\epsilon' < \epsilon$ of the (ϵ',α)-approximate degree of f. That is, $deg_{<\epsilon,\alpha}(f) = \min_{\epsilon'<\epsilon} deg_{\epsilon',\alpha}(f)$. As we write $thr(f) = deg_{<1}(f)$, we will usually say "α-threshold degree" for $(<1,\alpha)$-approximate degree.

We will use this definition to prove our main technical theorem on the application of the general correlation method. To prove the theorem, we first need the following consequence of large (ϵ,α)-approximate degree. We postpone its proof to Section 5.

Lemma 4.1 (extension of Lemma 2.5)**.** *Given $0 < \epsilon, \alpha \leq 1$. If $f : \{0,1\}^m \to \{-1,1\}$ is a Boolean function with $(<\epsilon,\alpha)$-approximate degree at least d, then there exist a function $g : \{0,1\}^m \to \{-1,1\}$ and a distribution μ on $\{0,1\}^m$ such that:*

1. $\mathrm{Cor}_\mu(g,f) \geq \epsilon$*;*
2. *for every $T \subseteq [m]$ with $|T| < d$ and every function $h : \{0,1\}^{|T|} \to \mathbb{R}$, $\mathbf{E}_{x\sim\mu}[g(x) \cdot h(x|T)] = 0$; and*
3. *for any restriction ρ with $|\rho| \geq d$, $\mu(C_\rho) \leq 2^{|\rho|^\alpha - |\rho|}/\epsilon$.*

Note that, although the upper bound on $\mu(C_\rho)$ may seem quite weak, it will be sufficient to obtain an exponential improvement in the dependence of communication complexity lower bounds on k. Moreover, we note in Section 5 that for any function f computed by an AC^0 circuit the assumption and the upper bound are essentially the best possible for d polynomial in m.

We now use Lemma 4.1 to prove an improvement of Proposition 3.2. This is the key to our improved lower bounds.

Lemma 4.2. *If $f : \{0,1\}^m \to \{1,-1\}$ has $(<\epsilon,\alpha)$-approximate degree at least d, if g and μ are given by the application of Lemma 4.1 to f, and if $r = r_\psi(y^0,y^1) \geq d$, then*

$$H(y^0,y^1) \leq \frac{2^{(2^{k-1}-1)r^\alpha}}{2^{2^{k-1}m}\epsilon^{2^{k-1}-1}}.$$

Proof. Note that by definition of $R_\psi(y^0, y^1)$, conditioned on each fixed value of $x_{R_\psi(y^0,y^1)} = (x_i)_{i \in R_\psi(y^0,y^1)}$ the random variable $z^u = z^u(x, y^0, y^1)$ is statistically independent of all z^v for $v \neq u$. For convenience of notation we assume without loss of generality that $R_\psi(y^0, y^1) = \{1, \ldots, r\}$.

Since g is 1/-1 valued,

$$\begin{aligned}
H(y^0, y^1) &= \Big|\mathbf{E}_x\big[\prod_{u \in \{0,1\}^{k-1}} \mu(z^u) g(z^u)\big]\Big| \\
&\leq \mathbf{E}_x\Big|\prod_{u \in \{0,1\}^{k-1}} \mu(z^u) g(z^u)\Big| \\
&= \mathbf{E}_x\big[\prod_{u \in \{0,1\}^{k-1}} \mu(z^u)\big] \\
&\leq \mathbf{E}_x[\mu(z^{0\ldots0})] \\
&\quad \times \max_{x_1,\ldots,x_r} \mathbf{E}_{x_{r+1},\ldots,x_m}\big[\prod_{u \neq 0\ldots0} \mu(z^u)\big] \\
&= \mathbf{E}_x[\mu(z^{0\ldots0})] && (1) \\
&\quad \times \max_{x_1,\ldots,x_r} \prod_{u \neq 0\ldots0} \mathbf{E}_{x_{r+1}\ldots x_m}\big[\mu(z^u)\big] && (2)
\end{aligned}$$

where $z_i^u = \psi(x_i, y_{*i}^u)$ for all $i \in [m]$.

We first consider line (1). For x chosen uniformly from $\{0,1\}^{ms}$, by assumption on ψ, for any $u \in \{0,1\}^{k-1}$ the random variable z^u is uniform in $\{0,1\}^m$. Therefore, in particular, $\mathbf{E}_x[\mu(z^{0\ldots0})] = \mathbf{E}_{z \in \{0,1\}^m}[\mu(z)]$. Further, since μ is a distribution, $\mathbf{E}_{z \in \{0,1\}^m}[\mu(z)] = 2^{-m}$.

We now bound the remaining terms. First we have

$$\max_{x_1,\ldots,x_r} \prod_{u \neq 0\ldots0} \mathbf{E}_{x_{r+1}\ldots x_m}\big[\mu(z^u)\big] \leq \prod_{u \neq 0\ldots0} \max_{x_1,\ldots,x_r} \mathbf{E}_{x_{r+1}\ldots x_m}\big[\mu(z^u)\big].$$

Fixing $x_1, \ldots, x_r$ fixes the values of $z_1^u, \ldots, z_r^u$ and by our assumption on ψ, for random $x_{r+1}, \ldots, x_m$ the values of of $z_{r+1}^u, \ldots, z_m^u$ are uniformly random. Therefore the value in line (2) is upper bounded by

$$\prod_{u \neq 0\ldots0} \max_{z_1^u,\ldots,z_r^u} \mathbf{E}_{z_{r+1}^u\ldots z_m^u}\big[\mu(z^u)\big] = \big(\max_{z_1,\ldots,z_r} \mathbf{E}_{z_{r+1}\ldots z_m}\big[\mu(z)\big]\big)^{2^{k-1}-1}.$$

Since $r \geq d$, by the property of μ implied by Lemma 4.1,

$$\max_{z_1,\ldots,z_r} \sum_{z_{r+1},\ldots,z_m} \mu(z) \leq 2^{r^\alpha - r}/\epsilon$$

and therefore line (2) is upper bounded by $(2^{r^\alpha - r}/(\epsilon 2^{m-r}))^{2^{k-1}-1} = (2^{r^\alpha - m}/\epsilon)^{2^{k-1}-1}$. The lemma follows immediately by combining the bounds for lines (1) and (2). □

Putting Proposition 3.1 and Lemma 4.2 together with the general correlation method yields our main technical theorem.

Theorem 4.3. *If $f : \{0,1\}^m \to \{1,-1\}$ has $(< 1-\epsilon, \alpha)$-approximate degree at least d, ψ is a selector function on $\{0,1\}^s$ with domain $D_\psi = D_{\psi,1} \times \cdots \times D_{\psi,(k-1)}$ then*

$$R^k_{1/2-\epsilon}(f \circ \psi^m) \geq \log_2(\epsilon(1-\epsilon)) - \frac{1}{2^{k-1}} \log_2 \Big(\sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot \Pr_{y^0,y^1 \in D_\psi^{(m)}} [r_\psi(y^0,y^1) = r] \Big).$$

Proof. We recap the argument from Section 3 with a more general choice of parameters. We first apply Lemma 4.1 to f to produce function g and distribution μ. By construction $\mathrm{Cor}_\mu(f,g) \geq 1-\epsilon$. Then we define the distribution λ based on μ and ψ by $\lambda(x,y) = \dfrac{\mu(z_1,\ldots,z_m)}{2^{n-m}|D_\psi|^m}$ where $z_i = \psi(x_i, y_{*i})$ for $y \in D_\psi^{(m)}$ and 0 otherwise. To prove a lower bound c on $R^k_{1/2-\epsilon}(f \circ \psi^m)$ we show that $\mathrm{Cor}_\lambda(f \circ \psi^m, \Pi^c_k) \leq 2\epsilon$. As before we have $\mathrm{Cor}_\lambda(f \circ \psi^m, g \circ \psi^m) = \mathrm{Cor}_\mu(f,g) \geq 1-\epsilon$, hence $\mathrm{Cor}_\lambda(f \circ \psi^m, \Pi^c_k) \leq \epsilon + \mathrm{Cor}_\lambda(g \circ \psi^m, \Pi^c_k)$ by the triangle inequality. It therefore suffices to show that $\mathrm{Cor}_\lambda(g \circ \psi^m, \Pi^c_k) \leq \epsilon$. Putting together the bounds of Section 3 with Lemma 4.2 we have

$$\begin{aligned} \mathrm{Cor}_\lambda(g \circ \psi^m, \Pi^c_k)^{2^{k-1}} &\leq 2^{(c+m)\cdot 2^{k-1}} \cdot \sum_{r=d}^{m} \frac{2^{(2^{k-1}-1)r^\alpha}}{2^{2^{k-1}m}(1-\epsilon)^{2^{k-1}-1}} \cdot \Pr_{y^0,y^1 \in D_\psi^{(m)}} [r_\psi(y^0,y^1) = r] \\ &< (\frac{2^c}{1-\epsilon})^{2^{k-1}} \cdot \sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot \Pr_{y^0,y^1 \in D_\psi^{(m)}} [r_\psi(y^0,y^1) = r] \end{aligned}$$

Taking 2^{k-1}-st roots we have

$$\mathrm{Cor}_\lambda(g \circ \psi^m, \Pi^c_k) < \frac{2^c}{1-\epsilon} \cdot \Big(\sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot \Pr_{y^0,y^1 \in D_\psi^{(m)}} [r_\psi(y^0,y^1) = r] \Big)^{1/2^{k-1}}.$$

Therefore we obtain that $R^k_{1/2-\epsilon}(f \circ \psi^m) \geq c$ if

$$\epsilon \geq \frac{2^c}{1-\epsilon} \cdot \Big(\sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot \Pr_{y^0,y^1 \in D_\psi^{(m)}} [r_\psi(y^0,y^1) = r] \Big)^{1/2^{k-1}}.$$

Rewriting and taking logarithms yields the claimed bound. □

We now show an example of how to apply this techical theorem to derive lower bounds for functions based on the pattern tensor selector function. Although we will find that other selector functions allow us to produce larger lower bounds, this highlights the consequences of (ϵ, α)-approximate degree alone and will let us obtain results for simpler functions than with the other selector functions. We will apply this bound to specific functions in Appendix A.

Theorem 4.4. *For $0 \leq \alpha < 1$ and any Boolean function f on m bits with $(5/6, \alpha)$-approximate degree d, the function $f \circ \psi^m_{k,\ell}$ defined on nk bits, where $n = ms$ for $s \geq \lceil \frac{4e(k-1)m}{d} \rceil^{k-1}$, requires $R^k_{1/3}(f \circ \psi^m_{k,\ell}) > d/2^k - 3$ for $k \leq (1-\alpha)\log_2 d$.*

Proof. By Proposition 3.3, $\Pr_{y^0,y^1 \in D^{(m)}_{\psi_{k,\ell}}} [r_{\psi_{k,\ell}}(y^0,y^1) = r] \leq (\frac{e(k-1)m}{r\ell})^r$ so

$$\sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot \Pr_{y^0,y^1 \in D^{(m)}_{\psi_{k,\ell}}} [r_{\psi_{k,\ell}}(y^0,y^1) = r] \leq \sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot (\frac{e(k-1)m}{r\ell})^r \tag{3}$$

Since $k \leq (1-\alpha)\log_2 d$, we have $(2^{k-1}-1)r^\alpha < d^{1-\alpha} r^\alpha \leq r$ for $r \geq d$ so (3) is

$$
\begin{aligned}
&\leq \sum_{r=d}^{m} (\frac{2e(k-1)m}{r\ell})^r \\
&\leq \sum_{r=d}^{m} 2^{-r} < 2^{-(d-1)} \qquad \text{for } \ell \geq \tfrac{4e(k-1)m}{d}.
\end{aligned}
$$

Plugging this in to Theorem 4.3 we obtain that

$$R^k_{1/3}(f \circ \psi^m) \geq \log_2(5/36) - \frac{1}{2^{k-1}} \log_2 2^{-(d-1)} > d/2^k - 3$$

as required. □

5 Proof of Lemma 4.1

Proof. As in the proof for Lemma 2.5, we write the requirements down as a linear program and study its dual. The lemma is implied by proving that the following linear program $\mathcal{P}$ has optimal value 1:

Minimize η subject to

$$
\begin{array}{rrrr}
y_S: & \displaystyle\sum_{x\in\{0,1\}^m} h(x)\chi_S(x) = 0 & |S| < d & (4) \\
\beta: & \displaystyle\sum_{x\in\{0,1\}^m} h(x)f(x) \geq \epsilon & & (5) \\
v_x: & \mu(x) - h(x) \geq 0 & x \in \{0,1\}^m & (6) \\
w_x: & \mu(x) + h(x) \geq 0 & x \in \{0,1\}^m & (7) \\
a_\rho: & \eta - 2^{|\rho|-|\rho|^\alpha} \displaystyle\sum_{x\in C_\rho} \mu(x) \geq 0 & \rho \in \{0,1,*\}^m, |\rho| \geq d & (8) \\
\gamma: & \displaystyle\sum_{x\in\{0,1\}^m} \mu(x) = 1 & & (9)
\end{array}
$$

Suppose that we have optimum $\eta = 1$. In this LP formulation, inequality γ ensures that the function μ is a probability distribution, and inequalities v_x and w_x ensure that $\mu(x) \geq |h(x)|$ so $||h||_1 \leq 1$. If $||h||_1 = 1$, then we must have $\mu(x) = |h(x)|$ and we can write $h(x) = \mu(x)g(x)$ as in the proof of Lemma 2.5 and then the inequalities y_S will ensure that $\mathrm{Cor}_\mu(g, \chi_S) = 0$ for $|S| < d$ and inequality β will ensure that $\mathrm{Cor}_\mu(f, g) \geq \epsilon$ as required. Finally, each inequality a_ρ ensures that $\mu(C_\rho) \leq 2^{-|\rho|+|\rho|^\alpha} = 2^{-|\rho|+|\rho|^\alpha}$ which is actually a little stronger than our claim.

The only issue is that an optimal solution might have $||h||_1 < 1$. However in this case inequality β ensures that $||h||_1 \geq \epsilon$. Therefore, for any solution of the above LP with function h, we can define another function $h'(x) = h(x)/||h||_1$ with $||h'||_1 = 1$ and a new probability distribution μ' by $\mu'(x) = |h'(x)| \leq \mu(x)/||h||_1 \leq \mu(x)/\epsilon$. This new h' and μ' still satisfy all the inequalities as before except possibly inequality a_ρ but in this case if we increase η by a $1/||h||_1$ factor it will also be satisfied. Therefore, the $\mu'(C_\rho) \leq 2^{-|\rho|+|\rho|^\alpha}/\epsilon$.

Here is the dual LP:

Maximize $\beta \cdot \epsilon + \gamma$ subject to

$$\eta: \quad \sum_{\rho \in \{0,1,*\}^m, |\rho| \geq d} a_\rho = 1 \tag{10}$$

$$\mu(x): \quad v_x + w_x + \gamma - \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho = 0 \qquad x \in \{0,1\}^m \tag{11}$$

$$h(x): \quad \beta f(x) + \sum_{|S| < d} y_S \chi_S(x) + w_x - v_x = 0 \qquad x \in \{0,1\}^m \tag{12}$$

$$\beta, v_x, w_x, a_\rho \geq 0 \qquad x \in \{0,1\}^m \tag{13}$$

Since y_S are arbitrary we can replace $\sum_{|S|<d} y_S \chi_S(x)$ by $p_d(x)$ where p_d is an arbitrary polynomial of degree $< d$ to obtain the modified dual:

Maximize $\beta \cdot \epsilon + \gamma$ subject to

$$\eta: \quad \sum_{\rho \in \{0,1,*\}^m, |\rho| \geq d} a_\rho = 1 \tag{14}$$

$$\mu(x): \quad v_x + w_x + \gamma - \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho = 0 \qquad x \in \{0,1\}^m \tag{15}$$

$$h(x): \quad \beta f(x) + p_d(x) + w_x - v_x = 0 \qquad x \in \{0,1\}^m \tag{16}$$

$$\beta, v_x, w_x, a_\rho \geq 0 \qquad x \in \{0,1\}^m \tag{17}$$

Equations (15) and (16) for $x \in \{0,1\}^m$ together are equivalent to:

$$2w_x + \beta f(x) + p_d(x) + \gamma - \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho = 0$$

and

$$2v_x - \beta f(x) - p_d(x) + \gamma - \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho = 0.$$

Since these are the only constraints on v_x and w_x respectively other than non-negativity these can be satisfied by any solution to

$$\beta f(x) + p_d(x) + \gamma \leq \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho$$

and

$$-\beta f(x) - p_d(x) + \gamma \leq \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho,$$

which together are equivalent to

$$|\beta f(x) + p_d(x)| + \gamma \leq \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho.$$

Since $p_d(x)$ is an arbitrary polynomial function of degree less than d we can write $p_d = -\beta q_d$ where q_d is another arbitrary polynomial function of degree less than d and we can replace the terms $|\beta f(x) + p_d(x)|$ by $\beta|f(x) - q_d(x)|$.

Therefore the dual program $\mathcal{D}$ is equivalent to maximizing $\beta \cdot \epsilon + \gamma$ subject to

$$\beta|f(x) - q_d(x)| + \gamma \leq \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho$$

for all $x \in \{0,1\}^m$, a_ρ is probability distribution on the set of all restrictions of size at least d, and q_d is a real-valued function of degree $< d$.

Now, let B be the set of points at which $|f(x) - q_d(x)| \geq \epsilon$. For any $x \in B$, the value of the objective function of $\mathcal{D}$, which is $\beta \cdot \epsilon + \gamma$, is not more than

$$\beta|f(x) - q_d(x)| + \gamma \leq \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho. \tag{18}$$

Let $R(x)$ denote the right-hand side of inequality (18). It suffices to prove that $R(x) \leq 1$ for some $x \in B$. This is, in turn, equivalent to proving that

$$\min_{x \in B} R(x) \leq 1,$$

for any distribution a_ρ. Suppose, by contradiction, that there exists a distribution a_ρ such that $R(x) > 1$ for any $x \in B$. Let Π, the set of restrictions, and ν, a distribution on Π, be the witnesses for the (ϵ, α)-approximate degree of f. Picking $\pi \in \Pi$ randomly according to ν, we define the random variable

$$I_\pi := \sum_{\rho: |\rho| \geq d,\ C_\rho \cap C_\pi \neq \emptyset} 2^{|\rho| - |\rho|^\alpha} a_\rho.$$

Then,

$$\mathbf{E}_{\pi \sim \nu}(I_\pi) = \sum_{\rho: |\rho| \geq d} \Pr[C_\rho \cap C_\pi \neq \emptyset] \cdot 2^{|\rho| - |\rho|^\alpha} a_\rho \leq \sum_{\rho: |\rho| \geq d} 2^{|\rho|^\alpha - |\rho|} \cdot 2^{|\rho| - |\rho|^\alpha} a_\rho \leq 1.$$

Therefore there exists $\pi \in \Pi$ for which $I_\pi \leq 1$. If there exists $x \in B$ such that $x \in C_\pi$, then since

$$R(x) = \sum_{C_\rho \ni x, |\rho| \geq d} 2^{|\rho| - |\rho|^\alpha} a_\rho > 1,$$

we would have $I_\pi > 1$. Thus $C_\pi \cap B = \emptyset$. So for any $x \in C_\pi$, we have $|f(x) - q_d(x)| < \epsilon$. But since the degree of q_d is less than d this contradicts the fact that $deg_\epsilon(f|_\pi) \geq d$. (Note that we actually only need the weaker assumption that that $deg_{\epsilon'}(f|_\pi) \geq d$ for all $\epsilon' < \epsilon$.) Thus the lemma follows. □

We note that the bounds in Lemma 4.1 are essentially the best possible for AC^0 functions: By results of Linial, Mansour, and Nisan [19], for any AC^0 function f and constant $0 < \lambda < 1$, there is a function p_d of degree $d < m^\lambda$, such that $||f - p_d||_2^2 \leq 2^{m - m^\delta}$ for some constant $\delta > 0$. Let B_m be the set of x such that $|f(x) - p_d(x)| \geq \epsilon$. Then $|B_m|\epsilon^2 \leq \sum_{x \in B_m} |f(x) - p_d(x)|^2 \leq ||f - p_d(x)||_2^2 \leq 2^{m - m^\delta}$ so $|B_m| \leq 2^{m - m^\delta}/\epsilon^2$. If we tried to replace the upper bound on $\mu(C_\rho)$ by some $c(|\rho|)$ where $c(m)$ is $\omega(1/|B_m|)$ then we could choose $a_x = 1/|B_m|$ for $x \in B_m$ and $a_\rho = 0$ for all other ρ and for these values β would be unbounded.

6 AC^0 functions with large (ϵ, α)-approximate degree

It is not obvious that any function, let alone a function in AC^0, has large (ϵ, α)-approximate degree for $\alpha < 1$. In this section we will show that for $\alpha < 1$, AC^0 contains functions with large $(5/6, \alpha)$-approximate degree and functions with large α-threshold degree (which is $(< 1, \alpha)$-approximate degree).

Before going into the details, we describe the general framework of our constructions. Parameterized by three integers $q > r > p > 0$, where r can be thought of as polynomial in q and p as logarithmic in q, the construction will produce an AC^0 circuit computing a function f on $n = pq$ bits that has (ϵ, α)-approximate degree polynomial in r, which will therefore also be polynomial in n.

First, we find an AC^0 circuit G on q bits that is a $\mathsf{\Sigma_{2i-1}}$ or $\mathsf{\Pi_{2i}}$ circuit for $i \geq 1$ (and hence has bottom level gates that are $\vee$ gates), such that for any set S of r input bits, the projection of G on S, denoted by G_S, computes a function of high ϵ-approximate degree (i.e. degree that is polynomial in r). Here the "projection" of a circuit on a subset S of input bits is defined as a new circuit obtained from the original by keeping only those nodes on some path from an input bit in S to the output gate. For $\epsilon = 5/6$ (or indeed any constant ϵ sufficiently less than 1) finding such a circuit is easy: G can be $\textsc{Or}_q$.

Next, we obtain another circuit $H := G \circ \textsc{And}_p^q$ by replacing every input bit of G by an $\wedge$-gate on p bits. Thus H has $n = pq$ input bits in blocks of p bits, each block corresponding to the inputs to a single $\wedge$-gate. We define a family of restrictions on these bits as follows. Each restriction first chooses r out of the q blocks of inputs. It will leave the rp input bits in these blocks unset, and for each of the remaining $q - r$ blocks, it will assign values to the p bits in this block uniformly at random from the set $\{0, 1\}^p - \{0^p, 1^p\}$. We will argue that this family of restrictions is spread out evenly, in the sense of the definition of (ϵ, α)-approximate degree.

Since G has large ϵ-approximate degree even when projected on any large enough subset of inputs, and each restriction in this family leaves a subset of r blocks unset, one might hope that for any restriction π in this family, $H|_\pi$ would have some projection of G on r input bits as a subfunction. In the simple case that G is $\textsc{Or}_q$ this indeed enough. However, in the case that G is more complex it is easy to see that this is not the case since the values assigned by π might force the values of the bottom level $\vee$ to 0 and hence its parent $\wedge$-gate to 0 which would produce a much simpler restriction of G rather than a projection of G on r inputs bits.

To produce the desired property we add some extra "checking" sub-circuits to H that will allow us to circumvent this danger. Fortunately for our applications, these sub-circuits are very small AC^0 circuits and the resulting circuit H', which has the same input bits as H, is thus also in AC^0.

First of all, we verify that the above family of restriction is indeed spread out evenly and then apply it for $G = \textsc{Or}_q$. Afterwards we describe the details of the construction of H' from H when G is more complex.

Lemma 6.1. *Let q, r, p, and d be positive integers with $q > r > p \geq 2$ and let $1 > \alpha > \beta > 0$ be such that $q^\beta \geq rp$, $2^{p-1} - 1 \geq q^{1-\beta}$, $q^\alpha \geq \frac{6}{\ln 2} 2^p r$, and $d^{\alpha-\beta} \geq 3p/\ln 2$. Fix any partition of a set of $m = pq$ input bits into q blocks of p bits each. Define distribution ν on R_{pq}^{pr} as follows: pick uniformly at random a subset of $q - r$ of the q blocks; then for each of these blocks, assign values to the variables in the block uniformly at random from $\{0, 1\}^p - \{0^p, 1^p\}$. Then for any $\rho \in \{0, 1, *\}^m$*

with $|\rho| \geq d$, the following holds[1]

$$\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] \leq 2^{|\rho|^\alpha - |\rho|}.$$

Proof. Fix any restriction ρ of size $i = |\rho| \geq d$. We have

$$\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] = \frac{1}{\binom{q}{q-r}} \sum_{S\subset[q],|S|=q-r} \prod_{j\in S} p_j,$$

where p_j is the probability that π and ρ agree on the variables in the j-th block. Write $i = i_1+\ldots+i_q$, where i_j is the number of assignments ρ makes to variables in the j-th block. Then

$$p_j \leq \frac{2^{p-i_j}}{2^p-2} = 2^{-i_j}(1+\frac{1}{2^{p-1}-1}).$$

Let $i_S = \sum_{j\in S} i_j$ be the number of assignments ρ makes to variables in blocks in S and $k_S = |\{j \in S : i_j > 0\}|$ be the number of blocks in S in which ρ assigns least one value. Hence,

$$\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] \quad < \quad \frac{1}{\binom{q}{q-r}} \sum_{S\subset[q],|S|=q-r} 2^{-i_S}(1+\frac{1}{2^{p-1}-1})^{k_S}. \tag{19}$$

Let $k = |\{j : i_j > 0\}|$ be the total number of blocks in which ρ assigns at least one value. There are 2 cases: (I) $k \geq q/2$, and (II) $k < q/2$.

Now consider case (I). Thus $i \geq q/2$. In Equation 19, we have $k_S \leq q$ for every S. Thus,

$$\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] \leq \frac{1}{\binom{q}{q-r}} \sum_{S\subset[q],|S|=q-r} 2^{-i_S}(1+\frac{1}{2^{p-1}-1})^q.$$

It is easy to see that $i_S \geq i - pr$ for every such S. Hence we get

$$\frac{1}{\binom{q}{q-r}} \sum_{S\subset[q],|S|=q-r} 2^{-i_S} \leq 2^{pr-i} \leq 2^{(2i)^\beta - i},$$

since $pr \leq q^\beta \leq (2i)^\beta$ in this case. Thus,

$$\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] \leq 2^{(2i)^\beta - i}(1+\frac{1}{2^{p-1}-1})^q \leq 2^{(2i)^\beta - i}e^{q^\beta} \leq 2^{2^\beta(1+1/\ln 2)i^\beta - i},$$

since $q^{1-\beta} \leq 2^{p-1}-1$ and $i \geq q/2$. We upper bound the term $2^\beta(1+1/\ln 2)\ i^\beta$ by i^α as follows: Since $i \geq d$,

$$i^{\alpha-\beta} \geq d^{\alpha-\beta} \geq 3p/\ln 2 \tag{20}$$

[1] Note that one can interpret this condition intuitively as requiring that, in a multiplicative amortized sense, each bit assigned by ρ contributes not much more than 1/2 to the probability of being consistent with a random π. For $\pi \sim \nu$, ρ could assign p bits arbitrarily in one of the r terms not selected by π and would be consistent with π. However, in our amortized sense these p bits should not contribute much more than a 2^{-p} factor to the probability of being consistent with π. On the other hand the probability that a given term is not selected by π is r/q. It is therefore necessary in our argument that 2^{-p} not be much smaller than r/q; this is the main motivation for the relationship between p and q that we use.

by our assumption in the statement of the lemma. Since $p \geq 2$, we have $i^{\alpha-\beta} > 6 > 2^{\beta}(1+1/\ln 2)$ which is all that we need to derive that $\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] < 2^{i^{\alpha}-i}$ in case I.

Next, we consider case (II). We must have $k \leq p^{1-\beta}(2^{p-1}-1)\, i^{\beta}$, because otherwise

$$i \geq k > p^{1-\beta}(2^{p-1}-1)i^{\beta} \geq p^{1-\beta}q^{1-\beta}i^{\beta},$$

which implies $i^{1-\beta} > (pq)^{1-\beta}$ and hence $i > pq = m$ which is impossible. Therefore

$$(1+\frac{1}{2^{p-1}-1})^{k_S} \leq e^{\frac{k_S}{2^{p-1}-1}} \leq e^{\frac{k}{2^{p-1}-1}} \leq e^{p^{1-\beta}i^{\beta}}.$$

So,

$$\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] < e^{p^{1-\beta}i^{\beta}}\mathcal{S} \quad \text{where} \quad \mathcal{S} = \frac{1}{\binom{q}{q-r}} \sum_{S\subset[q],|S|=q-r} 2^{-i_S} = E_{S\sim U}[2^{-i_S}].$$

and U is the uniform distribution on subsets of $[q]$ of size $q-r$.

Now we continue by upper bounding $\mathcal{S}$. For the moment let us assume that i is divisible by p. If we view the blocks as the bins, and the assigned positions by ρ as balls placed in corresponding bins, then we observe that $\mathcal{S}$ can only increase if we move one ball from a bin A of $x > 0$ balls to another bin B of $y \geq x$ balls. This is because only those i_S with S containing exactly one of these two bins are affected by this move. Then, we can write the contribution of these S's to $\mathcal{S}$ before the move as

$$\mathcal{S}' = \sum_{S\subset[q],\ |S|=q-r,\ S\cap\{A,B\}=1} 2^{-i_S} = \sum_{S'\subset[q]-\{A,B\},\ |S'|=q-r-1} 2^{-i_{S'}}(2^{-x}+2^{-y}),$$

and after the move as

$$\mathcal{S}'' = \sum_{S'\subset[q]-\{A,B\},\ |S'|=q-r-1} 2^{-i_{S'}}(2^{-x+1}+2^{-y-1}).$$

Since $y \geq x$, $\mathcal{S}'' > \mathcal{S}'$.

Hence w.l.o.g. and with the assumption that p divides i, we can assume that the balls are distributed such that every bin is either full (containing p balls) or empty. Hence $k = i/p$ and for any $1 \leq j \leq q$, either $i_j = 0$ or $i_j = p$.

Claim 6.2. *If i is divisible by p then $\mathcal{S} \leq 2^{-i}\ e^{2^{p+1}rk/q}$.*

We first see how the claim suffices to prove the lemma. If i is not divisible by p then we note that $\mathcal{S}$ is a decreasing function of i and apply the claim for the first $i' = p\lfloor i/p \rfloor > i-p$ positions set by ρ to obtain an upper bound of $\mathcal{S} < 2^{p-i}e^{2^{p+1}ri/(pq)}$ that applies for all choices of i. The overall bound we obtain in this case is then

$$\begin{aligned}\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] &< e^{p^{1-\beta}i^{\beta}}2^{p}e^{2^{p+1}ri/(pq)}2^{-i} \\ &= 2^{i^{\beta}p^{1-\beta}/\ln 2+p+2^{p+1}ri/(pq\ln 2)}2^{-i}.\end{aligned}$$

We now consider the exponent $i^\beta p^{1-\beta}/\ln 2 + p + 2^{p+1}ri/(pq\ln 2)$ and show that it is at most i^α. For the first term observe that by (20), $i^{\alpha-\beta} \geq 3p/\ln 2$ so $i^\beta p^{1-\beta}/\ln 2 \leq i^\alpha/3$. For the second term again by (20) we have $p \leq i^{\alpha-\beta}/3 \leq i^\alpha/3$. For the last term, since $q^\alpha \geq \frac{6}{\ln 2}2^p r$, we have

$$\frac{2^{p+1}ri}{pq\ln 2} \leq \frac{q^\alpha i}{3pq} \leq i(pq)^{\alpha-1}/3 \leq i^\alpha/3,$$

since $i \leq pq$. Therefore in case II we have $\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] < 2^{i^\alpha - i}$ as required. It only remains to prove the claim.

Proof of Claim: Let $T = \{t \mid i_t = p\}$ be the subset of k blocks assigned by ρ. Therefore $i_S = |S \cap T|p$ where S is a random set of size $q-r$ and T is a fixed set of size k and both are in $[q]$. We have two subcases: (IIa) when $k \leq r$ and (IIb) when $q/2 \geq k > r$.

If $k \leq r$ then we analyze $\mathcal{S}$ based on the number j of elements of S contained in T. There are $\binom{k}{j}$ choices of elements of T to choose from and $q-r-j$ elements to select from the $q-k$ elements of $\overline{T}$. Therefore

$$\mathcal{S} = \frac{\sum_{j=0}^k \binom{r}{j}\binom{q-k}{q-r-j}2^{-jp}}{\binom{q}{q-r}}.$$

Now since

$$\frac{\binom{q-k}{q-r-j}}{\binom{q}{q-r}} = \frac{(q-k)!(q-r)!r!}{q!(q-r-j)!(r-(k-j))!} < \frac{(q-r)^j r^{k-j}}{(q-k)^k} = \Big(\frac{r}{q-k}\Big)^k\Big(\frac{q-r}{r}\Big)^j,$$

we can upper bound $\mathcal{S}$ by

$$\begin{aligned}
\Big(\frac{r}{q-k}\Big)^k \sum_{j=0}^k \binom{k}{j} 2^{-pj}\Big(\frac{q-r}{r}\Big)^j &= \Big(\frac{r}{q-k}\Big)^k\Big(1+\frac{q-r}{2^p r}\Big)^k \\
&= 2^{-pk}\ \Big(\frac{r}{q-k}\Big)^k\Big(\frac{2^p r + (q-r)}{r}\Big)^k \\
&= 2^{-i}\ \Big(\frac{q+(2^p-1)r}{q-k}\Big)^k \\
&= 2^{-i}\ \Big(1+\frac{(2^p-1)r+k}{q-k}\Big)^k \\
&\leq 2^{-i}\ \Big(1+\frac{2^p r}{q-k}\Big)^k \\
&\leq 2^{-i}\ e^{2^p rk/(q-k)} \\
&\leq 2^{-i}\ e^{2^{p+1}rk/q}.
\end{aligned}$$

since $k \leq q/2$.

In the case that $r \leq k \leq q/2$ we observe that by symmetry we can equivalently view the expectation $\mathcal{S}$ as the result of an experiment in which the set S of size $q-r$ is chosen first and the set T of size k is chosen uniformly at random. We analyze this case based on the number j of elements of $\overline{S}$ contained in T. There are $\binom{r}{j}$ choices of elements of $\overline{S}$ to choose from and $k-j$ elements to select from the $q-r \geq q/2 \geq k$ elements of S. Therefore

$$\mathcal{S} = \frac{\sum_{j=0}^r \binom{r}{j}\binom{q-r}{k-j}2^{-(k-j)p}}{\binom{q}{k}}.$$

Using the fact that

$$\frac{\binom{q-r}{k-j}}{\binom{q}{k}} = \frac{(q-r)!(q-k)!k!}{q!(k-j)!(q-r-k+j)!} < \frac{(q-k)^{r-j}k^j}{(q-r)^r} = \Big(\frac{q-k}{q-r}\Big)^r\Big(\frac{k}{q-k}\Big)^j,$$

we upper bound $\mathcal{S}$ by

$$\begin{aligned}
2^{-pk}\Big(\frac{q-k}{q-r}\Big)^r\sum_{j=0}^{r}\binom{r}{j}\Big(\frac{2^pk}{q-k}\Big)^j &= 2^{-pk}\Big(\frac{q-k}{q-r}\Big)^r\Big(1+\frac{2^pk}{(q-k)}\Big)^r \\
&= 2^{-i}\Big(\frac{q-k}{q-r}\Big)^r\Big(\frac{q+(2^p-1)k}{q-k}\Big)^r \\
&= 2^{-i}\Big(\frac{q+(2^p-1)k}{q-r}\Big)^r \\
&= 2^{-i}\Big(1+\frac{(2^p-1)k+r}{q-r}\Big)^r \\
&\le 2^{-i}\Big(1+\frac{2^pk}{q-r}\Big)^r \\
&\le 2^{-i}e^{2^prk/(q-r)} \\
&\le 2^{-i}e^{2^{p+1}rk/q}
\end{aligned}$$

since $r \le q/2$. □

Recall that the function $\textsc{Tribes}_{p,q}$ on $m = pq$ bits is defined by

$$\textsc{Tribes}_{p,q}(x) = \vee_{i=1}^{q} \wedge_{j=1}^{p} x_{i,j}.$$

Usually the function $\textsc{Tribes}$ is defined so that 2^p is linear or nearly-linear in q. Using Lemma 6.1, we will show that, with a different relationship in which $q \gg 2^p$ but p is still $\Theta(\log q)$, the $(5/6, \alpha)$-approximate degree of $\textsc{Tribes}_{p,q}$ is large.

Corollary 6.3. *Given any $1 > \epsilon > 0$, let q, p be positive integers with $q > p \ge 2$ such that $2\lceil q^{1-\beta}\rceil < 2^p \le \frac{1}{6}q^{\alpha+\epsilon-1}\ln 2$ for some fixed constants $1 > \alpha > \beta > 1-\epsilon$. Then for large enough q, $\textsc{Tribes}_{p,q}$ has $(5/6,\alpha)$-approximate degree at least $\sqrt{q^{1-\epsilon}/12}$.*

Proof. Define the distribution ν as in the statement of Lemma 6.1, where a p-block corresponds to a p-term in $\textsc{Tribes}_{p,q}$. Applying Lemma 6.1 with $r := \lceil q^{1-\epsilon}\rceil$ and $d = \lceil\sqrt{r/12}\rceil$, it is clear that for any π with $\nu(\pi) > 0$, $\textsc{Or}_r$ is a subfunction of $\textsc{Tribes}_{p,q}|_\pi$ so $deg_{5/6}(\textsc{Tribes}_{p,q}|_\pi) \ge deg_{5/6}(\textsc{Or}_r) \ge \sqrt{r/12}$.

All conditions in the statement of the lemma would then be satisfied for q large enough. In particular, for q large enough,

$$q^\beta/r \ge q^{\beta+\epsilon-1} > \log q > p,$$

and

$$d^{\alpha-\beta} \ge (r/12)^{(\alpha-\beta)/2} \ge q^{(1-\epsilon)(\alpha-\beta)/2}/12 > 3\log q/\ln 2 > 3p/\ln 2.$$

□

Corollary 6.4. *For $p \geq 2$ a sufficiently large integer and $q = 2^{4p}$, $\text{TRIBES}_{p,q}$ has $(5/6, 0.9)$-approximate degree at least $q^{3/10}/\sqrt{12} = 2^{6p/5}/\sqrt{12}$.*

Proof. From the last corollary with $\epsilon = 0.4$, $\alpha = 0.9$, and $\beta = 0.8$ if $\lceil q^{0.2} \rceil < 2^p \leq \frac{1}{6} q^{0.3} \ln 2$ then we obtain a $(5/6, 0.9)$-approximate degree lower bound of $\sqrt{q^{0.6}/12}$ for q sufficiently large. For p sufficiently large the required conditions are satisfied with $q = 2^{4p}$. □

Next, we will construct an AC^0 function with large α-threshold degree. We first need a function that has large threshold degree even when projected on any sufficiently large subset of inputs.

Lemma 6.5. *Let d and s be positive integers such that $s \geq 8d^2$. Let t be an integer such that $s/(4d) \geq t \geq 2d$. Let $q = st$ and r be any integer such that $q \geq r \geq 2ds$. Then there is an explicit read-once $\mathsf{AND} \circ \mathsf{OR}$ formula G on q bits such that for any set S of r input bits, the function computed by G_S has threshold degree at least d.*

Proof. Let G be the $\mathsf{AND} \circ \mathsf{OR}$ formula with fan-in t at the top $\wedge$ gate and fan-in of s at each of the $\vee$ gates. Let S be any subset of input bits with $|S| = r$.

Let A be the set of $\vee$ gates in G that contain at least $4d^2$ elements of S. By Markov's inequality, $r \leq s|A| + 4d^2(t - |A|)$, and hence

$$|A| \geq \frac{r - 4d^2 t}{s - 4d^2} > \frac{r - 4d^2 t}{s} \geq d$$

since $r \geq 2ds$ and $4d^2 t \leq ds$. Hence G_S contains at least d $\vee$-gates, each having at least $4d^2$ inputs. This implies that G_S computes $\text{MP}_{d,4d^2}$ as a subfunction. By Proposition 2.4, $thr(G_S) \geq thr(\text{MP}_{d,4d^2}) \geq d$. □

We now describe how to produce the circuit H' defined earlier. We could define this for any AC^0 circuit G with a bottom level of $\vee$ gates and the next level having $\wedge$ gates but we restrict ourselves to the $\mathsf{AND} \circ \mathsf{OR}$ formula given by Lemma 6.5. Let $H = G \circ \text{AND}_p^q$ be the circuit obtained from G by replacing each of its input bits by an AND gate over p bits, for some $p > 0$. In particular for the choice of G from Lemma 6.5, H is a read-once $\text{AND} \circ \text{OR} \circ \text{AND}_p$ circuit on pq bits. We then obtain another circuit H' by applying the following operation to each bottom OR gate φ of H: let t be the number of AND_p gates that are inputs to φ; for every $i \in [t]$ denote the inputs to the i-th AND_p gate that feeds into φ by $z_{i,1}, \ldots, z_{i,p}$; for each such i, create two new OR gates $B_i = \vee_{j=1}^{p} z_{i,j}$ and $B_i' = \vee_{j=1}^{p} (\neg z_{i,j})$; then, create a new AND gate $A_\varphi = \bigwedge_{i=1}^{t} (B_i \wedge B_i')$; finally, add a new edge feeding the output of A_φ to φ.

The following lemma justifies the above construction.

Lemma 6.6. *Let G be any $\mathsf{AND} \circ \mathsf{OR}$ circuit on q bits. For some integer $p > 0$, let H' be the circuit constructed from G by following the process described above. Then the following hold:*

- *H' is a Π_4 circuit of size at most 4 times the size of G;*
- *Let π be any restriction that chooses a subset S of the blocks of inputs to H' to leave unset and assigns values from $\{\{0,1\}^p - \{0^p, 1^p\}$ to each other block. Then $H'|_\pi$ computes a function that contains the function computed by G_S as a subfunction.*

Proof. A sub-circuit of depth 2 with 3 gates is added for each bottom level $\textsc{Or}$ in G so the first part is immediate.

For the second part, let S be the set of blocks in G that are left unset by π. First note that for any block not in S, the associated $\textsc{And}_p$ gate in H' is forced to 0. Also observe that for any $\textsc{Or}$ gate φ' in H' corresponding to a bottom level $\textsc{Or}$ gate φ in G,

if φ has an input that corresponds to a block in S then setting the values of any such block in S to 0^p or 1^p will force A_φ to output 0,

if φ does not have any inputs corresponding to any block of S then A_φ outputs 1 and hence φ' outputs 1.

It follows that we can use $H'|_\pi$ to compute G_S by assigning 0^p in place of 0 and 1^p in place of 1 for each block in S. □

Finally we show that, with suitable parameters, H' has high α-threshold degree.

Theorem 6.7. *Let $p > 0$ be a sufficiently large integer multiple of 15 and let $q = 2^{4p}$. Then there is an explicit depth 4 AC^0 function on pq bits that has 0.9-threshold degree at least $q^{1/15}$.*

Proof. Let $d = q^{1/15}$, $s = 2q^{8/15}$, $t = q^{7/15}/2$, and $r = 4q^{3/5}$. Observe that, by our choice of p and q, all of these are integral and they satisfy the conditions of Lemma 6.5. We can apply that lemma to derive an $\mathsf{AND} \circ \mathsf{OR}$ circuit G with the property that for every S with $|S| = r$, $thr(G_S) \geq d$. Define the distribution ν as in the statement of Lemma 6.1 given the value of r. We can then apply Lemma 6.6 to G to derive the Π_4 circuit H' based on G with the property that for every π in the support of ν, $H'|_\pi$ computes as a subfunction the function G_S for some subset S of inputs with $|S| = r$ and therefore $thr(H'|_\pi) \geq thr(G_S) \geq d$.

Observe that for $\alpha = 0.9$ and $\beta = 0.8$ we have $q^\beta = q^{0.8} \geq q^{3/5} \log_2 q = rp$, $2^{p-1} - 1 = q^{1/4}/2 - 1 \geq q^{0.2} = q^{1-\beta}$, $q^\alpha = q^{0.9} \geq \frac{24}{\ln 2} q^{17/20} = \frac{6}{\ln 2} 2^p r$, and $d^{\alpha-\beta} = q^{0.1} \geq 3p/\ln 2$ for p sufficiently large. Therefore we may apply Lemma 6.1 to derive that $\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] \leq 2^{|\rho|^\alpha - |\rho|}$. It therefore follows that H' has α-threshold degree at least d as required. □

7 Multiparty communication complexity lower bounds for AC^0 functions for $O(\log n)$ players

In this section we use a different selector function ψ, which we denote by $\textsc{Index}_{\oplus^a_{k-1}}$ where $a > 0$ is an integer. This function has $s = 2^a$ and $D_{\textsc{Index}_{\oplus^a_{k-1}},j} = \{0,1\}^s$ for all j. For $X \in \{0,1\}^s$ and $Y \in \{0,1\}^{(k-1)s}$ define

$$\textsc{Index}_{\oplus^a_{k-1}}(X, Y) = X_{(Y_1 \oplus \ldots \oplus Y_{k-1})_{[a]}}$$

where the bits in X are indexed by a-bit vectors and $Y_{[a]}$ is the vector of the first a bits of Y. This function clearly satisfies the selector function requirement that the output be unbiased for each fixed value of Y.

Although the definition of $\textsc{Index}_{\oplus^a_{k-1}}$ uses the parity function, the number of players k will be $O(\log n)$ and hence these parity functions will be computable in AC^0. We can express the parity of $k-1$ items in DNF as an $\vee$ of 2^{k-2} conjunctions each of length $k-1$. Thus for any $w \in \{0,1\}^a$, we can check whether $(Y_1 \oplus \ldots \oplus Y_{k-1})_{[a]} = w$ by a Π_3 formula where the gates are, from top to bottom, $\wedge$ with fan-in a, $\vee$ with fan-in 2^{k-2}, and $\wedge$ with fan-in $k-1$. If we add x_w as an additional input to the top $\wedge$ gate, we can make this formula output x_w if the check returns true. Therefore we can write $\text{INDEX}_{\oplus^a_{k-1}}$ as a Σ_4 formula where the fan-ins are, from top to bottom, 2^a, $a+1$, 2^{k-2}, and $k-1$. The top $\vee$ gate is to do the check for every possible value of $w \in \{0,1\}^a$. Alternatively, we could dually write parity using CNF and express $\text{INDEX}_{\oplus^a_{k-1}}$ as a Σ_3 formula where the fan-ins are, from top to bottom, 2^a, $a2^{k-2}+1$, and $k-1$, where the inputs to each of the $(a2^{k-2}+1)$-fan-in $\wedge$ gates are the one bit of x and $a2^{k-2}$ $\vee$ gates with fan-in $k-1$.

When ψ is $\text{INDEX}_{\oplus^a_{k-1}}$, the variables $\text{INDEX}_{\oplus^a_{k-1}}(x_i, y^u_{*i})$ for $u \in \{0,1\}^{k-1}$ will be independent if and only if for every $u \neq v$, y^u_{*i} and y^v_{*i} select different bits of x_i. Hence we can easily prove an analogue of Proposition 3.3 for $\text{INDEX}_{\oplus^a_{k-1}}$.

Lemma 7.1. *If* $\psi = \text{INDEX}_{\oplus^a_{k-1}}$ *then* $\Pr_{y^0,y^1\in D^{(m)}_\psi}[r_\psi(y^0,y^1)=r] \le \binom{m}{r}2^{(2k-a-3)r} \le (\frac{em2^{2k-a-3}}{r})^r$.

Proof. Note that $D^{(m)}_\psi$ in this case is simply $\{0,1\}^{(k-1)ms}$. For each fixed $i \in [m]$ and each fixed pair of $u \neq v \in \{0,1\}^{k-1}$, the probability that y^u_{*i} and y^v_{*i} select the same bit of x_i is the probability that $(y^{u_1}_{*i} \oplus \cdots \oplus y^{u_{k-1}}_{*i})_{[a]} = (y^{v_1}_{*i} \oplus \cdots y^{v_{k-1}}_{*i})_{[a]}$. Since $u \neq v$, this is a homogeneous full rank system of a equations over $\mathbb{F}_2$ which is satisfied with probability precisely 2^{-a}. By a union bound over all of the $\binom{2^{k-1}}{2} < 2^{2k-3}$ pairs $u, v \in \{0,1\}^{k-1}$, it follows that the probability that i is bad for (y^0, y^1) is at most $2^{2k-3}2^{-a} = 2^{2k-a-3}$. The bound follows by the independence of the choices of (y^0, y^1) for different values of $i \in [m]$. □

Now we are ready to prove the main theorem for functions composed using this new selector function.

Theorem 7.2. *For* $0 \le \alpha < 1$ *and any Boolean function* f *on* m *bits with* (ϵ,α)*-approximate degree* d*, the function* $f \circ \text{INDEX}^m_{\oplus^a_{k-1}}$ *defined on* nk *bits, where* $n = ms$ *and* $s = 2^a \ge e2^{2k-1}m/d$*, requires that* $R^k_{1/2-\epsilon}(f \circ \text{INDEX}^m_{\oplus^a_{k-1}}) \ge d/2^k + \log_2(\epsilon(1-\epsilon))$ *for* $k \le (1-\alpha)\log_2 d$.

Proof. For $\psi = \text{INDEX}_{\oplus^a_{k-1}}$, by Lemma 7.1,

$$\sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot \Pr_{y^0,y^1\in D^{(m)}_\psi}[r_\psi(y^0,y^1)=r] \le \sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot (\frac{em2^{2k-a-3}}{r})^r \tag{21}$$

Since $k \le (1-\alpha)\log_2 d$, we have $(2^{k-1}-1)r^\alpha < d^{1-\alpha}r^\alpha \le r$ for $r \ge d$ so (21) is

$$\le \sum_{r=d}^{m} (\frac{em2^{2k-a-2}}{r})^r$$

$$\le \sum_{r=d}^{m} 2^{-r} < 2^{-(d-1)} \qquad \text{for } 2^a \ge e2^{2k-1}m/d.$$

Plugging this in to Theorem 4.3 we obtain that

$$R^k_{1/2-\epsilon}(f \circ \psi^m) \ge \log_2(\epsilon(1-\epsilon)) - \frac{1}{2^{k-1}}\log_2 2^{-(d-1)} > d/2^k + \log_2(\epsilon(1-\epsilon))$$

as required since $s = 2^a$. □

Let $\textsc{Tribes}'_{p,q}$ be the dual of the $\textsc{Tribes}_{p,q}$ function on $m = pq$ bits. Obviously the (ϵ, α)-degree of $\textsc{Tribes}'_{p,q}$ is the same as that of $\textsc{Tribes}_{p,q}$ for any ϵ and α. We first directly apply Theorem 7.2 to $\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}}$ and $\textsc{Tribes}'_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}}$ for suitable values of a.

Lemma 7.3. *Given any $1 > \epsilon > 0$, let q, p be positive integers with $q > p \geq 2$ such that $2\lceil q^{1-\beta}\rceil < 2^p \leq \frac{1}{6}q^{\alpha+\epsilon-1}\ln 2$ for some fixed constants $1 > \alpha > \beta > 1-\epsilon$. Let $a \geq \log_2(4epq^{(1+\epsilon)/2}) + 2k$ and $s = 2^a$. Then, for q sufficiently large, $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}})$ and $R^k_{1/3}(\textsc{Tribes}'_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}})$ are both $\Omega(q^{(1-\epsilon)/2}/2^k)$ for $k \leq \frac{1}{2}(1-\alpha)(1-\epsilon)\log_2 q - 2$.*

Proof. Let $m = pq$. By Corollary 6.3, for q sufficiently large, the $(5/6, \alpha)$-approximate degree d of both $\textsc{Tribes}_{p,q}$ and $\textsc{Tribes}'_{p,q}$ is at least $q^{(1-\epsilon)/2}/\sqrt{12}$. Thus $em/d \leq 4epq^{(1+\epsilon)/2}$ so by the choice of a we have $s = 2^a \geq e2^{2k-1}m/d$. Also $k \leq \frac{1}{2}(1-\alpha)(1-\epsilon)\log_2 q - 2$ implies that $k \leq (1-\alpha)\log_2 d$. Applying Theorem 7.2, we see that $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}})$ and $R^k_{1/3}(\textsc{Tribes}'_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}})$ are both $\Omega(q^{(1-\epsilon)/2}/2^k)$. □

In particular we obtain the following:

Corollary 7.4. *Let $p \geq 2$ be a sufficiently large integer. Let $q = 2^{4p}$, $k \leq p/10$, and $s = 2^{p+2k}$. Let $F = \textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ and $F' = \textsc{Tribes}'_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$. Let $n = pqs = p2^{5p+2k}$ be the number of input bits given to each player in computing F or F'. Then $R^k_{1/3}(F)$ and $R^k_{1/3}(F')$ are both $\Omega(q^{0.3}/2^k) = \Omega(2^{6p/5}/2^k)$ which is $n^{\Omega(1)}/O(4^k)$. Furthermore, F has polynomial-size depth 5 AC^0 formulas and F' has polynomial-size depth 4 AC^0 formulas.*

Proof. Let $\epsilon = 0.4$, $\alpha = 0.9$, and $\beta = 0.8$ and $a = p + 2k$. Observe that with these values and sufficiently large p, the conditions on the relationship between p and q are met for sufficiently large values of p as is the bound on a and the upper bound on k.

As noted above, $\textsc{Index}_{\oplus^a_{k-1}}$ has Σ_3 formulas with fan-in, top to bottom, of $2^a = 2^{p+2k}$, $a2^{k-2}+1 = (p+2k)2^{k-2}+1$, and $k-1$. Since $\textsc{Tribes}_{p,q}$ is given by a Σ_2 formula, $\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ is computable by a Σ_5 formula with fan-in top to bottom of q, p, 2^{p+2k}, $(p+2k)2^{k-2}+1$, and $k-1$. The total formula size of F is $n(p+2k+1)(k-1)2^{k-2}$ which is less than n^2.

The proof for F' goes similarly, except that since the second layer of $\textsc{Tribes}'_{p,q}$ can be merged with the top layer of $\textsc{Index}_{\oplus^{(p+2k)}_{k-1}}$, it has a polynomial-size depth 4 AC^0 formulas. □

Lemma 7.5. *$N^k(\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}})$ is $O(\log q + pa)$.*

Proof. Using the Σ_3 formula for $\textsc{Index}_{\oplus^a_{k-1}}$ we see that $\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}}$ can be expressed as a Σ_5 formula where the fan-ins from top to bottom are q, p, 2^a, $a2^{k-2}+1$, and $k-1$. The players use this formula to evaluate $\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}}$.

Observe that the fan-ins of the $\wedge$ gates are p, $a2^{k-2}+1$, and $k-1$ respectively, and the input to each of the $(a2^{k-2}+1)$-fan-in $\wedge$ gates at the middle $\wedge$ is one bit of x and $a2^{k-2}$ $\vee$ gates with fan-in $k-1$. Moreover, the 0-th player (who holds x), can evaluate each of these $\vee$ gates since it can see all of the input to these gates.

Player 0 guesses the top part of an accepting subtree by guessing a child of the root and, for each of the p children of that node, guesses which of the 2^a bits is selected, and broadcasts this information. This costs $\log_2 q + pa$ bits to send. Thus now there are p $\wedge$ gates with fan-in $a2^{k-2}+1$

that need to be evaluated. For each of these p gates, player 0 broadcasts a bit which is 1 if and only if all of the $a2^{k-2}$ feeding $\vee$ gates that depend on the bits of $y_1, \ldots, y_{k-1}$ evaluate to true. Given this information, player 1 can then evaluate all p $\wedge$ gates. $\square$

Corollary 7.6. *There is a function G in read-once depth 5 AC^0 such that G is in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for $k \le a' \log n$ for some constant $a' > 0$.*

Proof. Observe that, by Lemma 7.5, $F = \textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ with the parameters from Corollary 7.4 has $N^k(F)$ that is $O(\log^2 n)$ and thus satisfies all the conditions except for being read-once. To obtain the read-once property note that F is a restriction of the following function G

$$\bigvee_{u=1}^{q} \bigwedge_{v=1}^{p} \bigvee_{w=1}^{2^{p+2k}} \left(z_{0,u,v,w} \wedge \bigwedge_{\ell=1}^{(p+2k)2^{k-2}} \bigvee_{j=1}^{k-1} z_{j,u,v,w,\ell}\right)$$

and that the same $O(\log^2 n)$ upper bound from Lemma 7.5 applies equally well to G. $\square$

We now reduce the depth 5 function $f = \textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ to $\textsc{Disj}_{k,n}$ for a suitable value of n to obtain a NOF communication complexity lower bound on $\textsc{Disj}_{k,n}$ for k up to $\Theta(\log^{1/3} n)$ players. This is an exponential improvement in the number of players for which non-trivial lower bounds can be shown for $\textsc{Disj}_{k,n}$.

Theorem 7.7. *$R^k_{1/3}(\textsc{Disj}_{n,k})$ is $\Omega(2^{\sqrt{\log_2 n}/\sqrt{k}})$ for $k \le \frac{1}{5} \log_2^{1/3} n$.*

Proof. Recall that $\textsc{Disj}_{k,n}(x) = \vee_{i=1}^{n} \wedge_{j=0}^{k-1} x_{j,i}$. As in Corollary 7.6 start with $F = \textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ with the parameters from from Corollary 7.4. Unlike Corollary 7.6, however, we use the $\mathbf{\Sigma_4}$ circuit for $\textsc{Index}_{\oplus^{(p+2k)}_{k-1}}$ and reduce F to a $\mathbf{\Sigma_6}$ read-once formula G with $n = qp2^a(a+1)2^{k-2}k$ variables where $a = 2k + p$ given by

$$G(z) = \bigvee_{i=1}^{q} \bigwedge_{u=1}^{p} \bigvee_{v=1}^{2^a} \bigwedge_{w=1}^{a+1} \bigvee_{\ell=1}^{2^{k-2}} \bigwedge_{j=0}^{k-1} z_{j,i,u,v,w,\ell}.$$

Distributing the $\wedge$ gates through the $\vee$ gates, we have

$$G(z) \quad = \quad \bigvee_{i=1}^{q} \bigvee_{I \in [2^a]^p} \bigwedge_{u=1}^{p} \bigwedge_{w=1}^{a+1} \bigvee_{\ell=1}^{2^{k-2}} \bigwedge_{j=0}^{k-1} z_{j,i,u,I(u),w,\ell}$$

by distributing over the second "$\vee$", where $I(u)$ is the u-th index of I. This in turn equals

$$= \quad \bigvee_{i=1}^{q} \bigvee_{I \in [2^a]^p} \bigvee_{J \in [2^{k-2}]^{p(a+1)}} \bigwedge_{u=1}^{p} \bigwedge_{w=1}^{a+1} \bigwedge_{j=0}^{k-1} z_{j,i,u,I(u),w,J(u,w)}$$

by distributing over the third "$\vee$", where $J(u,w)$ is the entry of J indexed by (u,w). This in turn equals

$$\begin{aligned}
&= \bigvee_{i=1}^{q} \bigvee_{I\in[2^a]^p} \bigvee_{J\in[2^{k-2}]^{p(a+1)}} \bigwedge_{j=0}^{k-1} \bigwedge_{u=1}^{p} \bigwedge_{w=1}^{a+1} z_{j,i,u,I(u),w,J(u,w)} \\
&= \bigvee_{i=1}^{q} \bigvee_{I\in[2^a]^p} \bigvee_{J\in[2^{k-2}]^{p(a+1)}} \bigwedge_{j=0}^{k-1} y_{j,i,I,J} \\
&= \textsc{Disj}_{n,k}(y),
\end{aligned}$$

where the bits of vector $y \in \{0,1\}^{nk}$ for $n = q2^{ap+(k-2)p(a+1)}$ are indexed by $j \in \{0,\ldots,k-1\}$, $i \in [q]$, $I \in [2^a]^p$ and $J \in [2^{k-2}]^{p(a+1)}$ are given by

$$y_{j,i,I,J} = \bigwedge_{u=1}^{p} \bigwedge_{w=1}^{a+1} z_{j,i,u,I(u),w,J(u,w)}.$$

Observe that for any two players $j \neq j'$, player j' can compute any value $y_{j,i,I,J}$. Thus the k players can compute $\textsc{Tribes}_{p,q} \circ \psi^m_{k,\ell}$ by executing a NOF randomized communication protocol for $\textsc{Disj}_{n,k}$ on y of length nk, where $n = q2^{ap+(k-2)p(a+1)} = q2^{ap(k-1)+k-2}$. Plugging in $q = 2^{4p}$ and $a = 2k+p$ for $k \leq p/10$ we have that $R^k_{1/3}(\textsc{Disj}_{n,k})$ is $\Omega(2^{6p/5-k})$. Now for these values of k and a, we have $ap \geq k-2+4p$ and hence we have that $n \leq 2^{apk} \leq 2^{6p^2k/5}$. Therefore we have $p \geq \sqrt{5\log_2 n}/\sqrt{6k}$. It follows that $R^k_{1/3}(\textsc{Disj}_{n,k})$ is $\Omega(2^{\sqrt{\log_2 n}/\sqrt{k}})$ provided that $k \leq \frac{1}{10}\sqrt{5\log_2 n}/\sqrt{6k}$ which holds if $k \leq \frac{1}{5}\log_2^{1/3} n$. □

We note that the same lower bound for disjointness can be derived even more simply using the above technique for the simpler function $\textsc{Tribes}_{p,q} \circ \psi^m_{k,\ell}$ that uses the pattern tensor selector. This function is analyzed in Appendix A. Note that our bound also shows the following:

Corollary 7.8. *There is a depth-2 read-once* AC^0 *formula in* $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ *for k up to* $\Theta(\log^{1/3} n)$.

Although we have shown non-trivial lower bounds for $\textsc{Disj}_{k,n}$ for k up to $\Theta(\log^{1/3} n)$, it is open whether one can prove stronger lower bounds for $k = \omega(\log^{1/3} n)$ players for $\textsc{Disj}_{k,n}$ or any other depth-2 AC^0 function. The difficulty of extending our lower bound methods is our inability to apply Lemma 4.1 to $\textsc{Or}$ since the constant function 1 approximates $\textsc{Or}$ on all but one point.

For the functions we have considered so far we only have obtained lower bounds for protocols that have some fixed constant advantage. Using Theorem 6.7, which produces a function of large α-threshold degree, we obtain a lower bound on the randomized multiparty communication complexity of AC^0 functions for protocols that succeed with probability barely better than that of random guessing. This lower bound will be useful in proving lower bounds for $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuits in the next section.

Theorem 7.9. *There exist explicit constants $c, c' > 0$ and a depth 6* AC^0 *function $H : \{0,1\}^* \to \{0,1\}$ such that for $1/2 > \epsilon > 0$, $R^k_{1/2-\epsilon}(H_n)$ is $\Omega(n^c + \log \epsilon)$ for any $k \leq c' \log_2 n$.*

Proof. Let f' be the Π_4 function on $m = pq$ bits with 0.9-threshold degree at least $m^{1/15}/\log_2 m$ as given by Theorem 6.7. We use the dual function f to f' which is therefore a Σ_4 function of the same approximate degree. Since f has 0.9-threshold degree at least $m^{1/15}/\log_2 m$, it has $(< 1-\epsilon, 0.9)$-approximate degree at least $d = \lceil m^{1/15}/\log_2 m \rceil$ for any $\epsilon > 0$. For $k \leq 0.1 \log_2 d$, let

$a = \lceil \log_2(e2^{2k-1}m/d) \rceil$, and $s = 2^a$. By Theorem 7.2, the function $H_n = f \circ \textsc{Index}^m_{\oplus^a_{k-1}}$ defined on $n = msk$ bits requires that

$$R^k_{1/2-\epsilon}(H_n) \geq d/2^k + \log_2(\epsilon(1-\epsilon)).$$

Since d is $m^{\Omega(1)}$ and $k \leq 0.1 \log_2 d$, $n = msk = m2^a k$ is $d^{O(1)}$ and since $\epsilon < 1/2$ the lower bound on $R^k_{1/2-\epsilon}(H_n)$ is $\Omega(n^c + \log \epsilon)$ for some explicit constant $c > 0$. Combining the $\mathsf{\Pi}_3$ circuit for $\textsc{Index}_{\oplus^a_{k-1}}$ with that for f yields depth 6. □

8 Threshold circuit lower bounds for AC^0

Following the approach used by Viola [29], we show quasipolynomial lower bounds on the simulation of AC^0 functions by unrestricted $\mathsf{SYMM} \circ \mathsf{AND}$ and $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuits. This relies on the connection between multiparty communication complexity and threshold circuit complexity given by Håstad and Goldmann.

Proposition 8.1 (Håstad and Goldmann [14])**.**

- *If f is computed by a $\mathsf{SYMM} \circ \mathsf{AND}_{k-1}$ circuit of size S, then $D^k(f)$ is $O(k \log S)$;*
- *If f is computed by a $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}_{k-1}$ circuit of size S, then $R^k_{1/2-1/(2S)}(f)$ is $O(k \log S)$.*

We find it convenient to prove these results first for the simpler case of $\mathsf{SYMM} \circ \mathsf{AND}$ circuits. The proofs are almost identical to the argument in [29] with our hard AC^0 functions replacing the generalized inner product.

8.1 Lower bound for simulation by $\mathsf{SYMM} \circ \mathsf{AND}$ circuits

Theorem 8.2. *There is a function $G : \{0,1\}^* \to \{0,1\}$ in read-once depth-6 AC^0 such that G_N requires $\mathsf{SYMM} \circ \mathsf{AND}$ circuit size $N^{\Omega(\log \log N)}$.*

Proof. For any $k \geq 2$ let $p = 10k$, $q = 2^{4p}$, $s = 2^{p+2k}$, and $n = kpqs = pk2^{5p+2k} = p^2 2^{5.2p}/10$. Note that the function $F_p = \textsc{Tribes}'_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}} : \{0,1\}^n \to \{0,1\}$ is given by

$$F_p(X,Y) = \bigwedge_{u=1}^{q} \bigvee_{v=1}^{p} \textsc{Index}_{\oplus^{(p+2k)}_{k-1}}(X_{u,v}, Y_{u,v})$$

for any $X \in \{0,1\}^{pqs}$ and $Y \in \{0,1\}^{(k-1)pqs}$. Next, for $N = 49p^2 n$ and $Z = Z_1 \cdots Z_n$, where each $Z_i \in \{0,1\}^{49p^2}$, we define our hard function $G_N : \{0,1\}^N \to \{0,1\}$ as

$$G'_N(Z) = F_p\Big(\bigoplus_{j=1}^{49p^2} Z_{1j}, \ldots, \bigoplus_{j=1}^{49p^2} Z_{nj}\Big).$$

By Corollary 7.4, F_p is in depth-4 AC^0. Moreover, the parity on $O(p^2) = O(\log^2 N)$ bits can be computed by an AC^0 circuit of depth 3. It follows that G'_N is computable by depth-6 AC^0 circuits. It remains to show that G'_N is not computable by "small" $\mathsf{SYMM} \circ \mathsf{AND}$ circuits.

Suppose by contradiction that for some sufficiently small constant $\delta > 0$, there is such a circuit C of size $N^{\delta \log \log N}$ that computes G'_N. Let $\rho \in \{0,1,*\}^N$ be a random restriction such that $|\text{unset}(\rho)| = \lfloor \frac{N}{7p} \rfloor = 7pn$. We denote by $C|_\rho$ the circuit obtained from C after substituting all the values as prescribed by ρ. We consider the following two events:

- Event E_1: the function computed by $C|_\rho$ is computed by a $\mathsf{SYMM} \circ \mathsf{AND}$ circuit of size at most $|C| \cdot 2^k$ where the fan-in of each AND-gate is strictly less than k, and
- Event E_2: there is at least one bit that is left unassigned by ρ in every Z_i for $1 \leq i \leq n$.

First, we show that $\Pr[\neg E_1] < 1/2$ for sufficiently small $\delta > 0$: Fix any AND-gate φ in the second layer of C. By the decision tree version of Håstad's Switching Lemma[2] (cf. [5]), the probability over ρ that $\varphi|_\rho$ cannot be computed by a decision tree of depth strictly less than k is at most

$$(\frac{7|\text{unset}(\rho)|}{N})^k \leq (\frac{1}{p})^{p/10} = 2^{-0.1p\log_2 p}.$$

Since p is $\Theta(\log N)$ this quantity is $N^{-\Omega(\log\log N)}$. Thus by a union bound over all AND-gates in C and for sufficiently small δ, with probability strictly less than $1/2$, the function computed by $C|_\rho$ is computable by a symmetric function of at most $|C|$ decision trees of height strictly less than k. Any decision tree of height $< k$ can be written as a DNF of less than 2^k disjointly satisfied ANDs, each of size less than k. We can merge each of these terms into the top symmetric gate and conclude that the function computed by $C|_\rho$ is computed by a $\mathsf{SYMM} \circ \mathsf{AND}$ circuit of size at most $|C| \cdot 2^k$ where the fan-in of each AND-gate is strictly less than k.

Next, we also show that $\Pr[\neg E_2] < 1/2$. Fix any Z_i for some $i \in [n]$. It is easy to see that the probability that ρ assigns values to all of the bits in Z_i is the probability that all of unset(ρ) is outsize of Z_i which is at most

$$\Big(1 - \frac{|Z_i|}{N}\Big)^{\lfloor N/(7p) \rfloor} = \Big(1 - \frac{1}{n}\Big)^{7pn} \leq \exp(-7p) < 1/(2n)$$

for sufficiently large p. By union bound over all $i \in [n]$, we conclude that $\Pr[\neg E_2] < 1/2$.

Thus $\Pr[E_1 \wedge E_2] > 0$. Hence there exists a restriction ρ such that both E_1 and E_2 hold. By Proposition 8.1, the fact that E_1 holds implies that there is a k-party deterministic communication protocol computing $C|_\rho$ exchanging at most $O(k\log(|C| \cdot 2^k))$ which is $O(\log^3 N)$ bits. On the other hand, the fact that E_2 holds implies that $C|_\rho$ computes F_p as a subfunction. By Corollary 7.4, there is an assignment of the input to F_p, and therefore $C|_\rho$, to k players such that any k-party deterministic communication protocol computing F_p, and therefore $C|_\rho$, must exchange at least

$$\Omega(2^{6p/5}/2^k) = N^{\Omega(1)}$$

bits. Hence for sufficiently large N, we arrive at a contradiction. □

8.2 Lower bound for simulation by $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuits

Theorem 8.3. *There is a function $G : \{0,1\}^* \to \{0,1\}$ in AC^0 such that G_N requires $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuit size $N^{\Omega(\log\log N)}$.*

Proof. We follow the approach in proving Theorem 8.2, except that we use the function H_n from Theorem 7.9 instead of F_p. Let c, c' be the constants and H_n be the function as guaranteed by

[2] One could also use the original form [13] with suitable additional argument about the result of applying it to a conjunction, though the decision tree version is more convenient here.

Theorem 7.9. Let $k = \lfloor c' \log_2 n \rfloor$, $r = \lfloor \log_2 n \rfloor$, and $N = 49r^2 n$. For any $Z = Z_1 \cdots Z_n$, where each $Z_i \in \{0,1\}^{49r^2}$, we define our hard function $G_N : \{0,1\}^N \to \{0,1\}$ as

$$G'_N(Z) = H_n\big(\bigoplus_{j=1}^{49r^2} Z_{1j}, \ldots, \bigoplus_{j=1}^{49r^2} Z_{nj}\big).$$

As in the proof of Theorem 8.2, suppose by contradiction that for some sufficiently small constant $\delta > 0$, there is a circuit C of size $N^{\delta \log\log N}$ that computes G'_N. Let $\rho \in \{0,1,*\}^N$ be a random restriction such that $|\text{unset}(\rho)| = \lfloor \frac{N}{7r} \rfloor = 7rn$ and consider the two events:

- Event E_1: the function computed by $C|_\rho$ is computed by a $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuit of size at most $|C| \cdot 2^k$ where the fan-in of each AND-gate is strictly less than k, and
- Event E_2: there is at least one bit that is left unassigned by ρ in every Z_i for $1 \le i \le n$.

The fact that, for $\delta > 0$ sufficiently small, $\Pr[E_1 \wedge E_2] > 0$ is essentially identical to that in the proof of Theorem 8.2.

Hence there exists a restriction ρ such that both E_1 and E_2 hold. By Proposition 8.1, the fact that E_1 holds implies that for any partition of the input to k players and $\epsilon = 1/(|C| \cdot 2^{k+1})$, $R^k_{1/2-\epsilon}(C|_\rho)$ is $O(k \log(|C| \cdot 2^k)) = O(\log^3 N) = O(\log^3 n)$. On the other hand, the fact that E_2 holds implies that $C|_\rho$ computes H_n as a subfunction. By Theorem 7.9, there is an assignment of the input bits of H_n, and therefore of $C|_\rho$, to k players such that $R^k_{1/2-\epsilon}(C|_\rho) \ge R^k_{1/2-\epsilon}(H_n)$ which is $\Omega(n^c + \log \epsilon)$. Since $-\log_2 \epsilon$ is $O(k + \log |C|) = O(\log^2 N) = O(\log^2 n)$, $\Omega(n^c + \log \epsilon)$ is $\Omega(n^c)$ for sufficiently large N (and hence n), we arrive at a contradiction. □

Acknowledgements

We would like to thank Emanuele Viola for suggesting the application of our multiparty communication complexity lower bounds to circuit complexity. We would also like to thank Sasha Sherstov and Arkadev Chattopadhyay for pointing out the error in our calculation of the bound required for ℓ and therefore s in the proof of Theorem 4.4 in a previous version of this paper, and also for other helpful comments.

References

[1] Eric W. Allender. A note on the power of threshold circuits. In *30th Annual Symposium on Foundations of Computer Science*, pages 580–584, Research Triangle Park, NC, October 1989. IEEE.

[2] L. Babai, A. Gál, P. G. Kimmel, and S. V. Lokam. Communication complexity of simultaneous messages. *SIAM Journal on Computing*, 33(1):137–166, 2003.

[3] L. Babai, T. P. Hayes, and P. G. Kimmel. The cost of the missing bit: Communication complexity with help. *Combinatorica*, 21(4):455–488, 2001.

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

[5] P. Beame. A switching lemma primer. Technical Report UW-CSE-95–07–01, Department of Computer Science and Engineering, University of Washington, November 1994.

[6] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of set disjointness. *Computational Complexity*, 15(4):391–432, 2006.

[7] A. Ben-Aroya, O. Regev, and R. de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing. In *Proceedings 49th Annual Symposium on Foundations of Computer Science*, Philadelphia,PA, October 2008. IEEE. To appear.

[8] Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proceedings 48th Annual Symposium on Foundations of Computer Science*, pages 449–458, Berkeley,CA, October 2007. IEEE.

[9] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. Technical Report TR08-002, Electronic Colloquium in Computation Complexity, `http://www.eccc.uni-trier.de/eccc/`, 2008.

[10] F. R. K. Chung. Quasi-random classes of hypergraphs. *Random Structures and Algorithms*, 1(4):363–382, 1990.

[11] Matei David, Toniann Pitassi, and Emanuele Viola. Improved separations between nondeterministic and randomized multiparty communication. In *RANDOM 2008, 12th International Workshop on Randomization and Approximization Techniques in Computer Science*, pages 371–384, 2008.

[12] F. Green, J. Köbler, K. W. Regan, T. Schwentick, and J. Torán. The power of the middle bit of a #p function. *Journal of Computer and System Sciences*, 50(3):456–467, 1995.

[13] J. Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 6–20, Berkeley, CA, May 1986.

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

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

[16] R. Jain, H. Klauck, and A. Nayak. Direct product theorems for classical communication complexity via subdistribution bounds. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 599–608, Victoria, BC, May 2008.

[17] E. Kushilevitz and N. Nisan. *Communication Complexity.* Cambridge University Press, Cambridge, England ; New York, 1997.

[18] T. Lee and A. Shraibman. Disjointness is hard in the multi-party number-on-the-forehead model. In *Proceedings Twenty-Third Annual IEEE Conference on Computational Complexity*, pages 81–91, College Park, Maryland, June 2008.

[19] M. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability. In *30th Annual Symposium on Foundations of Computer Science*, Research Triangle Park,NC, October 1989.

[20] M. Minsky and S. Papert. *Perceptrons.* MIT Press, Cambridge, MA, 1988. Expanded Edition. The first edition appeared in 1968.

[21] N. Nisan and M. Szegedy. On the degree of boolean functions as real polynomials. *Computational Complexity*, 4:301–314, 1994.

[22] A. Razborov and A. Wigderson. Lower bounds on the size of depth 3 threshold circuits with AND gates at the bottom. *Information Processing Letters*, 45:303–307, 1993.

[23] A. A. Razborov and A. A. Sherstov. The sign-rank of AC^0. In *Proceedings 49th Annual Symposium on Foundations of Computer Science*, Philadelphia,PA, October 2008. IEEE. To appear.

[24] A. A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, pages 294–301, San Diego, CA, June 2007.

[25] A. A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the European Association for Theoretical Computer Science*, 95:59–93, 2008.

[26] A. A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 85–94, Victoria, BC, May 2008.

[27] A. A. Sherstov. Unbounded-error communication complexity of symmetric functions. In *Proceedings 49th Annual Symposium on Foundations of Computer Science*, Philadelphia,PA, October 2008. IEEE. To appear.

[28] P. Tesson. *Communication Complexity Questions Related to Finite Monoids and Semigroups.* PhD thesis, McGill University, 2002.

[29] E. Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM Journal on Computing*, 36(5):1387–1403, 2007.

[30] E. Viola and A. Wigderson. One-way multi-party communication lower bound for pointer jumping with applications. In *Proceedings 48th Annual Symposium on Foundations of Computer Science*, pages 427–437, Berkeley,CA, October 2007. IEEE.

[31] A. C. Yao. On ACC and threshold circuits. In *Proceedings 31st Annual Symposium on Foundations of Computer Science*, pages 619–627, St. Louis, MO, October 1990. IEEE.

A Multiparty communication complexity bounds for AC^0 using the pattern tensor method

A.1 Lower bounds and separations for depth 3 and 4 AC^0 functions for $O(\sqrt{\log n})$ players

We first show that $\textsc{Tribes}_{p,q} \circ \psi^m_{k,\ell}$ separates $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}}$ and $\mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for $k = O(\sqrt{\log n})$ for some appropriately chosen values of p and q.

Here we apply the (ϵ, α) degree bound for the $\textsc{Tribes}$ function shown in the previous section with Theorem 4.4 for the pattern tensor selector function $\psi_{k,\ell}$. Note that

$$\textsc{Tribes}_{p,q} \circ \psi^m_{k,\ell}(x) = \vee_{i\in[q]} \wedge_{u\in[p]} \vee_{u\in[s]} \wedge_{j\in[k]} x_{j,u,v,i}$$

is a depth 4 read-once formula. Recall that $\textsc{Tribes}'_{p,q}$ is the dual of the $\textsc{Tribes}_{p,q}$ function on $m = pq$ bits and has the same (ϵ, α)-degree of $\textsc{Tribes}'_{p,q}$ is the same as that of $\textsc{Tribes}_{p,q}$ for any ϵ and α. Observe also that

$$\textsc{Tribes}'_{p,q}\circ,^m \psi_{k,\ell}(x) = \wedge_{i\in[q]} \vee_{u\in[p],\ u\in[s]} \wedge_{j\in[k]} x_{j,u,v,i}$$

is a depth 3 read-once formula since the bottom layer of $\vee$ gates in $\textsc{Tribes}'_{p,q}$ can be combined with the top layer of $\psi_{k,\ell}$.

Lemma A.1. *Let $0 < \epsilon < 1/2$. Let q, p be sufficiently large positive integers with $q > p \geq 2$ such that $2\lceil q^{1-\beta}\rceil < 2^p \leq \frac{1}{6} q^{\alpha+\epsilon-1} \ln 2$ for some fixed constants $1 > \alpha > \beta > 1-\epsilon$. Let $s = \lceil 8\sqrt{3}e(k-1)pq^{(1+\epsilon)/2}\rceil^{k-1}$ and $n = pqs$. Then $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \psi^m_{k,\ell})$ and $R^k_{1/3}(\textsc{Tribes}'_{p,q} \circ \psi^m_{k,\ell})$ are both $\Omega(q^{(1-\epsilon)/2}/2^k)$, which is $\Omega(n^{1/(4k)}/2^k)$ for $k^2 \leq a \log_2 n$ for some constant $a > 0$ depending only on α, ϵ.*

In particular, for any $\delta > 0$, one can choose an $\epsilon > 0$ and other parameters as above to obtain a lower bound on $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \psi^m_{k,\ell})$ and $R^k_{1/3}(\textsc{Tribes}'_{p,q} \circ \psi^m_{k,\ell})$ of $\Omega(n^{(1-\delta)/(k+1)}/(2^k \log n))$.

Proof. We state the proof for $\textsc{Tribes}_{p,q} \circ \psi^m_{k,\ell}$. The same proof applies for $\textsc{Tribes}'_{p,q} \circ \psi^m_{k,\ell}$.

By Corollary 6.3, for q sufficiently large $\textsc{Tribes}_{p,q}$ has $(5/6, \alpha)$-approximate degree d at least $q^{(1-\epsilon)/2}/\sqrt{12}$. Letting $m = pq$ we observe that $4e(k-1)m/d \leq 8\sqrt{3}e(k-1)m/q^{(1-\epsilon)/2}$ and hence $s \geq \lceil 4e(k-1)m/d\rceil^{k-1}$. Then we can apply Theorem 4.4 to derive that $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \psi^m_{k,\ell})$ is $\Omega(q^{(1-\epsilon)/2}/2^k)$, when $k \leq b \log_2 q$, for some constant $b > 0$ depending only on α, ϵ.

We now bound the value of q as a function of n, k and ϵ. Since $\epsilon > 0$, $n > qs > q^{(k+1)/2}$ so $q \leq n^{2/(k+1)}$. Therefore $p < \log_2 q \leq \frac{2}{k+1} \log_2 n$. We now have $n = pqs \leq (ck)^{k-1} p^k q^{1+(1+\epsilon)(k-1)/2}$ for some constant $c > 0$ and thus

$$n \leq q^{(k+1)/2+\epsilon(k-1)/2} (c' \log_2 n)^k \tag{22}$$

for some constant $c' > 0$. Since $\epsilon < 1$ it follows that $q^k \geq n/(c' \log_2 n)^k$ and therefore $q \geq n^{1/k}/(c' \log_2 n)$ so $\log_2 q > \frac{1}{k} \log_2 n - \log_2 \log_2 n - c''$ for some constant c''. Therefore there is an a depending on c'' and b such that for q sufficiently large (which implies that n is) the assumption $k^2 \leq a \log_2 n$ implies that $k \leq b \log_2 q$ as required.

It remains to derive an expression for the complexity lower bound as a function of n. By (22), $q^{(1-\epsilon)/2}$ is at least

$$n^{\frac{1-\epsilon}{k+1+\epsilon(k-1)}} / (c \log_2 n)^{\frac{k(1-\epsilon)}{k+1+\epsilon(k-1)}},$$

which is $\Omega(n^{1/(3k+1)}/(\log n)^{1/3})$ for $\epsilon < 1/2$ and thus $\Omega(n^{1/(4k)})$ since $k^2 \leq a \log_2 n$ and n is sufficiently large. Moreover, since $\frac{1-\epsilon}{k+1+\epsilon(k-1)}$ is of the form $1/(k+1) - 2\epsilon k/(k+1)^2 + O(\epsilon^2/(k+1))$ we obtain the claimed asymptotic complexity bound as ϵ approaches 0. □

Choosing $\epsilon = 0.4$, $\alpha = 0.9$, and $\beta = 0.8$ in the above lemma we obtain the following less cluttered lower bound statement.

Corollary A.2. *Let $p \geq 2$ be a sufficiently large integer, $q = 2^{4p}$, and $m = pq$. Let $k \geq 2$ be an integer, $s = \lceil 8\sqrt{3}e(k-1)pq^{0.7}\rceil^{k-1}$, and $n = ms = pqs$. Then $R^k_{1/3}(\text{TRIBES}_{p,q} \circ \psi^m_{k,\ell})$ and $R^k_{1/3}(\text{TRIBES}'_{p,q} \circ \psi^m_{k,\ell})$ are both $\Omega(q^{0.3}/2^k)$ for $k^2 \leq b \log_2 n$ for some constant $b > 0$ which is $\Omega(n^{1/(4k)}/2^k)$ when k is at most $O(\sqrt{\log n})$.*

B Multiparty communication complexity lower bounds for AC^0 functions for $O(\log n/\log\log n)$ players

In this section we use a different selector function ψ, which we denote by $\psi^{\oplus b}_{k,\ell}$. This function has $s = b\ell^{k-1}$ and is the $\oplus$ of b independent copies of the pattern tensor $\psi_{k,\ell}$. Therefore the domain $D_{\psi^{\oplus b}_{k,\ell},j}$ is simply $D^b_{\psi^{\oplus b}_{k,\ell},j}$, the set of b-tuples of vectors in the domain for the pattern tensor. In particular for $X \in \{0,1\}^s$ and $Y \in \{0,1\}^{(k-1)s}$

$$\psi^{\oplus b}_{k,\ell}(X,Y) = \bigoplus_{b'=1}^{b} \bigvee_{s'=1}^{\ell^{k-1}} (X_{b's'} \wedge \bigwedge_{j=1}^{k-1} Y_{jb's'}).$$

This function clearly satisfies the selector function requirement that the output be unbiased for each fixed value of Y.

Although the definition of $\psi^{\oplus b}_{k,\ell}$ uses the parity function, in applications we will choose values of b that will be $O(\log n)$ and hence these parity functions will be computable in AC^0. We can express the parity of b items in a DNF formula as an $\vee$ of 2^{b-1} conjunctions each of length b. In $\psi^{\oplus b}_{k,\ell}$ the b inputs to these terms are each pattern tensors of the form $\psi_{k,b'}(X,Y) = \bigvee_{s'=1}^{\ell^{k-1}}(X_{b's'} \wedge \bigwedge_{j=1}^{k-1} Y_{jb's'})$ and their negations. Because of the special form of the promise for the inputs to each of these pattern tensors, we see that the negation of a pattern tensor is $\overline{\psi}_{k,b'}(X,Y) = \bigvee_{s'=1}^{\ell^{k-1}}(\overline{X}_{b's'} \wedge \bigwedge_{j=1}^{k-1} Y_{jb's'})$.

Therefore we can write $\psi^{\oplus b}_{k,\ell}$ as a $\mathsf{\Sigma}_4$ formula where the fan-ins are, from top to bottom, 2^{b-1}, b, s, and k. We could dually write parity using CNF form and express $\psi^{\oplus b}_{k,\ell}$ as a $\mathsf{\Pi}_3$ formula where the fan-ins are, from top to bottom, 2^{b-1}, bs, and k. The former will be useful for small non-deterministic communication complexity whereas the latter will be useful for small circuit depth.

When ψ is $\psi^{\oplus b}_{k,\ell}$, the variables $\psi^{\oplus b}_{k,\ell}(x_i, y^u_{*i})$ for $u \in \{0,1\}^{k-1}$ will be independent if and only if for every $u \neq v$ there is some $b' \in [b]$ such that $y^u_{*ib'}$ and $y^v_{*ib'}$ select different bits of $x_{ib'}$. (This follows since random variables $\oplus_{b' \in [b]} w_{b'}$ and $\oplus_{b' \in [b]} w'_{b'}$ are independent if there is some b' such that $w_{b'}$ and $w'_{b'}$ are independent.) It follows that in this case $r_{\psi^{\oplus b}_{k,\ell}}(y^0, y^1)$ is the number of $i \in [m]$ such that for every $b' \in [b]$, $y^0_{jib'} = y^1_{jib'}$ for some $j \in [k-1]$.

The key to the improvement possible with $\psi^{\oplus b}_{k,\ell}$ is that we can prove a sharper analogue of Proposition 3.3.

Lemma B.1. *If $\psi = \psi_{k,\ell}^{\oplus b}$ then* $\Pr_{y^0,y^1\in D_\psi^{(m)}}[r_\psi(y^0,y^1) = r] \leq \binom{m}{r}\big(\frac{k-1}{\ell}\big)^{br} \leq \big(\frac{em(k-1)^b}{r\ell^b}\big)^r$.

Proof. In this case $r_{\psi_{k,\ell}^{\oplus b}}(y^0,y^1)$ is the number of $i \in [m]$ such that for every $b' \in [b]$, $y^0_{jib'} = y^1_{jib'}$ for some $j \in [k-1]$. As in the case of Proposition 3.3, for each fixed i and b' the probability that $y^0_{jib'} = y^1_{jib'}$ for some $j \in [k-1]$ is bounded above by $(k-1)/\ell$. Since the values of (y^0,y^1) are independently chosen for different values of $b' \in [b]$ the probability for each fixed i that this holds for all $b' \in [b]$ is at most $\big(\frac{k-1}{\ell}\big)^b$. The bound follows by the independence of the choices of (y^0,y^1) for different values of $i \in [m]$. □

Now we are ready to prove the main theorem for functions composed using this selector function.

Theorem B.2. *For $0 \leq \alpha < 1$ and any Boolean function f on m bits with $(5/6,\alpha)$-approximate degree d, the function $f\circ(\psi_{k,\ell}^{\oplus b})^m$ defined on nk bits, where $n = ms$ and $s = b\lceil(k-1)(4em/d)^{1/b}\rceil^{k-1}$, requires that $R^k_{1/3}(f \circ (\psi_{k,\ell}^{\oplus b})^m \geq d/2^k - 3$ for $k \leq (1-\alpha)\log_2 d$.*

Proof. For $\psi = \psi_{k,\ell}^{\oplus b}$, by Lemma B.1,

$$\sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot \Pr_{y^0,y^1\in D_\psi^{(m)}}[r_\psi(y^0,y^1) = r] \;\leq\; \sum_{r=d}^{m} 2^{(2^{k-1}-1)r^\alpha} \cdot \big(\frac{em(k-1)^b}{r\ell^b}\big)^r \tag{23}$$

Since $k \leq (1-\alpha)\log_2 d$, we have $(2^{k-1}-1)r^\alpha < d^{1-\alpha}r^\alpha \leq r$ for $r \geq d$ so (23) is

$$\begin{aligned} &\leq \sum_{r=d}^{m}\big(\frac{2em(k-1)^b}{r\ell^b}\big)^r \\ &\leq \sum_{r=d}^{m} 2^{-r} < 2^{-(d-1)} \qquad \text{for } \ell \geq (k-1)[d/(4em)]^{1/b}. \end{aligned}$$

Plugging this in to Theorem 4.3 we obtain that

$$R^k_{1/3}(f\circ\psi^m) \geq \log_2(5/36) - \frac{1}{2^{k-1}}\log_2 2^{-(d-1)} > d/2^k - 3$$

as required since $s = b\ell^{k-1}$. □

We first directly apply Theorem B.2 to $\textsc{Tribes}_{p,q} \circ (\psi_{k,\ell}^{\oplus b})^m$ for suitable values of b.

Lemma B.3. *Given any $1 > \epsilon > 0$, let q, p be positive integers with $q > p \geq 2$ such that $2\lceil q^{1-\beta}\rceil < 2^p \leq \frac{1}{6}q^{\alpha+\epsilon-1}\ln 2$ for some fixed constants $1 > \alpha > \beta > 1-\epsilon$. Let $b \geq \lceil\log_2(16epq^{(1+\epsilon)/2})\rceil$ and $s = b(2k)^{k-1}$. Then, for q sufficiently large, $R^k_{1/3}(\textsc{Tribes}_{p,q}\circ(\psi_{k,\ell}^{\oplus b})^m)$ is $\Omega(q^{(1-\epsilon)/2}/2^k)$ for $n = pqs$ and $k \leq \frac{1}{2}(1-\alpha)(1-\epsilon)\log_2 q - 2$.*

Proof. Let $m = pq$. By Corollary 6.3, for q sufficiently large, the $(5/6,\alpha)$-approximate degree d of $\textsc{Tribes}_{p,q}$ is at least $q^{(1-\epsilon)/2}/\sqrt{12}$. Thus $4em/d \leq 16epq^{(1+\epsilon)/2}$ so by the choice of b we have $(4em/d)^{1/b} \leq 2$. Therefore $s = b(2k)^{k-1} \geq b\lceil(k-1)(4em/d)^{1/b}\rceil^{k-1}$. Also $k \leq \frac{1}{2}(1-\alpha)(1-\epsilon)\log_2 q - 2$ implies that $k \leq (1-\alpha)\log_2 d$. Applying Theorem B.2, we see that $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ (\psi_{k,\ell}^{\oplus b})^m)$ is $\Omega(q^{(1-\epsilon)/2}/2^k)$. □

In particular we obtain the following:

Corollary B.4. *Let $p \geq 2$ be a sufficiently large integer. Let $q = 2^{4p}$, $k \leq p/40$, and $s = p(2k)^{k-1}$. Let $n = pqs = p^2 2^{4p}(2k)^{k-1}$ be the number of input bits given to each player in computing $F = \textsc{Tribes}_{p,q} \circ (\psi_{k,\ell}^{\oplus b})^m$. Then $R_{1/3}^k(F)$ is $\Omega(q^{0.3}/2^k) = \Omega(2^{6p/5}/2^k)$ which is $n^{\Omega(1)}/k^{O(k)}$. Further, F has polynomial-size depth 4 AC^0 formulas.*

Proof. We apply Corollary 6.4 instead of Corollary 6.3. As noted above, $\psi_{k,\ell}^{\oplus b}$ has $\mathsf{\Pi}_3$ formulas with fan-in, top to bottom, of $2^{b-1} = 2^{p-1}$, $bs = ps$, and k. Since $\textsc{Tribes}_{p,q}$ is given by a $\mathsf{\Sigma}_2$ formula, $\textsc{Tribes}_{p,q} \circ (\psi_{k,\ell}^{\oplus b})^m$ is computable by a $\mathsf{\Sigma}_4$ formula with fan-in top to bottom of q, $p2^{p-1}$, ps, and k. The total formula size of F is $np2^{p-1}$ which is less than $n^{5/4}\log_2 n$. □

Lemma B.5. *$N^k(\textsc{Tribes}_{p,q} \circ (\psi_{k,\ell}^{\oplus b})^m)$ is $O(\log q + pb\log s)$.*

Proof. Using the $\mathsf{\Sigma}_4$ formula for $\psi_{k,\ell}^{\oplus b}$ we see that $\textsc{Tribes}_{p,q} \circ (\psi_{k,\ell}^{\oplus b})^m$ can be expressed as a $\mathsf{\Sigma}_6$ formula where the fan-ins from top to bottom are q, p, 2^{b-1}, b, s, and k. Observe that the fan-ins of the $\wedge$ gates are p, b, and k respectively. The players use this formula to evaluate $\textsc{Tribes}_{p,q} \circ (\psi_{k,\ell}^{\oplus b})^m$.

The 0-th player (who holds x), guesses an accepting subtree of this formula and sends both the the description of the subtree and the values of the bits of x at the leaves of this subtree. Player 1 can then evaluate the subtree and sends 1 if and only if it evaluates to true. The total number of bits needed to specify the subtree is $\log_2 q + p[\log_2 2^{b-1} + b\log_2 s] \leq \log_2 q + pb(\log_2 s + 1)$ and the number of bits of x at the leaves is pb. □

Corollary B.6. *There is a function G in read-once depth 4 AC^0 such that G is in $\mathsf{NP}_\mathsf{k}^\mathsf{cc} - \mathsf{BPP}_\mathsf{k}^\mathsf{cc}$ for $k\log k \leq a\log n$ for some constant $a > 0$.*

Proof. Observe that $F = \textsc{Tribes}_{p,q} \circ (\psi_{k,\ell}^{\oplus b})^m$ with the parameters from Corollary B.4 by Lemma B.5 has $N^k(F)$ that is $O(\log^3 n)$ and thus satisfies all the conditions except for being read-once. To obtain the read-once property note that F is a projection of the following function G.

$$\bigvee_{u=1}^{q} \bigwedge_{v=1}^{p2^{p-1}} \bigvee_{w=1}^{ps} \bigwedge_{j=1}^{k} z_{j,u,v,w}$$

and that the same $O(\log^3 n)$ upper bound from Lemma B.5 applies equally well to G. □