Correlation bounds for polynomials over $\{0,1\}$[1]

Emanuele Viola[2]

Abstract

This article is a unified treatment of the state-of-the-art on the fundamental challenge of exhibiting explicit functions that have small correlation with low-degree polynomials over $\{0,1\}$. It discusses long-standing results and recent developments, related proof techniques, and connections with pseudorandom generators. It also suggests several research directions.

1 Introduction

This article is about one of the most basic computational models: low-degree polynomials over the field $\{0,1\} = \mathrm{GF}(2)$. For example, the following is a polynomial of degree 2 in 3 variables

$$p(x_1, x_2, x_3) := x_1 \cdot x_2 + x_2 + x_3 + 1,$$

given by the sum of the 4 monomials x_1x_2, x_2, x_3, and 1, of degree $2, 1, 1$, and 0, respectively. This polynomial computes a function from $\{0,1\}^3$ to $\{0,1\}$, which we also denote p, by performing the arithmetic over $\{0,1\}$. Thus the sum "+" is modulo 2 and is the same as "xor," while the product "$\cdot$" is the same as "and." For instance, $p(1,1,0) = 1$. Being complexity theorists rather than algebraists, we are only interested in the function computed by a polynomial, not in the polynomial itself; therefore we need not bother with variables raised to powers bigger than 1, since for $x \in \{0,1\}$ one has $x = x^2 = x^3$ and so on. In general, a polynomial p of degree d in n Boolean variables $x_1, \ldots, x_n \in \{0,1\}$ is a sum of monomials of degree at most d:

$$p(x_1, \ldots, x_n) = \sum_{M \subseteq \{1,\ldots,n\}, |M| \leq d} c_M \prod_{i \in M} x_i,$$

where $c_M \in \{0,1\}$ and we let $\prod_{i \in \emptyset} x_i := 1$; such a polynomial p computes a function $p : \{0,1\}^n \to \{0,1\}$, interpreting again the sum modulo 2. We naturally measure the complexity of a polynomial by its degree d: the maximum number of variables appearing in any monomial. Since every function $f : \{0,1\}^n \to \{0,1\}$ can be computed by a polynomial of degree n, specifically $f(x_1, \ldots, x_n) = \sum_{a_1,\ldots,a_2} f(a_1, \ldots, a_n) \prod_{1 \leq i \leq n} (1 + a_i + x_i)$, we are interested in polynomials of low degree $d \ll n$.

Low-degree polynomials constitute a fundamental model of computation that arises in a variety of contexts, ranging from error-correcting codes to circuit lower bounds. As for any computational model, a first natural challenge is to exhibit explicit functions that cannot be computed in the model. This challenge is easily won: the monomial $\prod_{i=1}^{d} x_i$ requires degree d. A second, natural challenge has baffled researchers, and is the central topic of this article. One now asks for functions that not only cannot be computed by low-degree polynomials, but do not even *correlate* with them.

[2] Northeastern University, Boston MA 02115. viola@ccs.neu.edu.

2 Correlation bounds

We start by defining the correlation between a function f and polynomials of degree d. This quantity captures how well we can approximate f by polynomials of degree d, and is also known as the average-case hardness of f against polynomials of degree d.

Definition 1 (Correlation)**.** *Let $f : \{0,1\}^* \to \{0,1\}$ be a function, n an integer, and D a distribution on $\{0,1\}^n$. We define the* correlation between f and a polynomial $p : \{0,1\}^n \to \{0,1\}$ with respect to D *as*

$$\mathrm{Cor}_D(f,p) := \left| \Pr_{x\sim D}[f(x) = p(x)] - \Pr_{x\sim D}[f(x) \neq p(x)] \right| = 2\left|1/2 - \Pr_{x\sim D}[f(x) \neq p(x)]\right| \in [0,1].$$

We define the correlation between f and polynomials of degree d with respect to D *as*

$$\mathrm{Cor}_D(f,d) := \max_p \mathrm{Cor}_D(f,p) \in [0,1],$$

where the maximum is over all polynomials $p : \{0,1\}^n \to \{0,1\}$ of degree d.

Unless specified otherwise, D is the uniform distribution and we simply write $\mathrm{Cor}(f,d)$.

For more than two decades, researchers have sought to exhibit explicit functions that have small correlation with high-degree polynomials, an enterprise we refer to as obtaining, or proving, "correlation bounds." A dream setting of parameters would be to exhibit a function $f \in \mathrm{P}$ such that for every n, and for D the uniform distribution over $\{0,1\}^n$, $\mathrm{Cor}_D(f, \epsilon \cdot n) \leq \exp(-\epsilon \cdot n)$, where $\epsilon > 0$ is an absolute constant and $\exp(x) := 2^x$ throughout this article. The original motivation for seeking correlation bounds comes from circuit complexity, because functions with small correlation with polynomials require large constant-depth circuits with various gates, see e.g. [Raz2, Sm, HMP$^+$, Be]. An additional motivation comes from pseudorandomness: as we will see, sufficiently strong correlation bounds can be used to construct pseudorandom generators [Ni, NW], which in turn have myriad applications. But as this article also aims to put forth, today the challenge of proving correlation bounds is interesting per se, and its status is a fundamental benchmark for our understanding of complexity theory: it is not known how to achieve the dream setting of parameters mentioned above, and in fact nobody can even achieve the following strikingly weaker setting of parameters.

Open question 1. *Is there a function $f \in \mathrm{NP}$ such that for arbitrarily large n there is a distribution D on $\{0,1\}^n$ with respect to which $\mathrm{Cor}_D(f, \log_2 n) \leq 1/n$?*

Before discussing known results in the next sections, we add to the above concise motivation for tackling correlation bounds the following discussion of their relationship with other open problems.

Correlation bounds' place in the hierarchy of open problems. We point out that a negative answer to Question 1 implies that NP has circuits of quasipolynomial size $s = n^{O(\log n)}$. This relatively standard fact can be proved via boosting [F, Section 2.2] or min-max/linear-programming duality [GHR, Section 5]. Thus, an affirmative answer to Question 1 is necessary to prove that NP does not have circuits of quasipolynomial size, a leading goal of theoretical computer science. Of course, this connection can be strengthened in various ways, for example noting that the circuits for NP given by a negative answer to Question 1 can be written on inputs of length n as a majority of $n^{O(1)}$ polynomials of degree $\log_2 n$; thus, an affirmative answer to Question 1 is necessary even to prove that NP does not have circuits of the latter type. On the other hand, Question 1 cannot easily be related to polynomial-size lower bounds such as $\mathrm{NP} \not\subseteq \mathrm{P/poly}$, because a polynomial of degree $\log n$ may have a quasipolynomial number of monomials.

While there are many other open questions in complexity theory, arguably Question 1 is among those having remarkable yet not dramatic consequences, and therefore should be attacked first. To illustrate, let us consider some of the major open problems in the area of unbounded-fan-in constant-depth circuits AC^0. One such problem is to exhibit an explicit function that requires AC^0 circuits of depth d and size $s \geq \exp(n^{\epsilon})$ for some $\epsilon \gg 1/d$ (current lower bounds give $\epsilon = O(1/d)$, see [Hå]). However, via a guess-and-verify construction usually credited to [Ne], one can show that any function $f \in \mathrm{NL}$ has AC^0 circuits of depth d and size $\exp\left(n^{c/d}\right)$ where c depends only on f. This means that a strong enough progress on this problem would separate NP from NL. Furthermore, techniques by Valiant (see [Va, Section 5], also pointing to an earlier related work by Erdős, Graham, and Szemerédi) entail that improving the known lower bounds for AC^0 circuits of depth 3 to size $s = \exp(\Omega(n))$ would result in a super-linear size lower bound for (fan-in 2) circuits of logarithmic depth. On the other hand, even an answer to Question 1 with strong parameters is not known to have such daunting consequences, nor, if that is of concern, is known to require "radically new" ideas [BGS, RR, AW].

Progress on Question 1 is also implied by a corresponding progress in number-on-forehead communication complexity. Specifically, a long-standing open question in communication complexity is to exhibit an explicit function $f : (\{0,1\}^n)^k \to \{0,1\}$ that cannot be computed by number-on-forehead k-party protocols exchanging $O(k)$ bits, for some $k \geq \log_2 n$ [KN, Problem 6.21]. A lower bound on the communication required to compute some function f is usually proved by establishing that f has low correlation with k-party protocols – a technique also known as the discrepancy method, cf. [KN, VW]. The connection with polynomials is given by a beautiful observation of Håstad and Goldmann [HG, Proof of Lemma 4], which implies that if f has low correlation with k-party protocols then f also has low correlation with polynomials of degree $d := k - 1$. But the converse connection is not known.

Finally, we note that polynomials over $\{0,1\}$ constitute a simple model of algebraic computation, and so Question 1 can also be considered a basic question in algebraic complexity. In fact, an interesting special case – to which we will return in §2.2, §2.4 – is whether one can take f in Question 1 to be an explicit low-degree polynomial over $\{0,1\}$.

After this high-level discussion, we now move to presenting the known correlation bounds.

It is a remarkable state of affairs that, while we are currently unable to make the correlation small and the degree large *simultaneously*, as required by Question 1, we can make the correlation small and the degree large *separately*. And in fact we can even achieve this for the same explicit function $f = \mathrm{mod}_3$. We examine these two types of results in turn.

2.1 Large degree $d \gg \log n$ but noticeable correlation $\epsilon \gg 1/n$

Razborov [Raz2] (also in [CK, Section 2.7.1]) proves the existence of a symmetric function $f : \{0,1\}^n \to \{0,1\}$ that has correlation at most $1 - 1/n^{O(1)}$ with polynomials of degree $\Omega(\sqrt{n})$ (a function is symmetric when its value only depends on the number of input bits that are '1').

Smolensky [Sm] obtains a refined bound for the explicit function $\mathrm{mod}_3 : \{0,1\}^n \to \{0,1\}$ which evaluates to 1 if and only if the number of input bits that are '1' is of the form $3k+1$ for some integer k, i.e., it is congruent to 1 modulo 3:

$$\mathrm{mod}_3(x_1, \ldots, x_n) = 1 \Leftrightarrow \sum_i x_i = 1 (\mathrm{mod}\ 3).$$

For example, $\mathrm{mod}_3(1,0,0) = \mathrm{mod}_3(0,1,0) = 1 \neq \mathrm{mod}_3(1,0,1)$.

Theorem 2 ([Sm]). *For any n that is divisible by 3, and for U the uniform distribution over $\{0,1\}^n$, $\mathrm{Cor}_U(\mathrm{mod}_3, \epsilon\sqrt{n}) \leq 2/3$, where $\epsilon > 0$ is an absolute constant.*

While the proof of Smolensky's result has appeared several times, e.g. [Sm, BS, Be, AB2], we are unaware of a source that directly proves Theorem 2, and thus we include next a proof for completeness (the aforementioned sources either focus on polynomials over the field with three elements, or prove the bound for one of the three functions $\mathrm{mod}_{i,3}(x_1, \ldots, x_n) = 1 \Leftrightarrow \sum_i x_i = i (\mathrm{mod}\ 3)$ for $i = 0, 1, 2$).

Proof. The idea is to consider the set of inputs $X \subseteq \{0,1\}^n$ where the polynomial computes the mod_3 function correctly, and use the polynomial to represent any function defined on X by a polynomial of degree $n/2 + d$. This means that the number of functions defined on X should be smaller than the number of polynomials of degree $n/2 + d$, which leads to the desired tradeoff between $|X|$ and d. To carry through this argument, one works over a field F that extends $\{0,1\}$.

We start by noting that, since n is divisible by 3, one has

$$\sum_i x_i = 2 (\mathrm{mod}\ 3) \Leftrightarrow \sum_i 1 - x_i = 1 (\mathrm{mod}\ 3) \Leftrightarrow \mathrm{mod}_3(1 + x_1, \ldots, 1 + x_n) = 1, \qquad (1)$$

where the sums $1 + x_i$ in the input to mod_3 are modulo 2. Let F be the field of size 4 that extends $\{0,1\}$, which we can think of as $F = \{0,1\}[t]/(t^2 + t + 1)$: the set of polynomials over $\{0,1\}$ modulo the irreducible polynomial $t^2 + t + 1$. Note that $t \in F$ has order 3, since $t^2 = t + 1 \neq 1$, while $t^3 = t^2 + t = 1$. Let $h : \{1, t\} \to \{0,1\}$ be the "change of domain" linear map $h(\alpha) := (\alpha + 1)/(t + 1)$; this satisfies $h(1) = 0$ and $h(t) = 1$.

Observe that for every $y \in \{1,t\}^n$ we have, using Equation (1):

$$y_1 \cdots y_n = 1 + (t+1) \cdot \mathrm{mod}_3(h(y_1), \ldots, h(y_n)) + (t^2+1) \cdot \mathrm{mod}_3(1+h(y_1), \ldots, 1+h(y_n)). \quad (2)$$

Now fix any polynomial $p : \{0,1\}^n \to \{0,1\}$ and let

$$\Pr_{x \in \{0,1\}^n}[p(x) \neq \mathrm{mod}_3(x)] =: \delta,$$

which we aim to bound from below. Let $p' : \{1,t\}^n \to F$ be the polynomial

$$p'(y_1, \ldots, y_n) := 1 + (t+1) \cdot p(h(y_1), \ldots, h(y_n)) + (t^2+1) \cdot p(1+h(y_1), \ldots, 1+h(y_n));$$

note p' has the same degree d of p. By the definition of p' and δ, a union bound, and Equation (2) we see that

$$\Pr_{y \in \{1,t\}^n}[y_1 \cdots y_n = p'(y_1, \ldots, y_n)] \geq 1 - 2\delta. \quad (3)$$

Now let $S \subseteq \{1,t\}^n$ be the set of $y \in \{1,t\}^n$ such that $y_1 \cdots y_n = p'(y_1, \ldots, y_n)$; we have just shown that $|S| \geq 2^n(1-2\delta)$. Any function $f : S \to F$ can be written as a polynomial over F where no variable is raised to powers bigger than 1: $f(y_1, \ldots, y_n) = \sum_{a_1,\ldots,a_n} f(a_1, \ldots, a_n) \prod_{1 \leq i \leq n} (1 + h(y_i) + h(a_i))$. In any such polynomial we can replace any monomial M of degree $|M| > n/2$ by a polynomial of degree at most $n/2 + d$ as follows, without affecting the value on any input $y \in S$:

$$\prod_{i \in M} y_i = y_1 \cdots y_n \prod_{i \notin M} (y_i(t+1) + t) = p'(y_1, \ldots, y_n) \prod_{i \notin M} (y_i(t+1) + t),$$

where the first equality is not hard to verify. Doing this for every monomial we can write $f : S \to F$ as a polynomial over F of degree $\lfloor n/2 + d \rfloor$.

The number of functions from S to F is $|F|^{|S|}$, while the number of polynomials over F of degree $\lfloor n/2 + d \rfloor$ is $|F|^{\sum_{i=0}^{\lfloor n/2+d \rfloor} \binom{n}{i}}$. Thus

$$\log_{|F|} \#\text{functions} = |S| = 2^n(1-2\delta) \leq \sum_{i=0}^{\lfloor n/2+d \rfloor} \binom{n}{i} = \log_{|F|} \#\text{polynomials}.$$

Since $d = \epsilon\sqrt{n}$, we have

$$\sum_{i=0}^{\lfloor n/2+d \rfloor} \binom{n}{i} \leq 2^n/2 + d \cdot \binom{n}{\lfloor n/2 \rfloor} \leq 2^n/2 + \epsilon\sqrt{n} \cdot \Theta\left(\frac{2^n}{\sqrt{n}}\right) = (1/2 + \Theta(\epsilon))2^n,$$

where the second inequality follows from standard estimates on binomial coefficients. The standard estimate for even n is for example in [CT2, Lemma 17.5.1]; for odd $n = 2k+1$ one can first note $\binom{n}{\lfloor n/2 \rfloor} = \binom{2k+1}{k} < \binom{2k+2}{k+1} = \binom{n+1}{(n+1)/2}$ and then again apply [CT2, Lemma 17.5.1]. Therefore $1 - 2\delta \leq 1/2 + \Theta(\epsilon)$ and the theorem is proved. □

The limitation of the argument. There are two reasons why we get a poor correlation bound in the above proof of Theorem 2. The first is the union bound in (3), which immediately puts us in a regime where we cannot obtain subconstant correlation. This regime is unavoidable as the polynomial $p = 0$ of degree 0 has constant correlation with mod_3 with respect to the uniform distribution. (Later we will see a different distribution with respect to which mod_3 has vanishing, exponentially small correlation with polynomials of degree $\ll \log n$.) Nevertheless, let us pretend that the union bound in (3) is not there. This is not pointless because this step is indeed not present in related correlation bounds, which do however suffer from the second limitation we are about to discuss. The related correlation bounds are those between the parity function

$$\mathrm{parity}(x_1, \ldots, x_n) := x_1 + \cdots + x_n \qquad \mathrm{parity} : \{0,1\}^n \to \{0,1\}$$

and polynomials over the field with three elements, see e.g. [BS, AB2], or between the parity function and the sign of polynomials over the integers [ABFR]. If we assume that the union bound in (3) is missing, then we get $2^n(1-\delta) \leq \sum_{i=0}^{\lfloor n/2+d \rfloor} \binom{n}{i}$. Even if $d = 1$, this only gives $1 - \delta \leq 1/2 + \Theta(1/\sqrt{n})$, which means that the correlation is $\Theta(1/\sqrt{n})$: this argument does not give a correlation bound of the form $o(1/\sqrt{n})$. More generally, to our knowledge Question 1 is also open when replacing $1/n$ with $1/\sqrt{n}$.

Xor lemma. A striking feature of the above results ([Raz2] and Theorem 2) is that they prove non-trivial correlation bounds for polynomials of very high degree $d = n^{\Omega(1)}$. In this sense these results address the computational model which is the subject of Question 1, they "just" fail to provide a strong enough bound on the correlation. For other important computational models this would not be a problem: the extensive study of *hardness amplification* has developed many techniques to improve correlation bounds in the following sense: given an explicit function $f : \{0,1\}^n \to \{0,1\}$ that has correlation ϵ with some class $\mathcal{C}_n$ of functions on n bits, construct another explicit function $f' : \{0,1\}^{n'} \to \{0,1\}$, where $n' \approx n$, that has correlation $\epsilon' \ll \epsilon$ with a closely related class $\mathcal{C}_{n'}$ of functions on n' bits (see [SV] for a comprehensive list of references to research in hardness amplification). While the following discussion holds for any hardness amplification, for concreteness we focus on the foremost: Yao's xor lemma. Here $f' : (\{0,1\}^n)^k \to \{0,1\}$ is defined as the xor (or parity, or sum modulo 2) of k independent outputs of f:

$$f'(x^1, \ldots, x^k) := f(x^1) + \cdots + f(x^k) \in \{0,1\}, \qquad x^i \in \{0,1\}^n.$$

The compelling intuition is that, since functions from $\mathcal{C}_n$ have correlation at most ϵ with f, and f' is the xor of k independent evaluations of f, the correlation should decay exponentially with k: $\epsilon' \approx \epsilon^k$. This is indeed the case if one tries to compute $f'(x^1, \ldots, x^k)$ as $g_1(x^1) + \cdots + g_k(x^k)$ where $g_i : \{0,1\}^n \to \{0,1\}, g_i \in \mathcal{C}_n, 1 \leq i \leq k$, but in general a function $g : (\{0,1\}^n)^k \to \{0,1\}, g \in \mathcal{C}_{n'}$, needs not have this structure, making proofs of Yao's xor lemma more subtle. If we could prove this intuition true for low-degree polynomials, we could combine this with Theorem 2 to answer affirmatively Question 1 via the function

$$f(x^1, \ldots, x^k) := \mathrm{mod}_3(x^1) + \cdots + \mathrm{mod}_3(x^k) \tag{4}$$

for $k = n$. Of course the obstacle is that nobody knows whether Yao's xor lemma holds for polynomials.

Open question 2. *Does Yao's xor lemma hold for polynomials of degree $d \geq \log_2 n$? For example, let $f : \{0,1\}^n \to \{0,1\}$ satisfy $\mathrm{Cor}(f, n^{1/3}) \leq 1/3$, and for $n' := n^2$ define $f' : \{0,1\}^{n'} \to \{0,1\}$ as $f'(x^1, \ldots, x^n) := f(x^1) + \cdots + f(x^n)$. Is $\mathrm{Cor}(f', \log_2 n') \leq 1/n'$?*

We now discuss why, despite the many alternative proofs of Yao's xor lemma that are available (e.g., [GNW]), we cannot apply any of them to the computational model of low-degree polynomials. To prove that f' has correlation at most ϵ' with some class of functions, all known proofs of the lemma need (a slight modification of) the functions in the class to compute the majority function on about $1/\epsilon'$ bits. However, the majority function on $1/\epsilon'$ bits requires polynomials of degree $\Omega(1/\epsilon')$. This means that known proofs can only establish correlation bounds $\epsilon' \gg 1/n$, failing to answer Question 2. More generally, the works [Vi1, SV] show that computing the majority function on $1/\epsilon'$ bits is necessary for a central class of hardness amplification proofs.

An xor lemma is however known for polynomials of small degree $d \ll \log n$ [VW] (this and the other results on polynomials in [VW] appeared also in [Vi2]). In general, the picture for polynomials of small degree is different, as we now describe.

2.2 Negligible correlation $\epsilon \ll 1/n$ but small degree $d \ll \log n$

It is easy to prove exponentially small correlation bounds for polynomials of degree 1, for example the *inner product* function $\mathrm{IP} : \{0,1\}^n \to \{0,1\}$, defined for even n as

$$\mathrm{IP}(x_1, \ldots, x_n) := x_1 \cdot x_2 + x_3 \cdot x_4 + \cdots + x_{n-1} \cdot x_n, \tag{5}$$

satisfies $\mathrm{Cor}(\mathrm{IP}, 1) = 2^{-n/2}$. Already obtaining exponentially small bounds for polynomials of constant degree appears to be a challenge. The first such bounds come from the famed work by Babai, Nisan, and Szegedy [BNS] proving exponentially small correlation bounds between polynomials of degree $d := \epsilon \log n$ and, for $k := d + 1$, the *generalized inner product* function $\mathrm{GIP}_k : \{0,1\}^n \to \{0,1\}$,

$$\mathrm{GIP}_k(x_1, \ldots, x_n) := \prod_{i=1}^{k} x_i + \prod_{i=k+1}^{2k} x_i + \cdots + \prod_{i=n-k+1}^{n} x_i,$$

assuming for simplicity that n is a multiple of k. The intuition for this correlation bound is precisely that behind Yao's xor lemma (cf. §2.1): (i) any polynomial of degree d has correlation that is bounded away from 1 with any monomial of degree $k = d + 1$ in the definition of GIP, and (ii) since the monomials in the definition of GIP are on disjoint sets of variables, the correlation decays exponentially. (i) is not hard to establish formally. With some work, (ii) can also be established to obtain the following theorem.

Theorem 3 ([BNS])**.** *For every n, d, $\mathrm{Cor}(\mathrm{GIP}_{d+1}, d) \leq \exp\left(-\Omega(n/4^d \cdot d)\right)$.*

When $k \gg \log n$, GIP is almost always 0 on a uniform input, and thus GIP is not a candidate for having small correlation with respect to the uniform distribution with polynomials of degree $d \gg \log n$.

Our exposition of the results in [BNS] differs in multiple ways from the original. First, [BNS] does not discuss polynomials but rather number-on-forehead multiparty protocols. The results for polynomials are obtained via the observation of Håstad and Goldmann [HG, Proof of Lemma 4] mentioned in the subsection "Correlation bounds' place in the hierarchy of open problems" of §2. Second, [BNS] presents the proof with a different language. Alternative languages have been put forth in a series of papers [CT1, Raz1, VW], with the last one stressing the above intuition and the connections between multiparty protocols and polynomials.

Bourgain [Bo] later proves bounds similar to those in Theorem 3 but for the mod_3 function. A minor mistake in his proof is corrected by F. Green, Roy, and Straubing [GRS].

Theorem 4 ([Bo, GRS])**.** *For every n, d there is a distribution D on $\{0,1\}^n$ such that $\mathrm{Cor}_D(\mathrm{mod}_3, d) \leq \exp\left(-n/c^d\right)$, where c is an absolute constant.*

A random sample from the distribution D in Theorem 4 is obtained as follows: toss a fair coin, if "heads" then output a uniform $x \in \{0,1\}^n$ such that $\mathrm{mod}_3(x) = 1$, if "tails" then output a uniform $x \in \{0,1\}^n$ such that $\mathrm{mod}_3(x) = 0$. The value $c = 8$ in [Bo, GRS] is later improved to $c = 4$ in [VW, C]. [VW] also presents the proof in a different language.

Theorem 4 appears more than a decade after Theorem 3. However, Noam Nisan (personal communication) observes that in fact the first easily follows from the latter.

Sketch of Nisan's proof of Theorem 4. Grolmusz's [Gro] extends the results in [BNS] and shows that there is a distribution D' on $\{0,1\}^n$ such that for $k := d + 1$ the function

$$\mathrm{mod}_3 \wedge_k (x_1, \ldots, x_n) := \mathrm{mod}_3 \left(\prod_{i=1}^{k} x_i, \prod_{i=k+1}^{2k} x_i, \ldots, \prod_{i=n-k+1}^{n} x_i \right)$$

has correlation $\exp(-n/c^d)$ with polynomials of degree d, for an absolute constant c. A proof of this can also be found in [VW, §3.3]. An inspection of the proof reveals that, with respect to another distribution D'' on $\{0,1\}^n$, the same bound applies to the function

$$\mathrm{mod}_3\mathrm{mod}_2(x_1, \ldots, x_n) := \mathrm{mod}_3(x_1 + \cdots + x_k, x_{k+1} + \cdots + x_{2k}, \ldots, x_{n-k+1} + \cdots + x_n)$$

where we replace "and" with "parity;" the sums in the input to mod_3 are modulo 2.

Now consider the distribution D on $\{0,1\}^{n/k}$ that D'' induces on the input to mod_3 of length n/k. (To sample from D one can sample from D'', perform the n/k sums modulo 2, and return the string of length n/k.) Suppose that a polynomial $p(y_1, \ldots, y_{n/k})$ of degree d has correlation ϵ with the mod_3 function with respect to D. Then the polynomial

$$p'(x_1, \ldots, x_n) := p(x_1 + \cdots + x_k, x_{k+1} + \cdots + x_{2k}, \ldots, x_{n-k+1} + \cdots + x_n)$$

has degree d and correlation ϵ with the $\mathrm{mod}_3\mathrm{mod}_2$ function with respect to the distribution D'' on $\{0,1\}^n$. Therefore $\epsilon \leq \exp(-n/c^d)$. □

The mod_m functions have recently got even more attention because as shown in [GT2, LMS] they constitute a counterexample to a conjecture independently made in [GT3] and [Sa]. The main technical step in the counterarguments in [GT2, LMS] is to show an upper bound on the correlation between polynomials of degree 3 and the function

$$\mathrm{mod}_{\{4,5,6,7\},8}(x_1,\ldots,x_n) := 1 \Leftrightarrow \sum_i x_i \in \{4,5,6,7\} (\mathrm{mod}\ 8).$$

The strongest bound is given by Lovett, Meshulam, and Samorodnitsky [LMS] who prove the following theorem.

Theorem 5 ([LMS]). *For every n, $\mathrm{Cor}(\mathrm{mod}_{\{4,5,6,7\},8}, 3) \leq \exp(-\epsilon \cdot n)$, where $\epsilon > 0$ is an absolute constant.*

In fact, to disprove the conjecture in [GT3, Sa] any bound of the form $\mathrm{Cor}(\mathrm{mod}_{\{4,5,6,7\},8}, 3) \leq o(1)$ is sufficient. Such a bound was implicit in the clever 2001 work by Alon and Beigel [AB1]. With an unexpected use of Ramsey's theorem for hypergraphs, they were the first to establish that the parity function has vanishing correlation with constant-degree polynomials over $\{0,1,2\}$. A slight modification of their proof gives $\mathrm{Cor}(\mathrm{mod}_{\{4,5,6,7\},8}, 3) \leq o(1)$, and can be found in the paper by B. Green and Tao [GT2].

It is interesting to note that the function $\mathrm{mod}_{\{4,5,6,7\},8}$ is in fact a polynomial of degree 4 over $\{0,1\}$, the so-called elementary symmetric polynomial of degree 4

$$s_4(x_1,\ldots,x_n) := \sum_{1 \leq i_1 < i_2 < i_3 < i_4 \leq n} x_{i_1} \cdot x_{i_2} \cdot x_{i_3} \cdot x_{i_4}.$$

For suitable input lengths, elementary symmetric polynomials of higher degree d are candidates for having small correlation with polynomials of degree less than d. To our knowledge, even $d = n^{\Omega(1)}$ is a possibility.

The "squaring trick." Many of the results in this section, and all those that apply to degree $d \approx \log n$ (Theorems 3 and 4) use a common technique which we now discuss also to highlight its limitation. The idea is to reduce the challenge of proving a correlation bound for a polynomial of degree d to that of proving related correlation bounds for polynomials of degree $d-1$, by *squaring*. To illustrate, let $f : \{0,1\}^n \to \{0,1\}$ be any function and $p : \{0,1\}^n \to \{0,1\}$ a polynomial of degree d. Using the extremely convenient notation $e[z] := (-1)^z$, we write the correlation between f and p with respect to the uniform distribution as

$$\mathrm{Cor}(f,p) = \left| \Pr_{x \in \{0,1\}^n}[f(x) = p(x)] - \Pr_{x \in \{0,1\}^n}[f(x) \neq p(x)] \right| = \left| E_{x \in \{0,1\}^n} e[f(x) + p(x)] \right|.$$

If we square this quantity, and use that $E_Z[g(Z)]^2 = E_{Z,Z'}[g(Z) \cdot g(Z')]$, we obtain

$$\mathrm{Cor}(f,p)^2 = E_{x,y \in \{0,1\}^n} e[f(x) + f(y) + p(x) + p(y)].$$

Letting now $y = x + h$ we can rewrite this as

$$\mathrm{Cor}(f,p)^2 = E_{x,h\in\{0,1\}^n} e[f(x) + f(x+h) + p(x) + p(x+h)].$$

The crucial observation is now that, for every fixed h, the polynomial $p(x) + p(x+h)$ has degree $d-1$ in x, even though $p(x)$ has degree d. For example, if $d = 2$ and $p(x) = x_1x_2 + x_3$, we have $p(x)+p(x+h) = x_1x_2+x_3+(x_1+h_1)(x_2+h_2)+(x_3+h_3) = x_1h_2+h_1x_2+h_1h_2+h_3$, which indeed has degree $d-1 = 1$ in x. Thus we managed to reduce our task of bounding from above $\mathrm{Cor}(f,p)$ to that of bounding from above a related quantity which involves polynomials of degree $d-1$, specifically the average over h of the correlation between the function $f(x) + f(x+h)$ and polynomials of degree $d-1$. To iterate, we apply the same trick, this time coupled with the Cauchy-Schwarz inequality $E[Z]^2 \le E[Z^2]$:

$$\mathrm{Cor}(f,p)^4 = E_{x,h} e[f(x)+f(x+h)+p(x)+p(x+h)]^2 \le E_h\left[E_x e[f(x) + f(x+h) + p(x) + p(x+h)]^2\right].$$

We can now repeat the argument in the inner expectation, further reducing the degree of the polynomial. After d repetitions, the polynomial p becomes a constant. After one more, a total of $d+1$ repetitions, the polynomial p "disappears" and we are left with a certain expectation involving f, known as the "Gowers norm" of f and introduced independently in [Go1, Go2] and in [AKK+]:

$$\mathrm{Cor}(f,p)^{2^{d+1}} \le E_{x,h_1,h_2,\ldots,h_{d+1}} e\left[\sum_{S\subseteq[d+1]} f\left(x + \sum_{i\in S} h_i\right)\right]. \tag{6}$$

For interesting functions f, the expectation in the right-hand side of (6) can be easily shown to be small, sometimes exponentially in n, yielding correlation bounds. For example, applying this method to the generalized inner product function gives Theorem 3, while a complex-valued generalization of the method can be applied to the mod_3 function to obtain Theorem 4. This concludes the exposition of this technique; see, e.g., [VW] for more details.

This "squaring trick" for reducing the analysis of a polynomial of degree d to that of an expression involving polynomials of lower degree $d-1$ dates back at least to the work by Weyl at the beginning of the 20th century; for an exposition of the relevant proof by Weyl, as well as pointers to his work, the reader may consult [GR1]. This method was introduced in computer science by Babai, Nisan, and Szegedy [BNS], and employed later by various researchers [Go1, Go2, Bo, GT3, VW] in different contexts.

The obvious limitation of this technique is that, to bound the correlation with polynomials of degree d, it squares the correlation d times; this means that the bound on the correlation will be $\exp(-n/2^d)$ at best: nothing for degree $d = \log_2 n$. This bound is almost achieved by [BNS] which gives an explicit function f such that $\mathrm{Cor}(f,d) \le \exp(-\Omega(n/2^d \cdot d))$. The extra factor of d in the exponent arises because of the different context of multiparty protocols in [BNS], but a similar argument, given in [VW], establishes the following stronger bound.

Theorem 6 ([BNS, VW])**.** *There is an explicit $f \in P$ such that for every n and d, and U the uniform distribution over $\{0,1\}^n$, $\mathrm{Cor}_U(f,d) \le \exp(-\Omega(n/2^d))$.*

The function $f : \{0,1\}^n \to \{0,1\}$ in Theorem 6 takes as input an index $i \in \{0,1\}^{\epsilon n}$ and a seed $s \in \{0,1\}^{(1-\epsilon)n}$, and outputs the i-th output bit of a certain pseudorandom generator on seed s [NN] (Theorem 10 in §3). The natural question of whether these functions have small correlation with polynomials of degree $d \gg \log_2 n$ has been answered negatively in [VW] building on the results in [Raz2, GV, HV, He]: it can be shown that, for a specific implementation of the generator, the associated function $f : \{0,1\}^n \to \{0,1\}$ satisfies $\mathrm{Cor}(f, \log^c n) \geq 1 - o(1)$ for an absolute constant c. Determining how small c can be is an open problem whose solution might be informative, given that such functions are of great importance, as we will also see in §3.

2.3 Symmetric functions correlate well with degree $O(\sqrt{n})$

Many of the correlation bounds discussed in §2.1 and §2.2 are given by functions that are symmetric: their value depends only on the number of input bits that are '1.' In this section we show that any symmetric function $f : \{0,1\}^n \to \{0,1\}$ correlates well with polynomials of degree $O(\sqrt{n})$, matching the degree obtained in Theorem 2 up to constant factors, and excluding symmetric functions from the candidates to the dream setting of parameters $\mathrm{Cor}(f, \epsilon \cdot n) \leq \exp(-\epsilon \cdot n)$. While there is a shortage of such candidates, we point out that techniques in hardness amplification such as [IW] may be relevant. It also seems worthwhile to investigate whether the result in this section extends to other distributions.

Theorem 7. *For every n, every symmetric function $f : \{0,1\}^n \to \{0,1\}$ satisfies $\mathrm{Cor}(f, c\sqrt{n}) \geq 99/100$, where c is an absolute constant.*

We present below a proof of Theorem 7 that was communicated to us by Avi Wigderson and simplifies an independent argument of ours. It relies on a result by Bhatnagar, Gopalan, and Lipton, stated next, which in turn follows from well-known facts about the divisibility of binomial coefficients by 2, such as Lucas' theorem.

Lemma 8 (Corollary 2.7 in [BGL])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a function such that $f(x)$ depends only on the Hamming weight of x modulo 2^ℓ. Then f is computable by a polynomial of degree $d < 2^\ell$*

In §2.2 we saw an example of Lemma 8 when we noticed that the function $\mathrm{mod}_{\{4,5,6,7\},8}$ is computable by a polynomial of degree $4 < 2^3 = 8$. This polynomial was symmetric, and more generally the polynomials in Lemma 8 and Theorem 7 can be taken to be symmetric.

Proof of Theorem 7. The idea is to exhibit a polynomial of degree $O(\sqrt{n})$ that computes f on every input of Hamming weight between $n/2 - a\sqrt{n}$ and $n/2 + a\sqrt{n}$; for a suitable constant a this gives correlation 99/100 by a Chernoff bound.

Let a be a sufficiently large universal constant to be determined later, and let 2^ℓ be the smallest power of 2 bigger than $2a\sqrt{n} + 1$, thus $2^\ell \leq c\sqrt{n}$ for a constant c that depends only on a. Now take any function $f' : \{0,1\}^n \to \{0,1\}$ such that (i) f' agrees with f on every input of Hamming weight between $n/2 - a\sqrt{n}$ and $n/2 + a\sqrt{n}$, and (ii) the value of

$f'(x)$ depends only on the Hamming weight of x modulo 2^ℓ. Such an f' exists because f is symmetric and we ensured that $2^\ell > 2a\sqrt{n}+1$.

Applying first Lemma 8 and then a Chernoff bound (e.g., [DP, Theorem 1.1]) for a sufficiently large constant a, we have for $d < 2^\ell \le c\sqrt{n}$

$$\mathrm{Cor}(f,d) \ge \mathrm{Cor}(f,f') \ge 99/100,$$

which concludes the proof of the theorem. □

2.4 Other works

There are many papers on correlation bounds we have not discussed. F. Green [Gre, Theorem 3.10] manages to compute exactly the correlation between the parity function and quadratic ($d=2$) polynomials over $\{0,1,2\}$, which is $(3/4)^{\lceil n/4\rceil -1}$. [Gre] further discusses the difficulties currently preventing an extension of the result to degree $d>2$ or polynomials over fields different from $\{0,1,2\}$, while [GR2] studies the structure of quadratic polynomials over $\{0,1,2\}$ that correlate with the parity function best.

The work [Vi3] gives an explicit function that, with respect to the uniform distribution over $\{0,1\}^n$, has correlation $1/n^{\omega(1)}$ with polynomials of arbitrary degree but with at most $n^{\alpha\cdot\log n}$ monomials, for a small absolute constant $\alpha>0$. This is obtained by combining a switching lemma with Theorem 3, a technique from [RW]. The result does not answer Question 1 because a polynomial of degree $\log_2 n$ can have $\binom{n}{\log_2 n} \gg n^{\alpha\cdot\log n}$ monomials, and in fact the function in [Vi3] is a polynomial of degree $(0.3)\log_2 n$. For polynomials over $\{0,1,2\}$, the same correlation bounds hold for the parity function [Ha].

Other special classes of polynomials, for example symmetric polynomials, are studied in [CGT, GT1, BEHL]. We finally mention that many of the works we discussed, such as Theorems 2, 4, and 7 can be suitably extended to polynomials modulo $m \neq 2$. We chose to focus on $m=2$ because it is clean.

3 Pseudorandom generators vs. correlation bounds

In this section we discuss pseudorandom generators for polynomials and their connections to correlation bounds. Pseudorandom generators are fascinating algorithms that stretch short input seeds into much longer sequences that "look random;" naturally, here we interpret "look random" as "look random to polynomials," made formal in the next definition.

Definition 9 (Generator)**.** *We say that a map $G:\{0,1\}^* \to \{0,1\}^*$ is a* generator *$G:\{0,1\}^s \to \{0,1\}^n$ that fools polynomials of degree $d=d(n)$ with error $\epsilon=\epsilon(n)$ and seed length $s=s(n)$ if $x\in\{0,1\}^s$ implies $G(x)\in\{0,1\}^n$ and (i) for any n and polynomial $p:\{0,1\}^n\to\{0,1\}$ of degree d we have*

$$\left| \Pr_{S\in\{0,1\}^s}[p(G(S))=1] - \Pr_{U\in\{0,1\}^n}[p(U)=1] \right| \le \epsilon, \tag{7}$$

and (ii) G is computable in time polynomial in its output length.

Ideally, we would like generators that fool polynomials of large degree d with small error ϵ and small seed length s. We discuss below various connections between obtaining such generators and correlation bounds, but first we point out a notable difference: while for correlation bounds we do have results for large degree $d \gg \log n$ (e.g., Theorem 2), we know of no generator that fools polynomials of degree $d \geq \log_2 n$, even with constant error ϵ.

Open question 3. *Is there a generator $G : \{0,1\}^{n/2} \to \{0,1\}^n$ that fools polynomials of degree $\log_2 n$ with error $1/3$?*

While the smaller the error ϵ the better, generators for constant error are already of great interest; for example, a constant-error generator that fools small circuits is enough to derandomize BPP. However, we currently seem to be no better at constructing generators that fool polynomials with constant error than generators with shrinking error, such as $1/n$.

We now discuss the relationship between generators and correlation bounds, and then present the known generators.

From pseudorandomness to correlation. It is easy to see and well-known [NW] that a generator implies a worst-case lower bound, i.e., an explicit function that cannot be computed by (essentially) the class of functions fooled by the generator. The following simple observation, which does not seem to have appeared before [Vi4, §3], shows that in fact a generator implies even a correlation bound. We will use it later to obtain new candidates for answering Question 1.

Observation 1. *Suppose that there is a generator $G : \{0,1\}^{n-\log n-1} \to \{0,1\}^n$ that fools polynomials of degree $\log_2 n$ with error $0.5/n$. Then the answer to Question 1 is affirmative.*

Proof sketch. Let D be the distribution on $\{0,1\}^n$ where a random $x \in D$ is obtained as follows: toss a fair coin, if "heads" then let x be uniformly distributed over $\{0,1\}^n$, if "tails" then let $x := G(S)$ for a uniformly chosen $S \in \{0,1\}^{n-\log n-1}$. Define the function $f : \{0,1\}^n \to \{0,1\}$ as $f(x) = 1$ if and only if there is $s \in \{0,1\}^{n-\log n-1}$ such that $G(s) = x$; f is easily seen to be in NP. It is now a routine calculation to verify that any function $t : \{0,1\}^n \to \{0,1\}$ that satisfies $\mathrm{Cor}_D(f,t) \geq 1/n$ has advantage at least $0.5/n$ in distinguishing the output of the generator from random. Letting t range over polynomials of degree $\log_2 n$ concludes the proof. □

From correlation to pseudorandomness. The celebrated construction by Nisan and Wigderson [Ni, NW] shows that a sufficiently strong correlation bound with respect to the uniform distribution can be used to obtain a generator that fools polynomials. However, to obtain a generator $G : \{0,1\}^s \to \{0,1\}^n$ against polynomials of degree d, [NW] in particular needs a function f on $m \leq n$ input bits that has correlation at most $1/n$ with polynomials of degree d. Thus, the current correlation bounds are not strong enough to obtain generators for polynomials of degree $d \geq \log_2 n$. It is a pressing open problem to determine whether alternative constructions of generators are possible, ideally based on constant correlation

bounds such as Theorem 2. Here, an uncharted direction is to understand which distributions D enable one to construct generators starting from correlation bounds with respect to D.

The Nisan-Wigderson construction is however sharp enough to give non-trivial generators based on the current correlation bounds such as Theorem 3. Specifically, Luby, Veličković, and Wigderson [LVW, Theorem 2] obtain generators for polynomials that have arbitrary degree but at most $n^{\alpha \cdot \log n}$ terms for a small absolute constant $\alpha > 0$; a different proof of this result appears in the paper [Vi3] which we already mentioned in §2.4. Albeit falling short of answering Question 3 (cf. §2.4), this generator [LVW, Theorem 2] does fool polynomials of constant degree. However, its seed length, satisfying $n = s^{O(\log s)}$, has been superseded in this case by recent developments, which we now discuss.

Generators for degree $d \ll \log n$. Naor and Naor [NN] construct a generator that fools polynomials of degree 1 (i.e., linear) with a seed length that is optimal up to constant factors – a result with a surprising range of applications (cf. references in [BSSVW]).

Theorem 10 ([NN])**.** *There is a generator $G : \{0,1\}^{O(\log n)} \to \{0,1\}^n$ that fools polynomials of degree 1 with error $1/n$.*

Later, Alon et al. [AGHP] give three alternative constructions. A nice one is $G(a,b)_i := \langle a^i, b \rangle$ where $\langle \cdot, \cdot \rangle$ denotes inner product modulo 2, $a, b \in \{0,1\}^\ell$ for $\ell = O(\log n)$, and a^i denotes the result of considering a as an element of the field with 2^ℓ elements, and raising it to the power i.

Recent progress by Bogdanov, Lovett, and the author [BV, L, Vi4] has given generators for higher degree. The high-level idea in these works is simple: to fool polynomials of degree d, just sum together d generators for polynomials of degree 1.

Theorem 11 ([Vi4])**.** *Let $G : \{0,1\}^s \to \{0,1\}^n$ be a generator that fools polynomials of degree 1 with error ϵ. Then $G_d : (\{0,1\}^s)^d \to \{0,1\}^n$ defined as*

$$G_d(x^1, x^2, \ldots, x^d) := G(x^1) + G(x^2) + \cdots + G(x^d)$$

fools polynomials of degree d with error $\epsilon_d := 16 \cdot \epsilon^{1/2^{d-1}}$, where '+' denotes bit-wise xor.

In particular, the combination of Theorems 10 and 11 yields generators $G : \{0,1\}^s \to \{0,1\}^n$ that fool polynomials of degree d with error $\epsilon_d = 1/n$ and seed length $s = O(d \cdot 2^d \cdot \log(n))$.

Proof sketch of Theorem 11. This proof uses the notation $e[z] := (-1)^z$ and some of the techniques presented at the end of §2.2. First, let us rewrite Inequality (7) in the Definition 9 of a generator as

$$\left| E_{S \in \{0,1\}^s} e[p(G_d(S))] - E_{U \in \{0,1\}^n} e[p(U)] \right| \leq \epsilon_d / 2. \tag{8}$$

To prove Inequality (8), we proceed by induction on the degree d of the polynomial $p : \{0,1\}^n \to \{0,1\}$ to be fooled. The inductive step is structured as a case analysis based on the value $\tau := \mathrm{Cor}_U(p, 0) = |E_{U \in \{0,1\}^n} e[p(U)]|$.

If τ is large then p correlates with a constant, which is a polynomial of degree lower than d, and one can prove the intuitive fact that by induction G_{d-1} fools p. This concludes the overview of the proof in this case.

If τ is small we reason as follows. Let us denote by W the output of G_{d-1} and by Y an independent output of G, so that the output of G_d is $W + Y$. We start by an application of the Cauchy-Schwarz inequality:

$$\mathrm{E}_{W,Y}\, e\left[p(W+Y)\right]^2 \leq E_W\left[\mathrm{E}_Y\, e\left[p(W+Y)\right]^2\right] = \mathrm{E}_{W,Y,Y'}\, e\left[p(W+Y)+p(W+Y')\right], \quad (9)$$

where Y' is independent from and identically distributed to Y. Now we observe that for every fixed Y and Y', the polynomial $p(U+Y)+p(U+Y')$ has degree $d-1$ in U, though p has degree d. Since by induction W fools polynomials of degree $d-1$ with error ϵ_{d-1}, we can replace W with the uniform distribution $U \in \{0,1\}^n$:

$$\mathrm{E}_{W,Y,Y'}\, e\left[p(W+Y)+p(W+Y')\right] \leq \mathrm{E}_{U,Y,Y'}\, e\left[p(U+Y)+p(U+Y')\right] + \epsilon_{d-1}. \quad (10)$$

At this point, a standard argument shows that

$$\mathrm{E}_{U,Y,Y'}\, e\left[p(U+Y)+p(U+Y')\right] \leq \mathrm{E}_{U,U'}\, e\left[p(U)+p(U')\right] + \epsilon^2 = \tau^2 + \epsilon^2. \quad (11)$$

Therefore, chaining Equations (9), (10), and (11), we have that

$$\left|\mathrm{E}_{W,Y}\, e\left[p(W+Y)\right] - \mathrm{E}_U\, e\left[p(U)\right]\right| \leq \left|\mathrm{E}_{W,Y}\, e\left[p(W+Y)\right]\right| + \tau \leq \sqrt{\tau^2+\epsilon^2+\epsilon_{d-1}} + \tau.$$

This proves Inequality (8) for a suitable choice of ϵ_d, concluding the proof in this remaining case. □

Observe that Theorem 11 gives nothing for polynomials of degree $d = \log_2 n$. The reason is that its proof again relies on the "squaring trick" discussed in §2.2. But it is still not known whether the construction in Theorem 11 fools polynomials of degree $d \geq \log_2 n$.

Open question 4. *Does the sum of $d \gg \log n$ copies of a generator $G : \{0,1\}^s \to \{0,1\}^n$ that fools polynomials of degree 1 with error $1/n$ fools polynomials of degree d with error $1/3$?*

Finally, note that Observation 1 combined with the construction in Theorem 11 gives a new candidate function for an affirmative answer to Question 1: the function that on input $x \in \{0,1\}^n$ evaluates to 1 if and only if x is the bit-wise xor of $d \gg \log n$ outputs of generators that fool polynomials of degree 1.

4 Conclusion

We have discussed the challenge of proving correlation bounds for polynomials. We have seen that winning this challenge is necessary for proving lower bounds such as "NP does not have quasipolynomial-size circuits," that a great deal is known for various settings of parameters, and that there are many interesting research directions. The clock is ticking...

Acknowledgments. We are grateful to Noam Nisan for his permission to include his proof of Theorem 4, Rocco Servedio for a discussion on boosting, and Avi Wigderson for suggesting a simpler proof of Theorem 7. We also thank Frederic Green, Shachar Lovett, Rocco Servedio, and Avi Wigderson for helpful comments on a draft of this report.

References

[AB1] Noga Alon and Richard Beigel. Lower bounds for approximations by low degree polynomials over Z_m. In *16th Annual Conference on Computational Complexity*, pages 184–187. IEEE, June 18–21 2001. 9

[AB2] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach (Draft)*. January 2007. 4, 6

[ABFR] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994. 6

[AGHP] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992. 14

[AKK+] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing low-degree polynomials over GF(2). In *7th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, volume 2764 of *Lecture Notes in Computer Science*, pages 188–199. Springer, 2003. 10

[AW] Scott Aaronson and Avi Wigderson. Algebrization: a new barrier in complexity theory. In *40th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 731–740, 2008. 3

[Be] Richard Beigel. The polynomial method in circuit complexity. In *8th Annual Structure in Complexity Theory Conference*, pages 82–95. IEEE, 1993. 2, 4

[Bo] Jean Bourgain. Estimation of certain exponential sums arising in complexity theory. *C. R. Math. Acad. Sci. Paris*, 340(9):627–631, 2005. 8, 10

[BEHL] Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. Random low degree polynomials are hard to approximate. Manuscript, 2008. 12

[BGL] 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. 11

[BGS] Theodore Baker, John Gill, and Robert Solovay. Relativizations of the P=?NP question. *SIAM J. Comput.*, 4(4):431–442, 1975. 3

[BNS] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. 7, 8, 10

[BS] Ravi B. Boppana and Michael Sipser. The complexity of finite functions. In *Handbook of theoretical computer science, Vol. A*, pages 757–804. Elsevier, Amsterdam, 1990. 4, 6

[BSSVW] Eli Ben-Sasson, Madhu Sudan, Salil Vadhan, and Avi Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *35th Annual Symposium on Theory of Computing*, pages 612–621. ACM, 2003. 14

[BV] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. In *48th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 41–51. IEEE, 2007. 14

[C] Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *48th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 449–458. IEEE, 2007. 8

[CGT] Jin-Yi Cai, Frederic Green, and Thomas Thierauf. On the correlation of symmetric functions. *Mathematical Systems Theory*, 29(3):245–258, 1996. 12

[CK] Peter Clote and Evangelos Kranakis. *Boolean Functions and Computation Models*. Springer, 2002. 4

[CT1] Fan R. K. Chung and Prasad Tetali. Communication complexity and quasi randomness. *SIAM J. Discrete Math.*, 6(1):110–123, 1993. 8

[CT2] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006. 5

[DP] Devdatt Dubhashi and Alessandro Panconesi. Concentration of measure for the analysis of randomised algorithms, 2005. http://www.dsi.uniroma1.it/~ale/papers.html. 12

[F] Y. Freund. Boosting a weak learning algorithm by majority. *Information and Computation*, 121(2):256–285, September 1995. 3

[Go1] W. T. Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998. 10

[Go2] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001. 10

[Gre] Frederic Green. The correlation between parity and quadratic polynomials mod 3. *J. Comput. System Sci.*, 69(1):28–44, 2004. 12

[Gro] Vince Grolmusz. Separating the communication complexities of MOD m and MOD p circuits. *J. Comput. System Sci.*, 51(2):307–313, 1995. 8

[GHR] Mikael Goldmann, Johan Håstad, and Alexander A. Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992. 3

[GNW] Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's XOR lemma. Technical Report TR95–050, *Electronic Colloquium on Computational Complexity*, March 1995. www.eccc.uni-trier.de/. 7

[GR1] Andrew Granville and Zeév Rudnick. Uniform distribution. In *Equidistribution in Number Theory, An Introduction*, volume 237 of *NATO Science Series II: Mathematics, Physics and Chemistry*, pages 1–13. Springer, 2007. 10

[GR2] Frederic Green and Amitabha Roy. Uniqueness of optimal mod 3 circuits for parity. In *Dagstuhl Seminar Proceedings, Algebraic Methods in Computational Complexity*, volume 07411, 2007. 12

[GRS] Frederic Green, Amitabha Roy, and Howard Straubing. Bounds on an exponential sum arising in Boolean circuit complexity. *C. R. Math. Acad. Sci. Paris*, 341(5):279–282, 2005. 8

[GT1] Anna Gál and Vladimir Trifonov. On the correlation between parity and modular polynomials. In *31st International Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 4162 of *Lecture Notes in Computer Science*, pages 387–398. Springer, 2006. 12

[GT2] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms, 2007. arXiv:0711.3191v1. 9

[GT3] Ben Green and Terence Tao. An inverse theorem for the Gowers $U^3(G)$ norm. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 51(01):73–153, 2008. 9, 10

[GV] Dan Gutfreund and Emanuele Viola. Fooling parity tests with parity gates. In *8th International Workshop on Randomization and Computation (RANDOM)*, Lecture Notes in Computer Science, Volume 3122, pages 381–392. Springer, 2004. 11

[Ha] Kristoffer Arnsfelt Hansen. Lower bounds for circuits with few modular gates using exponential sums. *Electronic Colloquium on Computational Complexity*, Technical Report TR06-079, 2006. www.eccc.uni-trier.de/. 12

[Hå] Johan Håstad. *Computational limitations of small-depth circuits*. MIT Press, 1987. 3

[He] Alexander Healy. Randomness-efficient sampling within NC^1. *Computational Complexity*, 17(1):3–37, April 2008. 11

[HG] Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Comput. Complexity*, 1(2):113–129, 1991. 3, 8

[HMP+] András Hajnal, Wolfgang Maass, Pavel Pudlák, Mário Szegedy, and György Turán. Threshold circuits of bounded depth. *J. Comput. System Sci.*, 46(2):129–154, 1993. 2

[HV] Alexander Healy and Emanuele Viola. Constant-depth circuits for arithmetic in finite fields of characteristic two. In *23rd International Symposium on Theoretical Aspects of Computer Science (STACS)*, Lecture Notes in Computer Science, Volume 3884, pages 672–683. Springer, 2006. 11

[IW] Russell Impagliazzo and Avi Wigderson. $P = BPP$ if E requires exponential circuits: Derandomizing the XOR lemma. In *29th Annual Symposium on Theory of Computing (STOC)*, pages 220–229. ACM, 1997. 11

[KN] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, Cambridge, 1997. 3

[L] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. In *40th Annual Symposium on the Theory of Computing (STOC)*, pages 557–562. ACM, 2008. 14

[LMS] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. In *40th Annual Symposium on the Theory of Computing*

(STOC), pages 547–556. ACM, 2008. 9

[LVW] Michael Luby, Boban Velickovic, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *2nd Israeli Symposium on Theoretical Computer Science (ISTCS)*, pages 18–24, 1993. 14

[Ne] Valery A. Nepomnjaščiĭ. Rudimentary predicates and Turing calculations. *Soviet Mathematics-Doklady*, 11(6):1462–1465, 1970. 3

[Ni] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991. 2, 13

[NN] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993. 11, 14

[NW] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Computer & Systems Sciences*, 49(2):149–167, 1994. 2, 13

[Raz1] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Comput. Complexity*, 9(2):113–122, 2000. 8

[Raz2] Alexander Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Mat. Zametki*, 41(4):598–607, 623, 1987. 2, 4, 6, 11

[RR] Alexander Razborov and Steven Rudich. Natural proofs. *Journal of Computer and System Sciences*, 55(1):24–35, August 1997. 3

[RW] Alexander Razborov and Avi Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. *Inform. Process. Lett.*, 45(6):303–307, 1993. 12

[Sa] Alex Samorodnitsky. Low-degree tests at large distances. In *39th Annual Symposium on Theory of Computing (STOC)*, pages 506–515, 2007. 9

[Sm] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *19th Annual Symposium on Theory of Computing*, pages 77–82. ACM, 1987. 2, 4

[SV] Ronen Shaltiel and Emanuele Viola. Hardness amplification proofs require majority. In *40th Annual Symposium on the Theory of Computing (STOC)*, pages 589–598. ACM, 2008. 6, 7

[Va] Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. In *6th Symposium on Mathematical Foundations of Computer Science*, volume 53 of *Lecture Notes in Computer Science*, pages 162–176. Springer, 1977. 3

[Vi1] Emanuele Viola. *The Complexity of Hardness Amplification and Derandomization.* PhD thesis, Harvard University, 2006. www.eccc.uni-trier.de/. 7

[Vi2] Emanuele Viola. New correlation bounds for GF(2) polynomials using Gowers uniformity. *Electronic Colloquium on Computational Complexity*, Technical Report TR06-097, 2006. www.eccc.uni-trier.de/. 7

[Vi3] Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM Journal on Computing*, 36(5):1387–1403, 2007. 12, 14

[Vi4] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. In *23nd Annual Conference on Computational Complexity*. IEEE, June 23–26

2008. http://www.ccs.neu.edu/home/viola/. 13, 14

[VW] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. *Theory of Computing*, 4:137–168, 2008. 3, 7, 8, 10, 11