2009 50th Annual IEEE Symposium on Foundations of Computer Science

Multiparty Communication Complexity and Threshold Circuit Size of AC^0

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

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

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

Our lower bounds imply the first superpolynomial lower bounds for the simulation of AC^0 by MAJ ∘ SYMM ∘ AND circuits, showing that the well-known quasipolynomial simulations of AC^0 by such circuits are qualitatively optimal, even for formulas of small constant depth.

We also exhibit a depth 5 formula in NP$_k^{cc}$ − BPP$_k^{cc}$ for k up to $\Theta(\log n)$ and derive an $\Omega(2^{\sqrt{\log n}/\sqrt{k}})$ lower bound on the randomized k-party NOF communication complexity of set disjointness for up to $\Theta(\log^{1/3} n)$ players which is significantly larger than the $O(\log\log n)$ players allowed in the best previous lower bounds for multiparty set disjointness. We prove other strong results for depth 3 and 4 AC^0 functions.

Keywords-communication complexity, constant-depth circuits, lower bounds

1. Introduction

The multiparty communication complexity of AC^0 in the number-on-forehead (NOF) model has been an open question since Håstad and Goldmann [11] showed that any AC^0 or ACC^0 function has polylogarithmic randomized multiparty NOF communication complexity when its input bits are divided arbitrarily among a polylogarithmic number of players. This result is based on the simulations, due to Allender and Yao, of AC^0 circuits [1] and ACC^0 circuits [27] by quasipolynomial-size depth-3 circuits that consist of two layers of MAJORITY gates whose inputs are polylogarithmic-size AND gates of literals. These protocols may even be simultaneous NOF protocols in which the players in parallel send their information to a referee who computes the answer [2].

It is natural to ask whether these upper bounds can be improved. In the case of ACC^0, Razborov and Wigderson [18] showed that quasipolynomial size is required to simulate ACC^0 based on the result of Babai, Nisan, and Szegedy [4] that the Generalized Inner Product function in ACC^0 requires k-party NOF communication complexity $\Omega(n/4^k)$ which is polynomial in n for k up to $\Theta(\log n)$.

However, for AC^0 functions much less has been known. For the communication complexity of the set disjointness function with k players (which is in AC^0) there are lower bounds of the form $\Omega(n^{1/(k-1)}/(k-1))$ in the simultaneous NOF [24], [5] and $n^{\Omega(1/k)}/k^{O(k)}$ in the one-way NOF model [26]. These are sub-polynomial lower bounds for all non-constant values of k and, at best, polylogarithmic when k is $\Omega(\log n/\log\log n)$.

Until recently, there were no lower bounds for general multiparty NOF communication complexity of any AC^0 function. That changed with recent lower bounds for set disjointness by Lee and Shraibman [14] and Chattopadhyay and Ada [8] but no lower bounds apply for $\omega(\log\log n)$ players. As for circuit simulations of AC^0, Sherstov [20] recently showed that AC^0 cannot be simulated by polynomial-size MAJ ∘ MAJ circuits. However, there have been no non-trivial size lower bounds for the simulation of AC^0 by MAJ ∘ MAJ ∘ AND or even SYMM ∘ AND circuits with $\omega(\log\log n)$ bottom fan-in. As shown by Viola [25], sufficiently strong lower bounds for AC^0 in the multiparty NOF communication model, even for sub-logarithmic numbers of players, can yield quasipolynomial circuit size lower bounds.

We indeed produce such strong lower bounds. We show that there is an explicit linear-size fixed-depth AC^0 function that requires randomized k-party NOF communication complexity of $n^{\Omega(1)}/4^k$ even for error exponentially close to 1/2. For $\omega(1)$ players this bound is larger than all previous multiparty NOF communication complexity lower bounds for AC^0 functions, even those in the weaker simultaneous model. The bound is non-trivial for up to $\Theta(\log n)$ players and is sufficient to apply Viola's arguments to produce fixed-depth AC^0 functions that require MAJ∘SYMM∘AND circuits of $n^{\Omega(\log\log n)}$ size, showing that quasipolynomial size is necessary for the simulation of AC^0.

The function for which we derive our strongest communication complexity lower bound is computable in depth 6 AC^0. In the case of protocols with error 1/3, we exhibit a hard function computable by simple depth 4 formulas. We further show that the same lower bound applies to a

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

0272-5428/09 $26.00 © 2009 IEEE
DOI 10.1109/FOCS.2009.12

function having depth 5 formulas that also has $O(\log^2 n)$ nondeterministic communication complexity which shows that AC^0 contains functions in $\mathsf{NP}^{\mathsf{cc}}_{\mathsf{k}} - \mathsf{BPP}^{\mathsf{cc}}_{\mathsf{k}}$ for k up to $\Theta(\log n)$. As a consequence of the lower bound for this depth 5 function, we obtain $\Omega(2^{\sqrt{\log n}/\sqrt{k}-k})$ lower bounds on the k-party NOF communication complexity of set disjointness which is non-trivial for up to $\Theta(\log^{1/3} n)$ players. The best previous lower bounds for set disjointness only apply for $k \leq \log\log n - o(\log\log n)$ players (though these bounds are stronger than ours for $o(\log\log n)$ players).

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

Methods and Related Work: Recently, Sherstov introduced the pattern matrix method, a general method to use analytic properties of Boolean functions to derive communication lower bounds for related Boolean functions [20], [22]. In [20], this analytic property was large threshold degree, and the resulting communication lower bounds yielded lower bounds for simulations of AC^0 by $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits. Sherstov [22] extended this to large approximate degree, yielding a strong new method for lower bounds for two-party randomized and quantum communication complexity.

Chattopadhyay [7] generalized [20] to pattern tensors for $k \geq 2$ players to yield the first lower bounds for the general NOF multiparty communication complexity of any AC^0 function for $k \geq 3$, implying exponential lower bounds for computation of AC^0 functions by $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{ANY}$ circuits with $o(\log\log n)$ input fan-in – our results extend this to fan-in $\Omega(\log n)$. Lee and Schraibman [14] and Chattopadhyay and Ada [8] applied the full method in [22] to pattern tensors to yield the first lower bounds for the general NOF multiparty communication complexity of set disjointness for $k > 2$ players, improving on a long line of research on the problem [3], [24], [5], [26], [12], [6] and obtaining a lower bound of $\Omega(n^{\frac{1}{k+1}})/2^{2^{O(k)}}$. This yields a separation between randomized and nondeterministic k-party models for $k = o(\log\log n)$, which David, Pitassi, and Viola [10] 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 [8], [21] is as follows. To prove k-party lower bounds for a function F, we first show that F has $f \circ \psi^m$ as a subfunction where ψ is a bit-selection function and f has large approximate degree. For such an f there exists another function g and a distribution μ on inputs such that, with respect to μ, g is both highly correlated with f and orthogonal to all low-degree polynomials. It follows that $f \circ \psi^m$ is highly correlated with $g \circ \psi^m$ and, by the discrepancy method for communication complexity, it suffices to prove a discrepancy lower bound for $g \circ \psi^m$. Thanks to the orthogonality of g to all low degree polynomials this is possible using the bound in [4], [9], [17] derived from the iterated application of the Cauchy-Schwartz inequality. For example, the bound for set disjointness $\textsc{Disj}_{k,n}(x) = \vee_{i=1}^{n} \wedge_{j=1}^{k} x_{ji}$ corresponds to a particular selector ψ and $f = \textsc{Or}$ which has approximate degree $\Omega(\sqrt{n})$.

In the two party case, Sherstov [23] and Razborov and Sherstov [19] extended the pattern matrix method to yield sign-rank lower bounds for some simple functions. A key idea for their arguments is the existence of orthogonalizing distributions μ for their functions that are "min-smooth" in that they assign at least some fixed positive probability to any x such that $f(x) = 1$.

By contrast we show that any function f for which approximating f within ϵ on only a subset S of inputs requires large degree, there is an orthogonalizing distribution μ for f that is "max-smooth" – the probability of subsets defined by partial assignments is never much larger than under the uniform distribution. The smoothness quality and the properties of the constrained subset S are determined by a function α so we call the degree bound the (ϵ, α)-approximate degree. We show that for any function this degree bound is large if there is a diverse collection of partial assignments ρ such that each subfunction $f|_\rho$ of f requires large approximate degree. This property is somewhat delicate but we are able to exhibit simple AC^0 functions with large (ϵ, α)-approximate degree.

2. Preliminaries and the Generalized Discrepancy/Correlation Method

Circuit complexity: Let AND denote the class of all unbounded fan-in $\wedge$ functions (of literals), SYMM denote the class of all symmetric functions and $\mathsf{MAJ} \subset \mathsf{SYMM}$ denote the class of all majority functions. AC^0 is the class of functions $f : \{0,1\}^* \to \{0,1\}$ computed by polynomial size circuits (or formulas) of constant depth having $\neg$ gates and unbounded fan-in $\wedge$ and $\vee$ gates. Given classes of functions $\mathsf{C}_1, \mathsf{C}_2, \ldots \mathsf{C}_\mathsf{d}$, we let $\mathsf{C}_1 \circ \mathsf{C}_2 \circ \cdots \circ \mathsf{C}_\mathsf{d}$ be the class of all circuits of depth d whose inputs are given by variables and their negations and whose gates at the i-th level from the top are chosen from C_i.

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

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

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

Fact 2.1 (cf. [13])**.** *If there exists a distribution μ such that $\mathrm{Cor}_\mu(f, \Pi^c_k) \leq \epsilon$ then $R^k_{1/2-\epsilon/2}(f) \geq c$.*

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

Lemma 2.2 ([4], [9], [17])**.** *Let $f : \{0,1\}^{m \times k} \to \mathbb{R}$ and $\mathcal{U}$ be the uniform distribution over $X \times Y$ where $Y = Y_1 \times \cdots \times Y_{k-1}$. Then,*

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

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

Approximate and threshold degree: Given $0 \leq \epsilon < 1$, the ϵ-approximate degree of f, $deg_\epsilon(f)$, is the smallest d for which $||f - p||_\infty = \max_x |f(x) - p(x)| \leq \epsilon$ for some real-valued polynomial p of degree d. Following [16] we have the following property of the approximate degree of OR.

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

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

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

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

Lemma 2.4 ([22])**.** *If $f : \{0,1\}^m \to \{-1, 1\}$ has $deg_\epsilon(f) \geq d$ then there exists a function $g : \{0,1\}^m \to \{-1, 1\}$ and a distribution μ on $\{0,1\}^m$ such that:*

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

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

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

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

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

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

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

One example of a selector function ψ is the pattern tensor function $\psi_{k,\ell}$ used in [8], [14] which generalizes the pattern matrix function. In this example, $s = \ell^{k-1}$ and the s bits are arranged in a $(k-1)$-dimensional array indexed by $[\ell]^{k-1}$. $D_{\psi_{k,\ell},j}$ consists of the ℓ vectors $Y_j \in \{0,1\}^s$ that are 1 in all entries in one of the ℓ slices along the j-th dimension of this array and are 0 in every other entry. For $X \in \{0,1\}^s$ and such a $Y = (Y_1, \ldots, Y_{k-1}) \in \{0,1\}^{(k-1)s}$ the array $\wedge_{i=1}^{k-1} Y_i$ contains precisely one 1 which selects the bit of X to pass to f. This function is expressible by a small 2-level $\vee$ of $\wedge$s. As described in [10] 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 the full paper.

We give a brief overview of the remainder of the argument in [8], [10], which extends ideas of [20], [22] from 2-party to k-party communication complexity.

- Start with a Boolean function f on m bits having large $(1-\delta)$-approximate degree d.
- Apply the Orthogonality/Approximation Lemma to f to obtain a g that is $(1-\delta)$-correlated with f and a distribution μ under which g is not correlated with any low degree polynomial.
- Observe that from μ one can define a natural λ under which $g \circ \psi^m$ and $f \circ \psi^m$ have the same high correlation as g and f so to prove that $f \circ \psi^m$ is uncorrelated with low communication protocols, by the triangle inequality it suffices to prove this for $g \circ \psi^m$.

- The BNS-Chung bound/Gowers' norm used in Lemma 2.2 is based on the expectation of a function's correlation with itself on randomly chosen hypercubes of points. Use the orthogonality of g under μ to all polynomials of degree $< d$ to show that all low degree self-correlations of $g \circ \psi^m$ under λ disappear. The remaining high-degree self-correlations are bounded by analyzing overlaps in the choices of bits in different inputs among the hypercube of inputs. The argument repeatedly bounds the probability mass that μ assigns to small sub-cubes of the input by 1.
- The final lower bound is limited both by the upper bound on correlation in the high degree case and by the number of input bits required for each selector function.

Our argument follows this basic outline but improves it in two different ways. First, by considering a new measure that strengthens $(1-\delta)$-approximate degree we are able to obtain a much sharper upper bound on the high-degree self-correlations and second, we use a selector function that requires many fewer bits. We also show that some simple functions require large values for our strengthened measure (which turns out to be fairly non-trivial to prove).

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

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

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

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

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

This definition is an obvious weakening of the usual ℓ_∞ approximation of f since the non-light points can be ignored in the approximation. We will use this definition to prove our main technical theorem on the application of the general correlation method. To prove the theorem, we need the following lemma which generalizes Lemma 2.4 and is the first key to our substantially improved lower bounds. Its proof, which is based on LP duality, is given in Section 7.

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

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

Although the upper bound on $\mu(C_\rho)$ can be much larger than the $2^{-|\rho|}$ probability under the uniform distribution, we can use it to obtain an exponential improvement in the dependence of communication complexity lower bounds on k if $\alpha(r)$ is bounded below r^{α_0} for $r \geq d$ and $\alpha_0 < 1$. As we note in Section 7, for any function f computed by an AC^0 circuit this assumption and the upper bound are essentially the best possible for d polynomial in m.

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

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

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

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

Proof: The pattern of the argument follows the outline from Section 2. We first apply Lemma 3.1 to f to produce function g and distribution μ. By construction $\mathrm{Cor}_\mu(f,g) \geq 1-\epsilon$. Then we define a distribution λ on $\{0,1\}^{mks}$ based on μ and ψ by $\lambda(x,y) = \dfrac{\mu(z_1,\dots,z_m)}{2^{n-m}|D_\psi|^m}$ where $z_i = \psi(x_i, y_{*i})$

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

for $y \in D_{\psi}^{(m)}$ and 0 otherwise. To prove a lower bound c on $R^{k}_{1/2-\epsilon}(f \circ \psi^{m})$ we show that $\mathrm{Cor}_{\lambda}(f \circ \psi^{m}, \Pi_{k}^{c}) \leq 2\epsilon$.

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

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

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

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

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

where $z_i^u = \psi(x_i, y_{*i}^u)$. We now compute bounds on the self-correlation $H(y^0, y^1)$ that depend on the value of $r = r_{\psi}(y^0, y^1)$. The first bound is from [8] and is the key to the original method.

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

Proof: Let $\mathcal{Z} = \mathcal{Z}^{0\ldots0}\mathcal{Z}^{0\ldots1} \cdots \mathcal{Z}^{1\ldots1}$ be the joint distribution induced on $\{z^u\}_{u \in \{0,1\}^{k-1}}$ by taking x uniformly at random. By construction, z^u is uniformly distributed in $\{0,1\}^m$ for any $u \in \{0,1\}^{k-1}$ so each $\mathcal{Z}^u$ is a uniform distribution. For each choice of $z^{0\ldots0}$ we will also consider the conditional distribution $\mathcal{Z}^{\neq 0\ldots0} | z^{0\ldots0}$ on $\{z^u\}_{u \neq 0\ldots0}$ which is derived from $\mathcal{Z}$ by conditioning on $\mathcal{Z}^{0\ldots0} = z^{0\ldots0}$. Then,

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

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

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

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

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

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

The following bound for $r = r_{\psi}(y^0, y^1) \geq d$ is the key to the sharper bound that yields our exponentially better results. A weaker version in [8] applies only when $\alpha(r) = r$.

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

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

Since g is ±1-valued,

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

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

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

We now bound the remaining terms. First we have

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

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

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

By the property of μ implied by Lemma 3.1,

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

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

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

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

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

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

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

4. AC^0 FUNCTIONS WITH LARGE (ϵ, α)-APPROXIMATE DEGREE

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

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

Definition Given $\alpha : \{0, \dots, m\} \to \mathbb{R}$, we say that a probability distribution ν on $\{0,1,*\}^m$ is *α-spread* iff for every restriction $\rho \in \{0,1,*\}^m$, $\mathrm{Pr}_{\pi \sim \nu}[\pi \parallel \rho] \leq 2^{\alpha(|\rho|)-|\rho|}$. 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 $deg_\epsilon(f|_\pi) \leq d$. Note that for $\alpha(r) = r$, $deg_\epsilon(f) = deg_{\epsilon,\alpha}^*(f)$ since every distribution on restrictions is α-spread. We define $deg_{<\epsilon,\alpha}^*(f) = \min_{\epsilon' < \epsilon} deg_{\epsilon',\alpha}^*(f)$.

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

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

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 $\leq d$ and a distribution λ on $\{0,1,*\}^m$ such that $R(x) = \sum_{\rho \parallel x} 2^{|\rho|-\alpha(|\rho|)}\lambda_\rho > 1$ whenever $x \in B'$, where $B' = \{x : |f(x) - q(x)| > \epsilon\}$.

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

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

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

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

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

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

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

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

The proof of Lemma 4.2 is surprisingly involved and requires quite precise tail bounds. It is in the full paper.

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

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

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

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

It is clear that for any π with $\nu(\pi) > 0$, $\textsc{Or}_r$ is a subfunction of $\textsc{Tribes}_{p,q}|_\pi$ so $deg_{5/6}(\textsc{Tribes}_{p,q}|_\pi) \geq deg_{5/6}(\textsc{Or}_r) \geq \sqrt{r/12}$. 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:

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

Corollary 4.4 suffices for most of our communication complexity lower bounds. However our results for threshold circuit size require a function in AC^0 having large α-threshold degree, which is more difficult to produce. The proof of the next lemma, which involves more complex G and H, and our generic construction are in the full paper.

Lemma 4.5. *For any p sufficiently large multiple of 15 and $q = 2^{4p}$, if $\alpha : \{0, \ldots, m\} \to \mathbb{R}$ is defined as $\alpha(z) = z^{0.9}$ for $r \geq (3p \ln 2)^{10}$ and $\alpha(z) = z$ otherwise, then there is an explicit depth 4 AC^0 function on pq bits that has α-threshold degree at least $q^{1/15}$.*

5. Multiparty communication complexity lower bounds for AC^0

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

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

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

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

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

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

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

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

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

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

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

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

$$\leq \sum_{r=d}^{m} 2^{(2^{k-1}-1)\alpha(r)} \cdot \Big(\frac{em2^{2k-a-3}}{r}\Big)^r \tag{4}$$

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 (4) is

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

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

Plugging this into Theorem 3.2 we obtain that

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

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. The proofs of the remaining results in this section are in the full paper.

Theorem 5.3. *Let p be a sufficiently large integer and $q = 2^{4p}$, $k \leq p/10$, and $s = 2^{p+2k}$. Let $F = \textsc{Tribes}_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$ and $F' = \textsc{Tribes}'_{p,q} \circ \textsc{Index}^m_{\oplus^{(p+2k)}_{k-1}}$. Let $n = pqs = p2^{5p+2k}$ be the number of input bits given to each player in computing F or F'. Then $R^k_{1/3}(F)$ and $R^k_{1/3}(F')$ are both $\Omega(q^{0.3}/2^k)$ which is $n^{\Omega(1)}/4^k$. Furthermore, F has polynomial-size depth 5 AC^0 formulas and F' has polynomial-size depth 4 AC^0 formulas.*

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

Corollary 5.5. *There is a function G in depth 5 AC^0 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 $a' > 0$.*

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

Theorem 5.6. *$R^k_{1/3}(\textsc{Disj}_{n,k})$ is $\Omega(2^{\sqrt{\log_2 n}/\sqrt{k}})$ for $k \leq \frac{1}{5}\log_2^{1/3} n$.*

Although our bound for $\textsc{Disj}_{n,k}$ applies to exponentially more players than do the bounds in [14], [8], 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 $\Theta(\log^{1/3} n)$.*

Although we have shown non-trivial lower bounds for $\textsc{Disj}_{k,n}$ for k up to $\Theta(\log^{1/3} n)$, it is open whether one can prove stronger lower bounds for $k = \omega(\log^{1/3} n)$ players for $\textsc{Disj}_{k,n}$ or any other depth-2 AC^0 function. The difficulty of extending our lower bound methods is our inability to apply Lemma 3.1 to $\textsc{Or}$ since the constant function 1 approximates $\textsc{Or}$ on all but one point.

To prove lower bounds for $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuits we need lower bounds on protocols that succeed with probability barely better than that of random guessing. Using the function with large α-threshold degree given by Lemma 4.5 in place of $\textsc{Tribes}_{p,q}$ we obtain the following theorem.

Theorem 5.8. *There exist explicit constants $c, c' > 0$ and a depth 6 AC^0 function $H : \{0,1\}^* \to \{0,1\}$ such that for $1/2 > \epsilon > 0$, $R^k_{1/2-\epsilon}(H_n)$ is $\Omega(n^c + \log \epsilon)$ for any $k \leq c' \log_2 n$.*

6. Threshold circuit lower bounds for AC^0

Following the approach of Viola [25], which extends the ideas of Razborov and Wigderson [18], we show quasipolynomial lower bounds on the simulation of AC^0 functions by unrestricted $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuits.

Theorem 6.1. *There is a function $G : \{0,1\}^* \to \{0,1\}$ in AC^0 such that G_N requires $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}$ circuit size $N^{\Omega(\log\log N)}$.*

Proof Sketch: The proof is almost identical to an argument in [25] with our hard AC^0 functions replacing the generalized inner product. It relies on the following connection between multiparty communication complexity and threshold circuit complexity given by Håstad and Goldmann.

Proposition 6.2. *[11] If f is computed by a $\mathsf{MAJ} \circ \mathsf{SYMM} \circ \mathsf{AND}_{\mathsf{k-1}}$ circuit of size S, then $R^k_{1/2-1/(2S)}(f)$ is $O(k \log S)$.*

We use the function H_n from Theorem 5.8 and replace each input by an $\oplus$ of $\Theta(\log^2 n)$ new input bits to obtain a function G of $N = \Theta(n \log^2 n)$ inputs. This adds 2 to the depth and keeps the polynomial size. If G is computed by such a circuit C of size $N^{o(\log\log N)}$ then using random restrictions that leave bits unset with probability $\Theta(1/\log N)$ we can ensure both that all bottom-level AND gates of C are reduced to fan-in at most $\delta \log_2 N$ and that every $\oplus$ block of inputs in G contains at least one unset input bit. Applying Proposition 6.2 yields a contradiction to Theorem 5.8. ■

7. Proof of Lemma 3.1

Proof: As in the proof for Lemma 2.4, we write the requirements as a linear program and study its dual. The lemma is implied by proving that the following linear program $\mathcal{P}$ has optimal value ≤ 1:

Minimize η subject to

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

Suppose that we have optimum $\eta \leq 1$. In this LP formulation, inequality γ ensures that the function μ is a probability distribution, and inequalities v_x and w_x ensure that $\mu(x) \geq |h(x)|$ so $||h||_1 \leq 1$. If $||h||_1 = 1$, then we must have $\mu(x) = |h(x)|$ and we can write $h(x) = \mu(x)g(x)$ as in the proof of Lemma 2.4 and then the inequalities y_S will ensure that $\mathrm{Cor}_\mu(g, \chi_S) = 0$ for $|S| < d$ and inequality β will ensure that $\mathrm{Cor}_\mu(f, g) \geq \epsilon$ as required. Finally, each inequality λ_ρ ensures that $\mu(C_\rho) \leq 2^{-|\rho|+\alpha(\rho|)}$ which is actually a little stronger than our claim.

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

Here is the dual LP:

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

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

$$\mu(x) : \quad v_x + w_x + \gamma - \sum_{\rho\|x} 2^{|\rho|-\alpha(|\rho|)}\lambda_\rho = 0 : x \in \{0,1\}^m \tag{5}$$

$$h(x) : \beta f(x) + \sum_{|S|<d} y_S\chi_S(x) + w_x - v_x = 0 : x \in \{0,1\}^m \tag{6}$$

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

Since y_S are arbitrary we can replace $\sum_{|S|<d} y_S\chi_S(x)$ by $p_d(x)$ where p_d is an arbitrary polynomial of degree $< d$ and rewrite (6) as:

$$h(x) : \quad \beta f(x) + p_d(x) + w_x - v_x = 0 : x \in \{0,1\}^m \tag{7}$$

Equations (5) and (7) for $x \in \{0,1\}^m$ together are equivalent to:

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

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

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

which together are equivalent to

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

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

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

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

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

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

$$\beta|f(x) - q_d(x)| + \gamma \leq \sum_{\rho\|x} 2^{|\rho|-\alpha(|\rho|)}\lambda_\rho. \tag{8}$$

Let $R(x)$ denote the right-hand side of inequality (8). It suffices to prove that $R(x) \leq 1$ for some $x \in B$. This is, in turn, equivalent to proving that

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

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

We note that the bounds in Lemma 3.1 will require $\alpha(r) \geq r^\delta$ for some $\delta > 0$ when applied to AC^0 functions: By results of Linial, Mansour, and Nisan [15], for any AC^0 function f

and constant $0 < \eta < 1$, there is a function p_d of degree $d < m^\eta$, such that $||f - p_d||_2^2 \le 2^{m-m^\delta}$ for some constant $\delta > 0$. Let B_m be the set of x such that $|f(x) - p_d(x)| \ge \epsilon$. Then $|B_m|\epsilon^2 \le \sum_{x \in B_m} |f(x) - p_d(x)|^2 \le ||f - p_d(x)||_2^2 \le 2^{m-m^\delta}$ so $|B_m| \le 2^{m-m^\delta}/\epsilon^2$. If we tried to replace the upper bound on $\mu(C_\rho)$ by some $c(|\rho|)$ where $1/c(m)$ is $\omega(|B_m|)$ then in the dual program $\mathcal{D}$, we could choose $\lambda_x = 1/|B_m|$ for $x \in B_m$ and $\lambda_\rho = 0$ for all other ρ and for these values β would be unbounded.

Acknowledgements

We thank Emanuele Viola for suggesting the circuit complexity application and Alexander Sherstov, Arkadev Chattopadhyay, and the anonymous referees for many helpful comments.

References

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

[2] L. Babai, A. Gál, P. G. Kimmel, and S. V. Lokam, "Communication complexity of simultaneous messages," *SIAM Journal on Computing*, vol. 33, no. 1, pp. 137–166, 2003.

[3] L. Babai, T. P. Hayes, and P. G. Kimmel, "The cost of the missing bit: Communication complexity with help," *Combinatorica*, vol. 21, no. 4, pp. 455–488, 2001.

[4] L. Babai, N. Nisan, and M. Szegedy, "Multiparty protocols, pseudorandom generators for logspace, and time-space tradeoffs," *Journal of Computer and System Sciences*, vol. 45, no. 2, pp. 204–232, Oct. 1992.

[5] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson, "A strong direct product theorem for corruption and the multiparty communication complexity of set disjointness," *Computational Complexity*, vol. 15, no. 4, pp. 391–432, 2006.

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

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

[8] A. Chattopadhyay and A. Ada, "Multiparty communication complexity of disjointness," Electronic Colloquium in Computation Complexity, Tech. Rep. TR08-002, 2008.

[9] F. R. K. Chung, "Quasi-random classes of hypergraphs," *Random Structures and Algorithms*, vol. 1, no. 4, pp. 363–382, 1990.

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

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

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

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

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

[15] M. Linial, Y. Mansour, and N. Nisan, "Constant depth circuits, Fourier transform, and learnability," in *30th Annual Symposium on Foundations of Computer Science*, Research Triangle Park, NC, Oct. 1989, pp. 574–579.

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

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

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

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

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

[21] ——, "Communication lower bounds using dual polynomials," *Bulletin of the European Association for Theoretical Computer Science*, vol. 95, pp. 59–93, 2008.

[22] ——, "The pattern matrix method for lower bounds on quantum communication," in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, Victoria, BC, May 2008, pp. 85–94.

[23] ——, "Unbounded-error communication complexity of symmetric functions," in *Proceedings 49th Annual Symposium on Foundations of Computer Science*. Philadelphia,PA: IEEE, Oct. 2008, pp. 384–393.

[24] P. Tesson, "Communication complexity questions related to finite monoids and semigroups," Ph.D. dissertation, McGill University, 2002.

[25] E. Viola, "Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates," *SIAM Journal on Computing*, vol. 36, no. 5, pp. 1387–1403, 2007.

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

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