58th Annual IEEE Symposium on Foundations of Computer Science

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

Mark Bun
Department of Computer Science
Princeton University
Princeton, NJ, USA
mbun@cs.princeton.edu

Justin Thaler
Department of Computer Science
Georgetown University
Washington, DC, USA
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 constant-depth polynomial-size Boolean circuit, 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.**

***Keywords*-approximate degree; certificate complexity; communication complexity; polynomial approximation; quantum communication complexity; secret sharing**

I. 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], [34]–[37], [43], [48], algorithmic approximations for the inclusion-exclusion principle [33], [51], and algorithms for differentially private data release [25], [68]. Very recently, approximate degree upper bounds have also been used to show new complexity-theoretic *lower 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 [64]–[66].

Meanwhile, lower bounds on approximate degree have enabled significant progress in quantum query complexity [2], [4], [11], communication complexity [19], [26], [28]–[30], [46], [50], [53], [54], [56], circuit complexity [41], [52], oracle separations [14], [18], and secret-sharing [17]. In particular, approximate degree has been established as one of the most promising tools available for understanding the complexity of constant-depth polynomial-size Boolean circuits (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 [13], [17], [21]–[24], [49], [60].

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).

0272-5428/17 $31.00 © 2017 IEEE
DOI 10.1109/FOCS.2017.10

Applications. We describe several consequences of the above results in complexity theory and cryptography. We state these results somewhat informally in this introduction, leaving details to Section V. Specifically:

- For any constant $\delta > 0$, we obtain an $\Omega(n^{1-\delta})$ lower bound on the bounded-error quantum communication complexity of AC^0. This nearly matches the trivial $O(n)$ upper bound that holds for any function, and improves on the previous best lower bound of $\Omega(n^{2/3})$. This lower bound also applies to the multiparty number-on-the-forehead model, where the previous best lower bound was again $\Omega(n^{2/3})$, even for classical randomized protocols.
- 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 [17], [53], the second application requires some additional effort.

Anshu et al. [10] have also observed that our second application combines in a black-box manner with the techniques of their recent work [9] to yield a nearly quadratic separation between quantum communication complexity and the logarithm of matrix rank. In addition, they have extended our techniques to obtain a nearly power-4 separation between quantum communication complexity and the logarithm of *approximate* rank. Both results improve on their earlier reported power-2 separation between quantum communication complexity and the logarithm of approximate rank [9].

A. Prior Work on Approximate Degree

1) Early Results via Symmetrization: The notion of approximate degree was introduced in seminal work of Nisan and Szegedy [42], who proved a tight $\Omega(n^{1/2})$ lower bound on the approximate degree of OR_n and AND_n. Nisan and Szegedy's proof exploited a powerful technique known as *symmetrization*, which was introduced in the late 1960's by Minsky and Papert [41]. Until recently, symmetrization was the primary tool available for proving approximate degree lower bounds [4], [5], [14], [44], [45], [48].

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 $\deg(q) \leq \deg(p)$. 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], [50], [57]). A relatively new such lower-bound technique called the *method of dual polynomials* plays an essential role in our paper.

B. 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], [42], [57], [61]. 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., $f \circ g = f(g,\ldots,g)$.

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 [32], who designed a quantum query algorithm to prove that $\widetilde{\deg}(\mathsf{AND}_M \circ \mathsf{OR}_N) = O(\sqrt{MN})$. Later, Sherstov [58] proved the following more general result.

Theorem 2 (Sherstov [58])**.** *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 [58] 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 $p^* := p_{\text{robust}}(q, \ldots, q)$ is an $(\varepsilon+\delta)$-approximating polynomial for $f_M \circ g_N$. Note that the degree of p^* is at most the product of the degrees of p_{robust} and q.

Sherstov [55] and the authors [20] 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 [20] and implicitly in [55]) 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 [62], Lee [38] and Sherstov [57]. Specifically, writing $x = (x_1, \ldots, x_M) \in \left(\{-1,1\}^N\right)^M$,

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

This technique of combining dual witnesses, which we call the "dual block" method, will also be central to this work. The lower bound of [20], [55] refined the analysis of $\psi_f \star \psi_g$ from [57] in the case where $f = \mathsf{AND}_M$ and $g = \mathsf{OR}_N$.

As argued in subsequent work of Thaler [67, 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) Additional Prior Work: The method of dual polynomials has recently been used to establish a number of new lower bounds for approximate degree [18], [21], [27], [43], [57], [59], [67]. 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 about the complexity of AC^0, as well as oracle separations involving the polynomial hierarchy and 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 [22], [45]. 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).

C. 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 $\tilde{\Omega}(n^{23/27}) = \tilde{\Omega}(n^{0.851\ldots})$.

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 $O(\log^{2k-1}(n))$ (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) 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. 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 $(s_1, \ldots, s_N)$ 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. $\mathsf{SURJ}_{N,R}$ and related functions have been extensively studied in quantum query complexity. In particular, Beame and Machmouchi [13] showed that computing $\mathsf{SURJ}_{N,R}$ for $R = N/2 + 1$ requires $\tilde{\Omega}(n)$ quantum queries, making it a natural candidate for improved approximate degree lower bounds for AC^0.

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.

Getting to Know SURJECTIVITY*:* It is known that $\widetilde{\deg}(\mathsf{SURJ}_{N,R}) = \tilde{\Omega}(n^{2/3})$ when $R \leq N$ and $R = \Theta(N)$ [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

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

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:

$$\mathsf{SURJ}_{N,R}(s) = \mathsf{AND}_R(\mathsf{OR}_N(y_1), \ldots, \mathsf{OR}_N(y_R)). \quad (4)$$

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 I-B). 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 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 [38], [57], [62], as this method is tailored to showing optimality of robustification-based approximations (cf. Section I-B). However, there are several technical challenges to overcome, owing to the fact that Equation (4) does not express SURJ as a genuine block composition (as a single bit of the input $s \in \{-1,1\}^{N \cdot \log R}$ affects R of the variables y_{ij}).

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 $\mathsf{SURJ}_{N,R} = \mathsf{AND}_R(\mathsf{OR}_N(y_{1,1}, \ldots, y_{1,N}), \ldots, \mathsf{OR}_N(y_{R,1} \ldots, y_{R,N}))$. This suggests that for general functions $f\colon \{-1,1\}^R \to \{-1,1\}$, one should consider the transformed function

$$F(s) := f(\mathsf{OR}_N(y_1), \ldots, \mathsf{OR}_N(y_R)).$$

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

$$g(s) = (f \circ \mathsf{AND}_{\log R})(\mathsf{OR}_N(y_1), \ldots, \mathsf{OR}_N(y_{R'})).$$

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 almost 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 III), 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 IV), 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 $N = \tilde{O}(R)$, it is necessary and sufficient for us to 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 only 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 I-B) used in our prior work [20] 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 [47] 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 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

$$\psi_{\text{OR}}(x) = 0 \text{ whenever } |x| > R^{1/3}. \tag{6}$$

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

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

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 $y = (y_1, \ldots, y_R) \in \{-1,1\}^{N \cdot R}$ 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 $R^{-\Omega(R^{2/3})}$.

II. 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 $p \colon \mathbb{R}^n \to \mathbb{R}$ with $|p(x) - f(x)| \leq \epsilon$ 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 $\widetilde{\deg}_{1/3}(f)$.*

Strong LP duality implies the following characterization of approximate degree (see, e.g., [53]).

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

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

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

$$\forall p \colon \mathcal{X} \to \mathbb{R}, \deg p < d, \sum_{x \in \mathcal{X}} p(x) \cdot \psi(x) = 0. \tag{10}$$

For functions $\psi_1 \colon \mathcal{X} \to \mathbb{R}$ and $\psi_2 \colon \mathcal{X}' \to \mathbb{R}$ 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 $\mathcal{X}' \subseteq \mathcal{X}$.)

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 $\{1, \ldots, N\}$ by $[N]$ and the set $\{0, \ldots, N\}$ 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.

A. The Dual Block Method

This section collects definitions and preliminary results on the dual block method [38], [57], [62] for constructing dual witnesses for a block composed function $F \circ f$ by combining dual witnesses for F and f respectively.

Definition 8. *Let $\Psi : \{-1,1\}^M \to \mathbb{R}$ and $\psi : \{-1,1\}^m \to \mathbb{R}$ be functions that are not identically zero. Let $x = (x_1, \ldots, x_M) \in (\{-1,1\}^m)^M$. The* dual block composition *of Ψ and ψ, denoted $\Psi \star \psi : (\{-1,1\}^m)^M \to \mathbb{R}$ is*

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

Proposition 9. *The dual block composition satisfies the following properties:*

Preservation of ℓ_1-norm [57]*: If $\|\Psi\|_1 = 1$ and $\|\psi\|_1 = 1$, then*

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

Multiplicativity of pure high degree [57]*: 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*

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

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 10 (Bun and Thaler [21])**.** *Let $m, M \in \mathbb{N}$. There is 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 11 (Sherstov [57])**.** *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*

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

III. 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. 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 $f : \{-1,1\}^n \to \{-1,1\}$.)

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

Definition 12. *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 13. *Let $G^{\leq N} \colon \{-1,1\}^{N \cdot R}_{\leq N} \to \{-1,1\}$ be as in Definition 12. 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)

The proof of Theorem 13 appears in the full version of this work, and builds on a symmetrization argument due to Ambainis [5].

IV. Analyzing Block Composed Functions On Low Hamming Weight Inputs

To complete the proof of Theorem 5, we combine the following theorem with Theorem 13.

Theorem 14. *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 $G^{\leq N} \colon \{-1,1\}^{10 \cdot N \cdot n \cdot \log n}_{\leq N} \to \{-1,1\}$ 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 12). 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 14. Before embarking on this proof, we use it to complete the proofs of Theorems 3-5 from Section I-C.

Proof of Theorem 5 assuming Theorem 14.: 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 Theorem 13 to F_R yields a function g on $O(N \log R) = O(n \log^4 n)$ variables satisfying

$$\begin{aligned}\widetilde{\deg_\varepsilon}(g) &\geq \widetilde{\deg_\varepsilon}(G^{\leq N}) \cdot \lceil \log(R+1) \rceil \\ &\geq \Omega(n^{1/3} \cdot \widetilde{\deg}(f_n)^{2/3} \cdot \log n),\end{aligned}$$

where the final inequality holds by Theorem 14. Properties (1), (2), and (3) now follow from Properties (18), (19), and (20) of Theorem 13, respectively. ∎

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 15 (Sherstov [49], 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 $O(\log n)$.*

Lemma 15 immediately implies Theorems 3 and 4 in the case $k = 1$. Theorems 3 and 4 now follow by induction via Properties (2) and (3) of Theorem 5, respectively. ∎

A. Organization of the Proof of Theorem 14

Our proof of Theorem 14 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 $\Omega(\sqrt{k})$-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 [63] and Bun and Thaler [21]. 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 10 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 11 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 [47], 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 $G^{\leq N}$.

B. Step 1: A Dual Witness for OR_N

Proposition 16. *Let $k, N \in \mathbb{N}$ with $k \leq N$. There exist $c_1 \in (0,1)$ and $\psi : \{-1,1\}^N_{\leq k} \to \{-1,1\}$ such that:*

$$\langle \psi, \mathsf{OR}_N \rangle \geq 1/3 \tag{21}$$

$$\|\psi\|_1 = 1 \tag{22}$$

$$\forall p \colon \{-1,1\}^N \to \mathbb{R},\ \deg p < c_1 \sqrt{k} \Longrightarrow \langle \psi, p \rangle = 0 \tag{23}$$

$$\psi(1^N) > 0 \tag{24}$$

$$\sum_{|x|=t} |\psi(x)| \leq 5/(t+1)^2 \qquad \forall t = 0, 1, \ldots, k \tag{25}$$

For intuition, we mention that Properties (21)-(24) 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. Property (25) is an additional metric condition that we require later in the proof. The construction closely follows previous work of Špalek [63] and Bun and Thaler [21], and appears in the full version of this work.

C. 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 14. Here, we begin by defining a preliminary dual witness ζ. While ζ itself is insufficient to witness the veracity of Theorem 14, we will

ultimately "post-process" ζ into the desired dual witness $\hat{\zeta}$. We start by fixing choices of several key parameters:

- $d = \widetilde{\deg}_{2/3}(f_n)$.
- $k = \lfloor (n/d)^{1/3} \rfloor^2$
- $D = c_1\sqrt{k}\cdot d = O(n^{1/3}\cdot d^{2/3})$, where c_1 is the constant from Proposition 16
- $R = 10n\log n$
- $N = \lceil c_2 R \log^2 R \rceil$, where c_2 is a universal constant to be determined later (cf. Proposition 18)
- $m = R \cdot N$

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.

$$\langle \varphi, f_n \rangle \geq 2/3 \tag{26}$$

$$\|\varphi\|_1 = 1 \tag{27}$$

$$\forall p\colon \{-1,1\}^n \to \mathbb{R},\ \deg p < d \implies \langle \varphi, p \rangle = 0 \tag{28}$$

- The function $\Psi : \{-1,1\}^{10\log n} \to \mathbb{R}$ whose existence is guaranteed by Proposition 10.
- The dual witness $\psi : \{-1,1\}^N \to \mathbb{R}$ for OR_N guaranteed by Proposition 16, 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 9).

Proposition 17. *The dual witness $\zeta = \varphi \star \Psi \star \psi$ satisfies the following properties:*

$$\langle \zeta, G \rangle \geq 1/2 \tag{29}$$

$$\|\zeta\|_1 = 1 \tag{30}$$

$$\forall p\colon ((\{-1,1\}^N)^{10\log n})^n \to \mathbb{R},\ \deg p < D \implies \langle \zeta, p \rangle = 0. \tag{31}$$

Proposition 17 follows by regarding ζ as $\varphi \star (\Psi \star \psi)$ and applying the hardness amplification results Proposition 10 and Proposition 11.

D. 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 12 when applied to the function $F_R := f_n \circ \mathsf{AND}_{10\log n}$ on $R = 10n\log n$ variables.

Proposition 18. *Let $\zeta : ((\{-1,1\}^N)^{10\log n})^n \to \mathbb{R}$ be as constructed in Proposition 17. Then there exists a constant $c_2 > 0$ such that, for $N = \lceil c_2 R \log^2 R \rceil$ and sufficiently large n,*

$$\sum_{x \notin X} |\zeta(x)| \leq (2NR)^{-2R/k} \leq (2NR)^{-2D}. \tag{32}$$

Proof: For the proof of Proposition 18, 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 $R = 10n\log n$.

By symmetry, the function $\psi(x)$ may be written as $\omega(|x|)/\binom{N}{|x|}$ where $\omega\colon [k]_0 \to \mathbb{R}$. We may decompose $\omega = \omega_{+1} - \omega_{-1}$ where ω_{+1} and ω_{-1} are non-negative functions satisfying

$$\sum_{t=0}^{k} \omega_{+1}(t) = \sum_{t=0}^{k} \omega_{-1}(t) = 1/2. \tag{33}$$

By the definition of dual block composition, we have

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

A calculation reveals that

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

where

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

To control this quantity, we appeal to the following combinatorial lemma, whose proof appears in the full version of this work.

Lemma 19. *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,*

$$\sum_{r=0}^{k} \eta_i(r) \leq 1/2 \tag{34}$$

$$\eta_i(r) \leq 5/(r+1)^2 \qquad \forall r = 0,1,\dots,k. \tag{35}$$

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

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

Observe that the functions ω_{z_i} satisfy Condition (34) (cf. Equation (33)) and Condition (35) (cf. Property (25)). We complete the proof of Proposition 18 by letting c_2 equal the constant α appearing in the statement of Lemma 19, 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 9), 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 IV-C. ∎

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 [47].

Lemma 20 (cf. [47, 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 is a function $\phi_y : \{-1,1\}^m \to \mathbb{R}$ such that*

$$\phi_y(y) = 1 \tag{36}$$

$$|x| > D, x \neq y \implies \phi_y(x) = 0 \tag{37}$$

$$\deg p < D \implies \langle \phi_y, p \rangle = 0 \tag{38}$$

$$\sum_{|x| \leq D} |\phi_y(x)| \leq 2^D \binom{|y|}{D}. \tag{39}$$

Proposition 21. *There exists a function $\nu : (\{-1,1\}^N)^{10\log n})^n \to \mathbb{R}$ such that*

$$\forall p\colon (\{-1,1\}^N)^{10\log n})^n \to \mathbb{R},\ \deg p < D \implies \langle \nu, p \rangle = 0 \tag{40}$$

$$\|\nu\|_1 \leq 1/10 \tag{41}$$

$$|x| > N \implies \nu(x) = \zeta(x), \tag{42}$$

where ζ is as in Proposition 17.

Proof: Define

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

where ϕ_y is as in Lemma 20 with m and D set as at the beginning of Section IV-C. Property (40) follows immediately from Property (38) and linearity. Property (41) follows from Proposition 18 and Properties (36), (37), and (39) of Proposition 21. Finally, Property (42) follows from (36) and (37), together with the fact that $D < N$. ∎

Combining Proposition 21 with Proposition 17 allows us to complete the proof of Theorem 14, which was the goal of this section.

Proof of Theorem 14: Let $\zeta = \varphi \star \Psi \star \psi$ be as defined in Section IV-C, and let ν be the correction object constructed in Proposition 21. Observe that $\|\zeta - \nu\|_1 > 0$, as $\|\zeta\|_1 = 1$ (cf. Equality (30)) and $\|\nu\|_1 \leq 1/10$ (cf. Inequality (41)). Define the function

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

Since $\nu(x) = \zeta(x)$ whenever $|x| > N$ (cf. Equation (42)), 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:

$$\langle \hat{\zeta}, G^{\leq N} \rangle \geq 1/3 \tag{43}$$

$$\|\hat{\zeta}\|_1 = 1 \tag{44}$$

$$\forall p\colon (\{-1,1\}^N)^{10\log n})^n \to \mathbb{R},\ \deg p < D \implies \langle \hat{\zeta}, p \rangle = 0. \tag{45}$$

Inequality (43) follows from the fact that $\zeta = \nu$ outside X, together with Properties (29), (30), and (41). Equation (44) is immediate from the definition of $\hat{\zeta}$. Finally, (45) follows from (31), (40), and linearity. ∎

V. Applications

A. 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 [53] 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 [39]. By combining the pattern matrix method with Theorems 3 and 4, we obtain the following corollary.

Corollary 22. *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 $\Omega(n^{1-\delta})$.*

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 15), with the pattern matrix method [53].

Subsequent to [53], a number of works [12], [26], [29], [40], [54], [56] generalized the pattern matrix method to the multiparty number-on-the-forehead model. Combining our new approximate degree bounds with sharpest version of these results [56] yields the following corollary.

Corollary 23. *For any integer $k \geq 1$ and any constant $\delta > 0$, there is an AC^0 function $F\colon (\{-1,1\}^n)^k \to \{-1,1\}$ such that the k-party quantum number-on-the-forehead communication complexity of F (with arbitrary prior entanglement), denoted $Q^*_{1/3}(F)$, is $\Omega((n/4^k k^2)^{1-\delta})$.*

The previous best lower bound for an AC^0 function was $\Omega_k(n^{2/3})$, again by applying the pattern matrix method to the Element Distinctness function. Moreover, this was the best-known lower bound even for *classical* randomized number-on-the-forehead communication complexity.

B. 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 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 [31], 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 the full version of this work, we study the relationship between certificate complexity, denoted $C(f)$, 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. 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 24. *There is a Boolean function $F : \{-1,1\}^n \to \{-1,1\}$ such that $\widetilde{\deg}(F) \geq C(F)^{2-o(1)}$.*

C. Secret Sharing Schemes

Bogdanov et al. [17] observed that for any $f\colon \{-1,1\}^n \to \{-1,1\}$ and integer $d > 0$, any dual polynomial μ for the fact that $\widetilde{\deg}_\varepsilon(f) \geq d$ leads to a scheme for sharing a single secret bit $b \in \{-1,1\}$ among n parties as follows. Decompose μ as $\mu_+ - \mu_-$, where μ_+ and μ_- are non-negative functions with $\|\mu_+\|_1 = \|\mu_0\|_1 = 1/2$. Then in order to split b among n parties, one draws an input $x = (x_1, \ldots, x_n) \in \{-1,1\}^n$ from the distribution $2 \cdot \mu_b$, and gives bit x_i to the ith party. In order to reconstruct b, one simply applies f to $(x_1, \ldots, x_n)$.

Because μ is ε-correlated with f, the probability of correct reconstruction if the bit is chosen at random is at least $(1+\varepsilon)/2$ (and the the *reconstruction advantage*, defined to equal $\Pr_{x\sim\mu_+}[f(x)=1] - \Pr_{x\sim\mu_-}[f(x)=1]$, is at least ε). The fact that μ has pure high degree at least d means that any subset of shares of size less than d provides no information about the secret bit b. We direct the interested reader to [17] for further details.

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

Corollary 25. *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) Reconstruction 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. [17], who used the Element Distinctness lower bound (cf. Theorem 15) 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.

VI. Future Directions

A. 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 full 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 SURJECTIVITY function remains unresolved. No approximating polynomial of degree $o(n)$ is known, 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], [16]. 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 $k \to \infty$ [15].)

B. 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., [18], [22], [60] and the references therein).

Despite attention by many researchers, the best known lower bounds in the directions of Problems 2 and 3 are:

(a) For any constant $\Gamma > 0$, a depth-3 circuit with ε-approximate degree $\Omega(n^{1/2-\delta})$ for $\varepsilon = 1 - 2^{-n^{\Gamma}}$ [22],
(b) A depth-3 circuit with threshold degree $\Omega(n^{3/7})$ [49], and
(c) A depth-4 circuit with threshold degree $\Omega(n^{1/2})$ [49].

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 I-B1 that our recent work [22] also proved stronger hardness amplification results for approximate degree by moving beyond block composed functions. The methods of [22] 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 [21]. Moreover, this should be provable via a modest extension of our techniques. Combining such a lower bound with a result of Sherstov [60] 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 [60] constructed a dual polynomial ψ witnessing the fact that

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

(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 III 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.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] S. Aaronson, "The polynomial method in quantum and classical computing," in *FOCS*, 2008.
[2] ——, "Impossibility of succinct quantum proofs for collision-freeness," *Quantum Information & Computation*, vol. 12, no. 1-2, pp. 21–28, 2012.
[3] S. Aaronson, S. Ben-David, and R. Kothari, "Separations in query complexity using cheat sheets," in *STOC*, 2016.
[4] S. Aaronson and Y. Shi, "Quantum lower bounds for the collision and the element distinctness problems," *J. ACM*, vol. 51, no. 4, pp. 595–605, 2004.
[5] A. Ambainis, "Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range," *Theory of Computing*, vol. 1, no. 1, pp. 37–46, 2005.
[6] ——, "Quantum walk algorithm for element distinctness," *SIAM J. Comput.*, vol. 37, no. 1, pp. 210–239, 2007.
[7] A. Ambainis, K. Balodis, A. Belovs, T. Lee, M. Santha, and J. Smotrovs, "Separations in query complexity based on pointer functions," in *STOC*, 2016.
[8] A. Ambainis, A. M. Childs, B. Reichardt, R. Spalek, and S. Zhang, "Any and-or formula of size n can be evaluated in time $n^{1/2+o(1)}$ on a quantum computer," *SIAM J. Comput.*, vol. 39, no. 6, pp. 2513–2530, 2010.
[9] A. Anshu, S. Ben-David, A. Garg, R. Jain, R. Kothari, and T. Lee, "Separating quantum communication and approximate rank," *CoRR*, vol. abs/1611.05754, 2016.
[10] ——, Personal communication, 2017.
[11] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf, "Quantum lower bounds by polynomials," *J. ACM*, vol. 48, no. 4, pp. 778–797, 2001.
[12] P. Beame and T. Huynh, "Multiparty communication complexity and threshold circuit size of AC^0," *SIAM J. Comput.*, vol. 41, no. 3, pp. 484–518, 2012.
[13] P. Beame and W. Machmouchi, "The quantum query complexity of AC^0," *Quantum Information & Computation*, vol. 12, no. 7-8, pp. 670–676, 2012.
[14] R. Beigel, "Perceptrons, PP, and the Polynomial Hierarchy," *Computational Complexity*, vol. 4, pp. 339–349, 1994.
[15] A. Belovs, "Learning-graph-based quantum algorithm for k-distinctness," in *FOCS*, 2012.
[16] A. Belovs and R. Spalek, "Adversary lower bound for the k-sum problem," in *ITCS*, 2013.
[17] A. Bogdanov, Y. Ishai, E. Viola, and C. Williamson, "Bounded indistinguishability and the complexity of recovering secrets," in *CRYPTO*, 2016.

[18] A. Bouland, L. Chen, D. Holden, J. Thaler, and P. N. Vasudevan, "On SZK and PP," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 23, p. 140, 2016.

[19] H. Buhrman, N. K. Vereshchagin, and R. de Wolf, "On computation and communication with small bias," in *CCC*, 2007.

[20] M. Bun and J. Thaler, "Dual lower bounds for approximate degree and Markov-Bernstein inequalities," in *ICALP*, 2013.

[21] ——, "Hardness amplification and the approximate degree of constant-depth circuits," in *ICALP*, 2015.

[22] ——, "Approximate degree and the complexity of depth three circuits," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 23, p. 121, 2016.

[23] ——, "Dual polynomials for Collision and Element Distinctness," *Theory of Computing*, vol. 12, no. 16, pp. 1–34, 2016.

[24] ——, "Improved bounds on the sign-rank of AC^0," in *ICALP*, 2016.

[25] K. Chandrasekaran, J. Thaler, J. Ullman, and A. Wan, "Faster private release of marginals on small databases," in *ITCS*, 2014.

[26] A. Chattopadhyay and A. Ada, "Multiparty communication complexity of disjointness," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 15, no. 002, 2008.

[27] L. Chen, "Adaptivity vs. postselection, and hardness amplification for polynomial approximation," in *ISAAC*, 2016.

[28] M. David and T. Pitassi, "Separating NOF communication complexity classes RP and NP," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 15, no. 014, 2008.

[29] M. David, T. Pitassi, and E. Viola, "Improved separations between nondeterministic and randomized multiparty communication," *TOCT*, vol. 1, no. 2, 2009.

[30] D. Gavinsky and A. A. Sherstov, "A separation of NP and coNP in multiparty communication complexity," *Theory of Computing*, vol. 6, no. 1, pp. 227–245, 2010.

[31] M. Göös, T. Pitassi, and T. Watson, "Deterministic communication vs. partition number," in *FOCS*, 2015.

[32] P. Høyer, M. Mosca, and R. de Wolf, "Quantum search on bounded-error inputs," in *ICALP*, 2003.

[33] J. Kahn, N. Linial, and A. Samorodnitsky, "Inclusion-exclusion: Exact and approximate," *Combinatorica*, vol. 16, no. 4, pp. 465–477, 1996.

[34] A. T. Kalai, A. R. Klivans, Y. Mansour, and R. A. Servedio, "Agnostically learning halfspaces," *SIAM J. Comput.*, vol. 37, no. 6, pp. 1777–1805, 2008.

[35] V. Kanade and J. Thaler, "Distribution-independent reliable learning," in *COLT*, 2014.

[36] A. R. Klivans and R. A. Servedio, "Learning DNF in time $2^{\tilde{O}(n^{1/3})}$," *J. Comput. Syst. Sci.*, vol. 68, no. 2, pp. 303–318, 2004.

[37] ——, "Toward attribute efficient learning of decision lists and parities," *Journal of Machine Learning Research*, vol. 7, pp. 587–602, 2006.

[38] T. Lee, "A note on the sign degree of formulas," *CoRR*, vol. abs/0909.4607, 2009.

[39] T. Lee and A. Shraibman, "An approximation algorithm for approximation rank," in *CCC*, 2009.

[40] ——, "Disjointness is hard in the multiparty number-on-the-forehead model," *Computational Complexity*, vol. 18, no. 2, pp. 309–336, 2009.

[41] M. Minsky and S. Papert, *Perceptrons - an introduction to computational geometry*. MIT Press, 1969.

[42] N. Nisan and M. Szegedy, "On the degree of boolean functions as real polynomials," *Computational Complexity*, vol. 4, pp. 301–313, 1994.

[43] R. O'Donnell and R. A. Servedio, "New degree bounds for polynomial threshold functions," *Combinatorica*, vol. 30, no. 3, pp. 327–358, 2010.

[44] R. Paturi, "On the degree of polynomials that approximate symmetric boolean functions (preliminary version)," in *STOC*, 1992.

[45] V. V. Podolskii, "A uniform lower bound on weights of perceptrons," in *Computer Science Symposium in Russia (CSR)*, 2008.

[46] A. Rao and A. Yehudayoff, "Simplified lower bounds on the multiparty communication complexity of disjointness," in *CCC*, 2015.

[47] A. A. Razborov and A. A. Sherstov, "The sign-rank of AC^0," *SIAM J. Comput.*, vol. 39, no. 5, pp. 1833–1855, 2010.

[48] R. A. Servedio, L.-Y. Tan, and J. Thaler, "Attribute-efficient learning and weight-degree tradeoffs for polynomial threshold functions," in *COLT*, 2012.

[49] A. A. Sherstov, "The power of asymmetry in constant-depth circuits," in *FOCS*, 2015.

[50] ——, "Communication lower bounds using dual polynomials," *Bulletin of the EATCS*, vol. 95, pp. 59–93, 2008.

[51] ——, "Approximate inclusion-exclusion for arbitrary symmetric functions," *Computational Complexity*, vol. 18, no. 2, pp. 219–247, 2009.

[52] ——, "Separating AC^0 from depth-2 majority circuits," *SIAM J. Comput.*, vol. 38, no. 6, pp. 2113–2129, 2009.

[53] ——, "The pattern matrix method," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1969–2000, 2011.

[54] ——, "The multiparty communication complexity of set disjointness," in *STOC*, 2012.

[55] ——, "Approximating the AND-OR Tree," *Theory of Computing*, vol. 9, no. 20, pp. 653–663, 2013.

[56] ——, "Communication lower bounds using directional derivatives," in *STOC*, 2013.

[57] ——, "The intersection of two halfspaces has high threshold degree," *SIAM J. Comput.*, vol. 42, no. 6, pp. 2329–2374, 2013.

[58] ——, "Making polynomials robust to noise," *Theory of Computing*, vol. 9, pp. 593–615, 2013.

[59] ——, "Optimal bounds for sign-representing the intersection of two halfspaces by polynomials," *Combinatorica*, vol. 33, no. 1, pp. 73–96, 2013.

[60] ——, "Breaking the Minsky-Papert barrier for constant-depth circuits," in *STOC*, 2014.

[61] Y. Shi, "Approximating linear restrictions of boolean functions," 2002, manuscript.

[62] Y. Shi and Y. Zhu, "Quantum communication complexity of block-composed functions," *Quantum Information & Computation*, vol. 9, no. 5, pp. 444–460, 2009.

[63] R. Spalek, "A dual polynomial for OR," *CoRR*, vol. abs/0803.4516, 2008.

[64] A. Tal, "Shrinkage of de morgan formulae from quantum query complexity," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 21, p. 48, 2014.

[65] ——, "The bipartite formula complexity of inner-product is quadratic," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 23, p. 181, 2016.

[66] ——, "Computing requires larger formulas than approximating," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 23, p. 179, 2016.

[67] J. Thaler, "Lower bounds for the approximate degree of block-composed functions," in *ICALP*, 2016.

[68] J. Thaler, J. Ullman, and S. P. Vadhan, "Faster algorithms for privately releasing marginals," in *ICALP*, 2012.