Guest Column: Approximate Degree in Classical and Quantum Computing[1]

Mark Bun[2] *and Justin Thaler*[3]

Abstract

The approximate degree of a Boolean function f captures how well f can be approximated pointwise by low-degree polynomials. This article surveys what we know about approximate degree and illustrates some of its applications in theoretical computer science.

1 Introduction

The ability (or inability) to represent or approximate Boolean functions by polynomials is a central concept in complexity theory, underlying interactive and probabilistically checkable proof systems, circuit lower bounds, quantum complexity theory, and more. In this column, we survey some of what is known about our personal favorite notion of approximation by polynomials. The ε-*approximate degree* of a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$, denoted $\widetilde{\deg}_\varepsilon(f)$, is the least total degree of a real polynomial $p : \{-1,1\}^n \to \mathbb{R}$ such that

$$|f(x) - p(x)| \leq \varepsilon \text{ for all } x \in \{-1,1\}^n. \tag{1}$$

Every Boolean function is approximated to error $\varepsilon = 1$ by the constant 0 function, implying that $\widetilde{\deg}_1(f) = 0$ for all such f. However, whenever ε is strictly less than 1, $\widetilde{\deg}_\varepsilon(f)$ is a fascinating notion with a rich theory and applications throughout theoretical computer science.

Applications of approximate degree lower bounds. The study of approximate degree is itself a "proto-complexity theory" [Aar08], with pointwise approximation by real polynomials serving as a rudimentary model of computation, and degree acting as a measure of complexity. Moreover, when f has large (say, $n^{\Omega(1)}$) approximate degree, it is also hard to compute in a variety of other computational models. Different models correspond to different settings of the error parameter ε with two regimes of particular interest. First, if $\widetilde{\deg}_{1/3}(f)$ is large, then f cannot be efficiently

[2] Department of Computer Science, Boston University, Boston, MA 02215, USA. mbun@bu.edu. Supported by NSF grant CCF-1947889.

[3] Department of Computer Science, Georgetown University, Washington, DC 20057, USA. justin.thaler@georgetown.edu. Supported by NSF CAREER award CCF-184512.

evaluated by *bounded-error* quantum query algorithms [BBC+01].[4] This connection is often referred to as the "polynomial method in quantum computing." Second, if $\widetilde{\deg}_\varepsilon(f)$ is large *for every* $\varepsilon < 1$, then f is difficult to compute by *unbounded-error* randomized (or quantum) query algorithms. These are randomized algorithms that are only required to do slightly better than random guessing, and correspond to the complexity class **PP** (short for probabilistic polynomial time) defined by Gill [Gil77]. This connection has recently been used to answer long-standing questions in relativized complexity, e.g., in studying the power of statistical zero-knowledge proofs.

Applications of approximate degree upper bounds. We've seen how lower bounds on $\widetilde{\deg}_\varepsilon(f)$ imply hardness results for computing f. There are also many applications of upper bounds on $\widetilde{\deg}_\varepsilon(f)$, typically in the design of fast algorithms in areas such as learning theory [KS04, KKMS08] and differential privacy [TUV12, CTUW14]. Approximate degree upper bounds have also been used to prove complexity *lower bounds*. Here is an illustrative example. Suppose one shows that every circuit over n-bit inputs in a class $\mathcal{C}$ can be approximated to error $\varepsilon < 1$ by a polynomial of degree $o(n)$. We know that simple functions f such as Majority and Parity require approximate degree $\Omega(n)$, and therefore cannot be computed by circuits in $\mathcal{C}$. In fact, if $\varepsilon = 1/3$, then one can even conclude that $\mathcal{C}$ is not powerful enough to compute these functions *on average*, meaning that for every circuit $C \in \mathcal{C}$, we have $\Pr_{x \sim \{-1,1\}^n}[C(x) = f(x)] \leq 1/2 + \frac{1}{n^{\omega(1)}}$ [Tal17, BKT19]. This principle underlies several state-of-the-art lower bounds for frontier problems in circuit complexity.

Goals of this survey. This survey covers recent progress on proving approximate degree lower and upper bounds and describes some applications of the new bounds to oracle separations, quantum query and communication complexity, and circuit complexity. On the lower bounds side, progress has followed from an approach called the *method of dual polynomials*, which seeks to prove approximate degree lower bounds by constructing solutions to (the dual of) a certain linear program that captures the approximate degree of any function. This survey explains how several of these advances have been unlocked by a particularly simple and elegant technique—called *dual block composition*—for constructing solutions to this dual linear program.

2 Preliminaries

We often use a subscript after a function to clarify the number of variables over which it is defined. For example, OR_n denotes the function over domain $\{-1,1\}^n$ that evaluates to -1 if at least one of its inputs equals -1, and otherwise evaluates to 1. (Throughout this survey, -1 is interpreted as logical TRUE and $+1$ is interpreted as logical FALSE.) For any input $x \in \{-1,1\}^n$, we let $|x| = \sum_{i=1}^n (1 - x_i)/2$ denote the number of coordinates of x equal to -1 and refer to $|x|$ as the *Hamming weight* of x. We let

$$\mathcal{A}(x) = \sum_{i=1}^n x_i.$$

Note that $|x|$ and $\mathcal{A}(x)$ are both degree-1 polynomials in x. We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$, $[n]^*$ to denote $\{0, 1, \ldots, n\}$, and $\mathbf{1}_n$ to denote the input in $\{-1,1\}^n$ in which all entries are 1. For

[4] The choice of constant 1/3 is made for aesthetic reasons. Replacing $\varepsilon = 1/3$ with any other constant in $(0,1)$ changes the ε-approximate degree of f by at most a constant factor.

a real number t, we let $\mathrm{sgn}(t)$ equal 1 if t is nonnegative and equal -1 if t is negative.

For two functions f_m, g_b, we denote by $f_m \circ g_b$ the block-composed function over domain $\left(\{-1,1\}^b\right)^m$, i.e., $(f_m \circ g_b)(x_1,\ldots,x_m) \coloneqq f(g(x_1),\ldots,g(x_m))$. Given probability distributions $\mu_1,\ldots,\mu_m$ over $\{-1,1\}^b$, we let $\otimes_{i=1}^m \mu_i$ denote the product distribution over $(x_1,\ldots,x_m) \in \left(\{-1,1\}^b\right)^m$ where $x_i \sim \mu_i$. Given a single probability distribution μ, the distribution $\mu^{\otimes n}$ is the product distribution over $(x_1,\ldots,x_n)$ where each $x_i \sim \mu$.

We assume that all polynomials with domain $\{-1,1\}^n$ are multilinear. This is without loss of generality because $x_i^2 = 1$ whenever $x_i \in \{-1,1\}$. We denote the degree of a univariate polynomial q by $\deg(q)$. For a multivariate polynomial p, we denote by $\deg(p)$ the total degree of p, i.e., the maximum sum of variable degrees over all monomials of p with nonzero coefficients.

Recall that in applications of approximate degree, two regimes for the error parameter ε are of particular relevance. The first is $\varepsilon = 1/3$. For brevity, we use $\widetilde{\deg}(f)$ as a shorthand for $\widetilde{\deg}_{1/3}(f)$, and refer to this quantity without qualification as the *approximate degree* of f. The latter regime of special interest considers all ε arbitrarily close to, but strictly less than, 1. This regime is equivalent to a notion called the *threshold degree* of f, denoted $\deg_{\pm}(f)$, which is the least degree of a polynomial p such that

$$p(x) \cdot f(x) > 0 \text{ for all } x \in \{-1,1\}^n. \tag{2}$$

It is not hard to see that the threshold degree of f is greater than d if and only if for *every* $\varepsilon < 1$, f cannot be approximated to error ε by any degree-d polynomial. Any function p that satisfies Condition (2) is said to *sign-represent* f. If a nonzero p satisfies Condition (2) with weak rather than strict inequality, p is said to *weakly* sign-represent f.

3 Warm-Up 1: The Approximate Degree of OR_n

To get acquainted with approximate degree, let us study the function OR_n, which is well-known to have approximate degree $\Theta(\sqrt{n})$ [NS94].

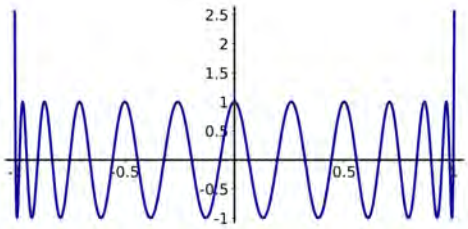
3.1 $O(\sqrt{n})$ Degree Upper Bound

The upper bound is almost an immediate consequence of the existence of *Chebyshev polynomials*. These polynomials naturally arise in the context of approximate degree because they are extremal for a classical result in approximation theory called *Markov's inequality* [Mar90].[5] This inequality states that if G is degree-d polynomial that is bounded over the interval $[-1,1]$, then its derivative $G'(t)$ cannot be too large at any point within the interval.

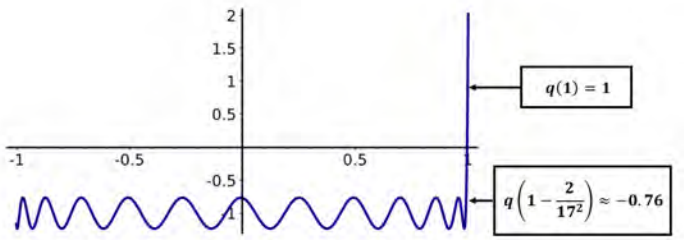
Lemma 1. (Markov's inequality) Let $G\colon [-1,1] \to [-1,1]$ be a real polynomial of degree at most d. Then $\max_{t\in[-1,1]} |G'(t)| \leq d^2$.

The Chebyshev polynomials are exactly extremal for Markov's inequality: For any integer $d > 0$, the degree-d Chebyshev polynomial T_d satisfies $|T_d(t)| \leq 1$ for all $t \in [-1,1]$, while $T_d'(1) = d^2$. In particular, for $d = \lfloor\sqrt{2n}\rfloor$, $T_d'(1) \approx 2n$. By shifting and scaling T_d without increasing its degree (i.e., by performing an affine transformation), we can obtain a univariate polynomial q that maps

[5] Not to be confused with Markov's inequality from probability theory, but rather a special case of the "Markov brothers' inequality" attributed jointly to A.A. Markov and V.A. Markov.



(a) The degree-24 Chebyshev polynomial T_{24}, which has derivative $24^2 = 576$ at input 1.



(b) The polynomial q from Equation (3) with $d = 24$ (used to approximate OR_n for $n = 17^2 = 289$).

1 to 1 and maps the entire interval $[-1, 1 - 2/n]$ to $[-1, -2/3]$. That is, as its input t decreases from 1, the polynomial q "jumps" very quickly from 1 down toward -1, and stays near -1 until t leaves the unit interval. Specifically, the following polynomial achieves the desired behavior:

$$q(t) = 2 \cdot \frac{T_d(t + 4/d^2)}{T_d(1 + 4/d^2)} - 1. \tag{3}$$

See Figures (a) and (b) for illustrations of $T_d(t)$ and $q(t)$ when $n = 17^2 = 289$ and $d = \lfloor \sqrt{2n} \rfloor = 24$.

The behavior displayed by q is exactly what we need to approximate the OR function. This is because OR_n evaluates to 1 on the unique input x with $\mathcal{A}(x)/n = 1$ and evaluates to -1 on all other inputs. Indeed, recalling that $\mathcal{A}(x) = \sum_{i=1}^{n} x_i$, the n-variate polynomial

$$p(x) \coloneqq q\left(\mathcal{A}(x)/n\right) \tag{4}$$

has total degree at most $\deg(q)$, and approximates the OR function pointwise to error at most $1/3$.

3.2 $\Omega(\sqrt{n})$ Lower Bound via Symmetrization

Prior to the method of dual polynomials, approximate degree lower bounds were typically proved via a technique called symmetrization. The ethos of this technique is that univariate polynomials are generally easier to understand than multivariate polynomials. Hence, symmetrization seeks to reduce the task of lower bounding the ε-approximate degree of a multivariate function f to a question about univariate polynomials. This is usually done by generically transforming an n-variate polynomial p into a univariate polynomial q without increasing its degree. One then argues that if p satisfies Condition (1), then q exhibits some behavior that forces it to have large degree. Since $\deg(q)$ lower bounds $\deg(p)$, one concludes that p must have large degree as well.

The transformation giving q is often built from a sequence of probability distributions D_t over $\{-1, 1\}^n$, where t ranges over a (finite or infinite) subset S of $\mathbb{R}$. One then shows that for any n-variate polynomial p, its *symmetrization* $q(t) = \mathbb{E}_{x \sim D_t}[p(x)]$ is a univariate polynomial of degree at most $\deg(p)$ over $t \in S$. Here are two classic examples.

- (t-biased symmetrization): Let S be the interval $[-1, 1]$. For $t \in S$, let μ_t be the distribution over $\{-1, 1\}$ with expected value t. Let B_t be the product distribution $\mu_t^{\otimes n}$ on $\{-1, 1\}^n$.

- (Minsky–Papert symmetrization): Let $S = [n]^*$. For each $t \in S$, define H_t to be the uniform distribution over $x \in \{-1,1\}^n$ with Hamming weight t (i.e., exactly t entries equal to -1).

The next two lemmas show that both of these classic symmetrization techniques are indeed degree non-increasing maps from n-variate to univariate polynomials.

Lemma 2. For any polynomial $p\colon \{-1,1\}^n \to \mathbb{R}$ of total degree at most d, the univariate function $q(t) = \mathbb{E}_{x\sim B_t}[p(x)]$ is a polynomial of degree at most d over $[-1,1]$.

Proof. By linearity of expectation, it is without loss of generality to consider a polynomial p consisting of a single monomial, e.g., $p(x) = x_1 x_2 \dots x_d$. Then since B_t is a product distribution,

$$\mathbb{E}_{x\sim B_t}[p(x)] = \mathbb{E}_{x_1\sim\mu_t}[x_1] \cdot \mathbb{E}_{x_2\sim\mu_t}[x_2] \cdot \ldots \cdot \mathbb{E}_{x_d\sim\mu_t}[x_d] = t^d.$$

□

Lemma 3. For any polynomial $p\colon \{-1,1\}^n \to \mathbb{R}$ of total degree at most d, the univariate function $q(t) = \mathbb{E}_{x\sim H_t}[p(x)]$ is a polynomial of degree at most d over $[n]^*$.

Proof. Again, it is without loss of generality to assume p is a single monomial, e.g., $p(x) = x_1 x_2 \dots x_d$. Applying the variable transformation $x_i \mapsto (1 - x_i)$ does not alter the degree of p, and hence it is also without loss of generality to assume that $p(x) = (1 - x_1) \cdot (1 - x_2) \cdots (1 - x_d)$. In this case, for $x \in \{-1,1\}^n$ we have $p(x) = 2^d$ if $x_1 = x_2 = \cdots = x_d = -1$ and $p(x) = 0$ otherwise.

For any $t \in [n]^*$, the number of n-bit inputs with Hamming weight exactly t is $\binom{n}{t}$, while the number of such inputs that additionally satisfy $x_1 = x_2 = \cdots = x_d = -1$ is $\binom{n-d}{t-d}$. (We are using the convention that if $t - d$ is negative, then $\binom{n-d}{t-d}$ is 0.) It follows that for any $t \in [n]^*$,

$$\mathbb{E}_{x\sim H_t}[p(x)] = 2^d \cdot \frac{\binom{n-d}{t-d}}{\binom{n}{t}} = \left(2^d \cdot \frac{(n-d)!}{n!}\right) \cdot t(t-1)(t-2)\cdots(t-d+1).$$

This is a polynomial of degree d in t. □

The tight $\Omega(\sqrt{n})$ lower bound for the approximate degree of OR_n follows easily from Lemma 2 and Markov's inequality (Lemma 1). In short, the proof applies Lemma 2 to any polynomial p approximating OR_n to derive a univariate polynomial $q(t)$ that is bounded on the whole interval $[-1,1]$ but has a large "jump" in the vicinity of $t = 1$ (quantitatively, a derivative of $\Omega(n)$). Markov's inequality then implies that q has degree $\Omega(\sqrt{n})$.

Theorem 4. The approximate degree of OR_n is $\Omega(\sqrt{n})$.

Proof. Let p approximate OR_n to error at most $1/3$, and let q be the univariate polynomial whose existence is guaranteed by Lemma 2. Since the distribution B_1 assigns probability 1 to the input $\mathbf{1}_n$, we may conclude that $q(1) = p(\mathbf{1}_n) \in [2/3, 4/3]$.

Now let $t = 1 - 4/n$. Then B_t assigns probability mass at most $(1 - 2/n)^n < 1/e^2$ to $\mathbf{1}_n$. Hence,

$$q(1 - 4/n) = \mathbb{E}_{x\sim B_t}[p(x)] \leq (4/3) \cdot 1/e^2 + (-2/3) \cdot (1 - 1/e^2) \leq -1/3.$$

The Mean Value Theorem now implies that there is some $t^* \in (1 - 4/n, 1)$ such that $q'(t^*) \geq n/4$.

Finally, since it approximates OR_n, the polynomial p has magnitude at most $4/3$ over the entire Boolean hypercube $\{-1,1\}^n$. Hence $q(t) \in [-4/3, 4/3]$ for all $t \in [-1,1]$ as well. Applying Markov's inequality to $\frac{3}{4}q$ now implies that $\deg(p) \geq \deg(q) \geq \sqrt{3n/16}$. □

Application: Quantum query and communication lower bounds. In (deterministic) query complexity, an algorithm is given oracle access to the bits of an unknown input $x \in \{-1,1\}^n$. Its goal is to evaluate a known function f on x by making as few queries to the oracle as possible. Quantum query complexity is a generalization of this model wherein the algorithm is allowed to make queries in superposition, and must output $f(x)$ with probability at least $2/3$. We refer the reader to [Amb18] for details of the model and a recent survey of results. While quantum query complexity is an information-theoretic model (i.e., the query algorithm is allowed to spend as long as it wants to decide which bits of x to query and to process the oracle's responses to the queries), it turns out to capture much of the power of quantum computing: Most query-efficient quantum algorithms can be realized as time-efficient algorithms and vice versa.

Beals et al. [BBC$^+$01] proved a result that is central to our understanding of quantum query complexity. They showed that any quantum query algorithm for f making at most T queries on every input can be transformed into an approximating polynomial for f of degree at most $2T$. Hence, if $\widetilde{\deg}_{1/3}(f) \geq d$, then the quantum query complexity of f is at least $\Omega(d)$.

Theorem 4 thus implies that the quantum query complexity of OR_n is $\Omega(\sqrt{n})$, matching an upper bound achievable via Grover's search algorithm. Equivalently, to quote the most recent tagline of Scott Aaronson's blog, "quantum computers need $\sim \sqrt{n}$ queries to search a list of size n." While this did not give the first tight quantum query lower bound proof for OR_n—it was first proved by Bennett et al. [BBBV97] using different techniques—the proof via approximate degree has other consequences in quantum complexity. Approximate degree lower bounds extend in a black-box manner from quantum query to quantum communication lower bounds [She11, SZ09]. In particular, the lower bound for OR_n transfers to a tight $\Omega(\sqrt{n})$ lower bound on the quantum communication complexity of the Disjointness function, an important result (first proved by Razborov [Raz03]) that is not known to follow from other techniques for lower bounding quantum query complexity.

4 Warm-Up 2: The Threshold Degree of the Minsky–Papert CNF

The function OR_n considered in the previous section is an interesting case study for (1/3)-approximate degree. However, the ε-approximate degree of OR_n is uninteresting for values of ε that are substantially closer to 1. In particular, for $\varepsilon \geq 1 - 1/n$, $\widetilde{\deg}_\varepsilon(\mathsf{OR}_n) = 1$. To see this, let $p(x) = (\mathcal{A}(x) + 1)/n - 1$. It is easy to check that $|p(x) - \mathsf{OR}_n(x)| \leq 1 - 1/n$ for all $x \in \{-1,1\}^n$.

In this section, we describe a classic function called the Minsky–Papert CNF that is much harder to approximate than OR_n when the error parameter ε is close to 1. Specifically, consider the block-composed function $\mathsf{AND}_m \circ \mathsf{OR}_b$, which is known to have $\deg_\pm(\mathsf{AND}_m \circ \mathsf{OR}_b) = \tilde{\Theta}(\min\{m, b^{1/2}\})$. The Minsky–Papert CNF is the function on n variables obtained by setting $m = n^{1/3}$ and $b = n^{2/3}$, for which the threshold degree bound becomes $\tilde{\Theta}(n^{1/3})$.

The upper bound is attained via two different polynomials that sign-represent $\mathsf{AND}_m \circ \mathsf{OR}_b$, one with degree $\tilde{O}(b^{1/2})$ and one with degree m. Throughout this section, let $x = (x_1, \ldots, x_m) \in \left(\{-1,1\}^b\right)^m$ denote an arbitrary input to $\mathsf{AND}_m \circ \mathsf{OR}_b$.

First upper bound. The first upper bound construction uses Chebyshev polynomials to approximate each OR_b to very small error, and then combines these approximations using a linear sign-representation of AND_m. Specifically, a slight generalization of the construction in Section 3.2 yields a polynomial p of degree $O(b^{1/2} \log m)$ that approximates OR_b to error $1/(3m)$.[6] Then the

[6] In fact, OR_b can be approximated to error $1/m$ with degree $\Theta(\sqrt{b \log m})$ [KLS96, BCDWZ99].

following degree-$O(b^{1/2}\log m)$ polynomial sign-represents $\mathsf{MP}(x)$:

$$p^*(x_1,\ldots,x_m) \coloneqq -1 + \sum_{i=1}^{m}(1+p(x_i)). \tag{5}$$

Indeed, if $\mathsf{OR}(x_i) = -1$ for all i, then $|1+p(x_i)| \leq 1/(3m)$ for all i, and hence $p^*(x) \leq -2/3 < 0$. Meanwhile, if $\mathsf{OR}(x_i) = 1$ for even a single i, then $(1+p(x_i)) \geq 2 - 1/(3m)$ and hence $p^*(x) > 0$.

Second upper bound. The second upper bound closely approximates each OR_b using a *ratio* of low-degree polynomials, combines these as before using a linear sign-representation of AND_m, and then "clears the denominator" to obtain a polynomial. This is a key idea underlying Beigel, Reingold, and Spielman's famous result that **PP** is closed under intersection [BRS95]. Specifically, there are degree-1 polynomials $p(x), q(x)$ over domain $\{-1,1\}^b$ such that the ratio p/q approximates OR_b to error $1/(3m)$ as follows. For $M \geq 6m$, let $p(x) = 1 - M\cdot|x|$ and $q(x) = 1 + M\cdot|x|$. Then if $x = \mathbf{1}_b$, we have $p(x)/q(x) = \frac{1}{1} = 1$, while if $x \neq \mathbf{1}_b$, we have $\frac{p(x)}{q(x)} \in \left[-1, \frac{1-M}{1+M}\right] \subseteq \left[-1, -1+\frac{1}{3m}\right]$.

Since $p(x)/q(x)$ approximates OR_b to error $1/(3m)$, by analogy with Equation (5), the following quantity sign-represents $\mathsf{AND}_m \circ \mathsf{OR}_b$:

$$-1 + \sum_{i=1}^{m}\left(1 + \frac{p(x_i)}{q(x_i)}\right). \tag{6}$$

Unfortunately, Expression (6) is not itself a low-degree polynomial; rather, it is a sum of ratios of linear polynomials. To get a polynomial that sign-represents $\mathsf{AND}_m \circ \mathsf{OR}_b$, we place all terms in the sum of Expression (6) over the common denominator $r(x) = \prod_{j=1}^{m} q(x_j)$. That is, for $i = 1,\ldots,m$, let $s_i(x) \coloneqq p(x_i)\cdot\prod_{j=1,\ldots,m:\, j\neq i} q(x_j)$. Then Expression (6) becomes

$$\frac{1}{r(x)}\cdot\sum_{i=1}^{m} s_i(x).$$

Finally, observe that $r(x) > 0$ for all $x \in \{-1,1\}^n$. Hence, multiplication by the denominator $r(x)$ does not alter the sign of the expression. This means that $p^*(x) \coloneqq \sum_{i=1}^{m} s_i(x)$ sign-represents $\mathsf{AND}_m \circ \mathsf{OR}_b$ and is clearly a polynomial of degree at most m.

Minsky and Papert [MP69] gave a classic symmetrization argument showing that one of these approximation techniques is always optimal for $\mathsf{AND}_m \circ \mathsf{OR}_b$.

Theorem 5. $\deg_{\pm}(\mathsf{AND}_m \circ \mathsf{OR}_b) \geq \Omega(\min\{m, b^{1/2}\})$.

Their proof used a generalization of Lemma 3 to show that if p sign-represents $\mathsf{AND}_m \circ \mathsf{OR}_{4m^2}$, then there exist a polynomial $q : ([4m^2]^*)^m \to \mathbb{R}$ such that $q(t_1,\ldots,t_m) > 0$ iff $t_i = 0$ for some index i. The polynomial q can then by symmetrized once again into a univariate polynomial $r : [2m]^* \to \mathbb{R}$ that changes sign m times as its input increases from 0 to $2m$. Such a polynomial requires degree at least m, so q and hence the original polynomial p do as well.

Applications: Learning, circuits, and communication. En route to their discovery of the fastest known algorithm for PAC learning CNF formulas, Klivans and Servedio [KS04] showed that the threshold degree of *any* polynomial size CNF is at most $\tilde{O}(n^{1/3})$. Up to logarithmic factors, the Minsky–Papert CNF thus has the largest possible threshold degree amongst all CNFs.

As with bounded-error approximate degree, generic lifting theorems are known which translate threshold degree lower bounds into communication lower bounds. For example, Sherstov [She09, She11] showed how Theorem 5 implies an inverse exponential upper bound on the *discrepancy* of a communication problem computed by a polynomial-size depth-3 circuit (hence, an exponential lower bound on its **PP** communication complexity). As a consequence, such a circuit cannot be computed by depth-2 majority circuits of subexponential size, despite the fact that quasipolynomial-size depth-3 majority circuits can compute all of AC^0 [All89]. This result was later strengthened by Razborov and Sherstov [RS10] to show that the same polynomial-size depth-3 circuit cannot be computed efficiently in the even more powerful **UPP** communication model, answering an old open question of Babai, Frankl, and Simon [BFS86] regarding the relationship between **UPP** and the communication analog of the polynomial hierarchy.

5 The Method of Dual Polynomials

Symmetrization arguments are quite powerful and have been used to determine the ε-approximate degree of many important functions. This includes all symmetric functions—those which depend only on the Hamming weight of the input [Pat92]. More sophisticated (and ad hoc) symmetrization arguments have also been applied to classes of non-symmetric functions such as halfspaces [She13b, She13c] and other functions central to quantum computing, cryptography, and circuit complexity [MP69, AS04], including the Minsky–Papert CNF described in Section 4.

Nevertheless, we should not expect symmetrization arguments to yield tight lower bounds for arbitrary functions. Approximating an n-variate function f is inherently a multivariate question. Unless f itself exhibits symmetric structure, it seems unlikely that a univariate function could fully capture the resistance of f to approximation by low-degree n-variate polynomials.

In contrast, a more recent lower bound technique called the *method of dual polynomials* is "lossless" in the sense that for any function f and any setting of the error parameter ε, the method is in principle capable of proving a tight lower bound on $\deg_\varepsilon(f)$. Here is how the method works. Fix a function $f\colon \{-1,1\}^n \to \{-1,1\}$ of interest and a degree bound d. What is the smallest error to which any polynomial of degree less than d can approximate f? The answer to this question is the value of the following linear program. It has $\binom{n}{<d} + 1$ variables, one for each coefficient of p and one for the error parameter ϵ, and $2 \cdot 2^n$ linear constraints that force p to approximate f to error at most ε at each input $x \in \{-1,1\}^n$.

$$\begin{aligned} \min_{p,\epsilon} \quad & \epsilon \\ \text{s.t.} \quad & |p(x) - f(x)| \le \epsilon && \text{for all } x \in \{-1,1\}^n \\ & \deg p < d \end{aligned}$$

Taking the dual yields the following.

$$
\begin{aligned}
\max_{\psi} \quad & \sum_{x \in \{-1,1\}^n} \psi(x) f(x) \\
\text{s.t.} \quad & \sum_{x \in \{-1,1\}^n} |\psi(x)| = 1 \\
& \sum_{x \in \{-1,1\}^n} \psi(x) p(x) = 0 \qquad\qquad \text{for all } p \text{ with } \deg p < d
\end{aligned}
$$

Weak LP duality implies that in order to prove that $\deg_\varepsilon(f) \geq d$, it suffices to identify a function $\psi\colon \{-1,1\}^n \to \mathbb{R}$ satisfying the following three conditions. Strong LP duality, moreover, implies that *every* approximate degree lower bound on f is witnessed by such a ψ.

$$\sum_{x \in \{-1,1\}^n} \psi(x) f(x) > \varepsilon, \tag{7}$$

$$\sum_{x \in \{-1,1\}^n} |\psi(x)| = 1 \tag{8}$$

$$\sum_{x \in \{-1,1\}^n} \psi(x) p(x) = 0 \text{ for all polynomials } p \text{ of degree less than } d. \tag{9}$$

Such a ψ is called a *dual polynomial* for f. We refer to Condition (7) by saying that ψ has correlation at least ε with f, to Condition (8) by saying that ψ has ℓ_1-norm 1, and to Condition (9) by saying that ψ has *pure high degree* at least d, denoting the largest such d by $\mathrm{phd}(\psi)$. This terminology comes from the fact that ψ satisfies Condition (9) if and only if its representation as a multilinear polynomial is a sum only of monomials with degree at least d. We use $\|\psi\|_1 = \sum_{x \in \{-1,1\}^n} |\psi(x)|$ to denote the ℓ_1-norm of ψ and $\langle \psi, \varphi \rangle = \sum_{x \in \{-1,1\}^n} \psi(x)\varphi(x)$ to denote the correlation of any two functions $\psi, \varphi\colon \{-1,1\}^n \to \mathbb{R}$.

One may find it helpful to think of ψ as capturing the "component" of f that is "completely missed" by polynomials of degree less than d. Indeed, the pure high degree condition means that every such polynomial p is totally uncorrelated with ψ. If ψ is well-correlated with f, then it means that ψ is a "big part" of f and hence such p must incur a lot of error when approximating f.

Decomposing dual polynomials into pieces. It can be fruitful to think of ψ as consisting of two pieces. There are in fact two natural ways to perform such a decomposition.

- We can think of $\psi = \frac{1}{2}(\psi_{+1} - \psi_{-1})$ where $\psi_{-1} = 2\max\{-\psi(x), 0\}$ and $\psi_{+1} = 2\max\{\psi(x), 0\}$ are non-negative functions. The factor of 2 is chosen to ensure that ψ_{-1} and ψ_{+1} are probability mass functions. Indeed, so long as ψ has pure high degree at least 1 (implying it is uncorrelated with the constant-1 function), then since ψ has ℓ_1-norm 1, it must be the case that $\|\psi_{-1}\|_1 = \|\psi_{+1}\|_1 = 1$. The pure high degree condition ensures that no degree-d polynomial can distinguish the distributions ψ_{-1} and ψ_{+1} with any advantage over random guessing, while the correlation condition guarantees that f *can* distinguish ψ_{-1} and ψ_{+1} with advantage ε. This perspective has been helpful in using approximate degree lower bounds to design low-complexity secret-sharing schemes [BIVW16].

- Alternatively, we can think of $\psi(x)$ as consisting of a sign, $\mathrm{sgn}(\psi(x)) \in \{-1,1\}$, and a magnitude $|\psi(x)|$. The sign $\mathrm{sgn}(\psi(x))$ can be thought of as ψ's "prediction" for $f(x)$ and the magnitude $|\psi(x)|$ as a measure of ψ's confidence in its prediction. The correlation requirement (Equation (7)) ensures that ψ's predictions, when weighted by its confidence, are accurate on average. With this in mind, we say that ψ *makes an error* at x if $\mathrm{sgn}(\psi(x)) \cdot f(x) < 0$.

  When ψ has ℓ_1-norm 1, we use $|\psi|$ to denote the probability distribution under which x is assigned probability $|\psi(x)|$. Observe that the correlation $\langle \psi, f \rangle$ equals

$$\Pr_{x\sim|\psi|}[\mathrm{sgn}(\psi(x)) = f(x)] - \Pr_{x\sim|\psi|}[\mathrm{sgn}(\psi(x)) \neq f(x)] = 1 - 2\Pr_{x\sim|\psi|}[\mathrm{sgn}(\psi(x)) \neq f(x)].$$

  If ψ weakly sign-represents f (i.e., ψ *never* makes an error), then $\langle \psi, f \rangle = 1$. In this case we say that ψ is *perfectly correlated* with f. This means that for *every* $\varepsilon < 1$, ψ demonstrates that the ε-approximate degree of f is at least $\mathrm{phd}(\psi)$; equivalently, $\deg_{\pm}(f)$ is at least $\mathrm{phd}(\psi)$.

A simple example of a dual polynomial. Consider the parity function PAR_n on n bits. Minsky and Papert famously[7] used symmetrization to prove that $\deg_{\pm}(\mathsf{PAR}_N) = n$. A dual polynomial for this fact is simply $\psi := 2^{-n} \cdot \mathsf{PAR}_n$. Clearly ψ has perfect correlation with PAR_n (since it is just a rescaling of PAR_n itself) and has ℓ_1-norm 1. Finally, as PAR_n is a monomial of degree n, it is uncorrelated with any polynomial of degree at most $n-1$.

5.1 A Dual Polynomial for OR_n

A more complicated example is to construct a dual polynomial for the fact that $\widetilde{\deg}(\mathsf{OR}_n) \geq \Omega(\sqrt{n})$. Here is a construction from [BT15a], slightly refining an earlier dual polynomial of Špalek [Špa08] and in turn building on ideas of Harry Buhrman and Mario Szegedy, and Kahn et al. [KLS96]. For any subset $S \subseteq [n]^*$, define the univariate polynomial $q_S(t) = \prod_{i\in[n]^*, i\notin S}(t-i)$. Let c be a sufficiently large constant, and let

$$S = \{0,1\} \cup \{ci^2 \colon i = 1,2,\ldots,\lfloor\sqrt{n/c}\rfloor\}. \tag{10}$$

Define $\psi\colon \{-1,1\}^n \to \mathbb{R}$ as $\psi(x) = (-1)^{|x|} \cdot q_S(|x|)$, and finally define the dual polynomial for OR_n to be $\psi_{\mathsf{OR}}(x) = \psi(x)/\|\psi\|_1$. By design, ψ_{OR} has ℓ_1-norm 1, so to show that it is a dual polynomial for OR_n, we must show it has pure high degree $\lfloor\sqrt{n/c}\rfloor$ and that it has correlation at least $1/3$ with OR_n. The former holds by the following fact.

Fact 6. If Q is any univariate polynomial of degree at most $n-1$, then $\sum_{t=0}^{n}(-1)^t\binom{n}{t}Q(t) = 0$.

Proof. We again use the fact that the parity function on n bits is uncorrelated with every polynomial of total degree at most $n-1$. The n-variate polynomial $Q(|x|)$ has degree at most $n-1$ and its correlation with the parity function is $\sum_{t=0}^{n}(-1)^t\binom{n}{t}Q(t)$. □

Lemma 7. ψ_{OR} has pure high degree at least $d = \lfloor\sqrt{n/c}\rfloor$.

Proof. Let $p\colon \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree less than d. Then Lemma 3 guarantees that $Q(t) = \mathbb{E}_{x\sim H_t}[p(t)]$ is a univariate polynomial of degree less than d over $[n]^*$. The correlation of p with ψ_{OR} is $\frac{1}{\|\psi'\|_1}\sum_{t=0}^{n}(-1)^t\binom{n}{t}q_S(t)\cdot Q(t) = 0$ using Fact 6 and the fact that $q_S \cdot Q$ is a univariate polynomial of degree at most $\deg(q_S) + \deg(Q) \leq (n-d-1) + d = n-1$. □

[7] Or perhaps infamously, as this result contributed to the first "AI winter" for neural network research.

The first conceptual step to showing that ψ_{OR} has correlation at least $1/3$ with OR_n is the following fact.

Fact 8. The correlation of ψ_{OR} with OR_n is $\langle \psi_{\mathsf{OR}}, \mathsf{OR}_n \rangle = 2 \cdot \psi_{\mathsf{OR}}(\mathbf{1}_n)$.

Proof. Since ψ_{OR} has pure high degree at least 1, it is uncorrelated with the constant-1 function. Hence, $\sum_{x \in \{-1,1\}^n} \psi_{\mathsf{OR}}(x) \cdot \mathsf{OR}_n(x) = 2 \cdot \psi_{\mathsf{OR}}(\mathbf{1}_n) + \sum_{x \in \{-1,1\}^n} \psi_{\mathsf{OR}}(x) \cdot (-1) = 2 \cdot \psi_{\mathsf{OR}}(\mathbf{1}_n)$. □

Hence, to show that $\langle \psi_{\mathsf{OR}}, \mathsf{OR} \rangle \geq 1/3$, it suffices to show that $\psi_{\mathsf{OR}}(\mathbf{1}_n) \geq 1/6$. In other words, ψ_{OR} places a constant fraction of its mass on this single input. We will not show it here, but this follows from an elementary, albeit lengthy, calculation.

5.1.1 Where did this dual come from?

A common complaint about dual polynomial constructions is that their definitions appear as if by magic, with lengthy calculations needed to show they are well-correlated with the target function f. But there is one source of intuition regarding their construction: complementary slackness. One can think of a dual polynomial ψ as assigning weights to the *constraints* of the primal linear program, with $\psi(x)$ being the weight assigned to the constraint $|p(x) - f(x)| \leq \varepsilon$. Complementary slackness asserts that if p is an optimal solution to the primal linear program, there must be an optimal solution ψ^* to the dual that only assigns nonzero weight to the constraints *made tight* by p, i.e., $\psi^*(x) \neq 0$ only for those x such that $|p(x) - f(x)| = \varepsilon$.

For the function $f = \mathsf{OR}_n$, we know roughly what an optimal solution to the primal looks like—see Equation (4), which gave an approximation $p(x) = q(\mathcal{A}(x)/n)$ for OR_n, where q is the transformed degree-d Chebyshev polynomial from Equation (3). The values of $\mathcal{A}(x)/n$ where $|q(\mathcal{A}(x)/n) - \mathsf{OR}_n(x)|$ is maximized are closely approximated by the extreme points of the degree-d Chebyshev polynomial: $\cos\left(\frac{i\pi}{d}\right) \approx 1 - \frac{1}{2} \cdot \left(\frac{i\pi}{d}\right)^2$ for $i = 1, 2, \ldots, d$. When $d = \Theta(\sqrt{n})$ we have $1 - \frac{1}{2} \cdot \left(\frac{i\pi}{d}\right)^2 \approx 1 - 2ci^2/n$ for some constant c. Inputs x for which $\mathcal{A}(x)/n = 1 - 2ci^2/n$ are precisely those inputs with Hamming weight $|x| = ci^2$. And these in turn are exactly those inputs (other than those with $|x| \in \{0, 1\}$) in S that are assigned nonzero values by ψ per Equation (10).

5.1.2 Two additional properties of ψ_{OR}

The dual polynomial ψ_{OR} we constructed satisfies additional properties beyond what is needed (Conditions (7)-(9)) to ensure that $\widetilde{\deg}(\mathsf{OR}) \geq \Omega(\sqrt{n})$. As we will see later, these properties play essential roles in constructing and analyzing dual polynomials for functions derived from OR_n via composition, e.g., $\mathsf{AND}_m \circ \mathsf{OR}_n$.

First, any dual polynomial for ψ_{OR} has an important one-sided error property [GS10]. Fact 8 implies that $\psi_{\mathsf{OR}}(\mathbf{1}_n)$ must be positive if ψ_{OR} is to have positive correlation with OR_n. Since $\mathsf{OR}_n^{-1}(+1) = \{\mathbf{1}_n\}$, this means that the only inputs on which ψ_{OR} makes an error are in $\mathsf{OR}_n^{-1}(-1)$.

Corollary 9. $\{x \colon \psi_{\mathsf{OR}}(x) \cdot \mathsf{OR}(x) < 0\} \subseteq \mathsf{OR}^{-1}(-1)$.

Second, as shown in [BT19b, BKT18], the calculation used to show that $\psi_{\mathsf{OR}}(\mathbf{1}_n) \geq 1/6$ in fact establishes the following stronger property, showing that the total mass that $|\psi_{\mathsf{OR}}|$ places on inputs of Hamming weight t decreases very rapidly with t, especially once $t \gg \sqrt{n}$.

Theorem 10. There are constants $c_1, c_2 > 0$ such that $\sum_{|x|=t} |\psi_{\mathsf{OR}}(x)| \leq c_1 \cdot \exp(-c_2 \cdot t/\sqrt{n})/t^2$ for all $t \in [n]^*$.

The extra properties satisfied by the dual polynomial ψ_{OR} captured in Theorem 10 and Corollary 9 both have natural "primal" interpretations, which readers might find more intuitive.

Primal interpretation of Corollary 9: One-sided approximate degree. Let ψ be a dual polynomial for the ε-approximate degree of f, such that ψ satisfies the additional property that

$$\{x \colon \psi(x) \cdot f(x) < 0\} \subseteq f^{-1}(-1). \tag{11}$$

Then ψ in fact witnesses that the *one-sided* approximate degree of g is at least $d = \mathrm{phd}(\psi)$. Here, one-sided approximate degree is an intermediate notion between approximate degree and threshold degree. Specifically, a real polynomial p is a *one-sided* ε-approximation for f if

$$|p(x) - (-1)| \leq \epsilon \quad \forall x \in f^{-1}(-1) \qquad \text{and} \qquad p(x) \geq 1 - \varepsilon \quad \forall x \in f^{-1}(1).$$

The one-sided approximate degree of f, denoted $\widetilde{\mathrm{odeg}}_{\epsilon}(g)$, is the minimum degree of a one-sided ϵ-approximation for f. Note that $\deg_{\pm}(f) \leq \widetilde{\mathrm{odeg}}_{\varepsilon}(f) \leq \widetilde{\deg}_{\varepsilon}(f)$ for every $\varepsilon > 0$, but there can be huge gaps in either inequality. For instance, we've seen that OR_n has one-sided approximate degree equal to its approximate degree (namely, $\Theta(\sqrt{n})$), which is vastly larger than its threshold degree, which is 1. Meanwhile $\widetilde{\mathrm{odeg}}_{1/3}(\mathsf{AND}_n) = 1$, with the one-sided approximation being $\mathcal{A}(x) + (n-1)$. This equals the threshold degree of AND_n and is vastly smaller than its approximate degree $\Theta(\sqrt{n})$.

Claim 11. For every $\varepsilon > 0$ and degree d, we have $\widetilde{\mathrm{odeg}}_{\varepsilon}(f) \geq d$ if and only if there exists a dual polynomial ψ satisfying Conditions (7)-(9) as well as Condition (11).

One can prove Claim 11 by expressing one-sided approximate degree as a linear program analogous to approximate degree, and observing that a ψ satisfying the assumptions of Claim 11 is equivalent to a solution to the dual linear program with value ε.

Primal interpretation of Theorem 10. Suppose f has a dual polynomial ψ placing very little mass on a subset $S \subseteq \{-1,1\}^n$, i.e., $|\psi(S)| \coloneqq \sum_{x \in S} |\psi(x)|$ is small. Then f cannot be approximated even by polynomials p that are allowed to be *very* large on inputs in S, so long as p approximates f well outside of S.

Claim 12. Let $0 < \delta < 1$. Suppose that ψ satisfies Conditions (7)-(9) and additionally that $|\psi(S)| \leq \varepsilon\delta/3$. Then for any polynomial p such that

$$|p(x) - f(x)| \leq \varepsilon/3 \text{ for all } x \notin S \quad \text{and} \quad |p(x)| \leq 1/\delta \text{ for all } x \in S, \tag{12}$$

we have $\deg(p) \geq d$.

Proof. Let p be a polynomial of degree less than d satisfying Condition (12). Then because ψ has pure high degree at least d, we have $\langle \psi, p \rangle = 0$. On the other hand,

$$\begin{aligned}\langle \psi, p \rangle &= \sum_{x \notin S} \psi(x)p(x) + \sum_{x \in S} \psi(x)p(x) \geq \left(\sum_{x \notin S} \psi(x)f(x) - |\psi(x)| \cdot \frac{\epsilon}{3} \right) - \sum_{x \in S} |\psi(x)| \cdot \frac{1}{\delta} \\ &\geq \langle \psi, f \rangle - \varepsilon\delta/3 - \varepsilon/3 - \varepsilon/3 > 0.\end{aligned}$$

Here, the first inequality used Condition (12), the second used that the ℓ_1-norm of ψ is 1 (Equation (8)) and that $|\psi(S)| \leq \varepsilon\delta/3$, and the final inequality used that $\langle \psi, f \rangle > \varepsilon$ (Condition (7)). □

A similar argument to Claim 12 shows that Theorem 10 implies that there is some constant $c > 0$ such that no polynomial of degree $d \leq c\sqrt{n}$ can satisfy the following condition:

$$|p(x) - \mathsf{OR}_n(x)| \leq \exp(c \cdot |x|/\sqrt{n}) \text{ for all } x \in \{-1,1\}^n.$$

6 Approximate Degree Under Function Composition

An beautiful and important result of Sherstov (refining earlier work of Buhrman et al. [BNRdW07]) shows that ε-approximate degree can increase at most multiplicatively under block composition.

Theorem 13 ([She12a])**.** $\widetilde{\deg}(f \circ g) \leq O(\widetilde{\deg}(f) \cdot \widetilde{\deg}(g))$.

In contrast to the situation for related measures such as quantum query complexity, it is still open whether the bound in Theorem 13 is tight for every pair of functions f, g.

Open Problem 14. For every pair of total Boolean functions f, g, is it the case that $\widetilde{\deg}(f \circ g) \geq \Omega(\widetilde{\deg}(f) \cdot \widetilde{\deg}(g))$?

Prior to 2012, Problem 14 was open even for the special case that $f = \mathsf{AND}$ and $g = \mathsf{OR}$. This case was eventually resolved via the method of dual polynomials [She13a, BT15a] using a simple yet powerful technique called dual block composition. Dual block composition tries to take dual polynomials witnessing the high approximate degrees of f and g individually, and combine them in a very specific manner to obtain a dual polynomial for the (even higher) approximate degree of $f \circ g$. The combining technique was proposed by several authors [She13b, Lee09, SZ09]. Here it is:

Definition 15. Given dual polynomials $\psi\colon \{-1,1\}^m \to \mathbb{R}$ and $\phi\colon \{-1,1\}^b \to \mathbb{R}$ such that ϕ has pure high degree at least 1, define the dual block composition $\psi \star \phi$ by

$$(\psi \star \phi)(x_1, \ldots, x_m) = \psi(\operatorname{sgn}(\phi(x_1)), \ldots, \operatorname{sgn}(\phi(x_m))) \cdot \prod_{i=1}^{m} (2|\phi(x_i)|).$$

Intuition for Definition 15. There are two ways to think about Definition 15, corresponding to the two ways of decomposing dual polynomials as discussed in Section 5. The first way to view $\psi \star \phi$ is as half the difference between two distributions $(\psi \star \phi)_{+1}$ and $(\psi \star \phi)_{-1}$ constructed as follows. To sample from $(\psi \star \phi)_{+1}$, first choose z from ψ_{+1} and then choose $x = (x_1, \ldots, x_m)$ from the product distribution $\otimes_{i=1}^m \phi_{z_i}$. Similarly, to sample from $(\psi \star \phi)_{-1}$, first choose z from ψ_{-1} and then choose $x = (x_1, \ldots, x_m)$ from the product distribution $\otimes_{i=1}^m \phi_{z_i}$.

The second interpretation is to get a prediction $\operatorname{sgn}((\psi \star \phi)(x))$ for $(f \circ g)(x)$ as follows. First, construct the vector $z = (\operatorname{sgn}(\phi(x_1)), \ldots, \operatorname{sgn}(\phi(x_m)))$ consisting of ϕ's predictions for each evaluation of g on $x_1, \ldots, x_m$. The final prediction $\operatorname{sgn}(\psi(z))$ for $(f \circ g)(x)$ is then simply ψ's prediction on input z. The confidence assigned to this prediction is proportional to the product of the confidences of all of the constituent predictions, namely $|\psi(z)| \cdot \prod_{i=1}^m |\phi(x_i)|$.

When is $\psi \star \phi$ a good dual witness? The hope is that if ψ is a dual witness to the fact that $\widetilde{\deg}(f) \geq d_f$ and ϕ is a dual witness to $\widetilde{\deg}(g) \geq d_g$, then $\psi \star \phi$ is a dual witness to the fact that $\widetilde{\deg}_\varepsilon(f \circ g) \geq d_f \cdot d_g$ for some constant $\varepsilon \in (0,1)$. This requires showing that $\psi \star \phi$ satisfies Conditions (7)-(9) for $d = d_f \cdot d_g$. In fact, as we prove below (Lemmas 16 and 17), $\psi \star \phi$ does

always satisfy the second and third (Conditions (8) and (9)). Unfortunately, it is *not* always true that $\psi \star \phi$ satisfies Condition (7). An example is when $f = \mathsf{AND}_m$ and $g = \mathsf{AND}_b$. That is, if ψ is a dual witness for $\widetilde{\deg}(\mathsf{AND}_m) \geq \Omega(\sqrt{m})$ and $\widetilde{\deg}(\mathsf{AND}_b) \geq \Omega(\sqrt{b})$, then $\psi \star \phi$ is *not* a dual witness for the fact that $\widetilde{\deg}(\mathsf{AND}_m \circ \mathsf{AND}_b) \geq \Omega(\sqrt{mb})$. While the latter statement is true (since $\mathsf{AND}_m \circ \mathsf{AND}_b$ is simply AND_{mb}), the function $\psi \star \phi$ is sadly not a dual witness to this fact.

However, there are a variety of special cases in which $\psi \star \phi$ *is* known to witness that $\widetilde{\deg}_\varepsilon(f \circ g) \geq d_f \cdot d_g$ for some constant $\varepsilon \in (0,1)$. Section 6.1 describes the proof for $\mathsf{AND}_m \circ \mathsf{OR}_b$.

Lemma 16. If ψ has pure high degree d_f and ϕ has pure high degree d_g, then the pure high degree of $\psi \star \phi$ is at least $d_f \cdot d_g$.

Proof. Let us consider the representation of $\psi \star \phi \colon \{-1,1\}^{m \cdot b} \to \mathbb{R}$ as a multilinear polynomial. The lemma is equivalent to showing that the coefficient of every monomial of degree less than $d_f \cdot d_g$ is 0 (i.e., all Fourier coefficients of $\psi \star \phi$ of degree less than $d_f \cdot d_g$ are 0).

By linearity, it is without loss of generality to assume that $\psi(z_1, \ldots, z_m)$ is itself a monomial. By assumption, the degree of this monomial is at least d_f; say, $\psi(z_1, \ldots, z_m) = z_1 z_2 \ldots z_{d_f}$ (larger degree can be handled similarly). Then

$$2^{-m} \cdot (\psi \star \phi)(x) = \left(\prod_{i=1}^{d_f} \operatorname{sgn}(\phi(x_i)) \right) \left(\prod_{i=1}^{m} |\phi(x_i)| \right) = \left(\prod_{i=1}^{d_f} \phi(x_i) \right) \left(\prod_{i=d_f+1}^{m} |\phi(x_i)| \right).$$

By assumption, all monomials of ϕ have degree at least d_g. Since $x_1, \ldots, x_{d_f}$ are disjoint blocks of variables, every monomial appearing in $\prod_{i=1}^{d_f} \phi(x_i)$ has degree at least $d_f \cdot d_g$. For example, if $\phi(x_i) = \prod_{j=1}^{d_g} x_{i,j}$, then $\prod_{i=1}^{d_f} \phi(x_i) = \prod_{i=1}^{d_f} \prod_{j=1}^{d_g} x_{i,j}$. Since the blocks of variables $x_{d_f+1}, \ldots, x_m$ are disjoint from $x_1, \ldots, x_{d_f}$, multiplying this expression by $\prod_{i=d_f+1}^{m} |\phi(x_i)|$ (or any other function of $x_{d_f+1}, \ldots, x_m$ for that matter) does not decrease the degree of any appearing monomial. This proves the lemma. □

Lemma 17. If ϕ has pure high degree at least 1, then the ℓ_1-norm of $\psi \star \phi$ is 1.

Proof. Since ψ has ℓ_1-norm 1, $|\psi|$ is a probability distribution. Recall that we can think of $|\psi \star \phi|$ as first choosing z according to the probability distribution $|\psi|$, and then choosing $x = (x_1, \ldots, x_m) \in (\{-1,1\}^b)^m$ from the product distribution $\otimes_{i=1}^m \phi_{z_i}$. Hence, $|\psi \star \phi|$ is a convex combination of probability distributions, and thus is itself a probability distribution. □

6.1 The Approximate Degree of $\mathsf{AND}_m \circ \mathsf{OR}_b$ is $\Omega(\sqrt{m \cdot b})$

We've seen that whenever ψ and ϕ are dual witnesses to the high approximate degrees of f and g, respectively, then $\psi \star \phi$ has two of the three properties needed to prove that $f \circ g$ has high approximate degree (large pure high degree, and ℓ_1-norm 1). We now sketch why the third property, namely high correlation with $f \circ g$, holds in the special case of $f = \mathsf{AND}_m$ and $g = \mathsf{OR}_b$.

Lemma 18. Let ψ have correlation at least 7/8 with AND_m and ϕ have correlation at least 7/8 with OR_b. Then $\psi \star \phi$ has correlation at least 1/3 with $\mathsf{AND}_m \circ \mathsf{OR}_b$.

Proof. Recall that to sample from $|\psi \star \phi|$, one chooses a vector $z \in \{-1,1\}^m$ according to $|\psi|$ and then chooses an input $x = (x_1, \ldots, x_m) \in (\{-1,1\}^b)^m$ from the product distribution $\otimes_{i=1}^m \phi_{z_i}$. Taking this perspective, a short calculation shows that $\langle \psi \star \phi, \mathsf{AND}_m \circ \mathsf{OR}_b \rangle$ equals

$$\sum_{z \in \{-1,1\}^m} \psi(z) \cdot \mathsf{AND}_m(z) \cdot \left(1 - 2 \cdot \underbrace{\Pr_{x \sim \otimes_{i=1}^m \phi_{z_i}} [(\mathsf{AND}_m \circ \mathsf{OR}_b)(x) \neq \mathsf{AND}_m(z)]}_{:= E(z)} \right). \tag{13}$$

In other words, $\langle \psi \star \phi, \mathsf{AND}_m \circ \mathsf{OR}_b \rangle$ is the same as $\langle \psi, \mathsf{AND}_m \rangle$, but each term in the sum is adjusted by an error term $E(z)$. Since we know that ψ has high correlation with AND_m, it is enough to show that these error terms are small. Quantitatively, it will be enough to show that $E(z) \leq 1/8$ for every z.

Case 1: $z \neq -\mathbf{1}_m$**.** In this case, $(\mathsf{AND}_m \circ \mathsf{OR}_b)(x) = \mathsf{AND}_m(z)$ so long as there is *at least* one x_i such that $\mathsf{OR}_b(x_i) = 1$. Let i be any index with $z_i = 1$. Then Fact 8 combined with the assumption that ϕ has correlation at least $7/8$ with OR_b implies that $\phi_{+1}(\mathbf{1}_b) \geq 7/8$ and hence $E(z) \leq 1/8$.

Case 2: $z = -\mathbf{1}_m$**.** In this case, $(\mathsf{AND}_m \circ \mathsf{OR}_b)(x) = \mathsf{AND}_m(z)$ only if $\mathsf{OR}_b(x_i) = -1$ for *all* $i = 1, 2, \ldots, m$, i.e., if $x_i \neq \mathbf{1}_b$ for all $i = 1, 2, \ldots, m$. In this case, Corollary 9 implies that $\phi_{-1}(\mathbf{1}_b) = 0$. It follows that for all x in the support of $\otimes_{i=1}^m \phi_{-1}$, we have $x_i \neq \mathbf{1}_b$ for all $i = 1, 2, \ldots, m$. Hence, $E(-\mathbf{1}_m) = 0$. □

Lemmas 16-18, together with $\widetilde{\deg}_{7/8}(\mathsf{AND}_m) = \Theta(\sqrt{m})$ and $\widetilde{\deg}_{7/8}(\mathsf{OR}_b) = \Theta(\sqrt{b})$, imply:

Theorem 19. $\widetilde{\deg}(\mathsf{AND}_m \circ \mathsf{OR}_b) \geq \Omega(\sqrt{mb})$.

The key to the proof of Lemma 18 was Case 2, which exploited the fact that the dual witness ϕ for the inner function $g = \mathsf{OR}_b$ had one-sided error: $\{x \colon \phi(x) \cdot g(x) < 0\} \subseteq g^{-1}(-1)$ (Corollary 9), i.e., ϕ is actually a dual witness for $\widetilde{\mathrm{odeg}}_{7/8}(\mathsf{OR}_b) \geq \Omega(\sqrt{b})$. In fact, the proof of Theorem 19 shows more generally that $\widetilde{\deg}(\mathsf{AND}_m \circ g) \geq \Omega(\sqrt{m} \cdot \widetilde{\mathrm{odeg}}_{1/3}(g))$. In contrast, recall that $\widetilde{\mathrm{odeg}}_{7/8}(\mathsf{AND}_b) = 1$. This explains why dual block composition yields a good dual witness for $\mathsf{AND}_m \circ \mathsf{OR}_b$ but not for $\mathsf{AND}_m \circ \mathsf{AND}_b$, even though both functions have approximate degree $\Theta(\sqrt{mb})$.

6.2 Hardness Amplification via Dual Block Composition

Hardness amplification theorems for approximate degree show that the block composition $f \circ g$ is harder to approximate by low-degree polynomials than is g alone. Theorem 19 is an example of such a result, with $f = \mathsf{AND}_m$ and $g = \mathsf{OR}_b$, showing that the degree required to approximate $f \circ g$ to error $1/3$ is larger than the degree required to approximate g to the same error. Sherstov [She12b] used a refined version of dual block composition to prove an XOR Lemma for approximate degree showing that $\mathsf{PAR}_m \circ g$ requires both higher degree *and larger error* to approximate than g itself.

Theorem 20. ([She12b]) Let g be a Boolean function with $\widetilde{\deg}_{1/2}(g) \geq d$ and $F = \mathsf{PAR}_m \circ g$. Then $\widetilde{\deg}_{1-2^{-m}}(F) \geq \Omega(m \cdot d)$.

When using approximate degree to study AC^0, the class of constant-depth $\{\mathsf{AND}, \mathsf{OR}, \mathsf{NOT}\}$-circuits, one would like the "hardness-amplified" function F to be a constant-depth circuit whenever g is. More recent work has shown that error amplification within AC^0 is possible by taking the outer function to be AND, so long as the inner function has high one-sided approximate degree.

Theorem 21. ([BT15b]) Let g be a Boolean function with $\widetilde{\mathrm{odeg}}_{1/2}(g) \geq d$ and $F = \mathsf{AND}_m \circ g$. Then $\widetilde{\deg}_{1-2^{-m}}(F) \geq d$.

Theorem 22. ([She18b]) Let g be a Boolean function with $\widetilde{\mathrm{odeg}}_{1/2}(g) \geq d$ and $F = \mathsf{AND}_m \circ g$. Then $\deg_{\pm}(F) \geq \min\{d, m\}$.

Note that since $\widetilde{\mathrm{odeg}}_{1/2}(\mathsf{OR}_b) \geq \Omega(\sqrt{b})$, Theorem 5 is a special case of Theorem 22. That is, Minsky and Papert's threshold degree lower bound for their CNF is a special case of a far more general result that can be proved using dual block composition as opposed to symmetrization.

Proof of Theorem 21. Here, we define a simple dual witness ψ for the fact that AND_m has approximate degree at least 1 by taking $\psi(\mathbf{1}_m) = 1/2$, $\psi(-\mathbf{1}_m) = -1/2$, and $\psi(x) = 0$ otherwise. Let ϕ be any dual witness to the fact that $\widetilde{\mathrm{odeg}}_{1/2}(f) \geq d$. We claim that $\psi \star \phi = \frac{1}{2} \cdot \left(\phi_{+1}^{\otimes m} - \phi_{-1}^{\otimes m}\right)$ witnesses that $\widetilde{\deg}_{1-2^{-m}}(F) \geq d$. Note that $\psi \star \phi$ has ℓ_1-norm 1 by Lemma 17, and pure high degree d by Lemma 18 and the fact that $\mathrm{phd}(\psi) \geq 1$ and $\mathrm{phd}(\phi) \geq d$.

To show that $\langle \psi \star \phi, \mathsf{AND}_m \circ g \rangle \geq 1 - 2^{-m}$, recall from the proof of Lemma 18 (Equation (13)) that the key to showing that $\langle \psi \star \phi, \mathsf{AND}_m \circ g \rangle \approx \langle \psi, \mathsf{AND}_m \rangle = 1$ is to upper bound

$$E(z) = \Pr_{x \sim \otimes_{i=1}^m \phi_{z_i}} [(\mathsf{AND}_m \circ g)(x) \neq \mathsf{AND}_m(z)] \tag{14}$$

for the two points $z = -\mathbf{1}_m, \mathbf{1}_m$ in the support of $|\psi|$.

Case 1: $z = \mathbf{1}_m$**.** In this case, $(\mathsf{AND}_m \circ g)(x) \neq \mathsf{AND}_m(z)$ only if $g(x_1) = g(x_2) = \cdots = g(x_m) = -1$. It can be seen that since ϕ has correlation at least $1/2$ with g, $\phi_{+1}(g^{-1}(-1)) \leq 1/2$. Hence, for $z = \mathbf{1}_m$, Expression (14) is at most 2^{-m}.

Case 2: $z = -\mathbf{1}_m$**.** Since ϕ is a dual witness for the *one-sided* approximate degree of g, the support of ϕ_{-1} is a subset of $g^{-1}(-1)$, and hence the support of $\otimes_{i=1}^m \phi_{-1}$ is a subset of $(\mathsf{AND}_m \circ g)^{-1}(-1)$. Hence, for $z = -\mathbf{1}_m$, Expression (14) is 0. □

The proof of Theorem 22 builds on this construction, adding to $\psi \star \phi$ an additional "correction term" ζ of pure high degree m such that $\psi \star \phi - \zeta$ is perfectly correlated with $\mathsf{AND}_m \circ g$.

Application: Oracle separations for statistical zero knowledge. Certain applications require the "hardness-amplifying function" to be still simpler than AND_m. Define $\mathsf{GAPMAJ}_m \colon \{-1,1\}^m \to \{-1,1\}$ to be the partial function that equals -1 if at least $2/3$ of its inputs are -1, equals $+1$ if at least $2/3$ of its inputs are $+1$, and is undefined otherwise.

Theorem 23 ([BCH+19])**.** Let f be a Boolean function with $\widetilde{\deg}_{1/2}(f) \geq d$. Let $F = \mathsf{GAPMAJ}_m \circ f$. Then $\widetilde{\deg}_{1-2^{-\Omega(t)}}(F) \geq d$ and $\deg_{\pm}(F) \geq \Omega(\min\{d, m\})$.

Bouland et al. [BCH+19] used this result to exhibit an oracle $\mathcal{O}$ relative to which $\mathbf{SZK}^{\mathcal{O}} \not\subseteq \mathbf{PP}^{\mathcal{O}}$. Here **SZK** is the class of languages with efficient statistical zero knowledge proofs—proofs of membership that reveal no information other than their own validity. As **PP** is a very powerful complexity class, this separation gives some evidence for the prevailing belief that **SZK** contains intractable problems. The proof of the oracle separation proceeds as follows. Using a standard diagonalization argument, it suffices to establish a separation in the analogous query complexity models:

Fact 24. To obtain an oracle $\mathcal{O}$ such that $\mathbf{SZK}^{\mathcal{O}} \not\subseteq \mathbf{PP}^{\mathcal{O}}$, it suffices to identify an F such that $\mathbf{SZK}^{\mathbf{dt}}(F) = O(\log n)$ and $\mathbf{PP}^{\mathbf{dt}}(F) = n^{\Omega(1)}$.

Here $\mathbf{SZK}^{\mathbf{dt}}(F)$ denotes the least cost of a statistical zero knowledge *query* protocol computing F. Similarly, $\mathbf{PP}^{\mathbf{dt}}(F)$ is the least d for which a randomized algorithm making at most d queries computes $F(x)$ with probability at least $1/2 + 2^{-d}$. Since the acceptance probability of any d-query randomized algorithm is a polynomial of degree at most d, we have that if $\mathbf{PP}^{\mathbf{dt}}(F) \leq d$, then $\widetilde{\deg}_{\varepsilon}(F) \leq d$ for $\varepsilon = 1 - 2^{-d}$. So to prove a $\mathbf{PP}^{\mathbf{dt}}$ lower bound, it is enough to prove an approximate degree lower bound for an error parameter that is exponentially close to 1.

The Permutation Testing Problem (PTP) is a partial function that interprets its input x as a list of (the binary representations of) $N = \Theta(n/\log n)$ numbers from range $[N]$. The list can itself be interpreted as a function $\pi\colon [N] \to [N]$. The function $\mathsf{PTP}(x) = -1$ if π is a permutation and $\mathsf{PTP}(x) = -1$ if π is "far" from every permutation. Aaronson [Aar12] used a sophisticated symmetrization argument (building on work of Aaronson and Shi [AS04]) to show that PTP has large $(1/3)$-approximate degree. Meanwhile, Permutation Testing has a non-interactive zero-knowledge protocol with logarithmic cost: A common random string samples a range item $i \in [N]$, and the prover is required to provide a preimage j of i. The verifier can confirm that $\pi(j) = i$ by querying $\log N$ bits of x. This protocol is perfectly complete and has soundness error bounded away from 1. It is perfect zero knowledge because, when the input is a permutation, the verifier learns only a random pair (i, j) such that $\pi(j) = i$; the verifier could compute this information on its own by picking j at random from $[N]$ and making $O(\log N)$ queries to learn $i = \pi(j)$.

To get a $\mathbf{PP}^{\mathbf{dt}}$ lower bound, we need a function with high ε-approximate degree even for ε exponentially close to 1. We can transform PTP into such a function by composing it with a function that preserves **SZK** query complexity, yet amplifies hardness against polynomial approximation. Specifically, let $F = \mathsf{GAPMAJ}_{n^{1/4}} \circ \mathsf{PTP}_{n^{3/4}}$. One can show that composition with GAPMAJ preserves logarithmic **SZK** query complexity. Meanwhile, Theorem 23 implies that $\widetilde{\deg}_{1-2^{-n^{1/4}}}(F) = \Omega(n^{1/4})$.

7 Beyond Block-Composed Functions

Section 6 showed how dual block composition can yield tight lower bounds for the approximate degree of a variety of block-composed functions. However, many functions of great interest in quantum computing and complexity theory are not block composed. Can dual block composition be used to determine the approximate degree of such functions?

This turns out to be possible. For many *non*-block-composed functions f_n on n-bit inputs, the approximate degree of f_n is *equivalent* to the approximate degree of a related block-composed function F_m defined over inputs of size $m \gg n$, but under the promise that the input to F has Hamming weight at most n. That is, approximating f to error ε by a degree d polynomial is equivalent to constructing a degree d polynomial p over domain $\{-1, 1\}^m$ such that

$$|p(x) - F(x)| \leq \varepsilon \text{ for all } |x| \leq n. \tag{15}$$

Note, crucially, that p is allowed to behave arbitrarily on inputs of Hamming weight larger than n.

Let us denote by $F^{\leq n}$ the partial function obtained by restricting the domain of F to inputs of Hamming weight at most n, and by $\widetilde{\deg}_\varepsilon(F^{\leq n})$ the least degree of a polynomial p satisfying Condition (15). As we will see, if $F = f \circ g$ is a block composition of two functions whose approximate degree is understood, then dual block composition can sometimes prove tight lower bounds on $\widetilde{\deg}_\varepsilon(F^{\leq n})$.

7.1 Surjectivity: A Case Study

The above connection between a non-block-composed function f and a block composed function F is best demonstrated with an example. Let $N \geq R$ with R a power of 2. The Surjectivity function (SURJ) takes as input a vector in $x \in \{-1,1\}^n$ with $n = N \log_2 R$. It interprets the vector as a list of (the binary representations of) N numbers $(k_1, \ldots, k_N)$ from range $[R] = \{1, \ldots, R\}$, and it outputs -1 if and only if for every $i \in [R]$, there is at least one index j such that $k_j = i$.

Approximate degree upper bound. We now relate SURJ to the block composition $\mathsf{AND}_R \circ \mathsf{OR}_N$. A natural way to do this is to consider representing the list $(k_1, \ldots, k_N) \in [R]^N$ via a set of $N \cdot R$ variables $y(x) = \{y_{i,j} \colon i \in [R], j \in [N]\}$ in which $y_{i,j} = -1$ if $k_j = i$ and $y_{i,j} = 1$ otherwise. Observe that each variable $y_{i,j}$ depends on only $\log_2 R$ bits of x, and moreover

$$\mathsf{SURJ}(x) = (\mathsf{AND}_R \circ \mathsf{OR}_N)\,(y(x)).$$

One can think of the input x to SURJ as a compressed representation of the input $y(x)$ to $\mathsf{AND}_R \circ \mathsf{OR}_N$, in that $y(x)$ consists of $N \cdot R$ bits while x consists of just $N \log_2 R$ bits.

A key observation is that for *any* input x to SURJ, the Hamming weight of the corresponding vector $y(x)$ is exactly N. This means that if p approximates $(\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}$ to error ε then $p(y(x))$ approximates SURJ to error ε, and has degree at most $\deg(p) \cdot \log_2 R$. Crucially, this holds regardless of how p behaves on inputs in $\{-1,1\}^{R \cdot N}$ of Hamming weight more than N.

Observation 25. $\widetilde{\deg}_\varepsilon(\mathsf{SURJ}) \leq \widetilde{\deg}_\varepsilon\left((\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}\right) \cdot \log_2(R)$.

We've already seen that $\widetilde{\deg}(\mathsf{AND}_R \circ \mathsf{OR}_N) = \Theta(\sqrt{RN})$. It turns out that $\mathsf{AND}_R \circ \mathsf{OR}_N$ is substantially easier to approximate when the approximation only needs to be accurate on inputs of Hamming weight at most N. Multiple proofs of this upper bound are known [She18a, BKT18]. Here we describe the approximating polynomial from [She18a].

Theorem 26. $\widetilde{\deg}\left((\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}\right) \leq O(R^{1/4} \cdot N^{1/2})$.

Proof. Let q be a polynomial over domain $\{-1,1\}^R$ of degree $O(\sqrt{R})$ that approximates AND_R to error $1/4$. A change of basis argument allows us to express q as a linear combination of *disjunctions*, i.e, terms of the form $\mathsf{OR}_S(x) = \vee_{i \in S} x_i$ for some subset $S \subseteq [R]$. Moreover, the sum of the magnitudes of the coefficients in the linear combination is at most $2^{O(\sqrt{R})}$.

Clearly $|s \circ \mathsf{OR}_N - \mathsf{AND}_R \circ \mathsf{OR}_N| \leq 1/4$. Because the composition of any two disjunctions is itself a disjunction, $s \circ \mathsf{OR}_N$ is itself a linear combination of disjunctions over domain $\{-1,1\}^{RN}$ in which the sum of the magnitudes of the coefficients is at most $W \leq 2^{O(\sqrt{R})}$. Let us write this linear combination as

$$(q \circ \mathsf{OR}_N)(y) = \sum_{S \subseteq \{-1,1\}^{RN}} c_S \cdot \mathsf{OR}_S(y). \tag{16}$$

Here is where we exploit the fact that we only require our final approximation to accurately approximate $\mathsf{AND}_R \circ \mathsf{OR}_N$ on inputs of Hamming weight at most N. A generalization of the Chebyshev-based construction in Section 3 shows that $\widetilde{\deg}_\varepsilon(\mathsf{OR}_{R\cdot N}^{\leq N}) \leq O\left(\sqrt{N\log(1/\varepsilon)}\right)$ for any $\varepsilon > 0$ (see Footnote 6), regardless of $R \cdot N$. Note that the approximating polynomial may be exponentially large in its degree for inputs x of Hamming weight more than N.

Now set $\varepsilon = 1/(12W)$, and let us replace each disjunction OR_S on the right hand side of Equation (16) with an ε-approximation to $\mathsf{OR}_S^{\leq N}$. The resulting polynomial p has degree $O(\sqrt{N\log(1/W)}) = O(R^{1/4}N^{1/2})$. On any input y of Hamming weight at most N, we have $|\,(s \circ \mathsf{OR}_N)\,(y) - p(y)| \leq 1/12$ and hence $|\,(\mathsf{AND}_R \circ \mathsf{OR}_N)\,(y) - p(y)| \leq 1/12 + 1/4 = 1/3$. □

Approximate degree lower bound. One might suspect that the approximation for SURJ constructed above is unnecessarily tying its own hands by ignoring all structure in the vector $y(x)$ besides the fact that $y(x)$ has Hamming weight at most N. For example, it is ignoring the fact that for each $j \in [N]$, $y_{i,j} = -1$ for *exactly* one index $i \in [R]$. It turns out that this additional structure in the vector $y(x)$ cannot be leveraged by low-degree polynomials. That is, the approximate degree of SURJ is not just upper bounded by that of $(\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}$, but in fact is equivalent to it.

Lemma 27. $\widetilde{\deg}(\mathsf{SURJ}) \geq \widetilde{\deg}((\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N})$.

Lemma 27 was shown in [BT19b] using a symmetrization argument due to Ambainis [Amb05]. A tight lower bound on the approximate degree of SURJ now follows from one for $\widetilde{\deg}((\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N})$, which can be proved by dual block composition.

Theorem 28 ([BKT18])**.** $\widetilde{\deg}((\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}) \geq \tilde{\Omega}(R^{1/4} \cdot N^{1/2})$.

Proof sketch. Let ψ be any dual polynomial for the fact that $\widetilde{\deg}_{7/8}(\mathsf{AND}_R) \geq \Omega(R^{1/4})$, and let $N' \coloneqq N/R^{1/2}$. It turns out to be useful to focus on the function $(\mathsf{AND}_R \circ \mathsf{OR}_{N'})^{\leq N}$ rather than $(\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}$ (the former is a subfunction of the latter, so a lower bound for the former will imply our desired lower bound for the latter).

Let ϕ be the dual polynomial for $\deg_{7/8}(\mathsf{OR}_{N'}) \geq \Omega(\sqrt{N'})$ constructed in Section 5.1. Lemmas 16-18 show that $\psi \star \phi$ is a dual polynomial for the fact that $\widetilde{\deg}(\mathsf{AND}_R \circ \mathsf{OR}_{N'}) \geq \Omega(\sqrt{R \cdot N'}) \geq \Omega(R^{1/4} \cdot N^{1/2})$. Unfortunately, this is not enough, as we need our degree lower bound to hold against polynomials that can behave arbitrarily on inputs of Hamming weight larger than N, i.e., we must lower bound $\widetilde{\deg}((\mathsf{AND}_R \circ \mathsf{OR}_{N'})^{\leq N})$.

The property making this possible is that $|\psi \star \phi|$ places *very little* mass on inputs of Hamming weight larger than N. Quantitatively,

$$\sum_{y \in \{-1,1\}^{R\cdot N'}\colon |y| > N} |(\psi \star \phi)(y)| \leq 2^{-\Omega(N/\sqrt{N'})} = 2^{-\Omega(R^{1/4}N^{1/2})}. \tag{17}$$

At a high level, this bound arises as follows. Theorem 10 shows that $|\phi|$ places most of its mass on inputs of very low Hamming weight. In particular, an exponentially small fraction of its mass lies on inputs of Hamming weight more than $\sqrt{N'}$. Recall that the probability distribution $|\psi \star \phi|$ can be thought of as first choosing z according to the distribution $|\psi|$, and then choosing $y = (y_1, \ldots, y_R) \in \left(\{-1,1\}^{N'}\right)^R$ from the product distribution $\otimes_{i=1}^R \phi_{z_i}$. Because $|\phi|$ (and hence also ϕ_{+1} and ϕ_{-1}) places such little mass on inputs of Hamming weight more $\sqrt{N'}$, it turns out that

for $y = (y_1, \ldots, y_R) \sim \otimes_{i=1}^{R} \phi_{z_i}$, the probability that y has Hamming weight greater than N is dominated by the probability of the following event: there are at least $\ell := N/\sqrt{N'}$ values of i for which $|y_i| \approx \sqrt{N'}$. And this probability is exponentially small in ℓ. We now explain how Condition (17) implies that $\widetilde{\deg}((\mathsf{AND}_R \circ \mathsf{OR}_{N'})^{\leq N}) \geq d$ for $d = \mathrm{phd}(\psi \star \phi)/\log N$. Suppose p approximates $\mathsf{AND}_R \circ \mathsf{OR}_{N'}$ for all inputs of Hamming weight at most N. Then in particular, $|p(y)| \leq 4/3$ for all $|y| \leq d < N$. An interpolation argument of Razborov and Sherstov shows that this implies p is bounded in magnitude by $\exp(\tilde{O}(d))$ for *all* inputs, even those of very large Hamming weight.

Lemma 29 ([RS10])**.** Let $p\colon \{-1,1\}^{R \cdot N} \to \mathbb{R}$ be a polynomial of degree at most d. If $|p(y)| \leq O(1)$ for all $|y| \leq N$, then $|p(y)| \leq (RN)^{O(d)}$ for all $y \in \{-1,1\}^{RN}$.

Hence, we conclude that $|p(y)| \leq (RN)^{O(d)} = 2^{O(R^{1/4}N^{1/4})}$ for all $y \in \{-1,1\}^{RN}$. Now recall that, as captured in Claim 12, if a dual polynomial for a function F places mass at most δ on a set S, then it in fact lower bounds the degree of polynomial approximations p to F that are permitted to be as large as roughly $1/\delta$ at inputs in S. Taking S to be the set of all inputs of Hamming weight greater than N and $\delta = 2^{-\Omega(R^{1/4}N^{1/2})}$, Condition (17) thus implies that p requires degree at least d. This completes the proof. □

7.2 Other Functions and Applications to Quantum Query Complexity

A number of other problems that arise in quantum query complexity can be related to block-composed functions under a Hamming weight promise. For example, the k-distinctness function ED^k interprets its input as a list of N numbers from a range of size R and outputs -1 if and only if there is some range item that appears at least k times in the list. It is easy to see that $\mathsf{ED}^k(x) = (\mathsf{OR}_R \circ \mathsf{THR}_N^k)(y(x))$ where THR^k denotes the symmetric k-threshold function that outputs -1 iff its input has Hamming weight at least k. As before, we have:

Lemma 30. For $k \geq 2$, $\widetilde{\deg}(\mathsf{ED}^k) = \tilde{\Theta}(\widetilde{\deg}(\big(\mathsf{OR} \circ \mathsf{THR}_N^k\big)^{\leq N}))$.

Dual block composition can be used to show that $\widetilde{\deg}(\big(\mathsf{OR} \circ \mathsf{THR}_N^k\big)^{\leq N}) \geq \Omega(N^{3/4-1/(4k)})$ for any constant $k \geq 2$ [BKT18, MTZ20]. For large k, this nearly matches a known upper bound of $O\left(n^{3/4-\frac{1}{2^{k+2}-4}}\right)$ on the quantum query complexity, and hence also approximate degree, of ED^k [Bel12]. Similar connections give tight lower bounds (up to logarithmic factors) on both the approximate degree and quantum query complexity of various property testing problems, including junta testing, statistical distance estimation, entropy approximation, and image size testing [BKT18].

7.3 Approximate Degree of AC^0

One of our favorite open questions in the study of approximate degree is to ascertain whether there are AC^0 circuits of approximate degree $\Omega(n)$. The Parity and Majority functions have linear approximate degree, but they are not in AC^0. For a long time, the best known lower bound on the approximate degree of an AC^0 function was $\Omega(n^{2/3})$, proved by Aaronson and Shi [AS04]. Analyzing non-block-composed functions, as described above, brings us a lot closer to answering this question. In particular, SURJ is in AC^0 and has approximate degree $\tilde{\Theta}(n^{3/4})$. In fact, the key to the SURJ lower bound (Theorem 28) can be seen as another hardness amplification theorem, showing that the function $(\mathsf{AND}_R \circ \mathsf{OR}_N)^{\leq N}$ requires higher degree to approximate than does AND_R itself. The main property of AND_R used in Theorem 28 is that it has approximate degree $\Omega(\sqrt{R})$. Simplifying

the actual construction slightly, replacing AND_R with SURJ_R yields a function $(\mathsf{SURJ}_R \circ \mathsf{OR}_N)^{\leq N}$ that has even larger approximate degree $\widetilde{\Omega}(n^{7/8})$.

By iteratively applying this hardness amplification technique, for any $\delta > 0$, one can obtain a family of AC^0 circuits with approximate degree $\Omega(n^{1-\delta})$ [BT19b, BKT18]. This was further improved by the authors from (1/3)-approximate degree to $(1 - 2^{-n^{1-\delta}})$-approximate degree [BT19a], and finally by Sherstov and Wu [SW19] to a $\Omega(n^{1-\delta})$ lower bound on the threshold degree of AC^0.

8 Open Questions

A direct sum theorem for approximate degree? Open Problem 14 asks whether, for every pair of functions f, g, $\widetilde{\deg}(f \circ g) \geq \Omega(\widetilde{\deg}(f) \cdot \widetilde{\deg}(g))$. Two partial results are state of the art. One result applies to arbitrary functions f, g.

Theorem 31 ([She12b])**.** For any Boolean functions $f : \{-1,1\}^m \to \{-1,1\}$ and g,

$$\widetilde{\deg}(f \circ g) \geq \Omega\left(\widetilde{\deg}(f) \cdot \widetilde{\deg}_{1-\widetilde{\deg}(f)/m}(g)\right).$$

This result resolves Open Problem 14 whenever the outer function f has linear approximate degree. It is proved using a similar variant of dual block composition underlying the "XOR lemma," Theorem 20. The other result does *not* apply to arbitrary functions f, g, but resolves Open Problem 14 (up to a logarithmic factor) in the special case that the outer function f is symmetric.

Theorem 32 (Bouland et al. [BDBGK18])**.** Let $f : \{-1,1\}^m \to \{-1,1\}$ be a symmetric Boolean function and g be an arbitrary function. Then $\widetilde{\deg}(f \circ g) \cdot \log m \geq \Omega(\widetilde{\deg}(f) \cdot \widetilde{\deg}(g))$.

Theorem 32 is not proved using the method of dual polynomials, but rather indirectly relies on a sophisticated quantum algorithm for combinatorial group testing, due to Belovs [Bel15].

Circuit lower bounds. Our recent work with Kothari [BKT19] built on the techniques used to prove Theorem 26 to show that any linear-size constant-depth circuit has sublinear approximate degree. This in turn yielded state-of-the-art lower bounds on the size of $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ circuits computing the Boolean inner product function on average. (These are constant-depth $\mathsf{AND}/\mathsf{OR}/\mathsf{NOT}$ circuits augmented with a layer of parity gates adjacent to the inputs—fully understanding their power seems to require new circuit lower bound techniques, i.e., beyond random restrictions.) However, our circuit lower bound is still *very* far from what we believe to be true. It is only slightly superlinear, while it is widely believed that exponential-size $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ circuits are needed to compute the Boolean inner product function.

A barrier to further progress is that we do not know a sublinear upper bound on the approximate degree of quadratic-size AC^0 circuits (even depth-2 circuits, i.e., CNFs), or on the threshold degree of quadratic-size depth-3 circuits. At the same time, our lower bound techniques for the approximate degree of AC^0 appear stuck at $\Omega(n^{1-\delta})$ for any constant $\delta > 0$.

While the potential space for improving the known lower bound of $\Omega(n^{1-\delta})$ appears small, we believe that significant progress in circuit complexity would follow from a matching upper bound. Could there be a positive function $\delta(d, c)$ such that every family of depth-d, size-n^c AC^0 circuits has approximate degree—or threshold degree—$O(n^{1-\delta(d,c)})$?

New paradigms in quantum algorithm design? Approximate degree has been useful in quantum computing because it lower bounds quantum query complexity [BBC$^+$01]. The converse is not true: An approximate degree upper bound of d does not imply a quantum query algorithm making $O(d)$ queries. However, recent work has identified variants of approximate degree that are closer to quantum query complexity. In fact, Arunachalam, Briët, and Palazuelos [ABP19] showed that quantum query complexity is *characterized* by one of these variants, called approximation by completely-bounded forms. Still, no one has yet been able to use this characterization to give a new quantum algorithm for any problem. Can known constructions of approximating polynomials be modified to yield completely-bounded forms? If so, this has the potential to be a powerful new paradigm in quantum algorithm design.

Acknowledgements. We are grateful to Lane Hemaspaandra for inviting us to write this column and to both Lane and Shuchen Zhu for valuable comments on this manuscript.

References

[Aar08] Scott Aaronson. The polynomial method in quantum and classical computing. In *Foundations of Computer Science*, page 3, 2008.

[Aar12] Scott Aaronson. Impossibility of succinct quantum proofs for collision-freeness. *Quantum Information & Computation*, 12(1-2):21–28, 2012.

[ABP19] Srinivasan Arunachalam, Jop Briët, and Carlos Palazuelos. Quantum query algorithms are completely bounded forms. *SIAM Journal on Computing*, 48(3):903–925, 2019.

[All89] Eric Allender. A note on the power of threshold circuits. In *Foundations of Computer Science*, pages 580–584, 1989.

[Amb05] 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.

[Amb18] Andris Ambainis. Understanding quantum algorithms via query complexity. In *Proceedings of the International Congress of Mathematicians*, May 2018.

[AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *Journal of the ACM*, 51(4):595–605, 2004.

[BBBV97] Charles H Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997.

[BBC$^+$01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald De Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001.

[BCDWZ99] Harry Buhrman, Richard Cleve, Ronald De Wolf, and Christof Zalka. Bounds for small-error and zero-error quantum algorithms. In *Foundations of Computer Science*, pages 358–368, 1999.

[BCH+19] Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On the power of statistical zero knowledge. *SIAM Journal on Computing*, 49(4):1–58, 2019.

[BDBGK18] Shalev Ben-David, Adam Bouland, Ankit Garg, and Robin Kothari. Classical lower bounds from quantum upper bounds. In *Foundations of Computer Science*, pages 339–349, 2018.

[Bel12] Aleksandrs Belovs. Learning-graph-based quantum algorithm for k-distinctness. In *Foundations of Computer Science*, pages 207–216, 2012.

[Bel15] Aleksandrs Belovs. Quantum algorithms for learning symmetric juntas via the adversary bound. *Computational Complexity*, 24(2):255–293, 2015.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Foundations of Computer Science*, pages 337–347, 1986.

[BIVW16] Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded indistinguishability and the complexity of recovering secrets. In *International Cryptology Conference*, volume 9816, pages 593–618, 2016.

[BKT18] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. In *Symposium on Theory of Computing*, pages 297–310, 2018.

[BKT19] Mark Bun, Robin Kothari, and Justin Thaler. Quantum algorithms and approximating polynomials for composed functions with shared inputs. In *Symposium on Discrete Algorithms*, pages 662–678, 2019. Extended and updated version available at `https://eccc.weizmann.ac.il/report/2018/156/`.

[BNRdW07] Harry Buhrman, Ilan Newman, Hein Rohrig, and Ronald de Wolf. Robust polynomials and quantum algorithms. *Theory of Computing Systems*, 40(4):379–395, 2007.

[BRS95] Richard Beigel, Nick Reingold, and Daniel A. Spielman. PP is closed under intersection. *Journal of Computer and System Sciences*, 50(2):191–202, 1995.

[BT15a] Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and Markov–Bernstein inequalities. *Information and Computation*, 243:2–25, 2015.

[BT15b] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *International Colloquium on Automata, Languages, and Programming*, pages 268–280, 2015.

[BT19a] Mark Bun and Justin Thaler. The large-error approximate degree of AC^0. In *International Conference on Randomization and Computation*, volume 145 of *LIPIcs*, pages 55:1–55:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[BT19b] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of AC^0. *SIAM Journal on Computing*, 49(4):59–96, 2019.

[CTUW14] Karthekeyan Chandrasekaran, Justin Thaler, Jonathan Ullman, and Andrew Wan. Faster private release of marginals on small databases. In *Innovations in Theoretical Computer Science*, pages 387–402, 2014.

[Gil77] John Gill. Computational complexity of probabilistic Turing machines. *SIAM Journal on Computing*, 6(4):675–695, 1977.

[GS10] Dmitry Gavinsky and Alexander A. Sherstov. A separation of NP and coNP in multiparty communication complexity. *Theory of Computing*, 6(1):227–245, 2010.

[KKMS08] Adam Tauman Kalai, Adam R Klivans, Yishay Mansour, and Rocco A Servedio. Agnostically learning halfspaces. *SIAM Journal on Computing*, 37(6):1777–1805, 2008.

[KLS96] Jeff Kahn, Nathan Linial, and Alex Samorodnitsky. Inclusion-exclusion: Exact and approximate. *Combinatorica*, 16(4):465–477, 1996.

[KS04] Adam R Klivans and Rocco A Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *Journal of Computer and System Sciences*, 68(2):303–318, 2004.

[Lee09] Troy Lee. A note on the sign degree of formulas. *arXiv preprint arXiv:0909.4607*, 2009.

[Mar90] Andrei Andreyevich Markov. On a question by DI Mendeleev. *Zapiski Imperatorskoi Akademii Nauk*, 62(1-24):12, 1890.

[MP69] Marvin Minsky and Seymour Papert. *Perceptrons: An introduction to computational geometry*. MIT Press, 1969.

[MTZ20] Nikhil S. Mande, Justin Thaler, and Shuchen Zhu. Improved approximate degree bounds for k-distinctness. In *Theory of Quantum Computation, Communication and Cryptography*, volume 158, pages 2:1–2:22, 2020.

[NS94] Noam Nisan and Mario Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4(4):301–313, 1994.

[Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *Symposium on Theory of Computing*, pages 468–474, 1992.

[Raz03] Alexander A Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145, 2003.

[RS10] Alexander A Razborov and Alexander A Sherstov. The sign-rank of AC^0. *SIAM Journal on Computing*, 39(5):1833–1855, 2010.

[She09] Alexander A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM Journal on Computing*, 38(6):2113–2129, 2009.

[She11] Alexander A Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

[She12a] Alexander A Sherstov. Making polynomials robust to noise. In *Symposium on Theory of Computing*, pages 747–758, 2012.

[She12b] Alexander A Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM Journal on Computing*, 41(5):1122–1165, 2012.

[She13a] Alexander A Sherstov. Approximating the AND-OR tree. *Theory of Computing*, 9(1):653–663, 2013.

[She13b] Alexander A Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM Journal on Computing*, 42(6):2329–2374, 2013.

[She13c] Alexander A Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica*, 33(1):73–96, 2013.

[She18a] Alexander A Sherstov. Algorithmic polynomials. In *Symposium on Theory of Computing*, pages 311–324, 2018.

[She18b] Alexander A Sherstov. Breaking the Minsky–Papert barrier for constant-depth circuits. *SIAM Journal on Computing*, 47(5):1809–1857, 2018.

[Špa08] Robert Špalek. A dual polynomial for OR. *arXiv preprint arXiv:0803.4516*, 2008.

[SW19] Alexander A Sherstov and Pei Wu. Near-optimal lower bounds on the threshold degree and sign-rank of AC^0. In *Symposium on Theory of Computing*, pages 401–412, 2019.

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009.

[Tal17] Avishay Tal. Formula lower bounds via the quantum method. In *Symposium on Theory of Computing*, pages 1256–1268, 2017.

[TUV12] Justin Thaler, Jonathan Ullman, and Salil Vadhan. Faster algorithms for privately releasing marginals. In *International Colloquium on Automata, Languages, and Programming*, pages 810–821, 2012.