Sparsity Lower Bounds for Probabilistic Polynomials

Josh Alman
Columbia University, New York, NY, USA

Arkadev Chattopadhyay
TIFR Mumbai, India

Ryan Williams
CSAIL and EECS, MIT, Cambridge, MA, USA

Abstract

Probabilistic polynomials over commutative rings offer a powerful way of representing Boolean functions. Although many degree lower bounds for such representations have been proved, sparsity lower bounds (counting the number of monomials in the polynomials) have not been so common. Sparsity upper bounds are of great interest for potential algorithmic applications, since sparse probabilistic polynomials are the key technical tool behind the best known algorithms for many core problems, including dense All-Pairs Shortest Paths, and the existence of sparser polynomials would lead to breakthrough algorithms for these problems.

In this paper, we prove several strong lower bounds on the sparsity of probabilistic and approximate polynomials computing Boolean functions when 0 means "false". Our main result is that the AND of n ORs of $c \log n$ variables requires probabilistic polynomials (over any commutative ring which isn't too large) of sparsity $n^{\Omega(\log c)}$ to achieve even $1/4$ error. The lower bound is tight, and it rules out a large class of polynomial-method approaches for refuting the APSP and SETH conjectures via matrix multiplication. Our other results include:

- Every probabilistic polynomial (over a commutative ring) for the disjointness function on two n-bit vectors requires exponential sparsity in order to achieve exponentially low error.
- A generic lower bound that any function requiring probabilistic polynomials of degree d must require probabilistic polynomials of sparsity $\Omega(2^d)$.
- Building on earlier work, we consider the probabilistic rank of Boolean functions which generalizes the notion of sparsity for probabilistic polynomials, and prove separations of probabilistic rank and probabilistic sparsity.

Some of our results and lemmas are basis independent. For example, over any basis $\{a, b\}$ for true and false where $a \neq b$, and any commutative ring R, the AND function on n variables has no probabilistic R-polynomial with $2^{o(n)}$ sparsity, $o(n)$ degree, and $1/2^{o(n)}$ error simultaneously. This AND lower bound is our main technical lemma used in the above lower bounds.

2012 ACM Subject Classification Theory of computation → Algebraic complexity theory; Theory of computation → Models of computation; Mathematics of computing → Discrete mathematics

Keywords and phrases Probabilistic Polynomials, Sparsity, Orthogonal Vectors, Probabilistic Rank

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.3

Funding *Josh Alman*: Work supported in part by NSF Grant CCF-2238221.
Arkadev Chattopadhyay: Supported by funds of Department of Atomic Energy, Govt. of India, under project RTI4001, and a Google India Research Award.
Ryan Williams: Work supported in part by NSF CCF-2127597 and NSF CCF-2420092.

1 Introduction

Let R be a ring. The sparsity of a polynomial $p \in R[x_1, \ldots, x_n]$, denoted by $\text{sparsity}(p)$, is the number of monomials of p. A *probabilistic polynomial in n variables* is a distribution $\mathcal{P}$ on n-variate polynomials. Its *degree* and *sparsity* are the maximum degree and sparsity, respectively, of any polynomial in its support. The error of $\mathcal{P}$ on a function $f : \{0,1\}^n \to \{0,1\}$ is the function from $\mathbb{N}$ to $[0,1]$ defined as

$$\text{error}(\mathcal{P}, f)(n) = \max_{x \in \{0,1\}^n} \Pr_{p \sim \mathcal{P}} [p(x) \neq f(x)] .$$

Probabilistic polynomials with low sparsity are the main technical component in a number of recent randomized algorithms for core problems. This includes the best known randomized algorithm for the dense case of All-Pairs Shortest Paths (APSP) [36], the best known randomized algorithm for Orthogonal Vectors (OV) [1], the best known deterministic algorithms for these problems (which *derandomizes* the aforementioned probabilistic polynomials) [11], as well as Constraint Satisfaction problems [37, 15], All-Pairs Nearest Neighbor problems [3, 2], systems of polynomial equations over finite fields [23], online matrix-vector multiplication [22], and Stable Matching [26]. APSP and OV are especially of interest, as substantially faster algorithms for them would have many applications (see the survey of Vassilevska-Williams [38]). Moreover, it has become popular to conjecture that it is impossible to improve the best-known runtimes of these problems by polynomial factors.

For all of these algorithms, the running time improvement is completely determined by the sparsity of a certain probabilistic polynomial: if we can find a sparser polynomial, we would directly improve the runtime. This leads to the main motivating question of this paper:

Is it possible to design sparser probabilistic polynomials to speed up these algorithms further, and in particular, can we design polynomials sparse enough to refute some of the popular conjectures?

In this paper, we prove unconditionally that several known probabilistic polynomial constructions are already optimal, in a broad sense. The particular Boolean functions of interest for APSP and OV are as follows.

For a function $c : \mathbb{N} \to \mathbb{N}$, the Boolean function $OV_c : \{0,1\}^{2 \cdot n \cdot c(n) \log n} \to \{0,1\}$, is defined by:

$$OV_c(x, y) = \bigvee_{i_1, i_2 \in [n]^2} \bigwedge_{j=1}^{c(n) \log n} (\overline{x_{i_1,j}} \vee \overline{y_{i_2,j}}).$$

The best known algorithm for OV [1] uses a probabilistic $\mathbb{F}_2$-polynomial representation of the OV_c function with $n^{O(\log c)}$ sparsity to quickly check the orthogonality of many pairs of vectors. The OV problem for n input vectors of dimension $c \log n$ could be solved in truly sub-quadratic time, if a probabilistic polynomial with sparsity $n^{O(1)}$ could be designed for OV_c. In particular, if there were a polynomial for OV_c with sparsity n^k (for some universal k) that worked for every $c \geq 1$, that would yield a fast enough algorithm for OV to refute the *Strong Exponential Time Hypothesis* (SETH) [18, 10]. For completeness, we prove this in Theorem 27 in Appendix B.

Chan and Williams [36, 11] showed that $OV_{n/\log n}$ can be used to solve APSP as well. Their algorithm, the best known randomized algorithm for APSP in dense graphs, crucially uses a probabilistic $\mathbb{F}_2$-polynomial representation of the $OV_{n/\log n}$ function with $n^{O(\log n)}$ sparsity, in order to get a $n^3/2^{\Omega(\sqrt{\log n})}$ time algorithm. If the sparsity could be improved to $n^{O(1)}$, then APSP would be solvable in truly subcubic time – an algorithmic breakthrough.

Our main result shows that in a broad sense, the sparsities of the known probabilistic polynomials for OV_c are already **optimal**, up to a constant factor in the exponent:

▶ **Theorem 1.** *For every $c : \mathbb{N} \to \mathbb{N}$, every $\varepsilon > 0$ and every finite commutative ring R such that $|R| \leq 2^{O(n^{1/2-\varepsilon})}$, R-probabilistic polynomials with error $1/4$ for OV_c which use the 0 of R to represent "false" require sparsity $n^{\Omega(\log c(n))}$.*

Theorem 1 is quite powerful, as it holds for all *finite commutative rings* rather than just finite fields. For instance, many simple functions like OR are known to have smaller polynomial representations when working modulo a composite number m rather than over a finite field [6], but we rule out a sparsity improvement over such rings as well.

Theorem 1 also rules out sparser representations over larger rings (even infinite rings like $\mathbb{R}$) without very large coefficients. For instance, given a sparse probabilistic polynomial for OV_c over the integers, we could compute all of its coefficients modulo 2 to get a probabilistic polynomial over $\mathbb{F}_2$ with at most the same degree, sparsity, and error. Even if we are working over a commutative ring R where such a trick doesn't work, it takes $n^{1/2-o(1)}$ bits to describe elements of R when $|R| \geq 2^{O(n^{1/2-\varepsilon})}$. Working with such large coefficients is prohibitive when trying to solve OV or APSP, since it would presumably multiply the runtime by such a large polynomial factor, unless polynomial evaluation techniques substantially different from the current known techniques (FFT and fast matrix multiplication) are used.

The (only) caveat of our lower bound in Theorem 1 is that it relies on using the 0 of R for false (which is very natural when working with $\wedge$ and $\vee$). We leave open the problem of proving a "basis independent" lower bound (where false can correspond to any value in R).

A General Lower Bound for AND and OR

Our main technical tool is a general (basis independent) lower bound for probabilistic polynomials representing the AND and OR functions, which is also of independent interest. The celebrated probabilistic polynomial constructions of Razborov and Smolensky, which are the key component behind the aforementioned probabilistic polynomials in many recent algorithms (including APSP and OV) yield:

▶ **Theorem 2** ([29, 32])**.** *For all primes p and integers $t \geq 1$, OR and AND on n variables have $\mathbb{F}_p$-probabilistic polynomials of degree $(p-1)t$, sparsity $O(n^{(p-1)t})$, and error $1/p^t$.*

Using this construction, there are $\mathbb{F}_p$-probabilistic polynomials for AND on n variables with low values for any *two choices* of the sparsity, degree, and error measures:

- Sparsity $(p-1)n$ and degree $p-1$ (but error $\Omega(1/p)$) follows from Theorem 2 with $t = 1$.
- Degree $\log n$ and error $1/p^{(\log n)/(p-1)}$ (but sparsity $n^{\Omega(\log n)}$) follows from Theorem 2 with $t = (\log n)/(p-1)$.
- Sparsity 1 and error 0 (but degree $\Omega(n)$) is achieved by the exact polynomial $x_1 x_2 \cdots x_n$.

A natural question arises: is there a probabilistic polynomial for AND with low degree, low sparsity, and low error *simultaneously*?

▶ **Question 1.** *Is there a probabilistic polynomial for AND on n variables over any commutative ring which has $O(\log n)$ degree, $1/poly(n)$ error, and $poly(n)$ sparsity?*

While Question 1 is fundamental in its own right, a positive answer would also contradict a non-uniform version of SETH, by using it to construct sparse probabilistic polynomials for OV_c; see the discussion in Appendix B for more details. In constructing these polynomials, there is a choice of which values in our ring R map to true and false. We could use $\{0, 1\}$,

or $\{1, -1\}$, or even a basis like $\{1, 2\}$ [16]. While the degrees of polynomials do not change under such transformations, it is known that in general, the sparsity of polynomials can be very sensitive to this map [21]. Could this be exploited to refute SETH?

We unconditionally prove a strong **no** answer to Question 1:

▶ **Theorem 3.** *For every commutative ring R, every distinct $a, b \in R$, and all $d \geq 1$ and $e > 0$, there is a $c \geq 1$ such that for all sufficiently large m, the AND function on m variables does not have a R-probabilistic polynomial, where a represents false and b represents true, of sparsity $2^{dm/c}$, degree m/c, and error $2^{-em/c}$.*

Indeed, a polynomial giving a yes answer to Question 1, upon setting all but $m = \omega(\log n)$ variables to 1 (or whichever element of the ring R corresponds to true), would contradict Theorem 3, since sparsity, degree, and error cannot increase upon setting variables. We emphasize that Theorem 3 holds over *any* commutative ring R and *any* choice of basis $\{a, b\}$.

Further Results

Our technical lemmas can be applied to yield several sparsity lower bounds for various algebraic representations of natural and well-studied Boolean functions. First, we give an exponential sparsity-error tradeoff lower bound for probabilistic polynomials for the disjointness function $DISJ$ and the "complements" function $COMP$. Informally, $DISJ$ determines whether two given bit-vectors have zero inner product, and $COMP$ determines whether two given bit-vectors are binary complements of each other; it is very closely related to the equality function (see Section 2 for formal definitions).

▶ **Theorem 4.** *For every commutative ring R, function $f : \mathbb{N} \to \mathbb{N}$, constants $e > 0$ and $d \geq 1$, there is a $c \geq 1$ such that the disjointness of two n bit vectors does not have a R-probabilistic polynomial of sparsity $f(c)2^{dn/c}$ and error $f(c)/2^{en/c}$ which uses the 0 of R to represent "false".*

▶ **Theorem 5.** *Theorem 4 also holds with disjointness replaced by $COMP$.*

That is, we cannot obtain a probabilistic polynomial for disjointness with subexponential sparsity and subexponential error simultaneously, *regardless* of the degree. (Observe that the exact representation of disjointness of two n-bit vectors has 2^n sparsity and 0 error.)

In the appendices, we also apply the above results in a number of other ways.

In Appendix A, we give a simple equivalence between probabilistic polynomials and certain depth-3 circuits, which makes some prior work relevant. In Appendix B, we discuss how sparse probabilistic polynomials for simple AC^0 functions would refute SETH:

▶ **Theorem 6** (Informal; see Theorems 24 and 27)**.** *For any commutative ring R, either of the following would refute SETH:*

- *A low-degree, low-sparsity, and low-error R-probabilistic polynomial for AND (over 0/1), or*
- *An R-probabilistic polynomial for $OR_n \circ AND_{c \log n} \circ OR_2$ with sparsity n^k and constant error for any fixed integer k.*

In Appendix C, we investigate probabilistic polynomials over the $\{-1, 1\}$ basis instead of $\{0, 1\}$. While the degree of a polynomial does not depend on the basis, the sparsity can be very sensitive to such a change, and we give examples where probabilistic polynomial sparsity lower bounds in the $\{0, 1\}$ setting do not hold in the $\{-1, 1\}$ setting. Building on our lower bound Theorem 5 over the $\{0, 1\}$ basis, we prove:

▶ **Theorem 7.** *Over any commutative ring R which doesn't have characteristic 2, the complements function $COMP$ has a probabilistic polynomial with sparsity poly(n) and error $1/$poly(n) over the basis $\{-1, 1\}$, and does not over the basis $\{0, 1\}$.*

In Appendix D, we consider a notion related to the probabilistic sparsity of a function, introduced by [4], called the *probabilistic rank*. The probabilistic rank of a function f measures the extent to which the truth table matrix, or communication matrix, of f can be probabilistically represented by low rank matrices. Probabilistic sparsity upper bounds give corresponding probabilistic rank upper bounds, a fact which is used by the aforementioned "polynomial method" algorithms to reduce polynomial evaluation to fast matrix multiplication. For some functions, the probabilistic rank could potentially be much smaller than the probabilistic sparsity. This is interesting, both for the prospect of being able to design faster algorithms using rank instead of sparsity, and given the connection between probabilistic rank and the matrix rigidity problem [4]. In Appendix D, we prove a separation between the two notions in a number of settings over $\mathbb{F}_2$, including showing that the two notions can be arbitrarily far apart for a natural function:

▶ **Theorem 8.** *For any $k > \Omega(n/\log(1/\varepsilon))$, the ε-probabilistic sparsity of $MAJ \circ XOR$ over $\mathbb{F}_2$, where MAJ has fan-in n and each XOR has fan-in k, is strictly greater than its ε-probabilistic rank. In particular, as k increases, the ε-probabilistic sparsity grows unboundedly while ε-probabilistic rank remains fixed.*

We note that there are more straightforward constructions of functions f which separate probabilistic sparsity and probabilistic rank, such as any function of large probabilistic sparsity which depends only on the first half of its input bits and thus has rank 1. However, the functions which are most interesting in algorithmic applications typically depend on both the first and second halves of the input variables similarly.

In Appendix E, we observe how bounds on one of the sparsity or degree of probabilistic polynomials for a function can give bounds on the other as well. We notably show that, if a function requires probabilistic degree d, then it must require probabilistic sparsity $\Omega(2^d)$. Hence, probabilistic degree lower bounds, which are more common in the literature, imply probabilistic sparsity lower bounds as well. We also relate probabilistic degree to probabilistic rank in a better-than-trivial way: Any n-variate Boolean function with a probabilistic polynomial of degree d and error ε has ε-probabilistic rank most $O\left(\binom{n/2}{d/2}\right)$, compared to the trivial $O\left(\binom{n/2}{d}\right)$.

Intuition for our Results

Our main proof technique, which we use to prove Theorem 3, and then again in some of our additional results, is a novel combination of two well-known ideas from the literature on lower bounds for Boolean functions and their polynomial representations.

The first idea is *carefully chosen random restrictions*. Random restrictions are convenient for studying probabilistic polynomial sparsity, since we can study each monomial of a probabilistic polynomial separately, and analyze the distribution of its resulting degree after applying a restriction. We give random restrictions that do not restrict too many variables, but that substantially lower the degree of any sparse polynomial.

The second idea is a variant of the *Schwartz-Zippel Lemma* (Lemma 13 below), which roughly says that low degree multivariate polynomials must have many roots on the Boolean hypercube. This is especially interesting, for instance, when applied to the OR function, which has only one Boolean root; it says that low degree probabilistic polynomials for OR must have high error.

Our proof strategy combines these two ideas in a new way. Given a sparse probabilistic polynomial for AND with low degree and error, we design a random restriction so that the restriction of at least one polynomial in the support of the probabilistic polynomial violates the Schwartz-Zippel Lemma, giving a contradiction. Previous work using random restrictions to get sparsity lower bounds for other types of polynomials has typically combined the random restrictions with simple degree lower bounds. Here, we need some novel technical arguments in order to apply the Schwartz-Zippel Lemma with the trade-off between the error and degree of the resulting restriction.

Of course, our description here (necessarily) glosses over many important details. In addition to random restrictions and the Schwartz-Zippel Lemma, our techniques use multi-party communication complexity, algebraic results about commutative rings (especially in our proof that Theorem 3 holds over any basis), and prior work in the area of polynomial representations of Boolean functions. Our proof extending Theorem 3 to any basis, which we give in Section G, is very general, and could be of independent interest for proving basis-independent versions of other results. It shows that for any function f, a simultaneous lower bound on the degree and sparsity of a probabilistic polynomial for f over any one basis implies a sparsity lower bound over any other basis as well.

Prior Work

Work on particular depth-three circuit size lower bounds (namely, $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits) can be seen as lower bounds on the sparsity of probabilistic polynomials (see Appendix A for an overview of the simple connection). For example:

- Razborov and Wigderson [30] showed that a simple depth-three $\mathsf{AC}^0[2]$ function f requires $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits of size $n^{\Omega(\log n)}$. This implies $f \notin \mathbb{F}_p\text{-SDE}_{0,1}[n^{o(\log n)}, n, 1/2 - \varepsilon]$ for any fixed prime p, a sparsity/degree/error lower bound for probabilistic polynomials computing f.
- Chattopadhyay [12] gave an AC^0 function requiring exponential $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{ANY}$ circuits when the bottom fan-in is $o(\log \log n)$; this corresponds to an $\Omega(\log \log n)$ degree lower bound for probabilistic polynomials computing an AC^0 function.
- Beame and Huynh [8] give a depth-9 AC^0 circuit that needs $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits of size $n^{\Omega(\log n)}$, corresponding to a probabilistic $n^{\Omega(\log n)}$-sparsity lower bound for this AC^0 function.
- Sherstov [31] gives a depth-3 AC^0 circuit with tighter $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuit size lower bound. Essentially, if the bottom fan-in is $(1/2 - \varepsilon) \log n$ for some $\varepsilon > 0$, then the circuit size must be exponential.

Our Lemma 20, which we prove on the way to proving Theorem 1, can be viewed as extending this line of work by showing a probabilistic $n^{\Omega(\log n)}$-sparsity lower bound for a depth-2 AC^0 circuit.

Polynomial sparsity has been studied in a stronger setting, where the polynomials in question only need to correlate with a target function f rather than probabilistically represent f, by Lovett and Srinivasan [24]. Their main result is that polynomials over $\mathbb{F}_2$ with $n^{o(\log n)}$ monomials have exponentially low correlation with the $\mathsf{MOD3}$ function. Note that a sparse probabilistic polynomial for a Boolean function f implies the existence of a sparse polynomial with high correlation with f, but not vice versa. Hence, Lovett and Srinivasan's lower bound applies to probabilistic polynomials as well. Their technique has similarities to some of ours – they also use a type of variable restriction to reduce low sparsity polynomials to low degree

polynomials – but their "tree restriction" technique and analysis isn't powerful enough to apply in our setting where, as we will see, we are constrained in how we are allowed to restrict our variables based on the function we are computing.

Sparsity has also been studied in the setting of polynomial threshold functions, where the sign of a polynomial dictates its output (as a Boolean function). Krause and Pudlak [21, 20] give a function f that has exponentially-high sparsity as a polynomial threshold function in the $\{-1, 1\}$ basis, but polynomially-low sparsity over $\{0, 1\}$. Other references on sparsity in the polynomial threshold function setting include Basu *et al.* [7], O'Donnell and Servedio [28], and Hansen and Poldoskii [16]. It should be noted that these polynomial threshold function sparsity lower bounds *do not suffice* to prove probabilistic polynomial sparsity lower bounds: they correspond to lower bounds against depth-2, $\mathsf{MAJ} \circ \mathsf{AND}$ circuits, compared to our lower bounds against probabilistic polynomials which correspond to depth-3, $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuits.

The degrees of probabilistic polynomials are more well-studied, starting with the work of Razborov and Smolensky [29, 32], although even some basic questions remain open. For instance, while AND is known to have constant-degree probabilistic polynomials for constant error over any fixed finite field, it is unknown what degree is needed over $\mathbb{R}$: there is a gap between the best upper bound of $O(\log n)$ [9, 34, 5] and the best lower bound of $\tilde{\Omega}(\sqrt{\log n})$ [25, 17]. Our connection between sparsity and degree which we prove in Appendix E makes use of the probabilistic degree of AND, and is thus only (close to) tight over a finite field.

2 Preliminaries

Definitions and Notation

As usual, all logarithms are assumed to be base-two. For gates $\mathcal{G}_1$ and $\mathcal{G}_2$, we write $\mathcal{G}_1 \circ \mathcal{G}_2$ to denote the circuit whose output gate is a $\mathcal{G}_1$ whose inputs are all copies of $\mathcal{G}_2$ on disjoint variables. We write AND_k, OR_k, and XOR_k to denote the AND, OR, and XOR functions on k inputs, respectively, and MAJ is the majority function which computes whether at least half of its inputs are 1.

Let $\mathcal{F} = \{f_n : \{0,1\}^n \to \{0,1\}\}$ be a decision problem, construed as an infinite family of Boolean functions (one for each $n \in \mathbb{N}$). To facilitate the presentation, we define a (non-uniform) complexity class for problems which are computable by sparse, low-degree, and low-error probabilistic polynomials over R:

▶ **Definition 9.** *A decision problem $\mathcal{F}$ is in the class R-**SDE**$[s(n), d(n), e(n)]$ if for all but finitely many n, there is a probabilistic R-polynomial $\mathcal{P}_n$ for $f_n \in \mathcal{F}$ with* **sparsity** *at most $s(n)$,* **degree** *at most $d(n)$, and* **error** *at most $e(n)$. We analogously define R-**SE**$[s(n), e(n)]$ and R-**DE**$[d(n), e(n)]$ when there is no bound on the third parameter, and R-**SDE**$_{a,b}[s(n), d(n), e(n)]$ for probabilistic polynomials over the basis where "false" means $a \in R$ and "true" means $b \in R$ (when omitted, the default is that $a = 0$ and $b = 1$).*

With this notation, we can succinctly state results about probabilistic polynomials for Boolean functions.

Many of our lower bounds concern two different Boolean functions, the *disjointness function*, and the *complements function*.

▶ **Definition 10.** *For vectors $x, y \in \{0,1\}^n$, the* disjointness function*, $DISJ : \{0,1\}^{2n} \to \{0,1\}$, is given by $DISJ(x_1, \ldots, x_n, y_1, \ldots, y_n) = \bigwedge_{i=1}^{n}(x_i \vee y_i)$. (Note that this definition negates the variables compared to how the disjointness function often appears in the literature.)*

▶ **Definition 11.** *For vectors $y \in \{0,1\}^n$, we write $\bar{y}$ to denote the* complement *of y, given by $\bar{y}_i = 1 - y_i$ for each $1 \leq i \leq n$. The* equality function*, $EQ : \{0,1\}^{2n} \to \{0,1\}$, is given by, for $x, y \in \{0,1\}^n$, $EQ(x,y) = 1$ if $x = y$ and $EQ(x,y) = 0$ otherwise. In other words, $EQ(x_1, \ldots, x_n, y_1, \ldots, y_n) := \bigwedge_{i=1}^n (x_i \oplus y_i \oplus 1)$.*

The complements function*, $COMP : \{0,1\}^{2n} \to \{0,1\}$, is given by, for $x, y \in \{0,1\}^n$, $COMP(x,y) = EQ(x,\bar{y})$. In other words, $COMP(x_1, \ldots, x_n, y_1, \ldots, y_n) := \bigwedge_{i=1}^n (x_i \oplus y_i)$.*

Chernoff Bound

In a few of our proofs, we will use a standard Chernoff bound for sums of independent Bernoulli random variables.

▶ **Lemma 12** (Chernoff bound)**.** *If $X_1, \ldots, X_n$ are independent Bernoulli random variables with sum $X = X_1 + \cdots + X_n$, and μ is the expected value of X, then we have:*

$\Pr[X \leq (1-\delta)\mu] \leq \exp\{-\delta^2\mu/2\}$ *for all* $0 < \delta < 1$*, and*

$\Pr[X \geq (1+\delta)\mu] \leq \exp\{-\delta^2\mu/3\}$ *for all* $0 < \delta$*.*

Useful Facts About Polynomials

We state two key facts from past work which we will use in the proofs of our main results. First, we will need the following variant of the Schwartz-Zippel-DeMillo-Lipton Lemma (which appears, for instance, in [27, Lemma 2.6]).

▶ **Lemma 13** (The 0-1 Schwartz-Zippel Lemma)**.** *Let p be a nonzero multilinear n-variate polynomial over any commutative ring R, of total degree at most d. Then $\Pr_{x \sim \{0,1\}^n}[p(x) \neq 0] \geq 1/2^d$.*

A proof can be found in Appendix G for completeness. Note that Lemma 13 is tight for $p(x_1, \ldots, x_d) = \prod_{i=1}^d x_i$. Second, we need a simple lower bound on the probabilistic degree of $AND \circ OR$ circuits which can be derived from a communication complexity lower bound. We refer the reader to the introduction of Sherstov's paper [31] for the relevant definitions (of the number-on-forehead communication model, and the k-party set disjointness problem), as we will only need the statement of Corollary 15 for this paper.

▶ **Theorem 14** ([31] Theorem 1.1)**.** *The number-on-forehead communication complexity of k-party set disjointness on n elements with error $1/3$ is at least $\Omega(\sqrt{n}/(2^k k))$.*

▶ **Corollary 15.** *For every $\varepsilon > 0$ and every $a > 0$, there is a $c > 0$ such that: For every sufficiently large positive integer n, every positive integer k with $k < c \cdot \log n$, and every finite commutative ring R with $|R| \leq 2^{a \cdot n^{1/2-\varepsilon}}$, the function $AND \circ OR$, where the AND has fan-in n and each OR has fan-in k, does not have a probabilistic polynomial of error $1/3$ and degree less than k.*

Proof (sketch). From such a polynomial, we can design a $1/3$ error number-on-forehead communication protocol for k-party set disjointness: draw a polynomial from the probabilistic polynomial, then map each monomial to a player whose variable is not in that monomial. Each player can evaluate their monomials and report the sum, from which they can compute set disjointness with the desired error. Each player only needs to communicate one element of R, using $O(\log(|R|)) = O(a \cdot n^{1/2-\varepsilon})$ bits of communication, so the total communication is $O(a \cdot k \cdot n^{1/2-\varepsilon})$. If $k < c \cdot \log n$ for sufficiently small $c > 0$, then this contradicts Theorem 14. ◀

3 Sparsity Lower Bounds for Probabilistic Polynomials

We begin by proving a probabilistic polynomial lower bound for AND:

▶ **Theorem 16.** *For every commutative ring R, and all $d \geq 1$ and $e > 0$, there is a $c \geq 1$ such that for all sufficiently large m, the AND function on m variables is not in R-**SDE**$[2^{dm/c}, m/c, 2^{-em/c}]$.*

For notational simplicity, we use the substitution $m = c \log n$ in our proof:

▶ **Restatement of Theorem 16.** *For every commutative ring R, and all $d \geq 1$ and $e > 0$, there is a $c \geq 1$ such that for all sufficiently large n, the AND function on $c \log n$ variables is not in R-**SDE**$[n^d, \log n, 1/n^e]$.*

Theorem 16 is a special case of Theorem 3, restricted to when the basis is $\{0, 1\}$. Theorems 4 and 5 will follow from this. We will later prove the more general Theorem 3, which holds for any basis over R, in Appendix G. In other words, *we are currently focusing on the $\{0, 1\}$ basis, but we later generalize this statement to hold over any basis.* Before we give the proof, let us show that Theorem 16 implies Theorems 4 and 5.

Proof of Theorem 4. Let P be an R-**SE**$[f(c)2^{dn/c}, f(c)/2^{en/c}]$ representation of $\text{AND}_n \circ \text{OR}_2$. From each of the OR_2 gates, pick at random one variable feeding into the gate and fix it 0. Thus, for each c', every monomial in P of degree at least $c'n/c$ survives with probability at most $1/2^{c'n/c}$. The probability that P under this restriction has degree greater than $c'n/c$ is therefore at most $f(c)2^{dn/c} \cdot \frac{1}{2^{c'n/c}}$. We set $c > c' > d$.

Note that under this restriction, the disjointness function reduces to AND_n. Thus, we have constructed a probabilistic polynomial for AND_n of degree $c'n/c$ with error $\frac{f(c)2^{dn/c}}{2^{c'n/c}} + \frac{f(c)}{2^{en/c}}$. If c is large enough, this contradicts Theorem 16. Hence our assumption about the existence of P must be false. ◀

Proof of Theorem 5. The proof is identical, since restricting one input of each XOR_2 gate to 0 reduces $COMP$ to AND as well. ◀

We now begin proving Theorem 16. We start with a simple case to illustrate our use of the 0-1 Schwartz-Zippel Lemma (Lemma 13):

▶ **Lemma 17.** *For every commutative ring R, and all $e > 1$, there is a $c > 1$ such that for all sufficiently large n, the AND function on $c \log n$ variables is not in R-**DE**$[\log n, 1/n^e]$.*

Proof. Let $\mathcal{P}$ be a probabilistic R-polynomial with error at most $1/n^e$ computing AND of $c \log n$ bits, such that every polynomial in the distribution has at most $\log n$ degree.

First we claim that, without loss of generality, the support of $\mathcal{P}$ may be assumed to not contain the identically zero polynomial z. Since $\mathcal{P}$ is a polynomial for AND with error at most $1/n^e$, on the point $y = (1, \ldots, 1)$ we must have $\Pr_{p \sim \mathcal{P}}[p(y) = 1] \geq 1 - 1/n^e$. Therefore, if z is in the support of $\mathcal{P}$, it must have probability at most $1/n^e$ of being chosen. Hence if we simply replace z in $\mathcal{P}$ with the polynomial which is identically the constant 1, the new probability of error is at most $2/n^e$.

Next, we prove there is no distribution $\mathcal{P}$ satisfying the above properties, if $e > 1$. Otherwise, by fixing randomness in $\mathcal{P}$ (i.e., by picking any polynomial in the support of $\mathcal{P}$ which achieves at most the average error), we can identify a fixed polynomial p of degree at most $\log n$ which disagrees with the AND function on at most a $1/n^e < 1/n$ fraction of inputs. By the previous paragraph, p is *not identically zero*. Since the AND of $c \log n$ variables is nonzero on only one point in $\{0, 1\}^{c \log n}$, it follows that the polynomial p is nonzero on at most $n^c/n^e + 1$ points in $\{0, 1\}^{c \log n}$. Hence the nonzero degree-$\log n$ polynomial p satisfies

$$\Pr_{x \sim \{0,1\}^{c \log n}}[p(x) \neq 0] \leq 1/n^e + 1/n^c,$$

contradicting Lemma 13 when $e, c > 1$. ◀

Before we turn to the harder part of the proof, where $e \leq 1$, we sketch the main ideas. Our first step is to hit a presumed probabilistic polynomial for AND with a random restriction of 1-inputs. The probabilities are chosen carefully so that the degree of the polynomial decreases considerably, yet the number of variables restricted is not too small, with high probability. That is, we shrink the degree of the polynomial, but we do not force it to a constant. Next, we select a non-zero polynomial from the remaining probabilistic polynomial distribution, and argue that with high probability, its degree has dropped significantly below $e \log n$, but its error on the AND function remains about $O(1/n^e)$. This contradicts Lemma 13, which says that a non-zero polynomial of degree-d must be non-zero on at least $1/2^d$ points. While the general idea of the proof is not too complicated, we need to be careful with the details, and choose parameters very particularly so that all of the required constraints hold. Let us now give the details:

Proof of Theorem 16. Assume there is a probabilistic polynomial $\mathcal{P}$ over R for AND with $d, e > 0$ such that for all $c \geq 1$, the sparsity is at most n^d, the degree is $\log n$, and the error is $1/n^e$. As in the proof of Lemma 17, we may assume that the zero polynomial is not in the support of $\mathcal{P}$. Let $\alpha, \beta, \gamma > 0$ be real-valued parameters to be set later. Define

$$g := (1+\alpha)(\log n)/\gamma, \text{ and } \ell := (1-\beta)(c \log n)/\gamma.$$

Consider the following construction $\mathcal{Q}$ of a probabilistic polynomial for AND on ℓ bits:

- Given $(x_1, \ldots, x_\ell) \in \{0,1\}^\ell$, sample $p(y_1, \ldots, y_{c \log n}) \sim \mathcal{P}$.
- For $i = 1, \ldots, c \log n$, independently and with probability $1 - 1/\gamma$, assign y_i to 1 in the polynomial p. Let p' be the remaining polynomial in $\text{vars}(p') \leq c \log n$ variables.
- If $\text{vars}(p') \geq \ell$ and $\deg(p') \leq g$, then output $p'(x_1, \ldots, x_\ell, 1, \cdots, 1)$ (where the $\cdots$ indicate $\text{vars}(p') - \ell$ ones).
- Otherwise, output a random bit.

Our central claim is that, for appropriate setting of α, β, γ, the probabilistic polynomial $\mathcal{Q}$ has degree at most g and computes the AND of ℓ bits with error at most $O(1/n^e)$. The degree of $\mathcal{Q}$ and AND functionality follow by construction; we need to bound the error. This will follow from positing a series of inequalities in the analysis that imply a small error bound, then proving at the end that the inequalities can be satisfied.

Define $\text{err} := \Pr[(\deg(p') > g) \vee (\text{vars}(p') < \ell)]$, where the probability is over the sampling in $\mathcal{P}$ (of step 1) and the random restriction to p' (of step 2). Observe that err is precisely the probability that case 4 is reached in the above procedure. Further observe that the error of $\mathcal{Q}$ is at most $1/n^e + \text{err}$: we have $1/n^e$ probability of error from $\mathcal{P}$ which is $1/n^e$, and probability err of reaching case 4.

We now claim that $\text{err} \leq 2/n^e$, when $\alpha, \beta, \gamma > 0$ are set appropriately. In particular, provided that

$$\frac{\alpha^2(1+\alpha)}{3\gamma^2} \geq d + e, \quad \text{and} \tag{1}$$

$$\frac{\beta^2 c}{2\ln(2)\gamma} \geq e, \tag{2}$$

we will have that both $\Pr[\deg(p') > g]$ and $\Pr[\text{vars}(p') < \ell]$ are at most $1/n^e$.

Observe that the number of variables $\text{vars}(p')$ in p' can be seen as a sum of independent $0-1$ random variables $Y_1, \ldots, Y_{c \log n}$, such that $\Pr[Y_i] = 1/\gamma$. Hence $\text{vars}(p')$ has expectation $(c \log n)/\gamma$ and by a Chernoff bound (Lemma 12),

$$\Pr[\text{vars}(p') < \ell] = \Pr[\sum_i Y_i < (1-\beta)(c \log n)/\gamma] \leq \exp\{-\beta^2 (c \log n)/(2\gamma)\}.$$

But if (2) holds, then $\Pr[\text{vars}(p') < \ell] < \exp\{-\beta^2 (c \log n)/(2\gamma)\} \leq \exp\{-e \ln n\} = 1/n^e$.

We will analyze $\Pr[\deg(p') > g]$ by considering the monomials of p. Let m be a monomial in $p \sim \mathcal{P}$ of degree $\deg(m) \leq \log n$. If $\deg(m) \leq g$, then the monomial m cannot possibly contribute to the event that $\deg(p') > g$ (the degree cannot increase by setting variables). So we may assume $\deg(m) > g$. There is a corresponding randomly reduced monomial m' in p', obtained by setting each of the variables in monomial m to 1 with probability $1/\gamma$. (Note this is the worst case; in some commutative rings R, the monomial m' could also cancel with another monomial and disappear from p'.) The degree $\deg(m')$ can be seen as a sum of independent 0-1 random variables $Z_1, \ldots, Z_{\deg(m)}$ where $\Pr[Z_i] = 1/\gamma$. By a Chernoff bound (Lemma 12),

$$\begin{aligned}
\Pr[\deg(m') > (1+\alpha)\deg(m)/\gamma] &= \Pr[\sum_i Z_i > (1+\alpha)\deg(m)/\gamma] \\
&\leq \exp\{-\alpha^2 \deg(m)/(3\gamma)\} \\
&< \exp\{-\alpha^2 g/(3\gamma)\} = \exp\{-\alpha^2(1+\alpha)(\log n)/(3\gamma^2)\}.
\end{aligned}$$

Now, if (1) holds, then $(\alpha^2(1+\alpha))/(3\ln(2)\gamma^2) \geq d+e$ and therefore

$$\Pr[\deg(m') > (1+\alpha)\deg(m)/\gamma] < \exp\{-\alpha^2(1+\alpha)(\log n)/(3\gamma^2)\} \leq \exp\{-(d+e)\ln n\} = 1/n^{d+e}.$$

This bound and $\deg(m) \leq \log n$ holds for every monomial m in p, and there are at most n^d monomials, so

$$\begin{aligned}
\Pr[\deg(p') > (1+\alpha)(\log n)/\gamma] &\leq \Pr\left[(\exists \text{ monomial } m')[\deg(m') > (1+\alpha)(\log n)/\gamma]\right] \\
&\leq \Pr\left[(\exists \text{ monomial } m')[\deg(m') > (1+\alpha)\deg(m)/\gamma]\right] \\
&\leq n^d/n^{d+e} = 1/n^e,
\end{aligned}$$

by the union bound.

We have proved that, assuming α, β, γ are set properly, $\mathcal{Q}$ has degree at most g and computes the AND of ℓ bits with error at most $O(1/n^e)$. Finally, suppose that

$$\frac{1+\alpha}{\gamma} < e. \tag{3}$$

Then, $\mathcal{Q}$ is a probabilistic polynomial for the AND on ℓ variables, with degree at most $g = (1+\alpha)(\log n)/\gamma < e \log n$ and error $O(1/n^e)$.

Fixing the randomness in $\mathcal{Q}$, we are in an analogous situation as the earlier case of $e > 1$: we can find a single non-zero degree-g polynomial q that differs from the AND of ℓ variables on an $O(1/n^e)$-fraction of points. Since the AND is nonzero on a $1/2^\ell$ fraction of points, the polynomial q is non-zero on at most an $O(1/n^e) + 1/2^\ell$ fraction of points. Provided that

$$\frac{(1-\beta)c}{\gamma} > e, \tag{4}$$

we will have $\ell = (1-\beta)(c \log n)/\gamma > e \log n$ and $1/n^e > 1/2^\ell$, hence q is non-zero on an $O(1/n^e)$ fraction of points. But then the degree of q is $\deg(q) \leq g \leq (e-\varepsilon)\log n$, for some $\varepsilon > 0$. This contradicts Lemma 13.

It remains to show that (1), (2), (3), and (4) can be simultaneously satisfied to yield the above contradiction, which we do in Lemma 46 in Appendix F. ◀

The full proof of Theorem 3, which generalizes the above Theorem 16 to any basis $\{a, b\}$ over R, can be found in Appendix G. It shows that picking representatives in R for true and false other than 1 and 0 cannot help to overcome the lower bound of Theorem 16. It is unfortunately not so straightforward to prove from here, since our proof of Theorem 16 critically uses the Schwartz-Zippel lemma, Lemma 13, which does not hold for general a, b over any commutative ring R. For a simple example, if $(a, b) = (0, 3)$ over $R = \mathbb{Z}/6\mathbb{Z}$, then the nonzero linear polynomial $p(x) = 2x$ is nonetheless zero on both a and b. We solve this issue by taking any probabilistic polynomial $\mathcal{P}$ over R with a basis $\{a, b\}$ where the Schwartz-Zippel lemma does not hold, and performing a combinatorial transformation to yield a new probabilistic polynomial $\mathcal{Q}$ over R with a different basis where the Schwartz-Zippel lemma does hold, and such that $\mathcal{Q}$ has the same degree and error, and only mildly greater sparsity, than $\mathcal{P}$.

3.1 Probabilistic Sparsity Lower Bounds for Compositions

We next show how probabilistic degree lower bounds for Boolean functions f can lead to probabilistic sparsity lower bounds for compositions of f with simple functions like OR and XOR.

▶ **Theorem 18.** *Let R be any commutative ring, and let $f : \{0,1\}^n \to \{0,1\}$ be any Boolean function such that any probabilistic polynomial of error ε for f over R requires degree d. Then, for every $0 < a < \varepsilon$, every $(\varepsilon - a)$-error probabilistic polynomial over R for the kn-variate function $f \circ OR_k$, the composition of f with ORs of k disjoint variables, requires sparsity at least $a \cdot k^d$ over the $\{0,1\}$ basis.*

Proof. Label the variables of our function of interest, $f \circ OR_k$, by $x_{i,j}$ for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, k\}$, so that our function can be written $f\left(\bigvee_{j=1}^{k} x_{1,j}, \ldots, \bigvee_{j=1}^{k} x_{n,j}\right)$. Consider the following random restriction ρ: for each $i = 1, \ldots, n$, choose one $j \in \{1, \ldots, k\}$ uniformly and independently at random, and set $x_{i,j'}$ to 0 for each $j' \neq j$. Notice in particular that after applying ρ to the inputs of our function $f \circ OR_k$, the result is the function f on the remaining n variables which were not set to 0.

Assume to the contrary that a probabilistic polynomial P exists for $f \circ OR_k$ with sparsity less than $a \cdot k^d$. We construct a probabilistic polynomial for f as follows:

1. Draw a $p \sim P$, and a random restriction ρ as described above.
2. Let p_ρ be the polynomial obtained by substituting 0s chosen by ρ into the variables of p.
3. If $deg(p_\rho) < d$ then output p_ρ, otherwise output a random bit.

Since the restriction ρ transforms $f \circ OR_k$ into f, we know that p_ρ as constructed in step 2 is a probabilistic polynomial for f with error $(\varepsilon - a)$. If we can show that we actually return p_ρ with probability at least $1 - a$ in step 3, then we know our probabilistic polynomial has error at most $(\varepsilon - a) + a = \varepsilon$, and it by definition has degree less than d, which will contradict our assumption about f as desired.

Observe that the monomials of p_ρ are a subset of the monomials of p. Moreover, a monomial of p appears in p_ρ if and only if none of its variables were set to 0. For a monomial m in p of degree at least d, the probability that m appears in p_ρ is at most $1/k^d$: if both x_{i,j_1} and x_{i,j_2} appear in m for some i and $j_1 \neq j_2$, then m is definitely set to zero upon restriction ρ, and otherwise, each variable in m is set to 0 with probability $1 - 1/k$ independently of the rest.

Therefore, if p has sparsity $|p|$, then the expected number of monomials of degree at least d in p_ρ is at most $|p|/k^d \leq a$. Hence, by Markov's inequality, the probability that p_ρ has degree at least d, meaning it has at least one monomial of degree at least d, is at most a, as desired. ◀

▶ **Corollary 19.** *Theorem 18 also holds with $f \circ OR_k$ replaced by $f \circ XOR_k$.*

Applying Theorem 18 and Corollary 15, we can prove that a simple depth-2 AC^0 circuit requires probabilistic sparsity $n^{\Omega(\log n)}$. (Note the lower bound is tight, as mentioned in the introduction.)

▶ **Lemma 20.** *For every $\varepsilon > 0$ and every finite commutative ring R such that $|R| \leq 2^{O(n^{1/2-\varepsilon})}$, and every function $g : \mathbb{N} \to \mathbb{N}$, R-probabilistic polynomials with error $1/4$ for $AND \circ OR$, where the AND has fan-in n and each OR has fan-in $g(n) \cdot \log(n)$, require sparsity $n^{\Omega(\log g(n))}$ over the $\{0,1\}$ basis.*

Proof. Pick a sufficiently small $c > 0$ such that Corollary 15 holds, meaning $f := AND_n \circ OR_{c \log n}$ requires R-probabilistic degree $\Omega(\log n)$ for error $1/3$. Applying Theorem 18 to f with $a = 1/12$ and $k = g(n)$, we see that $f \circ OR_{g(n)}$ requires $\mathbb{F}_2$-probabilistic sparsity $(g(n))^{\Omega(\log n)} = n^{\Omega(\log g(n))}$ for error $1/4$. Collapsing the two bottom layers of ORs, we see that $f \circ OR_{g(n)} = AND_n \circ OR_{c \cdot g(n) \log n}$, which implies the desired result since $c \cdot g(n) \log n < g(n) \log(n)$. ◀

We can finally prove our main theorem:

Proof of Theorem 1. Setting $y_{i,j} = 1$ for all i, j changes OV_g into the negation of the function from Lemma 20. The same lower bounds hold since setting variables does not increase the degree, sparsity, or error. ◀

4 Conclusion

By combining multiple techniques (random restrictions, Schwartz-Zippel, communication complexity lower bounds, known degree lower bounds, etc.), we have shown sparsity lower bounds for probabilistic and approximate polynomials computing several natural functions of interest in algorithm design. Perhaps the main question left open by our work is whether there is a polynomially-sparse constant-error probabilistic polynomial for the depth-three $OR_n \circ AND_n \circ OR_2$ circuit over *some* basis $\{a, b\}$ where $a \neq b$, working in *some* ring R. Even more generally, does this circuit have low *probabilistic rank* over some ring? In this paper, we proved several strong sparsity lower bounds over the basis $\{0, 1\}$ which significantly narrows the search space for such polynomials, but some of steps in our proofs require that the basis contains 0 ("killing" monomials by setting variables to 0).

For the complements ($COMP$) function, there is a nice sparsity upper bound over $\{-1, 1\}$ (compared to the lower bound over $\{0, 1\}$), so it is still possible (but unlikely) that the $\{-1, 1\}$ basis could support sparse probabilistic polynomials for depth-3 AC^0 functions. More basis-independent methods for sparsity lower bounds (such as Theorem 3) would be of great interest, as well as new techniques for constructing sparse polynomial representations.

References

1 Amir Abboud, Ryan Williams, and Huacheng Yu. More applications of the polynomial method to algorithm design. In *SODA*, pages 218–230, 2015.

2 Josh Alman, Timothy M Chan, and Ryan Williams. Polynomial representations of threshold functions and algorithmic applications. In *FOCS*, pages 467–476, 2016. doi:10.1109/FOCS.2016.57.

3 Josh Alman and Ryan Williams. Probabilistic polynomials and hamming nearest neighbors. In *FOCS*, pages 136–150, 2015. doi:10.1109/FOCS.2015.18.

4 Josh Alman and Ryan Williams. Probabilistic rank and matrix rigidity. In *STOC*, pages 641–652, 2017. doi:10.1145/3055399.3055484.

5 James Aspnes, Richard Beigel, Merrick Furst, and Steven Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994. doi:10.1007/BF01215346.

6 David A Mix Barrington, Richard Beigel, and Steven Rudich. Representing boolean functions as polynomials modulo composite numbers. *Computational Complexity*, 4(4):367–382, 1994. doi:10.1007/BF01263424.

7 Saugata Basu, Nayantara Bhatnagar, Parikshit Gopalan, and Richard J. Lipton. Polynomials that sign represent parity and descartes' rule of signs. *Computational Complexity*, 17(3):377–406, 2008. doi:10.1007/S00037-008-0244-2.

8 Paul Beame and Trinh Huynh. Multiparty communication complexity and threshold circuit size of AC^0. *SIAM Journal on Computing*, 41(3):484–518, 2012. doi:10.1137/100792779.

9 Richard Beigel, Nick Reingold, and Daniel Spielman. The perceptron strikes back. In *Structure in Complexity Theory Conference*, pages 286–291. IEEE, 1991. doi:10.1109/SCT.1991.160270.

10 Chris Calabro, Russell Impagliazzo, and Ramamohan Paturi. The complexity of satisfiability of small depth circuits. In *Parameterized and Exact Complexity (IWPEC)*, pages 75–85, 2009. doi:10.1007/978-3-642-11269-0_6.

11 Timothy M Chan and Ryan Williams. Deterministic apsp, orthogonal vectors, and more: Quickly derandomizing razborov-smolensky. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1246–1255. Society for Industrial and Applied Mathematics, 2016. doi:10.1137/1.9781611974331.CH87.

12 Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *FOCS*, pages 449–458. IEEE, 2007. doi:10.1109/FOCS.2007.30.

13 Ernie Croot, Vsevolod F Lev, and Péter Pál Pach. Progression-free sets in Zn. *Annals of Mathematics*, 185:331–337, 2017.

14 Zeev Dvir and Benjamin L Edelman. Matrix rigidity and the croot-lev-pach lemma. *Theory Of Computing*, 15(8):1–7, 2019. doi:10.4086/TOC.2019.V015A008.

15 Jiawei Gao, Russell Impagliazzo, Antonina Kolokolova, and R. Ryan Williams. Completeness for first-order properties on sparse structures with algorithmic applications. In *SODA*, pages 2162–2181, 2017. doi:10.1137/1.9781611974782.141.

16 Kristoffer Arnsfelt Hansen and Vladimir V Podolskii. Polynomial threshold functions and boolean threshold circuits. *Information and Computation*, 240:56–73, 2015. doi:10.1016/J.IC.2014.09.008.

17 Prahladh Harsha and Srikanth Srinivasan. On polynomial approximations to ac. *Random Structures & Algorithms*, 54(2):289–303, 2019. doi:10.1002/RSA.20786.

18 Russell Impagliazzo and Ramamohan Paturi. On the complexity of k-sat. *JCSS*, 62(2):367–375, 2001. doi:10.1006/JCSS.2000.1727.

19 Swastik Kopparty and Srikanth Srinivasan. Certifying polynomials for AC^0(parity) circuits, with applications. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2012*, pages 36–47, 2012. doi:10.4230/LIPICS.FSTTCS.2012.36.

20 Matthias Krause and Pavel Pudlák. On the computational power of depth-2 circuits with threshold and modulo gates. *Theoretical Computer Science*, 174(1-2):137–156, 1997. doi:10.1016/S0304-3975(96)00019-9.

21 Matthias Krause and Pavel Pudlák. Computing boolean functions by polynomials and threshold circuits. *computational complexity*, 7(4):346–370, 1998. doi:10.1007/S000370050015.

22 Kasper Green Larsen and R. Ryan Williams. Faster online matrix-vector multiplication. In *SODA*, pages 2182–2189, 2017. doi:10.1137/1.9781611974782.142.

23 Daniel Lokshtanov, Ramamohan Paturi, Suguru Tamaki, R. Ryan Williams, and Huacheng Yu. Beating brute force for systems of polynomial equations over finite fields. In *SODA*, pages 2190–2202, 2017. doi:10.1137/1.9781611974782.143.

24 Shachar Lovett and Srikanth Srinivasan. Correlation bounds for poly-size AC^0 circuits with $n^{1-o(1)}$ symmetric gates. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 640–651. Springer, 2011.

25 Raghu Meka, Oanh Nguyen, and Van Vu. Anti-concentration for polynomials of independent random variables. *arXiv preprint*, 2015. arXiv:1507.00829.

26 Daniel Moeller, Ramamohan Paturi, and Stefan Schneider. Subquadratic algorithms for succinct stable matching. In *Computer Science Symposium in Russia*, pages 294–308, 2016. doi:10.1007/978-3-319-34171-2_21.

27 Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational complexity*, 4(4):301–313, 1994. doi:10.1007/BF01263419.

28 Ryan O'Donnell and Rocco A Servedio. Extremal properties of polynomial threshold functions. In *IEEE Conference on Computational Complexity*, pages 3–12. Citeseer, 2003. doi:10.1109/CCC.2003.1214406.

29 A. A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

30 Alexander Razborov and Avi Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold cicuits with and gates at the bottom. *Information Processing Letters*, 45(6):303–307, 1993. doi:10.1016/0020-0190(93)90041-7.

31 Alexander A Sherstov. Communication lower bounds using directional derivatives. *Journal of the ACM (JACM)*, 61(6):34, 2014.

32 Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *STOC*, pages 77–82, 1987. doi:10.1145/28395.28404.

33 Srikanth Srinivasan. On improved degree lower bounds for polynomial approximation. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS*, pages 201–212, 2013. doi:10.4230/LIPICS.FSTTCS.2013.201.

34 Jun Tarui. Probabilistic polynomials, ac0 functions and the polynomial-time hierarchy. *Theoretical computer science*, 113(1):167–183, 1993. doi:10.1016/0304-3975(93)90214-E.

35 Ryan Williams. A new algorithm for optimal 2-constraint satisfaction and its implications. *Theoretical Computer Science*, 348(2-3):357–365, 2005. doi:10.1016/J.TCS.2005.09.023.

36 Ryan Williams. Faster all-pairs shortest paths via circuit complexity. In *STOC*, pages 664–673, 2014. doi:10.1145/2591796.2591811.

37 Ryan Williams. The polynomial method in circuit complexity applied to algorithm design (invited talk). In *34th International Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS*, pages 47–60, 2014.

38 Virginia Vassilevska Williams. On some fine-grained questions in algorithms and complexity. In *Proceedings of the ICM*, 2018.

A Depth-3 Circuits and Probabilistic Polynomials

The study of probabilistic polynomials over $\mathbb{F}_p$ is equivalent to studying depth-three circuits of a certain form; let us recall the simple connection. In the following, let $||x||$ denote the number of 1s in the binary string x.

▶ **Definition 21.** *The* approximate majority of separation δ on n bits, *denoted* MAJ_δ, *is defined only on* $x \in \{0,1\}^n$ *with* $||x|| \notin ((1/2-\delta)n, (1/2+\delta)n)$, *and is given by* $MAJ_\delta(x) = 1$ *when* $||x|| \geq (1/2+\delta)n$, *and* $f(x) = 0$ *when* $||x|| \leq (1/2-\delta)n$.

▶ **Lemma 22.** *If f is computable by a depth-three circuit with an approximate majority of separation δ at the output, PARITY gates with fan-in at most s on the middle layer, and AND gates of fan-in at most d at the bottom layer, then there is an $\mathbb{F}_2$ probabilistic polynomial for f with degree d, sparsity s, and error δ.*

Proof. An $\mathbb{F}_2$ polynomial can be viewed as a $PARITY \circ AND$ circuit, and vice versa: each monomial in such a polynomial is computing the AND of its constituent variables, and then the output of the polynomial is the $\mathbb{F}_2$ sum of these monomials, which is a $PARITY$. We thus get our probabilistic polynomial by selecting a uniformly random $PARITY \circ AND$ subcircuit which feeds into the top MAJ_δ gate of our circuit, and outputting it as an $\mathbb{F}_2$ polynomial. ◀

▶ **Lemma 23.** *For any constant $c > 0$, if there is an $\mathbb{F}_2$ probabilistic polynomial for f with degree d, sparsity s, and error δ, then f is computable by a depth-three circuit with an approximate majority of separation $(1-c)\delta$ and fan-in $O(n/\delta)$ at the output, PARITY gates with fan-in at most s on the middle layer, and AND gates of fan-in at most d at the bottom layer.*

If we didn't bound the fan-in of the approximate majority gate, then the proof would be almost identical to the proof of Lemma 22. In order to bound the fan-in by $O(n)$, we need to use a Chernoff bound:

Proof. Randomly sample t polynomials $p_1, \ldots, p_t$ from the probabilistic polynomial, for a parameter t to be selected later. For each $x \in \{0,1\}^n$, by the Chernoff bound, the probability that less than a $(1-c)\delta$ fraction of the polynomials $p_1, \ldots, p_t$ output the correct answer on x is at most $exp(-c^2\delta t)$. By setting $t = a \cdot n/(c^2\delta)$ for a sufficiently large constant a, this can be made less than 2^{-n}. Then, by the union bound, there is a probability less than 1 that, for any $x \in \{0,1\}^n$, less than a $(1-c)\delta$ fraction of the polynomials $p_1, \ldots, p_t$ output the correct answer on x. Hence, by the probabilistic method, there must be a choice of $p_1, \ldots, p_t$ such that at least a $(1-c)\delta$ fraction give the correct answer on every x. We can convert these polynomials into the desired circuit as in the proof of Theorem 22. ◀

It is interesting to note, via Lemmas 22 and 23, that we may assume that any probabilistic polynomial with constant error over $\mathbb{F}_2$ is the uniform distribution on only $O(n)$ different polynomials.

B SETH Predictions about Probabilistic Polynomials

Several algorithms based on the polynomial method in circuit complexity (such as [36, 37, 1, 3, 2]) have the following form: (a) we identify a "subcircuit" C which, if C could be evaluated on many inputs rapidly, we could solve a desired problem faster, (b) efficiently convert C into a polynomial (probabilistic, approximate, etc.) so that the evaluation task becomes algebraic, and (c) solve the evaluation task rapidly using algebraic algorithms, such as a fast Fourier transform or a matrix multiplication. How far can this approach be pushed? A significant question has been whether this kind of approach can be used to solve CNF-SAT fast enough to refute the pesky Strong Exponential Time Hypothesis (SETH).

We observe in this section that a low-degree, low-sparsity, and low-error probabilistic polynomial for AND would have contradicted SETH (recall that we unconditionally prove in this paper that there is no such polynomial). Along the way, we demonstrate other "predictions" about polynomial representations obtained by assuming SETH.

▶ **Theorem 24.** *For any commutative ring R, a low-degree, low-sparsity, and low-error R-probabilistic polynomial for AND (over 0/1) implies that there is a universal $\varepsilon > 0$ such that for all $k \geq 3$, there is a non-uniform circuit family of size $2^{(1-\varepsilon)n}$ for solving k-SAT on n variables.*[1]

The proof of Theorem 24, follows from a few claims. First, we note that a low-degree/low-sparsity/low-error probabilistic polynomial for AND over 0/1 is equivalent to low-sparsity/low-error probabilistic polynomial for $DISJ$ over 0/1. Let R be any commutative ring.

▶ **Proposition 25.** *For any $n, s, d, e \geq 0$, if $AND_n \in R$-**SDE**$[s, d, e]$ then $AND_n \circ OR_2 \in R$-**SE**$[s \cdot 3^d, e]$.*

Proof. From a probabilistic polynomial for AND_n, substitute the exact OR of two variables, $OR(x, y) = x + y - xy$, to get a probabilistic polynomial of the same error for $AND_n \circ OR_2$. In this substitution, each original monomial is replaced by a product of at most d polynomials, each of which is the OR of two variables and has three monomials; so the new polynomial has at most $s \cdot 3^d$ monomials. ◀

Next, we note that a probabilistic polynomial for disjointness ($AND_n \circ OR_2$) with low sparsity and error implies the probabilistic polynomial needed for refuting SETH:

▶ **Theorem 26.** *For any $n, s \geq 0$, if $AND_{c \log n} \circ OR_2 \in R$-**SE**$[s, 1/(8n)]$ then $OR_n \circ AND_{c \log n} \circ OR_2$ has a R-probabilistic polynomial $\mathcal{P}$ of sparsity $O(n^2 \cdot s^2)$ and error 11/24 which (when there is no error) outputs 0 for false and a nonzero value for true.*

Proof. We define $\mathcal{P}$. Let $\mathcal{Q}$ be the R-**SE**$[s, 1/(8n)]$ probabilistic polynomial for $AND_{c \log n} \circ OR_2$. Draw a $q \sim \mathcal{Q}$. By a union bound, with probability at least 7/8, q correctly computes the 0 or 1 value of $AND_{c \log n} \circ OR_2$ on the n different inputs which we want to compute the OR of. Hence, there are n polynomials $q_1, \ldots, q_n$ which each have at most s monomials, and we want to compute their OR assuming they all evalutate to 0 or 1.

The subring R' of R generated by 1 consists of all integer multiples of 1. It is isomorphic to $\mathbb{Z}/m\mathbb{Z}$ for some nonnegative integer $m \neq 1$. (The $m = 0$ case is when no positive integer multiple of 1 is 0 in R, and hence $R' = \mathbb{Z}$.) We consider three cases depending on whether $m = 0$, $m = 2$, or $m > 2$.

If $m = 0$, then simply $q_1 + \cdots + q_n$ suffices to compute OR over $\mathbb{Z}$. This has sparsity at most $n \cdot s$, and no additional error, for a total error of 1/8.

If $m = 2$, then let $S_1, S_2 \subseteq \{1, \ldots, n\}$ be two uniformly random subsets, and we pick

$$-1 + \prod_{i=1}^{2} \left(1 + \sum_{j \in S_i} q_j\right).$$

This is 0 when all the q_j are 0, and each $\sum_{j \in S_i} q_j$ is 1 with probability 1/2 otherwise, so the entire polynomial is 1 with probability at least 3/4. Hence the total error is 3/8, and the sparsity is $O(n^2 \cdot s^2)$.

If $m > 2$, then pick uniformly and independently random $a_1, \ldots, a_n \in \{1, \ldots, m\}$, and pick the polynomial $\sum_{i=1}^{n} a_i q_i$. If each $q_i = 0$ then this is 0, and otherwise it is 0 with probability at most $1/m$. The sparsity is at most $n \cdot s$, and the error is at most $1/8 + 1/m \leq 11/24$. ◀

[1] Moreover, an efficiently-samplable probabilistic polynomial for AND with these properties would refute the original Strong Exponential Time Hypothesis. All interesting probabilistic polynomial constructions we are aware of, such as those found in the works [29, 32, 9, 5, 19, 33, 3, 2], are efficiently samplable.

Finally, we note that a $n^{O(1)}$-sparse constant-error probabilistic polynomial (in the sense of Theorem 26) for $OR_n \circ AND_{c \log n} \circ OR_2$ for *every* $c \geq 1$ would refute (non-uniform) SETH:

▶ **Theorem 27.** *Suppose there is a fixed $k \geq 1$ such that for all $c \geq 1$, $OR_n \circ AND_{c \log n} \circ OR_2 \in R\text{-}\mathbf{SE}[n^k, 11/24]$. Then there is a universal $\varepsilon > 0$ such that for all $k \geq 3$, there is a non-uniform circuit family of size $2^{(1-\varepsilon)n}$ for solving k-SAT on n variables.*

Proof (sketch). The Orthogonal Vectors problem asks: *given a collection of bit vectors, is there a pair of them which are orthogonal?* We show that the hypothesis implies that for all $c \geq 1$, the Orthogonal Vectors problem with n vectors from $\{0,1\}^{c \log n}$ can be solved in $n^{2-\delta}$ time with $O(n^{2-\delta})$ advice, for a universal $\delta > 0$. The theorem then follows from a (by now) standard reduction (as in [35, 3]).

By our hypothesis, we can store (as non-uniform advice) a probabilistic polynomial $\mathcal{P}$ with sparsity $n^{\alpha k}$ and error $11/24$ for the function $OR_{n^\alpha} \circ AND_{c \log n} \circ OR_2$, where $\alpha \in (0, 1/k)$ is a sufficiently small constant. Note this requires storing a distribution of only $O(n)$ distinct $n^{\alpha k}$-sparse polynomials (see Section A).

To solve Orthogonal Vectors on n vectors, we proceed just as in Abboud-Williams-Yu [1]: partition the set of vectors into $O(n^{1-\alpha/2})$ parts, with each part containing at most $n^{\alpha/2}$ vectors. For all $O(n^{2-\alpha})$ pairs (P_1, P_2) of parts, we want to evaluate the probabilistic polynomial $\mathcal{P}$ on all pairs of vectors in the union $P_1 \cup P_2$ of the two parts. This gives an input of n^α pairs of vectors; feeding this input to a polynomial $p \sim \mathcal{P}$, we can estimate whether there is an orthogonal pair among vectors in $P_1 \cup P_2$.

The problem of solving Orthogonal Vectors now reduces to: *evaluate a random $p \sim \mathcal{P}$ with sparsity $n^{\alpha k}$, on $O(n^{2-\alpha})$ different pairs of input vectors.* When $\alpha k < 0.3$ (for example), this problem can be solved in $n^{2-\alpha+o(1)}$ time, via fast rectangular matrix multiplication. To have a high probability of success, we only need to sample $O(\log n)$ random p from our probabilistic polynomial $\mathcal{P}$. ◀

For similar reasons, many types of low sparsity polynomial representations for depth-3 AC^0, including low sparsity approximate polynomials over $\mathbb{R}$, would refute this form of SETH.

C Sparsity for −1/1 Probabilistic Polynomials

In many settings, the choice of basis (which field element corresponds to "true" and which to "false") can have a large impact on the sparsity of polynomial representations of Boolean functions. For instance, over $\{0,1\}$, computing XOR exactly on n inputs requires $\Omega(2^n)$ monomials, whereas over $\{-1,1\}$ it requires only one monomial.

We will show in this section that the choice of basis can make a big difference in the sparsity of probabilistic polynomials for the complements function (defined in Section 2). Although Theorem 16 holds over any choice of basis for our polynomials, our proof of Theorem 5 relies heavily on the fact that we are working over a basis where "false" corresponded to 0, and hence setting a variable to false eliminated every monomial that it was a part of.

Suppose we look instead at the $\{-1,1\}$ basis, with the standard convention that "true" corresponds to -1. In this setting, the statement of Theorem 5 is no longer true, since we can construct a probabilistic polynomial for the complements function with polynomial sparsity and polynomial error:

▶ **Lemma 28.** *For any $\varepsilon > 0$, and any commutative ring R which doesn't have characteristic 2, the AND of n variables has a probabilistic polynomial over R of sparsity $O(1/\varepsilon)$ and error ε over the basis $\{-1,1\}$.*

Proof. Pick a random subset $S \subseteq \{1, \ldots, n\}$, and consider the polynomial $E(x_1, \ldots, x_n) = 1 + (-1)^{|S|} \prod_{i \in S} x_i$. If $AND(x) = 1$, then some nonempty subset of the x_i's are 1's, and the probability that $\prod_{i \in S} x_i = (-1)^{|S|}$ is $1/2$. Hence, in this case, $E = 0$ with probability $1/2$. If $AND(x) = -1$, then all x_i's are -1, and we always have that $E = 2$. Hence, if we take $k = \lceil \log(1/\varepsilon) \rceil$ independent copies $E_1, \ldots, E_k$ of the above E, then the polynomial $P := 1 - \frac{1}{2^{k-1}} \prod_{i=1}^{k} E_i$ has the desired properties. ◀

▶ **Theorem 29.** *For any $\varepsilon > 0$, and any commutative ring R which doesn't have characteristic 2, $COMP$ has a probabilistic polynomial over R of sparsity $O(1/\varepsilon)$ and error ε over the basis $\{-1, 1\}$.*

Proof. The XOR of two variables is a single monomial over the $\{-1, 1\}$ basis, and so we simply substitute these monomials into the probabilistic polynomial from Lemma 28. ◀

Theorem 29 with $\varepsilon = 1/\text{poly}(n)$ shows that $COMP$ has a probabilistic polynomial with polynomial sparsity and any polynomial error over the $\{-1, 1\}$ basis, whereas Theorem 5 says that no such probabilistic polynomial can exist over the $\{0, 1\}$ basis.

Finally, we remark that just as in the $\{0, 1\}$ case, we can still construct probabilistic polynomials for OR over $\{-1, 1\}$ that achieve any two of the three goals of low sparsity, low degree, and low error:

- Low sparsity and error comes from Theorem 29,
- Low degree and error still follows from Theorem 2 with $t = (\log n)/(p-1)$, since degree and error are unchanged upon changing basis, and
- Low sparsity and degree still follows from Theorem 2 with $t = 1$, since then the degree is still $(p-1)$ and so by exploiting multilinearity, the sparsity must be $O(n^{p-1})$.

D Separation Between Probabilistic Rank and Probabilistic Sparsity

▶ **Definition 30.** *Let $a : [2^{n/2}] \to \{0,1\}^{n/2}$ be the canonical bijection, R be any ring, $f : \{0,1\}^n \to \{0,1\}$ be any Boolean function, and $A \cup B = \{x_1, \ldots, x_n\}$ be any partition of the input variables to f into two sets of size $|A| = |B| = n/2$. For $x, y \in \{0,1\}^{n/2}$, let $f(A|_x, B|_y)$ denote f evaluated at the assignment where the variables in A are set to x and the variables in B are set to y. The* truth table matrix *$M_{f,A,B}$ of f is the $2^{n/2} \times 2^{n/2}$ matrix given by $M_{f,A,B}(i,j) = f(A|_{a(i)}, B|_{b(j)})$.*

For $\varepsilon \geq 0$, and any $2^{n/2} \times 2^{n/2}$ matrix M over R, the ε-probabilistic rank of M is the minimum r for which there is a distribution $\mathcal{D}$ on $2^{n/2} \times 2^{n/2}$ matrices over R of rank at most r such that for all i, j:

$$\Pr_{D \sim \mathcal{D}}[D(i,j) \neq M(i,j)] \leq \varepsilon.$$

The ε-probabilistic rank *of f is the maximum, over all such partitions $A \cup B$, of the ε-probabilistic rank of $M_{f,A,B}$ over R.*

In this section, we compare the probabilistic sparsity of a function with its probabilistic rank. We know that the probabilistic sparsity is always an upper bound on the probabilistic rank:

▶ **Lemma 31** ([4] Corollary 2.1)**.** *For any ring R, if $f \in R$-$\mathbf{SE}[m, \varepsilon]$, then f has ε-probabilistic rank at most m over R.*

Using our probabilistic polynomial sparsity lower bounds, we are able to prove a separation in two different settings.

D.1 Complements

First, by using Theorem 5, we can give a simple explicit function, the *complements function* (defined in Section 2), which has substantially lower probabilistic rank than probabilistic sparsity over $\mathbb{F}_2$.

Recall from Theorem 5 that $COMP$ requires superpolynomial sparsity for any polynomial error over $\mathbb{F}_2$. In contrast, EQ is known to have low probabilistic rank, by simulating a communication complexity protocol for disjointness (similar to our proof of Theorem 29), and it follows that $COMP$ has low probabilistic rank as well:

▶ **Proposition 32** ([4] Lemma D.2)**.** *EQ has ε-probabilistic rank $O(1/\varepsilon)$ over any field, including $\mathbb{F}_2$.*

▶ **Proposition 33.** *$COMP$ has ε-probabilistic rank $O(1/\varepsilon)$ over any field, including $\mathbb{F}_2$.*

Proof. The communication matrix of $COMP$ is the same as that of EQ, up to a permutation of columns that swaps y with $\bar{y}$ for each $y \in \{0,1\}^{\log n}$. ◀

▶ **Corollary 34.** *$COMP$ has polynomial rank for any polynomial error over $\mathbb{F}_2$, whereas it requires superpolynomial sparsity for any polynomial error over $\mathbb{F}_2$.*

Proof. Follows from Corollary 33 and Theorem 5. ◀

D.2 Compositions with high fan-in XORs

Second, we combine Corollary 19 with Lemma 31 and a simple upper bound construction to prove a separation between probabilistic rank and probabilistic sparsity for functions of the form $f \circ XOR_k$, over $\mathbb{F}_2$.

Let $f : \{0,1\}^n \to \{0,1\}$ be any Boolean function, and $\varepsilon > 0$, $k \geq 1$ be any values. Let d be the smallest value such that f has an $\mathbb{F}_2$-probabilistic polynomial of error ε and degree d. We will prove a separation for the nk-variate function

$$f \circ XOR_k(x_{1,1}, \ldots, x_{n,k}) := f\left(\bigoplus_{j=1}^{k} x_{1,j}, \ldots, \bigoplus_{j=1}^{k} x_{n,j}\right).$$

▶ **Proposition 35.** *Over $\mathbb{F}_2$, the function $f \circ XOR_k$ has ε-probabilistic rank at most*

$$r := \sum_{i=0}^{2d} \binom{2n}{i}.$$

Proof. Suppose we have partitioned our nk variables into parts A, B, and for $1 \leq i \leq n$, let $A_i \subseteq A$ and $B_i \subseteq B$ be the corresponding partition of the variables $x_{i,1}, \ldots, x_{i,k}$. Define $y_i := \bigoplus_{x_{i,j} \in A_i} x_{i,j} \in \{0,1\}$ and $z_i := \bigoplus_{x_{i,j} \in B_i} x_{i,j} \in \{0,1\}$, and note that y_i and z_i are each functions of the variables in only one of the parts.

By assumption, f has an $\mathbb{F}_2$-probabilistic polynomial P of error ε and degree d. Suppose we draw a $p \in P$, and consider the expression

$$p(\chi(y_1, z_1), \chi(y_2, z_2), \ldots, \chi(y_n, z_n)), \tag{5}$$

where $\chi : \{0,1\}^2 \to \{0,1\}$ is the exact degree-2 polynomial for computing XOR on two variables. Expression (5) is therefore a probabilistic expression for $f \circ XOR_k$, and it is a degree $2d$ polynomial in the $2n$ variables $y_1, \ldots, y_n$ and $z_1, \ldots, z_n$, where each depends only on the variables in one side of the partition. The probabilistic rank is hence at most r by Lemma 31. ◀

Meanwhile, recall from Corollary 19 that $f \circ XOR_k$ requires $\mathbb{F}_2$-probabilistic sparsity $\varepsilon \cdot k^d$. Combining these two bounds can show a separation, for instance:

▶ **Theorem 36.** *For any*

$$k > \left(\frac{n \cdot e}{d}\right)^2,$$

the ε-probabilistic sparsity of $f \circ XOR_k$ is strictly greater than its ε-probabilistic rank. In particular, as k increases, the ε-probabilistic sparsity grows unboundedly while ε-probabilistic rank remains fixed.

Proof. With this bound on k, we have

$$k^d > \left(\frac{n \cdot e}{d'}\right)^{2d} \geq \sum_{i=0}^{2d} \binom{2n}{i},$$

as desired, where the second inequality is a standard bound on binomial coefficients. We can make the separation arbitrarily big by making k sufficiently large, since increasing k does not change the rank upper bound (the bound from Proposition 35 has no dependence on k), but increases the sparsity lower bound to as large as we would like. ◀

We can now apply known upper and lower bounds to get separations for other explicit functions. Consider, for instance, the majority function:

▶ **Theorem 37** ([29, 32])**.** *$\mathbb{F}_2$-probabilistic polynomials for MAJ on n bits with error ε require degree $\Omega(\sqrt{n \log(1/\varepsilon)})$.*

▶ **Theorem 38** ([3])**.** *There is an $\mathbb{F}_2$-probabilistic polynomial for MAJ on n bits with error ε and degree $O(\sqrt{n \log(1/\varepsilon)})$.*

Combining these with Theorem 36 yields:

▶ **Theorem 39.** *There is a constant $c > 0$ such that, for any*

$$k > \frac{c}{\log(1/\varepsilon)} \cdot n,$$

the ε-probabilistic sparsity of $MAJ \circ XOR$ over $\mathbb{F}_2$, where MAJ has fan-in n and each XOR has fan-in k, is strictly greater than its ε-probabilistic rank. In particular, as k increases, the ε-probabilistic sparsity grows unboundedly while ε-probabilistic rank remains fixed.

E Relationship between Degree, Sparsity, and Rank

Since the degree and sparsity of a probabilistic polynomial are related concepts, we are able to bound one using bounds on the other. Bounding the degree of a probabilistic polynomial will bound the sparsity of the polynomial, since there are only so many monomials of a given degree:

▶ **Proposition 40.** *For any decision problem $\mathcal{F}$, any $d, \varepsilon > 0$, and any prime p, if $\mathcal{F} \in \mathbb{F}_p$-$DE[d, \varepsilon]$, then $\mathcal{F} \in \mathbb{F}_p$-$SDE[m, d, \varepsilon]$, where*

$$m := \sum_{i=0}^{d} \binom{n}{i} \leq (1+n)^d.$$

Proof. Let P be a $\mathbb{F}_p$-DE$[d, \varepsilon]$ representation of $\mathcal{F}$. Since $x_i^2 = x_i$ whenever $x_i \in \{0, 1\}$, we may assume that any polynomial in the support of P is multilinear. Hence, each polynomial in the support of P is a multilinear polynomial of degree at most d, which has at most m monomials. P is therefore a $\mathbb{F}_p$-SDE$[m, d, \varepsilon]$ representation of $\mathcal{F}$. ◀

Perhaps more surprisingly, decision problems with low sparsity probabilistic polynomials must have low degree representations as well:

▶ **Proposition 41.** *For any decision problem $\mathcal{F}$, any $m, \varepsilon > 0$, and any prime p, if $\mathcal{F} \in \mathbb{F}_p$-SE$[m, \varepsilon]$, then $\mathcal{F} \in \mathbb{F}_p$-DE$[d, 2\varepsilon]$, where*

$$d := (p-1)\log_p(m/\varepsilon).$$

Proof. Let P be a $\mathbb{F}_p$-SE$[m, \varepsilon]$ representation of $\mathcal{F}$. We design a new probabilistic polynomial for $\mathcal{F}$, defined as follows:

1. Draw a polynomial $p(x_1, \ldots, x_n)$ from P. We can write p as a sum of monomials as $p(x_1, \ldots, x_n) = \sum_{i=1}^{m'} a_i \cdot m_i$, where $m' \leq m$, each $a_i \in \mathbb{F}_p$, and each m_i is a monomial.
2. Recall by Theorem 2 that AND $\in \mathbb{F}_p$-SDE$[O(n^{(p-1)t}), (p-1)t, 1/p^t]$ for all $t \geq 1$. Draw a polynomial q from the corresponding probabilistic polynomial on n variables with $t = \log_p(m/\varepsilon)$.
3. Each monomial m_i computes the AND of some set $x_{i,1}, x_{i,2}, \ldots, x_{i,n_i}$ of variables. It can therefore be probabilistically computed as $q(x_{i,1}, \ldots, x_{i,n_i}, 1, 1, \ldots, 1)$, where there are $(n - n_i)$ '1's. We can thus output the polynomial

$$\sum_{i=1}^{m'} a_i \cdot q(x_{i,1}, \ldots, x_{i,n_i}, 1, 1, \ldots, 1).$$

This has degree $(p-1)t = d$, and by a union bound over the error of the replacement for each monomial, it has error at most $\varepsilon + m/p^t = 2\varepsilon$. ◀

At most a logarithmic factor is lost when converting between these two bounds; if $\mathcal{F}$ has probabilistic degree d then it can have probabilistic sparsity $\leq n^d$, and if it has probabilistic sparsity n^d then it can have probabilistic degree $\leq d \cdot (p-1) \cdot \log_p(n)$.

Proposition 41 allows us to prove probabilistic sparsity lower bound from probabilistic degree lower bounds, which are much more common in the literature. For instance:

▶ **Corollary 42.** *If $f : \{0,1\}^n \to \{0,1\}$ is any Boolean function which requires $\mathbb{F}_2$-probabilistic degree d for error ε, then f requires $\mathbb{F}_2$-probabilistic sparsity $\varepsilon \cdot 2^d$ for error ε.*

Proof. A lower sparsity representation would yield a probabilistic degree less than d using Proposition 41. ◀

Combining Proposition 40 with Lemma 31, we see that if f has a probabilistic polynomial of degree d, then it has probabilistic rank at most $\sum_{i=1}^{d} \binom{n}{i}$. Applying ideas from the recent breakthrough work of Croot, Lev, and Pach [13] on Roth's theorem over $\mathbb{Z}_4^n$, we can derive an improved bound on the rank from the degree:

▶ **Proposition 43.** *Let n be even. Let $f : \{0,1\}^n \to \{0,1\}$ be any Boolean function with a probabilistic polynomial of degree d and error ε. Then, the ε-probabilistic rank of f is at most $m := 2\sum_{i=0}^{\lfloor d/2 \rfloor} \binom{n/2}{i}$.*

We note that Dvir and Edelman [14] also used this observation when studying the rigidity of a certain family of matrices. Proposition 43 will follow immediately from Theorem 45 below, but first we need a definition.

▶ **Definition 44.** *Let R be and ring, n be an even integer, and $a : [2^{n/2}] \to \{0,1\}^{n/2}$ be any bijection between the two sets. For any function $f : \{0,1\}^n \to R$, the* Boolean rank of f over R *is the minimum r such that there are matrices A, B of dimensions $2^{n/2} \times r$ and $r \times 2^{n/2}$ with $(A \cdot B)[i,j] = f(a(i), a(j))$ for all i, j.*

▶ **Theorem 45.** *Let n be even. Let P be a multilinear polynomial in n variables with degree at most d over a ring R. Then, the Boolean rank of P over R is at most $m := 2\sum_{i=0}^{\lfloor d/2 \rfloor} \binom{n/2}{i}$.*

Proof. Think of P as being over two sets of variables, $\{x_1, \ldots, x_{n/2}\}$ and $\{y_1, \ldots, y_{n/2}\}$. We begin with some notation: For sets $I, J \subseteq [n/2]$, let x^I denote $\prod_{i \in T} x_i$, and let $c_{I,J}$ be the coefficient in P of the monomial $x^I y^J$. Let $I_1, \ldots, I_t$ be a list of all subsets of $[n/2]$ of cardinality at most $\lfloor d/2 \rfloor$. Given a monomial M and a variable assignment $A \in \{0,1\}^{n/2}$, we let $M|_A$ denote the evaluation of M on the assignment A.

We want to prepare two matrices A ($2^{n/2} \times r$) and B ($r \times 2^{n/2}$), and prove for all i, j that

$$(A \cdot B)[i,j] = P(a(i), a(j)).$$

For each $i \in [2^{n/2}]$, we make a $2t$-dimensional vector

$$A[i,:] := \left(x^{I_1}|_{a(i)}, \ldots, x^{I_t}|_{a(i)}, \sum_{J \subseteq [n/2], |J| \leq |I_1|} c_{J,I_1} x^J|_{a(i)}, \ldots, \sum_{J \subseteq [n/2], |J| \leq |I_t|} c_{J,I_t} x^J|_{a(i)} \right).$$

Correspondingly, for all $j \in [2^{n/2}]$ we make an $2t$-dimensional vector

$$B[:,j] := \left(\sum_{J \subseteq [n/2], |J| < |I_1|} c_{I_1,J} y^J|_{a(j)}, \ldots, \sum_{J \subseteq [n/2], |J| < |I_t|} c_{I_t,J} y^J|_{a(j)}, y^{I_1}|_{a(j)}, \ldots, y^{I_t}|_{a(j)} \right).$$

Since every d-set of $\{x_1, \ldots, x_{n/2}, y_1, \ldots, y_{n/2}\}$ contains either at most $\lfloor d/2 \rfloor$ x_i's or at most $\lfloor d/2 \rfloor$ y_i's, we observe as desired that

$$\langle A[i,:], B[:,j] \rangle = \sum_{I,J \in \{I_1, \ldots, I_t\}} a_{I,J} x^I|_{a(i)} y^J|_{a(j)} = P(a(i), a(j)). \qquad \blacktriangleleft$$

F Missing Proof

▶ **Lemma 46.** *There is a setting to α, β, and γ which satisfies equations* (1), (2), (3), *and* (4) *from the proof of Theorem 16.*

Proof. Recall that d, e are fixed, and we may set $c \geq 1$ to be arbitrarily large.

Setting

$$\beta := \sqrt{\frac{2\ln(2)e\gamma}{c}},$$

inequality (2) immediately holds. Setting $\gamma := \alpha + 2$, we have $\frac{\alpha+1}{\gamma} = \frac{\alpha+1}{\alpha+2} < 1$, so (3) holds. To satisfy (1), we need

$$\frac{\alpha^2(1+\alpha)}{3\ln(2)\gamma^2} = \frac{\alpha^2 + \alpha^3}{\ln(2)(3\alpha^2 + 12\alpha + 12)} \geq d + e.$$

Setting $\alpha > 1$ to be sufficiently large satisfies this inequality. Finally, we turn to (4). Substituting in our above solution for β, we have

$$\frac{(1-\beta)c}{\gamma} = \frac{c - \sqrt{2\ln(2)}e^{1/2}\gamma^{1/2}c^{1/2}}{\gamma}.$$

As γ is only a function of d and e which are fixed constants, and c can be made arbitrarily large, the above expression can be made larger than e. This completes the proof. ◀

G Generalization to a-b basis

In this section, we generalize Theorem 16 to hold for a general basis, where we can choose any two elements $a, b \in R$ from our commutative ring R to represent true and false. What we prove is:

▶ **Reminder of Theorem 3.** *For every commutative ring R, every pair of distinct $a, b \in R$, and all $d \geq 1$ and $e > 0$, there is a $c \geq 1$ such that for all sufficiently large m, the AND function on m variables is not in R-$\mathbf{SDE}_{a,b}[2^{dm/c}, m/c, 2^{-em/c}]$.*

We begin by recalling and proving the version of the Schwartz-Zippel Lemma, Lemma 13, which we use in the proof of Theorem 16.

▶ **Reminder of Lemma 13.** *Let p be a nonzero multilinear n-variate polynomial over any commutative ring R, of total degree at most d. Then $\Pr_{x\sim\{0,1\}^n}[p(x) \neq 0] \geq 1/2^d$.*

Proof. Let $d' \leq d$ be the total degree of p. Without loss of generality, we may assume p has the monomial $x_1x_2\cdots x_{d'}$ with a nonzero coefficient. Consider any $\{0,1\}$ assignment to the remaining $n - d'$ variables. The resulting polynomial p' still has the monomial $x_1x_2\cdots x_{d'}$ with the same coefficient, and so it is a nonzero multilinear polynomial of degree d' in d' variables. It remains to show that p' is nonzero on at least one of the $2^{d'}$ assignments to the remaining d' variables, which will imply as desired that $\Pr_{x\sim\{0,1\}^n}[p(x) \neq 0] \geq 1/2^d$.

Let m be a monomial in p' of lowest degree with a nonzero coefficient, let $c \neq 0$ be its coefficient, and let $S \subseteq \{x_1, \ldots, x_{d'}\}$ be the set of variables in m (it may be that m is the constant term, in which case $S = \emptyset$). Then, consider the assignment which sets the variables in S to 1, and the remaining variables to 0. Our polynomial p' must evaluate to c on this assignment, since monomial m will have all variables set to 1, and all other monomials will have at least one variable set to 0 by how we chose m. This is the desired assignment since $c \neq 0$. ◀

By carefully examining our proof of Theorem 16, we see that we only used the fact that we were working with the basis $\{0,1\}$ in three ways:

- We need that substituting a value in for one variable in a monomial decreases the degree of that monomial by at least one; this is true regardless of what basis we are working over.
- We need that the correct output when the AND is false is 0, and that the correct output when AND is true is any nonzero value; by subtracting b from our polynomial we can always assume this is the case, while changing the number of monomials by at most one and leaving the degree unchanged.
- We need the Schwartz-Zippel lemma (Lemma 13) to hold for our choice of basis.

If we could generalize Lemma 13 to hold for any basis $\{a, b\}$, then our same proof would work as stated to prove Theorem 3. Unfortunately this is impossible in general: If $a, b \in R$ are such that $a - b$ is a zero-divisor, then Lemma 13 does not hold when we use the $\{a, b\}$ basis. For example, working over $R = \mathbb{Z}/6\mathbb{Z}$, with $(a, b) = (0, 3)$, the single variable linear polynomial $p(x) = 2x$ is a nonzero polynomial, but is zero on all inputs from the basis.

Although it is possible to salvage this method, at least for most choices of $\{a, b\}$, we will instead proceed in a different way, which does not rely on how our original proof works, except for the fact (as remarked) that Theorem 3 holds when the *input basis* (what values we input to the polynomial for true or false) is $\{0, 1\}$, and the *output basis* (what values the polynomial outputs for true or false) is $\{0, b\}$ for any nonzero $b \in R$.

We begin with the input basis $\{0, b\}$ for any nonzero $b \in R$, where true maps to b but false still maps to 0.

▶ **Lemma 47.** *For every commutative ring R, every nonzero $a \in R$, and all $d \geq 1$ and $e > 0$, there is a $c \geq 1$ such that for all sufficiently large m, the AND function on m variables is not in R-**SDE**$_{0,b}[2^{dm/c}, m/c, 2^{-em/c}]$.*

Proof. For a given d and e, we make the same choice of $c \geq 1$ as in Theorem 16. Assume to the contrary that AND is in R-**SDE**$_{0,b}[2^{dm/c}, m/c, 2^{-em/c}]$, and let p be the corresponding probabilistic polynomial. We will convert p into a R-**SDE**$_{0,1}[2^{dm/c}, m/c, 2^{-em/c}]$ representation, which will contradict Theorem 16.

Draw a polynomial q from p, and consider any nonzero monomial $\mathfrak{m}$ of q, with degree $\mathfrak{d}$ and nonzero coefficient $\mathfrak{a}$. If any of the variables in $\mathfrak{m}$ is assigned to 0, then $\mathfrak{m}$ evaluates to 0, and otherwise it evaluates to $\mathfrak{a} \cdot b^{\mathfrak{d}}$.

Consider the new polynonial q', given by

$$q'(x) = \sum_{\mathfrak{a} \cdot \mathfrak{m} \in q} \mathfrak{a} \cdot b^{\mathfrak{d}} \cdot \mathfrak{m}.$$

In other words, we have multiplied each monomial $\mathfrak{m}$ from q by $b^{\mathfrak{d}}$. Now, for any $x \in \{0, 1\}^n$, let $x' \in \{0, b\}^n$ be given by $x'_i = b \cdot x_i$. Then we always have that $q'(x) = q(x')$. Since x corresponds to the same true/false assignment over the $\{0, 1\}$ basis as x' does over the $\{0, b\}$ basis, this means that the resulting distribution on q' is a probabilistic polynomial with input basis $\{0, 1\}$ and output basis $\{0, b\}$ for AND with the same sparsity, degree, and error as p, as desired. ◀

Finally, we generalize to all input bases $\{a, b\}$.

Proof of Theorem 3. For a given $d \geq 1$ and $e > 0$, let $c(d, e)$ be the choice of c which is made by Lemma 47. We will choose $c = c(d+1, e)$ here. Similar to before, our plan is to convert from a R-**SDE**$_{a,b}[2^{dm/c}, m/c, 2^{-em/c}]$ representation to a R-**SDE**$_{0,b-a}[2^{(d+1)m/c}, m/c, 2^{-em/c}]$ representation, and then to invoke Lemma 47.

Assume to the contrary that AND is in R-**SDE**$_{a,b}[2^{dm/c}, m/c, 2^{-em/c}]$, and let p be the corresponding probabilistic polynomial. Draw a polynomial q from p, and consider the new polynomial q' given by

$$q'(x_1, x_2, \ldots, x_n) = q(x_1 + a, x_2 + a, \ldots, x_n + a).$$

Our desired probabilistic polynomial over the $\{0, b - a\}$ input basis will be the resulting distribution on q'. The polynomial q' clearly has the same degree as q, and errs on the same true/false inputs as q.

Consider any nonzero monomial $\mathfrak{m}$ of q, with degree $\mathfrak{d}$ and nonzero coefficient $\mathfrak{a}$. Each variable in $\mathfrak{m}$ has been replaced by the sum of two terms, and so $\mathfrak{m}$ has been replaced by $2^{\mathfrak{d}}$ monomials when expanded out. Since p had degree at most m/c, we have that $2^{\mathfrak{d}} \leq 2^{m/c}$, and so since q had at most $2^{dm/c}$ monomials, we have that q' has at most $2^{dm/c} \cdot 2^{m/c} = 2^{(d+1)m/c}$ monomials. Our probabilistic polynomial is therefore a R-**SDE**$_{0,b-a}[2^{(d+1)m/c}, m/c, 2^{-em/c}]$ representation, as desired. ◀

We note finally that, although we have been limiting ourselves to the same input basis $\{a, b\}$ for all of the n variables, we could actually choose a different input basis for each variable, and then another output basis like usual, and the above proof still holds.