THEORY OF COMPUTING, Volume 17 (7), 2021, pp. 1–46
www.theoryofcomputing.org

SPECIAL ISSUE: APPROX-RANDOM 2019

The Large-Error Approximate Degree of AC^0

Mark Bun* Justin Thaler†

Received January 24, 2020; Revised August 18, 2020; Published September 23, 2021

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'15). 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'16).

Second, for any $\delta > 0$, we exhibit a function $f : \{-1,1\}^n \to \{-1,1\}$ 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 $n^{1-\delta}$. In our FOCS'17 paper we 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 measures, including discrepancy, margin complexity, and threshold weight, that are central

An extended abstract of this paper appeared in the Proceedings of the 23rd International Conference on Randomization and Computation, 2019 [17].

*Supported by NSF Grant CCF-1947889. This work was done while the author was at Princeton University, and at the Simons Institute for the Theory of Computing supported by a Google Research Fellowship.

†Supported by NSF Grant CCF-1845125.

ACM Classification: F.1.2, F.1.3

AMS Classification: 68Q15, 68Q17, 68Q32

Key words and phrases: approximate degree, polynomial approximation, polynomial threshold, polynomial method, dual polynomials, surjectivity, discrepancy

 DOI: 10.4086/toc.2021.v017a007

to communication complexity and learning theory. 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'15). Additional applications in learning theory, communication complexity, and cryptography are described.

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 $|p(x) - f(x)| \le \varepsilon$ 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 lower bounds on the ε-approximate degree 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 (see 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. Briefly, both $\mathbf{UPP}^{\mathrm{dt}}$ and $\mathbf{PP}^{\mathrm{dt}}$ capture the minimum number of queries required to evaluate a Boolean function with probability at least $1/2$ on every input; $\mathbf{UPP}^{\mathrm{cc}}, \mathbf{PP}^{\mathrm{cc}}$ do the same for two-party communication problems. It is known that the threshold degree of f is equivalent to its complexity $\mathbf{UPP}^{\mathrm{dt}}(f)$, while the logarithm of 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 $d > 0$, $\widetilde{\deg}_{1-2^{-d}}(f) \ge \Omega(d) \iff \mathbf{PP}^{\mathrm{dt}}(f) \ge \Omega(d)$. As described in Section 2, $\mathbf{PP}^{\mathrm{dt}}$ is also closely related to the notion of *threshold weight*. Section 2 elaborates on these models and their many applications in learning theory, circuit complexity, and cryptography.

[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 $1 - 2^{-\tilde{\Omega}(n^d)}$ [37], and this bound is tight [11]. 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.

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 $\mathsf{SURJECTIVITY}$, 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 [44]. Our analysis is much simpler than Sherstov's, which takes up the bulk of a (70+)-page manuscript [44]. An additional advantage of our analysis is that our lower bound on the threshold degree of $\mathsf{SURJECTIVITY}$ "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 $\Omega(n^{2/5})$ [15].

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.

To summarize our results succinctly:

- We prove an $\tilde{\Omega}(n^{1/2})$ lower bound on the $\mathbf{UPP}$ complexity of $\mathsf{SURJECTIVITY}$ in the query setting, and of a related AC^0 function in the communication setting.
- We prove an $\Omega(n^{1-\delta})$ lower bound on the $\mathbf{PP}$ 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 work 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 [27] runs in time $\exp(\tilde{O}(n^{1/3}))$ owing to an $\tilde{O}(n^{1/3})$ 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 [33] 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 [39, 40] and Buhrman et al. [11] 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 [3] classic simulation of AC^0 functions by depth-three majority circuits is optimal. (This resolved an open problem of Krause and Pudlák [30].) A subsequent, related breakthrough of Razborov and Sherstov [38] 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. [5].

These breakthrough lower bounds raised the intriguing possibility that AC^0 functions could be *maximally* hard for the $\mathbf{UPP}^{\mathrm{cc}}$ and $\mathbf{PP}^{\mathrm{cc}}$ communication models, as well as for related complexity measures. Nevertheless, the quantitative parameters achieved in these papers are far from actually

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

showing that this is the case. Indeed, the following basic questions about the complexity of AC^0 remain open.

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

Open Problem 2. Is there an AC^0 function $F\colon \{-1,1\}^{n\times n} \to \{-1,1\}$ with $\mathbf{PP}^{\mathrm{cc}}$ 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 $\mathbf{UPP}^{\mathrm{cc}}$ and $\mathbf{PP}^{\mathrm{cc}}$ complexity at most n. Obtaining an affirmative answer to Open Problem 1 is harder than for Open Problem 2, since $\mathbf{UPP}^{\mathrm{cc}}(f) \le \mathbf{PP}^{\mathrm{cc}}(f)$ for all f.

Guided by these open problems, a sequence of papers has established quantitatively stronger and more general lower bounds for AC^0 functions [13–16, 18, 43, 44]. In addition to making partial progress toward resolving these questions, the techniques developed in this body of work have found fruitful applications in new domains. For example, Bouland et al. [9] built on techniques from some of these papers [13–15, 43] to resolve several old open questions about the relativized power of statistical zero knowledge proofs and their variants. As another example, our recent papers [12, 18] 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 [7, 8, 19]. We thus believe that the new techniques developed in this article will find further applications, perhaps in unexpected areas.

Prior to our work, the best known result toward a resolution of Open Problem 1 was an $\Omega(n^{2/5})$ lower bound on $\mathbf{UPP}^{\mathrm{cc}}$ complexity of an AC^0 function [15], while the best known result toward Open Problem 2 was an $\Omega(n^{1/2})$ bound on the $\mathbf{PP}^{\mathrm{cc}}$ complexity of a very complicated AC^0 circuit [44].

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 $\mathsf{SURJECTIVITY}_{R,N}$ ($\mathsf{SURJ}_{R,N}$ 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 $(s_1,\dots,s_N)$ from a range $[R] := \{1,\dots,R\}$, 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 [6] 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 recently shown to be $\tilde{\Theta}(R^{1/4}\cdot N^{1/2})$. The lower bound is from our prior work [12], while the upper bound was shown by Sherstov [45], with a different proof given in [12]. In particular, when $R = N/2$, $\widetilde{\deg}_{1/3}(\mathsf{SURJ}_{R,N}) = \tilde{\Theta}(N^{3/4})$. Our prior work [12, 18] 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).

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

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 (we prove this in Section 6 for completeness). 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 1.1. *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 1.1 matches the best prior threshold degree lower bound for AC^0, previously proved in [44] 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 an $\tilde{\Omega}(n^{1/2})$ lower bound on the $\mathbf{UPP}^{\mathrm{cc}}$ complexity of a related AC^0 function, yielding progress on Open Question 1 (see Section 1). In contrast, Sherstov's $\Omega(n^{1/2})$ threshold degree lower bound for AC^0 [44] 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 1.2. *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 $\mathbf{PP}^{\mathrm{cc}}$ 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 earlier results [12,18], which gave a similar result for $\varepsilon = 1/3$. By combining this large-error approximate-degree lower bound with a "query-to-communication lifting theorem" for **PP** [40], we obtain an $\Omega(n^{1-\delta})$ bound on the $\mathbf{PP}^{\mathrm{cc}}$ complexity of an AC^0 function, nearly resolving Open Question 2 from the previous section.

Theorem 1.3. *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 $\Omega(n^{1/2})$ [44].

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.

Reference	$\mathbf{PP}^{\mathrm{dt}}$ and $\log(\text{threshold weight})$	$\mathbf{PP}^{\mathrm{cc}}$ $\approx \log(1/\text{discrepancy})$	$\mathbf{UPP}^{\mathrm{dt}}$ $=$ threshold degree	$\mathbf{UPP}^{\mathrm{cc}}$ $\approx \log(\text{sign-rank})$	Circuit Depth
[33]	—	—	$\Omega(n^{1/3})$	—	2
[30]	$\Omega(n^{1/3})$	—	—	—	3
[21]	—	$\Omega(\log^k(n))$	—	$\Omega(\log^k(n))$	$O(k)$
[35]	$\Omega(n^{1/3}\log^k n)$	—	$\Omega(n^{1/3}\log^k(n))$	—	$O(k)$
[39]	—	$\Omega(n^{1/5})$	—	—	3
[11,40]	—	$\Omega(n^{1/3})$	—	—	3
[38]	—	—	—	$\Omega(n^{1/3})$	3
[14]	$\Omega(n^{2/5})$	$\Omega(n^{2/5})$	—	—	3
[43]	$\Omega(n^{1/2-\delta})$	$\Omega(n^{1/2-\delta})$	$\Omega(n^{1/2-\delta})$	—	$O(1/\delta)$
[44]	$\Omega(n^{3/7})$	—	$\Omega(n^{3/7})$	—	3
[44]	$\Omega(n^{1/2})$	$\Omega(n^{1/2})$	$\Omega(n^{1/2})$	—	4
[15]	—	—	—	$\Omega(n^{2/5})$	3
[16]	$\Omega(n^{1/2-\delta})$	$\Omega(n^{1/2-\delta})$	—	—	3
This paper	$\tilde{\Omega}(n^{1/2})$	$\tilde{\Omega}(n^{1/2})$	$\tilde{\Omega}(n^{1/2})$	—	3
This paper	—	—	—	$\tilde{\Omega}(n^{1/2})$	7
This paper	$\Omega(n^{1-\delta})$	$\Omega(n^{1-\delta})$	—	—	$O(1/\delta)$

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 of polynomial size.

- $\mathbf{PP}^{\mathrm{dt}}(f)$ 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 **PP**-cost of the algorithm is the sum of the query cost and $\log(1/\beta)$. The $\log(1/\beta)$ term here is a proxy for the number of bits of randomness used by the algorithm. Hence, the **PP**-cost of a query algorithm is a lower bound on the runtime, assuming both a query and a random coin toss cost at least one time step. $\mathbf{PP}^{\mathrm{dt}}(f)$ is the minimum **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).

- $\mathbf{UPP}^{\mathrm{cc}}(f)$ [36] 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).
- $\mathbf{PP}^{\mathrm{cc}}(f)$ [5] is the minimum **PP**-cost of a protocol for f, where the **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 1.3 and Corollary 1.2. Further technical background and details are given in Section 9.

2.1 Applications of Theorem 1.3

$\mathbf{PP}^{\mathrm{cc}}$ is known to be equivalent to two measures of central importance in learning theory and communication complexity, namely *margin complexity* [32] and *discrepancy* [26]. Hence, Theorem 1.3 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. [7] 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 1.3 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 [48] introduced the *evolvability* model to quantify how complex functions can emerge through an evolutionary process over realistic population sizes within realistic time periods. Feldman [20] showed that the "weak evolvability" of a class of functions $\mathcal{F} = \{\phi_1, \dots, \phi_{|\mathcal{F}|}\}$ is characterized by the $\mathbf{PP}^{\mathrm{cc}}$ complexity of the function $F(x,y) = \phi_x(y)$. Hence, a consequence of Theorem 1.3 is that there are AC^0 functions that are nearly maximally hard to evolve. That is, 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 evolution to produce a function that has advantage just $2^{-n^{1-\delta}}$ over random guessing. See [20] for additional details.
- **Circuit Complexity**. If $\mathbf{PP}^{\mathrm{cc}}(f) \geq d$, then f is not computable by Majority-of-Threshold circuits of size $2^{\Omega(d)}$ [34]. Hence, by showing that AC^0 has nearly maximal $\mathbf{PP}^{\mathrm{cc}}$ 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 [3]).

2.2 Applications of Corollary 1.2

As indicated in Section 1, $\mathbf{UPP}^{\mathrm{cc}}(F)$ is known to be characterized by (the logarithm of) of the *sign-rank* of the matrix $[F(x,y)]_{x,y\in\{-1,1\}^{n\times n}}$ [36].[4] Hence, Corollary 1.2 implies an $\exp(\tilde{\Omega}(n^{1/2}))$ lower bound on

[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.

the sign-rank of AC^0 function. Below, we highlight two additional applications of Corollary 1.2, based on the following connections between communication complexity, circuit complexity, and learning theory.

In communication complexity, $\mathbf{UPP}^{\mathrm{cc}}$ is the most powerful two-party model against which we know how to prove lower bounds. In circuit complexity, if $\mathbf{UPP}^{\mathrm{cc}}(f) \geq d$, then f cannot be computed by Threshold-of-Majority circuits of size $2^{\Omega(d)}$ [22]. (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 $\mathbf{UPP}^{\mathrm{cc}}$-complexity of a concept class precisely captures how many dimensions are needed. To connect this to a previously mentioned example, Klivans and Servedio [27] observed that an upper bound of d on the $\mathbf{UPP}^{\mathrm{cc}}$ 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 1.2 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.** $\mathbf{UPP}^{\mathrm{cc}}$-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

$$|p(x) - f(x)| \leq \varepsilon \text{ for all inputs } x \text{ of Hamming weight at most } N. \tag{3.1}$$

Note that Property (3.1) 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 [18] showed that 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 3.1. *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 3.1 is a substantial strengthening of the classic result of Minsky and Papert [33] 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 3.1 establishes that Minsky and Papert's lower bound holds even under the promise that the n-bit input to $\mathsf{MP}_{N^{1/2},N}$ 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 3.1 and comparison to prior work. Like much recent work on approximate-degree 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 conditions:

- $\psi(x) \cdot f(x) \geq 0$ 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 [44] 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 [14], 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 [31,42,47] (see Section 5.8 for details of this very important technique for combining dual witnesses).

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 [44], 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 3.1 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 3.1, 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:

$$\psi(x) = 0 \text{ for all inputs } x \text{ of Hamming weight more than } N. \tag{3.2}$$

(Here, the Hamming weight of x is the number of -1 entries.) 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:

$$|\psi'_{\mathsf{base}}(x)| \gg n^{-d} \text{ for all inputs } x \text{ of Hamming weight at most } d\,. \tag{3.3}$$

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 work [12, 18], 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 (3.2) 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 in our prior work we 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 [44]. For example, prior work of Bouland et al. [9] 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 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 of this lower bound on threshold degree. We expect 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 have the property we need (see Equation (3.3)) 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 earlier paper [14]. 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).

3.2 Extension to UPP$^{\text{cc}}$: Proof of Corollary 1.2

Building on the celebrated framework of Forster [21], Razborov and Sherstov [38] 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 **UPP**$^{\text{cc}}$ 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:

$$|\phi(x)| \geq 2^{-O(d)} \cdot 2^{-n} \text{ for all but a } 2^{-\Omega(d)} \text{ fraction of inputs } x \in \{-1,1\}^n. \tag{3.4}$$

Note that this is a different smoothness condition than the one satisfied by the dual witness ψ'_{base} discussed above for $\mathsf{MP}_{N^{1/2},N}$ (Equation (3.3)): 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 (3.4) 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 (3.4), 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 (3.4). 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 $\mathsf{PP}^{\mathsf{cc}}$ bound: Proof of Theorem 1.3

As mentioned in Section 1.1.2, the core of Theorem 1.3 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 [18].

As stated in Section 3.1, for $\varepsilon = 1/3$, our prior work [18] 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 $\operatorname{poly}(n)$ inputs such that for some $\ell = n \cdot \operatorname{polylog}(n)$, $\widetilde{\deg}_{\varepsilon'}(G^{\leq \ell}) \gg d$. [5] Second, we exhibited such a G. In our prior work [12, 18], 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 earlier papers [12, 18] 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 $\operatorname{polylog}(n)$ for some $\varepsilon = 1 - 1/n^{\operatorname{polylog}(n)}$.

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. (For simplicity of exposition, some of these parameters are described differently here than in the proof in Section 8.) 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 [9, 18].

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 papers [12, 18]. After all, those papers 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, in [12, 18] we were able to get away with exclusively using the simple and clean technique called dual block composition (described in Section 5.8) for constructing dual witnesses for block composed functions by combining dual witnesses for the constituent functions. In this article, we instead use a closely related but more involved construction introduced by Sherstov [41]. (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 intermediate parameters $t = n^{1/(k+2)}$, $z = n^{(k+1)/(k+2)}$, $r = 10 \log n$, and $m = n^{2/(k+2)}$, and we build a dual witness for $G^{\leq \ell}$ via a multi-step construction.

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 [41], to give a dual witness γ for $f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_m$ satisfying the following conditions: 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

[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 $\mathsf{AND}_R \circ \mathsf{OR}_N$.

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 conditions: 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 for $G^{\leq \ell}$, must evaluate to 0 on all inputs of Hamming weight larger than ℓ, as described in Equation (3.2)). Nonetheless, as in our prior work [12, 18], we are able to argue that η places *very little* mass on inputs of Hamming weight more than ℓ, and thereby invoke techniques from these papers to zero out this mass. The reason this final step of the argument is not immediate from [12, 18] is as follows. Although these papers 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 1.3. 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 [12, 18].

4 Subsequent work and discussion

Subsequent to our work, Sherstov and Wu [46] 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 1.3 (see Theorem 8.2 in Section 8), as well as our earlier paper [18], Sherstov and Wu obtain their lower bound on the threshold degree of 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 1.3, in order to obtain such a g, it suffices to construct a function G with $\deg_\pm(G^{\le n}) \gg \deg_\pm(f)$. 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 SURJECTIVITY. Hence, their construction in full can be viewed as a generalization of our Theorem 3.1 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 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 paper, 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 [18], 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 [18] 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 work [12, 45], 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.

Outline for the rest of the paper. Section 5 covers technical preliminaries. Section 7 contains the proof of our tight threshold-degree lower bound for SURJECTIVITY (Theorem 1.1) and its extension to a sign-rank lower bound for a related function in AC^0 (Corollary 1.2). Section 8 proves our nearly optimal bound on the discrepancy of AC^0 (Theorem 1.3). Section 9 elaborates on applications of these results in communication complexity, circuit complexity, learning theory, and cryptography.

5 Preliminaries

5.1 Notation

For a natural number N, let $[N] = \{1,2,\dots,N\}$ and $[N]_0 = \{0,1,2,\dots,N\}$. All logarithms in this paper are taken in base 2 unless otherwise noted via the notation ln, which refers to base e. For $x \in \mathbb{R}$, define $\mathrm{sgn}(x) = -1$ if $x < 0$ and 1 otherwise. For any set S, the notation $x \sim S$ means that x is drawn uniformly at random from S.

As mentioned in Section 1, we sometimes use subscripts to indicate the number of variables on which a function is defined. For example, we denote $\mathsf{OR}\colon \{-1,1\}^n \to \{-1,1\}$ by OR_n.

For any input $x \in \{-1,1\}^n$, $|x| = |\{i\colon x_i = -1\}|$ denotes the Hamming weight of x. $\mathcal{H}_n$ denotes the set $\{-1,1\}^n$, and for any $d \le n$, we define $\mathcal{H}_n^{\le d} := \{x \in \{-1,1\}^n\colon |x| \le d\}$. Similarly, $\mathcal{H}_n^{> d} := \{x \in \{-1,1\}^n\colon |x| > d\}$. Given a Boolean function f_n, we denote by $f_n^{\le d}$ the *partial* function obtained by restricting the domain of f_n to $\mathcal{H}_n^{\le d}$.

5.2 Threshold degree and its dual formulation

Definition 5.1. Let $D \subseteq \{-1,1\}^n$, and let f be a function mapping D to $\{-1,1\}$. The *threshold degree* of f, denoted $\deg_{\pm}(f)$, is the least degree of a real polynomial $p\colon \{-1,1\}^n \to \mathbb{R}$ such that $p(x) \cdot f(x) > 0$ when $x \in \mathcal{D}$. No constraint is placed on $p(x)$ for any $x \in (\{-1,1\}^n \setminus D)$.

In this paper, we make essential use of a (standard) dual formulation of threshold degree. To describe this dual formulation, we need to introduce some terminology.

Definition 5.2. Let $\psi\colon \{-1,1\}^n \to \mathbb{R}$ be any real-valued function on the Boolean hypercube. Given another function $p\colon \{-1,1\}^n \to \mathbb{R}$, we let $\langle \psi, p\rangle := \sum_{x\in\{-1,1\}^n} \psi(x)\cdot p(x)$, and refer to $\langle \psi, p\rangle$ as the *correlation* of ψ with p. If $\langle \psi, p\rangle = 0$ for all polynomials p of degree at most d, we say that ψ has *pure high degree* at least d. We let $\|\psi\|_1 := \sum_{x\in\{-1,1\}^n} |\psi(x)|$, and refer to $\|\psi\|_1$ as the ℓ_1-norm of ψ.

The following theorem provides the aforementioned dual formulation of threshold degree.

Theorem 5.3. *Let $f : D \to \{-1,1\}$ with $D \subseteq \{-1,1\}^n$. Then $\deg_{\pm}(f) > d$ if and only if there is a real function $\psi : \{-1,1\}^n \to \mathbb{R}$ such that:*

1. *(Pure high degree): ψ has pure high degree at least d.*
2. *(Non-triviality): $\|\psi\|_1 > 0$.*
3. *(Sign Agreement): $\psi(x)\cdot f(x) \ge 0$ for all $x \in \{-1,1\}^n$.*
4. *(Appropriate Support): $\psi(x) = 0$ for all $x \in \{-1,1\}^n \setminus D$.*

We refer to functions mapping $\{-1,1\}^n \to \mathbb{R}$ as *dual polynomials*. For a function $f\colon D \to \{-1,1\}$, we refer to any ψ satisfying the conditions of Theorem 5.3 as a *threshold degree dual polynomial for* f, or alternatively as a *dual witness to the fact that* $\deg_{\pm}(f) \ge d$. Along the way to constructing a threshold-degree dual polynomial for f, we will often construct dual polynomials that *almost* satisfy the Sign Agreement and Appropriate Support conditions, and it will be convenient to give names to the inputs on which these two conditions are violated. To this end, given a dual polynomial $\psi\colon \{-1,1\}^n \to \mathbb{R}$, let $\mathcal{E}(\psi, f) = \{x \in D\colon \psi(x)\cdot f(x) < 0\}$, and let $\mathcal{W}(\psi, f) := \{x \in \{-1,1\}^n \setminus D\colon |\psi(x)| > 0\}$. For each input $x \in \mathcal{E}$, we say that ψ *makes an error* on x. We will let $E(\psi, f) := \sum_{x\in\mathcal{E}(\psi,f)} |\psi(x)|$ and $W(\psi, f) := \sum_{x\in\mathcal{W}(\psi,f)} |\psi(x)|$.

5.3 Approximate degree and its dual formulation

Definition 5.4. Let $D \subseteq \{-1,1\}^n$, and let f be a function mapping D to $\{-1,1\}$. The ε-*approximate degree* of f, denoted $\widetilde{\deg}_\varepsilon(f)$, is the least degree of a real polynomial $p\colon \{-1,1\}^n \to \mathbb{R}$ such that $|p(x) - f(x)| \le \varepsilon$ when $x \in \mathcal{D}$. No constraint is placed on $p(x)$ for any $x \in (\{-1,1\}^n \setminus D)$.[6]

The following theorem provides a standard dual formulation of approximate degree.

Theorem 5.5. *Let $f : D \to \{-1,1\}$ with $D \subseteq \{-1,1\}^n$. Then $\widetilde{\deg}_\varepsilon(f) > d$ if and only if there is a real function $\psi : \{-1,1\}^n \to \mathbb{R}$ such that:*

1. *(Pure high degree): ψ has pure high degree at least d.*

2. *(Non-triviality): $\|\psi\|_1 > 0$.*

3. *(Correlation): $\sum_{x\in\{-1,1\}^n} \psi(x)\cdot f(x) \ge \varepsilon \cdot \|\psi\|_1$.*

4. *(Appropriate Support): $\psi(x) = 0$ for all $x \in \{-1,1\}^n \setminus D$.*

We will also frequently use the following simple fact, which is immediate from linearity.

Fact 5.6. *Suppose that $\psi_1, \dots, \psi_k\colon \{-1,1\}^n \to \mathbb{R}$ all have pure high degree at least d. Then their sum, $\psi_1 + \cdots + \psi_k$, has pure high degree at least d.*

5.4 The Surjectivity function

Definition 5.7. For $N \ge R$, define the function $\mathsf{SURJ}_{R,N} : [R]^N \to \{-1,1\}$ by $\mathsf{SURJ}_{R,N}(s_1, \dots, s_N) = -1$ iff for every $j \in [R]$, there exists an i such that $s_i = j$.

When N and R are clear from context, we will refer to the function SURJ without the explicit dependence on these parameters. It will often be convenient to think of the input to $\mathsf{SURJ}_{R,N}$ as a function mapping $\{-1,1\}^n \to \{-1,1\}$ rather than $[R]^N \to \{-1,1\}$. When needed, we assume that R is a power of 2 and an element of $[R]$ is encoded in binary using $\log R$ bits. In this case we will view Surjectivity as a function on $n = N \log R$ bits, i. e., $\mathsf{SURJ} : \{-1,1\}^n \to \{-1,1\}$.

For technical reasons, when proving lower bounds, it will be more convenient to work with a variant of SURJ where the range $[R]$ is augmented by a "dummy element" 0 that is simply ignored by the function. That is, while any of the items $s_1, \dots, s_N$ may take the dummy value 0, the presence of a 0 in the input is not required for the input to be deemed surjective. We denote this variant of Surjectivity by dSURJ. More formally:

[6] Prior work (e. g., [1, 12]) has also considered another natural definition of approximate degree for promise problems, that does require $p(x)$ to be bounded in magnitude outside of the domain D on which f is defined. For our purposes in this paper, the definition that does not constrain p on inputs outside of D is most useful.

Definition 5.8. For $N \geq R$, define the function $\mathsf{dSURJ}_{R,N} : [R]_0^N \to \{-1,1\}$ by $\mathsf{dSURJ}_{R,N}(s_1,\ldots,s_N) = -1$ iff for every $j \in [R]$, there exists an $i \in [N]$ such that $s_i = j$.

The following simple proposition shows that a lower bound on the approximate degree of dSURJ implies a lower bound for SURJ itself.

Proposition 5.9 (Bun, Kothari, and Thaler [12])**.** *Let $\varepsilon > 0$ and $N \geq R$. Then*

$$\widetilde{\deg}_\varepsilon(\mathsf{dSURJ}_{R,N}) \leq \widetilde{\deg}_\varepsilon(\mathsf{SURJ}_{R+1,N+1}) \cdot \log(R+1)\,.$$

We will also use two additional simple properties of $\mathsf{SURJ}_{R,N}$. The first roughly says that increasing the range size can only make $\mathsf{SURJECTIVITY}$ harder to approximate (so long as the domain size remains significantly larger than the range size).

Proposition 5.10. *Let $\varepsilon > 0$. Then for any $R' \geq R$, $\widetilde{\deg}_\varepsilon(\mathsf{SURJ}_{R,N}) \leq \widetilde{\deg}_\varepsilon(\mathsf{SURJ}_{R',N+R'-R})$.*

Proof. Let $p : \{-1,1\}^{(N+R'-R)\cdot\lceil\log(R')\rceil} \to \{-1,1\}$ be a polynomial of degree d that ε-approximates $\mathsf{SURJ}_{R',N+R'-R}$. We will use to p to construct a polynomial of degree d that ε-approximates $\mathsf{SURJ}_{R,N}$. Recall that an input to $\mathsf{SURJ}_{R,N}$ takes the form $(s_1,\ldots,s_N)$ where each s_i is the binary representation of a number in $[R]$. For every $(s_1,\ldots,s_N) \in [R]^N$, observe that

$$\mathsf{SURJ}_{R,N}(s_1,\ldots,s_N) = \mathsf{SURJ}_{R',N+R'-R}(s_1,\ldots,s_N,R+1,R+2,\ldots,R')\,.$$

Hence, the polynomial

$$p(s_1,\ldots,s_N,R+1,R+2,\ldots,R')$$

has degree at most d and ε-approximates $\mathsf{SURJ}_{R,N}$. This assumes that for each element r of $[R]$, each bit of the encoding of r in $\{-1,1\}^{\lceil\log R'\rceil}$ (i. e., when r is viewed as an element of $[R']$), is a degree-1 function of its encoding in $\{-1,1\}^{\lceil\log R\rceil}$ (i. e., when r is viewed as an element of $[R]$). This property holds for the natural binary encoding. Even if some encoding were used that did not satisfy this condition, then $p(s_1,\ldots,s_N,R+1,R+2,\ldots,R')$ will have degree no larger than $d\cdot(\log(R')+1)$. □

The second says that increasing the domain size can only make $\mathsf{SURJECTIVITY}$ harder to approximate.

Proposition 5.11. *Let $\varepsilon > 0$. If $N' > N$, then $\widetilde{\deg}_\varepsilon(\mathsf{SURJ}_{R,N}) \leq \widetilde{\deg}_\varepsilon(\mathsf{SURJ}_{R+1,N'})$.*

Proof. For every $(s_1,\ldots,s_N) \in [R]^N$, observe that

$$\mathsf{SURJ}_{R,N}(s_1,\ldots,s_N) = \mathsf{SURJ}_{R+1,N'}(s_1,\ldots,s_N,R+1,R+1,\ldots,R+1),$$

where the element $R+1$ is repeated $N'-N$ times. Hence, if $p : \{-1,1\}^{N'\cdot\lceil\log(R+1)\rceil} \to \{-1,1\}$ is a polynomial of degree d that ε-approximates $\mathsf{SURJ}_{R+1,N'}$, then $p(s_1,\ldots,s_N,R+1,R+1,\ldots,R+1)$ is a degree d polynomial that ε-approximates $\mathsf{SURJ}_{R,N}$. □

5.5 A useful auxiliary function

The following lemma, due to Razborov and Sherstov [38], gives a function that we will use many times in this paper to "zero out" mass that intermediate dual witnesses ψ place on the "bad" sets $\mathcal{E}(\psi, f)$ and $\mathcal{W}(\psi, f)$.

Lemma 5.12 ([38, Proof of Lemma 3.2]). *Let $D, n \in \mathbb{N}$ with $0 \le D \le n-1$. Then for every $y \in \{-1,1\}^n$ with $|y| > D$, there exists a function $\phi_y : \{-1,1\}^n \to \mathbb{R}$ such that*

$$\phi_y(y) = 1 \tag{5.1}$$

$$|x| > D, x \neq y \implies \phi_y(x) = 0 \tag{5.2}$$

$$\deg p \le D \implies \langle \phi_y, p \rangle = 0 \tag{5.3}$$

$$\sum_{|x| \le D} |\phi_y(x)| \le 2^D \binom{|y|}{D}. \tag{5.4}$$

5.6 A dual witness for OR

We will repeatedly make use of a dual witness for the high approximate degree of the OR function. The dual witness itself was essentially first constructed by Špalek [50], but we require an analysis of it due to Bun, Kothari, and Thaler [12].

Proposition 5.13 ([12]). *There exist constants $c_1, c_2 \in (0,1)$ for which the following holds. Let $T \in \mathbb{N}$ and $1/T \le \delta \le 1/2$. Then there is a function $\omega : [T]_0 \to \mathbb{R}$ such that*

$$\omega(0) - \sum_{t=1}^{T} \omega(t) \ge 1 - \delta, \tag{5.5}$$

$$\omega(0) \ge (1-\delta)/2, \tag{5.6}$$

$$\sum_{t=0}^{T} |\omega(t)| = 1, \tag{5.7}$$

$$\text{for all univariate polynomials } q\colon \mathbb{R} \to \mathbb{R},\ \deg q < c_1\sqrt{\delta T} \implies \sum_{t=0}^{T} \omega(t) \cdot q(t) = 0, \tag{5.8}$$

$$|\omega(t)| \le \frac{170 \exp(-c_2 t \sqrt{\delta}/\sqrt{T})}{\delta \cdot t^2} \qquad \forall t = 1, \ldots, T. \tag{5.9}$$

Proposition 5.14. *Let $T, N \in \mathbb{N}$ with $T \le N$, and let $\delta > 1/T$. Define ω as in Proposition 5.13. Define the function $\psi : \{-1,1\}^N \to \mathbb{R}$ by $\psi(x) = \omega(|x|)/\binom{N}{|x|}$ for $x \in \mathcal{H}_N^{\le T}$ and $\psi(x) = 0$ otherwise. Then ψ*

satisfies:

$$\langle \psi, \mathsf{OR}_N \rangle \geq 1 - \delta\,, \tag{5.10}$$

$$\psi(\mathbf{1}_N) \geq (1-\delta)/2\,, \tag{5.11}$$

$$\|\psi\|_1 = 1\,, \tag{5.12}$$

$$\text{for any polynomial } p\colon \{-1,1\}^N \to \mathbb{R},\ \deg p < c_1\sqrt{\delta T} \implies \langle \psi, p \rangle = 0\,, \tag{5.13}$$

$$\sum_{|x|=t} |\psi(x)| \leq \frac{170\exp(-c_2 t\sqrt{\delta}/\sqrt{T})}{\delta \cdot t^2} \qquad \forall t = 1,\dots,T\,. \tag{5.14}$$

5.7 Block composition of functions

Definition 5.15. For functions $f : Y^n \to Z$ and $g : X \to Y$, define the *block composition* $f \circ g : X^n \to Z$ by $(f \circ g)(x_1,\dots,x_n) = f(g(x_1),\dots,g(x_n))$, for all $x_1,\dots,x_n \in X$.

5.8 Dual block composition

We now define the dual block method [31,42,47] for combining dual witnesses for functions f, g in order to obtain a dual witness for $f \circ g$. We then state two basic properties of the method.

Definition 5.16. Let $\Psi : \{-1,1\}^M \to \mathbb{R}$ and $\psi : \{-1,1\}^m \to \mathbb{R}$ be functions that are not identically zero. Let $x = (x_1,\dots,x_M) \in (\{-1,1\}^m)^M$. Define the *dual block composition* of Ψ and ψ, denoted $\Psi \star \psi : (\{-1,1\}^m)^M \to \mathbb{R}$, by

$$(\Psi \star \psi)(x_1,\dots,x_M) = 2^M \cdot \Psi(\dots,\operatorname{sgn}(\psi(x_i)),\dots) \cdot \prod_{i=1}^{M} |\psi(x_i)|\,.$$

Proposition 5.17 ([18,42])**.** *The dual block composition has the following properties:*

***Preservation of ℓ_1-norm*:** *Assume that ψ has pure high degree at least 1. If $\|\Psi\|_1 = 1$ and $\|\psi\|_1 = 1$, then*

$$\|\Psi \star \psi\|_1 = 1\,. \tag{5.15}$$

***Multiplicativity of pure high degree*:** *If $\langle \Psi, P \rangle = 0$ for every polynomial $P\colon \{-1,1\}^M \to \{-1,1\}$ of degree less than D, and $\langle \psi, p \rangle = 0$ for every polynomial $p\colon \{-1,1\}^m \to \{-1,1\}$ of degree less than d, then for every polynomial $q\colon \{-1,1\}^{m\cdot M} \to \{-1,1\}$,*

$$\deg q < D \cdot d \implies \langle \Psi \star \psi, q \rangle = 0\,. \tag{5.16}$$

***Associativity*:** *For every $\zeta : \{-1,1\}^{m_\zeta} \to \mathbb{R}$, $\varphi : \{-1,1\}^{m_\varphi} \to \mathbb{R}$, and $\psi : \{-1,1\}^{m_\psi} \to \mathbb{R}$, we have*

$$(\zeta \star \varphi) \star \psi = \zeta \star (\varphi \star \psi)\,. \tag{5.17}$$

Given three dual witnesses ζ, φ, and ψ, the associativity property allows us to write $\zeta \star \varphi \star \psi$ without ambiguity.

Next, we state an important lemma that follows directly from an analysis of Bun, Kothari, and Thaler [12, Proof of Proposition 31]. This lemma shows that if ψ is the dual witness for OR_N from Proposition 5.14, then for *any* $\Phi\colon \{-1,1\}^R \to \{-1,1\}$, $\xi = \Phi \star \psi$ places an exponentially small amount of mass on inputs outside of $\mathcal{H}^{\leq N}_{N\cdot R}$.[7]

Lemma 5.18 ([12]). *Fix a parameter* $1 \leq \alpha \leq R^2$*, and assume that* $N \geq \lceil 20\sqrt{\alpha}\rceil R$*. For any* $\beta \in (0,1)$ *assume that* $\Phi : \{-1,1\}^R \to \mathbb{R}$ *satisfies* $\|\Phi\|_1 = 1$*. Furthermore, let* $\psi : \{-1,1\}^N \to \mathbb{R}$ *be symmetric (meaning* $\psi(x)$ *depends only on* $|x|$*) and assume that* ψ *satisfies the following conditions:*

$$\sum_{t=0}^{N} \sum_{|x|=t} \psi(x) = 0\,, \tag{5.18}$$

$$\sum_{t=0}^{N} \sum_{|x|=t} |\psi(x)| = 1\,, \tag{5.19}$$

$$\sum_{|x|=t} |\psi(x)| \leq \alpha \exp(-\beta t)/t^2 \quad \forall t = 1,2,\ldots,N\,. \tag{5.20}$$

Then for sufficiently large R,

$$\sum_{x \notin \mathcal{H}^{\leq N}_{N\cdot R}} |(\Phi \star \psi)(x)| \leq \frac{2}{\beta}\exp\left(-\frac{\beta N}{6\ln R}\right)\,. \tag{5.21}$$

In particular, if ψ *is the dual witness for* OR_N *obtained from Proposition 5.14 with constant parameter* $\delta \in (0,1)$ *and with* $T = N/\log^3(N)$*, then there is a constant* $c_3 > 0$ *such that*

$$\sum_{x \notin \mathcal{H}^{\leq N}_{N\cdot R}} |(\Phi \star \psi)(x)| \leq 2^{-c_3 \cdot \sqrt{N \log N}}\,. \tag{5.22}$$

5.9 Hard functions and Hamming weight promises

In our earlier paper [18], we identified a natural class of functions satisfying the following condition: for any function $h\colon \{-1,1\}^n \to \{-1,1\}$ in the class, there is a related function $H\colon \{-1,1\}^{\mathrm{poly}(n)} \to \{-1,1\}$ such that $\widetilde{\deg}_\varepsilon(h) = \tilde{\Theta}(\widetilde{\deg}_\varepsilon(H^{\leq N}))$ for some $N = \tilde{\Theta}(n)$. This enables one to prove ε-approximate degree (respectively, threshold degree) lower bounds on h by lower bounding the ε-approximate degree (threshold degree) of the related function H, *which is defined on significantly more variables than h itself*, under the promise that the Hamming weight of H's input is at most N.

In more detail, this connection is as follows. Fix $R,N \in \mathbb{N}$, let $f : \{-1,1\}^R \to \{-1,1\}$ and let $g : \{-1,1\}^N \to \{-1,1\}$. Suppose g is a *symmetric* function, in the sense that for any $x \in \{-1,1\}^N$ and any permutation $\sigma : [N] \to [N]$, we have

$$g(x_1,\ldots,x_N) = g(x_{\sigma(1)},\ldots,x_{\sigma(N)})\,.$$

[7] Our Lemma 5.18 follows directly from an intermediate calculation in the proof of Proposition 31 of [12].

Equivalently, the value of g on any input x depends only on its Hamming weight $|x|$.

The functions f and g give rise to two functions. The first, denoted by $F^{\mathsf{prop}} : [R]_0^N \to \{-1,1\}$, is a certain property of a list of numbers $s_1,\dots,s_N \in [R]_0$. The second, which we denote by $F^{\le N} : \mathcal{H}_{N\cdot R}^{\le N} \to \{-1,1\}$, is the block composition of f and g restricted to inputs of Hamming weight at most N. Formally, these functions are defined as:

$$F^{\mathsf{prop}}(s_1,\dots,s_N) = f(g(\mathbb{1}[s_1=1],\dots,\mathbb{1}[s_N=1]),\dots,g(\mathbb{1}[s_1=R],\dots,\mathbb{1}[s_N=R]))$$
$$F^{\le N}(x_1,\dots,x_R) = \begin{cases} f(g(x_1),\dots,g(x_R)) & \text{if } x_1,\dots,x_R \in \{-1,1\}^N, |x_1|+\cdots+|x_R| \le N \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Observe that for $f = \mathsf{AND}_R$ and $g = \mathsf{OR}_N$, $F^{\mathsf{prop}} = \mathsf{dSURJ}_{R,N}$. That is, $\mathsf{dSURJ}_{R,N}$ can be expressed as an AND_R (over all range items $r \in [R]$) of the OR_N (over all inputs $i \in [N]$) of "Is input x_i equal to r?".

The following proposition relates the approximate degrees of the two functions F^{prop} and $F^{\le N}$.

Theorem 5.19 (Bun and Thaler [18])**.** *Let $f : \{-1,1\}^R \to \{-1,1\}$ be any function and let $g : \{-1,1\}^N \to \{-1,1\}$ be a symmetric function. Then for F^{prop} and $F^{\le N}$ defined above, we have*

$$\widetilde{\deg}_\varepsilon(F^{\mathsf{prop}}) \ge \widetilde{\deg}_\varepsilon(F^{\le N}).$$

In addition,

$$\deg_\pm(F^{\mathsf{prop}}) \ge \deg_\pm(F^{\le N}).$$

6 The threshold degree of Surjectivity: Upper bound

For completeness, we prove a tight upper bound on the threshold degree of $\mathsf{SURJ}_{R,N}$. This upper bound follows from standard techniques.

Claim 6.1. *The function $\mathsf{SURJ}_{R,N}$ has threshold degree $O\left(\min\left\{R\cdot\log R, N^{1/2}\cdot\log^{3/2} R\right\}\right)$.*

Proof. We assume throughout that $R \le N$, the claim being trivial otherwise. This claim follows from two well-known facts, stated as Expressions (6.1) and (6.2) below.

$$\deg_\pm(\mathsf{SURJ}_{R,N}) \le \deg_\pm(\mathsf{AND}_R \circ \mathsf{OR}_N)\cdot\log R \tag{6.1}$$

$$\deg_\pm(\mathsf{AND}_R \circ \mathsf{OR}_N) = O\left(\min\left\{R, (N\log R)^{1/2}\right\}\right) \tag{6.2}$$

Expression (6.1) holds because, as mentioned in Section 1.1.1, $\mathsf{SURJ}_{R,N}$ 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?", and the quoted question is computed by a conjunction of width $\log R$ over the input bits. That is, if

$$y_{i,j} := \begin{cases} -1 \text{ if } s_i = j \\ 1 \text{ otherwise,} \end{cases}$$

then

$$\mathsf{SURJ}_{R,N}(x) = \mathsf{AND}_R(\mathsf{OR}_N(y_{1,1},\ldots,y_{N,1}),\ldots,\mathsf{OR}(y_{1,R},\ldots,y_{N,R})).$$

Hence, if p sign-represents $\mathsf{AND}_R \circ \mathsf{OR}_N$, then $p(y_{1,1},\ldots,y_{N,R})$ is a polynomial of degree $\deg(p)\cdot\log R$ that sign-represents $\mathsf{SURJ}_{R,N}$.

Equation (6.2) holds by the following standard analysis. To show that

$$\deg_\pm(\mathsf{AND}_R \circ \mathsf{OR}_N) \le O((N\log N)^{1/2}),$$

we exploit the well-known fact that $\widetilde{\deg}_{1/(3R)}(\mathsf{OR}_N) = \Theta(\sqrt{N\log R})$ [10,25]. Let p be a minimal degree polynomial that uniformly approximates OR to error $1/(3R)$. For inputs $x=(x_1,\ldots,x_R)\in\left(\{-1,1\}^N\right)^R$, the polynomial $\sum_{j=1}^R p(x_j)+R-1$ sign-represents $\mathsf{AND}_R \circ \mathsf{OR}_N$, and has degree at most $\deg(p) = O(\sqrt{N\log R})$.

To show that $\deg_\pm(\mathsf{AND}_R \circ \mathsf{OR}_N)\le O(R)$, we use the fact that there exist two linear functions p and q such that the ratio

$$\frac{p(x)}{q(x)} = \frac{-\left(N-\sum_{i=1}^N x_i\right)+1/(3R)}{\left(N-\sum_{i=1}^N x_i\right)+1/(3R)}$$

satisfies $|p(x)/q(x)-\mathsf{OR}_N(x)|\le 1/(3R)$ for every x. Then the following sum of rational functions agrees in sign with $\mathsf{AND}_R \circ \mathsf{OR}_N$ at all inputs $x=(x_1,\ldots,x_R)\in\left(\{-1,1\}^R\right)^N$: $\sum_{j=1}^R p(x_j)/q(x_j)+R-1 = \sum_{j=1}^R p(x_j)q(x_j)/q^2(x_j)+R-1$. Placing everything over the non-negative common denominator $\prod_{j=1}^R q^2(x_j)$, it follows that the polynomial

$$(R-1)\cdot\prod_{j=1}^R q^2(x_j) + \sum_{j=1}^R p(x_j)q(x_j)\prod_{k=1,\ldots,R:\ k\neq j} q(x_j)^2$$

sign-represents $\mathsf{AND}_R \circ \mathsf{OR}_N$. The degree of this polynomial is $O(R)$. □

7 The threshold degree of Surjectivity: Lower bound

The key technical result established in this section is Theorem 3.1, restated here for the reader's convenience.

Theorem 3.1. *Let $R=N^{1/2}$. Then $\deg_\pm\left((\mathsf{AND}_R\circ\mathsf{OR}_N)^{\le N}\right)=\Omega(N^{1/2}/\log^{3/2}N)$.*

By combining Theorems 5.19, 5.9, and 3.1, we obtain the desired lower bound on the threshold degree of SURJ.

Corollary 7.1. *For $R=N^{1/2}$, $\deg_\pm(\mathsf{SURJ}_{R,N})=\Omega(N^{1/2}/\log^{5/2}N)$.*

In fact, combining Corollary 7.1 with Proposition 5.10 determines the threshold degree of $\mathsf{SURJ}_{R,N}$ up to logarithmic factors, for all settings of R and N.

Corollary 7.2 (Restatement of Theorem 1.1)**.** *For any constant $c<1$, if $1<R<cN$, then*

$$\deg_\pm(\mathsf{SURJ}_{R,N}) = \tilde{\Omega}\left(\min(R,N^{1/2})\right).$$

Proof. If $N^{1/2} \le R \le cN$, let $N' = \Theta(N)$ be such that $N = N' + R - (N')^{1/2}$. Then Proposition 5.10 implies that $\deg_{\pm}(\mathsf{SURJ}_{R,N}) \ge \deg_{\pm}\left(\mathsf{SURJ}_{(N')^{1/2},N'}\right) = \tilde{\Omega}(N^{1/2})$, where the final equality holds by Corollary 7.1.

If $R < N^{1/2}$, then Proposition 5.11 implies that $\deg_{\pm}(\mathsf{SURJ}_{R,N}) \ge \deg_{\pm}(\mathsf{SURJ}_{R-1,(R-1)^2}) = \tilde{\Omega}(R)$, where the final equality holds by Corollary 7.1. $\square$

Proof of Theorem 3.1. Let $d = N^{1/2}/(c \log^{3/2} N)$ for a sufficiently large constant c to be chosen later. Let ψ be the dual witness for OR_N from Proposition 5.14 with $T = N/\log^3 N$ and $\delta = 1/4$. Then ψ has pure high degree at least d. Fix any positive integer r, and define $\Phi\colon \{-1,1\}^r \to \mathbb{R}$ as follows:

$$\Phi_r(w) = \begin{cases} -1/2 & \text{if } w = (-1,-1,\dots,-1) \\ 1/2 & \text{if } w = (1,1,\dots,1) \\ 0 & \text{otherwise.} \end{cases}$$

Observe that Φ_r has pure high degree 1, and also $\|\Phi_r\|_1 = 1$. We remark that the dual block composition $\Phi_r \star \psi$ is essentially the same dual witness constructed in [14] to show that $\widetilde{\deg}_{\varepsilon}(\mathsf{AND}_r \circ \mathsf{OR}_N) = \Omega(N^{1/2})$ for $\varepsilon = 1 - 2^{-r}$ (see Inequality (7.3) below). This dual witness was also used in subsequent work [9, 12, 43].

We observe that $\Phi_r \star \psi$ has the following properties.

$$\|\Phi_r \star \psi\|_1 = 1\,. \tag{7.1}$$

$$\Phi_r \star \psi \text{ has pure high degree at least } d\,. \tag{7.2}$$

$$E(\Phi_r \star \psi, \mathsf{AND}_r \circ \mathsf{OR}_N) \le 2^{-r}\,. \tag{7.3}$$

$$(\Phi_r \star \psi)(\mathbf{1}_{r\cdot N}) \ge \frac{1}{2} \cdot (3/4)^r\,. \tag{7.4}$$

$$(\Phi_r \star \psi)(y) \ge 0 \text{ for all } y \in \mathcal{H}^{\le r}_{r\cdot N}\,. \tag{7.5}$$

$$\text{There is some constant } c_3 > 0 \text{ such that } W(\Phi_r \star \psi, (\mathsf{AND}_r \circ \mathsf{OR}_N)^{\le N}) \le 2^{-c_3\sqrt{N \log N}}\,. \tag{7.6}$$

Justification for Properties (7.1)-(7.6). Equation (7.1) follows from Proposition 5.17 (see Equation (5.15)) and the fact that $\|\Phi_r\|_1 = \|\psi\|_1 = 1$. Equation (7.2) follows from Proposition 5.17 (see Equation (5.16)), and the fact that the pure high degree of Φ_r is at least 1, and the pure high degree of ψ is at least d. Inequality (7.3) is an immediate consequence of [14, Theorem 1]. Inequality (7.4) is immediate from the definition of dual block composition and the fact that $\psi(\mathbf{1}_N) \ge 3/8$ (see Inequality (5.11)).

Property (7.5) holds because $(\Phi_r \star \psi)(x) < 0$ only if $\psi(x_i) < 0$ for all $i = 1,\dots,r$, and $\psi(x_i) < 0$ implies that $|x_i| \ge 1$ (see Inequality (5.11)). Hence, $(\Phi_r \star \psi)(x) < 0 \Longrightarrow |x| \ge r$.

Finally, Property (7.6) is immediate from Lemma 5.18 (see Inequality (5.22)) by choosing c_3 sufficiently large.

Definitions of auxiliary functions v_x. Fix any $x \in \mathcal{H}^{\le d}_{R\cdot N}$. We think of x as consisting of R blocks, each of length N. Accordingly, denote $x = (x_1,\dots,x_R)$, where each $x_i \in \{-1,1\}^N$.

Let $\mathcal{I} = \{i \colon x_i = \mathbf{1}_N\}$, and let $r(x) = |\mathcal{I}|$. Enumerate the set $\mathcal{I}$ as $\mathcal{I} = \{i_1, i_2, \dots, i_{r(x)}\}$. Since $|x| \le d$, $r(x) \ge R - d$. Define the function $\nu_x \colon \left(\{-1,1\}^N\right)^R \to \mathbb{R}$ by

$$\nu_x(y_1, \dots, y_R) = \begin{cases} (\Phi_{r(x)} \star \psi)(y_{i_1}, \dots, y_{i_{r(x)}}) & \text{if } y_i = x_i \text{ for all } i \notin \mathcal{I} \\ 0 & \text{otherwise.} \end{cases}$$

We observe that ν_x has the following properties.

$$\|\nu_x\|_1 = 1\,. \tag{7.7}$$

$$\nu_x \text{ has pure high degree at least } d\,. \tag{7.8}$$

$$E\left(\nu_x, \mathsf{AND}_R \circ \mathsf{OR}_N\right) = E\left(\Psi_{r(x)} \star \psi, \mathsf{AND}_{r(x)} \circ \mathsf{OR}_N\right) \le 2^{-r(x)}\,. \tag{7.9}$$

$$\nu_x(x) \cdot \left(\mathsf{AND}_R \circ \mathsf{OR}_N\right)(x) = \left(\Phi_{r(x)} \star \psi\right)\left(\mathbf{1}_{R\cdot N}\right) \ge 1/2 \cdot (3/4)^{r(x)} \ge 1/2 \cdot (3/4)^R\,. \tag{7.10}$$

$$\nu_x(y) \ge 0 \text{ for all } y \in \mathcal{H}^{\le d}_{RN}\,. \tag{7.11}$$

$$W\left(\nu_x, \left(\mathsf{AND}_R \circ \mathsf{OR}_N\right)^{\le N+d}\right) \le 2^{-c_3\sqrt{N \log N}}\,. \tag{7.12}$$

Justification for Properties (7.7)-(7.12). Equations (7.7) and (7.8) are immediate from Equations (7.1) and (7.2) respectively.

We now derive Expression (7.9). The first equality holds by the following reasoning. Since $x_i \ne \mathbf{1}_N$ for all $i \notin \mathcal{I}$, $z_i = x_i$ for all $i \notin \mathcal{I} \Longrightarrow \mathsf{AND}_R \circ \mathsf{OR}_N(z) = \left(\mathsf{AND}_{r(x)} \circ \mathsf{OR}_N\right)(z_{i_1}, \dots, z_{i_{r(x)}})$. The inequality in Expression (7.9) follows from Inequality (7.3).

Expression (7.10) is immediate from the definition of dual block composition and the fact that $\operatorname{sgn}(\psi(\mathbf{1}_N)) > 0$ (see Expression (5.11)). Property (7.11) is immediate from Property (7.5) and the fact that $2d \le R$. Inequality (7.12) is immediate from Inequality (7.6), the definition of ν_x, and the fact that $|x| \le d$.

Another intermediate dual witness. Let $T = |\mathcal{H}^{\le d}_{RN}| = \sum_{i=0}^{d} \binom{RN}{i}$. Consider the dual witness $\eta = \frac{1}{T} \cdot \sum_{x \in \mathcal{H}^{\le d}_{RN}} \nu_x$. Let $r_{\min} = \min_{x \in \mathcal{H}^{\le d}_{RN}} r(x) \ge R - d \ge R/2$.

We observe that η has the following properties.

$$\|\eta\|_1 \le 1\,. \tag{7.13}$$

$$\|\eta\|_1 \ge 1 - 2^{1 - r_{\min}}\,. \tag{7.14}$$

$$\eta \text{ has pure high degree at least } d\,. \tag{7.15}$$

$$E(\eta, \mathsf{AND}_R \circ \mathsf{OR}_N) \le 2^{-r_{\min}}\,. \tag{7.16}$$

$$\text{For all } y \in \mathcal{H}^{\le d}_{RN}, \eta(y) \ge (1/T) \cdot 1/2 \cdot (3/4)^R\,. \tag{7.17}$$

$$W(\eta, (\mathsf{AND}_R \circ \mathsf{OR}_N)^{\le N+d}) \le 2^{-c_3\sqrt{N \log N}}\,. \tag{7.18}$$

Justification for Properties (7.13)-(7.18). To justify Inequality (7.13), observe that

$$\sum_{y\in\{-1,1\}^{RN}}|\eta(y)| \le \sum_{y\in\{-1,1\}^{RN}}\frac{1}{T}\sum_{x\in\mathcal{H}^{\le d}_{RN}}|\nu_x(y)| = \frac{1}{T}\sum_{x\in\mathcal{H}^{\le d}_{RN}}\sum_{y\in\{-1,1\}^{RN}}|\nu_x(y)|$$
$$= \frac{1}{T}\sum_{x\in\mathcal{H}^{\le d}_{RN}}\|\nu_x\|_1 = 1\,,$$

where the final equality is an immediate consequence of Equation (7.7).

To justify Inequality (7.14), observe that

$$\sum_{y\in\{-1,1\}^{RN}}|\eta(y)| = \frac{1}{T}\sum_{y\in\{-1,1\}^{RN}}\left|\sum_{x\in\mathcal{H}^{\le d}_{RN}}\nu_x(y)\right|$$
$$\ge \frac{1}{T}\left(\left(\sum_{y\in\{-1,1\}^{RN}}\sum_{\substack{x\in\mathcal{H}^{\le d}:\\ y\notin\mathcal{E}(\nu_x,\mathsf{AND}_R\circ\mathsf{OR}_N)}}|\nu_x(y)|\right)\right)-\left(\left(\sum_{y\in\{-1,1\}^{RN}}\sum_{\substack{x\in\mathcal{H}^{\le d}:\\ y\in\mathcal{E}(\nu_x,\mathsf{AND}_R\circ\mathsf{OR}_N)}}|\nu_x(y)|\right)\right)$$
$$= \frac{1}{T}\left(\left(\sum_{x\in\mathcal{H}^{\le d}_{RN}}\sum_{\substack{y\in\{-1,1\}^{RN}:\\ y\notin\mathcal{E}(\nu_x,\mathsf{AND}_R\circ\mathsf{OR}_N)}}|\nu_x(y)|\right)\right)-\left(\left(\sum_{x\in\mathcal{H}^{\le d}_{RN}}\sum_{\substack{y\in\{-1,1\}^{RN}:\\ y\in\mathcal{E}(\nu_x,\mathsf{AND}_R\circ\mathsf{OR}_N)}}|\nu_x(y)|\right)\right)$$
$$= \frac{1}{T}\left(\left(\sum_{x\in\mathcal{H}^{\le d}_{RN}}(1-E(\nu_x,\mathsf{AND}_R\circ\mathsf{OR}_N))\right)-\left(\sum_{x\in\mathcal{H}^{\le d}_{RN}}E(\nu_x,\mathsf{AND}_R\circ\mathsf{OR}_N)\right)\right)$$
$$= \frac{1}{T}\sum_{x\in\mathcal{H}^{\le d}_{RN}}1-2\cdot E(\nu_x,\mathsf{AND}_R\circ\mathsf{OR}_N) \ge 1-2^{1-r_{\min}}\,,$$

where the final statement holds by Expression (7.9).

Equation (7.15) is immediate from Equation (7.8) and Fact 5.6.

Expressions (7.16) and (7.18) are respectively immediate from Expressions (7.9) and (7.12) and the triangle inequality.

Expression (7.17) follows from Expressions (7.10) and (7.11).

Let $\zeta := \eta/\|\eta\|_1$. By the definition of ζ, and Equations (7.13)-(7.18), ζ has the following properties.

$$\|\zeta\| = 1\,. \tag{7.19}$$

$$\zeta \text{ has pure high degree at least } d\,. \tag{7.20}$$

$$E(\zeta,\mathsf{AND}_R\circ\mathsf{OR}_N) \le 2^{-r_{\min}}/\left(1-2^{-r_{\min}}\right) \le 2^{1-r_{\min}}\,. \tag{7.21}$$

$$W\left(\zeta,(\mathsf{AND}_R\circ\mathsf{OR}_N)^{\le N+d}\right) \le 2^{-c_3\sqrt{N\log N}}/(1-2^{-r_{\min}}) \le 2^{1-c_3\sqrt{N\log N}}\,. \tag{7.22}$$

$$\text{For all } x\in\mathcal{H}^{\le d}_{RN}, \zeta(x) \ge (1/T)\cdot 1/2\cdot(3/4)^R\,. \tag{7.23}$$

Let $s = (1/T) \cdot 1/2 \cdot (3/4)^R$ denote the right hand side of Expression (7.23). Recalling that $T = \sum_{i=0}^{d} \binom{RN}{i} \leq (1+RN)^d$, and that $d = N^{1/2}/(c \log^{3/2} N)$ for a constant c of our choosing, taking c sufficiently large ensures that

$$s \geq 2^{-R/5}. \tag{7.24}$$

The final dual witness. Let $f = (\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N+d}$. We now modify ζ using Lemma 5.12 to zero out the mass on the "bad sets" $\mathcal{E}(\zeta, \mathsf{AND}_R \circ \mathsf{OR}_N)$ and $\mathcal{W}(\zeta, f)$, thereby constructing a threshold-degree dual witness μ for f. Specifically, let $B = \mathcal{W}(\zeta, f) \cup \mathcal{E}(\zeta, \mathsf{AND}_R \circ \mathsf{OR}_N)$. By Expression (7.23), and the fact that $f(x) = 1$ for all $x \in \mathcal{H}_{RN}^{\leq d}$, it holds that

$$B \subseteq \{-1,1\}^{RN} \setminus \mathcal{H}_{RN}^{\leq d}. \tag{7.25}$$

Hence, for every $x \in B$, we can invoke Lemma 5.12 with $D = d$ to obtain a function ϕ_x having Properties (5.1)-(5.4). For every $x \in B$, define $\psi_{\mathsf{corr},x} := \zeta(x) \cdot \phi_x$. Our final dual witness μ is

$$\mu := \zeta - \sum_{x \in B} \psi_{\mathsf{corr},x}. \tag{7.26}$$

Analysis of μ: A brief overview. Since the correction terms in the definition of μ are specifically designed to have pure high degree d and zero out all of the "bad" mass of ζ, all that remains in order to show that μ witnesses a degree d lower bound on the threshold degree of f is show that the correction terms do not disturb the sign of any inputs in their support. To accomplish this, we start by observing that the total ℓ_1-mass of all the correction objects is at most $(W(\zeta, f) + E(\zeta, \mathsf{AND}_R \circ \mathsf{OR}_N)) \cdot 2^d \cdot \binom{RN}{d}$. Recalling that $R = N^{1/2}$ and $d = N^{1/2}/(c \log^{3/2} N)$ for a constant c of our choosing, by Properties (7.21) and (7.22) we can set c sufficiently large to ensure that both $R - d - 1 > 4R/5$ and $c_3\sqrt{N \log N} - 1 > 4R/5$. Then

$$(W(\zeta, f) + E(\zeta, \mathsf{AND}_R \circ \mathsf{OR}_N)) \cdot 2^d \cdot \binom{RN}{d} \leq 2^{-4R/5} \cdot (RN)^d \leq 2^{-3R/5} \leq s/2, \tag{7.27}$$

where the final inequality holds by Property (7.24). Hence, for each $x \in \mathcal{H}_{RN}^{\leq d}$, $\mu(x) \cdot f(x) \geq \zeta(x) \cdot f(x) - s/2 > s/2$, where the final inequality holds by Property (7.23). That is, the correction terms do not disturb the sign of any points in their support.

Analysis of μ: Details. We need to prove that μ satisfies the conditions of Theorem 5.3 with $f = (\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N+d}$.

- **(Pure high degree).** That μ has pure high degree at least d follows from Fact 5.6 and the fact that ζ and each $\psi_{\mathsf{corr},x}$ have pure high degree at least d (Equations (7.20) and (5.3)).
- **(Sign-agreement).** To establish sign-agreement, we consider three cases:
  - $y \in B$. In this case $\mu(y) = \zeta(y) - \psi_{\mathsf{corr},y}(y) = \zeta(y) - \zeta(y) = 0$. Here, the first equality holds because $B \subseteq \mathcal{H}_{RN}^{>d}$ (see Equation (7.25)), and for each $x \in B \setminus \{y\}$, the point y is not in the support of $\psi_{\mathsf{corr},x}$ (as Equation (5.2) states that the support of $\psi_{\mathsf{corr},x}$ is a subset of $\mathcal{H}_{RN}^{\leq d} \cup \{x\}$).

- $y \notin (B \cup \mathcal{H}_{RN}^{\leq d})$. In this case, $\mu(y) \cdot f(y) = \zeta(y) \cdot f(y) \geq 0$. The first equality holds because $y \notin B$, and since $y \notin \mathcal{H}_{RN}^{\leq d}$, y is not in the support of any $\psi_{\mathsf{corr},x}$ for any $x \in B$ (see Equation (5.2)).
- $y \in \mathcal{H}_{RN}^{\leq d}$. In this case,

$$\mu(y) = \zeta(y) - \sum_{x \in B} \psi_{\mathsf{corr},x}(y) \geq \zeta(y) - \left| \sum_{x \in B} \psi_{\mathsf{corr},x}(y) \right| \geq \zeta(y) - \sum_{x \in B} \|\psi_{\mathsf{corr},x}\|_1$$
$$\geq \zeta(y) - (E(\zeta, \mathsf{AND}_R \circ \mathsf{OR}_N) + W(\zeta, f)) \cdot 2^d \cdot \binom{RN}{d} > s - s/2 > 0 \,. \qquad (7.28)$$

  Here, the third to last inequality follows from the definition of $\psi_{\mathsf{corr},x}$ and Expression (5.4), and the penultimate inequality follows by Expressions (7.23) and (7.27).

- **(Appropriate support).** The first case considered in the sign-agreement analysis above also establishes the appropriate support condition, since it shows that $\mu(x) = 0$ for all $x \in B$, and by definition of B, $\mathcal{H}_{RN}^{>N+d} \subseteq B$.
- **(Non-triviality).** The third case considered in the sign-agreement analysis also establishes non-triviality, since $\mu(x) > 0$ for all $x \in \mathcal{H}_{RN}^{\leq d}$.

By Theorem 5.3, μ witnesses the fact that for $R = N^{1/2}$,

$$\deg_{\pm}\left((\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N+d} \right) = \Omega\left(N^{1/2} / \log^{3/2} N \right) .$$

For any $t > 0$,

$$\deg_{\pm}\left((\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq t} \right)$$

is clearly nondecreasing with N. Hence,

$$\deg_{\pm}\left((\mathsf{AND}_R \circ \mathsf{OR}_{N+d})^{\leq N+d} \right) = \Omega\left(N^{1/2} / \log^{3/2} N \right) .$$

Since $d = o(N)$, setting $N' = N + d$ implies that

$$\deg_{\pm}\left((\mathsf{AND}_R \circ \mathsf{OR}_{N'})^{\leq N'} \right) = \Omega\left((N')^{1/2} / \log^{3/2}(N') \right) ,$$

as desired. $\square$

8 The large-error approximate degree of AC^0 is nearly linear

Theorem 8.1. *For any constant $\delta > 0$, there is a function $h \colon \{-1,1\}^n \to \{-1,1\}$ computed by an AC^0 circuit of depth $O(1/\delta)$ such that $\widetilde{\deg}_{\varepsilon}(h) = \tilde{\Omega}(n^{1-\delta})$, for some $\varepsilon = 1 - 2^{-\tilde{\Omega}(n^{1-\delta})}$.*

Theorem 8.1 is a consequence of the following hardness amplification result, which shows how to transform any function f that is hard to approximate by low-degree polynomials into a function F that is even harder to approximate, in terms of *both* the degree and the error that is achievable by polynomials of said degree.

Theorem 8.2. *Fix constants* $k \geq 1, c \geq 0$. *Let* $f_n \colon \{-1,1\}^{n \log^c n} \to \{-1,1\}$ *be any (infinite family of) functions satisfying* $\widetilde{\deg}_\varepsilon(f_n) \geq \Omega\left(n^{k/(k+1)}\right)$ *for some* $\varepsilon = 1 - 2^{-\Omega(n^{k/(k+1)})}$ *and sufficiently large n. Then there is an explicitly given sequence of functions* $F_n \colon \{-1,1\}^{n \log^{c+O(1)} n} \to \{-1,1\}$ *such that* $\widetilde{\deg}_\varepsilon(F_n) \geq \Omega\left(n^{(k+1)/(k+2)}\right)$ *for some* $\varepsilon = 1 - 2^{-\Omega\left(n^{(k+1)/(k+2)}\right)}$ *and sufficiently large n. Moreover, if* f_n *is computed by polynomial-size circuits of depth* Δ, *then* F_n *is computed by polynomial-size circuits of depth at most* $\Delta + O(1)$.

Proof of Theorem 8.1 assuming Theorem 8.2. $\mathsf{SURJ}_{N^{1/2},N}$ is a function on $n = \Theta(N \log N)$ input bits, computed by a polynomial-size circuit of depth 3, and Corollary 7.1 implies that $\deg_\pm(\mathsf{SURJ}_{N^{1/2},N}) = \tilde{\Omega}(N^{1/2})$. Applying Theorem 8.2 to $f = \mathsf{SURJ}_{N^{1/2},N}$ with $k = 1$ yields a function F_1 on $n \cdot \text{polylog}(n)$ inputs that is computed by a polynomial-size circuit of depth $3 + O(1) = O(1)$, and satisfies $\widetilde{\deg}_\varepsilon(F_1) \geq \Omega(n^{2/3})$ for some $\varepsilon = 1 - 2^{-\Omega(n^{2/3})}$. Applying Theorem 8.2 yet again, with $f = F_1$ and $k = 2$ yields a function F_2 in AC^0 that is also defined on $n \cdot \text{polylog}(n)$ inputs, is computed by a polynomial-size circuit of depth $O(1)$, and satisfies $\widetilde{\deg}_\varepsilon(F_2) \geq \Omega(n^{3/4})$ for some $\varepsilon = 1 - 2^{-\Omega(n^{3/4})}$.

In general, for any constant $\delta \in (0,1)$, let k be a constant such that $1 - \delta \leq k/(k+1)$, i. e., $k \geq 1/\delta - 1$. Then iteratively applying Theorem 8.2 $k-1$ times, starting with $f = \mathsf{SURJ}_{N^{1/2},N}$, yields a function h computed by an AC^0 circuit of depth $O(k) = O(1/\delta)$, defined on $n' = n \log^{O(k)} n = \tilde{O}(n)$ input bits, such that $\widetilde{\deg}_\varepsilon(h) = \Omega(n^{1-\delta})$ for some $\varepsilon = 1 - 2^{-\Omega(n^{1-\delta})}$. Theorem 8.1 follows. □

Proof of Theorem 8.2. Theorem 5.19 implies that, in order to prove Theorem 8.2, it is sufficient to identify a block-composed function G defined on $\text{poly}(n)$ inputs that is computed by a polynomial-size circuit of depth $\Delta + O(1)$ such that for some $\ell = n \cdot \text{polylog}(n)$, we have $\deg_\varepsilon(G^{\leq \ell}) \geq \Omega(n^{(k+1)/(k+2)})$ for some $\varepsilon = 1 - 2^{-\Omega\left(n^{(k+1)/(k+2)}\right)}$. Indeed, if we accomplish this, then Theorem 8.2 follows by setting $F = G^{\mathsf{prop}}$.

To define G, we need the following lemma, which follows from the techniques of [2] (see [28] for an exposition).

Lemma 8.3. *There exists a Boolean circuit* $C_n \colon \{-1,1\}^n \to \{-1,1\}$ *with n inputs, depth 3, and size* $\tilde{O}(n^2)$ *satisfying the following two conditions:*

- $C_n(x) = 1$ *for all x of Hamming weight at most* $n/3$.
- $C_n(x) = -1$ *for all x of Hamming weight at least* $2n/3$.

We refer to the function computed by the circuit C of Lemma 8.3 as GAPMAJ, short for a gapped majority function (such a function is sometimes also called an *approximate majority* function).[8] We remark that while the circuit C from Lemma 8.3 is not explicitly constructed, explicit constructions of AC^0 circuits satisfying the two bulleted conditions of Lemma 8.3 are known [49].

Definition of G. Let $t = n^{1/(k+2)}$, $z = n^{(k+1)/(k+2)}$, $r = 10 \log n$, and $m = n^{2/(k+2)}$. Let $N = t \cdot Z \cdot r \cdot M$, where $Z = z \log^c z$, and $M = 10n \log^{c+4} n$. We define $G \colon \{-1,1\}^N \to \{-1,1\}$ to equal $\mathsf{GAPMAJ}_t \circ f_z \circ$

[8] In prior related work [9], GAPMAJ referred to the *promise* function that equals 1 for all inputs x of Hamming weight at most $n/3$, equals -1 for all inputs x of Hamming weight at least $2n/3$, and is undefined otherwise. In contrast, we use GAPMAJ to refer to any total function in AC^0 that agrees with the partial function from [9] at all points in the partial function's domain.

$\mathsf{AND}_r \circ \mathsf{OR}_M$. Here, f_z denotes the function f on $Z = z\log^c z$ variables whose existence is assumed by the hypothesis of the theorem.

We now begin the process of constructing a dual witness μ showing that $\widetilde{\deg}_\varepsilon(G^{\le 10n\log^{c+4} n}) = \Omega\left(n^{(k+1)/(k+2)}\right)$ for some $\varepsilon = 1 - 2^{-\Omega\left(n^{(k+1)/(k+2)}\right)}$.

For any appropriate constant $c > 0$, let $\phi' \colon \{-1,1\}^Z \to \mathbb{R}$ be a dual witness for the fact that $\widetilde{\deg}_\varepsilon(f_z) \ge C \cdot z^{k/(k+1)} = C \cdot n^{k/(k+2)} := d$ (for some constant C) as per Theorem 5.5. Then $\phi := \phi'/\|\phi\|_1$ has the following properties.

$$\|\phi\|_1 = 1\,. \tag{8.1}$$

$$\phi \text{ has pure high degree at least } d\,. \tag{8.2}$$

$$\sum_{x\in\{-1,1\}^Z} \phi(x)\cdot f_z(x) \ge 1-\delta' \text{ for some } \delta' = 2^{-\Omega(d)}\,. \tag{8.3}$$

Similar to the proof of Theorem 3.1, let ψ be the dual witness for OR_M from Proposition 5.14 with $M = 10n\log^{c+4} n$, $T = m$, and $\delta = 1/4$, and let

$$\Phi_r(w) = \begin{cases} -1/2 & \text{if } w = (-1,-1,\dots,-1) \\ 1/2 & \text{if } w = (1,1,\dots,1) \\ 0 & \text{otherwise.} \end{cases}$$

We remind the reader that Equations (7.1)-(7.3) from the proof of Theorem 3.1 guarantee that $\Phi_r \star \psi$ has the following properties.

$$\|\Phi_r \star \psi\|_1 = 1\,. \tag{8.4}$$

$$\Phi_r \star \psi \text{ has pure high degree at least that of } \psi\text{, which is } D' := \Omega\left(m^{1/2}\right) = \Omega\left(n^{1/(k+2)}\right)\,. \tag{8.5}$$

$$E\left(\Phi_r \star \psi, \mathsf{AND}_r \circ \mathsf{OR}_M\right) \le 2^{-r} = 1/n^{10}\,. \tag{8.6}$$

We now combine ϕ and $\Phi_r \star \psi$ to obtain an intermediate dual witness γ, which will witness the fact that $\widetilde{\deg}_{1-2^{-\Omega(d)}}(f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_M) = \Omega(d\cdot\sqrt{m})$. The combining technique that we use is precisely the one introduced by Sherstov [41, Theorem 6.1] to establish a direct-sum theorem for approximate degree. Roughly speaking, [41, Theorem 6.1] showed that for any Boolean functions f and g, if $\widetilde{\deg}_{\delta_1}(f) \ge d$ and $\widetilde{\deg}_{1/n}(g) \ge d'$, then $\widetilde{\deg}_{\delta_2}(f \circ g) = \Omega(d\cdot d')$ for some $\delta_1 \approx \delta_2$. The γ that we construct below is (a normalized version of) the dual witness constructed in the proof of [41, Theorem 6.1] when applied with outer function f_z and inner function $g = \mathsf{AND}_r \circ \mathsf{OR}_M$.

For a fixed *even* integer $j \in (d/2-2, d/2]$, let P_j be the degree j univariate polynomial given by $P_j(a) = \prod_{i=1}^{j}(a-i)$, and define $p_j \colon \{-1,1\}^Z \to \mathbb{R}$ to be the unique multilinear polynomial such that $p_j(x) = P_j(|x|)$. We will need the following properties of p_j.

$$\text{The sum of the absolute values of the Fourier coefficients of } p_j \text{ is at most } j!\binom{Z+j}{j}\,. \tag{8.7}$$

$$\text{For any } c \in [1-1/Z^2, 1]\text{, it holds that } p_j(c,\dots,c) \ge \frac{1}{2}j! \ge 1\,. \tag{8.8}$$

Property (8.7) is precisely Lemma 3.1 (specifically, Property 3.2) of [41]. To see that Property (8.8) holds, first observe that $p_j(x) \geq 0$ for all $x \in \{-1,1\}^Z$, owing to the fact that j is even. Second, by the multilinearity of p_j, the value of $p_j(c,\dots,c)$ is the expected value of p_j under the distribution over $x \in \{-1,1\}^Z$ in which each coordinate i of x is chosen independently from $\{-1,1\}$ such that the expected value of x_i is c. Denoting this distribution as B_c, $\Pr_{x\sim B_c}[x = \mathbf{1}_Z] \geq 1/2$. Hence,

$$p_j(c,\dots,c) \geq \Pr_{x\sim B_c}[x = \mathbf{1}_Z] \cdot p_j(\mathbf{1}_Z) \geq \frac{1}{2} j!\,. \tag{8.9}$$

Let $\delta'' = 2 \cdot E(\mathsf{AND}_r \circ \mathsf{OR}_M, \Phi_r \star \psi) \leq 2 \cdot 2^{-r} = 2/n^{10}$. For $y \in \{-1,1\}^{r\cdot M}$, define[9]:

$$\alpha(y) = \begin{cases} 1 & \text{if } (\mathsf{AND}_r \circ \mathsf{OR}_M)(y) = \operatorname{sgn}((\Phi_r \star \psi)(y)) = +1 \\ 1-2\delta'' & \text{if } (\mathsf{AND}_r \circ \mathsf{OR}_M)(y) = \operatorname{sgn}((\Phi_r \star \psi)(y)) = -1 \\ -1 & \text{otherwise.} \end{cases}$$

For an input x to $f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_M$, write $x = (x_1,\dots,x_Z)$ with each $x_i \in \{-1,1\}^{r\cdot M}$. Define:

$$\gamma'(x_1,\dots,x_Z) = p_j\left(\alpha(x_1),\dots,\alpha(x_Z)\right) \cdot \left(\phi \star \Phi_r \star \psi\right)(x),$$

and let $\gamma = \gamma'/\|\gamma'\|_1$.

We observe that γ' has the following properties.

$$\|\gamma'\|_1 = p_j(1-2\delta'',\dots,1-2\delta'') > 0\,. \tag{8.10}$$

$$\gamma' \text{ has pure high degree at least } (d/2)\cdot D'\,. \tag{8.11}$$

$$\sum_{x\in\{-1,1\}^{Z\cdot r\cdot M}} \gamma'(x)\cdot\left(f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_M\right)(x) \geq p_j\left(\dots,1-2\delta'',\dots\right)\cdot\left((1-2\delta') - \frac{2(\delta'')^{j+1}}{(1-\delta'')^n}\cdot\binom{n}{j+1}\right). \tag{8.12}$$

Property (8.10) is precisely Claim 6.2 from [41]. Equation (8.11) is immediate from [41, Equation 6.7], combined with the fact that $\Phi_r \star \psi$ has pure high degree at least D' (Equation (8.5)) and ϕ has pure high degree at least d (Equation (8.2)). Property (8.12) is precisely Claim 6.3 from [41].

Recall that we defined $\gamma = \gamma'/\|\gamma'\|_1$. Properties (8.10)-(8.12) imply that γ satisfies the following conditions:

$$\|\gamma\|_1 = 1\,. \tag{8.13}$$

$$\gamma \text{ has pure high degree at least } (d/2)\cdot D'\,. \tag{8.14}$$

$$\begin{aligned}\sum_{x\in\{-1,1\}^{Z\cdot r\cdot M}} \gamma(x)\cdot\left(f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_M\right)(x) &\geq (1-2\delta') - \frac{2(\delta'')^{j+1}}{(1-\delta'')^n}\cdot\binom{n}{j+1} \\ &\geq \left(1-2\delta' - 2^{-3d}\right) = 1-2^{-\Omega(d)}\,. \end{aligned} \tag{8.15}$$

[9] Our definition of α specializes the definition in [41, Page 39] to our setting: our definition and the definition in [41, Page 39] coincide owing to the fact that (as established in [14, Theorem 1]) $(\Phi_r \star \psi)(y) < 0 \Longrightarrow (\mathsf{AND}_r \circ \mathsf{OR}_M)(y) < 0$.

In Property (8.15), the penultimate inequality holds because $\delta'' \le 2 \cdot 2^{-r} = 2/n^{10}$, and $j > d/2 - 2$.

The final steps in the construction of μ. Next, define $\zeta \colon \{-1,1\}^{t \cdot Z \cdot r \cdot M} \to \mathbb{R}$ via:

$$\zeta = \Phi_t \star \gamma.$$

For every $x \in \mathcal{W}\left(\zeta, G^{\le 10 n \log^{c+4} n}\right)$, invoke Lemma 5.12 with $D = C' \cdot n^{(k+1)/(k+2)}$ (for a sufficiently small constant C' to be chosen later) to obtain a function ϕ_x satisfying Conditions (5.1)-(5.4), and define $\psi_{\mathsf{corr},x} \coloneqq \zeta(x) \cdot \phi_x$. We define our final dual witness to be $\mu \coloneqq \zeta - \sum_{x \in \mathcal{W}\left(\zeta, G^{\le 10 n \log^{c+4} n}\right)} \psi_{\mathsf{corr},x}$.

Analysis of ζ and μ. We observe that ζ has the following properties.

$$\|\zeta\|_1 = 1. \tag{8.16}$$

$$\zeta \text{ has pure high degree at least that of } \gamma\text{, which is at least } D'' \coloneqq (d/2) \cdot D'. \tag{8.17}$$

$$\sum_{x \in \{-1,1\}^{t \cdot Z \cdot r \cdot M}} \zeta(x) \cdot (\mathsf{GAPMAJ}_t \circ f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_M) \ge 1 - 2^{-\Omega(d \cdot t)}. \tag{8.18}$$

$$W(\zeta, G^{\le 10 n \log^{c+4} n}) \le 2^{-\omega(n^{(k+1)/(k+2)} \log n)}. \tag{8.19}$$

Justification for Properties (8.16)-(8.19). Equation (8.16) follows from Proposition 5.17 (see Equation (5.15)), and the facts that $\|\Phi_t\| = 1$ and $\|\gamma\| = 1$ (see Equation (8.13)), and γ has pure high degree at least 1. Equation (8.17) follows from Proposition 5.17 (see Equation (5.16)), the fact that Φ_t has pure high degree 1, and the fact that γ has pure high degree at least $(d/2) \cdot D'$ (see Equation (8.14)). The validity of Inequality (8.18) is established via a standard, but somewhat lengthy, analysis that we defer to Section 8.1.

We justify Inequality (8.19) as follows. Denote an input x in $\{-1,1\}^{t \cdot Z \cdot r \cdot M}$ as $x = (\ldots, x_{i,s}, \ldots)$ where i ranges over $1, \ldots, t$, s ranges over $1, \ldots, Z$, and each $x_{i,s} \in \{-1,1\}^{r \cdot M}$. Since p_j is non-negative on all inputs in $[-1,1]^Z$ (this follows from the same reasoning as in the proof of Property (8.8)),

$$\begin{aligned}
|\zeta(x)| &= \left| (\Phi_t \star \phi \star \Phi_r \star \psi)(x) \cdot \left(\|\gamma'\|_1\right)^{-t} \cdot \prod_{i=1}^{t} p_j\left(\alpha(x_{i,1}), \ldots, \alpha(x_{i,Z})\right) \right| \\
&= \left| (\Phi_t \star \phi \star \Phi_r \star \psi)(x) \cdot p_j(\ldots, 1 - 2\delta'', \ldots)^{-t} \cdot \prod_{i=1}^{t} p_j(\alpha(x_{i,1}), \ldots, \alpha(x_{i,Z})) \right| \\
&\le \left| (\Phi_t \star \phi \star \Phi_r \star \psi)(x) \cdot \prod_{i=1}^{t} p_j(\alpha(x_{i,1}), \ldots, \alpha(x_{i,Z})) \right| \\
&\le |(\Phi_t \star \phi \star \Phi_r \star \psi)(x)| \left(j! \binom{Z+j}{j} \right)^t \\
&\le |(\Phi_t \star \Phi_r \star \psi)(x)| \cdot Z^{O(jt)} \\
&\le |(\Phi_t \star \phi \star \Phi_r \star \psi)(x)| \cdot 2^{O(dt \log n)}.
\end{aligned} \tag{8.20}$$

Here, the second equality holds by Equation (8.10), the first inequality holds by Property (8.8), and the second inequality follows from Property (8.7) and the fact that $|\alpha(x_{i,s})| \le 1$ for all $i = 1,\dots,t$ and $s = 1,\dots,Z$.

Now invoke Lemma 5.18 (and associativity of dual block composition, see Equation (5.17)) to conclude that

$$W\left((\Phi_t \star \phi \star \Phi_r) \star \psi, (\mathsf{GAPMAJ}_t \circ f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_M)^{\le 10n\log^{c+4} n}\right) \le 2^{-\Omega\left(n\log^2 n/\sqrt{m}\right)} \le 2^{-\Omega(n^{(k+1)/(k+2)}\log^2 n)}. \tag{8.21}$$

Combining Inequalities (8.20) and (8.21), we conclude that

$$\begin{aligned} W\left(\zeta, (\mathsf{GAPMAJ}_t \circ f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_M)^{\le 10n\log^{c+4} n}\right) &\le 2^{-\Omega\left(n^{(k+1)/(k+2)}\log^2 n\right)} \cdot 2^{O(dt\log n)} \\ &\le 2^{-\Omega\left(n^{(k+1)/(k+2)}\log^2 n\right)} \cdot 2^{O\left(n^{(k+1)/(k+2)}\log n\right)} \\ &\le 2^{-\Omega\left(n^{(k+1)/(k+2)}\log^2 n\right)}, \end{aligned}$$

where the final inequality holds for large enough n.

Finally, we observe the following properties of μ.

$$\|\mu\|_1 \le 1 + 2^{-\omega(n^{(k+1)/(k+2)})}. \tag{8.22}$$

$$\mu \text{ has pure high degree at least } \min(D'', D) = \Omega\left(n^{(k+1)/(k+2)}\right). \tag{8.23}$$

$$\sum_{x \in \{-1,1\}^{t\cdot Z\cdot r\cdot M}} \mu(x)\cdot G(x) \ge 1 - 2^{-\Omega(t\cdot d)} = 1 - 2^{-\Omega\left(n^{(k+1)/(k+2)}\right)}. \tag{8.24}$$

$$\mu(x) = 0 \text{ for all } x \in \mathcal{H}_{t\cdot Z\cdot r\cdot M}^{>10n\log^{c+4} n}. \tag{8.25}$$

Justification for Properties (8.22)-(8.25). For Inequality (8.22), observe that

$$\begin{aligned} \|\mu\|_1 &\le \|\zeta\|_1 + \sum_{x \in \mathcal{W}(\zeta, G^{\le 10n\log^{c+4} n})} \|\psi_{\mathsf{corr},x}\|_1 \\ &\le 1 + W(\zeta, G^{\le 10n\log^{c+4} n}) \cdot 2^D \cdot \binom{t\cdot Z\cdot r\cdot M}{D} \\ &\le 1 + 2^{-\omega\left(n^{(k+1)/(k+2)}\log n\right)}, \end{aligned}$$

where the penultimate inequality holds by Expressions (5.4) and (8.16), and the final inequality holds by Inequality (8.19).

Equation (8.23) follows from Fact 5.6, Equation (7.20), and the fact that each term $\psi_{\mathsf{corr},x}$ has pure high degree at least D (Equation (5.3)).

Expression (8.24) holds because

$$\sum_{x\in\{-1,1\}^{t\cdot Z\cdot r\cdot M}} \mu(x)\cdot G(x) \geq \sum_{x\in\{-1,1\}^{t\cdot Z\cdot r\cdot M}} \zeta(x)\cdot G(x) - \sum_{x\in\mathcal{W}\left(\zeta, G^{\leq 10n\log^{c+4} n}\right)} \|\psi_{\mathsf{corr},x}\|_1$$
$$\geq 1-2^{-\Omega(td)} - 2^{-\omega(n^{(k+1)/(k+2)})} \geq 1-2^{-\Omega(td)},$$

where the penultimate inequality invokes Expressions (5.4) and (8.19).

Expression (8.25) holds because for any $x \in \mathcal{H}_{t\cdot Z\cdot r\cdot M}^{>10n\log^{c+4} n}$, we have $\mu(x) = \zeta(x) - \psi_{\mathsf{corr},x}(x) = \zeta(x) - \zeta(x) = 0$. Here, the first equality holds because Property (5.2) ensures that, for any $x' \in \mathcal{W}\left(\zeta, G^{\leq 10n\log^{c+4} n}\right)$ such that $x \neq x'$, we have $\psi_{\mathsf{corr},x'}(x) = 0$.

It is immediate from Properties (8.22)-(8.25) that μ satisfies the conditions required by Theorem 5.5 to witness the fact that $\widetilde{\deg}_\varepsilon(G^{\leq 10n\log^{c+4} n}) = \Omega(n^{(k+1)/(k+2)})$ for some $\varepsilon = 1-2^{-\Omega(n^{(k+1)/(k+2)})}$. $\square$

8.1 Justification for Expression (8.18)

All notation in this section the same as in the proof of Theorem 8.2. Let $F = f_z \circ \mathsf{AND}_r \circ \mathsf{OR}_M$. Recall that $\zeta = \Phi_t \star \gamma$, where γ has the following properties.

$$\|\gamma\|_1 = 1. \tag{8.26}$$

$$\gamma \text{ has pure high degree at least } (d/2)\cdot D' \geq 1. \tag{8.27}$$

$$\sum_{y\in\{-1,1\}^{Z\cdot r\cdot M}} \gamma(y)\cdot F(y) \geq \varepsilon := 1-2^{-\Omega(d)}. \tag{8.28}$$

Equation (8.27) implies that

$$\sum_{y:\ \mathrm{sgn}(\gamma(y))=1} |\gamma(y)| = \sum_{y:\ \mathrm{sgn}(\gamma(y))=-1} |\gamma(y)| = 1/2. \tag{8.29}$$

Let $E_1(\gamma,F) = \sum_{y:\ \gamma(y)>0 \text{ and } F(y)<0} |\gamma(y)|$ and $E_{-1}(\gamma,F) = \sum_{y:\ \gamma(y)<0 \text{ and } F(y)>0} |\gamma(y)|$. Combining Expressions (8.28) and (8.29) implies that $E_1(\gamma,F)$ and $E_{-1}(\gamma,F) \leq 2^{-\Omega(d)}$.

Let τ be the product distribution on $\left(\{-1,1\}^{Z\cdot r\cdot M}\right)^t$ given by $\tau(x_1,\ldots,x_t) = \prod_{i=1}^t |\gamma(x_i)|$. Given a string $b = (b_1,\ldots,b_t) \in \{-1,1\}^t$, let τ_b be the distribution over $x \in \{-1,1\}^{t\cdot Z\cdot r\cdot M}$ equal to τ conditioned on $(\ldots,\mathrm{sgn}(x_i),\ldots,) = b$. Equation (8.29) implies that when $x = (x_1,\ldots,x_t)$ is drawn from τ, the string $(\ldots,\mathrm{sgn}(\gamma(x_i),\ldots)$ is uniformly distributed in $\{-1,1\}^t$.

Moreover, for any given $b \in \{-1,1\}^t$, the following two random variables are identically distributed:

- The string $(\ldots,F(x_i),\ldots)$ when one chooses $(\ldots,x_i,\ldots)$ from the conditional distribution τ_b.
- The string $(\ldots,\delta_i b_i,\ldots)$, where $\delta \in \{-1,1\}^t$ is a random string whose ith bit independently takes on value -1 with probability $2\sum_{x\in E_{b_i}} |\psi(x)| < 2^{-\Omega(d)}$.

Thus,

$$\begin{aligned}
&\sum_{x\in\{-1,1\}^{t\cdot Z\cdot r\cdot M}} \zeta(x)\cdot \mathsf{GAPMAJ}(\dots,F(x_i),\dots) \\
&= 2^t\cdot \sum_{b\in\{-1,1\}^t} \Phi_t(b)\cdot \mathbb{E}_{x\sim\tau_b}[\mathsf{GAPMAJ}(\dots,F(x_i),\dots)] \\
&= 2^t\left(\frac{1}{2}\mathbb{E}_{x\sim\tau_{1_t}}[\mathsf{GAPMAJ}(\dots,F(x_i),\dots)] - \frac{1}{2}\mathbb{E}_{x\sim\tau_{-1_t}}[\mathsf{GAPMAJ}(\dots,F(x_i),\dots)\right) \\
&= \frac{1}{2}\mathbb{E}_{\delta^{(1)}}[\mathsf{GAPMAJ}_t(\delta_1^{(1)},\dots,\delta_t^{(1)})] - \frac{1}{2}\mathbb{E}_{\delta^{(-1)}}[\mathsf{GAPMAJ}_t(-\delta_1^{(-1)},\dots,-\delta_t^{(-1)})],
\end{aligned}$$

where for $j\in\{-1,1\}$, $\delta^{(j)}\in\{-1,1\}^t$ is a random string whose ith coordinate takes value -1 with probability $2\sum_{x\in E_j}|\psi(x)| < 2^{-\Omega(d)}$.

Note that $\mathbb{E}_{\delta^{(1)}}[\mathsf{GAPMAJ}_t(\delta_1^{(1)},\dots,\delta_t^{(1)})] \geq 1-2^{-\Omega(td)}$: since each bit of $\delta^{(1)}$ equals 1 with probability $1-2^{-\Omega(d)}$, the probability that more than $t/3$ of the coordinates of $\delta^{(1)}$ are -1 is at most $\binom{t}{t/3}\cdot 2^{-\Omega(td)} = 2^{-\Omega(td)}$. Similarly, $\mathbb{E}_{\delta^{(-1)}}[\mathsf{GAPMAJ}_t(-\delta_1^{(-1)},\dots,-\delta_t^{(-1)})] \leq -1+2^{-\Omega(td)}$. We conclude that the correlation of ζ with $\mathsf{GAPMAJ}_t\circ F$ is at least $1-2^{-\Omega(td)}$ as claimed.

9 Algorithmic and complexity-theoretic applications

9.1 Complexity measures

Throughout this section, for a function $F\colon\{-1,1\}^n\times\{-1,1\}^n\to\{-1,1\}$, we also view F as a $2^n\times 2^n$ matrix whose (x,y)'th entry is given by $F(x,y)$.

Approximate rank. For a matrix $F\in\{-1,1\}^{N\times N}$, the ε-approximate rank of F, denoted $\mathrm{rank}_\varepsilon(F)$, is the least rank of a matrix $A\in\mathbb{R}^{N\times N}$ such that $|A_{ij}-F_{ij}|\leq\varepsilon$ for all $(i,j)\in[N]\times[N]$. Approximate rank is a fundamental notion in learning theory and communication complexity. Meanwhile, the *sign-rank* of F is the least rank of a matrix $A\in\mathbb{R}^{N\times N}$ such that $|A_{ij}-F_{ij}|<1$ for all $(i,j)\in[N]\times[N]$. Clearly, for *any* F, the *exact* (not just approximate) rank of F is at most N.

Discrepancy. The discrepancy of F, denoted $\mathrm{disc}(F)$, is a combinatorial measure of the complexity of F that roughly captures F's correlation with combinatorial rectangles (small discrepancy corresponds to high complexity).

Definition 9.1. A combinatorial rectangle of $X\times Y$ is a set of the form $A\times B$ with $A\subseteq X$ and $B\subseteq Y$. For a distribution μ over $X\times Y$, the discrepancy of F with respect to μ is defined to be the maximum over all rectangles R of the *bias* of F on R. That is:

$$\mathrm{disc}_\mu(F) = \max_R\left|\sum_{(x,y)\in R}\mu(x,y)F(x,y)\right|.$$

The discrepancy of F, $\mathrm{disc}(F)$ is defined to be $\min_\mu \mathrm{disc}_\mu(F)$.

It is known that for any function $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$, the discrepancy of F is at least $2^{-O(n)}$. Discrepancy plays a central role in communication complexity because it characterizes $\mathsf{PP}^{\mathsf{cc}}$, the class of communication problems efficiently solvable by small-bias protocols [26]. Discrepancy is also important in circuit complexity, where it lower bounds the size of Majority-of-Threshold circuits computing F [23, 24, 34, 39]. Finally, discrepancy is known to be equivalent (up to a constant factor) to margin complexity [32], which is itself a fundamental notion in learning theory.

Threshold weight. The weight of an n-variate polynomial p is the sum of the absolute value of its coefficients. The length of p is the number of non-zero coefficients of p. The threshold weight (respectively, length) of $F\colon \{-1,1\}^N \times \{-1,1\}^N \to \{-1,1\}$ is the least weight (respectively, length) of a polynomial p with integer coefficients such that $p(x,y)\cdot F(x,y) > 0$ for all $(x,y) \in \{-1,1\}^N \times \{-1,1\}^N$ (note that no restriction is placed on the degree of p). The threshold weight of F is always at most $2^{O(n)}$, since Parseval's inequality implies that every Boolean function is always *exactly* computed by a polynomial of weight $2^{O(n)}$.

9.2 Nearly optimal bounds on discrepancy, threshold weight, and more

Via well-known techniques, we now translate our approximate-degree lower bounds into approximate-rank, discrepancy, and threshold-weight bounds in a black-box manner.

We start with the following theorem (from which Theorem 1.3 from Section 1.1.2 follows).

Theorem 9.2. *For any constant $\delta > 0$, there is an AC^0 function $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ with discrepancy at most $\exp(-\Omega(n^{1-\delta}))$ and ε-approximate rank $\exp(\Omega(n^{1-\delta}))$ for some $\varepsilon = 1-2^{-\Omega(n^{1-\delta})}$.*

Proof. Let f be the AC^0 function with ε-approximate degree at least $n^{1-\delta}$ for some $\varepsilon = 1-2^{-\Omega(n^{1-\delta})}$ whose existence is guaranteed by Theorem 8.1. The pattern matrix method [40, Theorem 8.1] implies that for some $C = O(1)$, the function $F\colon \{-1,1\}^{Cn} \times \{-1,1\}^{Cn} \to \{-1,1\}$ given by

$$F(x,y) = f\left(\ldots, \vee_{j=1}^{C} (x_{i,j} \wedge y_{i,j}) \ldots\right)$$

satisfies $\mathrm{rank}_{\varepsilon'}(F) \geq \exp(\Omega(n^{1-\delta}))$, for some $\varepsilon' = 1-2^{-\Omega(n^{1-\delta})}$. Moreover, by [40, Theorem 7.3] F also satisfies $\mathrm{disc}(F) \leq \exp(-\Omega(n^{1-\delta}))$.[10] Moreover, if f is computed by a Boolean circuit of depth k and polynomial size, then F is computed by a Boolean circuit of polynomial size and depth $k+2$. This completes the proof of the theorem for approximate rank and discrepancy.

Meanwhile, a result of Krause [29] implies that the following function F' on $3n$ inputs has threshold weight $2^{\Omega(n^{1-\delta})}$: $F'(x,y,z) = f(\ldots,(x_i \wedge \bar{z}_i) \vee (y_i \wedge z_i),\ldots)$ is at least $2^{\Omega(n^{1-\delta})}$.[11] Clearly, since f is computed by a Boolean circuit of depth k and polynomial size, F' is computed by a Boolean circuit of polynomial size and depth $k+2$. $\square$

[10] [40, Theorem 7.3] is actually expressed in terms of the degree d threshold weight of f, rather than the ε-approximate degree of f, but our lower bound on the ε-approximate degree of f is easily seen to imply the lower bound on the degree d threshold weight of f required to apply [40, Theorem 7.3] (see, e. g., [14, Lemma 20]).

[11] As in Footnote 10, Krause's result is expressed in terms of the degree d threshold weight of f, rather than the ε-approximate degree of f, but our lower bound on the ε-approximate degree of f is easily seen to imply the lower bound on the degree d threshold weight of f required to apply Krause's result (see, e. g., [14, Lemma 20]).

Via established applications of discrepancy, we conclude from Theorem 9.2 that for any constant $\delta > 0$, there is an AC^0 function F with margin complexity $\exp(\Omega(n^{1-\delta}))$ [32], Majority-of-Threshold circuit size $\exp(\Omega(n^{1-\delta}))$ [34, 39], and $\mathsf{PP}^{\mathsf{cc}}$ communication complexity $\Omega(n^{1-\delta})$ [26]. These bounds are all nearly tight, in the sense that any function has margin complexity $2^{O(n)}$, is computed by Majority-of-Threshold circuits of size $2^{O(n)}$, and has $\mathsf{PP}^{\mathsf{cc}}$ communication complexity at most n.

9.3 Lower bounds for sign-rank and threshold length

9.3.1 Threshold length

Theorem 9.3. *There is a function $F\colon \{-1,1\}^n \to \{-1,1\}$ computed by a polynomial-size circuit of depth 3 and logarithmic bottom fan-in such that the threshold length of F is* $\exp(\Omega(n^{1/2}))$.

Proof. Recall (Corollary 7.1) that for $n = N\log(N^{1/2})$, we have $\mathsf{SURJ}_{N^{1/2},N}\colon \{-1,1\}^n \to \{-1,1\}$, and $\deg_{\pm}(\mathsf{SURJ}_{N^{1/2},N}) = \Omega(n^{1/2})$. Moreover, $\mathsf{SURJ}_{N^{1/2},N}$ is computed by a quadratic-size circuit of depth 3, with an AND gate at the top, and logarithmic bottom fan-in.

Krause and Pudlák showed that if $\deg_{\pm}(f) \geq d$, then the function $F\colon \{-1,1\}^{3n} \to \{-1,1\}$ defined via $F(x,y,z) := f(\dots,(x_i \wedge \bar{z}_i) \vee (y_i \wedge z_i),\dots)$ has threshold length $2^{\Omega(d)}$. If $f = \mathsf{SURJ}$, then F is clearly computed by a polynomial-size circuit of depth 5, with an AND where gates at the bottom two layers (just above the inputs) have fan-in 2, and gates at the third-to-bottom layer have fan-in $O(\log n)$. Hence, each gate at the third layer from the bottom computes a function of just $O(\log n)$ inputs, and any such function can be computed by a polynomial-size DNF and logarithmic width. Replacing each gate at the third-to-bottom layer with an equivalent polynomial-size DNF of polynomial size and logarithmic width, and collapsing the two adjacent layers of OR gates, yields a depth-three circuit of logarithmic bottom fan-in. $\square$

9.3.2 Sign-rank

Recall that Corollary 1.2 from the Introduction asserted the existence of an AC^0 function

$$F(x,y)\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$$

such that the sign-rank of the matrix $[F(x,y)]$ is $\exp(\tilde{\Omega}(n^{1/2}))$. We prove Corollary 1.2 below.

Via established applications of sign-rank, we conclude as a consequence that there is a communication problem in AC^0 with $\mathbf{UPP}^{\mathrm{cc}}$ communication complexity $\tilde{\Omega}(n^{1/2})$ [36], and such that all Threshold-of-Majority circuits computing the function have size $2^{\tilde{\Omega}(n^{1/2})}$ [22].

Proof of Corollary 1.2. Suppose that $\deg_{\pm}(f) \geq d$. In order to establish sign-rank lower bounds for a certain matrix A derived from f, Razborov and Sherstov [38] extended a lemma of Forster [21] to show that it is enough to construct a dual witness μ for the fact that $\deg_{\pm}(f) \geq d$ that additionally satisfies a *smoothness* condition. Specifically, to show that the sign-rank of A is $\exp(\Omega(d))$, it suffices to show that there is a threshold degree dual witness μ for f satisfying $\mu(x) = \exp(-O(d))$ for all but an $\exp(-O(d))$ fraction of inputs in $x \in \{-1,1\}^n$. Formally, we have the following theorem, which is implicit in [38] (the statement here is taken from [15, Theorem 4.1]).

Theorem 9.4 (Implicit in [38, Theorem 1.1])**.** *Let $h\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function, and suppose there exists a function $\tau\colon \{-1,1\}^n \to \{-1,1\}$ of pure high degree at least d such that $\tau(x)\cdot h(x) \geq 0$ for all $x \in \{-1,1\}^n$, and $\|\tau\|_1 = 1$. Moreover, suppose that $\tau(x) \geq 2^{-cd}\cdot 2^{-n}$ for all but a 2^{-cd} fraction of inputs $x \in \{-1,1\}^n$. Then there exists a constant C (depending only on c) such that if $F(x,y) := h(\dots,\wedge_{j=1}^{C}(x_{ij}\vee y_{ij}),\dots)$, then the matrix $[F(x,y)]_{x,y}$ has sign-rank $\exp(\Omega(d))$.*

Let $R = N^{1/2}$, and assume $R+1$ is a power of 2. Letting $n = N\log(R+1)$, recall $\mathsf{dSURJ}_{R,N}$ is a function on n bits, where any $x \in \{-1,1\}^n$ is interpreted as specifying a list of N numbers from $[R]_0$. Here, 0 denotes a "dummy item" that is ignored by dSURJ. We assume without loss of generality in this section that:

$$\text{The dummy element is represented by the string } \mathbf{1}_{\log(R+1)}\,. \tag{9.1}$$

This ensures that strings $x \in \{-1,1\}^n$ that encode mostly dummy items have low Hamming weight.

Recall that within the proof of Theorem 1.1 and Corollary 7.1, we proved that $\deg_\pm(\mathsf{dSURJ}_{R,N}) \geq d$ for some $d = \Omega(N^{1/2}/\log^{3/2}N)$, via a two-step process. First, we borrowed a result (Theorem 5.19) from our prior work [18], which showed that for any range size R, $\deg_\pm(\mathsf{dSURJ}_{R,N})$ is equivalent to $\deg_\pm\left((\mathsf{AND}_R\circ\mathsf{OR}_N)^{\leq N}\right)$. Second, we constructed a dual witness $\mu\colon \mathcal{H}_{NR}^{\leq N}\to\mathbb{R}$ showing that the latter quantity is at least d.

Unfortunately, the construction of μ is not sufficient to apply Theorem 9.4 to $\mathsf{dSURJ}_{R,N}$, for two reasons. First, to apply Theorem 9.4 to $\mathsf{dSURJ}_{R,N}$, we need to give a smooth dual witness for $\mathsf{dSURJ}_{R,N}$ itself, rather than for $(\mathsf{AND}_R\circ\mathsf{OR}_N)^{\leq N}$. Note that $\mathsf{dSURJ}_{R,N}$ is defined over the domain $\{-1,1\}^n$ where $n = N\log(R+1)$, while $(\mathsf{AND}_R\circ\mathsf{OR}_N)^{\leq N}$ is defined over the domain $\mathcal{H}_{NR}^{\leq N}$. Second, the dual witness μ for $(\mathsf{AND}_R\circ\mathsf{OR}_N)^{\leq N}$ constructed in the proof of Corollary 7.1 is not smooth in the sense required by Razborov and Sherstov, as it is only "large" on inputs of Hamming weight at most d (see the first paragraph of Section 3.2 for further discussion of this point).

We will address both of the above issues as follows. First, we will show how to turn μ into a dual witness $\hat{\sigma}$ for the fact that $\deg_\pm(\mathsf{dSURJ}_{R,N}) \geq d$, 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 $\deg_\pm\left(\mathsf{dSURJ}_{R,N}\circ\mathsf{AND}_{\log^2 n}\circ\mathsf{PARITY}_{\log^3 n}\right) \geq d$, such that τ satisfies the smoothness condition required to apply Theorem 9.4.

Construction and analysis of $\hat{\sigma}$. In the proof of Theorem 3.1, we constructed a dual witness μ for $F = \mathsf{AND}_R\circ\mathsf{OR}_N$ satisfying the following conditions.

$$\mu(x)\cdot F(x) \geq 0 \text{ for all } x \in \{-1,1\}^{R\cdot N}\,. \tag{9.2}$$

$$\mu(x) = 0 \text{ for all } x \in \mathcal{H}_{R\cdot N}^{>N}\,. \tag{9.3}$$

$$\mu \text{ has pure high degree at least } d\,. \tag{9.4}$$

$$\mu(x) \geq 2^{-1-R/5} \text{ for all } x \in \mathcal{H}_{NR}^{\leq d}\,. \tag{9.5}$$

$$\mu(x) \text{ has } \ell_1\text{-norm at most } 2\,. \tag{9.6}$$

Expressions (9.2)-(9.4) are explicitly established in the proof of Theorem 3.1 (specifically, the part entitled "Analysis of μ"). Expression (9.5) is immediate from the penultimate inequality in Expression (7.28). Equation (9.6) holds because we defined $\mu = \zeta - \sum_{x \in B} \psi_{\mathsf{corr},x}$ (Expression (7.26)), and hence

$$\|\mu\|_1 \leq \|\zeta\|_1 + \sum_{x \in B} \|\psi_{\mathsf{corr},x}\|_1 \leq 1 + (E(\zeta, \mathsf{AND}_R \circ \mathsf{OR}_N) + W(\zeta, f)) \cdot 2^d \cdot \binom{RN}{d} \leq 2.$$

Defining $\mu' = \mu / \|\mu\|_1$ yields a function that also satisfies Equations (9.2)-(9.4), and such that:

$$\mu'(x) \geq 2^{-R/4} \text{ for all } x \in \mathcal{H}_{NR}^{\leq d}. \tag{9.7}$$

$$\mu'(x) \text{ has } \ell_1\text{-norm } 1\,. \tag{9.8}$$

Let $\mathcal{Z}$ denote the subset of $([N]_0)^R$ defined as follows: $\mathcal{Z} := \{z \in ([N]_0)^R : z_1 + \cdots + z_R \leq N\}$. Define $G(z_1, \ldots, z_R) \colon \mathcal{Z} \to \{-1, 1\}$ to equal -1 if and only if $z_i \geq 1$ for all $i = 1, \ldots, R$.

We will now turn μ' (which is defined over domain $\{-1,1\}^{NR}$) into a function over domain $\mathcal{Z}$.

Define

$$\sigma(z_1, \ldots, z_R) = \sum_{x = (x_1, \ldots, x_R) \in (\{-1,1\}^N)^R : |x_i| = z_i \text{ for all } i} \mu'(x).$$

We claim that σ has the following properties.

$$\sigma(z) \cdot G(z) \geq 0 \text{ for all } z \in \mathcal{Z}. \tag{9.9}$$

$$\sum_{z \in \mathcal{Z}} |\sigma(z)| = 1\,. \tag{9.10}$$

$$\deg(q) \leq d \Longrightarrow \sum_{z \in \mathcal{Z}} \sigma(z) \cdot q(z) = 0\,. \tag{9.11}$$

$$\sigma(z) \geq 2^{-R/4} \text{ for all } z \in \mathcal{Z} \text{ such that } \sum_{i=1}^{R} z_i \leq d\,. \tag{9.12}$$

Expression (9.9) is immediate from Expression (9.2). Equation (9.10) is immediate from Equations (9.8) and (9.2), and the fact that $F(x) = F(x')$ if $|x_i| = |x_i'|$ for $i = 1, \ldots, R$. Equation (9.11) holds because

$$\sum_{z \in \mathcal{Z}} \sigma(z) q(z) = \sum_{x = (x_1, \ldots, x_R) \in \{-1,1\}^{R \cdot N}} \mu'(x) \cdot q(|x_1|, \ldots, |x_R|) = 0,$$

where the last equality holds because $q(|x_1|, \ldots, |x_R|)$ is a polynomial over $\left(\{-1,1\}^N\right)^R$ of degree at most $\deg(q)$. Expression (9.12) is immediate from Expression (9.7).

Finally, we are in a position to define our desired dual witness $\hat{\sigma} \colon \{-1,1\}^n \to \mathbb{R}$, which will witness the fact that $\deg_{\pm}(\mathsf{dSURJ}_{R,N}) \geq d / \log R$. For an $x \in \{-1,1\}^n$ interpreted as a sequence $(s_1, \ldots, s_N)$ in $[R]_0^N$, let $z_i(x)$ denote $\left|\{j \colon s_j = i\}\right|$, and let $z(x) = (z_1(x), \ldots, z_R(x)) \in \mathcal{Z}$. For a $z^* \in \mathcal{Z}$, let $N(z^*) = |\{x \in \{-1,1\}^n \colon z(x) = z^*\}|$. Define $\hat{\sigma} \colon \{-1,1\}^n \to \mathbb{R}$ via:

$$\hat{\sigma}(x) = \frac{1}{N(z(x))} \cdot \sigma(z(x))\,. \tag{9.13}$$

We observe that $\hat{\sigma}$ has the following properties.

$$\hat{\sigma}(x) \cdot \mathsf{dSURJ}(x) \geq 0 \text{ for all } x \in \{-1,1\}^n . \tag{9.14}$$

$$\hat{\sigma}(x) \text{ has } \ell_1\text{-norm } 1 . \tag{9.15}$$

$$\hat{\sigma}(x) \geq 2^{-R/3} \text{ for all } x \in \mathcal{H}_n^{\leq d} . \tag{9.16}$$

$$\hat{\sigma}(x) \text{ has pure high degree at least } d/\log R . \tag{9.17}$$

Expression (9.14) follows from Expression (9.9) and the definition of $\hat{\sigma}$. Equation (9.15) is immediate from the definition of $\hat{\sigma}$ and Equation (9.10). To see that Expression (9.16) holds, observe that if $x \in \mathcal{H}_n^{\leq d}$, then Equation (9.1) implies that $\sum_{i=1}^R z_i(x) \leq d$. Hence, Expression (9.12) implies that $\sigma(z(x)) \geq 2^{-R/4}$. Moreover, for any $x \in \mathcal{H}_n^{\leq d}$, $N(z(x)) \leq \binom{n}{d}$ (this is because for any x, x', if $z(x) = z(x')$, then $|x| = |x'|$, so $N(z(x)) \leq \binom{n}{|x|} \leq \binom{n}{d}$). Combining these two inequalities with the definition of $\hat{\sigma}$ (Equation (9.13)) yields that $\hat{\sigma}(x) \geq 2^{-R/4} \cdot \binom{n}{d}^{-1} \geq 2^{-R/3}$.

Equation (9.17) follows from Proposition 9.5 below, combined with Equation (9.11).

Proposition 9.5. *Let $\sigma \colon : \mathcal{Z} \to \mathbb{R}$ satisfy Equation (9.11), and let $\hat{\sigma}$ be as per Equation (9.13). Then $\hat{\sigma}$ has pure high degree at least $d/\log R$.*

Proposition 9.5 is essentially a dual formulation of an important lemma of Ambainis [4], as we now explain.

Lemma 9.6 (Ambainis [4])**.** *Let $n = N\log(R+1)$. Let $p \colon \{-1,1\}^n \to \{-1,1\}$ be any polynomial of degree at most d. Then there is a polynomial $q \colon [R]_0^N \to \mathbb{R}$ of degree at most $d \log R$ such that for all $z^* \in \mathcal{Z}$, $q(z^*) = \mathbb{E}_{x \colon z(x)=z^*}[p(x)]$.*

Proposition 9.5 follows from Lemma 9.6 by the following reasoning. If $p \colon \{-1,1\}^n \to \{-1,1\}$ is polynomial of degree at most $d/\log R$, then

$$\sum_{x \in \{-1,1\}^n} \hat{\sigma}(x) \cdot p(x) = \sum_{z \in \mathcal{Z}} \sigma(z) \cdot q(z) = 0 , \tag{9.18}$$

where the polynomial q is as in Lemma 9.6, and the second equality holds by Equation (9.11).

Construction and analysis of τ. Now that we have constructed a dual witness $\hat{\sigma}$ for the high threshold degree of dSURJ (captured by Equations (9.14), (9.15), and (9.17)), that additionally satisfies the extra condition of "largeness on low-Hamming-weight inputs" (Equation (9.16)), we can turn to constructing a dual witness for $\mathsf{dSURJ} \circ \mathsf{AND}_{\log^2 n} \circ \mathsf{PARITY}_{\log^3 n}$ that satisfies the smoothness condition needed by Razborov's and Sherstov's sign-rank analysis.

Specifically, let $a = \log^2 n$, and define $\psi \colon \{-1,1\}^a \to \mathbb{R}$ via:

$$\psi(x) = \begin{cases} -1/2 \text{ if } x = -\mathbf{1}_a \\ 1/(2 \cdot (2^a - 1)) \text{ otherwise.} \end{cases}$$

Clearly, ψ has pure high degree at least 1, has ℓ_1-norm 1, and $\psi(x) \cdot \mathsf{AND}_a(x) \geq 0$ for all $x \in \{-1,1\}^a$. Consider the dual witness $\zeta = \hat{\sigma} \star \psi$. Then ζ has the following properties:

$$\zeta \text{ has } \ell_1\text{-norm } 1\,. \tag{9.19}$$

$$\zeta \text{ has pure high degree at least } d/\log R\,. \tag{9.20}$$

$$\zeta(x) \cdot \left(\mathsf{dSURJ} \circ \mathsf{AND}_{\log^2 n}\right)(x) \geq 0 \text{ for all } x \in \{-1,1\}^{n \cdot a}\,. \tag{9.21}$$

Here, Equation (9.19) follows from Equation (9.15), the fact that ψ has ℓ_1-norm 1, and Property (5.15). Equation (9.20) follows from Equation (9.17), Equation (5.16), and the fact that ψ has pure high degree at least 1. Expression (9.21) follows from the definition of dual block composition, Expression (9.14), and the fact that $\psi(x) \cdot \mathsf{AND}_a(x) \geq 0$ for all $x \in \{-1,1\}^a$.

Let S be the set of all inputs $x = (x_1, \ldots, x_n) \in (\{-1,1\}^a)^n$ such that fewer than d of the x_i's are equal to $-\mathbf{1}_a$. By Expression (9.16) and the definition of dual block composition,

$$\text{For any input } x \in S,\ |\zeta(x)| \geq 2^{-R/3} \cdot (2^a - 1)^{-n} \geq 2^{-R/3} \cdot 2^{-an}\,. \tag{9.22}$$

Moreover,

$$\Pr_{x \sim (\{-1,1\}^a)^n}[x \notin S] \leq 2^{-ad} \cdot \binom{n}{d} \leq n^{-\Omega(d \log n)} \leq 2^{-\Omega(R)}\,. \tag{9.23}$$

Finally, letting $\ell = \log^3 n$, let $\eta \colon \{-1,1\}^\ell \to \mathbb{R}$ be defined by $\eta(x) = 2^{-\ell} \cdot \mathsf{PARITY}_\ell(x)$. Observe that η has ℓ_1-norm 1, has pure high degree ℓ, and $\operatorname{sgn}(\eta(x)) = \operatorname{sgn}(\mathsf{PARITY}_\ell(x))$ for all $x \in \{-1,1\}^\ell$. Then Equation (5.15) implies that $\tau := \zeta \star \eta$ has ℓ_1-norm 1, Equation (5.16) implies that τ has pure high degree at least $(d/\log R) \cdot \ell \geq R$, and it follows from Equations (9.22) and (9.23) that

$$\Pr_{x \sim \left(\left(\{-1,1\}^\ell\right)^a\right)^n}\left[|\tau(x)| < 2^{-R/3} \cdot 2^{-\ell \cdot a \cdot n}\right] \leq 2^{-\Omega(R)}\,. \tag{9.24}$$

Hence, we can apply Theorem 9.4 to the function $h = \mathsf{dSURJ} \circ \mathsf{AND}_{\log^2 n} \circ \mathsf{PARITY}_{\log^3 n}$, which is defined on $\tilde{O}(n)$ variables, to obtain an AC^0 function $F(x,y)$ that is also defined on $\tilde{O}(n)$ variables, such that $[F(x,y)]_{x,y}$ has sign-rank $\exp(\tilde{\Omega}(n^{1/2}))$. $\square$

9.4 Secret sharing schemes

Bogdanov et al. [7] observed that for any $f \colon \{-1,1\}^n \to \{-1,1\}$ and integer $d > 0$, any dual polynomial μ for the fact that $\widetilde{\deg}_\varepsilon(f) \geq d$ leads to a scheme for sharing a single secret bit $b \in \{-1,1\}$ among n parties as follows. Decompose μ as $\mu_+ - \mu_-$, where μ_+ and μ_- are non-negative functions with $\|\mu_+\|_1 = \|\mu_0\|_1 = 1/2$. Then in order to split b among n parties, one draws an input $x = (x_1, \ldots, x_n) \in \{-1,1\}^n$ from the distribution $2 \cdot \mu_b$, and gives bit x_i to the ith party. In order to reconstruct b, one applies f to $(x_1, \ldots, x_n)$.

Because μ is ε-correlated with f, the probability of correct reconstruction if the bit is chosen at random is at least $(1+\varepsilon)/2$ (and the the *reconstruction advantage*, defined to equal $\Pr_{x \sim \mu_+}[f(x) = 1] - \Pr_{x \sim \mu_-}[f(x) = 1]$, is at least ε). The fact that μ has pure high degree at least d means that any subset of shares of size less than d provides no information about the secret bit b. Hence, an immediate corollary of Theorem 8.1 is the following.

Corollary 9.7. *For any arbitrarily small constant $\delta > 0$, there is a secret sharing scheme that shares a single bit b among n parties by assigning a bit x_i to each party i. The scheme has the following properties.*

1. *The reconstruction procedure is computed by an AC^0 circuit.*
2. *The reconstruction advantage is at least $1 - 2^{-\Omega(n^{1-\delta})}$.*
3. *Any subset of shares of size less than $d = \Omega(n^{1-\delta})$ provides no information about the secret bit b.*

The best previous results could only guarantee security against subsets of shares of size $d = \Theta(n^{1/2})$, or could only guarantee reconstruction advantage bounded away from 1 [7,18]. Cheng et al. [19] recently considered a relaxed notion of security, where even very small subsets of shares are allowed to provide (a bounded amount of) information about the secret bit b. Under this relaxed notion of security, they achieved perfect reconstruction and security against subsets of size $\Omega(n)$.

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

References

[1] SCOTT AARONSON AND YAOYUN SHI: Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004. [doi:10.1145/1008731.1008735] 16

[2] MIKLOS AJTAI AND MICHAEL BEN-OR: A theorem on probabilistic constant depth computations. In *Proc. 16th STOC*, pp. 471–474. ACM Press, 1984. [doi:10.1145/800057.808715] 28

[3] ERIC ALLENDER: A note on the power of threshold circuits. In *Proc. 30th FOCS*, pp. 580–584. IEEE Comp. Soc., 1989. [doi:10.1109/SFCS.1989.63538] 3, 7

[4] ANDRIS AMBAINIS: Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(3):37–46, 2005. [doi:10.4086/toc.2005.v001a003] 39

[5] LÁSZLÓ BABAI, PÉTER FRANKL, AND JANOS SIMON: Complexity classes in communication complexity theory (preliminary version). In *Proc. 27th FOCS*, pp. 337–347. IEEE Comp. Soc., 1986. [doi:10.1109/SFCS.1986.15] 3, 6

[6] PAUL BEAME AND WIDAD MACHMOUCHI: The quantum query complexity of AC^0. *Quantum Inf. Comput.*, 12(7-8):670–676, 2012. [doi:10.26421/QIC12.7-8-11] 4

[7] ANDREJ BOGDANOV, YUVAL ISHAI, EMANUELE VIOLA, AND CHRISTOPHER WILLIAMSON: Bounded indistinguishability and the complexity of recovering secrets. In *Proc. 36th Ann. Internat. Cryptology Conf. (CRYPTO'16)*, pp. 593–618. Springer, 2016. [doi:10.1007/978-3-662-53015-3_21, ECCC:TR15-182] 4, 7, 40, 41

[8] ANDREJ BOGDANOV AND CHRISTOPHER WILLIAMSON: Approximate bounded indistinguishability. In *Proc. 44th Internat. Colloq. on Automata, Languages, and Programming (ICALP'17)*, pp. 53:1–11. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017. [doi:10.4230/LIPIcs.ICALP.2017.53] 4

[9] ADAM BOULAND, LIJIE CHEN, DHIRAJ HOLDEN, JUSTIN THALER, AND PRASHANT NALINI VASUDEVAN: On the power of statistical zero knowledge. *SIAM J. Comput.*, 49(4):FOCS17:1–58, 2019. Preliminary version in FOCS'17. [doi:10.1137/17M1161749, ECCC:TR16-140, arXiv:1609.02888] 4, 10, 12, 23, 28

[10] HARRY BUHRMAN, RICHARD CLEVE, RONALD DE WOLF, AND CHRISTOF ZALKA: Bounds for small-error and zero-error quantum algorithms. In *Proc. 40th FOCS*, pp. 358–368. IEEE Comp. Soc., 1999. [doi:10.1109/SFFCS.1999.814607, arXiv:cs/9904019] 22

[11] HARRY BUHRMAN, NIKOLAI K. VERESHCHAGIN, AND RONALD DE WOLF: On computation and communication with small bias. In *Proc. 22nd IEEE Conf. on Comput. Complexity (CCC'07)*, pp. 24–32. IEEE Comp. Soc., 2007. [doi:10.1109/CCC.2007.18] 2, 3, 6

[12] MARK BUN, ROBIN KOTHARI, AND JUSTIN THALER: The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *Theory of Computing*, 16(10):1–71, 2020. Preliminary version in STOC'18. [doi:10.4086/toc.2020.v016a010] 4, 5, 10, 12, 13, 14, 16, 17, 18, 20, 23

[13] MARK BUN AND JUSTIN THALER: Dual lower bounds for approximate degree and Markov–Bernstein inequalities. *Inform. Comput.*, 243:2–25, 2015. Preliminary version in ICALP'13. [doi:10.1016/j.ic.2014.12.003] 4

[14] MARK BUN AND JUSTIN THALER: Hardness amplification and the approximate degree of constant-depth circuits. In *Proc. 42nd Internat. Colloq. on Automata, Languages, and Programming (ICALP'15)*, pp. 268–280. Springer, 2015. [doi:10.1007/978-3-662-47672-7_22, arXiv:1311.1616] 4, 6, 9, 11, 23, 30, 35

[15] MARK BUN AND JUSTIN THALER: Improved bounds on the sign-rank of AC^0. In *Proc. 43rd Internat. Colloq. on Automata, Languages, and Programming (ICALP'16)*, pp. 37:1–14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. [doi:10.4230/LIPIcs.ICALP.2016.37] 3, 4, 6, 36

[16] MARK BUN AND JUSTIN THALER: Approximate degree and the complexity of depth three circuits. In *Proc. 22nd Internat. Workshop on Randomization and Computation (RANDOM'18)*, pp. 35:1–18. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. [doi:10.4230/LIPIcs.APPROX-RANDOM.2018.35, ECCC:TR16-121] 4, 6

[17] MARK BUN AND JUSTIN THALER: The large-error approximate degree of AC^0. In *Proc. 23rd Internat. Workshop on Randomization and Computation (RANDOM'19)*, pp. 55:1–16. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.APPROX-RANDOM.2019.55] 1

[18] MARK BUN AND JUSTIN THALER: A nearly optimal lower bound on the approximate degree of AC^0. *SIAM J. Comput.*, 49(4):FOCS17:59–96, 2019. Preliminary version in FOCS'17. [doi:10.1137/17M1161737] 4, 5, 8, 10, 12, 13, 14, 19, 20, 21, 37, 41

[19] KUAN CHENG, YUVAL ISHAI, AND XIN LI: Near-optimal secret sharing and error correcting codes in AC^0. In *Proc. Theory of Cryptography Conf. (TCC'17)*, pp. 424–458. Springer, 2017. [doi:10.1007/978-3-319-70503-3_14] 4, 41

[20] VITALY FELDMAN: Evolvability from learning algorithms. In *Proc. 40th STOC*, pp. 619–628. ACM Press, 2008. [doi:10.1145/1374376.1374465] 7

[21] JÜRGEN FORSTER: A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. System Sci.*, 65(4):612–625, 2002. Preliminary version in CCC'01. [doi:10.1016/S0022-0000(02)00019-3] 6, 11, 36

[22] 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 *Proc. 21st Found. Softw. Techn. Theoret. Comp. Sci. Conf. (EFSTTCS'01)*, pp. 171–182. Springer, 2001. [doi:10.1007/3-540-45294-X_15] 8, 36

[23] MIKAEL GOLDMANN, JOHAN HÅSTAD, AND ALEXANDER A. RAZBOROV: Majority gates vs. general weighted threshold gates. *Comput. Complexity*, 2(4):277–300, 1992. [doi:10.1007/BF01200426] 35

[24] ANDRÁS HAJNAL, WOLFGANG MAASS, PAVEL PUDLÁK, MARIO SZEGEDY, AND GYÖRGY TURÁN: Threshold circuits of bounded depth. *J. Comput. System Sci.*, 46(2):129–154, 1993. [doi:10.1016/0022-0000(93)90001-D] 35

[25] JEFF KAHN, NATHAN LINIAL, AND ALEX SAMORODNITSKY: Inclusion-exclusion: Exact and approximate. *Combinatorica*, 16(4):465–477, 1996. [doi:10.1007/BF01271266] 22

[26] HARTMUT KLAUCK: Lower bounds for quantum communication complexity. *SIAM J. Comput.*, 37(1):20–46, 2007. Preliminary version in FOCS'01. [doi:10.1137/S0097539702405620, arXiv:quant-ph/0106160] 7, 35, 36

[27] ADAM R. KLIVANS AND ROCCO A. SERVEDIO: Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. System Sci.*, 68(2):303–318, 2004. [doi:10.1016/j.jcss.2003.07.007] 3, 8

[28] SWASTIK KOPPARTY: AC^0 lower bounds and pseudorandomness, 2013. Lecture notes of "Topics in Complexity Theory and Pseudorandomness (Spring 2013)" at Rutgers University. 28

[29] MATTHIAS KRAUSE: On the computational power of Boolean decision lists. *Comput. Complexity*, 14(4):362–375, 2006. [doi:10.1007/s00037-005-0203-0] 35

[30] MATTHIAS KRAUSE AND PAVEL PUDLÁK: On the computational power of depth-2 circuits with threshold and modulo gates. *Theoret. Comput. Sci.*, 174(1–2):137–156, 1997. [doi:10.1016/S0304-3975(96)00019-9] 3, 6

[31] TROY LEE: A note on the sign degree of formulas, 2009. [arXiv:0909.4607] 9, 19

[32] NATI LINIAL AND ADI SHRAIBMAN: Learning complexity vs communication complexity. *Combin. Probab. Comput.*, 18(1–2):227–245, 2009. [doi:10.1017/S0963548308009656] 7, 35, 36

[33] MARVIN MINSKY AND SEYMOUR PAPERT: *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1969. [doi:10.7551/mitpress/11301.001.0001] 3, 6, 8

[34] NOAM NISAN: The communication complexity of threshold gates. In *Combinatorics, Paul Erdős is Eighty*, pp. 301–315. János Bolyai Math. Soc., Budapest, 1994. 7, 35, 36

[35] RYAN O'DONNELL AND ROCCO A. SERVEDIO: New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010. Preliminary version in STOC'03. [doi:10.1007/s00493-010-2173-3] 6

[36] RAMAMOHAN PATURI AND JANOS SIMON: Probabilistic communication complexity. *J. Comput. System Sci.*, 33(1):106–123, 1986. [doi:10.1016/0022-0000(86)90046-2] 6, 7, 36

[37] VLADIMIR V. PODOLSKII: Perceptrons of large weight. *Probl. Info. Transmission*, 45:46–53, 2009. Preliminary version in CSR'07. [doi:10.1134/S0032946009010062] 2

[38] 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] 3, 6, 11, 18, 36, 37

[39] ALEXANDER A. SHERSTOV: Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. [doi:10.1137/08071421X] 3, 6, 35, 36

[40] ALEXANDER A. SHERSTOV: The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. Preliminary version in STOC'08. [doi:10.1137/080733644] 3, 5, 6, 35

[41] ALEXANDER A. SHERSTOV: Strong direct product theorems for quantum communication and query complexity. *SIAM J. Comput.*, 41(5):1122–1165, 2012. Preliminary version in STOC'14. [doi:10.1137/110842661] 12, 29, 30

[42] ALEXANDER A. SHERSTOV: The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013. Preliminary version in FOCS'09. [doi:10.1137/100785260] 9, 19

[43] ALEXANDER A. SHERSTOV: Breaking the Minsky-Papert barrier for constant-depth circuits. *SIAM J. Comput.*, 47(5):1809–1857, 2018. Preliminary version in STOC'14. [doi:10.1137/15M1015704] 4, 6, 23

[44] ALEXANDER A. SHERSTOV: The power of asymmetry in constant-depth circuits. *SIAM J. Comput.*, 47(6):2362–2434, 2018. Preliminary version in FOCS'15. [doi:10.1137/100785260] 3, 4, 5, 6, 9, 10

[45] ALEXANDER A. SHERSTOV: Algorithmic polynomials. *SIAM J. Comput.*, 49(6):1173–1231, 2020. Preliminary version in STOC'18. [doi:10.1137/19M1278831, arXiv:1801.04607] 4, 14

[46] ALEXANDER A. SHERSTOV AND PEI WU: Near-optimal lower bounds on the threshold degree and sign-rank of AC^0. *SIAM J. Comput.*, online:STOC19:1–86, 2021. Preliminary version in STOC'19. [doi:10.1137/20M1312459, arXiv:1901.00988] 13

[47] YAOYUN SHI AND YUFAN ZHU: Quantum communication complexity of block-composed functions. *Quantum Inf. Comput.*, 9(5):444–460, 2009. [doi:10.26421/QIC9.5-6-7] 9, 19

[48] LESLIE G. VALIANT: Evolvability. *J. ACM*, 56(1):3:1–21, 2009. [doi:10.1145/1462153.1462156] 7

[49] EMANUELE VIOLA: On approximate majority and probabilistic time. *Comput. Complexity*, 18(3):337–375, 2009. Preliminary version in CCC'07. [doi:10.1007/s00037-009-0267-3] 28

[50] ROBERT ŠPALEK: A dual polynomial for OR, 2008. [arXiv:0803.4516] 18

AUTHORS

Mark Bun
Assistant professor
Department of Computer Science
Boston University
Boston, MA, USA
mbun@bu.edu
http://cs-people.bu.edu/mbun/

Justin Thaler
Assistant professor
Department of Computer Science
Georgetown University
Washington, DC, USA
justin.thaler@georgetown.edu
http://people.cs.georgetown.edu/jthaler/

ABOUT THE AUTHORS

MARK BUN is an Assistant Professor in the Department of Computer Science at Boston University. He completed his Ph. D. in the Theory of Computation group at Harvard University, where he was advised by Salil Vadhan. After that, he was a postdoctoral researcher at Princeton University and a Google Research Fellow at the Simons Institute. His research interests include computational complexity, differential privacy, cryptozoology, and learning theory. As a native Seattleite, he enjoys coffee, rain, composting, 4-way stop signs, hiking, cheese rolling, and competitive duck herding.

JUSTIN THALER, known to his friends as Justin, is an Assistant Professor in the Department of Computer Science at Georgetown University. He completed his Ph. D. in the Theory of Computation group at Harvard University, where he was advised by Michael Mitzenmacher. After that, he was a postdoctoral researcher at the Simons Institute, and a Research Scientist at Yahoo Labs in New York City. His research interests include computational complexity, verifiable computing, and algorithms for massive datasets. In his spare time, he enjoys running and playing low-quality tennis.