Electronic Colloquium on Computational Complexity, Revision 2 of 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

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

May 6, 2010

Abstract

We prove an $n^{\Omega(1)}/4^k$ lower bound on the randomized k-party communication complexity of 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 $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuits, showing that the well-known quasipolynomial simulations of AC^0 by such circuits due to Allender (1989) and Yao (1990) are qualitatively optimal, even for formulas of small constant depth.

We also exhibit a depth 5 formula in $\mathsf{NP}^{\mathsf{cc}}_k - \mathsf{BPP}^{\mathsf{cc}}_k$ for k up to $\Theta(\log n)$ and derive $\Omega(2^{\sqrt{\log n}/\sqrt{k}})$ 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. We prove other strong results for depth 3 and 4 AC^0 functions.

1 Introduction

The number-on-forehead (NOF) multiparty communication complexity of AC^0 has been an open question since Håstad and Goldmann [13] 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, due to Allender and Yao, of AC^0 circuits [1] and 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. These protocols may even be 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 [21] showed that quasipolynomial size is required to simulate ACC^0 based on the

*Research supported by NSF grants CCF-0514870 and CCF-0830626
†Also known as Dang-Trinh Huynh-Ngoc. Research supported by NSF grants CCF-0514870 and CCF-0830626 and a Vietnam Education Foundation Fellowship

ISSN 1433-8092

result of Babai, Nisan, and Szegedy [4] that the Generalized Inner Product function in ACC^0 requires k-party NOF communication complexity $\Omega(n/4^k)$ which is polynomial in n for k up to $\Theta(\log n)$.

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, 5] 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. That changed with recent lower bounds for set disjointness by Lee and Shraibman [16] and Chattopadhyay and Ada [9] but no lower bounds apply for $\omega(\log\log n)$ players. As for circuit simulations of AC^0, Sherstov [23] 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 $n^{\Omega(1)}/4^k$ even for protocols with error exponentially close to $1/2$. For $\omega(1)$ 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, showing that quasipolynomial size is 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 depth 4 formulas. We further show that the same lower bound applies to a function having 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 for set disjointness, due to Lee and Shraibman [16] and Chattopadhyay and Ada [9], only apply for $k \leq \log\log n - o(\log\log n)$ players (though these bounds are stronger than ours for $o(\log\log n)$ 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 pattern matrix method, a general method to use analytic properties of Boolean functions to derive communication lower bounds for related Boolean functions [23, 25]. In [23], this analytic property was large threshold degree, and the resulting communication lower bounds yielded lower bounds for simulations of AC^0 by $\mathsf{MAJ}\circ\mathsf{MAJ}$ circuits. Sherstov [25] extended this to large approximate degree, yielding a strong new method for lower bounds for two-party randomized and quantum communication complexity.

Chattopadhyay [8] generalized [23] to pattern tensors for $k \geq 2$ players to yield the first lower bounds for the general NOF multiparty communication complexity of any AC^0 function for $k \geq 3$, implying exponential lower bounds for computation of AC^0 functions by $\mathsf{MAJ}\circ\mathsf{SYMM}\circ\mathsf{ANY}$ circuits

with $o(\log\log n)$ input fan-in – our results extend this to fan-in $\Omega(\log n)$. Lee and Schraibman [16] and Chattopadhyay and Ada [9] applied the full method in [25] to pattern tensors to yield the first lower bounds for the general NOF multiparty communication complexity of set disjointness for $k > 2$ players, improving on a long line of research on the problem [3, 28, 5, 30, 14, 6] and obtaining a lower bound of $\Omega(n^{\frac{1}{k+1}})/2^{2^{O(k)}}$. This yields a separation between randomized and nondeterministic k-party models for $k = o(\log\log n)$, which David, Pitassi, and Viola [11] improved to $\Omega(\log n)$ players for other functions based on pseudorandom generators. They asked whether there was a separation for $\Omega(\log n)$ players for AC^0 functions since their functions are 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 as described in [9, 24] is as follows. To prove k-party lower bounds for a function F, we first show that F has $f \circ \psi^m$ as a subfunction where ψ is a bit-selection function and f has large approximate degree. For such an f 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 $f \circ \psi^m$ is highly correlated with $g \circ \psi^m$ and, by the discrepancy method for communication complexity, it suffices to prove a discrepancy lower bound for $g \circ \psi^m$. Thanks to the orthogonality of g to all low degree polynomials this is possible using the bound in [4, 10, 20] derived from the iterated application of the Cauchy-Schwartz inequality. 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 a particular selector ψ and $f = \textsc{Or}$ which has approximate degree $\Omega(\sqrt{n})$.

In the two party case, Sherstov [26] and Razborov and Sherstov [22] extended the pattern matrix method to yield sign-rank lower bounds for some simple functions. A key idea for their arguments is the existence of orthogonalizing distributions μ for their functions that are "min-smooth" in that they assign at least some fixed positive probability to any x such that $f(x) = 1$.

By contrast we show that any function f for which approximating f within ϵ on only a subset S of inputs requires large degree, there is an orthogonalizing distribution μ for f that is "max-smooth" – the probability of subsets defined by partial assignments is never much larger than under the uniform distribution. The smoothness quality and the properties of the constrained subset S are determined by a function α so we call the degree bound the (ϵ, α)-approximate degree. We show that for any function this degree bound is large if there is a diverse collection of partial assignments ρ such that each subfunction $f|_\rho$ of f requires large approximate degree. This property is somewhat delicate, and does not hold for $\textsc{Or}$, but we are able to exhibit simple AC^0 functions with large (ϵ, α)-approximate degree.

Organization In Section 2 we review the relevant properties of correlation and its connection to multiparty communication complexity. We also describe a general form of the method of [25, 9, 11] based on selector functions and orthogonalizing distributions for functions of large ϵ-approximate degree and briefly discuss its limitations.

In Section 3 we introduce our new definition of (ϵ, α)-approximate degree and derive the additional "max-smoothness" property of the orthogonalizing distributions for functions of large (ϵ, α)-approximate degree. Using this additional max-smoothness property we derive our main technical theorem which gives communication complexity lower bounds based on the (ϵ, α)-degree lower bound and the properties of the selector function used.

In Section 4 we give a method for producing functions of large (ϵ, α)-approximate degree based on certain kinds of 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 an $\mathsf{AND} \circ \mathsf{OR}$ circuit has large (ϵ, α)-approximate degree for every $\epsilon < 1$. We use this function in our lower bounds for protocols having exponentially small advantage.

In Section 5, we introduce the $\textsc{Index}_{\oplus^a_{k-1}}$ selector function and combine it with the functions from Section 4 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 disjointness.

In Section 6 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 5. In Appendix A.1 we apply the lower bound from Section 3 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 $k = O(\sqrt{\log n})$. As part of this we also review earlier methods in more detail and which shows the value of moving from ϵ-approximate degree to (ϵ, α)-approximate degree. In Appendix A.2 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 and the generalized discrepancy/correlation method

Circuit complexity Let AND denote the class of all unbounded fan-in $\wedge$ functions (of literals), SYMM denote the class of all symmetric functions and $\mathsf{MAJ} \subset \mathsf{SYMM}$ denote the class of all majority functions. 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 Σ_1 formula if it is a clause and a Π_1 formula if it is a term. For $i \geq 1$, a Σ_{i+1} formula is an unbounded fan-in $\vee$ of Π_i formulas and a Π_{i+1} formula is an unbounded fan-in $\wedge$ of Σ_i formulas. The output gate of F is at the top and its inputs are at the bottom of the circuit. Given classes of functions $\mathsf{C}_1, \mathsf{C}_2, \ldots \mathsf{C}_d$, we let $\mathsf{C}_1 \circ \mathsf{C}_2 \circ \cdots \circ \mathsf{C}_d$ 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 C_i. Thus, for example, $\Pi_{i+1} = \mathsf{AND} \circ \Sigma_i$.

We will assume that Boolean functions on m bits are maps $f : \{0,1\}^m \to \{-1,1\}$.

Correlation Let μ be a distribution on $\{0,1\}^m$. The correlation between two real-valued functions f and g under μ is defined as $\mathrm{Cor}_\mu(f,g) := \mathbf{E}_{x\sim\mu}[f(x)g(x)]$. If $\mathcal{G}$ is a class of functions, the correlation between f and $\mathcal{G}$ under μ is defined as $\mathrm{Cor}_\mu(f,G) := \max_{g\in\mathcal{G}} \mathrm{Cor}_\mu(f,g)$.

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

Fact 2.1 (cf. [15]). *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, henceforth we 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, 20]). *Let $f : \{0,1\}^{m\times k} \to \mathbb{R}$ and $\mathcal{U}$ be the uniform distribution over $X \times Y$ where $Y = Y_1 \times \cdots \times Y_{k-1}$. Then,*

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

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 $0 \leq \epsilon < 1$, the ϵ-approximate degree of f, $deg_\epsilon(f)$, is the smallest d for which $||f - p||_\infty = \max_x |f(x) - p(x)| \leq \epsilon$ for some real-valued polynomial p of degree d. Following [19] 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 $thr(f) = deg_{<1}(f)$.

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.

The following Orthogonality-Approximation Lemma is the key to lower bounds using the pattern matrix (and pattern tensor) method. It is easily proved by duality of ℓ_1 and ℓ_∞ norms or by LP duality.

Lemma 2.4 ([25]). *If $f : \{0,1\}^m \to \{-1,1\}$ has $deg_\epsilon(f) \geq d$ 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 3 using more general LP duality.

The second major component of the pattern matrix/tensor method is the use of particular selector functions to provide inputs to functions f with large ϵ-approximate degree.

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

- There exist sets $D_{\psi,1}, \ldots, D_{\psi,(k-1)} \subseteq \{0,1\}^s$ such that for any $Y = (Y_1, \ldots, Y_{k-1}) \in D_\psi := D_{\psi,1} \times \cdots \times D_{\psi,(k-1)}$, $\Pr_{X \in \{0,1\}^s}[\psi(X,Y) = 0] = \Pr_{X \in \{0,1\}^s}[\psi(X,Y) = 1] = 1/2$.

Let $D_\psi^{(m)} := D_{\psi,1}^m \times \cdots \times D_{\psi,(k-1)}^m$. For any function $f : \{0,1\}^m \to \{1,-1\}$ and any selector function ψ we define a new function $f \circ \psi^m$ on $\{0,1\}^{kms}$ bits by, on any $x \in \{0,1\}^{ms}$ and $y = (y_1, \ldots, y_{k-1}) \in D_\psi^{(m)}$,

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

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 $x, y_1, \ldots, y_{k-1} \in \{0,1\}^{ms}$, 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 $f \circ \psi^m(x, y_1, \ldots, y_{k-1})$.

One example of a selector function ψ is the pattern tensor function $\psi_{k,\ell}$ used in [9, 16] which generalizes the pattern matrix function. In this example, $s = \ell^{k-1}$ and the s bits are arranged in a $(k-1)$-dimensional array indexed by $[\ell]^{k-1}$. $D_{\psi_{k,\ell},j}$ 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 such a $Y = (Y_1, \ldots, Y_{k-1}) \in \{0,1\}^{(k-1)s}$ the array $\wedge_{i=1}^{k-1} Y_i$ contains precisely one 1 which selects the bit of X to pass to f. This function is expressible by a small 2-level $\vee$ of $\wedge$s. As described in [11] the generalized discrepancy/correlation arguments work for any selector function that uses the inputs for players 1 to $k-1$ to select which bits from player 0's input to pass on to f, but we need our more general formulation for some examples we consider in Appendix A.2.

We give a brief overview of the remainder of the argument in [9, 11], which extends ideas of [23, 25] from 2-party to k-party communication complexity.

- Start with a Boolean function f on m bits having large $(1-\delta)$-approximate degree d.
- Apply the Orthogonality/Approximation Lemma to f to obtain a g that is $(1-\delta)$-correlated with f and a distribution μ under which g is not correlated with any low degree polynomial.
- Observe that from μ one can define a natural λ under which $g \circ \psi^m$ and $f \circ \psi^m$ have the same high correlation as g and f so to prove that $f \circ \psi^m$ is uncorrelated with low communication protocols, by the triangle inequality it suffices to prove this for $g \circ \psi^m$.
- The BNS-Chung bound/Gowers' norm used in Lemma 2.2 is based on the expectation of a function's correlation with itself on randomly chosen hypercubes of points. Use the orthogonality of g under μ to all polynomials of degree $< d$ to show that all low degree self-correlations of $g \circ \psi^m$ under λ disappear. The remaining high-degree self-correlations are bounded by analyzing overlaps in the choices of bits in different inputs among the hypercube of inputs. The argument repeatedly bounds the probability mass that μ assigns to small sub-cubes of the input by 1.
- The final lower bound is limited both by the upper bound on correlation in the high degree case and by the number of input bits required for each selector function.

Our argument follows this basic outline but improves it in two different ways. We first address the weakness of the upper bound on the high-degree self-correlations, which is implied by how little can be assumed about the orthogonalizing distribution μ given by Lemma 2.4. In particular, the

arguments in [25, 9, 16] all allow that μ may assign all of its probability mass to small subsets of points defined by partial assignments. Indeed, for the function OR_m, this is not far from tight. However, we will show that for other very simple functions 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 2.4 by considering a new measure that strengthens $(1-\delta)$-approximate degree. We also show that some simple functions require large values for our strengthened measure (which turns out to be fairly non-trivial to prove).

We also address the inefficiency of the pattern tensor selector function by defining a new selector function that requires many fewer bits. 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 $\Theta(\log n)$ players.

In the body of the paper we include our results containing both of these improvements. In Appendix A.1 we discuss certain other results that rely on the pattern tensor selector rather than our more efficient selector functions. This allows us to discuss more precisely how the addition of the max-smoothness property of the orthogonalizing distribution μ on its own already yields improved lower bounds without any change to the selector function.

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

We introduce (ϵ, α)-approximate degree and show how it implies our main technical theorem on the general correlation method.

A *restriction* is a $\rho \in \{0,1,*\}^m$, and we let $|\rho| = |\{i : \rho_i \neq *\}|$. Two restrictions π and ρ are *compatible*, $\pi \parallel \rho$, iff they agree on all non-star positions. Let $C_\rho = \{x \in \{0,1\}^m : x \parallel \rho\}$.

Definition Let $\alpha : \{0,\ldots,m\} \to \mathbb{R}$. Given a probability distribution λ on the set of restrictions $\{0,1,*\}^m$, we say that $x \in \{0,1\}^m$ is *α-light for λ* iff $\sum_{\rho \parallel x} 2^{|\rho| - \alpha(|\rho|)}\lambda(\rho) \leq 1$. Note that when $\alpha(r) = r$, every point is α-light for every distribution λ.

Definition Let $\alpha : \{0,\ldots,m\} \to \mathbb{R}$. The *$(\epsilon,\alpha)$-approximate degree*[1] of f, denoted as $deg_{\epsilon,\alpha}(f)$, is defined to be the minimum integer $d \geq 0$ such that there is some polynomial q of degree $\leq d$ and some probability distribution λ on restrictions such that for every $x \in \{0,1\}^m$ if x is α-light for λ then $|f(x) - q(x)| \leq \epsilon$. Note that this reduces to $deg_\epsilon(f)$ if $\alpha(r) \geq r$ for all r. Also define $deg_{<\epsilon,\alpha}(f) = \inf_{\epsilon'<\epsilon} deg_{\epsilon',\alpha}(f)$. As we write $thr(f) = deg_{<1}(f)$, we will usually say "α-threshold degree" for $(<1,\alpha)$-approximate degree.

This definition is an obvious weakening of the usual ℓ_∞ approximation of f since the non-light points can be ignored in the approximation. We will use this definition to prove our main technical theorem on the application of the general correlation method. To prove the theorem, we need the following lemma which generalizes Lemma 2.4 and is the first key to our substantially improved lower bounds.

[1] We use the same notation for a somewhat different and more general definition than that in earlier versions of this paper. First, α previously was a constant analogous to $\log_r \alpha(r)$ though this was not defined for all r. Second, the old definition was closer to that of a related quantity that we now call $deg^*_{\epsilon,\alpha}$ and define later.

Lemma 3.1 (Max-Smooth Orthogonality-Approximation Lemma)**.** *Let* $0 < \epsilon \leq 1$ *and* $\alpha : \{0, \ldots, m\} \to \mathbb{R}$*. If* $f : \{0,1\}^m \to \{-1,1\}$ *has* $deg_{<\epsilon,\alpha}(f) \geq d$*, then there exists a function* $g : \{0,1\}^m \to \{-1,1\}$ *and a distribution* μ *on* $\{0,1\}^m$ *such that:*

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

Proof. As in one of the proofs for Lemma 2.4, we write the requirements 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

$$
\begin{aligned}
y_S : &\quad \sum_{x\in\{0,1\}^m} h(x)\chi_S(x) = 0 \ : \ |S| < d \\
\beta : &\quad \sum_{x\in\{0,1\}^m} h(x)f(x) \geq \epsilon \\
v_x : &\quad \mu(x) - h(x) \geq 0 \ : \ x \in \{0,1\}^m \\
w_x : &\quad \mu(x) + h(x) \geq 0 \ : \ x \in \{0,1\}^m \\
\lambda_\rho : &\quad \eta - 2^{|\rho|-\alpha(|\rho|)} \sum_{x\in C_\rho} \mu(x) \geq 0 \ : \ \rho \in \{0,1,*\}^m \\
\gamma : &\quad \sum_{x\in\{0,1\}^m} \mu(x) = 1
\end{aligned}
$$

Suppose that we have optimum $\eta \leq 1$. 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 2.4 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 λ_ρ ensures that $\mu(C_\rho) \leq 2^{-|\rho|+\alpha(\rho|)}$ 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 λ_ρ but in this case if we increase η by a $1/||h||_1$ factor it will also be satisfied. Therefore, $\mu'(C_\rho) \leq 2^{-|\rho|+\alpha(|\rho|)}/\epsilon$.

Here is the dual LP:

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

$$\eta : \qquad \sum_{\rho \in \{0,1,*\}^m} \lambda_\rho = 1$$

$$\mu(x) : \qquad v_x + w_x + \gamma - \sum_{\rho \| x} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho = 0 : x \in \{0,1\}^m \tag{1}$$

$$h(x) : \qquad \beta f(x) + \sum_{|S|<d} y_S \chi_S(x) + w_x - v_x = 0 : x \in \{0,1\}^m \tag{2}$$

$$\beta, v_x, w_x, \lambda_\rho \geq 0 : x \in \{0,1\}^m$$

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 $< d$ and rewrite (2) as:

$$h(x) : \qquad \beta f(x) + p_d(x) + w_x - v_x = 0 : x \in \{0,1\}^m \tag{3}$$

Equations (1) and (3) for $x \in \{0,1\}^m$ together are equivalent to:

$$2w_x + \beta f(x) + p_d(x) + \gamma - \sum_{\rho \| x} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho = 0 \text{ and}$$
$$2v_x - \beta f(x) - p_d(x) + \gamma - \sum_{\rho \| x} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho = 0.$$

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

$$\beta f(x) + p_d(x) + \gamma \leq \sum_{\rho \| x} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho \text{ and}$$
$$-\beta f(x) - p_d(x) + \gamma \leq \sum_{\rho \| x} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho,$$

which together are equivalent to

$$|\beta f(x) + p_d(x)| + \gamma \leq \sum_{\rho \| x} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho.$$

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

$$\beta |f(x) - q_d(x)| + \gamma \leq \sum_{\rho \| x} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho$$

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

Now, let B be the set of points $x \in \{0,1\}^m$ 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

$$\beta |f(x) - q_d(x)| + \gamma \leq \sum_{\rho \| x} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho. \tag{4}$$

Let $R(x)$ denote the right-hand side of inequality (4). 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 λ. Since $deg_{<\epsilon,\alpha}(f)$ is larger than the degree of q_d, there must exist $x \in \{0,1\}^m$ that is both α-light for λ and $|f(x) - q_d(x)| \geq \epsilon$. Since $|f(x) - q_d(x)| \geq \epsilon$ we have $x \in B$ and since x is α-light for λ we have $R(x) \leq 1$ which is what we need to prove. □

Although the upper bound on $\mu(C_\rho)$ in Lemma 2.4 can be much larger than the $2^{-|\rho|}$ probability under the uniform distribution, we can use it to obtain an exponential improvement in the dependence of communication complexity lower bounds on k if $\alpha(r)$ is bounded below r^{α_0} for $r \geq d$ and $\alpha_0 < 1$. For any AC^0 function f we see that this assumption and the upper bound are essentially the best possible for d polynomial in m as follows:

By results of Linial, Mansour, and Nisan [17], for any AC^0 function f and constant $0 < \eta < 1$, there is a function p_d of degree $d < m^\eta$, 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 $1/c(m)$ is $\omega(|B_m|)$ then in the dual program $\mathcal{D}$, we could choose $\lambda_x = 1/|B_m|$ for $x \in B_m$ and $\lambda_\rho = 0$ for all other ρ and for these values β would be unbounded.

We now see how to apply Lemma 2.4 to obtain communication complexity lower bounds.

Definition Let ψ be a selector function with $D_\psi = D_{\psi,1} \times \cdots \times D_{\psi,(k-1)}$. For fixed $y^0, y^1 \in D_\psi^{(m)}$, $i \in [m]$ and uniformly random x_i, we call i *good for* (y^0, y^1) if the set of 2^{k-1} random variables $z_i^u = \psi(x_i, y_{*i}^u)$ for $u \in \{0,1\}^{k-1}$ are mutually independent, where y^u is defined as in Lemma 2.2; 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 $r_\psi(y^0, y^1) = |R_\psi(y^0, y^1)|$.

Now we are ready to state the main technical consequence of the Max-Smooth Orthogonality-Approximation Lemma. A similar version with $\alpha(r) = r$ follows from earlier work but the ability to have $\alpha(r) < r^{\alpha_0}$ for large r yields exponentially better lower bounds than in previous work.

Theorem 3.2. *Let $\alpha : \{0, \ldots, m\} \to \mathbb{R}$. If a function $f : \{0,1\}^m \to \{1,-1\}$ has $deg_{<1-\epsilon,\alpha}(f) \geq d$ and ψ is a selector function on $\{0,1\}^{ks}$ with $D_\psi = D_{\psi,1} \times \cdots \times D_{\psi,(k-1)}$ then*

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

Proof. The pattern of the argument follows the outline from Section 2. We first apply Lemma 3.1 to f to produce function g and distribution μ. By construction $\mathrm{Cor}_\mu(f, g) \geq 1 - \epsilon$. Then we define a distribution λ on $\{0,1\}^{mks}$ 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 $\mathrm{Cor}_\lambda(f \circ \psi^m, \Pi_k^c) \leq 2\epsilon$.

Since ψ is a selector function, each $z_i = \psi(x_i, y_{*i})$ is a uniformly random bit for each fixed $y_{*i} \in D_\psi$ and random x_i. We therefore have $\mathrm{Cor}_\lambda(f \circ \psi^m, g \circ \psi^m) = \mathrm{Cor}_\mu(f, g) \geq 1 - \epsilon$, hence $\mathrm{Cor}_\lambda(f \circ \psi^m, \Pi_k^c) \leq \epsilon + \mathrm{Cor}_\lambda(g \circ \psi^m, \Pi_k^c)$ by the triangle inequality. It therefore suffices to show that $\mathrm{Cor}_\lambda(g \circ \psi^m, \Pi_k^c) \leq \epsilon$.

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

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

where $H(y^0, y^1)$ is the self-correlation in the hypercube defined by y^0 and y^1:

$$H(y^0, y^1) := \Big| \mathbf{E}_x \big[\prod_{u \in \{0,1\}^{k-1}} \mu(z_1^u, \ldots, z_m^u) g(z_1^u, \ldots, z_m^u) \big] \Big|,$$

where $z_i^u = \psi(x_i, y_{*i}^u)$. We now compute bounds on the self-correlation $H(y^0, y^1)$ that depend on the value of $r = r_\psi(y^0, y^1)$. The first bound is from [9] and is the key to the original method.

Proposition 3.3. *If $r = r_\psi(y^0, y^1) < d$, then $H(y^0, y^1) = 0$.*

Proof. Taking x uniformly at random, 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 construction, 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 $\mathcal{Z}^{\neq 0\ldots0} | z^{0\ldots0}$ on $\{z^u\}_{u \neq 0\ldots0}$ which is derived from $\mathcal{Z}$ by conditioning on $\mathcal{Z}^{0\ldots0} = z^{0\ldots0}$. Then,

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

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 conditioning of $\mathcal{Z}^{\neq 0\ldots0}$ 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

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

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

$$H(y^0, y^1) = \Big| \mathbf{E}_{z^{0\ldots0}} \big[\mu(z^{0\ldots0}) g(z^{0\ldots0}) Q(z^{0\ldots0}) \big] \Big| = 0$$

by the orthogonality property of μ and g since $r < d$. □

The following bound for $r = r_\psi(y^0, y^1) \geq d$ is the key to the sharper bound that yields our exponentially better results. A weaker version in [9] applies only when $\alpha(r) = r$ (but does not have the $\epsilon^{2^{k-1}-1}$ in the denominator).

Lemma 3.4. $H(y^0, y^1) \leq \dfrac{2^{(2^{k-1}-1)\alpha(r)}}{2^{2^{k-1}m}\epsilon^{2^{k-1}-1}}$.

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-valued,

$$\begin{aligned}
H(y^0, y^1) &= \Big|\mathbf{E}_x\big[\prod_{u\in\{0,1\}^{k-1}} \mu(z^u)g(z^u)\big]\Big| \\
&\leq \mathbf{E}_x\Big|\prod_{u\in\{0,1\}^{k-1}} \mu(z^u)g(z^u)\Big| \\
&= \mathbf{E}_x\big[\prod_{u\in\{0,1\}^{k-1}} \mu(z^u)\big] \\
&\leq \mathbf{E}_x[\mu(z^{0\ldots0})] \\
&\quad \times \max_{x_1,\ldots,x_r} \mathbf{E}_{x_{r+1},\ldots,x_m}\big[\prod_{u\neq 0\ldots0} \mu(z^u)\big] \\
&= \mathbf{E}_x[\mu(z^{0\ldots0})] \qquad (5) \\
&\quad \times \max_{x_1,\ldots,x_r} \prod_{u\neq 0\ldots0} \mathbf{E}_{x_{r+1}\ldots x_m}\big[\mu(z^u)\big] \qquad (6)
\end{aligned}$$

where $z^u_i = \psi(x_i, y^u_{*i})$ for all $i \in [m]$.

We first consider line (5). 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 $\{0,1\}^m$. 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 $z^u_1, \ldots, z^u_r$ and by our assumption on ψ, for random $x_{r+1}, \ldots, x_m$ the values of $z^u_{r+1}, \ldots, z^u_m$ are uniformly random. Therefore the value in line (6) is upper bounded by

$$\prod_{u\neq 0\ldots0} \max_{z^u_1,\ldots,z^u_r} \mathbf{E}_{z^u_{r+1}\ldots z^u_m}\big[\mu(z^u)\big] = \big(\max_{z_1,\ldots,z_r} \mathbf{E}_{z_{r+1}\ldots z_m}\big[\mu(z)\big]\big)^{2^{k-1}-1}.$$

By the property of μ implied by Lemma 3.1,

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

and therefore line (6) is at most $(2^{\alpha(r)-r}/(\epsilon 2^{m-r}))^{2^{k-1}-1} = (2^{\alpha(r)-m}/\epsilon)^{2^{k-1}-1}$. (This is the one place where we use the max-smoothness property of the distribution μ.) The lemma follows immediately by combining the bounds for lines (5) and (6). □

Plugging in the bounds of Proposition 3.3 and Lemma 3.4 we obtain that

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

Taking 2^{k-1}-st roots and using Fact 2.1 we obtain that $R^k_{1/2-\epsilon}(f\circ\psi^m)\ge c$ if

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

Rewriting and taking logarithms yields the claimed bound of Theorem 3.2. □

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

Given $\epsilon<1$ and α, it is not obvious that any function, let alone a function in AC^0, has large (ϵ,α)-approximate degree. This section shows that AC^0 does contain functions with large $(5/6,\alpha)$-approximate degree and functions with large α-threshold degree where $\alpha(z)\le z^{\alpha_0}$ for $\alpha_0<1$ and all large z.

We first reduce this new notion of approximate degree to a more tractable notion, which is only large if many widely distributed restrictions of f also require large approximate degree. Given a function f on $\{0,1\}^m$ and a restriction ρ, we define $f|_\rho$ on $\{0,1\}^{m-|\rho|}$ in the natural way. We also define $\mathcal{R}^r_m:=\{\rho\in\{0,1,*\}^m:|\rho|=m-r\}$.

Definition Given $\alpha:\{0,\dots,m\}\to\mathbb{R}$, we say that a probability distribution ν on $\{0,1,*\}^m$ is *α-spread* iff for every restriction $\rho\in\{0,1,*\}^m$, $\Pr_{\pi\sim\nu}[\pi\parallel\rho]\le 2^{\alpha(|\rho|)-|\rho|}$. Let $deg^*_{\epsilon,\alpha}(f)$ be the minimum d such that for any α-spread distribution ν on $\{0,1,*\}^m$, there is some π with $\nu(\pi)>0$ and $deg_\epsilon(f|_\pi)\le d$. Note that for $\alpha(r)=r$, $deg_\epsilon(f)=deg^*_{\epsilon,\alpha}(f)$ since every distribution on restrictions is α-spread. We define $deg^*_{<\epsilon,\alpha}(f)=\min_{\epsilon'<\epsilon}deg^*_{\epsilon',\alpha}(f)$.

Given the following lemma, to show that $deg_{\epsilon,\alpha}(f)$ is large, it suffices to show that $deg^*_{\epsilon,\alpha}(f)$ is large.

Lemma 4.1. *Let $f:\{0,1\}^m\to\{-1,1\}$ and $\alpha:\{0,\dots,m\}\to\mathbb{R}$. For $0<\epsilon\le 1$, $deg_{\epsilon,\alpha}(f)\ge deg^*_{\epsilon,\alpha}(f)$.*

Proof. Suppose, by contradiction, that for some d, (i) $deg^*_{\epsilon,\alpha}(f)>d$, and (ii) $deg_{\epsilon,\alpha}(f)=d$. Then by definition, (i') there exists an α-spread distribution ν on $\{0,1,*\}^m$ such that $deg_\epsilon(f|_\pi)>d$ for every π with $\nu(\pi)>0$, and (ii') there exists a polynomial q of degree $\le d$ and a distribution λ on $\{0,1,*\}^m$ such that $R(x)=\sum_{\rho\parallel x}2^{|\rho|-\alpha(|\rho|)}\lambda_\rho>1$ whenever $x\in B'$, where $B'=\{x:|f(x)-q(x)|>\epsilon\}$.

Choosing $\pi\sim\nu$, we define the random variable

$$I_\pi:=\sum_{\rho\parallel\pi}2^{|\rho|-\alpha(|\rho|)}\lambda_\rho.$$

Then,

$$\mathbf{E}_{\pi\sim\nu}(I_\pi)=\sum_\rho\Pr_{\pi\sim\nu}[\rho\parallel\pi]\cdot 2^{|\rho|-\alpha(|\rho|)}\lambda_\rho\le\sum_\rho 2^{\alpha(|\rho|)-|\rho|}\cdot 2^{|\rho|-\alpha(|\rho|)}\lambda_\rho\le 1.$$

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

$$R(x) = \sum_{\rho \| x} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho > 1,$$

we would have $I_\pi > 1$. Thus $C_\pi \cap B' = \emptyset$. So for any $x \in C_\pi$, we have $|f(x) - q(x)| \leq \epsilon$. But since the degree of q_d is $\leq d$ this contradicts the fact that $deg_\epsilon(f|_\pi) > d$. The lemma follows. □

For the rest of this section we always take $\alpha(z) \leq z^{\alpha_0}$ for some $\alpha_0 < 1$ for large enough z and $\alpha(z) = z$ otherwise. By definition, to show that $deg^*_{\epsilon,\alpha}(f)$ is large, we need to exhibit an α-spread distribution ν such that for any restriction ρ with $\nu(\rho) > 0$, $deg_\epsilon(f|_\rho)$ is large. An obvious choice for such ν is the uniform distribution on $\mathcal{R}^r_m$ where $r \approx m^{\alpha_0}$. Indeed, it is not hard to show with this distribution that the parity function has large (ϵ, α)-approximate degree. However this simple ν cannot be used for AC^0 circuits since these circuits shrink rapidly under such restrictions. Thus in Lemma 4.2 we define a more involved α-spread family of restrictions. With this family, we give a generic construction that takes a circuit G on q bits and produces another circuit H on $m = pq$ bits such that for any restriction π in the family, $H|_\pi$ contains the *projection* of G on some set S of r bits – a new function denoted by G_S and obtained from G by keeping only those nodes on paths from the inputs in S to the output gate – as a subfunction. If each such projection of G has ϵ-approximate degree $r^{\Omega(1)}$ and if p is $O(\log q)$ and r is polynomial in q and hence in $m = pq$, then we derive that H has (ϵ, α)-approximate degree $m^{\Omega(1)}$.

Lemma 4.2. *Let q, r, p, and w be integers with $q > r > p \geq 2$ and let $1 > \alpha_0 > \beta > 0$ be such that $q^\beta \geq rp$, $2^{p-1} - 1 \geq q^{1-\beta}$, $q^{\alpha_0} \geq \frac{6}{\ln 2} 2^p r$, and $w^{\alpha_0 - \beta} \geq 3p/\ln 2$. Fix any partition of a set of $m = pq$ bits into q blocks of p bits each. Define distribution ν on $\mathcal{R}^{pr}_{pq}$ as follows: choose a subset of $q - r$ blocks uniformly at random; then assign values to the variables in each of these blocks uniformly at random from $\{0,1\}^p - \{0^p, 1^p\}$. Then for any $\rho \in \{0,1,*\}^m$ with $|\rho| \geq w$, we have $\Pr_{\pi \sim \nu}[\rho \parallel \pi] \leq 2^{|\rho|^{\alpha_0} - |\rho|}$.*

The proof of Lemma 4.2 is surprisingly involved and requires quite precise tail bounds. We defer the proof to Section 7. Intuitively, we need the parameter choices given here because the conclusion requires that, in an amortized sense, each bit assigned by ρ contributes not much more than $1/2$ to the probability of being consistent with a random $\pi \sim \nu$. Hence, in our amortized sense the p bits in any one of the q terms should not contribute much more than a 2^{-p} factor to the probability of being consistent. However, ρ and π are consistent in any term that is not selected by π which happens for any fixed term with probability r/q. It is therefore necessary in our argument that r/q not be much larger than 2^{-p}.

For $\epsilon = 5/6$, a simple candidate for G is $G = \textsc{Or}_q$. With this G and the family of restrictions given by Lemma 4.2, the next lemma constructs $H = \textsc{Tribes}_{p,q}$ that has large $(5/6, \alpha)$-approximate degree. Recall that $\textsc{Tribes}_{p,q}(x) = \vee_{i=1}^q \wedge_{j=1}^p x_{i,j}$.

Lemma 4.3. *Given any constants $0 < \epsilon, \alpha_0, \beta < 1$ with $\beta > 1 - \epsilon$ and $\alpha_0 - \beta \geq 0.1$. Let $q > p \geq 2$ be integers such that $2\lceil q^{1-\beta} \rceil < 2^p \leq \frac{1}{6} q^{\alpha_0 + \epsilon - 1} \ln 2$. Define $\alpha(z) = z^{\alpha_0}$ for $z^{\alpha_0 - \beta} \geq 3p/\ln 2$ and $\alpha(z) = z$ otherwise. Then for large enough q, we have $deg_{5/6,\alpha}(\textsc{Tribes}_{p,q}) \geq \sqrt{q^{1-\epsilon}/12}$.*

Proof. Define the distribution ν as in the statement of Lemma 4.2, where a p-block corresponds to a p-term in $\textsc{Tribes}_{p,q}$, by applying this lemma with $r := \lceil q^{1-\epsilon} \rceil$ and $w = (3p/\ln 2)^{1/(\alpha_0 - \beta)}$. For q large enough,

$$q^\beta / r \geq q^{\beta + \epsilon - 1} > \log q > p, \text{ and } w^{\alpha_0 - \beta} \geq 3p/\ln 2.$$

For any π with $\nu(\pi) > 0$, OR_r is a subfunction of $\text{TRIBES}_{p,q}|_\pi$ so $deg_{5/6}(\text{TRIBES}_{p,q}|_\pi) \geq deg_{5/6}(\text{OR}_r) \geq \sqrt{r/12}$. Thus, $deg_{5/6,\alpha}(\text{TRIBES}_{p,q}) \geq deg^*_{5/6,\alpha}(\text{TRIBES}_{p,q}) \geq \sqrt{r/12}$. □

In particular, with $\epsilon = 0.4, \beta = 0.8, \alpha_0 = 0.9$, we get:

Corollary 4.4. *For sufficiently large p and $q = 2^{4p}$, if $\alpha : \{0, \ldots, m\} \to \mathbb{R}$ satisfies $\alpha(z) = z^{0.9}$ for $r \geq (3p \ln 2)^{10}$ and $\alpha(z) = z$ otherwise, then $deg_{5/6,\alpha}(\text{TRIBES}_{p,q}) \geq q^{3/10}/\sqrt{12} = 2^{6p/5}/\sqrt{12}$.*

Corollary 4.4 suffices for most of our communication complexity lower bounds. However our results for threshold circuit size require a function in AC^0 having large α-threshold degree. In the rest of this section we show such a function, whose construction involves more complex G and H.

We first construct, in Lemma 4.6, a circuit G on q bits that has large threshold degree when projected on any r bits for sufficiently large r. The lemma uses the following property of the threshold degree of (the dual of) the Minsky-Papert function shown in [18].

Proposition 4.5. *Let $\text{MP'}_{q,q'} : \{0,1\}^{q \cdot q'} \to \{-1, 1\}$ be defined by $\text{MP'}_{q,q'}(x) := \wedge_{i=1}^{q} \vee_{j=1}^{q'} x_{ij}$. Let $m > 0$ be such that $m \leq q$ and $4m^2 \leq q'$. Then $thr(\text{MP'}_{q,q'}) \geq m$.*

Lemma 4.6. *Let q, r, d, s and t be positive integers such that $q = st, q \geq r \geq 2ds$, and $s/(4d) \geq t \geq 2d$. 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 4.5, $thr(G_S) \geq thr(\text{MP}_{d,4d^2}) \geq d$. □

Given G, we now construct the circuit H of large α-threshold degree. We could define this for any AC^0 circuit G but we restrict ourselves to the $\mathsf{AND} \circ \mathsf{OR}$ formula given by Lemma 4.6. 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 4.6, 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 4.7. *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 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 AND_p gate in H is forced to 0. Also observe that for any OR gate φ' in H corresponding to a bottom level 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.

Lemma 4.8. *For any p sufficiently large multiple of 15 and $q = 2^{4p}$, if $\alpha : \{0, \ldots, m\} \to \mathbb{R}$ is defined as $\alpha(z) = z^{0.9}$ for $r \geq (3p \ln 2)^{10}$ and $\alpha(z) = z$ otherwise, then there is an explicit depth 4 AC^0 function on pq bits that has α-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 4.6. 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 4.2 given the value of r and $w = \lceil \log^{20} pq \rceil$. We can then apply Lemma 4.7 to G to derive the $\mathsf{\Pi_4}$ 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 $thr(H|_\pi) \geq thr(G_S) \geq d$.

Note that for $\alpha_0 = 0.9$ and $\beta = 0.8$, all the conditions of Lemma 4.2 are satisfied. In particular, for p sufficiently large, $q^\beta = q^{0.8} \geq q^{3/5} \log_2 q = rp$, $2^{p-1} - 1 = q^{1/4}/2 - 1 \geq q^{0.2} = q^{1-\beta}$, $q^{\alpha_0} = q^{0.9} \geq \frac{24}{\ln 2} q^{17/20} = \frac{6}{\ln 2} 2^p r$, and $w^{\alpha_0 - \beta} \geq \log^2 q \geq 3p/\ln 2$.

It follows that H has α-threshold degree at least d as required. □

5 Multiparty communication complexity lower bounds for AC^0

Together with the functions from the previous section, Theorem 3.2 is sufficient to improve the lower bounds for AC^0 functions based on pattern tensor selectors from $O(\log \log n)$ players to $\Omega(\sqrt{\log n})$ players. These results, which show the power of our introduction of (ϵ, α)-approximate degree on its own, are described in Appendix A.1. We need one more ingredient to obtain our strongest lower bounds, namely, a different selector function ψ, which we denote by $\text{INDEX}_{\oplus^a_{k-1}}$ where $a > 0$ is an integer. This function has $s = 2^a$ and $D_{\text{INDEX}_{\oplus^a_{k-1}}, j} = \{0,1\}^s$ for all j. For $X \in \{0,1\}^s$ and $Y \in \{0,1\}^{(k-1)s}$ define

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

where the bits in X are indexed by a-bit vectors and $Y_{[a]}$ denotes 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 parity, the number of players k will be $O(\log n)$ and hence it is computable in AC^0. We can either write $\textsc{Index}_{\oplus^a_{k-1}}$ as an $\vee \circ \wedge \circ \vee \circ \wedge$ formula where the fan-ins are 2^a, $a+1$, 2^{k-2}, and $k-1$, respectively, or as an $\vee \circ \wedge \circ \vee$ formula where the fan-ins are 2^a, $a2^{k-2}+1$, and $k-1$, respectively.

With $\psi = \textsc{Index}_{\oplus^a_{k-1}}$, the variables $z^u_i = \textsc{Index}_{\oplus^a_{k-1}}(x_i, y^u_{*i})$ for $u \in \{0,1\}^{k-1}$ are independent iff for every $u \neq v$, y^u_{*i} and y^v_{*i} select different bits of x_i.

Lemma 5.1. *If $\psi = \textsc{Index}_{\oplus^a_{k-1}}$ then*

$$\Pr_{y^0,y^1 \in D^{(m)}_\psi}[r_\psi(y^0,y^1) = r] \leq \binom{m}{r} 2^{(2k-a-3)r} \leq \big(\frac{em2^{2k-a-3}}{r}\big)^r.$$

Proof. In this case $D^{(m)}_\psi$ 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 $i \in [m]$. □

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

Theorem 5.2. *Let $\alpha : \{0,\ldots,m\} \to \mathbb{R}$ and $0 < \alpha_0 < 1$. For any Boolean function f on m bits such that $deg_{1-\epsilon,\alpha}(f) \geq d$ and $\alpha(r) \leq r^{\alpha_0}$ for all $r \geq d$, the function $f \circ \textsc{Index}^m_{\oplus^a_{k-1}}$ defined on nk bits, where $n = ms$ and $s = 2^a \geq e2^{2k-1}m/d$, requires that $R^k_{1/2-\epsilon}(f \circ \textsc{Index}^m_{\oplus^a_{k-1}}) \geq d/2^k + \log_2(\epsilon(1-\epsilon))$ for $k \leq (1-\alpha_0)\log_2 d$.*

Proof. For $\psi = \textsc{Index}_{\oplus^a_{k-1}}$, by Lemma 5.1,

$$\sum_{r=d}^{m} 2^{(2^{k-1}-1)\alpha(r)} \cdot \Pr_{y^0,y^1 \in D^{(m)}_\psi}[r_\psi(y^0,y^1) = r] \leq \sum_{r=d}^{m} 2^{(2^{k-1}-1)\alpha(r)} \cdot \big(\frac{em2^{2k-a-3}}{r}\big)^r \tag{7}$$

Since $k \leq (1-\alpha_0)\log_2 d$, we have $(2^{k-1}-1)\alpha(r) < d^{1-\alpha_0}\alpha(r) \leq r$ for $r \geq d$ so (7) is

$$\leq \sum_{r=d}^{m} \big(\frac{em2^{2k-a-2}}{r}\big)^r \leq \sum_{r=d}^{m} 2^{-r} < 2^{-(d-1)} \qquad \text{for } 2^a \geq e2^{2k-1}m/d.$$

Plugging this into Theorem 3.2 we obtain that

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

as required. □

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. By applying the above theorem for $f = \textsc{Tribes}_{p,q}$ and $f = \textsc{Tribes}'_{p,q}$, we obtain the following result.

Theorem 5.3. *Let p be a sufficiently large integer and $q = 2^{4p}$, $k \leq p/10$, 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 $\Omega(q^{0.3}/2^k)$ which is $n^{\Omega(1)}/4^k$. Furthermore, F has polynomial-size depth 5 $\mathsf{AC^0}$ formulas and F' has polynomial-size depth 4 $\mathsf{AC^0}$ formulas.*

Proof. Let $\epsilon = 0.4$, $\alpha_0 = 0.9$, 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 $\mathsf{\Sigma_3}$ 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 $\mathsf{\Sigma_2}$ formula, $\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ is computable by a $\mathsf{\Sigma_5}$ 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 $\mathsf{AC^0}$ formulas. □

Lemma 5.4. *$N^k(\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^a_{k-1}})$ is $O(\log q + pa)$.*

Proof. Using the $\mathsf{\Sigma_3}$ 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 $\mathsf{\Sigma_5}$ 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. □

Corollary 5.5. *There is a function G in depth 5 $\mathsf{AC^0}$ such that G is in $\mathsf{NP^{cc}_k - BPP^{cc}_k}$ for $k \leq a' \log n$ for some constant $a' > 0$.*

Proof. Observe that, by Lemma 5.4, $F = \textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ with the parameters from Theorem 5.3 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 5.4 applies equally well to G. □

Applying distributive law to the depth 5 function $f = \textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ we derive the following exponential improvement in the number of players for which non-trivial lower bounds can

be shown for $\textsc{Disj}_{k,n}$. (The same lower bound for disjointness can be derived even more simply using the above technique for the simpler function $\textsc{Tribes}_{p,q} \circ \psi^m_{k,\ell}$ using the pattern tensor selector analyzed in Appendix A.1.)

Theorem 5.6. *$R^k_{1/3}(\textsc{Disj}_{n,k})$ 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 5.5 start with $F = \textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ with the parameters from from Theorem 5.3. Unlike Corollary 5.5, however, we use the Σ_4 circuit for $\textsc{Index}_{\oplus^{(p+2k)}_{k-1}}$ and reduce F to a Σ_6 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

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

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

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

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 $\textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ 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 $k \le p/10$ 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 $ap \ge k - 2 + 4p$ and hence we have that $n \le 2^{apk} \le 2^{6p^2k/5}$. Therefore we have $p \ge \sqrt{5\log_2 n}/\sqrt{6k}$. It follows that $R^k_{1/3}(\textsc{Disj}_{n,k})$ is $\Omega(2^{\sqrt{\log_2 n}/\sqrt{k}})$ provided that $k \le \frac{1}{10}\sqrt{5\log_2 n}/\sqrt{6k}$ which holds if $k \le \frac{1}{5}\log_2^{1/3} n$. □

Although our bound for $\textsc{Disj}_{n,k}$ applies to exponentially more players than do the bounds in [16, 9], the previous bounds are stronger for $k \leq \log\log n - o(\log\log n)$ players.

Corollary 5.7. *There is a depth-2* AC^0 *formula in* $\mathsf{NP}^{cc}_k - \mathsf{BPP}^{cc}_k$ *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 3.1 to Or since the constant function 1 approximates Or on all but one point.

To prove lower bounds for $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuits we need lower bounds on protocols that succeed with probability barely better than that of random guessing. Using the function with large α-threshold degree given by Lemma 4.8 in place of $\textsc{Tribes}_{p,q}$ we obtain the following theorem.

Theorem 5.8. *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 Lemma 4.8. 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 $a = \lceil \log_2(e 2^{2k-1} m/d) \rceil$, and $s = 2^a$. By Theorem 5.2, 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 Π_3 circuit for $\textsc{Index}_{\oplus^a_{k-1}}$ with that for f yields depth 6. □

6 Threshold circuit lower bounds for AC^0

Following the approach of Viola [29], which extends the ideas of Razborov and Wigderson [21], we show quasipolynomial lower bounds on the simulation of AC^0 functions by unrestricted $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuits.

Theorem 6.1. *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* $N^{\Omega(\log N)}$.

In the rest of this section we prove the above theorem.[2] An important ingredient in our construction is the following function F^m_t, first defined by Sipser [27].

Definition For $t \geq 2$, the Sipser function F^m_t is defined by a depth t read-once circuit. The root of this circuit is an Or gate with fan-in $\frac{1}{2}(m \log m)^{1/4}$. Below are alternating levels of And and Or gates with fan-in m. The bottom level has fan-in $\sqrt{\frac{1}{2} t m \log m}$. Therefore for t constant and large enough m, F^m_t is an AC^0 function on $O(m^t)$ inputs.

[2] This theorem is a stronger version of the one proved in earlier versions of this paper, which only gave a size lower bound of $N^{\Omega(\log\log N)}$.

The circuit defining F_t^m partitions the input into $B = \{B_i\}_{i=1}^r$, for some r, where each block B_i consists of all variables that are fed to the same bottom gate. If $\mathcal{R}$ is a distribution of restrictions of the variables in the same block, we define $\mathcal{R}^B := (\mathcal{R})^r$, which is a distribution of restrictions of all variables. Håstad showed that there exists a distribution with the following useful property.

Proposition 6.2. *[12] Let $0 \leq q \leq 1$ be a real number, F be some function, and $B = \{B_i\}$ be any partition of the input of F to equal-size blocks. There exists a distribution $\mathcal{R}_q$ of restrictions in each block B_i such that the following holds.*

- *If F is a CNF with clause size at most w, and $s > 0$, then with probability at least $1 - (6qw)^s$ over the choice of $\rho \sim \mathcal{R}_q^B$, $F|_\rho$ can be written as a DNF with term size at most s, and moreover, any input assignment satisfies at most one of these terms.*
- *For any odd constant $t \geq 3$ and large enough m, if $F = F_t^m$, $q = (\frac{2t}{m} \log m)^{1/2}$, and B is the partition of the input to F_t^m as mentioned above, then with probability at least $2/3$ over the choice of $\rho \sim \mathcal{R}_q^B$, $F|_\rho$ contains F_{t-1}^m as a subfunction.*

We note that in the above proposition, in the first item, the property that any input assignment satisfies at most one of the terms is implicit in [12]. The observation that this property holds is made explicit by Berg and Ulfberg [7].

The proof of our theorem also relies on the following connection between multiparty communication complexity and threshold circuit complexity given by Håstad and Goldmann.

Proposition 6.3. *[13]*

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

(b) *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 are now ready to prove our theorem.

Proof of Theorem 6.1. We first give a brief overview of the proof: We use the function H_n from Theorem 5.8 and replace each input by a $\oplus$ of $\Theta(\log n)$ disjoint copies of F_3^n to obtain an AC^0 function G on $N = O(n^4 \log n)$ inputs. If G is computed by such a circuit C of size $N^{o(\log N)}$ then using suitable random restrictions as described in Proposition 6.2, we can ensure that all bottom-level AND gates of C are reduced to fan-in at most $\delta \log_2 N$ and at the same time that every $\oplus$ block of inputs in G is still nontrivial. Applying Proposition 6.3 yields a contradiction to Theorem 5.8.

More precisely, let c, c' be the constants and H_n be the function given by Theorem 5.8. Let $k = \lfloor c' \log_2 n \rfloor$, $r = \lceil \log_3 2n \rceil$, and $m' < n^3$ be the input size of F_3^n. For any $Z = Z_1 \cdots Z_n$, where each $Z_i \in \{0,1\}^{r \times m'}$, we define our hard function $G_N : \{0,1\}^N \to \{0,1\}$ as

$$G_N(Z) := H_n\big(A_1, \ldots, A_n\big), \text{ where each } A_i := \bigoplus_{j=1}^{r} F_3^n(Z_{i,j}).$$

Suppose by contradiction that for some sufficiently small constant $\delta > 0$, there is a $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuit C of size $N^{\delta \log N}$ that computes G_N. Let B be the partition of the input of G_N that is the union of all the input partition of the F_3^n functions. Let $q = (\frac{6 \log n}{n})^{1/2}$ and $\mathcal{R}_q^B$ be the distribution as described in Proposition 6.2. Let $\rho \sim \mathcal{R}_q^B$. Consider the following two events:

- Event E_1: $C|_\rho$ is computed by a $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuit of size at most $|C| \cdot (2N)^k$ where the fan-in of each AND-gate is strictly less than k, and

- Event E_2: A_i contains F_2^n as a subfunction for all $1 \le i \le n$.

First, we show that $\Pr[\neg E_1] < 1/2$ for sufficiently small $\delta > 0$: Fix any AND-gate φ in C. By Proposition 6.2, the probability over ρ that $\varphi|_\rho$ cannot be written as a DNF with term size less than k is at most

$$(6q)^k < (\frac{216 \log n}{n})^{k/2}.$$

This quantity is $N^{-\Omega(\log N)}$. Thus by a union bound over all AND-gates in C and for sufficiently small δ, with probability strictly less than $1/2$, every AND gate in $C|_\rho$ can be written as a DNF with term size less than k. Any such DNF has size at most $(2N)^k$. Also by Proposition 6.2, in any such DNF, no two terms can be satisfied at the same time. Thus we can merge each of these terms into the next-level symmetric gate and conclude that the function computed by $C|_\rho$ is computed by a $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuit of size at most $|C| \cdot (2N)^k$ where the fan-in of each AND gate is strictly less than k.

Next, we show that $\Pr[\neg E_2] < 1/2$. By Proposition 6.2, with probability at least $1 - (1/3)^r > 1 - 1/(2n)$, each A_i contain F_2^n as a subfunction. By union bound over all $i \in [n]$, we conclude that $\Pr[\neg E_2] < 1/2$.

Hence there exists a restriction ρ such that both E_1 and E_2 hold. By Proposition 6.3, the fact that E_1 holds implies that for any partition of the input to k players and $\epsilon = 1/(2|C| \cdot (2N)^k)$, $R^k_{1/2-\epsilon}(C|_\rho)$ is $O(k \log(|C| \cdot (2N)^k)) = O(\log^3 N) = O(\log^3 n)$. On the other hand, the fact that E_2 holds implies that $C|_\rho$ computes H_n as a subfunction. By Theorem 5.8, 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 $O(\log^2 N) = O(\log^2 n)$, $\Omega(n^c + \log \epsilon)$ is $\Omega(n^c)$ for sufficiently large N (and hence n), we arrive at a contradiction. □

Remark Although the proof for Theorem 6.1 uses the second part of Proposition 6.3 and the function given by Theorem 5.8, the same proof that instead uses the first part of the proposition and the simpler function given by Theorem 5.3 would yield a simpler AC^0 function that requires quasipolynomial size to be simulated by $\mathsf{SYMM} \circ \mathsf{AND}$ circuits.

7 Proof of Lemma 4.2

Proof. Fix any restriction ρ of size $i = |\rho| \ge w$. 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

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

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,

$$\Pr_{\pi \sim \nu}[C_\rho \cap C_\pi \neq \emptyset] \quad < \quad \frac{1}{\binom{q}{q-r}} \sum_{S \subset [q], |S| = q-r} 2^{-i_S}(1 + \frac{1}{2^{p-1} - 1})^{k_S}. \tag{8}$$

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 8, 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_0} as follows: Since $i \geq w$,

$$i^{\alpha_0-\beta} \geq w^{\alpha_0-\beta} \geq 3p/\ln 2 \tag{9}$$

by our assumption in the statement of the lemma. Since $p \geq 2$, we have $i^{\alpha_0-\beta} > 6 > 2^\beta(1+1/\ln 2)$ which is all that we need to derive that $\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] < 2^{i^{\alpha_0}-i}$ 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 7.1. *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

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

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_0}. For the first term observe that by (9), $i^{\alpha_0-\beta} \geq 3p/\ln 2$ so $i^\beta p^{1-\beta}/\ln 2 \leq i^{\alpha_0}/3$. For the second term again by (9) we have $p \leq i^{\alpha_0-\beta}/3 \leq i^{\alpha_0}/3$. For the last term, since $q^{\alpha_0} \geq \frac{6}{\ln 2}2^p r$, we have

$$\frac{2^{p+1}ri}{pq\ln 2} \leq \frac{q^{\alpha_0}i}{3pq} \leq i(pq)^{\alpha_0-1}/3 \leq i^{\alpha_0}/3,$$

since $i \leq pq$. Therefore in case II we have $\Pr_{\pi\sim\nu}[C_\rho \cap C_\pi \neq \emptyset] < 2^{i^{\alpha_0}-i}$ 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

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

since $r \leq q/2$. □

8 Discussion

In this paper we have proven strong randomized communication complexity lower bounds for AC^0 functions for up to $\Theta(\log n)$ players. For protocols of constant error, functions computed by polynomial-size depth-4 circuits suffice, and for protocols of error exponentially close to that of random guessing, functions computed by polynomial-size depth-6 circuits suffice. It would be nice to reduce the circuit depths required for these lower bounds.

A particularly interesting and useful function for further investigation is the depth-2 function set disjointness. The best lower bounds for set disjointness are non-trivial only for $O(\log^{1/3} n)$ players and are not particular large. It is still consistent with our knowledge that set-disjointness requires polynomial communication complexity even for $\Omega(\log n)$ players. Such lower bounds would imply a depth-2 separation between $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}}$ and $\mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for the same numbers of players.

Acknowledgements

We thank Emanuele Viola for suggesting the circuit complexity application (Section 6) and Alexander Sherstov, Arkadev Chattopadhyay, and the anonymous FOCS referees for many helpful comments. We also thank Avi Wigderson for suggesting using the Sipser function in Section 6 to improve our circuit lower bounds from earlier versions of this paper.

References

[1] E. 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, 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.

[6] 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*, pages 477–486, Philadelphia,PA, October 2008. IEEE.

[7] C. Berg and S. Ulfberg. A lower bound for perceptrons and an oracle separation of the $\mathrm{PP}^{\mathrm{PH}}$ hierarchy. *J. Comput. Syst. Sci.*, 56(3):263–271, 1998.

[8] A. 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] A. Chattopadhyay and A. 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] M. David, T. Pitassi, and E. 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] J. Håstad. *Computational Limitations of Small-Depth Circuits*. MIT Press, 1987. ACM Doctoral Dissertation Award Series (1986).

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

[14] 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.

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

[16] 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.

[17] M. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability. In *30th Annual Symposium on Foundations of Computer Science*, pages 574–579, Research Triangle Park, NC, October 1989.

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

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

[20] R. Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9:113–122, 2000.

[21] 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.

[22] A. A. Razborov and A. A. Sherstov. The sign-rank of AC^0. In *Proceedings 49th Annual Symposium on Foundations of Computer Science*, pages 57–66, Philadelphia,PA, October 2008. IEEE.

[23] 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.

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

[25] 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.

[26] A. A. Sherstov. Unbounded-error communication complexity of symmetric functions. In *Proceedings 49th Annual Symposium on Foundations of Computer Science*, pages 384–393, Philadelphia,PA, October 2008. IEEE.

[27] M. Sipser. Borel sets and circuit complexity. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, pages 61–69, Boston, MA, April 1983.

[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 Other communication complexity bounds for AC^0 circuits

In Section 5 we exhibit a depth-4 AC^0 function that has nontrivial communication lower bounds for up to $\Theta(\log n)$ players and a depth-2 and a depth-5 AC^0 functions that are in $\mathsf{NP}^{\mathsf{cc}}_k - \mathsf{BPP}^{\mathsf{cc}}_k$ for k up to $\Theta(\log^{1/3} n)$ and $\Theta(\log n)$, respectively. In this section we prove a number of related results, namely, a depth-3 AC^0 function that has nontrivial communication lower bounds for up to $\Theta(\sqrt{\log n})$ players and a depth-4 AC^0 functions that is in $\mathsf{NP}^{\mathsf{cc}}_k - \mathsf{BPP}^{\mathsf{cc}}_k$ for k up to $\Theta(\log n/\log\log n)$.

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

Using the pattern selector function $\psi_{k,\ell}$ the results of this section will let us obtain results for simpler functions than with the other selector functions we consider. This also allows us to review the details of the methods from prior work and highlight the consequences of (ϵ, α)-approximate degree alone.

We first review the independence properties of the patthen tensor selection function $\psi_{k,\ell}$ as captured using the definition of r_ψ from Section 3.

Proposition A.1. *[9, 16] If $\psi = \psi_{k,\ell}$, then*

$$\Pr_{y^0,y^1\in D_\psi^{(m)}}[r_\psi(y^0,y^1)=r] \le \big(\frac{e(k-1)m}{r\ell}\big)^r.$$

Proof. In the case, $D_\psi^{(m)} = D_{\psi_{k,\ell}}^{(m)}$ is $[\ell]^{m(k-1)s}$. $z_i^u = \psi(x_i, y_{*i}^u)$ 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 $j \in [k-1]$. There are ℓ elements in $D_{\psi_{k,\ell},j}$ for each j so the probability that $y_{ji}^0 = y_{ji}^1$ is $1/\ell$. Therefore the probability that $y_{ji}^0 = y_{ji}^1$ for some $j \in [k-1]$ is at most $(k-1)/\ell$. By the independence of the choices for different $i \in [m]$ $\Pr_{y^0,y^1\in D_\psi^{(m)}}[r_\psi(y^0,y^1)=r] \le \binom{m}{r}\big(\frac{k-1}{\ell}\big)^r \le \big(\frac{em(k-1)}{r\ell}\big)^r$. □

Remark The lower bounds in [9, 16] use the above property of $\psi = \psi_{k,\ell}$ and follow the same general outline as in Theorem 3.2 but instead of being able to use Lemma 3.4, they use the following bound. This is weaker because it only relies on the assumption of large approximate degree of the function f.

Proposition A.2. *[9, 16] If $r = r_\psi(y^0, y^1)$ then $H(y^0, y^1) \leq \dfrac{2^{(2^{k-1}-1)r}}{2^{2^{k-1}m}}$.*

In [9, 16], to prove the lower bound for $\text{DISJ}_{k,n}$, the function f is set to OR_m and ψ is $\psi_{k,\ell}$. By Proposition 2.3, $d = deg_{5/6}(\text{OR}_m) \geq \sqrt{m/12}$. Plugging the bound in Proposition A.1 together with the bounds from Proposition 3.3 for $r < d$ and from Proposition A.2 when $r \geq d$ into the correlation inequality it is not hard to show that $R^k_{1/3}(f \circ \psi^m) \geq d/2^k - O(1)$ for $\ell > \frac{2^{2^k} kem}{d}$. Hence for suitable $k = O(\log\log n)$ they derive lower bounds on $R^k_{1/3}(\text{DISJ}_{k,n})$.

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 A.2 and from the inefficiency of the selector function $\psi_{k,\ell}$.

The following theorem yields the stronger results that follow from using the pattern tensor selector and a function of large $(5/6, \alpha)$-approximate degree rather than simply large 5/6-approximate degree.

Theorem A.3. *For any Boolean function f on m bits with $deg_{5/6,\alpha}(f) \geq d$ for some $\alpha : \{0, \ldots, m\} \to \mathbb{R}$ such that $\alpha(r) \leq r^{\alpha_0}$ for $r \geq d$, 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 $k \leq (1 - \alpha_0)\log_2 d$.*

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

$$\sum_{r=d}^{m} 2^{(2^{k-1}-1)\alpha(r)} \Pr_{y^0, y^1 \in D^{(m)}_{\psi_{k,\ell}}}[r_{\psi_{k,\ell}}(y^0, y^1) = r] \leq \sum_{r=d}^{m} 2^{(2^{k-1}-1)\alpha(r)} \cdot \left(\frac{e(k-1)m}{r\ell}\right)^r \qquad (10)$$

Since $k \leq (1 - \alpha_0)\log_2 d$, we have $(2^{k-1} - 1)\alpha(r) < d^{1-\alpha_0} r^{\alpha_0} \leq r$ for $r \geq d$ so (10) is

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

Plugging this in to Theorem 3.2 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. □

Here we apply the (ϵ, α) degree bound for the TRIBES function with Theorem A.3 for the pattern tensor selector function $\psi_{k,\ell}$. Note that

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

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

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

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

Lemma A.4. *Given any constants $0 < \epsilon, \alpha_0, \beta < 1$ with $\beta > 1 - \epsilon$ and $\alpha_0 - \beta \geq 0.1$. Let $q > p \geq 2$ be integers such that $2\lceil q^{1-\beta} \rceil < 2^p \leq \frac{1}{6} q^{\alpha_0 + \epsilon - 1} \ln 2$. Let $s = \lceil 8\sqrt{3} e (k-1) p q^{(1+\epsilon)/2} \rceil^{k-1}$ 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_0, 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 4.3, for q sufficiently large $\textsc{Tribes}_{p,q}$ has $(5/6, \alpha)$-approximate degree d at least $q^{(1-\epsilon)/2}/\sqrt{12}$ where $\alpha(r) = r^{\alpha_0}$ for $r \geq d$. 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 A.3 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_0, 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

$$n \leq q^{(k+1)/2 + \epsilon(k-1)/2} (c' \log_2 n)^k \tag{11}$$

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 (11), $q^{(1-\epsilon)/2}$ 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 $\epsilon = 0.4$, $\alpha_0 = 0.9$, and $\beta = 0.8$ in the above lemma we obtain the following less cluttered lower bound statement.

Corollary A.5. *Let p be a sufficiently large integer, $q = 2^{4p}$, and $m = pq$. Let $k \geq 2$ be an integer, $s = \lceil 8\sqrt{3} e (k-1) p q^{0.7} \rceil^{k-1}$, and $n = ms = pqs$. 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 $\Omega(q^{0.3}/2^k)$ 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 $O(\sqrt{\log n})$.*

A.2 A depth-4 AC^0 functions that is in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for k up to $\Theta(\log n / \log\log n)$

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 $D_{\psi^{\oplus b}_{k,\ell}, j}$ 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 $\psi_{k,\ell}^{\oplus b}$ 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 $\psi_{k,\ell}^{\oplus b}$ 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 $\psi_{k,\ell}^{\oplus b}$ as a $\mathbf{\Sigma_4}$ 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 $\psi_{k,\ell}^{\oplus b}$ as a $\mathbf{\Pi_3}$ 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 $\psi_{k,\ell}^{\oplus b}$, the variables $\psi_{k,\ell}^{\oplus b}(x_i, y_{*i}^u)$ 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 $y_{*ib'}^u$ and $y_{*ib'}^v$ 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 $r_{\psi_{k,\ell}^{\oplus b}}(y^0, y^1)$ is the number of $i \in [m]$ such that for every $b' \in [b]$, $y_{jib'}^0 = y_{jib'}^1$ for some $j \in [k-1]$.

The key to the improvement possible with $\psi_{k,\ell}^{\oplus b}$ is that we can prove a sharper analogue of Proposition A.1.

Lemma A.6. *If $\psi = \psi_{k,\ell}^{\oplus b}$ then $\Pr_{y^0,y^1 \in D_\psi^{(m)}}[r_\psi(y^0,y^1) = r] \leq \binom{m}{r}\left(\frac{k-1}{\ell}\right)^{br} \leq \left(\frac{em(k-1)^b}{r\ell^b}\right)^r$.*

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 $b' \in [b]$, $y_{jib'}^0 = y_{jib'}^1$ for some $j \in [k-1]$. As in the case of Proposition A.1, for each fixed i and b' the probability that $y_{jib'}^0 = y_{jib'}^1$ 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 $\left(\frac{k-1}{\ell}\right)^b$. 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 A.7. *For $0 < \alpha_0 < 1$ and any Boolean function f on m bits with $deg_{5/6,\alpha}(f) \geq d$ where $\alpha(r) \leq r^{\alpha_0}$ for $r \geq d$, 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 $R_{1/3}^k(f \circ (\psi_{k,\ell}^{\oplus b})^m \geq d/2^k - 3$ for $k \leq (1-\alpha_0)\log_2 d$.*

Proof. For $\psi = \psi_{k,\ell}^{\oplus b}$, by Lemma A.6,

$$\sum_{r=d}^{m} 2^{(2^{k-1}-1)\alpha(r)} \Pr_{y^0,y^1 \in D_\psi^{(m)}}[r_\psi(y^0,y^1) = r] \leq \sum_{r=d}^{m} 2^{(2^{k-1}-1)\alpha(r)} \left(\frac{em(k-1)^b}{r\ell^b}\right)^r \tag{12}$$

Since $k \leq (1-\alpha_0)\log_2 d$, we have $(2^{k-1}-1)\alpha(r) < d^{1-\alpha_0}\alpha(r) \leq r$ for $r \geq d$ so (12) is

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

Plugging this in to Theorem 3.2 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 A.7 to $\textsc{Tribes}_{p,q} \circ (\psi^{\oplus b}_{k,\ell})^m$ for suitable values of b.

Lemma A.8. *Given any constants $0 < \epsilon, \alpha_0, \beta < 1$ with $\beta > 1 - \epsilon$ and $\alpha_0 - \beta \geq 0.1$. Let $q > p \geq 2$ be integers such that $2\lceil q^{1-\beta} \rceil < 2^p \leq \frac{1}{6} q^{\alpha_0+\epsilon-1} \ln 2$. Let $b \geq \lceil \log_2(16epq^{(1+\epsilon)/2}) \rceil$ and $s = b(2k)^{k-1}$. Then, for q sufficiently large, $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ (\psi^{\oplus b}_{k,\ell})^m)$ is $\Omega(q^{(1-\epsilon)/2}/2^k)$ for $n = pqs$ and $k \leq \frac{1}{2}(1-\alpha_0)(1-\epsilon)\log_2 q - 2$.*

Proof. Let $m = pq$. By Corollary 4.3, for q sufficiently large, the $(5/6, \alpha)$-approximate degree d of $\textsc{Tribes}_{p,q}$ is at least $q^{(1-\epsilon)/2}/\sqrt{12}$ where $\alpha(r) = r^{\alpha_0}$ for $r \geq d$. 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 $k \leq \frac{1}{2}(1-\alpha_0)(1-\epsilon)\log_2 q - 2$ implies that $k \leq (1-\alpha_0)\log_2 d$. Applying Theorem A.7, we see that $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ (\psi^{\oplus b}_{k,\ell})^m)$ is $\Omega(q^{(1-\epsilon)/2}/2^k)$. □

In particular we obtain the following:

Corollary A.9. *Let p be a sufficiently large integer, $q = 2^{4p}$, $k \leq p/40$, 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 $F = \textsc{Tribes}_{p,q} \circ (\psi^{\oplus b}_{k,\ell})^m$. Then $R^k_{1/3}(F)$ 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 4.4 instead of Corollary 4.3. As noted above, $\psi^{\oplus b}_{k,\ell}$ has Π_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 Σ_2 formula, $\textsc{Tribes}_{p,q} \circ (\psi^{\oplus b}_{k,\ell})^m$ is computable by a Σ_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 A.10. *$N^k(\textsc{Tribes}_{p,q} \circ (\psi^{\oplus b}_{k,\ell})^m)$ is $O(\log q + pb \log s)$.*

Proof. Using the Σ_4 formula for $\psi^{\oplus b}_{k,\ell}$ we see that $\textsc{Tribes}_{p,q} \circ (\psi^{\oplus b}_{k,\ell})^m$ can be expressed as a Σ_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 $\textsc{Tribes}_{p,q} \circ (\psi^{\oplus b}_{k,\ell})^m$.

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 A.11. *There is a function G in depth 4 AC^0 such that G is in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for $k \log k \leq a \log n$ for some constant $a > 0$.*

Proof. Observe that $F = \textsc{Tribes}_{p,q} \circ (\psi^{\oplus b}_{k,\ell})^m$ with the parameters from Corollary A.9 by Lemma A.10 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 A.10 applies equally well to G. □

ECCC ISSN 1433-8092

http://eccc.hpi-web.de