Near-Optimal Lower Bounds on the Threshold Degree and Sign-Rank of AC^0

Alexander A. Sherstov
University of California, Los Angeles
Los Angeles, California, USA
sherstov@cs.ucla.edu

Pei Wu
University of California, Los Angeles
Los Angeles, California, USA
pwu@cs.ucla.edu

ABSTRACT

The *threshold degree* of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is the minimum degree of a real polynomial p that represents f in sign: $\operatorname{sgn} p(x) = (-1)^{f(x)}$. A related notion is *sign-rank*, defined for a Boolean matrix $F = [F_{ij}]$ as the minimum rank of a real matrix M with $\operatorname{sgn} M_{ij} = (-1)^{F_{ij}}$. Determining the maximum threshold degree and sign-rank achievable by constant-depth circuits (AC^0) is a well-known and extensively studied open problem, with complexity-theoretic and algorithmic applications.

We give an essentially optimal solution to this problem. For any $\epsilon > 0$, we construct an AC^0 circuit in n variables that has threshold degree $\Omega(n^{1-\epsilon})$ and sign-rank $\exp(\Omega(n^{1-\epsilon}))$, improving on the previous best lower bounds of $\Omega(\sqrt{n})$ and $\exp(\tilde{\Omega}(\sqrt{n}))$, respectively. Our results subsume *all* previous lower bounds on the threshold degree and sign-rank of AC^0 circuits of any given depth, with a strict improvement starting at depth 4. As a corollary, we also obtain near-optimal bounds on the discrepancy, threshold weight, and threshold density of AC^0, strictly subsuming previous work on these quantities. Our work gives some of the strongest lower bounds to date on the communication complexity of AC^0.

CCS CONCEPTS

• **Theory of computation → Complexity classes**; **Communication complexity**; **Circuit complexity**; **Algebraic complexity theory**.

KEYWORDS

constant-depth circuits, threshold degree, sign-representation by polynomials, sign-rank, communication complexity, unbounded-error communication

ACM Reference Format:
Alexander A. Sherstov and Pei Wu. 2019. Near-Optimal Lower Bounds on the Threshold Degree and Sign-Rank of AC^0. In *Proceedings of the 51st Annual ACM SIGACT Symposium on the Theory of Computing (STOC '19), June 23–26, 2019, Phoenix, AZ, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3313276.3316408

STOC '19, June 23–26, 2019, Phoenix, AZ, USA

ACM ISBN 978-1-4503-6705-9/19/06…$15.00
https://doi.org/10.1145/3313276.3316408

1 INTRODUCTION

A real polynomial p is said to *sign-represent* the Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ if $\operatorname{sgn} p(x) = (-1)^{f(x)}$ for every input $x \in \{0,1\}^n$. The *threshold degree* of f, denoted $\deg_{\pm}(f)$, is the minimum degree of a multivariate real polynomial that sign-represents f. Equivalent terms in the literature include *strong degree*, *voting polynomial degree*, *PTF degree*, and *sign degree*. Since any function $f\colon \{0,1\}^n \to \{0,1\}$ can be represented exactly by a real polynomial of degree at most n, the threshold degree of f is an integer between 0 and n. Viewed as a computational model, sign-representation is remarkably powerful because it corresponds to the strongest form of pointwise approximation. The formal study of threshold degree began in 1969 with the pioneering work of Minsky and Papert [21] on limitations of perceptrons. The authors of [21] famously proved that the parity function on n variables has the maximum possible threshold degree, n. They obtained lower bounds on the threshold degree of several other functions, including DNF formulas and intersections of halfspaces. Since then, sign-representing polynomials have found applications far beyond artificial intelligence. In theoretical computer science, applications of threshold degree include circuit lower bounds, size-depth trade-offs, communication complexity, structural complexity, and computational learning; see the full version [39] of this paper for a bibliographic overview.

The notion of threshold degree has been especially influential in the study of AC^0, the class of constant-depth polynomial-size circuits with $\wedge, \vee, \neg$ gates of unbounded fan-in. The first such result was obtained by Aspnes et al. [3], who used sign-representing polynomials to give a beautiful new proof of classic lower bounds for AC^0. In communication complexity, the notion of threshold degree played a central role in the first construction [28, 30] of an AC^0 circuit with exponentially small discrepancy and hence large communication complexity in nearly every model. That discrepancy result was used in [28] to show the optimality of Allender's classic simulation of AC^0 by majority circuits, solving the open problem [18] on the relation between the two circuit classes. Subsequent work [5, 13, 36, 38] resolved other questions in communication complexity and circuit complexity related to constant-depth circuits by generalizing the threshold degree method of [28, 30].

Sign-representing polynomials also paved the way for *algorithmic* breakthroughs in the study of constant-depth circuits. Specifically, any function of threshold degree d can be viewed as a halfspace in $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d}$ dimensions, corresponding to the monomials in a sign-representation of f. As a result, a class of functions of threshold degree at most d can be learned in the standard PAC model under arbitrary distributions in time polynomial in $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d}$. Klivans and Servedio [16] used this threshold degree approach to give what is currently the fastest algorithm

Table 1: Known bounds on the maximum threshold degree of $\wedge, \vee, \neg$-circuits of polynomial size and constant depth. In all bounds, n denotes the number of variables, and k denotes an arbitrary positive integer.

Depth	Threshold degree	Reference
2	$\Omega(n^{1/3})$	Minsky and Papert [21]
2	$O(n^{1/3}\log n)$	Klivans and Servedio [16]
$k+2$	$\Omega(n^{1/3}\log^{2k/3} n)$	O'Donnell and Servedio [23]
k	$\Omega(n^{\frac{k-1}{2k-1}})$	Sherstov [35]
4	$\Omega(\sqrt{n})$	Sherstov [37]
3	$\tilde{\Omega}(\sqrt{n})$	Bun and Thaler [12]
k	$\tilde{\Omega}(n^{\frac{k-1}{k+1}})$	This paper

Table 2: Known lower bounds on the maximum sign-rank of $\wedge, \vee, \neg$-circuits $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ of polynomial size and constant depth. In all bounds, k denotes an arbitrary positive integer.

Depth	Sign-rank	Reference
3	$\exp(\Omega(n^{1/3}))$	Razborov and Sherstov [26]
3	$\exp(\tilde{\Omega}(n^{2/5}))$	Bun and Thaler [10]
7	$\exp(\tilde{\Omega}(\sqrt{n}))$	Bun and Thaler [12]
$3k$	$\exp(\tilde{\Omega}(n^{1-\frac{1}{k+1}}))$	This paper
$3k+1$	$\exp(\tilde{\Omega}(n^{1-\frac{1}{k+1.5}}))$	This paper

for learning polynomial-size DNF formulas, with running time $\exp(\tilde{O}(n^{1/3}))$. Another learning-theoretic breakthrough based on threshold degree is the fastest algorithm for learning Boolean formulas, obtained by O'Donnell and Servedio [23] for formulas of constant depth and by Ambainis et al. [2] for arbitrary depth. Their algorithm runs in time $\exp(\tilde{O}(n^{(2^{k-1}-1)/(2^k-1)}))$ for formulas of size n and constant depth k, and in time $\exp(\tilde{O}(\sqrt{n}))$ for formulas of unbounded depth. In both cases, the bound on the running time follows from the corresponding upper bound on the threshold degree.

A far-reaching generalization of threshold degree is the matrix-analytic notion of *sign-rank*, which allows sign-representation out of arbitrary low-dimensional subspaces rather than the subspace of low-degree polynomials. The contribution of this paper is to prove essentially optimal lower bounds on the threshold degree and sign-rank of AC^0, which in turn imply lower bounds on other fundamental complexity measures of interest in communication complexity and learning theory. In the remainder of this section, we give a detailed overview of the previous work, present our main results, and discuss our proofs.

1.1 Threshold Degree of AC^0

Determining the maximum threshold degree of an AC^0 circuit in n variables is a longstanding open problem in the area. It is motivated by algorithmic and complexity-theoretic applications [8, 16, 17, 23, 26], in addition to being a natural question in its own right. Table 1 gives a quantitative summary of the results obtained to date. In their seminal monograph, Minsky and Papert [21] proved a lower bound of $\Omega(n^{1/3})$ on the threshold degree of the following DNF formula in n variables: $f(x) = \bigwedge_{i=1}^{n^{1/3}} \bigvee_{j=1}^{n^{2/3}} x_{i,j}$. Three decades later, Klivans and Servedio [16] obtained an $O(n^{1/3}\log n)$ upper bound on the threshold degree of any polynomial-size DNF formula in n variables, essentially matching Minsky and Papert's result and resolving the problem for depth 2. Determining the threshold degree of circuits of depth $k \geqslant 3$ proved to be challenging. The only upper bound known to date is the trivial $O(n)$, which follows directly from the definition of threshold degree. In particular, it is consistent with our knowledge that there are AC^0 circuits with linear threshold degree. On the lower bounds side, the only progress for a long time was due to O'Donnell and Servedio [23], who constructed for any $k \geqslant 1$ a circuit of depth $k+2$ with threshold degree $\Omega(n^{1/3}\log^{2k/3} n)$. The authors of [23] formally posed the problem of obtaining a polynomial improvement on Minsky and Papert's lower bound. Such an improvement was obtained in [35], with a threshold degree lower bound of $\Omega(n^{(k-1)/(2k-1)})$ for circuits of depth k. A polynomially stronger result was obtained in [37], with a lower bound of $\Omega(\sqrt{n})$ on the threshold degree of an explicit circuit of depth 4. Bun and Thaler [12] recently used a different, depth-3 circuit to give a much simpler proof of the $\tilde{\Omega}(\sqrt{n})$ lower bound for AC^0. We obtain a quadratically stronger, and near-optimal, lower bound on the threshold degree of AC^0.

Theorem 1.1. *Fix an integer $k \geqslant 1$. Then there is an (explicitly given) Boolean circuit family $\{f_n\}_{n=1}^{\infty}$, where $f_n\colon \{0,1\}^n \to \{0,1\}$ has polynomial size, depth k, and threshold degree*

$$\deg_{\pm}(f_n) = \Omega\left(n^{\frac{k-1}{k+1}} \cdot (\log n)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right).$$

Moreover, f_n has bottom fan-in $O(\log n)$ for all $k \neq 2$.

For large k, Theorem 1.1 essentially matches the trivial upper bound of n on the threshold degree of any function. For any fixed depth k, Theorem 1.1 subsumes all previous lower bounds on the threshold degree of AC^0, with a polynomial improvement starting at depth $k = 4$. In particular, the lower bounds due to Minsky and Papert [21] and Bun and Thaler [12] are subsumed as the special cases $k = 2$ and $k = 3$, respectively. From a computational learning perspective, Theorem 1.1 definitively rules out the threshold degree approach to learning constant-depth circuits. By well-known reductions, Theorem 1.1 implies a number of other lower bounds for the representation of AC^0 circuits by polynomials; see the full version [39] of this paper for details.

1.2 Sign-rank of AC^0

The *sign-rank* of a matrix $A = [A_{ij}]$ without zero entries, denoted $\mathrm{rk}_{\pm}(A)$, is the least rank of a real matrix $M = [M_{ij}]$ with $\operatorname{sgn} M_{ij} = \operatorname{sgn} A_{ij}$ for all i, j. In other words, the sign-rank of A

is the minimum rank of a matrix that can be obtained by making arbitrary sign-preserving changes to the entries of A. The sign-rank of a Boolean function $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined in the natural way as the sign-rank of the matrix $[(-1)^{F(x,y)}]_{x,y}$. In particular, the sign-rank of F is an integer between 1 and 2^n. This fundamental notion has been studied in contexts as diverse as matrix analysis, communication complexity, circuit complexity, and learning theory; see [26] for a bibliographic overview. To a complexity theorist, sign-rank is a vastly more challenging quantity to analyze than threshold degree. Indeed, a sign-rank lower bound rules out sign-representation out of *every* linear subspace of given dimension, whereas a threshold degree lower bound rules out sign-representation specifically by linear combinations of monomials up to a given degree.

Unsurprisingly, progress in understanding sign-rank has been slow and difficult. No nontrivial lower bounds were known for any explicit matrices until the breakthrough work of Forster [14], who proved strong lower bounds on the sign-rank of Hadamard matrices and more generally all sign matrices with small spectral norm. The sign-rank of constant-depth circuits $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ has since seen considerable work, as summarized in Table 2. The first exponential lower bound on the sign-rank of an AC^0 circuit was obtained by Razborov and Sherstov [26], solving a 22-year-old problem due to Babai, Frankl, and Simon [4]. The authors of [26] constructed a polynomial-size circuit of depth 3 with sign-rank $\exp(\Omega(n^{1/3}))$. In follow-up work, Bun and Thaler [10] constructed a polynomial-size circuit of depth 3 with sign-rank $\exp(\tilde{\Omega}(n^{2/5}))$. A more recent and incomparable result, also due to Bun and Thaler [12], is a sign-rank lower bound of $\exp(\tilde{\Omega}(\sqrt{n}))$ for a circuit of polynomial size and depth 7. No nontrivial upper bounds are known on the sign-rank of AC^0. Closing this gap between the best lower bound of $\exp(\tilde{\Omega}(\sqrt{n}))$ and the trivial upper bound of 2^n has been a challenging open problem. We solve this problem almost completely, by constructing for any $\epsilon > 0$ a constant-depth circuit with sign-rank $\exp(\Omega(n^{1-\epsilon}))$. In quantitative detail, our results on the sign-rank of AC^0 are the following two theorems.

Theorem 1.2. *Fix an integer $k \geqslant 1$. Then there is an (explicitly given) Boolean circuit family $\{F_n\}_{n=1}^{\infty}$, where $F_n\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ has polynomial size, depth $3k$, and sign-rank*

$$\mathrm{rk}_{\pm}(F_n) = \exp\left(\Omega\left(n^{1-\frac{1}{k+1}} \cdot (\log n)^{-\frac{k(k-1)}{2(k+1)}}\right)\right).$$

Theorem 1.3. *Fix an integer $k \geqslant 1$. Then there is an (explicitly given) Boolean circuit family $\{G_n\}_{n=1}^{\infty}$, where $G_n\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ has polynomial size, depth $3k+1$, and sign-rank*

$$\mathrm{rk}_{\pm}(G_n) = \exp\left(\Omega\left(n^{1-\frac{1}{k+1.5}} \cdot (\log n)^{-\frac{k^2}{2k+3}}\right)\right).$$

For large k, the lower bounds of Theorems 1.2 and 1.3 approach the trivial upper bound of 2^n on the sign-rank of any Boolean function $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. For any given depth, Theorems 1.2 and 1.3 subsume all previous lower bounds on the sign-rank of AC^0, with a strict improvement starting at depth 3. From a computational learning perspective, Theorems 1.2 and 1.3 state that AC^0 has near-maximum *dimension complexity* [12, 26, 27, 29], namely, $\exp(\Omega(n^{1-\epsilon}))$ for any constant $\epsilon > 0$. This rules out the possibility of learning AC^0 circuits via dimension complexity [26], a far-reaching generalization of the threshold degree approach from the monomial basis to arbitrary bases.

1.3 Communication Complexity

Theorems 1.1–1.3 imply strong new lower bounds on the communication complexity of AC^0. We adopt the standard randomized model of Yao [19], with players Alice and Bob and a Boolean function $F\colon X \times Y \to \{0,1\}$. On input $(x,y) \in X \times Y$, Alice and Bob receive the arguments x and y, respectively, and communicate back and forth according to an agreed-upon protocol. Each player privately holds an unlimited supply of uniformly random bits that he or she can use when deciding what message to send at any given point in the protocol. The *cost* of a protocol is the total number of bits communicated in a worst-case execution. The *ϵ-error randomized communication complexity* $R_\epsilon(F)$ of F is the least cost of a protocol that computes F with probability of error at most ϵ on every input.

Of particular interest to us are communication protocols with error probability close to that of random guessing, $1/2$. There are two standard ways to formalize the complexity of a communication problem F in this setting, both inspired by probabilistic polynomial time PP for Turing machines:

$$\mathrm{UPP}(F) = \inf_{0 \leqslant \epsilon < 1/2} R_\epsilon(F),$$

$$\mathrm{PP}(F) = \inf_{0 \leqslant \epsilon < 1/2} \left\{ R_\epsilon(F) + \log_2\left(\frac{1}{\frac{1}{2} - \epsilon}\right) \right\}.$$

The former quantity, introduced by Paturi and Simon [25], is called the *communication complexity of F with unbounded error*, in reference to the fact that the error probability can be arbitrarily close to $1/2$. The latter quantity is called the *communication complexity of F with weakly unbounded error*. Proposed by Babai et al. [4], it features an additional penalty term that depends on the error probability. It is clear that $\mathrm{UPP}(F) \leqslant \mathrm{PP}(F) \leqslant n+2$ for every communication problem $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, with an exponential gap achievable between the two complexity measures [6, 27]. These two models occupy a special place in the study of communication because they are more powerful than any other standard model (deterministic, nondeterministic, randomized, quantum with or without entanglement). Moreover, unbounded-error protocols represent a frontier in communication complexity theory in that they are the most powerful protocols for which explicit lower bounds are known. Our results imply that even for such protocols, AC^0 has near-maximal communication complexity. To begin with, combining Theorem 1.1 with the *pattern matrix method* [28, 30] gives:

Theorem 1.4. *Let $k \geqslant 3$ be a fixed integer. Then there is an (explicitly given) Boolean circuit family $\{F_n\}_{n=1}^{\infty}$, where $F_n\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ has polynomial size, depth k, communication complexity*

$$\mathrm{PP}(F_n) = \Omega\left(n^{\frac{k-1}{k+1}} \cdot (\log n)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right)$$

and discrepancy

$$\mathrm{disc}(F_n) = \exp\left(-\Omega\left(n^{\frac{k-1}{k+1}} \cdot (\log n)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right)\right).$$

Discrepancy is a combinatorial complexity measure of interest in communication complexity theory and other research areas; see the

full version [39] of this paper for a formal definition. As k grows, the bounds of Theorem 1.4 approach the best possible bounds for any communication problem $F_n\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. The same *qualitative* behavior was achieved in previous work by Bun and Thaler [12], who constructed, for any constant $\epsilon > 0$, a constant-depth circuit $F_n\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with communication complexity $\mathrm{PP}(F_n) = \Omega(n^{1-\epsilon})$ and discrepancy $\mathrm{disc}(F_n) = \exp(-\Omega(n^{1-\epsilon}))$. Theorem 1.4 strictly subsumes the result of Bun and Thaler [12] and all other prior work on the discrepancy and PP-complexity of constant-depth circuits [6, 28, 30, 35, 37]. For any fixed depth greater than 3, the bounds of Theorem 1.4 are a polynomial improvement in n over all previous work. We further show that Theorem 1.4 carries over to the *number-on-the-forehead model*, the strongest formalism of multiparty communication. This result, presented in detail in Section 3.4, uses the multiparty version [36] of the pattern matrix method.

Our work also gives near-optimal lower bounds for AC^0 in the much more powerful unbounded-error model. Specifically, it is well-known [25] that the unbounded-error communication complexity of any Boolean function $F\colon X \times Y \to \{0,1\}$ coincides up to an additive constant with the logarithm of the sign-rank of F. As a result, Theorems 1.2 and 1.3 imply:

Theorem 1.5. *Let $k \geqslant 1$ be a given integer. Let $\{F_n\}_{n=1}^{\infty}$ and $\{G_n\}_{n=1}^{\infty}$ be the polynomial-size circuit families of depth $3k$ and $3k+1$, respectively, constructed in Theorems 1.2 and 1.3. Then*

$$\mathrm{UPP}(F_n) = \Omega\left(n^{1-\frac{1}{k+1}} \cdot (\log n)^{-\frac{k(k-1)}{2(k+1)}}\right),$$

$$\mathrm{UPP}(G_n) = \Omega\left(n^{1-\frac{1}{k+1.5}} \cdot (\log n)^{-\frac{k^2}{2k+3}}\right).$$

For large k, the lower bounds of Theorem 1.5 essentially match the trivial upper bound of $n+1$ on the unbounded-error communication complexity of any function $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Theorem 1.5 strictly subsumes all previous lower bounds on the unbounded-error communication complexity of AC^0, with a polynomial improvement for any depth greater than 2. The best lower bound on the unbounded-error communication complexity of AC^0 prior to our work was $\tilde{\Omega}(\sqrt{n})$ for a circuit of depth 7, due to Bun and Thaler [12]. Finally, we remark that Theorem 1.5 gives essentially the strongest possible separation of the communication complexity classes PH and UPP. We refer the reader to the work of Babai et al. [4] for definitions and detailed background on these classes.

Qualitatively, Theorem 1.5 is stronger than Theorem 1.4 because communication protocols with unbounded error are significantly more powerful than those with weakly unbounded error. On the other hand, Theorem 1.4 is stronger quantitatively for any fixed depth and has the additional advantage of generalizing to the multiparty setting.

1.4 Previous Approaches

In the remainder of this section, we discuss our proofs of Theorems 1.1–1.3. The notation that we use here is standard, and we defer its formal review to Section 2. We start with necessary approximation-theoretic background, then review relevant previous work, and finally contrast it with the approach of this paper. To sidestep minor technicalities, we will represent Boolean functions in this overview as mappings $\{-1,+1\}^n \to \{-1,+1\}$. We alert the reader that we will revert to the standard $\{0,1\}^n \to \{0,1\}$ representation starting with Section 2.

Background. Recall that our results concern the sign-representation of Boolean functions and matrices. To properly set the stage for our proofs, however, we need to consider the more general notion of pointwise approximation [22]. Let $f\colon \{-1,+1\}^n \to \{-1,+1\}$ be a Boolean function of interest. The *ϵ-approximate degree of f*, denoted $\deg_\epsilon(f)$, is the minimum degree of a real polynomial that approximates f within ϵ pointwise: $\deg_\epsilon(f) = \min\{\deg p\colon \|f-p\|_\infty \leqslant \epsilon\}$. The regimes of most interest are *bounded-error approximation*, corresponding to constants $\epsilon \in (0,1)$; and *large-error approximation*, corresponding to $\epsilon = 1 - o(1)$. In the former case, the choice of error parameter $\epsilon \in (0,1)$ is immaterial and affects the approximate degree of a Boolean function by at most a multiplicative constant. It is clear that pointwise approximation is a stronger requirement than sign-representation, and thus $\deg_\pm(f) \leqslant \deg_\epsilon(f)$ for all $0 \leqslant \epsilon < 1$. A moment's thought reveals that threshold degree is in fact the limiting case of ϵ-approximate degree as the error parameter approaches 1:

$$\deg_\pm(f) = \lim_{\epsilon \nearrow 1} \deg_\epsilon(f). \tag{1}$$

Both approximate degree and threshold degree have dual characterizations [30], obtained by appeal to linear programming duality. Specifically, $\deg_\epsilon(f) \geqslant d$ if and only if there is a function $\phi\colon \{-1,+1\}^n \to \mathbb{R}$ with the following two properties: $\langle \phi, f\rangle > \epsilon\|\phi\|_1$; and $\langle \phi, p\rangle = 0$ for every polynomial p of degree less than d. Rephrasing, ϕ must have large correlation with f but zero correlation with every low-degree polynomial. By weak linear programming duality, ϕ constitutes a proof that $\deg_\epsilon(f) \geqslant d$ and for that reason is said to *witness* the lower bound $\deg_\epsilon(f) \geqslant d$. In view of (1), this discussion generalizes to threshold degree. The dual characterization here states that $\deg_\pm(f) \geqslant d$ if and only if there is a nonzero function $\phi\colon \{-1,+1\}^n \to \mathbb{R}$ with the following two properties: $\phi(x)f(x) \geqslant 0$ for all x; and $\langle \phi, p\rangle = 0$ for every polynomial p of degree less than d. In this dual characterization, ϕ agrees in sign with f and is additionally orthogonal to polynomials of degree less than d. The sign-agreement property can be restated in terms of correlation, as $\langle \phi, f\rangle = \|\phi\|_1$. As before, ϕ is called a threshold degree *witness* for f.

What distinguishes the dual characterizations of approximate degree and threshold degree is how the dual object ϕ relates to f. Specifically, a threshold degree witness must agree in sign with f at every point. An approximate degree witness, on the other hand, need only exhibit such sign-agreement with f at *most* points, in that the points where the sign of ϕ is correct should account for most of the ℓ_1 norm of ϕ. As a result, constructing dual objects for threshold degree is significantly more difficult than for approximate degree. This difficulty is to be expected because the gap between threshold degree and approximate degree can be arbitrary, e.g., 1 versus $\Theta(n)$ for the majority function on n bits [24].

Hardness amplification via block-composition. Much of the recent work on approximate degree and threshold degree is concerned with composing functions in ways that amplify their hardness. Of particular significance here is *block-composition*, defined for

functions $f\colon \{-1,+1\}^n \to \{-1,+1\}$ and $g\colon X \to \{-1,+1\}$ as the function $f \circ g\colon X^n \to \{-1,+1\}$ given by $(f \circ g)(x_1,\dots,x_n) = f(g(x_1),\dots,g(x_n))$. Block-composition works particularly well for threshold degree. To use an already familiar example, the block-composition $\mathrm{AND}_{n^{1/3}} \circ \mathrm{OR}_{n^{2/3}}$ has threshold degree $\Omega(n^{1/3})$ whereas the constituent functions $\mathrm{AND}_{n^{1/3}}$ and $\mathrm{OR}_{n^{2/3}}$ have threshold degree 1. As a more extreme example, Sherstov [34] obtained a lower bound of $\Omega(n)$ on the threshold degree of the conjunction $h_1 \wedge h_2$ of two halfspaces $h_1, h_2\colon \{0,1\}^n \to \{0,1\}$, each of which by definition has threshold degree 1. The fact that threshold degree can increase spectacularly under block-composition is the basis of much previous work, including the best previous lower bounds [35, 37] on the threshold degree of AC^0 . Apart from threshold degree, block-composition has yielded strong results for approximate degree in various error regimes, including direct sum theorems [32], direct product theorems [31], and error amplification results [8, 9, 31, 40].

How, then, does one prove lower bounds on the threshold degree or approximate degree of a composed function $f \circ g$? It is here that the dual characterizations take center stage: they make it possible to prove lower bounds *algorithmically*, by constructing the corresponding dual object for the composed function. Such algorithmic proofs run the gamut in terms of technical sophistication, from straightforward to highly technical, but they have some structure in common. In most cases, one starts by obtaining dual objects ϕ and ψ for the constituent functions f and g, respectively, either by direct construction or by appeal to linear programming duality. They are then combined to yield a dual object Φ for the composed function, using *dual block-composition* [20, 32]:

$$\Phi(x_1,x_2,\dots,x_n) = \phi(\operatorname{sgn}\psi(x_1),\dots,\operatorname{sgn}\psi(x_n)) \prod_{i=1}^{n} |\psi(x_i)|. \quad (2)$$

This composed dual object often requires additional work to ensure sign-agreement or correlation with the composed Boolean function. Among the generic tools available to assist in this process is a "corrector" object ζ due to Razborov and Sherstov [26], with the following four properties: (i) ζ is orthogonal to low-degree polynomials; (ii) ζ takes on 1 at a prescribed point of the hypercube; (iii) ζ is bounded on inputs of low Hamming weight; and (iv) ζ vanishes on all other points of the hypercube. Using the Razborov–Sherstov object, suitably shifted and scaled, one can surgically correct the behavior of a given dual object Φ on a substantial fraction of inputs, thus modifying its metric properties without affecting its orthogonality to low-degree polynomials. This technique has played an important role in recent work, e.g., [7, 10–12].

Hardness amplification for approximate degree. Block-composition has produced a treasure trove of results on polynomial representations of Boolean functions, yet it is of little use when it comes to constructing functions with high *bounded-error* approximate degree. To illustrate the issue, consider arbitrary functions $f\colon \{-1,+1\}^{n_1} \to \{-1,+1\}$ and $g\colon \{-1,+1\}^{n_2} \to \{-1,+1\}$ with 1/3-approximate degrees $n_1^{\alpha_1}$ and $n_2^{\alpha_2}$, respectively, for some $0 < \alpha_1 < 1$ and $0 < \alpha_2 < 1$. It is well-known [33] that the composed function $f \circ g$ on $n_1 n_2$ variables has 1/3-approximate degree $O(n_1^{\alpha_1} n_2^{\alpha_2}) = O(n_1 n_2)^{\max\{\alpha_1,\alpha_2\}}$. This means that relative to the new number of variables, the block-composed function $f \circ g$ is asymptotically no harder to approximate to bounded error than the constituent functions f and g. In particular, one cannot use block-composition to transform functions on n bits with 1/3-approximate degree at most n^α into functions on $N \geqslant n$ bits with 1/3-approximate degree $\omega(N^\alpha)$.

Until recently, the best lower bound on the bounded-error approximate degree of AC^0 was $\Omega(n^{2/3})$, due to Aaronson and Shi [1]. Breaking this $n^{2/3}$ barrier was a fundamental problem in its own right, in addition to being a hard prerequisite for *threshold* degree lower bounds for AC^0 better than $\Omega(n^{2/3})$. This barrier was overcome in a brilliant paper of Bun and Thaler [11], who proved, for any constant $\epsilon > 0$, an $\Omega(n^{1-\epsilon})$ lower bound on the 1/3-approximate degree of AC^0. Their hardness amplification for approximate degree works as follows. Let $f\colon \{-1,+1\}^n \to \{-1,+1\}$ be given, with 1/3-approximate degree n^α for some $0 \leqslant \alpha < 1$. Bun and Thaler consider the block-composition $F = f \circ \mathrm{AND}_{\Theta(\log m)} \circ \mathrm{OR}_m$, for an appropriate parameter $m = \mathrm{poly}(n)$. As shown in earlier work [8, 32] on approximate degree, dual block-composition witnesses the lower bound $\deg_{1/3}(F) = \Omega(\deg_{1/3}(\mathrm{OR}_m)\deg_{1/3}(f)) = \Omega(\sqrt{m}\deg_{1/3}(f))$. Next, Bun and Thaler make the crucial observation that the dual object for OR_m has most of its ℓ_1 mass on inputs of Hamming weight $O(1)$, which in view of (2) implies that the dual object for F places most of its ℓ_1 mass on inputs of Hamming weight $\tilde{O}(n)$. The authors of [11] then use the Razborov–Sherstov corrector object to transfer the small amount of ℓ_1 mass that the dual object for F places on inputs of high Hamming weight, to inputs of low Hamming weight. The resulting dual object for F is supported entirely on inputs of low Hamming weight and therefore witnesses a lower bound on the 1/3-approximate degree of the *restriction* F' of F to inputs of low Hamming weight. By re-encoding the input to F', one finally obtains a function F'' on $\tilde{O}(n)$ variables with 1/3-approximate degree polynomially larger than that of f. This passage from f to F'' is the desired hardness amplification for approximate degree. We find it helpful to think of Bun and Thaler's technique as block-composition followed by input compression, to reduce the number of input variables in the block-composed function. To obtain an $\Omega(n^{1-\epsilon})$ lower bound on the approximate degree of AC^0, the authors of [11] start with a trivial circuit and iteratively apply the hardness amplification step a constant number of times, until approximate degree $\Omega(n^{1-\epsilon})$ is reached.

In follow-up work, Bun, Kothari, and Thaler [7] refined the technique of [11] by deriving optimal concentration bounds for the dual object for OR_m. They thereby obtained tight or nearly tight lower bounds on the 1/3-approximate degree of *surjectivity*, *element distinctness*, and other important problems. The most recent contribution to this line of work is due to Bun and Thaler [12], who prove an $\Omega(n^{1-\epsilon})$ lower bound on the $(1 - 2^{-n^{1-\epsilon}})$-approximate degree of AC^0 by combining the method of [11] with Sherstov's work [31] on direct product theorems for approximate degree. This near-linear lower bound substantially strengthens the authors' previous result [11] on the *bounded-error* approximate degree of AC^0, but does not address the threshold degree.

1.5 Our Approach

Threshold Degree of AC^0. Bun and Thaler [12] refer to obtaining an $\Omega(n^{1-\epsilon})$ threshold degree lower bound for AC^0 as the "main glaring open question left by our work." It is important to note

here that lower bounds on approximate degree, even with the error parameter exponentially close to 1 as in [12], have no implications for threshold degree. For example, there are functions [34] with $(1 - 2^{-\Theta(n)})$-approximate degree $\Theta(n)$ but threshold degree 1. Our proof of Theorem 1.1 is unrelated to the most recent work of Bun and Thaler [12] on the large-error approximate degree of AC^0 and instead builds on their earlier and simpler "block-composition followed by input compression" approach [11]. The centerpiece of our proof is a hardness amplification result for threshold degree, whereby any function f with threshold degree n^α for a constant $0 \leqslant \alpha < 1$ can be transformed efficiently and within AC^0 into a function F with polynomially larger threshold degree.

In more detail, let $f\colon \{-1,+1\}^n \to \{-1,+1\}$ be a function of interest, with threshold degree n^α. We consider the block-composition $f \circ \mathrm{MP}_m$, where $m = n^{O(1)}$ is an appropriate parameter and $\mathrm{MP}_m = \mathrm{AND}_m \circ \mathrm{OR}_{m^2}$ is the Minsky–Papert function with threshold degree $\Omega(m)$. We construct the dual object for MP_m from scratch to ensure concentration on inputs of Hamming weight $\tilde{O}(m)$. By applying dual block-composition to the threshold degree witnesses of f and MP_m, we obtain a dual object Φ witnessing the $\Omega(mn^\alpha)$ threshold degree of $f \circ \mathrm{MP}_m$. So far in the proof, our differences from [11] are as follows: (i) since our goal is amplification of threshold degree, we work with witnesses of threshold degree rather than approximate degree; (ii) to ensure rapid growth of threshold degree, we use block-composition with inner function $\mathrm{MP}_m = \mathrm{AND}_m \circ \mathrm{OR}_{m^2}$ of threshold degree $\Theta(m)$, in place of Bun and Thaler's inner function $\mathrm{AND}_{\Theta(\log m)} \circ \mathrm{OR}_m$ of threshold degree $\Theta(\log m)$.

Since the dual object for MP_m by construction has most of its ℓ_1 norm on inputs of Hamming weight $\tilde{O}(m)$, the dual object Φ for the composed function has most of its ℓ_1 norm on inputs of Hamming weight $\tilde{O}(nm)$. Analogous to [7, 11, 12], we would like to use the Razborov–Sherstov corrector object to *remove* the ℓ_1 mass that Φ has on the inputs of high Hamming weight, transferring it to inputs of low Hamming weight. This brings us to the novel and technically demanding part of our proof. Previous works [7, 11, 12] transferred the ℓ_1 mass from the inputs of high Hamming weight to the neighborhood of the all-zeroes input $(0, 0, \ldots, 0)$. An unavoidable feature of the Razborov–Sherstov transfer process is that it amplifies the ℓ_1 mass being transferred. When the transferred mass finally reaches its destination, it overwhelms Φ's original values at the local points, destroying Φ's sign-agreement with the composed function $f \circ \mathrm{MP}_m$. It is this difficulty that most prevented earlier works [7, 11, 12] from obtaining a strong threshold degree lower bound for AC^0.

We proceed differently. Instead of transferring the ℓ_1 mass of Φ from the inputs of high Hamming weight to the neighborhood of $(0, 0, \ldots, 0)$, we transfer it simultaneously to *exponentially many* strategically chosen neighborhoods. Split this way across many neighborhoods, the transferred mass does not overpower the original values of Φ and in particular does not change any signs. Working out the details of this transfer scheme requires subtle and lengthy calculations; it was not clear to us until the end that such a scheme exists. Once the transfer process is complete, we obtain a witness for the $\Omega(mn^\alpha)$ threshold degree of $f \circ \mathrm{MP}_m$ restricted to inputs of low Hamming weight. Compressing the input as in [7, 11], we obtain an amplification theorem for threshold degree. With this work behind us, the proof of Theorem 1.1 for any depth k amounts to starting with a trivial circuit and amplifying its threshold degree $O(k)$ times.

Sign-rank of AC^0. It is not known how to "lift" a threshold degree lower bound in a black-box manner to a sign-rank lower bound. In particular, Theorem 1.1 has no implications a priori for the sign-rank of AC^0. Our proofs of Theorems 1.2 and 1.3 are completely disjoint from Theorem 1.1 and are instead based on a stronger approximation-theoretic quantity that we call *γ-smooth threshold degree*. Formally, the γ-smooth threshold degree of a Boolean function $f\colon X \to \{-1,+1\}$ is the largest d for which there is a nonzero function $\phi\colon X \to \mathbb{R}$ with the following two properties: $\phi(x)f(x) \geqslant \gamma \cdot \|\phi\|_1/|X|$ for all $x \in X$; and $\langle \phi, p \rangle = 0$ for every polynomial p of degree less than d. Taking $\gamma = 0$ in this formalism, one recovers the standard dual characterization of the threshold degree of f. In particular, threshold degree is synonymous with 0-smooth threshold degree. The general case of γ-smooth threshold degree for $\gamma > 0$ requires threshold degree witnesses ϕ that are *min-smooth*, in that the absolute value of ϕ at any given point is at least a γ fraction of the average absolute value of ϕ over all points. A substantial advantage of *smooth* threshold degree is that it has immediate sign-rank implications. Specifically, any lower bound of d on the $2^{-O(d)}$-smooth threshold degree can be converted efficiently and in a black-box manner into a sign-rank lower bound of $2^{\Omega(d)}$, using a combination of the pattern matrix method [28, 30] and Forster's spectral lower bound on sign-rank [14, 15]. Accordingly, we obtain Theorems 1.2 and 1.3 by proving an $\Omega(n^{1-\epsilon})$ lower bound on the $2^{-O(n^{1-\epsilon})}$-smooth threshold degree of AC^0, for any constant $\epsilon > 0$.

At the core of our result is an amplification theorem for smooth threshold degree, whose repeated application makes it possible to prove arbitrarily strong lower bounds for AC^0. Amplifying smooth threshold degree is a complex juggling act due to the presence of two parameters—degree and smoothness—that must evolve in coordinated fashion. The approach of Theorem 1.1 is not useful here because the threshold degree witnesses that arise from the proof of Theorem 1.1 are highly nonsmooth. In more detail, when amplifying the threshold degree of a function f as in the proof of Theorem 1.1, two phenomena adversely affect the smoothness parameter. The first is block-composition itself as a composition technique, which in the regime of interest to us transforms *every* threshold degree witness for f into a hopelessly nonsmooth witness for the composed function. The other culprit is the input compression step, which re-encodes the input and thereby affects the smoothness in ways that are hard to control. To overcome these difficulties, we develop a novel approach based on what we call *local smoothness*.

Formally, let $\Phi\colon \mathbb{N}^n \to \mathbb{R}$ be a function of interest. For a subset $X \subseteq \mathbb{N}^n$ and a real number $K \geqslant 1$, we say that Φ is *K-smooth on X* if $|\Phi(x)| \leqslant K^{|x-x'|}|\Phi(x')|$ for all $x, x' \in X$. Put another way, for any two points of X at ℓ_1 distance d, the corresponding values of Φ differ in magnitude by a factor of at most K^d. In and of itself, a locally smooth function Φ need not be min-smooth because for a pair of points that are far from each other, the corresponding Φ-values can differ by many orders of magnitude. However, locally smooth functions exhibit extraordinary plasticity. Specifically, we show

how to modify a locally smooth function's metric properties—such as its support or the distribution of its ℓ_1 mass—without the change being detectable by low-degree polynomials. This apparatus makes it possible to restore min-smoothness to the dual object Φ that results from the block-composition step and preserve that min-smoothness throughout the input compression step, eliminating the two obstacles to min-smoothness in the earlier proof of Theorem 1.1. The new block-composition step uses a *locally smooth* witness for the threshold degree of MP_m, which needs to be built from scratch and is quite different from the witness in the proof of Theorem 1.1.

Our described approach departs considerably from previous work on the sign-rank of constant-depth circuits [10, 12, 26]. The analytic notion in those earlier papers is weaker than γ-smooth threshold degree and in particular allows the dual object to be *arbitrary* on a γ fraction of the inputs. This weaker property is acceptable when the main result is proved in one shot, with a closed-form construction of the dual object. By contrast, we must construct dual objects iteratively, with each iteration increasing the degree parameter and proportionately decreasing the smoothness parameter. This iterative process requires that the dual object in each iteration be min-smooth on the entire domain. Perhaps unexpectedly, we find γ-smooth threshold degree easier to work with than the weaker notion in previous work [10, 12, 26]. In particular, we are able to give a new and short proof of the $\exp(\Omega(n^{1/3}))$ lower bound on the sign-rank of AC^0, originally obtained by Razborov and Sherstov [26] with a much more complicated approach. The new proof can be found in the full version of our paper [39], where it serves as a prelude to our main result on the sign-rank of AC^0.

2 PRELIMINARIES

2.1 General

For a string $x \in \{0,1\}^n$ and a set $S \subseteq \{1,2,\ldots,n\}$, we let $x|_S$ denote the restriction of x to the indices in S. In other words, $x|_S = x_{i_1}x_{i_2}\ldots x_{i_{|S|}}$, where $i_1 < i_2 < \cdots < i_{|S|}$ are the elements of S. The *characteristic function* of a set $S \subseteq \{1,2,\ldots,n\}$ is given by

$$\mathbf{1}_S(x) = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{otherwise.} \end{cases}$$

For a logical condition C, we use the Iverson bracket

$$\mathbf{I}[C] = \begin{cases} 1 & \text{if } C \text{ holds,} \\ 0 & \text{otherwise.} \end{cases}$$

We let $\mathbb{N} = \{0,1,2,3,\ldots\}$ denote the set of natural numbers.

We adopt the extended real number system $\mathbb{R} \cup \{-\infty,\infty\}$ in all calculations, with the additional convention that $0/0 = 0$. We use the comparison operators in a unary capacity to denote one-sided intervals of the real line. Thus, $<a$, $\leqslant a$, $>a$, $\geqslant a$ stand for $(-\infty,a)$, $(-\infty,a]$, (a,∞), $[a,\infty)$, respectively. We let $\ln x$ and $\log x$ stand for the natural logarithm of x and the logarithm of x to base 2, respectively. We use the following two versions of the sign function:

$$\operatorname{sgn} x = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0, \end{cases} \qquad \widetilde{\operatorname{sgn}}\, x = \begin{cases} -1 & \text{if } x < 0, \\ 1 & \text{if } x \geqslant 0. \end{cases}$$

The term *Euclidean space* refers to $\mathbb{R}^n$ for some positive integer n. We let e_i denote the vector whose ith component is 1 and the others are 0. Thus, the vectors $e_1, e_2, \ldots, e_n$ form the standard basis for $\mathbb{R}^n$. For vectors x and y, we write $x \leqslant y$ to mean that $x_i \leqslant y_i$ for each i. The relations $\geqslant$, $<$, $>$ on vectors are defined analogously.

We frequently omit the argument in equations and inequalities involving functions, as in $\operatorname{sgn} p = (-1)^f$. Such statements are to be interpreted pointwise. For example, the statement "$f \geqslant 2|g|$ on X" means that $f(x) \geqslant 2|g(x)|$ for every $x \in X$. The positive and negative parts of a function $f\colon X \to \mathbb{R}$ are denoted $\operatorname{pos} f = \max\{f,0\}$ and $\operatorname{neg} f = \max\{-f,0\}$, respectively.

2.2 Boolean Functions and Circuits

We view Boolean functions as mappings $X \to \{0,1\}$ for some finite set X. More generally, we consider *partial* Boolean functions $f\colon X \to \{0,1,*\}$, with the output value $*$ used for don't-care inputs. The negation of a Boolean function f is denoted as usual by $\overline{f} = 1 - f$. The familiar functions $\mathrm{OR}_n\colon \{0,1\}^n \to \{0,1\}$ and $\mathrm{AND}_n\colon \{0,1\}^n \to \{0,1\}$ are given by $\mathrm{OR}_n(x) = \bigvee_{i=1}^n x_i$ and $\mathrm{AND}_n(x) = \bigwedge_{i=1}^n x_i$. We abbreviate $\mathrm{NOR}_n = \neg\mathrm{OR}_n$. The generalized *Minsky–Papert function* $\mathrm{MP}_{m,r}\colon (\{0,1\}^r)^m \to \{0,1\}$ is given by $\mathrm{MP}_{m,r}(x) = \bigwedge_{i=1}^m \bigvee_{j=1}^r x_{i,j}$. We abbreviate $\mathrm{MP}_m = \mathrm{MP}_{m,m^2}$, which is the right setting of parameters for most of our applications.

We adopt the standard notation for function composition, with $f \circ g$ defined by $(f \circ g)(x) = f(g(x))$. In addition, we use the $\circ$ operator to denote the *componentwise* composition of Boolean functions. Formally, the componentwise composition of $f\colon \{0,1\}^n \to \{0,1\}$ and $g\colon X \to \{0,1\}$ is the function $f \circ g\colon X^n \to \{0,1\}$ given by $(f \circ g)(x_1,x_2,\ldots,x_n) = f(g(x_1),g(x_2),\ldots,g(x_n))$. To illustrate, $\mathrm{MP}_{m,r} = \mathrm{AND}_m \circ \mathrm{OR}_r$. Componentwise composition is consistent with standard composition, which in the context of Boolean functions is only defined for $n = 1$. Thus, the meaning of $f \circ g$ is determined by the range of g and is never in doubt. Compositions $f_1 \circ f_2 \circ \cdots \circ f_k$ of three or more functions, where each instance of the $\circ$ operator can be standard or componentwise, are well-defined by associativity and do not require parenthesization.

For Boolean strings $x, y \in \{0,1\}^n$, we let $x \oplus y$ denote their bitwise XOR. The strings $x \wedge y$ and $x \vee y$ are defined analogously, with the binary connective applied bitwise. A *Boolean circuit* C in variables $x_1, x_2, \ldots, x_n$ is a circuit with inputs $x_1, \neg x_1, \ldots, x_n, \neg x_n$ and gates $\wedge$ and $\vee$. The circuit C is *monotone* if it does not use any of the negated inputs $\neg x_1, \neg x_2, \ldots, \neg x_n$. The *fan-in* of C is the maximum in-degree of any $\wedge$ or $\vee$ gate. Unless stated otherwise, we place no restrictions on the gate fan-in. The *size* of C is the number of $\wedge$ and $\vee$ gates. The *depth* of C is the maximum number of $\wedge$ and $\vee$ gates on any path from an input to the circuit output. With this convention, the circuit that computes $(x_1,x_2,\ldots,x_n) \mapsto x_1$ has depth 0. The circuit class AC^0 consists of function families $\{f_n\}_{n=1}^\infty$ such that $f_n\colon \{0,1\}^n \to \{0,1\}$ is computed by a Boolean circuit of size at most cn^c and depth at most c, for some constant $c \geqslant 1$ and all n. We specify small-depth layered circuits by indicating the type of gate used in each layer. For example, an *AND-OR-AND circuit* is a depth-3 circuit with the top and bottom layers composed of $\wedge$ gates, and middle layer composed of $\vee$ gates. A *Boolean formula* is a Boolean circuit in which every gate has fan-out 1. Common examples of Boolean formulas are DNF and CNF formulas.

2.3 Norms and Products

For a set X, we let $\mathbb{R}^X$ denote the linear space of real-valued functions on X. The *support* of a function $f \in \mathbb{R}^X$ is denoted $\operatorname{supp} f = \{x \in X : f(x) \neq 0\}$. For real-valued functions with finite support, we adopt the usual norms $\|f\|_\infty = \max_{x \in \operatorname{supp} f} |f(x)|$ and $\|f\|_1 = \sum_{x \in \operatorname{supp} f} |f(x)|$, as well as inner product $\langle f, g \rangle = \sum_{x \in \operatorname{supp} f \cap \operatorname{supp} g} f(x)g(x)$. This covers as a special case functions on finite sets. The *tensor product* of $f \in \mathbb{R}^X$ and $g \in \mathbb{R}^Y$ is denoted $f \otimes g \in \mathbb{R}^{X \times Y}$ and given by $(f \otimes g)(x, y) = f(x)g(y)$. The tensor product $f \otimes f \otimes \cdots \otimes f$ (n times) is abbreviated $f^{\otimes n}$. For a subset $S \subseteq \{1, 2, \ldots, n\}$ and a function $f : X \to \mathbb{R}$, we define $f^{\otimes S} : X^n \to \mathbb{R}$ by $f^{\otimes S}(x_1, x_2, \ldots, x_n) = \prod_{i \in S} f(x_i)$. As extremal cases, we have $f^{\otimes \varnothing} \equiv 1$ and $f^{\otimes \{1,2,\ldots,n\}} = f^{\otimes n}$. Tensor product notation generalizes naturally to *sets* of functions: $F \otimes G = \{f \otimes g : f \in F, g \in G\}$ and $F^{\otimes n} = \{f_1 \otimes f_2 \otimes \cdots \otimes f_n : f_1, f_2, \ldots, f_n \in F\}$.

Analogous to functions, we adopt the familiar norms for vectors $x \in \mathbb{R}^n$ in Euclidean space: $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$ and $\|x\|_1 = \sum_{i=1}^n |x_i|$. The latter norm is particularly prominent in this paper, and to avoid notational clutter we use $|x|$ interchangeably with $\|x\|_1$. We refer to $|x| = \|x\|_1$ as the *weight* of x. For any sets $X \subseteq \mathbb{N}^n$ and $W \subseteq \mathbb{R}$, we define $X|_W = \{x \in X : |x| \in W\}$. In the case of a one-element set $W = \{w\}$, we further shorten $X|_{\{w\}}$ to $X|_w$. To illustrate, $\mathbb{N}^n|_{\leqslant w}$ denotes the set of vectors whose components are natural numbers and sum to at most w, whereas $\{0,1\}^n|_w$ denotes the set of Boolean strings of length n and Hamming weight exactly w. For a function $f : X \to \mathbb{R}$ on a subset $X \subseteq \mathbb{N}^n$, we let $f|_W$ denote the restriction of f to $X|_W$. A typical instance of this notation would be $f|_{\leqslant w}$ for some real number w.

2.4 Orthogonal Content

For a multivariate real polynomial $p : \mathbb{R}^n \to \mathbb{R}$, we let $\deg p$ denote the total degree of p, i.e., the largest degree of any monomial of p. We use the terms *degree* and *total degree* interchangeably in this paper. It will be convenient to define the degree of the zero polynomial by $\deg 0 = -\infty$. For a real-valued function ϕ supported on a finite subset of $\mathbb{R}^n$, we define the *orthogonal content of* ϕ, denoted $\operatorname{orth} \phi$, to be the minimum degree of a real polynomial p for which $\langle \phi, p \rangle \neq 0$. We adopt the convention that $\operatorname{orth} \phi = \infty$ if no such polynomial exists. It is clear that $\operatorname{orth} \phi \in \mathbb{N} \cup \{\infty\}$, with the extremal cases $\operatorname{orth} \phi = 0 \iff \langle \phi, 1 \rangle \neq 0$ and $\operatorname{orth} \phi = \infty \iff \phi = 0$. Our next result, whose proof is available in the full version [39], records additional facts about orthogonal content.

Proposition 2.1. *Let X and Y be nonempty finite subsets of Euclidean space. Then:*

(i) $\operatorname{orth}(\phi + \psi) \geqslant \min\{\operatorname{orth} \phi, \operatorname{orth} \psi\}$ *for all* $\phi, \psi : X \to \mathbb{R}$;
(ii) $\operatorname{orth}(\phi \otimes \psi) = \operatorname{orth}(\phi) + \operatorname{orth}(\psi)$ *for* $\phi : X \to \mathbb{R}$ *and* $\psi : Y \to \mathbb{R}$;
(iii) $\operatorname{orth}(\phi^{\otimes n} - \psi^{\otimes n}) \geqslant \operatorname{orth}(\phi - \psi)$ *for* $\phi, \psi : X \to \mathbb{R}$ *and* $n \geqslant 1$.

Let $f : X \to \{0, 1\}$ be a given Boolean function, for a finite subset $X \subset \mathbb{R}^n$. The *threshold degree* of f, denoted $\deg_\pm(f)$, is the least degree of a real polynomial p that represents f in sign: $\operatorname{sgn} p(x) = (-1)^{f(x)}$ for each $x \in X$. One of the first results on polynomial representations of Boolean functions was the following tight lower bound on the threshold degree of MP_m, due to Minsky and Papert [21].

Theorem 2.2 (Minsky and Papert). $\deg_\pm(\mathrm{MP}_m) = \Omega(m)$.

2.5 Basic Dual Objects

As described in the introduction, we prove our main results constructively, by building explicit dual objects that witness the corresponding lower bounds. An important tool in this process is the following lemma, which is used to adjust a dual object's metric properties while preserving its orthogonality to low-degree polynomials.

Lemma 2.3. *Fix a point $u \in \mathbb{N}^n$ and a natural number $d < |u|$. Then there is $\zeta_u : \mathbb{N}^n \to \mathbb{R}$ such that*

$$\begin{aligned} &\operatorname{supp} \zeta_u \subseteq \{u\} \cup \{v \in \mathbb{N}^n : v \leqslant u \text{ and } |v| \leqslant d\}, \\ &\zeta_u(u) = 1, \\ &\|\zeta_u\|_1 \leqslant 1 + 2^d \binom{|u|}{d}, \\ &\operatorname{orth} \zeta_u > d. \end{aligned}$$

This result is a symmetrized version of Lemma 3.2 of Razborov and Sherstov [26]; a detailed proof is available in the full version [39] of this paper. The next lemma is an adaptation of a result due to Bun and Thaler [11] that serves to identify the dominant components of a vector. Its proof is also available in the full version [39] of this paper.

Lemma 2.4. *Fix $\theta > 0$ and let $v \in \mathbb{R}^n$ be an arbitrary vector with $\|v\|_1 \geqslant \theta$. Then there is $S \subseteq \{1, 2, \ldots, n\}$ such that*

$$\begin{aligned} &|S| \geqslant \frac{\|v\|_1}{2\|v\|_\infty}, \\ &\min_{i \in S} |v_i| \geqslant \frac{1}{|S|} \cdot \frac{\theta}{2(1 + \ln n)}, \\ &\sum_{i \notin S} |v_i| < \theta. \end{aligned}$$

3 THE THRESHOLD DEGREE OF AC^0

This section is devoted to our results on threshold degree. While we are mainly interested in the threshold degree of AC^0, the techniques developed here apply to a much broader class of functions. Specifically, we prove an *amplification theorem* that takes an arbitrary function f and builds from it a function F with higher threshold degree. The transformation $f \mapsto F$ is efficient with regard to circuit depth and size and in particular preserves membership in AC^0. To deduce our main results for AC^0, we start with a single-gate circuit and iteratively apply the amplification theorem to produce constant-depth circuits of higher and higher threshold degree. We develop this general machinery in Sections 3.1–3.3, followed by the applications to AC^0 in Section 3.4.

3.1 Shifting Probability Mass in Product Distributions

Consider a product distribution Λ on $\mathbb{N}^n$ whereby every component is concentrated near 0. The centerpiece of our threshold degree analysis, presented here, is the construction of an associated probability distribution $\tilde{\Lambda}$ that is supported entirely on inputs of low weight and cannot be distinguished from Λ by a low-degree polynomial. More formally, define $\mathfrak{B}(r, c, \alpha)$ to be the family of probability distributions λ on $\mathbb{N}$ such that $\operatorname{supp} \lambda = \{0, 1, 2, \ldots, r'\}$ for some

nonnegative integer $r' \leqslant r$, and in addition

$$\frac{c^{t+1}}{(t+1)^2\, 2^{\alpha t}} \leqslant \lambda(t) \leqslant \frac{1}{c(t+1)^2\, 2^{\alpha t}}, \qquad t \in \operatorname{supp} \lambda. \tag{3}$$

Distributions in this family are subject to pointwise constraints, hence the symbol $\mathfrak{B}$ for "bounded." In this notation, our analysis handles any distribution $\Lambda \in \mathfrak{B}(r, c, \alpha)^{\otimes n}$. The precise statement of our result is as follows.

Theorem 3.1. *Let $\Lambda \in \mathfrak{B}(r, c, \alpha)^{\otimes n}$ be given, for some integer $r \geqslant 0$ and reals $c > 0$ and $\alpha \geqslant 0$. Let d and θ be positive integers with*

$$\theta \geqslant 2d, \tag{4}$$

$$\theta \geqslant \frac{4en(1 + \ln n)}{c^2}. \tag{5}$$

Then there is a function $\tilde{\Lambda} \colon \mathbb{N}^n \to \mathbb{R}$ such that

$$\operatorname{supp} \tilde{\Lambda} \subseteq (\operatorname{supp} \Lambda)|_{<2\theta}, \tag{6}$$

$$\operatorname{orth}(\Lambda - \tilde{\Lambda}) > d, \tag{7}$$

$$|\Lambda - \tilde{\Lambda}| \leqslant \left(\frac{8nr}{c}\right)^d 2^{-\lceil \theta/r \rceil - \alpha \lceil \theta/2 \rceil + 2} \Lambda \qquad \text{on } \mathbb{N}^n|_{<2\theta}. \tag{8}$$

In general, the function $\tilde{\Lambda}$ constructed in Theorem 3.1 may not be a probability distribution. However, when θ is large enough relative to the other parameters, the pointwise property (8) forces $|\Lambda - \tilde{\Lambda}| \leqslant \Lambda$ on the support of $\tilde{\Lambda}$, and in particular $\tilde{\Lambda} \geqslant 0$. Since $\operatorname{orth}(\Lambda - \tilde{\Lambda}) > 0$ by construction, $\tilde{\Lambda}$ *is* a probability distribution in that case.

Proof of Theorem 3.1. For $c > 1$, we have $\mathfrak{B}(r, c, \alpha) = \varnothing$ and the theorem holds vacuously. Another degenerate possibility is $r = 0$, in which case Λ is the single-point distribution on 0^n, and therefore it suffices to take $\tilde{\Lambda} = \Lambda$. In what follows, we treat the general case when $c \in (0, 1], r \geqslant 1$.

For every vector $v \in \mathbb{N}^n$ with $\|v\|_1 \geqslant \theta$, let $S(v) \subseteq \{1, 2, \ldots, n\}$ denote the corresponding subset identified by Lemma 2.4. To restate the lemma's guarantees,

$$|S(v)| \geqslant \frac{\theta}{r}, \qquad v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}, \tag{9}$$

$$\min_{i \in S(v)} v_i \geqslant \frac{\theta}{2|S(v)|(1 + \ln n)}, \qquad v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}, \tag{10}$$

$$\|v|_{\overline{S(v)}}\|_1 < \theta. \qquad v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}. \tag{11}$$

Property (11) implies that

$$\|v|_{S(v)}\|_1 > \theta, \qquad v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}, \tag{12}$$

and in particular

$$\|v|_{S(v)}\|_1 > d, \qquad v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}. \tag{13}$$

For each $i = 1, 2, \ldots, n$ and each $u \in \mathbb{N}^i|_{>d}$, Lemma 2.3 gives a function $\zeta_u \colon \mathbb{N}^i \to \mathbb{R}$ such that

$$\operatorname{supp} \zeta_u \subseteq \{u\} \cup \{v \in \mathbb{N}^i : v \leqslant u \text{ and } |v| \leqslant d\}, \tag{14}$$

$$\zeta_u(u) = 1, \tag{15}$$

$$\|\zeta_u\|_1 \leqslant 1 + 2^d \binom{\|u\|_1}{d}, \tag{16}$$

$$\operatorname{orth} \zeta_u > d, \tag{17}$$

and in particular

$$\begin{aligned} \|\zeta_u\|_\infty &\leqslant \max\{|\zeta_u(u)|, \|\zeta_u\|_1 - |\zeta_u(u)|\} \\ &\leqslant 2^d \binom{\|u\|_1}{d} \\ &\leqslant 2\|u\|_1{}^d. \end{aligned} \tag{18}$$

The central object of study in our proof is the following function $\zeta \colon \mathbb{N}^n \to \mathbb{R}$, built from the auxiliary objects $S(v)$ and ζ_u just introduced:

$$\zeta(x) = \sum_{v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}} \Lambda(v)\, \zeta_{v|_{S(v)}}(x|_{S(v)})\, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}]. \tag{19}$$

The expression on the right-hand side is well-formed because, to restate (13), each string $v|_{S(v)}$ has weight greater than d and can therefore be used as a subscript in $\zeta_{v|_{S(v)}}$. Specializing (17) and (18),

$$\operatorname{orth} \zeta_{v|_{S(v)}} > d, \qquad v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}, \tag{20}$$

$$\|\zeta_{v|_{S(v)}}\|_\infty \leqslant 2(nr)^d, \qquad v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}. \tag{21}$$

Property (14) ensures that $\zeta_{v|_{S(v)}}(x|_{S(v)})\, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \neq 0$ only when $x \leqslant v$. It follows that

$$\begin{aligned} \operatorname{supp} \zeta &\subseteq \bigcup_{v \in \operatorname{supp} \Lambda} \{x \in \mathbb{N}^n : x \leqslant v\} \\ &= \operatorname{supp} \Lambda, \end{aligned} \tag{22}$$

where second step is valid because $\Lambda \in \mathfrak{B}(r, c, \alpha)^{\otimes n}$.

Before carrying on with the proof, we take a moment to simplify the defining expression for ζ. For any $v \in \mathbb{N}^n|_{\geqslant 2\theta}$, we have

$$\begin{aligned} &\zeta_{v|_{S(v)}}(x|_{S(v)})\, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \\ &\quad = \zeta_{v|_{S(v)}}(x|_{S(v)})\, \mathbf{I}[x|_{S(v)} = v|_{S(v)} \text{ or } \|x|_{S(v)}\|_1 \leqslant d] \\ &\qquad\qquad \times \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \\ &\quad = \zeta_{v|_{S(v)}}(x|_{S(v)})(\mathbf{I}[x|_{S(v)} = v|_{S(v)}] \\ &\qquad\qquad + \mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d])\mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \\ &\quad = \zeta_{v|_{S(v)}}(x|_{S(v)})\mathbf{I}[x = v] \\ &\qquad\qquad + \zeta_{v|_{S(v)}}(x|_{S(v)})\mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d]\, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \\ &\quad = \mathbf{I}[x = v] + \zeta_{v|_{S(v)}}(x|_{S(v)})\mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d]\, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}], \end{aligned}$$

where the first, second, and fourth steps are valid by (14), (13), and (15), respectively. Making this substitution in the defining equation for ζ,

$$\begin{aligned} \zeta(x) = \Bigg(&\sum_{v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}} \Lambda(v) \zeta_{v|_{S(v)}}(x|_{S(v)}) \mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d] \\ &\times \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \Bigg) + \sum_{v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}} \Lambda(v) \mathbf{I}[x = v]. \end{aligned} \tag{23}$$

We proceed to establish key properties of ζ.

Step 1: Orthogonality. By Proposition 2.1(ii), each term in the summation on the right-hand side of (19) is a function orthogonal to polynomials of degree less than $\operatorname{orth}\zeta_{v|_{S(v)}}$. Therefore,

$$\begin{aligned}\operatorname{orth}\zeta &\geqslant \min_{v\in(\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} \operatorname{orth}\zeta_{v|_{S(v)}} \\ &> d, \end{aligned} \tag{24}$$

where the first step uses Proposition 2.1(i), and the second step applies (20).

Step 2: Heavy inputs. We now examine the behavior of ζ on inputs of weight at least 2θ, which we think of as "heavy." For any string $v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}$, we have

$$\begin{aligned} x \in \mathbb{N}^n|_{\geqslant 2\theta} &\implies \|x\|_1 > d + \theta \\ &\implies \|x|_{S(v)}\|_1 > d \quad \vee \quad \|x|_{\overline{S(v)}}\|_1 > \theta \\ &\implies \|x|_{S(v)}\|_1 > d \quad \vee \quad x|_{\overline{S(v)}} \neq v|_{\overline{S(v)}}, \end{aligned}$$

where the final implication uses (11). We conclude that the first summation in (23) vanishes on $\mathbb{N}^n|_{\geqslant 2\theta}$, so that

$$\zeta(x) = \Lambda(x), \qquad x \in \mathbb{N}^n|_{\geqslant 2\theta}. \tag{25}$$

This completes the analysis of heavy inputs.

Step 3: Light inputs. We now turn to inputs of weight less than 2θ, the most technical part of the proof. Fix an arbitrary string $x \in (\operatorname{supp}\Lambda)|_{<2\theta}$. Then

$$\begin{aligned} \frac{|\zeta(x)|}{\Lambda(x)} &= \left| \sum_{v\in(\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} \frac{\Lambda(v)}{\Lambda(x)} \zeta_{v|_{S(v)}}(x|_{S(v)})\, \mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d] \right. \\ &\qquad \left. \times\, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \right| \\ &\leqslant \sum_{v\in(\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} \frac{\Lambda(v)}{\Lambda(x)} |\zeta_{v|_{S(v)}}(x|_{S(v)})| \mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d] \\ &\qquad \times\, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \\ &\leqslant 2(nr)^d \sum_{v\in(\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} \frac{\Lambda(v)}{\Lambda(x)} \mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d] \\ &\qquad \times\, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \\ &= 2(nr)^d \sum_{\substack{S\subseteq\{1,\ldots,n\}:\\ |S|\geqslant\theta/r}} \mathbf{I}[\|x|_S\|_1 \leqslant d] \sum_{\substack{v\in(\operatorname{supp}\Lambda)|_{\geqslant 2\theta}:\\ S(v)=S}} \frac{\Lambda(v)}{\Lambda(x)} \\ &\qquad \times\, \mathbf{I}[x|_{\overline{S}} = v|_{\overline{S}}] \\ &\leqslant 2(nr)^d \sum_{\substack{S\subseteq\{1,\ldots,n\}:\\ |S|\geqslant\theta/r}} \mathbf{I}[\|x|_S\|_1 \leqslant d] \sum_{\substack{v\in\mathbb{N}^n:\\ \sum_{i\in S} v_i \geqslant \theta,\\ \min_{i\in S} v_i \geqslant \frac{\theta}{2|S|(1+\ln n)}}} \frac{\Lambda(v)}{\Lambda(x)} \\ &\qquad \times\, \mathbf{I}[x|_{\overline{S}} = v|_{\overline{S}}], \end{aligned} \tag{26}$$

where the first step uses (23); the second step applies the triangle inequality; the third step is valid by (21); the fourth step amounts to collecting terms according to $S(v)$, which by (9) has cardinality at least θ/r; and the fifth step uses (10) and (12).

Bounding (26) requires a bit of work. To start with, write $\Lambda = \bigotimes_{i=1}^n \lambda_i$ for some $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathfrak{B}(r, c, \alpha)$. In the full version [39] of this paper, we prove the following for every nonempty $S \subseteq \{1, 2, \ldots, n\}$:

$$\mathbf{I}[\|x|_S\|_1 \leqslant d] \prod_{i\in S} \frac{1}{\lambda_i(x_i)} \leqslant c^{-|S|} \left(\frac{2^\alpha \mathrm{e}^2}{c}\right)^d, \tag{27}$$

$$\begin{aligned} &\sum_{\substack{v\in\mathbb{N}^n:\\ \sum_{i\in S} v_i \geqslant \theta,\\ \min_{i\in S} v_i \geqslant \frac{\theta}{2|S|(1+\ln n)}}} \frac{\Lambda(v)}{\Lambda(x)} \mathbf{I}[x|_{\overline{S}} = v|_{\overline{S}}] \\ &\qquad \leqslant 2^{-\alpha\theta} \left(\frac{2|S|(1+\ln n)}{c\theta}\right)^{|S|} \prod_{i\in S} \frac{1}{\lambda_i(x_i)}. \end{aligned} \tag{28}$$

It remains to put together the bounds obtained so far. We have:

$$\begin{aligned} \frac{|\zeta(x)|}{\Lambda(x)} &\leqslant 2(nr)^d \sum_{\substack{S\subseteq\{1,\ldots,n\}:\\ |S|\geqslant\theta/r}} \mathbf{I}[\|x|_S\|_1 \leqslant d] \cdot 2^{-\alpha\theta} \left(\frac{2|S|(1+\ln n)}{c\theta}\right)^{|S|} \\ &\qquad \times \prod_{i\in S} \frac{1}{\lambda_i(x_i)} \\ &\leqslant 2(nr)^d \sum_{\substack{S\subseteq\{1,\ldots,n\}:\\ |S|\geqslant\theta/r}} 2^{-\alpha\theta} \left(\frac{2|S|(1+\ln n)}{c^2\theta}\right)^{|S|} \cdot \left(\frac{2^\alpha \mathrm{e}^2}{c}\right)^d \\ &\leqslant 2 \cdot \frac{(\mathrm{e}^2 nr/c)^d}{2^{\alpha\lceil\theta/2\rceil}} \sum_{\substack{S\subseteq\{1,\ldots,n\}:\\ |S|\geqslant\theta/r}} \left(\frac{2|S|(1+\ln n)}{c^2\theta}\right)^{|S|} \\ &= 2 \cdot \frac{(\mathrm{e}^2 nr/c)^d}{2^{\alpha\lceil\theta/2\rceil}} \sum_{s=\lceil\theta/r\rceil}^{\infty} \binom{n}{s} \left(\frac{2s(1+\ln n)}{c^2\theta}\right)^s \\ &\leqslant 2 \cdot \frac{(\mathrm{e}^2 nr/c)^d}{2^{\alpha\lceil\theta/2\rceil}} \sum_{s=\lceil\theta/r\rceil}^{\infty} \left(\frac{\mathrm{e}n}{s} \cdot \frac{2s(1+\ln n)}{c^2\theta}\right)^s \\ &\leqslant 2 \cdot \frac{(\mathrm{e}^2 nr/c)^d}{2^{\alpha\lceil\theta/2\rceil}} \sum_{s=\lceil\theta/r\rceil}^{\infty} 2^{-s} \\ &= 4 \cdot \frac{(\mathrm{e}^2 nr/c)^d}{2^{\alpha\lceil\theta/2\rceil+\lceil\theta/r\rceil}}, \end{aligned}$$

where the first step follows from (26) and (28); the second step substitutes the bound from (27); the third step uses (4); and the next-to-last step uses (5). In summary, we have shown that

$$|\zeta(x)| \leqslant 4 \cdot \frac{(\mathrm{e}^2 nr/c)^d}{2^{\alpha\lceil\theta/2\rceil+\lceil\theta/r\rceil}} \Lambda(x), \qquad x \in (\operatorname{supp}\Lambda)|_{<2\theta}. \tag{29}$$

Step 4: Finishing the proof. Define $\tilde{\Lambda} = \Lambda - \zeta$. Then the support property (6) follows from (22) and (25); the analytic indistinguishability property (7) follows from (24); and the pointwise property (8) follows from (22) and (29). ∎

3.2 A Bounded Dual Polynomial for MP

We now turn to the construction of a gadget for our amplification theorem. Let $\mathfrak{B}^*(r, c, \alpha)$ denote the family of probability distributions λ on $\mathbb{N}$ such that $\operatorname{supp}\lambda = \{0, 1, 2, \ldots, r'\}$ for some nonnegative integer $r' \leqslant r$, and moreover

$$\frac{c}{(t+1)^2\, 2^{\alpha t}} \leqslant \lambda(t) \leqslant \frac{1}{c(t+1)^2\, 2^{\alpha t}}, \qquad t \in \operatorname{supp}\lambda.$$

In this family, a distribution's weight at any given point is prescribed up to the multiplicative constant c, in contrast to the exponentially large range allowed in the definition of $\mathfrak{B}(r, c, \alpha)$. For all parameter settings, we have $\mathfrak{B}^*(r, c, \alpha) \subseteq \mathfrak{B}(r, c, \alpha)$. Indeed, the containment holds trivially for $c \leqslant 1$, and remains valid for $c > 1$ because the left-hand side and right-hand side are both empty in that case. It will be helpful to have shorthand notation for *translates* of distributions in $\mathfrak{B}(r, c, \alpha)$: we define $\mathfrak{B}^*(r, c, \alpha, \Delta)$ for $\Delta \geqslant 0$ to be the family of probability distributions λ on $\mathbb{N}$ such that $\lambda(t) = \lambda'(t - a)$ for some $\lambda' \in \mathfrak{B}^*(r, c, \alpha)$ and $a \in [0, \Delta]$.

As our next step toward analyzing the threshold degree of AC^0, we will construct a dual object that witnesses the high threshold degree of $\mathrm{MP}^*_{m,r}$ and possesses additional metric properties in the sense of $\mathfrak{B}^*$.

Theorem 3.2. *For some absolute constants $c_1, c_2 \in (0, 1)$ and all positive integers m and r, there are probability distributions Λ_0, Λ_1 such that*

$$\Lambda_i \in \operatorname{conv}\left(\mathfrak{B}^*\left(r, c_1, \frac{c_2}{\sqrt{r}}, 1\right)^{\otimes m}\right), \qquad i = 0, 1, \tag{30}$$

$$\operatorname{supp}\Lambda_i \subseteq (\mathrm{MP}^*_{m,r})^{-1}(i), \qquad i = 0, 1, \tag{31}$$

$$\operatorname{orth}(\Lambda_1 - \Lambda_0) \geqslant \min\{m, c_1\sqrt{r}\}. \tag{32}$$

A proof of this theorem is available in the full version [39] of our paper.

3.3 Hardness Amplification

We now present a black-box transformation that takes any given circuit in n variables with threshold degree $n^{1-\epsilon}$ into a circuit with polynomially larger threshold degree, $\Omega(n^{1-\frac{\epsilon}{1+\epsilon}})$. This hardness amplification procedure increases the circuit size additively by $n^{O(1)}$ and the circuit depth by 2, preserving membership in AC^0.

Theorem 3.3. *Let n, m, N be positive integers. Then there is an (explicitly given) transformation $H\colon \{0,1\}^N \to \{0,1\}^n$, computable by an AND-OR-AND circuit of size $(Nnm)^{O(1)}$ with bottom fan-in $O(\log(nm))$, such that for all functions $f\colon \{0,1\}^n \to \{0,1\}$,*

$$\min\{\deg_\pm(f \circ H), \deg_\pm(f \circ \neg H)\} \geqslant \min\left\{cm\deg_\pm(f), \frac{cN}{50m\log^2(n+m)} - n\right\}\log(n+m),$$

where $0 < c < 1$ is an absolute constant independent of n, m, N.

The proof of this central result combines Theorems 3.1 and 3.2 and can be found in the full version [39]. For the function H with multi-bit output, the notation $\neg H$ above refers to the function obtained by negating *each* of H's outputs.

3.4 Threshold Degree and Discrepancy of AC^0

We have reached our main result on the sign-representation of constant-depth circuits. For any $\epsilon > 0$, the next theorem constructs a circuit family in AC^0 with threshold degree $\Omega(n^{1-\epsilon})$. The proof amounts to a recursive application of the hardness amplification procedure of Section 3.3.

Theorem 3.4. *Fix an integer $k \geqslant 1$. Then there is an (explicitly given) family of functions $\{f_{k,n}\}_{n=1}^\infty$, where $f_{k,n}\colon \{0,1\}^n \to \{0,1\}$ has threshold degree*

$$\deg_\pm(f_{k,n}) = \Omega\left(n^{\frac{k-1}{k+1}} \cdot (\log n)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right) \tag{33}$$

and is computable by a monotone Boolean circuit of size $n^{O(1)}$ and depth k. In addition, the circuit for $f_{k,n}$ has bottom fan-in $O(\log n)$ for all $k \neq 2$.

Proof. The proof is by induction on k. The base cases $k = 1$ and $k = 2$ correspond to the families

$$f_{1,n}(x) = x_1, \qquad n = 1, 2, 3, \ldots,$$
$$f_{2,n}(x) = \mathrm{MP}_{\lfloor n^{1/3}\rfloor}, \qquad n = 1, 2, 3, \ldots.$$

For the former, the threshold degree lower bound (33) is trivial. For the latter, it follows from Theorem 2.2.

For the inductive step, fix $k \geqslant 3$. Due to the asymptotic nature of (33), it is enough to construct the functions in $\{f_{k,n}\}_{n=1}^\infty$ for n larger than a certain constant of our choosing. As a starting point, the inductive hypothesis gives an explicit family $\{f_{k-2,n}\}_{n=1}^\infty$ in which $f_{k-2,n}\colon \{0,1\}^n \to \{0,1\}$ has threshold degree

$$\deg_\pm(f_{k-2,n}) = \Omega\left(n^{\frac{k-3}{k-1}} \cdot (\log n)^{-\frac{1}{k-1}\lceil\frac{k-4}{2}\rceil\lfloor\frac{k-4}{2}\rfloor}\right) \tag{34}$$

and is computable by a monotone Boolean circuit of size $n^{O(1)}$ and depth $k - 2$. We view the circuit for $f_{k-2,n}$ as composed of $k - 2$ layers of alternating gates, where without loss of generality the bottom layer consists of AND gates. This last property can be forced by using $\neg f_{k-2,n}(\neg x_1, \neg x_2, \ldots, \neg x_n)$ instead of $f_{k-2,n}(x_1, x_2, \ldots, x_n)$, which interchanges the circuit's AND and OR gates without affecting the threshold degree, circuit depth, or circuit size.

Now, let $c > 0$ be the absolute constant from Theorem 3.3. For every N larger than a certain constant, we apply Theorem 3.3 with

$$n = \left\lceil N^{\frac{k-1}{k+1}}(\log N)^{-\frac{1}{k+1}\lceil\frac{k-4}{2}\rceil\lfloor\frac{k-4}{2}\rfloor - \frac{2(k-1)}{k+1}} \cdot \frac{c}{100}\right\rceil, \tag{35}$$

$$m = \left\lceil N^{\frac{2}{k+1}}(\log N)^{\frac{1}{k+1}\lceil\frac{k-4}{2}\rceil\lfloor\frac{k-4}{2}\rfloor - \frac{4}{k+1}}\right\rceil, \tag{36}$$

$$f = f_{k-2,n} \tag{37}$$

to obtain a function $H_N\colon \{0,1\}^N \to \{0,1\}^n$ such that the composition $F_N = f_{k-2,n} \circ H_N$ has threshold degree

$$\begin{aligned}\deg_\pm(F_N) &\geqslant \min\left\{cm\deg_\pm(f_{k-2,n}), \frac{cN}{50m\log^2(n+m)} - n\right\}\log(n+m)\\ &= \Theta\left(N^{\frac{k-1}{k+1}}(\log N)^{-\frac{1}{k+1}\lceil\frac{k-4}{2}\rceil\lfloor\frac{k-4}{2}\rfloor - \frac{k-3}{k+1}}\right)\\ &= \Theta\left(N^{\frac{k-1}{k+1}}(\log N)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right),\end{aligned} \tag{38}$$

where the second step uses (34)–(36). Moreover, Theorem 3.3 ensures that H_N is computable by an AND-OR-AND circuit of polynomial size and bottom fan-in $O(\log N)$. The bottom layer of $f_{k-2,n}$ consists of AND gates, which can be merged with the top layer of H_N to give a circuit for $F_N = f_{k-2,n} \circ H_N$ of depth $(k-2)+3-1 = k$.

We have thus constructed, for some constant N_0, a family of functions $\{F_N\}_{N=N_0}^{\infty}$ in which each $F_N \colon \{0,1\}^N \to \{0,1\}$ has threshold degree (38) and is computable by a Boolean circuit of polynomial size, depth k, and bottom fan-in $O(\log N)$. Now, take the circuit for F_N and replace the negated inputs in it with N new, unnegated inputs. The resulting *monotone* circuit on $2N$ variables clearly has threshold degree at least that of F_N. This completes the inductive step. □

Theorem 3.4 settles Theorem 1.1 from the introduction. In the full version of this paper, we use the pattern matrix method to "lift" Theorem 3.4 to communication complexity.

4 THE SIGN-RANK OF AC^0

We now turn to our second main result, a near-optimal lower bound on the sign-rank of constant-depth circuits. Our approach here is based on the notion of *local smoothness* and is unrelated to the threshold degree analysis. We start by defining local smoothness and recording basic properties of locally smooth functions. Next, we develop techniques for manipulating locally smooth functions to achieve desired global behavior, without the manipulations being detectable by low-degree polynomials. To apply this machinery to constant-depth circuits, we design a locally smooth dual polynomial for the Minsky–Papert function. We then use this dual object to prove an amplification theorem for *smooth* threshold degree. We apply the amplification theorem iteratively to construct, for any $\epsilon > 0$, a constant-depth circuit with $\exp(-O(n^{1-\epsilon}))$-smooth threshold degree $\Omega(n^{1-\epsilon})$. Finally, we conclude with a proof of our main result on the sign-rank of AC^0.

Due to the page limit, we defer the remainder of this section to the full version of this paper [39].

ACKNOWLEDGMENTS

This work was supported in part by NSF grants CCF-1814947 and CCF-1149018 and an Alfred P. Sloan Foundation Research Fellowship. The authors are thankful to Mark Bun and Justin Thaler for valuable comments on an earlier version of this paper.

REFERENCES

[1] Scott Aaronson and Yaoyun Shi. 2004. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM* 51, 4 (2004), 595–605.
[2] Andris Ambainis, Andrew M. Childs, Ben Reichardt, Robert Špalek, and Shengyu Zhang. 2010. Any AND-OR Formula of Size N can be Evaluated in time $N^{1/2+o(1)}$ on a Quantum Computer. *SIAM J. Comput.* 39, 6 (2010), 2513–2530.
[3] James Aspnes, Richard Beigel, Merrick L. Furst, and Steven Rudich. 1994. The Expressive Power of Voting Polynomials. *Combinatorica* 14, 2 (1994), 135–148.
[4] László Babai, Peter Frankl, and Janos Simon. 1986. Complexity classes in communication complexity theory. In *FOCS*. 337–347.
[5] Paul Beame and Trinh Huynh. 2012. Multiparty Communication Complexity and Threshold Circuit Size of AC^0. *SIAM J. Comput.* 41, 3 (2012), 484–518.
[6] Harry Buhrman, Nikolai K. Vereshchagin, and R. de Wolf. 2007. On computation and communication with small bias. In *CCC*. 24–32.
[7] Mark Bun, Robin Kothari, and Justin Thaler. 2018. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. In *STOC*. 297–310.
[8] Mark Bun and Justin Thaler. 2015. Hardness Amplification and the Approximate Degree of Constant-Depth Circuits. In *ICALP*. 268–280.
[9] Mark Bun and Justin Thaler. 2016. Approximate Degree and the Complexity of Depth Three Circuits. In *Electronic Colloquium on Computational Complexity (ECCC)*. Report TR16-121.
[10] Mark Bun and Justin Thaler. 2016. Improved Bounds on the Sign-Rank of AC^0. In *ICALP*. 37:1–37:14.
[11] Mark Bun and Justin Thaler. 2017. A Nearly Optimal Lower Bound on the Approximate Degree of AC^0. In *FOCS*. 1–12.
[12] Mark Bun and Justin Thaler. 2018. The Large-Error Approximate Degree of AC^0. In *Electronic Colloquium on Computational Complexity (ECCC)*. Report TR18-143.
[13] Arkadev Chattopadhyay. 2007. Discrepancy and the power of bottom fan-in in depth-three circuits. In *FOCS*. 449–458.
[14] Jürgen Forster. 2002. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.* 65, 4 (2002), 612–625.
[15] Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans-Ulrich Simon. 2001. Relations between communication complexity, linear arrangements, and computational complexity. In *Proc. of the 21st Conf. on Foundations of Software Technology and Theoretical Computer Science (FST TCS)*. 171–182.
[16] Adam R. Klivans and Rocco A. Servedio. 2004. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. Syst. Sci.* 68, 2 (2004), 303–318.
[17] Adam R. Klivans and Rocco A. Servedio. 2006. Toward Attribute Efficient Learning of Decision Lists and Parities. *J. Machine Learning Research* 7 (2006), 587–602.
[18] Matthias Krause and Pavel Pudlák. 1997. On the Computational Power of Depth-2 Circuits with Threshold and Modulo Gates. *Theor. Comput. Sci.* 174, 1–2 (1997), 137–156.
[19] Eyal Kushilevitz and Noam Nisan. 1997. *Communication complexity*. Cambridge University Press.
[20] Troy Lee. 2009. A note on the sign degree of formulas. Available at http://arxiv.org/abs/0909.4607.
[21] Marvin L. Minsky and Seymour A. Papert. 1969. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, Mass.
[22] Noam Nisan and Mario Szegedy. 1994. On the degree of Boolean functions as real polynomials. *Computational Complexity* 4 (1994), 301–313.
[23] Ryan O'Donnell and Rocco A. Servedio. 2010. New degree bounds for polynomial threshold functions. *Combinatorica* 30, 3 (2010), 327–358.
[24] Ramamohan Paturi. 1992. On the degree of polynomials that approximate symmetric Boolean functions. In *STOC*. 468–474.
[25] Ramamohan Paturi and Janos Simon. 1986. Probabilistic communication complexity. *J. Comput. Syst. Sci.* 33, 1 (1986), 106–123.
[26] Alexander A. Razborov and Alexander A. Sherstov. 2010. The sign-rank of AC^0. *SIAM J. Comput.* 39, 5 (2010), 1833–1855. Preliminary version in FOCS, 2008.
[27] Alexander A. Sherstov. 2008. Halfspace matrices. *Computational Complexity* 17, 2 (2008), 149–178. Preliminary version in CCC, 2007.
[28] Alexander A. Sherstov. 2009. Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.* 38, 6 (2009), 2113–2129. Preliminary version in STOC, 2007.
[29] Alexander A. Sherstov. 2010. Communication Complexity Under Product and Nonproduct Distributions. *Computational Complexity* 19, 1 (2010), 135–150. Preliminary version in CCC, 2008.
[30] Alexander A. Sherstov. 2011. The pattern matrix method. *SIAM J. Comput.* 40, 6 (2011), 1969–2000. Preliminary version in STOC, 2008.
[31] Alexander A. Sherstov. 2012. Strong Direct Product Theorems for Quantum Communication and Query Complexity. *SIAM J. Comput.* 41, 5 (2012), 1122–1165. Preliminary version in STOC, 2011.
[32] Alexander A. Sherstov. 2013. The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.* 42, 6 (2013), 2329–2374. Preliminary version in FOCS, 2009.
[33] Alexander A. Sherstov. 2013. Making polynomials robust to noise. *Theory of Computing* 9 (2013), 593–615. Preliminary version in STOC, 2012.
[34] Alexander A. Sherstov. 2013. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica* 33, 1 (2013), 73–96. Preliminary version in STOC, 2010.
[35] Alexander A. Sherstov. 2014. Breaking the Minsky–Papert barrier for constant-depth circuits. In *STOC*. 223–232. Full version available as ECCC Report TR14-009, January 2014.
[36] Alexander A. Sherstov. 2014. Communication lower bounds using directional derivatives. *J. ACM* 61, 6 (2014), 1–71. Preliminary version in STOC, 2013.
[37] Alexander A. Sherstov. 2015. The Power of Asymmetry in Constant-Depth Circuits. In *FOCS*. 431–450.
[38] Alexander A. Sherstov. 2016. The multiparty communication complexity of set disjointness. *SIAM J. Comput.* 45, 4 (2016), 1450–1489. Preliminary version in STOC, 2009.
[39] Alexander A. Sherstov and Pei Wu. 2019. Near-optimal lower bounds on the threshold degree and sign-rank of AC^0. In *Electronic Colloquium on Computational Complexity (ECCC)*. Report TR19-003.
[40] Justin Thaler. 2016. Lower Bounds for the Approximate Degree of Block-Composed Functions. In *ICALP*. 17:1–17:15.