Lower Bounds for the Approximate Degree of Block-Composed Functions[*]

Justin Thaler

Yahoo! Research, New York, NY, USA

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 $\mathbf{PP}$ in both the query and communication complexity settings. Our separations improve on work of Beigel (1994) and Buhrman, Vereshchagin, and de Wolf (CCC, 2007).

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases approximate degree, one-sided approximate degree, polynomial approximations, threshold degree, communication complexity

Digital Object Identifier 10.4230/LIPIcs.ICALP.2016.17

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 polynomial that point-wise approximates f to error ε. The threshold degree, denoted $\deg_\pm(f)$, is the least degree of a real polynomial that agrees in sign with f point-wise.

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 [2, 3, 25, 6, 41], and have proven instrumental in resolving a host of long-standing open problems in communication and circuit complexity [40, 39, 35, 42, 16, 44, 34, 41, 14, 12, 13, 27, 8]. 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, 24, 20, 36]. They also yield fast algorithms for private data release [49, 11].

[*] The full version of this paper is available at `http://eccc.hpi-web.de/report/2014/150/`.

43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016).
Editors: Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi;
Article No. 17; pp. 17:1–17:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

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 [32, 38, 15] 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 [43, 9, 10, 46, 45, 26, 47]. 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 [43, 9, 10, 46, 45, 26, 47] 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 [40, 42, 16, 44, 34, 41, 14].

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 [46, 10, 9, 43].[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.

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. 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.

[1] Strictly speaking, the terms positive and negative one-sided approximate degree were introduced by Kanade and Thaler [21], 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 [10, 46]. For our purposes, the distinction between positive and negative one-sided approximate degree is crucial.

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 $x = (x_1, \dots, x_n)$, let i^* denote the largest index such that $x_{i^*} = -1$, and let $i^* = 0$ if no such index exists. We define $\mathrm{OMB}(x_1, \dots, x_n) = -1$ if i^* is odd, and $\mathrm{OMB}(x_1, \dots, x_n) = 1$ otherwise. 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., $F = \mathrm{OMB}_M(f, \dots, f)$. 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 $F = \mathrm{OMB}_M(f, \dots, f)$. 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

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

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 $\mathrm{OMB}_M(f, \dots, f)$, 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.

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

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

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

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

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. (1) 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 $\widetilde{\deg}_{+,2/3}(f) \geq d$. Conditions 1 and 1 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 [10, 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 $f = \text{OR}$ (see Section 1.2.2), and $f = \overline{\text{ED}}$, where ED is the well-studied Element Distinctness function that we use in our applications to communication and query complexity. 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 [46, 10, 9, 43]?[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 $f = \text{OR}_N$. It is known that $\widetilde{\deg}_{+,2/3}(\text{OR}_N) = \Omega(N^{1/2})$ (see, e.g., [30, 10, 17]), yet $\text{OR}_M(\text{OR}_N, \ldots, \text{OR}_N) = \text{OR}_{N \cdot M}$ 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 $F = \text{OMB}_M(f, \ldots, f)$, it holds that $\widetilde{deg}_{\varepsilon}(\text{OMB}_M(f, \ldots, f)) \geq d$, 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.* (Due to space constraints, we defer further discussion of the duality theory to the full version of the paper.)

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

[2] One may also ask about replacing OMB_M with AND_M in the statement of Theorem 1. Analyses from prior works [10, 46] 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).

dual polynomial ψ_{comb} witnessing the claimed lower bound on the composed function $F = \text{OMB}(f, \ldots, f)$.

1.2.2 Prior Work on the Approximate Degree of OMB

Beigel [7] proved that for any $d > 0$, there is an $\varepsilon \in 1 - 2^{-\Omega(n/d^2)}$ such that $\widetilde{\deg}_{\varepsilon}(\text{OMB}_n) \geq d$, and used this result[3] to give an oracle separating the (Turing Machine) complexity class **PP** from $\mathbf{P}^{\mathbf{NP}}$. Note that $\text{OMB}_M(\text{OR}_N, \ldots, \text{OR}_N)$ is a sub-function of $\text{OMB}_{M\cdot(2N)}$. As mentioned in Section 1.1, it is known that $\widetilde{\deg}_{+,2/3}(\text{OR}_N) = \Omega(N^{1/2})$. 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 $f = \text{OR}_{d^2}$. 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 [24] 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 $f = \text{OR}_N$) is asymptotically tight for all $d > 0$.[4]

1.2.3 Earlier Constructions of Dual Polynomials

Given functions g_M, f_N, Sherstov [45] and Lee [26] 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

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

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

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 [43] and independently Bun and Thaler [9] 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 [10] 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 [10]). *Suppose $\widetilde{deg}_{-,2/3}(f) \geq d$. Then $\widetilde{deg}_{-,\varepsilon}(OR_M(f, \ldots, f)) \geq d$, for $\varepsilon = 1 - 2^{-M}$.*

[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 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 [46].

▶ **Theorem 5** (Sherstov [46]). *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 proof 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} "zeros out" ψ_{comb} on points x such that $0 \neq \widetilde{\mathsf{sgn}}(\psi_{\text{comb}}(x)) \neq \widetilde{\mathsf{sgn}}(\text{OR}_M(f,\ldots,f))$. 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 [46]). *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 [10]. Rather it relies on an elaborate inductive construction of a dual polynomial (which is nonetheless reminiscent of Eq. (3)).

In the full version of the paper, we explain why any dual witness establishing Theorem 1 must qualitatively depart from the dual witnesses constructed in prior work. 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.

1.3 Applications

This section gives an overview of our applications to query and communication complexity. Due to space constraints, formal definitions of the complexity classes involved in these applications, and statements and proofs of the relevant theorems, are deferred to the full version of the paper.

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. [4], 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.

[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 $c > 0$.

1.3.1 Query Complexity

Connecting 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 [5]) 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 $\mathbf{PH}$), and counting (as captured by the complexity class $\#\mathbf{P}$ and its decisional variant $\mathbf{PP}$). Both $\mathbf{PH}$ and $\mathbf{PP}$ generalize $\mathbf{NP}$ in natural ways. Toda famously showed that their power is related: $\mathbf{PH} \subseteq \mathbf{P}^{\mathbf{PP}}$ [50].

Beigel [7] was interested in determining how much of the Polynomial Hierarchy is contained in $\mathbf{PP}$ itself, and he set out to give an oracle separating $\mathbf{P}^{\mathbf{NP}}$ from $\mathbf{PP}$. 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 $\mathbf{NP}$ 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 $\mathbf{PP}$, Beigel proved a lower bound on the approximate degree of OMB. (Recall from Section 1.2.2 that in [7] 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 $\mathbf{PP}$ 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

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

where $\overline{\mathrm{ED}}$ is the negation of the well-studied Element Distinctness function (due to space constraints, we defer a formal definition of the Element Distinctness function to the full version of the paper). Prior work has shown that $\overline{\mathrm{ED}}_N$ satisfies $\widetilde{\deg}_{+,2/3}(\overline{\mathrm{ED}}_N) = \tilde{\Omega}(N^{2/3})$ [10], 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 [4] 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 [1] 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 $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$ and $\mathbf{PP}^{\mathrm{cc}}$ 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 $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$. $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$ 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 [19] were the first to prove lower bounds against this class, and Papakonstantinou et al. [31] characterized the class in terms of limited memory communication models. Göös et al. [18] related $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$ to various other communication classes near the frontier of understanding.

The class $\mathbf{PP}^{\mathrm{cc}}$. $\mathbf{PP}^{\mathrm{cc}}$ captures the difficulty of computing functions to small-bias, and it turns out to be characterized by an important combinatorial quantity known as *discrepancy* [22]. Motivated in part by this characterization, $\mathbf{PP}^{\mathrm{cc}}$ has received intense study (cf. [40, 39, 8, 18, 22, 29, 48] and many others).

An improved separation between $\mathbf{PP}^{\mathrm{cc}}$ and $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$. Buhrman, Vereshchagin, and de Wolf [8] gave the first separation between $\mathbf{PP}^{\mathrm{cc}}$ and $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$.[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 $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}(G) = O(\log^2 n)$, but $\mathbf{PP}^{\mathrm{cc}}(G) = \Omega(n^{1/3})$. 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 [40] to the function F of Eq. (4), we obtain a communication problem F' that satisfies $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}(F') = O(\log^2 n)$, but $\mathbf{PP}^{\mathrm{cc}}(F') = \tilde{\Omega}(n^{2/5})$.

An improved separation between $\mathbf{PP}^{\mathrm{cc}}$ and $\mathbf{UPP}^{\mathrm{cc}}$ for an $\mathbf{AC}^0$ function. Buhrman et al.'s function G also exhibited the first separation between $\mathbf{PP}^{\mathrm{cc}}$ and a related communication class called $\mathbf{UPP}^{\mathrm{cc}}$, which captures the difficulty of computing f to strictly positive bias (Sherstov [37] independently separated these two classes). In more detail, the function G used by Buhrman et al. satisfies $\mathbf{UPP}^{\mathrm{cc}}(G) = O(\log n)$, while $\mathbf{PP}^{\mathrm{cc}}(G) = \Omega(n^{1/3})$, and until our work this remained the best known separation between $\mathbf{PP}^{\mathrm{cc}}$ and $\mathbf{UPP}^{\mathrm{cc}}$ for any function in AC^0. Our communication problem F' improves on this separation, giving a function F' in AC^0 satisfying $\mathbf{UPP}^{\mathrm{cc}}(F') = O(\log n)$, but $\mathbf{PP}^{\mathrm{cc}}(F') = \tilde{\Omega}(n^{2/5})$.

To further motivate this application, we mention that $\mathbf{PP}^{\mathrm{cc}}$ is characterized not only by discrepancy, but also by the learning-theoretic notion of *margin complexity* [29, 28], while $\mathbf{UPP}^{\mathrm{cc}}$ is characterized by the notion of *dimension complexity* [33]. Both margin complexity

[6] Buhrman et al. framed their result as an exponential separation between the $\mathbf{PP}^{\mathrm{cc}}$ and a related class called $\mathbf{UPP}^{\mathrm{cc}}$. As pointed out in subsequent work [18], their result also separates $\mathbf{P}^{\mathbf{NP}^{\mathrm{cc}}}$ and $\mathbf{PP}^{\mathrm{cc}}$.

and dimension complexity underly state-of-the-art learning algorithms for constant-depth circuits in a variety of learning models (for details, see [24, 35, 10, 40, 23] 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

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. In the full version of the paper, we collect formal definitions of approximate degree and its one-sided variants, along with their dual characterizations, and formalize 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: $\widetilde{\mathsf{sgn}}(t) = 1$ if $t > 0$, $\widetilde{\mathsf{sgn}}(t) = -1$ if $t < 0$, and $\widetilde{\mathsf{sgn}}(t) = 0$ if $t = 0$. 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 7.** *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 8.** *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 $\psi(x_1, \ldots, x_M) = \prod_{i=1}^M \psi_i(x_i)$.*

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** (restated). *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., $\psi_{\mathrm{in}}(x_i) \geq 0$ for all $x_i \in f^{-1}(+1)$. Let E denote the set of all $x_i \in \{-1,1\}^N$ on which $\psi_{\mathrm{in}}(x_i)$ is in error, i.e., $0 \neq \widetilde{\mathsf{sgn}}(\psi_{\mathrm{in}}(x_i)) \neq \widetilde{\mathsf{sgn}}(f(x_i))$.

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 creates 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 7, 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 17, 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 9.** $\sum_{x_i \in \{-1,1\}^N} \mathbb{I}_{Q^-}(x_i) \cdot |\psi_{in}(x_i)| = \sum_{x_i \in \{-1,1\}^N} \mathbb{I}_{Q^+}(x_i) \cdot |\psi_{in}(x_i)| = 1/2.$

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 what 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 as follows.

▶ **Fact 10.** *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 11.** $W_{\psi_{in}}(E) = (1-\varepsilon')/2.$

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

Proof. This holds by the following sequence of equalities:

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

The first equality holds by Fact 9, the second because ψ_{in} satisfies Property (c), and the third by Fact 11. ◀

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

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

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

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): (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 13, 14, and 17.

▶ **Proposition 13.** *ψ_{comb} has pure high degree at least d.*

Proof. Since ψ_{in} has pure high degree at least d, Fact 8 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 7. ◀

▶ **Proposition 14.** $\sum_{x \in (\{-1,1\}^N)^M} \psi_{comb}(x) \cdot F(x) \geq \|\psi\|_1 \cdot \varepsilon$.

The proof of Proposition 14 will make use of the following two lemmas.

▶ **Lemma 15.** $\|\psi\|_1 \geq 1/2$.

Proof. Consider the set $S = \{(x_1, \dots, x_M) : x_1 \in Q^- \text{ and } x_2, \dots, x_M \in f^{-1}(+1)\}$. 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 $x = (x_1, \dots, x_M) \in S$. We first note that for all $i \geq 2$, $\psi^{(i)}(x) = 0$. Indeed, $Q^- \cap E = \emptyset$ (cf. Fact 10), 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 9 and 12. ◀

▶ **Lemma 16.** *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 $\{(x_1,\dots,x_M) : x_i \in E \text{ for all } i\}$. 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 $x = (x_1,\dots,x_M) \in E^M$. Fact 10 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,

$$F(x) = \text{OMB}_M(-1,-1,\dots,-1) = (-1)^M.$$

Thus, $x \in E_{\text{comb}}$ as claimed.

Fix any $x = (x_1,\dots,x_M) \in \left(\{-1,1\}^N\right)^M$ such that there exists an $i \in \{1,\dots,M\}$ 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

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

  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 $\widetilde{\mathsf{sgn}}(\psi_{\text{comb}}(x)) = \widetilde{\mathsf{sgn}}(\psi^{(1)}(x)) = F(x_1,\dots,x_M)$. 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 10), 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 10), 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 10), it holds that $\widetilde{\mathsf{sgn}}(\psi_{\text{in}}(x_{i^*-1})) = +1$. Meanwhile, since $x_{i^*} \notin E$, $\widetilde{\mathsf{sgn}}(\psi_{\text{in}}(x_{i^*})) = f(x_{i^*}) = +1$. Hence, $\widetilde{\mathsf{sgn}}(\psi^{(i^*-1)}(x)) = (-1)^{i^*-2}$, while $\widetilde{\mathsf{sgn}}(\psi^{(i^*)}(x)) = (-1)^{i^*-1}$, completing the proof.

Combining all of the above, we conclude that $\psi_{\text{comb}}(x) = \sum_{j=1}^{M} \psi^{(j)}_{\text{comb}}(x) = \psi^{(i^*-1)}_{\text{comb}}(x) + \psi^{(i^*)}_{\text{comb}}(x)$, 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 16. ◀

Proof of Proposition 14. 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 16 that $E_{\text{comb}} = E^M$ is the set of points on which ψ_{comb} makes an error. Observe that for each j:
$$\sum_{x \in E^M} \psi^{(j)}(x) \leq (2/\varepsilon')^{M-1} \prod_{i=1}^{M} \left(\sum_{x_i \in E} |\psi_{\text{in}}(x_i)| \right) \leq (2/\varepsilon')^{M-1} \cdot \prod_{i=1}^{M} ((1-\varepsilon')/2) \leq 3^{M-1}/6^M < 2^{-M-1} \tag{11}$$
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 10) (see also the $E^M \subseteq E_{\text{comb}}$ direction in the proof of Lemma 16). The second inequality holds by Fact 11, and the third because $\varepsilon' \geq 2/3$. Combining Lemma 15 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 17.** *$\psi_{comb}(x) \geq 0$ for all $x \in F^{-1}(+1)$.*

Proof. Lemma 16 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 10), 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 from Propositions 13, 14, 17 and the dual characterization of $\widetilde{\deg}_{+,\varepsilon}$. ◀

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 S. Aaronson and A. Wigderson. Algebrization: A new barrier in complexity theory. *TOCT*, 1(1), 2009.

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

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 László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *FOCS*, pages 337–347, 1986.

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

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

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

8 Harry Buhrman, Nikolai K. Vereshchagin, and Ronald de Wolf. On computation and communication with small bias. In *CCC*, pages 24–32, 2007.

9 Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and markov-bernstein inequalities. In *ICALP (1)*, pages 303–314, 2013.

10 Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *ICALP, Part I*, pages 268–280, 2015.

11 Karthekeyan Chandrasekaran, Justin Thaler, Jonathan Ullman, and Andrew Wan. Faster private release of marginals on small databases. In *ITCS*, pages 387–402, 2014.

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

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

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

15 R. 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.

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

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

18 Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:49, 2015. To appear in ICALP, 2016.

19 Russell Impagliazzo and Ryan Williams. Communication complexity with synchronized clocks. In *CCC*, pages 259–269, 2010.

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

21 Varun Kanade and Justin Thaler. Distribution-independent reliable learning. In *COLT*, pages 3–24, 2014.

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

23 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.

24 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.

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

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

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

28 Nati Linial and Adi Shraibman. Learning complexity vs. communication complexity. In *CCC*, pages 53–63, 2008.

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

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

31 Periklis A. Papakonstantinou, Dominik Scheder, and Hao Song. Overlays and limited memory communication. In *CCC*, pages 298–308, 2014.

32 Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *STOC*, pages 468–474, 1992.

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

34 Anup Rao and Amir Yehudayoff. Simplified lower bounds on the multiparty communication complexity of disjointness. In *CCC*, pages 88–101, 2015.

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

36 Rocco A. Servedio, Li-Yang Tan, and Justin Thaler. Attribute-efficient learning and weight-degree tradeoffs for polynomial threshold functions. In *COLT*, pages 14.1–14.19, 2012.

37 A. A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

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

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

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

41 A. A. Sherstov. Strong direct product theorems for quantum communication and query complexity. In *STOC*, pages 41–50, 2011.

42 A. A. Sherstov. The multiparty communication complexity of set disjointness. In *STOC*, pages 525–548, 2012.

43 A. A. Sherstov. Approximating the AND-OR tree. *Theory of Computing*, 9(20):653–663, 2013.

44 A. A. Sherstov. Communication lower bounds using directional derivatives. In *STOC*, pages 921–930, 2013.

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

46 A. A. Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In *STOC*, pages 223–232, 2014.

47 A. A. Sherstov. The power of asymmetry in constant-depth circuits. In *FOCS*, pages 431–450, 2015.

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

49 Justin Thaler, Jonathan Ullman, and Salil P. Vadhan. Faster algorithms for privately releasing marginals. In *ICALP, Part I*, pages 810–821, 2012.

50 S. Toda. On the computational power of PP and +P. In *FOCS*, pages 514–519, 1989.