The Large-Error Approximate Degree of AC^0

Mark Bun
Boston University, Boston, MA, USA
http://cs-people.bu.edu/mbun/
mbun@bu.edu

Justin Thaler
Georgetown University, Washington, DC, USA
http://people.cs.georgetown.edu/jthaler/
justin.thaler@georgetown.edu

Abstract

We prove two new results about the inability of low-degree polynomials to uniformly approximate constant-depth circuits, even to slightly-better-than-trivial error. First, we prove a tight $\tilde{\Omega}(n^{1/2})$ lower bound on the threshold degree of the SURJECTIVITY function on n variables. This matches the best known threshold degree bound for any AC^0 function, previously exhibited by a much more complicated circuit of larger depth (Sherstov, FOCS 2015). Our result also extends to a $2^{\tilde{\Omega}(n^{1/2})}$ lower bound on the sign-rank of an AC^0 function, improving on the previous best bound of $2^{\Omega(n^{2/5})}$ (Bun and Thaler, ICALP 2016).

Second, for any $\delta > 0$, we exhibit a function $f\colon \{-1,1\}^n \to \{-1,1\}$ that is computed by a circuit of depth $O(1/\delta)$ and is hard to approximate by polynomials in the following sense: f cannot be uniformly approximated to error $\varepsilon = 1 - 2^{-\Omega(n^{1-\delta})}$, even by polynomials of degree $n^{1-\delta}$. Our recent prior work (Bun and Thaler, FOCS 2017) proved a similar lower bound, but which held only for error $\varepsilon = 1/3$.

Our result implies $2^{\Omega(n^{1-\delta})}$ lower bounds on the complexity of AC^0 under a variety of basic measures such as discrepancy, margin complexity, and threshold weight. This nearly matches the trivial upper bound of $2^{O(n)}$ that holds for every function. The previous best lower bound on AC^0 for these measures was $2^{\Omega(n^{1/2})}$ (Sherstov, FOCS 2015). Additional applications in learning theory, communication complexity, and cryptography are described.

2012 ACM Subject Classification Mathematics of computing → Approximation; Theory of computation → Communication complexity; Theory of computation → Circuit complexity

Keywords and phrases approximate degree, discrepancy, margin complexity, polynomial approximations, secret sharing, threshold circuits

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2019.55

Category RANDOM

Related Version The full version of this work appears at http://eccc.weizmann.ac.il/report/2018/143/.

Funding *Mark Bun*: This work was done while author was at Princeton University and the Simons Institute for the Theory of Computing, supported by a Google Research Fellowship.
Justin Thaler: Supported by NSF Grant CCF-1845125.

Acknowledgements The authors are grateful to Robin Kothari, Nikhil Mande, Jonathan Ullman, and the anonymous reviewers for valuable comments on earlier versions of this manuscript.

1 Introduction

The *threshold degree* of a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$, denoted $\deg_{\pm}(f)$, is the least degree of a real polynomial p that sign-represents f, i.e., $p(x) \cdot f(x) > 0$ for all $x \in \{-1,1\}^n$. A closely related notion is the ε-approximate degree of f, denoted $\widetilde{\deg}_{\varepsilon}(f)$, which is the least degree of a real polynomial p such that $|p(x) - f(x)| \leq \varepsilon$ for all $x \in \{-1,1\}^n$.

The parameter setting $\varepsilon = 1$ is a degenerate case: $\widetilde{\deg}_1(f) = 0$ because the constant 0 function approximates any Boolean f to error $\varepsilon = 1$. However, as soon as ε is strictly less than 1, ε-approximate degree is a highly non-trivial notion with a rich mathematical theory. In particular, it is easily seen that

$$\deg_{\pm}(f) = \lim_{\varepsilon \nearrow 1} \widetilde{\deg}_{\varepsilon}(f).$$

In other words, threshold degree is equivalent to the notion of ε-approximate degree when ε is permitted to be *arbitrarily* close to (but strictly less than) 1.[1]

In this paper, we are concerned with proving ε-approximate degree lower bounds when either:

- ε is *arbitrarily* close to 1, or
- ε is *exponentially* close to 1 (i.e., $\varepsilon = 1 - 2^{-n^{1-\delta}}$ for some constant $\delta > 0$).

The former parameter regime captures threshold degree, while we refer to the latter as *large-error* approximate degree. While the approximate and threshold degree of a function f capture simple statements about its approximability by polynomials, these quantities relate intimately to the complexity of computing f in concrete computational models. Specifically, the query complexity models $\mathbf{UPP}^{\mathrm{dt}}$ and $\mathbf{PP}^{\mathrm{dt}}$, and the communication models $\mathbf{UPP}^{\mathrm{cc}}, \mathbf{PP}^{\mathrm{cc}}$, are all defined (cf. Section 2) as natural analogs of the Turing machine class $\mathbf{PP}$, which in turn captures probabilistic computation with arbitrarily small advantage over random guessing. It is known that the threshold degree of f is equivalent to its complexity $\mathbf{UPP}^{\mathrm{dt}}(f)$, while a fundamental matrix-analytic analog of threshold degree known as *sign-rank* characterizes $\mathbf{UPP}^{\mathrm{cc}}$. Similarly, large-error approximate degree characterizes the query complexity measure $\mathbf{PP}^{\mathrm{dt}}$, in the following sense: for any $d > 0$, $\widetilde{\deg}_{1-2^{-d}}(f) \geq \Omega(d) \iff \mathbf{PP}^{\mathrm{dt}}(f) \geq \Omega(d)$. Section 2 elaborates on these models and their many applications in learning theory, circuit complexity, and cryptography.

Our Results in a Nutshell. We prove two results about the threshold degree and large-error approximate degree of functions in AC^0.[2] First, we prove a tight $\tilde{\Omega}(n^{1/2})$ lower bound on the threshold degree (i.e., $\mathbf{UPP}^{\mathrm{dt}}$ complexity) of a natural function called SURJECTIVITY, which is computed by a depth three circuit with logarithmic bottom fan-in. This matches the previous best threshold degree lower bound for any AC^0 function, due to Sherstov [34]. Our analysis is much simpler than Sherstov's, which takes up the bulk of a (70+)-page manuscript [34]. An additional advantage of our analysis is that our lower bound on the threshold degree of SURJECTIVITY "lifts" to give a lower bound for the communication analog $\mathbf{UPP}^{\mathrm{cc}}$ as well. In particular, we obtain an $\Omega(n^{1/2})$ $\mathbf{UPP}^{\mathrm{cc}}$ lower bound for a related AC^0 function; this improves over the previous best $\mathbf{UPP}^{\mathrm{cc}}$ lower bound for AC^0, of $\Omega(n^{2/5})$ [12].

Second, we give nearly optimal bounds on the large-error approximate degree (and hence, $\mathbf{PP}^{\mathrm{dt}}$ complexity) of AC^0. For any constant $\delta > 0$, we show that there is an AC^0 function with ε-approximate degree $\Omega(n^{1-\delta})$, where $\varepsilon = 1 - 2^{-\Omega(n^{1-\delta})}$. This result lifts to an analogous $\mathbf{PP}^{\mathrm{cc}}$ lower bound.

---

[1] It is known that for any $d > 0$, there are functions of threshold degree d that cannot be approximated by degree d polynomials to error better than $1 - 2^{-\tilde{\Omega}(n^d)}$ [27], and this bound is tight [7]. Hence, threshold degree is also equivalent to the notion of ε-approximate degree for some value of ε that is *doubly*-exponentially close to 1.

[2] AC^0 is the non-uniform class of sequences of functions computed by polynomial size Boolean circuits of constant depth.

Table 1 Comparison of our new bounds for AC^0 to prior work in roughly chronological order. The circuit depth column lists the depth of the Boolean circuit used to exhibit the bound, δ denotes an arbitrarily small positive constant, and k an arbitrary positive integer. All Boolean circuits are polynomial size.

Reference	$\mathbf{PP}^{\mathrm{dt}}$ log(threshold weight)	$\mathbf{PP}^{\mathrm{cc}}$ log(1/discrepancy)	$\mathbf{UPP}^{\mathrm{dt}}$ threshold degree	$\mathbf{UPP}^{\mathrm{cc}}$ log(sign-rank)	Circuit Depth
[23]	—	—	$\Omega(n^{1/3})$	—	2
[20]	$\Omega(n^{1/3})$	—	—	—	3
[16]	—	$\Omega(\log^k(n))$	—	$\Omega(\log^k(n))$	$O(k)$
[25]	$\Omega(n^{1/3}\log^k n)$	—	$\Omega(n^{1/3}\log^k(n))$	—	$O(k)$
[29]	—	$\Omega(n^{1/5})$	—	—	3
[7,31]	—	$\Omega(n^{1/3})$	—	—	3
[28]	—	—	—	$\Omega(n^{1/3})$	3
[10]	$\Omega(n^{2/5})$	$\Omega(n^{2/5})$	—	—	3
[33]	$\Omega(n^{1/2-\delta})$	$\Omega(n^{1/2-\delta})$	$\Omega(n^{1/2-\delta})$	—	$O(1/\delta)$
[34]	$\Omega(n^{3/7})$	—	$\Omega(n^{3/7})$	—	3
[34]	$\Omega(n^{1/2})$	$\Omega(n^{1/2})$	$\Omega(n^{1/2})$	—	4
[12]	—	—	—	$\Omega(n^{2/5})$	3
[11]	$\Omega(n^{1/2-\delta})$	$\Omega(n^{1/2-\delta})$	—	—	3
This work	$\tilde{\Omega}(n^{1/2})$	$\tilde{\Omega}(n^{1/2})$	$\tilde{\Omega}(n^{1/2})$	—	3
This work	—	—	—	$\tilde{\Omega}(n^{1/2})$	7
This work	$\Omega(n^{1-\delta})$	$\Omega(n^{1-\delta})$	—	—	$O(1/\delta)$

To summarize our results succinctly:

- We prove a $\tilde{\Omega}(n^{1/2})$ lower bound on the **UPP** complexity of SURJECTIVITY in the query setting, and of a related AC^0 function in the communication setting.
- We prove a $\Omega(n^{1-\delta})$ lower bound on the **PP** complexity of some AC^0 circuit of depth $O(1/\delta)$, in both the query and communication settings.

Table 1 compares our new lower bounds for AC^0 to the long line of prior works with similar goals.

Context and Prior Work. The study of both large-error approximate degree and threshold degree has led to many breakthrough results in theoretical computer science, especially in the algorithmic and complexity-theoretic study of constant depth circuits. For example, threshold degree upper bounds are at the core of many of the fastest known PAC learning algorithms. This includes the notorious case of polynomial size CNF formulas on n variables, for which the fastest known algorithm [19] runs in time $\exp(\tilde{O}(n^{1/3}))$ owing to a $\tilde{O}(n^{1/3})$ upper bound on the threshold degree of any such formula. This upper bound is tight, matching a classic $\Omega(n^{1/3})$ lower bound of Minsky and Papert [23] for the following read-once CNF: $\mathsf{AND}_{n^{1/3}} \circ \mathsf{OR}_{n^{2/3}}$ (here, we use subscripts to clarify the number of inputs on which a function is defined).

In complexity theory, breakthrough results of Sherstov [29,31] and Buhrman et al. [7] used lower bounds on large-error approximate degree to show that there are AC^0 functions with polynomial $\mathbf{PP}^{\mathrm{cc}}$ complexity. One notable implication of these results is that Allender's [1] classic simulation of AC^0 functions by depth-three majority circuits is optimal. (This resolved an open problem of Krause and Pudlák [20].) A subsequent, related breakthrough of Razborov and Sherstov [28] used Minsky and Papert's lower bound on the threshold degree of $\mathsf{AND}_{n^{1/3}} \circ \mathsf{OR}_{n^{2/3}}$ to prove the first polynomial $\mathbf{UPP}^{\mathrm{cc}}$ lower bound for a function in AC^0, answering an old open question of Babai et al. [2].

These breakthrough lower bounds raised the intriguing possibility that AC^0 functions could be *maximally* hard for the $\mathbf{UPP}^{cc}$ and $\mathbf{PP}^{cc}$ communication models, as well as for related complexity measures. Nevertheless, the quantitative parameters achieved in these works are far from actually showing that this is the case. Indeed, the following basic questions about the complexity of AC^0 remain open.

▶ **Problem 1.** *Is there an AC^0 function $F\colon \{-1,1\}^{n\times n} \to \{-1,1\}$ with $\mathbf{UPP}^{cc}$ complexity $\Omega(n)$?*

▶ **Problem 2.** *Is there an AC^0 function $F\colon \{-1,1\}^{n\times n} \to \{-1,1\}$ with $\mathbf{PP}^{cc}$ complexity $\Omega(n)$?*

An affirmative answer to either question would be tight: *Every* function $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ has $\mathbf{UPP}^{cc}$ and $\mathbf{PP}^{cc}$ complexity at most n. Obtaining an affirmative answer to Open Problem 1 is harder than for Open Problem 2, since $\mathbf{UPP}^{cc}(f) \leq \mathbf{PP}^{cc}(f)$ for all f.

Guided by these open problems, a sequence of works has established quantitatively stronger and more general lower bounds for AC^0 functions [9–13, 33, 34]. In addition to making partial progress toward resolving these questions, the techniques developed in these works have found fruitful applications in new domains. For example, Bouland et al. [6] built on techniques from a number of aforementioned works [9,10,12,33] to resolve several old open questions about the relativized power of statistical zero knowledge proofs and their variants. As another example, our recent prior works [8,13] built on the same line of work to resolve or nearly resolve a number of longstanding open questions in quantum query complexity. Finally, large-error and threshold degree lower bounds on AC^0 functions have recently proved instrumental in the development of cryptographic secret-sharing schemes with reconstruction procedures in AC^0 [4,5,14]. We thus believe that the new techniques developed in this work will find further applications, perhaps in unexpected areas.

Prior to our work, the best known result toward a resolution of Open Problem 1 was a $\Omega(n^{2/5})$ lower bound on $\mathbf{UPP}^{cc}$ complexity of an AC^0 function [12], while the best known result toward Open Problem 2 was a $\Omega(n^{1/2})$ bound on the $\mathbf{PP}^{cc}$ complexity of a very complicated AC^0 circuit [34].

1.1 Our Results In Detail

1.1.1 Resolving the Threshold Degree of SURJECTIVITY

Surjectivity and its History. Let R be a power of 2 and $n = N \log R$. The function $\mathsf{SURJECTIVITY}_n$ ($\mathsf{SURJ}_{R,N}$ for short) is defined as follows. Given an input in $\{-1,1\}^n$, $\mathsf{SURJ}_{R,N}$ interprets the input as a list of N numbers $(s_1,\ldots,s_N)$ from a range $[R] := \{1,\ldots,R\}$, and evaluates to -1 if and only if every element of the range $[R]$ appears at least once in the list.[3] $\mathsf{SURJ}_{R,N}$ is computed by an AC^0 circuit of depth three and logarithmic bottom fan-in, since it is equivalent to the AND_R (over all range items $r \in [R]$) of the OR_N (over all inputs $i \in [N]$) of "Is input s_i equal to r?", where the quoted question is computed by a conjunction of width $\log R$ over the input bits.

$\mathsf{SURJ}_{R,N}$ has been studied extensively in the contexts of quantum query complexity and approximate degree. Beame and Machmouchi [3] showed that computing $\mathsf{SURJ}_{R,N}$ for $R = N/2 + 1$ requires $\tilde{\Omega}(n)$ quantum queries, making it the only known AC^0 function with linear quantum query complexity. Meanwhile, the $(1/3)$-approximate degree of $\mathsf{SURJ}_{R,N}$ was

[3] As is standard, we associate -1 with logical TRUE and $+1$ with logical FALSE throughout.

recently shown to be $\tilde{\Theta}(R^{1/4} \cdot N^{1/2})$. The lower bound is from our prior work [8], while the upper bound was shown by Sherstov [35], with a different proof given in [8]. In particular, when $R = N/2$, $\widetilde{\deg}_{1/3}(\mathsf{SURJ}_{R,N}) = \tilde{\Theta}(N^{3/4})$. Our prior works [8, 13] built directly on the approximate degree lower bound for $\mathsf{SURJ}_{R,N}$ to give near-optimal lower bounds on the (1/3)-approximate degree of AC^0 (see Section 3.3 for details).

Our Result. In spite of the progress described above, the threshold degree $\mathsf{SURJ}_{R,N}$ remained open. For $R < N/2$, an upper bound of $\tilde{O}(\min\{R, N^{1/2}\})$ follows from standard techniques. The best known lower bound was $\Omega(\min\{R, N^{1/3}\})$, obtained by a reduction to Minsky and Papert's threshold degree lower bound for $\mathsf{AND}_{n^{1/3}} \circ \mathsf{OR}_{n^{2/3}}$. In this work, we settle the threshold degree of $\mathsf{SURJ}_{R,N}$, showing that the known upper bound is tight up to logarithmic factors.

▶ **Theorem 3.** *For $R < N/2$, the threshold degree of $\mathsf{SURJ}_{R,N}$ is $\tilde{\Theta}(\min\{R, N^{1/2}\})$. In particular, if $R = N^{1/2}$, $\deg_{\pm}(\mathsf{SURJ}_{R,N}) = \tilde{\Theta}(N^{1/2})$.*

In addition to resolving a natural question in its own right, Theorem 3 matches the best prior threshold degree lower bound for AC^0, previously proved in [34] for a much more complicated function computed by a circuit of strictly greater depth. Furthermore, with some extra effort, our lower bound for $\mathsf{SURJ}_{R,N}$ extends to give a $\tilde{\Omega}(n^{1/2})$ lower bound on the $\mathbf{UPP}^{\mathrm{cc}}$ complexity of a related AC^0 function, yielding progress on Open Question 1 (cf. Section 1). In contrast, Sherstov's $\Omega(n^{1/2})$ threshold degree lower bound for AC^0 [34] is not known to extend to $\mathbf{UPP}^{\mathrm{cc}}$ complexity. As stated in Section 1, the best previous $\mathbf{UPP}^{\mathrm{cc}}$ lower bound for an AC^0 function was $\Omega(n^{2/5})$.

▶ **Corollary 4.** *There is an AC^0 function $F\colon \{-1,1\}^{n\times n} \to \{-1,1\}$ such that $\mathbf{UPP}^{cc}(F) \geq \tilde{\Omega}(n^{1/2})$.*

1.1.2 AC^0 Has Nearly Maximal $\mathrm{PP}^{\mathrm{cc}}$ Complexity

In our second result, for any constant $\delta > 0$, we exhibit an AC^0 function $f\colon \{-1,1\}^n \to \{-1,1\}$ with $\widetilde{\deg}_{\varepsilon}(f) = \Omega(n^{1-\delta})$ for some $\varepsilon = 1 - 2^{-\Omega(n^{1-\delta})}$. This is a major strengthening of our prior works [8, 13], which proved a similar result for $\varepsilon = 1/3$. By combining this large-error approximate degree lower bound with a "query-to-communication lifting theorem" for $\mathbf{PP}$ [31], we obtain a $\Omega(n^{1-\delta})$ bound on the $\mathbf{PP}^{\mathrm{cc}}$ complexity of an AC^0 function, nearly resolving Open Question 2 from the previous section.

▶ **Theorem 5.** *For any constant $\delta > 0$, there is an AC^0 function $F\colon \{-1,1\}^{n\times n} \to \{-1,1\}$ with $\mathbf{PP}^{cc}(F) = \Omega(n^{1-\delta})$.*

The best previous lower bound for the $\mathbf{PP}^{\mathrm{cc}}$ complexity of an AC^0 function was $\Omega(n^{1/2})$ [34].

2 Algorithmic and Complexity-Theoretic Applications

To introduce the applications of our results, we begin by defining the query complexity quantities $\mathbf{UPP}^{\mathrm{dt}}$ and $\mathbf{PP}^{\mathrm{dt}}$ and the communication complexity quantities $\mathbf{UPP}^{\mathrm{cc}}$ and $\mathbf{PP}^{\mathrm{cc}}$.

Query Models. In randomized query complexity, an algorithm aims to evaluate a known Boolean function f on an unknown input $x \in \{-1,1\}^n$ by reading as few bits of x as possible. We say that the *query cost* of a randomized algorithm is the maximum number of bits it queries for any input x.

- $\mathbf{UPP}^{\mathrm{dt}}$ considers "unbounded error" randomized algorithms, which means that on any input x, the algorithm outputs $f(x)$ with probability strictly greater than $1/2$. $\mathbf{UPP}^{\mathrm{dt}}(f)$ is the minimum query cost of any unbounded error algorithm for f.
- $\mathbf{PP}^{\mathrm{dt}}(f)$ captures "large" (rather than unbounded) error algorithms. If a randomized query algorithm outputs $f(x)$ with probability $1/2+\beta$ for all x, then the $\mathbf{PP}$-cost of the algorithm is the sum of the query cost and $\log(1/\beta)$. $\mathbf{PP}^{\mathrm{dt}}(x)$ is the minimum $\mathbf{PP}$-cost of any randomized query algorithm for f.

Communication Models. $\mathbf{UPP}^{\mathrm{cc}}$ and $\mathbf{PP}^{\mathrm{cc}}$ consider the standard two-party setup where Alice holds an input x and Bob holds an input y, and they run a private-coin randomized communication protocol to compute a function $f(x,y)$, while minimizing the number of bits they exchange. In direct analogy to the query complexity measures above, we say that the *communication cost* of a randomized protocol is the maximum number of bits Alice and Bob exchange on any input (x,y).

- $\mathbf{UPP}^{\mathrm{cc}}(f)$ [26] is the minimum communication cost of any randomized protocol that outputs $f(x,y)$ with probability strictly greater than $1/2$ on all inputs (x,y).
- $\mathbf{PP}^{\mathrm{cc}}(f)$ [2] is the minimum $\mathbf{PP}$-cost of a protocol for f, where the $\mathbf{PP}$-cost of a protocol that outputs $f(x,y)$ with probability $1/2+\beta$ for all (x,y) is the sum of the communication cost and $\log(1/\beta)$.

We now give an overview of the applications of Theorem 5 and Corollary 4.

2.1 Applications of Theorem 5

$\mathbf{PP}^{\mathrm{cc}}$ is known to be equivalent to two measures of central importance in learning theory and communication complexity, namely *margin complexity* [22] and *discrepancy* [18]. Hence, Theorem 5 implies that AC^0 has nearly maximal complexity under both measures. Below, we highlight four additional applications.

- **Communication Complexity**. The $\mathbf{PP}^{\mathrm{cc}}$ communication model can efficiently simulate almost every two-party communication model, including $\mathbf{P}$ (i.e., deterministic communication), $\mathbf{BPP}$ (randomized communication), $\mathbf{BQP}$ (quantum), and $\mathbf{P}^{\mathbf{NP}}$. The only well-studied exceptions are $\mathbf{UPP}^{\mathrm{cc}}$, and communication analogs of the polynomial hierarchy (the latter of which we do not know how to prove lower bounds against). Hence, in showing that AC^0 has essentially maximal $\mathbf{PP}^{\mathrm{cc}}$ complexity, we subsume or nearly subsume all previous results on the communication complexity of AC^0.
- **Cryptography**. Bogdanov et al. [4] observed that for any $f\colon \{-1,1\}^n \to \{-1,1\}$ and $d>0$, if one shows that $\widetilde{\deg}_{\varepsilon}(f) \geq d$, then one obtains a scheme for sharing a secret bit $b \in \{-1,1\}$ among n parties such that any subset of d shares provides no reconstruction advantage, yet applying f to all n shares yields b with probability at least $1/2+\varepsilon/2$. They combined this with known approximate degree lower bounds for AC^0 functions to get secret sharing schemes with reconstruction procedures in AC^0. Via this connection, an immediate corollary of Theorem 5 is a nearly optimal secret sharing scheme in AC^0: for any desired constant $\delta>0$, any subset of $n^{1-\delta}$ shares provides no reconstruction advantage, yet all n shares can be successfully reconstructed (by applying an AC^0 function) with probability $1-2^{-n^{1-\delta}}$.
- **Learning Theory**. Valiant [38] introduced the *evolvability* model in an effort to quantify how (and which) mechanisms can evolve in realistic population sizes within realistic time periods. Feldman [15] showed that the "weak evolvability" of a class of functions $\mathcal{F}=\{\phi_1,\ldots,\phi_{|\mathcal{F}|}\}$ is characterized by the $\mathbf{PP}^{\mathrm{cc}}$ complexity of the function $F(x,y)=\phi_x(y)$. Hence, a consequence of Theorem 5 is that there are AC^0 functions that are nearly

maximally hard to evolve (i.e., for any constant $\delta > 0$, there are AC^0 functions that require either $2^{n^{1-\delta}}$ generations, or populations of size $2^{n^{1-\delta}}$ to evolve, even if one only wants to evolve a mechanism that has advantage just $2^{-n^{1-\delta}}$ over random guessing). We also obtain a nearly optimal $2^{n^{1-\delta}}$ lower bound on the *threshold weight* of an AC^0 function. Threshold weight is another central quantity underlying many algorithmic results in learning theory. Our results rule out the possibility that algorithms based on threshold weight bounds can PAC learn AC^0 in time significantly faster than 2^n.

- **Circuit Complexity**. If $\mathbf{PP}^{cc}(f) \geq d$, then f is not computable by Majority-of-Threshold circuits of size $2^{\Omega(d)}$ [24]. Hence, by showing that AC^0 has nearly maximal $\mathbf{PP}^{cc}$ complexity, we show that there are AC^0 functions that are not computed by Majority-of-Threshold circuits of size $2^{n^{1-\delta}}$. That is, AC^0 has essentially no non-trivial simulation by Majority-of-Threshold circuits (in contrast, AC^0 can be efficiently simulated by depth-three Majority circuits [1]).

2.2 Applications of Corollary 4

As indicated in Section 1, $\mathbf{UPP}^{cc}(F)$ is known to be characterized by (the logarithm of) of the *sign-rank* of the matrix $[F(x,y)]_{x,y\in\{-1,1\}^{n\times n}}$ [26].[4] Hence, Corollary 4 implies an $\exp(\tilde{\Omega}(n^{1/2}))$ lower bound on the sign-rank of AC^0 function. Below, we highlight two additional applications of Corollary 4, based on the following connections between communication complexity, circuit complexity, and learning theory.

In communication complexity, $\mathbf{UPP}^{cc}$ is the most powerful two-party model against which we know how to prove lower bounds. In circuit complexity, if $\mathbf{UPP}^{cc}(f) \geq d$, then f cannot be computed by Threshold-of-Majority circuits of size $2^{\Omega(d)}$ [17]. (Threshold-of-Majority circuits represent the most powerful class of threshold circuits against which we can prove superpolynomial lower bounds.) In learning theory, it is commonly assumed that data can be classified by a halfspace in many dimensions; the $\mathbf{UPP}^{cc}$-complexity of a concept class precisely captures how many dimensions are needed. To connect this to a previously mentioned example, Klivans and Servedio [19] observed that an upper bound of d on the $\mathbf{UPP}^{cc}$ complexity of a concept class $\mathcal{C}$ yields a PAC learning for $\mathcal{C}$ running in time $2^{O(d)}$. They used this result to give a $2^{\tilde{O}(n^{1/3})}$-time algorithm for PAC-learning CNFs. This remains the state-of-the-art algorithm for this fundamental problem. Accordingly, Corollary 4 has the following implications.

- **Circuit Complexity.** There are AC^0 functions that are not computable by Threshold-of-Majority Circuits of size $2^{\tilde{\Omega}(n^{1/2})}$.
- **Learning Theory.** $\mathbf{UPP}^{cc}$-based learning algorithms cannot learn AC^0 in time better than $2^{\tilde{\Omega}(n^{1/2})}$.

3 Techniques

3.1 The SURJECTIVITY Lower Bound

For a function f_n, let $f^{\leq N}$ denote the partial function obtained by restricting f to the domain of inputs of Hamming weight at most N. The ε-approximate degree of $f^{\leq N}$, denoted $\widetilde{\deg}_{\varepsilon}(f^{\leq N})$, is the least degree of a real polynomial p such that

$$|p(x) - f(x)| \leq \varepsilon \text{ for all inputs } x \text{ of Hamming weight at most } N. \tag{1}$$

[4] The sign-rank of a matrix M with entries in $\{\pm 1\}$ is the least rank of a real matrix M' that agrees in sign with M entry-wise.

Note that Property (1) allows p to *behave arbitrarily* on inputs x of Hamming weight more than N. Similarly, the threshold degree of $f^{\leq N}$ is the least degree of a real polynomial p such that

$$p(x) \cdot f(x) > 0 \text{ for all inputs } x \text{ of Hamming weight at most } N.$$

Our prior work [13] showed the ε-approximate (respectively, threshold) degree of $\mathsf{SURJ}_{R,N}$ is *equivalent* to the ε-approximate (respectively, threshold) degree of $(\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}$. Hence, the main technical result underpinning our threshold degree lower bound for SURJ is the following theorem about the threshold degree of $(\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}$ (we have made no effort to optimize the logarithmic factors).

▶ **Theorem 6.** *Let* $R = N^{1/2}$. *Then* $\deg_{\pm}\left((\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}\right) = \Omega(N^{1/2}/\log^{3/2} N)$.

Discussion. Theorem 6 is a substantial strengthening of the classic result of Minsky and Papert [23] mentioned above, which established that the total function $\mathsf{MP}_{N^{1/2},N} := \mathsf{AND}_{N^{1/2}} \circ \mathsf{OR}_N$ on $n = N^{3/2}$ inputs has threshold degree $\Omega(N^{1/2})$. Theorem 6 establishes that Minsky and Papert's lower bound holds even under the promise that the input has Hamming weight at most $N = n^{2/3}$. That is, any polynomial that sign-represents $\mathsf{AND}_{n^{1/3}} \circ \mathsf{OR}_{n^{2/3}}$ on inputs of Hamming weight at most $n^{2/3}$ has degree $\tilde{\Omega}(n^{1/3})$, *even when* p *is allowed to behave arbitrarily on inputs of Hamming weight larger than* $n^{2/3}$.

Proof overview for Theorem 6 and comparison to prior work. Like much recent work on approximate and threshold degree lower bounds, our proof makes use of *dual polynomials*. A dual polynomial is a dual solution to a certain linear program capturing the approximate or threshold degree of any function, and acts as a certificate of the high approximate or threshold degree of the function.

A dual polynomial that witnesses the fact that $\deg_{\pm}(f_M) \geq d$ is a function $\psi\colon \{-1,1\}^M \to \{-1,1\}$ satisfying three properties:

- $\psi(x) \cdot f(x) \geq 0$ for all $x \in \{-1,1\}^M$. If ψ satisfies this condition, we say ψ agrees in sign with f.
- $\sum_{x \in \{-1,1\}^M} |\psi(x)| = 1$. If ψ satisfies this condition, it is said to have ℓ_1-norm equal to 1.
- For all polynomials $p\colon \{-1,1\}^M \to \mathbb{R}$ of degree at most d, $\sum_{x \in \{-1,1\}^M} p(x) \cdot \psi(x) = 0$. If ψ satisfies this condition, it is said to have *pure high degree* at least d.

A dual witness for the fact that $\widetilde{\deg}_{\varepsilon}(f_M) \geq d$ is similar, except that the first condition is replaced with:

- $\sum_{x \in \{-1,1\}^M} \psi(x) \cdot f(x) > \varepsilon$. If ψ satisfies this condition, it is said to be ε-*correlated* with f. If $\psi(x) \cdot f(x) < 0$, we say that ψ *makes an error* at x.

Sherstov [34] reproved Minsky and Papert's result by constructing an explicit dual witness for $\mathsf{MP}_{N^{1/2},N}$, via a two-step process. First, Sherstov started with a dual witness ψ_{base} for the fact that

$$\widetilde{\deg}_{\varepsilon}(\mathsf{MP}_{N^{1/2},N}) = \Omega(N^{1/2}), \text{ for } \varepsilon = 1 - 2^{-N^{1/2}}.$$

The function ψ_{base} was introduced in our prior work [10], where it was constructed by combining a dual witness for $\mathsf{AND}_{N^{1/2}}$ with a dual witness for OR_N via a technique called dual block composition [21, 32, 37].

Unfortunately, ψ_{base} falls short of witnessing Minsky and Papert's threshold degree lower bound because it makes errors on some inputs. In the second step of Sherstov's construction [34], he adds in a correction term that zeros out the errors of ψ_{base}, without disturbing the sign of ψ_{base} on any other inputs, and without lowering its pure high degree.

Theorem 6 asserts that $\mathsf{MP}^{\leq N}_{N^{1/2},N}$ satisfies the same threshold degree lower bound as $\mathsf{MP}_{N^{1/2},N}$ itself. To prove Theorem 6, we need to construct a dual witness ψ that not only reproves Minsky and Papert's classic lower bound for $\mathsf{MP}_{N^{1/2},N}$, but also satisfies the extra condition that:

$$\psi(x) = 0 \text{ for all inputs } x \text{ of Hamming weight more than } N. \tag{2}$$

To accomplish this, we apply a novel strategy that can be thought of as a three-step process. First, like Sherstov, we start with ψ_{base}. Second, we modify ψ_{base} to obtain a dual witness ψ'_{base} that places significant mass on all inputs of Hamming weight at most d, for some $d = \tilde{\Omega}(N^{1/2})$ (details of the construction of ψ'_{base} are described two paragraphs hence). More specifically, we ensure that ψ'_{base} satisfies:

$$|\psi'_{\mathsf{base}}(x)| \gg n^{-d} \text{ for all inputs } x \text{ of Hamming weight at most } d. \tag{3}$$

We refer to this property by saying that ψ'_{base} is "smooth" or "large" on all inputs of Hamming weight at most d. Note that, in modifying ψ_{base} to obtain ψ'_{base}, we do *not* correct the errors that ψ_{base} makes, nor do we ensure that ψ'_{base} is supported on inputs of Hamming weight at most N.

Third, we add in a correction term, very different than Sherstov's correction term, that not only zeros out the errors of ψ'_{base}, but also zeros out any mass it places on inputs of Hamming weight more than N. While the general technique we use to construct this correction term appeared in our prior works [8, 13], the novelty in our construction and analysis is two-fold. First, the technique was used in our prior work only to zero out mass placed on inputs of Hamming weight more than N (i.e., to ensure that Equation (2) is satisfied), not to correct errors. Second, and more importantly, we crucially exploit the largeness of ψ'_{base} on inputs of Hamming weight at most d to ensure that the correction term does not disturb the sign of ψ'_{base} on any inputs other than those on which it is deliberately being zeroed out. This is what enables us to obtain a threshold degree lower bound, whereas our prior works [8, 13] were only able to obtain ε-approximate degree lower bounds for ε bounded away from 1.

Our "smoothing followed by correction" approach appears to be significantly more generic than the correction technique of [34]. For example, prior work of Bouland et al. [6] proved an $\Omega(n^{1/4})$ lower bound on the threshold degree of a certain function denoted $\mathsf{GAPMAJ}_{n^{1/4}} \circ \mathsf{PTP}_{n^{3/4}}$, and used this result to give an oracle separating the oracle complexity classes SZK and UPP, thereby answering an open question of Watrous from 2002. Our techniques can be used to give a much simpler proof of this result, as well as several others appearing in the literature (for brevity, we omit the details of these simpler proofs of prior results). We are confident that our technique will find additional applications in the future.

Details of the smoothing step. As stated above, the dual witness ψ_{base} from our prior work does not satisfy the property we need (cf. Equation (3)) of being "large" on all inputs of Hamming weight at most $d = \tilde{\Omega}(N^{1/2})$.

Fortunately, we observe that although ψ_{base} is *not* large on all inputs of Hamming weight at most d, it *is* large on one very special input of low Hamming weight, namely the $\mathsf{ALL\text{-}FALSE}$ input. That is, $\psi_{\mathsf{base}}(\mathbf{1}) \geq 2^{-d}$. So we just need a way to "bootstrap" this largeness property

on $\mathbf{1}$ to a largeness property on all inputs of Hamming weight at most d. Put another way, we need to be able to treat other inputs of Hamming weight at most d as if they actually have Hamming weight 0. But $\mathsf{MP}_{N^{1/2},N} := \mathsf{AND}_{N^{1/2}} \circ \mathsf{OR}_N$ has a property that enables precisely this: we can fix the inputs to any constant fraction c of the OR gates to an arbitrary value in $\mathsf{OR}^{-1}(-1)$, and the remaining function of the unrestricted inputs is $\mathsf{AND}_{(1-c)\cdot R} \circ \mathsf{OR}_N$. This is "almost" the same function as $\mathsf{AND}_R \circ \mathsf{OR}_N$; we have merely slightly reduced the top fan-in, which does not substantially lower the threshold degree of the resulting function.

We exploit the above observation to achieve the following: for each input x of Hamming weight at most d, we build a dual witness ν_x targeted at x (i.e., that essentially treats x as if it is the $\mathsf{ALL\text{-}FALSE}$ input). We do this as follows. Let T be the set of all OR gates that are fed one or more -1s by x, and let $S \subseteq [N^{1/2} \cdot N]$ be the union of the inputs to each of the OR gates in T. Let ψ_{base} be the dual witness for $\mathsf{AND}_{N^{1/2}-|T|} \circ \mathsf{OR}_N$ given in our prior work [10]. We let

$$\nu_x(y) = \begin{cases} \psi_{\mathsf{base}}(y_{\bar{S}}) & \text{if } y_S = x_S \\ 0 & \text{otherwise,} \end{cases}$$

where $y_{\bar{S}}$ denotes the set of all the coordinates of y other than those in S.

The dual witness ψ'_{base} is then defined to be the average of the ν_x's, over all inputs x of Hamming weight at most d. This averaged dual witness ψ'_{base} has all of the same useful properties as ψ_{base}, and additionally satisfies the key requirement captured by Equation (3).

3.2 Extension to $\mathrm{UPP}^{\mathrm{cc}}$: Proof of Corollary 4

Building on the celebrated framework of Forster [16], Razborov and Sherstov [28] developed techniques to translate threshold degree lower bounds into sign-rank lower bounds. Specifically, they showed that, in order for a threshold degree lower bound of the form $\deg_{\pm}(f_n) \geq d$ to translate into a $\mathbf{UPP}^{\mathrm{cc}}$ lower bound for a related function F, it suffices for the threshold degree lower bound for f_n to be exhibited by a dual witness ϕ satisfying the following smoothness condition:

$$|\phi(x)| \geq 2^{-O(d)} \cdot 2^{-n} \text{ for all but a } 2^{-\Omega(d)} \text{ fraction of inputs } x \in \{-1,1\}^n. \tag{4}$$

Note that this is a different smoothness condition than the one satisfied by the dual witness ψ'_{base} discussed above for $\mathsf{MP}_{N^{1/2},N}$ (cf. Equation (3)): on inputs x of Hamming weight at most d, $|\psi'_{\mathsf{base}}(x)|$ is always at least $n^{-d} \gg 2^{-d} \cdot 2^{-n}$, whereas on inputs x of Hamming weight more than d, $|\psi'_{\mathsf{base}}(x)|$ may be 0. In words, $|\psi'_{\mathsf{base}}(x)|$ is *very large* on inputs x of Hamming weight at most d, but may not be large at all on inputs of larger Hamming weight. In contrast, Equation (4) requires a dual witness to be "somewhat large" (within a $2^{-O(d)}$ factor of uniform) on *nearly all* inputs.

In summary, our construction of a dual witness for $\mathsf{MP}^{\leq N}_{N^{1/2},N}$ that is sketched in the previous subsection is not sufficient to apply Razborov and Sherstov's framework to $\mathsf{SURJ}_{R,N}$, for two reasons. First, the dual witness we construct for $\mathsf{MP}^{\leq N}_{N^{1/2},N}$ is not smooth in the sense of Equation (4), as it is only "large" on inputs of Hamming weight at most d. Second, to apply Razborov and Sherstov's framework to $\mathsf{SURJ}_{R,N}$, we actually need to give a smooth dual witness for $\mathsf{SURJ}_{R,N}$ itself, not for $\mathsf{MP}^{\leq N}_{N^{1/2},N}$. Note that $\mathsf{SURJ}_{R,N}$ is defined over the domain $\{-1,1\}^n$ where $n = N \log R$, while $\mathsf{MP}^{\leq N}_{N^{1/2},N}$ is defined over subset of $\{-1,1\}^{NR}$ consisting of inputs of Hamming weight at most N.

We address both of the above issues as follows. First, we show how to turn our dual witness μ for $\mathsf{MP}^{\leq N}_{N^{1/2},N}$ into a dual witness $\hat{\sigma}$ for the fact that $\deg_{\pm}(\mathsf{SURJ}_{R,N}) \geq d$, such that $\hat{\sigma}$ inherits the "largeness" property of μ on inputs of Hamming weight at most d. Second, we transform

$\hat{\sigma}$ into a dual witness τ for the fact that $\deg_{\pm}\left(\mathsf{SURJ}_{R,N} \circ \mathsf{AND}_{\log^2 n} \circ \mathsf{PARITY}_{\log^3 n}\right) \geq d$, such that τ satisfies the smoothness condition given in Equation (4). We conclude that $\mathsf{SURJ}_{R,N} \circ \mathsf{AND}_{\log^2 n} \circ \mathsf{PARITY}_{\log^3 n}$ can be transformed into a related function F (on $\tilde{O}(n)$ inputs, and which is also in AC^0) that has sign-rank $\exp(\tilde{\Omega}(n^{1/2}))$.

3.3 The $\mathrm{PP}^{\mathrm{cc}}$ Bound: Proof of Theorem 5

As mentioned in Section 1.1.2, the core of Theorem 5 is to exhibit an AC^0 function f such that $\widetilde{\deg}_{\varepsilon}(f) = \Omega(n^{1-\delta})$ for some $\varepsilon = 1 - 2^{-\Omega(n^{1-\delta})}$. To accomplish this, we prove a hardness amplification theorem that should be understood in the context of a weaker result from our prior work [13].

As stated in Section 3.1, for $\varepsilon = 1/3$, our prior work [13] showed how to take any Boolean function f_n in AC^0 with ε-approximate degree d and transform it into a related function g on roughly the same number of variables, such that g is still in AC^0, and g has significantly higher ε'-approximate degree for some $\varepsilon' \approx 1/3$. This was done in a two-step process. First, we showed that in order to construct a "harder" function g, it is sufficient to identify an AC^0 function G defined on $\mathrm{poly}(n)$ inputs such that for some $\ell = n \cdot \mathrm{polylog}(n)$, $\widetilde{\deg}_{\varepsilon'}(G^{\leq \ell}) \gg d$.[5] Second, we exhibited such a G. In our prior works [8, 13], for general functions f_n, the function G was $f_n \circ \mathsf{AND}_r \circ \mathsf{OR}_{m'}$, where $r = 10 \log n$, and $m' = \Theta(n/d)$.

We would like to prove a similar result, but we require that G have larger ε'-approximate degree than f_n, *where ε' is exponentially closer to 1 than is ε itself.* Unfortunately, the definition of G from our prior works [8, 13] does not necessarily result in such a function. For example, if $f_n = \mathsf{OR}_n$ (or any polylogarithmic DNF for that matter), then the function $G = f_n \circ \mathsf{AND}_r \circ \mathsf{OR}_{m'}$ is also a DNF of polylogarithmic width, and it is not hard to see that all such DNFs have ε-approximate degree at most $\mathrm{polylog}(n)$ for some $\varepsilon = 1 - 1/n^{\mathrm{polylog}(n)}$.

To address this situation, we change the definition of G. Rather than defining $G := f_n \circ \mathsf{AND}_r \circ \mathsf{OR}_{m'}$, we define $G = \mathsf{GAPMAJ}_t \circ f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_m$ for appropriately chosen settings of the parameters t, z, r, and m. Here, GAPMAJ_t denotes any function evaluating to 1 on inputs of Hamming weight at most $t/3$, -1 on inputs of Hamming weight at least $2t/3$, and taking any value in $\{-1, 1\}$ on all other inputs (such functions are also called *approximate majorities*, and it is known that there are approximate majorities computable in AC^0). GAPMAJ has also played an important role in related prior work [6, 13].

In order to show that $\widetilde{\deg}_{\varepsilon'}(G^{\leq \ell}) \gg \widetilde{\deg}_{\varepsilon}(f_n)$ for an ε' that is exponentially closer to 1 than is ε, we require a more delicate construction of a dual witness than our prior works [8, 13]. After all, our prior works only required a dual witness for $G^{\leq \ell}$ with correlation at least $1/3$ with G^{ℓ}, while we require a dual witness achieving correlation with $G^{\leq \ell}$ that is exponentially close to 1. Roughly speaking, whereas our prior works [8, 13] were able to get away with exclusively using the simple and clean technique called dual block composition for constructing dual witnesses, we use a closely related but more involved construction introduced by Sherstov [30]. (Sherstov introduced his construction to prove that approximate degree satisfies a type of direct-sum theorem.)

More specifically, suppose that for some positive integer k, f_z has $\varepsilon(z)$-approximate degree at least $d(z) = z^{k/(k+1)}$, where $\varepsilon(z) = 1 - 2^{-z^{k/(k+1)}}$. In our definition of G, we set $t = n^{1/(k+2)}$, $z = n^{(k+1)/(k+2)}$, $r = 10 \log n$, and $m = n^{2/(k+2)}$, and we build a dual witness for $G^{\leq \ell}$ via a multi-step construction.

[5] This step was also used in the analysis of $\mathsf{SURJ}_{R,N}$ outlined in Section 3.2 above, where G was the function $\mathsf{AND}_R \circ \mathsf{OR}_N$.

In Step 1, we take dual witnesses ψ_{f_z}, ψ_{AND_r}, and ψ_{OR_m} for f_z, AND_r, and OR_m respectively, and we combine them using the technique of Sherstov [30], to give a dual witness γ for $f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_m$ satisfying the following properties: γ has pure high degree at least $D(n) = n^{(k+1)/(k+2)} = d(n)^{(k+1)/k} \gg d(n)$, and γ's correlation with with $f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_m$ is $\varepsilon'' \approx \varepsilon(z)$. That is, γ witnesses the fact that the ε''-approximate degree of $f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_m$ is much larger than the $\varepsilon(n)$-approximate degree of f_n itself.

This step of the construction is in contrast to our prior work, which constructed a dual witness for $f_n \circ \mathsf{AND}_r \circ \mathsf{OR}_{m'}$ via direct dual block composition of ψ_{f_n}, ψ_{AND_r}, and ψ_{OR_m}. Direct dual block composition does not suffice for us because it would yield a dual witness with significantly worse correlation with $f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_m$ than $\varepsilon(z)$.

While achieving correlation $\varepsilon'' \approx \varepsilon(z)$ is an improvement over what would obtain from direct dual block composition, it is still significantly farther from 1 than is $\varepsilon(n)$, i.e., $1 - \varepsilon'' \gg 1 - \varepsilon(n)$. And we ultimately need to construct a dual witness for $G^{\leq \ell}$ that is significantly *closer* to 1 than is $\varepsilon(n)$. To address this issue, in Step 2 of our construction, we use dual block composition to turn γ into a dual witness η for $G = \mathsf{GAPMAJ}_t \circ f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_m$ satisfying the following properties: η has the same pure high degree as γ, and moreover η has correlation at least $\varepsilon' = 1 - 2^{-\Omega(n^{(k+1)/(k+2)})}$ with G.

However, after Step 2, we are still not done, because η places some mass on inputs of Hamming weight as large as $t \cdot z \cdot r \cdot m \gg \ell$. Hence η is only a dual witness to the high ε'-approximate degree of G, not the high ε'-approximate degree of $G^{\leq \ell}$ (recall that any dual witness witness for $G^{\leq \ell}$, must evaluate to 0 on all inputs of Hamming weight larger than ℓ, cf. Equation (2)). Nonetheless, as in our prior work [8, 13], we are able to argue that η places *very little* mass on inputs of Hamming weight more than ℓ, and thereby invoke techniques from our prior work [8, 13] to zero out this mass. The reason this final step of the argument is not immediate from our prior work [8, 13] is as follows. Although prior work has developed a precise understanding of how much mass is placed on inputs of Hamming weight more than ℓ by dual witnesses constructed via basic dual block composition, the dual witness γ for $f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_m$ that we constructed in Step 1 was *not* built by invoking pure dual block composition. Our key observation is that Sherstov's technique that we invoked to construct γ is "similar enough" to vanilla dual block composition that the precise understanding of dual block composition developed in our prior work can be brought to bear on our dual witness η.

In summary, there are two main technical contributions in our proof of Theorem 5. The first is the identification of a hardness amplification construction for ε-approximate degree that not only amplifies the degree against which the lower bound holds, but also the error parameter ε. The second is constructing a dual polynomial to witness the claimed lower bound, using techniques more involved and delicate than the vanilla dual block composition technique that sufficed in our prior works [8, 13].

4 Subsequent Work and Discussion

Subsequent to our work, Sherstov and Wu [36] have made major progress toward resolving Open Problem 1 by showing nearly optimal threshold degree and sign-rank lower bounds for AC^0. Specifically, for every $k \geq 1$, they exhibit a family of depth-k AC^0 circuits with threshold degree $\tilde{\Omega}(n^{(k-1)/(k+1)})$. This generalizes Minsky and Papert's lower bound of $\Omega(n^{1/3})$ on the threshold degree of DNF, as well as our lower bound of $\tilde{\Omega}(n^{1/2})$ for the depth-3 $\mathsf{SURJECTIVITY}$ function. Sherstov and Wu, moreover, show that for any positive constant $\delta > 0$ there is a family of AC^0 circuits with depth $O(1/\delta)$ and sign-rank $\exp\left(\tilde{\Omega}(n^{1-\delta})\right)$. This gives an almost optimal improvement to our sign-rank lower bound of $\exp\left(\tilde{\Omega}(n^{1/2})\right)$ on an AC^0 function.

As in our proof of Theorem 5, as well as our prior work [13], Sherstov and Wu obtain their threshold degree lower bound for AC^0 by recursively applying a new hardness amplification theorem. Their hardness amplification theorem shows how to convert a function f_z into a new function g_n, computable by circuits with slightly higher depth and roughly the same size, but with polynomially larger threshold degree. Again as in the proof of Theorem 5, in order to obtain such a g, it suffices to construct a function G with $\deg_{\pm}(G^{\leq n}) \gg \deg_{\pm}(f)$. Starting from a function f_z with threshold degree $z^{(k-1)/(k+1)}$, the function G that they identify as sufficient for this purpose is $G = f_z \circ \mathsf{MP}_{r,r^2}$, where $z = n^{(k+1)/(k+3)}$ and $r = n^{2/(k+3)}$. When f_z is a trivial function, this recovers our lower bound of $\tilde{\Omega}(n^{1/2})$ for $\mathsf{SURJECTIVITY}$. Hence, their construction in full can be viewed as a generalization of our Theorem 6 that is amenable to recursive application. This requires several technical new ideas in the construction of the dual witness. However, we remain optimistic that the simplicity of our analysis for $\mathsf{SURJECTIVITY}$ will nonetheless lead to future applications of our techniques.

Sherstov and Wu's sign-rank lower bound follows from a similar high-level (though more technically demanding) strategy, where they show that *smooth* threshold degree also obeys such a hardness amplification theorem.

While these new results resolve the most glaring question raised in the initial version of this work, a number of interesting directions remain for further study. A common feature of our large-error approximate degree lower bound and Sherstov and Wu's threshold degree and sign-rank lower bounds for AC^0 is that, in order to obtain lower bounds of the form $\Omega(n^{1-\delta})$, we must consider functions computed by circuits of depth $\Theta(1/\delta)$. This contrasts with the situation for bounded error approximate degree [13], where a lower bound of $\Omega(n^{1-\delta})$ can be obtained at depth only $O(\log(1/\delta))$. Can one show that there are AC^0 functions f of depth $O(\log(1/\delta))$ with $\widetilde{\deg}_{\varepsilon}(f) = \Omega(n^{1-\delta})$ for $\varepsilon = 1 - 2^{-\Omega(n^{1-\delta})}$ or with $\deg_{\pm}(f) = \Omega(n^{1-\delta})$? There is a common underlying reason why our construction and Sherstov and Wu's construction both require circuits of depth $\Theta(1/\delta)$ and not $\Theta(\log(1/\delta))$: a component of the hardness amplifier in both constructions (in our case, $\mathsf{GAPMAJ}_{n^{1/(k+1)}}$, and in Sherstov and Wu's case, the top gate of MP_{r,r^2}) is used to amplify error but does not amplify degree. In contrast, in the construction of [13] for lower bounding bounded-error approximate degree, up to a logarithmic factor, all of the hardness amplifier is used to amplify degree.

We would also like to highlight the question of proving sublinear *upper bounds* on the threshold degree of AC^0. Given the surprising $O(R^{1/4} \cdot N^{1/2})$ upper bound on the $(1/3)$-approximate degree of $\mathsf{SURJ}_{R,N}$ from recent works [8,35], we have begun to seriously entertain the possibility that for every function f computable by AC^0 of depth k, there is some constant $\delta(k) > 0$ such that the threshold degree (and possibly even $(1/3)$-approximate degree) of f is $O(n^{1-\delta})$. Unfortunately, we cannot currently even show that this is true for depth three circuits of quadratic size. Any progress in this direction would be very interesting, and we believe that such progress would likely lead to new circuit lower bounds.

References

1 Eric Allender. A note on the power of threshold circuits. In *Foundations of Computer Science, 1989., 30th Annual Symposium on*, pages 580–584. IEEE, 1989.

2 László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*, pages 337–347. IEEE Computer Society, 1986. doi:10.1109/SFCS.1986.15.

3 Paul Beame and Widad Machmouchi. The quantum query complexity of AC^0. *Quantum Information & Computation*, 12(7-8):670–676, 2012. URL: http://www.rintonpress.com/xxqic12/qic-12-78/0670-0676.pdf.

4 Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded Indistinguishability and the Complexity of Recovering Secrets. In Matthew Robshaw and Jonathan Katz, editors, *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part III*, volume 9816 of *Lecture Notes in Computer Science*, pages 593–618. Springer, 2016. doi:10.1007/978-3-662-53015-3_21.

5 Andrej Bogdanov and Christopher Williamson. Approximate Bounded Indistinguishability. In Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl, editors, *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*, volume 80 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 53:1–53:11, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.ICALP.2017.53.

6 Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On The Power of Statistical Zero Knowledge. In *To Appear In Proceedings of IEEE Symposium on Foundations of Computer Science (FOCS)*, 2017. Preliminary version available at http://eccc.hpi-web.de/report/2016/140.

7 Harry Buhrman, Nikolai K. Vereshchagin, and Ronald de Wolf. On Computation and Communication with Small Bias. In *22nd Annual IEEE Conference on Computational Complexity (CCC 2007), 13-16 June 2007, San Diego, California, USA*, pages 24–32. IEEE Computer Society, 2007. doi:10.1109/CCC.2007.18.

8 Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 297–310. ACM, 2018.

9 Mark Bun and Justin Thaler. Dual Lower Bounds for Approximate Degree and Markov-Bernstein Inequalities. In Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg, editors, *ICALP (1)*, volume 7965 of *Lecture Notes in Computer Science*, pages 303–314. Springer, 2013. doi:10.1007/978-3-642-39206-1_26.

10 Mark Bun and Justin Thaler. Hardness Amplification and the Approximate Degree of Constant-Depth Circuits. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, volume 9134 of *Lecture Notes in Computer Science*, pages 268–280. Springer, 2015. Full version available at http://eccc.hpi-web.de/report/2013/151. doi:10.1007/978-3-662-47672-7_22.

11 Mark Bun and Justin Thaler. Approximate Degree and the Complexity of Depth Three Circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:121, 2016. URL: http://eccc.hpi-web.de/report/2016/121.

12 Mark Bun and Justin Thaler. Improved Bounds on the Sign-Rank of AC^0. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 37:1–37:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016. doi:10.4230/LIPIcs.ICALP.2016.37.

13 Mark Bun and Justin Thaler. A Nearly Optimal Lower Bound on the Approximate Degree of AC^0. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 1–12, 2017. doi:10.1109/FOCS.2017.10.

14 Kuan Cheng, Yuval Ishai, and Xin Li. Near-Optimal Secret Sharing and Error Correcting Codes in AC^0. In *Theory of Cryptography Conference*, pages 424–458. Springer, 2017.

15 Vitaly Feldman. Evolvability from learning algorithms. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 619–628. ACM, 2008.

16 Jürgen Forster. A Linear Lower Bound on the Unbounded Error Probabilistic Communication Complexity. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity, Chicago, Illinois, USA, June 18-21, 2001*, pages 100–106. IEEE Computer Society, 2001. doi:10.1109/CCC.2001.933877.

17 Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans Ulrich Simon. Relations Between Communication Complexity, Linear Arrangements, and Computational Complexity. In Ramesh Hariharan, Madhavan Mukund, and V. Vinay, editors, *FST TCS 2001: Foundations of Software Technology and Theoretical Computer Science, 21st Conference, Bangalore, India, December 13-15, 2001, Proceedings*, volume 2245 of *Lecture Notes in Computer Science*, pages 171–182. Springer, 2001. doi:10.1007/3-540-45294-X_15.

18 Hartmut Klauck. Lower bounds for quantum communication complexity. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 288–297. IEEE, 2001.

19 Adam R. Klivans and Rocco A. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. Syst. Sci.*, 68(2):303–318, 2004. doi:10.1016/j.jcss.2003.07.007.

20 Matthias Krause and Pavel Pudlák. On the Computational Power of Depth-2 Circuits with Threshold and Modulo Gates. *Theor. Comput. Sci.*, 174(1-2):137–156, 1997. doi:10.1016/S0304-3975(96)00019-9.

21 Troy Lee. A note on the sign degree of formulas. *CoRR*, abs/0909.4607, 2009. arXiv:0909.4607.

22 Nati Linial and Adi Shraibman. Learning complexity vs communication complexity. *Combinatorics, Probability and Computing*, 18(1-2):227–245, 2009.

23 Marvin Minsky and Seymour Papert. *Perceptrons - an introduction to computational geometry*. MIT Press, 1969.

24 Noam Nisan. The Communication Complexity of Threshold Gates. In *Combinatorics, Paul Erdos is Eighty*, pages 301–315, 1994.

25 Ryan O'Donnell and Rocco A. Servedio. New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010. Preliminary version in STOC 2003. doi:10.1007/s00493-010-2173-3.

26 Ramamohan Paturi and Janos Simon. Probabilistic Communication Complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986. doi:10.1016/0022-0000(86)90046-2.

27 Vladimir V Podolskii. Perceptrons of large weight. In *International Computer Science Symposium in Russia*, pages 328–336. Springer, 2007.

28 Alexander A. Razborov and Alexander A. Sherstov. The Sign-Rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010. doi:10.1137/080744037.

29 Alexander A. Sherstov. Separating AC^0 from Depth-2 Majority Circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. doi:10.1137/08071421X.

30 Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 41–50. ACM, 2011. doi:10.1145/1993636.1993643.

31 Alexander A. Sherstov. The Pattern Matrix Method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. Preliminary version in *STOC* 2008. doi:10.1137/080733644.

32 Alexander A. Sherstov. The Intersection of Two Halfspaces Has High Threshold Degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013. Preliminary version in FOCS 2009. doi:10.1137/100785260.

33 Alexander A. Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 223–232. ACM, 2014. doi:10.1145/2591796.2591871.

34 Alexander A. Sherstov. The Power of Asymmetry in Constant-Depth Circuits. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 431–450, 2015. doi:10.1109/FOCS.2015.34.

35 Alexander A. Sherstov. Algorithmic polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:10, 2018. To appear in STOC 2018. URL: `https://eccc.weizmann.ac.il/report/2018/010`.

36 Alexander A. Sherstov and Pei Wu. Near-Optimal Lower Bounds on the Threshold Degree and Sign-Rank of AC^0. *Electronic Colloquium on Computational Complexity (ECCC)*, 26:3, 2019. URL: `https://eccc.weizmann.ac.il/report/2019/003`.

37 Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009. URL: `http://www.rintonpress.com/xxqic9/qic-9-56/0444-0460.pdf`.

38 Leslie G Valiant. Evolvability. *Journal of the ACM (JACM)*, 56(1):3, 2009.