NEAR-OPTIMAL LOWER BOUNDS ON THE THRESHOLD DEGREE AND SIGN-RANK OF AC^0

ALEXANDER A. SHERSTOV AND PEI WU

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 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.

* Computer Science Department, UCLA, Los Angeles, CA 90095. Supported by NSF grant CCF-1814947 and an Alfred P. Sloan Foundation Research Fellowship.
✉ {sherstov,pwu}@cs.ucla.edu .

Contents

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* [5], *voting polynomial degree* [30], *PTF degree* [37], and *sign degree* [12]. 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 [35] on limitations of perceptrons. The authors of [35] 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 range from circuit lower bounds [30, 31] and size-depth trade-offs [40, 58] to computational learning [28, 27, 38, 4, 50, 52, 15, 53, 59] and structural complexity theory [10].

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. [5], 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 critical role in the first construction [45, 47] of an AC^0 circuit with exponentially small discrepancy and hence large communication complexity in nearly every model. That discrepancy result was used in [45] to show the optimality of Allender's classic simulation of AC^0 by majority circuits, solving the open problem [30] on the relation between the two circuit classes. Subsequent work [21, 8, 56, 54] resolved other questions in communication complexity and circuit complexity related to constant-depth circuits by generalizing the threshold degree method of [45, 47].

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 [28] used this threshold degree approach to give what is currently the fastest algorithm 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 [38] for formulas of constant depth and by Ambainis et al. [4] 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 $\mathbf{AC^0}$. 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 the algorithmic and complexity-theoretic applications discussed above [28, 38, 29, 42, 15], 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 [35] 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 [28] 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 [38], who constructed circuits of depth k with threshold degree $\Omega(n^{1/3}\log^{2(k-2)/3} n)$. The authors of [38] formally posed the problem of obtaining a polynomial improvement on Minsky and Papert's lower bound. Such an improvement was obtained in [53], with a threshold degree lower bound of

Depth	**Threshold degree**	**Reference**
2	$\Omega(n^{1/3})$	Minsky and Papert [35]
2	$O(n^{1/3}\log n)$	Klivans and Servedio [28]
k	$\Omega(n^{1/3}\log^{\frac{2(k-2)}{3}} n)$	O'Donnell and Servedio [38]
k	$\Omega(n^{\frac{k-1}{2k-1}})$	Sherstov [53]
4	$\Omega(\sqrt{n})$	Sherstov [55]
3	$\tilde{\Omega}(\sqrt{n})$	Bun and Thaler [19]
k	$\tilde{\Omega}(n^{\frac{k-1}{k+1}})$	This paper

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.

$\Omega(n^{(k-1)/(2k-1)})$ for circuits of depth k. A polynomially stronger result was obtained in [55], with a lower bound of $\Omega(\sqrt{n})$ on the threshold degree of an explicit circuit of depth 4. Bun and Thaler [19] recently used a different, depth-3 circuit to give a much simpler proof of an $\tilde{\Omega}(\sqrt{n})$ 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. *Let $k \geqslant 1$ be a fixed integer. Then there is an (explicitly given) Boolean circuit family $\{f_n\}_{n=1}^{\infty}$, where $f_n \colon \{0,1\}^n \to \{0,1\}$ 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 [35] and Bun and Thaler [19] 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.

1.2. Sign-rank of AC^0. The *sign-rank* of a matrix $A = [A_{ij}]$ without zero entries 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

Depth	**Sign-rank**	**Reference**
3	$\exp(\Omega(n^{1/3}))$	Razborov and Sherstov [42]
3	$\exp(\Omega(n^{2/5}))$	Bun and Thaler [17]
7	$\exp(\tilde{\Omega}(\sqrt{n}))$	Bun and Thaler [19]
$3k$	$\exp(\tilde{\Omega}(n^{1-\frac{1}{k+1}}))$	This paper
$3k+1$	$\exp(\tilde{\Omega}(n^{1-\frac{1}{k+1.5}}))$	This paper

Table 2: Known lower bounds on the maximum sign-rank 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.

learning theory [41, 2, 11, 23, 24, 28, 34, 44, 48, 42, 17, 19]. 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 a sign-representation out of *every* linear subspace of given dimension, whereas a threshold degree lower bound rules out a 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 available for any explicit matrices until the breakthrough work of Forster [23], 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 [42], solving a 22-year-old problem due to Babai, Frankl, and Simon [6]. The authors of [42] constructed a polynomial-size circuit of depth 3 with sign-rank $\exp(\Omega(n^{1/3}))$. In follow-up work, Bun and Thaler [17, 19] 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 [17, 19], 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. *Let $k \geqslant 1$ be a given 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 $3k$, and sign-rank*

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

As a companion result, we prove the following qualitatively similar but quantitatively incomparable theorem.

THEOREM 1.3. *Let $k \geqslant 1$ be a given integer. 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 fixed depth k, 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* [44, 46, 42, 19], namely, $\exp(\Omega(n^{1-\epsilon}))$ for any constant $\epsilon > 0$. This rules out the possibility of learning AC^0 circuits via dimension complexity [42], 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 [32], 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:

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

and

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

The former quantity, introduced by Paturi and Simon [41], 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. [6], it features an additional penalty term that depends on the error probability. It is clear that

$$1 \leqslant \mathrm{UPP}(F) \leqslant \mathrm{PP}(F) \leqslant n+2$$

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 [12, 44]. These two models occupy a special place in the study of communication because they are more powerful than almost 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 currently 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* [45, 47] 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

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

Discrepancy is a combinatorial complexity measure of interest in communication complexity theory and other research areas; see Section 2.8 for a formal definition. As k grows, the bounds of Theorem 1.4 approach the best possible bounds for any communication problem $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. The same *qualitative* behavior was achieved in previous work by Bun and Thaler [19], 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 $\mathrm{PP}(F) = \Omega(n^{1-\epsilon})$ and discrepancy $\operatorname{disc}(F) = \exp(-\Omega(n^{1-\epsilon}))$. Theorem 1.4 strictly subsumes the result of Bun and Thaler [19] and all other prior work on the discrepancy and PP-complexity of constant-depth circuits [45, 47, 8, 56, 54]. For any fixed depth $k \geqslant 4$, the bounds of Theorem 1.4 are a polynomial improvement in n over all previous work. We further obtain a counterpart of Theorem 1.4 for *number-on-the-forehead model*, the strongest formalism of multiparty communication. This result, presented in detail in Section 4.5, uses the multiparty version [54] 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 [41] 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 work on the unbounded-error communication complexity of AC^0, with a polynomial improvement for any depth $k \geqslant 3$. 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 [19]. 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. [6] 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 k and has the additional advantage of generalizing to the multiparty setting.

1.4. Threshold weight and threshold density. By well-known reductions, Theorem 1.1 implies a number of other lower bounds for the representation of

AC^0 circuits by polynomials. For the sake of completeness, we mention two such consequences. The *threshold density* of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$, denoted $\mathrm{dns}(f)$, is the minimum size of a set family $\mathscr{S} \subseteq \mathscr{P}(\{1,2,\ldots,n\})$ such that

$$\operatorname{sgn}\left(\sum_{S\in\mathscr{S}} \lambda_S(-1)^{\sum_{i\in S} x_i}\right) \equiv (-1)^{f(x)}$$

for some reals λ_S. A related complexity measure is *threshold weight*, denoted $W(f)$ and defined as the minimum sum $\sum_{S\subseteq\{1,2,\ldots,n\}} |\lambda_S|$ over all integers λ_S such that

$$\operatorname{sgn}\left(\sum_{S\subseteq\{1,2,\ldots,n\}} \lambda_S(-1)^{\sum_{i\in S} x_i}\right) \equiv (-1)^{f(x)}.$$

It is not hard to see that the threshold density and threshold weight of f correspond to the minimum size of a threshold-of-parity and majority-of-parity circuit for f, respectively. The definitions imply that $\mathrm{dns}(f) \leqslant W(f)$ for every f, and a little more thought reveals that $1 \leqslant \mathrm{dns}(f) \leqslant 2^n$ and $1 \leqslant W(f) \leqslant (2\sqrt{2})^n$. These complexity measures have seen extensive work, motivated by applications to computational learning and circuit complexity. For a bibliographic overview, we refer the reader to [53, Section 8.2].

Krause and Pudlák [30, Proposition 2.1] gave an ingenious method for transforming threshold degree lower bounds into lower bounds on threshold density and thus also threshold weight. Specifically, let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function of interest. The authors of [30] considered the related function $F\colon (\{0,1\}^n)^3 \to \{0,1\}$ given by $F(x,y,z) = f(\ldots, (\overline{z_i} \wedge x_i) \vee (z_i \wedge y_i), \ldots)$, and proved that $\mathrm{dns}(F) \geqslant 2^{\deg_\pm(f)}$. In this light, Theorem 1.1 implies that the threshold density of AC^0 is $\exp(\Omega(n^{1-\epsilon}))$ for any constant $\epsilon > 0$.

Corollary 1.6. *Let $k \geqslant 3$ be a fixed integer. Then there is an (explicitly given) Boolean circuit family $\{F_n\}_{n=1}^\infty$, where $F_n\colon \{0,1\}^n \to \{0,1\}$ has polynomial size and depth k and satisfies*

$$\begin{aligned} W(F_n) &\geqslant \mathrm{dns}(F_n) \\ &= \exp\left(\Omega\left(n^{\frac{k-1}{k+1}} \cdot (\log n)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right)\right). \end{aligned}$$

Observe that the circuit family $\{F_n\}_{n=1}^\infty$ of Corollary 1.6 has the same depth as the circuit family $\{f_n\}_{n=1}^\infty$ of Theorem 1.1. This is because f_n has bottom fan-in $O(\log n)$, and thus the Krause-Pudlák transformation $f_n \mapsto F_n$ can be "absorbed" into the bottom two levels of f_n. Corollary 1.6 subsumes all previous lower bounds [30, 15, 53, 55, 19] on the threshold weight and density of AC^0, with a polynomial improvement for every $k \geqslant 4$. The improvement is particularly noteworthy in the case of threshold density, where the best previous lower bound [55, 19] was $\exp(\Omega(\sqrt{n}))$.

1.5. 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 [36]. 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 the 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:

$$\deg_\pm(f) = \lim_{\epsilon \nearrow 1} \deg_\epsilon(f). \tag{1.1}$$

Both approximate degree and threshold degree have dual characterizations [47], 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 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 (1.1), this discussion carries over to the case of 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 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 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 [39].

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 Boolean 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 [52] 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 was the basis of much previous work, including the best previous lower bounds [53, 55] 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 [50] and direct product theorems [49] for approximate degree and error amplification for approximate degree [49, 15, 59, 16].

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 ϕ for the function of interest. Such algorithmic proofs run the gamut in terms of technical sophistication, from straightforward to lengthy and 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* [50, 33]:

$$\Phi(x_1,x_2,\dots,x_n) = \phi(\operatorname{sgn}\psi(x_1),\dots,\operatorname{sgn}\psi(x_n)) \prod_{i=1}^{n} |\psi(x_i)|. \tag{1.2}$$

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 [42], 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., [17, 18, 13, 19].

Hardness amplification for approximate degree. While block-composition has produced a treasure trove of results on the polynomial representation of Boolean functions, it is of limited 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 [51] 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 no harder to approximate to bounded error than either of 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 any future *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 [18], who proved, for any constant $\epsilon > 0$, an $\Omega(n^{1-\epsilon})$ lower bound on the 1/3-approximate degree of AC^0. In more detail, let $f\colon \{-1,+1\}^n \to \{-1,+1\}$ be a function of interest, 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 [50, 15] on approximate degree, dual block-composition witnesses the lower bound $\deg_{1/3}(F) = \Omega(\deg_{1/3}(\mathrm{OR}_m)\deg_{1/3}(f)) = \Omega(\sqrt{m}\deg_{1/3}(f))$. Here, 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 (1.2) implies that the dual object for F places most of it ℓ_1 mass on inputs of Hamming weight $O(n \log n)$. The authors of [18] 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 $n(\log n)^{O(1)}$ 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 [18] 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 [13] refined the technique of [18] by deriving optimal concentration bounds for the dual object for OR_m. They thereby obtained 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 [19], 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 [18] with Sherstov's work [49] on direct product theorems for approximate degree. This new result substantially strengthens the authors' previous result [18] on the *bounded-error* approximate degree of AC^0 but falls short of a threshold degree lower bound.

1.6. Our approach.

Threshold degree of AC^0. Bun and Thaler [19] 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 [19], have no implications for threshold degree. For example, there are functions [52] 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 [19] on the large-error approximate degree of AC^0 and instead builds on the earlier and simpler "block-composition followed by

input compression" approach of [18]. 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 $0 \leqslant \alpha < 1$ is 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 [18] 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 [18, 13, 19], we would like to use the Razborov–Sherstov corrector object to *remove* the ℓ_1 mass that Φ has on inputs on 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 [18, 13, 19] transferred the ℓ_1 mass from inputs of high Hamming weight to the neighborhood of the all-zeroes input $(0, 0, \ldots, 0)$. An unavoidable downside 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 various points, destroying Φ's sign-agreement with the composed function $f \circ \mathrm{MP}_m$. It is this difficulty that prevented earlier works [18, 13, 19] from obtaining a strong threshold degree lower bound for AC^0.

We proceed differently. Instead of transferring the ℓ_1 mass of Φ from inputs of high Hamming weight to the neighborhood of $(0, 0, \ldots, 0)$, we transfer it simultaneously to *exponentially many* neighborhoods of inputs with low Hamming weight. 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 calculations; it is in fact surprising that such a scheme exists. Once the transfer process is complete, we obtain a witness for the $\Omega(mn^\alpha)$ threshold degree of $f \circ \mathrm{MP}_m$ even for the restriction of the domain to inputs of low Hamming weight. Compressing the input as in [18, 13], 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 transform a threshold degree lower bound in a black-box manner into a sign-rank lower bound. In particular, Theorem 1.1 has no implications a priori for the sign-rank of AC^0. Instead, our proofs of Theorems 1.2 and 1.3 are 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 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.

The 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 transformed 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 [45, 47] and Forster's spectral lower bound on sign-rank [23, 24]. Accordingly, we obtain Theorems 1.2 and 1.3 by proving an $\Omega(n^{1-\epsilon})$ lower bound on the $2^{-n^{1-\epsilon}}$-smooth threshold degree of AC^0, for any constant $\epsilon > 0$. At the core of this 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.

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 unrelated to our proof of Theorem 1.1.

Central to our work is an analytic property that 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 block-composition step here 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 is quite different from previous work on the sign-rank of constant-depth circuits [42, 17, 19]. 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 [42, 17, 19]. 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 [42] with a much more complicated approach. The new proof can be found in Section 5.1, 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. The following well-known bound [26, Proposition 1.4] is used in our proofs without further mention:

$$\sum_{i=0}^{k} \binom{n}{i} \leqslant \Big(\frac{\mathrm{e}n}{k}\Big)^k, \qquad k = 0,1,2,\ldots,n, \tag{2.1}$$

where $\mathrm{e} = 2.7182\ldots$ denotes Euler's number.

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 $e_1, e_2, \ldots, e_n$ correspond to 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. Componentwise composition generalizes in the natural manner to partial Boolean functions $f\colon \{0,1\}^n \to \{0,1,*\}$ and $g\colon X \to \{0,1,*\}$, as follows:

$$(f \circ g)(x_1, \ldots, x_n) = \begin{cases} f(g(x_1), \ldots, g(x_n)) & \text{if } x_1, \ldots, x_n \in g^{-1}(0 \cup 1), \\ * & \text{otherwise.} \end{cases}$$

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 $x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n$ 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 output gate. 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 each $f_n\colon \{0,1\}^n \to \{0,1\}$ is computed 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 and inner product:

$$\begin{aligned}\|f\|_\infty &= \max_{x \in \operatorname{supp} f} |f(x)|,\\ \|f\|_1 &= \sum_{x \in \operatorname{supp} f} |f(x)|,\\ \langle f, g\rangle &= \sum_{x \in \operatorname{supp} f \cap \operatorname{supp} g} f(x)g(x).\end{aligned}$$

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 $f\colon X \to \mathbb{R}$, we define $f^{\otimes S}\colon X^n \to \mathbb{R}$ 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 $F^{\otimes n} = \{f_1 \otimes f_2 \otimes \cdots \otimes f_n : f_1, f_2, \ldots, f_n \in F\}$. A *conical combination* of $f_1, f_2, \ldots, f_k \in \mathbb{R}^X$ is any function of the form $\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_k f_k$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are nonnegative reals. A *convex combination* of $f_1, f_2, \ldots, f_k \in \mathbb{R}^X$ is any function $\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_k f_k$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are nonnegative reals that sum to 1. The *conical hull* of $F \subseteq \mathbb{R}^X$, denoted $\operatorname{cone} F$, is the set of all conical combinations of functions in F. The *convex hull*, denoted $\operatorname{conv} F$, is defined analogously as the set of all convex combinations of functions in F. For any set of functions $F \subseteq \mathbb{R}^X$, we have

$$(\operatorname{conv} F)^{\otimes n} \subseteq \operatorname{conv}(F^{\otimes n}). \tag{2.2}$$

Throughout this manuscript, we view probability distributions as real functions. This convention makes available the shorthands introduced above. In particular, for probability distributions μ and λ, the symbol $\operatorname{supp} \mu$ denotes the support of μ, and $\mu \otimes \lambda$ denotes the probability distribution given by $(\mu \otimes \lambda)(x, y) = \mu(x)\lambda(y)$. If μ is a probability distribution on X, we consider μ to be defined also on any superset of X with the understanding that $\mu = 0$ outside X. We let $\mathfrak{D}(X)$ denote the family of all finitely supported probability distributions on X. Most of this paper is concerned with the distribution family $\mathfrak{D}(\mathbb{N}^n)$ and its subfamilies, each of which we denote with a Fraktur letter.

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 $f\colon X \to \mathbb{R}$ on a subset $X \subseteq \mathbb{N}^n$, we let $f|_W$ denote the restriction of f to $X|_W$. A typical use of this notation would be $f|_{\leqslant w}$ for some real number w.

2.4. Orthogonal content. For a multivariate real polynomial $p\colon \mathbb{R}^n \to \mathbb{R}$, 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 $\operatorname{orth}\phi = 0 \;\Leftrightarrow\; \langle \phi, 1\rangle \neq 0$ and $\operatorname{orth}\phi = \infty \;\Leftrightarrow\; \phi = 0$. Our next three results record 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* $\phi, \psi\colon X \to \mathbb{R}$;
(ii) $\operatorname{orth}(\phi \otimes \psi) = \operatorname{orth}(\phi) + \operatorname{orth}(\psi)$ *for all* $\phi\colon X \to \mathbb{R}$ *and* $\psi\colon Y \to \mathbb{R}$;
(iii) $\operatorname{orth}(\phi^{\otimes n} - \psi^{\otimes n}) \geqslant \operatorname{orth}(\phi - \psi)$ *for all* $\phi, \psi\colon X \to \mathbb{R}$ *and all* $n \geqslant 1$.

Proof. Item (i) is immediate, as is the upper bound in (ii). For the lower bound in (ii), simply note that the linearity of inner product makes it possible to restrict attention to factored polynomials $p(x)q(y)$, where p and q are polynomials on X and Y, respectively. For (iii), use a telescoping sum to write

$$\begin{aligned}\phi^{\otimes n} - \psi^{\otimes n} &= \sum_{i=0}^{n-1} (\phi^{\otimes(n-i)} \otimes \psi^{\otimes i} - \phi^{\otimes(n-i-1)} \otimes \psi^{\otimes(i+1)}) \\ &= \sum_{i=0}^{n-1} \phi^{\otimes(n-i-1)} \otimes (\phi - \psi) \otimes \psi^{\otimes i}.\end{aligned}$$

By (ii), each term in the final expression has orthogonal content at least $\operatorname{orth}(\phi-\psi)$. By (i), then, the sum has orthogonal content at least $\operatorname{orth}(\phi - \psi)$ as well. □

PROPOSITION 2.2. *Let $\phi_0, \phi_1\colon X \to \mathbb{R}$ be given functions on a finite subset X of Euclidean space. Then for every polynomial $p\colon X^n \to \mathbb{R}$, the mapping $z \mapsto \langle \bigotimes_{i=1}^n \phi_{z_i}, p\rangle$ is a polynomial on $\{0,1\}^n$ of degree at most $(\deg p)/\operatorname{orth}(\phi_1 - \phi_0)$.*

Proof. We may assume that $\operatorname{orth}(\phi_1 - \phi_0) > 0$ since the proposition holds trivially otherwise. By linearity, it suffices to consider factored polynomials $p(x_1, \ldots, x_n) = \prod_{i=1}^n p_i(x_i)$, where each p_i is a nonzero polynomial on X. In this setting, we have

$$\left\langle \bigotimes_{i=1}^n \phi_{z_i}, p \right\rangle = \prod_{i=1}^n \langle \phi_{z_i}, p_i \rangle . \tag{2.3}$$

By definition, $\langle \phi_0, p_i\rangle = \langle \phi_1, p_i\rangle$ for any index i with $\deg p_i < \operatorname{orth}(\phi_1 - \phi_0)$. As a result, such indices do not contribute to the degree of the right-hand side of (2.3) as a function of z. The contribution of any other index to the degree is clearly at most 1. Summarizing, the right-hand side of (2.3) is a polynomial in $z \in \{0,1\}^n$ of degree at most $|\{i : \deg p_i \geqslant \operatorname{orth}(\phi_1 - \phi_0)\}| \leqslant (\deg p)/\operatorname{orth}(\phi_1 - \phi_0)$. □

COROLLARY 2.3. *Let X be a finite subset of Euclidean space. Then for any functions $\phi_0, \phi_1 \colon X \to \mathbb{R}$ and $\psi \colon \{0,1\}^n \to \mathbb{R}$,*

$$\operatorname{orth}\left(\sum_{z\in\{0,1\}^n} \psi(z) \bigotimes_{i=1}^{n} \phi_{z_i}\right) \geqslant \operatorname{orth}(\psi) \cdot \operatorname{orth}(\phi_1 - \phi_0).$$

Proof. We may assume that $\operatorname{orth}(\psi)\cdot\operatorname{orth}(\phi_1-\phi_0) > 0$ since the claim holds trivially otherwise. Fix a polynomial any polynomial P of degree less than $\operatorname{orth}(\psi)\cdot\operatorname{orth}(\phi_1-\phi_0)$. The linearity of inner product leads to

$$\left\langle \sum_{z\in\{0,1\}^n} \psi(z) \bigotimes_{i=1}^{n} \phi_{z_i}, P \right\rangle = \sum_{z\in\{0,1\}^n} \psi(z) \left\langle \bigotimes_{i=1}^{n} \phi_{z_i}, P \right\rangle.$$

By Proposition 2.2, the right-hand side is the inner product of ψ with a polynomial of degree less than $\operatorname{orth}\psi$ and is therefore zero. □

Observe that Corollary 2.3 gives an alternate proof of Proposition 2.1(iii). Our next proposition uses orthogonal content to give a useful criterion for a real-valued function to be a probability distribution.

PROPOSITION 2.4. *Let Λ be a probability distribution on a finite subset X of Euclidean space. Let $\tilde{\Lambda} \colon X \to \mathbb{R}$ be given with $\tilde{\Lambda} \geqslant 0$ and $\operatorname{orth}(\Lambda - \tilde{\Lambda}) > 0$. Then $\tilde{\Lambda}$ is a probability distribution on X.*

Proof. By hypothesis, $\tilde{\Lambda}$ is a nonnegative function. Moreover, $\|\tilde{\Lambda}\|_1 = \langle \tilde{\Lambda}, 1\rangle = \langle \Lambda, 1\rangle - \langle \Lambda - \tilde{\Lambda}, 1\rangle = \langle \Lambda, 1\rangle = 1$, where the third step uses $\operatorname{orth}(\Lambda - \tilde{\Lambda}) > 0$. □

2.5. Sign-representation. Let $f \colon X \to \{0,1\}$ 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 $x \in X$. The term "threshold degree" appears to be due to Saks [43]. Equivalent terms in the literature include "strong degree" [5], "voting polynomial degree" [30], "polynomial threshold function degree" [38], and "sign degree" [12]. 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 [35].

THEOREM 2.5 (Minsky and Papert). $\deg_{\pm}(\mathrm{MP}_m) = \Omega(m)$.

Three new proofs of this lower bound, unrelated to Minsky and Papert's original proof, were discovered recently in [53]. Threshold degree admits the following dual characterization, obtained by appeal to linear programming duality.

FACT 2.6. *Let $f \colon X \to \{0,1\}$ be a given Boolean function on a finite subset X of Euclidean space. Then $\deg_{\pm}(f) \geqslant d$ if and only if there exists $\psi \colon X \to \mathbb{R}$ such that*

$$\begin{aligned} &(-1)^{f(x)}\psi(x) \geqslant 0, && x \in X,\\ &\operatorname{orth}\psi \geqslant d,\\ &\psi \not\equiv 0. \end{aligned}$$

The function ψ acts as a *witness* for the threshold degree of f, and is called a *dual polynomial* due to its origin in a dual linear program. We refer the reader to [5, 38, 50] for a proof of Fact 2.6. The following equivalent statement is occasionally more convenient to work with.

FACT 2.7. *For every Boolean function $f\colon X \to \{0,1\}$ on a finite subset X of Euclidean space,*

$$\deg_{\pm}(f) = \max_{\mu \in \mathfrak{D}(X)} \operatorname{orth}((-1)^f \cdot \mu). \tag{2.4}$$

We now define a generalization of threshold degree inspired by the dual view in Fact 2.7. For a function $f\colon X \to \{0,1\}$ and a real number $0 \leqslant \gamma \leqslant 1$, let

$$\deg_{\pm}(f,\gamma) = \max_{\substack{\mu \in \mathfrak{D}(X)\colon \\ \mu \geqslant \gamma/|X| \text{ on } X}} \operatorname{orth}((-1)^f \cdot \mu). \tag{2.5}$$

We call this quantity the *γ-smooth threshold degree of f*, in reference to the fact that the maximization in (2.5) is over probability distributions μ that place on every point of the domain at least a γ fraction of the weight the point would receive under the uniform distribution. A glance at (2.4) and (2.5) reveals that $\deg_{\pm}(f,\gamma)$ is monotonically nonincreasing in γ, with the limiting case $\deg_{\pm}(f,0) = \deg_{\pm}(f)$.

FACT 2.8. *For every nonconstant function $f\colon X \to \{0,1\}$,*

$$\deg_{\pm}\left(f, \frac{1}{2}\right) \geqslant 1.$$

Proof. Define $\mu = \frac{1}{2}\mu_0 + \frac{1}{2}\mu_1$, where μ_i be the uniform probability distribution on $f^{-1}(i)$. Then clearly $\operatorname{orth}((-1)^f \cdot \mu) \geqslant 1$ and $\mu \geqslant \frac{1}{2}\max\{\mu_0, \mu_1\} \geqslant \frac{1}{2|X|}$ on X. □

2.6. Symmetrization. Let S_n denote the symmetric group on n elements. For a permutation $\sigma \in S_n$ and an arbitrary sequence $x = (x_1, x_2, \ldots, x_n)$, we adopt the shorthand $\sigma x = (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$. A function $f(x_1, x_2, \ldots, x_n)$ is called *symmetric* if it is invariant under permutation of the input variables: $f(x_1, x_2, \ldots, x_n) = f(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$ for all x and σ. Symmetric functions on $\{0,1\}^n$ are intimately related to univariate polynomials, as was first observed by Minsky and Papert in their *symmetrization argument* [35].

PROPOSITION 2.9 (Minsky and Papert). *Let $p\colon \mathbb{R}^n \to \mathbb{R}$ be a given polynomial. Then the mapping*

$$t \mapsto \mathop{\mathbf{E}}_{x \in \{0,1\}^n|_t} p(x)$$

is a univariate polynomial on $\{0, 1, 2, \ldots, n\}$ of degree at most $\deg p$.

Minsky and Papert's result generalizes to block-symmetric functions:

PROPOSITION 2.10. *Let $n_1, \dots, n_k$ be positive integers. Let $p\colon \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k} \to \mathbb{R}$ be a given polynomial. Then the mapping*

$$(t_1, t_2, \dots, t_k) \mapsto \mathop{\mathbf{E}}_{x_1 \in \{0,1\}^{n_1}|_{t_1}} \mathop{\mathbf{E}}_{x_2 \in \{0,1\}^{n_2}|_{t_2}} \cdots \mathop{\mathbf{E}}_{x_k \in \{0,1\}^{n_k}|_{t_k}} p(x_1, x_2, \dots, x_k)$$

is a polynomial on $\{0, 1, \dots, n_1\} \times \{0, 1, \dots, n_2\} \times \cdots \times \{0, 1, \dots, n_k\}$ of degree at most $\deg p$.

Proposition 2.10 follows in a straightforward manner from Proposition 2.9 by induction on the number of blocks k, as pointed out in [42, Proposition 2.3]. The next result is yet another generalization of Minsky and Papert's symmetrization technique, this time to the setting when $x_1, x_2, \dots, x_n$ are vectors rather than bits.

PROPOSITION 2.11. *Let $p\colon (\mathbb{R}^m)^n \to \mathbb{R}$ be a polynomial of degree d. Then there is a polynomial $p^*\colon \mathbb{R}^n \to \mathbb{R}$ of degree at most d such that for all $x_1, x_2, \dots, x_n \in \{e_1, e_2, \dots, e_m, 0^m\}$,*

$$\mathop{\mathbf{E}}_{\sigma \in S_n} p(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = p^*(x_1 + x_2 + \cdots + x_n).$$

Proof. We closely follow an argument due to Ambainis [3, Lemma 3.4], who proved a related result. Since the components of $x_1, x_2, \dots, x_n$ are Boolean-valued, we have $x_{i,j} = x_{i,j}^2 = x_{i,j}^3 = \cdots$ and therefore we may assume that p is multilinear. By linearity, it further suffices to consider the case when p is a single monomial:

$$p(x_1, x_2, \dots, x_n) = \prod_{j=1}^{m} \prod_{i \in S_j} x_{i,j} \tag{2.6}$$

for some sets $S_1, S_2, \dots, S_m \subseteq \{1, 2, \dots, n\}$ with $\sum_{j=1}^m |S_j| \leqslant d$. If some pair of sets $S_j, S_{j'}$ with $j \neq j'$ have nonempty intersection, then the right-hand side of (2.6) contains a product of the form $x_{i,j}x_{i,j'}$ for some i and thus $p \equiv 0$ on the domain in question. As a result, the proposition holds with $p^* = 0$. In the complementary case when $S_1, S_2, \dots, S_m$ are pairwise disjoint, we calculate

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\sigma \in S_n} p(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) \\
&\quad = \prod_{j=1}^{m} \mathop{\mathbf{E}}_{\sigma \in S_n} \left[\prod_{i \in S_j} x_{\sigma(i),j} \,\middle|\, \prod_{i \in S_{j'}} x_{\sigma(i),j'} = 1 \text{ for all } j' < j \right] \\
&\quad = \prod_{j=1}^{m} \binom{x_{1,j} + x_{2,j} + \cdots + x_{n,j}}{|S_j|} \binom{n - |S_1| - |S_2| - \cdots - |S_{j-1}|}{|S_j|}^{-1}.
\end{aligned}$$

Expanding out the binomial coefficients shows that the final expression is an m-variate polynomial whose argument is the vector sum $x_1 + x_2 + \cdots + x_n \in \mathbb{R}^m$. Moreover, the degree of this polynomial is $\sum |S_j| \leqslant d$. ∎

COROLLARY 2.12. *Let $p\colon (\mathbb{R}^m)^n \to \mathbb{R}$ be a polynomial of degree d. Then the mapping*

$$v \mapsto \mathop{\mathbf{E}}_{\substack{x \in \{0^m, e_1, e_2, \ldots, e_m\}^n \colon \\ x_1 + x_2 + \cdots + x_n = v}} p \tag{2.7}$$

is a polynomial on $\mathbb{N}^m|_{\leqslant n}$ of degree at most $\deg p$.

Minsky and Papert's symmetrization corresponds to $m = 1$ in Corollary 2.12.

Proof of Corollary 2.12. Let $v \in \mathbb{N}^m|_{\leqslant n}$ be given. Then all representations $v = x_1 + x_2 + \cdots + x_n$ with $x_1, x_2, \ldots, x_n \in \{0^m, e_1, e_2, \ldots, e_m\}$ are the same up to the order of the summands. As a result, (2.7) is the same mapping as

$$v \mapsto \mathop{\mathbf{E}}_{\sigma \in S_n} p(\sigma(\underbrace{e_1, \ldots, e_1}_{v_1}, \underbrace{e_2, \ldots, e_2}_{v_2}, \ldots, \underbrace{e_m, \ldots, e_m}_{v_m}, \underbrace{0^m, 0^m \ldots, 0^m}_{n - v_1 - \cdots - v_m})),$$

which by Proposition 2.11 is a polynomial in

$$\underbrace{e_1 + \cdots + e_1}_{v_1} + \underbrace{e_2 + \cdots + e_2}_{v_2} + \cdots + \underbrace{e_m + \cdots + e_m}_{v_m} + \underbrace{0^m + \cdots + 0^m}_{n - v_1 - \cdots - v_m} = v$$

of degree at most $\deg p$. □

Analogous to symmetrized polynomials, it will be also helplful to work with symmetrized versions of Boolean functions. We define $\mathrm{AND}_n^*, \mathrm{OR}_n^* \colon \{0, 1, 2, \ldots, n\} \to \{0, 1\}$ by

$$\mathrm{AND}_n^*(t) = \begin{cases} 1 & \text{if } t = n, \\ 0 & \text{otherwise,} \end{cases} \qquad \mathrm{OR}_n^*(t) = \begin{cases} 0 & \text{if } t = 0, \\ 1 & \text{otherwise.} \end{cases}$$

The symmetrized variant of the Minsky–Papert function is $\mathrm{MP}_{m,r}^* = \mathrm{AND}_m \circ \mathrm{OR}_r^*$.

2.7. Communication complexity. An excellent reference on communication complexity is the monograph by Kushilevitz and Nisan [32]. In this overview, we will limit ourselves to key definitions and notation. We adopt the standard randomized model of multiparty communication, due to Chandra et al. [20]. The model features ℓ communicating players, tasked with computing a Boolean function $F\colon X_1 \times X_2 \times \cdots \times X_\ell \to \{0, 1\}$ for some finite sets $X_1, X_2, \ldots, X_\ell$. A given input $(x_1, x_2, \ldots, x_\ell) \in X_1 \times X_2 \times \cdots \times X_\ell$ is distributed among the players by placing x_i, figuratively speaking, on the forehead of the ith player (for $i = 1, 2, \ldots, \ell$). In other words, the ith player knows the arguments $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_\ell$ but not x_i. The players communicate by sending broadcast messages, taking turns according to a protocol agreed upon in advance. Each of them privately holds an unlimited supply of uniformly random bits, which he can use along with his available arguments when deciding what message to send at any given point in the protocol. The players' objective to compute $F(x_1, x_2, \ldots, x_\ell)$. An *$\epsilon$-error protocol* for F is one which, on every input $(x_1, x_2, \ldots, x_\ell)$, produces the correct answer $F(x_1, x_2, \ldots, x_\ell)$ with probability at least $1 - \epsilon$. The *cost* of a protocol is the total bit length of the

messages broadcast by all the players in the worst case.[1] The *ϵ-error randomized communication complexity* of F, denoted $R_\epsilon(F)$, is the least cost of an ϵ-error randomized protocol for F. As a special case of this model for $\ell = 2$, one recovers the original two-party model of Yao [61] reviewed in the introduction.

Our work focuses on randomized protocols with error probability close to that of random guessing, $1/2$. There are two natural ways to define the communication complexity of a multiparty problem F in this setting. The *communication complexity of F with unbounded error*, introduced by Paturi and Simon [41], is the quantity

$$\mathrm{UPP}(F) = \min_{0<\epsilon<1/2} R_\epsilon(F). \tag{2.8}$$

Here, the error is unbounded in the sense that it can be arbitrarily close to $1/2$. Babai et al. [6] proposed an alternate quantity, which includes an additive penalty term that depends on the error probability:

$$\mathrm{PP}(F) = \min_{0<\epsilon<1/2} \left\{ R_\epsilon(F) + \log \frac{1}{\frac{1}{2} - \epsilon} \right\}. \tag{2.9}$$

This quantity is known as the *communication complexity of F with weakly unbounded error*.

2.8. Discrepancy and sign-rank. An *ℓ-dimensional cylinder intersection* is a function $\chi\colon X_1 \times X_2 \times \cdots \times X_\ell \to \{0,1\}$ of the form

$$\chi(x_1, x_2, \ldots, x_\ell) = \prod_{i=1}^{\ell} \chi_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_\ell),$$

where $\chi_i\colon X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_\ell \to \{0,1\}$. In other words, an ℓ-dimensional cylinder intersection is the product of ℓ functions with range $\{0,1\}$, where the ith function does not depend on the ith coordinate but may depend arbitrarily on the other $\ell - 1$ coordinates. Introduced by Babai et al. [7], cylinder intersections are the fundamental building blocks of communication protocols and for that reason play a central role in the theory. For a Boolean function $F\colon X_1 \times X_2 \times \cdots \times X_\ell \to \{0,1\}$ and a probability distribution P on $X_1 \times X_2 \times \cdots \times X_\ell$, the *discrepancy of F with respect to P* is given by

$$\mathrm{disc}_P(F) = \max_{\chi} \left| \sum_{x \in X_1 \times X_2 \times \cdots \times X_\ell} (-1)^{F(x)} P(x) \chi(x) \right|,$$

where the maximum is over cylinder intersections χ. The minimum discrepancy over all distributions is denoted

$$\mathrm{disc}(F) = \min_P \mathrm{disc}_P(F).$$

The *discrepancy method* [22, 7, 32] is a classic technique that bounds randomized communication complexity from below in terms of discrepancy.

[1] The contribution of a b-bit broadcast to the protocol cost is b rather than $\ell \cdot b$.

THEOREM 2.13 (Discrepancy method). *Let $F\colon X_1 \times X_2 \times \cdots \times X_\ell \to \{0,1\}$ be a given communication problem. Then*

$$2^{R_\epsilon(F)} \geqslant \frac{1-2\epsilon}{\operatorname{disc}(F)}.$$

Combining this theorem with the definition of $\mathrm{PP}(F)$ gives the following corollary.

COROLLARY 2.14. *Let $F\colon X_1 \times X_2 \times \cdots \times X_\ell \to \{0,1\}$ be a given communication problem. Then*

$$\mathrm{PP}(F) \geqslant \log \frac{2}{\operatorname{disc}(F)}.$$

The *sign-rank* of a real matrix $A \in \mathbb{R}^{n\times m}$ with nonzero entries is the least rank of a matrix $B \in \mathbb{R}^{n\times m}$ such that $\operatorname{sgn} A_{i,j} = \operatorname{sgn} B_{i,j}$ for all i, j. In general, the sign-rank of a matrix can be vastly smaller than its rank. For example, consider the following nonsingular matrices of order $n \geqslant 3$:

$$\begin{bmatrix} 1 & & & & & \\ & 1 & & & 1 & \\ & & 1 & & & \\ & & & \ddots & & \\ -1 & & & & 1 & \\ & & & & & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & & & & & \\ & 1 & & & -1 & \\ & & 1 & & & \\ & & & \ddots & & \\ -1 & & & & 1 & \\ & & & & & 1 \end{bmatrix}.$$

These matrices have sign-rank at most 2 and 3, respectively. Indeed, the first matrix has the same sign pattern as $[2(j-i)+1]_{i,j}$. The second has the same sign pattern as $[\langle v_i, v_j\rangle - (1-\epsilon)]_{i,j}$, where $v_1, v_2, \ldots, v_n \in \mathbb{R}^2$ are arbitrary pairwise distinct unit vectors and ϵ is a suitably small positive real, cf. [41, Section 5]. As a matter of notational convenience, we extend the notion of sign-rank to Boolean functions $f\colon X \times Y \to \{0,1\}$ by defining $\mathrm{rk}_\pm(f) = \mathrm{rk}_\pm(M_f)$, where $M_f = [(-1)^{f(x,y)}]_{x\in X, y\in Y}$ is the matrix associated with f. A remarkable fact, due to Paturi and Simon [41], is that the sign-rank of a two-party communication problem fully characterizes its unbounded-error communication complexity.

THEOREM 2.15 (Paturi and Simon). *Let $F\colon X \times Y \to \{0,1\}$ be a given communication problem. Then*

$$\log \mathrm{rk}_\pm(F) \leqslant \mathrm{UPP}(F) \leqslant \log \mathrm{rk}_\pm(F) + 2.$$

As Corollary 2.14 and Theorem 2.15 show, the study of communication with unbounded and weakly unbounded error is in essence the study of discrepancy and sign-rank. These quantities are difficult to analyze from first principles. The *pattern matrix method*, developed in [45, 47], is a technique that transforms lower bounds for polynomial approximation into bounds on discrepancy, sign-rank, and various other quantities in communication complexity. For our discrepancy bounds, we use the following special case of the pattern matrix method [54, Theorem 5.7 and equation (119)].

THEOREM 2.16 (Sherstov). *Let $f\colon \{0,1\}^n \to \{0,1\}$ be given. Consider the ℓ-party communication problem $F\colon (\{0,1\}^{nm})^\ell \to \{0,1\}$ given by $F = f \circ \mathrm{NOR}_m \circ \mathrm{AND}_\ell$. Then*

$$\mathrm{disc}(F) \leqslant \left(\frac{c2^\ell \ell}{\sqrt{m}}\right)^{\deg_\pm(f)/2},$$

where $c > 0$ is a constant independent of n, m, ℓ, f.

We note that the case $\ell = 2$ of Theorem 2.16 is vastly easier to prove than the general statement; this two-party result can be found in [56, Theorem 7.3 and equation (7.3)]. For our sign-rank lower bounds, we use the following theorem implicit in [48].

THEOREM 2.17 (Sherstov, implicit). *Let $f\colon \{0,1\}^n \to \{0,1\}$ be given. Suppose that $\deg_\pm(f, \gamma) \geqslant d$, where γ and d are positive reals. Fix an integer $m \geqslant 2$ and define $F\colon \{0,1\}^{mn} \times \{0,1\}^{mn} \to \{0,1\}$ by $F = f \circ \mathrm{OR}_m \circ \mathrm{AND}_2$. Then*

$$\mathrm{rk}_\pm(F) \geqslant \gamma \left\lfloor \frac{m}{2} \right\rfloor^{d/2}.$$

For the reader's convenience, we give a detailed proof of Theorem 2.17 in Appendix B.

3. Auxiliary results

In this section, we collect a number of supporting results on approximate degree that have appeared in one form or another in previous work. For the reader's convenience, we provide self-contained proofs whenever the precise formulation that we need departs from published work.

3.1. 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 due to Razborov and Sherstov [42]. Informally, it is used to adjust a dual object's metric properties while preserving its orthogonality to low-degree polynomials. The lemma plays a basic role in several recent papers [42, 18, 13] as well as our work.

LEMMA 3.1 (Razborov and Sherstov). *Fix integers d and n, where $0 \leqslant d < n$. Then there is an (explicitly given) function $\zeta\colon \{0,1\}^n \to \mathbb{R}$ such that*

$$\begin{aligned}
&\operatorname{supp} \zeta \subseteq \{0,1\}^n|_{\leqslant d} \cup \{1^n\},\\
&\zeta(1^n) = 1,\\
&\|\zeta\|_1 \leqslant 1 + 2^d \binom{n}{d},\\
&\operatorname{orth} \zeta > d.
\end{aligned}$$

In more detail, this result corresponds to taking $k = d$ and $\zeta = (-1)^n g$ in the proof of Lemma 3.2 of [42]. We will need the following symmetrized version of Lemma 3.1.

LEMMA 3.2. *Fix a point $u \in \mathbb{N}^n$ and a natural number $d < |u|$. Then there is $\zeta_u \colon \mathbb{N}^n \to \mathbb{R}$ such that*

$$\operatorname{supp} \zeta_u \subseteq \{u\} \cup \{v \in \mathbb{N}^n : v \leqslant u \text{ and } |v| \leqslant d\}, \tag{3.1}$$
$$\zeta_u(u) = 1, \tag{3.2}$$
$$\|\zeta_u\|_1 \leqslant 1 + 2^d \binom{|u|}{d}, \tag{3.3}$$
$$\operatorname{orth} \zeta_u > d. \tag{3.4}$$

Proof. Lemma 3.1 gives a function $\zeta \colon \{0,1\}^{|u|} \to \mathbb{R}$ such that

$$\operatorname{supp} \zeta \subseteq \{0,1\}^{|u|}|_{\leqslant d} \cup \{1^{|u|}\}, \tag{3.5}$$
$$\zeta(1^{|u|}) = 1, \tag{3.6}$$
$$\|\zeta\|_1 \leqslant 1 + 2^d \binom{|u|}{d}, \tag{3.7}$$
$$\operatorname{orth} \zeta > d. \tag{3.8}$$

Now define $\zeta_u \colon \mathbb{N}^n \to \mathbb{R}$ by

$$\zeta_u(v) = \sum_{x_1 \in \{0,1\}^{|u_1|}|_{|v_1|}} \cdots \sum_{x_n \in \{0,1\}^{|u_n|}|_{|v_n|}} \zeta(x_1 \dots x_n).$$

Then (3.1)–(3.3) are immediate from (3.5)–(3.7), respectively. To verify the remaining property (3.4), fix a polynomial $p \colon \mathbb{R}^n \to \mathbb{R}$ of degree at most d. Then

$$\begin{aligned}
\langle \zeta_u, p \rangle &= \sum_{v : v \leqslant u} \left(\sum_{x_1 \in \{0,1\}^{|u_1|}|_{|v_1|}} \cdots \sum_{x_n \in \{0,1\}^{|u_n|}|_{|v_n|}} \zeta(x_1 \dots x_n) \right) p(v_1, \dots, v_n) \\
&= \sum_{v : v \leqslant u} \left(\sum_{x_1 \in \{0,1\}^{|u_1|}|_{|v_1|}} \cdots \sum_{x_n \in \{0,1\}^{|u_n|}|_{|v_n|}} \zeta(x_1 \dots x_n) p(|x_1|, \dots, |x_n|) \right) \\
&= \sum_{x_1 \in \{0,1\}^{|u_1|}} \cdots \sum_{x_n \in \{0,1\}^{|u_n|}} \zeta(x_1 \dots x_n) p(|x_1|, \dots, |x_n|) \\
&= 0,
\end{aligned}$$

where the last step uses (3.8). ∎

When constructing a dual polynomial for a complicated constant-depth circuit, it is natural to start with a dual polynomial for the OR function or, equivalently, its counterpart AND. The first such dual polynomial was constructed by Špalek [60], with many refinements and generalizations [14, 53, 55, 18, 13] obtained in follow-up work. We augment this line of work with yet another construction, which delivers the exact combination of analytic and metric properties that we need.

THEOREM 3.3. *Let $0 < \epsilon < 1$ be given. Then for some constants $c', c'' \in (0,1)$ and all integers $N \geqslant n \geqslant 1$, there is an (explicitly given) function $\psi\colon \{0,1,2,\ldots,N\} \to \mathbb{R}$ such that*

$$\begin{aligned}
&\psi(0) > \frac{1-\epsilon}{2},\\
&\|\psi\|_1 = 1,\\
&\operatorname{orth}\psi \geqslant c'\sqrt{n},\\
&\operatorname{sgn}\psi(t) = (-1)^t, && t = 0,1,2,\ldots,N,\\
&|\psi(t)| \in \left[\frac{c'}{(t+1)^2\, 2^{c''t/\sqrt{n}}},\ \frac{1}{c'(t+1)^2\, 2^{c''t/\sqrt{n}}}\right], && t = 0,1,2,\ldots,N.
\end{aligned}$$

A self-contained proof of Theorem 3.3 is available in Appendix A.

3.2. Dominant components. We now recall a lemma due to Bun and Thaler [18] that serves to identify the dominant components of a vector. Its primary use [18, 13] is to prove concentration-of-measure results for product distributions on $\mathbb{N}^n$.

LEMMA 3.4 (Bun and Thaler). *Let $v \in \mathbb{R}^n$ be given, $v \neq 0^n$. Then there is $S \subseteq \{1,2,\ldots,n\}$ such that*

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

Proof (adapted from [18]). By renumbering the indices if necessary, we may assume that $|v_1| \geqslant |v_2| \geqslant \cdots \geqslant |v_n| \geqslant 0$. For the sake of contradiction, suppose that no such set S exists. Then

$$|v_i| < \frac{1}{i} \cdot \frac{\|v\|_1}{2(1+\ln n)}$$

for every index $i \geqslant \frac{\|v\|_1}{2\|v\|_\infty}$. As a result,

$$\begin{aligned}
\|v\|_1 &= \sum_{i < \frac{\|v\|_1}{2\|v\|_\infty}} |v_i| + \sum_{i=\left\lceil \frac{\|v\|_1}{2\|v\|_\infty} \right\rceil}^{n} |v_i|\\
&\leqslant \sum_{i < \frac{\|v\|_1}{2\|v\|_\infty}} \|v\|_\infty + \sum_{i=\left\lceil \frac{\|v\|_1}{2\|v\|_\infty} \right\rceil}^{n} \frac{1}{i} \cdot \frac{\|v\|_1}{2(1+\ln n)}\\
&< \frac{\|v\|_1}{2} + \frac{\|v\|_1}{2(1+\ln n)} \sum_{i=1}^{n} \frac{1}{i}\\
&\leqslant \|v\|_1,
\end{aligned}$$

where the final step uses

$$\sum_{i=1}^{n} \frac{1}{i} = 1 + \sum_{i=2}^{n} \frac{1}{i} \leqslant 1 + \int_1^n \frac{di}{i} = 1 + \ln n.$$

We have arrived at $\|v\|_1 < \|v\|_1$, a contradiction. □

We will need a slightly more general statement, which can be thought of as an extremal analogue of Lemma 3.4.

LEMMA 3.5. *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*

$$|S| \geqslant \frac{\|v\|_1}{2\|v\|_\infty}, \tag{3.9}$$

$$\min_{i \in S} |v_i| \geqslant \frac{1}{|S|} \cdot \frac{\theta}{2(1 + \ln n)}, \tag{3.10}$$

$$\sum_{i \notin S} |v_i| < \theta. \tag{3.11}$$

Proof. Fix n, v, and θ for the remainder of the proof. We will refer to a subset $S \subseteq \{1, 2, \ldots, n\}$ as *regular* if S satisfies (3.9) and (3.10). Lemma 3.4 along with $\|v\|_1 \geqslant \theta$ ensures the existence of at least one regular set. Now, let S be a *maximal* regular set. For the sake of contradiction, suppose that (3.11) fails. Applying Lemma 3.4 to $v|_{\overline{S}}$ produces a nonempty set $T \subseteq \overline{S}$ with

$$\min_{i \in T} |v_i| \geqslant \frac{1}{|T|} \cdot \frac{\theta}{2(1 + \ln n)}.$$

But then $S \cup T$ is regular, contradicting the maximality of S. □

Lemmas 3.4 and 3.5 imply the following concentration-of-measure result for product distributions on $\mathbb{N}^n$, due to Bun and Thaler [18].

LEMMA 3.6 (Bun and Thaler). *Let $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathfrak{D}(\mathbb{N})$ be given with*

$$\lambda_i(t) \leqslant \frac{C\alpha^t}{(t+1)^2}, \qquad t \in \mathbb{N}, \tag{3.12}$$

where $C \geqslant 0$ and $0 \leqslant \alpha \leqslant 1$. Then for all $\theta \geqslant 8Ce n(1 + \ln n)$,

$$\mathop{\mathbf{P}}_{v \sim \lambda_1 \times \lambda_2 \times \cdots \times \lambda_n} [\|v\|_1 \geqslant \theta] \leqslant \alpha^{\theta/2}.$$

Proof (adapted from [18]). For a nonempty subset $S \subseteq \{1, 2, \ldots, n\}$ and a vector $v \in \mathbb{N}^n$, we say that v *is S-heavy* if the following conditions are simultaneously

satisfied:

$$|v_i| \geqslant \frac{1}{|S|} \cdot \frac{\theta}{4(1+\ln n)}, \qquad i \in S, \tag{3.13}$$

$$\sum_{i \in S} |v_i| > \frac{\theta}{2}. \tag{3.14}$$

Now, consider a random vector $v \in \mathbb{N}^n$ distributed according to $\lambda_1 \times \lambda_2 \times \cdots \times \lambda_n$. We have

$$\begin{aligned}
\mathop{\mathbf{P}}_{v}[\|v\|_1 \geqslant \theta] &\leqslant \mathop{\mathbf{P}}_{v}[v \text{ is } S\text{-heavy for some nonempty } S \neq \varnothing] \\
&\leqslant \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ S \neq \varnothing}} \mathop{\mathbf{P}}_{v}[v \text{ is } S\text{-heavy}] \\
&\leqslant \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ S \neq \varnothing}} \alpha^{\theta/2} \left(\sum_{t \geqslant \frac{1}{|S|} \cdot \frac{\theta}{4(1+\ln n)}} \frac{C}{(t+1)^2} \right)^{|S|} \\
&\leqslant \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ S \neq \varnothing}} \alpha^{\theta/2} \left(C \int_{\frac{1}{|S|} \cdot \frac{\theta}{4(1+\ln n)}}^{\infty} \frac{dt}{t^2} \right)^{|S|} \\
&= \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ S \neq \varnothing}} \alpha^{\theta/2} \left(\frac{C|S| \cdot 4(1+\ln n)}{\theta} \right)^{|S|} \\
&= \sum_{s=1}^{n} \binom{n}{s} \cdot \alpha^{\theta/2} \left(\frac{Cs \cdot 4(1+\ln n)}{\theta} \right)^{s} \\
&\leqslant \sum_{s=1}^{n} \alpha^{\theta/2} \left(\frac{\mathrm{e}n}{s} \cdot \frac{Cs \cdot 4(1+\ln n)}{\theta} \right)^{s} \\
&\leqslant \alpha^{\theta/2},
\end{aligned}$$

where the first inequality holds by Lemma 3.5; the second step applies the union bound; the third step uses $0 \leqslant \alpha \leqslant 1$ and the upper bound (3.12) for the λ_i; and the last two steps use (2.1) and the hypothesis that $\theta \geqslant 8C\mathrm{e}n(1+\ln n)$, respectively. □

3.3. Input transformation. We work almost exclusively with Boolean functions on $\mathbb{N}^n|_{\leqslant\theta}$, where the dimension parameter n is polynomially larger than the Hamming weight parameter θ. This choice of domain is admittedly unusual but greatly simplifies the analysis. Fortunately, approximation-theoretic results obtained in this setting carry over in a blackbox manner to the hypercube. In more detail, we will now prove that every function on $\mathbb{N}^n|_{\leqslant\theta}$ can be transformed into a function on $O(\theta \log n)$ Boolean variables with similar approximation-theoretic properties. Analogous input transformations, with similar proofs, have been used in previous work to translate results from $\{0,1\}^n|_\theta$ or $\{0,1\}^n|_{\leqslant\theta}$ to the hypercube setting [18, 13]. The presentation below seems more economical than previous treatments.

Recall that $e_1, e_2, \ldots, e_n$ denote the standard basis for $\mathbb{R}^n$. The following encoding lemma was proved in [55, Lemma 3.1].

LEMMA 3.7 (Sherstov). *Let $n \geqslant 1$ be a given integer. Then there is a surjection $g\colon \{0,1\}^{6\lceil \log(n+1)\rceil} \to \{0^n, e_1, e_2, \ldots, e_n\}$ such that*

$$\mathop{\mathbf{E}}_{g^{-1}(0^n)} p = \mathop{\mathbf{E}}_{g^{-1}(e_1)} p = \mathop{\mathbf{E}}_{g^{-1}(e_2)} p = \cdots = \mathop{\mathbf{E}}_{g^{-1}(e_n)} p$$

for every polynomial p of degree at most $\lceil \log(n+1)\rceil$. Moreover, g can be constructed deterministically in time polynomial in n.

Observe that the points $0^n, e_1, e_2, \ldots, e_n$ in this lemma act simply as labels and can be replaced with any other tuple of $n+1$ distinct points. Indeed, this result was originally stated in [55] for a different choice of points. A tensor version of Lemma 3.7 is as follows.

LEMMA 3.8. *Let $g\colon \{0,1\}^{6\lceil \log(n+1)\rceil} \to \{0^n, e_1, e_2, \ldots, e_n\}$ be as constructed in Lemma 3.7. Then for any integer $\theta \geqslant 1$ and for any polynomial $p\colon (\mathbb{R}^{6\lceil \log(n+1)\rceil})^\theta \to \mathbb{R}$, the mapping*

$$(y_1, y_2, \ldots, y_\theta) \mapsto \mathop{\mathbf{E}}_{g^{-1}(y_1)\times g^{-1}(y_2)\times\cdots\times g^{-1}(y_\theta)} p$$

is a polynomial in $y \in \{0^n, e_1, e_2, \ldots, e_n\}^\theta$ of degree at most $(\deg p)/\lceil \log(n+1)+1\rceil$.

Proof. By linearity, it suffices to prove consider factored polynomials of the form $p(x_1, x_2, \ldots, x_\theta) = p_1(x_1)p_2(x_2)\cdots p_\theta(x_\theta)$, where $p_1, p_2, \ldots, p_\theta$ are real polynomials on $\{0,1\}^{6\lceil \log(n+1)\rceil}$. For such a polynomial, the defining equation simplifies to

$$\mathop{\mathbf{E}}_{g^{-1}(y_1)\times g^{-1}(y_2)\times\cdots\times g^{-1}(y_\theta)} p \;= \prod_{i=1}^{n} \mathop{\mathbf{E}}_{g^{-1}(y_i)} p_i. \tag{3.15}$$

We now examine the individual contributions of $p_1, p_2, \ldots, p_\theta$ to the degree of the right-hand side as a real polynomial in y. For any polynomial p_i of degree at most $\lceil \log(n+1)\rceil$, Lemma 3.7 ensures that the corresponding expectation $\mathbf{E}_{g^{-1}(y_i)} p_i$ is a constant independent of the input y_i. Thus, polynomials p_i of degree at most $\lceil \log(n+1)\rceil$ do not contribute to the degree of the right-hand side of (3.15). For the other polynomials p_i, the expectation $\mathbf{E}_{g^{-1}(y_i)} p_i$ is a linear polynomial in y_i, namely,

$$\begin{aligned}\mathop{\mathbf{E}}_{g^{-1}(y_i)} p_i = y_{i,1} \mathop{\mathbf{E}}_{g^{-1}(e_1)} p_i + y_{i,2} \mathop{\mathbf{E}}_{g^{-1}(e_2)} p_i + \cdots + y_{i,n} \mathop{\mathbf{E}}_{g^{-1}(e_n)} p_i \\ + \left(1 - \sum_{j=1}^{n} y_{i,j}\right) \mathop{\mathbf{E}}_{g^{-1}(0^n)} p_i,\end{aligned}$$

where we are crucially exploiting the fact that $y_i \in \{0^n, e_1, e_2, \ldots, e_n\}$. Thus, polynomials p_i of degree greater than $\lceil \log(n+1)\rceil$ contribute at most 1 each to the degree. Summarizing, the right-hand side of (3.15) is a real polynomial in

$y_1, y_2, \dots, y_\theta$ of degree at most

$$|\{i : \deg p_i \geqslant \lceil \log(n+1) \rceil + 1\}| \leqslant \frac{\deg p}{\lceil \log(n+1) \rceil + 1}. \qquad \square$$

We have reached the claimed result on input transformation.

THEOREM 3.9. *Let $n, \theta \geqslant 1$ be given integers. Set $N = 6\lceil \log(n+1) \rceil \theta$. There is a surjection $G\colon \{0,1\}^N \to \mathbb{N}^n|_{\leqslant \theta}$ such that:*

(i) *for every polynomial $p\colon \mathbb{R}^N \to \mathbb{R}$, the mapping $v \mapsto \mathbf{E}_{G^{-1}(v)} p$ is a polynomial on $\mathbb{N}^n|_{\leqslant \theta}$ of degree at most $(\deg p)/\lceil \log(n+1) + 1 \rceil$;*
(ii) *for every coordinate $i = 1, 2, \dots, n$, the mapping $x \mapsto \mathrm{OR}^*_\theta(G(x)_i)$ is computable by an explicitly given DNF formula with $O(\theta n^6)$ terms, each with at most $6\lceil \log(n+1) \rceil$ variables.*

Applying Theorem 3.9 to a function $f\colon \mathbb{N}^n|_{\leqslant \theta} \to \{0,1\}$ produces a composed function $f \circ G\colon \{0,1\}^{6\lceil \log(n+1) \rceil \theta} \to \{0,1\}$ in the hypercube setting. The theorem ensures that lower bounds for the pointwise approximation, or sign-representation, of f apply to $f \circ G$ as well. Moreover, the circuit complexity of $f \circ G$ is only slightly higher than that of f. This way, Theorem 3.9 efficiently transfers approximation-theoretic results from $\mathbb{N}^n|_{\leqslant \theta}$ (or any subset thereof, such as $\{0,1\}^n|_{\leqslant \theta}$ or $\mathbb{N}^n|_\theta$) to the traditional setting of the hypercube.

Proof of Theorem 3.9. Define $G\colon (\{0,1\}^{6\lceil \log(n+1) \rceil})^\theta \to \mathbb{N}^n|_{\leqslant \theta}$ by

$$G(x_1, x_2, \dots, x_\theta) = g(x_1) + g(x_2) + \cdots + g(x_\theta),$$

where $g\colon \{0,1\}^{6\lceil \log(n+1) \rceil} \to \{0^n, e_1, e_2, \dots, e_n\}$ is as constructed in Lemma 3.7. The surjectivity of G follows trivially from that of g. We proceed to verify the additional properties required of G.

(i) For $v \in \mathbb{N}^n|_{\leqslant \theta}$, we have the partition

$$G^{-1}(v) = \bigcup_{\substack{y \in \{0^n, e_1, e_2, \dots, e_n\}^\theta\colon \\ y_1 + y_2 + \cdots + y_\theta = v}} g^{-1}(y_1) \times g^{-1}(y_2) \times \cdots \times g^{-1}(y_\theta). \tag{3.16}$$

All representations $v = y_1 + y_2 + \cdots + y_\theta$ with $y_1, y_2, \dots, y_\theta \in \{0^n, e_1, e_2, \dots, e_n\}$ are the same up to the order of the summands. As a result, each part $g^{-1}(y_1) \times g^{-1}(y_2) \times \cdots \times g^{-1}(y_\theta)$ in the partition on the right-hand side of (3.16) has the same cardinality. We conclude that for any given polynomial p,

$$\mathop{\mathbf{E}}_{G^{-1}(v)} p = \mathop{\mathbf{E}}_{\substack{y \in \{0^n, e_1, e_2, \dots, e_n\}^\theta\colon \\ y_1 + y_2 + \cdots + y_\theta = v}} \quad \mathop{\mathbf{E}}_{g^{-1}(y_1) \times g^{-1}(y_2) \times \cdots \times g^{-1}(y_\theta)} p. \tag{3.17}$$

Recall from Lemma 3.8 that the rightmost expectation in this equation is a polynomial in $y_1, y_2, \dots, y_\theta \in \{0^n, e_1, e_2, \dots, e_n\}$ of degree at most $(\deg p)/\lceil \log(n+1) + 1 \rceil$. As a result, Corollary 2.12 implies that the right-hand side of (3.17) is a polynomial in v of degree at most $(\deg p)/\lceil \log(n+1) + 1 \rceil$.

(ii) Fix an index i. Then

$$\mathrm{OR}^*_\theta(G(x)_i) = \bigvee_{j=1}^{\theta} \mathbf{I}[g(x_j) = e_i].$$

Each of the disjuncts on the right-hand side is a function of $6\lceil\log(n+1)\rceil$ Boolean variables. Therefore, $\mathrm{OR}^*_\theta(G(x)_i)$ is representable by a DNF formula with $O(\theta n^6)$ terms, each with at most $6\lceil\log(n+1)\rceil$ variables. □

4. 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. We give analogous amplification theorems for various other approximation-theoretic quantities. 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 4.1–4.3, followed by the application to AC^0 in Section 4.5.

4.1. 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 work, 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

$$\frac{c^{t+1}}{(t+1)^2\, 2^{\alpha t}} \leqslant \lambda(t) \leqslant \frac{1}{c(t+1)^2\, 2^{\alpha t}}, \qquad t \in \operatorname{supp}\lambda. \tag{4.1}$$

Distributions in this family are subject to pointwise constraints, hence the symbol $\mathfrak{B}$ for "bounded." Our choice of bounding functions is motivated mainly by the metric properties of the dual polynomial for OR_n, constructed in Theorem 3.3.

In this notation, our analysis handles any distribution $\Lambda \in \mathfrak{B}(r, c, \alpha)^{\otimes n}$. It would be possible to generalize our work further, but the lower and upper bounds in (4.1) are already exponentially far apart and capture a much larger class of probability distributions than what we need for the applications to AC^0. The precise statement of our result is as follows.

THEOREM 4.1. *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*

$$\theta \geqslant 2d, \tag{4.2}$$

$$\theta \geqslant \frac{4en(1+\ln n)}{c^2}. \tag{4.3}$$

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

$$\operatorname{supp}\tilde{\Lambda} \subseteq (\operatorname{supp}\Lambda)|_{<2\theta}, \tag{4.4}$$

$$\operatorname{orth}(\Lambda - \tilde{\Lambda}) > d, \tag{4.5}$$

$$|\Lambda - \tilde{\Lambda}| \leqslant \left(\frac{8nr}{c}\right)^d 2^{-\lceil\theta/r\rceil - \alpha\lceil\theta/2\rceil + 2}\,\Lambda \qquad \text{on } \operatorname{supp}\tilde{\Lambda}. \tag{4.6}$$

In general, the function $\tilde{\Lambda}$ constructed in Theorem 4.1 may not be a probability distribution. However, when θ is large enough relative to the other parameters, the pointwise property (4.6) forces $|\Lambda - \tilde{\Lambda}| \leqslant \Lambda$ and in particular $\tilde{\Lambda} \geqslant 0$. Since $\operatorname{orth}(\Lambda - \tilde{\Lambda}) > 0$ by construction, Proposition 2.4 guarantees that $\tilde{\Lambda}$ *is* a probability distribution in that case.

Proof of Theorem 4.1. 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 3.5. To restate the lemma's guarantees,

$$|S(v)| \geqslant \frac{\theta}{r}, \qquad v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}, \tag{4.7}$$

$$\min_{i \in S(v)} v_i \geqslant \frac{\theta}{2|S(v)|(1+\ln n)}, \qquad v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}, \tag{4.8}$$

$$\|v|_{\overline{S(v)}}\|_1 < \theta. \qquad v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}. \tag{4.9}$$

Property (4.9) implies that

$$\|v|_{S(v)}\|_1 > \theta, \qquad v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}, \tag{4.10}$$

and in particular

$$\|v|_{S(v)}\|_1 > d, \qquad v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}. \tag{4.11}$$

For each $i = 1, 2, \ldots, n$ and each $u \in \mathbb{N}^i|_{>d}$, Lemma 3.2 gives a function $\zeta_u \colon \mathbb{N}^i \to \mathbb{R}$ such that

$$\operatorname{supp} \zeta_u \subseteq \{u\} \cup \{v \in \mathbb{N}^i : v \leqslant u \text{ and } |v| \leqslant d\}, \tag{4.12}$$

$$\zeta_u(u) = 1, \tag{4.13}$$

$$\|\zeta_u\|_1 \leqslant 1 + 2^d \binom{\|u\|_1}{d}, \tag{4.14}$$

$$\operatorname{orth} \zeta_u > d, \tag{4.15}$$

and in particular

$$\begin{aligned} \|\zeta_u\|_\infty &\leqslant \max\{|\zeta_u(u)|, \|\zeta_u\|_1 - |\zeta_u(u)|\} \\ &\leqslant 2^d \binom{\|u\|_1}{d} \\ &\leqslant 2\|u\|_1{}^d. \end{aligned} \tag{4.16}$$

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:

$$\zeta(x) = \sum_{v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}} \Lambda(v)\, \zeta_{v|_{S(v)}}(x|_{S(v)})\, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}]. \tag{4.17}$$

The expression on the right-hand side is well-formed because, to restate (4.11), each string $v|_{S(v)}$ has weight greater than d and can therefore be used as a subscript in $\zeta_{v|_{S(v)}}$. Specializing (4.15) and (4.16),

$$\operatorname{orth} \zeta_{v|_{S(v)}} > d, \qquad v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}, \tag{4.18}$$

$$\|\zeta_{v|_{S(v)}}\|_\infty \leqslant 2(nr)^d, \qquad v \in (\operatorname{supp} \Lambda)|_{\geqslant 2\theta}. \tag{4.19}$$

Property (4.12) 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

$$\begin{aligned} \operatorname{supp} \zeta &\subseteq \bigcup_{v \in \operatorname{supp} \Lambda} \{x \in \mathbb{N}^n : x \leqslant v\} \\ &= \operatorname{supp} \Lambda, \end{aligned} \tag{4.20}$$

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

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

where the first, second, and fourth steps are valid by (4.12), (4.11), and (4.13), respectively. Making this substitution in the defining equation for ζ,

$$\begin{aligned}
\zeta(x) = \sum_{v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} &\Lambda(v)\zeta_{v|_{S(v)}}(x|_{S(v)})\mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d]\,\mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \\
&+ \sum_{v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} \Lambda(v)\mathbf{I}[x = v]. \qquad (4.21)
\end{aligned}$$

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 (4.17) is a function orthogonal to polynomials of degree less than $\operatorname{orth}\zeta_{v|_{S(v)}}$. Therefore,

$$\begin{aligned}
\operatorname{orth}\zeta &\geqslant \min_{v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} \operatorname{orth}\zeta_{v|_{S(v)}} \\
&> d, \qquad (4.22)
\end{aligned}$$

where the first step uses Proposition 2.1(i) and the second step applies (4.18).

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 (4.9). We conclude that the first summation in (4.21) vanishes on $\mathbb{N}^n|_{\geqslant 2\theta}$, so that

$$\zeta(x) = \Lambda(x), \qquad x \in \mathbb{N}^n|_{\geqslant 2\theta}. \qquad (4.23)$$

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

$$\begin{aligned}
\frac{|\zeta(x)|}{\Lambda(x)} &= \left| \sum_{v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} \frac{\Lambda(v)}{\Lambda(x)} \zeta_{v|_{S(v)}}(x|_{S(v)}) \, \mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d] \, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \right| \\
&\leqslant \sum_{v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} \frac{\Lambda(v)}{\Lambda(x)} |\zeta_{v|_{S(v)}}(x|_{S(v)})| \, \mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d] \, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \\
&\leqslant 2(nr)^d \sum_{v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}} \frac{\Lambda(v)}{\Lambda(x)} \, \mathbf{I}[\|x|_{S(v)}\|_1 \leqslant d] \, \mathbf{I}[x|_{\overline{S(v)}} = v|_{\overline{S(v)}}] \\
&= 2(nr)^d \sum_{\substack{S \subseteq \{1,\ldots,n\}: \\ |S| \geqslant \theta/r}} \mathbf{I}[\|x|_S\|_1 \leqslant d] \sum_{\substack{v \in (\operatorname{supp}\Lambda)|_{\geqslant 2\theta}: \\ S(v) = S}} \frac{\Lambda(v)}{\Lambda(x)} \, \mathbf{I}[x|_{\overline{S}} = v|_{\overline{S}}] \\
&\leqslant 2(nr)^d \sum_{\substack{S \subseteq \{1,\ldots,n\}: \\ |S| \geqslant \theta/r}} \mathbf{I}[\|x|_S\|_1 \leqslant d] \sum_{\substack{v \in \mathbb{N}^n: \\ \sum_{i \in S} v_i \geqslant \theta, \\ \min_{i \in S} v_i \geqslant \frac{\theta}{2|S|(1+\ln n)}}} \frac{\Lambda(v)}{\Lambda(x)} \, \mathbf{I}[x|_{\overline{S}} = v|_{\overline{S}}],
\end{aligned} \tag{4.24}$$

where the first step uses (4.21); the second step applies the triangle inequality; the third step is valid by (4.19); the fourth step amounts to collecting terms according to $S(v)$, which by (4.7) has cardinality at least θ/r; and the fifth step uses (4.8) and (4.10).

Bounding (4.24) requires a bit of work. To start with, write $\Lambda = \bigotimes_{i=1}^n \lambda_i$ for some $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathfrak{B}(r, c, \alpha)$. Then for every nonempty set $S \subseteq \{1, 2, \ldots, n\}$,

$$\begin{aligned}
\mathbf{I}[\|x|_S\|_1 \leqslant d] \prod_{i \in S} \lambda_i(x_i) &\geqslant \mathbf{I}[\|x|_S\|_1 \leqslant d] \prod_{i \in S} \frac{c^{x_i+1}}{(x_i+1)^2 \, 2^{\alpha x_i}} \\
&= \mathbf{I}[\|x|_S\|_1 \leqslant d] \, c^{|S|} \left(\frac{c}{2^\alpha} \right)^{\sum_{i \in S} x_i} \prod_{i \in S} \frac{1}{(x_i+1)^2} \\
&\geqslant \mathbf{I}[\|x|_S\|_1 \leqslant d] \, c^{|S|} \left(\frac{c}{2^\alpha} \right)^{\sum_{i \in S} x_i} \left(\frac{|S|}{\sum_{i \in S}(x_i+1)} \right)^{2|S|} \\
&\geqslant c^{|S|} \left(\frac{c}{2^\alpha} \right)^d \left(\frac{|S|}{|S|+d} \right)^{2|S|} \\
&\geqslant c^{|S|} \left(\frac{c}{2^\alpha \mathrm{e}^2} \right)^d,
\end{aligned} \tag{4.25}$$

where the first step applies the definition of $\mathfrak{B}(r, c, \alpha)$; the third step is valid by the arithmetic-geometric mean inequality; and the last step uses the bound $1 + t \leqslant \mathrm{e}^t$

for real t. Continuing,

$$\begin{aligned}
&\sum_{\substack{v\in\mathbb{N}^n:\\ \sum_{i\in S} v_i \geqslant \theta,\\ \min_{i\in S} v_i \geqslant \frac{\theta}{2|S|(1+\ln n)}}} \frac{\Lambda(v)}{\Lambda(x)}\, \mathbf{I}[x|_{\overline{S}} = v|_{\overline{S}}] \\
&\qquad = \sum_{\substack{v\in\mathbb{N}^n:\\ \sum_{i\in S} v_i \geqslant \theta,\\ \min_{i\in S} v_i \geqslant \frac{\theta}{2|S|(1+\ln n)},\\ v_i = x_i \text{ for } i\notin S}} \prod_{i\in S} \frac{\lambda_i(v_i)}{\lambda_i(x_i)} \\
&\qquad \leqslant \sum_{\substack{v\in\mathbb{N}^n:\\ \sum_{i\in S} v_i \geqslant \theta,\\ \min_{i\in S} v_i \geqslant \frac{\theta}{2|S|(1+\ln n)},\\ v_i = x_i \text{ for } i\notin S}} 2^{-\alpha \sum_{i\in S} v_i} \prod_{i\in S} \frac{1}{c(v_i+1)^2 \lambda_i(x_i)} \\
&\qquad \leqslant \sum_{\substack{v\in\mathbb{N}^n:\\ \min_{i\in S} v_i \geqslant \frac{\theta}{2|S|(1+\ln n)},\\ v_i = x_i \text{ for } i\notin S}} 2^{-\alpha\theta} \prod_{i\in S} \frac{1}{c(v_i+1)^2 \lambda_i(x_i)} \\
&\qquad = 2^{-\alpha\theta} \left(\sum_{t=\lceil \frac{\theta}{2|S|(1+\ln n)} \rceil}^{\infty} \frac{1}{c(t+1)^2} \right)^{|S|} \prod_{i\in S} \frac{1}{\lambda_i(x_i)} \\
&\qquad \leqslant 2^{-\alpha\theta} \left(\int_{\lceil \frac{\theta}{2|S|(1+\ln n)} \rceil}^{\infty} \frac{dt}{ct^2} \right)^{|S|} \prod_{i\in S} \frac{1}{\lambda_i(x_i)} \\
&\qquad \leqslant 2^{-\alpha\theta} \left(\frac{2|S|(1+\ln n)}{c\theta} \right)^{|S|} \prod_{i\in S} \frac{1}{\lambda_i(x_i)}, \qquad (4.26)
\end{aligned}$$

where the first step uses $\Lambda = \bigotimes_{i=1}^n \lambda_i$, and the second step applies the definition of $\mathfrak{B}(r, c, \alpha)$.

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

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

where the first step follows from (4.24) and (4.26); the second step substitutes the bound from (4.25); the third step uses (4.2); and the next-to-last step uses (4.3). In summary, we have shown that

$$|\zeta(x)| \leqslant 4 \cdot \frac{(\mathrm{e}^2 nr/c)^d}{2^{\alpha\lceil\theta/2\rceil + \lceil\theta/r\rceil}} \Lambda(x), \qquad\qquad x \in (\operatorname{supp}\Lambda)|_{<2\theta}. \tag{4.27}$$

Step 4: Finishing the proof. Define $\tilde{\Lambda} = \Lambda - \zeta$. Then the support property (4.4) follows from (4.20) and (4.23); the analytic indistinguishability property (4.5) follows from (4.22); and the pointwise property (4.6) follows from (4.4) and (4.27). □

We record a generalization of Theorem 4.1 to translates of probability distributions in $\mathfrak{B}(r,c,\alpha)^{\otimes n}$, and further to convex combinations of such distributions. Formally, define $\mathfrak{B}(r,c,\alpha,\Delta)$ for $\Delta \geqslant 0$ to be the family of probability distributions λ on $\mathbb{N}$ such that $\lambda(t) \equiv \lambda'(t-a)$ for some $\lambda' \in \mathfrak{B}(r,c,\alpha)$ and $a \in [0,\Delta]$. We have:

Corollary 4.2. *Let $\Lambda \in \operatorname{conv}(\mathfrak{B}(r,c,\alpha,\Delta)^{\otimes n})$ be given, for some integers $r, \Delta \geqslant 0$ and reals $c > 0$ and $\alpha \geqslant 0$. Let d and θ be positive integers with*

$$\theta \geqslant 2d, \tag{4.28}$$

$$\theta \geqslant \frac{4\mathrm{e}n(1+\ln n)}{c^2}. \tag{4.29}$$

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

$$\operatorname{supp}\tilde{\Lambda} \subseteq (\operatorname{supp}\Lambda)|_{<2\theta+n\Delta}, \tag{4.30}$$

$$\operatorname{orth}(\Lambda - \tilde{\Lambda}) > d, \tag{4.31}$$

$$|\Lambda - \tilde{\Lambda}| \leqslant \left(\frac{8nr}{c}\right)^d 2^{-\lceil\theta/r\rceil - \alpha\lceil\theta/2\rceil + 2}\,\Lambda \qquad \text{on } \operatorname{supp}\tilde{\Lambda}. \tag{4.32}$$

Proof. We first consider the special case when $\Lambda \in \mathfrak{B}(r,c,\alpha,\Delta)^{\otimes n}$. Then by definition, $\Lambda(t_1,\dots,t_n) = \Lambda'(t_1 - a_1,\dots,t_n - a_n)$ for some probability distribution $\Lambda' \in \mathfrak{B}(r,c,\alpha)^{\otimes n}$ and integers $a_1,\dots,a_n \in [0,\Delta]$. Applying Theorem 4.1 to Λ' yields a function $\tilde{\Lambda}'\colon \mathbb{N}^n \to \mathbb{R}$ with

$$\operatorname{supp}\tilde{\Lambda}' \subseteq (\operatorname{supp}\Lambda)|_{<2\theta}, \tag{4.33}$$

$$\operatorname{orth}(\Lambda' - \tilde{\Lambda}') > d, \tag{4.34}$$

$$|\Lambda' - \tilde{\Lambda}'| \leqslant \left(\frac{8nr}{c}\right)^d 2^{-\lceil\theta/r\rceil - \alpha\lceil\theta/2\rceil + 2}\,\Lambda' \qquad \text{on } \operatorname{supp}\tilde{\Lambda}'. \tag{4.35}$$

Then properties (4.30)–(4.32) follow from (4.33)–(4.35), respectively, for the function $\tilde{\Lambda}\colon \prod_{i=1}^n \{a_i, a_i+1, \dots\} \to \mathbb{R}$ given by $\tilde{\Lambda}(t_1,\dots,t_n) = \tilde{\Lambda}'(t_1 - a,\dots,t_n - a_n)$.

In the general case of a convex combination $\Lambda = \lambda_1\Lambda_1 + \cdots + \lambda_k\Lambda_k$ of probability distributions $\Lambda_1,\dots,\Lambda_k \in \mathfrak{B}(r,c,\alpha,\Delta)^{\otimes n}$, one uses the technique of the previous paragraph to transform $\Lambda_1,\dots,\Lambda_k$ individually into corresponding functions $\tilde{\Lambda}_1,\dots,\tilde{\Lambda}_k$, and takes $\tilde{\Lambda} = \lambda_1\tilde{\Lambda}_1 + \cdots + \lambda_k\tilde{\Lambda}_k$. □

4.2. 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

$$\operatorname{supp}\lambda = \{0,1,2,\dots,r'\}$$

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. As before, 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 a first 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 $\mathfrak{B}^*$. To simplify the exposition, we start with an auxiliary construction.

LEMMA 4.3. *Let $0 < \epsilon < 1$ be given. Then for some constants $c_1, c_2 \in (0,1)$ and all integers $R \geqslant r \geqslant 1$, there are (explicitly given) probability distributions $\lambda_0, \lambda_1, \lambda_2$ such that:*

$$\operatorname{supp} \lambda_0 = \{0\}, \tag{4.36}$$

$$\operatorname{supp} \lambda_i = \{1, 2, \ldots, R\}, \qquad i = 1, 2, \tag{4.37}$$

$$\lambda_i \in \mathfrak{B}^* \left(R, c_1, \frac{c_2}{\sqrt{r}}, 1 \right), \qquad i = 0, 1, 2, \tag{4.38}$$

$$\operatorname{orth}((1-\epsilon)\lambda_0 + \epsilon\lambda_2 - \lambda_1) \geqslant c_1 \sqrt{r}. \tag{4.39}$$

Our analysis of the threshold degree of AC^0 only uses the special case $R = r$ of Lemma 4.3. The more general formulation with $R \geqslant r$ will be needed much later, in the analysis of the sign-rank of AC^0.

Proof. Theorem 3.3 constructs a function $\psi \colon \{0, 1, 2, \ldots, R\} \to \mathbb{R}$ such that

$$\psi(0) > \frac{1 - \frac{\epsilon}{2}}{2}, \tag{4.40}$$

$$\|\psi\|_1 = 1, \tag{4.41}$$

$$\operatorname{orth} \psi \geqslant c' \sqrt{r}, \tag{4.42}$$

$$|\psi(t)| \in \left[\frac{c'}{(t+1)^2\, 2^{c''t/\sqrt{r}}},\ \frac{1}{c'(t+1)^2\, 2^{c''t/\sqrt{r}}} \right], \qquad t = 0, 1, \ldots, r, \tag{4.43}$$

for some absolute constants $c', c'' \in (0,1)$. Property (4.41) makes it possible to view $|\psi|$ as a probability distribution on $\{0, 1, 2, \ldots, R\}$. Let μ_0, μ_1, μ_2 be the probability distributions induced by $|\psi|$ on $\{0\}, \{t \neq 0 : \psi(t) < 0\}$, and $\{t \neq 0 : \psi(t) > 0\}$, respectively. It is clear from (4.40) that the negative part of ψ is a multiple of μ_1, whereas the positive part of ψ is a nonnegative linear combination of μ_0 and μ_2. Moreover, it follows from $\langle \psi, 1 \rangle = 0$ and $\|\psi\|_1 = 1$ that the positive and negative parts of ψ both have ℓ_1-norm $1/2$. Summarizing,

$$\psi = \frac{1-\delta}{2}\mu_0 - \frac{1}{2}\mu_1 + \frac{\delta}{2}\mu_2 \tag{4.44}$$

for some $0 \leqslant \delta \leqslant 1$. In view of (4.40), we infer the more precise bound

$$0 \leqslant \delta < \frac{\epsilon}{2}. \tag{4.45}$$

We define

$$\lambda_0 = \mu_0, \tag{4.46}$$

$$\lambda_1 = \frac{1-\epsilon\delta}{1-\delta^2}\mu_1 + \delta \cdot \frac{\epsilon-\delta}{1-\delta^2}\mu_2, \tag{4.47}$$

$$\lambda_2 = \frac{\epsilon-\delta}{\epsilon(1-\delta^2)}\mu_1 + \delta \cdot \frac{1-\epsilon\delta}{\epsilon(1-\delta^2)}\mu_2. \tag{4.48}$$

It follows from $0 \leqslant \delta \leqslant \epsilon$ that λ_1 and λ_2 are convex combinations of μ_1 and μ_2 and are therefore probability distributions with support

$$\operatorname{supp} \lambda_i \subseteq \{1, 2, \ldots, R\}, \qquad i = 1, 2. \tag{4.49}$$

Recall from (4.44) that $|\psi| = \frac{1}{2}\mu_1 + \frac{\delta}{2}\mu_2$ on $\{1, 2, \ldots, R\}$. Comparing the coefficients in $|\psi| = \frac{1}{2}\mu_1 + \frac{\delta}{2}\mu_2$ with the corresponding coefficients in the defining equations for λ_1 and λ_2, where $0 \leqslant \delta \leqslant \epsilon/2$ by (4.45), we conclude that $\lambda_1, \lambda_2 \in [c'''|\psi|, |\psi|/c''']$ on $\{1, 2, \ldots, R\}$ for some constant $c''' = c'''(\epsilon) \in (0, 1)$. In view of (4.43), we arrive at

$$|\lambda_i(t)| \in \left[\frac{c'c'''}{(t+1)^2\, 2^{c''t/\sqrt{r}}},\ \frac{1}{c'c'''(t+1)^2\, 2^{c''t/\sqrt{r}}}\right], \qquad i = 1, 2; \quad t = 1, 2, \ldots, R. \tag{4.50}$$

Continuing,

$$\begin{aligned} \operatorname{orth}((1-\epsilon)\lambda_0 + \epsilon\lambda_2 - \lambda_1) &= \operatorname{orth}\left(2 \cdot \frac{1-\epsilon}{1-\delta}\left(\frac{1-\delta}{2}\mu_0 - \frac{1}{2}\mu_1 + \frac{\delta}{2}\mu_2\right)\right) \\ &= \operatorname{orth}\left(2 \cdot \frac{1-\epsilon}{1-\delta}\,\psi\right) \\ &\geqslant c'\sqrt{r}, \end{aligned} \tag{4.51}$$

where the first step follows from the defining equations (4.46)–(4.48), the second step uses (4.44), and the final step is a restatement of (4.45).

We are now in a position to verify the claimed properties of $\lambda_0, \lambda_1, \lambda_2$ in the theorem statement. Property (4.36) follows from (4.46), whereas property (4.37) is immediate from (4.49) and (4.50). The remaining properties (4.38) and (4.39) for small enough constants $c_1, c_2 \in (0, 1)$ now follow from (4.50) and (4.51), respectively. ∎

We are now in a position to construct our desired dual polynomial for the Minsky–Papert function.

THEOREM 4.4. *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*

$$\Lambda_i \in \operatorname{conv}\left(\mathfrak{B}^*\left(r, c_1, \frac{c_2}{\sqrt{r}}, 1\right)^{\otimes m}\right), \qquad i = 0, 1, \tag{4.52}$$

$$\operatorname{supp} \Lambda_i \subseteq (\mathrm{MP}^*_{m,r})^{-1}(i), \qquad i = 0, 1, \tag{4.53}$$

$$\operatorname{orth}(\Lambda_1 - \Lambda_0) \geqslant \min\{m, c_1\sqrt{r}\}. \tag{4.54}$$

The last two properties in the theorem statement are equivalent, in the sense of linear programming duality, to the lower bound $\deg_{\pm}(\mathrm{MP}^*_{m,r}) \geqslant \min\{m, c_1\sqrt{r}\}$ and can be recovered in a black-box manner from many previous papers, e.g., [35, 45, 53]. The key new property that we prove is (4.52), with the newly established Lemma 4.3 playing an essential role.

Proof of Theorem 4.4. Take $\epsilon = 1/2$ and $R = r$ in Lemma 4.3, and let $\lambda_0, \lambda_1, \lambda_2$ be the resulting probability distributions. Let

$$\begin{aligned}
\Lambda_0 &= \mathop{\mathbf{E}}_{\substack{S \subseteq \{1,2,\ldots,m\} \\ |S| \text{ odd}}} \lambda_0^{\otimes S} \cdot \lambda_2^{\otimes \overline{S}}, \\
\Lambda_1 &= \lambda_1^{\otimes m}.
\end{aligned}$$

Then (4.52) is immediate from (4.38), whereas (4.53) follows from (4.36) and (4.37). To verify the remaining property (4.54), rewrite

$$\begin{aligned}
\Lambda_0 &= 2^{-m+1} \sum_{\substack{S \subseteq \{1,2,\ldots,m\} \\ |S| \text{ odd}}} \lambda_0^{\otimes S} \cdot \lambda_2^{\otimes \overline{S}} \\
&= \left(\frac{1}{2}\lambda_0 + \frac{1}{2}\lambda_2\right)^{\otimes m} - \left(-\frac{1}{2}\lambda_0 + \frac{1}{2}\lambda_2\right)^{\otimes m}.
\end{aligned}$$

Observe that

$$\operatorname{orth}(\lambda_i - \lambda_j) \geqslant 1, \qquad i, j = 0, 1, 2, \tag{4.55}$$

which can be seen from $\langle \lambda_i - \lambda_j, 1\rangle = \langle \lambda_i, 1\rangle - \langle \lambda_j, 1\rangle = 1 - 1 = 0$. Now

$$\begin{aligned}
&\operatorname{orth}(\Lambda_1 - \Lambda_0) \\
&\quad = \operatorname{orth}\left(\lambda_1^{\otimes m} - \left(\frac{1}{2}\lambda_0 + \frac{1}{2}\lambda_2\right)^{\otimes m} + \left(-\frac{1}{2}\lambda_0 + \frac{1}{2}\lambda_2\right)^{\otimes m}\right) \\
&\quad \geqslant \min\left\{\operatorname{orth}\left(\lambda_1^{\otimes m} - \left(\frac{1}{2}\lambda_0 + \frac{1}{2}\lambda_2\right)^{\otimes m}\right), \operatorname{orth}\left(-\frac{1}{2}\lambda_0 + \frac{1}{2}\lambda_2\right)^{\otimes m}\right\} \\
&\quad \geqslant \min\left\{\operatorname{orth}\left(\lambda_1 - \frac{1}{2}\lambda_0 - \frac{1}{2}\lambda_2\right), \operatorname{orth}\left(-\frac{1}{2}\lambda_0 + \frac{1}{2}\lambda_2\right)^{\otimes m}\right\} \\
&\quad = \min\left\{\operatorname{orth}\left(\lambda_1 - \frac{1}{2}\lambda_0 - \frac{1}{2}\lambda_2\right), m \operatorname{orth}\left(-\frac{1}{2}\lambda_0 + \frac{1}{2}\lambda_2\right)\right\} \\
&\quad = \min\left\{\operatorname{orth}\left(\lambda_1 - \frac{1}{2}\lambda_0 - \frac{1}{2}\lambda_2\right), m\right\} \\
&\quad \geqslant \min\{c\sqrt{r}, m\},
\end{aligned}$$

where the last four steps are valid by Proposition 2.1(i), Proposition 2.1(iii), Proposition 2.1(ii), equation (4.55), and equation (4.39), respectively. □

4.3. Hardness amplification for threshold degree and beyond. We now present a blackbox transformation that takes any given circuit 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. We obtain analogous hardness amplification results for a host of other approximation-theoretic complexity measures. For this reason, we adopt the following abstract view of polynomial approximation. Let I_0, I_1, I_* be nonempty convex subsets of the real line, i.e., any kind of nonempty intervals (closed, open, or half-open; bounded or unbounded). Let $f\colon X \to \{0, 1, *\}$ be a (possibly partial) Boolean function on a finite subset X of Euclidean space. We define an *(I_0, I_1, I_*)-approximant for f* to be any real polynomial p that maps $f^{-1}(0), f^{-1}(1), f^{-1}(*)$ into I_0, I_1, I_*, respectively. The *(I_0, I_1, I_*)-approximate degree of f*, denoted $\deg_{I_0,I_1,I_*}(f)$, the least degree of an (I_0, I_1, I_*)-approximant for f. Threshold degree corresponds to the special case

$$\deg_{\pm} = \deg_{(0,\infty),(-\infty,0),(-\infty,\infty)}. \tag{4.56}$$

Other notable cases include *ϵ-approximate degree* and *one-sided ϵ-approximate degree*, given by

$$\deg_{\epsilon} = \deg_{[-\epsilon,\epsilon],[1-\epsilon,1+\epsilon],[-\epsilon,1+\epsilon]}, \tag{4.57}$$

$$\deg^{+}_{\epsilon} = \deg_{[-\epsilon,\epsilon],[1-\epsilon,\infty),(-\infty,\infty)}, \tag{4.58}$$

respectively. Our hardness amplification result applies to (I_0, I_1, I_*)-approximate degree for any nonempty convex $I_0, I_1, I_* \subseteq \mathbb{R}$, with threshold degree being a special case. The centerpiece of our argument is the following lemma.

LEMMA 4.5. *Let $c, c', c'' > 0$ be the absolute constants from Theorem 4.4. Let n, m, r, d, θ be positive integers such that*

$$\theta \geqslant 2d, \tag{4.59}$$

$$\theta \geqslant \frac{4enm(1+\ln(nm))}{c'^2}, \tag{4.60}$$

$$\theta \geqslant \frac{2\sqrt{r}}{c''}\left(d\log\left(\frac{8nmr}{c'}\right)+2\right). \tag{4.61}$$

Then for each $z \in \{0,1\}^n$, there is a probability distribution $\tilde{\Lambda}_z$ on $\mathbb{N}^{nm}$ such that:

(i) *the support of $\tilde{\Lambda}_z$ is contained in $(\prod_{i=1}^n (\mathrm{MP}^*_{m,r})^{-1}(z_i))|_{<2\theta+nm}$;*
(ii) *for every polynomial $p\colon \mathbb{R}^{nm} \to \mathbb{R}$ of degree at most d, the mapping $z \mapsto \mathbf{E}_{\tilde{\Lambda}_z} p$ is a polynomial on $\{0,1\}^n$ of degree at most $\frac{1}{\min\{m, c\sqrt{r}\}} \cdot \deg p$.*

Proof. Theorem 4.4 constructs probability distributions Λ_0 and Λ_1 such that

$$\Lambda_i \in \operatorname{conv}\left(\mathfrak{B}^*\left(r, c', \frac{c''}{\sqrt{r}}, 1\right)^{\otimes m}\right), \qquad i = 0, 1, \tag{4.62}$$

$$\operatorname{supp}\Lambda_i \subseteq (\mathrm{MP}^*_{m,r})^{-1}(i), \qquad i = 0, 1, \tag{4.63}$$

$$\operatorname{orth}(\Lambda_1 - \Lambda_0) \geqslant \min\{m, c\sqrt{r}\}. \tag{4.64}$$

As a result, the probability distributions $\Lambda_z = \bigotimes_{i=1}^n \Lambda_{z_i}$ for $z \in \{0,1\}^n$ obey

$$\begin{aligned}\Lambda_z &\in \left(\operatorname{conv}\left(\mathfrak{B}^*\left(r, c', \frac{c''}{\sqrt{r}}, 1\right)^{\otimes m}\right)\right)^{\otimes n} \\ &\subseteq \operatorname{conv}\left(\mathfrak{B}^*\left(r, c', \frac{c''}{\sqrt{r}}, 1\right)^{\otimes nm}\right) \\ &\subseteq \operatorname{conv}\left(\mathfrak{B}\left(r, c', \frac{c''}{\sqrt{r}}, 1\right)^{\otimes nm}\right).\end{aligned} \tag{4.65}$$

By (4.59)–(4.61), (4.65), and Corollary 4.2, there are functions $\tilde{\Lambda}_z\colon \mathbb{N}^{nm} \to \mathbb{R}$ for $z \in \{0,1\}^n$ such that

$$\operatorname{supp}\tilde{\Lambda}_z \subseteq (\operatorname{supp}\Lambda_z)|_{<2\theta+nm}, \tag{4.66}$$

$$\operatorname{orth}(\Lambda_z - \tilde{\Lambda}_z) > d, \tag{4.67}$$

$$|\Lambda_z - \tilde{\Lambda}_z| \leqslant \Lambda_z \qquad \text{on } \operatorname{supp}\tilde{\Lambda}_z. \tag{4.68}$$

We now verify the properties claimed in the statement of the lemma. The pointwise bound (4.68) implies that each $\tilde{\Lambda}_z$ is a probability distribution. By (4.63) and (4.66), each $\tilde{\Lambda}_z$ has support contained in $(\prod_{i=1}^n (\mathrm{MP}^*_{m,r})^{-1}(z_i))|_{<2\theta+nm}$. Finally, let p be any polynomial of degree at most d. Then (4.67) guarantees that $\mathbf{E}_{\tilde{\Lambda}_z} p = \mathbf{E}_{\Lambda_z} p$, where the right-hand side is by (4.64) and Proposition 2.2 a polynomial in $z \in \{0,1\}^n$ of degree at most $\deg p / \operatorname{orth}(\Lambda_1 - \Lambda_0) \leqslant \deg p / \min\{m, c\sqrt{r}\}$. □

At its core, a hardness amplification result is a lower bound on the complexity of a composed function in terms of the complexities of its constituent parts. We now prove such a composition theorem for (I_0, I_1, I_*)-approximate degree.

THEOREM 4.6. *There is an absolute constant $0 < c < 1$ such that*

$$\deg_{I_0,I_1,I_*}((f \circ \mathrm{MP}_m^*)|_{\leqslant\theta}) \geqslant \min\left\{cm \deg_{I_0,I_1,I_*}(f), \frac{c\theta}{m\log(n+m)} - n\right\},$$
$$\deg_{I_0,I_1,I_*}((f \circ \neg\mathrm{MP}_m^*)|_{\leqslant\theta}) \geqslant \min\left\{cm \deg_{I_0,I_1,I_*}(f), \frac{c\theta}{m\log(n+m)} - n\right\}$$

for all positive integers n, m, θ, all functions $f\colon \{0,1\}^n \to \{0,1,*\}$, and all nonempty convex sets $I_0, I_1, I_* \subseteq \mathbb{R}$.

As a practical matter, note that the left-hand sides of the inequalities in Theorem 4.6 are monotonic functions of m. Therefore, the theorem implies that $(f \circ \mathrm{MP}_m^*)|_{\leqslant\theta}$ and $(f \circ \neg\mathrm{MP}_m^*)|_{\leqslant\theta}$ have (I_0, I_1, I_*)-approximate degree at least

$$\max_{m'=1,2,\ldots,m} \min\left\{cm' \deg_{I_0,I_1,I_*}(f), \frac{c\theta}{m'\log(n+m')} - n\right\}.$$

Proof of Theorem 4.6. Negating a function's input has no effect on the (I_0, I_1, I_*)-approximate degree, so that $f(x_1, x_2, \ldots, x_n)$ and $f(\neg x_1, \neg x_2, \ldots, \neg x_n)$ both have (I_0, I_1, I_*)-approximate degree $\deg_{I_0,I_1,I_*}(f)$. Therefore, it suffices to prove the lower bound on $\deg_{I_0,I_1,I_*}((f \circ \mathrm{MP}_m^*)|_{\leqslant\theta})$ for all f.

Let $c \in (0,1)$ be an absolute constant that is sufficiently small relative to the constants in Lemma 4.5. For $\theta \leqslant \frac{1}{c} \cdot nm\log(n+m)$, the lower bounds in the statement of the theorem are nonpositive and therefore trivially true. In the complementary case $\theta > \frac{1}{c} \cdot nm\log(n+m)$, Lemma 4.5 applies to the positive integers n', m', r', d', θ', where

$$\begin{aligned} n' &= n, \\ m' &= m, \\ r' &= m^2, \\ \theta' &= \left\lfloor \frac{\theta - nm}{2} \right\rfloor, \\ d' &= \left\lfloor \frac{c\theta}{m\log(n+m)} \right\rfloor. \end{aligned}$$

We thus obtain, for each $z \in \{0,1\}^n$, a probability distribution $\tilde{\Lambda}_z$ on $\mathbb{N}^{nm}$ such that:

(i) the support of $\tilde{\Lambda}_z$ is contained in $(\prod_{i=1}^n (\mathrm{MP}_m^*)^{-1}(z_i))|_{\leqslant\theta}$;
(ii) for every polynomial $p\colon \mathbb{R}^{nm} \to \mathbb{R}$ of degree at most d', the mapping $z \mapsto \mathbf{E}_{\tilde{\Lambda}_z} p$ is a polynomial on $\{0,1\}^n$ of degree at most $\frac{1}{cm} \cdot \deg p$.

Now, let $p\colon \mathbb{R}^{nm} \to \mathbb{R}$ be an (I_0, I_1, I_*)-approximant for $(f \circ \mathrm{MP}_m^*)|_{\leqslant\theta}$ of degree at most d'. Consider the mapping $p^*\colon z \mapsto \mathbf{E}_{\tilde{\Lambda}_z} p$, which we view as a polynomial in $z \in \{0,1\}^n$. Then (i) along with the convexity of I_0, I_1, I_* ensures that p^* is an (I_0, I_1, I_*)-approximant for f, whence $\deg p^* \geqslant \deg_{I_0,I_1,I_*}(f)$. At the same time, (ii)

guarantees that $\deg p^* \leqslant \frac{1}{cm} \cdot \deg p$. This pair of lower and upper bounds force

$$\deg p \geqslant cm \deg_{I_0,I_1,I_*}(f).$$

Since p was chosen arbitrarily from among (I_0, I_1, I_*)-approximants of $(f \circ \mathrm{MP}^*_m)|_{\leqslant\theta}$ that have degree at most d', we conclude that

$$\begin{aligned}\deg_{I_0,I_1,I_*}((f \circ \mathrm{MP}_m)|_{\leqslant\theta}) &\geqslant \min\{cm \deg_{I_0,I_1,I_*}(f),\ d'+1\} \\ &\geqslant \min\left\{cm \deg_{I_0,I_1,I_*}(f),\ \frac{c\theta}{m\log(n+m)}\right\}. \qquad \square\end{aligned}$$

The previous composition theorem has the following analogue for Boolean inputs.

THEOREM 4.7. *Let $0 < c < 1$ be the absolute constant from Theorem 4.6. 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 $f\colon \{0,1\}^n \to \{0,1,*\}$ and all nonempty convex sets $I_0, I_1, I_* \subseteq \mathbb{R}$,*

$$\deg_{I_0,I_1,I_*}(f \circ H) \geqslant \min\left\{cm \deg_{I_0,I_1,I_*}(f), \frac{cN}{50m\log^2(n+m)} - n\right\}\log(n+m),$$
$$\deg_{I_0,I_1,I_*}(f \circ \neg H) \geqslant \min\left\{cm \deg_{I_0,I_1,I_*}(f), \frac{cN}{50m\log^2(n+m)} - n\right\}\log(n+m).$$

Proof. As in the previous proof, settling the first lower bound for all f will automatically settle the second lower bound, due to the invariance of (I_0, I_1, I_*)-approximate degree under negation of the input bits. In what follows, we focus on $f \circ H$.

We may assume that $N \geqslant 50mn\log^2(n+m)$ since otherwise the lower bounds in the theorem statement are nonpositive and hence trivially true. Define

$$\theta = \left\lceil \frac{N}{50\log(n+m)} \right\rceil.$$

Theorem 3.9 gives a surjection $G\colon \{0,1\}^{6\theta\lceil\log(nm+1)\rceil} \to \mathbb{N}^{nm}|_{\leqslant\theta}$ with the following two properties:

(i) for every coordinate $i = 1, 2, \ldots, nm$, the mapping $x \mapsto \mathrm{OR}^*_\theta(G(x)_i)$ is computable by an explicit DNF formula of size $(nm\theta)^{O(1)} = N^{O(1)}$ with bottom fan-in $O(\log(nm))$;

(ii) for any polynomial p, the map $v \mapsto \mathbf{E}_{G^{-1}(v)}\, p$ is a polynomial on $\mathbb{N}^{nm}|_{\leqslant\theta}$ of degree at most $(\deg p)/\lceil\log(nm+1)+1\rceil \leqslant (\deg p)/\log(n+m)$.

Consider the composition $F = (f \circ \mathrm{MP}^*_{m,\theta}) \circ G$. Then

$$\begin{aligned}F &= (f \circ (\mathrm{AND}_m \circ \mathrm{OR}^*_\theta)) \circ G \\ &= f \circ (\underbrace{(\mathrm{AND}_m \circ \mathrm{OR}^*_\theta, \ldots, \mathrm{AND}_m \circ \mathrm{OR}^*_\theta)}_{n} \circ G),\end{aligned}$$

which by property (i) of G means that F is the composition of f and an AND-OR-AND circuit H on $6\theta\lceil\log(nm+1)\rceil \leqslant N$ variables of size $(nmN)^{O(1)} = N^{O(1)}$ with bottom fan-in $O(\log(nm))$. Hence, the proof will be complete once we show that

$$\deg_{I_0,I_1,I_*}(F) \geqslant \min\left\{cm\deg_{I_0,I_1,I_*}(f), \frac{cN}{50m\log^2(n+m)} - n\right\}\log(n+m). \tag{4.69}$$

For this, fix an (I_0, I_1, I_*)-approximant p for F of degree $\deg_{I_0,I_1,I_*}(F)$. Consider the polynomial $p^*\colon \mathbb{N}^{nm}|_{\leqslant\theta} \to \mathbb{R}$ given by $p^*(v) = \mathbf{E}_{G^{-1}(v)}\, p$. Since I_0, I_1, I_* are convex and p is an (I_0, I_1, I_*)-approximant for $F = (f \circ \mathrm{MP}^*_{m,\theta}) \circ G$, it follows that p^* is an (I_0, I_1, I_*)-approximant for $(f \circ \mathrm{MP}^*_{m,\theta})|_{\leqslant\theta}$. Therefore,

$$\begin{aligned}
\deg p^* &\geqslant \deg_{I_0,I_1,I_*}((f \circ \mathrm{MP}^*_{m,\theta})|_{\leqslant\theta}) \\
&\geqslant \deg_{I_0,I_1,I_*}((f \circ \mathrm{MP}^*_{m})|_{\leqslant\theta}) \\
&\geqslant \min\left\{cm\deg_{I_0,I_1,I_*}(f), \frac{c\theta}{m\log(n+m)} - n\right\} \\
&\geqslant \min\left\{cm\deg_{I_0,I_1,I_*}(f), \frac{cN}{50m\log^2(n+m)} - n\right\},
\end{aligned}$$

where the second step is valid because $\mathrm{MP}^*_{m,\theta}$ contains $\mathrm{MP}^*_m = \mathrm{MP}^*_{m,m^2}$ as a subfunction, and the third step is legitimate by Theorem 4.6. However, property (ii) of G states that

$$\begin{aligned}
\deg p^* &\leqslant \frac{\deg p}{\log(n+m)} \\
&= \frac{\deg_{I_0,I_1,I_*}(F)}{\log(n+m)}.
\end{aligned}$$

Comparing these lower and upper bounds on the degree of p^* settles (4.69). □

At last, we illustrate the use of the previous two composition results to amplify hardness for polynomial approximation.

THEOREM 4.8 (Hardness amplification). *Let $I_0, I_1, I_* \subseteq \mathbb{R}$ be any nonempty convex subsets. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a given function with*

$$\deg_{I_0,I_1,I_*}(f) \geqslant n^{1-\frac{1}{k}},$$

for some real number $k \geqslant 1$. Suppose further that f is computable by a Boolean circuit of size s and depth d, where $d \geqslant 1$. Then there is a function $F\colon \{0,1\}^N \to \{0,1\}$ on $N = \Theta(n^{1+\frac{1}{k}}\log^2 n)$ variables with

$$\deg_{I_0,I_1,I_*}(F) \geqslant \Omega\left(\frac{N^{1-\frac{1}{k+1}}}{\log^{1-\frac{2}{k+1}} N}\right).$$

Moreover, F is computable by a Boolean circuit of size $s + n^{O(1)}$, bottom fan-in $O(\log n)$, depth $d+2$ if the circuit for f is monotone, and depth $d+3$ otherwise.

Proof. Take

$$\begin{aligned} m &= \lceil n^{1/k} \rceil, \\ N &= \left\lceil \frac{100}{c} mn \log^2(n+m) \right\rceil, \end{aligned}$$

where $0 < c < 1$ is the absolute constant from Theorem 4.6. Then Theorem 4.7 gives an explicit transformation $H\colon \{0,1\}^N \to \{0,1\}^n$, computable by an AND-OR-AND circuit of size $n^{O(1)}$ with bottom fan-in $O(\log n)$, such that

$$\begin{aligned} &\min\{\deg_{I_0,I_1,I_*}(f \circ H), \deg_{I_0,I_1,I_*}(f \circ \neg H)\} \\ &\qquad \geqslant \min\left\{ cm \deg_{I_0,I_1,I_*}(f), \frac{cN}{50m \log^2(n+m)} - n \right\} \log(n+m) \\ &\qquad \geqslant cn \log n \\ &\qquad = \Theta\left(\frac{N^{1-\frac{1}{k+1}}}{\log^{1-\frac{2}{k+1}} N} \right). \end{aligned}$$

Now, fix a circuit for f of size s and depth $d \geqslant 1$. Composing the circuits for f and H results in circuits for $f \circ H$ and $f \circ \neg H$ of size $s + n^{O(1)}$, bottom fan-in $O(\log n)$, and depth at most $d+3$. Thus, F can be taken to be either of $f \circ H$ and $f \circ \neg H$.

When the circuit for f is monotone, the depth of F can be reduced to $d+2$ as follows. After merging like gates if necessary, the circuit for f can be viewed as composed of d layers of alternating gates ($\wedge$ and $\vee$). The bottom layer of f can therefore be merged with the top layer of either H or $\neg H$, resulting in a circuit of depth at most $d + 3 - 1 = 2$. □

We emphasize that in view of (4.56), the symbol $\deg_{I_0,I_1,I_*}$ in Theorems 4.6–4.8 can be replaced with the threshold degree symbol $\deg_\pm$. The same goes for any other special case of (I_0, I_1, I_*)-approximate degree.

4.4. Threshold degree of surjectivity. We start with the simplest application of our amplification theorem, in which the outer function f is the identity map $f\colon \{0,1\} \to \{0,1\}$ on a single bit.

THEOREM 4.9. *For any integer $m \geqslant 1$,*

$$\deg_\pm(\mathrm{MP}^*_m|_{\leqslant m^2 \log m}) = \Omega(m).$$

Proof. Let $f\colon \{0,1\} \to \{0,1\}$ be the identity function, so that $\deg_\pm(f) = 1$. Invoking Theorem 4.6 with $n = 1$ and $\theta = \lfloor m^2 \log m \rfloor$, one obtains the claimed lower bound. □

Theorem 4.9 has a useful interpretation. For positive integers n and r, the *surjectivity problem* is the problem of determining whether a given mapping $\{1,2,\ldots,n\} \to \{1,2,\ldots,r\}$ is surjective. This problem is trivial for $r > n$, and the standard regime studied in previous work is $r \leqslant cn$ for some constant $0 < c < 1$. The input to the surjectivity problem is represented by a Boolean matrix $x \in \{0,1\}^{r \times n}$ with precisely

one nonzero entry in every column. More formally, let $e_1, e_2, \ldots, e_r$ be the standard basis for $\mathbb{R}^n$. The surjectivity function $\mathrm{SURJ}_{n,r} \colon \{e_1, e_2, \ldots, e_r\}^n \to \{0,1\}$ is given by

$$\mathrm{SURJ}_{n,r}(x_1, x_2, \ldots, x_n) = \bigwedge_{j=1}^{r} \bigvee_{i=1}^{n} x_{i,j}.$$

It is clear that $\mathrm{SURJ}_{n,r}(x_1, x_2, \ldots, x_n)$ is uniquely determined by the vector sum $x_1 + x_2 + \cdots + x_n \in \mathbb{N}^r|_n$. It is therefore natural to consider a symmetric counterpart of the surjectivity function, with domain $\mathbb{N}^r|_n$ instead of $\{e_1, e_2, \ldots, e_r\}^n$. This symmetric version is $(\mathrm{AND}_r \circ \mathrm{OR}_n^*)|_n = \mathrm{MP}_{r,n}^*|_n$, and Proposition 2.11 ensures that

$$\deg_\pm(\mathrm{SURJ}_{n,r}) = \deg_\pm(\mathrm{MP}_{r,n}^*|_n). \tag{4.70}$$

The surjectivity problem has seen much work recently [9, 57, 13, 19]. In particular, Bun and Thaler [19] have obtained an essentially tight lower bound of $\tilde{\Omega}(\min\{r, \sqrt{n/\log n}\})$ on the threshold degree of $\mathrm{SURJ}_{n,r}$ in the standard regime $r \leqslant (1 - \Omega(1))n$. As a corollary to Theorem 4.9, we give a new proof of Bun and Thaler's result, sharpening their bound by a polylogarithmic factor.

Corollary 4.10. *For any integers $n > r \geqslant 1$,*

$$\deg_\pm(\mathrm{SURJ}_{n,r}) \geqslant \Omega\left(\min\left\{r, \sqrt{\frac{n-r}{1+\log(n-r)}}\right\}\right). \tag{4.71}$$

Proof. Define

$$r' = \min\left\{r - 1, \left\lfloor \sqrt{\frac{n-r}{1+\log(n-r)}} \right\rfloor\right\}. \tag{4.72}$$

We may assume that $r' \geqslant 1$ since (4.71) holds trivially otherwise. The identity

$$\begin{aligned}&\mathrm{MP}_{r',n}^*(x_1, x_2, \ldots, x_{r'}) \\ &\qquad = \mathrm{MP}_{r,n}^*\left(x_1, x_2, \ldots, x_{r'}, \underbrace{1, 1, \ldots, 1}_{r-r'-1}, 1 + n - (r - r') - \sum_{i=1}^{r'} x_i\right)\end{aligned}$$

holds for all $(x_1, x_2, \ldots, x_{r'}) \in \mathbb{N}^{r'}|_{\leqslant n-(r-r')}$, whence

$$\deg_\pm(\mathrm{MP}_{r',n}^*|_{\leqslant n-(r-r')}) \leqslant \deg_\pm(\mathrm{MP}_{r,n}^*|_n). \tag{4.73}$$

Now

$$\begin{aligned}\deg_{\pm}(\mathrm{SURJ}_{n,r}) &= \deg_{\pm}(\mathrm{MP}^{*}_{r,n}|_{n}) \\ &\geqslant \deg_{\pm}(\mathrm{MP}^{*}_{r',n}|_{\leqslant n-(r-r')}) \\ &\geqslant \deg_{\pm}(\mathrm{MP}^{*}_{r',r'^{2}}|_{\leqslant r'^{2}\log r'}) \\ &\geqslant \Omega(r'),\end{aligned}$$

where the four steps use (4.70), (4.73), (4.72), and Theorem 4.9, respectively. □

4.5. Threshold degree and discrepancy of $\mathbf{AC^0}$. We now turn to 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 4.3.

THEOREM 4.11. *Let $k \geqslant 1$ be a fixed integer. 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*

$$\deg_{\pm}(f_{k,n}) = \Omega\left(n^{\frac{k-1}{k+1}} \cdot (\log n)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right) \tag{4.74}$$

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

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

For the former, the threshold degree lower bound (4.74) is trivial. For the latter, it follows from Theorem 2.5.

For the inductive step, fix $k \geqslant 3$. Due to the asymptotic nature of (4.74), 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

$$\deg_{\pm}(f_{k-2,n}) = \Omega\left(n^{\frac{k-3}{k-1}} \cdot (\log n)^{-\frac{1}{k-1}\lceil\frac{k-4}{2}\rceil\lfloor\frac{k-4}{2}\rfloor}\right) \tag{4.75}$$

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 4.6. For every N larger that a certain constant, we apply Theorem 4.7 with

$$n = \left\lceil N^{\frac{k-1}{k+1}} (\log N)^{-\frac{1}{k+1}\lceil\frac{k-4}{2}\rceil\lfloor\frac{k-4}{2}\rfloor - \frac{2(k-1)}{k+1}} \cdot \frac{c}{100} \right\rceil, \tag{4.76}$$

$$m = \left\lceil N^{\frac{2}{k+1}} (\log N)^{\frac{1}{k+1}\lceil\frac{k-4}{2}\rceil\lfloor\frac{k-4}{2}\rfloor - \frac{4}{k+1}} \right\rceil, \tag{4.77}$$

$$f = f_{k-2,n}, \tag{4.78}$$

$$I_0 = (0, \infty), \tag{4.79}$$

$$I_1 = (-\infty, 0), \tag{4.80}$$

$$I_* = (-\infty, \infty) \tag{4.81}$$

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

$$\begin{aligned}
\deg_\pm(F_N) &\geqslant \min\left\{cm \deg_\pm(f_{k-2,n}), \frac{cN}{50m\log^2(n+m)} - n\right\} \log(n+m) \\
&= \Theta\left(N^{\frac{k-1}{k+1}} (\log N)^{-\frac{1}{k+1}\lceil\frac{k-4}{2}\rceil\lfloor\frac{k-4}{2}\rfloor - \frac{k-3}{k+1}}\right) \\
&= \Theta\left(N^{\frac{k-1}{k+1}} (\log N)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right),
\end{aligned} \tag{4.82}$$

where the second step uses (4.75)–(4.77). Moreover, Theorem 4.7 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 produce 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 (4.82) 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 computes F_N as a subfunction and therefore has threshold degree at least that of F_N. This completes the inductive step. □

Using the pattern matrix method, we now lift the previous theorem to multiparty communication complexity.

THEOREM 4.12. *Let $k \geqslant 3$ be a fixed integer. Let $\ell \colon \mathbb{N} \to \mathbb{N}$ be a given function. Then there is an (explicitly given) family $\{F_n\}_{n=1}^\infty$, where $F_n \colon (\{0,1\}^n)^{\ell(n)} \to \{0,1\}$ is an $\ell(n)$-party communication problem with discrepancy*

$$\operatorname{disc}(F_n) \leqslant 2\exp\left(-\Omega\left(\left(\frac{n}{4^{\ell(n)}\ell(n)^2}\right)^{\frac{k-1}{k+1}} \cdot (\log n)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right)\right) \tag{4.83}$$

and communication complexity

$$\mathrm{PP}(F_n) = \Omega\left(\left(\frac{n}{4^{\ell(n)}\ell(n)^2}\right)^{\frac{k-1}{k+1}} \cdot (\log n)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right). \tag{4.84}$$

Moreover, F_n is computable by a Boolean circuit of polynomial size and depth $k+2$ in which the bottom three layers have fan-in $O(\log n)$, $O(4^{\ell(n)}\ell(n)^2)$, and $\ell(n)$, in that order. In particular, if $\ell(n) = O(1)$, then F_n is computable by a Boolean circuit of polynomial size, depth k, and bottom fan-in $O(\log n)$.

Proof. Theorem 4.11 constructs a family of functions $\{f_n\}_{n=1}^{\infty}$, where $f_n \colon \{0,1\}^n \to \{0,1\}$ has threshold degree

$$\deg_{\pm}(f_n) = \Omega\left(n^{\frac{k-1}{k+1}} \cdot (\log n)^{-\frac{1}{k+1}\lceil\frac{k-2}{2}\rceil\lfloor\frac{k-2}{2}\rfloor}\right) \tag{4.85}$$

and is computable by a Boolean circuit of polynomial size, depth k, and bottom fan-in $O(\log n)$. Now, let $c > 0$ be the absolute constant from Theorem 2.16. For any given n, define

$$F_n = \begin{cases} \mathrm{AND}_{\ell(n)} & \text{if } n \leqslant 2m, \\ f_{\lfloor n/m \rfloor} \circ \mathrm{NOR}_m \circ \mathrm{AND}_{\ell(n)} & \text{otherwise,} \end{cases}$$

where $m = 2\lceil c4^{\ell(n)}\ell(n)^2\rceil$. Then the discrepancy bound (2.16) is trivial for $n \leqslant 2m$, and follows from (4.85) and Theorem 2.16 for $n > 2m$. The lower bound (4.84) on the communication complexity of F_n with weakly unbounded error is now immediate by the discrepancy method (Corollary 2.14).

It remains to examine the circuit complexity of F_n. Since f_n is computable by a circuit of polynomial size, depth k, and bottom fan-in $O(\log n)$, it follows that F_n is computable by a circuit of polynomial size and depth $k+2$ in which the bottom three levels have fan-in $O(\log n)$, $O(4^{\ell(n)}\ell(n)^2)$, and $\ell(n)$, in that order. This means that for $\ell(n) = O(1)$, any gate of the bottom four levels can be computed by a circuit of polynomial size, depth 2, and bottom fan-in $O(\log n)$, which in turn yields a circuit for F_n of polynomial size, depth $(k+2)-4+2 = k$, and bottom fan-in $O(\log n)$. □

Theorems 4.11 and 4.12 settle Theorems 1.1 and 1.4, respectively, from the introduction.

5. The sign-rank of AC^0

We now turn to the second main result of this paper, a near-linear lower bound on the sign-rank of constant-depth circuits. To start with, we show that our smoothing technique from Theorem 4.4 already gives an exponential lower bound on the sign-rank of AC^0. Specifically, we prove in Section 5.1 that the Minsky–Papert function $\mathrm{MP}_{n^{1/3}}$ has $\exp(-O(n^{1/3}))$-smooth threshold degree $\Omega(n^{1/3})$, which by Theorem 2.17 immediately implies an $\exp(\Omega(n^{1/3}))$ lower bound on the sign-rank of an AC^0 circuit family of depth 3. This result was originally obtained, with a longer and more demanding proof, by Razborov and Sherstov [42].

To obtain a near-optimal lower bound of $\exp(\Omega(n^{1-\epsilon}))$, we use a completely different approach. It is based on the notion of *local smoothness* and is unrelated to the threshold degree analysis. In Section 5.2, we define local smoothness and record

basic properties of locally smooth functions. In Sections 5.3 and 5.4, 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 in Section 5.5 a locally smooth dual polynomial for the Minsky–Papert function. We use this dual object in Section 5.6 to prove an amplification theorem for *smooth* threshold degree. We apply the amplification theorem iteratively in Section 5.7 to construct, for any $\epsilon > 0$, a constant-depth circuit with $\exp(-n^{1-\epsilon})$-smooth threshold degree $\Omega(n^{1-\epsilon})$. Finally, we present our main result on the sign-rank of AC^0 in Section 5.8.

In the remainder of this section, we adopt the following additional notation. For an arbitrary subset X of Euclidean space, we write $\operatorname{diam} X = \sup_{x,x' \in X} |x - x'|$, with the convention that $\operatorname{diam} \varnothing = 0$. For a vector $x \in \mathbb{Z}^n$ and a natural number d, we let $B_d(x) = \{v \in \mathbb{Z}^n : |x - v| \leqslant d\}$ denote the set of *integer-valued* vectors within distance d of x. For all x,

$$|B_d(x)| = |B_d(0)| \leqslant 2^d \binom{n+d}{d}, \tag{5.1}$$

where the binomial coefficient corresponds to the number of *nonnegative* integer vectors of weight at most d. Finally, for vectors $u, v \in \mathbb{N}^n$, we define $\operatorname{cube}(u, v)$ to be the smallest Cartesian product of integer intervals that contains both u and v. Specifically,

$$\begin{aligned}\operatorname{cube}(u,v) &= \{w \in \mathbb{N}^n : \min\{u_i, v_i\} \leqslant w_i \leqslant \max\{u_i, v_i\} \text{ for all } i\} \\ &= \prod_{i=1}^{n} \{\min\{u_i, v_i\}, \min\{u_i, v_i\} + 1, \ldots, \max\{u_i, v_i\}\}.\end{aligned}$$

5.1. A simple lower bound for depth 3. We start by presenting a new proof of Razborov and Sherstov's exponential lower bound [42] on the sign-rank of AC^0. More precisely, we prove the following stronger result that was not known before.

THEOREM 5.1. *There is a constant $0 < c < 1$ such that for all positive integers m and r,*

$$\deg_{\pm}(\mathrm{MP}_{m,r}, 12^{-m-1}) \geqslant \min\{m, c\sqrt{r}\}.$$

Theorem 5.1 is asymptotically optimal, and it is the first lower bound on the smooth threshold degree of the Minsky–Papert function. As we will discuss shortly, this theorem implies an $\exp(\Omega(n^{1/3}))$ lower bound on the sign-rank of AC^0. In addition, we will use Theorem 5.1 as the base case in the inductive proof of Theorem 1.3.

Proof of Theorem 5.1. It is well-known [36, 39, 60] that for some constant $c > 0$ and all r, any real polynomial $p\colon \{0,1\}^r \to \mathbb{R}$ with $\|p - \mathrm{OR}_r\|_\infty \leqslant 0.49$ has degree at least $c\sqrt{r}$. By linear programming duality [53, Theorem 2.5], this approximation-theoretic fact is equivalent to the existence of a function $\psi\colon \{0,1\}^m \to \mathbb{R}$ with

$$\psi(0) > 0.49, \tag{5.2}$$

$$\|\psi\|_1 = 1, \tag{5.3}$$

$$\operatorname{orth} \psi \geqslant c\sqrt{r}. \tag{5.4}$$

The rest of the proof is a reprise of Section 4.2. To begin with, property (5.3) makes it possible to view $|\psi|$ as a probability distribution on $\{0,1\}^r$. Let μ_0, μ_1, μ_2 be the probability distributions induced by $|\psi|$ on the sets $\{0^r\}, \{x \neq 0^r : \psi(x) < 0\}$, and $\{x \neq 0^r : \psi(x) > 0\}$, respectively. It is clear from (5.2) that the negative part of ψ is a multiple of μ_1, whereas the positive part of ψ is a nonnegative linear combination of μ_0 and μ_2. Moreover, it follows from $\langle \psi, 1 \rangle = 0$ and $\|\psi\|_1 = 1$ that the positive and negative parts of ψ both have ℓ_1-norm $1/2$. Summarizing,

$$\psi = \frac{1-\delta}{2}\mu_0 - \frac{1}{2}\mu_1 + \frac{\delta}{2}\mu_2 \tag{5.5}$$

for some $0 \leqslant \delta \leqslant 1$. In view of (5.2), we infer the more precise bound

$$0 \leqslant \delta < \frac{1}{50}. \tag{5.6}$$

Let υ be the uniform probability distribution on $\{0,1\}^r \setminus \{0^r\}$. We define

$$\lambda_0 = \mu_0, \tag{5.7}$$

$$\lambda_1 = \frac{2}{3(1-\delta)}\mu_1 + \left(1 - \frac{2}{3(1-\delta)}\right)\upsilon, \tag{5.8}$$

$$\lambda_2 = \frac{2\delta}{1-\delta}\mu_2 + \left(1 - \frac{2\delta}{1-\delta}\right)\upsilon. \tag{5.9}$$

It is clear from (5.6) that λ_1 and λ_2 are convex combinations of υ, μ_1, μ_2 and therefore are probability distributions with support

$$\operatorname{supp}\lambda_i \subseteq \{0,1\}^r \setminus \{0^r\}, \qquad i = 1, 2, \tag{5.10}$$

whereas

$$\operatorname{supp}\lambda_0 = \{0^r\} \tag{5.11}$$

by definition. Moreover, (5.6) implies that

$$\lambda_i \geqslant \frac{1}{4}\upsilon, \qquad i = 1, 2. \tag{5.12}$$

The defining equations (5.7)–(5.9) further imply that

$$\frac{2}{3}\lambda_0 + \frac{1}{3}\lambda_2 - \lambda_1 = \frac{4}{3(1-\delta)}\psi,$$

which along with (5.4) gives

$$\operatorname{orth}\left(\frac{2}{3}\lambda_0 + \frac{1}{3}\lambda_2 - \lambda_1\right) \geqslant c\sqrt{r}. \tag{5.13}$$

With this work behind us, define

$$\Lambda = \frac{1}{2}\left(\frac{2}{3}\lambda_0 + \frac{1}{3}\lambda_2\right)^{\otimes m} - \frac{1}{2}\left(-\frac{1}{3}\lambda_0 + \frac{1}{3}\lambda_2\right)^{\otimes m} + \frac{1}{2}\lambda_1^{\otimes m}.$$

Multiplying out the tensor products in the definition of Λ and collecting like terms, we obtain

$$\begin{aligned}
\Lambda &= \frac{1}{2}\sum_{\substack{S \subseteq \{1,2,\ldots,m\} \\ S \neq \varnothing}} \frac{2^{|S|} - (-1)^{|S|}}{3^m} \lambda_0^{\otimes S} \cdot \lambda_2^{\otimes \overline{S}} + \frac{1}{2}\lambda_1^{\otimes m} && (5.14)\\
&\geqslant \frac{1}{4}\sum_{\substack{S \subseteq \{1,2,\ldots,m\} \\ S \neq \varnothing}} \frac{2^{|S|}}{3^m} \lambda_0^{\otimes S} \cdot \lambda_2^{\otimes \overline{S}} + \frac{1}{2}\lambda_1^{\otimes m} \\
&\geqslant \frac{1}{4}\sum_{\substack{S \subseteq \{1,2,\ldots,m\} \\ S \neq \varnothing}} \frac{2^{|S|}}{3^m} \lambda_0^{\otimes S} \cdot \left(\frac{1}{4}\upsilon\right)^{\otimes \overline{S}} + \frac{1}{2}\left(\frac{1}{4}\upsilon\right)^{\otimes m} \\
&\geqslant \frac{1}{4}\sum_{S \subseteq \{1,2,\ldots,m\}} \frac{2^{|S|}}{3^m} \lambda_0^{\otimes S} \cdot \left(\frac{1}{4}\upsilon\right)^{\otimes \overline{S}} \\
&= \frac{1}{4}\left(\frac{2}{3}\lambda_0 + \frac{1}{3}\cdot\frac{1}{4}\upsilon\right)^{\otimes m} \\
&\geqslant \frac{1}{4}\left(\frac{1}{12 \cdot 2^r}\right)^m \mathbf{1}_{(\{0,1\}^r)^m}, && (5.15)
\end{aligned}$$

where the third step uses (5.12). In particular, Λ is a nonnegative function. We further calculate

$$\begin{aligned}
\langle \Lambda, 1 \rangle &= \frac{1}{2}\left\langle \frac{2}{3}\lambda_0 + \frac{1}{3}\lambda_2, 1 \right\rangle^m - \frac{1}{2}\left\langle -\frac{1}{3}\lambda_0 + \frac{1}{3}\lambda_2, 1 \right\rangle^m + \frac{1}{2}\langle \lambda_1, 1 \rangle^m \\
&= \frac{1}{2}\left\langle \frac{2}{3}\lambda_0 + \frac{1}{3}\lambda_2, 1 \right\rangle^m + \frac{1}{2}\langle \lambda_1, 1 \rangle^m \\
&= \frac{1}{2} + \frac{1}{2} \\
&= 1, && (5.16)
\end{aligned}$$

which makes Λ a probability distribution on $(\{0,1\}^r)^m$.

It remains to examine the orthogonal content of $\Lambda \cdot (-1)^{\mathrm{MP}_{m,r}}$. We have

$$\begin{aligned}
\Lambda \cdot (-1)^{\mathrm{MP}_{m,r}} &= \frac{1}{2} \sum_{\substack{S \subseteq \{1,2,\ldots,m\} \\ S \neq \varnothing}} \frac{2^{|S|} - (-1)^{|S|}}{3^m} \lambda_0^{\otimes S} \cdot \lambda_2^{\otimes \overline{S}} \cdot (-1)^{\mathrm{MP}_{m,r}} \\
&\qquad + \frac{1}{2} \lambda_1^{\otimes m} \cdot (-1)^{\mathrm{MP}_{m,r}} \\
&= \frac{1}{2} \sum_{\substack{S \subseteq \{1,2,\ldots,m\} \\ S \neq \varnothing}} \frac{2^{|S|} - (-1)^{|S|}}{3^m} \lambda_0^{\otimes S} \cdot \lambda_2^{\otimes \overline{S}} - \frac{1}{2} \lambda_1^{\otimes m} \\
&= \frac{1}{2} \left(\frac{2}{3} \lambda_0 + \frac{1}{3} \lambda_2 \right)^{\otimes m} - \frac{1}{2} \left(-\frac{1}{3} \lambda_0 + \frac{1}{3} \lambda_2 \right)^{\otimes m} - \frac{1}{2} \lambda_1^{\otimes m},
\end{aligned}$$

where the first step uses (5.14); the second step uses (5.10) and (5.11); and the final equality can be verified by multiplying out the tensor powers and collecting like terms. Now

$$\begin{aligned}
&\operatorname{orth}(\Lambda \cdot (-1)^{\mathrm{MP}_{m,r}}) \\
&\qquad = \min \left\{ \operatorname{orth} \left(\frac{1}{2} \left(\frac{2}{3} \lambda_0 + \frac{1}{3} \lambda_2 \right)^{\otimes m} - \frac{1}{2} \lambda_1^{\otimes m} \right), \right. \\
&\qquad\qquad \left. \operatorname{orth} \left(-\frac{1}{2} \left(-\frac{1}{3} \lambda_0 + \frac{1}{3} \lambda_2 \right)^{\otimes m} \right) \right\} \\
&\qquad \geqslant \min \left\{ \operatorname{orth} \left(\frac{2}{3} \lambda_0 + \frac{1}{3} \lambda_2 - \lambda_1 \right), m \operatorname{orth} \left(-\frac{1}{3} \lambda_0 + \frac{1}{3} \lambda_2 \right) \right\} \\
&\qquad \geqslant \min \left\{ c\sqrt{r}, m \operatorname{orth} \left(-\frac{1}{3} \lambda_0 + \frac{1}{3} \lambda_2 \right) \right\} \\
&\qquad \geqslant \min\{c\sqrt{r}, m\},
\end{aligned}$$

where the first step applies Proposition 2.1(i); the second step applies Proposition 2.1(ii), (iii); the third step substitutes the lower bound from (5.13); and the last step uses $\langle -\lambda_0 + \lambda_2, 1 \rangle = -\langle \lambda_0, 1 \rangle + \langle \lambda_2, 1 \rangle = -1 + 1 = 0$. Combining this conclusion with (5.15) and (5.16) completes the proof. □

We now lift the approximation-theoretic result just obtained to a sign-rank lower bound, reproving a result of Razborov and Sherstov [42].

THEOREM 5.2 (Razborov and Sherstov). *Define* $F_n \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *by*

$$F_n = \mathrm{AND}_{n^{1/3}} \circ \mathrm{OR}_{n^{2/3}} \circ \mathrm{AND}_2.$$

Then

$$\mathrm{rk}_{\pm}(F_n) \geqslant 2^{\Omega(n^{1/3})}.$$

Proof. Theorem 5.1 states that

$$\deg_{\pm}(\mathrm{AND}_{n^{1/3}} \circ \mathrm{OR}_{n^{2/3}}, \exp(-c'n^{1/3})) \geqslant c''n^{1/3}$$

for some absolute constants $c', c'' > 0$ and all n. This lower bound along with Theorem 2.17 implies that the composition

$$H_n = \mathrm{AND}_{n^{1/3}} \circ \mathrm{OR}_{n^{2/3}} \circ \mathrm{OR}_{2^{\lceil \exp(\frac{4c'}{c''}) \rceil}} \circ \mathrm{AND}_2$$

has sign-rank $\mathrm{rk}_{\pm}(H_n) = \exp(\Omega(n^{1/3}))$. This completes the proof because for some integer constant $c \geqslant 1$, each H_n is a subfunction of F_{cn}. □

5.2. Local smoothness. The remainder of this paper focuses on our $\exp(\Omega(n^{1-\epsilon}))$ lower bound on the sign-rank of AC^0, whose proof is unrelated to the work in Section 5.1. Central to our approach is an analytic notion that 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

$$|\Phi(x)| \leqslant K^{|x-x'|}|\Phi(x')| \quad \text{for all } x, x' \in X.$$

Put another way, for any two points of X at distance d, the corresponding values of Φ differ in magnitude by a factor of at most K^d. For any set X, we let $\mathrm{Smooth}(K, X)$ denote the family of functions that are smooth on X. The following proposition collects basic properties of local smoothness, to which we refer as the restriction property, scaling property, tensor property, and conical property.

PROPOSITION 5.3. *Let $K \geqslant 1$ be given.*

(i) *If $\Phi \in \mathrm{Smooth}(K, X)$ and $X' \subseteq X$, then $\Phi \in \mathrm{Smooth}(K, X')$.*
(ii) *If $\Phi \in \mathrm{Smooth}(K, X)$ and $a \in \mathbb{R}$, then $a\Phi \in \mathrm{Smooth}(K, X)$.*
(iii) $\mathrm{Smooth}(K, X) \otimes \mathrm{Smooth}(K, Y) \subseteq \mathrm{Smooth}(K, X \times Y)$.
(iv) *If $\Phi, \Psi \in \mathrm{Smooth}(K, X)$ and Φ, Ψ are nonnegative on X, then* $\mathrm{cone}\{\Phi, \Psi\} \subseteq \mathrm{Smooth}(K, X)$.

Proof. Properties (i) and (ii) are immediate from the definition of K-smoothness. For (iii), fix $(x, y), (x', y') \in X \times Y$ arbitrarily. Then

$$\begin{aligned} |\Phi(x)\Psi(y)| &\leqslant K^{|x-x'|}|\Phi(x')|\, K^{|y-y'|}|\Psi(y')| \\ &= K^{|(x,y)-(x',y')|}|\Phi(x')\Psi(y')|, \end{aligned}$$

where the first step uses the K-smoothness of Φ and Ψ. Finally, for (iv), let a and b be nonnegative reals. Then

$$\begin{aligned} |a\Phi(x) + b\Psi(x)| &= a|\Phi(x)| + b|\Psi(x)| \\ &\leqslant aK^{|x-x'|}|\Phi(x')| + bK^{|x-x'|}|\Psi(x')| \\ &= K^{|x-x'|}|a\Phi(x') + b\Psi(x')| \end{aligned}$$

for all $x, x' \in X$, where the second step uses the K-smoothness of Φ and Ψ. □

We will take a special interest in locally smooth functions that are probability distributions. For our purposes, it will be sufficient to consider locally smooth distributions whose support is the Cartesian product of integer intervals. By way of notation, for an integer $n \geqslant 1$ and a real number $K \geqslant 1$, we let $\mathfrak{S}(n, K)$ denote the set of probability distributions Λ such that:

(i) Λ is supported on $\prod_{i=1}^{n} \{0, 1, 2, \ldots, r_i\}$, for some $r_1, r_2, \ldots, r_n \in \mathbb{N}$;
(ii) Λ is K-smooth on its support.

Analogous to the development in Section 4.1, it will be helpful to have notation for translates of distributions in $\mathfrak{S}(n, K)$. For $\Delta \geqslant 0$, we let $\mathfrak{S}(n, K, \Delta)$ denote the set of probability distributions $\Lambda \in \mathfrak{D}(\mathbb{N}^n)$ such that $\Lambda(t_1, \ldots, t_n) \equiv \Lambda'(t_1 - a_1, \ldots, t_n - a_n)$ for some fixed $\Lambda' \in \mathfrak{S}(n, K)$ and $a \in \mathbb{N}^n|_{\leqslant \Delta}$. As a special case, $\mathfrak{S}(n, K, 0) = \mathfrak{S}(n, K)$. Specializing Proposition 5.3(iii) to this context, we obtain:

PROPOSITION 5.4. *For any $n', n'', \Delta', \Delta'', K$, one has*

$$\mathfrak{S}(n', K, \Delta') \otimes \mathfrak{S}(n'', K, \Delta'') \subseteq \mathfrak{S}(n' + n', K, \Delta' + \Delta'').$$

Proof. The only nontrivial property to verify is K-smoothness, which follows from Proposition 5.3(iii). □

5.3. Metric properties of locally smooth distributions. If Λ is a locally smooth distribution on $X = \prod_{i=1}^{n} \{0, 1, 2, \ldots, r_i\}$, then a moment's thought reveals that $\Lambda(x) > 0$ at every point $x \in X$. In general, local smoothness provides one with considerable control of the metric behavior of X, making it possible to prove nontrivial upper and lower bounds on $\Lambda(S)$ for various sets $S \subseteq X$. We now record two such results, as regards our work on the sign-rank on AC^0.

PROPOSITION 5.5. *Let Λ be a probability distribution on $X = \prod_{i=1}^{n} \{0, 1, 2, \ldots, r_i\}$. Let θ and d be nonnegative integers with $\theta \geqslant d$. If Λ is K-smooth on $X|_{\leqslant \theta}$, then*

$$\Lambda(X|_{\leqslant \theta}) \leqslant K^d \binom{n+d}{d} \Lambda(X|_{\leqslant \theta - d}).$$

Proof. Consider an arbitrary vector $x \in X|_{\leqslant \theta}$. By definition, the components of x are nonnegative integers that sum to at most θ. By decreasing the components of x as needed, one can obtain a vector x' with

$$x' \in X|_{\leqslant \theta - d},$$
$$x' \leqslant x,$$
$$|x' - x| \leqslant d.$$

In particular, the K-smoothness of Λ implies that

$$\Lambda(x) \leqslant K^d \Lambda(x').$$

Summing on both sides over $x \in X|_{\leqslant\theta}$, we obtain

$$\begin{aligned}\Lambda(X|_{\leqslant\theta}) &\leqslant K^d \Lambda(X|_{\leqslant\theta-d}) \max_{x' \in X|_{\leqslant\theta-d}} |\{x \in X|_{\leqslant\theta} : x \geqslant x' \text{ and } |x-x'| \leqslant d\}| \\ &\leqslant K^d \Lambda(X|_{\leqslant\theta-d}) \max_{x' \in \mathbb{N}^n} |\{x \in \mathbb{N}^n : x \geqslant x' \text{ and } |x-x'| \leqslant d\}| \\ &= K^d \Lambda(X|_{\leqslant\theta-d}) \binom{n+d}{d}. \qquad \square\end{aligned}$$

PROPOSITION 5.6. *Let Λ be a probability distribution on $X = \prod_{i=1}^n \{0, 1, 2, \ldots, r_i\}$. Let θ and d be nonnegative integers with*

$$d < \frac{1}{2} \min\left\{\theta, \sum_{i=1}^n r_i\right\}. \tag{5.17}$$

If Λ is K-smooth on $X|_{\leqslant\theta}$, then

$$\Lambda(X|_{\leqslant\theta}) \leqslant 2^{d+1} K^{2d+1} \binom{n+d}{d} \Lambda(X|_{\leqslant\theta} \setminus B_d(u))$$

for every $u \in X$.

Proof. Fix $u \in X$ for the rest of the proof. If $|u| > \theta + d$, then $X|_{\leqslant\theta} \setminus B_d(u) = X|_{\leqslant\theta}$ and the statement holds trivially. In what follows, we treat the complementary case $|u| \leqslant \theta + d$. Here, the key is to find a vector u' with

$$|u - u'| = d + 1, \tag{5.18}$$

$$u' \in X|_{\leqslant\theta}. \tag{5.19}$$

The algorithm for finding u' depends on $|u|$, as follows.

(i) If $|u| > d$, decrease one or more of the components of u as needed to obtain a vector u' whose components are nonnegative integers that sum to exactly $|u| - d - 1$. Then (5.18) is immediate, whereas (5.19) follows in view of $|u| \leqslant \theta + d$.

(ii) If $|u| \leqslant d$, the analysis is more subtle. Recall that $u \in \prod_{i=1}^n \{0, 1, 2, \ldots, r_i\}$ and therefore $|(r_1, \ldots, r_n) - u| = \sum r_i - |u| \geqslant \sum r_i - d > d$, where the last step uses (5.17). As a result, by increasing the components of u as necessary, one can obtain a vector $u' \in \prod_{i=1}^n \{0, 1, 2, \ldots, r_i\}$ with $|u'| = |u| + d + 1$. Then property (5.18) is immediate. Property (5.19) follows from $|u'| = |u| + d + 1 \leqslant 2d + 1 < \theta + 1$, where the last step uses (5.17).

Now that u' has been constructed, apply the K-smoothness of Λ to conclude that for every $x \in X|_{\leqslant\theta} \cap B_d(u)$,

$$\begin{aligned}\Lambda(x) &\leqslant K^{|x-u'|} \Lambda(u') \\ &\leqslant K^{|x-u|+|u-u'|} \Lambda(u') \\ &\leqslant K^{2d+1} \Lambda(u'),\end{aligned} \tag{5.20}$$

where the last step uses (5.18). As a result,

$$\begin{aligned}\Lambda(X|_{\leqslant\theta} \cap B_d(u)) &\leqslant \big|\, X|_{\leqslant\theta} \cap B_d(u)\big|\, K^{2d+1}\Lambda(u') \\ &\leqslant |B_d(u)|\, K^{2d+1}\Lambda(u') \\ &\leqslant |B_d(u)|\, K^{2d+1}\Lambda(X|_{\leqslant\theta} \setminus B_d(u)) \\ &\leqslant 2^d \binom{n+d}{d} K^{2d+1}\Lambda(X|_{\leqslant\theta} \setminus B_d(u)), \end{aligned} \tag{5.21}$$

where the first inequality is the result of summing (5.20) over $x \in X|_{\leqslant\theta} \cap B_d(u)$; the third step uses (5.18) and (5.19); and the last step applies (5.1). To complete the proof, add $\Lambda(X|_{\leqslant\theta} \setminus B_d(u))$ to both sides of (5.21). □

5.4. Weight transfer in locally smooth distributions. Locally smooth functions exhibit great plasticity. In what follows, we will show that a locally smooth function on $\prod_{i=1}^n \{0, 1, 2, \ldots, r_i\}$ can be modified to achieve a broad range of global metric behaviors—without the modification being detectable by low-degree polynomials. Among other things, we will be able to take any locally smooth distribution and make it globally min-smooth. Our starting point is a generalization of Lemma 3.2, which corresponds to taking $v = 0^n$ in the new result.

LEMMA 5.7. *Fix points $u, v \in \mathbb{N}^n$ and a natural number $d < |u - v|$. Then there is a function $\zeta_{u,v}\colon \operatorname{cube}(u, v) \to \mathbb{R}$ such that*

$$\operatorname{supp} \zeta_{u,v} \subseteq \{u\} \cup \{x \in \operatorname{cube}(u, v) : |x - v| \leqslant d\}, \tag{5.22}$$

$$\zeta_{u,v}(u) = 1, \tag{5.23}$$

$$\|\zeta_{u,v}\|_1 \leqslant 1 + 2^d \binom{|u - v|}{d}, \tag{5.24}$$

$$\operatorname{orth} \zeta_{u,v} > d. \tag{5.25}$$

Proof. Abbreviate $u^* = (|u_1 - v_1|, |u_2 - v_2|, \ldots, |u_n - v_n|)$. Lemma 3.2 constructs a function $\zeta_{u^*}\colon \mathbb{N}^n \to \mathbb{R}$ such that

$$\operatorname{supp} \zeta_{u^*} \subseteq \{u^*\} \cup \{x \in \mathbb{N}^n : x \leqslant u^* \text{ and } |x| \leqslant d\}, \tag{5.26}$$

$$\zeta_{u^*}(u^*) = 1, \tag{5.27}$$

$$\|\zeta_{u^*}\|_1 \leqslant 1 + 2^d \binom{|u^*|}{d}, \tag{5.28}$$

$$\operatorname{orth} \zeta_{u^*} > d. \tag{5.29}$$

Define $\zeta_{u,v}\colon \operatorname{cube}(u, v) \to \mathbb{R}$ by

$$\zeta_{u,v}(x) = \zeta_{u^*}(|x_1 - v_1|, |x_2 - v_2|, \ldots, |x_n - v_n|).$$

Then (5.22) and (5.23) are immediate from (5.26) and (5.27), respectively. Property (5.24) can be verified as follows:

$$\begin{aligned}\|\zeta_{u,v}\|_1 &= \sum_{x \in \mathrm{cube}(u,v)} \zeta_{u^*}(|x_1 - v_1|, |x_2 - v_2|, \ldots, |x_n - v_n|) \\ &= \sum_{\substack{w \in \mathbb{N}^n: \\ w \leqslant u^*}} \zeta_{u^*}(w) \\ &\leqslant 1 + 2^d \binom{|u^*|}{d},\end{aligned}$$

where the last step uses (5.28). For (5.25), fix an arbitrary polynomial p of degree at most d. Then at every point $x \in \mathrm{cube}(u, v)$, we have

$$\begin{aligned}p(x) &= p((x_1 - v_1) + v_1, \ldots, (x_n - v_n) + v_n) \\ &= p(\mathrm{sgn}(u_1 - v_1)|x_1 - v_1| + v_1, \ldots, \mathrm{sgn}(u_n - v_n)|x_n - v_n| + v_n) \\ &= q(|x_1 - v_1|, \ldots, |x_n - v_n|), \end{aligned} \tag{5.30}$$

where q is some polynomial of degree at most d. As a result,

$$\begin{aligned}\langle \zeta_{u,v}, p \rangle &= \sum_{x \in \mathrm{cube}(u,v)} \zeta_{u^*}(|x_1 - v_1|, \ldots, |x_n - v_n|)\, p(x) \\ &= \sum_{x \in \mathrm{cube}(u,v)} \zeta_{u^*}(|x_1 - v_1|, \ldots, |x_n - v_n|)\, q(|x_1 - v_1|, \ldots, |x_n - v_n|) \\ &= \sum_{\substack{w \in \mathbb{N}^n: \\ w \leqslant u^*}} \zeta_{u^*}(w)\, q(w) \\ &= \langle \zeta_{u^*}, q \rangle \\ &= 0,\end{aligned}$$

where the second, fourth, and fifth steps are valid by (5.30), (5.26), and (5.29), respectively. □

Our next result is a smooth analogue of Lemma 5.7. The smoothness offers a great deal of flexibility when using the lemma to transfer weight from one region of $\mathbb{N}^n$ to another, in a way that cannot be detected by a low-degree polynomial.

LEMMA 5.8. *Let $X = \prod_{i=1}^n \{0, 1, 2, \ldots, r_i\}$, where each $r_i \geqslant 0$ is an integer. Let θ and d be nonnegative integers with*

$$d < \frac{1}{3} \min \left\{ \theta, \sum_{i=1}^n r_i \right\}.$$

Let Λ be a probability distribution on $X|_{\leqslant\theta}$. Suppose further that Λ is K-smooth on $X|_{\leqslant\theta}$. Then for every $u \in X$, there is a function $Z_u \colon \mathbb{N}^n \to \mathbb{R}$ with

$$Z_u(u) = 1, \tag{5.31}$$

$$\operatorname{orth} Z_u > d, \tag{5.32}$$

$$\|Z_u\|_1 \leqslant 2^d \binom{\operatorname{diam}(\{u\} \cup \operatorname{supp}\Lambda)}{d} + 1, \tag{5.33}$$

$$|Z_u(x)| \leqslant 2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\{u\} \cup \operatorname{supp}\Lambda)}{d} \Lambda(x), \qquad x \neq u. \tag{5.34}$$

Proof. We have

$$\begin{aligned} 1 &= \Lambda(X|_{\leqslant\theta}) \\ &\leqslant K^d \binom{n+d}{d} \Lambda(X|_{\leqslant\theta-d}) \\ &\leqslant 2^{d+1} K^{3d+1} \binom{n+d}{d}^2 \Lambda(X|_{\leqslant\theta-d} \setminus B_d(u)), \end{aligned} \tag{5.35}$$

where the last two step apply Propositions 5.5 and 5.6, respectively.

We now move on to the construction of Z_u. For any $v \in X|_{\leqslant\theta-d} \setminus B_d(u)$, Lemma 5.7 gives a function $\zeta_{u,v} \colon \mathbb{N}^n \to \mathbb{R}$ with

$$\operatorname{supp}\zeta_{u,v} \subseteq X|_{\leqslant\theta} \cup \{u\}, \tag{5.36}$$

$$\zeta_{u,v}(u) = 1, \tag{5.37}$$

$$\operatorname{orth}\zeta_{u,v} > d, \tag{5.38}$$

$$\|\zeta_{u,v}\|_1 \leqslant 2^d \binom{|u-v|}{d} + 1. \tag{5.39}$$

The last inequality can be simplified as follows:

$$\begin{aligned} \|\zeta_{u,v}\|_1 &\leqslant 2^d \binom{\operatorname{diam}(X|_{\leqslant\theta} \cup \{u\})}{d} + 1 \\ &\leqslant 2^d \binom{\operatorname{diam}(\{u\} \cup \operatorname{supp}\Lambda)}{d} + 1, \end{aligned} \tag{5.40}$$

where the first step uses $v \in X|_{\leqslant\theta}$, and the second step is legitimate because Λ is a K-smooth probability distribution on $X|_{\leqslant\theta}$ and therefore $\Lambda \neq 0$ at every point of $X|_{\leqslant\theta}$. Combining (5.37) and (5.40),

$$\|\zeta_{u,v}\|_\infty \leqslant 2^d \binom{\operatorname{diam}(\{u\} \cup \operatorname{supp}\Lambda)}{d}. \tag{5.41}$$

We define $Z_u \colon \mathbb{N}^n \to \mathbb{R}$ by

$$Z_u(x) = \frac{1}{\Lambda(X|_{\leqslant \theta-d} \setminus B_d(u))} \sum_{v \in X|_{\leqslant \theta-d} \setminus B_d(u)} \Lambda(v)\, \zeta_{u,v}(x),$$

which is legitimate since $\Lambda(X|_{\leqslant \theta-d} \setminus B_d(u)) > 0$ by (5.35). Then properties (5.31), (5.32), and (5.33) for Z_u are immediate from the corresponding properties (5.37), (5.38), and (5.40) of $\zeta_{u,v}$.

It remains to verify (5.34). Fix $x \neq u$. If $x \notin X|_{\leqslant \theta}$, then (5.36) implies that $Z_u(x) = 0$ and therefore (5.34) holds in that case. In the complementary case when $x \in X|_{\leqslant \theta}$, we have

$$\begin{aligned}
|Z_u(x)| &= \sum_{v \in X|_{\leqslant \theta-d} \setminus B_d(u)} \frac{\Lambda(v)}{\Lambda(X|_{\leqslant \theta-d} \setminus B_d(u))} \cdot |\zeta_{u,v}(x)| \\
&= \sum_{\substack{v \in X|_{\leqslant \theta-d} \setminus B_d(u): \\ |v-x| \leqslant d}} \frac{\Lambda(v)}{\Lambda(X|_{\leqslant \theta-d} \setminus B_d(u))} \cdot |\zeta_{u,v}(x)| \\
&\leqslant \sum_{\substack{v \in X|_{\leqslant \theta-d} \setminus B_d(u): \\ |v-x| \leqslant d}} \frac{K^d \Lambda(x)}{\Lambda(X|_{\leqslant \theta-d} \setminus B_d(u))} \cdot 2^d \binom{\operatorname{diam}(\{u\} \cup \operatorname{supp} \Lambda)}{d} \\
&\leqslant 2^d \binom{n+d}{d} \cdot \frac{K^d \Lambda(x)}{\Lambda(X|_{\leqslant \theta-d} \setminus B_d(u))} \cdot 2^d \binom{\operatorname{diam}(\{u\} \cup \operatorname{supp} \Lambda)}{d},
\end{aligned}$$

where the first step applies the triangle inequality to the definition of Z_u; the second step uses (5.36) and $x \neq u$; the third step applies the K-smoothness of Λ and substitutes the bound from (5.41); and the final step uses (5.1). In view of (5.35), this completes the proof of (5.34). □

We now show how to efficiently zero out a locally smooth function on points of large Hamming weight. The modified function is pointwise close to the original and cannot be distinguished from it by any low-degree polynomial.

LEMMA 5.9. *Define $X = \prod_{i=1}^n \{0, 1, 2, \ldots, r_i\}$, where each $r_i \geqslant 0$ is an integer. Let θ and d be nonnegative integers with*

$$d < \frac{\theta}{3}. \tag{5.42}$$

Let $\Phi \colon X \to \mathbb{R}$ be a function that is K-smooth on $X|_{\leqslant \theta}$, with $\Phi|_{\leqslant \theta} \not\equiv 0$. Then there is $\tilde{\Phi} \colon X \to \mathbb{R}$ such that

$$\operatorname{orth}(\Phi - \tilde{\Phi}) > d, \tag{5.43}$$

$$\operatorname{supp} \tilde{\Phi} \subseteq X|_{\leqslant \theta}, \tag{5.44}$$

$$|\Phi - \tilde{\Phi}| \leqslant 2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\operatorname{supp} \Phi)}{d} \frac{\|\Phi|_{>\theta}\|_1}{\|\Phi|_{\leqslant \theta}\|_1} \cdot |\Phi| \quad \text{on } X|_{\leqslant \theta}. \tag{5.45}$$

Proof. If $\theta > \sum_{i=1}^n r_i$, the lemma holds trivially for $\tilde{\Phi} = \Phi$. In what follows, we treat the complementary case $\theta \leqslant \sum_{i=1}^n r_i$. By (5.42),

$$d < \frac{1}{3} \min \left\{ \theta, \sum_{i=1}^n r_i \right\}.$$

Since Φ is K-smooth on $X|_{\leqslant\theta}$, the probability distribution Λ on $X|_{\leqslant\theta}$ given by $\Lambda(x) = |\Phi(x)|/\|\Phi|_{\leqslant\theta}\|_1$ is also K-smooth. As a result, Lemma 5.8 gives for every $u \in X$ a function $Z_u \colon X \to \mathbb{R}$ with

$$Z_u(u) = 1, \tag{5.46}$$

$$|Z_u(x)| \leqslant 2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\{u\} \cup \operatorname{supp} \Lambda)}{d} \frac{|\Phi(x)|}{\|\Phi|_{\leqslant\theta}\|_1} \quad \text{for } x \neq u, \tag{5.47}$$

$$\operatorname{orth} Z_u > d, \tag{5.48}$$

$$\operatorname{supp} Z_u \subseteq X|_{\leqslant\theta} \cup \{u\}. \tag{5.49}$$

Now define

$$\tilde{\Phi} = \Phi - \sum_{u \in X|_{>\theta}} \Phi(u) Z_u.$$

Then (5.43) is immediate from (5.48). To verify (5.44), fix any point $x \in X|_{>\theta}$. Then

$$\begin{aligned} \tilde{\Phi}(x) &= \Phi(x) - \sum_{u \in X|_{>\theta}} \Phi(u) Z_u(x) \\ &= \Phi(x) - \Phi(x) Z_x(x) \\ &= 0, \end{aligned}$$

where the last two steps use (5.49) and (5.46), respectively.

It remains to verify (5.45) on $X|_{\leqslant\theta}$:

$$\begin{aligned} |\Phi - \tilde{\Phi}| &\leqslant \sum_{\substack{u \in X|_{>\theta}: \\ \Phi(u) \neq 0}} \Phi(u) \, |Z_u| \\ &\leqslant 2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\operatorname{supp} \Phi)}{d} \sum_{\substack{u \in X|_{>\theta}: \\ \Phi(u) \neq 0}} |\Phi(u)| \cdot \frac{|\Phi|}{\|\Phi|_{\leqslant\theta}\|_1} \\ &= 2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\operatorname{supp} \Phi)}{d} \frac{\|\Phi|_{>\theta}\|_1}{\|\Phi|_{\leqslant\theta}\|_1} \cdot |\Phi|, \end{aligned}$$

where the second step uses (5.47). ∎

For technical reasons, we need a generalization of the previous lemma to functions on $\prod_{i=1}^n \{\Delta_i, \Delta_i + 1, \ldots, \Delta_i + r_i\}$ for nonnegative integers Δ_i and r_i, and further

to convex combinations of such functions. We obtain these generalizations in the two corollaries that follow.

COROLLARY 5.10. *Define* $X = \prod_{i=1}^{n} \{\Delta_i, \Delta_i + 1, \ldots, \Delta_i + r_i\}$, *where all* Δ_i *and* r_i *are nonnegative integers. Let* θ *and* d *be nonnegative integers with*

$$d < \frac{1}{3}\left(\theta - \sum_{i=1}^{n} \Delta_i\right).$$

Let $\Phi\colon X \to \mathbb{R}$ *be a function that is* K*-smooth on* $X|_{\leqslant\theta}$, *with* $\Phi|_{\leqslant\theta} \not\equiv 0$. *Then there is a function* $\tilde{\Phi}\colon X \to \mathbb{R}$ *such that*

$$\operatorname{orth}(\Phi - \tilde{\Phi}) > d, \tag{5.50}$$

$$\operatorname{supp} \tilde{\Phi} \subseteq X|_{\leqslant\theta}, \tag{5.51}$$

$$|\Phi - \tilde{\Phi}| \leqslant 2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\operatorname{supp}\Phi)}{d} \frac{\|\Phi|_{>\theta}\|_1}{\|\Phi|_{\leqslant\theta}\|_1} \cdot |\Phi| \quad \text{on } X|_{\leqslant\theta}. \tag{5.52}$$

Proof. Abbreviate $X' = \prod_{i=1}^{n} \{0, 1, 2, \ldots, r_i\}$ and $\theta' = \theta - \sum_{i=1}^{n} \Delta_i$. In this notation,

$$d < \frac{\theta'}{3}. \tag{5.53}$$

Consider the function $\Phi'\colon X' \to \mathbb{R}$ given by $\Phi'(x) = \Phi(x + (\Delta_1, \Delta_2 \ldots, \Delta_n))$. Then any two points $u, v \in X'|_{\leqslant\theta'}$ obey

$$\begin{aligned} |\Phi'(u)| &= |\Phi(u + (\Delta_1, \Delta_2, \ldots, \Delta_n))| \\ &\leqslant K^{|u-v|} |\Phi(v + (\Delta_1, \Delta_2, \ldots, \Delta_n))| \\ &= K^{|u-v|} |\Phi'(v)|, \end{aligned}$$

where the second step uses the K-smoothness of Φ on $X|_{\leqslant\theta}$ As a result, Φ' is K-smooth on $X'|_{\leqslant\theta'}$. Moreover, $\|\Phi'|_{\leqslant\theta'}\|_1 = \|\Phi|_{\leqslant\theta}\|_1 > 0$. In view of (5.53), Lemma 5.9 gives a function a function $\tilde{\Phi}'\colon X' \to \mathbb{R}$ such that

$$\operatorname{orth}(\Phi' - \tilde{\Phi}') > d,$$
$$\operatorname{supp} \tilde{\Phi}' \subseteq X'|_{\leqslant\theta'},$$

and

$$\begin{aligned} |\Phi' - \tilde{\Phi}'| &\leqslant 2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\operatorname{supp}\Phi')}{d} \frac{\|\Phi'|_{>\theta'}\|_1}{\|\Phi'|_{\leqslant\theta'}\|_1} \cdot |\Phi'| \\ &= 2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\operatorname{supp}\Phi)}{d} \frac{\|\Phi|_{>\theta}\|_1}{\|\Phi|_{\leqslant\theta}\|_1} \cdot |\Phi'| \end{aligned}$$

on $X'|_{\leqslant\theta'}$. As a result, (5.50)–(5.52) hold for the real-valued function $\tilde{\Phi}\colon X \to \mathbb{R}$ given by $\tilde{\Phi}(x) = \tilde{\Phi}'(x - (\Delta_1, \Delta_2, \dots, \Delta_n))$. □

COROLLARY 5.11. *Fix integers $\Delta, d, \theta \geqslant 0$ and $n \geqslant 1$, and a real number δ, where*

$$\begin{aligned}&\delta \in [0,1),\\ &d < \frac{1}{3}(\theta - \Delta).\end{aligned}$$

Then for every

$$\Lambda \in \operatorname{conv}(\mathfrak{S}(n,K,\Delta) \cap \{\Lambda' \in \mathfrak{D}(\mathbb{N}^n) : \Lambda'(\mathbb{N}^n|_{>\theta}) \leqslant \delta\}),$$

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

$$\begin{aligned}&\operatorname{orth}(\Lambda - \tilde{\Lambda}) > d,\\ &\operatorname{supp}\tilde{\Lambda} \subseteq \mathbb{N}^n|_{\leqslant\theta} \cap \operatorname{supp}\Lambda,\\ &|\Lambda - \tilde{\Lambda}| \leqslant 2^{3d+1}K^{4d+1}\binom{n+d}{d}^3\binom{\operatorname{diam}(\operatorname{supp}\Lambda)}{d}\frac{\delta}{1-\delta}\cdot\Lambda \qquad \text{on } \mathbb{N}^n|_{\leqslant\theta}.\end{aligned}$$

Proof. Write Λ out explicitly as

$$\Lambda = \sum_{i=1}^{N} \lambda_i \Lambda_i$$

for some positive reals $\lambda_1, \dots, \lambda_N$ with $\sum \lambda_i = 1$, where $\Lambda_i \in \mathfrak{S}(n,K,\Delta)$ and $\Lambda_i(\mathbb{N}^n|_{>\theta}) \leqslant \delta$. Then clearly

$$\operatorname{supp}\Lambda = \bigcup_{i=1}^{n} \operatorname{supp}\Lambda_i. \tag{5.54}$$

For $i = 1, 2, \dots, N$, Corollary 5.10 constructs $\tilde{\Lambda}_i\colon \mathbb{N}^n \to \mathbb{R}$ with

$$\operatorname{orth}(\Lambda_i - \tilde{\Lambda}_i) > d, \tag{5.55}$$

$$\operatorname{supp}\tilde{\Lambda}_i \subseteq \mathbb{N}^n|_{\leqslant\theta}, \tag{5.56}$$

$$|\Lambda_i - \tilde{\Lambda}_i| \leqslant 2^{3d+1}K^{4d+1}\binom{n+d}{d}^3\binom{\operatorname{diam}(\operatorname{supp}\Lambda_i)}{d}\frac{\delta}{1-\delta}\cdot\Lambda_i \quad \text{on } \mathbb{N}^n|_{\leqslant\theta}, \tag{5.57}$$

$$\operatorname{supp}\tilde{\Lambda}_i \subseteq \operatorname{supp}\Lambda_i. \tag{5.58}$$

In view of (5.54)–(5.58), the proof is complete by taking $\tilde{\Lambda} = \sum_{i=1}^{N} \lambda_i \tilde{\Lambda}_i$. □

Our next result uses local smoothness to achieve something completely different. Here, we show how to start with a locally smooth function and make it globally min-smooth. The new function has the same sign pointwise as the original, and

cannot be distinguished from it by any low-degree polynomial. Crucially for us, the global min-smoothness can be achieved relative to any distribution on the domain.

LEMMA 5.12. *Define* $X = \prod_{i=1}^{n} \{0, 1, 2, \ldots, r_i\}$, *where each* $r_i \geqslant 0$ *is an integer. Let* θ *and* d *be nonnegative integers with*

$$d < \frac{1}{3} \min \left\{ \theta, \sum_{i=1}^{n} r_i \right\}.$$

Let $\Phi \colon X|_{\leqslant \theta} \to \mathbb{R}$ *be a function that is* K*-smooth on* $X|_{\leqslant \theta}$. *Then for every probability distribution* Λ^* *on* $X|_{\leqslant \theta}$, *there is* $\Phi^* \colon X|_{\leqslant \theta} \to \mathbb{R}$ *such that*

$$\operatorname{orth}(\Phi - \Phi^*) > d, \tag{5.59}$$

$$\|\Phi^*\|_1 \leqslant 2\|\Phi\|_1, \tag{5.60}$$

$$\Phi \cdot \Phi^* \geqslant 0, \tag{5.61}$$

$$|\Phi^*| \geqslant \left(2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\operatorname{supp} \Phi)}{d} \right)^{-1} \|\Phi\|_1 \, \Lambda^*. \tag{5.62}$$

Proof. If $\Phi \equiv 0$, the lemma holds trivially with $\Phi^* = \Phi$. In the complementary case, abbreviate

$$N = 2^{3d+1} K^{4d+1} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\operatorname{supp} \Phi)}{d}.$$

We will view $|\Phi|/\|\Phi\|_1$ as a probability distribution on $X|_{\leqslant \theta}$. By hypothesis, this probability distribution is K-smooth on $X|_{\leqslant \theta}$. In particular, $X|_{\leqslant \theta} \subseteq \operatorname{supp} |\Phi| = \operatorname{supp} \Phi$. Therefore, Lemma 5.8 gives for every $u \in X|_{\leqslant \theta}$ a function $Z_u \colon X|_{\leqslant \theta} \to \mathbb{R}$ with

$$Z_u(u) = 1, \tag{5.63}$$

$$\|Z_u\|_1 \leqslant \frac{N}{2} + 1, \tag{5.64}$$

$$|Z_u(x)| \leqslant N \cdot \frac{|\Phi(x)|}{\|\Phi\|_1}, \qquad x \neq u, \tag{5.65}$$

$$\operatorname{orth} Z_u > d. \tag{5.66}$$

Now, define $\Phi^* \colon X|_{\leqslant \theta} \to \mathbb{R}$ by

$$\Phi^* = \Phi + \frac{\|\Phi\|_1}{N} \sum_{u \in X|_{\leqslant \theta}} \widetilde{\operatorname{sgn}}(\Phi(u)) \Lambda^*(u) Z_u.$$

Then (5.59) follows directly from (5.66). For (5.60), we have:

$$\begin{aligned}
\|\Phi^*\|_1 &\leqslant \|\Phi\|_1 + \frac{\|\Phi\|_1}{N} \sum_{u \in X|_{\leqslant\theta}} \Lambda^*(u)\, \|Z_u\|_1 \\
&\leqslant \|\Phi\|_1 + \frac{\|\Phi\|_1}{N} \cdot \left(\frac{N}{2} + 1\right) \sum_{u \in X|_{\leqslant\theta}} \Lambda^*(u) \\
&= \frac{3N+2}{2N} \|\Phi\|_1 \\
&\leqslant 2\, \|\Phi\|_1, \qquad (5.67)
\end{aligned}$$

where the second step uses (5.64). The remaining properties (5.61) and (5.62) can be established simultaneously as follows: for every $x \in X|_{\leqslant\theta}$,

$$\begin{aligned}
&\widetilde{\operatorname{sgn}}(\Phi(x)) \cdot \Phi^*(x) \\
&\qquad = |\Phi(x)| + \frac{\|\Phi\|_1}{N} \sum_{u \in X|_{\leqslant\theta}} \Lambda^*(u) Z_u(x) \\
&\qquad \geqslant |\Phi(x)| + \frac{\|\Phi\|_1}{N}\, \Lambda^*(x) Z_x(x) - \frac{\|\Phi\|_1}{N} \sum_{\substack{u \in X|_{\leqslant\theta}: \\ u \neq x}} \Lambda^*(u)\, |Z_u(x)| \\
&\qquad = |\Phi(x)| + \frac{\|\Phi\|_1}{N}\, \Lambda^*(x) - \frac{\|\Phi\|_1}{N} \sum_{\substack{u \in X|_{\leqslant\theta}: \\ u \neq x}} \Lambda^*(u)\, |Z_u(x)| \\
&\qquad \geqslant |\Phi(x)| + \frac{\|\Phi\|_1}{N}\, \Lambda^*(x) - \frac{\|\Phi\|_1}{N} \cdot N \cdot \frac{|\Phi(x)|}{\|\Phi\|_1} \sum_{\substack{u \in X|_{\leqslant\theta}: \\ u \neq x}} \Lambda^*(u) \\
&\qquad = |\Phi(x)| + \frac{\|\Phi\|_1}{N}\, \Lambda^*(x) - |\Phi(x)|\, (1 - \Lambda^*(x)) \\
&\qquad \geqslant \frac{\|\Phi\|_1}{N}\, \Lambda^*(x), \qquad (5.68)
\end{aligned}$$

where the third and fourth steps use (5.63) and (5.65), respectively. □

5.5. A locally smooth dual polynomial for MP. As Sections 5.2–5.4 show, local smoothness implies several useful metric and analytic properties. To tap into this resource, we now construct a locally smooth dual polynomial for the Minsky–Papert function. It is helpful to view this new result as a counterpart of Theorem 4.4 from our analysis of the threshold degree of AC^0. The new proof is considerably more technical because local smoothness is a delicate property to achieve.

THEOREM 5.13. *For some absolute constant $0 < c < 1$ and all positive integers m, r, R with $r \leqslant R$, there are probability distributions Λ_0 and Λ_1 such that*

$$\operatorname{supp} \Lambda_0 = (\mathrm{MP}^*_{m,R})^{-1}(0), \tag{5.69}$$

$$\operatorname{supp} \Lambda_1 = (\mathrm{MP}^*_{m,R})^{-1}(1), \tag{5.70}$$

$$\operatorname{orth}(\Lambda_0 - \Lambda_1) \geqslant \min\{m, c\sqrt{r}\}, \tag{5.71}$$

$$\frac{\Lambda_0 + \Lambda_1}{2} \in \operatorname{Smooth}\left(\frac{m}{c}, \{0, 1, 2, \ldots, R\}^m\right), \tag{5.72}$$

$$\Lambda_0, \Lambda_1 \in \operatorname{conv}\left(\left\{\lambda \in \mathfrak{S}\left(1, \frac{1}{c}, 1\right) : \lambda(t) \leqslant \frac{1}{c(t+1)^2\, 2^{ct/\sqrt{r}}} \text{ for } t \in \mathbb{N}\right\}^{\otimes m}\right). \tag{5.73}$$

Our proof of Theorem 5.13 repeatedly employs the following simple but useful criterion for K-smoothness: a probability distribution λ is K-smooth on an integer interval $I = \{i, i+1, i+2, \ldots, j\}$ if and only if the probabilities of any two *consecutive* integers in I are within a factor of K.

Proof of Theorem 5.13. Abbreviate $\epsilon = 1/6$. For some absolute constants $c', c'' \in (0, 1)$, Lemma 4.3 constructs probability distributions $\lambda_0, \lambda_1, \lambda_2$ such that

$$\operatorname{supp} \lambda_0 = \{0\}, \tag{5.74}$$

$$\operatorname{supp} \lambda_i = \{1, 2, \ldots, R\}, \qquad i = 1, 2, \tag{5.75}$$

$$\lambda_i(t) \in \left[\frac{c'}{t^2\, 2^{c''t/\sqrt{r}}},\ \frac{1}{c' t^2\, 2^{c''t/\sqrt{r}}}\right], \qquad i = 1, 2; \quad t = 1, 2, \ldots, R, \tag{5.76}$$

$$\operatorname{orth}((1-\epsilon)\lambda_0 + \epsilon\lambda_2 - \lambda_1) \geqslant c'\sqrt{r}. \tag{5.77}$$

We infer that

$$\lambda_0 \in \mathfrak{S}(1, K), \tag{5.78}$$

$$\lambda_1 \in \mathfrak{S}(1, K, 1), \tag{5.79}$$

$$\lambda_2 \in \mathfrak{S}(1, K, 1), \tag{5.80}$$

$$(1-\epsilon)\lambda_0 + \epsilon\lambda_2 \in \mathfrak{S}(1, K), \tag{5.81}$$

$$\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}\lambda_1 \in \mathfrak{S}(1, Km) \tag{5.82}$$

for some large constant $K = K(c', c'') \geqslant 1$. Indeed, (5.78) is trivial since λ_0 is the single-point distribution on the origin; (5.79) holds because by (5.75) and (5.76), the probabilities of any pair of consecutive integers in $\operatorname{supp} \lambda_1 = \{1, 2, \ldots, R\}$ are the same up to a constant factor; and (5.80)–(5.82) can be seen analogously, by comparing the probabilities of any pair of consecutive integers. Combining (5.78)–(5.82)

with Proposition 5.4, we obtain

$$\{\lambda_0, \lambda_1, \lambda_2\}^{\otimes m} \subseteq \mathfrak{S}(m, K, m), \tag{5.83}$$

$$((1-\epsilon)\lambda_0 + \epsilon\lambda_2)^{\otimes m} \in \mathfrak{S}(m, K), \tag{5.84}$$

$$\left(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}\lambda_1\right)^{\otimes m} \in \mathfrak{S}(m, Km). \tag{5.85}$$

The proof centers around the dual objects $\Psi_1, \Psi_2 \colon \{0, 1, 2, \ldots, R\}^m \to \mathbb{R}$ given by

$$\Psi_1 = \left(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}\lambda_1\right)^{\otimes m} - 2\lambda_1^{\otimes m}$$

and

$$\Psi_2 = 2((1-\epsilon)\lambda_0 + \epsilon\lambda_2)^{\otimes m} - 2(-\epsilon\lambda_0 + \epsilon\lambda_2)^{\otimes m} - \left(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}((1-\epsilon)\lambda_0 + \epsilon\lambda_2)\right)^{\otimes m}.$$

The next four claims establish key properties of Ψ_1 and Ψ_2.

Claim 5.14. *Ψ_1 satisfies*

$$\operatorname{pos}\Psi_1 \in \operatorname{cone}(\{\lambda_0, \lambda_1\}^{\otimes m} \setminus \{\lambda_1^{\otimes m}\}), \tag{5.86}$$

$$\operatorname{neg}\Psi_1 \in \operatorname{cone}\{\lambda_1^{\otimes m}\}, \tag{5.87}$$

$$\frac{1}{5}|\Psi_1| \leqslant \left(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}\lambda_1\right)^{\otimes m} \leqslant |\Psi_1|. \tag{5.88}$$

Claim 5.15. *Ψ_2 satisfies*

$$\operatorname{pos}\Psi_2 \in \operatorname{cone}(\{\lambda_0, \lambda_2\}^{\otimes m} \setminus \{\lambda_2^{\otimes m}\}), \tag{5.89}$$

$$\operatorname{neg}\Psi_2 \in \operatorname{cone}\{\lambda_2^{\otimes m}\}, \tag{5.90}$$

$$\frac{1}{3}|\Psi_2| \leqslant ((1-\epsilon)\lambda_0 + \epsilon\lambda_2)^{\otimes m} \leqslant 3|\Psi_2|. \tag{5.91}$$

Claim 5.16. *Ψ_1 and Ψ_2 satisfy*

$$\operatorname{supp}(\operatorname{pos}\Psi_i) = (\mathrm{MP}^*_{m,R})^{-1}(0), \qquad i = 1, 2, \tag{5.92}$$

$$\operatorname{supp}(\operatorname{neg}\Psi_i) = (\mathrm{MP}^*_{m,R})^{-1}(1), \qquad i = 1, 2. \tag{5.93}$$

Claim 5.17. $\operatorname{orth}(\Psi_1 + \Psi_2) \geqslant \min\{m, c'\sqrt{r}\}$.

We will settle Claims 5.14–5.17 shortly, once we complete the main proof. Define

$$\Lambda_0 = \frac{2}{\|\Psi_1\|_1 + \|\Psi_2\|_1} \operatorname{pos}(\Psi_1 + \Psi_2),$$
$$\Lambda_1 = \frac{2}{\|\Psi_1\|_1 + \|\Psi_2\|_1} \operatorname{neg}(\Psi_1 + \Psi_2),$$

where the denominators are nonzero by (5.88). We proceed to verify the properties required of Λ_0 and Λ_1 in the theorem statement.

Support. Recall from Claim 5.16 that the positive parts of Ψ_1 and Ψ_2 are supported on $(\mathrm{MP}^*_{m,R})^{-1}(0)$. Therefore, the positive part of $\Psi_1 + \Psi_2$ is supported on $(\mathrm{MP}^*_{m,R})^{-1}(0)$ as well, which in turn implies that

$$\operatorname{supp} \Lambda_0 = (\mathrm{MP}^*_{m,R})^{-1}(0). \tag{5.94}$$

Analogously, Claim 5.16 states that the negative parts of Ψ_1 and Ψ_2 are supported on $(\mathrm{MP}^*_{m,R})^{-1}(1)$. As a result, the negative part of $\Psi_1 + \Psi_2$ is also supported on $(\mathrm{MP}^*_{m,R})^{-1}(1)$, whence

$$\operatorname{supp} \Lambda_1 = (\mathrm{MP}^*_{m,R})^{-1}(1). \tag{5.95}$$

Orthogonality. The defining equations for Λ_0 and Λ_1 imply that

$$\Lambda_0 - \Lambda_1 = \frac{2}{\|\Psi_1\|_1 + \|\Psi_2\|_1} (\Psi_1 + \Psi_2),$$

which along with Claim 5.17 forces

$$\operatorname{orth}(\Lambda_0 - \Lambda_1) \geqslant \min\{m, c'\sqrt{r}\}. \tag{5.96}$$

Nonnegativity and norm. By definition, Λ_0 and Λ_1 are nonnegative functions. We calculate

$$\begin{aligned} \|\Lambda_0\|_1 - \|\Lambda_1\|_1 &= \langle \Lambda_0, 1 \rangle - \langle \Lambda_1, 1 \rangle \\ &= \langle \Lambda_0 - \Lambda_1, 1 \rangle \\ &= 0, \end{aligned} \tag{5.97}$$

where the first step uses the nonnegativity of Λ_0 and Λ_1, and the last step applies (5.96). In addition,

$$\begin{aligned} \|\Lambda_0\|_1 + \|\Lambda_1\|_1 &= \frac{2}{\|\Psi_1\|_1 + \|\Psi_2\|_1} (\|\operatorname{pos}(\Psi_1 + \Psi_2)\|_1 + \|\operatorname{neg}(\Psi_1 + \Psi_2)\|_1) \\ &= \frac{2}{\|\Psi_1\|_1 + \|\Psi_2\|_1} \|\Psi_1 + \Psi_2\|_1 \\ &= 2, \end{aligned} \tag{5.98}$$

where the last step uses Claim 5.16. A consequence of (5.97) and (5.98) is that $\|\Lambda_0\|_1 = \|\Lambda_1\|_1 = 1$, which makes Λ_0 and Λ_1 probability distributions. In view of (5.94) and (5.95), we conclude that

$$\Lambda_i \in \mathfrak{D}((\mathrm{MP}^*_{m,R})^{-1}(i)), \qquad i = 0, 1. \tag{5.99}$$

In particular,

$$\frac{\Lambda_0 + \Lambda_1}{2} \in \mathfrak{D}(\{0, 1, 2, \ldots, R\}^m). \tag{5.100}$$

Smoothness. We have

$$\begin{aligned}\frac{\Lambda_0 + \Lambda_1}{2} &= \frac{|\Psi_1 + \Psi_2|}{\|\Psi_1\|_1 + \|\Psi_2\|_1} \\ &= \frac{1}{\|\Psi_1\|_1 + \|\Psi_2\|_1} |\Psi_1| + \frac{1}{\|\Psi_1\|_1 + \|\Psi_2\|_1} |\Psi_2|, \end{aligned} \tag{5.101}$$

where the first step follows from the defining equations for Λ_0 and Λ_1, and the second step uses Claim 5.16. Inequality (5.88) shows that at every point, $|\Psi_1|$ is within a factor of 5 of the tensor product $(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}\lambda_1)^{\otimes m}$, which by (5.85) is Km-smooth on its support. It follows that $|\Psi_1|$ is $5Km$-smooth on $\{0, 1, 2, \ldots, R\}^m$. By an analogous argument, (5.91) and (5.84) imply that $|\Psi_2|$ is $3K$-smooth (and hence also $5Km$-smooth) on $\{0, 1, 2, \ldots, R\}^m$. Now (5.101) shows that $\frac{1}{2}(\Lambda_0+\Lambda_1)$ is a conical combination of two nonnegative $5Km$-smooth functions on $\{0, 1, 2, \ldots, R\}^m$. By Proposition 5.3(iv),

$$\frac{\Lambda_0 + \Lambda_1}{2} \in \mathrm{Smooth}(5Km, \{0, 1, 2, \ldots, R\}^m). \tag{5.102}$$

Having examined the convex combination $\frac{\Lambda_0+\Lambda_1}{2}$, we now turn to the individual distributions Λ_0 and Λ_1. We have

$$\begin{aligned}\Lambda_0 &= \frac{2}{\|\Psi_1\|_1 + \|\Psi_2\|_1} \mathrm{pos}(\Psi_1 + \Psi_2) \\ &= \frac{2}{\|\Psi_1\|_1 + \|\Psi_2\|_1} (\mathrm{pos}(\Psi_1) + \mathrm{pos}(\Psi_2)) \\ &\in \mathrm{cone}(\{\lambda_0, \lambda_1, \lambda_2\}^{\otimes m}),\end{aligned}$$

where the first equation restates the definition of Λ_0, the second step applies (5.92), and the last step uses (5.86) and (5.89). Analogously,

$$\begin{aligned}\Lambda_1 &= \frac{2}{\|\Psi_1\|_1 + \|\Psi_2\|_1} \mathrm{neg}(\Psi_1 + \Psi_2) \\ &= \frac{2}{\|\Psi_1\|_1 + \|\Psi_2\|_1} (\mathrm{neg}(\Psi_1) + \mathrm{neg}(\Psi_2)) \\ &\in \mathrm{cone}(\{\lambda_1^{\otimes m}, \lambda_2^{\otimes m}\}),\end{aligned}$$

where the first equation restates the definition of Λ_1, the second step applies (5.93), and the last step uses (5.87) and (5.90). Thus, Λ_0 and Λ_1 are conical combinations of probability distributions in $\{\lambda_0, \lambda_1, \lambda_2\}^{\otimes m}$. Since Λ_0 and Λ_1 are themselves probability distributions, we conclude that

$$\Lambda_0, \Lambda_1 \in \operatorname{conv}(\{\lambda_0, \lambda_1, \lambda_2\}^{\otimes m}).$$

By (5.74)–(5.76),

$$\lambda_i(t) \leqslant \frac{1}{c'''(t+1)^2\, 2^{c'''t/\sqrt{r}}} \qquad (t \in \mathbb{N};\ i = 0, 1, 2)$$

for some constant $c''' > 0$. The last two equations along with (5.78)–(5.80) yield

$$\Lambda_0, \Lambda_1 \in \operatorname{conv}\left(\left\{\lambda \in \mathfrak{S}(1, K, 1) : \lambda(t) \leqslant \frac{1}{c'''(t+1)^2\, 2^{c'''t/\sqrt{r}}} \text{ for } t \in \mathbb{N}\right\}^{\otimes m}\right). \quad (5.103)$$

Now (5.94)–(5.96), (5.102), and (5.103) imply (5.69)–(5.73) for a small enough constant $c > 0$. □

We now settle the four claims made in the proof of Theorem 5.13.

Proof of Claim 5.14. Multiplying out the tensor product in the definition of Ψ_1 and collecting like terms, we obtain

$$\Psi_1 = -\left(2 - \left(\frac{m}{m+1}\right)^m\right)\lambda_1^{\otimes m} + \sum_{\substack{S \subseteq \{1,2,\ldots,m\} \\ S \neq \varnothing}} \left(\frac{1}{m+1}\right)^{|S|}\left(\frac{m}{m+1}\right)^{m-|S|} \lambda_0^{\otimes S} \cdot \lambda_1^{\otimes \overline{S}}. \quad (5.104)$$

Recall from (5.74) and (5.75) that λ_0 and λ_1 are supported on $\{0\}$ and $\{1, 2, \ldots, R\}$, respectively. Therefore, the right-hand side of (5.104) is the sum of 2^m nonzero functions whose supports are pairwise disjoint. Now (5.86) and (5.87) follow directly from (5.104). One further obtains that

$$|\Psi_1| = \left(2 - \left(\frac{m}{m+1}\right)^m\right)\lambda_1^{\otimes m} + \sum_{\substack{S \subseteq \{1,2,\ldots,m\} \\ S \neq \varnothing}} \left(\frac{1}{m+1}\right)^{|S|}\left(\frac{m}{m+1}\right)^{m-|S|} \lambda_0^{\otimes S} \cdot \lambda_1^{\otimes \overline{S}}.$$

From first principles,

$$\left(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}\lambda_1\right)^{\otimes m} = \left(\frac{m}{m+1}\right)^m \lambda_1^{\otimes m} + \sum_{\substack{S\subseteq\{1,2,\ldots,m\}\\ S\neq\varnothing}} \left(\frac{1}{m+1}\right)^{|S|}\left(\frac{m}{m+1}\right)^{m-|S|} \lambda_0^{\otimes S}\cdot\lambda_1^{\otimes\overline{S}}.$$

Comparing the right-hand sides of the last two equations settles (5.88). □

Proof of Claim 5.15. Multiplying out the tensor powers in the definition of Ψ_2 and collecting like terms, we obtain

$$\Psi_2 = -\left(\frac{m}{m+1}\right)^m \epsilon^m \lambda_2^{\otimes m} + \sum_{\substack{S\subseteq\{1,2,\ldots,m\}\\ S\neq\varnothing}} a_{|S|}\,\lambda_0^{\otimes S}\cdot\lambda_2^{\otimes\overline{S}}, \tag{5.105}$$

where the coefficients $a_1, a_2, \ldots, a_m$ are given by

$$\begin{aligned} a_i &= \left(2(1-\epsilon)^i\epsilon^{m-i} - 2(-1)^i\epsilon^m - \left(1-\frac{\epsilon m}{m+1}\right)^i\left(\frac{\epsilon m}{m+1}\right)^{m-i}\right) \\ &= (1-\epsilon)^i\epsilon^{m-i}\left(2 - 2\left(\frac{-\epsilon}{1-\epsilon}\right)^i - \left(\frac{m+1-\epsilon m}{(m+1)(1-\epsilon)}\right)^i\left(\frac{m}{m+1}\right)^{m-i}\right) \\ &\in \left[\frac{1}{3}(1-\epsilon)^i\epsilon^{m-i},\, 3(1-\epsilon)^i\epsilon^{m-i}\right]. \end{aligned} \tag{5.106}$$

As in the proof of the previous claim, recall from (5.74) and (5.75) that λ_0 and λ_2 have disjoint support. Therefore, the right-hand side of (5.105) is the sum of 2^m nonzero functions whose supports are pairwise disjoint. Now (5.89) and (5.90) are immediate from (5.106). The disjointness of the supports of the summands on the right-hand side of (5.105) also implies that

$$|\Psi_2| = \left(\frac{m}{m+1}\right)^m \epsilon^m \lambda_0^{\otimes m} + \sum_{\substack{S\subseteq\{1,2,\ldots,m\}\\ S\neq\varnothing}} |a_{|S|}|\,\lambda_0^{\otimes S}\cdot\lambda_2^{\otimes\overline{S}}.$$

In view of (5.106), we conclude that $|\Psi_2|$ coincides up to a factor of 3 with the function

$$\sum_{S\subseteq\{1,2,\ldots,m\}} (1-\epsilon)^{|S|}\epsilon^{m-|S|}\lambda_0^{\otimes S}\cdot\lambda_2^{\otimes\overline{S}} = ((1-\epsilon)\lambda_0 + \epsilon\lambda_2)^{\otimes m}.$$

This settles (5.91) and completes the proof. □

Proof of Claim 5.16. Recall from (5.74) and (5.75) that $\operatorname{supp}\lambda_0 = \{0\}$ and $\operatorname{supp}\lambda_1 = \operatorname{supp}\lambda_2 = \{1, 2, \ldots, R\}$. In this light, (5.86)–(5.88) imply

$$\begin{aligned}\operatorname{supp}(\operatorname{pos}\Psi_1) &\subseteq (\mathrm{MP}^*_{m,R})^{-1}(0),\\ \operatorname{supp}(\operatorname{neg}\Psi_1) &\subseteq (\mathrm{MP}^*_{m,R})^{-1}(1),\\ \operatorname{supp}(\Psi_1) &= (\mathrm{MP}^*_{m,R})^{-1}(0) \cup (\mathrm{MP}^*_{m,R})^{-1}(1),\end{aligned}$$

respectively. Analogously, (5.89)–(5.91) imply

$$\begin{aligned}\operatorname{supp}(\operatorname{pos}\Psi_2) &\subseteq (\mathrm{MP}^*_{m,R})^{-1}(0),\\ \operatorname{supp}(\operatorname{neg}\Psi_2) &\subseteq (\mathrm{MP}^*_{m,R})^{-1}(1),\\ \operatorname{supp}(\Psi_2) &= (\mathrm{MP}^*_{m,R})^{-1}(0) \cup (\mathrm{MP}^*_{m,R})^{-1}(1).\end{aligned}$$

Since the support of each Ψ_i is the disjoint union of the supports of its positive and negative parts, (5.92) and (5.93) follow. □

Proof of Claim 5.17. Write $\Psi_1 + \Psi_2 = A + B + C$, where

$$\begin{aligned}A &= \left(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}\lambda_1\right)^{\otimes m} - \left(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}((1-\epsilon)\lambda_0 + \epsilon\lambda_2)\right)^{\otimes m},\\ B &= 2((1-\epsilon)\lambda_0 + \epsilon\lambda_2)^{\otimes m} - 2\lambda_1^{\otimes m},\\ C &= -2(-\epsilon\lambda_0 + \epsilon\lambda_2)^{\otimes m}.\end{aligned}$$

As a result, Proposition 2.1(i) guarantees that

$$\operatorname{orth}(\Psi_1 + \Psi_2) \geqslant \min\{\operatorname{orth} A, \operatorname{orth} B, \operatorname{orth} C\}. \tag{5.107}$$

We have

$$\begin{aligned}\operatorname{orth} A &\geqslant \operatorname{orth}\left(\left(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}\lambda_1\right)\right.\\ &\qquad\qquad \left. - \left(\frac{1}{m+1}\lambda_0 + \frac{m}{m+1}((1-\epsilon)\lambda_0 + \epsilon\lambda_2)\right)\right)\\ &= \operatorname{orth}\left(-\frac{m}{m+1}((1-\epsilon)\lambda_0 + \epsilon\lambda_2 - \lambda_1)\right)\\ &\geqslant c'\sqrt{r},\end{aligned} \tag{5.108}$$

where the first step uses Proposition 2.1(iii), and the last step is a restatement of (5.77). Analogously,

$$\begin{aligned}\operatorname{orth} B &\geqslant \operatorname{orth}(((1-\epsilon)\lambda_0 + \epsilon\lambda_2) - \lambda_1)\\ &\geqslant c'\sqrt{r},\end{aligned} \tag{5.109}$$

where the first and second steps use Proposition 2.1(iii) and (5.77), respectively. Finally,

$$\begin{aligned}\operatorname{orth} C &= \operatorname{orth}((-\epsilon\lambda_0 + \epsilon\lambda_2)^{\otimes m}) \\ &= m \operatorname{orth}(-\epsilon\lambda_0 + \epsilon\lambda_2) \\ &\geqslant m, \end{aligned} \tag{5.110}$$

where the second step applies Proposition 2.1(ii), and the third step is valid because $\langle -\epsilon\lambda_0 + \epsilon\lambda_2, 1\rangle = -\epsilon\langle\lambda_0, 1\rangle + \epsilon\langle\lambda_2, 1\rangle = -\epsilon + \epsilon = 0$. By (5.107)–(5.110), the proof is complete. □

5.6. An amplification theorem for smooth threshold degree. We have reached the technical centerpiece of our sign-rank analysis, an amplification theorem for smooth threshold degree. This result is considerably stronger than the amplification theorems for threshold degree in Section 4.3, which does not preserve smoothness. We prove the new amplification theorem by manipulating locally smooth distributions to achieve the desired global behavior, an approach unrelated to our work in Section 4.3. A detailed statement of our result follows.

THEOREM 5.18. *There is an absolute constant $C \geqslant 1$ such that*

for all:

> *positive integers n, m, r, R, θ with $R \geqslant r$ and $\theta \geqslant Cnm\log(2nm)$;*
> *real numbers $\gamma \in [0, 1]$;*
> *functions $f\colon \{0,1\}^n \to \{0,1\}$;*
> *probability distributions Λ^* on $\{0, 1, 2, \ldots, R\}^{mn}|_{\leqslant\theta}$; and*
> *positive integers d with*
>
> $$d \leqslant \frac{1}{C} \min\left\{ m \deg_{\pm}(f, \gamma),\ \sqrt{r} \deg_{\pm}(f, \gamma),\ \frac{\theta}{\sqrt{r}\log(2nmR)} \right\}, \tag{5.111}$$

one has:

$$\operatorname{orth}((-1)^{f\circ \mathrm{MP}^*_{m,R}} \cdot \Lambda) \geqslant d, \tag{5.112}$$

$$\Lambda \geqslant \gamma \cdot (CnmR)^{-8d}\, \Lambda^* \tag{5.113}$$

for some $\Lambda \in \mathfrak{D}(\{0, 1, 2, \ldots, R\}^{mn}|_{\leqslant\theta})$.

Proof. Let $0 < c < 1$ be the constant from Theorem 5.13. Take $C \geqslant 1/c$ to be a sufficiently large absolute constant. By hypothesis,

$$\theta \geqslant Cnm\log(2nm). \tag{5.114}$$

Abbreviate

$$\begin{aligned} X &= \{0, 1, 2, \ldots, R\}^{nm}, \\ \delta &= 2^{-c\theta/(2\sqrt{r})}. \end{aligned} \tag{5.115}$$

The following inequalities are straightforward to verify:

$$d < \frac{1}{3}\min\{\theta - nm, nmR\}, \tag{5.116}$$

$$\theta \geqslant \frac{8enm(1+\ln(nm))}{c}, \tag{5.117}$$

$$\frac{2^{3d+1}}{c^{4d+1}}\binom{n+d}{d}^3\binom{nmR}{d}\frac{\delta}{1-\delta} < \frac{1}{2}, \tag{5.118}$$

$$2^{3d+1}\left(\frac{3m}{c}\right)^{4d+1}\binom{n+d}{d}^3\binom{nmR}{d} \leqslant \frac{(CnmR)^{8d}}{4}. \tag{5.119}$$

For example, (5.116) holds because $d \leqslant nm/C$ by (5.111) and $\theta \geqslant Cnm\log(2nm)$ by (5.114). Inequalities (5.117)–(5.119) follow analogously from (5.111) and (5.114) for a large enough constant C. The rest of the proof splits neatly into four major steps.

Step 1: Key distributions. Theorem 5.13 provides probability distributions Λ_0 and Λ_1 such that

$$\operatorname{supp}\Lambda_i = (\mathrm{MP}^*_{m,R})^{-1}(i), \qquad i = 0, 1, \tag{5.120}$$

$$\operatorname{orth}(\Lambda_0 - \Lambda_1) \geqslant \min\{m, c\sqrt{r}\}, \tag{5.121}$$

$$\frac{\Lambda_0 + \Lambda_1}{2} \in \operatorname{Smooth}\left(\frac{m}{c}, \{0, 1, 2, \ldots, R\}^m\right), \tag{5.122}$$

$$\Lambda_0, \Lambda_1 \in \operatorname{conv}\left(\left\{\lambda \in \mathfrak{S}\left(1, \frac{1}{c}, 1\right) : \lambda(t) \leqslant \frac{1}{c(t+1)^2\, 2^{ct/\sqrt{r}}} \text{ for } t \in \mathbb{N}\right\}^{\otimes m}\right). \tag{5.123}$$

Consider the probability distributions

$$\Lambda_z = \bigotimes_{i=1}^n \Lambda_{z_i}, \qquad z \in \{0,1\}^n.$$

Then

$$\begin{aligned}
\Lambda_z &\in \operatorname{conv}\left(\left\{\lambda \in \mathfrak{S}\left(1, \frac{1}{c}, 1\right) : \lambda(t) \leqslant \frac{1}{c(t+1)^2\, 2^{ct/\sqrt{r}}} \text{ for } t \in \mathbb{N}\right\}^{\otimes mn}\right) \\
&\subseteq \operatorname{conv}\left(\mathfrak{S}\left(1, \frac{1}{c}, 1\right)^{\otimes mn} \cap \right. \\
&\qquad\qquad \left. \left\{\lambda \in \mathfrak{D}(\mathbb{N}) : \lambda(t) \leqslant \frac{1}{c(t+1)^2\, 2^{ct/\sqrt{r}}} \text{ for } t \in \mathbb{N}\right\}^{\otimes mn}\right) \\
&\subseteq \operatorname{conv}\left(\mathfrak{S}\left(1, \frac{1}{c}, 1\right)^{\otimes mn} \cap \left\{\Lambda \in \mathfrak{D}(\mathbb{N}^{mn}) : \Lambda(\mathbb{N}^{nm}|_{>\theta}) \leqslant 2^{-c\theta/(2\sqrt{r})}\right\}\right) \\
&\subseteq \operatorname{conv}\left(\mathfrak{S}\left(1, \frac{1}{c}, 1\right)^{\otimes mn} \cap \left\{\Lambda \in \mathfrak{D}(\mathbb{N}^{mn}) : \Lambda(\mathbb{N}^{nm}|_{>\theta}) \leqslant \delta\right\}\right) \\
&\subseteq \operatorname{conv}\left(\mathfrak{S}\left(nm, \frac{1}{c}, nm\right) \cap \left\{\Lambda \in \mathfrak{D}(\mathbb{N}^{mn}) : \Lambda(\mathbb{N}^{nm}|_{>\theta}) \leqslant \delta\right\}\right), \quad (5.124)
\end{aligned}$$

where the first step uses (2.2) and (5.123); the third step is valid by (5.117) and Lemma 3.6; the fourth step is a substitution from (5.115); and the last step is an application of Proposition 5.4.

Step 2: Restricting the support. By (5.116), (5.124), and Corollary 5.11, there is a real function $\tilde{\Lambda}_z \colon \mathbb{N}^{nm} \to \mathbb{R}$ such that

$$\operatorname{orth}(\Lambda_z - \tilde{\Lambda}_z) > d, \tag{5.125}$$

$$\operatorname{supp} \tilde{\Lambda}_z \subseteq \mathbb{N}^{nm}|_{\leqslant \theta}, \tag{5.126}$$

$$\operatorname{supp} \tilde{\Lambda}_z \subseteq \operatorname{supp} \Lambda_z, \tag{5.127}$$

and

$$|\Lambda_z - \tilde{\Lambda}_z| \leqslant \frac{2^{3d+1}}{c^{4d+1}} \binom{n+d}{d}^3 \binom{\operatorname{diam}(\operatorname{supp} \Lambda_z)}{d} \frac{\delta}{1-\delta} \cdot \Lambda_z \qquad \text{on } \mathbb{N}^{nm}|_{\leqslant \theta}.$$

In view of (5.118) and $\operatorname{diam}(\operatorname{supp} \Lambda_z) \leqslant nmR$, the last equation simplifies to

$$|\Lambda_z - \tilde{\Lambda}_z| \leqslant \frac{1}{2} \Lambda_z \quad \text{on } \mathbb{N}^{nm}|_{\leqslant \theta}. \tag{5.128}$$

Properties (5.126) and (5.128) imply that $\tilde{\Lambda}_z$ is a nonnegative function, which along with (5.125) and Proposition 2.4 implies that $\tilde{\Lambda}_z$ is a probability distribution. Combining this fact with (5.120), (5.126), and (5.127) gives

$$\tilde{\Lambda}_z \in \mathfrak{D}\left(\mathbb{N}^{nm}|_{\leqslant \theta} \cap \prod_{i=1}^{n} (\mathrm{MP}^*_{m,R})^{-1}(z_i)\right), \qquad z \in \{0,1\}^n. \tag{5.129}$$

In particular, the $\tilde{\Lambda}_z$ are supported on disjoint sets of inputs.

Step 3: Ensuring min-smoothness. Recall from (5.129) that each of the probability distributions $\tilde{\Lambda}_z$ is supported on a subset of $X|_{\leqslant\theta}$. Consider the function $\Phi\colon X|_{\leqslant\theta} \to \mathbb{R}$ given by

$$\Phi = 2^{-n} \sum_{z\in\{0,1\}^n} (-1)^{f(z)}\tilde{\Lambda}_z.$$

Again by (5.129), the support of $\tilde{\Lambda}_z$ is contained in $\prod_{i=1}^n (\mathrm{MP}^*_{m,R})^{-1}(z_i)$. This means in particular that $f \circ \mathrm{MP}^*_{m,R} = f(z)$ on the support of $\tilde{\Lambda}_z$, whence

$$(-1)^{f(z)}\tilde{\Lambda}_z = (-1)^{f\circ\mathrm{MP}^*_{m,R}} \cdot \tilde{\Lambda}_z \tag{5.130}$$

everywhere on $X|_{\leqslant\theta}$. Making this substitution in the defining equation for Φ, we find that

$$(-1)^{f\circ\mathrm{MP}^*_{m,R}} \cdot \Phi \geqslant 0. \tag{5.131}$$

The fact that the $\tilde{\Lambda}_z$ are supported on pairwise disjoint sets of inputs forces

$$|\Phi| = 2^{-n} \sum_{z\in\{0,1\}^n} \tilde{\Lambda}_z \tag{5.132}$$

and in particular

$$\|\Phi\|_1 = 1. \tag{5.133}$$

We now examine the smoothness of Φ. For this, consider the probability distribution

$$\Lambda = 2^{-n} \sum_{z\in\{0,1\}^n} \Lambda_z. \tag{5.134}$$

Comparing equations (5.132) and (5.134) term by term and using the upper bound (5.128), we find that $|\Lambda - |\Phi|| \leqslant \frac{1}{2}\Lambda$ on $X|_{\leqslant\theta}$. Equivalently,

$$\frac{1}{2}\Lambda \leqslant |\Phi| \leqslant \frac{3}{2}\Lambda \quad \text{on } X|_{\leqslant\theta.} \tag{5.135}$$

But

$$\begin{aligned}\Lambda &= \left(\frac{1}{2}\Lambda_0 + \frac{1}{2}\Lambda_1\right)^{\otimes n} \\ &\in \mathrm{Smooth}\left(\frac{m}{c}, \{0,1,2,\dots,R\}^m\right)^{\otimes n} \\ &\subseteq \mathrm{Smooth}\left(\frac{m}{c}, \{0,1,2,\dots,R\}^{mn}\right),\end{aligned} \tag{5.136}$$

where the last two steps are valid by (5.122) and Proposition 5.3(iii), respectively. Combining (5.135) and (5.136), we conclude that Φ is $(3m/c)$-smooth on $X|_{\leqslant\theta}$. As a result, (5.116) and Lemma 5.12 provide a function $\Phi^*\colon X|_{\leqslant\theta} \to \mathbb{R}$ with

$$\operatorname{orth}(\Phi - \Phi^*) > d, \tag{5.137}$$

$$\|\Phi^*\|_1 \leqslant 2\|\Phi\|_1, \tag{5.138}$$

$$\Phi \cdot \Phi^* \geqslant 0, \tag{5.139}$$

$$|\Phi^*| \geqslant \left(2^{3d+1}\left(\frac{3m}{c}\right)^{4d+1}\binom{n+d}{d}^3\binom{\operatorname{diam}(\operatorname{supp}\Phi)}{d}\right)^{-1}\|\Phi\|_1\,\Lambda^*. \tag{5.140}$$

In view of (5.133), the second property simplifies to

$$\|\Phi^*\|_1 \leqslant 2. \tag{5.141}$$

Recall that on $X|_{\leqslant\theta}$, the function Φ is $(3m/c)$-smooth and not identically zero. Therefore, Φ must be nonzero at every point of $X|_{\leqslant\theta}$, which includes the support of Φ^*. As a result, (5.131) and (5.139) imply that

$$(-1)^{f\circ \mathrm{MP}_{m,R}} \cdot \Phi^* \geqslant 0. \tag{5.142}$$

Finally, using $\operatorname{diam}(\operatorname{supp}\Phi) \leqslant nmR$ along with the bounds (5.119) and (5.133), we can restate (5.140) as

$$|\Phi^*| \geqslant 4(CnmR)^{-8d}\Lambda^*. \tag{5.143}$$

Step 4: The final construction. By the definition of smooth threshold degree, there is a probability distribution μ on $\{0,1\}^n$ such that

$$\operatorname{orth}((-1)^f \cdot \mu) \geqslant \deg_\pm(f,\gamma), \tag{5.144}$$

$$\mu(z) \geqslant \gamma \cdot 2^{-n}, \qquad z \in \{0,1\}^n. \tag{5.145}$$

Define

$$\Phi_{\text{final}} = \sum_{z\in\{0,1\}^n} \mu(z)(-1)^{f(z)}\tilde{\Lambda}_z - \gamma\Phi + \gamma\Phi^*.$$

The right-hand side is a linear combination of functions on $X|_{\leqslant\theta}$, whence

$$\operatorname{supp}(\Phi_{\text{final}}) \subseteq X|_{\leqslant\theta}. \tag{5.146}$$

Moreover,

$$\begin{aligned}\|\Phi_{\text{final}}\|_1 &\leqslant \sum_{z\in\{0,1\}^n} \mu(z)\|\tilde{\Lambda}_z\|_1 + \gamma\|\Phi\|_1 + \gamma\|\Phi^*\|_1 \\ &\leqslant 1 + 3\gamma \\ &\leqslant 4, \end{aligned} \tag{5.147}$$

where the first step applies the triangle inequality, and the second step uses (5.129), (5.133) and (5.141). Continuing,

$$\begin{aligned}&(-1)^{f\circ \mathrm{MP}^*_{m,R}} \cdot \Phi_{\text{final}} \\ &\quad= (-1)^{f\circ \mathrm{MP}^*_{m,R}} \cdot \left(\sum_{z\in\{0,1\}^n} (\mu(z) - \gamma 2^{-n})(-1)^{f(z)}\tilde{\Lambda}_z + \gamma\Phi^* \right) \\ &\quad= \sum_{z\in\{0,1\}^n} (\mu(z) - \gamma 2^{-n})(-1)^{f\circ \mathrm{MP}^*_{m,R}} \cdot (-1)^{f(z)}\tilde{\Lambda}_z + \gamma(-1)^{f\circ \mathrm{MP}^*_{m,R}} \cdot \Phi^* \\ &\quad= \sum_{z\in\{0,1\}^n} (\mu(z) - \gamma 2^{-n})\tilde{\Lambda}_z + \gamma|\Phi^*| \qquad (5.148) \\ &\quad\geqslant \gamma|\Phi^*| \\ &\quad\geqslant 4\gamma(CnmR)^{-8d}\Lambda^*, \qquad (5.149)\end{aligned}$$

where the first step applies the definition of Φ; the third step uses (5.130) and (5.142); the fourth step follows from (5.145); and the fifth step substitutes the lower bound from (5.143). Now

$$\Phi_{\text{final}} \not\equiv 0 \tag{5.150}$$

follows from (5.148) if $\gamma = 0$, and from (5.149) if $\gamma > 0$.

It remains to examine the orthogonal content of Φ_{final}. For this, write

$$\begin{aligned}\Phi_{\text{final}} = \sum_{z\in\{0,1\}^n} \mu(z)(-1)^{f(z)}\Lambda_z + \sum_{z\in\{0,1\}^n} \mu(z)(-1)^{f(z)}(\tilde{\Lambda}_z - \Lambda_z) \\ + \gamma(\Phi^* - \Phi).\end{aligned}$$

Then

$$\begin{aligned}
\operatorname{orth}(\Phi_{\text{final}}) &\geqslant \min\left\{\operatorname{orth}\left(\sum_{z\in\{0,1\}^n} \mu(z)(-1)^{f(z)}\Lambda_z\right),\right. \\
&\qquad\qquad \left.\min_z\{\operatorname{orth}(\tilde{\Lambda}_z - \Lambda_z)\},\ \operatorname{orth}(\Phi^* - \Phi)\right\} \\
&\geqslant \min\left\{\operatorname{orth}\left(\sum_{z\in\{0,1\}^n} \mu(z)(-1)^{f(z)}\Lambda_z\right), d\right\} \\
&\geqslant \min\left\{\operatorname{orth}\left(\sum_{z\in\{0,1\}^n} \mu(z)(-1)^{f(z)}\bigotimes_{i=1}^n \Lambda_{z_i}\right), d\right\} \\
&\geqslant \min\left\{\operatorname{orth}(\mu\cdot(-1)^f)\operatorname{orth}(\Lambda_1 - \Lambda_0), d\right\} \\
&\geqslant \min\{\deg_\pm(f,\gamma)\min\{m, c\sqrt{r}\}, d\} \\
&= d, \qquad (5.151)
\end{aligned}$$

where the first step applies Proposition 2.1(i); the second step follows from (5.125) and (5.137); the third step is valid by the definition of Λ_z; the fourth step applies Corollary 2.3; the fifth step substitutes the lower bounds from (5.121) and (5.144); and the final step uses (5.111).

To complete the proof, let

$$\Lambda = \frac{\Phi_{\text{final}}}{\|\Phi_{\text{final}}\|_1}\cdot(-1)^{f\circ \mathrm{MP}^*_{m,R}},$$

where the right-hand side is well-defined by (5.150). Then $\|\Lambda\|_1 = 1$ by definition. Moreover, (5.146) and (5.149) guarantee that Λ is a nonnegative function with support contained in $X|_{\leqslant\theta}$, so that $\Lambda \in \mathfrak{D}(X|_{\leqslant\theta})$. The orthogonality property (5.112) follows from (5.151), whereas the min-smoothness property (5.113) follows from (5.147) and (5.149). □

We now translate the new amplification theorem from $\mathbb{N}^n|_{\leqslant\theta}$ to the hypercube, using the input transformation scheme of Theorem 3.9.

THEOREM 5.19. *Let $C \geqslant 1$ be the absolute constant from Theorem 5.18. Fix positive integers n, m, θ with $\theta \geqslant Cnm\log(2nm)$. Then there is an (explicitly given) transformation $H\colon \{0,1\}^{6\theta\lceil\log(nm+1)\rceil} \to \{0,1\}^n$, computable by an AND-OR-AND circuit of polynomial size with bottom fan-in at most $6\lceil\log(nm+1)\rceil$, such that*

$$\deg_\pm(f\circ H, \gamma\theta^{-24d}) \geqslant d\lceil\log(nm+1)+1\rceil, \qquad (5.152)$$

$$\deg_\pm(f\circ \neg H, \gamma\theta^{-24d}) \geqslant d\lceil\log(nm+1)+1\rceil \qquad (5.153)$$

for all Boolean functions $f\colon \{0,1\}^n \to \{0,1\}$, *all real numbers* $\gamma \in [0,1]$, *and all positive integers*

$$d \leqslant \frac{1}{C} \min\left\{ m \deg_{\pm}(f,\gamma), \frac{\theta}{4m \log \theta} \right\}.$$

Proof. Negating a function's input bits has no effect on its γ-smooth threshold degree for any $0 \leqslant \gamma \leqslant 1$, so that $f(x_1, x_2, \ldots, x_n)$ and $f(\neg x_1, \neg x_2, \ldots, \neg x_n)$ both have γ-smooth threshold degree $\deg_{\pm}(f,\gamma)$. Therefore, proving (5.152) for all f will also settle (5.153) for all f. In what follows, we focus on the former.

Theorem 3.9 constructs an explicit surjection $G\colon \{0,1\}^N \to \mathbb{N}^{nm}|_{\leqslant\theta}$ on $N = 6\theta\lceil \log(nm+1)\rceil$ variables with the following two properties:

(i) for every coordinate $i = 1, 2, \ldots, nm$, the mapping $x \mapsto \mathrm{OR}^*_\theta(G(x)_i)$ is computable by a DNF formula of size $(nm\theta)^{O(1)} = \theta^{O(1)}$ with bottom fan-in at most $6\lceil \log(nm+1)\rceil$;

(ii) for any polynomial p, the map $v \mapsto \mathbf{E}_{G^{-1}(v)}\, p$ is a polynomial on $\mathbb{N}^{nm}|_{\leqslant\theta}$ of degree at most $(\deg p)/\lceil \log(nm+1)+1\rceil$.

Consider the composition $F = (f \circ \mathrm{MP}^*_{m,\theta}) \circ G$. Then

$$\begin{aligned} F &= (f \circ (\mathrm{AND}_m \circ \mathrm{OR}^*_\theta)) \circ G \\ &= f \circ (\underbrace{(\mathrm{AND}_m \circ \mathrm{OR}^*_\theta, \ldots, \mathrm{AND}_m \circ \mathrm{OR}^*_\theta}_{n}) \circ G), \end{aligned}$$

which by property (i) of G means that F is the composition of f and an AND-OR-AND circuit H of size $(nm\theta)^{O(1)} = \theta^{O(1)}$ and bottom fan-in $6\lceil \log(nm+1)\rceil$. Hence, the proof will be complete once we show that

$$\deg_{\pm}(F, \gamma\theta^{-24d}) \geqslant d\lceil \log(nm+1)+1\rceil. \tag{5.154}$$

Define $r = m^2$ and $R = \max\{\theta, r\}$, and consider the probability distribution on $\{0,1,2,\ldots,R\}^{nm}|_{\leqslant\theta} = \mathbb{N}^{nm}|_{\leqslant\theta}$ given by $\Lambda^*(v) = |G^{-1}(v)|/2^N$. Then Theorem 5.18 constructs a probability distribution Λ on $\mathbb{N}^{nm}|_{\leqslant\theta}$ such that

$$\mathrm{orth}((-1)^{f\circ \mathrm{MP}^*_{m,R}} \cdot \Lambda) \geqslant d, \tag{5.155}$$

$$\Lambda \geqslant \gamma\theta^{-24d}\, \Lambda^*. \tag{5.156}$$

In view of $R \geqslant \theta$, inequality (5.155) can be restated as

$$\mathrm{orth}((-1)^{f\circ \mathrm{MP}^*_{m,\theta}} \cdot \Lambda) \geqslant d. \tag{5.157}$$

Define

$$\lambda = \sum_{v \in \mathbb{N}^{nm}|_{\leqslant\theta}} \Lambda(v) \cdot \frac{\mathbf{1}_{G^{-1}(v)}}{|G^{-1}(v)|},$$

where $\mathbf{1}_{G^{-1}(v)}$ denotes as usual the characteristic function of the set $G^{-1}(v)$. Clearly, λ is a probability distribution on $\{0,1\}^N$. Moreover,

$$\begin{aligned}\lambda &\geqslant \gamma\theta^{-24d} \sum_{v \in \mathbb{N}^{nm}|_{\leqslant\theta}} \Lambda^*(v) \cdot \frac{\mathbf{1}_{G^{-1}(v)}}{|G^{-1}(v)|} \\ &= \gamma\theta^{-24d} \sum_{v \in \mathbb{N}^{nm}|_{\leqslant\theta}} \frac{|G^{-1}(v)|}{2^N} \cdot \frac{\mathbf{1}_{G^{-1}(v)}}{|G^{-1}(v)|} \\ &= \gamma\theta^{-24d} \cdot \frac{\mathbf{1}_{\{0,1\}^N}}{2^N}, \end{aligned} \tag{5.158}$$

where the first two steps use (5.156) and the definition of Λ^*, respectively.

Finally, we examine the orthogonal content of $(-1)^F \cdot \lambda$. Let $p\colon \mathbb{R}^N \to \mathbb{R}$ be any polynomial of degree less than $d\lceil \log(nm+1)+1\rceil$. Then by property (ii) of G, the mapping $p^*\colon v \mapsto \mathbf{E}_{G^{-1}(v)}\, p$ is a polynomial on $\mathbb{N}^{nm}|_{\leqslant\theta}$ of degree less than d. As a result,

$$\begin{aligned}\langle (-1)^F \cdot \lambda, p\rangle &= \langle (-1)^{(f\circ \mathrm{MP}^*_{m,\theta})\circ G} \cdot \lambda, p\rangle \\ &= \sum_{v \in \mathbb{N}^{nm}|_{\leqslant\theta}} \sum_{G^{-1}(v)} (-1)^{(f\circ \mathrm{MP}^*_{m,\theta})\circ G} \cdot \lambda \cdot p \\ &= \sum_{v \in \mathbb{N}^{nm}|_{\leqslant\theta}} (-1)^{(f\circ \mathrm{MP}^*_{m,\theta})(v)} \sum_{G^{-1}(v)} \lambda \cdot p \\ &= \sum_{v \in \mathbb{N}^{nm}|_{\leqslant\theta}} (-1)^{(f\circ \mathrm{MP}^*_{m,\theta})(v)} \Lambda(v) \mathop{\mathbf{E}}_{G^{-1}(v)} p \\ &= \langle (-1)^{f\circ \mathrm{MP}^*_{m,\theta}} \cdot \Lambda, p^*\rangle \\ &= 0, \end{aligned}$$

where the last step uses (5.157) and $\deg p^* < d$. We conclude that $\operatorname{orth}((-1)^F \cdot \lambda) \geqslant d\lceil \log(nm+1)+1\rceil$, which along with (5.158) settles (5.154). □

5.7. The smooth threshold degree of $\mathbf{AC^0}$. We now construct, for any $\epsilon > 0$, a constant-depth circuit $f\colon \{0,1\}^n \to \{0,1\}$ with $\exp(-n^{1-\epsilon})$-smooth threshold degree $\Omega(n^{1-\epsilon})$. This result may find applications in future work, in addition to its use in this paper to obtain a lower bound on the sign-rank of AC^0. The proof proceeds by induction, with the amplification theorem for smooth threshold degree (Theorem 5.19) applied repeatedly to construct increasingly harder circuits. To simplify the exposition, we isolate the inductive step in the following lemma.

LEMMA 5.20. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean circuit of size s, depth d, and smooth threshold degree*

$$\deg_\pm \left(f, \exp\left(-c' \cdot \frac{n^{1-\alpha}}{\log^\beta n} \right) \right) \geqslant c'' \cdot \frac{n^{1-\alpha}}{\log^\beta n},$$

for some real numbers $\alpha \in [0,1]$, $\beta \geqslant 0$, and $c', c'' > 0$. Then f can be transformed in polynomial time into a Boolean circuit $F\colon \{0,1\}^N \to \{0,1\}$ on $N = \Theta(n^{1+\alpha}\log^{2+\beta} n)$ variables that has size $s + N^{O(1)}$, depth at most $d+3$, bottom

fan-in $O(\log n)$, and smooth threshold degree

$$\deg_{\pm}\left(F, \exp\left(-C' \cdot \frac{N^{\frac{1}{1+\alpha}}}{\log^{\frac{1-\alpha+\beta}{1+\alpha}} N}\right)\right) \geqslant C'' \cdot \frac{N^{\frac{1}{1+\alpha}}}{\log^{\frac{1-\alpha+\beta}{1+\alpha}} N}, \tag{5.159}$$

where $C', C'' > 0$ are real numbers that depend on c', c'' only. Moreover, if the circuit for f is monotone with AND gates at the bottom, then the depth of F is at most $d+2$.

Proof. Let $C \geqslant 1$ be the absolute constant from Theorem 5.18. Apply Theorem 5.19 with

$$\begin{aligned} m &= \lceil n^{\alpha} \log^{\beta} n \rceil, \\ \theta &= \lceil Cmn \log(2nm) \rceil, \\ \gamma &= \exp\left(-c' \cdot \frac{n^{1-\alpha}}{\log^{\beta} n}\right) \end{aligned}$$

to obtain a function $H\colon \{0,1\}^N \to \{0,1\}^n$ on $N = \Theta(n^{1+\alpha} \log^{2+\beta} n)$ variables such that the composition $F = f \circ H$ satisfies (5.159) for some $C', C'' > 0$ that depend only on c', c'', and furthermore H is computable by an AND-OR-AND circuit of polynomial size and bottom fan-in $O(\log N)$. Clearly, the composition $F = f \circ H$ is a circuit of size $s + N^{O(1)}$, depth $d+3$, and bottom fan-in $O(\log N)$. Moreover, if the circuit for f is monotone with AND gates at the bottom level, then the bottom level of f can be merged with the top level of H to reduce the depth of $F = f \circ H$ to at most $(d+3) - 1 = d+2$. □

We now obtain our lower bounds on the smooth threshold degree of AC^0. We present two incomparable theorems here, both of which apply Lemma 5.20 in a recursive manner but with different base cases.

THEOREM 5.21. *Let $k \geqslant 0$ be a given integer. Then there is an (explicitly given) circuit family $\{f_n\}_{n=1}^{\infty}$, where $f_n\colon \{0,1\}^n \to \{0,1\}$ has polynomial size, depth $3k$, bottom fan-in $O(\log n)$, and smooth threshold degree*

$$\deg_{\pm}\left(f_n, \exp\left(-c' \cdot \frac{n^{1-\frac{1}{k+1}}}{\log^{\frac{k(k-1)}{2(k+1)}} n}\right)\right) \geqslant c'' \cdot \frac{n^{1-\frac{1}{k+1}}}{\log^{\frac{k(k-1)}{2(k+1)}} n} \tag{5.160}$$

for some constants $c', c'' > 0$ and all $n \geqslant 2$.

Proof. The proof is by induction on k. The base case $k = 0$ corresponds to the family of "dictator" functions $x \mapsto x_1$, each of which has 1/2-smooth threshold degree 1 by Fact 2.8. For the inductive step, fix an explicit circuit family $\{f_n\}_{n=1}^{\infty}$ in which $f_n\colon \{0,1\}^n \to \{0,1\}$ has polynomial size, depth $3k$, and smooth threshold degree (5.160) for some constants $c', c'' > 0$. Then taking $\alpha = \frac{1}{k+1}$ and $\beta = \frac{k(k-1)}{2(k+1)}$ in Lemma 5.20 produces an explicit circuit family $\{F_n\}_{n=1}^{\infty}$ in which $F_n\colon \{0,1\}^n \to$

$\{0,1\}$ has polynomial size, depth $3k+3 = 3(k+1)$, and smooth threshold degree

$$\deg_{\pm}\left(F_n, \exp\left(-C' \cdot \frac{n^{\frac{k+1}{k+2}}}{\log^{\frac{k(k+1)}{2(k+2)}} n}\right)\right) \geqslant C'' \cdot \frac{n^{\frac{k+1}{k+2}}}{\log^{\frac{k(k+1)}{2(k+2)}} n}$$

for some constants $C', C'' > 0$. This completes the inductive step. □

THEOREM 5.22. *Let $k \geqslant 1$ be a given integer. Then there is an (explicitly given) circuit family $\{f_n\}_{n=1}^{\infty}$, where $f_n\colon \{0,1\}^n \to \{0,1\}$ has polynomial size, depth $3k+1$, bottom fan-in $O(\log n)$, and smooth threshold degree*

$$\deg_{\pm}\left(f_n, \exp\left(-c' \cdot \frac{n^{1-\frac{2}{2k+3}}}{\log^{\frac{k^2}{2k+3}} n}\right)\right) \geqslant c'' \cdot \frac{n^{1-\frac{2}{2k+3}}}{\log^{\frac{k^2}{2k+3}} n} \tag{5.161}$$

for some constants $c', c'' > 0$ and all $n \geqslant 2$.

Proof. As with Theorem 5.21, the proof is by induction on k. For the base case $k = 1$, consider the family $\{g_n\}_{n=1}^{\infty}$ in which $g_n\colon \{0,1\}^n \to \{0,1\}$ is given by

$$g_n(x) = \bigvee_{i=1}^{\lfloor n^{1/3} \rfloor} \bigwedge_{j=1}^{\lfloor n^{2/3} \rfloor} x_{i,j}.$$

Then

$$\begin{aligned}\deg_{\pm}(g_n, 12^{-\lfloor n^{1/3} \rfloor - 1}) &= \deg_{\pm}(\mathrm{MP}_{\lfloor n^{1/3} \rfloor, \lfloor n^{2/3} \rfloor}, 12^{-\lfloor n^{1/3} \rfloor - 1}) \\ &\geqslant cn^{1/3}\end{aligned}$$

for some absolute constant $c > 0$, where the first step is valid because a function's smooth threshold degree remains unchanged when one negates the function or its input variables, and the second step uses Theorem 5.1. Applying Lemma 5.20 to the circuit family $\{g_n\}_{n=1}^{\infty}$ with $\alpha = 2/3$ and $\beta = 0$ yields an explicit circuit family $\{G_n\}_{n=1}^{\infty}$ in which $G_n\colon \{0,1\}^n \to \{0,1\}$ has polynomial size, depth $2 + 2 = 4$, bottom fan-in $O(\log n)$, and smooth threshold degree

$$\deg_{\pm}\left(G_n, \exp\left(-C' \cdot \frac{n^{3/5}}{\log^{1/5} n}\right)\right) \geqslant C'' \cdot \frac{n^{3/5}}{\log^{1/5} n}$$

for some constants $C', C'' > 0$. This new circuit family $\{G_n\}_{n=1}^{\infty}$ establishes the base case.

For the inductive step, fix an integer $k \geqslant 1$ and an explicit circuit family $\{f_n\}_{n=1}^{\infty}$ in which $f_n\colon \{0,1\}^n \to \{0,1\}$ has polynomial size, depth $3k+1$, and smooth threshold degree (5.161) for some constants $c', c'' > 0$. Applying Lemma 5.20 with $\alpha = 2/(2k+3)$ and $\beta = k^2/(2k+3)$ yields an explicit circuit family $\{F_n\}_{n=1}^{\infty}$, where $F_n\colon \{0,1\}^n \to \{0,1\}$ has polynomial size, depth $(3k+1)+3 = 3(k+1)+1$, bottom

fan-in $O(\log n)$, and smooth threshold degree

$$\deg_{\pm}\left(F_n, \exp\left(-C''' \cdot \frac{n^{\frac{2k+3}{2k+5}}}{\log^{\frac{(k+1)^2}{2k+5}} n}\right)\right) \geqslant C'''' \cdot \frac{n^{\frac{2k+3}{2k+5}}}{\log^{\frac{(k+1)^2}{2k+5}} n}$$

for some constants $C''', C'''' > 0$. This completes the inductive step. □

5.8. The sign-rank of AC^0. We have reached our main result on the sign-rank and unbounded-error communication complexity of constant-depth circuits. The proof amounts to lifting, by means of Theorem 2.17, the lower bounds on the smooth threshold degree in Theorems 5.21 and 5.22 to sign-rank lower bounds.

THEOREM 5.23. *Let $k \geqslant 1$ be a given 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 $3k$, bottom fan-in $O(\log n)$, sign-rank*

$$\mathrm{rk}_{\pm}(F_n) = \exp\left(\Omega\left(n^{1-\frac{1}{k+1}} \cdot (\log n)^{-\frac{k(k-1)}{2(k+1)}}\right)\right), \tag{5.162}$$

and unbounded-error communication complexity

$$\mathrm{UPP}(F_n) = \Omega\left(n^{1-\frac{1}{k+1}} \cdot (\log n)^{-\frac{k(k-1)}{2(k+1)}}\right). \tag{5.163}$$

Proof. Theorem 5.21 constructs a circuit family $\{f_n\}_{n=1}^{\infty}$ in which $f_n \colon \{0,1\}^n \to \{0,1\}$ has polynomial size, depth $3k$, bottom fan-in $O(\log n)$, and smooth threshold degree (5.160) for some constants $c', c'' > 0$ and all $n \geqslant 2$. Abbreviate $m = 2\lceil \exp(4c'/c'') \rceil$. For any $n \geqslant m$, define $F_n = f_{\lfloor n/m \rfloor} \circ \mathrm{OR}_m \circ \mathrm{AND}_2$. Then (5.162) is immediate from (5.160) and Theorem 2.17. Combining (5.163) with Theorem (2.15) settles (5.163).

It remains to analyze the circuit complexity of F_n. We defined F_n formally as a circuit of depth $3k+2$ in which the bottom four levels have fan-ins $n^{O(1)}$, $O(\log n)$, $2m$, and 2, in that order. Since m is a constant independent of n, these four levels can be computed by a circuit of polynomial size, depth 2, and bottom fan-in $O(\log n)$. This optimization reduces the depth of F_n to $(3k+2) - 4 + 2 = 3k$ while keeping the bottom fan-in at $O(\log n)$. □

We now similarly lift Theorem 5.22 to a lower bound on sign-rank and unbounded-error communication complexity.

THEOREM 5.24. *Let $k \geqslant 1$ be a given 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 $3k+1$, bottom fan-in $O(\log n)$, sign-rank*

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

and unbounded-error communication complexity

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

Proof. The proof is analogous to that of Theorem 5.23, with the only difference that the appeal to Theorem 5.21 should be replaced with an appeal to Theorem 5.22. □

Theorems 5.23 and 5.24 settle Theorems 1.2, 1.3, and 1.5 in the introduction.

Acknowledgments

The authors are thankful to Mark Bun and Justin Thaler for valuable comments on an earlier version of this paper.

References

[1] S. Aaronson and Y. Shi, *Quantum lower bounds for the collision and the element distinctness problems*, J. ACM, 51 (2004), pp. 595–605, doi:10.1145/1008731.1008735.

[2] N. Alon, P. Frankl, and V. Rödl, *Geometrical realization of set systems and probabilistic communication complexity*, in *Proceedings of the Twenty-Sixth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 1985, pp. 277–280, doi:10.1109/SFCS.1985.30.

[3] A. Ambainis, *Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range*, Theory of Computing, 1 (2005), pp. 37–46, doi:10.4086/toc.2005.v001a003.

[4] A. Ambainis, A. M. Childs, B. Reichardt, R. Špalek, and S. Zhang, *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 (2010), pp. 2513–2530, doi:10.1137/080712167.

[5] J. Aspnes, R. Beigel, M. L. Furst, and S. Rudich, *The expressive power of voting polynomials*, Combinatorica, 14 (1994), pp. 135–148, doi:10.1007/BF01215346.

[6] L. Babai, P. Frankl, and J. Simon, *Complexity classes in communication complexity theory*, in *Proceedings of the Twenty-Seventh Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 1986, pp. 337–347, doi:10.1109/SFCS.1986.15.

[7] L. Babai, N. Nisan, and M. Szegedy, *Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs*, J. Comput. Syst. Sci., 45 (1992), pp. 204–232, doi:10.1016/0022-0000(92)90047-M.

[8] P. Beame and T. Huynh, *Multiparty communication complexity and threshold circuit size of* AC^0, SIAM J. Comput., 41 (2012), pp. 484–518, doi:10.1137/100792779.

[9] P. Beame and W. Machmouchi, *The quantum query complexity of* AC^0, Quantum Information & Computation, 12 (2012), pp. 670–676.

[10] R. Beigel, N. Reingold, and D. A. Spielman, PP *is closed under intersection*, J. Comput. Syst. Sci., 50 (1995), pp. 191–202, doi:10.1006/jcss.1995.1017.

[11] S. Ben-David, N. Eiron, and H. U. Simon, *Limitations of learning via embeddings in Euclidean half spaces*, J. Mach. Learn. Res., 3 (2003), pp. 441–461.

[12] H. Buhrman, N. K. Vereshchagin, and R. de Wolf, *On computation and communication with small bias*, in *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity* (CCC), 2007, pp. 24–32, doi:10.1109/CCC.2007.18.

[13] M. Bun, R. Kothari, and J. Thaler, *The polynomial method strikes back: Tight quantum query bounds via dual polynomials*. ECCC Report TR17-169, 2017.

[14] M. Bun and J. Thaler, *Dual lower bounds for approximate degree and Markov–Bernstein inequalities*, Inf. Comput., 243 (2015), pp. 2–25, doi:10.1016/j.ic.2014.12.003.

[15] M. Bun and J. Thaler, *Hardness amplification and the approximate degree of constant-depth circuits*, in *Proceedings of the Forty-Second International Colloquium on Automata, Languages and Programming* (ICALP), 2015, pp. 268–280, doi:10.1007/978-3-662-47672-7_22.

[16] M. Bun and J. Thaler, *Approximate degree and the complexity of depth three circuits*, in Electronic Colloquium on Computational Complexity (ECCC), 2016. Report TR16-121.

[17] M. Bun and J. Thaler, *Improved bounds on the sign-rank of* AC^0, in *Proceedings of the Forty-Third International Colloquium on Automata, Languages and Programming* (ICALP), 2016, pp. 37:1–37:14, doi:10.4230/LIPIcs.ICALP.2016.37.

[18] M. Bun and J. Thaler, *A nearly optimal lower bound on the approximate degree of* AC^0, in *Proceedings of the Fifty-Eighth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2017, pp. 1–12, doi:10.1109/FOCS.2017.10.

[19] M. Bun and J. Thaler, *The large-error approximate degree of* AC^0, in Electronic Colloquium on Computational Complexity (ECCC), August 2018. Report TR18-143.

[20] A. K. Chandra, M. L. Furst, and R. J. Lipton, *Multi-party protocols*, in *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing* (STOC), 1983, pp. 94–99, doi:10.1145/800061.808737.

[21] A. Chattopadhyay and A. Ada, *Multiparty communication complexity of disjointness*, in Electronic Colloquium on Computational Complexity (ECCC), January 2008. Report TR08-002.

[22] B. Chor and O. Goldreich, *Unbiased bits from sources of weak randomness and probabilistic communication complexity*, SIAM J. Comput., 17 (1988), pp. 230–261, doi:10.1137/0217015.

[23] J. Forster, *A linear lower bound on the unbounded error probabilistic communication complexity*, J. Comput. Syst. Sci., 65 (2002), pp. 612–625, doi:10.1016/S0022-0000(02)00019-3.

[24] J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H.-U. Simon, *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), 2001, pp. 171–182, doi:10.1007/3-540-45294-X_15.

[25] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, 2nd ed., 1994.

[26] S. Jukna, *Extremal Combinatorics with Applications in Computer Science*, Springer-Verlag Berlin Heidelberg, 2nd ed., 2011, doi:10.1007/978-3-642-17364-6.

[27] A. R. Klivans, R. O'Donnell, and R. A. Servedio, *Learning intersections and thresholds of halfspaces*, J. Comput. Syst. Sci., 68 (2004), pp. 808–840, doi:10.1016/j.jcss.2003.11.002.

[28] A. R. Klivans and R. A. Servedio, *Learning DNF in time* $2^{\tilde{O}(n^{1/3})}$, J. Comput. Syst. Sci., 68 (2004), pp. 303–318, doi:10.1016/j.jcss.2003.07.007.

[29] A. R. Klivans and R. A. Servedio, *Toward attribute efficient learning of decision lists and parities*, J. Machine Learning Research, 7 (2006), pp. 587–602.

[30] M. Krause and P. Pudlák, *On the computational power of depth-2 circuits with threshold and modulo gates*, Theor. Comput. Sci., 174 (1997), pp. 137–156, doi:10.1016/S0304-3975(96)00019-9.

[31] M. Krause and P. Pudlák, *Computing Boolean functions by polynomials and threshold circuits*, Comput. Complex., 7 (1998), pp. 346–370, doi:10.1007/s000370050015.

[32] E. Kushilevitz and N. Nisan, *Communication complexity*, Cambridge University Press, 1997.

[33] T. Lee, *A note on the sign degree of formulas*, 2009. Available at `http://arxiv.org/abs/0909.4607`.

[34] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman, *Complexity measures of sign matrices*, Combinatorica, 27 (2007), pp. 439–463, doi:10.1007/s00493-007-2160-5.

[35] M. L. Minsky and S. A. Papert, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge, Mass., 1969.

[36] N. Nisan and M. Szegedy, *On the degree of Boolean functions as real polynomials*, Computational Complexity, 4 (1994), pp. 301–313, doi:10.1007/BF01263419.

[37] R. O'Donnell and R. A. Servedio, *Extremal properties of polynomial threshold functions*, J. Comput. Syst. Sci., 74 (2008), pp. 298–312, doi:10.1016/j.jcss.2007.06.021.

[38] R. O'Donnell and R. A. Servedio, *New degree bounds for polynomial threshold functions*, Combinatorica, 30 (2010), pp. 327–358, doi:10.1007/s00493-010-2173-3.

[39] R. Paturi, *On the degree of polynomials that approximate symmetric Boolean functions*, in *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 1992, pp. 468–474, doi:10.1145/129712.129758.

[40] R. Paturi and M. E. Saks, *Approximating threshold circuits by rational functions*, Inf. Comput., 112 (1994), pp. 257–272, doi:10.1006/inco.1994.1059.

[41] R. Paturi and J. Simon, *Probabilistic communication complexity*, J. Comput. Syst. Sci., 33 (1986), pp. 106–123, doi:10.1016/0022-0000(86)90046-2.

[42] A. A. Razborov and A. A. Sherstov, *The sign-rank of* AC^0, SIAM J. Comput., 39 (2010), pp. 1833–1855, doi:10.1137/080744037. Preliminary version in *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2008.

[43] M. E. Saks, *Slicing the hypercube*, Surveys in Combinatorics, (1993), pp. 211–255, doi:10.1017/CBO9780511662089.009.

[44] A. A. Sherstov, *Halfspace matrices*, Computational Complexity, 17 (2008), pp. 149–178, doi:10.1007/s00037-008-0242-4. Preliminary version in *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity* (CCC), 2007.

[45] A. A. Sherstov, *Separating* AC^0 *from depth-2 majority circuits*, SIAM J. Comput., 38 (2009), pp. 2113–2129, doi:10.1137/08071421X. Preliminary version in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing* (STOC), 2007.

[46] A. A. Sherstov, *Communication complexity under product and nonproduct distributions*, Computational Complexity, 19 (2010), pp. 135–150, doi:10.1007/s00037-009-0285-1. Preliminary version in *Proceedings of the Twenty-Third Annual IEEE Conference on Computational Complexity* (CCC), 2008.

[47] A. A. Sherstov, *The pattern matrix method*, SIAM J. Comput., 40 (2011), pp. 1969–2000, doi:10.1137/080733644. Preliminary version in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing* (STOC), 2008.

[48] A. A. Sherstov, *The unbounded-error communication complexity of symmetric functions*, Combinatorica, 31 (2011), pp. 583–614, doi:10.1007/s00493-011-2580-0. Preliminary version in *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2008.

[49] A. A. Sherstov, *Strong direct product theorems for quantum communication and query complexity*, SIAM J. Comput., 41 (2012), pp. 1122–1165, doi:10.1137/110842661. Preliminary version in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (STOC), 2011.

[50] A. A. Sherstov, *The intersection of two halfspaces has high threshold degree*, SIAM J. Comput., 42 (2013), pp. 2329–2374, doi:10.1137/100785260. Preliminary version in *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2009.

[51] A. A. Sherstov, *Making polynomials robust to noise*, Theory of Computing, 9 (2013), pp. 593–615, doi:10.4086/toc.2013.v009a018. Preliminary version in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 2012.

[52] A. A. Sherstov, *Optimal bounds for sign-representing the intersection of two halfspaces by polynomials*, Combinatorica, 33 (2013), pp. 73–96, doi:10.1007/s00493-013-2759-7. Preliminary version in *Proceedings of the Forty-Second Annual ACM Symposium on Theory of Computing* (STOC), 2010.

[53] A. A. Sherstov, *Breaking the Minsky–Papert barrier for constant-depth circuits*, in *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* (STOC), 2014, pp. 223–232, doi:10.1145/2591796.2591871. Full version available as ECCC Report TR14-009, January 2014.

[54] A. A. Sherstov, *Communication lower bounds using directional derivatives*, J. ACM, 61 (2014), pp. 1–71, doi:10.1145/2629334. Preliminary version in *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing* (STOC), 2013.

[55] A. A. Sherstov, *The power of asymmetry in constant-depth circuits*, in *Proceedings of the Fifty-Sixth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2015, pp. 431–450, doi:10.1109/FOCS.2015.34.

[56] A. A. Sherstov, *The multiparty communication complexity of set disjointness*, SIAM J. Comput., 45 (2016), pp. 1450–1489, doi:10.1137/120891587. Preliminary version in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 2009.

[57] A. A. Sherstov, *Algorithmic polynomials*, in *Proceedings of the Fiftieth Annual ACM Symposium on Theory of Computing* (STOC), 2018, pp. 311–324, doi:10.1145/3188745.3188958.

[58] K.-Y. Siu, V. P. Roychowdhury, and T. Kailath, *Rational approximation techniques for analysis of neural networks*, IEEE Transactions on Information Theory, 40 (1994), pp. 455–466, doi:10.1109/18.312168.

[59] J. Thaler, *Lower bounds for the approximate degree of block-composed functions*, in *Proceedings of the Forty-Third International Colloquium on Automata, Languages and Programming* (ICALP), 2016, pp. 17:1–17:15, doi:10.4230/LIPIcs.ICALP.2016.17.

[60] R. Špalek, *A dual polynomial for OR*. Available at `http://arxiv.org/abs/0803.4516`, 2008.

[61] A. C.-C. Yao, *Some complexity questions related to distributive computing*, in *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing* (STOC), 1979, pp. 209–213, doi:10.1145/800135.804414.

Appendix A. A dual object for OR

The purpose of this appendix is to prove Theorem 3.3, which gives a dual polynomial for the OR function with a number of additional properties. The treatment here closely follows earlier work by Špalek [60], Bun and Thaler [14, 18, 13], and Sherstov [53, 55]. We start with a well-known binomial identity [25].

FACT A.1. *For every univariate polynomial p of degree less than n,*

$$\sum_{t=0}^{n}(-1)^t\binom{n}{t}p(t)=0.$$

The next lemma constructs a dual polynomial for OR that has the sign behavior claimed in Theorem 3.3 but may lack some of the metric properties. The lemma is an adaptation of [53, Lemma A.2].

LEMMA A.2. *Let ϵ be given, $0<\epsilon<1$. Then for some constant $c=c(\epsilon)\in(0,1)$ and every integer $n\geqslant 1$, there is an (explicitly given) function $\omega\colon\{0,1,2,\ldots,n\}\to\mathbb{R}$ such that*

$$\omega(0)>\frac{1-\epsilon}{2}\cdot\|\omega\|_1, \tag{A.1}$$

$$|\omega(t)|\leqslant\frac{1}{ct^2\,2^{ct/\sqrt{n}}}\cdot\|\omega\|_1 \qquad (t=1,2,\ldots,n), \tag{A.2}$$

$$(-1)^t\omega(t)\geqslant 0 \qquad (t=0,1,2,\ldots,n), \tag{A.3}$$

$$\operatorname{orth}\omega\geqslant c\sqrt{n}. \tag{A.4}$$

REMARK A.3. It is helpful to keep in mind that properties (A.1)–(A.4) are logically monotonic in c. In other words, establishing these properties for a given constant $c>0$ also establishes them for all smaller positive constants.

Proof of Lemma A.2. Let $\Delta=8\lceil 1/\epsilon\rceil+3$. If $n\leqslant\Delta$, the requirements of the lemma hold for the function $\omega:(0,1,2,3\ldots,n)\mapsto(1,-1,0,0,\ldots,0)$ and all $c\in(0,1/\Delta]$. In what follows, we treat the complementary case $n>\Delta$.

Define $d=\lfloor\sqrt{n/\Delta}\rfloor$ and let $S=\{1,\frac{\Delta+1}{2}\}\cup\{i^2\Delta:i=0,1,2,\ldots,d\}$, so that $S\subseteq\{0,1,2,\ldots,n\}$. Consider the function $\omega\colon\{0,1,2,\ldots,n\}\to\mathbb{R}$ given by

$$\omega(t)=\frac{(-1)^{n+t+|S|+1}}{n!}\binom{n}{t}\prod_{\substack{i=0,1,2,\ldots,n:\\ i\notin S}}(t-i).$$

Fact A.1 implies that

$$\begin{aligned}\operatorname{orth}\omega&>d+1\\&\geqslant\sqrt{\frac{n}{\Delta}}.\end{aligned} \tag{A.5}$$

A routine calculation reveals that

$$\omega(t)=\begin{cases}(-1)^{|\{i\in S:i<t\}|}\prod_{i\in S\setminus\{t\}}\frac{1}{|t-i|} & \text{if } t\in S,\\ 0 & \text{otherwise.}\end{cases} \tag{A.6}$$

It follows that

$$\begin{aligned}\frac{\omega(0)}{|\omega(1)|} &= \frac{\Delta-1}{\Delta+1}\prod_{i=1}^{d}\frac{i^2\Delta-1}{i^2\Delta} \\ &\geqslant 1-\frac{2}{\Delta+1}-\sum_{i=1}^{d}\frac{1}{i^2\Delta} \\ &> 1-\frac{2}{\Delta+1}-\frac{1}{\Delta}\sum_{i=1}^{\infty}\frac{1}{i^2} \\ &> 1-\frac{4}{\Delta}. \end{aligned} \tag{A.7}$$

An analogous application of (A.6) shows that

$$\begin{aligned}\frac{|\omega(\frac{\Delta+1}{2})|}{|\omega(0)|} &= \frac{\frac{\Delta+1}{2}}{\frac{\Delta+1}{2}\cdot(\frac{\Delta+1}{2}-1)}\frac{\Delta^d d!\,d!}{(\Delta-\frac{\Delta+1}{2})\cdot\frac{1}{2}\Delta^{d-1}(d-1)!\,(d+1)!} \\ &= \frac{8\Delta d}{(\Delta-1)^2(d+1)} \\ &\leqslant \frac{8\Delta}{(\Delta-1)^2}. \end{aligned} \tag{A.8}$$

Finally, for $i = 1, 2, \ldots, d$,

$$\begin{aligned}\frac{|\omega(i^2\Delta)|}{|\omega(0)|} &= \frac{\frac{\Delta+1}{2}}{(i^2\Delta-1)(i^2\Delta-\frac{\Delta+1}{2})}\cdot\frac{d!\,d!\,\Delta^d}{\frac{1}{2}\cdot(d-i)!\,(d+i)!\,\Delta^d} \\ &\leqslant \frac{2(\Delta+1)}{i^4(\Delta-1)^2}\cdot\frac{d!\,d!}{(d-i)!\,(d+i)!} \\ &= \frac{2(\Delta+1)}{i^4(\Delta-1)^2}\cdot\frac{d}{d+i}\cdot\frac{d-1}{d+i-1}\cdot\cdots\cdot\frac{d-i+1}{d+1} \\ &\leqslant \frac{2(\Delta+1)}{i^4(\Delta-1)^2}\cdot\left(1-\frac{i}{d+i}\right)^i \\ &\leqslant \frac{2(\Delta+1)}{i^4(\Delta-1)^2}\cdot\exp\left(-\frac{i^2}{d+i}\right) \\ &\leqslant \frac{2(\Delta+1)}{i^4(\Delta-1)^2}\cdot\exp\left(-\frac{i^2}{2d}\right) \\ &\leqslant \frac{2(\Delta+1)}{i^4(\Delta-1)^2}\cdot\exp\left(-\frac{i^2}{2\sqrt{n/\Delta}}\right). \end{aligned} \tag{A.9}$$

Now,

$$\begin{aligned}\frac{\|\omega\|_1}{\omega(0)} &= 1 + \frac{|\omega(1)|}{\omega(0)} + \frac{|\omega(\frac{\Delta+1}{2})|}{\omega(0)} + \sum_{i=1}^{d} \frac{|\omega(i^2\Delta)|}{\omega(0)} \\ &\leqslant 1 + \left(1 - \frac{4}{\Delta}\right)^{-1} + \frac{8\Delta}{(\Delta-1)^2} + \sum_{i=1}^{\infty} \frac{2(\Delta+1)}{i^4(\Delta-1)^2} \\ &= 1 + \left(1 - \frac{4}{\Delta}\right)^{-1} + \frac{8\Delta}{(\Delta-1)^2} + \frac{\pi^4(\Delta+1)}{45(\Delta-1)^2} \\ &\leqslant \frac{2}{1-\frac{8}{\Delta}} \\ &< \frac{2}{1-\epsilon}, \end{aligned} \tag{A.10}$$

where the second step uses (A.7)–(A.9), and the last step substitutes the definition of Δ. Now (A.1) follows from (A.10). Moreover, for $c = c(\Delta) > 0$ small enough, (A.4) follows from (A.5), whereas (A.2) follows from (A.9) and the fact that ω vanishes outside the union $\{1, \frac{\Delta+1}{2}\} \cup \{i^2\Delta : i = 0, 1, 2, \ldots, d\}$.

It remains to verify that ω has the desired sign behavior. Since ω vanishes outside S, the requirement (A.3) holds trivially at those points. For $t \in S$, it follows from (A.6) that

$$\begin{aligned} &\operatorname{sgn}\omega(1) = -1, \\ &\operatorname{sgn}\omega\big(\tfrac{\Delta+1}{2}\big) = 1, \\ &\operatorname{sgn}\omega(i^2\Delta) = (-1)^i, \qquad\qquad i = 0, 1, 2, \ldots, d. \end{aligned}$$

Since $\Delta \in 4\mathbb{Z} + 3$ by definition, we conclude that $\operatorname{sgn}\omega(t) = (-1)^t$ for all $t \in S$. This settles (A.3) and completes the proof. □

We have reached the main result of this section.

THEOREM (restatement of Theorem 3.3). *Let $0 < \epsilon < 1$ be given. Then for some constants $c', c'' \in (0,1)$ and all integers $N \geqslant n \geqslant 1$, there is an (explicitly given) function $\psi\colon \{0, 1, 2, \ldots, N\} \to \mathbb{R}$ such that*

$$\psi(0) > \frac{1-\epsilon}{2}, \tag{A.11}$$

$$\|\psi\|_1 = 1, \tag{A.12}$$

$$\operatorname{orth}\psi \geqslant c'\sqrt{n}, \tag{A.13}$$

$$\operatorname{sgn}\psi(t) = (-1)^t, \qquad t = 0, 1, 2, \ldots, N, \tag{A.14}$$

$$|\psi(t)| \in \left[\frac{c'}{(t+1)^2\, 2^{c''t/\sqrt{n}}},\ \frac{1}{c'(t+1)^2\, 2^{c''t/\sqrt{n}}}\right], \quad t = 0, 1, 2, \ldots, N. \tag{A.15}$$

Proof. The degenerate case $N = 1$ holds for the function $\omega : (0,1) \mapsto (1/2, -1/2)$ and all $c', c'' \in (0, 1/4)$. In the rest of the proof, we treat the complementary case $N \geqslant 2$.

For some sufficiently small constant $c \in (0, 1/4)$ and all $n \geqslant 1$, Lemma A.2 and Remark A.3 ensure the existence of a function $\omega\colon \{0, 1, 2, \ldots, \lceil n/2 \rceil\} \to \mathbb{R}$ such that

$$\|\omega\|_1 = 1, \tag{A.16}$$

$$\omega(0) > \frac{1}{2}\left(1 - \frac{\epsilon}{6}\right), \tag{A.17}$$

$$|\omega(t)| \leqslant \frac{1}{ct^2\, 2^{ct/\sqrt{n}}} \qquad (t = 1, 2, \ldots, \lceil n/2 \rceil), \tag{A.18}$$

$$(-1)^t \omega(t) \geqslant 0 \qquad (t = 0, 1, 2, \ldots, \lceil n/2 \rceil), \tag{A.19}$$

$$\operatorname{orth} \omega \geqslant c\sqrt{n}. \tag{A.20}$$

For convenience, extend ω to all of $\mathbb{Z}$ by defining it to be zero outside its original domain. Define $\Psi\colon \{0, 1, 2, \ldots, N\} \to \mathbb{R}$ by

$$\Psi(t) = \omega(t) + \delta \left(\sum_{i=1}^{N - \lceil n/2 \rceil} \frac{(-1)^i}{i^2\, 2^{ci/\sqrt{n}}} \omega(t - i) + \sum_{i = N - \lceil n/2 \rceil + 1}^{N} \frac{(-1)^i}{i^2\, 2^{ci/\sqrt{n}}} \omega(-t + i) \right),$$

where

$$\delta = \frac{5\epsilon}{\pi^2(1 - \epsilon)}.$$

By (A.20) and Proposition 2.1(i),

$$\operatorname{orth} \Psi \geqslant c\sqrt{n}. \tag{A.21}$$

We now move on to metric properties of Ψ. Multiplying the defining equation for Ψ on both sides by $(-1)^t$ and applying (A.19), we arrive at

$$(-1)^t \Psi(t) = |\omega(t)| + \delta \left(\sum_{i=1}^{N - \lceil n/2 \rceil} \frac{|\omega(t - i)|}{i^2\, 2^{ci/\sqrt{n}}} + \sum_{i = N - \lceil n/2 \rceil + 1}^{N} \frac{|\omega(-t + i)|}{i^2\, 2^{ci/\sqrt{n}}} \right), \qquad t = 0, 1, 2, \ldots, N. \tag{A.22}$$

Summing over t gives

$$\begin{aligned}\|\Psi\|_1 &= \|\omega\|_1 + \delta \sum_{i=1}^{N} \frac{1}{i^2\, 2^{ci/\sqrt{n}}} \|\omega\|_1 \\ &= 1 + \delta \sum_{i=1}^{N} \frac{1}{i^2\, 2^{ci/\sqrt{n}}} \\ &\in \left[1, 1 + \delta \sum_{i=1}^{\infty} \frac{1}{i^2}\right] \\ &= \left[1, \frac{6-\epsilon}{6(1-\epsilon)}\right], \end{aligned} \tag{A.23}$$

where the second step uses (A.16). We also have

$$\begin{aligned}\Psi(0) &\geqslant \omega(0) \\ &> \frac{6-\epsilon}{12}, \end{aligned} \tag{A.24}$$

where the first and second steps use (A.22) and (A.17), respectively.

We now estimate $|\Psi(t)|$ for each $t = 1, 2, \ldots, N$. For a lower bound, we have

$$\begin{aligned}|\Psi(t)| &= |\omega(t)| + \delta \left(\sum_{i=1}^{N-\lceil n/2 \rceil} \frac{|\omega(t-i)|}{i^2\, 2^{ci/\sqrt{n}}} + \sum_{i=N-\lceil n/2 \rceil + 1}^{N} \frac{|\omega(-t+i)|}{i^2\, 2^{ci/\sqrt{n}}} \right) \\ &\geqslant \delta \cdot \frac{|\omega(0)|}{t^2\, 2^{ct/\sqrt{n}}} \\ &\geqslant \frac{5\epsilon}{\pi^2(1-\epsilon)} \cdot \frac{6-\epsilon}{12} \cdot \frac{1}{t^2\, 2^{ct/\sqrt{n}}}, \end{aligned} \tag{A.25}$$

where the first and last steps use (A.22) and (A.17), respectively. The upper bound on $|\Psi(t)|$ is somewhat more technical. To begin with, we have the following bound for every positive integer t:

$$\begin{aligned}\sum_{i=1}^{t-1} \frac{1}{(t-i)^2\, i^2} &= \sum_{i=1}^{t-1} \frac{1}{\max\{(t-i)^2, i^2\}\, \min\{(t-i)^2, i^2\}} \\ &\leqslant \frac{1}{(t/2)^2} \sum_{i=1}^{t-1} \frac{1}{\min\{(t-i)^2, i^2\}} \\ &\leqslant \frac{1}{(t/2)^2} \cdot 2 \sum_{i=1}^{\infty} \frac{1}{i^2} \\ &= \frac{4\pi^2}{3t^2}. \end{aligned} \tag{A.26}$$

Continuing,

$$\begin{aligned}\sum_{i=1}^{\infty} \frac{|\omega(t-i)|}{i^2\, 2^{ci/\sqrt{n}}} &= \frac{|\omega(0)|}{t^2\, 2^{ct/\sqrt{n}}} + \sum_{i=1}^{t-1} \frac{|\omega(t-i)|}{i^2\, 2^{ci/\sqrt{n}}} \\ &\leqslant \frac{1}{t^2\, 2^{ct/\sqrt{n}}} + \sum_{i=1}^{t-1} \frac{1}{c(t-i)^2\, i^2\, 2^{ct/\sqrt{n}}} \\ &\leqslant \frac{1}{t^2\, 2^{ct/\sqrt{n}}} \left(1 + \frac{4\pi^2}{3c}\right), \end{aligned} \tag{A.27}$$

where the second step uses (A.16) and (A.18), and the third step substitutes the bound from (A.26). Analogously,

$$\begin{aligned}\sum_{i=1}^{\infty} \frac{|\omega(-t+i)|}{i^2\, 2^{ci/\sqrt{n}}} &= \frac{|\omega(0)|}{t^2\, 2^{ct/\sqrt{n}}} + \sum_{i=t+1}^{\infty} \frac{|\omega(-t+i)|}{i^2\, 2^{ci/\sqrt{n}}} \\ &\leqslant \frac{1}{t^2\, 2^{ct/\sqrt{n}}} + \sum_{i=t+1}^{\infty} \frac{1}{c(t-i)^2\, i^2\, 2^{ci/\sqrt{n}}} \\ &\leqslant \frac{1}{t^2\, 2^{ct/\sqrt{n}}} \left(1 + \sum_{i=t+1}^{\infty} \frac{1}{c(t-i)^2}\right) \\ &\leqslant \frac{1}{t^2\, 2^{ct/\sqrt{n}}} \left(1 + \frac{\pi^2}{6c}\right), \end{aligned} \tag{A.28}$$

where the second step uses (A.16) and (A.18). Now for every integer $t \geqslant 1$,

$$\begin{aligned}|\Psi(t)| &\leqslant |\omega(t)| + \delta \left(\sum_{i=1}^{\infty} \frac{|\omega(t-i)|}{i^2\, 2^{ci/\sqrt{n}}} + \sum_{i=1}^{\infty} \frac{|\omega(-t+i)|}{i^2\, 2^{ci/\sqrt{n}}}\right) \\ &\leqslant \frac{1}{ct^2\, 2^{ct/\sqrt{n}}} \left(1 + 2c\delta + \frac{4\pi^2\delta}{3} + \frac{\pi^2\delta}{6}\right), \end{aligned} \tag{A.29}$$

where the first step is immediate from the defining equation for Ψ, and the second step uses (A.18), (A.27), and (A.28). To complete the proof, let $\psi\colon \{0, 1, 2, \ldots, N\} \to \mathbb{R}$ be given by $\psi = \Psi/\|\Psi\|_1$. Then for a small enough constant $c' = c'(c, \epsilon, \delta) > 0$ and $c'' = c$, properties (A.11)–(A.15) follow directly from (A.21)–(A.25) and (A.29). □

Appendix B. Sign-rank and smooth threshold degree

The purpose of this appendix is to prove Theorem 2.17, implicit in [48, 42]. We closely follow the treatment in those earlier papers. Sections B.1–B.3 cover necessary technical background, followed by the proof proper in Section B.4.

B.1. Fourier transform. Consider the real vector space of functions $\{0,1\}^n \to \mathbb{R}$. For $S \subseteq \{1, 2, \ldots, n\}$, define $\chi_S\colon \{0,1\}^n \to \{-1, +1\}$ by $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. Then

$$\langle \chi_S, \chi_T \rangle = \begin{cases} 2^n & \text{if } S = T, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\{\chi_S\}_{S\subseteq\{1,2,\ldots,n\}}$ is an orthogonal basis for the vector space in question. In particular, every function $\phi\colon \{0,1\}^n \to \mathbb{R}$ has a unique representation of the form

$$\phi = \sum_{S\subseteq\{1,2,\ldots,n\}} \hat{\phi}(S)\chi_S$$

for some reals $\hat{\phi}(S)$, where by orthogonality $\hat{\phi}(S) = 2^{-n}\langle \phi, \chi_S\rangle$. The reals $\hat{\phi}(S)$ are called the *Fourier coefficients of* ϕ, and the mapping $\phi \mapsto \hat{\phi}$ is the *Fourier transform of* f. The following fact is immediate from the definition of $\hat{\phi}(S)$.

PROPOSITION B.1. *Let $\phi\colon \{0,1\}^n \to \mathbb{R}$ be given. Then*

$$\max_{S\subseteq\{1,2,\ldots,n\}} |\hat{\phi}(S)| \leqslant 2^{-n}\|\phi\|_1.$$

The linear subspace of real polynomials on $\{0,1\}^n$ of degree at most d is easily seen to be $\operatorname{span}\{\chi_S : |S| \leqslant d\}$. Its orthogonal complement, $\operatorname{span}\{\chi_S\colon |S| > d\}$, is then the linear subspace of functions that have zero inner product with every polynomial of degree at most d. As a result, the orthogonal content of a nonzero function $\phi\colon \{0,1\}^n \to \mathbb{R}$ is given by

$$\operatorname{orth}\phi = \min\{|S| : \hat{\phi}(S) \neq 0\}, \qquad \phi \not\equiv 0. \tag{B.1}$$

B.2. Forster's bound. The *spectral norm* of a real matrix $A = [A_{xy}]_{x\in X, y\in Y}$ is given by

$$\|A\| = \max_{v\in\mathbb{R}^{|Y|},\ \|v\|_2=1} \|Av\|_2,$$

where $\|\cdot\|_2$ is the Euclidean norm on vectors. The first strong lower bound on the sign-rank of an explicit matrix was obtained by Forster [23], who proved that

$$\operatorname{rk}_{\pm}(A) \geqslant \frac{\sqrt{|X|\,|Y|}}{\|A\|}$$

for any matrix $A = [A_{xy}]_{x\in X, y\in Y}$ with ± 1 entries. Forster's result has seen a number of generalizations, including the following theorem due to Forster et al. [24, Theorem 3].

THEOREM B.2 (Forster et al.). *Let $A = [A_{xy}]_{x\in X, y\in Y}$ be a real matrix without zero entries. Then*

$$\operatorname{rk}_{\pm}(A) \geqslant \frac{\sqrt{|X|\,|Y|}}{\|A\|} \min_{x,y} |A_{xy}|.$$

B.3. Spectral norm of pattern matrices. *Pattern matrices* were introduced in [45, 47] and proved useful in obtaining strong lower bounds on communication complexity. Relevant definitions and results from [47] follow. Let n and N be positive integers with $n \mid N$. Partition $\{1, 2, \ldots, N\}$ into n contiguous blocks, each

with N/n elements:

$$\{1,2,\ldots,N\} = \left\{1,2,\ldots,\frac{N}{n}\right\} \cup \left\{\frac{N}{n}+1,\ldots,\frac{2N}{n}\right\} \\ \cup\cdots\cup \left\{\frac{(n-1)N}{n}+1,\ldots,N\right\}.$$

Now, let $\mathscr{V}(N,n)$ denote the family of subsets $V \subseteq \{1,2,\ldots,N\}$ that have exactly one element in each of these blocks (in particular, $|V| = n$). Clearly, $|\mathscr{V}(N,n)| = (N/n)^n$. For a function $\phi\colon \{0,1\}^n \to \mathbb{R}$, the (N,n,ϕ)-*pattern matrix* is the real matrix A given by

$$A = \Big[\phi(x|_V \oplus w)\Big]_{x\in\{0,1\}^N,\,(V,w)\in\mathscr{V}(N,n)\times\{0,1\}^n}.$$

In words, A is the matrix of size 2^N by $(N/n)^n 2^n$ whose rows are indexed by strings $x \in \{0,1\}^N$, whose columns are indexed by pairs $(V,w) \in \mathscr{V}(N,n)\times\{0,1\}^n$, and whose entries are given by $A_{x,(V,w)} = \phi(x|_V \oplus w)$. We will need the following expression for the spectral norm of a pattern matrix [47, Theorem 4.3].

THEOREM B.3 (Sherstov). *Let $\phi\colon \{0,1\}^n \to \mathbb{R}$ be given. Let A be the (N,n,ϕ)-pattern matrix. Then*

$$\|A\| \;=\; \sqrt{2^{N+n}\left(\frac{N}{n}\right)^n}\;\max_{S\subseteq\{1,2,\ldots,n\}}\left\{|\hat{\phi}(S)|\left(\frac{n}{N}\right)^{|S|/2}\right\}.$$

B.4. Proof of Theorem 2.17. We are now in a position to prove Theorem 2.17. We will derive it from the following more general result, stated in terms of pattern matrices.

THEOREM B.4. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be given. Suppose that $\deg_\pm(f,\gamma) \geqslant d$, where γ and d are positive reals. Then for any integer $T \geqslant 1$, the $(Tn,n,(-1)^f)$-pattern matrix has sign-rank at least $\gamma T^{d/2}$.*

Proof. By the definition of smooth threshold degree, there is a probability distribution μ on $\{0,1\}^n$ such that

$$\mu(x) \geqslant \gamma\, 2^{-n}, \qquad x \in \{0,1\}^n, \tag{B.2}$$

$$\operatorname{orth}((-1)^f \cdot \mu) \geqslant d. \tag{B.3}$$

Abbreviate $\phi = (-1)^f \cdot \mu$. Let F and Φ denote the $(Tn,n,(-1)^f)$- and (Tn,n,ϕ)-pattern matrices, respectively. By (B.1) and (B.3),

$$\hat{\phi}(S) = 0, \qquad |S| < d. \tag{B.4}$$

The remaining Fourier coefficients of ϕ can be bounded using Proposition B.1:

$$|\hat{\phi}(S)| \leqslant 2^{-n}, \qquad S \subseteq \{1,2,\ldots,n\}. \tag{B.5}$$

Now

$$\begin{aligned}\mathrm{rk}_{\pm}(F) &= \mathrm{rk}_{\pm}(\Phi) \\ &\geqslant \frac{\sqrt{2^{Tn+n}\,T^n}}{\|\Phi\|} \cdot \gamma\, 2^{-n} \\ &= \frac{\gamma\, 2^{-n}}{\max_S\{|\hat{\phi}(S)|\, T^{-|S|/2}\}} \\ &\geqslant \gamma T^{d/2},\end{aligned}$$

where the first step is valid because F and Φ have the same sign pattern; the second step uses (B.2) and Theorem B.2; the third step applies Theorem B.3; and the final step substitutes the upper bounds from (B.4) and (B.5). □

We have reached the main result of this appendix.

THEOREM (restatement of Theorem 2.17). *Let $f\colon \{0,1\}^n \to \{0,1\}$ be given. Suppose that $\deg_{\pm}(f,\gamma) \geqslant d$, where γ and d are positive reals. Fix an integer $m \geqslant 2$ and define $F\colon \{0,1\}^{mn} \times \{0,1\}^{mn} \to \{0,1\}$ by $F(x,y) = f \circ \mathrm{OR}_m \circ \mathrm{AND}_2$. Then*

$$\mathrm{rk}_{\pm}(F) \geqslant \gamma \left\lfloor \frac{m}{2} \right\rfloor^{d/2}.$$

Proof. The result is immediate from Theorem B.4 since the $(\lfloor m/2 \rfloor n, n, (-1)^f)$-pattern matrix is a submatrix of $[(-1)^{F(x,y)}]_{x,y}$. □