Electronic Colloquium on Computational Complexity, Report No. 51 (2017)

A Nearly Optimal Lower Bound on the Approximate Degree of AC^0

Mark Bun*
mbun@cs.princeton.edu

Justin Thaler†
justin.thaler@georgetown.edu

Abstract

The approximate degree of a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$ is the least degree of a real polynomial that approximates f pointwise to error at most $1/3$. We introduce a generic method for increasing the approximate degree of a given function, while preserving its computability by constant-depth circuits.

Specifically, we show how to transform any Boolean function f with approximate degree d into a function F on $O(n \cdot \mathrm{polylog}(n))$ variables with approximate degree at least $D = \Omega(n^{1/3} \cdot d^{2/3})$. In particular, if $d = n^{1-\Omega(1)}$, then D is polynomially larger than d. Moreover, if f is computed by a polynomial-size Boolean circuit of constant depth, then so is F.

By recursively applying our transformation, for any constant $\delta > 0$ we exhibit an AC^0 function of approximate degree $\Omega(n^{1-\delta})$. This improves over the best previous lower bound of $\Omega(n^{2/3})$ due to Aaronson and Shi (J. ACM 2004), and nearly matches the trivial upper bound of n that holds for any function. Our lower bounds also apply to (quasipolynomial-size) DNFs of polylogarithmic width.

We describe several applications of these results. We give:

- For any constant $\delta > 0$, an $\Omega(n^{1-\delta})$ lower bound on the quantum communication complexity of a function in AC^0.
- A Boolean function f with approximate degree at least $C(f)^{2-o(1)}$, where $C(f)$ is the *certificate complexity* of f. This separation is optimal up to the $o(1)$ term in the exponent.
- Improved secret sharing schemes with reconstruction procedures in AC^0.

1 Introduction

The ε-approximate degree of a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$, denoted $\widetilde{\deg}_\varepsilon(f)$, is the least degree of a real polynomial that approximates f pointwise to error at most ε. By convention, $\widetilde{\deg}(f)$ is used to denote $\widetilde{\deg}_{1/3}(f)$, and this quantity is referred to without qualification as the *approximate degree* of f. The choice of the constant $1/3$ is arbitrary, as $\widetilde{\deg}(f)$ is related to $\widetilde{\deg}_\varepsilon(f)$ by a constant factor for any constant $\varepsilon \in (0,1)$. Any Boolean function f has an exact representation as a multilinear polynomial of degree at most n, so the approximate degree of f is always at most n.

Approximate degree is a natural measure of the complexity of a Boolean function, with a wide variety of applications throughout theoretical computer science. For example, upper bounds on approximate degree underly many state-of-the-art learning algorithms [8, 32–35, 40, 45], algorithmic approximations for the inclusion-exclusion principle [31, 48], and algorithms for differentially private data release [22, 65]. Very recently, approximate degree upper bounds have also been used to show new complexity-theoretic *lower*

*Princeton University.
†Georgetown University.

ISSN 1433-8092

bounds. In particular, upper bounds on the approximate degree of Boolean formulae underly the best known lower bounds on the formula complexity and graph complexity of explicit functions [61–63].

Meanwhile, lower bounds on approximate degree have enabled significant progress in quantum query complexity [2,4,9], communication complexity [16,23,26–28,43,47,50,51,53], circuit complexity [38,49], oracle separations [11, 15], and secret-sharing [14]. In particular, approximate degree has been established as one of the most promising tools available for understanding the complexity of constant-depth Boolean circuits[1] (captured by the complexity class AC^0). Indeed, approximate degree lower bounds lie at the heart of the best known bounds on the complexity of AC^0 under measures such as sign-rank, discrepancy and margin complexity, Majority-of-Threshold and Threshold-of-Majority circuit size, and more.

Despite all of these applications, progress in understanding approximate degree has been slow and difficult. As noted by many authors, the following basic problem remains unresolved [10, 14, 18–21, 46, 57].

Problem 1. *Is there a constant-depth circuit in n variables with approximate degree $\Omega(n)$?*

Prior to this work, the best result in this direction was Aaronson and Shi's well-known $\Omega(n^{2/3})$ lower bound on the approximate degree of the Element Distinctness function (ED for short). In this paper, we nearly resolve Open Problem 1. Specifically, for any constant $\delta > 0$, we exhibit an explicit constant-depth circuit $\mathcal{C}$ with approximate degree $\Omega(n^{1-\delta})$. Moreover, the circuit $\mathcal{C}$ that we exhibit has depth $O(\log(1/\delta))$. Our lower bound also applies to DNF formulae of polylogarithmic width (and quasipolynomial size).

Applications. We describe several consequences of the above results in complexity theory and cryptography. (Nevertheless, the list of applications we state here is not exhaustive.) We state these results somewhat informally in this introduction, leaving details to Section 5. Specifically:

- For any constant $\delta > 0$, we obtain an $\Omega(n^{1-\delta})$ lower bounds on the quantum communication complexity of AC^0. This nearly matches the trivial $O(n)$ upper bound that holds for any function.
- We exhibit a function f with approximate degree at least $C(f)^{2-o(1)}$, where $C(f)$ is the *certificate complexity* of f. This separation is optimal up to the $o(1)$ term in the exponent. The previous best result was a power-7/6 separation, reported by Aaronson et al. [3].
- We give improved secret sharing schemes with reconstruction procedures in AC^0.

While the first and third applications follow by combining our approximate degree lower bounds with prior works in a black box manner [14, 50], the second application requires some additional effort.

1.1 Prior Work on Approximate Degree

1.1.1 Early Results via Symmetrization

The notion of approximate degree was introduced in seminal work of Nisan and Szegedy [39], who proved a tight $\Omega(n^{1/2})$ lower bound on the approximate degree of the functions OR_n and AND_n.[2] Nisan and Szegedy's proof exploited a powerful technique known as *symmetrization*, which was introduced in the late 1960's by Minsky and Papert [38]. Until recently, symmetrization was the primary tool available for proving approximate degree lower bounds [4, 5, 11, 41, 42, 45].

Symmetrization arguments proceed in two steps. First, a polynomial p on n variables (which is assumed to approximate the target function f) is transformed into a univariate polynomial q in such a way that

[1] In this paper, all circuits are Boolean and of polynomial size unless otherwise specified.

[2] Whenever it is not clear from context, we use subscripts to denote the number of variables on which a function is defined.

$\deg(q) \leq \deg(p)$. Second, a lower bound on $\deg(q)$ is proved, using techniques tailored to the analysis of univariate polynomials.

Although powerful, symmetrization is inherently lossy: by turning a polynomial p on n variables into a univariate polynomial q, information about p is necessarily thrown away. Hence, several works identified the development of non-symmetrization techniques for lower bounding the approximate degree of Boolean functions as an important research direction (e.g., [1, 47, 54]). A relatively new such lower-bound technique called the *method of dual polynomials* plays an essential role in our paper.

1.2 The Method of Dual Polynomials and the AND-OR Tree

A dual polynomial is a dual solution to a certain linear program capturing the approximate degree of any function. These polynomials act as certificates of the high approximate degree of a function. Strong LP duality implies that the technique is lossless, in contrast to symmetrization. That is, for any function f and any ε, there is always some dual polynomial ψ that witnesses a tight ε-approximate degree lower bound for f.

A dual polynomial that witnesses the fact that $\widetilde{\deg}_\varepsilon(f_n) \geq d$ is a function $\psi\colon \{-1,1\}^n \to \{-1,1\}$ satisfying three properties:

- $\sum_{x \in \{-1,1\}^n} \psi(x) \cdot f(x) > \varepsilon$. If ψ satisfies this condition, it is said to be *well-correlated* with f.
- $\sum_{x \in \{-1,1\}^n} |\psi(x)| = 1$. If ψ satisfies this condition, it is said to have ℓ_1-norm equal to 1.
- For all polynomials $p\colon \{-1,1\}^n \to \mathbb{R}$ of degree less than d, we have $\sum_{x \in \{-1,1\}^n} p(x) \cdot \psi(x) = 0$. If ψ satisfies this condition, it is said to have *pure high degree* at least d.

One success story for the method of dual polynomials is the resolution of the approximate degree of the two-level AND-OR tree. For many years, this was the simplest function whose approximate degree resisted characterization by symmetrization methods [5, 39, 54, 58]. Given two functions f_M, g_N, let $f \circ g\colon \{-1,1\}^{M \cdot N} \to \{-1,1\}$ denote their *block composition*, i.e., $f \circ g = f(g, \dots, g)$.

Theorem 1. *The approximate degree of the function* $\mathsf{AND}_M \circ \mathsf{OR}_N$ *is* $\Theta(\sqrt{M \cdot N})$.

Ideas pertaining to both the upper and lower bounds of Theorem 1 will be useful to understanding the results in this paper. The upper bound of Theorem 1 was established by Høyer, Mosca, and de Wolf [30], who designed a quantum query algorithm to prove that $\widetilde{\deg}(\mathsf{AND}_M \circ \mathsf{OR}_N) = O(\sqrt{MN})$. Later, Sherstov [55] proved the following more general result.

Theorem 2 (Sherstov [55])**.** *For* any *Boolean functions* f, g*, we have* $\widetilde{\deg}(f \circ g) = O(\widetilde{\deg}(f) \cdot \widetilde{\deg}(g))$.

Sherstov's remarkable proof of Theorem 2 is via a technique we call *robustification*. This approximation technique will be an important source of intuition for our new results.

Robustification. Sherstov [55] showed that for any polynomial $p\colon \{-1,1\}^M \to \{-1,1\}$, and every $\delta > 0$, there is a polynomial p_{robust} of degree $O(\deg(p) + \log(1/\delta))$ that is robust to noise in the sense that $|p(y) - p_{\text{robust}}(y + \mathbf{e})| < \delta$ for all $y \in \{-1,1\}^M$, and $\mathbf{e} \in [-1/3, 1/3]^M$. Hence, given functions f_M, g_N, one can obtain an $(\varepsilon + \delta)$-approximating polynomial for the block composition $f_M \circ g_N$ as follows. Let p be an ε-approximating polynomial for f_M, and q a $(1/3)$-approximating polynomial for g_N. Then the block composition $p^* := p_{\text{robust}}(q, \dots, q)$ is an $(\varepsilon + \delta)$- approximating polynomial for $f_M \circ g_N$. Notice that the degree of p^* is at most the product of the degrees of p_{robust} and q.

Sherstov [52] and the authors [17] independently used the method of dual polynomials to obtain the matching $\Omega(\sqrt{M \cdot N})$ lower bound of Theorem 1. These lower bound proofs work by constructing (explicitly in [17] and implicitly in [52]) an optimal dual polynomial $\psi_{\text{AND-OR}}$ for the AND-OR tree. Specifically, $\psi_{\text{AND-OR}}$ is obtained by taking dual polynomials $\psi_{\text{AND}}, \psi_{\text{OR}}$ respectively witnessing the fact that $\widetilde{\deg}(\mathsf{AND}_M) = \Omega(\sqrt{M})$ and $\widetilde{\deg}(\mathsf{OR}_N) = \Omega(\sqrt{N})$, and combining them in a precise manner.

For arbitrary Boolean functions f and g, this method of combining dual polynomials ψ_f and ψ_g to obtain a dual polynomial $\psi_f \star \psi_g$ for $f \circ g$ was introduced in earlier line of work by Shi and Zhu [59], Lee [36] and Sherstov [54]. Specifically, writing $x = (x_1, \dots, x_M) \in \left(\{-1,1\}^N\right)^M$,

$$(\psi_f \star \psi_g)(x) := 2^M \cdot \psi_f(\dots, \operatorname{sgn}(\psi_g(x_i)), \dots) \cdot \prod_{i=1}^{M} |\psi_g(x_i)|.$$

This technique of combining dual witnesses, which we call the "dual block" method, will also be central to this paper. The lower bound of [17, 52] was obtained by refining the analysis of $\psi_f \star \psi_g$ from [54] in the case where $f = \mathsf{AND}_M$ and $g = \mathsf{OR}_N$.

As argued in subsequent work of Thaler [64, Section 1.2.4], the combining method $\psi_f \star \psi_g$ is specifically tailored to showing optimality of the polynomial approximation p^* for $f \circ g$ obtained via robustification. This assertion can be made precise via complementary slackness: the dual solution $\psi_f \star \psi_g$ can be shown to obey complementary slackness in an approximate (yet precise) sense with respect to the solution to the primal linear program corresponding to p^*.

1.2.1 Additional Prior Work

The method of dual polynomials has recently been used to establish a number of new lower bounds for approximate degree [15, 18, 24, 40, 54, 56, 64]. All of these results focus on block composed functions, and can be viewed as *hardness amplification* results. Specifically, they show that the block composition $f \circ g$ is strictly harder to approximate by low-degree polynomials (requiring either higher degree or higher error) than either f or g individually. These results have enabled progress on a number of open questions regarding the complexity of AC^0, as well as oracle separations involving the polynomial hierarchy and various notions of statistical zero-knowledge proofs.

Recently, a handful of works have proved stronger hardness amplification results for approximate degree by moving beyond block composed functions [19, 42]. These papers use very different techniques than the ones we introduce in this work, as they are focused on a different form of hardness amplification for polynomial approximation (specifically, they amplify approximation error instead of degree).

1.3 Our Results and Techniques

A major technical hurdle to progress on Problem 1 is the need to go beyond the block composed functions that were the focus of prior work. Specifically, Theorem 2 implies that the approximate degree of $f_M \circ g_N$ (viewed as a function of the number of inputs $M \cdot N$) is *never* higher than the approximate degree of f_M or g_N individually (viewed as a function of M and N respectively). For example, if f_M and g_N both have approximate degree equal to the square root of the number of inputs (i.e., $\widetilde{\deg}(f_M) = O(\sqrt{M})$ and $\widetilde{\deg}(g_N) = O(\sqrt{N})$), then the block composition $f_M \circ g_N$ has the same property (i.e., $\widetilde{\deg}(f_M \circ g_N) = O(\sqrt{M \cdot N})$). Our results introduce an analysis of non-block-composed functions that overcomes this hurdle.

Quantitatively, our main lower bounds for constant-depth circuits and DNFs are as follows. To obtain the tightest possible results for a given circuit depth, our analysis pays close attention to whether a circuit $\mathcal{C}$ is monotone ($\mathcal{C}$ is said to be monotone if it contains no NOT gates).

Theorem 3. *Let $k \geq 1$ be any constant integer. Then there is an (explicitly given) monotone circuit on $n \cdot \log^{4k-4}(n)$ variables of depth $2k$, with AND gates at the bottom, which computes a function with approximate degree $\Omega(n^{1-2^{k-1}/3^k} \cdot \log^{3-2^{k+2}/3^k}(n))$.*

For example, Theorem 3 implies a Boolean circuit of depth 6 on n variables with approximate degree $\tilde{\Omega}(n^{23/27}) = \tilde{\Omega}(n^{0.851\dots})$.

Theorem 4. *Let $k \geq 1$ be any constant integer. Then there is an (explicitly given) monotone DNF on $n \cdot \log^{4k-4}(n)$ variables of width $O\left(\log^{2k-1}(n)\right)$ (and size $2^{O(\log^{2k}(n))}$) which computes a function with approximate degree $\Omega(n^{1-2^{k-1}/3^k} \cdot \log^{3-2^{k+2}/3^k}(n))$.*

Theorems 3 and 4 are in fact corollaries of a more general hardness amplification theorem. This result shows how to take any Boolean function f and transform it into a related function g on roughly the same number of variables that has significantly higher approximate degree (unless the approximate degree of f is already $\tilde{\Omega}(n)$). Moreover, if f is computed by a low-depth circuit, then g is as well.

Theorem 5. *Let $f \colon \{-1,1\}^n \to \{-1,1\}$ with $\widetilde{\deg}(f) = d$. Then f can be transformed into a related function $g \colon \{-1,1\}^m \to \{-1,1\}$ with $m = O(n \log^4 n)$ and $\widetilde{\deg}(g) = \Omega(n^{1/3} \cdot d^{2/3} \cdot \log n)$. Moreover, g satisfies the following additional properties.*

- *If f is computed by a circuit of depth k, then g is computed by a circuit of depth $k+3$.* (1)
- *If f is computed by a monotone circuit of depth k with AND gates at the bottom, then g is computed by a monotone circuit of depth $k+2$ with AND gates at the bottom.* (2)
- *If f is computed by monotone DNF of width w, then g is computed by monotone DNF of width $O(w \cdot \log^2 n)$.* (3)

1.3.1 Hardness Amplification Construction

The goal of this subsection is to convey the main ideas underlying the transformation of f into the harder-to-approximate function g in the statement of Theorem 5. We focus on illustrating these ideas when we start with the function $f = \mathsf{AND}_R$, where we assume for simplicity that R is a power of 2. Let $n = N \log R$ for a parameter N to be determined later.[3] Consider the function

$$\mathsf{SURJECTIVITY} \colon \{-1,1\}^n \to \{-1,1\}$$

($\mathsf{SURJ}_{N,R}$ for short) defined as follows. $\mathsf{SURJ}_{N,R}$ interprets its input s as a list of N numbers $(s_1, \dots, s_N)$ from a range $[R]$. The function $\mathsf{SURJ}_{N,R}(s) = -1$ if and only if every element of the range $[R]$ appears at least once in the list.[4]

When we apply Theorem 5 to $f = \mathsf{AND}_R$, the harder function g we construct is precisely $\mathsf{SURJ}_{N,R}$ (for a suitable choice of $N \leq \tilde{O}(R)$). Before describing our transformation for general f, we provide some intuition for why $\mathsf{SURJ}_{N,R}$ is harder to approximate than AND_R.

[3] All logarithms in this paper are taken in base 2.

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

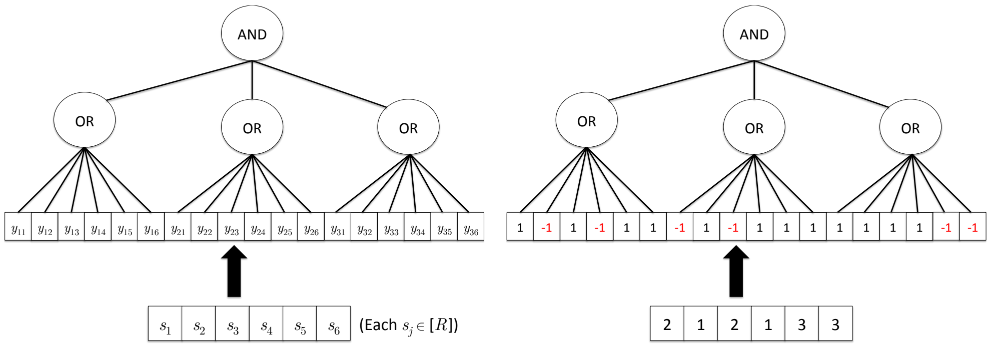

Figure 1: Depiction of Equation (4) when $N = 6$ and $R = 3$.

Getting to Know SURJECTIVITY. It is known that $\widetilde{\deg}(\mathsf{SURJ}_{N,R}) = \tilde{\Omega}(n^{2/3})$ when $R = N/2$ [4]. We do not improve this lower bound for $\mathsf{SURJ}_{N,R}$, but we give a much more general and intuitive proof for it. The best known upper bound on $\widetilde{\deg}(\mathsf{SURJ}_{N,R})$ is the trivial $O(n)$ that holds for any function on n variables.

Although this upper bound is trivial, the following is an instructive way to achieve it. For $(i, j) \in [R] \times [N]$, let[5]

$$y_{ij}(s) = \begin{cases} -1 \text{ if } s_j = i \\ 1 \text{ otherwise.} \end{cases}$$

Observe that $y_{ij}(s)$ is exactly computed by a polynomial in s of degree at most $\log R$, as $y_{ij}(s)$ depends on only $\log R$ bits of s. For brevity, we will typically denote $y_{ij}(s)$ by y_{ij}, but the reader should always bear in mind that y_{ij} is a function of s.

Clearly, it holds that:

$$\mathsf{SURJ}_{N,R}(s) = \mathsf{AND}_R(\mathsf{OR}_N(y_{1,1}, \ldots, y_{1,N}), \ldots, \mathsf{OR}_N(y_{R,1}, \ldots, y_{R,N})). \tag{4}$$

Equality (4) is depicted in Figure 1 in the special case $N = 6$, $R = 3$. Let p^* be the polynomial approximation of degree $O(\sqrt{R \cdot N})$ for the block composed function $\mathsf{AND}_R \circ \mathsf{OR}_N$ obtained via robustification (cf. Section 1.2). Then

$$p^*(y_{1,1}, \ldots, y_{1,N}, \ldots, y_{R,1}, \ldots, y_{R,N})$$

approximates $\mathsf{SURJ}_{N,R}$, and has degree $O(\deg(p^*) \cdot \log R)$. If $N = O(R)$, then this degree bound is $O(N \log R) = O(n)$.

Our analysis in the proof of Theorem 5 is tailored to showing a sense in which this robustification-based approximation method is nearly optimal. Unsurprisingly, our analysis makes heavy use of the dual block method of combining dual witnesses [36, 54, 59], as this method is tailored to showing optimality of robustification-based approximations (cf. Section 1.2). However, there are several technical challenges to overcome, owing to the fact that Equation (4) does not express SURJ as a genuine block composition (since a single bit of the input $s \in \{-1, 1\}^{N \cdot \log R}$ affects R of the variables y_{ij}).

[5]We clarify that earlier work [5] using similar notation reverses the roles of i and j in the definition of $y_{ij}(s)$. We depart from the convention of earlier work because it simplifies the expression of the harder function g exhibited in Theorem 5.

The Transformation for General Functions. Recall from the preceding discussion that when applying our hardness-amplifying transformation to the function $f = \mathsf{AND}_R$, the harder function (on $n = N \cdot \log R$ bits, for some $N = \tilde{O}(R)$) takes the form $\mathsf{SURJ}_{N,R} = \mathsf{AND}_R(\mathsf{OR}_N(y_{1,1}, \dots, y_{1,N}), \dots, \mathsf{OR}_N(y_{R,1} \dots, y_{R,N}))$. This suggests that for general functions $f \colon \{-1,1\}^R \to \{-1,1\}$, one should consider the transformed function

$$F(s) := f(\mathsf{OR}_N(y_{1,1}, \dots, y_{1,N}), \dots, \mathsf{OR}_N(y_{R,1}, \dots, y_{R,N})).$$

Unfortunately, this simple candidate fails spectacularly. Consider the particular case where $f = \mathsf{OR}_R$. It is easy to see that

$$\mathsf{OR}_R(\mathsf{OR}_N(y_{1,1}, \dots, y_{1,N}), \dots, \mathsf{OR}_N(y_{R,1}, \dots, y_{R,N}))$$

evaluates to -1 on *all* inputs $s \in \{-1,1\}^{N \cdot \log R}$. Hence, it has (exact) degree equal to 0.

Fortunately, we are able to show that a modification of the above candidate does work for general functions f_R. Let $R' = R \log R$. Still simplifying, but only slightly, the harder function that we exhibit is $g \colon \{-1,1\}^{N \cdot \log(R')} \to \{-1,1\}$ defined via:

$$g(s) = (f \circ \mathsf{AND}_{\log R})(\mathsf{OR}_N(y_{1,1}, \dots, y_{1,N}), \dots, \mathsf{OR}_N(y_{R',1}, \dots, y_{R',N})).$$

1.3.2 Hardness Amplification Analysis

For expository purposes, we again describe the main ideas of our analysis in the case where $f = \mathsf{AND}_R$. Recall that in this case, the harder function g exhibited in Theorem 5 is $\mathsf{SURJ}_{N,R}$ on $n = N \cdot \log R$ bits. Moreover, in order to approximate $\mathsf{SURJ}_{N,R}$, it is *sufficient* to approximate the *block composed function* $\mathsf{AND}_R \circ \mathsf{OR}_N$. This can be done by a polynomial of degree $O(\sqrt{R \cdot N})$ using robustification.

The goal of our analysis is to show that there is a sense in which this approximation method for $\mathsf{SURJ}_{N,R}$ is essentially optimal. Quantitatively, our analysis yields an $\Omega(R^{2/3})$ lower bound on the approximate degree of $\mathsf{SURJ}_{N,R}$.

At a high level, our analysis proceeds in two stages. In the first stage (Section 3), we give a reduction showing that to approximate $\mathsf{SURJ}_{N,R}(x)$, it is *necessary* to approximate $\mathsf{AND}_R \circ \mathsf{OR}_N$, under the promise that the input has Hamming weight *at most* N. This reduction is somewhat subtle, but conceptually crucial to our results. Nevertheless, at the technical level, it is a straightforward application of a symmetrization argument due to Ambainis [5].

In the second stage (Section 4), we prove that approximating $\mathsf{AND}_R \circ \mathsf{OR}_N$ under the above promise requires degree $\Omega(R^{2/3})$. Executing this second stage is the more technically involved part of our proof, and we devote the remainder of this informal overview to it. Specifically, for some $N = \tilde{O}(R)$, we must construct a dual polynomial $\psi_{\text{AND-OR}}$ witnessing the fact that $\widetilde{\deg}(\mathsf{AND}_R \circ \mathsf{OR}_N) = \Omega(R^{2/3})$, such that $\psi_{\text{AND-OR}}$ is supported exclusively on inputs of Hamming weight at most N.

As a first attempt, one could consider the dual polynomial $\psi_{\text{AND}} \star \psi_{\text{OR}}$ (cf. Section 1.2) used in our prior work [17] to lower bound the approximate degree of the AND-OR tree. Unfortunately, this dual polynomial has inputs of Hamming weight as large as $\Omega(R \cdot N)$ in its support.

Our strategy for handling this issue is to modify $\psi_{\text{AND}} \star \psi_{\text{OR}}$ by post-processing it to zero out all of the mass it places on inputs of Hamming weight more than N. This must be done without significantly affecting its pure high degree, its ℓ_1-norm, or its correlation with $\mathsf{AND}_R \circ \mathsf{OR}_N$. In more detail, let $|y|$ denote the Hamming weight of an input $y \in \{-1,1\}^{R \cdot N}$, and suppose that we can show

$$\sum_{|y|>N} |(\psi_{\text{AND}} \star \psi_{\text{OR}})(y)| \ll R^{-D}. \tag{5}$$

Intuitively, if Inequality (5) holds for a large value of D, then inputs of Hamming weight greater than N are not very important to the dual witness $\psi_{\text{AND}} \star \psi_{\text{OR}}$, and hence it is plausible that the lower bound witnessed by $\psi_{\text{AND}} \star \psi_{\text{OR}}$ holds even if such inputs are ignored completely.

To make the above intuition precise, we use a result of Razborov and Sherstov [44] to establish that Inequality (5) implies the existence of a (explicitly given) function $\psi_{\text{corr}} \colon \{-1,1\}^{N\cdot R} \to \{-1,1\}$ such that:

- $\psi_{\text{corr}}(y) = \psi_{\text{AND}} \star \psi_{\text{OR}}(y)$ for all $|y| > N$,
- ψ_{corr} has pure high degree D, and
- $\sum_{|y|>N} |\psi_{\text{corr}}(y)| \ll R^{-D}$.

Let $\psi_{\text{AND-OR}} = C \cdot (\psi_{\text{AND}} \star \psi_{\text{OR}} - \psi_{\text{corr}})$, where $C \geq 1 - o(1)$ is chosen so that the resulting function has ℓ_1-norm equal to 1. Then $\psi_{\text{AND-OR}}$ has:

1. Pure high degree $\min\{D, \sqrt{R \cdot N}\}$,
2. The same correlation, up to a factor of $1 - o(1)$, as $\psi_{\text{AND}} \star \psi_{\text{OR}}$ has with $\mathsf{AND}_R \circ \mathsf{OR}_N$, and
3. Support restricted to inputs of Hamming weight at most N.

Hence, Step 2 of the proof is complete if we can show that Inequality (5) holds for $D = \Omega(R^{2/3})$. Unfortunately, Inequality (5) does *not* hold unless we modify the dual witness ψ_{OR} to satisfy additional properties. First, we modify ψ_{OR} so that

$$\psi_{\text{OR}} \text{ is supported only on inputs of Hamming weight at most } R^{1/3}. \tag{6}$$

Moreover, we further ensure that ψ_{OR} is biased toward inputs of low Hamming weight in the sense that

$$\text{For all } t \geq 0, \ \sum_{|x|=t} |\psi_{\text{OR}}(x)| \lesssim 1/(t+1)^2. \tag{7}$$

We can guarantee that both Conditions (6) and (7) hold while still ensuring that ψ_{OR} has pure high degree $\Omega(R^{1/6})$, as well as the same ℓ_1-norm and correlation with OR_N. (The fact that this modified dual polynomial ψ_{OR} has pure high degree $\Omega(R^{1/6})$ rather than $\Omega(R^{1/2})$ is the reason we are only able to establish an $\Omega(R^{2/3})$ lower bound on the approximate degree of $\mathsf{SURJ}_{N,R}$, rather than $\Omega(R)$.)

We now explain why these modifications imply that Inequality (5) holds for $D = \Omega(R^{2/3})$. Recall that

$$(\psi_{\text{AND}} \star \psi_{\text{OR}})(y_1, \dots, y_R) = 2^R \cdot \psi_{\text{AND}}(\dots, \operatorname{sgn}(\psi_{\text{OR}}(y_i)), \dots) \cdot \prod_{i=1}^{R} |\psi_{\text{OR}}(y_i)|.$$

For intuition, let us focus on the final factor in this expression, $\prod_{i=1}^R |\psi_{\text{OR}}(y_i)|$. Since ψ_{OR} has ℓ_1-norm equal to 1, the function $|\psi_{\text{OR}}|$ is a probability distribution, and $\prod_{i=1}^R |\psi_{\text{OR}}(y_i)|$ is a product distribution over $\left(\{-1,1\}^N\right)^R$. At a high level, our analysis shows that this product distribution is "exponentially more biased" toward inputs of low Hamming weight than is ψ_{OR} itself.

More specifically, Conditions (6) and (7) together imply that, if $y = (y_1, \dots, y_R) \in \{-1,1\}^{N\cdot R}$ is drawn from the product distribution $\prod_{i=1}^R |\psi_{\text{OR}}(y_i)|$, then the probability that y has Hamming weight more than $N = \tilde{O}(R)$ is dominated by the probability that roughly $R^{2/3}$ of the y_i's each have Hamming weight close to $R^{1/3}$ (and the remaining y_i's have low Hamming weight). But then Condition (7) ensures that the probability that this occurs is at most $R^{-\Omega(R^{2/3})}$.

1.4 Paper Organization

Section 2 covers technical preliminaries. Stage 1 of the proof of our main hardness amplification theorem, Theorem 5, is completed in Section 3. In Section 4, we execute Stage 2 of the proof of Theorem 5, and use it to establish Theorems 3 and 4 from the introduction. Finally, Section 5 describes applications of our results to complexity theory and cryptography.

2 Preliminaries

We begin by formally defining the notion of approximate degree of any partial function defined on a subset of $\mathbb{R}^n$. Throughout, for any subset $\mathcal{X} \subseteq \mathbb{R}^n$ and polynomial $p\colon \mathcal{X} \to \mathbb{R}$, we use $\deg(p)$ to denote the total degree of p, and refer to this without qualification as the degree of p.

Definition 6. *Let $\mathcal{X} \subseteq \mathbb{R}^n$, and let $f : \mathcal{X} \to \{-1, 1\}$. The ε-*approximate degree *of f, denoted $\widetilde{\deg}_\epsilon(f)$, is the least degree of a real polynomial $p\colon \mathbb{R}^n \to \mathbb{R}$ such that $|p(x) - f(x)| \le \epsilon$ for all $x \in \mathcal{X}$. We refer to such a p as an ε-*approximating polynomial *for f. We use $\widetilde{\deg}(f)$ to denote $\widetilde{\deg}_{1/3}(f)$.*

We highlight two slightly non-standard aspects of Definition 6. The first is that it considers subsets of $\mathbb{R}^n$ rather than $\{-1, 1\}^n$. This level of generality has been considered in some prior works [5, 21, 54], and we will require it in our proof of Theorem 5 (cf. Section 3). Second, our definition of an ε-approximating polynomial p for f above does not place any restriction on $p(x)$ for x outside of the domain of f. This is in contrast to some other works (e.g. [15, 20, 28, 53]) that do require $p(x)$ to be bounded for some inputs x outside of the domain of f. Our definition is the most natural and convenient for the purposes of our analyses.

Strong LP duality implies the following characterization of approximate degree (see, e.g., [50]).

Theorem 7. *Let $\mathcal{X}$ be a finite subset of $\mathbb{R}^n$, and let $f : \mathcal{X} \to \{-1, 1\}$. Then $\widetilde{\deg}_\varepsilon(f) \ge d$ if and only if there exists a function $\psi\colon \mathcal{X} \to \mathbb{R}$ satisfying the following properties.*

$$\sum_{x \in \mathcal{X}} \psi(x) \cdot f(x) > \varepsilon, \tag{8}$$

$$\sum_{x \in \mathcal{X}} |\psi(x)| = 1, and \tag{9}$$

$$\textit{For every polynomial } p\colon \mathcal{X} \to \mathbb{R} \textit{ of degree less than } d, \sum_{x \in \mathcal{X}} p(x) \cdot \psi(x) = 0. \tag{10}$$

For functions $\psi_1 : \mathcal{X} \to \mathbb{R}$ and $\psi_2 : \mathcal{X}' \to \mathbb{R}$ defined on finite domains $\mathcal{X}, \mathcal{X}'$ with $\mathcal{X} \subseteq \mathcal{X}'$, we define

$$\langle \psi_1, \psi_2 \rangle := \sum_{x \in \mathcal{X}} \psi_1(x) \cdot \psi_2(x),$$

and we refer to this as the correlation of ψ_1 with ψ_2. (We define $\langle \psi_1, \psi_2 \rangle$ similarly if instead $\mathcal{X}' \subseteq \mathcal{X}$.) An equivalent way to define $\langle \psi_1, \psi_2 \rangle$ is to first extend the domain of ψ_1 to $\mathcal{X}'$ by setting $\psi_1(x) = 0$ for all $x \in \mathcal{X}' \setminus \mathcal{X}$, and then define

$$\langle \psi_1, \psi_2 \rangle := \sum_{x \in \mathcal{X}'} \psi_1(x) \cdot \psi_2(x).$$

We refer to the right hand side of Equation (9) as the ℓ_1-norm of ψ, and denote this quantity by $\|\psi\|_1$. If ψ satisfies Equation (10), it is said to have *pure high degree* at least d.

Additional Notation. For an input $x \in \{-1,1\}^n$, we use $|x|$ to denote the Hamming weight of x, i.e., $|x| := \sum_{i=1}^n (1-x_i)/2$. Let $\{-1,1\}^N_{\leq k} := \{x \in \{-1,1\}^N : |x| \leq k\}$. We denote the set $\{1,\dots,N\}$ by $[N]$ and the set $\{0,\dots,N\}$ by $[N]_0$. Given $t \in \mathbb{R}$, we define $\mathrm{sgn}(t)$ to equal 1 if $t > 0$ and to equal -1 otherwise. The function $\mathbf{1}_N \colon \{-1,1\}^N \to \{-1,1\}$ denotes the constant function that always evaluates to 1. We denote by 1^N the N-dimensional vector with all entries equal to 1.

Minsky-Papert Symmetrization. The following well-known lemma is due to Minsky and Papert [38].

Lemma 8. *Let $p\colon \{-1,1\}^n \to \{-1,1\}$ be an arbitrary polynomial. Then there is a univariate polynomial $q\colon \mathbb{R} \to \mathbb{R}$ of degree at most $\deg(p)$ such that*

$$q(t) = \frac{1}{\binom{n}{t}} \sum_{x \in \{-1,1\}^n \colon |x|=t} p(x)$$

for all $t \in [n]_0$.

2.1 The Dual Block Method

This section collects definitions and preliminary results on the dual block method [36,54,59] for constructing dual witnesses for a block composed function $F \circ f$ by combining dual witnesses for F and f respectively.

Definition 9. *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*

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

Proposition 10. *The dual block composition satisfies the following properties:*

***Preservation of ℓ_1-norm*:** *If $\|\Psi\|_1 = 1$ and $\|\psi\|_1 = 1$, then*

$$\|\Psi \star \psi\|_1 = 1. \tag{11}$$

***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 $q\colon \{-1,1\}^{m\cdot M} \to \{-1,1\}$,*

$$\deg q < D \cdot d \implies \langle \Psi \star \psi, q\rangle = 0. \tag{12}$$

***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*

$$(\zeta \star \varphi) \star \psi = \zeta \star (\varphi \star \psi). \tag{13}$$

Proof. Properties (11) and (12) appear in [54, Proof of Theorem 3.3]. Proving that Property (13) holds is a straightforward if tedious calculation that we now perform. Below, we will write an input

$$x = (x_1, \ldots, x_{m_\zeta}) = ((x_{1,1}, \ldots, x_{1,m_\varphi}), \ldots, (x_{m_\zeta,1}, \ldots, x_{m_\zeta,m_\varphi})),$$

where each $x_{i,j} \in \{-1,1\}^{m_\psi}$. We expand

$$((\zeta \star \varphi) \star \psi)(x_{1,1}, \ldots, x_{m_\zeta,m_\varphi}) = 2^{m_\zeta \cdot m_\varphi} \cdot (\zeta \star \varphi)(\ldots, \operatorname{sgn}(\psi(x_{i,j})), \ldots) \cdot \prod_{i=1}^{m_\zeta} \prod_{j=1}^{m_\varphi} |\psi(x_{i,j})|$$

$$= 2^{m_\zeta \cdot m_\varphi} \cdot \left(2^{m_\zeta} \cdot \zeta(\ldots, \operatorname{sgn}(\varphi(\ldots, \operatorname{sgn}(\psi(x_{i,j})), \ldots)), \ldots) \cdot \prod_{i=1}^{m_\zeta} |\varphi(\ldots, \operatorname{sgn}(\psi(x_{i,j})), \ldots)| \right) \cdot \prod_{i=1}^{m_\zeta} \prod_{j=1}^{m_\varphi} |\psi(x_{i,j})|$$

$$= 2^{m_\zeta} \cdot \zeta(\ldots, \operatorname{sgn}(\varphi(\ldots, \operatorname{sgn}(\psi(x_{i,j})), \ldots), \ldots) \cdot \prod_{i=1}^{m_\zeta} \left(2^{m_\varphi} \cdot |\varphi(\ldots, \operatorname{sgn}(\psi(x_{i,j})), \ldots)| \prod_{j=1}^{m_\varphi} |\psi(x_{i,j})| \right)$$

$$= 2^{m_\zeta} \cdot \zeta(\ldots, \operatorname{sgn}((\varphi \star \psi)(x_i)), \ldots) \cdot \prod_{i=1}^{m_\zeta} |(\varphi \star \psi)(x_i)|$$

$$= (\zeta \star (\varphi \star \psi))(x_1, \ldots, x_{m_\zeta}).$$

□

The following proposition identifies conditions under which a dual witness ψ for the large $(1/3)$-approximate degree of a function f can be transformed, via dual block composition with a certain function $\Psi \colon \{-1,1\}^M \to \{-1,1\}$, into a dual witness for the large $(1 - 2^{-\Omega(M)})$-approximate degree of the block composition $\mathsf{AND}_M \circ f$.

Proposition 11 (Bun and Thaler [18])**.** *Let $m, M \in \mathbb{N}$. There exists a function $\Psi : \{-1,1\}^M \to \mathbb{R}$ with the following properties. Let $f : \{-1,1\}^m \to \{-1,1\}$ be any function. Let $\psi : \{-1,1\}^m \to \mathbb{R}$ be any function such that $\langle \psi, f \rangle \geq 1/3$, $\|\psi\|_1 = 1$, and $\psi(x) \geq 0$ whenever $f(x) = 1$. Then*

$$\langle \Psi \star \psi, \mathsf{AND}_M \circ f \rangle \geq 1 - (2/3)^M, \tag{14}$$

$$\|\Psi \star \psi\|_1 = 1, \tag{15}$$

$$\langle \Psi, \mathbf{1}_M \rangle = 0. \tag{16}$$

The following proposition roughly states that if ψ and Ψ are dual polynomials that are well-correlated with f and F respectively, then the dual block composition $\Psi \star \psi$ is well-correlated with the block composed function $F \circ f$. There is, however, a potential loss in correlation that is proportional to the number of variables on which F is defined.

Proposition 12 (Sherstov [54])**.** *Let $f : \{-1,1\}^m \to \{-1,1\}$ and $F \colon \{-1,1\}^M \to \{-1,1\}$, and let $\varepsilon, \delta > 0$. Let $\psi \colon \{-1,1\}^m \to \{-1,1\}$ be a function with $\|\psi\|_1 = 1$ and $\langle \psi, f \rangle \geq 1 - \delta$. Let $\Psi \colon \{-1,1\}^M \to \{-1,1\}$ be a function with $\|\Psi\|_1 = 1$ and $\langle \Psi, F \rangle \geq \varepsilon$. Then*

$$\langle \Psi \star \psi, F \circ f \rangle \geq \varepsilon - 4M\delta.$$

3 Connecting Symmetric Properties and Block Composed Functions

In this section, we execute Stage 1 of our program for proving our main hardness amplification theorem, Theorem 5. Throughout this entire section, fix an arbitrary function $F_R : \{-1,1\}^R \to \{-1,1\}$. (In order to prove Theorem 5, we will ultimately set $R = 10 \cdot n \cdot \log n$, and take $F_R = f \circ \mathsf{AND}_{10 \log n}$ for $f : \{-1,1\}^n \to \{-1,1\}$.)

Our analysis relies on several intermediate functions, which we now define and analyze. All of these functions are variants of the function $F_R \circ \mathsf{OR}_N$.

3.1 The First Function: Block Composition Under a Promise

We define a promise variant of the function $F_R \circ \mathsf{OR}_N$ as follows.

Definition 13. *Fix positive numbers N and R. Recall that $\{-1,1\}^{N \cdot R}_{\leq N}$ denotes the subset of $\{-1,1\}^{N \cdot R}$ consisting of vectors of Hamming weight at most N. Define $G^{\leq N}$ to be the partial function obtained from $F_R \circ \mathsf{OR}_N$ by restricting its domain to $\{-1,1\}^{N \cdot R}_{\leq N}$.*

Our goal is to reduce establishing Theorem 5 to establishing a lower bound on the approximate degree of $G^{\leq N}$. Specifically, we prove the following theorem relating the approximate degree of $G^{\leq N}$ to that of a function g which is not much more complex than F_R:

Theorem 14. *Let $G^{\leq N} \colon \{-1,1\}^{N \cdot R}_{\leq N} \to \{-1,1\}$ be as in Definition 13. There exists a function $g \colon \{-1,1\}^{12 \cdot N \cdot \lceil \log(R+1) \rceil} \to \{-1,1\}$ such that*

$$\widetilde{\deg}_\varepsilon(g) \geq \widetilde{\deg}_\varepsilon(G^{\leq N}) \cdot \lceil \log(R+1) \rceil. \tag{17}$$

Moreover:

- *If F_R is computed by a circuit of depth k, then g is computed by a circuit of depth $k+2$.* (18)

- *If F_R is computed by a monotone circuit of depth k, then g is computed by a monotone circuit of depth $k+2$ with AND gates at the bottom.* (19)

- *If F_R is computed by a monotone DNF of width w, then g is computed by a monotone DNF of width $O(w \cdot \log R)$.* (20)

3.2 The Second Function: A Property of Evaluation Tables

Consider a vector $s = (s_1, \ldots, s_N) \in [R]_0^N$. Observe that s can be thought of as the *evaluation table* of a function $f_s \colon [N] \to [R]_0$ defined via $f_s(i) = s_i$. The second function G^{prop} that we define (cf. Definition 16 below) can be thought of as a *property* of such a function f_s.

In order to define G^{prop}, it is useful to describe such a function f_s as follows.

Definition 15. *Fix any $s \in [R]_0^N$. For $(i,j) \in [R]_0 \times [N]$, define $Y(s) = (\ldots, Y_{ij}(s), \ldots) \in \left(\{-1,1\}^N\right)^{R+1}$ where*

$$Y_{ij}(s) = \begin{cases} -1 \text{ if } s_j = i \\ 1 \text{ otherwise.} \end{cases}$$

Observe that any vector $y = (\dots, y_{ij}, \dots) \in \left(\{-1,1\}^N\right)^{R+1}$ equals $Y(s)$ for some $s \in [N]^{R+1}$ if and only if y satisfies the following condition:

$$\text{For every } j \in [N], \text{ there exists exactly one value of } i \text{ in } [R]_0 \text{ such that } y_{ij} = -1. \tag{21}$$

Accordingly, the domain of our second function G^{prop} is the subset of $\left(\{-1,1\}^N\right)^{R+1}$ satisfying Condition (21).

Definition 16. *Let $\mathcal{D}_{N,R}$ be the subset of $\left(\{-1,1\}^N\right)^{R+1}$ of vectors satisfying Condition* (21). *We refer to any function from $\mathcal{D}_{N,R}$ to $\{-1,1\}$ as a* property *of functions $[N] \to [R]_0$. Define the property $G^{\mathrm{prop}} \colon \mathcal{D}_{N,R} \to \{-1,1\}$ via: $G^{\mathrm{prop}}(y_0, y_1, \dots, y_R) := F_R(\mathsf{OR}_N(y_1), \dots, \mathsf{OR}_N(y_R))$.*

One may view G^{prop} as a property of functions $f_s \colon [N] \to [R]_0$ as follows. The property G^{prop} first obtains a vector of R bits $(b_1, \dots, b_R)$, one for each of the R non-zero range items $1, \dots, R$, and then feeds these bits into F_R. Here, the bit b_i for range item i is obtained by testing whether i appears in the image of f_s (any occurrences of range item 0 are effectively ignored by G^{prop}).

The following lemma establishes that G^{prop} satisfies a basic symmetry condition. This holds regardless of the base function F_R used to define G^{prop}.

Lemma 17. *For a permutation $\sigma : [N] \to [N]$ and a vector $y_i \in \{-1,1\}^N$, let $\sigma(y_i) := (y_{i,\sigma(1)}, \dots, y_{i,\sigma(N)})$. Then $G^{\mathrm{prop}}(y_0, \dots, y_R) = G^{\mathrm{prop}}(\sigma(y_0), \dots, \sigma(y_R))$.*

Proof. Immediate from Definition 16 and the fact that OR_N depends only on the Hamming weight of its input. □

Viewing G^{prop} as a property of functions $f_s \colon [N] \to [R]_0$, Lemma 17 simply states that G^{prop} is invariant under permutations of the domain of f_s.

3.3 The Third Function: A Symmetrized Property

To define our third function $\tilde{G}^{\mathrm{prop}}$, it is useful to consider yet another representation of a function $f_s \colon [N] \to [R]_0$.

Definition 18. *Given $s \in [R]_0^N$, and its associated function f_s, let $Z_i(s) = |f_s^{-1}(i)|$, and define $Z(s) = (Z_0(s), \dots, Z_R(s))$.*

That is, each function $Z_i(s)$ counts the number of of inputs $j \in [N]$ such that $f_s(j) = i$. Observe that a vector $z = (z_0, \dots, z_R) \in [N]_0^{R+1}$ equals $Z(s)$ for some $s \in [R]_0^N$ if and only if

$$z_0 + \cdots + z_R = N. \tag{22}$$

Accordingly, the domain upon which our third function $\tilde{G}^{\mathrm{prop}}$ is defined is the subset of $[N]_0^{R+1}$ satisfying Equation (22).

Definition 19. *Let $\tilde{\mathcal{D}}_{N,R}$ be the subset of $[N]_0^{R+1} \subset \mathbb{R}^{R+1}$ consisting of all vectors $z = (z_0, \dots, z_R)$ satisfying Equation* (22). *Define $\tilde{G}^{\mathrm{prop}} \colon \tilde{\mathcal{D}}_{N,R} \to \{-1,1\}$ as follows. For any $z \in \tilde{\mathcal{D}}_{N,R}$, let s be an arbitrary vector such that $Z(s) = z$. Define $\tilde{G}^{\mathrm{prop}}(z) = G^{\mathrm{prop}}(Y(s))$, where $Y(s)$ is as in Definition 15.*

The function $\tilde{G}^{\mathrm{prop}}$ is well-defined, as Lemma 17 implies that for any pair $s, s' \in [R]_0^N$ such that $Z(s) = Z(s')$, it holds that $G^{\mathrm{prop}}(Y(s)) = G^{\mathrm{prop}}(Y(s'))$. It is straightforward to see that the following is an alternative definition of $\tilde{G}^{\mathrm{prop}}$ on its domain $\tilde{\mathcal{D}}_{N,R}$:

$$\tilde{G}^{\mathrm{prop}}(z_0, \dots, z_R) = F_R(\mathbb{I}_{>0}(z_1), \dots, \mathbb{I}_{>0}(z_R)), \tag{23}$$

where $\mathbb{I}_{>0}(z_i) = 1$ if $z_i = 0$ and is equal to -1 otherwise.

Relating the Approximate Degrees of G^{prop} and $G^{\leq N}$. The following lemma is implicit in the proof of [5, Lemma 3.4]. It states that $\tilde{G}^{\text{prop}}$ is no harder to approximate by low-degree polynomials than is G^{prop}.

Lemma 20 (Ambainis [5]). *Let $G^{\text{prop}} \colon \mathcal{D}_{N,R} \to \{-1,1\}$ be any property of functions $f_s \colon [N] \to [R]_0$ that is symmetric with respect to permutations of the domain of f_s. Let p be a polynomial of degree d that ε-approximates G^{prop} on its domain $\mathcal{D}_{N,R}$. Then there exists a polynomial $\tilde{p} \colon \mathbb{R}^{R+1} \to \mathbb{R}$ of degree at most d that ε-approximates $\tilde{G}^{\text{prop}}$ on its domain $\tilde{\mathcal{D}}_{N,R}$.*

The following theorem is the technical heart of this section.

Theorem 21. *Let $\varepsilon > 0$. Then $\widetilde{\deg}_\varepsilon(G^{\text{prop}}) \geq \widetilde{\deg}_\varepsilon(G^{\leq N})$.*

Proof. Recall that the domain of G^{prop} is the subset $\mathcal{D}_{N,R}$ of $\left(\{-1,1\}^N\right)^{R+1}$ of vectors satisfying Condition (21), and the domain of $G^{\leq N}$ is $\{-1,1\}^{N \cdot R}_{\leq N} = \{x \in (\{-1,1\}^N)^R : |x| \leq N\}$. Let $p \colon \mathcal{D}_{N,R} \to \mathbb{R}$ be a polynomial of degree d that ε-approximates G^{prop}. We will construct a polynomial $q \colon \{-1,1\}^{N \cdot R}_{\leq N} \to \mathbb{R}$ of degree at most d that ε-approximates $G^{\leq N}$.

By Lemma 20, there exists a polynomial $\tilde{p} : \mathbb{R}^{R+1} \to \mathbb{R}$ of degree at most d such that

$$|\tilde{p}(z_0, \ldots, z_R) - \tilde{G}^{\text{prop}}(z_0, \ldots, z_R)| \leq \varepsilon \qquad \text{whenever } (z_0, \ldots z_R) \in \tilde{\mathcal{D}}_{N,R}. \tag{24}$$

For each $i = 0, \ldots, R$, define a function $T_i : \{-1,1\}^{N \cdot R}_{\leq N} \to [N]_0$ by $T_i(x) = |\{j \colon x_{ij} = -1\}| = \frac{1}{2}(N - \sum_{j=1}^N x_{ij})$. Now define the polynomial $q \colon \{-1,1\}^{N \cdot R}_{\leq N} \to \mathbb{R}$ by

$$q(x) = \tilde{p}\left(N - \sum_{i=1}^R T_i(x), T_1(x), \ldots, T_R(x)\right).$$

Since each of the functions T_i is linear, the polynomial q has degree at most d.

We now verify that

$$|q(x) - G^{\leq N}(x)| \leq \varepsilon \text{ for all } x \in \{-1,1\}^{N \cdot R}_{\leq N}. \tag{25}$$

Fix some $x = (x_1, \ldots, x_R) \in \{-1,1\}^{N \cdot R}_{\leq N}$. Then

$$\begin{aligned}
\tilde{G}^{\text{prop}}\left(N - \sum_{i=1}^R T_i(x), T_1(x), \ldots, T_R(x)\right) &= F_R(\mathbb{I}_{>0}(T_1(x)), \ldots, \mathbb{I}_{>0}(T_R(x))) \\
&= F_R(\mathsf{OR}_N(x_1), \ldots, \mathsf{OR}_N(x_R)) \\
&= G^{\leq N}(x).
\end{aligned}$$

Here, the first equality holds by combining Equation (23) with the fact that $\left(N - \sum_{i=1}^R T_i(x), T_1(x), \ldots, T_R(x)\right)$ is a sequence of non-negative numbers summing to N and hence is in the domain $\tilde{\mathcal{D}}_{N,R}$ of $\tilde{G}^{\text{prop}}$. The second equality holds by definition of $\mathbb{I}_{>0}$. The third equality holds by definition of $G^{\leq N}$ and $\{-1,1\}^{N \cdot R}_{\leq N}$.

Property (25) now follows by definition of q and Property (24). $\square$

3.4 The Final Function: From a Property to a Circuit

Recall that our goal in this section is to prove Theorem 14 reducing our main hardness amplification theorem (Theorem 5) to a lower bound on the approximate degree of $G^{\leq N}$. Theorem 14 refers to a total function g on $m = O(n \log^4 n)$ bits. But none of the first three functions defined in this section (i.e., $G^{\leq N}$, G^{prop}, and $\tilde{G}^{\mathrm{prop}}$) are total functions on bits. Hence, we still need to construct a function g with domain $\{-1,1\}^m$, with circuit depth or DNF width not much higher than that of F_R. (Recall that the function g referred to in Theorem 5 will ultimately be obtained in Section 4 by applying the construction here with $R = 10n \log n$, and $F_R = f \circ \mathsf{AND}_{10 \log n}$).

Our function g will interpret its input $u \in \{-1,1\}^m$ as specifying a list s of N numbers from the set $[R]_0$, and will output $G^{\mathrm{prop}}(Y(s))$. There are many ways to translate u into the list s. It turns out that a relatively simple translation method suffices to ensure Property (18) of Theorem 14, i.e., that if F_R is computed by a circuit of depth k, then g is computed by a circuit of depth $k+2$. We will begin by showing how to construct an auxiliary function g^* that is already enough to satisfy Property (18). Slightly more effort will then be required to modify g^* to construct g establishing Properties (19) and (20) of Theorem 14.

3.4.1 Definition of g^*

Definition 22. *Fix positive integers N, R, and k with $k \geq \lceil R+1 \rceil$. Let $m = N \cdot k$, and fix any function $\phi\colon \{-1,1\}^k \to [R]_0$. We associate an input $u = (u_1, \ldots, u_N) \in \left(\{-1,1\}^k\right)^N$ with the vector $s_u \in [R]_0^N$ defined as $s_u = (\phi(u_1), \ldots, \phi(u_N))$. Let $Y : [R]_0^N \to \mathcal{D}_{N,R}$ be as in Definition 15. Given any property $G\colon \mathcal{D}_{N,R} \to \{-1,1\}$, define $G_\phi\colon \{-1,1\}^m \to \{-1,1\}$ by $G_\phi(u) = G(Y(s_u))$.*

The following lemma is a restatement of [46, Theorem 3.2].

Lemma 23 (Sherstov [46])**.** *Let $k = 6\lceil \log(R+1) \rceil$. There exists an (explicitly given) surjection*

$$\phi : \{-1,1\}^k \to [R]_0$$

such that the following holds. For every property $G\colon \mathcal{D}_{N,R} \to \{-1,1\}$ and $\varepsilon > 0$,

$$\widetilde{\deg}_\varepsilon(G_\phi) \geq \widetilde{\deg}_\varepsilon(G) \cdot \lceil \log(R+1) \rceil.$$

The following corollary defines the function g^* and uses Lemma 23 to show that a lower bound on the approximate degree of $G^{\leq N}$ implies a lower bound on the approximate degree of g^*.

Corollary 24. *Fix an integer $N > 0$. Let $G^{\leq N}\colon \{-1,1\}^{N \cdot R}_{\leq N} \to \{-1,1\}$ be as in Definition 13, $G^{\mathrm{prop}}\colon \mathcal{D}_{N,R} \to \{-1,1\}$ be as in Definition 16, and ϕ be as in Lemma 23. Let $m = 6N \cdot \lceil \log(R+1) \rceil$, and define $g^*\colon \{-1,1\}^m \to \{-1,1\}$ to equal G^{prop}_ϕ as per Definition 22. Then for every $\varepsilon > 0$,*

$$\widetilde{\deg}_\varepsilon(g^*) \geq \widetilde{\deg}_\varepsilon(G^{\leq N}) \cdot \lceil \log(R+1) \rceil. \tag{26}$$

Moreover, if F_R is computed by Boolean circuit of depth k and size $\mathrm{poly}(R)$, then g^* is computed by circuit of depth $k+2$ and size $\mathrm{poly}(R, N)$.

Proof. Inequality (26) follows by combining Lemma 23 and Theorem 21.

We now turn to showing that if F_R is computed by a circuit of small depth then g^* is as well. Consider an input $u = (u_1, \ldots, u_N) \in \left(\{-1,1\}^{6\lceil \log(R+1) \rceil}\right)^N$, and recall that we associate u with the vector

$(\phi(u_1), \ldots, \phi(u_N)) \in [R]_0^N$. The output $g(u)$ is obtained by applying F_R to a sequence of bits $(b_1, \ldots, b_R)$, where $b_i = -1$ if and only if there exists a $j \in [N]$ such that $\phi(u_j) = i$.

Since ϕ is a function on just $6\lceil \log(R+1) \rceil$ bits, each bit b_i is computed by a DNF $\mathcal{C}_i$ of width $6\lceil \log(R+1) \rceil$, and hence size at most $O(N \cdot R^{12})$.

Hence, if F_R is computed by a Boolean circuit $\mathcal{C}$ of size $S = \mathrm{poly}(R)$ and depth k, then by replacing each input b_i to $\mathcal{C}$ with the DNF $\mathcal{C}_i$, one obtains a circuit $\mathcal{C}^*$ for g^* of size at most $O(S \cdot N \cdot R^{12}) = \mathrm{poly}(R, N)$ and depth $k + 2$. $\square$

3.4.2 Definition of g

Even if F_R is a DNF of polylogarithmic width, the function g^* defined in Corollary 24 may not be. However, it is not hard to see that if F_R is a *monotone* DNF of polylogarithmic width w, then g^* is a (non-monotone) DNF of width at most $O(w \cdot \log n)$. Indeed, in this case the circuit $\mathcal{C}^*$ for g^* constructed in the proof of Corollary 24 is an $\mathsf{OR} - \mathsf{AND} - \mathsf{OR} - \mathsf{AND}$ circuit *with all negations at the input level*. Each AND gate in the second level from the top has fan-in at most w, and the bottom AND gates each have fan-in at most $w' = 6\lceil \log(R+1) \rceil$. Any such circuit can be transformed into a (non-monotone) DNF of width at most $w \cdot w' = O(w \cdot \log R)$.

Unfortunately, this observation is still not enough for us to eventually obtain our desired $n^{1-\delta}$ lower bounds for polylogarithmic width DNFs (cf. Theorem 4). To obtain such lower bounds, we need to recursively apply our hardness amplification methods, and hence we need the harder function g to itself be a *monotone* DNF. Our definition of g achieves this by applying a simple transformation to g^*. This transformation has appeared in related contexts [25, Proof of Lemma 3].

Definition 25. *Fix $F_R \colon \{-1,1\}^R \to \{-1,1\}$, and let $g^* \colon \{-1,1\}^m \to \{-1,1\}$ be as in Definition 22. Let $\mathcal{C}^*$ be any circuit computing g^* such that all negations in $\mathcal{C}_g$ appear at the inputs. Let $g \colon \{-1,1\}^{2m} \to \{-1,1\}$ be the monotone function defined as follows. Associate each of the first m inputs to g with an input to g^*, and each of the last m inputs to g with the negation of an input to g^*. Then g is obtained from g^* by replacing each literal of $\mathcal{C}^*$ with the corresponding (unnegated) input to g.*

We now complete the proof of Theorem 14.

Proof of Theorem 14. We begin by establishing Expression (17). Let $p \colon \{-1,1\}^{2m} \to \mathbb{R}$ be a degree d polynomial approximating g to error ϵ. Then one can turn p into a polynomial $q \colon \{-1,1\}^m \to \{-1,1\}$ of degree at most d approximating g^* to the same error by simply replacing each input to p with the corresponding input (or its negation) to g^*. It follows that $\widetilde{\deg}_\varepsilon(g) \geq \widetilde{\deg}_\varepsilon(g^*)$. The inequality in Expression (17) follows from Corollary 24.

Property (18) is immediate from Corollary 24, since the construction of Definition 25 does not change the circuit depth of $\mathcal{C}^*$.

The discussion preceding the statement of Definition 25 revealed that, if F_R is a monotone DNF of polylogarithmic width w, then g^* is a (non-monotone) DNF of width at most $O(w \cdot \log R)$. It is then immediate from the definition of g that g is a monotone DNF of width $O(w \cdot \log R)$. This yields Property (20).

By similar reasoning, if F_R is computed by a monotone circuit of depth k, then g^* is computed by a circuit of depth $k+2$ with AND gates at the bottom, and all negations at the inputs. It is then immediate from the definition of g that g is computed by a monotone circuit of depth $k+2$ with AND gates at the bottom. This establishes Property (19), completing the proof. $\square$

4 Analyzing Block Composed Functions On Low Hamming Weight Inputs

To complete the proof of Theorem 5, we combine the following theorem with Theorem 14.

Theorem 26. *Let $f_n : \{-1,1\}^n \to \{-1,1\}$ be any function. Let $N = c \cdot n \log^3 n$ for a sufficiently large constant $c > 0$. Let $G^{\leq N} : \{-1,1\}^{10 \cdot N \cdot n \cdot \log n}_{\leq N} \to \{-1,1\}$ equal $f_n \circ \mathsf{AND}_{10 \log n} \circ \mathsf{OR}_N$ restricted to inputs in $\{-1,1\}^{10 \cdot N \cdot n \cdot \log n}_{\leq N} = \{x \in \{-1,1\}^{10 \cdot N \cdot n \cdot \log n} : |x| \leq N\}$ (cf. Definition 13). Then $\widetilde{\deg}(G^{\leq N}) \geq n^{1/3} \cdot \widetilde{\deg}(f_n)^{2/3}$.*

The primary goal of this section is to prove Theorem 26. Before embarking on this proof, we use it to complete the proofs of Theorems 3-5 from Section 1.3.

Proof of Theorem 5 assuming Theorem 26. We begin by establishing Property (1) in the conclusion of Theorem 5. Let $R = 10 \cdot n \cdot \log n$ and $F_R := f_n \circ \mathsf{AND}_{10 \cdot \log n}$. Applying Corollary 24 to F_R yields a function g on $O(N \log R) = O(n \log^4 n)$ variables satisfying

$$\widetilde{\deg}_\varepsilon(g) \geq \widetilde{\deg}_\varepsilon(G^{\leq N}) \cdot \lceil \log(R+1) \rceil \geq \Omega\left(n^{1/3} \cdot \widetilde{\deg}(f_n)^{2/3} \cdot \log n\right),$$

where the final inequality holds by Theorem 26. Suppose f_n is computed by polynomial size Boolean circuit $\mathcal{C}$ of depth k. Then F_R is computed by polynomial-size Boolean circuit of depth $k+1$, and Property (18) of Theorem 14 guarantees that g is computed by polynomial size Boolean circuit of depth $k+3$. Hence, g satisfies Property (1) as desired.

To establish Property (2), suppose that f_n is computed by a monotone circuit of depth k with AND gates at the bottom. Then F_R is computed by such a circuit as well. Property (19) of Theorem 14 then implies that g is computed by a circuit of depth $k+2$ with AND gates at the bottom.

To establish Property (3), observe that if f_n is computed by a monotone DNF of width w, then F_R is computed by a monotone DNF of width $O(w \cdot \log n)$, and Property (20) of Theorem 14 implies that g is computed by a monotone DNF of width $O(w \cdot \log^2 n)$.

□

Proof of Theorems 3 and 4 assuming Theorem 5. One can almost obtain Theorems 3 and 4 by recursively applying Theorem 5, starting in the base case with the function OR_n. However, to obtain stronger degree lower bounds for a given circuit depth or DNF width, we instead use the following well-known result of Aaronson and Shi [4] regarding the approximate degree of (the negation of) the well-known Element Distinctness function.

Lemma 27 (Sherstov [46], refining Aaronson and Shi [4])**.** *There is a function $\overline{\mathsf{ED}} \colon \{-1,1\}^n \to \{-1,1\}$ such that $\widetilde{\deg}(\overline{\mathsf{ED}}) = \Omega(n^{2/3} \log^{1/3} n)$. Moreover, $\overline{\mathsf{ED}}$ is computed by a monotone DNF of polynomial size and width $O(\log n)$.*

Proof. If the word monotone were omitted from the conclusion, this would be a restatement of [46, Theorem 3.3]. Using the same technique as in Definition 25, the non-monotone DNF constructed in [46, Theorem 3.3] can be transformed into a monotone DNF satisfying the same properties. □

Lemma 27 immediately implies Theorems 3 and 4 in the case $k = 1$. Assume by way of induction that Theorem 3 holds for an integer $k \geq 1$. That is, there exists a function $f^{(k)}$ on $n \cdot \log^{4k-4}(n)$ variables, computed by monotone circuit of depth $2k$ with AND gates at the bottom, with approximate degree $\Omega(n^{1-2^{k-1}/3^k} \cdot \log^{3-2^{k+2}/3^k}(n))$. By applying Theorem 5 to $f^{(k)}$, one obtains (by Property (2)) a function $f^{(k+1)}$ on $n \cdot \log^{4k}(n)$ variables, computed by a monotone circuit of depth $2k + 2$ with AND gates at the bottom, with approximate degree $\Omega(n^{1-2^k/3^{k+1}} \cdot \log^{3-2^{k+3}/3^{k+1}}(n))$. The function $f^{(k+1)}$ satisfies the conclusion of Theorem 3, completing the inductive proof of Theorem 3.

Similarly, assume by way of induction that Theorem 4 holds for an integer $k \geq 1$, for a DNF $f^{(k)}$. By applying Theorem 5 to $f^{(k)}$, one obtains (by Property (3)) a function $f^{(k+1)}$ satisfying the conclusion of Theorem 4 for integer $k + 1$. $\square$

4.1 Organization of the Proof of Theorem 26

Our proof of Theorem 26 entails using a dual witness for the approximate degree of f_n to construct a dual witness for the higher approximate degree of $G^{\leq N}$. For expository purposes, we think about the construction of a dual witness for $G^{\leq N}$ as consisting of four steps.

Step 1. Let $d = \widetilde{\deg}(f_n)$. We begin by constructing a dual witness φ for the $\Omega\left(\sqrt{k}\right)$-approximate degree of the OR_N function when restricted to inputs of Hamming weight at most $k = (n/d)^{2/3}$. This construction closely mirrors previous constructions of Špalek [60] and Bun and Thaler [18]. However, we need φ to satisfy an additional metric condition that is not guaranteed by these prior constructions. Specifically, we require that the total ℓ_1 weight that φ places on the t'th layer of the Hamming cube should be upper bounded by $O(1/(t+1)^2)$.

Step 2. We apply the error amplification construction of Proposition 11 to transform φ into a new dual polynomial ψ that witnesses the fact that the $(1-\delta)$-approximate degree of the function $\mathsf{AND}_{10\log n} \circ \mathsf{OR}_N$ remains $\Omega(\sqrt{k})$, even with error parameter $\delta \leq 1/N^2$.

Step 3. We appeal to the degree amplification construction of Proposition 12 to combine ψ from Step 2 with a dual witness Ψ for the high approximate degree of f_n. This yields a dual witness ζ showing that the approximate degree of the composed function $f_n \circ \mathsf{AND}_{10\log n} \circ \mathsf{OR}_N$ is $\Omega(d \cdot \sqrt{k}) = \Omega(n^{1/3} \cdot d^{2/3})$.

Step 4. Using a construction of Razborov and Sherstov [44], we zero out the mass that ζ places on inputs of Hamming weight larger than N, while maintaining its pure high degree and correlation with $G^{\leq N}$. This yields the final desired dual witness $\hat{\zeta}$ for $G^{\leq N}$.

4.2 Step 1: A Dual Witness for OR_N

Proposition 28. *Let $k, N \in \mathbb{N}$ with $k \leq N$. Then there exist a constant $c_1 \in (0,1)$ and a function $\psi : \{-1,1\}^N_{\leq k} \to \{-1,1\}$ such that:*

$$\langle \psi, \mathsf{OR}_N \rangle \geq 1/3 \tag{27}$$

$$\|\psi\|_1 = 1 \tag{28}$$

$$\text{For any polynomial } p\colon \{-1,1\}^N \to \mathbb{R},\ \deg p < c_1\sqrt{k} \implies \langle \psi, p \rangle = 0 \tag{29}$$

$$\psi(1^N) > 0 \tag{30}$$

$$\sum_{|x|=t} |\psi(x)| \leq 5/(t+1)^2 \qquad \forall t = 0, 1, \ldots, k \tag{31}$$

For intuition, we mention that Properties (27)-(30) amount to a dual formulation of the fact that the "one-sided" approximate degree of OR_N is $\Omega(\sqrt{k})$, even under the promise that the input has Hamming weight at most k.[6] Property (31) is an additional metric condition that we require later in the proof.

The key to proving Proposition 28 is the following explicit construction of a univariate function from first principles. The construction closely follows previous work of Špalek [60] and Bun and Thaler [18], and appears in Appendix A.

Lemma 29. *Let $k \in \mathbb{N}$. There exists a constant $c_1 \in (0,1)$ and a function $\omega : \{0, 1, \ldots, k\} \to \mathbb{R}$ such that*

$$\omega(0) - \sum_{t=1}^{k} \omega(t) \geq 1/3 \tag{32}$$

$$\sum_{t=0}^{k} |\omega(t)| = 1 \tag{33}$$

$$\text{For all univariate polynomials } q\colon \mathbb{R} \to \mathbb{R},\ \deg q < c_1\sqrt{k} \implies \sum_{t=0}^{k} \omega(t) \cdot q(t) = 0 \tag{34}$$

$$\omega(0) > 0 \tag{35}$$

$$\omega(t) \leq 5/(t+1)^2 \qquad \forall t = 0, 1, \ldots, k \tag{36}$$

Proof of Proposition 28. Let ω be the function guaranteed by Lemma 29. Consider the function $\psi : \{-1,1\}^N_{\leq k} \to \{-1,1\}$ defined by

$$\psi(x) = \frac{1}{\binom{N}{|x|}} \cdot \omega(|x|).$$

That ψ satisfies Conditions (27), (28), (30), and (31) is immediate from the definition of ψ and Properties (32), (33), (35), and (36) of ω. Property (29) is a consequence of Minsky-Papert symmetrization. Specifically, for any polynomial $p\colon \{-1,1\}^N \to \mathbb{R}$, Lemma 8 implies that there is a univariate polynomial q of degree at most $\deg(p)$ such that for all $t \in [N]_0$, we have $q(t) = \binom{N}{t}^{-1} \sum_{x \in \{-1,1\}^N :\, |x|=t} p(x)$. Hence, $\sum_{x \in \{-1,1\}^N} \psi(x) \cdot p(x) = \sum_{t=0}^{N} \omega(t) \cdot q(t) = 0$, where the final equality holds by Property (34). $\square$

4.3 Steps 2 and 3: A Preliminary Dual Witness for $G = f_n \circ \mathsf{AND}_{10 \log n} \circ \mathsf{OR}_N$

Recall that our ultimate goal in this section is to construct a dual witness for the veracity of Theorem 26. Here, we begin by defining a preliminary dual witness ζ. While ζ itself is insufficient to witness the veracity of Theorem 26, we will ultimately "post-process" ζ into the desired dual witness $\hat{\zeta}$. We start by fixing choices of several key parameters:

---

[6] One-sided approximate degree is a variant of approximate degree defined in, e.g., [18]. We will not need the primal formulation of approximate degree in this work, and therefore omit a formal definition of this notion.

- $d = \widetilde{\deg}_{2/3}(f_n)$.
- $k = \lfloor (n/d)^{1/3} \rfloor^2$
- $D = c_1\sqrt{k} \cdot d = O(n^{1/3} \cdot d^{2/3})$, where c_1 is the constant from Lemma 29
- $R = 10n \log n$
- $N = \lceil c_2 R \log^2 R \rceil$, where c_2 is a universal constant to be determined later (cf. Proposition 31)
- $m = R \cdot N$

To state our construction of a preliminary dual witness ζ, we begin with the following objects:

- A dual witness $\varphi : \{-1,1\}^n \to \mathbb{R}$ for the fact that $\widetilde{\deg}_{2/3}(f_n) \geq d$. By Theorem 7, φ satisfies the following conditions.

$$\langle \varphi, f_n \rangle \geq 2/3 \tag{37}$$

$$\|\varphi\|_1 = 1 \tag{38}$$

$$\text{For any polynomial } p\colon \{-1,1\}^n \to \mathbb{R},\ \deg p < d \implies \langle \varphi, p \rangle = 0 \tag{39}$$

- The function $\Psi : \{-1,1\}^{10 \log n} \to \mathbb{R}$ whose existence is guaranteed by Proposition 11.
- The dual witness $\psi : \{-1,1\}^N \to \mathbb{R}$ for OR_N guaranteed by Proposition 28, using the choice of the parameter k above.

We apply dual block composition sequentially to the three dual witnesses to obtain a function $\zeta = \varphi \star \Psi \star \psi$. This function is well-defined because dual block composition is associative (Proposition 10).

Proposition 30. *The dual witness $\zeta = \varphi \star \Psi \star \psi$ satisfies the following properties:*

$$\langle \zeta, G \rangle \geq 1/2 \tag{40}$$

$$\|\zeta\|_1 = 1 \tag{41}$$

$$\textit{For all polynomials } p\colon ((\{-1,1\}^N)^{10\log n})^n \to \mathbb{R},\ \deg p < D \implies \langle \zeta, p \rangle = 0. \tag{42}$$

Proof. It is easiest to reason about these properties by regarding ζ as $\varphi \star (\Psi \star \psi)$. To this end, let $\xi : (\{-1,1\}^N)^{10\log n} \to \mathbb{R}$ denote $\Psi \star \psi$. Then ξ satisfies the following properties:

$$\langle \xi, \mathsf{AND}_{10\log n} \circ \mathsf{OR}_N \rangle \geq 1 - \frac{1}{24n} \tag{43}$$

$$\|\xi\|_1 = 1 \tag{44}$$

$$\deg p < c_1\sqrt{k} \implies \langle \xi, p \rangle = 0. \tag{45}$$

Property (43) follows from Expression (14) of Proposition 11, together with Properties (27), (28), and (30) of the dual witness ψ for OR_N. Property (44) follows from Property (11) of dual block composition (cf. Proposition 10), and the fact that both Ψ and ψ have unit ℓ_1-norm (Equations (15) and (28)). Finally, Property (45) is a consequence of Property (12) of dual block composition (cf. Proposition 10), together with Properties (16) and (29), which state that Ψ and ψ have pure high degree at least 1 and $c_1\sqrt{k}$, respectively.

We now verify Properties (40)-(42) of $\zeta = \varphi \star \xi$. Property (40) follows from Proposition 12, together with Properties (37) and (38) of φ and Properties (43) and (44) of ξ. Property (41) follows from Property (11) of dual block composition (cf. Proposition 10), and the fact that both φ and ξ have unit ℓ_1-norm (Equations (38) and (44)). Finally, Property (42) follows from Property (12) (cf. Proposition 10) of dual block composition, together with Properties (39) and (45) of the pure high degrees of φ and ξ, respectively. $\square$

4.4 Step 4: Constructing the Final Dual Witness

For a fixed number $N \in \mathbb{N}$, let $X = \{-1,1\}^{N \cdot 10 \log n \cdot n}_{\leq N} = \{x \in ((\{-1,1\}^N)^{10 \log n})^n : |x| \leq N\}$. Recall that this set X is the same one that appears in Definition 13 when applied to the function $F_R := f_n \circ \mathsf{AND}_{10 \log n}$ on $R = 10n \log n$ variables.

Proposition 31. *Let $\zeta : ((\{-1,1\}^N)^{10 \log n})^n \to \mathbb{R}$ be as constructed in Proposition 30. Then there exists a constant $c_2 > 0$ such that, for $N = \lceil c_2 R \log^2 R \rceil$ and sufficiently large n,*

$$\sum_{x \notin X} |\zeta(x)| \leq (2NR)^{-2R/k} \leq (2NR)^{-2D}. \tag{46}$$

Proof. For the proof of Proposition 31, it is now useful to regard the dual witness ζ as the iterated dual block composition $(\varphi \star \Psi) \star \psi$. In this proof, let us denote $\Phi := \varphi \star \Psi$. Then $\Phi : \{-1,1\}^R \to \mathbb{R}$ where $R = 10n \log n$.

Write ψ as a difference of non-negative functions $\psi_{+1} - \psi_{-1}$. Since ψ has strictly positive pure high degree, it is in particular orthogonal to the constant function $\mathbf{1}_N$, and hence $\|\psi_{+1}\|_1 = \|\psi_{-1}\|_1 = 1/2$. Recalling that $\psi(x) = \omega(|x|)/\binom{N}{|x|}$ where $\omega\colon [k]_0 \to \mathbb{R}$ is given in Lemma 29, we may analogously write $\omega = \omega_{+1} - \omega_{-1}$ where ω_{+1} and ω_{-1} are non-negative functions satisfying

$$\sum_{t=0}^{k} \omega_{+1}(t) = \sum_{t=0}^{k} \omega_{-1}(t) = 1/2. \tag{47}$$

By the definition of dual block composition, we have

$$\zeta(x_1, \ldots, x_R) = 2^R \cdot \Phi(\ldots, \operatorname{sgn}(\psi(x_i)), \ldots) \cdot \prod_{i=1}^{R} |\psi(x_i)|.$$

Consequently,

$$\begin{aligned}
\sum_{x \notin X} |\zeta(x)| &= 2^R \sum_{z \in \{-1,1\}^R} |\Phi(z)| \left(\sum_{\substack{(x_1,\ldots,x_R) \notin X \text{ s.t.} \\ \operatorname{sgn}(\psi(x_1))=z_1,\ldots,\operatorname{sgn}(\psi(x_R))=z_R}} \prod_{i=1}^{R} |\psi(x_i)| \right) \\
&= 2^R \sum_{z \in \{-1,1\}^R} |\Phi(z)| \left(\sum_{(x_1,\ldots,x_R) \notin X} \prod_{i=1}^{R} \psi_{z_i}(x_i) \right) \\
&= 2^R \sum_{z \in \{-1,1\}^R} |\Phi(z)| \left(\sum_{(x_1,\ldots,x_R) \notin X} \prod_{i=1}^{R} \frac{\omega_{z_i}(|x_i|)}{\binom{N}{|x_i|}} \right).
\end{aligned} \tag{48}$$

Observe that for any $(t_1, \dots, t_R) \in [k]_0^R$, the number of inputs $(x_1, \dots, x_R) \in \left(\{-1,1\}^N\right)^R$ such that $|x_i| = t_i$ for all $i \in [R]$ is exactly $\prod_{i=1}^R \binom{N}{t_i}$. Hence, defining

$$P = \{(t_1, \dots, t_R) \in [k]_0^R : t_1 + \cdots + t_R > N\},$$

we may rewrite Expression (48) as

$$2^R \sum_{z \in \{-1,1\}^R} |\Phi(z)| \left(\sum_{(t_1,\dots,t_R) \in P} \prod_{i=1}^R \omega_{z_i}(t_i) \right).$$

To control this quantity, we appeal to the following combinatorial lemma, whose proof we defer to Section 4.5.

Lemma 32. *Let $k, R \in \mathbb{N}$ with $k \leq N$. There is a constant $\alpha > 0$ such that the following holds. Let $N = \lceil \alpha R \log^2 R \rceil$. Let $\eta_i : [k]_0 \to \mathbb{R}$, for $i = 1, \dots R$, be a sequence of non-negative functions where for every i,*

$$\sum_{r=0}^k \eta_i(r) \leq 1/2 \tag{49}$$

$$\eta_i(r) \leq 5/(r+1)^2 \qquad \forall r = 0, 1, \dots, k. \tag{50}$$

Let $P = \{\vec{t} = (t_1, \dots, t_R) \in [k]_0^R : t_1 + \cdots + t_R > N\}$. Then

$$\sum_{\vec{t} \in P} \prod_{i=1}^R \eta_i(t_i) \leq 2^{-R} \cdot (2NR)^{-2R/k}.$$

Observe that the functions ω_{z_i} satisfy Condition (49) (cf. Equation (47)) and Condition (50) (cf. Property (36)). We complete the proof of Proposition 31 by letting c_2 equal the constant α appearing in the statement of Lemma 32, and bounding

$$\begin{aligned} 2^R \sum_{z \in \{-1,1\}^R} |\Phi(z)| \left(\sum_{\vec{t} \in P} \prod_{i=1}^R \omega_{z_i}(t_i) \right) &\leq 2^R \sum_{z \in \{-1,1\}^R} |\Phi(z)| \cdot \left(2^{-R} \cdot (2NR)^{-2R/k} \right) \\ &= (2NR)^{-2R/k} \leq (2NR)^{-2D}. \end{aligned}$$

Here, the equality appeals to the fact that $\|\Phi\|_1 = 1$ (by Property (11) of Proposition 10), and the last inequality holds for sufficiently large n by virtue of the fact that $R/k = \Theta(n^{1/3} d^{2/3} \log n)$, while $D = O(n^{1/3} d^{2/3})$ for the values of R and D specified at the start of Section 4.3. $\square$

We are now in a position to construct our final dual witness for the high approximate degree of $G^{\leq N}$. This dual witness $\hat{\zeta}$ is obtained by modifying ζ to zero out all of the mass it places on inputs of total Hamming weight larger than N. This zeroing process is done in a careful way so as not to decrease the pure high degree of ζ, nor to significantly affect its correlation with $G^{\leq N}$. The technical tool that enables this process is a construction of Razborov and Sherstov [44].

Lemma 33 (cf. [44, Proof of Lemma 3.2])**.** *Let $D, m \in \mathbb{N}$ with $0 \leq D \leq m-1$. Then for every $y \in \{-1,1\}^m$ with $|y| > D$, there exists a function $\phi_y : \{-1,1\}^m \to \mathbb{R}$ such that*

$$\phi_y(y) = 1 \tag{51}$$

$$|x| > D, x \neq y \implies \phi_y(x) = 0 \tag{52}$$

$$\deg p < D \implies \langle \phi_y, p \rangle = 0 \tag{53}$$

$$\sum_{|x| \leq D} |\phi_y(x)| \leq 2^D \binom{|y|}{D}. \tag{54}$$

Proposition 34. *There exists a function $\nu : ((\{-1,1\}^N)^{10\log n})^n \to \mathbb{R}$ such that*

$$\textit{For all polynomials } p \colon ((\{-1,1\}^N)^{10\log n})^n \to \mathbb{R},\ \deg p < D \implies \langle \nu, p \rangle = 0 \tag{55}$$

$$\|\nu\|_1 \leq 1/10 \tag{56}$$

$$|x| > N \implies \nu(x) = \zeta(x), \tag{57}$$

where ζ is as in Proposition 30.

Proof. Define

$$\nu(x) = \sum_{y : |y| > N} \zeta(y)\phi_y(x),$$

where ϕ_y is as in Lemma 33 with m and D set as at the beginning of Section 4.3. Property (55) follows immediately from Property (53) and linearity. By Proposition 31 and Property (54), we have

$$\begin{aligned}
\|\nu\|_1 &\leq \sum_{y : |y| > N} |\zeta(y)| \cdot 2^D \cdot \binom{|y|}{D} \\
&\leq (2NR)^{-2D} \cdot 2^D \cdot m^D \\
&\leq (2m)^{-2D} \cdot (2m)^D \\
&\leq 1/10,
\end{aligned}$$

establishing Property (56). Finally, Property (57) follows from (51) and (52), together with the fact that $D < N$. □

Combining Proposition 34 with Proposition 30 allows us to complete the proof of Theorem 26, which was the goal of this section.

Proof of Theorem 26. Let $\zeta = \varphi \star \Psi \star \psi$ be as defined in Section 4.3, and let ν be the correction object constructed in Proposition 34. Observe that $\|\zeta - \nu\|_1 > 0$, as $\|\zeta\|_1 = 1$ (cf. Equality (41)) and $\|\nu\|_1 \leq 1/10$ (cf. Inequality (56)). Define the function

$$\hat{\zeta}(x) = \frac{\zeta(x) - \nu(x)}{\|\zeta - \nu\|_1}.$$

Since $\nu(x) = \zeta(x)$ whenever $|x| > N$ (cf. Equation (57)), the function $\hat{\zeta}$ is supported on the set X. By Theorem 7, to show that it is a dual witness for the high approximate degree of $G^{\leq N}$, it suffices to show that $\hat{\zeta}$ satisfies the following three properties:

$$\langle \hat{\zeta}, G^{\leq N} \rangle \geq 1/3 \tag{58}$$

$$\|\hat{\zeta}\|_1 = 1 \tag{59}$$

$$\text{For all polynomials } p\colon ((\{-1,1\}^N)^{10 \log n})^n \to \mathbb{R},\ \deg p < D \implies \langle \hat{\zeta}, p \rangle = 0. \tag{60}$$

We establish (58) by computing

$$\begin{aligned}
\langle \hat{\zeta}, G^{\leq N} \rangle &= \frac{1}{\|\zeta - \nu\|_1} \langle \zeta - \nu, G^{\leq N} \rangle \\
&= \frac{1}{\|\zeta - \nu\|_1} \langle \zeta - \nu, G \rangle && \text{since } \zeta = \nu \text{ outside } X \\
&= \frac{1}{\|\zeta - \nu\|_1} \left(\langle \zeta, G \rangle - \langle \nu, G \rangle \right) \\
&\geq \frac{1}{\|\zeta - \nu\|_1} \left(\langle \zeta, G \rangle - \|\nu\|_1 \right) \\
&\geq \frac{1}{\|\zeta - \nu\|_1} (1/2 - 1/10) && \text{by (40) and (56)} \\
&\geq \frac{1}{\|\zeta\|_1 + \|\nu\|_1} (1/2 - 1/10) \\
&\geq \frac{1}{1 + 1/10} (1/2 - 1/10) && \text{by (41) and (56)} \\
&\geq \frac{1}{3}.
\end{aligned}$$

Equation (59) is immediate from the definition of $\hat{\zeta}$. Finally, (60) follows from (42), (55), and linearity. □

4.5 Proof of Lemma 32

All that remains to complete the proof of Theorem 26 is to establish the deferred combinatorial lemma from Section 4.4. We begin by stating a two simple lemmas.

Lemma 35. *Let $k, n \in \mathbb{N}$ with $k \leq n$. Then $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$.*

Lemma 36. *Let $m \in \mathbb{N}$. Then*

$$\sum_{r=m}^{\infty} r^{-2} \leq \frac{2}{m}.$$

Proof. We calculate

$$\sum_{r=m}^{\infty} r^{-2} \leq \sum_{r=m}^{\infty} \frac{2}{r(r+1)} = 2 \sum_{r=m}^{\infty} \left(\frac{1}{r} - \frac{1}{r+1} \right) = \frac{2}{m}.$$

□

We are now ready to prove Lemma 32, which we restate below for the reader's convenience.

Lemma 32. *Let $k, R \in \mathbb{N}$ with $k \leq N$. There is a constant $\alpha > 0$ such that the following holds. Let $N = \lceil \alpha R \log^2 R \rceil$. Let $\eta_i : [k]_0 \to \mathbb{R}$, for $i = 1, \ldots R$, be a sequence of non-negative functions where for every i,*

$$\sum_{r=0}^{k} \eta_i(r) \leq 1/2 \tag{49}$$

$$\eta_i(r) \leq 5/(r+1)^2 \qquad \forall r = 0, 1, \ldots, k. \tag{50}$$

Let $P = \{\vec{t} = (t_1, \ldots, t_R) \in [k]_0^R : t_1 + \cdots + t_R > N\}$. Then

$$\sum_{\vec{t} \in P} \prod_{i=1}^{R} \eta_i(t_i) \leq 2^{-R} \cdot (2NR)^{-2R/k}.$$

Proof. Define a universal constant

$$C = \sum_{s=1}^{\infty} \frac{1}{s \log^2(2s)}.$$

Note that $C < \infty$ by, say, the Cauchy condensation test.

We begin with a simple, but important, structural observation about the set P. Let $t = (t_1, \ldots, t_R) \in [k]_0^R$ be a sequence such that $t_1 + \cdots + t_R > N$. Let $M = \lfloor \frac{N}{2k} \rfloor$. Then we claim that there exists an $s \in \{M \ldots, R\}$ such that $t_i \geq N/(2Cs \log^2(2s))$ for at least s indices $i \in [R]$. To see this, assume without loss of generality that the entries of $\vec{t}$ are sorted so that $t_1 \geq t_2 \geq \cdots \geq t_R$. Then there must exist an $s \geq M$ such that $t_s \geq N/(2Cs \log^2(2s))$. Otherwise, because no t_i can exceed k, we would have:

$$t_1 + \cdots + t_R < M \cdot k + \sum_{s=1}^{\infty} \frac{N}{2Cs \log^2(2s)} \leq \frac{N}{2} + \frac{N}{2C} \cdot \sum_{s=1}^{\infty} \frac{1}{s \log^2(2s)} = \frac{N}{2} + \frac{N}{2} = N.$$

Since the entries of $\vec{t}$ are sorted, the preceding values $t_1, \ldots, t_{s-1} \geq N/(2Cs \log^2(2s))$ as well.

For each subset $S \subseteq [R]$, define

$$P_S = \{\vec{t} \in P : t_i \geq N/\left(2C|S| \log^2(2|S|)\right) \text{ for all indices } i \in S\}.$$

The observations above guarantee that for every $\vec{t} = (t_1, \ldots, t_R) \in P$, there exists some set S of size at least $s \in \{M, \dots, R\}$ such that $t_i \geq N/(2Cs\log^2(2s))$ for all $i \in S$. Hence,

$$\begin{aligned}
\sum_{\vec{t}\in P}\prod_{i=1}^{R}\eta_i(t_i) &\leq \sum_{s=M}^{R}\sum_{S\subseteq[R]:|S|=s}\sum_{\vec{t}\in P_S}\prod_{i=1}^{R}\eta_i(t_i) \\
&\leq \sum_{s=M}^{R}\binom{R}{s}\left(\sum_{r=\lceil N/\left(2Cs\log^2(2s)\right)\rceil}^{k}\eta_i(r)\right)^{s}\left(\sum_{r=0}^{k}\eta_i(r)\right)^{R-s} \\
&\leq 2^{-R}\sum_{s=M}^{R}\binom{R}{s}\left(\sum_{r=\lceil N/\left(2Cs\log^2(2s)\right)\rceil}^{k}10(r+1)^{-2}\right)^{s} && \text{by Properties (49) and (50)} \\
&\leq 2^{-R}\sum_{s=M}^{R}\left(\frac{Re}{s}\right)^{s}\left(\frac{40Cs\log^2(2s)}{N}\right)^{s} && \text{by Lemmas 35 and 36} \\
&\leq 2^{-R}\sum_{s=M}^{R}4^{-s} && \text{setting } N=\lceil(160Ce)\cdot R\log^2(2R)\rceil \\
&\leq 2^{-R}\cdot 2^{-M}.
\end{aligned}$$

The claim follows as long as $M = \lfloor \frac{N}{2k} \rfloor \geq 2\log(2NR)\cdot R/k$, which is true for the setting of N chosen above. $\square$

5 Applications

5.1 Approximate Rank and Quantum Communication Complexity of AC^0

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 $(i,j) \in [N]\times[N]$. Sherstov's pattern matrix method [50] allows one to translate approximate degree lower bounds into approximate rank lower bounds in a black-box manner. Moreover, the logarithm of the approximate rank of a communication matrix is known to lower bound its quantum communication complexity, even when prior entanglement is allowed [37]. By combining the pattern matrix method with Theorems 3 and 4, we obtain the following corollary.

Corollary 37. *For any constant $\delta > 0$, there is an AC^0 function $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ such that $[F(x,y)]_{x,y}$ has approximate rank $\mathrm{rank}_{1/3}(F) \geq \exp(n^{1-\delta})$. Similarly, there is a DNF $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ of width $\mathrm{polylog}(n)$ (and quasipolynomial size) such that $[F(x,y)]_{x,y}$ has approximate rank at least $\exp(n^{1-\delta})$. Moreover, the quantum communication complexity of F (with arbitrary prior entanglement), denoted $Q^*_{1/3}(F)$, is $\Omega(n^{1-\delta})$.*

Proof. Let f be the AC^0 function, or low-width DNF, with $(1/2)$-approximate degree at least $n^{1-\delta}$ whose existence is guaranteed by Theorem 3 or Theorem 4 respectively. The pattern matrix method [50, Theorem 8.1] implies that the function $F\colon \{-1,1\}^{4n} \times \{-1,1\}^{4n} \to \{-1,1\}$ given by

$$F(x,y) = f\left(\dots, \vee_{j=1}^{4}(x_{i,j} \wedge y_{i,j})\dots\right)$$

satisfies $\mathrm{rank}_{1/3}(F) \geq \exp(\Omega(n^{1-\delta}))$. 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 $k+2$. Similarly, if

f is computed by a DNF formula of width w, then F is computed by a DNF formula of width $O(w)$. The claimed lower bound on $Q^*_{1/3}(F)$ follows from the fact that for any $2^n \times 2^n$ matrix F, we have $Q^*_{1/3}(F) \geq \Omega(\log \mathrm{rank}_{1/3}(F)) - O(\log n)$ [37]. $\square$

The best previous lower bound on the approximate rank and quantum communication complexity of an AC^0 function was $\exp\left(\tilde{\Omega}(n^{2/3})\right)$ and $\tilde{\Omega}(n^{2/3})$ respectively. This follows from combining the Element Distinctness lower bound (Theorem 27), with the pattern matrix method [50].

5.2 Nearly Optimal Separation Between Certificate Complexity and Approximate Degree

Certificate complexity, approximate degree, Fourier degree, block sensitivity, and deterministic, randomized, and quantum query complexities are all natural measures of the complexity of Boolean functions, with many applications in theoretical computer science. While it is known that all of these measures are polynomially related, much effort has been devoted to understanding the maximal possible separations between these measures. Ambainis et al. [7], building on techniques of Göös, Pitassi, and Watson [29], recently made remarkable progress in this direction, establishing a number of surprising separations between several of these measures. Subsequent work by Aaronson, Ben-David, and Kothari [3] unified and strengthened a number of these separations.

Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a (total) Boolean function. In this section, we study the relationship between certificate complexity, denoted $C(f)$ and defined below, and approximate degree. We build on Theorem 4 to construct a function $F\colon \{-1,1\}^n \to \{-1,1\}$ with $\widetilde{\deg}(F) = n^{1-o(1)}$ and certificate complexity $n^{1/2+o(1)}$. The function F exhibits what is essentially the maximal possible separation between these two measures, as it is known that $\widetilde{\deg}(f) = O(C(f)^2)$ for all Boolean functions f.[7] The best previous separation was reported by Aaronson et al. [3], who gave a function f with $\widetilde{\deg}(f) = \tilde{\Omega}(C(f)^{7/6})$.

Theorem 38. *There is a Boolean function $F : \{-1,1\}^n \to \{-1,1\}$ such that $\widetilde{\deg}(F) \geq C(F)^{2-o(1)}$.*

Certificate complexity definitions. Let $f\colon \{-1,1\}^n \to \{-1,1\}$, and let $x \in \{-1,1\}^n$. A subset $S \subseteq \{1,\dots,n\}$ is a (-1)-*certificate (respectively,* $(+1)$-*certificate) for f at x* if for all inputs $y \in \{-1,1\}^n$ such that $y_i = x_i$ for all $i \in S$, it holds that $f(y) = f(x) = -1$ (respectively, $f(y) = f(x) = 1$). For any $x \in \{-1,1\}^n$, let $C(f,x)$ denote the minimum size of a certificate for f at x. Define $C(f) := \max_{x\in\{-1,1\}^n} C(f,x)$. Define the (-1)-certificate complexity of f to be $C_{-1}(f) := \max_{x\in f^{-1}(-1)} C(f,x)$, and the $(+1)$-certificate complexity of f to be $C_{+1}(f) := \max_{x\in f^{-1}(+1)} C(f,x)$.

5.2.1 Warm-Up: A Power $3/2$ Separation

Before proving Theorem 38, we begin by proving a weaker separation that illustrates most of the ideas in our construction.

Proposition 39. *There is a Boolean function $F\colon \{-1,1\}^n \to \{-1,1\}$ such that $\widetilde{\deg}(F) \geq C(F)^{3/2-o(1)}$.*

Proof. While Theorem 4 is stated only for constant $k \geq 1$, the proof is easily seen to hold when k is a function of n. In particular, explicitly accounting for the constant factor loss that occurs in each step of the inductive proof of Theorem 4, we obtain the following statement that holds even if k grows with n.

[7] This follows by combining the relationship $D(f) \leq C(f)\cdot \mathrm{bs}(f)$ [9] with the relationships $\mathrm{bs}(f) \leq C(f)$ and $\widetilde{\deg}(f) \leq D(f)$.

Theorem 40 (Generalized Version of Theorem 4)**.** *For any integer $k \geq 1$, there is an (explicitly given) monotone DNF on $n \cdot \log^{4k-4}(n)$ variables of width $O(\log^{2k-1}(n))$ that computes a function with approximate degree $2^{-O(k)} \cdot n^{1-2^{k-1}/3^k} \cdot \log^{3-2^{k+2}/3^k}(n)$.*

Applying Theorem 40 for an appropriately chosen $k = O(\log \log n)$ yields a function $f : \{-1,1\}^M \to \{-1,1\}$ that is computed by a DNF on $M \leq n \cdot \log^{O(k)}(n) \leq n^{1+o(1)}$ variables with width $O(\log^{O(k)}(n)) \leq M^{o(1)}$. Equivalently, $C_{-1}(f) \leq M^{o(1)}$. Moreover, $\widetilde{\deg}(f) \geq n^{1-o(1)} \geq M^{1-o(1)}$.

Let $\mathsf{MAJ}_{10 \log M}$ denote the Majority function on $10 \log M$ bits. The following result is implicit in [15, Theorem 4.2].

Lemma 41 (Bouland et al. [15])**.** *Let $f \colon \{-1,1\}^M \to \{-1,1\}$. Then*

$$\widetilde{\deg}_\varepsilon(\mathsf{MAJ}_{10 \log M} \circ f) \geq \widetilde{\deg}(f)$$

for $\varepsilon = 1 - 1/M^2$.

Now consider the block-composed function $F = \mathsf{AND}_M \circ \mathsf{MAJ}_{10 \log M} \circ f$. This is a function on $10M^2 \log M$ variables, and Lemma 41 together with Proposition 12 implies that

$$\widetilde{\deg}(F) \geq M^{1/2} \cdot \widetilde{\deg}(f) \geq M^{3/2-o(1)}. \tag{61}$$

We now show that the function F has certificate complexity $C(F) \leq M^{1+o(1)}$. Let $\hat{F} = \mathsf{MAJ}_{10 \log M} \circ f$. Then $C_{-1}(\hat{F}) \leq 5 \log M \cdot C_{-1}(f) = M^{o(1)}$; this uses the fact that in order to certify that $\mathsf{MAJ}_{10 \log M}$ evaluates to -1, it is enough to certify that at least half of its inputs are equal to -1. And, trivially, $C_{+1}(\hat{F}) \leq 10M \log M$.

Any input z to AND_M has a certificate S such that $z_i = +1$ for at most one index $i \in S$. By composing certificates, it follows that

$$C(F) \leq C_{+1}(\hat{F}) + M \cdot C_{-1}(\hat{F}) \leq M^{1+o(1)}. \tag{62}$$

Combining (61) and (62) completes the proof of Proposition 39.

□

5.2.2 A Nearly Quadratic Separation

To improve Proposition 39 to a nearly quadratic separation, we replace AND_M in the definition of F with a function f^* defined on roughly M variables, such that $\widetilde{\deg}(f^*) \geq M^{1-o(1)}$. This function f^* must moreover possess certificates satisfying the same key property as the AND function. Namely, every input z to f^* must have a certificate S such that $z_i = +1$ for only a small (as we will see, $M^{o(1)}$) number of indices $i \in S$.

Construction of f^*. While the function $f \colon \{-1,1\}^M \to \{-1,1\}$ considered in the proof of Proposition 39 satisfies the requisite approximate degree bound, it lacks the key property regarding its certificates. Hence, we must modify the function f to obtain a suitable function f^*. The modification we use generalizes a technique introduced by Aaronson et al. [3, Theorem 8] to give separations between quantum query complexity and certificate complexity, Fourier degree, and approximate degree.

Definition 42. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be computed by a DNF formula $\mathcal{C}_f$ of width w. We define a function $f^*\colon \{-1,1\}^{2n} \to \{-1,1\}$ as follows. Let each of the first n inputs to f^* be associated with an input to f, and each of the last n inputs of f^* be associated with the negation of an input to f. For an $i \in [n]$, let (x_i, x_{n+i}) be the pair of inputs to f^* corresponding to the ith input to f, and say that the pair is* balanced *if exactly one of x_i, x_{n+i} is equal to -1 (and exactly one is equal to $+1$).*

For an input $x \in \{-1,1\}^{2n}$, define $\gamma(x) \in \{-1,+1,\perp\}^n$ by

$$(\gamma(x))_i = \begin{cases} -1 \text{ if } (x_i, x_{n+i}) \text{ is balanced and } x_i = -1, \\ +1 \text{ if } (x_i, x_{n+i}) \text{ is balanced and } x_{n+i} = -1, \\ \perp \ \text{ otherwise.} \end{cases}$$

We say a clause of the DNF formula $\mathcal{C}_f$ is satisfied *by a vector $y \in \{-1,+1,\perp\}^n$ if every literal in that clause is satisfied by y (if $y_i = \perp$, then any literal corresponding to an input i or its negation is automatically unsatisfied). Define $f^*\colon \{-1,1\}^{2n} \to \{-1,1\}$ by:*

$$f^*(x) = \begin{cases} -1 \text{ if there is a clause of } \mathcal{C}_f \text{ that is satisfied by } \gamma(x), \text{ and for all } i \in [n], (x_i, x_{n+i}) \neq (+1,+1). \\ +1 \text{ otherwise.} \end{cases}$$

Let $\varepsilon > 0$ and suppose $p : \{-1,1\}^{2n} \to \mathbb{R}$ is a polynomial with $|p(x) - f^*(x)| \leq \varepsilon$ for every $x \in \{-1,1\}^{2n}$. Then the polynomial $q : \{-1,1\}^n \to \mathbb{R}$ defined by $q(y) = p(y_1, \ldots, y_n, -y_1, \ldots, -y_n)$ satisfies $|q(y) - f(y)| \leq \varepsilon$ for all $y \in \{-1,1\}^n$, since the definition of f^* guarantees that $f(y) = f^*(y_1, \ldots, y_n, -y_1, \ldots, -y_n)$. Hence, for every $\varepsilon > 0$, we have

$$\widetilde{\deg}_\varepsilon(f^*) \geq \widetilde{\deg}_\varepsilon(f). \tag{63}$$

Completing the Proof of Theorem 38. Let f^* denote the function obtained by applying Definition 42 to the function $f\colon \{-1,1\}^M \to \{-1,1\}$ described in the proof of Proposition 39. Recall that f is computed by a DNF $\mathcal{C}_f$ of width $M^{o(1)}$. Hence f^* is a function on $2M$ variables, and by Inequality (63),

$$\widetilde{\deg}(f^*) \geq \widetilde{\deg}(f) \geq M^{1-o(1)}.$$

We now argue that every input x to f^* has a certificate S in which at most $M^{o(1)}$ entries of $x|_S$ are equal to $+1$. To see this, first let x be any input in $(f^*)^{-1}(-1)$. Then by definition of f^*, it suffices to certify that (a) there is a clause of $\mathcal{C}_f$ that is satisfied by $\gamma(x)$ and (b) there is no $i \in [M]$ such that $(x_i, x_{M+i}) = (+1,+1)$. Letting w denote the width of $\mathcal{C}_f$, condition (a) can be certified by providing at most $2w = M^{o(1)}$ indices of x, and condition (b) can be certified by supplying all of the coordinates of x that are equal to -1.

Now suppose that x is an input in $(f^*)^{-1}(+1)$. There are two kinds of such inputs to certify. The first kind is any input x with $(x_i, x_{M+i}) = (+1,+1)$ for some $i \in [M]$. Such an input can be certified by providing x_i and x_{M+i}. The second kind is an input such that for all $i \in [M]$, the pair (x_i, x_{M+i}) has at least one coordinate equal to -1, yet no clause of $\mathcal{C}_f$ is satisfied by $\gamma(x)$. This kind of input can be certified by providing the indices of all (-1)'s in the input x. Such a certificate is enough to reveal $(\gamma(x))_i$ for all i under the assumption that every pair (x_i, x_{M+i}) with exactly one (-1) provided is balanced. While this certificate does not prove that every pair (x_i, x_{M+i}) is actually balanced, it is still enough to prove that there is no clause of $\mathcal{C}_f$ that is satisfied. This is because changing a pair (x_i, x_{M+i}) from balanced to unbalanced cannot cause an unsatisfied clause of $\mathcal{C}_f$ to become satisfied.

To summarize, the value of $f^*(x)$ can always be certified by providing at most $M^{o(1)}$ indices of x that are equal to $+1$. To complete our construction, let $F = f^* \circ \hat{F}$, where $\hat{F} = \mathsf{MAJ}_{10 \log M} \circ f$. This is a function on $M^{2+o(1)}$ variables. With the aforementioned property of the certificates of f^* in hand, the argument that $\widetilde{\deg}(F) \geq C(F)^{2-o(1)}$ is identical to that of the previous section. Indeed, by composing certificates for f^* and $\hat{F}$, one obtains

$$C(F) \leq M^{o(1)} \cdot C_{+1}(\hat{F}) + M \cdot C_{-1}(\hat{F}) \leq M^{1+o(1)}. \tag{64}$$

Since $\widetilde{\deg}(f^*) \geq M^{1-o(1)}$, Lemma 41 and Proposition 12 imply that

$$\widetilde{\deg}(F) \geq M^{2-o(1)}. \tag{65}$$

Combining Equations (64) and (65) completes the proof of Theorem 38.

5.3 Secret Sharing Schemes

Bogdanov et al. [14] 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 $x = (x_1, \dots, x_n) \in \{-1,1\}^n$ from the distribution $2 \cdot \mu_b$, and gives bit x_i to the ith party. In order to reconstruct b, one simply applies f to $(x_1, \dots, x_n)$.

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. We direct the interested reader to [14] for further details.

Hence, an immediate corollary of our new approximate degree lower bounds for AC^0 is the following.

Corollary 43. *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 satisfies the following properties.*

(a) The reconstruction procedure is computed by an AC^0 circuit.

(b) The reconstruction advantage is at least 0.49.

(c) Any subset of shares of size less than $d = \Omega(n^{1-\delta})$ provides no information about the secret bit b.

The above corollary improves over an analogous result of Bogdanov et al. [14], who used the Element Distinctness lower bound (cf. Theorem 27) to give a scheme for which subsets of shares of size less than $d = \Omega(n^{2/3})$ provides no information about the secret bit b.

To make the secret sharing scheme of Corollary 43 explicit, one needs an explicit dual polynomial witnessing our new approximate degree lower bounds for AC^0 (cf. Theorem 3). Strictly speaking, our proof of Theorem 3 does not achieve this, owing to the *primal-based* symmetrization step of Section 3. However, this issue is easily addressed.

In more detail, recall that Theorem 26 establishes that there is an AC^0 function

$$G^{\leq N}\colon \{-1,1\}^{N'}_{\leq N} \to \{-1,1\}$$

for some $N' > N$, such that $G^{\leq N}$ has approximate degree at least $\Omega(N^{1-\delta})$. In fact, the proof of Theorem 26 constructs an explicit dual polynomial ψ witnessing this approximate degree bound. Definition 22 and Corollary 24 define an associated function $g^* \colon \{-1,1\}^m \to \{-1,1\}$, for $m = \tilde{O}(N)$, such that $\widetilde{\deg}_\varepsilon(g^*) \geq \widetilde{\deg}_\varepsilon(G^{\leq N}) \cdot \lceil \log(R+1) \rceil$. A natural averaging construction shows how to translate the dual polynomial $\psi \colon \{-1,1\}^{N'}_{\leq N} \to \mathbb{R}$ for $G^{\leq N}$ into a dual polynomial $\phi \colon \{-1,1\}^m \to \mathbb{R}$ for g^*. The analysis in the proof of Theorem 21 can then be used to show that this transformation preserves pure high degree, and that the correlation of ϕ with g^* is the same as the correlation of ψ with $G^{\leq N}$. (These remarks also apply to the augmented construction of g in Definition 25.)

We further believe that closer inspection of this dual witness should show that shares from the resulting scheme can be *sampled* by an AC^0 circuit.

6 Future Directions

6.1 Stronger Results for Constant Error Approximation

Throughout this section, δ denotes an arbitrarily small positive constant. While our $\Omega(n^{1-\delta})$ lower bound on the approximate degree of AC^0 comes close to resolving Problem 1 from the introduction, we fall short of a complete solution. Can our techniques be refined to give an $\Omega(n)$ lower bound on the approximate degree of a function in AC^0? Even the approximate degree of the $\mathsf{SURJECTIVITY}$ function remains unresolved. It is reasonable to conjecture that this function has essentially maximal approximate degree, $\Omega(n)$, yet our methods do not improve on the known $\Omega(n^{2/3})$ lower bound for this function.

It would also be very interesting to extend our $\Omega(n^{1-\delta})$ lower bounds for DNFs of polylogarithmic width and quasipolynomial size to DNFs of polynomial size (and ideally of logarithmic width). Currently, the best known lower bound on the approximate degree of polynomial size DNFs remains $\tilde{\Omega}(n^{2/3})$ for Element Distinctness.

For any constant integer $k > 0$, the k-sum function is a DNF of width $O(\log n)$ that might have approximate degree $\Omega(n^{k/(k+1)})$ [6, 13]. Another candidate DNF that might have approximate degree polynomially larger than $\Omega(n^{2/3})$ is the k-distinctness function for $k \geq 3$. (The best known upper bound on the approximate degree of the k-distinctness function is $O(n^{1-2^{k-2}/(2^k-1)})$; this bound approaches $n^{3/4}$ as $k \to \infty$ [12].) However, we believe that substantially new techniques will be required to resolve the approximate degree of these specific candidates. As explained in Section 1.3.1, our analysis is tailored to showing (near-)optimality of robustification-based approximating polynomials for the functions we consider, in a sense that can be made precise via complementary slackness. But the best known approximating polynomials for k-sum and k-distinctness are derived from sophisticated quantum algorithms [6, 12]. In particular, they are not constructed via robustification. Hence, we expect that any proof of a novel approximate degree lower bound for these functions will have to look very different than our own, as they will have to implicitly engage with non-robustification based approximating polynomials.

6.2 Stronger Results for Large Error Approximation

Another open direction is to strengthen our ε-approximate degree lower bounds on AC^0 from $\varepsilon = 1/3$ to ε much closer to 1. For example, the following two variants of Problem 1 from the introduction are open.

Problem 2. *Is there a constant-depth circuit in n variables with ε-approximate degree $\Omega(n)$, for (say) $\varepsilon = 1 - 2^{-\Omega(n)}$?*

Problem 3. *Is there a constant-depth circuit in n variables with ε-approximate degree $\Omega(n)$, for* any *$\varepsilon < 1$?*

Problem 3 is equivalent to asking whether there is an AC^0 function with linear *threshold degree*. Resolving Problems 2 and 3 would have a wide variety of consequences in computational learning theory, circuit complexity, and communication complexity (see, e.g., [15, 19, 57] and the references therein).

Despite attention by many researchers, the best known lower bounds in the directions of Problems 3 and 4 are:

(a) For any constant $\Gamma > 0$, a depth-3 circuit with ε-approximate degree $\Omega(n^{1/2-\delta})$ for $\varepsilon = 1 - 2^{-n^{\Gamma}}$ [19],

(b) A depth-3 circuit with threshold degree $\Omega(n^{3/7})$ [46], and

(c) A depth-4 circuit with threshold degree $\Omega(n^{1/2})$ [46].

We believe that the following three results in the directions of Problems 2 and 3 should be achievable via relatively modest extensions of our techniques.

First, it should be possible to nearly resolve Problem 2 as follows. Recall from Section 1.2.1 that our recent work [19] also proved stronger hardness amplification results for approximate degree by moving beyond block composed functions. The methods of [19] amplify approximation error but not degree, while in this paper we amplify degree but not approximation error. We believe that it is possible to combine the two sets of techniques to exhibit a function in AC^0 on n variables with ε-approximate degree at least $n^{1-\delta}$, even for $\varepsilon = 1-2^{-\Omega(n^{1-\delta})}$. Such a result would translate in a black-box manner into lower bounds of $2^{\Omega(n^{1-\delta})}$ on the margin complexity, (multiplicative inverse of) discrepancy, threshold weight, and Majority-of-Threshold circuit size of AC^0, nearly matching trivial $2^{O(n)}$ upper bounds.

Second, we are confident that the polylogarithmic width DNF $f\colon \{-1,1\}^n \to \{-1,1\}$ of approximate degree $\Omega(n^{1-\delta})$ exhibited in Theorem 4 in fact has large *one-sided* approximate degree [18]. Moreover, this should be provable via a modest extension of our techniques. Combining such a lower bound with a result of Sherstov [57] would imply that $\mathsf{AND}_{n^{1-\delta}} \circ f$ has threshold degree $\Omega(n^{1-\delta})$, thereby yielding a depth three circuit (of quasipolynomial size) on $N = n^{2-2\delta}$ variables with threshold degree $\Omega(N^{1/2-\delta})$.

Third, we believe that the following function g on $O(n \log^4 n)$ variables has threshold degree $\Omega(n^{3/5})$. Let $f_n = \mathsf{AND}_{n^{1/5}} \circ \mathsf{OR}_{n^{2/5}} \circ \mathsf{AND}_{n^{2/5}}$, and let g be the harder function obtained by applying the construction of Theorem 5 to f_n. Note that g is computed by a circuit of depth 5.

Sherstov [57] constructed a dual polynomial ψ witnessing the fact that

$$\deg_{\pm}\left(\mathsf{AND}_{n^{1/5}} \circ \mathsf{OR}_{n^{2/5}} \circ \mathsf{AND}_{n^{2/5}} \circ \mathsf{OR}_{n^{2/5}}\right) = \Omega(n^{3/5}).$$

(Note that this block composed function is defined over $n^{7/5}$ variables.) In order to show that g likewise has threshold degree $\Omega(n^{3/5})$, our results from Section 3 imply that it is enough to "zero out" the mass that ψ places on inputs of Hamming weight larger than a suitable threshold $N = \tilde{O}(n)$, without affecting the sign of ψ on the remaining inputs. We believe that is possible to achieve this via a refinement of the zeroing technique used in this work.

A final ambitious direction. A more ambitious direction toward resolving Problems 2 and 3 would be to obtain a version of our hardness amplification result (Theorem 5) that (a) applies to threshold degree rather than approximate degree and (b) can be applied recursively. This would allow one to obtain an $\Omega(n^{1-\delta})$ lower bound on the threshold degree of AC^0, nearly resolving Problem 3 above.

One might hope to obtain such a result by extending the above envisioned analysis for obtaining an $\Omega(n^{3/5})$ threshold degree lower bound, so as to allow recursive application of the construction and analysis.

However, we believe that achieving this goal will require substantial new ideas. The only available techniques for recursively amplifying threshold degree bounds are due to Sherstov [46, 57], who considers block composed functions of the form $\mathsf{OR} \circ f$. Specifically, he uses a dual witness for the *outer function* OR to "amplify the efficacy" of a dual witness for the *inner function* f.

In contrast, our recursive construction in this paper considers block compositions of the form $f \circ \mathsf{OR}$, and uses a dual witness for the *inner function* OR to "amplify the efficacy" of a dual witness for the *outer function* f. This difference appears to prevent us from combining the methods of [57] with our own in a manner that would enable recursive application. Finding a way to reconcile the two approaches may present a promising avenue for obtaining a (near-)resolution of Problem 3.

Acknowledgements. We are grateful to Shalev Ben-David for illuminating conversations regarding separations between approximate degree and certificate complexity, and to Robin Kothari and Sasha Sherstov for valuable comments on an earlier version of this manuscript.

References

[1] Scott Aaronson. The polynomial method in quantum and classical computing. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, page 3, 2008.

[2] Scott Aaronson. Impossibility of succinct quantum proofs for collision-freeness. *Quantum Information & Computation*, 12(1-2):21–28, 2012.

[3] Scott Aaronson, Shalev Ben-David, and Robin Kothari. Separations in query complexity using cheat sheets. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 863–876, 2016.

[4] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004.

[5] Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

[6] Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM J. Comput.*, 37(1):210–239, 2007.

[7] Andris Ambainis, Kaspars Balodis, Aleksandrs Belovs, Troy Lee, Miklos Santha, and Juris Smotrovs. Separations in query complexity based on pointer functions. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 800–813, 2016.

[8] Andris Ambainis, Andrew M. Childs, Ben Reichardt, Robert Spalek, and Shengyu Zhang. Any and-or formula of size n can be evaluated in time $\mathrm{n}^{1/2+\mathrm{o}(1)}$ on a quantum computer. *SIAM J. Comput.*, 39(6):2513–2530, 2010.

[9] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[10] Paul Beame and Widad Machmouchi. The quantum query complexity of AC^0. *Quantum Information & Computation*, 12(7-8):670–676, 2012.

[11] Richard Beigel. Perceptrons, PP, and the Polynomial Hierarchy. *Computational Complexity*, 4:339–349, 1994.

[12] Aleksandrs Belovs. Learning-graph-based quantum algorithm for k-distinctness. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 207–216. IEEE Computer Society, 2012.

[13] Aleksandrs Belovs and Robert Spalek. Adversary lower bound for the k-sum problem. In Robert D. Kleinberg, editor, *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013*, pages 323–328. ACM, 2013.

[14] Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded indistinguishability and the complexity of recovering secrets. In Matthew Robshaw and Jonathan Katz, editors, *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part III*, volume 9816 of *Lecture Notes in Computer Science*, pages 593–618. Springer, 2016.

[15] Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On SZK and PP. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:140, 2016.

[16] Harry Buhrman, Nikolai K. Vereshchagin, and Ronald de Wolf. On computation and communication with small bias. In *22nd Annual IEEE Conference on Computational Complexity (CCC 2007), 13-16 June 2007, San Diego, California, USA*, pages 24–32. IEEE Computer Society, 2007.

[17] Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and markov-bernstein inequalities. In Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg, editors, *ICALP (1)*, volume 7965 of *Lecture Notes in Computer Science*, pages 303–314. Springer, 2013.

[18] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, volume 9134 of *Lecture Notes in Computer Science*, pages 268–280. Springer, 2015. Full version available at `http://eccc.hpi-web.de/report/2013/151`.

[19] Mark Bun and Justin Thaler. Approximate degree and the complexity of depth three circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:121, 2016.

[20] Mark Bun and Justin Thaler. Dual polynomials for Collision and Element Distinctness. *Theory of Computing*, 12(16):1–34, 2016.

[21] Mark Bun and Justin Thaler. Improved bounds on the sign-rank of AC^0. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 37:1–37:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.

[22] Karthekeyan Chandrasekaran, Justin Thaler, Jonathan Ullman, and Andrew Wan. Faster private release of marginals on small databases. *CoRR*, abs/1304.3754, 2013.

[23] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(002), 2008.

[24] Lijie Chen. Adaptivity vs. postselection, and hardness amplification for polynomial approximation. In *27th International Symposium on Algorithms and Computation, ISAAC 2016, December 12-14, 2016, Sydney, Australia*, pages 26:1–26:12, 2016.

[25] Andrew M. Childs, Shelby Kimmel, and Robin Kothari. The quantum query complexity of read-many formulas. In Leah Epstein and Paolo Ferragina, editors, *Algorithms - ESA 2012 - 20th Annual European Symposium, Ljubljana, Slovenia, September 10-12, 2012. Proceedings*, volume 7501 of *Lecture Notes in Computer Science*, pages 337–348. Springer, 2012.

[26] Matei David and Toniann Pitassi. Separating NOF communication complexity classes RP and NP. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(014), 2008.

[27] Matei David, Toniann Pitassi, and Emanuele Viola. Improved separations between nondeterministic and randomized multiparty communication. *TOCT*, 1(2), 2009.

[28] Dmitry Gavinsky and Alexander A. Sherstov. A separation of NP and coNP in multiparty communication complexity. *Theory of Computing*, 6(1):227–245, 2010.

[29] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1077–1088, 2015.

[30] Peter Høyer, Michele Mosca, and Ronald de Wolf. Quantum search on bounded-error inputs. In Jos C. M. Baeten, Jan Karel Lenstra, Joachim Parrow, and Gerhard J. Woeginger, editors, *Automata, Languages and Programming, 30th International Colloquium, ICALP 2003, Eindhoven, The Netherlands, June 30 - July 4, 2003. Proceedings*, volume 2719 of *Lecture Notes in Computer Science*, pages 291–299. Springer, 2003.

[31] Jeff Kahn, Nathan Linial, and Alex Samorodnitsky. Inclusion-exclusion: Exact and approximate. *Combinatorica*, 16(4):465–477, 1996.

[32] Adam Tauman Kalai, Adam R. Klivans, Yishay Mansour, and Rocco A. Servedio. Agnostically learning halfspaces. *SIAM J. Comput.*, 37(6):1777–1805, 2008.

[33] Varun Kanade and Justin Thaler. Distribution-independent reliable learning. In Maria-Florina Balcan and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, volume 35 of *JMLR Proceedings*, pages 3–24. JMLR.org, 2014.

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

[35] Adam R. Klivans and Rocco A. Servedio. Toward attribute efficient learning of decision lists and parities. *Journal of Machine Learning Research*, 7:587–602, 2006.

[36] Troy Lee. A note on the sign degree of formulas. *CoRR*, abs/0909.4607, 2009.

[37] Troy Lee and Adi Shraibman. An approximation algorithm for approximation rank. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*, pages 351–357, 2009.

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

[39] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[40] Ryan O'Donnell and Rocco A. Servedio. New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010.

[41] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In S. Rao Kosaraju, Mike Fellows, Avi Wigderson, and John A. Ellis, editors, *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 468–474. ACM, 1992.

[42] Vladimir V. Podolskii. A uniform lower bound on weights of perceptrons. In Edward A. Hirsch, Alexander A. Razborov, Alexei L. Semenov, and Anatol Slissenko, editors, *Computer Science - Theory and Applications, Third International Computer Science Symposium in Russia, CSR 2008, Moscow, Russia, June 7-12, 2008, Proceedings*, volume 5010 of *Lecture Notes in Computer Science*, pages 261–272. Springer, 2008.

[43] Anup Rao and Amir Yehudayoff. Simplified lower bounds on the multiparty communication complexity of disjointness. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:60, 2014.

[44] Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010.

[45] Rocco A. Servedio, Li-Yang Tan, and Justin Thaler. Attribute-efficient learning and weight-degree tradeoffs for polynomial threshold functions. In Shie Mannor, Nathan Srebro, and Robert C. Williamson, editors, *COLT*, volume 23 of *JMLR Proceedings*, pages 14.1–14.19. JMLR.org, 2012.

[46] A. A. Sherstov. The power of asymmetry in constant-depth circuits. In *FOCS*, 2015. Full version available at `http://eccc.hpi-web.de/report/2015/147/`.

[47] Alexander A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008.

[48] Alexander A. Sherstov. Approximate inclusion-exclusion for arbitrary symmetric functions. *Computational Complexity*, 18(2):219–247, 2009.

[49] Alexander A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009.

[50] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

[51] Alexander A. Sherstov. The multiparty communication complexity of set disjointness. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 525–548. ACM, 2012.

[52] Alexander A. Sherstov. Approximating the AND-OR Tree. *Theory of Computing*, 9(20):653–663, 2013.

[53] Alexander A. Sherstov. Communication lower bounds using directional derivatives. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 921–930. ACM, 2013.

[54] Alexander A. Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013.

[55] Alexander A. Sherstov. Making polynomials robust to noise. *Theory of Computing*, 9:593–615, 2013.

[56] Alexander A. Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica*, 33(1):73–96, 2013.

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

[58] Yaoyun Shi. Approximating linear restrictions of boolean functions. 2002. Manuscript. Available online at: `web.eecs.umich.edu/ shiyy/mypapers/linear02-j.ps`.

[59] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009.

[60] Robert Spalek. A dual polynomial for OR. *CoRR*, abs/0803.4516, 2008.

[61] Avishay Tal. Shrinkage of de morgan formulae from quantum query complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:48, 2014.

[62] Avishay Tal. The bipartite formula complexity of inner-product is quadratic. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:181, 2016.

[63] Avishay Tal. Computing requires larger formulas than approximating. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:179, 2016.

[64] Justin Thaler. Lower bounds for the approximate degree of block-composed functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:150, 2014. To appear in *ICALP*, 2016.

[65] Justin Thaler, Jonathan Ullman, and Salil P. Vadhan. Faster algorithms for privately releasing marginals. In Artur Czumaj, Kurt Mehlhorn, Andrew M. Pitts, and Roger Wattenhofer, editors, *Automata, Languages, and Programming - 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part I*, volume 7391 of *Lecture Notes in Computer Science*, pages 810–821. Springer, 2012.

A A Refined Dual Witness for OR

Our goal is to prove the following equivalent formulation of Lemma 29.

Lemma 44. *Let $k \in \mathbb{N}$. There exists a constant $c_1 \in (0,1)$ and a function $\omega : \{0,1,\dots,k\} \to \mathbb{R}$ such that*

$$\omega(0) - \sum_{t=1}^{k} \omega(t) \geq \frac{1}{3}\|\omega\|_1 \tag{66}$$

$$\textit{For all univariate polynomials } q\colon \mathbb{R}\to\mathbb{R},\ \deg p < c_1\sqrt{k} \implies \sum_{t=0}^{k}\omega(t)\cdot q(t) = 0 \tag{67}$$

$$\omega(0) > 0 \tag{68}$$

$$\omega(t) \leq \frac{5\|\omega\|_1}{(t+1)^2} \qquad \forall t = 0,1,\dots,k \tag{69}$$

In the proof of Lemma 44, we make use of the following combinatorial identity.

Fact 45. *Let $k \in \mathbb{N}$, and let p be a polynomial of degree less than k. Then*

$$\sum_{t=0}^{k}(-1)^t\binom{k}{t}p(t) = 0.$$

Proof of Lemma 29. Let $c = 25$ below. Let $m = \lfloor\sqrt{k/c}\rfloor$ and define the set

$$T = \{1,2\} \cup \{ci^2 : 0 \leq i \leq m\}.$$

Note that $|T| = \Omega(\sqrt{k})$. Define the polynomial

$$\omega(t) = \frac{(-1)^{t+(k-m)}}{k!}\binom{k}{t}\prod_{r\in[k]_0\setminus T}(t-r).$$

It is immediate from Fact 45 that ω satisfies (67) for $c_1 = 1/\sqrt{c}$. By inspection, we have $\omega(0) > 0$ and $\omega(1) < 0$.

Expanding out the binomial coefficient reveals that

$$|\omega(t)| = \begin{cases}\prod\limits_{r\in T\setminus\{t\}}\frac{1}{|t-r|} & \text{for } t \in T,\\ 0 & \text{otherwise.}\end{cases}$$

For $t = 1$, we observe

$$\frac{|\omega(1)|}{\omega(0)} = \frac{2\prod_{i=1}^m ci^2}{\prod_{i=1}^m(ci^2-1)} = 2\prod_{i=1}^{m}\frac{i^2}{i^2-1/c} \geq 2.$$

On the other hand, for $t = 2$, we have

$$\begin{aligned}\frac{|\omega(2)|}{\omega(0)} &= \frac{2\prod_{i=1}^m ci^2}{2\prod_{i=1}^m(ci^2-2)}\\ &= \left(\prod_{i=1}^{m}\frac{i^2-2/c}{i^2}\right)^{-1}\\ &\leq \left(1-\sum_{i=1}^{m}\frac{2}{ci^2}\right)^{-1}\\ &\leq \left(1-\frac{\pi^2}{3c}\right)^{-1} = \frac{3c}{3c-\pi^2} \end{aligned}\tag{70}$$

where the first inequality follows from the fact that $\prod_{i=1}^m (1-a_i) \geq 1 - \sum_{i=1}^m a_i$ for $a_i \in (0,1)$.

For $t = cj^2$ with $j \geq 1$, we get

$$\begin{aligned}
\frac{|\omega(t)|}{\omega(0)} &= \frac{2\prod_{i=1}^m ci^2}{(cj^2-1)(cj^2-2)\prod_{i\in[m]\setminus\{j\}}|ci^2-cj^2|} \\
&= \frac{2(m!)^2}{(c^2j^4-3cj^2+2)\prod_{i\in[m]\setminus\{j\}}(i+j)|i-j|} \\
&= \frac{2(m!)^2}{(c^2j^4-3cj^2+2)(m+j)!(m-j)!} \\
&\leq \frac{2}{c^2j^4-3cj^2+2}
\end{aligned}$$

where the last inequality follows because

$$\frac{(m!)^2}{(m+j)!(m-j)!} = \frac{m}{m+j}\cdot\frac{m-1}{m+j-1}\cdot\ldots\cdot\frac{m-j+1}{m+1}$$

is a product of factors that are each smaller than 1. Since

$$\frac{|\omega(cj^2)|}{\|\omega\|_1} \leq \frac{|\omega(t)|}{\omega(0)} \leq \frac{2}{c^2j^4-3cj^2+2} \leq \frac{5}{(cj^2+1)^2}$$

for $c \geq 8$, this establishes (69).

What remains is to perform the correlation calculation to establish (66). First, observe that the total contribution of $t > 2$ to $\|\omega\|_1/\omega(0)$ is at most

$$\sum_{t>2}\frac{|\omega(t)|}{\omega(0)} = \sum_{j=1}^m \frac{2}{c^2j^4-3cj^2+2} < \sum_{j=1}^\infty \frac{2}{cj^2} < \frac{\pi^2}{3c}. \tag{71}$$

Next, we calculate

$$\begin{aligned}
\omega(0) - \sum_{t=1}^k \omega(t) &\geq \omega(0) - \omega(1) - \left(\sum_{t=2}^k |\omega(t)|\right) \\
&\geq \omega(0) - \omega(1) - \left(\omega(2) + \omega(0)\cdot\frac{\pi^2}{3c}\right) && \text{by (71)} \\
&\geq -\omega(1) + \omega(0)\left(1 - \frac{3c}{3c-\pi^2} - \frac{\pi^2}{3c}\right) && \text{by (70)} \\
&\geq -\omega(1) - \frac{1}{3}\omega(0) && \text{by our choice of } c = 25. \qquad (72)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\|\omega\|_1 &\leq \omega(0) - \omega(1) + \omega(2) + \omega(0)\cdot\frac{\pi^2}{3c} && \text{by (71)} \\
&\leq -\omega(1) + \omega(0)\left(1 + \frac{3c}{3c-\pi^2} + \frac{\pi^2}{3c}\right) && \text{by (70)} \\
&\leq -\omega(1) + \frac{7}{3}\omega(0) && \text{since } c = 25. \qquad (73)
\end{aligned}$$

Combining (72) and (73), and using the fact that $-\omega(1) \geq 2\omega(0)$ shows that

$$\frac{\omega(0) - \sum_{t=1}^{k} \omega(t)}{\|\omega\|_1} \geq \frac{-\omega(1) - \frac{1}{3}\omega(0)}{-\omega(1) + \frac{7}{3}\omega(0)} \geq \frac{1}{3}.$$

This establishes (66), completing the proof.

$\square$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il