SIAM J. Comput.
Vol. 41, No. 3, pp. 484–518

MULTIPARTY COMMUNICATION COMPLEXITY AND THRESHOLD CIRCUIT SIZE OF AC^0[*]

PAUL BEAME[†] AND TRINH HUYNH[†]

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 nontrivial lower bounds for general NOF multiparty communication complexity for any AC^0 function for $\omega(\log\log n)$ players. For nonconstant 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 $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits, showing that the well-known quasi-polynomial 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 $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ 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.

Key words. communication complexity, constant-depth circuits, generalized discrepancy method, pattern matrix method, approximation degree, threshold degree

AMS subject classifications. 03D15, 68Q05, 68Q15, 68Q17

DOI. 10.1137/100792779

1. Introduction.

The complexity class AC^0 of functions computable by families of polynomial-size constant-depth circuits of $\wedge$, $\vee$, and $\neg$ gates is well known for its limitations, that is, its inability to compute some simple functions such as the parity or majority functions [17, 1, 42, 18]. But how much simpler than majority, say, are functions computed by AC^0 circuits?

One way to compare the power of AC^0 circuits to that of majority was considered by Allender [2], who showed that any function computed by an AC^0 circuit can be computed by a quasi-polynomial-size depth 3 circuit of majority gates. With considerable extra work,Yao [43] showed that $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits of quasi-polynomial size, consisting of a majority gate receiving inputs from symmetric gates at the next level that each receive a polylog conjunction of input literals, can actually compute any function in ACC^0, the analogue of AC^0 for circuits which also allow unbounded fan-in modular counting gates for arbitrary fixed moduli. This simulation was further improved by Beigel and Tarui [9] to quasi-polynomial-size $\mathsf{SYM} \circ \mathsf{AND}$ circuits, consisting of a single symmetric gate whose inputs are polylogarithmic AND gates, which can be simulated by depth 3 circuits of majority gates. It is a very natural question to ask whether such a simulation can be made polynomial.

Razborov and Wigderson [30] have shown that quasi-polynomial size is indeed required to simulate these more powerful ACC^0 circuits by $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ cir-

[*] Received by the editors April 19, 2010; accepted for publication (in revised form) March 20, 2012; published electronically May 8, 2012. This work was supported by NSF grants CCF-0514870 and CCF-0830626. A preliminary version of this paper appeared in the *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, 2009.

http://www.siam.org/journals/sicomp/41-3/79277.html

[†] Computer Science and Engineering, University of Washington, Seattle, WA 98195-2350 (beame@cs.washington.edu, trinh@cs.washington.edu). The second author is also known as Dang-Trinh Huynh-Ngoc. The second author's research was supported by a Vietnam Education Foundation Fellowship.

cuits, but the question of whether the simulation for AC^0 circuits can be improved to polynomial has remained open.

An important approach to understanding the complexity of $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits (and hence ACC^0 circuits) has been through number-on-forehead (NOF) multiparty communication complexity. In this model, many players cooperate to compute some function of their joint inputs and every input value can be seen by all but one of the players (hence the notion that the input values that a player cannot see are written on that player's forehead). Håstad and Goldmann [20] showed that any AC^0 or ACC^0 function has at most polylogarithmic randomized multiparty NOF communication complexity when its input bits are divided arbitrarily among a polylogarithmic number of players. NOF multiparty communication complexity is of considerable interest in its own right as it has also been used to derive time-space tradeoff lower bounds and proof complexity lower bounds.

Håstad and Goldman's result was based on the above simulations of AC^0 and ACC^0 by $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits with polylog fan-in at the inputs and, using the stronger results of [9], implies the same upper bound for deterministic algorithms. These protocols may even be *simultaneous* NOF protocols, in which the players in parallel send their information to a referee who computes the answer [3]. The quasi-polynomial lower bound of Razborov and Wigderson for simulating ACC^0 follows from a lower bound on NOF multiparty communication complexity due to Babai, Nisan, and Szegedy [5]. They introduced the *discrepancy method for cylinder intersections* and showed, among other things, that the generalized inner product, a function in ACC^0 computable by a depth 2 circuit consisting of a single parity gate whose inputs are n AND gates with fan-in k, requires k-party NOF communication complexity $\Omega(n/4^k)$ which is polynomial in n for k up to $\Theta(\log n)$. This function, like all the functions to which they applied their discrepancy method, cannot be computed in AC^0.

The largest lower bounds on the NOF multiparty complexity of AC^0 functions have been in models with restricted interaction. For the communication complexity of the set disjointness function with k players (which is given by a polynomial-size disjunctive normal form (DNF) formula and hence is in AC^0) there are lower bounds of the form $\Omega(n^{1/(k-1)}/(k-1))$ in the simultaneous NOF [39, 8] and $n^{\Omega(1/k)}/k^{O(k)}$ in the one-way NOF model [41]. These are subpolynomial lower bounds for all nonconstant values of k and, at best, polylogarithmic when k is $\Omega(\log n/\log\log n)$.

Until recently, the general NOF multiparty communication complexity of AC^0 functions was completely open. One reason for interest in obtaining such a lower bound was a reduction by Beame, Pitassi, and Segerlind [7], which showed that polynomial lower bounds for the NOF communication complexity of the k-player set disjointness function would yield exponential lower bounds for small rank proofs involving degree $(k-1)$ polynomials, but no superlogarithmic lower bounds for general NOF protocols were known for AC^0 functions, even for three players. That changed with lower bounds for a depth 3 AC^0 function by Chattopadhyay [13] and for set disjointness by Lee and Shraibman [25] and Chattopadhyay and Ada [14] but no lower bounds apply for $\omega(\log\log n)$ players. As for circuit simulations of AC^0, the result of Chattopadhyay [13] implies that AC^0 cannot be simulated by polynomial-size $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{ANY}$ circuits with $o(\log\log n)$ input fan-in at every bottom gate. However, there have been no nontrivial-size lower bounds for the simulation of AC^0 by $\mathsf{MAJ} \circ \mathsf{MAJ} \circ \mathsf{AND}$ or even $\mathsf{SYM} \circ \mathsf{AND}$ circuits with $\omega(\log\log n)$ bottom fan-in. As shown by Viola [40], sufficiently strong lower bounds for AC^0 in the multiparty NOF communication model, even for sublogarithmic numbers of players, can yield quasi-polynomial circuit size lower bounds.

We indeed produce such strong lower bounds. We show that there is an explicit linear-size fixed-depth $\mathsf{AC^0}$ function that requires randomized k-party NOF communication complexity of $n^{\Omega(1)}/4^k$ even for protocols with error exponentially close to $1/2$ (as a function of the input size). For $\omega(1)$ players this bound is larger than all previous multiparty NOF communication complexity lower bounds for $\mathsf{AC^0}$ functions, even those in the weaker simultaneous model. The bound is nontrivial for up to $\Theta(\log n)$ players and is sufficient to produce fixed-depth $\mathsf{AC^0}$ functions that require $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits of $n^{\Omega(\log n)}$ size, showing that quasi-polynomial size is necessary for the simulation of $\mathsf{AC^0}$.

The function above for which we derive randomized communication complexity lower bounds for error exponentially close to $1/2$ is computable in depth 6 $\mathsf{AC^0}$. The remainder of our lower bounds apply to bounded-error randomized protocols—those with error at most some fixed constant $\epsilon < 1/2$. For such protocols, 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 $\mathsf{AC^0}$ contains functions in $\mathsf{NP^{cc}_k} - \mathsf{BPP^{cc}_k}$ for k up to $\Theta(\log n)$. We obtain an $\Omega(2^{\sqrt{\log n}/\sqrt{k}-k})$ lower bound on the randomized k-party NOF communication complexity of the set disjointness function, a bound that is nontrivial for up to $\Theta(\log^{1/3} n)$ players. The best previous lower bounds for set disjointness, due to Lee and Shraibman [25] and Chattopadhyay and Ada [14], only apply for $k \le \log\log n - o(\log\log n)$ players (though these bounds are stronger than ours for $o(\log\log n)$ players).

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

Results such as ours which prove lower bounds for simple functions that were previously known only for more complex functions can substantially broaden the applicability of such bounds. Alternatively, as in the case of our extension of the lower bound in [30], they can be seen as showing that our understanding of the hardness of the more complex functions is less based on that complexity than we might have thought.

Methods and related work. Building upon the generalized discrepancy method introduced by Klauck [22] and Razborov [31], Sherstov [33, 35] and Shi and Zhu [37] introduced general methods to use analytic properties of Boolean functions to derive communication lower bounds for related Boolean functions. These related functions are obtained by disjointly applying simple operations to select each bit to be passed on to the original functions. (Krause and Pudlák [23] and Raz and McKenzie [29] also employed simple selection operations to extend functions in order to strengthen the models for which lower bounds could be shown. However, they did not base their arguments on the analytic properties of the original functions as in these new methods.)

These new methods have been successfully applied to yield strong lower bounds for two-party randomized and quantum communication complexity. Sherstov called his method the *pattern matrix method.* Later, Chattopadhyay [13] generalized the ideas in this method for $k \ge 2$ players to yield the first lower bounds for the general NOF multiparty communication complexity of any $\mathsf{AC^0}$ function for $k \ge 3$, implying exponential lower bounds for computation of $\mathsf{AC^0}$ functions by $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{ANY}$ circuits with $o(\log\log n)$ input fan-in at every bottom gate—our results extend this to fan-in $\Omega(\log n)$. Then Lee and Schraibman [25] and Chattopadhyay and Ada [14] further generalized the pattern matrix method 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 [4, 39, 8, 41, 21, 10] 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 [16] 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 [14, 34] is as follows. To prove k-party lower bounds for a function F, we find two other functions f and ψ such that F computes the composition $f \circ \psi$ as a subfunction, and then we prove the communication lower bound for this composition. The function f is chosen so that it has large approximate degree. Using linear programming (LP) duality (or duality of norms) it follows that for such an f there exist 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.[1] The function ψ is also chosen to be some suitable *selector function*, which will be defined later, so that the correlation of $f \circ \psi$ and $g \circ \psi$ (under a suitable distribution) is the same as that of f and g and thus is also high. Thus it suffices to lower bound the communication complexity of $g \circ \psi$. By using the discrepancy method of [5] we then proceed to prove a communication lower bound for $g \circ \psi$. Given a well-chosen selector function ψ, thanks to the orthogonality of g to all low-degree polynomials, $g \circ \psi$ has low discrepancy and this is shown using the bound in [5, 15, 28] which is derived via iterated application of the Cauchy–Schwartz inequality.

As an example, the bound for set disjointness $\text{Disj}_{k,n}(\mathbf{x}) = \vee_{i=1}^{n} \wedge_{j=1}^{k} x_{hi}$, which more properly should be called set intersection, corresponds to the so-called pattern tensor selector function ψ and $f = \text{Or}$ which has approximate degree $\Omega(\sqrt{n})$.

Our key technical contribution to the above method is that we identify a stronger requirement on f, rather than just large approximate degree, that if satisfied produces much stronger communication lower bounds than previously possible. We show that for any function f for which even approximating f within ϵ on only a subset S of inputs requires large degree, there exist another function g and a distribution μ that satisfy the same conditions as above and moreover, μ is "max-smooth"—the probability of subsets defined by partial assignments is never much larger than under the uniform distribution.[2] The smoothness quality and the properties of the subset S are determined by a function α, so we call the degree bound the (ϵ, α)-approximate degree. This notion is defined precisely to that we can derive the smoothness bound we want via LP duality. We then 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 Or, but we are able to exhibit simple AC^0 functions with large (ϵ, α)-approximate degree, including a variant of the $\text{Tribes}_{p,q}(\mathbf{x}) = \vee_{i=1}^{q} \wedge_{j=1}^{p} x_{i,j}$ function defined (though not named) by Ben-Or and Linial [11].

[1] We note that this duality between approximability and orthogonality was also used in the work of Shi and Zhu [37], who also introduced another general method to relate analytic properties to communication lower bounds of Boolean functions.

[2] In the two-party case, Sherstov [36] and Razborov and Sherstov [32] 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 $f(x) = 1$.

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 [35, 14, 16] based on pattern tensor selector functions and orthogonalizing distributions for functions of large ϵ-approximate degree, and we 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 giving communication complexity lower bounds based on (ϵ, α)-degree lower bounds, 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 apply our construction to the $\textsc{Or}_q$ function to yield the function $\textsc{Tribes}_{p,q}(\mathbf{x})$ with large (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 a new 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 $\mathsf{NP}_\mathsf{k}^\mathsf{cc}$ and $\mathsf{BPP}_\mathsf{k}^\mathsf{cc}$ 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 $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits 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 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 demonstrate the advantage of using (ϵ, α)-approximate degree instead of ϵ-approximate degree. In Appendix A.2 we analyze a selector function that is a small parity of pattern tensor selector functions and use it to obtain depth 4 separating functions in $\mathsf{NP}_\mathsf{k}^\mathsf{cc} - \mathsf{BPP}_\mathsf{k}^\mathsf{cc}$ for $k = O(\log n / \log \log n)$.

2. Preliminaries and the pattern matrix method.

We will assume that Boolean functions on m bits are maps $f : \{0,1\}^m \mapsto \{0,1\}$, where 1 stands for "true." For the convenience of notation, especially in the Fourier analysis parts in this paper, we will sometimes adopt the output range $\{1, -1\}$ for Boolean functions, where 1 stands for "true." Which output range is to be used will be explicitly stated or be clear from the context.

We write vectors as boldface small letters, e.g., $\mathbf{x}$, with entries written as x_i. We denote by $|\mathbf{x}|$ the Hamming weight of a vector $\mathbf{x} \in \{0,1\}^*$. For $\mathbf{x} \in \{0,1\}^m$ and $S = \{i_1, \ldots, i_s\} \subseteq [m]$, we write $\mathbf{x}_S = (x_{i_1}, \ldots, x_{i_s})$.

Circuit complexity. Let AND denote the class of all unbounded fan-in $\wedge$ functions (of literals), SYM denote the class of all Boolean symmetric functions (i.e., $f \in \mathsf{SYM}$ iff $f(\mathbf{x})$ depends only on $|\mathbf{x}|$ for every input $\mathbf{x}$), and $\mathsf{MAJ} \subset \mathsf{SYM}$ denote the class of all majority functions (i.e., $\mathsf{MAJ}(\mathbf{x})$ output 1 iff $|\mathbf{x}|$ is at least half of the input bits). AC^0 is the class of functions $f : \{0,1\}^* \mapsto \{0,1\}$ computed by polynomial-size circuits (or formulas) of constant depth having $\neg$ gates and unbounded fan-in $\wedge$ and $\vee$ gates. A formula is a $\mathsf{\Sigma_1}$ formula if it is a clause and a $\mathsf{\Pi_1}$ formula if it is a term. For $i \geq 1$, a $\mathsf{\Sigma_{i+1}}$ formula is an unbounded fan-in $\vee$ of $\mathsf{\Pi_i}$ formulas and a $\mathsf{\Pi_{i+1}}$ formula is an unbounded fan-in $\wedge$ of $\mathsf{\Sigma_i}$ formulas.

We assume that the layout of circuits has the output gate at the top and the inputs at the bottom. Given classes of functions $\mathsf{C_1}, \mathsf{C_2}, \dots \mathsf{C_d}$, 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 ith level from the top are from $\mathsf{C_i}$.

Norms and correlation. Given a real-valued function f on $\{0,1\}^m$, for $p \geq 1$, define $||f||_p = (\sum_{\mathbf{x} \in \{0,1\}^m} |f(\mathbf{x})|^p)^{1/p}$ and $||f||_\infty = \max\{|f(\mathbf{x})| : \mathbf{x} \in \{0,1\}^m\}$. Let μ be a probability distribution on $\{0,1\}^m$, i.e., a nonnegative function with $||\mu||_1 = 1$. The correlation between two real-valued functions f and g on $\{0,1\}^m$ under μ is defined as $\mathrm{Cor}_\mu(f,g) := \mathbf{E}_{\mathbf{x}\sim\mu}[f(\mathbf{x})g(\mathbf{x})]$. 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. For $k \geq 2$ we consider the usual NOF model of k-party communication complexity. In this model, a function f is associated with a fixed k-partition of its inputs, and player i has access to all inputs *except* those in block i of the partition. (We can view block i as assigned to the forehead of player i in order to be read only by the other players.) The players communicate by broadcasting bits to the other players, which can be viewed as writing the bits on a common board, switching turns based on the content of the board, and the last bit written is the output of the protocol. 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—the minimum total number of bits communicated by the players when the protocol is of the given type.

Let Π^c_k be the class of all functions computable by deterministic k-party communication protocols of cost at most c. The following fact provides a means to prove randomized communication lower bounds.

FACT 2.1 (cf. [24]). *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 $\mathbf{x}$ and the inputs to players 1 through $k-1$ as $\mathbf{y}_1, \dots, \mathbf{y}_{k-1}$.

LEMMA 2.2 (see [5, 15, 28]). *Let $X \times Y$ be the input space, where $Y = Y_1 \times \cdots \times Y_{k-1}$, of a function $f : X \times Y \mapsto \mathbb{R}$ and let $\mathbf{U}$ be the uniform distribution over $X \times Y$. If player 0 receives an input from X and player $i > 0$ receives an input from Y_i, then*

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

where $\mathbf{y}^{\mathbf{u}} = (\mathbf{y}_1^{u_1}, \dots, \mathbf{y}_{k-1}^{u_{k-1}})$ for $\mathbf{u} \in \{0,1\}^{k-1}$.

Approximate and threshold degree. Given $0 \leq \epsilon < 1$, the ϵ-approximate degree of a function $f : X \to \mathbb{R}$, $deg_\epsilon(f)$, is the smallest d for which $||f - p||_\infty \leq \epsilon$ for some real-valued multivariate polynomial p of degree d. Following [27] we have the following property of the approximate degree of OR.

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

The threshold degree of a function $f : X \to \{1, -1\}$, $deg^{\pm}(f)$, is the smallest d for which there exists a real-valued multivariate polynomial p of degree d such that $f(x) = sign(p(x))$, where $p(x) \neq 0$ for all x. Hence it follows that $deg^{\pm}(f) = \min_{\epsilon<1} deg_{\epsilon}(f)$. For this reason, we write $deg^{\pm}(f) = deg_{<1}(f)$.

The pattern matrix (and pattern tensor) method. Define the inner product on the space of functions $\{0,1\}^m \mapsto \mathbb{R}$ by $\langle f, g\rangle = \mathbf{E}_{\mathbf{x}\in\{0,1\}^m}[f(\mathbf{x}) \cdot g(\mathbf{x})]$, which is $\mathrm{Cor}_U(f, g)$ for the uniform distribution U on $\{0,1\}^m$. For $S \subseteq [m]$, let $\chi_S : \{0,1\}^m \mapsto \{-1,1\}$ be the function $\chi_S(\mathbf{x}) = \prod_{i\in S}(-1)^{x_i}$. The χ_S'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 (see [35, 37]). *If $f : \{0,1\}^m \mapsto \{-1,1\}$ has $deg_{\epsilon}(f) \geq d$, then there exists a function $g : \{0,1\}^m \mapsto \{-1,1\}$ 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 $p : \{0,1\}^{|S|} \mapsto \mathbb{R}$, $\mathbf{E}_{\mathbf{x}\sim\mu}[g(\mathbf{x}) \cdot p(\mathbf{x}_S)] = 0$.*

Proof. Let Φ_d be the space of multivariate polynomials of degree less than d. By definition, $deg_{\epsilon}(f) \geq d$ iff $\min_{p\in\Phi_d} ||f - p||_\infty > \epsilon$. By duality of norms we have $\min_{p\in\Phi_d} ||f - p||_\infty = 2^m \cdot \max_{q\in\Phi_d^{\perp},\ ||q||_1=1}\langle f, q\rangle$. Let $h \in \Phi_d^{\perp}$ with $||h||_1 = 1$ achieve this maximum value of $\langle f, h\rangle$. Define $\mu(\mathbf{x}) = |h(\mathbf{x})|$. The condition $||h||_1 = 1$ implies that μ is a probability distribution. Letting $g(\mathbf{x}) = h(\mathbf{x})/\mu(\mathbf{x})$ for $\mu(\mathbf{x}) \neq 0$ and $g(\mathbf{x}) = 1$ for $\mu(\mathbf{x}) = 0$, we have $h(\mathbf{x}) = \mu(\mathbf{x})g(\mathbf{x})$. Therefore

$$\epsilon < 2^m\langle f, h\rangle = 2^m\mathbf{E}[f \cdot h] = 2^m\mathbf{E}[f \cdot g \cdot \mu] = \mathbf{E}_{\mathbf{x}\sim\mu}[f(\mathbf{x})g(\mathbf{x})] = \mathrm{Cor}_{\mu}(f, g).$$

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

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 2.5. *Any function $\psi : \{0,1\}^{k\times s} \mapsto \{0,1\}$ 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 $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}) \in D_{\psi} := D_{\psi,1} \times \cdots \times D_{\psi,(k-1)}$, $\Pr_{\mathbf{x}\in\{0,1\}^s}[\psi(\mathbf{x}, \mathbf{y}) = 0] = \Pr_{\mathbf{x}\in\{0,1\}^s}[\psi(\mathbf{x}, \mathbf{y}) = 1] = 1/2$.*

Now we define the general form of k-party NOF communication problems studied in this paper. Let $D_{\psi}^{(m)} := D_{\psi,1}^m \times \cdots \times D_{\psi,(k-1)}^m$, where each $D_{\psi,i}^m$ is the cross-product of m disjoint copies of $D_{\psi,i}$. For any function $f : \{0,1\}^m \mapsto \{1,-1\}$ and any selector function $\psi : \{0,1\}^{k\times s} \mapsto \{0,1\}$, we define a new function $f \circ \psi$ on $\{0,1\}^{k\times m\times s}$ bits as follows: on every $\mathbf{x} \in \{0,1\}^{m\times s}$ and $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}) \in D_{\psi}^{(m)}$,

$$(f \circ \psi)(\mathbf{x}, \mathbf{y}) = (f \circ \psi)(\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1}) = f(\psi(\mathbf{x}_1, \mathbf{y}_{*1}), \ldots, \psi(\mathbf{x}_m, \mathbf{y}_{*m})),$$

where $\mathbf{y}_{*i} = (\mathbf{y}_{1i}, \ldots, \mathbf{y}_{(k-1)i})$ for $i \in [m]$. We will write $z_i = \psi(\mathbf{x}_i, \mathbf{y}_{*i})$ for $i \in [m]$ and $\mathbf{z} = (z_1, \ldots, z_m)$ for the input to f in the above expression. In the k-party NOF communication problem computing $f \circ \psi$, on every input $\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1} \in \{0,1\}^{m\times s}$, player 0 has $\mathbf{x}$ on his forehead (and can see all the $\mathbf{y}_i$ but not $\mathbf{x}$), each other player

$i \in [k-1]$ has $\mathbf{y}_i$ on his forehead (and can only see $\mathbf{x}$ and those $\mathbf{y}_j$ for $j \neq i$), and they need to compute $f \circ \psi(\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1})$.

One example of a selector function ψ is the pattern tensor function $\psi_{k,\ell}$ used in [14, 25] 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 $\mathbf{y}_j \in \{0,1\}^{\ell^{k-1}}$ that are 1 in all entries in one of the ℓ slices along the jth dimension of this array and are 0 in every other entry. For $\mathbf{x} \in \{0,1\}^s$ and such a $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}) \in \{0,1\}^{(k-1)s}$ the array $\wedge_{i=1}^{k-1} \mathbf{y}_i$ contains precisely one 1 which selects the bit of $\mathbf{x}$ to pass to f. This function is expressible by a small two-level $\vee$ of $\wedge$s. As described in [16] 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 [14, 16], which extends ideas of [33, 35] from two-party to k-party communication complexity:

- Start with a function $f : \{0,1\}^m \mapsto \{1,-1\}$ 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 multivariate polynomial.
- Observe that from μ one can define a natural λ under which $g \circ \psi$ and $f \circ \psi$ have the same high correlation as g and f, therefore in order to prove that $f \circ \psi$ is uncorrelated with low communication protocols, it suffices to prove this for $g \circ \psi$ and apply the triangle inequality.
- 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 multivariate polynomials of degree $< d$ to show that all low-degree self-correlations of $g \circ \psi$ 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 subcubes 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 [35, 14, 25] all allow that μ may assign all its probability mass to small subcubes 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 nontrivial to prove).

We also address the inefficiency of the pattern tensor selector function by defining a new selector function that requires significantly fewer bits. David, Pitassi, and Viola [16] 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 define $|\rho| = |\{i : \rho_i \neq *\}|$. Two restrictions π and ρ are *compatible*, denoted as $\pi \parallel \rho$, iff every coordinate i with both $\pi_i, \rho_i \neq *$ satisfies $\pi_i = \rho_i$. Define $C_\rho = \{\mathbf{x} \in \{0,1\}^m : \mathbf{x} \parallel \rho\}$.

Definition 3.1. *Let $\alpha : \{0,\dots,m\} \mapsto \mathbb{R}$. Given a probability distribution λ on the set of restrictions $\{0,1,*\}^m$, we say that $\mathbf{x} \in \{0,1\}^m$ is* α-light *for λ iff $\sum_{\rho\parallel\mathbf{x}} 2^{|\rho|-\alpha(|\rho|)}\lambda(\rho) \leq 1$. Note that when α is the identity function, every point is α-light for every distribution λ.*[3]

Definition 3.2. *Let $\alpha : \{0,\dots,m\} \mapsto \mathbb{R}$ and $f : \{0,1\}^m \mapsto \mathbb{R}$. The (ϵ,α)-*approximate degree[4] *of f, denoted as $deg_{\epsilon,\alpha}(f)$, is defined to be the minimum integer $d \geq 0$ such that there is some multivariate polynomial q of degree d and some probability distribution λ on restrictions such that for every $\mathbf{x} \in \{0,1\}^m$ if $\mathbf{x}$ is α-light for λ, then $|f(\mathbf{x}) - q(\mathbf{x})| \leq \epsilon$. Note that this reduces to $deg_\epsilon(f)$ if $\alpha(r) \geq r$ for all r. Also define $deg_{<\epsilon,\alpha}(f) = \inf_{\epsilon'<\epsilon} deg_{\epsilon',\alpha}(f)$. For $f : \{0,1\}^m \to \{1,-1\}$ we write $deg^{\pm}(f) = deg_{<1}(f)$ and we will usually say "α-threshold degree" for $(<1,\alpha)$-approximate degree.*

This definition is obviously a relaxation of the usual ℓ_∞ approximation of f since the nonlight points can be ignored in the approximation. We will use this definition to prove our main technical theorem. That theorem relies essentially on the following lemma, which generalizes Lemma 2.4 and is the first key to our substantially improved lower bounds. The proof of this lemma is based on LP duality and, indeed, our definition of (ϵ,α)-approximate degree was derived by expressing the desired additional smoothness property over Lemma 2.4 in terms of a stronger linear program and then determining what extension of ϵ-approximate degree of f would be necessary to bound that linear program.

Lemma 3.3 (max-smooth orthogonality-approximation lemma). *Let $0 < \epsilon < 1$ and $\alpha : \{0,\dots,m\} \mapsto \mathbb{R}$. If $f : \{0,1\}^m \mapsto \{-1,1\}$ has $deg_{<\epsilon,\alpha}(f) \geq d$, then there exists a function $g : \{0,1\}^m \mapsto \{-1,1\}$ 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 $p : \{0,1\}^{|S|} \mapsto \mathbb{R}$, $\mathbf{E}_{\mathbf{x}\sim\mu}[g(\mathbf{x}) \cdot p(\mathbf{x}_S)] = 0$; and*
3. *for any restriction ρ, $\mu(C_\rho) \leq 2^{\alpha(|\rho|)-|\rho|}/\epsilon$.*

[3] We will be mostly interested when $\alpha(r) \leq r^{\alpha_0}$ for every large enough r and fixed $1 > \alpha_0 > 0$.

[4] 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.

Proof. 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

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

Suppose that we have optimum $\eta \leq 1$. In this LP formulation, inequalities $v_{\mathbf{x}}$ and $w_{\mathbf{x}}$ ensure that $\mu(\mathbf{x}) \geq |h(\mathbf{x})|$ and so together with inequality γ, this ensures that μ is a probability distribution, and $||h||_1 \leq 1$. If $||h||_1 = 1$, then we must have $\mu(\mathbf{x}) = |h(\mathbf{x})|$ and we can write $h(\mathbf{x}) = \mu(\mathbf{x})g(\mathbf{x})$ 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 $h'(\mathbf{x}) = h(\mathbf{x})/||h||_1$ with $||h'||_1 = 1$ and a new probability distribution μ' by $\mu'(\mathbf{x}) = |h'(\mathbf{x})| \leq \mu(\mathbf{x})/||h||_1 \leq \mu(\mathbf{x})/\epsilon$. 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

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

$$\mu(x): \quad v_{\mathbf{x}} + w_{\mathbf{x}} + \gamma - \sum_{\rho\|\mathbf{x}} 2^{|\rho|-\alpha(|\rho|)}\lambda_\rho = 0 : \mathbf{x}\in\{0,1\}^m, \tag{3.1}$$

$$h(\mathbf{x}): \quad \beta f(\mathbf{x}) + \sum_{|S|<d} y_S\chi_S(\mathbf{x}) + w_{\mathbf{x}} - v_{\mathbf{x}} = 0 : \mathbf{x}\in\{0,1\}^m, \tag{3.2}$$

$$\beta, v_{\mathbf{x}}, w_{\mathbf{x}}, \lambda_\rho \geq 0 : \mathbf{x}\in\{0,1\}.^m$$

Since y_S are arbitrary we can replace $\sum_{|S|<d} y_S\chi_S(\mathbf{x})$ by $p_d(\mathbf{x})$, where p_d is an arbitrary polynomial of degree $< d$ and rewrite (3.2) as

$$h(\mathbf{x}): \quad \beta f(\mathbf{x}) + p_d(\mathbf{x}) + w_{\mathbf{x}} - v_{\mathbf{x}} = 0 : \mathbf{x}\in\{0,1\}^m. \tag{3.3}$$

Equations (3.1) and (3.3) for $\mathbf{x} \in \{0,1\}^m$ together are equivalent to

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

Since these are the only constraints on $v_{\mathbf{x}}$ and $w_{\mathbf{x}}$, respectively, other than nonnegativity, these can be satisfied by any solution to

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

which together are equivalent to

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

Since $p_d(\mathbf{x})$ 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 $|\beta f(\mathbf{x}) + p_d(\mathbf{x})|$ by $\beta |f(\mathbf{x}) - q_d(\mathbf{x})|$.

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

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

for all $\mathbf{x} \in \{0,1\}^m$, 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 $\mathbf{x} \in \{0,1\}^m$ at which $|f(\mathbf{x}) - q_d(\mathbf{x})| \geq \epsilon$. For any $\mathbf{x} \in B$, the value of the objective function of $\mathcal{D}$, which is $\beta \cdot \epsilon + \gamma$, is not more than

$$\beta |f(\mathbf{x}) - q_d(\mathbf{x})| + \gamma \leq \sum_{\rho \| \mathbf{x}} 2^{|\rho| - \alpha(|\rho|)} \lambda_\rho. \tag{3.4}$$

Let $R(\mathbf{x})$ denote the right-hand side of inequality (3.4). It suffices to prove that $R(\mathbf{x}) \leq 1$ for some $\mathbf{x} \in B$. This, in turn, is equivalent to proving that

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

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

Although the upper bound on $\mu(C_\rho)$ in Lemma 3.3 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 above by r^{α_0} for $r \geq d$ and $\alpha_0 < 1$.

We now see how to apply Lemma 3.3 to obtain communication lower bounds. By Fact 2.1, it suffices to upper bound the discrepancy with low-cost communication protocols. We will do this by using Lemma 2.2, which intuitively bounds the discrepancy of a function by the expectation of the function's correlation with itself on randomly chosen hypercubes of points. Since our function is of the form $g \circ \psi$, we will look at how the output bits of ψ, which are then given to g, relate to each other. This is the purpose of the next technical definition, which is motivated by Lemma 2.2.

Definition 3.4. *Let ψ be a selector function with $D_\psi = D_{\psi,1} \times \cdots \times D_{\psi,(k-1)}$. For fixed $\mathbf{y}^0, \mathbf{y}^1 \in D_\psi^{(m)}$, $i \in [m]$ and uniformly random $\mathbf{x}_i$, we call i* good *for $(\mathbf{y}^0, \mathbf{y}^1)$ if the set of 2^{k-1} random variables $z_i^{\mathbf{u}} = \psi(\mathbf{x}_i, \mathbf{y}_{*i}^{\mathbf{u}})$ for $\mathbf{u} \in \{0,1\}^{k-1}$ are mutually independent, where $\mathbf{y}^{\mathbf{u}}$ is defined as in Lemma 2.2; otherwise we call i* bad *for $(\mathbf{y}^0, \mathbf{y}^1)$. Let $R_\psi(\mathbf{y}^0, \mathbf{y}^1)$ be the set of $i \in [m]$ that are bad for $(\mathbf{y}^0, \mathbf{y}^1)$ and let $r_\psi(\mathbf{y}^0, \mathbf{y}^1) = |R_\psi(\mathbf{y}^0, \mathbf{y}^1)|$.*

Given the above definition and using Lemma 2.2, we will show that on every $\mathbf{y}^0$ and $\mathbf{y}^1$, the expectation (over $\mathbf{x}$) of the correlation of $g \circ \psi$ with itself on the hypercube $\{\mathbf{y}^{\mathbf{u}}\}_{\mathbf{u} \in \{0,1\}^{k-1}}$ will be bounded by the correlation of g with a polynomial defined only on the bad coordinates of $\mathbf{y}^0$ and $\mathbf{y}^1$. We will also show that for some selectors ψ, the number of bad coordinates is small for random $\mathbf{y}^0$ and $\mathbf{y}^1$, and if this is the case, then the correlation is zero by the orthogonality of g to small-degree polynomials.

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 3.5. *Let $\alpha : \{0, \ldots, m\} \mapsto \mathbb{R}$. If a function $f : \{0,1\}^m \mapsto \{1,-1\}$ 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*

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

Proof. The pattern of the argument follows the outline from section 2. We first apply Lemma 3.3 to f to produce function g and distribution μ. By construction $\mathrm{Cor}_\mu(f, g) \geq 1 - \epsilon$.

Given μ, we define a distribution λ on $\{0,1\}^{k \times m \times s}$ such that $\mathrm{Cor}_\lambda(f \circ \psi, g \circ \psi) = \mathrm{Cor}_\mu(f, g)$ as follows: for any $\mathbf{x} \in \{0,1\}^{m \times s}$ and $\mathbf{y} \in D_\psi^{(m)}$ and $z_i = \psi(\mathbf{x}_i, \mathbf{y}_{*,i})$ for every $i \in [m]$,

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

and $\lambda(\mathbf{x}, \mathbf{y}) = 0$ if $\mathbf{y} \notin D_\psi^{(m)}$. That is, λ gives equal weight to every input that produces the same $(z_1, \ldots, z_m)$. Thus

$$\begin{aligned} \mathrm{Cor}_\lambda(f \circ \psi, g \circ \psi) &= \sum_{(\mathbf{x},\mathbf{y}) \in \{0,1\}^{m \times s} \times D_\psi^{(m)}} f \circ \psi(\mathbf{x}, \mathbf{y}) g \circ \psi(\mathbf{x}, \mathbf{y}) \lambda(\mathbf{x}, \mathbf{y}) \\ &= \sum_{\mathbf{x} \in \{0,1\}^{m \times s}, \mathbf{y} \in D_\psi^{(m)}} f(\mathbf{z}) g(\mathbf{z}) \lambda(\mathbf{x}, \mathbf{y}), \end{aligned}$$

where $\mathbf{z}$ is defined as above,

$$\begin{aligned} &= \sum_{\mathbf{z} \in \{0,1\}^m} \sum_{\substack{\mathbf{x} \in \{0,1\}^{m \times s}, \mathbf{y} \in D_\psi^{(m)}: \\ \forall\ i \in [m], z_i = \psi(\mathbf{x}_i, \mathbf{y}_{*,i})}} f(\mathbf{z}) g(\mathbf{z}) \frac{\mu(z_1, \ldots, z_m)}{2^{ms-m} |D_\psi|^m} \\ &= \sum_{\mathbf{z} \in \{0,1\}^m} f(\mathbf{z}) g(\mathbf{z}) \mu(z_1, \ldots, z_m) = \mathrm{Cor}_\mu(f, g) \leq 1 - \epsilon, \end{aligned}$$

where the last line follows because, by definition of the selector, for each fixed $\mathbf{y}_{*,i} \in D_\psi$, each $z_i = \psi(\mathbf{x}_i, \mathbf{y}_{*,i})$ is a uniformly random bit given uniformly random $\mathbf{x}_i \in \{0,1\}^s$.

To prove a lower bound c on $R^k_{1/2-\epsilon}(f \circ \psi)$ we show that $\mathrm{Cor}_\lambda(f \circ \psi, \Pi^c_k) \leq 2\epsilon$. Since $\mathrm{Cor}_\lambda(f \circ \psi, g \circ \psi) \geq 1 - \epsilon$, by the triangle inequality of correlation, it suffices to show that $\mathrm{Cor}_\lambda(g \circ \psi, \Pi^c_k) \leq \epsilon$.

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

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

where $H(y^0, y^1)$ is the self-correlation of g under μ in the hypercube defined by $\mathbf{y}^0$ and $\mathbf{y}^1$:

$$H(\mathbf{y}^0, \mathbf{y}^1) := \left| \mathbf{E}_{\mathbf{x}} \left[\prod_{\mathbf{u} \in \{0,1\}^{k-1}} \mu(z^{\mathbf{u}}_1, \ldots, z^{\mathbf{u}}_m) g(z^{\mathbf{u}}_1, \ldots, z^{\mathbf{u}}_m) \right] \right|,$$

where $z^{\mathbf{u}}_i = \psi(\mathbf{x}_i, \mathbf{y}^{\mathbf{u}}_{*i})$. We now state two bounds on the self-correlation $H(\mathbf{y}^0, \mathbf{y}^1)$ that depend on the value of $r = r_\psi(\mathbf{y}^0, \mathbf{y}^1)$. The first bound is a slightly more general version of a bound from [14] and was a key to the original method.

PROPOSITION 3.6. *If $r = r_\psi(\mathbf{y}^0, \mathbf{y}^1) < d$, then $H(\mathbf{y}^0, \mathbf{y}^1) = 0$.*

The following bound is the key to our exponentially better results than in previous work. A weaker version given in [14] corresponds to the case that $\alpha(r) = r$ (but does not have the $(1-\epsilon)^{2^{k-1}-1}$ in the denominator).

LEMMA 3.7. $H(\mathbf{y}^0, \mathbf{y}^1) \leq \frac{2^{(2^{k-1}-1)\alpha(r)}}{2^{2^{k-1}m}(1-\epsilon)^{2^{k-1}-1}}$.

Before we explain the intuition for the above two bounds and prove them, we assume them and plug them in to finish the proof:

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

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

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

Rewriting and taking logarithms yields the claimed bound of Theorem 3.5.

It remains to prove Propositions 3.6 and Lemma 3.7. Intuitively, we will divide the product inside the self-correlation $H(\mathbf{y}^0, \mathbf{y}^1)$ into two terms, where one term is $\mu(z^{\mathbf{u}}_1, \ldots, z^{\mathbf{u}}_m) g(z^{\mathbf{u}}_1, \ldots, z^{\mathbf{u}}_m)$ for some point $\mathbf{u} \in \{0,1\}^{k-1}$, and the other consists of all $\mu(z^{\mathbf{u}}_1, \ldots, z^{\mathbf{u}}_m) g(z^{\mathbf{u}}_1, \ldots, z^{\mathbf{u}}_m)$ for all the other $\mathbf{u}$'s. Then by Definition 3.4, for random $\mathbf{x}$, the input bits $z^{\mathbf{u}}_1, \ldots, z^{\mathbf{u}}_m$ given to the first term and the input bits given to the second term only depend on each other in the bad coordinates. Given this, Proposition 3.6

translates the self-correlation $H(\mathbf{y}^0, \mathbf{y}^1)$ into a correlation of g with a polynomial of degree at most $r_\psi(\mathbf{y}^0, \mathbf{y}^1) < d$ and hence this correlation is zero. Also, given this bound $r = r_\psi(\mathbf{y}^0, \mathbf{y}^1)$ on the number of dependent coordinates, Lemma 3.7 bounds each $\mu(\cdot)g(\cdot)$ by $\mu(\cdot)$ and uses the fact that μ is a max-smooth probability distribution to obtain a bound on the self-correlation as a function of r.

We derive an expansion for $H(\mathbf{y}^0, \mathbf{y}^1)$ that we will use to prove both bounds. For fixed $\mathbf{y}^0, \mathbf{y}^1 \in D_\psi^{(m)}$ and uniformly random $\mathbf{x} \in \{0,1\}^{m \times s}$, let $\mathbf{Z} = \mathbf{Z}^{0\ldots0}\mathbf{Z}^{0\ldots1} \cdots \mathbf{Z}^{1\ldots1}$ be the joint distribution induced on $\{\mathbf{z}^\mathbf{u}\}_{\mathbf{u}\in\{0,1\}^{k-1}}$, where $z_i^\mathbf{u} = \psi(\mathbf{x}_i, \mathbf{y}_{*,i}^\mathbf{u})$ for every $\mathbf{u} \in \{0,1\}^{k-1}$ and $i \in [m]$. By construction, $\mathbf{z}^\mathbf{u}$ is uniformly distributed in $\{0,1\}^m$ for any $\mathbf{u} \in \{0,1\}^{k-1}$ so each $\mathbf{Z}^\mathbf{u}$ is a uniform distribution. Also from Definition 3.4, for any index i at which $(\mathbf{y}^0, \mathbf{y}^1)$ is good, the set of 2^{k-1} random variables $\{z_i^\mathbf{u}\}_{\mathbf{u}\in\{0,1\}^{k-1}}$ are mutually independent. Since $R_\psi(\mathbf{y}^0, \mathbf{y}^1)$ consists of those $i \in [m]$ that are bad for $(\mathbf{y}^0, \mathbf{y}^1)$, conditioned on each fixed value of $\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)} = (\mathbf{x}_i)_{i\in R_\psi(\mathbf{y}^0,\mathbf{y}^1)}$, the random variables $\{\mathbf{z}^\mathbf{u}\}_{\mathbf{u}\in\{0,1\}^{k-1}}$ are mutually independent. Then, by taking the bad indices out of the expectations, we have

$$\begin{aligned}
H(\mathbf{y}^0, \mathbf{y}^1) &= \left| \mathbf{E}_{\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)}} \mathbf{E}_{\mathbf{z}^{0\ldots0}\ldots\mathbf{z}^{1\ldots1}\sim(\mathbf{Z}|\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)})} \left[\prod_{\mathbf{u}\in\{0,1\}^{k-1}} \mu(\mathbf{z}^\mathbf{u}) g(\mathbf{z}^\mathbf{u}) \right] \right| \\
&= \left| \mathbf{E}_{\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)}} \prod_{\mathbf{u}\in\{0,1\}^{k-1}} \mathbf{E}_{\mathbf{z}^u\sim(\mathbf{Z}^\mathbf{u}|\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)})} [\mu(\mathbf{z}^\mathbf{u}) g(\mathbf{z}^\mathbf{u})] \right| \\
&= \Biggl| \mathbf{E}_{\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)}} \mathbf{E}_{\mathbf{z}^{0\ldots0}\sim(\mathbf{Z}^{0\ldots0}|\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)})} [\mu(\mathbf{z}^{0\ldots0}) g(\mathbf{z}^{0\ldots0})] \\
&\qquad \times \prod_{u\neq0\ldots0} \mathbf{E}_{\mathbf{z}^u\sim(\mathbf{Z}^\mathbf{u}|\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)})} [\mu(\mathbf{z}^\mathbf{u}) g(\mathbf{z}^\mathbf{u})] \Biggr|. \qquad (3.5)
\end{aligned}$$

Proof of Proposition 3.6. Since $(\mathbf{y}^0, \mathbf{y}^1)$ will be fixed, for simplicity write $R = R_\psi(\mathbf{y}^0, \mathbf{y}^1)$ and define $\gamma_{\mathbf{x}_R} = \gamma_{\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)}} = \prod_{\mathbf{u}\neq0\ldots0} \mathbf{E}_{\mathbf{z}^\mathbf{u}\sim(\mathbf{Z}^\mathbf{u}|\mathbf{x}_{R_\psi(\mathbf{y}^0,\mathbf{y}^1)})}[\mu(\mathbf{z}^\mathbf{u})g(\mathbf{z}^\mathbf{u})]$. We can then rewrite (3.5) as

$$H(\mathbf{y}^0, \mathbf{y}^1) = \left| \mathbf{E}_{\mathbf{x}_R}\ \gamma_{\mathbf{x}_R} \mathbf{E}_{\mathbf{z}^{0\ldots0}\sim(\mathbf{Z}^{0\ldots0}|\mathbf{x}_R)} [\mu(\mathbf{z}^{0\ldots0}) g(\mathbf{z}^{0\ldots0})] \right|.$$

We also write $\psi(\mathbf{x}_R)$ as shorthand for the length-r vector $(\psi(\mathbf{x}_i, \mathbf{y}_{*,i}^{0\ldots0}))_{i\in R}$. The dependence of $\mathbf{Z}^{0\ldots0}$ on $\mathbf{x}_R$ is given entirely by $\mathbf{Z}_R^{0\ldots0}$ so we group the terms based on the value of $\mathbf{z}_R^{0\ldots0} = \psi(\mathbf{x}_R)$. For each choice of $\mathbf{z}_R^{0\ldots0}$, let $\Pr[\mathbf{z}_R^{0\ldots0}] = \sum_{\mathbf{x}_R,\psi(\mathbf{x}_R)=\mathbf{z}_R^{0\ldots0}} \Pr[\mathbf{x}_R]$, and define

$$h(\mathbf{z}_R^{0\ldots0}) = (1/\Pr[\mathbf{z}_R^{0\ldots0}]) \sum_{\mathbf{x}_R,\psi(\mathbf{x}_R)=\mathbf{z}_R^{0\ldots0}} \Pr[\mathbf{x}_R]\gamma_{\mathbf{x}_R}.$$

We now obtain

$$\begin{aligned}
H(\mathbf{y}^0, \mathbf{y}^1) &= \left| \sum_{\mathbf{z}_R^{0\ldots0}} \Pr[\mathbf{z}_R^{0\ldots0}] h(\mathbf{z}_R^{0\ldots0}) \mathbf{E}_{\mathbf{z}^{0\ldots0}\sim(\mathbf{Z}^{0\ldots0}|\mathbf{x}_R)} [\mu(\mathbf{z}^{0\ldots0}) g(\mathbf{z}^{0\ldots0})] \right. \\
&= \left| \sum_{\mathbf{z}_R^{0\ldots0}} \Pr[\mathbf{z}_R^{0\ldots0}] h(\mathbf{z}_R^{0\ldots0}) \mathbf{E}_{\mathbf{z}^{0\ldots0}\sim(\mathbf{Z}^{0\ldots0}|\mathbf{z}_R^{0\ldots0})} [\mu(\mathbf{z}^{0\ldots0}) g(\mathbf{z}^{0\ldots0})] \right| \\
&= \left| \mathbf{E}_{\mathbf{z}^{0\ldots0}\sim\mathbf{Z}^{0\ldots0}} [\mu(\mathbf{z}^{0\ldots0}) g(\mathbf{z}^{0\ldots0}) h(\mathbf{z}_R^{0\ldots0})] \right|.
\end{aligned}$$

Since $\mathbf{Z}^{0\dots0}$ is the uniform distribution and $h(\cdot)$ is a polynomial that depends only on $r = |R| < d$ coordinates, by the orthogonality property g under μ with low-degree polynomials, $H(\mathbf{y}^0, \mathbf{y}^1) = 0$. ◻

Proof of Lemma 3.7. For convenience of notation we assume without loss of generality that $R = R_\psi(\mathbf{y}^0, \mathbf{y}^1) = \{1, \dots, r\}$. Since g is ±1-valued, we simplify (3.5) to yield

$$\begin{aligned} H(\mathbf{y}^0, \mathbf{y}^1) &= \Bigg| \mathbf{E}_{\mathbf{x}_R} \mathbf{E}_{\mathbf{z}^{0\dots0} \sim (\mathbf{Z}^{0\dots0} | \mathbf{x}_R)} [\mu(\mathbf{z}^{0\dots0}) g(\mathbf{z}^{0\dots0})] \\ &\qquad \times \prod_{\mathbf{u} \neq 0\dots0} \mathbf{E}_{\mathbf{z}^u \sim (\mathbf{Z}^{\mathbf{u}} | \mathbf{x}_R)} [\mu(\mathbf{z}^{\mathbf{u}}) g(\mathbf{z}^{\mathbf{u}})] \Bigg| \\ &\leq \Bigg| \mathbf{E}_{\mathbf{x}_R} \mathbf{E}_{\mathbf{z}^{0\dots0} \sim (\mathbf{Z}^{0\dots0} | x_R)} [\mu(\mathbf{z}^{0\dots0}) g(\mathbf{z}^{0\dots0})] \Bigg| \\ &\qquad \times \max_{\mathbf{x}_R} \Bigg| \prod_{\mathbf{u} \neq 0\dots0} \mathbf{E}_{\mathbf{z}^u \sim (\mathbf{Z}^{\mathbf{u}} | \mathbf{x}_R)} [\mu(\mathbf{z}^{\mathbf{u}}) g(\mathbf{z}^{\mathbf{u}})] \Bigg| \\ &= \Big| \mathbf{E}_{\mathbf{x}} [\mu(\mathbf{z}^{0\dots0}) g(\mathbf{z}^{0\dots0})] \Big| \\ &\qquad \times \max_{\mathbf{x}_1, \dots, \mathbf{x}_r} \Bigg| \prod_{\mathbf{u} \neq 0\dots0} \mathbf{E}_{\mathbf{x}_{r+1}, \dots, \mathbf{x}_m} [\mu(\mathbf{z}^{\mathbf{u}}) g(\mathbf{z}^{\mathbf{u}})] \Bigg| \end{aligned}$$

now since g is ±1-valued,

$$\leq \mathbf{E}_{\mathbf{x}} [\mu(\mathbf{z}^{0\dots0})] \tag{3.6}$$

$$\times \max_{\mathbf{x}_1, \dots, \mathbf{x}_r} \prod_{\mathbf{u} \neq 0\dots0} \mathbf{E}_{\mathbf{x}_{r+1} \dots \mathbf{x}_m} \big[\mu(\mathbf{z}^{\mathbf{u}})\big], \tag{3.7}$$

where $\mathbf{z}_i^{\mathbf{u}} = \psi(\mathbf{x}_i, \mathbf{y}_{*,i}^{\mathbf{u}})$ for all $i \in [m]$.

We first consider line (3.6). For $\mathbf{x}$ chosen uniformly from $\{0,1\}^{ms}$, by assumption on ψ, for any $\mathbf{u} \in \{0,1\}^{k-1}$ the random variable $\mathbf{z}^{\mathbf{u}}$ is uniform in $\{0,1\}^m$. In particular, $\mathbf{E}_{\mathbf{x}}[\mu(\mathbf{z}^{0\dots0})] = \mathbf{E}_{\mathbf{z} \in \{0,1\}^m}[\mu(\mathbf{z})]$. Further, since μ is a distribution, $\mathbf{E}_{\mathbf{z} \in \{0,1\}^m}[\mu(\mathbf{z})] = 2^{-m}$.

We now bound the remaining terms using the max-smoothness property of the distribution μ. (This is the one place where we use this property.) First we have

$$\max_{\mathbf{x}_1, \dots, \mathbf{x}_r} \prod_{\mathbf{u} \neq 0\dots0} \mathbf{E}_{\mathbf{x}_{r+1} \dots \mathbf{x}_m} \big[\mu(\mathbf{z}^{\mathbf{u}})\big] \leq \prod_{\mathbf{u} \neq 0\dots0} \max_{\mathbf{x}_1, \dots, \mathbf{x}_r} \mathbf{E}_{\mathbf{x}_{r+1} \dots \mathbf{x}_m} \big[\mu(\mathbf{z}^{\mathbf{u}})\big].$$

Fixing $\mathbf{x}_1, \dots, \mathbf{x}_r$ fixes the values of $z_1^{\mathbf{u}}, \dots, z_r^{\mathbf{u}}$ and by our assumption on ψ, for uniformly random $\mathbf{x}_{r+1}, \dots, \mathbf{x}_m$ the values of $z_{r+1}^{\mathbf{u}}, \dots, z_m^{\mathbf{u}}$ are uniformly random. Therefore line (3.7) is upper bounded by

$$\prod_{\mathbf{u} \neq 0\dots0} \max_{z_1^{\mathbf{u}}, \dots, z_r^{\mathbf{u}}} \mathbf{E}_{z_{r+1}^{\mathbf{u}} \dots z_m^{\mathbf{u}}} \big[\mu(z^{\mathbf{u}})\big] = \big(\max_{z_1, \dots, z_r} \mathbf{E}_{z_{r+1} \dots z_m} \big[\mu(\mathbf{z})\big]\big)^{2^{k-1}-1}.$$

By the property of μ implied by Lemma 3.3,

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

and therefore line (3.7) is at most $(2^{\alpha(r)-m}/(1-\epsilon))^{2^{k-1}-1}$. The lemma follows immediately by combining the bounds for lines (3.6) and (3.7). ◻

We have completed the proof of Theorem 3.5. $\square$

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 large only if many widely spread restrictions of the function 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 $\mathcal{R}^r_m := \{\rho \in \{0,1,*\}^m : |\rho| = m-r\}$.

To motivate the name "α-spread" in the following definition, the reader can think of the function α as having $\alpha(z) \le z^{\alpha_0}$ for all large enough z and fixed constant α_0, and thus the probability bound below intuitively says that ν behaves like a uniform distribution on $\{0,1\}^m$. Thus, the support of ν is very spread out.

Definition 4.1. *Given $\alpha : \{0, \dots, m\} \mapsto \mathbb{R}$, 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 $\alpha(r) \ge r$, $deg_\epsilon(f) = deg^*_{\epsilon,\alpha}(f)$ since in this case 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 4.2. *Let $f : \{0,1\}^m \mapsto \{-1,1\}$ and $\alpha : \{0,\dots,m\} \mapsto \mathbb{R}$. 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 $R(\mathbf{x}) = \sum_{\rho \parallel \mathbf{x}} 2^{|\rho|-\alpha(|\rho|)}\lambda_\rho > 1$ whenever $|f(\mathbf{x}) - q(\mathbf{x})| > \epsilon$.

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

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

Then, since ν is α-spread,

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

Therefore there exists a restriction π in the support of ν for which $I_\pi \leq 1$. If there exists an input $\mathbf{x}$ such that $|f(\mathbf{x}) - q(\mathbf{x})| > \epsilon$ and $\mathbf{x} \in C_\pi$, then since

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

we would have $I_\pi > 1$. Thus, $|f(\mathbf{x}) - q(\mathbf{x})| \leq \epsilon$ for every $\mathbf{x} \in C_\pi$. But since the degree of q is $\leq d$ this violates our assumption that $deg_\epsilon(f|_\pi) > d$. Hence 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 the following lemma, we define a more involved α-spread family of restrictions.

Lemma 4.3. *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 $\mathcal{R}^{pr}_m$ 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 4.3 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}.

With the family of restrictions in Lemma 4.3, we will use the following approach in the rest of this section to obtain functions with large (ϵ, α)-approximate degree. Let $q > r > 0$, where r is polynomial in q, and G be any circuit on q bits such that any *projection* of G on any r input bits has large ϵ-approximate degree, where a projection of G on a set S of input bits, denoted by G_S, is a new circuit obtained from G by keeping only those gates on paths from the inputs in S to the output gate. For some $p = O(\log q)$, we will produce another circuit H on pq bits such that for any restriction π in the family of restrictions in Lemma 4.3, the circuit $H|_\pi$ contains as a subcircuit a projection of G on some r bits. Thus H has large (ϵ, α)-approximate degree. Next we give a simple demonstration of this construction.

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

Lemma 4.4. *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 $2\lceil q^{1-\beta} \rceil < 2^p \leq \frac{1}{6} q^{\alpha_0 + \epsilon - 1} \ln 2$. 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 4.3, 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 $\nu(\pi) > 0$, $\textsc{Or}_r$ is a subfunction of $\textsc{Tribes}_{p,q}|_\pi$ so $deg_{5/6}(\textsc{Tribes}_{p,q}|_\pi) \geq deg_{5/6}(\textsc{Or}_r) \geq \sqrt{r/12}$. Thus, $deg_{5/6,\alpha}(\textsc{Tribes}_{p,q}) \geq deg^*_{5/6,\alpha}(\textsc{Tribes}_{p,q}) \geq \sqrt{r/12}$. □

In particular, with $\epsilon = 0.4, \beta = 0.8, \alpha_0 = 0.9$, we get the following corollary.

COROLLARY 4.5. *For sufficiently large p and $q = 2^{4p}$, if $\alpha : \{0, \ldots, m\} \mapsto \mathbb{R}$ satisfies $\alpha(z) = z^{0.9}$ for $z \geq (3p \ln 2)^{10}$ and $\alpha(z) = z$ otherwise, then $deg_{5/6,\alpha}(\textsc{Tribes}_{p,q}) \geq q^{3/10}/\sqrt{12} = 2^{6p/5}/\sqrt{12}$.*

Corollary 4.5 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 4.7, a circuit G that has large threshold degree when projected on any sufficiently large set of input bits. The lemma uses the following property of the threshold degree of (the dual of) the Minsky–Papert function [26].

PROPOSITION 4.6. *Let $\mathrm{MP}_{q,q'} : \{0,1\}^{q \cdot q'} \mapsto \{-1, 1\}$ be defined by $\mathrm{MP}_{q,q'}(\mathbf{x}) := \wedge_{i=1}^{q} \vee_{j=1}^{q'} x_{ij}$. Then for any $d > 0$, $deg^{\pm}(\mathrm{MP}_{d,4d^2}) \geq d$.*

LEMMA 4.7. *Let r, d, s, and t be positive integers such that $st \geq r \geq 2ds$ and $s/(4d) \geq t$. Then there is an explicit read-once $\mathsf{AND} \circ \mathsf{OR}$ formula G on st 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 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. Then we can easily bound the size of S by $r \leq s|A| + 4d^2(t - |A|)$, 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 $\mathrm{MP}_{d,4d^2}$ as a subfunction. By Proposition 4.6, $deg^{\pm}(G_S) \geq deg^{\pm}(\mathrm{MP}_{d,4d^2}) \geq d$. □

Using the $\mathsf{AND} \circ \mathsf{OR}$ formula G on $q = st$ bits given by Lemma 4.7, we now construct the circuit H of large α-threshold degree. Let $H' = G \circ (\textsc{And}_p)$ be the circuit obtained from G by replacing each input bit with an $\textsc{And}$ gate on p bits for some $p > 0$. For the choice of G from Lemma 4.7, H' is a $\mathsf{AND} \circ \mathsf{OR} \circ \mathsf{AND_p}$ circuit on pq bits. We then obtain the circuit H by applying the following steps to each $\textsc{Or}$ gate φ of H':

1. Let s be the number of $\textsc{And}_p$ gates fed into φ. For every $i \in [s]$,
   - let the input bits to the ith $\textsc{And}_p$ gate be $z_{i,1}, \ldots, z_{i,p}$, then create two new $\textsc{Or}$ gates $B_i = \vee_{j=1}^{p} z_{i,j}$ and $B'_i = \vee_{j=1}^{p} (\neg z_{i,j})$.
2. Create a new $\textsc{And}$ gate $A_\varphi = \bigwedge_{i=1}^{s} (B_i \wedge B'_i)$.
3. Finally, add a new edge feeding the output of A_φ to φ.

The following lemma justifies the construction.

LEMMA 4.8. *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 as described above. Then the following hold:*

- *H has pq input bits divided into q blocks of p bits each, where each block corresponds to an input bit in G.*
- *H is a depth 4 $\mathsf{AND} \circ \mathsf{OR} \circ \mathsf{AND} \circ \mathsf{OR}$ circuit of size at most four times the size of G.*
- *Let π be any restriction that chooses a subset S' of the blocks of inputs to H, and hence the corresponding set S of input bits to G, to leave unset and assigns values from $\{0,1\}^p - \{0^p, 1^p\}$ to each other block. Then $H|_\pi$ contains G_S as a subcircuit.*

Proof. A subcircuit of depth 2 with three gates is added for each last level Or in G so the first two parts are immediate.

For the last part, note that for any block not in S', the associated And$_p$ gate in H is forced to 0, and the associated B and B' gates (in step 1 in the construction) are forced to 1. Let φ be any Or gate in G which becomes an Or gate φ_H in H. Then,

- if φ does not have any input bit in S, then all the B and B' gates under A_φ output 1, and hence A_φ outputs 1 and hence φ_H outputs 1 (to the top And gate);
- if φ has an input bit in S, then setting the values of the corresponding block in S' to 0^p or 1^p will force one of the associated B or B' gate to 0 and hence force A_φ to output 0.

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 4.9. *For any p sufficiently large multiple of 15 and $q = 2^{4p}$, if $\alpha : \{0, \ldots, m\} \mapsto \mathbb{R}$ is defined as $\alpha(z) = z^{0.9}$ for $z \geq (3p \ln 2)^{10}$ and $\alpha(z) = z$ otherwise, then there is an explicit depth 4 $\mathsf{AC^0}$ 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 these are integral and they satisfy the conditions of Lemma 4.7. We can apply that lemma to derive an $\mathsf{AND} \circ \mathsf{OR}$ circuit G with the property that for every S with $|S| = r$, $deg^{\pm}(G_S) \geq d$.

Define the distribution ν as in the statement of Lemma 4.3 given the value of r and $w = \lceil \log^{20} pq \rceil$. We can then apply Lemma 4.8 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 $deg^{\pm}(H|_\pi) \geq deg^{\pm}(G_S) \geq d$.

Note that for $\alpha_0 = 0.9$ and $\beta = 0.8$, all the conditions of Lemma 4.3 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 $\mathsf{AC^0}$. Together with the functions from the previous section, Theorem 3.5 is sufficient to improve the lower bounds for $\mathsf{AC^0}$ 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 new selector function ψ, which we denote by $\textsc{Index}_{\oplus^a_{k-1}}$ where $a > 0$ is an integer. This function has $s = 2^a$ and $D_{\textsc{Index}_{\oplus^a_{k-1}}, j} = \{0,1\}^s$ for all j. For $\mathbf{x} \in \{0,1\}^s$ and $\mathbf{y} \in \{0,1\}^{(k-1)s}$ define

$$\textsc{Index}_{\oplus^a_{k-1}}(\mathbf{x}, \mathbf{y}) = \mathbf{x}_{(\mathbf{y}_1 \oplus \ldots \oplus \mathbf{y}_{k-1})_{[a]}},$$

where the bits in $\mathbf{x}$ are indexed by a-bit vectors and $\mathbf{y}_{[a]}$ denotes the vector of the first a bits of $\mathbf{y}$. This function clearly satisfies the selector function requirement that the output be unbiased for each fixed value of $\mathbf{y}$.

Although the definition of $\textsc{Index}_{\oplus^a_{k-1}}$ uses the parity function, in applications we will choose the number of players k that will be $O(\log n)$ and hence these parity functions will be computable in AC^0. We can express the parity of $k-1$ items in DNF as an $\vee$ of 2^{k-2} conjunctions each of length $k-1$. Thus for any $\mathbf{w} \in \{0,1\}^a$, we can check whether $(\mathbf{y}_1 \oplus \ldots \oplus \mathbf{y}_{k-1})|_{[a]} = \mathbf{w}$ by a $\mathbf{\Pi}_3$ formula where the gates are, from top to bottom, $\wedge$ with fan-in a, $\vee$ with fan-in 2^{k-2}, and $\wedge$ with fan-in $k-1$. If we add $x_{\mathbf{w}}$ as an additional input to the top $\wedge$ gate, we can make this formula output $x_{\mathbf{w}}$ if the check returns true. Therefore we can write $\textsc{Index}_{\oplus^a_{k-1}}$ as a $\mathbf{\Sigma}_4$ formula where the fan-ins are, from top to bottom, 2^a, $a+1$, 2^{k-2}, and $k-1$. The top $\vee$ gate is to do the check for every possible value of $\mathbf{w} \in \{0,1\}^a$. Alternatively, we could dually write parity using conjunctive normal form (CNF) and express $\textsc{Index}_{\oplus^a_{k-1}}$ as a $\mathbf{\Sigma}_3$ formula where the fan-ins are, from top to bottom, 2^a, $a2^{k-2}+1$, and $k-1$, where the inputs to each of the $(a2^{k-2}+1)$-fan-in $\wedge$ gates are the one bit of $\mathbf{x}$ and $a2^{k-2}$ $\vee$ gates with fan-in $k-1$.

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

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

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

Proof. In this case $D_\psi^{(m)}$ is simply $\{0,1\}^{(k-1)ms}$. For each fixed $i \in [m]$ and each fixed pair of $\mathbf{u} \neq \mathbf{v} \in \{0,1\}^{k-1}$, the probability that $\mathbf{y}_{*i}^{\mathbf{u}}$ and $\mathbf{y}_{*i}^{\mathbf{v}}$ select the same bit of $\mathbf{x}_i$ is the probability that $(\mathbf{y}_{*i}^{u_1} \oplus \cdots \oplus \mathbf{y}_{*i}^{u_{k-1}})_{[a]} = (\mathbf{y}_{*i}^{v_1} \oplus \cdots \mathbf{y}_{*i}^{v_{k-1}})_{[a]}$. Since $\mathbf{u} \neq \mathbf{v}$, 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 the $\binom{2^{k-1}}{2} < 2^{2k-3}$ pairs $\mathbf{u}, \mathbf{v} \in \{0,1\}^{k-1}$, it follows that the probability that i is bad for $(\mathbf{y}^0, \mathbf{y}^1)$ is at most $2^{2k-3}2^{-a} = 2^{2k-a-3}$. The bound follows by the independence of the choices of $(\mathbf{y}^0, \mathbf{y}^1)$ for different values of $i \in [m]$. $\square$

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

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

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

(5.1)
$$\sum_{r=d}^{m} 2^{(2^{k-1}-1)\alpha(r)} \cdot \Pr_{\mathbf{y}^0,\mathbf{y}^1 \in D_\psi^{(m)}}[r_\psi(\mathbf{y}^0,\mathbf{y}^1) = r] \leq \sum_{r=d}^{m} 2^{(2^{k-1}-1)\alpha(r)} \cdot \left(\frac{em2^{2k-a-3}}{r}\right)^r$$

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 (5.1) is

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

Plugging this into Theorem 3.5 we obtain that

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

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 $q=2^{4p}$, $k\leq p/10$, $s=2^{p+2k}$, and $m=pq$. Let $F=\textsc{Tribes}_{p,q}\circ\textsc{Index}_{\oplus^{(p+2k)}_{k-1}}$ and $F'=\textsc{Tribes}'_{p,q}\circ\textsc{Index}_{\oplus^{(p+2k)}_{k-1}}$. 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 AC^0 formulas and F' has polynomial-size depth 4 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 Σ_3 formulas with fan-in, top to bottom, of $2^a=2^{p+2k}$, $a2^{k-2}+1=(p+2k)2^{k-2}+1$, and $k-1$. Since $\textsc{Tribes}_{p,q}$ is given by a Σ_2 formula, $\textsc{Tribes}_{p,q}\circ\textsc{Index}_{\oplus^{(p+2k)}_{k-1}}$ is computable by a Σ_5 formula with fan-in top to bottom of q, p, 2^{p+2k}, $(p+2k)2^{k-2}+1$, and $k-1$. The total formula size of F is $n(p+2k+1)(k-1)2^{k-2}$ which is less than n^2.

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

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

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

Observe that the fan-ins of the $\wedge$ gates are p and $a2^{k-2}+1$, and the input to each of the $(a2^{k-2}+1)$-fan-in $\wedge$ gates are one bit of $\mathbf{x}$ and $a2^{k-2}$ $\vee$ gates with fan-in $k-1$. Moreover, the 0th player (who holds x) can evaluate each of these $\vee$ gates since it can see all 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 iff all of the $a2^{k-2}$ feeding $\vee$ gates that depend only on $\mathbf{y}_1,\dots,\mathbf{y}_{k-1}$ evaluate to true. Given this information, player 1, who sees $\mathbf{x}$, can then evaluate all p $\wedge$ gates. □

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

Proof. Observe that by Lemma 5.4, $F=\textsc{Tribes}_{p,q}\circ\textsc{Index}_{\oplus^{(p+2k)}_{k-1}}$ 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 function G

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

and that the same $O(\log^2 n)$ upper bound from Lemma 5.4 applies equally well to G. $\square$

Applying distributive law to the depth 5 function $f=\textsc{Tribes}_{p,q}\circ\textsc{Index}_{\oplus_{k-1}^{(p+2k)}}$ we derive the following exponential improvement in the number of players for which nontrivial 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 $\textsc{Tribes}_{p,q}\circ\psi_{k,\ell}$ 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 $\textsc{Disj}_{k,n}(\mathbf{x})=\vee_{i=1}^{n}\wedge_{j=0}^{k-1}x_{j,i}$. As in Corollary 5.5 start with $F=\textsc{Tribes}_{p,q}\circ\textsc{Index}_{\oplus_{k-1}^{(p+2k)}}$ with the parameters from Theorem 5.3. Unlike Corollary 5.5, however, we use the $\mathsf{\Sigma_4}$ circuit for $\textsc{Index}_{\oplus_{k-1}^{(p+2k)}}$ and reduce F to a $\mathsf{\Sigma_6}$ 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 uth 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

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

where the bits of vector $\mathbf{y}\in\{0,1\}^{nk}$ for $n=q2^{ap+(k-2)p(a+1)}$ are indexed by $j\in\{0,\dots,k-1\}$, $i\in[q]$, $I\in[2^a]^p$ 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 $\textsc{Tribes}_{p,q} \circ \textsc{Index}_{\oplus_{k-1}^{(p+2k)}}$ by executing a NOF randomized communication protocol for $\textsc{Disj}_{n,k}$ on $\mathbf{y}$ of length nk, where $n = q2^{ap+(k-2)p(a+1)} = q2^{ap(k-1)+k-2}$. Plugging in $q = 2^{4p}$ and $a = 2k + p$ for $k \leq p/10$ we have that $R^k_{1/3}(\textsc{Disj}_{n,k})$ is $\Omega(2^{6p/5-k})$. Now for these values of k and a, we have $ap \geq k - 2 + 4p$ and hence we have that $n \leq 2^{apk} \leq 2^{6p^2k/5}$. Therefore we have $p \geq \sqrt{5\log_2 n}/\sqrt{6k}$. It follows that $R^k_{1/3}(\textsc{Disj}_{n,k})$ is $\Omega(2^{\sqrt{\log_2 n}/\sqrt{k}})$ provided that $k \leq \frac{1}{10}\sqrt{5\log_2 n}/\sqrt{6k}$ which holds if $k \leq \frac{1}{5}\log_2^{1/3} n$. □

Although our bound for $\textsc{Disj}_{n,k}$ applies to exponentially more players than do the bounds in [25, 14], 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* $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ *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 3.3 to $\textsc{Or}$ since the constant function 1 approximates $\textsc{Or}$ on all but one point.

To prove lower bounds for $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ 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 4.9 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 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 4.9. We use the dual function f to f' which is therefore a Σ_4 function of the same approximate degree. Since f has 0.9-threshold degree at least $m^{1/15}/\log_2 m$, it has $(1-\epsilon, 0.9)$-approximate degree at least $d = \lceil m^{1/15}/\log_2 m\rceil$ for any $\epsilon > 0$. For $k \leq 0.1\log_2 d$, let $a = \lceil\log_2(e2^{2k-1}m/d)\rceil$, and $s = 2^a$. By Theorem 5.2, the function $H_n = f \circ \textsc{Index}_{\oplus^a_{k-1}}$ 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 $k \leq 0.1\log_2 d$, $n = msk = m2^ak$ 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 [40], which extends the ideas of Razborov and Wigderson [30], we show quasipolynomial lower bounds on the simulation of AC^0 functions by unrestricted $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits.

Theorem 6.1. *For N sufficiently large, there is a Boolean function G_N on N bits in* AC^0 *such that G_N requires* $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ *circuit size* $N^{\Omega(\log N)}$.

In the rest of this section we prove the above theorem.[5] An important ingredient in our construction is the following function F^m_t, first defined by Sipser [38], with parameters as modified by Håstad [19].

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

DEFINITION 6.2. *For $t \geq 2$, the Sipser function F_t^m 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 $\sqrt{\frac{1}{2} tm \log m}$. Therefore for t constant and large enough m, F_t^m is an AC^0 function on $O(m^t)$ inputs.*

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.[6]

PROPOSITION 6.3 (see [19]). *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 hold:*

- *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.*[7]
- *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.*

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 6.4 (see [20]).

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

(b) *If f is computed by a $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}_{k-1}$ 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 a $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuit C of size $N^{o(\log N)}$, then using suitable random restrictions as described in Proposition 6.3, 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 6.4 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, and $N = nrm'$. We construct our hard circuit G_N on N bit by replacing each bit in H_n with a xorification of r disjoint copies of F_3^n. That is, for any $Z = Z_1 \cdots Z_n$, where each $Z_i \in \{0,1\}^{r \times m'}$,

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

[6] The distribution of restrictions in Proposition 6.3 is not the usual uniform distribution on restrictions setting a fixed number of variables but is suitable for the Sipser functions.

[7] The property that any input assignment satisfies at most one of the terms is implicit in [19]. The observation that this property holds is made explicit by Berg and Ulfberg [12].

Suppose by contradiction that for some sufficiently small constant $\delta > 0$, there is a $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ 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 6.3. Let $\rho \sim \mathcal{R}_q^B$. Consider the following two events:

- Event E_1: $C|_\rho$ is computed by a $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuit of size at most $|C| \cdot (2N)^k$ where the fan-in of each AND-gate is strictly less than k.
- Event E_2: the subcircuit computing A_i in $G_N|_\rho$ contains F_2^n as a subfunction for all $1 \leq i \leq n$.

First, we show that $\Pr[\neg E_1] < 1/2$ for sufficiently small $\delta > 0$. Fix any AND-gate φ in C. By Proposition 6.3, the probability over ρ that $\varphi|_\rho$ cannot be written as a DNF with term size less than k is at most

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

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 6.3, 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 $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ 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 6.3, with probability at least $1 - (1/3)^r > 1 - 1/(2n)$, each A_i contains 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 6.4, the fact that E_1 holds implies that for any partition of the input to k players and $\epsilon = 1/(2|C| \cdot (2N)^k)$,

$$R_{1/2-\epsilon}^k(C|_\rho) = O(k \log(|C| \cdot (2N)^k)) = O(\log^3 N) = O(\log^3 n).$$

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 $R_{1/2-\epsilon}^k(C|_\rho) \geq R_{1/2-\epsilon}^k(H_n)$ 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 6.4 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 $\mathsf{SYM} \circ \mathsf{AND}$ circuits.

7. Proof of Lemma 4.3. Fix any restriction ρ of size $i = |\rho| \geq w$. We have

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

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

jth block. Then

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

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 at least one value. Hence,

$$\Pr_{\pi\sim\nu}[\rho \parallel \pi] < \frac{1}{\binom{q}{q-r}} \sum_{S\subset[q],|S|=q-r} 2^{-i_S}\left(1+\frac{1}{2^{p-1}-1}\right)^{k_S}. \tag{7.1}$$

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

Now consider case (I). Thus $i \ge q/2$. In (7.1), we have $k_S \le q$ for every S. Thus,

$$\Pr_{\pi\sim\nu}[\rho \parallel \pi] \le \frac{1}{\binom{q}{q-r}} \sum_{S\subset[q],|S|=q-r} 2^{-i_S}\left(1+\frac{1}{2^{p-1}-1}\right)^{q}.$$

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

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

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

$$\Pr_{\pi\sim\nu}[\rho \parallel \pi] \le 2^{(2i)^\beta-i}\left(1+\frac{1}{2^{p-1}-1}\right)^{q} \le 2^{(2i)^\beta-i}e^{q^\beta} \le 2^{2^\beta(1+1/\ln 2)i^\beta - i},$$

since $q^{1-\beta} \le 2^{p-1}-1$ and $i \ge q/2$. We upper bound the term $2^\beta(1+1/\ln 2)\, i^\beta$ by i^{α_0} as follows. Since $i \ge w$,

$$i^{\alpha_0-\beta} \ge w^{\alpha_0-\beta} \ge 3p/\ln 2 \tag{7.2}$$

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

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

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

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

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

So,

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

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 coordinates in ρ as balls placed in corresponding bins, then we observe that $\mathcal{S}$ can increase only 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 without loss of generality 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

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

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 (7.2), $i^{\alpha_0-\beta} \geq 3p/\ln 2$ so $i^{\beta}p^{1-\beta}/\ln 2 \leq i^{\alpha_0}/3$. For the second term again by (7.2) 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 $\Pr_{\pi\sim\nu}[\rho \parallel \pi] < 2^{i^{\alpha_0}-i}$ as required. It only remains to prove the claim.

Proof of Claim 7.1. 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

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

since $k \le q/2$.

In the case that $r \le k \le q/2$ 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 $q-r \ge q/2 \ge k$ 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

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

since $r \le q/2$. □

8. Discussion. In this work we have proved the first 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 $1/2$, functions computed by polynomial-size depth 6 circuits suffice. It is interesting to reduce the circuit depth required for these lower bounds.

A particularly important function for further investigation is the depth 2 function set intersection. Our new randomized lower bound for the set-intersection function is nontrivial only for up to $\Theta(\log^{1/3} n)$ players and is subpolynomial. It is consistent with our current knowledge that the set-intersection function requires randomized polynomial communication complexity even for $\Omega(\log n)$ players. The difficulty of extending our lower bound methods to prove stronger bounds for this function is our inability to apply the "max-smooth" Lemma 3.3 to $f = \textsc{Or}$ since the constant function 1 approximates $\textsc{Or}$ on all but one point, and hence $\textsc{Or}$ has zero (ϵ, α)-approximate degree for any interesting α.

Appendix 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 depth 2 and depth 5 AC^0 functions that are in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ 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 function that is in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for k up to $\Theta(\log n / \log\log n)$.

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 pattern tensor selection function $\psi_{k,\ell}$ as captured using the definition of r_ψ from section 3.

Proposition A.1 (see [14, 25]). *If $\psi = \psi_{k,\ell}$, then*

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

Proof. In the case, $D_\psi^{(m)} = D_{\psi_{k,\ell}}^{(m)}$ is $[\ell]^{m(k-1)s}$. $z_i^{\mathbf{u}} = \psi(\mathbf{x}_i, \mathbf{y}_{*i}^{\mathbf{u}})$ for $\mathbf{u} \in \{0,1\}^{k-1}$ will be independent iff $\mathbf{y}_{*i}^{\mathbf{u}}$ and $\mathbf{y}_{*i}^{\mathbf{v}}$ select different bits of $\mathbf{x}_i$ for every $\mathbf{u} \ne \mathbf{v}$. This will be true for $\mathbf{u}$ and $\mathbf{v}$ iff there is some $j \in [k-1]$ such that $y_{ji}^{\mathbf{u}} \ne y_{ji}^{\mathbf{v}}$. However, since this must hold for every $\mathbf{u}$ and $\mathbf{v}$, in particular those that agree everywhere except for a single bit, it is necessary and sufficient for independence that $y_{ji}^0 \ne y_{ji}^1$ for every $j \in [k-1]$. Therefore $r_{\psi_{k,\ell}}(\mathbf{y}^0,\mathbf{y}^1)$ 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 $i \in [m]$ $\Pr_{\mathbf{y}^0,\mathbf{y}^1 \in D_\psi^{(m)}}[r_\psi(\mathbf{y}^0,\mathbf{y}^1) = r] \le \binom{m}{r}(\frac{k-1}{\ell})^r \le (\frac{em(k-1)}{r\ell})^r$. □

Remark. The lower bounds in [14, 25] use the above property of $\psi = \psi_{k,\ell}$ and follow the same general outline as in Theorem 3.5 but instead of being able to use Lemma 3.7, 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 (see [14, 25]). *If* $r = r_\psi(\mathbf{y}^0, \mathbf{y}^1)$, *then* $H(\mathbf{y}^0, \mathbf{y}^1) \leq \frac{2^{(2^{k-1}-1)r}}{2^{2^{k-1}}m}$.

In [14, 25], to prove the lower bound for $\textsc{Disj}_{k,n}$, the function f is set to $\textsc{Or}_m$ and ψ is $\psi_{k,\ell}$. By Proposition 2.3, $d = deg_{5/6}(\textsc{Or}_m) \geq \sqrt{m/12}$. Plugging the bound in Proposition A.1 together with the bounds from Proposition 3.6 for $r < d$ and from Proposition A.2 when $r \geq d$ into the correlation inequality, it is not hard to show that $R^k_{1/3}(f \circ \psi) \geq d/2^k - O(1)$ for $\ell > \frac{2^{2^k} kem}{d}$. Hence for suitable $k = O(\log\log n)$ they derive lower bounds on $R^k_{1/3}(\textsc{Disj}_{k,n})$.

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. *Let* $\alpha : \{0, \ldots, m\} \mapsto \mathbb{R}$, $1 > \alpha_0 > 0$, $d > 0$, *such that* $\alpha(r) \leq r^{\alpha_0}$ *for all* $r \geq d$. *For any function* $f : \{0,1\}^m \mapsto \{1, -1\}$ *with* $deg_{5/6,\alpha}(f) \geq d$, *the function* $f \circ \psi_{k,\ell}$ *defined on kn bits, where* $n = ms$ *for* $s \geq \lceil \frac{4e(k-1)m}{d} \rceil^{k-1}$, *requires* $R^k_{1/3}(f \circ \psi_{k,\ell}) > d/2^k - 3$ *for* $k \leq (1 - \alpha_0)\log_2 d$.

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

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

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 the last quantity 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 3.5 we obtain that

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

as required. ▯

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

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

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

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

is a depth 3 formula since the bottom layer of $\vee$ gates in $\textsc{Tribes}'_{p,q}$ 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 $s = \lceil 8\sqrt{3}e(k-1)pq^{(1+\epsilon)/2}\rceil^{k-1}$ and $n = pqs$. Then $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \psi_{k,\ell})$ and $R^k_{1/3}(\textsc{Tribes}'_{p,q} \circ \psi_{k,\ell})$ 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 $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \psi_{k,\ell})$ and $R^k_{1/3}(\textsc{Tribes}'_{p,q} \circ \psi_{k,\ell})$ of $\Omega(n^{(1-\delta)/(k+1)}/(2^k \log n))$.

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

By Lemma 4.4, 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 $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \psi_{k,\ell})$ 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

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

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 (A.1), $q^{(1-\epsilon)/2}$ 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, $s = \lceil 8\sqrt{3}e(k-1)pq^{0.7}\rceil^{k-1}$, and $n = ms = pqs$. Then $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \psi_{k,\ell})$ and $R^k_{1/3}(\textsc{Tribes}'_{p,q} \circ \psi_{k,\ell})$ 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 $\mathsf{AC^0}$ functions in $\mathsf{NP^{cc}_k}-\mathsf{BPP^{cc}_k}$ for $k = \Theta(\log n/\log\log n)$.

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 $\mathbf{x} \in \{0,1\}^s$ and $\mathbf{y} \in \{0,1\}^{(k-1)s}$

$$\psi_{k,\ell}^{\oplus b}(\mathbf{x},\mathbf{y}) = \bigoplus_{b'=1}^{b} \bigvee_{s'=1}^{\ell^{k-1}} \left(x_{b's'} \wedge \bigwedge_{j=1}^{k-1} y_{jb's'} \right).$$

This function clearly satisfies the selector function requirement that the output be unbiased for each fixed value of $\mathbf{y}$.

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 pattern tensors of the form $\psi_{k,b'}(\mathbf{x},\mathbf{y}) = \bigvee_{s'=1}^{\ell^{k-1}}(x_{b's'} \wedge \bigwedge_{j=1}^{k-1} y_{jb's'})$ 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 $\overline{\psi}_{k,b'}(\mathbf{x},\mathbf{y}) = \bigvee_{s'=1}^{\ell^{k-1}}(\overline{x}_{b's'} \wedge \bigwedge_{j=1}^{k-1} y_{jb's'})$.

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, ℓ^{k-1}, 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}, $b\ell^{k-1}$, and k.

When ψ is $\psi_{k,\ell}^{\oplus b}$, the variables $\psi_{k,\ell}^{\oplus b}(\mathbf{x}_i, \mathbf{y}_{*i}^{\mathbf{u}})$ for $\mathbf{u} \in \{0,1\}^{k-1}$ will be independent if and only if for every $\mathbf{u} \neq \mathbf{v}$ there is some $b' \in [b]$ such that $\mathbf{y}_{*ib'}^{\mathbf{u}}$ and $\mathbf{y}_{*ib'}^{\mathbf{v}}$ 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 $r_{\psi_{k,\ell}^{\oplus b}}(\mathbf{y}^0,\mathbf{y}^1)$ 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* $\Pr_{\mathbf{y}^0,\mathbf{y}^1 \in D_\psi^{(m)}}[r_\psi(\mathbf{y}^0,\mathbf{y}^1) = r] \leq \binom{m}{r}(\frac{k-1}{\ell})^{br} \leq (\frac{em(k-1)^b}{r\ell^b})^r$.

Proof. In this case $r_{\psi_{k,\ell}^{\oplus b}}(\mathbf{y}^0,\mathbf{y}^1)$ 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 $(\mathbf{y}^0,\mathbf{y}^1)$ 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 $(\frac{k-1}{\ell})^b$. The bound follows by the independence of the choices of $(\mathbf{y}^0,\mathbf{y}^1)$ 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. *Let $\alpha : \{0,\ldots,m\} \mapsto \mathbb{R}$, $1 > \alpha_0 > 0$, $d > 0$, such that $\alpha(r) \leq r^{\alpha_0}$ for all $r \geq d$. For any function $f : \{0,1\}^m \mapsto \{1,-1\}$ with $deg_{5/6,\alpha}(f) \geq d$, the function $f \circ \psi_{k,\ell}^{\oplus b}$ defined on kn bits, where $n = ms$ and $s = b\lceil (k-1)(4em/d)^{1/b}\rceil^{k-1}$, requires that $R_{1/3}^k(f \circ \psi_{k,\ell}^{\oplus b}) \geq d/2^k - 3$ for $k \leq (1-\alpha_0)\log_2 d$.*

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

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

Since $k \le (1-\alpha_0)\log_2 d$, we have $(2^{k-1}-1)\alpha(r) < d^{1-\alpha_0}\alpha(r) \le r$ for $r \ge d$ so (A.2) is

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

Plugging this into Theorem 3.5 we obtain that

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

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

We first directly apply Theorem A.7 to $\textsc{Tribes}_{p,q} \circ \psi^{\oplus b}_{k,\ell}$ for suitable values of b.

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

Proof. Let $m = pq$. By Lemma 4.4, 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 $r \ge d$. Thus $4em/d \le 16epq^{(1+\epsilon)/2}$, so by the choice of b we have $(4em/d)^{1/b} \le 2$. Therefore $s = b(2k)^{k-1} \ge b\lceil (k-1)(4em/d)^{1/b}\rceil^{k-1}$. Also $k \le \frac{1}{2}(1-\alpha_0)(1-\epsilon)\log_2 q - 2$ implies that $k \le (1-\alpha_0)\log_2 d$. Applying Theorem A.7, we see that $R^k_{1/3}(\textsc{Tribes}_{p,q} \circ \psi^{\oplus b}_{k,\ell})$ is $\Omega(q^{(1-\epsilon)/2}/2^k)$. □

In particular we obtain the following.

Corollary A.9. *Let p be a sufficiently large integer, $q = 2^{4p}$, $k \le p/40$, 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 $F = \textsc{Tribes}_{p,q} \circ \psi^{\oplus b}_{k,\ell}$. 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 4.5 instead of Lemma 4.4. As noted above, $\psi^{\oplus b}_{k,\ell}$ has $\mathsf{\Pi}_3$ formulas with fan-in, top to bottom, of $2^{b-1} = 2^{p-1}$, $bs = ps$, and k. Since $\textsc{Tribes}_{p,q}$ is given by a $\mathsf{\Sigma}_2$ formula, $\textsc{Tribes}_{p,q} \circ \psi^{\oplus b}_{k,\ell}$ is computable by a $\mathsf{\Sigma}_4$ formula with fan-in top to bottom of q, $p2^{p-1}$, ps, and k. The total formula size of F is $np2^{p-1}$, which is less than $n^{5/4}\log_2 n$. □

Lemma A.10. *$N^k(\textsc{Tribes}_{p,q} \circ \psi^{\oplus b}_{k,\ell})$ is $O(\log q + pb\log s)$.*

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

The 0th player (who holds x), guesses an accepting subtree of this formula and sends both the description of the subtree and the values of the bits of $\mathbf{x}$ 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 $\log_2 q + p[\log_2 2^{b-1} + b\log_2 s] \le \log_2 q + pb(\log_2 s + 1)$ and the number of bits of $\mathbf{x}$ 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 $k\log k \le a\log n$ for some constant $a > 0$.*

Proof. Observe that $F = \textsc{Tribes}_{p,q} \circ \psi_{k,\ell}^{\oplus b}$ with the parameters from Corollary A.9 by Lemma A.10 has $N^k(F)$ that is $O(\log^3 n \log\log n)$ 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 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. $\square$

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

REFERENCES

[1] M. Ajtai, Σ_1^1*-formulae on finite structures*, Ann. Pure Appl. Logic, 24 (1983), pp. 1–48.

[2] E. W. Allender, *A note on the power of threshold circuits*, in Proceedings of the 30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, NC, IEEE, 1989, pp. 580–584.

[3] L. Babai, A. Gál, P. G. Kimmel, and S. V. Lokam, *Communication complexity of simultaneous messages*, SIAM J. Comput., 33 (2003), pp. 137–166.

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

[5] L. Babai, N. Nisan, and M. Szegedy, *Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs*, J. Comput. System Sci., 45 (1992), pp. 204–232.

[6] P. Beame and D.-T. Huynh-Ngoc, *Multiparty communication complexity and threshold circuit complexity of* AC^0, in Proceedings of the 50th Annual Symposium on Foundations of Computer Science, Atlanta, GA, IEEE, 2009, pp. 53–62.

[7] P. Beame, T. Pitassi, and N. Segerlind, *Lower bounds for Lovász-Schrijver systems and beyond follow from multiparty communication complexity*, SIAM J. Comput., 37 (2007), pp. 845–869.

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

[9] R. Beigel and J. Tarui, *On ACC*, Comput. Complexity, 4 (1994), pp. 350–366.

[10] A. Ben-Aroya, O. Regev, and R. de Wolf, *A hypercontractive inequality for matrix-valued functions with applications to quantum computing*, in Proceedings of the 49th Annual Symposium on Foundations of Computer Science, Philadelphia, IEEE, 2008, pp. 477–486.

[11] M. Ben-Or and N. Linial, *Collective coin flipping, robust voting schemes and minima of Banzhaf values*, in Proceedings of the 26th Annual Symposium on Foundations of Computer Science, Portland, OR, 1985, IEEE, pp. 408–416.

[12] C. Berg and S. Ulfberg, *A lower bound for perceptrons and an oracle separation of the* PP^{PH} *hierarchy*, J. Comput. System Sci., 56 (1998), pp. 263–271.

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

[14] A. Chattopadhyay and A. Ada, *Multiparty Communication Complexity of Disjointness*, http://www.eccc.uni-trier.de/eccc/ (2008).

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

[16] M. David, T. Pitassi, and E. Viola, *Improved separations between nondeterministic and randomized multiparty communication*, in RANDOM 2008, Proceedings of the 12th International Workshop on Randomization and Approximization Techniques in Computer Science, 2008, pp. 371–384.

[17] M. Furst, J. B. Saxe, and M. Sipser, *Parity, circuits, and the polynomial-time hierarchy*, Math. Systems Theory, 17 (1984), pp. 13–27.

[18] J. Håstad, *Almost optimal lower bounds for small depth circuits*, in Proceedings of the 18th Annual ACM Symposium on Theory of Computing, Berkeley, CA, 1986, pp. 6–20.

[19] J. Håstad, *Computational Limitations of Small-Depth Circuits*, MIT Press, Cambridge, MA, 1987.

[20] J. Håstad and M. Goldmann, *On the power of small-depth threshold circuits*, Comput. Complexity, 1 (1991), pp. 113–129.

[21] R. Jain, H. Klauck, and A. Nayak, *Direct product theorems for classical communication complexity via subdistribution bounds*, in Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, BC, 2008, pp. 599–608.

[22] H. Klauck, *Lower bounds for quantum communication complexity*, in Proceedings of the 42nd Annual Symposium on Foundations of Computer Science, Las Vegas, NV, 2001, IEEE, pp. 288–297.

[23] M. Krause and P. Pudlák, *On the computational power of depth-*2 *circuits with threshold and modulo gates*, Theoret. Comput. Sci., 174 (1997), pp. 137–156.

[24] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, Cambridge, UK, 1997.

[25] T. Lee and A. Shraibman, *Disjointness is hard in the multi-party number-on-the-forehead model*, in Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, College Park, MD, 2008, pp. 81–91.

[26] M. Minsky and S. Papert, *Perceptrons*, MIT Press, Cambridge, MA, 1988.

[27] N. Nisan and M. Szegedy, *On the degree of boolean functions as real polynomials*, Comput. Complexity, 4 (1994), pp. 301–314.

[28] R. Raz, *The BNS-Chung criterion for multi-party communication complexity*, Comput. Complexity, 9 (2000), pp. 113–122.

[29] R. Raz and P. McKenzie, *Separation of the monotone* NC *hierarchy*, Combinatorica, 19 (1999), pp. 403–435.

[30] A. Razborov and A. Wigderson, *Lower bounds on the size of depth* 3 *threshold circuits with AND gates at the bottom*, Inform. Process. Lett., 45 (1993), pp. 303–307.

[31] A. A. Razborov, *Quantum communication complexity of symmetric predicates*, Izv. Math., 67 (2003), pp. 145–159.

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

[33] A. A. Sherstov, *Separating* AC^0 *from depth-*2 *majority circuits*, in Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, CA, 2007, pp. 294–301.

[34] A. A. Sherstov, *Communication lower bounds using dual polynomials*, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS, 95 (2008), pp. 59–93.

[35] A. A. Sherstov, *The pattern matrix method for lower bounds on quantum communication*, in Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, BC, 2008, pp. 85–94.

[36] A. A. Sherstov, *Unbounded-error communication complexity of symmetric functions*, in Proceedings of the 49th Annual Symposium on Foundations of Computer Science, Philadelphia, 2008, IEEE, pp. 384–393.

[37] Y. Shi and Y. Zhu, *The quantum communication complexity of block-composed functions*, Quantum Inf. Comput., 9 (2009), pp. 444–460.

[38] M. Sipser, *Borel sets and circuit complexity*, in Proceedings of the 50th Annual ACM Symposium on Theory of Computing, Boston, MA, 1983, pp. 61–69.

[39] P. Tesson, *Communication Complexity Questions Related to Finite Monoids and Semigroups*, Ph.D. thesis, McGill University, Montreal, 2002.

[40] E. Viola, *Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates*, SIAM J. Comput., 36 (2007), pp. 1387–1403.

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

[42] A. C. Yao, *Separating the polynomial hierarchy by oracles: Part* I, in Proceedings of the 26th Annual Symposium on Foundations of Computer Science, Portland, OR, 1985, IEEE, pp. 1–10.

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