On Improved Degree Lower Bounds for Polynomial Approximation

Srikanth Srinivasan

Department of Mathematics, IIT Bombay, Mumbai, India
srikanth@math.iitb.ac.in

Abstract

A polynomial $P \in \mathbb{F}[X_1, \ldots, X_n]$ is said to *ε-approximate* a boolean function $F : \{0,1\}^n \to \{0,1\}$ under distribution $\mathcal{D}$ over $\{0,1\}^n$ if $\Pr_{x \sim \mathcal{D}}[P(x) \neq F(x)] \leq \varepsilon$. Smolensky (1987) showed that for any constant primes $p \neq q$, any polynomial $P \in \mathbb{F}_p[x_1, \ldots, x_n]$ that $(\frac{1}{2q} - \Omega(1))$-approximates the boolean function $\mathrm{MOD}_q : \{0,1\}^n \to \{0,1\}$ – which accepts its input iff the number of ones is non-zero modulo q – under the uniform distribution must have degree $\Omega(\sqrt{n})$.

We consider the problem of finding an explicit function $f : \{0,1\}^n \to \{0,1\}$ that has no ε-approximating polynomial of degree less than $n^{1/2+\Omega(1)}$ under *some distribution*, for some constant $\varepsilon > 0$. We show a number of negative results in this direction: specifically, we show that many interesting classes of functions including symmetric functions and linear threshold functions do have approximating polynomials of degree $O(\sqrt{n} \cdot \mathrm{polylog}(n))$ under *every distribution*. This demonstrates the power of this model of computation.

The above results, in turn, provide further motivation for this lower bound question. Using the upper bounds obtained above, we show that finding such a function f would have applications to: lower bounds for $\mathrm{AC}^0 \circ \mathcal{F}$ where $\mathcal{F} = \mathrm{SYM} \cup \mathrm{THR}$; stronger lower bounds for 1-round compression by $\mathrm{ACC}^0[p]$ circuits; improved correlation lower bounds against low degree polynomials; and (under further conditions) showing that the Inner Product (over $\mathbb{F}_2$) function does not have small $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ circuits.

1998 ACM Subject Classification F.1.1 Models of Computation, F.1.2 Modes of Computation, F.1.3 Complexity Measures and Classes

Keywords and phrases Polynomials, Approximation, Compression, Circuit lower bounds

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2013.201

1 Introduction

Quite often, when trying to understand a model of computation (such as, say, a class of Boolean circuits), it is profitable to see if functions computed by that model can be approximated by functions that can be computed using a different, simpler, and in general, "nicer" model. Low-degree polynomials are an important example of such a nice model. Given the versatility of polynomials and the wealth of algebraic and combinatorial understanding we have of them, it is not surprising that approximations by low-degree polynomials have been used to prove a variety of results in Complexity theory [22, 24, 1, 4, 5], Learning Theory [19, 18], Derandomization [3, 8], etc..

The different applications listed above use various notions of "approximation", of what it means to be "low-degree", and which ring to choose our low-degree polynomials from. Here, we study one such means of approximation that has been useful in proving circuit lower bounds [22, 24]. We say that a polynomial $P \in \mathbb{F}[X_1, \ldots, X_n]$ (for some field $\mathbb{F}$) *ε-approximates* a Boolean function $F : \{0,1\}^n \to \{0,1\}$ over a probability distribution $\mathcal{D}$ on $\{0,1\}^n$ if $\Pr_{x \sim \mathcal{D}}[P(x) \neq F(x)] \leq \varepsilon$. We typically think of $\varepsilon > 0$ as a small constant, though

33rd Int'l Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2013).
Editors: Anil Seth and Nisheeth K. Vishnoi; pp. 201–212

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

the regime of ε close to $1/2$ is also interesting and is dealt with later on.

A seminal result of Razborov [22] showed, in the case $\mathbb{F} = \mathbb{F}_2$, that a certain symmetric function $F : \{0,1\}^n \to \{0,1\}$ has no ε-approximating polynomial of degree $o(\sqrt{n})$ over the uniform distribution $\mathcal{U}$ for some constant $\varepsilon > 0$; this fact was used to prove a lower bound for $\mathrm{ACC}^0[2]$ circuits computing the Majority function. Smolensky [24] further showed that for constant primes $p \neq q$ and $\mathbb{F} = \mathbb{F}_p$, the function $\mathrm{MOD}_q : \{0,1\}^n \to \{0,1\}$ – which accepts its input x iff $\sum_i x_i \not\equiv 0 \pmod q$ – does not have a $((1/2q) - \varepsilon)$-approximating polynomial over $\mathcal{U}$ of degree $o(\sqrt{n})$. In another work [25], Smolensky also showed a similar degree lower bounds for polynomials that $((1/2) - \varepsilon)$-approximate the Majority function over $\mathcal{U}$ for *any* field $\mathbb{F}$. All these lower bounds are known to be tight under $\mathcal{U}$ [6, 27]. Moreover, as far as we are aware, these remain the best degree lower bounds that we know of today for polynomials approximating *explicit functions*[1], even if we allow the probability distribution $\mathcal{D}$ to be arbitrary (even non-explicit).

Therefore, we ask if Smolensky's lower bound can be strengthened considerably, even just in this weaker regime. More formally,

▶ Problem 1. Come up with an explicit function $F : \{0,1\}^n \to \{0,1\}$ that has no $(1/10)$-approximating polynomials of degree less than $n^{\frac{1}{2}+\Omega(1)}$ under *some* distribution $\mathcal{D}$.

The reasons for considering a lower bound under an arbitrary (as opposed to the uniform) distribution are twofold: the first is that the circuit lower bound results mentioned above [22, 24] and other applications we will see later on only require this weaker lower bound result; and secondly, the above lower bound notion, by LP duality, has a nice dual *upper bound* notion as well. More precisely, if a Boolean function F has ε-approximating polynomials of degree d under *every* distribution $\mathcal{D}$, then there is a probability distribution over *polynomials* of degree d that computes F correctly on *every input* with probability $(1 - \varepsilon)$. Such a distribution is called a *probabilistic polynomial* [26]. Probabilistic polynomials for a function F are much easier to reason with than polynomials that approximate function F w.r.t. a fixed distribution; this is the same kind of distinction that exists between, say, randomized and distributional algorithms. Combination, composition, error-reduction, etc. are much easier with probabilistic polynomials than with their distributional counterparts.

Results. We begin our search for a suitable lower bound candidate $F : \{0,1\}^n \to \{0,1\}$ with the class of *symmetric* functions. Recall that $F : \{0,1\}^n \to \{0,1\}$ is symmetric if $F(x)$ depends only on the Hamming weight of x, denoted $|x|$. The degree lower bounds mentioned above [22, 24, 25] were all proved for symmetric functions over the uniform distribution. A natural question to ask is if we can improve these lower bounds for functions in this class.

It has been observed [6, 27] in the literature that as long as we only consider the uniform distribution, the lower bounds of Razborov and Smolensky are indeed tight. In other words, for any constant $\varepsilon \in (0, 1/2)$, any symmetric function $F : \{0,1\}^n \to \{0,1\}$ has an ε-approximating polynomial of degree $O(\sqrt{n})$ under the uniform distribution. The way to prove this is to note that the uniform distribution puts all but an ε-fraction of its mass on the central $O(\sqrt{n})$ layers of the hypercube $\{0,1\}^n$. Thus, by interpolating a polynomial of degree $O(\sqrt{n})$ that computes F exactly on these points of the hypercube (possible since F is symmetric), we obtain an $O(\sqrt{n})$ degree polynomial that ε-approximates F.

However, the strategy used above is crucially dependent on the underlying distribution being uniform (or at least having a nice product structure) and thus it is conceivable that

[1] Throughout, we are not very formal about what we mean by an *explicit* function, since any reasonable notion will do. The reader may take it to mean a function from a function family computable in polynomial time.

under other distributions, symmetric functions might not have such low-degree polynomial approximations. Somewhat surprisingly, this turns out to be false. We show that symmetric functions in fact have small error *probabilistic polynomials* of degree $O(\sqrt{n} \cdot \mathrm{polylog}(n))$ and hence they have small error approximations of the same degree over any distribution $\mathcal{D}$.

We describe a simple case of the proof here for intuition. Assume that F is the MOD_2 function and that the underlying field is $\mathbb{R}$. As we outlined above, F has an ε-approximation $P \in \mathbb{R}[X_1, \ldots, X_n]$ of degree $O(\sqrt{n})$ under the uniform distribution. We obtain an ε-error probabilistic polynomial for F as follows. Suppose $x \in \{0,1\}^n$ is an arbitrary input. Note that for any $y \in \{0,1\}^n$, we can write $\mathrm{MOD}_2(x) = \mathrm{MOD}_2(x \oplus y) \bigoplus \mathrm{MOD}_2(y)$, where $x \oplus y \in \{0,1\}^n$ is obtained by taking the bitwise sum of x and y modulo 2. Now say $\mathbf{y} \in \{0,1\}^n$ is chosen uniformly at random; we have $\mathrm{MOD}_2(x) = \mathrm{MOD}_2(x \oplus \mathbf{y}) \bigoplus \mathrm{MOD}_2(\mathbf{y})$. Since $\mathbf{y}$ is uniformly random, we see that $x \oplus \mathbf{y}$ is also uniformly distributed over $\{0,1\}^n$. Hence, we have $P(x \bigoplus \mathbf{y}) = \mathrm{MOD}_2(x \oplus \mathbf{y})$ with probability at least $1 - \varepsilon$. Thus, the probabilistic polynomial $\mathbf{Q}(x) := P(x \oplus \mathbf{y}) \bigoplus \mathrm{MOD}_2(\mathbf{y})$ computes $\mathrm{MOD}_2(x)$ correctly with probability at least $1 - \varepsilon$. It is easily argued that the degree of $\mathbf{Q}$ for every fixed value of $\mathbf{y}$ is at most the degree of P and this concludes the proof.

It turns out that the above idea, suitably modified, also gives low-degree polynomials for MOD_q as long as q is not too large. After having done this, we can handle the case of general symmetric functions by using the Chinese Remainder Theorem (Theorem 15), which tells us that the Hamming weight $|x| \in \{0, \ldots, n\}$ is completely determined by the values $|x| \pmod q$, where q ranges over all "small" primes. This allows us to obtain $O(\sqrt{n} \cdot \mathrm{poly}(\log n))$-degree probabilistic polynomials for all symmetric functions using the probabilistic polynomials we construct for MOD_q. In fact, we obtain a more general result: we get probabilistic polynomials of low degree for all functions that are determined by "small-weight" sums of the bits of the input x. This result is proved in Section 3.2.

In Section 3.3, we extend these results to certain functions that depend on "large weight" sums of the bits of x. The class of functions that we consider are the Linear Threshold functions that have received quite some attention in the literature [14, 15, 17]. Using the above result in conjunction with ideas due to Hofmeister [17], we show that any Linear Threshold function of n Boolean variables has a probabilistic polynomial of degree $O(\sqrt{n} \cdot \mathrm{poly}(\log n))$.

The above results show that probabilistic polynomials of degree $n^{1/2+o(1)}$ are a surprisingly powerful and robust model of computation. This suggests that solving Problem 1 might be a significant challenge. On the other hand, however, these results also imply that solving Problem 1 would have significant rewards. In Section 4, we show that an explicit function F as above would have applications to many problems including lower bounds for compression by constant-depth circuits [11], lower bounds for some strong constant-depth circuit classes, improved correlation lower bounds against low-degree polynomials over small fields [27], and to showing that the Inner Product function does not have small $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ circuits [23] (if the function F has some additional "nice" properties).

Due to lack of space, many proofs are postponed to the full version of the paper.

2 Definitions and preliminaries

Let $\mathbb{F}$ be an arbitrary field and $n \in \mathbb{N}$ be a growing parameter. We denote by $\mathcal{B}_n$ the set of all functions $f : \{0,1\}^n \to \{0,1\}$.

▶ **Definition 2.** A function $f \in \mathcal{B}_n$ is said to be *symmetric* if there exists a function $h : \{0, \ldots, n\} \to \{0,1\}$ such that for any $x \in \{0,1\}^n$, we have $f(x_1, \ldots, x_n) = h\left(\sum_{i=1}^n x_i\right)$.

That is, the value of $f(x)$ is determined by the Hamming weight of x, denoted $|x|$.

More generally, given $W \in \mathbb{N}$, we say that f is *W-sum determined* if there exist $w_1, w_2, \ldots, w_n \in \mathbb{N}$ such that $\sum_i w_i \leq W$ and a function $h : \{0, \ldots, W\} \to \{0,1\}$ such that for any $x \in \{0,1\}^n$, we have $f(x_1, \ldots, x_n) = h\left(\sum_{i=1}^n w_i x_i\right)$. Note that symmetric functions in $\mathcal{B}_n$ are n-sum determined. It can also be seen that *every* $f \in \mathcal{B}_n$ is $(2^n - 1)$-sum determined.

Some examples of symmetric functions in $\mathcal{B}_n$ are:

- the Majority function MAJ^n, which accepts inputs x such that $|x| > n/2$,
- the $\mathrm{MOD}^n_{m,r}$ function for $m \geq 2$ and $r \in \mathbb{N}$, which accepts x such that $|x| \equiv r \pmod m$.

We omit the superscript n above when it is clear from context.

▶ **Definition 3.** A function $f \in \mathcal{B}_n$ is said to be a *Linear Threshold function* if there exist $a_1, \ldots, a_n, \theta \in \mathbb{R}$ such that for any $x \in \{0,1\}^n$ $f(x) = 1$ iff $\sum_{i=1}^n a_i x_i \geq \theta$.

Notation. We denote by $\mathrm{SUM}^n_W, \mathrm{THR}^n$ the set of all W-sum determined functions and linear threshold functions in $\mathcal{B}_n$ respectively. When n is clear from context, we use $\mathrm{SUM}_W, \mathrm{THR}$ instead of $\mathrm{SUM}^n_W, \mathrm{THR}^n$.

▶ **Definition 4.** Let $n, d \in \mathbb{N}$ and $\varepsilon \in (0, \frac{1}{2})$. Fix a probability distribution $\mathcal{D}$ over $\{0,1\}^n$. We say that a function $f \in \mathcal{B}_n$ is *$(d, \varepsilon, \mathcal{D})_{\mathbb{F}}$-approximable* if there is a polynomial $P \in \mathbb{F}[X_1, \ldots, X_n]$ of degree at most d such that $\Pr_{x \sim \mathcal{D}}[P(x) \neq f(x)] \leq \varepsilon$.

Smolensky [24] proved the following degree lower bounds on ε-approximating polynomials for the $\mathrm{MOD}^n_{q,0}$ function, where q is a constant prime.

▶ **Theorem 5** ([24]). *For any $\varepsilon < 1/10q$, any polynomial that ε-approximates the $\mathrm{MOD}^n_{q,0}$ function w.r.t. the uniform distribution over $\{0,1\}^n$ has degree $\Omega(\sqrt{n \log(1/\varepsilon)})$.*

▶ **Definition 6** (Probabilistic polynomial [26])**.** A *probabilistic polynomial* of degree d in $\mathbb{F}[X_1, \ldots, X_n]$ is a probability distribution $\mathbf{P}$ over polynomials of degree at most d in $\mathbb{F}[X_1, \ldots, X_n]$. Given $f \in \mathcal{B}_n$ and an $\varepsilon \in (0, \frac{1}{2})$, we say that $\mathbf{P}$ is an *ε-error probabilistic polynomial* for f if $\forall x \in \{0,1\}^n$, we have $\Pr_{\mathbf{P}}[\mathbf{P}(x) \neq f(x)] \leq \varepsilon$. The *$\varepsilon$-error probabilistic degree of f* over $\mathbb{F}$ is the least d s.t. f has an ε-error probabilistic polynomial of degree d.

The following simple facts will be useful. The first is an easy consequence of LP duality and the second follows, for example, from polynomial interpolation or Möbius Inversion.

▶ Fact 7. Fix any $\varepsilon \in (0, \frac{1}{2})$, $n \in \mathbb{N}$, and any field $\mathbb{F}$. Then, for any $f \in \mathcal{B}_n$, f has ε-error probabilistic degree d iff f is $(d, \varepsilon, \mathcal{D})_{\mathbb{F}}$-approximable for *every* probability distribution $\mathcal{D}$.

▶ Fact 8. Every $f \in \mathcal{B}_n$ can be *exactly* represented by a polynomial in $\mathbb{F}[X_1, \ldots, X_n]$ of degree at most n. That is, there is a polynomial $P \in \mathbb{F}[X_1, \ldots, X_n]$ of degree at most n such that for any $x \in \{0,1\}^n$, we have $P(x) = f(x)$.

▶ Fact 9. Assume $f \in \mathcal{B}_s$ and $g_i \in \mathcal{B}_t$ $(i \in [s])$. Define $F \in \mathcal{B}_t$ as follows: $F(x) := f(g_1(x), \ldots, g_s(x))$. Say $\mathbf{P}$ is an ε-error probabilistic polynomial of degree d for f and $\mathbf{Q}_i$ $(i \in [s])$ is a δ_i-error probabilistic polynomial of degree d_i for g_i. Then, $\mathbf{P}(\mathbf{Q}_1, \ldots, \mathbf{Q}_s)$ is an $(\varepsilon + \sum_{i=1}^s \delta_i)$-error probabilistic polynomial for F of degree $d \cdot (\max_{i \in [s]} d_i)$.

We also recall here the definitions of some well-known constant-depth circuit classes. Throughout, the size of a circuit is the number of *wires* in the circuit. The class of polynomial sized constant-depth circuits made up of AND, OR, and NOT gates is called AC^0; for p a *constant* prime, $\mathrm{ACC}^0[p]$ is the class of polynomial sized constant depth circuits containing AND, OR, NOT and MOD_p gates (a MOD_p gate computes a $\mathrm{MOD}_{p,r}$ function applied to

its input bits, for some r). We will abuse notation and use "AC^0 circuits of size s" to refer to constant-depth circuits of size s with AND, OR, NOT gates (similarly, we will also say "$\mathrm{ACC}^0[p]$ circuits of size s").

The following is implied by works of Razborov [22], Smolensky [24], and Tarui [26].

▶ **Theorem 10** ([22, 24, 26]). *Let $s \in \mathbb{N}$ and $\varepsilon \in (0, 1/2)$ be arbitrary parameters. Any AC^0 circuit of size s and depth d has an ε-error probabilistic polynomial of degree $(\log s \cdot \log(1/\varepsilon))^{O(d)}$ over any field $\mathbb{F}$. Any $\mathrm{ACC}^0[p]$ circuit of size s and depth d has an ε-error probabilistic polynomial of degree $(\log s \cdot \log(1/\varepsilon))^{O(d)}$ over the field $\mathbb{F}_p$.*

We also consider constant-depth circuits with other gate types: a SYM gate can compute an arbitrary symmetric function of its inputs, a THR gate can compute an arbitrary linear threshold function, and a SYMTHR gate can do either. For a gate type $\mathcal{G}$ (such as SYM, THR, MOD_p, etc.), we use $\mathrm{AC}^0 \circ \mathcal{G}$ to denote the class of polynomial constant-depth circuits where the inputs feed into gates of type $\mathcal{G}$, which in turn feed into an AC^0 circuit.

3 Probabilistic polynomials for SUM_W and THR

Let $\mathbb{F}$ be an arbitrary field and $n \in \mathbb{N}$ a growing parameter. In this section, we prove the following theorems.

▶ **Theorem 11.** *Fix $\varepsilon \in (0, \frac{1}{2})$ and $W \in \mathbb{N}$. Any $f \in \mathrm{SUM}_W^n$ has ε-error probabilistic degree at most $O(\sqrt{n \log(1/\varepsilon)} \cdot (\log W)^{O(1)})$.*

▶ **Theorem 12.** *Fix $\varepsilon \in (0, \frac{1}{2})$. Any $f \in \mathrm{THR}^n$ has ε-error probabilistic degree at most $O(\sqrt{n} \cdot (\log n \log\left(\frac{1}{\varepsilon}\right))^{O(1)})$.*

We will prove these two theorems in three steps:

- We will first show that certain functions in $\mathcal{B}_n$ (variants of the modular functions $\mathrm{MOD}_{m,r}$) have ε-error probabilistic degree $O(\sqrt{n \log(1/\varepsilon)} \cdot (\log n)^{O(1)})$. This will establish a weak form of Theorem 11.
- Using some Chinese Remaindering ideas, we will then derive Theorem 11.
- Using some more Chinese Remaindering ideas due to Hofmeister [17] and Theorem 11, we will prove Theorem 12.

We first introduce a generalization of the symmetric $\mathrm{MOD}_{m,r}^n$ function. Given $w \in \mathbb{Z}_m^n$, we denote by $\mathrm{MOD}_{m,r,w}^n$ (or just $\mathrm{MOD}_{m,r,w}$ if n is clear from context) the Boolean function in $\mathcal{B}_n$ that accepts its input $x \in \{0,1\}^n$ iff $\sum_{i=1}^n w_i x_i \equiv r \pmod m$. We identify $\mathbb{Z}_m$ with $\{0, 1, \ldots, m-1\}$.

3.1 Probabilistic polynomials for $\mathrm{MOD}_{m,r,w}^n$

The main result of this section is the following.

▶ **Lemma 13.** *Fix integers $m \geq 2$ and $r \in \{0, 1, \ldots, m-1\}$. For any $\varepsilon \in (0, \frac{1}{2})$ and $w \in \mathbb{Z}_m^n$, the function $\mathrm{MOD}_{m,r,w}^n$ has an ε-error probabilistic degree $O(\sqrt{n \log(m/\varepsilon)} \cdot m)$.*

In particular for $m = (\log n)^{O(1)}$, the probabilistic degree is $O(\sqrt{n \log(1/\varepsilon)} \cdot (\log n)^{O(1)})$.

We first need a lemma regarding the approximability of symmetric functions in $\mathcal{B}_n$ with respect to some simple product distributions over $\{0,1\}^n$. For $\rho \in [0,1]$, let $\mathcal{D}_\rho$ denote the distribution over $\{0,1\}^n$ where each bit is set to 1 independently with probability ρ.

The following lemma is a slight extension of one from a survey of Viola [27] – where the proof is attributed to Avi Wigderson – based on a lemma due to Bhatnagar, Gopalan,

and Lipton [6]. Though the proof in [27] only works in finite characteristic, the lemma can actually be proved for any field $\mathbb{F}$. The proof is postponed to the full version of the paper.

▶ **Lemma 14.** *Let $f \in \mathcal{B}_n$ be an arbitrary symmetric function, $\rho \in [0,1]$, and $\varepsilon \in (0, \frac{1}{2})$. Then, f is $(O(\sqrt{n \log(1/\varepsilon)}), \varepsilon, \mathcal{D}_\rho)_{\mathbb{F}}$-approximable.*

Proof of Lemma 13. The idea behind the proof is to exploit the *random self reducibility* of $\mathrm{MOD}_{m,r,w}$. Let $x \in \{0,1\}^n$ be *arbitrary* and $w \circ x \in \mathbb{Z}_m^n$ be the vector whose i-th entry is $w_i x_i$. For any $y \in \mathbb{Z}_m^n$ and $t \in \mathbb{Z}_m$, we define $S_{y,t}(x) \in \{0,1\}^n$ to be such that the ith bit of $S_{y,t}(x)$ is 1 iff the i-th entry of the vector $(w \circ x) + y$ is congruent to t modulo m; that is, $w_i x_i + y_i \equiv t \pmod m$. The basic observation we will use is that to check if $\sum_i w_i x_i \equiv r \pmod m$, it suffices to check that $\sum_i (w_i x_i + y_i) \equiv r + \sum_i y_i \pmod m$, which in turn reduces to computing $\mathrm{MOD}_{m,r'}(S_{y,t}(x))$ for various r', t.

Formally, note that for *any* $y \in \mathbb{Z}_m^n$, we have

$$\mathrm{MOD}_{m,r,w}(x) = \sum_{(r_1,\ldots,r_{m-1}) \in T} \bigwedge_{t=1}^{m-1} \mathrm{MOD}_{m,r_t}(S_{y,t}(x)) \tag{1}$$

where $T = \left\{ (r_1, \ldots, r_{m-1}) \in \mathbb{Z}_m^{m-1} \;\middle|\; \sum_j j \cdot r_j \equiv r + \sum_i y_i \pmod m \right\}$. (Note that the function on the left hand side above is $\mathrm{MOD}_{m,r,w}$, where as on the right hand side, we have the symmetric functions $\mathrm{MOD}_{m,r'}$ for various r'.) For any fixed y, the i-th bit of $S_{y,t}(x)$ depends only on x_i and by Fact 8, can be represented exactly by a degree 1 polynomial in x_i. Now, for a uniformly random $\mathbf{y} \in \mathbb{Z}_m^n$, $(w \circ x) + \mathbf{y}$ is a uniformly random element of $\mathbb{Z}_m^n$. Hence, $S_{\mathbf{y},t}(x)$ is a random element from $\{0,1\}^n$ chosen according to the distribution $\mathcal{D}_{\frac{1}{m}}$ and its individual bits are degree 1 *probabilistic* polynomials in x.

Let $\delta \in (0, \frac{1}{2})$ be a parameter whose value we will fix later. Since the function $\mathrm{MOD}_{m,r'}$ is symmetric, by Lemma 14, we know that for all m, r' there is a polynomial $P_{m,r'}$ of degree $O(\sqrt{n \log(1/\delta)})$ such that $\Pr_{\mathbf{x} \sim \mathcal{D}_{\frac{1}{m}}}[P_{m,r'}(\mathbf{x}) \neq \mathrm{MOD}_{m,r'}(\mathbf{x})] \leq \delta$. Consider the *probabilistic* polynomial $\mathbf{Q}_{m,r,w}(X_1, \ldots, X_n)$ defined (based on the choice of $\mathbf{y} \in \mathbb{Z}_m^n$) as follows:

$$\mathbf{Q}_{m,r,w}(X_1, \ldots, X_n) = \sum_{(r_1,\ldots,r_{m-1}) \in T} \prod_{t=1}^{m-1} P_{m,r_t}(S_{\mathbf{y},t}(X_1, \ldots, X_n))$$

.

Observe that the degree of $\mathbf{Q}_{m,r,w}$ is $O(m\sqrt{n \log(1/\delta)})$.

We claim that for suitably small δ, $\mathbf{Q}_{m,r,w}$ is an ε-error probabilistic polynomial for $\mathrm{MOD}_{m,r,w}$. To see this, note that if $\mathbf{y} \in \mathbb{Z}_m^n$ satisfies $P_{m,r'}(S_{\mathbf{y},t}(x)) = \mathrm{MOD}_{m,r'}(S_{\mathbf{y},t}(x))$ for all choices of r' and t, then by (1), we have $\mathbf{Q}_{m,r,w}(x) = \mathrm{MOD}_{m,r,w}(x)$. Thus, we have

$$\Pr_{\mathbf{Q}}[Q_{m,r,w}(x) \neq \mathrm{MOD}_{m,r,w}(x)] \leq \Pr_{\mathbf{y}}[\exists r', t \; P_{m,r'}(S_{\mathbf{y},t}(x)) \neq \mathrm{MOD}_{m,r'}(S_{\mathbf{y},t}(x))]$$

For any fixed r', t, by our choice of $P_{m,r'}$, the probability that $P_{m,r'}(S_{\mathbf{y},t}(x))$ does not equal $\mathrm{MOD}_{m,r'}(S_{\mathbf{y},t}(x))$ is at most δ. By a union bound over all t, r' and using the inequality above, we have $\Pr_{\mathbf{Q}}[Q_{m,r,w}(x) \neq \mathrm{MOD}_{m,r,w}(x)] \leq m^2 \cdot \delta$. Setting $\delta = \varepsilon/m^2$, we get that $\mathbf{Q}_{m,r,w}$ is an ε-error probabilistic polynomial for $\mathrm{MOD}_{m,r,w}$. The degree of $\mathbf{Q}_{m,r,w}$ is $O(m\sqrt{n \log(1/\delta)}) = O(\sqrt{n \log(m/\varepsilon)} \cdot m)$. ◀

3.2 Proof of Theorem 11

By the definition of SUM_W, there exist $w_1, \ldots, w_n \in \mathbb{N}$ such that $\sum_{i=1}^n w_i \leq W$ and for any $x \in \{0,1\}^n$, the output $f(x)$ is determined by the value $\sum_{i=1}^n w_i x_i \in \{0, 1, \ldots, W\}$.

We reduce the problem of constructing small error probabilistic polynomials for f to that of constructing small error probabilistic polynomials for $\mathrm{MOD}_{m,r,\overline{w}}$ (for suitable $m, \overline{w}$) by using the Chinese Remainder Theorem (see, e.g., [13]), a special case of which follows.

▶ **Theorem 15.** *Fix a positive $W \in \mathbb{N}$ and any distinct primes $p_1, \ldots, p_\ell$ ($\ell \geq 1$) such that $\prod_{j=1}^{\ell} p_j > 2W$. Then, given any congruence classes $a_j \in \{0, \ldots, p_j - 1\}$ ($j \in [\ell]$) modulo these primes, there is at most one $k \in \{-W, -(W-1), \ldots, W\}$ s.t. $k \equiv a_j \pmod{p_j}$ for each $j \in [\ell]$.*

Let $\ell = \lceil \log W \rceil + 2$ and $p_1 \leq \cdots \leq p_\ell$ be the first ℓ distinct primes. By the Prime Number Theorem [13], $\max_{j \in [\ell]} p_j = p_\ell = O(\log W \log\log W)$. Since $\prod_{j=1}^{\ell} p_j \geq 2^\ell > 2W$, we see using Theorem 15 that each integer $k \in \{0, \ldots, W\}$ is uniquely determined (among the integers $\{0, \ldots, W\}$) by its congruence classes modulo each of the p_j ($j \in [\ell]$); that is, by the tuple $(k \pmod{p_j})_{j \in [\ell]}$. In particular, for any $x \in \{0,1\}^n$, the integer $\sum_{i=1}^{n} w_i x_i$ is uniquely determined by $((\sum_i w_i x_i) \pmod{p_j})_{j \in [\ell]}$. Moreover, note that for any $j \in [\ell]$, the congruence class of $\sum_i w_i x_i$ modulo p_j is determined uniquely by the tuple $(\mathrm{MOD}_{p_j,r,\overline{w}^j}(x))_{r \in \mathbb{Z}_{p_j}}$, where $\overline{w}^j$ is obtained from w by dropping each of its entries modulo p_j. Hence, for any $x \in \{0,1\}^n$, the value of $f(x)$ is determined by the tuple $(\mathrm{MOD}_{p_j,r,\overline{w}^j}(x))_{j \in [\ell], r \in \mathbb{Z}_{p_j}}$.

We summarize the discussion above in the form of the following claim. Let $s = \sum_{j=1}^{\ell} p_j$.

▶ Claim 16. Let $f, w_1, \ldots, w_n, p_1, \ldots, p_\ell$ be as above. Then, there is a function $g \in \mathcal{B}_s$ such that for all $x \in \{0,1\}^n$, we have $f(x) = g\left(\mathrm{MOD}_{p_j,r,\overline{w}^j}(x) : j \in [\ell], r \in \mathbb{Z}_{p_j}\right)$.

Let $\delta \in (0, \frac{1}{2})$ be a parameter that we will fix below. Now, by Lemma 13, we know that for each $j \in [\ell]$ and $r \in \mathbb{Z}_{p_j}$, the function $\mathrm{MOD}_{p_j,r,\overline{w}^j}$ has a δ-approximating probabilistic polynomial of degree $O(p_j \sqrt{n \log(p_j/\delta)}) = O(\log W \log\log W \sqrt{n \log(\log W/\delta)}) = O(\log^2 W \sqrt{n \log(1/\delta)})$. Moreover, by Fact 8, the function g from Claim 16 can be represented *exactly* by a polynomial of degree $s = \sum_j p_j = O(\log^3 W)$. Thus, by Fact 9, we see that f has a $((\sum_j p_j) \cdot \delta)$-error probabilistic polynomial of degree $O((\log W)^5 \sqrt{n \log(1/\delta)})$ for any $\delta \in (0, 1/2)$.

Since $\sum_j p_j = O(\log^3 W)$, we may set $\delta = \varepsilon / C \log^3 W$ for some large absolute constant C and thus obtain an ε-error probabilistic polynomial for f. The degree of this polynomial is $(\log W)^{O(1)} \sqrt{n \log(1/\varepsilon)}$. This completes the proof of Theorem 11.

3.3 Proof of Theorem 12

In this section, we construct probabilistic polynomials for an arbitrary $f \in \mathrm{THR}^n$. We use many ideas and observations of Hofmeister [17] (see also [14, 15]) regarding the structure of Linear Threshold functions.

Recall that given $f \in \mathrm{THR}^n$, there exist $a_1, \ldots, a_n, \theta \in \mathbb{R}$ such that for all $x \in \{0,1\}^n$, we have $f(x) = 1$ iff $\sum_{i=1}^{n} a_i x_i - \theta \geq 0$. Moreover, it is known by a result of Muroga [21] that we may in fact choose $a_1, \ldots, a_n$ and θ to be integers of magnitude at most $M = 2^{O(n \log n)}$. We fix such integers $a_1, \ldots, a_n, \theta$. Let N denote $M(n+1)$. Let $I = \{i \in \mathbb{Z} \mid -(n+1) \leq i \leq n+1\}$ and $I^{\geq 0}$ be the non-negative members of I.

Similar to [17], we define the following.

- For $j \in \{0, \ldots, \ell = \lceil \log N \rceil\}$, we define integers $a_i^{(j)}$ ($i \in [n]$) and $\theta^{(j)}$ as follows. We define $\theta^{(0)} = \theta$ and $a_i^{(0)} = a_i$ for $i \in [n]$. For $j \in [\ell]$, we define $\theta^{(j)} = \mathrm{trunc}(\theta^{(j-1)}/2)$ and $a_i^{(j)} = \mathrm{trunc}(a_i^{(j-1)}/2)$ for $i \in [n]$. Here, $\mathrm{trunc}(z) = \lfloor z \rfloor$ for $z \geq 0$ and $\lceil z \rceil$ for $z < 0$.
- For $j \in \{0, \ldots, \ell\}$ and any prime $p \in \mathbb{N}$, define

- $\mathrm{INS}_j \in \mathcal{B}_n$ so that $\mathrm{INS}_j(x) = 1$ iff $\sum_{i=1}^n a_i^{(j)} x_i - \theta^{(j)} \in I$.
- $\mathrm{INS}_j^p \in \mathcal{B}_n$ so that $\mathrm{INS}_j^p(x) = 1$ iff $\exists k \in I$ such that $\sum_{i=1}^n a_i^{(j)} x_i - \theta^{(j)} \equiv k \pmod p$.
- $\mathrm{POS}_j \in \mathcal{B}_n$ so that $\mathrm{POS}_j(x) = 1$ iff $\sum_{i=1}^n a_i^{(j)} x_i - \theta^{(j)} \in I^{\geq 0}$.
- $\mathrm{POS}_j^p \in \mathcal{B}_n$ so that $\mathrm{POS}_j^p(x) = 1$ iff $\exists k \in I^{\geq 0}$ such that $\sum_{i=1}^n a_i^{(j)} x_i - \theta^{(j)} \equiv k \pmod p$.

The following is implicit in [17]. We omit the proof.

▶ **Lemma 17.** *Let $f, \mathrm{POS}_j, \mathrm{INS}_j$ ($j \in \{0, \ldots, \ell\}$) be as defined above. Then, for any $x \in \{0,1\}^n$, $f(x) = \mathrm{POS}_0(x) + \sum_{j=1}^{\ell} \left(\overline{\mathrm{INS}_{j-1}(x)} \wedge \mathrm{POS}_j(x)\right)$, where the sum is taken over the integers (and hence, the equality also holds modulo the characteristic of $\mathbb{F}$).*

We construct a small error probabilistic polynomial for f by constructing small error probabilistic polynomials for the functions $\mathrm{INS}_j, \mathrm{POS}_j$ ($j \in \{0, \ldots, \ell\}$). The following lemma assures us that this is possible.

▶ **Lemma 18.** *For $j \in \{0, \ldots, \ell\}$ and any $\delta \in (0, 1/2)$, the functions INS_j and POS_j have δ-error probabilistic polynomials of degree $O(\sqrt{n} \cdot (\log n \log(1/\delta))^{O(1)})$.*

Assuming Lemma 18, we can prove Theorem 12 easily as follows. Fix $\delta = \varepsilon/(2\ell+1)$. Using Lemma 18, we have δ-error probabilistic polynomials $\mathbf{P}_j$ for POS_j and $\mathbf{Q}_j$ for INS_j of degree $O(\sqrt{n} \cdot (\log n \log(1/\varepsilon))^{O(1)})$. Consider the probabilistic polynomial $\mathbf{P} = \mathbf{P}_0 + \sum_{j=1}^{\ell} (1 - \mathbf{Q}_{j-1}) \cdot \mathbf{P}_j$. By Lemma 17, for any $x \in \{0,1\}^n$, we have $\mathbf{P}(x) = f(x)$ unless there is a $j \in \{0, \ldots, \ell\}$ such that $\mathbf{P}_j(x) \neq \mathrm{POS}_j(x)$ or a $j \in [\ell]$ such that $\mathbf{Q}_j(x) \neq \mathrm{INS}_j(x)$. By a union bound, the probability that this occurs is at most $(2\ell+1) \cdot \delta = \varepsilon$. Hence, $\mathbf{P}$ is an ε-error probabilistic polynomial for f of degree $O(\sqrt{n} \cdot (\log n \log(1/\varepsilon))^{O(1)})$. This finishes the proof of Theorem 12.

Proof of Lemma 18. Fix a $j \in \{0, \ldots, \ell\}$. We prove the claim for INS_j only. The proof for POS_j is very similar with $I^{\geq 0}$ replacing I throughout.

Recall that $\mathrm{INS}_j(x) = 1$ iff $\sum_{i=1}^n a_i^{(j)} x_i - \theta^{(j)} \in I$. Hence, $\mathrm{INS}_j(x)$ depends only on the sum $\sum_i a_i^{(j)} x_i$ and this might indicate that Theorem 11, that constructs probabilistic polynomials for functions in SUM_W, might be useful. However, the value of W here is too large: it may be as large as $M \cdot (n+1) = 2^{\Omega(n \log n)}$. We therefore first reduce the problem to the case of small-weight sums by going modulo small primes. This idea has been used before by Hofmeister [17], albeit for a different purpose from ours.

Consider any prime $p \in \mathbb{N}$. Note that $\mathrm{INS}_j^p(x)$ is a function only of $\sum_i b_i x_i$ where $b_i \in \{0, \ldots, p-1\}$ is chosen so that $b_i \equiv a_i^{(j)} \pmod p$. Thus, $\mathrm{INS}_j^p \in \mathrm{SUM}_W$ where $W = O(pn)$. Moreover, observe that:

- If $\mathrm{INS}_j(x) = 1$, then $\mathrm{INS}_j^p(x) = 1$ for any prime p.
- If $\mathrm{INS}_j(x) = 0$, then though $\mathrm{INS}_j^p(x)$ could be 1 for some values of p, this does not happen too often. The following makes this precise:

  ▶ Claim 19. If $\mathrm{INS}_j(x) = 0$, then $\left| \left\{ p \in \mathbb{N} \,\middle|\, p \text{ prime and } \mathrm{INS}_j^p(x) = 1 \right\} \right| \leq \log((2N)^{2n+3}) \leq Cn^2 \log n$ for some constant $C > 0$.

  Proof. Since $\mathrm{INS}_j(x) = 0$, the integer $K := \sum_i a_i^{(j)} x_i - \theta^{(j)}$ does not belong to I. Also, we know that $|K| \leq M(n+1) = N$. Now, let $S_x = \left\{ p \in \mathbb{N} \,\middle|\, p \text{ prime and } \mathrm{INS}_j^p(x) = 1 \right\}$. By the definition of S_x, $p \in S_x$ iff there is some $k \in I$ such that $K \equiv k \pmod p$ or in other words, $p | (K - k)$. Thus, every $p \in S_x$ divides $K - k$ for some $k \in I$ and hence the product $\prod_{k \in I}(K-k)$. Therefore, $|S_x| \leq \log(\prod_k |K-k|) \leq \log((2N)^{2n+3})$ as claimed. ◀

Now fix $r = \lceil (2Cn^2 \log n)/\delta \rceil$ where C is the constant from Claim 19. Let $p_1, \ldots, p_r$ be the first r primes. By our reasoning above, for any $x \in \{0,1\}^n$, we have

- $\mathrm{INS}_j(x) = 1 \Rightarrow \mathrm{INS}_j^{p_k}(x) = 1 \quad \forall k \in [r]$,
- $\mathrm{INS}_j(x) = 0 \Rightarrow \Pr_{k \in [r]}[\mathrm{INS}_j^{p_k}(x) = 1] \leq \frac{Cn^2 \log n}{r} \leq \frac{\delta}{2}$

We claim that the functions $\mathrm{INS}_j^{p_k}$ ($k \in [r]$) have probabilistic polynomials of low degree. As we argued above, the function $\mathrm{INS}_j^{p_k}(x) \in \mathrm{SUM}_{W_k}$ for $W_k = O(p_k n)$. Moreover, by the Prime Number Theorem [13], we have $p_k \leq O(r \log r)$ for each $k \in [r]$. Hence, for each $k \in [r]$, the function $\mathrm{INS}_j^{p_k} \in \mathrm{SUM}_W$ for $W = O(nr \log r)$. Applying Theorem 11, we see that the function $\mathrm{INS}_j^{p_k}$ has a $(\delta/2)$-error probabilistic polynomial $\mathbf{Q}_k$ of degree $O(\sqrt{n \log(1/\delta)}(\log W)^{O(1)}) = \sqrt{n} \cdot (\log n \log(1/\delta))^{O(1)}$, where the final equality follows from our choice of r.

Now, consider the probabilistic polynomial $\mathbf{Q}$: that is, we first pick $\mathbf{k} \in [r]$ uniformly at random and then sample a polynomial from the distribution $\mathbf{Q_k}$. We claim that $\mathbf{Q}$ is a δ-error probabilistic polynomial for INS_j. To see this, note that for any $x \in \{0,1\}^n$, we must have $\mathbf{Q}(x) = \mathrm{INS}_j(x)$ unless one of the following two events occurs:

- $\mathrm{INS}_j^{p_\mathbf{k}}(x) \neq \mathrm{INS}_j(x)$. As we saw above, this happens with probability at most $\delta/2$.
- $\mathbf{Q_k}(x) \neq \mathrm{INS}_j^{p_\mathbf{k}}(x)$. By our choice of $\mathbf{Q_k}$, this happens with probability at most $\delta/2$.

Hence, the probability that $\mathbf{Q}(x) \neq \mathrm{INS}_j(x)$ is at most δ. Thus, $\mathbf{Q}$ is a δ-error probabilistic polynomial for INS_j of degree $\sqrt{n} \cdot (\log n \log(1/\delta))^{O(1)}$. ◀

4 Connections to other problems

1-round Compression by $\mathrm{ACC}^0[p]$ circuits. Motivated by applications in Cryptography, parameterized complexity, and PCPs, Chattopadhyay and Santhanam [11] study the problem of proving lower bounds for compression by constant-depth circuits. We briefly describe the setup here, referring to [11] for details.

We define a *compression game*, between two players Alice and Bob, as follows. Let $f \in \mathcal{B}_n$ be known to both players. Alice, a computationally bounded player, is given $x \in \{0,1\}^n$ and wishes to compute $f(x)$ with the aid of Bob, who is computationally unbounded. The aim is to minimize the amount of communication between the players. We consider the case when Alice's computational power is restricted to a class of constant-depth circuits $\mathcal{C}$, which we call a $\mathcal{C}$-compression game. We also consider the special case of 1-*round compression games*, where Alice sends a message to Bob based on the input x and Bob declares the value $f(x)$. Note that for any reasonable $\mathcal{C}$, there is always a 1-round communication protocol with communication n: Alice simply sends x to Bob, who then outputs $f(x)$.

In the case that $\mathcal{C} = \mathrm{AC}^0$, [11] showed a close to optimum communication lower bound of $n/(\log n)^{O(1)}$ for the MOD_2 function. In the case that $\mathcal{C} = \mathrm{ACC}^0[p]$ (p prime),using Theorem 5, they show that any 1-*round* compression protocol for $\mathrm{MOD}_{q,0}$ must involve $\Omega(\sqrt{n}/(\log n)^{O(1)})$ bits of communication. Using the same idea, we can show that stronger inapproximability results (w.r.t. *any* distribution) would imply stronger communication lower bounds for 1-round $\mathrm{ACC}^0[p]$-compression game. The proof is omitted.

▶ **Theorem 20.** *Assume $f \in \mathcal{B}_n$ has a 1-round $\mathrm{ACC}^0[p]$-compression protocol with communication c. Then, f has a $(1/10)$-error probabilistic polynomial over $\mathbb{F}_p$ of degree $O(c \cdot (\log n \log c)^{O(1)})$. In particular, if f has no $(1/10)$-error probabilistic polynomial of degree $n^{1/2+\varepsilon}$ over $\mathbb{F}_p$ for some $\varepsilon > 0$, then any 1-round $\mathrm{ACC}^0[p]$-compression protocol for f requires $\Omega(n^{1/2+\varepsilon}/(\log n)^{O(1)})$ bits of communication.*

Lower bounds for $\mathrm{AC}^0 \circ \mathrm{SYMTHR}$. There has been considerable work on proving lower bounds for AC^0 augmented with more powerful gates in some specific way, as a way of making progress towards proving lower bounds for stronger circuit classes, such as TC^0. Some of these works [16, 9, 20] have considered the model where the AC^0 circuits are allowed to have a "few" SYM, THR, or other gates. We consider a different variant where there are no bounds on the *number* of SYM and THR gates but on their *position* in the circuit: we require that these gates appear just above the input variables. As far as we know, lower bounds even for simple special cases of this model such as $\mathrm{OR} \circ \mathrm{AND} \circ \mathrm{SYM}$ and $\mathrm{OR} \circ \mathrm{AND} \circ \mathrm{THR}$ are unknown. In the case that the SYM gates at the bottom involve modular computation, some recent progress [12, 10] has been made. We can show that any function in the circuit class $\mathrm{AC}^0 \circ \mathrm{SYMTHR}$ has small error probabilistic polynomials of degree $O(\sqrt{n} \cdot \mathrm{polylog}(n))$, and hence proving that an explicit function f does not have probabilistic polynomials of this degree will prove an explicit lower bound for this circuit class. The proof is omitted.

▶ **Theorem 21.** *Let $F \in \mathcal{B}_n$ be computed by an $\mathrm{AC}^0 \circ \mathrm{SYMTHR}$ circuit of size s. Then, F has a $(1/10)$-error probabilistic polynomial of degree $O(\sqrt{n}(\log n \log s)^{O(1)})$ over any field $\mathbb{F}$. In particular, if F has no $(1/10)$-error probabilistic polynomial of degree less than $n^{1/2+\Omega(1)}$ over some field, then any $\mathrm{AC}^0 \circ \mathrm{SYMTHR}$ circuit for F must have size $2^{n^{\Omega(1)}}$.*

Correlation bounds for low-degree polynomials. Let $p \in \mathbb{N}$ be a constant prime. We consider the problem of proving *correlation bounds* over $\mathbb{F}_p$: i.e., showing that an explicit Boolean function $f \in \mathcal{B}_n$ is $(d, \delta, \mathcal{D})_{\mathbb{F}_p}$-inapproximable for δ close to $1/2$. For brevity, we say that $f \in \mathcal{B}_n$ is $(d, \varepsilon, \mathcal{D})_{\mathbb{F}_p}$-*correlated* (resp. uncorrelated) if it is $(d, \frac{1}{2} - \varepsilon, \mathcal{D})_{\mathbb{F}_p}$-approximable (resp. inapproximable). Showing strong correlation bounds against low-degree polynomials, (say, for $\varepsilon \ll 1/n$, $d = \mathrm{polylog}(n)$, and $\mathcal{D}$ the uniform distribution) would have applications to constructing PRGs for $\mathrm{ACC}^0[p]$ (see [27]).

Smolensky [25] showed that the MAJ^n function is $(d, O\left(\frac{d}{\sqrt{n}}\right), \mathcal{U})_{\mathbb{F}_p}$-uncorrelated for any $d \in \mathbb{N}$ where $\mathcal{U}$ denotes the uniform distribution. For $d \ll \log n$, this bound has been strengthened by results of Bourgain [7] and Viola and Wigderson [28]. However, when $d \geq \log n$, Smolensky's bound has not been improved for any probability distribution $\mathcal{D}$. We can show that proving probabilistic polynomial degree lower bounds of $n^{1/2+\Omega(1)}$ over $\mathbb{F}_p$ would improve on Smolensky's correlation lower bound for *some* (possibly non-explicit) distribution. The proof is omitted.

▶ **Theorem 22.** *Assume that for some $d \in \mathbb{N}$ and $\delta > 0$, the function $f \in \mathcal{B}_n$ is $(d, \delta, \mathcal{D})_{\mathbb{F}_p}$-correlated for* every *distribution $\mathcal{D}$ over $\{0,1\}^n$. Then, for any $\varepsilon \in (0, \frac{1}{2})$, f has an ε-error probabilistic polynomials of degree $\frac{d \cdot (\log(1/\delta) \log(1/\varepsilon))^{O(1)}}{\delta}$ over $\mathbb{F}_p$. In particular, if f has no $(1/10)$-error probabilistic polynomial of degree $n^{1/2+\eta}$ for some constant $\eta > 0$, than there is some distribution $\mathcal{D}$ such that f is $(d, O\left(\frac{d \cdot (\log n)^{O(1)}}{n^{1/2+\eta}}\right), \mathcal{D})_{\mathbb{F}_p}$-uncorrelated.*

Lower bounds for $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ circuits computing Inner Product. We define the Inner product function $\mathrm{IP}^n \in \mathcal{B}_{2n}$ as follows: $\mathrm{IP}^n(x_1, \ldots, x_n, y_1, \ldots, y_n) = \bigoplus_{i=1}^n x_i \wedge y_i$. By definition, IP^n has a depth-2 $\mathrm{ACC}^0[2]$ circuit of linear size, which is made up of $2n$ AND gates feeding into a $\mathrm{MOD}_{2,1}$ gate. Servedio and Viola [23] consider the question of whether IP^n can be computed by a polynomial sized $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ circuit. This question has relations to Matrix Rigidity (see [23]) and Communication complexity [2].

Here, we show the following: if there is a constant-degree polynomial from $\mathbb{F}_2[X_1, \ldots, X_n]$ that does not have a $(1/10)$-error probabilistic polynomial of degree $n^{1/2+\Omega(1)}$ over *some*

field $\mathbb{F}$, then IP^n does not have polynomial-sized $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ circuits. Note that by Lemma 13, any degree-1 polynomial over $\mathbb{F}_2$ has a (1/10)-error probabilistic polynomial of degree $O(n^{1/2})$ over any field $\mathbb{F}$. For polynomials of degree 2 and above, as far as we are aware, there are no such results known.

▶ **Theorem 23.** *If* IP^n *has an* $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ *circuit of size* s*, then any constant-degree polynomial* $P \in \mathbb{F}_2[X_1, \ldots, X_n]$ *has a* (1/10)*-error probabilistic polynomial* $\mathbf{Q}$ *of degree* $O(\sqrt{n} \cdot (\log n \log s)^{O(1)})$ *over* any *field. In particular, if there is a constant-degree polynomial* $P \in \mathbb{F}_2[X_1, \ldots, X_n]$ *that has no* (1/10)*-error probabilistic polynomial of degree less than* $n^{1/2+\Omega(1)}$ *over some field, then any* $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ *circuit for* IP^n *must have size* $2^{n^{\Omega(1)}}$.

The proof of the above is postponed to the full version of the paper. Note that as opposed to the explicit lower bounds required for the applications in previous sections, the above theorem says that a degree lower bound for probabilistic polynomials over some $\mathbb{F}$ even for a somewhat *non-explicit* function suffices to prove a lower bound for the explicit Inner Product function.

Acknowledgements. I would like to thank Arkadev Chattopadhyay, Parikshit Gopalan, Kristoffer Hansen, Prahladh Harsha, Swastik Kopparty, Nutan Limaye, and Emanuele Viola for their valuable feedback and encouragement. Arkadev Chattopadhyay also suggested the application to 1-round compression (Section 4) and I thank him for his permission to include it here. I am also grateful to the anonymous referees for their corrections and suggestions. Finally, I acknowledge the support of DST Inspire grant IFA12-ENG-14.

References

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

2 László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *FOCS*, pages 337–347. IEEE Computer Society, 1986.

3 Louay M. J. Bazzi. Polylogarithmic independence can fool dnf formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009.

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

5 Richard Beigel. The polynomial method in circuit complexity. In *In Proceedings of the 8th IEEE Structure in Complexity Theory Conference*, pages 82–95. IEEE Computer Society Press, 1995.

6 Nayantara Bhatnagar, Parikshit Gopalan, and Richard J. Lipton. Symmetric polynomials over z_m and simultaneous communication protocols. *J. Comput. Syst. Sci.*, 72(2):252–285, 2006.

7 Jean Bourgain. Estimation of certain exponential sums arising in complexity theory. *Comptes Rendus Mathematique*, 340(9):627 – 631, 2005.

8 Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *J. ACM*, 57(5), 2010.

9 Arkadev Chattopadhyay and Kristoffer Arnsfelt Hansen. Lower bounds for circuits with few modular and symmetric gates. In Luís Caires, Giuseppe F. Italiano, Luís Monteiro, Catuscia Palamidessi, and Moti Yung, editors, *ICALP*, volume 3580 of *Lecture Notes in Computer Science*, pages 994–1005. Springer, 2005.

10 Arkadev Chattopadhyay and Shachar Lovett. Linear systems over finite abelian groups. In *IEEE Conference on Computational Complexity*, pages 300–308. IEEE Computer Society, 2011.

11 Arkadev Chattopadhyay and Rahul Santhanam. Lower bounds on interactive compressibility by constant-depth circuits. In *FOCS*, pages 619–628. IEEE Computer Society, 2012.

12 Arkadev Chattopadhyay and Avi Wigderson. Linear systems over composite moduli. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 43–52. IEEE Computer Society, 2009.

13 G. Everest and T. Ward. *An Introduction to Number Theory*. Graduate Texts in Mathematics. Springer, 2005.

14 Mikael Goldmann, Johan Håstad, and Alexander A. Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

15 Mikael Goldmann and Marek Karpinski. Simulating threshold circuits by majority circuits. *SIAM J. Comput.*, 27(1):230–246, 1998.

16 Kristoffer Arnsfelt Hansen and Peter Bro Miltersen. Some meet-in-the-middle circuit lower bounds. In Jirí Fiala, Václav Koubek, and Jan Kratochvíl, editors, *MFCS*, volume 3153 of *Lecture Notes in Computer Science*, pages 334–345. Springer, 2004.

17 Thomas Hofmeister. A note on the simulation of exponential threshold weights. In Jin yi Cai and C. K. Wong, editors, *COCOON*, volume 1090 of *Lecture Notes in Computer Science*, pages 136–141. Springer, 1996.

18 Adam R. Klivans and Rocco A. Servedio. Learning dnf in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. Syst. Sci.*, 68(2):303–318, 2004.

19 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993.

20 Shachar Lovett and Srikanth Srinivasan. Correlation bounds for poly-size AC^0 circuits with $n^{1-o(1)}$ symmetric gates. In Leslie Ann Goldberg, Klaus Jansen, R. Ravi, and José D. P. Rolim, editors, *APPROX-RANDOM*, volume 6845 of *Lecture Notes in Computer Science*, pages 640–651. Springer, 2011.

21 Saburo Muroga. *Threshold logic and its applications*. Wiley, 1971.

22 Alexander A. Razborov. Lower bounds on the size of constant-depth networks over a complete basis with logical addition. *Mathematicheskie Zametki*, 41(4):598–607, 1987.

23 Rocco A. Servedio and Emanuele Viola. On a special case of rigidity. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:144, 2012.

24 Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, pages 77–82, 1987.

25 Roman Smolensky. On representations by low-degree polynomials. In *FOCS*, pages 130–138. IEEE Computer Society, 1993.

26 Jun Tarui. Probablistic polynomials, ac0 functions, and the polynomial-time hierarchy. *Theoretical Computer Science*, 113(1):167–183, 1993.

27 Emanuele Viola. On the power of small-depth computation. *Foundations and Trends in Theoretical Computer Science*, 5(1):1–72, 2009.

28 Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008.