Electronic Colloquium on Computational Complexity, Revision 3 of Report No. 150 (2014)

Lower Bounds for the Approximate Degree of Block-Composed Functions

Justin Thaler*

Abstract

We describe a new hardness amplification result for point-wise approximation of Boolean functions by low-degree polynomials. Specifically, for any function f on N bits, define $F(x_1,\ldots,x_M) = \mathrm{OMB}(f(x_1),\ldots,f(x_M))$ to be the function on $M \cdot N$ bits obtained by block-composing f with a function known as ODD-MAX-BIT. We show that, if f requires large degree to approximate to error $2/3$ in a certain one-sided sense (captured by a complexity measure known as *positive one-sided approximate degree*), then F requires large degree to approximate even to error $1-2^{-M}$. This generalizes a result of Beigel (Computational Complexity, 1994), who proved an identical result for the special case $f = \mathrm{OR}$.

Unlike related prior work, our result implies strong approximate degree lower bounds even for many functions F that have low *threshold degree*. Our proof is constructive: we exhibit a solution to the dual of an appropriate linear program capturing the approximate degree of any function. We describe several applications, including improved separations between the complexity classes $\mathbf{P}^{\mathbf{NP}}$ and **PP** in both the query and communication complexity settings. Our separations improve on work of Beigel (1994) and Buhrman, Vereshchagin, and de Wolf (CCC, 2007).

*Yahoo Labs.

ISSN 1433-8092

1 Introduction

Approximate degree and threshold degree are two measures of Boolean function complexity that capture the difficulty of point-wise approximation by low-degree polynomials. The ε-approximate degree of a function $f\colon \{-1,1\}^n \to \{-1,1\}$, denoted $\widetilde{\deg}_\varepsilon(f)$, is the least degree of a real polynomial that point-wise approximates f to error ε. The threshold degree of f, denoted $\deg_\pm(f)$, is the least degree of a real polynomial that agrees in sign with f at all points.

Approximate degree and threshold degree have found a diverse array of algorithmic and complexity-theoretic applications. On the complexity side, approximate degree lower bounds underlie many tight lower bounds on quantum query complexity [1, 3, 7, 28, 45], and have proven instrumental in resolving a host of long-standing open problems in communication and circuit complexity [11, 15–17, 19, 30, 37, 38, 43–46, 48]. On the algorithms side, upper bounds on these complexity measures underlie the fastest known learning algorithms in a number of important models, including the PAC, agnostic, and mistake-bounded models [23, 26, 27, 39]. They also yield fast algorithms for private data release [14, 53].

Despite these applications, our understanding of approximate and threshold degree remains limited. While tight upper and lower bounds are known for some specific functions, including symmetric functions [18, 35, 42] and certain read-once formulae, few general results are known, and characterizing the approximate and threshold degrees of many simple functions remains open. However, a handful of recent works has established various forms of "hardness amplification" for approximate degree [12, 13, 29, 40, 47, 49, 51]. Roughly speaking, these results show how to take a function f which is hard to approximate by low-degree polynomials in a weak sense, and turn f into a related function F that is hard to approximate by low-degree polynomials in a much stronger sense.

Our Contributions. We extend this recent line of work by establishing a new, generic form of hardness amplification for approximate degree. Unlike prior work, our result implies strong lower bounds even for many functions F that have low threshold degree (e.g., halfspaces). In contrast, analogous hardness amplification results [12, 13, 29, 40, 47, 49, 51] apply only to functions with polynomially large threshold degree. We describe several applications of our result, including an improved separation between the complexity classes $\mathbf{P}^{\mathbf{NP}}$ and $\mathbf{PP}$ in both the query and communication complexity settings (see Section 1.3 for details).

We prove our results by constructing explicit *dual polynomials*, which are dual solutions to an appropriate linear program capturing the approximate degree of any function. This "method of dual polynomials" has proven to be a powerful technique for establishing lower bounds on approximate degree. Our construction departs qualitatively from earlier applications of the method, and we believe it to be of interest in its own right. In addition to implying approximate degree lower bounds, dual polynomials have been used to resolve several long-standing open problems in communication complexity, and they yield explicit distributions under which various communication problems are hard [17, 19, 37, 44–46, 48].

1.1 Overview of Our Results

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Our hardness amplification method relies heavily on a complexity measure known as *one-sided approximate degree*, or, more precisely, its "positive" and "negative" variants, denoted $\widetilde{\deg}_{+,\varepsilon}(f)$ and $\widetilde{\deg}_{-,\varepsilon}(f)$ respectively. These are intermediate complexity measures that lie between ε-approximate degree and threshold degree, and they have played a central role in recent prior work on hardness amplification for approximate degree [12, 13, 47, 51].[1] Unlike the latter two complexity measures, $\widetilde{\deg}_{+,\varepsilon}(f)$ and $\widetilde{\deg}_{-,\varepsilon}(f)$ treat inputs in $f^{-1}(+1)$ and inputs in $f^{-1}(-1)$ asymmetrically.

[1] Strictly speaking, the terms positive and negative one-sided approximate degree were introduced by Kanade and Thaler [24], who gave applications of these complexity measures to learning theory. Earlier works on hardness amplification for pointwise approximation by polynomials only used negative one-sided approximate degree, and referred to this complexity measure without qualification as one-sided approximate degree [13, 51]. For our purposes, the distinction between positive and negative one-sided approximate degree is crucial.

In more detail, a polynomial p is said to be a positive one-sided ε-approximation for a Boolean function f if $|p(x)-f(x)| \leq \varepsilon$ for all $x \in f^{-1}(-1)$, and $p(x) \geq 1-\varepsilon$ for all $x \in f^{-1}(+1)$. The positive one-sided ε-approximate degree of f is the least degree of a positive one-sided ε-approximation for f. Negative one-sided ε-approximate degree is defined analogously. (Appendix A contains formal definitions.) Notice that $\widetilde{\deg}_{+,\varepsilon}(f)$ and $\widetilde{\deg}_{-,\varepsilon}(f)$ are always at most $\widetilde{\deg}_{\varepsilon}(f)$, but can be much smaller. Similarly, $\widetilde{\deg}_{+,\varepsilon}(f)$ and $\widetilde{\deg}_{-,\varepsilon}(f)$ are always at least $\deg_{\pm}(f)$, but can be much larger.

Let $\mathrm{OMB}: \{-1,1\}^n \to \{-1,1\}$ denote a specific polynomial size DNF formula known as ODD-MAX-BIT, defined as follows. On input $x=(x_1,\ldots,x_n)$, let i^* denote the largest index such that $x_{i^*}=-1$, and let $i^*=0$ if no such index exists. We define

$$\mathrm{OMB}(x_1,\ldots,x_n) = \begin{cases} -1 \text{ if } i^* \text{ is odd} \\ 1 \text{ otherwise} \end{cases}$$

When appropriate, we also use subscripts after function symbols to indicate the number of variables over which the function is defined. Thus, OMB_M denotes the OMB function on M inputs.

For any function $f\colon \{-1,1\}^N \to \{-1,1\}$, define $F: \left(\{-1,1\}^N\right)^M \to \{-1,1\}$ to be the block-composition of OMB_M with f, i.e., $F=\mathrm{OMB}_M(f,\ldots,f)$. Our hardness amplification result establishes that if $\widetilde{\deg}_{+,\varepsilon}(f)$ is large for some ε bounded away from 1, then $\widetilde{\deg}_{+,\varepsilon}(F)$ is large even for ε exponentially close to 1.

Theorem 1. *Fix an $f\colon \{-1,1\}^N \to \{-1,1\}$, and let $F=\mathrm{OMB}_M(f,\ldots,f)$. If $\widetilde{deg}_{+,2/3}(f) \geq d$, then $\widetilde{deg}_{+,\varepsilon}(F) \geq d$ for $\varepsilon = 1-2^{-M}$.*

A Matching Upper Bound for Theorem 1. To understand the intuition underlying Theorem 1, it is instructive to consider (matching) upper bounds. We begin by giving the well-known sign-representing polynomial for OMB_M itself. Define $p\colon \{-1,1\}^M \to \mathbb{R}$ via

$$p(x_1,\ldots,x_M) := 1+\sum_{i=1}^{M}(-2)^i \cdot (1-x_i)/2.$$

It is easy to see that $\mathrm{OMB}_M(x)=\mathrm{sgn}\,(p(x))$, and in fact $2^{-M-1}\cdot p(x)$ approximates OMB_M to error $\varepsilon = 1-2^{-M-1}$.

We now turn to constructing approximants for $\mathrm{OMB}_M(f,\ldots,f)$, for an arbitrary inner function f. Fix a $W \geq 2$, and let $q\colon \{-1,1\}^N \to \mathbb{R}$ be any degree d polynomial satisfying the following two properties.

$$q(x) = 0 \text{ for all } x \in f^{-1}(+1). \tag{1}$$

$$1 \leq q(x) \leq W-1 \text{ for all } x \in f^{-1}(-1). \tag{2}$$

Denoting an $(M\cdot N)$-bit input as $(x_1,\ldots,x_M) \in \left(\{-1,1\}^N\right)^M$, it is easy to check that

$$F(x_1,\ldots,x_M) = \mathrm{sgn}(h(x_1,\ldots,x_M)), \text{ where } h(x_1,\ldots,x_M) = 1+\sum_{i=1}^{M}(-W)^i \cdot q(x_i).$$

In fact, $W^{-M-1}\cdot h(x)$ approximates F to error $1-W^{-M-1}$, and has degree equal to that of q. If $W=O(1)$, then this construction shows that F can be approximated to error $1-2^{-O(M)}$ by a degree d polynomial, which matches the error bound of Theorem 1 up to a constant factor in the exponent.

Observation 2. *If there exists a polynomial q of degree d satisfying Eq. (1) and Eq. (2) with $W=O(1)$, then $\widetilde{deg}_{\varepsilon}(F) \leq d$ for some $\varepsilon = 1-2^{-O(M)}$.*

A few words are in order regarding the relationship between the hypothesis of the upper bound (Observation 2), and the hypothesis of the lower bound (Theorem 1) that $\widetilde{\deg}_{+,2/3}(f) \geq d$. Notice that Conditions 1 and 2 together imply that $r(x) := \frac{1}{2W}\cdot(1-2q(x))$ is a positive one-sided approximation to f for error

parameter $\varepsilon = 1 - \frac{1}{2W}$. Moreover, r has the additional (crucial) property that this approximant is *constant* on inputs in $f^{-1}(+1)$. Observe that the smaller W is, the smaller the error of the one-sided approximant $r(x)$ for $f(x)$, and the smaller the error of the derived approximant $W^{-M-1} \cdot h(x)$ that we constructed for F.

In general, requiring that r be constant on inputs in $f^{-1}(+1)$ is a very stringent condition, which will not be satisfied by all one-sided approximations for f. However, Bun and Thaler [13, Theorem 2] have identified a large class of functions for which *any* one-sided approximation for f can be transformed into one that is constant on inputs in $f^{-1}(+1)$, without increasing its degree. This class includes important functions such as $f = \mathrm{OR}$ (see Section 1.2.2), and $f = \overline{\mathrm{ED}}$, where ED is the well-studied Element Distinctness function that we use in our applications to communication and query complexity (see Section 4.1.3 for the definition of ED). For such functions, Observation 2 implies that Theorem 1 is tight.

Can the Hypothesis in Theorem 1 be Weakened? There are two natural ways to weaken the hypothesis of Theorem 1, and it is natural to wonder whether Theorem 1 would continue to hold under these hypotheses. Specifically, we can ask:

- Does Theorem 1 hold if we replace the outer function OMB_M function with the simpler function OR_M, as in previous hardness amplification results for approximate degree [12, 13, 47, 51]?[2]
- Is a one-sided hardness assumption really essential for Theorem 1 to hold? That is, does OMB_M still amplify the hardness of f if we replace the assumption that $\widetilde{\deg}_{+,2/3}(f) \geq d$ with the weaker assumption that $\widetilde{\deg}_{2/3}(f) \geq d$?

The answer to the first question is no. A counterexample is given by $f = \mathrm{OR}_N$. It is known that $\widetilde{\deg}_{+,2/3}(\mathrm{OR}_N) = \Omega(N^{1/2})$ (see, e.g., [13, 20, 33]), yet $\mathrm{OR}_M(\mathrm{OR}_N, \ldots, \mathrm{OR}_N) = \mathrm{OR}_{N \cdot M}$ can be approximated to error $1 - 1/(MN) \ll 1 - 2^{-M}$ by a polynomial of degree 1. Thus, the use of OMB_M as the "hardness amplifier" is essential to Theorem 1.

The answer to the second question, unfortunately, remains unknown. Formally, we leave the resolution of the following conjecture as an open problem.

Conjecture 3. *Suppose that $f\colon \{-1,1\}^N \to \{-1,1\}$ satisfies $\widetilde{deg}_{2/3}(f) \geq d$. Then letting $F = \mathrm{OMB}_M(f, \ldots, f)$, it holds that $\widetilde{deg}_{\varepsilon}(\mathrm{OMB}_M(f, \ldots, f)) \geq d$, for some $\varepsilon = 1 - 2^{-\Omega(M)}$.*

1.2 Technical Comparison to Prior Work

1.2.1 The Method of Dual Polynomials

A dual witness to the statement $\widetilde{\deg}_{\varepsilon}(f) \geq d$ is a non-zero function $\psi : \{-1,1\}^N \to \mathbb{R}$ satisfying two conditions: (a) $\sum_{x \in \{-1,1\}^N} \psi(x) \cdot f(x) \geq \varepsilon \cdot \|\psi\|_1$, where $\|\psi\|_1 = \sum_{x \in \{-1,1\}^N} |\psi(x)|$, and (b) ψ has zero correlation with all polynomials of degree at most d. We refer to Property (a) by saying that ψ is *ε-correlated* with f. We refer to Property (b) by saying that ψ has *pure high degree d*. We refer to ψ as a *dual polynomial* for f.

A dual witness to the statement that $\widetilde{\deg}_{+,\varepsilon}(f) \geq d$ must satisfy an additional correlation condition, namely: (c) $\phi(x)$ agrees in sign with $f(x)$ for all $x \in f^{-1}(+1)$. We refer to Property (c) by saying that ϕ has *positive one-sided error*. (See Appendix A for details of the duality theory.)

We prove Theorem 1 by showing the following: given a dual polynomial ψ_{in} witnessing the assumed $\widetilde{\deg}_{+,2/3}$ lower bound on the inner function f, one can construct an explicit dual polynomial ψ_{comb} witnessing the claimed lower bound on the composed function $F = \mathrm{OMB}(f, \ldots, f)$.

[2] One may also ask about replacing OMB_M with AND_M in the statement of Theorem 1. Analyses from prior works [13, 51] apply in this case, but show that the resulting function in fact has high threshold degree, and hence is not suitable for our applications to query and communication complexity. We discuss this point in detail in the next section (see Theorem 5, Footnote 5, and the surrounding discussion).

1.2.2 Prior Work on the Approximate Degree of OMB

Beigel [8] proved that for any $d > 0$, there is an $\varepsilon \in 1 - 2^{-\Omega(n/d^2)}$ such that $\widetilde{\deg}_{\varepsilon}(\mathrm{OMB}_n) \geq d$, and used this result[3] to give an oracle separating the (Turing Machine) complexity class **PP** from $\mathbf{P}^{\mathbf{NP}}$. Note that $\mathrm{OMB}_M(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ is a sub-function of $\mathrm{OMB}_{M\cdot(2N)}$. As mentioned in Section 1.1, it is known that $\widetilde{\deg}_{+,2/3}(\mathrm{OR}_N) = \Omega(N^{1/2})$. Hence, Theorem 1 can be viewed as a substantial strengthening of Beigel's result: we recover Beigel's lower bound as a special case of Theorem 1 by letting $f = \mathrm{OR}_{d^2}$. Unlike Beigel's proof, which used a non-constructive symmetrization technique, our proof of Theorem 1 constructs an explicit dual polynomial witnessing the lower bound.

For any $\varepsilon > 0$, Klivans and Servedio [27] gave an optimal ε-approximating polynomial for the function OMB, showing that Beigel's lower bound (and hence also our Theorem 1 in the case $f = \mathrm{OR}_N$) is asymptotically tight for all $d > 0$.[4]

1.2.3 Earlier Constructions of Dual Polynomials for Block-Composed Functions

Given functions g_M, f_N, Sherstov [49] and Lee [29] independently described a powerful method for constructing a dual polynomial for the composed function $F = g_M(f_N, \ldots, f_N) : \{-1,1\}^{M\cdot N} \to \{-1,1\}$. This method takes a dual polynomial ψ_{in} for f_N, and a dual polynomial ψ_{out} for g, and combines them to obtain a dual polynomial ψ_{comb} for the composed function F.

Specifically, denoting an $(M \cdot N)$-bit input as $(x_1, \ldots, x_M) \in \left(\{-1,1\}^N\right)^M$, Sherstov and Lee defined

$$\psi_{\mathrm{comb}}(x_1, \ldots, x_M) = \psi_{\mathrm{out}}\left(\widetilde{\mathrm{sgn}}\left(\psi_{\mathrm{in}}(x_1)\right), \ldots, \widetilde{\mathrm{sgn}}\left(\psi_{\mathrm{in}}(x_M)\right)\right) \cdot \prod_{i=1}^{M} |\psi_{\mathrm{in}}(x_i)|. \tag{3}$$

Here, $\widetilde{\mathrm{sgn}} : \mathbb{R} \to \{-1,0,1\}$ denotes the function satisfying $\widetilde{\mathrm{sgn}}(t) = 1$ if $t > 0$, $\widetilde{\mathrm{sgn}}(t) = -1$ if $t < 0$, and $\widetilde{\mathrm{sgn}}(0) = 0$.

Recall that for ψ_{comb} to witness a good lower bound for the approximate degree of F, it must be well-correlated with F (Property (a) of Section 1.2.1), and it must have large pure high degree (Property (b) of Section 1.2.1). Sherstov and Lee showed that the pure high degree of ψ_{comb} is multiplicative in the pure high degrees of ψ_{in} and ψ_{out}. That is, if ψ_{in} has pure high degree d_1, and ψ_{out} has pure high degree d_2, then ψ_{comb} has pure high degree $d_1 \cdot d_2$. And while ψ_{comb} is not *in general* well-correlated with the composed function F, several important examples have been identified in which this is the case, as we now explain.

Sherstov [47] and independently Bun and Thaler [12] used the combining technique of Eq. (3) to resolve the (1/3)-approximate degree of the two-level AND-OR tree. Subsequent work by Bun and Thaler [13] used Eq. (3) to establish a hardness amplification result that looks similar to our Theorem 1. Specifically, Bun and Thaler proved:

Theorem 4 (Bun and Thaler [13])**.** *Suppose* $\widetilde{deg}_{-,2/3}(f) \geq d$. *Then* $\widetilde{deg}_{-,\varepsilon}(OR_M(f, \ldots, f)) \geq d$, *for* $\varepsilon = 1 - 2^{-M}$.

Theorem 4 is identical to our Theorem 1, but for two differences: first, in our Theorem 1, the outer function in the composition is OMB, while in Theorem 4 it is OR. Second, the hypothesis in Theorem 1 is that the inner function f satisfies $\widetilde{\deg}_{+,2/3}(f) \geq d$, while the assumption in Theorem 4 is that $\widetilde{\deg}_{-,2/3}(f) \geq d$. *Both* of these differences are crucial for obtaining a hardness amplification result that applies to functions with low threshold degree (which is essential for our applications to the communication and query complexity described in Section 1.3 below). Indeed, subsequent work by Sherstov refined Theorem 4 to yield a threshold degree lower bound, rather than a $\widetilde{\deg}_{-,\varepsilon}$ lower bound [51].

[3] Beigel describes his result as a lower bound on the *degree-d threshold weight* of OMB_n. However, his argument is easily seen to establish the claimed approximate degree lower bound.

[4] Like Beigel, Klivans and Servedio state their results in terms of degree-d threshold weight. However, their construction is easily seen to imply the claimed upper bound on the approximate degree of OMB_n.

Theorem 5 (Sherstov [51]). *Suppose* $\widetilde{deg}_{-,2/3}(f) \geq d$. *Then* $\deg_{\pm}(OR_M(f,\ldots,f)) \geq \min\{d, cM\}$ *for some constant* $c > 0$.[5]

Sherstov gives several proofs of Theorem 5. One of them draws heavily on Eq. (3): he constructs a dual witness of the form $\psi_{\text{comb}} + \psi_{\text{fix}}$, where ψ_{comb} is the dual witness constructed by Bun and Thaler using Eq. (3) to prove Theorem 4, and ψ_{fix} is used "zero out" ψ_{comb} on points x such that $0 \neq \widetilde{\text{sgn}}(\psi_{\text{comb}}(x)) \neq \widetilde{\text{sgn}}(\text{OR}_M(f,\ldots,f))$. This ensures that $\psi_{\text{comb}} + \psi_{\text{fix}}$ is perfectly correlated with F.

Sherstov used Theorem 5 to give a depth three circuit with threshold degree $\tilde{\Omega}(n^{2/5})$. He also established the following result, which yields a polynomially stronger lower bound for depth $k > 3$.

Theorem 6 (Sherstov [51]). *For any* $k \geq 2$, *there is a depth* k *(read-once) Boolean circuit computing a function* F *satisfying* $\deg_{\pm}(F) = \Omega(n^{(k-1)/(2k-1)})$.

Sherstov's proof of Theorem 6 is not a refinement of the proof Theorem 4 from [13]. Rather it relies on an elaborate inductive construction of a dual polynomial (which is nonetheless reminiscent of Eq. (3)).

1.2.4 Complementary Slackness and the Need for New Techniques

In this section, we explain why any dual witness establishing Theorem 1 must qualitatively depart from the dual witnesses constructed in prior work (cf. Section 1.2.3). In brief, we first argue that the dual witnesses constructed in prior work are implicitly tailored to show optimality of a specific technique for approximating block-composed functions. We then explain that this technique is far from optimal for the functions to which Theorem 1 applies.

Approximating Block-Composed Functions via "Robustification". Sherstov [50] provided a generic technique for approximating block-composed functions. Specifically, he showed that for any polynomial $p : \{-1,1\}^M \to [-1,1]$, and every $\delta > 0$, there is a polynomial $p_{\text{robust}} : \mathbb{R}^M \to \mathbb{R}$ 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 $g = g_M, f = f_N$, one can obtain an $(\varepsilon + \delta)$-approximating polynomial for the block-composition $g(f,\ldots,f)$ as follows: let p be an ε-approximating polynomial for g, and q a $(1/3)$-approximating polynomial for f. Then the block composition $p^* := p_{\text{robust}}(q,\ldots,q)$ is an $(\varepsilon+\delta)$-approximating polynomial for $g(f,\ldots,f)$. Notice that the degree of p^* is at most the product of the degrees of p_{robust} and q.

This generic construction yields asymptotically optimal ε-approximating polynomials for essentially all block-composed functions considered in prior work on hardness amplification. Indeed, this holds for the two-level AND-OR tree when $\varepsilon = 1/3$ [12, 47], as well as for the functions considered in Theorems 4, 5, and 6, for ε exponentially close to 1 (see e.g. [51, Theorem 1.2]).

Showing Robustification Is Optimal (Except When It's Not). Intuitively, the dual witness ψ_{comb} constructed via Eq. (3) is specifically tailored to show optimality of the above generic technique for approximating block-composed functions. Indeed, ψ_{comb} "almost" obeys complementary slackness with respect to p^* in the following sense.

Suppose that p_{robust} achieved *exactly* optimal error ε among all degree d polynomial approximations to the outer function g. Then p_{robust} yields an optimal solution to the relevant linear program capturing the ε-approximate degree of g (cf. Appendix A). Complementary slackness states that there is an optimal dual solution (i.e., a weighting of the constraints from the primal linear program) which places non-zero weight only on constraints that are made tight by the primal optimum. In our context, this means that there is an optimal dual polynomial ψ_{out} for g such that $\psi_{\text{out}}(y) \neq 0$ only for "maximal error points" $y \in \{-1,1\}^M$, i.e., points y satisfying $|p_{\text{robust}}(y) - g(y)| = \varepsilon$. Let ψ_{in} be any dual polynomial for the inner function f, and

[5] By De Morgan's laws and the observation that $\widetilde{\deg}_{-,\varepsilon}(f) = \widetilde{\deg}_{+,\varepsilon}(\overline{f})$, the following is an equivalent formulation of Theorem 5. Suppose that $\widetilde{\deg}_{+,2/3}(f) \geq d$. Then $\deg_{\pm}(\text{AND}_M(f,\ldots,f)) \geq \min\{d, cM\}$ for some constant $c > 0$.

suppose ψ_{out} is combined with ψ_{in} as per Eq. (3) to obtain a dual polynomial ψ_{comb} for $g(f,\ldots,f)$. If ψ_{in} were *perfectly* correlated with f, then $\psi_{\text{comb}}(x) \neq 0$ only for $x = (x_1,\ldots,x_M) \in \left(\{-1,1\}^N\right)^M$ such that

$$|p_{\text{robust}}(q(x_1),\ldots,q(x_M)) - g(f(x_1),\ldots,f(x_M))| =$$
$$|p_{\text{robust}}(f(x_1),\ldots,f(x_M)) - g(f(x_1),\ldots,f(x_M))| \pm \delta \geq \varepsilon - \delta \approx \varepsilon.$$

Put another way, ψ_{comb} places non-zero weight only on points on which $p_{\text{robust}}(q,\ldots,q)$ achieves "nearly maximal error" of at least $\varepsilon - \delta$. This is what we mean when we say that ψ_{comb} "almost" satisfies complementary slackness with respect to the primal solution corresponding to $p_{\text{robust}}(q,\ldots,q)$.

In general, ψ_{in} will not be perfectly correlated with f, but the analyses of ψ_{comb} from prior work identify settings in which ψ_{comb} still places "most" of its weight on points x such that $|p_{\text{robust}}(q,\ldots,q) - g(f,\ldots,f)| = \varepsilon \pm \delta \approx \varepsilon$.

When Robustification Is Sub-Optimal. In contrast to these earlier results, Theorem 1 applies to functions for which $p^* := p_{\text{robust}}(q,\ldots,q)$ is not an optimal approximating polynomial. To see this, recall the approximation for $\text{OMB}_M(f,\ldots,f)$ described in the discussion surrounding Observation 2. This approximation was of the form $p(q,\ldots,q)$, where p was a *non-robust* ε-approximating polynomial for OMB_M (for some $\varepsilon = 1 - 2^{-\Theta(M)}$), and q is an approximating polynomial for f.

Since $p_{\text{robust}}(q,\ldots,q)$ is not an optimal approximating polynomial for $\text{OMB}_M(f,\ldots,f)$, we do not expect there to be any dual witness obeying complementary slackness with respect to $p_{\text{robust}}(q,\ldots,q)$. Accordingly, the dual witness ψ_{comb} that we construct to prove Theorem 1 departs from Eq. (3).

Remark 7. *The reason that we did not need to use a robust approximating polynomial for the outer function* OMB_M *in Observation 2 is that we used an inner approximation q that is* constant *for inputs in $f^{-1}(+1)$. Hence, we can use an outer approximation p that is robust only to highly restricted noise vectors. Namely, for any input x, p needs to be robust only to noise vectors* $\mathbf{e}$ *such that* $\mathbf{e}_i$ *is constant on all coordinates i such that $x_i = +1$.*

1.3 Applications

This section gives an overview of our applications to query and communication complexity. Formal definitions of the complexity classes involved in these applications, and statements and proofs of the relevant theorems, are deferred to Section 4.

Notation. Given a query or communication model $\mathbf{C}$ and a function f, the notation $\mathbf{C}(f)$ denotes the least cost of a protocol computing f in the model $\mathbf{C}$. Following Babai et al. [5], we define a corresponding complexity class, also denoted $\mathbf{C}$, consisting of all problems that have polylogarithmic cost protocols in the model $\mathbf{C}$. Throughout, we use the superscript cc to denote communication complexity classes, and the subscript query to denote query complexity classes. Any complexity class without a subscript refers to a classical (Turing Machine) class.

1.3.1 Query Complexity

The Connection Between Query Complexity, Approximate Degree, and Oracle Separations. A significant motivation for studying query complexity is that separations of query complexity classes immediately yield oracle separations of their classical counterparts. Such oracle separations are sometimes construed as evidence that the same separation applies to the classes' classical counterparts. At a minimum, oracle separations imply a formal barrier (called the *relativization* barrier [6]) to disproving the corresponding Turing Machine separation.

It is well-known that approximate degree lower bounds imply lower bounds on (even quantum) query complexity. So to summarize, approximate degree lower bounds imply query complexity lower bounds, which in turn often imply oracle separations for classical complexity classes.

ODD-MAX-BIT, Counting, and the Polynomial Hierarchy. An important question in complexity theory is to determine the relative power of alternation (as captured by the polynomial-hierarchy **PH**), and counting (as captured by the complexity class **#P** and its decisional variant **PP**). Both **PH** and **PP** generalize **NP** in natural ways. Toda famously showed that their power is related: $\mathbf{PH} \subseteq \mathbf{P}^{\mathbf{PP}}$ [54].

Beigel [8] was interested in determining how much of the Polynomial Hierarchy is contained in **PP** itself, and he set out to give an oracle separating $\mathbf{P}^{\mathbf{NP}}$ from **PP**. To do so, he introduced the function OMB and observed that OMB is in the query complexity of analog of $\mathbf{P}^{\mathbf{NP}}$ — essentially, the query protocol uses the **NP** oracle to perform a binary search for the largest index i^* such that $x_{i^*} = -1$. Then, to show that OMB is not in the query complexity analog of **PP**, Beigel proved a lower bound on the approximate degree of OMB. (Recall from Section 1.2.2 that in [8] Beigel proved that for any $d > 0$, there is an $\varepsilon \in 1 - 2^{-\Omega(n/d^2)}$ such that $\widetilde{\deg}_\varepsilon(\mathrm{OMB}_n) \geq d$).

Thus, Beigel's result separated the query complexity classes $\mathbf{PP}_{\mathsf{query}}$ and $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$, and this in turn implied an oracle separating the classical classes **PP** from $\mathbf{P}^{\mathbf{NP}}$.

An Improved Separation for Query Complexity. Quantitatively, Beigel's analysis implies that $\mathbf{PP}_{\mathsf{query}}(\mathrm{OMB}) = \Omega(n^{1/3})$, and prior to our work, this was the best known separation between $\mathbf{PP}_{\mathsf{query}}(f)$ and $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}(f)$ for any function f. We improve on this separation by giving a function F in $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ such that $\mathbf{PP}_{\mathsf{query}}(F) = \tilde{\Omega}(n^{2/5})$.

Details of the separation. The function F we use to exhibit this improved separation is

$$F := \mathrm{OMB}_{n^{2/5}}(\overline{\mathrm{ED}}_{n^{3/5}}, \ldots, \overline{\mathrm{ED}}_{n^{3/5}}), \tag{4}$$

where $\overline{\mathrm{ED}}$ is a function computed by a polynomial size, logarithmic width DNF that we formally define in Section 4.1.3. Prior work has shown that $\overline{\mathrm{ED}}_N$ satisfies $\widetilde{\deg}_{+,2/3}(\overline{\mathrm{ED}}_N) = \tilde{\Omega}(N^{2/3})$ [13], so Theorem 1 implies that $\widetilde{\deg}_{+,\varepsilon}(F) = \tilde{\Omega}(n^{2/5})$ even for $\varepsilon = 1 - 2^{-n^{2/5}}$. This in turn implies the claimed lower bound $\mathbf{PP}_{\mathsf{query}}(F) = \tilde{\Omega}(n^{2/5})$. Meanwhile, $\overline{\mathrm{ED}}$ is in $\mathbf{NP}_{\mathrm{query}}$, and hence the same binary search-based $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ protocol that works for OMB also works for F.

1.3.2 Communication Complexity

Babai, Frankl, and Simon [5] defined the (two-party) communication analogs of many complexity classes from the Turing Machine world. Since their seminal paper, these communication classes have been studied intensely, with the following motivation.

Relationship to Turing Machine Complexity. Just as query complexity separations are sometimes construed as evidence that the same separation applies to the classes' classical counterparts, so too are communication complexity separations. In addition, Aaronson and Wigderson [2] showed that a separation of communication complexity classes implies a formal barrier (called the *algebraization* barrier) to disproving the analogous separation in the Turing Machine world. Their result is analogous to how query complexity separations imply that the relativization barrier applies in the Turing Machine world.

Thus, studying $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$ and $\mathbf{PP}^{\mathbf{cc}}$ sheds additional light on the relationship between their Turing Machine counterparts. These communication classes are also of interest in their own right, as we now explain.

The class $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$. $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$ lies near the frontier of our current understanding of communication complexity classes, in that it is one of the most powerful communication models against which we know how to prove lower bounds. This communication class has received considerable attention in recent years: Impagliazzo and Williams [22] were the first to prove lower bounds against this class, and Papakonstantinou et al. [34] characterized the class in terms of limited memory communication models. Göös et al. [21] related $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$ to various other communication classes near the frontier of understanding.

The class $\mathbf{PP}^{\mathbf{cc}}$. $\mathbf{PP}^{\mathbf{cc}}$ captures the difficulty of computing functions to small-bias, and it turns out to be characterized by an important combinatorial quantity known as *discrepancy* [25]. Motivated in part by this characterization, $\mathbf{PP}^{\mathbf{cc}}$ has received intense study (cf. [11, 21, 25, 32, 43, 44, 52] and many others).

An improved separation between $\mathbf{PP}^{\mathbf{cc}}$ and $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$. Buhrman, Vereshchagin, and de Wolf [11] gave the first separation between $\mathbf{PP}^{\mathbf{cc}}$ and $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$.[6] Specifically, they "lifted" Beigel's query complexity lower bound for OMB to the communication setting, showing that a certain communication problem G derived from OMB satisfies $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}(G) = O(\log^2 n)$, but $\mathbf{PP}^{\mathbf{cc}}(G) = \Omega(n^{1/3})$. Prior to our work, this was the best separation between these two communication classes.

We improve on this separation. By applying Sherstov's pattern matrix method [44] to the function F of Eq. (4), we obtain a communication problem F' that satisfies $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}(F') = O(\log^2 n)$, but $\mathbf{PP}^{\mathbf{cc}}(F') = \tilde{\Omega}(n^{2/5})$.

An improved separation between $\mathbf{PP}^{\mathbf{cc}}$ and $\mathbf{UPP}^{\mathbf{cc}}$ for an $\mathbf{AC}^{\mathbf{0}}$ function. Buhrman et al.'s function G also exhibited the first separation between $\mathbf{PP}^{\mathbf{cc}}$ and a related communication class called $\mathbf{UPP}^{\mathbf{cc}}$, which captures the difficulty of computing f to strictly positive bias (Sherstov [41] independently separated these two classes). In more detail, the function G used by Buhrman et al. satisfies $\mathbf{UPP}^{\mathbf{cc}}(G) = O(\log n)$, while $\mathbf{PP}^{\mathbf{cc}}(G) = \Omega(n^{1/3})$, and until our work this remained the best known separation between $\mathbf{PP}^{\mathbf{cc}}$ and $\mathbf{UPP}^{\mathbf{cc}}$ for any function in AC^0. Our communication problem F' improves on this separation, giving a function F' in AC^0 satisfying $\mathbf{UPP}^{\mathbf{cc}}(F') = O(\log n)$, but $\mathbf{PP}^{\mathbf{cc}}(F') = \tilde{\Omega}(n^{2/5})$.

To further motivate this application, we mention that $\mathbf{PP}^{\mathbf{cc}}$ is characterized not only by discrepancy, but also by the learning-theoretic notion of *margin complexity* [31, 32], while $\mathbf{UPP}^{\mathbf{cc}}$ is characterized by the notion of *dimension complexity* [36]. Both margin complexity and dimension complexity underly state-of-the-art learning algorithms for constant-depth circuits in a variety of learning models (for details, see [13, 26, 27, 38, 44] and the references therein). Separating these two quantities sheds light on the relative power of these algorithms.

1.3.3 Roadmap for the Rest of the Paper

For completeness, we collect formal definitions of approximate degree and its one-sided variants, along with their dual characterizations, in Appendix A. We introduce notation and establish preliminary lemmas in Section 2. Section 3 provides an intuitive overview of the dual witness we construct to prove Theorem 1, before providing proof details. Section 4 formalizes our applications to query and communication complexity.

2 Notation and Preliminary Facts

Given a set $T \subseteq \{-1,1\}^N$, we let $\mathbb{I}_T$ denote the indicator vector of T; that is, $\mathbb{I}_T(x) = 1$ if $x \in T$, and $\mathbb{I}_T(x) = 0$ otherwise. Given a dual polynomial $\psi : \{-1,1\}^N \to \mathbb{R}$, we define the *$L_1$-weight* of T under ψ to be $W_\psi(T) = \sum_{x \in T} |\psi(x)|$. We use the standard notation $\|\psi\|_1 := W_\psi(\{-1,1\}^N)$, and refer to $\|\psi\|_1$ as the *L_1-norm* of ψ. Define the function $\widetilde{\mathsf{sgn}} : \mathbb{R} \to \{-1,0,1\}$ via:

$$\widetilde{\mathsf{sgn}}(t) = \begin{cases} 1 \text{ if } t > 0 \\ -1 \text{ if } t < 0 \\ 0 \text{ otherwise.} \end{cases}$$

We say that a dual polynomial ψ for a function f *makes an error* on input x if $0 \neq \widetilde{\mathsf{sgn}}(\psi(x)) \neq \widetilde{\mathsf{sgn}}(f(x))$.

Crucial to our proof are the following two facts that provide methods of combining multiple dual witnesses while preserving their pure high degree.

Fact 8. *If $\psi_1, \psi_2 : \left(\{-1,1\}^N\right)^M \to \{-1,1\}$ both have pure high degree d, then so does $\psi_1 + \psi_2$.*

Fact 9. *Suppose that $\psi_1, \ldots, \psi_M : \{-1,1\}^N \to \{-1,1\}$ are each defined over disjoint sets of variables, and there is some i such that ψ_i has pure high degree d. Then so does the function $\psi : \left(\{-1,1\}^N\right)^M \to \{-1,1\}$ defined via $\psi(x_1, \ldots, x_M) = \prod_{i=1}^M \psi_i(x_i)$.*

[6] Buhrman et al. framed their result as an exponential separation between the $\mathbf{PP}^{\mathbf{cc}}$ and a related class called $\mathbf{UPP}^{\mathbf{cc}}$. As pointed out in subsequent work [21], their result also separates $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$ and $\mathbf{PP}^{\mathbf{cc}}$.

3 Proof of Theorem 1

This section proves Theorem 1, which we restate here for the reader's convenience. Recall from the introduction that for any Boolean function $f : \{-1,1\}^N \to \{-1,1\}$, F denotes the function $\mathrm{OMB}_M(f,\ldots,f)$ that maps $\{-1,1\}^{M\cdot N}$ to $\{-1,1\}$.

Theorem 1. *If $\widetilde{deg}_{+,2/3}(f) \geq d$, then $\widetilde{deg}_{+,\varepsilon}(F) \geq d$ for $\varepsilon = 1-2^{-M}$.*

Proof. Let ψ_{in} denote a dual witness for the fact that $\widetilde{\deg}_{+,2/3}(f) \geq d$, normalized to ensure that its L_1-norm is 1. Recall from Section 1.2.1 that ψ_{in} satisfies three properties: (a) ψ_{in} has pure high degree at least d, (b) ψ_{in} has correlation $\varepsilon' \geq 2/3$ with f, and (c) ψ_{in} has positive one-sided error for f, i.e., $\psi_{\text{in}}(x_i) \geq 0$ for all $x_i \in f^{-1}(+1)$. Let E denote the set of all $x_i \in \{-1,1\}^N$ on which $\psi_{\text{in}}(x_i)$ is in error, i.e., $0 \neq \widetilde{\mathsf{sgn}}(\psi_{\text{in}}(x_i)) \neq \widetilde{\mathsf{sgn}}(f(x_i))$.

Proof Overview. For any vector $x = (x_1,\ldots,x_M) \in \left(\{-1,1\}^N\right)^M$, we think of x_M as the "most significant" block in x, because if $f(x_M) = -1$, then F evaluates to -1 regardless of the values of the other blocks $x_1,\ldots,x_{M-1}$. Similarly, we think of x_1 as the "least significant block" of x.

We think of our dual witness ψ_{comb} as being constructed iteratively. The first iteration will create a dual witness $\psi^{(1)}$ that "uses" the least significant block x_1 to "achieve" pure high degree at least d. That is, $\psi^{(1)}$ will be uncorrelated with any polynomial p, unless the degree of p is at least d *even when restricted to the variables in the first block*. However, $\psi^{(1)}$ will only have correlation ε' with F, and hence it will make errors if $\varepsilon' < 1$. The second iteration creates a dual witness $\psi^{(2)}_{\text{comb}} = \psi^{(1)} + \psi^{(2)}$, where $\psi^{(2)}$ is a correction term that zeros out there errors of $\psi^{(1)}$. Moreover, $\psi^{(2)}$ will use the second block x_2 to achieve pure high degree at least d. By Fact 8, this ensures that $\psi^{(2)}_{\text{comb}}$ also has pure high degree at least d.

If $\psi^{(2)}$ zeroed out all of the errors of $\psi^{(1)}$ without introducing any new errors, then $\psi^{(2)}_{\text{comb}}$ would have perfect correlation with F, and we would be done. Unfortunately, $\psi^{(2)}$ does introduce new errors. But we have made tangible progress: we show that the number of errors $\psi^{(2)}$ makes, relative to $\psi^{(1)}$, falls by a factor of $W_{\psi_{\text{in}}}(f^{-1}(+1))/W_{\psi_{\text{in}}}(E) = \varepsilon'/(1-\varepsilon')$. Since $\varepsilon' \geq 2/3$, we conclude that $\varepsilon'/(1-\varepsilon') \geq 2$, and hence that $\psi^{(2)}$ makes at most half as many errors as $\psi^{(1)}$.

In general, the ith iteration adds in a correction term $\psi^{(i)}$ that zeros out all of the errors of the dual witness $\psi^{(i-1)}_{\text{comb}}$ constructed in the previous iteration. $\psi^{(i)}$ will use the ith input block x_i to achieve pure high degree at least d, and will introduce at most a $W_{\psi_{\text{in}}}(E)/W_{\psi_{\text{in}}}(f^{-1}(+1)) \leq 1/2$ fraction of the errors made by $\psi^{(i-1)}$. At the end of iteration M, we have constructed a dual witness $\psi_{\text{comb}} := \sum_{i=1}^{M} \psi^{(i)}$ that makes only a $\left(W_{\psi_{\text{in}}}(E)/W_{\psi_{\text{in}}}(f^{-1}(+1))\right)^M = ((1-\varepsilon')/\varepsilon')^M \leq 2^{-M}$ fraction of the errors made by $\psi^{(1)}$, and we are done.

Proof Details. Throughout, we assume without loss of generality that M is odd (we only exploit this assumption in the proof of Lemma 18, which shows that ψ_{comb} has positive one-sided error for F).

Properties of ψ_{in}. Throughout, we let $Q^-, Q^+ \subseteq \{-1,1\}^N$ denote the set of inputs x_i for which $\psi_{\text{in}}(x_i) < 0$ and $\psi_{\text{in}}(x_i) > 0$ respectively. We assume $d \geq 1$, as otherwise Theorem 1 holds trivially. We make use of the following simple facts about $\mathbb{I}_{Q^+}$ and $\mathbb{I}_{Q^-}$.

Fact 10. $\sum_{x_i\in\{-1,1\}^N} \mathbb{I}_{Q^-}(x_i)\cdot|\psi_{in}(x_i)| = \sum_{x_i\in\{-1,1\}^N} \mathbb{I}_{Q^+}(x_i)\cdot|\psi_{in}(x_i)| = 1/2$.

Proof. Since ψ_{in} witnesses the fact that $\widetilde{\deg}_{+,1/2}(f) \geq d$, ψ_{in} has pure high degree at least $d \geq 1$. In particular, ψ_{in} is uncorrelated with any constant function. Hence, $\sum_{x_i\in\{-1,1\}^N} \psi_{\text{in}}(x_i) = 0$. Since $\sum_{x_i\in\{-1,1\}^N} |\psi_{\text{in}}(x_i)| = 1$, it follows that $\sum_{x_i\in\{-1,1\}^N : x_i\in Q^+} |\psi_{\text{in}}(x_i)| = \sum_{x_i\in\{-1,1\}^N : x_i\in Q^-} |\psi_{\text{in}}(x_i)| = 1/2$, which is equivalent to the statement we wished to prove. □

A crucial implication of the fact that ψ_{in} has positive one-sided error is that if ψ_{in} outputs a negative value on input x_i, we can "trust" that $f(x_i) = -1$. This is formalized in the next fact.

Fact 11. *For all $x_i \in Q^-$, it holds that $f(x_i) = -1$. Equivalently, $E \subseteq f^{-1}(-1)$, or in other words $E \cap f^{-1}(+1) = \emptyset$.*

The following two facts relate the correlation of ψ_{in} with f to the L_1-weight of the sets E and $f^{-1}(+1)$ under ψ_{in}.

Fact 12. $W_{\psi_{in}}(E) = (1-\varepsilon')/2$.

Proof. By Property (a), $\varepsilon' = \sum_{x_i \in \{-1,1\}^N} \psi_{\text{in}}(x_i) \cdot f(x_i) = 1 - 2\sum_{x_i \in E} |\psi_{\text{in}}(x_i)|$. □

Fact 13. $W_{\psi_{in}}(f^{-1}(+1)) = \varepsilon'/2$.

Proof. This holds by the following sequence of equalities:

$$1/2 = \sum_{x_i \in Q^+} |\psi_{\text{in}}(x_i)| = \sum_{x_i \in E} |\psi_{\text{in}}(x_i)| + \sum_{x_i \in f^{-1}(+1)} |\psi_{\text{in}}(x_i)| = \left(1/2 - \varepsilon'/2\right) + \sum_{x_i \in f^{-1}(+1)} |\psi_{\text{in}}(x_i)|.$$

The first equality holds by Fact 10, the second because ψ_{in} satisfies Property (c), and the third by Fact 12. □

Construction of ψ_{comb}. The dual witness we construct is:

$$\psi_{\text{comb}}(x_1, \ldots, x_M) = \sum_{i=1}^{M} \psi^{(i)}, \text{ where} \tag{5}$$

$$\psi^{(i)} = (-1)^{i-1} \cdot (2/\varepsilon')^{M-1} \left(\prod_{j<i} \mathbb{I}_E(x_j) \cdot |\psi_{\text{in}}(x_j)| \right) \cdot \psi_{\text{in}}(x_i) \cdot \left(\prod_{j=i+1}^{M} \mathbb{I}_{f^{-1}(+1)}(x_j) \cdot |\psi_{\text{in}}(x_j)| \right). \tag{6}$$

Recall that, to show that ψ_{comb} is a dual witness for the property $\widetilde{\deg}_{+,\varepsilon}(F) \geq d$ for $\varepsilon = 1 - 2^{-M}$, it suffices to establish three properties of ψ_{comb} (cf. Section 1.2.1 or Lemma 30 in Appendix A): (a) it must have pure high degree at least d, (b) it must satisfy $\sum_{x \in (\{-1,1\}^N)^M} \psi_{\text{comb}}(x) \cdot F(x) \geq \|\psi\|_1 \cdot \varepsilon$, where $\|\psi\|_1 = \sum_{(x \in \{-1,1\}^N)^M} |\psi_{\text{comb}}(x)|$, and (c) it must have positive one-sided error. We establish each in turn below, in Propositions 14, 15, and 18.

Proposition 14. *ψ_{comb} has pure high degree at least d.*

Proof. Since ψ_{in} has pure high degree at least d, Fact 9 implies that each term $\psi^{(i)}$ in the sum within Eq. (5) also has pure high degree at least d. The lemma then follows by Fact 8. □

Proposition 15. $\sum_{x \in (\{-1,1\}^N)^M} \psi_{comb}(x) \cdot F(x) \geq \|\psi\|_1 \cdot \varepsilon$.

The proof of Proposition 15 will make use of the following two lemmas.

Lemma 16. $\|\psi\|_1 \geq 1/2$.

Proof. Consider the set $S = \{(x_1, \ldots, x_M) : x_1 \in Q^- \text{ and } x_2, \ldots, x_M \in f^{-1}(+1)\}$. We claim that the weight, $W_{\psi_{\text{comb}}}(S)$, that ψ_{comb} places on the set S is $1/2$. The lemma clearly follows.

To see this, fix $x = (x_1, \ldots, x_M) \in S$. We first note that for all $i \geq 2$, $\psi^{(i)}(x) = 0$. Indeed, $Q^- \cap E = \emptyset$ (cf. Fact 11), and hence $\mathbb{I}_E(x_1) = 0$. Thus, it is immediate from Eq. (6) that $\psi^{(i)}(x) = 0$ for $i \geq 2$.

So it suffices to show that $\sum_{x \in S} -\psi^{(1)}(x) \geq 1/2$. This follows from the following calculation:

$$\sum_{x \in S} -\psi^{(1)}(x) = (2/\varepsilon')^{M-1} \cdot \left(\sum_{x_1 \in Q^-} -\psi_{\text{in}}(x_1) \right) \cdot \left(\prod_{j=2}^{M} \left(\sum_{x_j \in \{-1,1\}^N} \mathbb{I}_{f^{-1}(+1)}(x_j) \cdot |\psi_{\text{in}}(x_j)| \right) \right)$$

$$= (2/\varepsilon')^{M-1} \cdot (1/2) \cdot \prod_{j=2}^{M} (\varepsilon'/2) = 1/2,$$

where the first equality holds by Eq. (6), and the second holds by Facts 10 and 13. □

Lemma 17. *Let $E_{comb} \subseteq \left(\{-1,1\}^N\right)^M$ denote the set of inputs on which ψ_{comb} makes an error, i.e., $0 \neq \widetilde{\mathsf{sgn}}(\psi_{comb}(x)) \neq \widetilde{\mathsf{sgn}}(F(x))$. Let $E^M \subseteq \left(\{-1,1\}^N\right)^M$ denote $\{(x_1,\ldots,x_M) : x_i \in E \text{ for all } i\}$. Then $E_{comb} = E^M$.*

Proof. We first show that $E^M \subseteq E_{\text{comb}}$ before showing that $E_{\text{comb}} \subseteq E^M$. Suppose that $x = (x_1,\ldots,x_M) \in E^M$. Fact 11 states that $E \subseteq f^{-1}(-1)$, and hence $\mathbb{I}_{f^{-1}(+1)}(x_M) = 0$. It is then immediate from Eq. (6) that $\psi^{(i)}(x) = 0$ for all $i < M$. Meanwhile, by Eq. (6) it holds that

$$\widetilde{\mathsf{sgn}}(\psi^{(M)}(x)) = (-1)^{M-1} \cdot \widetilde{\mathsf{sgn}}(\psi_{\text{in}}(x_M)) = (-1)^{M-1}.$$

Here, we used the fact that $\widetilde{\mathsf{sgn}}(\psi_{\text{in}}(x_M)) > 0$ if $x_M \in E$. (To see this, note that since $x_M \in E$, it holds that

$$0 \neq \widetilde{\mathsf{sgn}}(\psi_{\text{in}}(x_M)) \neq f(x_M) = -1,$$

where the final equality holds because $E \subseteq f^{-1}(-1)$.) At the same time, $F(x) = \text{OMB}_M(-1,-1,\ldots,-1) = (-1)^M$. Thus, $x \in E_{\text{comb}}$ as claimed.

Fix any $x = (x_1,\ldots,x_M) \in \left(\{-1,1\}^N\right)^M$ such that there exists an $i \in \{1,\ldots,M\}$ satisfying $x_i \notin E$. To show that $E_{\text{comb}} \subseteq E^M$, we must show that $x \notin E_{\text{comb}}$. To this end, let i^* be the smallest coordinate such that $x_{i^*} \notin E$. It is clear that $\psi_{\text{comb}}(x) = 0$ if $\psi_{\text{in}}(x_i) = 0$ for any $i \in [M]$, and hence $x \notin E_{\text{comb}}$. So assume throughout that $\psi_{\text{in}}(x_i) \neq 0$ for all i. The proof proceeds via a case analysis.

- Case 1: There exists a $j > i^*$ such that $x_j \notin f^{-1}(+1)$. In this case, $\mathbb{I}_{f^{-1}(+1)}(x_j) = 0$, so it is immediate from Eq. (6) that $\psi^{(k)}(x) = 0$ for all $k < j$. Meanwhile, since $\mathbb{I}_E(x_{i^*}) = 0$, it is immediate from Eq. (6) that $\psi^{(k)}(x) = 0$ for all $k \geq j$. Thus, $\psi_{\text{comb}}(x) = \sum_{k=0}^{M} \psi^{(k)}(x) = 0$, implying that $x \notin E_{\text{comb}}$.

- Case 2: $i^* = 1$, and $x_j \in f^{-1}(+1)$ for all $j > i^*$. In this case, it is clear by Eq. (6) that

$$\widetilde{\mathsf{sgn}}(\psi^{(1)}(x)) = (-1)^0 \cdot \widetilde{\mathsf{sgn}}(\psi_{\text{in}}(x_1)) = \widetilde{\mathsf{sgn}}(\psi_{\text{in}}(x_1)) = \widetilde{\mathsf{sgn}}(f(x_1)) = F(x_1,\ldots,x_M). \tag{7}$$

  Here, the third equality holds because $x_1 \notin E$, and the fourth equality exploits the fact that if $x_j \in f^{-1}(+1)$ for all $j > 1$, then $F(x) = f(x_1)$.

  Meanwhile, since $x_1 \notin E$, it holds that $\mathbb{I}_E(x_1) = 0$, and so it is clear by Eq. (6) that $\psi^{(k)}(x) = 0$ for all $k \geq 2$. Combining this with Eq. (7), we conclude that $\widetilde{\mathsf{sgn}}(\psi_{\text{comb}}(x)) = \widetilde{\mathsf{sgn}}(\psi^{(1)}(x)) = F(x_1,\ldots,x_M)$. Thus, $x \notin E_{\text{comb}}$.

- Case 3: $i^* \geq 2$, and $x_j \in f^{-1}(+1)$ for all $j > i^*$. First, we argue that $\psi^{(k)} = 0$ for all $k < i^* - 1$. Indeed, for all such k, $x_{k+1} \in E \subseteq f^{-1}(-1)$ (cf. Fact 11), and so it holds that $\mathbb{I}_{f^{-1}(+1)}(x_{k+1}) = 0$. Hence, it is immediate from Eq. (6) that $\psi^{(k)}(x) = 0$.

  Next, we argue that $\psi^{(k)} = 0$ for all $k \geq i^* + 1$. Indeed, $x_{i^*} \notin E$, so $\mathbb{I}_E(x_{i^*}) = 0$. It is then immediate from Eq. (6) that $\psi^{(k)}(x) = 0$ for all $k \geq i^* + 1$.

  Finally, we claim that either $\psi^{(i^*-1)}(x) + \psi^{(i^*)}(x) = 0$ or $\widetilde{\mathsf{sgn}}(\psi^{(i^*-1)}(x) + \psi^{(i^*)}(x)) = F(x)$. This follows from the following calculation.

  - Case 3a: Suppose $x_{i^*} \notin f^{-1}(+1)$, i.e., that $\mathbb{I}_{f^{-1}(+1)}(x_{i^*}) = 0$. Then is clear from Eq. (6) that $\psi^{(i^*-1)}(x) = 0$. Meanwhile, since $x_{i^*} \notin E$, it is clear from Eq. (6) that

$$\widetilde{\mathsf{sgn}}(\psi^{(i^*)}(x)) = (-1)^{i^*-1} \cdot \widetilde{\mathsf{sgn}}(\psi_{\text{in}}(x_{i^*})) = (-1)^{i^*-1} \cdot f(x_{i^*}) = F(x),$$

    where the final equality exploits the fact that if $x_j \in f^{-1}(+1)$ for all $j > i^*$, and $x_{i^*-1} \in E \subseteq f^{-1}(-1)$ (Fact 11), then $F(x) = (-1)^{i^*-1} \cdot f(x_{i^*})$.

– Case 3b: Suppose $x_{i^*} \in f^{-1}(+1)$. We claim that it holds that $\psi^{(i^*-1)}(x) = -\psi^{(i^*)}(x)$. To see this, note that in this case

$$\psi^{(i^*-1)}(x) = (-1)^{i^*-2} \cdot (2/\varepsilon')^{M-1} \cdot \psi_{\text{in}}(x_{i^*-1}) \cdot \prod_{j \neq i^*-1} |\psi_{\text{in}}(x_j)|, \text{ and} \tag{8}$$

$$\psi^{(i^*)}(x) = (-1)^{i^*-1} \cdot (2/\varepsilon')^{M-1} \cdot \psi_{\text{in}}(x_{i^*}) \cdot \prod_{j \neq i^*} |\psi_{\text{in}}(x_j)|. \tag{9}$$

Both of the above quantities are clearly equal in absolute value, but it remains to show that $\psi^{(i^*-1)}(x) = -\psi^{(i^*)}(x)$. Since $x_{i^*-1} \in E \subseteq f^{-1}(-1)$ (Fact 11), it holds that $\widetilde{\text{sgn}}(\psi_{\text{in}}(x_{i^*-1})) = +1$. Meanwhile, since $x_{i^*} \notin E$, $\widetilde{\text{sgn}}(\psi_{\text{in}}(x_{i^*})) = f(x_{i^*}) = +1$. Hence, $\widetilde{\text{sgn}}(\psi^{(i^*-1)}(x)) = (-1)^{i^*-2}$, while $\widetilde{\text{sgn}}(\psi^{(i^*)}(x)) = (-1)^{i^*-1}$, completing the proof.

Combining all of the above, we conclude that $\psi_{\text{comb}}(x) = \sum_{j=1}^{M} \psi_{\text{comb}}^{(j)}(x) = \psi_{\text{comb}}^{(i^*-1)}(x) + \psi_{\text{comb}}^{(i^*)}(x)$, and the latter expression is either equal to 0 or agrees in sign with $F(x)$. Thus, $x \notin E_{\text{comb}}$. This completes the proof of Lemma 17.

□

Proof of Proposition 15. Note that

$$\sum_{x \in (\{-1,1\}^N)^M} \psi_{\text{comb}}(x) \cdot F(x) = \sum_{x \in (\{-1,1\}^N)^M} |\psi_{\text{comb}}(x)| - 2 \sum_{x \in E_{\text{comb}}} |\psi_{\text{comb}}(x)| = \|\psi\|_1 - 2 \sum_{x \in E_{\text{comb}}} |\psi_{\text{comb}}(x)|, \tag{10}$$

where we recall from Lemma 17 that $E_{\text{comb}} = E^M$ is the set of points on which ψ_{comb} makes an error. Observe that for each j:

$$\sum_{x \in E^M} \psi^{(j)}(x) \quad \leq (2/\varepsilon')^{M-1} \prod_{i=1}^{M} \left(\sum_{x_i \in E} |\psi_{\text{in}}(x_i)| \right) \leq (2/\varepsilon')^{M-1} \cdot \prod_{i=1}^{M} \big((1-\varepsilon')/2\big) \leq 3^{M-1}/6^M < 2^{-M-1}. \tag{11}$$

Here, the first equality holds because, for all $x \in E^M$ and $j < M$, $\psi^{(j)}(x) = 0$; this follows by combining Eq. (6) with the fact that $E \cap f^{-1}(+1) = \emptyset$ (Fact 11) (see also the $E^M \subseteq E_{\text{comb}}$ direction in the proof of Lemma 17). The second inequality holds by Fact 12, and the third holds because $\varepsilon' \geq 2/3$. Combining Lemma 16 with Eq. (10) and Eq. (11), we conclude that $\sum_{x \in (\{-1,1\}^N)^M} \psi_{\text{comb}}(x) \cdot F(x) \geq \|\psi\|_1 - 2^{-M-1} \geq \|\psi\|_1(1-2^{-M})$, completing the proof. □

Proposition 18. *$\psi_{comb}(x) \geq 0$ for all $x \in F^{-1}(+1)$.*

Proof. Lemma 17 implies that the set E_{comb} on which ψ_{comb} makes an error is equal to E^M. Since $E \subseteq f^{-1}(-1)$ (cf. Fact 11), and we assumed that M is odd, it is obvious from the definition of F that $E^M \subseteq F^{-1}(-1)$. It follows that ψ_{comb} makes no errors on $F^{-1}(+1)$, implying the proposition. □

Theorem 1 follows by combining Propositions 14, 15, and 18 and the dual characterization of $\widetilde{\deg}_{+,\varepsilon}$ (cf. Lemma 30 in Appendix A). □

4 Details of Applications to Query and Communication Complexity

4.1 Query Complexity

4.1.1 Definitions

Deterministic Query Complexity. A deterministic decision tree is a binary tree T, in which each internal node is labeled with a variable x_i, and each leaf of T is labeled with a value in $\{-1,+1\}$. Given an input $x \in \{-1,1\}^n$, the tree is evaluated as follows. The tree queries the value of the variable x_i associated with the root. If $x_i = 1$ (respectively, $x_i = -1$) then the tree recursively evaluate the left (respectively, right) subtree. When a leaf is reached, the tree outputs the value of the leaf, and this output is denoted $T(x)$.

T is said to compute a function $f\colon \{-1,1\}^n \to \{-1,1\}$ if $T(x) = f(x)$ for all $x \in \{-1,1\}^n$. The deterministic query complexity of f, denoted $D(f)$, is the least depth of a decision tree computing f.

Randomized Query Complexity. A randomized decision tree T is a probability distribution μ over deterministic decision trees. T is evaluated by choosing a deterministic tree according to μ, and then evaluating the deterministic tree as above. The *complexity* of T is the largest depth of any deterministic tree in the support of μ. T is said to compute a function $f\colon \{-1,1\}^n \to \{-1,1\}$ with error ε if for all $x \in \{-1,1\}^n$, $\Pr[T(x) = f(x)] \geq 1-\varepsilon$. We use $R_\varepsilon(f)$ to denote the least complexity of a randomized decision tree computing f to error ε. We say T computes f if $\Pr[T(x) = f(x)] > 1/2$ for all $x \in \{-1,1\}^n$.

The class $\mathbf{PP}_{\mathsf{query}}$. Fix a function $f\colon \{-1,1\}^n \to \{-1,1\}$, and let T be a decision tree satisfying $\Pr[T(x) = f(x)] > 1/2$ for all $x \in \{-1,1\}^n$. In analogy with the communication complexity class $\mathbf{PP}^{\mathsf{cc}}$ defined by Babai, Frankl, and Simon [5] (see Section 4.2.1 below), we define the $\mathbf{PP}$ query complexity of T, denoted $\mathbf{PP}_{\mathsf{query}}(T)$ to be the complexity of T plus $\log(1/\beta)$, where $\beta := \min_x(\Pr[T(x) = f(x)] - 1/2)$ is the *bias* of T. We define $\mathbf{PP}_{\mathsf{query}}(f)$ to be the minimum of $\mathbf{PP}_{\mathsf{query}}(T)$ over all randomized decision trees T that compute f.

The class $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$. A $\mathbf{P}^{\mathbf{NP}}$ decision tree is a deterministic decision tree that is allowed, at any internal node, to query the output value of any DNF over x – if the DNF that is queried has size S and width k, then the decision tree is charged a cost of $k + \log S$.[7] Thus, the complexity of a $\mathbf{P}^{\mathbf{NP}}$ decision tree computing f is the maximum over all root-to-leaf paths of the sum of the (standard) input queries along the path and the query cost of the DNF queries along the path. We define $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}(f)$ to be the minimum of $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}(T)$ over all $\mathbf{P}^{\mathbf{NP}}$ decision trees T that compute f.

4.1.2 The Polynomial Method for Lower Bounding Decision Tree Complexity

It is well-known that approximate degree lower bounds (even quantum) decision tree complexity. Formally, we will use the following result that refers only to randomized decision tree complexity.

Lemma 19. *(cf. [10, Theorem 15]) Suppose $R_\varepsilon(f) \leq d$. Then $\widetilde{deg}_{2\varepsilon}(f) \leq d$.*

4.1.3 An Improved Separation Between $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ and $\mathbf{PP}_{\mathsf{query}}$.

The purpose of this section is to prove the following theorem.

Theorem 20. *There is an (explicitly given) function $F\colon \{-1,1\}^n \to \{-1,1\}$ such that $\mathbf{PP}_{\mathsf{query}}(F) = \tilde{\Omega}(n^{2/5})$, and $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}(F) = O(\log^2 n)$.*

[7] It would also be natural to charge the decision tree only k queries when querying the output value of a DNF of width k, since this is the number of queries to the input x required to check a certificate that the DNF evaluates to TRUE (here, a certificate consists of a clause of the DNF that is satisfied by the input x). We choose to charge the decision tree $k + \log S$ queries to account for the $\log S$ bits required to specify the certificate. Observe that $\log S \leq k \log n$, and hence the cost of a DNF query under the two definitions can differ by at most a logarithmic factor. Hence, the complexity class $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ of functions solvable by polylogarithmic cost $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ protocols is the same under both definitions.

The previous best separation was due to Beigel [8], who proved that $\mathbf{PP}_{\mathsf{query}}(\mathrm{OMB}_n) = \Omega(n^{1/3})$, while $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}(\mathrm{OMB}_n) = O(\log^2 n)$.

Proof of Theorem 20. Bun and Thaler [13, Corollary 3], building on work of Aaronson and Shi [1], exhibit a function known as $\mathrm{ED}_N : \{-1,1\}^N \to \{-1,1\}$ (short for ELEMENT DISTINCTNESS) that is computed by a polynomial size CNF formula of width $O(\log N)$, and satisfies $\widetilde{\deg}_{-,2/3}(\mathrm{ED}_N) = \Omega((N/\log N)^{2/3})$.[8] Specifically, ED_N is defined as follows: Fix an $R = \Theta(N)$ that is a power of 2, and let $N = m \cdot \log_2 R$ for some $m = \Theta(N/\log N)$. ED_N takes N bits as input, and interprets its input as m blocks $(x_1, \ldots, x_m)$ with each block consisting of $\log_2 R$ bits. Each block is interpreted as a number in the range $[R]$, and ED_N evaluates to -1 on x if and only if all m numbers are distinct.

It is easy to see that for any function f, $\widetilde{\deg}_{+,\varepsilon}(f) = \widetilde{\deg}_{-,\varepsilon}(\overline{f})$, where $\overline{f}$ denotes the negation of f. Hence, Bun and Thaler's result implies the following:

Lemma 21. *(Bun and Thaler) There is a function, $\overline{\mathrm{ED}}_N : \{-1,1\}^N \to \{-1,1\}$, computed by a DNF formula of polynomial size and width $O(\log N)$, such that $\widetilde{deg}_{+,2/3}(\overline{\mathrm{ED}}_N) = \Omega((N/\log N)^{2/3})$.*

Fix an $n > 0$. Let $F : \{-1,1\}^n \to \{-1,1\}$ be defined via: $F = \mathrm{OMB}_M(\overline{\mathrm{ED}}_N, \ldots, \overline{\mathrm{ED}}_N)$, where $M = n^{2/5}$ and $N = n^{3/5}$. Theorem 1, combined with Lemma 21, implies the following corollary.

Corollary 22.

$$\widetilde{deg}_{+,\varepsilon}(F) = \tilde{\Omega}(n^{2/5}), \textit{ for some } \varepsilon = 1 - 2^{-\tilde{\Omega}(n^{2/5})}.$$

Combining Corollary 22 with Lemma 19, we conclude that any randomized decision tree for F of complexity $n^{2/5}$ has bias at most $2^{-\tilde{\Omega}(n^{2/5})}$. This immediately implies that $\mathbf{PP}_{\mathsf{query}}(F) = \tilde{\Omega}(n^{2/5})$. Hence, the proof of Theorem 20 will be complete if we show that $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}(F) = O(\log^2 n)$.

The $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ protocol for F works as follows. The decision performs a binary search to find the largest input i^* coordinate of OMB_M that is -1. Namely, the decision tree maintains upper and lower bounds ℓ, u on i^*. It repeatedly asks an NP oracle questions of the form "is there an index i in the interval $[(\ell+u)/2, u]$ such that $\overline{\mathrm{ED}}_N(x_i) = -1$?", and updates ℓ and u based on the answer. Notice that since $\overline{\mathrm{ED}}_N$ is itself computed by a DNF of polynomial size and width $O(\log N)$, the answer to each such question is also computed by a DNF of polynomial size and width $O(\log N)$.

After $\log N$ queries, the decision tree will have ascertained the largest index i^* such that $\overline{\mathrm{ED}}_N(x_{i^*}) = -1$. At this point, the decision tree outputs -1 if i^* is odd, and outputs $+1$ if i^* is even. Since each query to the NP oracle has cost $O(\log n)$, and at most $O(\log n)$ such queries are made, the resulting protocol has cost $O(\log^2 n)$. □

4.2 Communication Complexity

4.2.1 Definitions

Let $f\colon X \times Y \to \{-1,1\}$ be a function. Consider a two-party communication problem in which Alice is given an input $x \in X$, Bob is given an input $y \in Y$, and their goal is to output $f(x,y)$ with probability at least $1/2 + \beta$ for some bias $\beta > 0$. Alice and Bob each have access to an arbitrarily long sequence of private random bits, and the cost $C(P)$ of a protocol P is the worst-case number of bits they must exchange over all inputs $(x,y) \in X \times Y$.

The classes $\mathbf{PP}^{\mathbf{cc}}$ and $\mathbf{UPP}^{\mathbf{cc}}$. Babai et al. [5] defined the $\mathbf{PP}^{\mathbf{cc}}$ to capture the complexity of computing f with small bias. The **PP** communication complexity of f, denoted by $\mathbf{PP}^{\mathbf{cc}}(f)$, is the minimum value of $C(P) + \log(1/\beta(P))$ over all protocols P that compute f with positive bias.The **UPP** communication complexity of f, denoted by $\mathbf{UPP}^{\mathbf{cc}}(f)$, is the minimum value of $C(P)$ over all protocols P that compute f with positive bias.

[8] This bound is tight up to a logarithmic factor, as $\widetilde{\deg}_{2/3}(\mathrm{ED}_N) = O(N^{2/3}\log^{1/3}(N))$ [4].

The class $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$. A $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$ protocol augments the standard two-party communication model to allow Alice and Bob not only to exchange messages, but in addition to ask "NP queries". Here, an NP query consists of a collection of combinatorial rectangles $\{S_w : w \in \{0,1\}^k\}$, where the output of the query (x,y) is determined by whether or not $(x,y) \in \cup_w S_w$. The cost of the NP query is k. The cost of a $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$ protocol is the total amount of communication, plus the cost of the NP queries made by the protocol.

4.2.2 An Improved Separation Between $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$ and $\mathbf{PP}^{\mathbf{cc}}$

The purpose of this section is to prove the following theorem.

Theorem 23. *There is an (explicitly given) function $F' : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ such that $\mathbf{PP}^{\mathbf{cc}}(F') = \tilde{\Omega}(n^{2/5})$, and $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}(F') = O(\log^2 n)$.*

The previous best separation was due to Buhrman, Vereshchagin, and de Wolf [11], who gave a communication problem G derived from OMB such that $\mathbf{PP}^{\mathbf{cc}}(G) = \Omega(n^{1/3})$, while $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}(G) = O(\log^2 n)$.

Proof of Theorem 23. Before describing the function F', we first introduce the concept of discrepancy.

Discrepancy. Consider a Boolean function $F : X \times Y \to \{-1,1\}$, and let $M^{(F)}$ be its communication matrix $M^{(F)} = [F(x,y)]_{x \in X, y \in Y}$. Recall that a combinatorial rectangle of $X \times Y$ is a set of the form $A \times B$ with $A \subseteq X$ and $B \subseteq Y$. For a distribution μ over $X \times Y$, the *discrepancy* of F with respect to μ is defined to be the maximum over all rectangles R of the *bias* of F on R. That is:

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

The discrepancy of F, $\mathrm{disc}(F)$, is defined to be $\min_\mu \mathrm{disc}_\mu(F)$. It is known that discrepancy characterizes the communication model $\mathbf{PP}^{\mathbf{cc}}$ in the sense that $\mathbf{PP}^{\mathbf{cc}}(F) = \Theta(\log(1/\mathrm{disc}(F)) + \log\log(|X| \cdot |Y|))$ [25].

Sherstov's pattern matrix method [44] shows how to generically transform any function F such that $\widetilde{\deg}_\varepsilon(F)$ is large into another function with low discrepancy, as long as ε is exponentially close to 1.

Lemma 24 ([44], adapted from Corollary 1.2 and Theorem 7.3)**.** *Let $F : \{-1,1\}^n \to \{-1,1\}$ be given, and define the communication problem $F' : \{-1,1\}^{4n} \times \{-1,1\}^{4n} \to \{-1,1\}$ by*

$$F'(x,y) = F(\vee_{j=1}^4 (x_{1,j} \wedge y_{1,j}), \ldots, \vee_{j=1}^4 (x_{n,j} \wedge y_{n,j})).$$

Suppose that $\widetilde{deg}_\varepsilon(F) \geq d$ for $\varepsilon = 1 - 2^{-d}$. Then $disc(F')^2 \leq 2n \cdot 2^{-d}$.

A Function F' with Small Discrepancy. Combining Corollary 22 with Lemma 24, we obtain a function F' satisfying $\mathrm{disc}(F') \leq 2^{-\tilde{\Omega}(n^{2/5})}$. Since $\mathbf{PP}^{\mathbf{cc}}(F') = \Theta(\log(1/\mathrm{disc}(F')) + \log\log(|X| \cdot |Y|))$, it follows that $\mathbf{PP}^{\mathbf{cc}}(F') = \tilde{\Omega}(n^{2/5})$.

Thus, to complete the proof of Theorem 23, it suffices to show that $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}(F') = O(\log^2 n)$. The $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$ protocol for F' simply simulates the $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ protocol for F, applied to the n-bit input

$$(\vee_{v=1}^4 (x_{1,v} \wedge y_{1,v}), \ldots, \vee_{v=1}^4 (x_{n,v} \wedge y_{n,v})) \in \{-1,1\}^n.$$

Observe that every time the $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ protocol queries an input bit i to F, Alice and Bob can simulate the query with a constant amount of communication, by computing $\vee_{j=1}^4 (x_{i,j} \wedge y_{i,j})$. Furthermore, recall that the $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ protocol makes $O(\log n)$ DNF queries, each to a polynomial size DNF of width $O(\log n)$. The communication protocol can simulate each DNF query with an NP communication query of cost $O(\log n)$, resulting in a $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}$ protocol of total cost $O(\log^2 n)$ that correctly computes the function F'. □

4.2.3 An Improved Separation Between $\mathbf{UPP}^{\mathbf{cc}}$ and $\mathbf{PP}^{\mathbf{cc}}$ for an AC^0 Function

The purpose of this section is to prove the following theorem.

Theorem 25. *There is an (explicitly given) function $F' \colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ computed by a Boolean circuit of polynomial size and constant depth such that $\mathbf{PP}^{\mathbf{cc}}(F') = \tilde{\Omega}(n^{2/5})$, and $\mathbf{UPP}^{\mathbf{cc}}(F') = O(\log n)$.*

The previous best separation between $\mathbf{PP}^{\mathbf{cc}}(G)$ and $\mathbf{UPP}^{\mathbf{cc}}(G)$ for any function G in AC^0 was also due to Buhrman, Vereshchagin, and de Wolf [11], who gave an AC^0 function G satisfying $\mathbf{PP}^{\mathbf{cc}}(G) = \Omega(n^{1/3})$, while $\mathbf{UPP}^{\mathbf{cc}}(G) = O(\log n)$.

Proof of Theorem 25. We use the same function F' as in Theorem 25. Theorem 23 already proved that $\mathbf{PP}^{\mathbf{cc}}(F') = \tilde{\Omega}(n^{2/5})$, so we only need to prove that $\mathbf{UPP}^{\mathbf{cc}}(F') = O(\log n)$. A polylogarithmic upper bound on $\mathbf{UPP}^{\mathbf{cc}}(F')$ follows directly from the fact that $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}}(F') = O(\log^2 n)$ (cf. Theorem 23), and the fact that $\mathbf{P}^{\mathbf{NP}^{\mathbf{cc}}} \subseteq \mathbf{UPP}^{\mathbf{cc}}$ (see, e.g., [21]). The core of our argument to obtain a tight $O(\log n)$ bound on $\mathbf{UPP}^{\mathbf{cc}}(F')$ is contained in the following lemma.

Lemma 26. *Let $F = \mathrm{OMB}_M(\overline{\mathrm{ED}}_N, \ldots, \overline{\mathrm{ED}}_N)$. Then $\deg_{\pm}(F) = d$ for some $d = O(\log n)$.*

Proof. Given two inputs $z_j, z_k \in \{-1,1\}^{\log_2 R}$, let $\mathrm{EQ}(z_j, z_k)$ denote the function that evaluates to 1 if $z_j = z_k$, and evaluates to 0 otherwise. Trivially, $\mathrm{EQ}(z_j, z_k)$ is exactly computed by a polynomial of degree at most $2\log_2 R$.

Let $z = (z_1, \ldots, z_m) \in \left(\{-1,1\}^{\log_2 R}\right)^m = \{-1,1\}^N$ denote an input to $\overline{\mathrm{ED}}$. Define

$$q(z) := \sum_{j,k \in [m], j \neq k} \mathrm{EQ}(z_j, z_k).$$

Let $K = \binom{m}{2}$. Notice that q satisfies the following two properties.

- Property 1: If $\overline{\mathrm{ED}}_N(z) = 1$, then $q(z) = 0$, because $z_j \neq z_k$ for all $j \neq k$.
- Property 2: If $\overline{\mathrm{ED}}_N(z) = -1$, then $q(z) \in \{1, \ldots, K\}$, because there is at least one pair $j \neq k$ such that $z_j = z_k$.

Hence, the discussion preceding Observation 2 from Section 1.1 implies that

$$F(x) = \widetilde{\mathsf{sgn}}(h(x)), \text{ where } h(x_1, \ldots, x_M) = 1 + \sum_{i=1}^{M} (-(K+1))^i \cdot q(x_i). \tag{12}$$

□

Recall from Lemma 24 that in the communication problem corresponding to F', Alice has input $x \in \{-1,1\}^{4n}$, Bob has input $y \in \{-1,1\}^{4n}$, and the goal is to output $F(\ldots, \vee_{v=1}^{4}(x_{u,v} \wedge y_{u,v}), \ldots)$, where u ranges over $\{1, \ldots, n\}$. We first use Lemma 26 in a standard way to give a simple $\mathbf{UPP}^{\mathbf{cc}}$ protocol P of cost $O(\log^2 n)$ that computes the function F', before giving a refined protocol with cost $O(\log n)$.

Let $h(x) = \sum_{S \subseteq \{-1,1\}^n, |S| \leq d} c_S \chi_S(x)$ be a polynomial of degree $d = O(\log n)$ that sign-represents F as per Lemma 26. Alice picks an S at random, with probability proportional to $|c_S|$. Alice then sends Bob the set S, and for each of the (at most) d indices $u \in S$, Alice sends Bob the values $\{x_{u,v} : u \in S, 1 \leq v \leq 4\}$. Notice that the total communication required is $\log \binom{n}{d} + 4 \cdot d = O(\log^2 n)$ bits. Bob uses this information to compute $\chi_S(\ldots, \vee_{v=1}^{4}(x_{u,v} \wedge y_{u,v}), \ldots)$, and outputs $\widetilde{\mathsf{sgn}}\left(c_S \cdot \chi_S(\ldots, \vee_{v=1}^{4}(x_{u,v} \wedge y_{u,v}), \ldots)\right)$.

It is easy to see that Bob outputs 1 with probability $1/2 + \frac{h(x)}{2\sum_{S \subseteq \{-1,1\}^n, |S| \leq d} |c_S|}$. Since $h(x)$ sign-represents F, this implies that P computes F' with positive bias, and hence P is a $\mathbf{UPP}^{\mathbf{cc}}$ protocol for f achieving communication cost $O(\log^2 n)$.

An $O(\log n)$ bound on $\mathbf{UPP}^{\mathsf{cc}}(F')$. Notice that there was nothing special about low-degree parities in the above $\mathbf{UPP}^{\mathsf{cc}}$ protocol of cost $O(\log^2 n)$: the only properties we exploited in the protocol were that (a) F is sign-represented as a a linear combination of $n^{O(\log n)}$ functions and (b) each such function depends on only $O(\log n)$ variables. The key to refining the $O(\log^2 n)$ upper bound on $\mathbf{UPP}^{\mathsf{cc}}(F')$ proved above is to observe that the sign-representation h for F given in Eq. (12) actually sign-represents F as a linear combination of just $\mathsf{poly}(n)$ functions (which are not low-degree parities), each of which depends on only $O(\log n)$ variables.

In more detail, let $x = (x_1, \ldots, x_M) \in \left(\{-1,1\}^N\right)^M$ denote an input to the function $F = \mathrm{OMB}_M(\overline{\mathrm{ED}}_N, \ldots, \overline{\mathrm{ED}}_N)$ as in the proof of Lemma 26. Let $x_i = (z_{i,1}, \ldots, z_{i,m}) \in \left(\{-1,1\}^{\log_2 R}\right)^m = \{-1,1\}^N$ for each i.

For each $i \in [M]$ and $j,k \in [m]$ such that $j < k$, let $\phi_{i,j,k}(x)$ denote the function $\mathrm{EQ}(z_{i,j}, z_{i,k})$. Let $\mathscr{T} = \cup_{i,j,k}\{\phi_{i,j,k}\}$. Notice that $|\mathscr{T}| = O(M \cdot m^2)$, and each function $\phi_{i,j,k} \in \mathscr{T}$ depends on only $2 \cdot \log_2 R = O(\log n)$ variables. Then Eq. (12) gave a function $h(x)$ satisfying the following two properties: (a) $h(x)$ sign-represents F and (b) $h(x)$ is a linear combination of functions in $\mathscr{T}$, i.e., $h(x) = \sum_{\phi \in \mathscr{T}} c_\phi \cdot \phi(x)$ for some reals $c_\phi \in \mathbb{R}$.

The remainder of the proof is essentially identical to the analysis of our earlier protocol. Consider the following communication protocol for F'. Alice picks a $\phi \in \mathscr{T}$ at random, with probability proportional to $|c_\phi|$. Alice then sends Bob ϕ; this requires $O(\log|\mathscr{T}|) = O(\log n)$ bits. In addition, for each bit $u \in [n]$ that ϕ depends on, Alice sends Bob the values $\{x_{u,v} : 1 \le v \le 4\}$. Since each ϕ depends on only $2 \cdot \log_2 R = O(\log n)$ variables u, this requires $O(\log n)$ bits of communication as well. Bob uses this information to compute $\phi(\ldots, \vee_{j=1}^4 (x_{u,v} \wedge y_{u,v}), \ldots)$. Bob outputs $\widetilde{\mathsf{sgn}}\left(c_S \cdot \chi_S(\ldots, \vee_{j=1}^4 (x_{i,j} \wedge y_{i,j}), \ldots)\right)$.

Just as in the analysis of the protocol of cost $O(\log^2 n)$, it is easy to see that Bob outputs 1 with probability $1/2 + \frac{h(x)}{2\sum_{\phi \in \mathscr{T}} |c_\phi|}$. Since $h(x)$ sign-represents F, this implies that P computes F' with positive bias, and hence P is a $\mathbf{UPP}^{\mathsf{cc}}$ protocol for F' achieving communication cost $O(\log n)$. □

5 Future Directions

Our analysis naturally suggests several directions for future work. Perhaps the primary question is to determine what is the "right" analog of Theorem 1 when the hypothesis that $\widetilde{\deg}_{+,2/3}(f) \ge d$ is replaced with the hypothesis that $\widetilde{\deg}_{-,2/3}(f) \ge d$. We conjecture that the following bound holds:

Conjecture 27. *Suppose that $f\colon \{-1,1\}^N \to \{-1,1\}$ satisfies $\widetilde{deg}_{-,2/3}(f) \ge d$. Then letting $F = \mathrm{OMB}_M(f, \ldots, f)$, it holds that $\deg_\pm(F) = \Omega(\min\{d \cdot M^{1/3}, M\})$.*

Recall that Bun and Thaler [13] proved that $\widetilde{\deg}_{-,2/3}(\mathrm{ED}_N) = \Omega((N/\log N)^{2/3})$ (cf. Lemma 21). Thus, we obtain the following special case of Conjecture 27, which we highlight separately.

Conjecture 28. *Let $F = \mathrm{OMB}_{n^{1/2}}(\mathrm{ED}_{n^{1/2}}, \ldots, \mathrm{ED}_{n^{1/2}})$. Then $\deg_\pm(F) = \tilde{\Omega}(n^{1/2})$.*

A proof of Conjecture 28 would yield a polynomial improvement over the current best threshold degree lower bound for a depth three Boolean circuit of polynomial size, which is $\Omega(n^{3/7})$ [40].[9] On the other hand, disproving Conjecture 28 would likely require the development of new techniques for constructing low-degree threshold representations for block-composed functions.

It would also be interesting to resolve Conjecture 3 (cf. Section 1.1), which we restate here informally for reference. Is block-composition with OMB is still an effective form of hardness amplification if the one-sided hypothesis that $\widetilde{\deg}_{+,2/3}(f) \ge d$ from Theorem 1 is replaced with the weaker hypothesis that $\widetilde{\deg}_{2/3}(f) \ge d$?

[9] Conjecture 28 appeared in an earlier version of this manuscript. At that time, the best threshold degree lower bound for any constant depth Boolean circuit of polynomial size was polynomially smaller than $n^{1/2}$ [51].

Finally, it would be interesting to determine the largest possible separation between $\mathbf{P}$ and $\mathbf{P}^{\mathbf{NP}}$ in the query and communication models. We identified functions in $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ and $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$ that have, respectively, $\mathbf{PP}_{\mathsf{query}}$ and $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$ cost $\tilde{\Omega}(n^{2/5})$. Our methods would translate improved one-sided approximate degree lower bounds into improved separations. Concretely, we conjecture that for any integer $k > 0$, there exists a function f computed by a DNF of width k such that $\widetilde{\deg}_{+,2/3}(f) = \Omega(n^{k/(k+1)})$. Indeed, for any k, the k-sum function is a natural candidate for this conjecture — see [9] for details. Our methods would translate such an f into functions in $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$ and $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$ that have $\mathbf{PP}_{\mathsf{query}}$ and $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$ cost $\Omega(n^{k/(2k+1)})$, which approaches $\Omega(n^{1/2})$ as $k \to \infty$. Is it possible that such a separation is essentially tight? That is, for *every* function f in $\mathbf{P}^{\mathbf{NP}}_{\mathsf{query}}$, is it the case that $\mathbf{PP}_{\mathsf{query}}(f) = O(n^{1/2})$?

Acknowledgements The author is grateful to Mark Bun, Mika Göös, and the anonymous reviewers for insightful comments on earlier versions of this manuscript.

References

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

[2] Scott Aaronson and Avi Wigderson. Algebrization: A new barrier in complexity theory. *TOCT*, 1(1), 2009.

[3] 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.

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

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

[6] Theodore P. Baker, John Gill, and Robert Solovay. Relativizations of the P =? NP question. *SIAM J. Comput.*, 4(4):431–442, 1975.

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

[8] Richard Beigel. Perceptrons, pp, and the polynomial hierarchy. *Computational Complexity*, 4:339–349, 1994.

[9] 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.

[10] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theor. Comput. Sci.*, 288(1):21–43, 2002.

[11] 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.

[12] 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.

[13] 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`.

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

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

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

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

[18] Ronald de Wolf. A note on quantum algorithms and the minimal degree of ε-error polynomials for symmetric functions. *Quantum Information & Computation*, 8(10):943–950, 2010.

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

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

[21] Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:49, 2015.

[22] Russell Impagliazzo and Ryan Williams. Communication complexity with synchronized clocks. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, Massachusetts, June 9-12, 2010*, pages 259–269. IEEE Computer Society, 2010.

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

[24] 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.

[25] Hartmut Klauck. Lower bounds for quantum communication complexity. *SIAM J. Comput.*, 37(1):20–46, 2007.

[26] 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.

[27] 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.

[28] Samuel Kutin. Quantum lower bound for the collision problem with small range. *Theory of Computing*, 1(1):29–36, 2005.

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

[30] Troy Lee and Adi Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[31] Nati Linial and Adi Shraibman. Learning complexity vs. communication complexity. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, CCC 2008, 23-26 June 2008, College Park, Maryland, USA*, pages 53–63. IEEE Computer Society, 2008.

[32] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Struct. Algorithms*, 34(3):368–394, 2009.

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

[34] Periklis A. Papakonstantinou, Dominik Scheder, and Hao Song. Overlays and limited memory communication. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 298–308. IEEE, 2014.

[35] 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.

[36] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.

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

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

[39] 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.

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

[41] Alexander A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

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

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

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

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

[46] 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.

[47] Alexander A. Sherstov. Approximating the and-or tree. *Theory of Computing*, 9(20):653–663, 2013.

[48] 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.

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

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

[51] 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.

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

[53] 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.

[54] S. Toda. On the computational power of PP and (+)P. In IEEE, editor, *30th annual Symposium on Foundations of Computer Science, October 30–November 1, 1989, Research Triangle Park, North Carolina*, pages 514–519, 1989. Formerly called the Annual Symposium on Switching and Automata Theory. IEEE catalog no. 89CH2808-4. Computer Society order no. 1982.

A Polynomial Approximations and their Dual Characterizations

The presentation in this section borrows heavily from our earlier work [13].

A.1 Approximate Degree

The ε-approximate degree of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\widetilde{\deg}_\varepsilon(f)$, is the minimum (total) degree of any real polynomial p such that $\|p - f\|_\infty \leq \varepsilon$, i.e., $|p(x) - f(x)| \leq \varepsilon$ for all $x \in \{-1,1\}^n$. Any polynomial p satisfying $\|p - f\|_\infty \leq \varepsilon$ is called an *ε-approximation* for f. By convention, $\widetilde{\deg}(f)$ denotes $\widetilde{\deg}_{1/3}(f)$, and this quantity is referred to with qualification as the *approximate degree* of a function. The choice of $1/3$ is arbitrary, as $\widetilde{\deg}(f)$ is related to $\widetilde{\deg}_\varepsilon(f)$ by a constant factor for any constant $\varepsilon \in (0,1)$.

Given a Boolean function f, let p be a real polynomial that minimizes $\|p - f\|_\infty$ among all polynomials of degree at most d. Since we work over $x \in \{-1,1\}^n$, we may assume without loss of generality that p is

multilinear with the representation $p(x) = \sum_{|S| \leq d} c_S \chi_S(x)$ where the coefficients c_S are real numbers. Then p is an optimum of the following linear program.

min	ε	
such that	$\left\lvert f(x) - \sum_{\lvert S\rvert \leq d} c_S \chi_S(x) \right\rvert \leq \varepsilon$	for each $x \in \{-1,1\}^n$
	$c_S \in \mathbb{R}$	for each $\lvert S\rvert \leq d$
	$\varepsilon \geq 0$	

The dual LP is as follows.

max	$\sum_{x \in \{-1,1\}^n} \phi(x) f(x)$	
such that	$\sum_{x \in \{-1,1\}^n} \lvert\phi(x)\rvert = 1$	
	$\sum_{x \in \{-1,1\}^n} \phi(x) \chi_S(x) = 0$	for each $\lvert S\rvert \leq d$
	$\phi(x) \in \mathbb{R}$	for each $x \in \{-1,1\}^n$

Strong LP-duality thus yields the following well-known dual characterization of approximate degree (cf. [44]).

Lemma 29. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then $\widetilde{deg}_\varepsilon(f) > d$ if and only if there is a polynomial $\phi : \{-1,1\}^n \to \mathbb{R}$ such that*

$$\sum_{x \in \{-1,1\}^n} f(x)\phi(x) > \varepsilon, \tag{13}$$

$$\sum_{x \in \{-1,1\}^n} |\phi(x)| = 1, \tag{14}$$

and

$$\sum_{x \in \{-1,1\}^n} \phi(x)\chi_S(x) = 0 \text{ for each } |S| \leq d. \tag{15}$$

If ϕ satisfies Eq. (15), we say ϕ has *pure high degree d*. We refer to any feasible solution ϕ to the dual LP as a *dual polynomial* for f.

A.2 Positive One-Sided Approximate Degree

Positive one-sided ε-approximate degree, denoted $\widetilde{\deg}_{+,\varepsilon}(f)$, is the least degree of a real polynomial p with that is an *positive one-sided ε-approximation* to f, meaning

1. $|p(x) + 1| \leq \varepsilon$ for all $x \in f^{-1}(-1)$.
2. $p(x) \geq 1 - \varepsilon$ for all $x \in f^{-1}(+1)$.

That is, we require p to be very accurate on inputs in $f^{-1}(-1)$, but only require "one-sided accuracy" on inputs in $f^{-1}(+1)$. The primal and dual LPs change in a simple but crucial way if we look at one-sided approximate degree rather than approximate degree. Let $p(x) = \sum_{|S| \leq d} c_S \chi_S(x)$ be a polynomial of degree d for which the positive one-sided ε-approximate degree of f is attained. Then p is an optimum of the following linear program.

min	ε	
such that	$\left\lvert f(x) - \sum_{\lvert S\rvert \leq d} c_S \chi_S(x) \right\rvert \leq \varepsilon$	for each $x \in f^{-1}(-1)$
	$\sum_{\lvert S\rvert \leq d} c_S \chi_S(x) \geq 1 - \varepsilon$	for each $x \in f^{-1}(+1)$
	$c_S \in \mathbb{R}$	for each $\lvert S\rvert \leq d$
	$\varepsilon \geq 0$	

The dual LP is as follows.

max	$\sum_{x\in\{-1,1\}^n} \phi(x)f(x)$	
such that	$\sum_{x\in\{-1,1\}^n} \lvert\phi(x)\rvert = 1$	
	$\sum_{x\in\{-1,1\}^n} \phi(x)\chi_S(x) = 0$	for each $\lvert S\rvert \leq d$
	$\phi(x) \geq 0$ for each $x \in f^{-1}(+1)$	
	$\phi(x) \in \mathbb{R}$	for each $x \in \{-1,1\}^n$

We again appeal to strong LP-duality for the following dual characterization of positive one-sided approximate degree.

Lemma 30. *Fix any constant $C > 0$. Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then $\widetilde{deg}_{+,\varepsilon}(f) > d$ if and only if there is a polynomial $\phi : \{-1,1\}^n \to \mathbb{R}$ such that*

$$\sum_{x\in\{-1,1\}^n} f(x)\phi(x) > C\cdot\varepsilon, \tag{16}$$

$$\sum_{x\in\{-1,1\}^n} |\phi(x)| = C, \tag{17}$$

$$\sum_{x\in\{-1,1\}^n} \phi(x)\chi_S(x) = 0 \text{ for each } |S| \leq d, \tag{18}$$

and

$$\phi(x) \geq 0 \text{ for each } x \in f^{-1}(+1). \tag{19}$$

Observe that a feasible solution ϕ to this dual LP is a feasible solution to the dual LP for approximate degree, with the additional constraint that $\phi(x)$ agrees in sign with $f(x)$ whenever $x \in f^{-1}(+1)$. We refer to any such feasible solution ϕ as a dual polynomial for f with *positive one-sided error*.

A.3 Negative One-Sided Approximate Degree

Negative one-sided ε-approximate degree, denoted $\widetilde{\deg}_{-,\varepsilon}(f)$, is defined analogously to positive one-sided ε-approximate degree. Specifically, it equals the least degree of a real polynomial p with that is an *negative one-sided ε-approximation* to f, meaning

1. $|p(x) - 1| \leq \varepsilon$ for all $x \in f^{-1}(+1)$.
2. $p(x) \leq -1 + \varepsilon$ for all $x \in f^{-1}(-1)$.

Negative one-sided approximate degree has a dual characterization analogous to Lemma 30. However, we do not make use of this dual characterization in this work, and therefore omit the details for brevity.

ECCC ISSN 1433-8092
http://eccc.hpi-web.de