Pseudorandom generators for $\mathrm{CC}^0[p]$ and the Fourier spectrum of low-degree polynomials over finite fields

Extended abstract

Shachar Lovett
Department of Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel
Email: shachar.lovett@weizmann.ac.il

Partha Mukhopadhyay
Department of Computer Science
Technion
Haifa 32000, Israel
Email: partha@cs.technion.ac.il

Amir Shpilka
Department of Computer Science
Technion
Haifa 32000, Israel
Email: shpilka@cs.technion.ac.il

***Abstract*—In this paper we give the first construction of a pseudorandom generator, with seed length $O(\log n)$, for $\mathrm{CC}^0[p]$, the class of constant-depth circuits with unbounded fan-in MOD_p gates, for some prime p. More accurately, the seed length of our generator is $O(\log n)$ for any constant error $\epsilon > 0$. In fact, we obtain our generator by fooling distributions generated by low degree polynomials, over $\mathbb{F}_p$, *when evaluated on the Boolean cube*. This result significantly extends previous constructions that either required a long seed [1] or that could only fool the distribution generated by linear functions over $\mathbb{F}_p$, when evaluated on the Boolean cube [2], [3].**

Enroute of constructing our PRG, we prove two structural results for low degree polynomials over finite fields that can be of independent interest.

1) **Let f be an n-variate degree d polynomial over $\mathbb{F}_p$. Then, for every $\epsilon > 0$ there exists a subset $S \subset [n]$, whose size depends only on d and ϵ, such that $\sum_{\alpha \in \mathbb{F}_p^n : \alpha \neq 0, \alpha_S = 0} |\hat{f}(\alpha)|^2 \leq \epsilon$. Namely, there is a constant size subset S such that the total weight of the nonzero Fourier coefficients that do not involve any variable from S is small.**
2) **Let f be an n-variate degree d polynomial over $\mathbb{F}_p$. If the distribution of f when applied to uniform zero-one bits is ϵ-far (in statistical distance) from its distribution when applied to biased bits, then for every $\delta > 0$, f can be approximated over zero-one bits, up to error δ, by a function of a small number (depending only on ϵ, δ and d) of lower degree polynomials.**

***Keywords*-constant depth circuits; pseudorandom generator; fourier spectrum; low degree polynomials;**

The first author is supported by the Israel Science Foundation (grant 1300/05) and ERC starting grant 239985. The second author is supported in part at the Technion by an Aly Kaufman Fellowship and by the Israel Science Foundation (grant 439/06). The third author is supported by the Israel Science Foundation (grant 439/06).

I. Introduction

A *pseudorandom generator* (PRG for short), over a domain D,[1] for a family of *tests* $\mathcal{T}$ is an explicit function $G : D^r \to D^n$ such that no test $T \in \mathcal{T}$ can distinguish a random output of G from truly uniform input elements in D^n. Namely,

$$\max_{T \in \mathcal{T}} \left| \Pr_{x \in D^r}[T(G(x)) = 0] - \Pr_{x \in D^n}[T(x) = 0] \right| \leq \epsilon \, .$$

Ideally, one would like to have the *seed* r as short as possible and the error ϵ to be as small as possible. A pseudorandom generator is considered efficient if the seed length is $O(\log n)$ (as in this case, for some applications, one can enumerate over all seeds to find a 'good' one). Pseudorandom generators have been a major object of study in theoretical computer science for several decades, and have found applications in the area of computational complexity, cryptography, algorithms design and more (see [4], [5]).

A family of tests that was widely considered in the literature is low degree polynomials over finite fields. Before stating the formal definition of a PRG for low degree polynomials we fix some notation: let f be a function and $\mathcal{D}$ a distribution over the inputs of f. We denote by $f(\mathcal{D})$ the output distribution of f given inputs sampled according to $\mathcal{D}$. For a set S we denote by $f(S)$ the output distribution given that the inputs are uniformly sampled in S (for example, $f(\{0,1\}^n)$ is the distribution of f over uniform input bits).

Definition 1 (Pseudorandom distributions for degree d polynomials)**.** A distribution $\mathcal{D}$ taking values in $\mathbb{F}_p^n$ is *pseudorandom for degree d polynomials over $\mathbb{F}_p$* with

[1] One should think of D as either over $\{0,1\}$ or over $\mathbb{F}_p$.

error ϵ if, for any degree d polynomial $f(x_1, \ldots, x_n)$ over $\mathbb{F}_p$, the distributions $f(\mathcal{D})$ and $f(\mathbb{F}_p^n)$ are ϵ-close in statistical distance. A function $G : \{0,1\}^r \to \mathbb{F}_p^n$ is a *pseudorandom generator for degree d polynomials over* $\mathbb{F}_p$, if the output distribution of G, given uniformly sampled seeds, is a pseudorandom distribution for degree d polynomials.

PRGs for linear polynomials over $\mathbb{F}_2$ were first constructed in [6] who gave PRGs with $O(\log n)$ seed length. The distributions constructed in [6] are also known as ϵ-biased distributions. Alon et al. extended this construction to work over arbitrary finite fields [7]. In [1] a pseudorandom generator for the class of bounded degree polynomials over finite fields was given.[2] The seed length of [1] was not optimal and was later improved in a sequence of works [8], [9], [10]. Note that all these generators take as input vectors from $\mathbb{F}_p^r$ and output vectors in $\mathbb{F}_p^n$. In [2], [3] a different kind of PRGs for linear polynomials were obtained. Both works constructed a PRG $G : \{0,1\}^r \to \{0,1\}^n$ that fools distributions generated by linear polynomials over $\mathbb{F}_p$, when evaluated on $\{0,1\}^n$, with seed length $r = O(\log n)$. Namely, if $f = \sum_{i=1}^n \alpha_i x_i$ is a linear polynomial over $\mathbb{F}_p$ then the two distributions $f(G(\{0,1\}^r))$ and $f(\{0,1\}^n)$ are close to each other. Thus, although f is a polynomial over $\mathbb{F}_p$ they restrict their attention to the behavior of f on Boolean inputs. We call such a generator a *bit-pseudorandom generator*. We shall later give a more formal definition of bit-PRGs.

Another family of tests that received a lot of attention is bounded depth circuits (i.e. AC^0 circuits). This is the class of constant-depth circuits with unbounded fan-in AND, OR and NOT gates. AC^0 is probably the most intensively studied amongst classes of small-depth circuits. Håstad [11] showed that the PARITY function cannot be approximated by any polynomial size AC^0 circuit. I.e., that no polynomial size AC^0 circuit agrees with parity on more than $\frac{1}{2} + \exp(-n)$ fraction of inputs. In other words, the *correlation* of PARITY with AC^0 is exponentially small. This result was later used by Nisan [12] for constructing efficient pseudorandom generators for AC^0 (these pseudorandom generators use $r = \mathrm{polylog(n)}$ bits). Recently, following a breakthrough by Bazzi [13], Braverman [14] showed that any polylog-wise independent distribution is pseudorandom for AC^0 circuits, thus settling a conjecture of Linial and Nisan [15]. $\mathrm{AC}^0[p]$ is another well studied class of circuits, consisting of all constant-depth circuits with unbounded fan-in AND, OR, NOT and MOD_p gates (a MOD_p gate outputs 1 if the sum of its inputs is divisible by p, and 0 otherwise). In contrast to the impressive success in constructing pseudorandom generators for AC^0, no PRGs are known for $\mathrm{AC}^0[p]$. One reason is that no strong correlation lower bounds are known for this class. Razborov and Smolensky [16], [17] proved exponential lower bounds for $\mathrm{AC}^0[p]$ circuits and their results also imply correlation lower bounds, albeit those are much weaker than the ones known for AC^0. Namely, [16], [17] showed that the MOD_q function has polynomially small correlation with $\mathrm{AC}^0[p]$ when p and q are coprime. The class of $\mathrm{AC}^0[m]$ where m is not a prime power is only very weakly understood; in particular, currently we cannot separate it from NP!

[2]This is not explicitly stated in [1], but it follows from their result for depth 2 circuits with a symmetric function at the top.

A. Our results

Motivated by the problem of constructing pseudorandom generators for $\mathrm{AC}^0[p]$, we study a natural subclass - $\mathrm{CC}^0[p]$ circuits. The class $\mathrm{CC}^0[p]$ is the class of constant depth circuits using only MOD_p gates. While exponential lower bounds for this class follow from the work of Smolensky [17], no pseudorandom generator better than the one constructed in [1] (whose seed length is $r = \exp(\sqrt{\log n})$) is known for it. Our main result is an explicit pseudorandom generator fooling any $\mathrm{CC}^0[p]$ circuit while using only $r = O(\log n)$ random bits, for any fixed error $\epsilon > 0$. Actually, our construction gives bit-pseudorandom generators for low-degree polynomials over finite fields, from which the result for $\mathrm{CC}^0[p]$ follows: Let $\mathbb{F}_p$ be a prime finite field. The MOD_p function can be computed by a degree $p-1$ polynomial over $\mathbb{F}_p$

$$\mathrm{MOD}_p(x_1, \ldots, x_n) = (x_1 + \ldots + x_n)^{p-1} \pmod p .$$

Hence, any depth k circuit in $\mathrm{CC}^0[p]$ can be computed by a polynomial over $\mathbb{F}_p$ of degree $d = (p-1)k$. Thus, in order to fool $\mathrm{CC}^0[p]$ it is sufficient to fool distributions induced by low degree polynomials over $\mathbb{F}_p$, when evaluated on inputs from the Boolean cube. In other words, we have to generalize the aforementioned results of [2], [3] from linear polynomials to any constant degree polynomials. This motivates the following definition of bit-pseudorandom generators for polynomials.

Definition 2 (Bit-pseudorandom distributions for degree d polynomials)**.** A distribution $\mathcal{D}$ taking values in $\{0,1\}^n$ is *bit-pseudorandom for degree d polynomials over* $\mathbb{F}_p$ with error ϵ if, for any degree d polynomial $f(x_1, \ldots, x_n)$ over $\mathbb{F}_p$, the distributions $f(\mathcal{D})$ and $f(\{0,1\}^n)$ are ϵ-close in statistical distance. A function $G : \{0,1\}^r \to \{0,1\}^n$ is a *bit-pseudorandom generator*

for degree d polynomials over $\mathbb{F}_p$ if the output distribution of G over a uniform seed is a bit-pseudorandom distribution for degree d polynomials.

Notice the difference between this definition and Definition 1 where one has to fool the distribution of the polynomial when evaluated over the entire space and not just over the Boolean cube. As mentioned above, PRGs for polynomials over small finite fields were studied in several works [1], [8], [9], [10]. The best result to date is by Viola.

Theorem 3 (Theorem 1 in [10])**.** *There exists an explicit and efficient function $G : \{0,1\}^r \to \mathbb{F}_p^n$ for $r = O(d \cdot \log(pn) + 2^d \cdot \log(1/\epsilon))$ such that $G(\{0,1\}^r)$ is pseudorandom for degree d polynomials over $\mathbb{F}_p$ with error ϵ.*

The problem of constructing bit-pseudorandom generators for linear polynomials (i.e. the case of $d = 1$) was first studied by [2], [3]. Before describing their generator we need a few notations. For $a = (a_1, \ldots, a_n) \in \mathbb{F}_p^n$ define $a^{p-1} = (a_1^{p-1}, \ldots, a_n^{p-1}) \in \{0,1\}^n$ to be the $p-1$ power of a. Similarly for a distribution $\mathcal{D} \subset \mathbb{F}_p^n$, define $\mathcal{D}^{p-1} \subset \{0,1\}^n$ by raising each element of $\mathcal{D}$ to the $p-1$ power. Both [2], [3] discovered the following construction for a bit-pseudorandom generator for linear polynomials over $\mathbb{F}_p$: the bitwise-XOR of the $p-1$ power of a pseudorandom distribution for degree $(p-1)$ polynomial over $\mathbb{F}_p$, and a k-wise independent distribution.

Theorem 4 (Bit-pseudorandom distribution for linear polynomials [2], [3])**.** *Let $\mathbb{F}_p$ be a prime finite field and $\epsilon > 0$ be an error parameter. Let $\mathcal{D} \subset \mathbb{F}_p^n$ be a pseudorandom distribution for degree $p-1$ polynomials over $\mathbb{F}_p$ with error ϵ. Let $K \subset \{0,1\}^n$ be a k-wise independent distribution for $k = O(p^3 \log 1/\epsilon)$. Then $\mathcal{D}^{p-1} \oplus K$ is bit-pseudorandom distribution for linear polynomials over $\mathbb{F}_p$ with error $O(\epsilon)$.*

Our main result extends Theorem 4 to any constant degree polynomial. We prove that the following is a bit-pseudorandom distribution for degree d polynomials over $\mathbb{F}_p$: the bitwise-XOR of the $p-1$ power of a pseudorandom distribution for degree $((p-1)d)$ polynomials over $\mathbb{F}_p$, and a k-wise independent distribution.

Theorem 5 (Main Theorem: Bit-pseudorandom distribution)**.** *Let $\mathbb{F}_p$ be an odd prime finite field, $d \geq 1$ an integer and $\epsilon > 0$ an error parameter. Then there exist $\delta = \delta(p,d,\epsilon)$ and $k = k(p,d,\epsilon)$ such that the following holds. Let $\mathcal{D} \subset \mathbb{F}_p^n$ be a pseudorandom distribution for degree $((p-1)d)$ polynomials with error δ. Let $K \subset \{0,1\}^n$ be a k-wise independent distribution. Then, the bitwise-XOR of the two distributions $\mathcal{D}^{p-1} \oplus K$ is a bit-pseudorandom distribution for degree d polynomials over $\mathbb{F}_p$ with error ϵ. The parameters k, δ satisfy*

$$k(p,d,\epsilon), \delta(p,d,\epsilon)^{-1} \leq \exp^{(2d+1)}(\epsilon^{-c_{p,d}})$$

where $\exp^{(t)}$ is the t-times iterated exponential function, and $c_{p,d} > 0$ is some constant which depends only on p and d.

An immediate corollary is that there exists an efficient and explicit pseudorandom generator $G : \{0,1\}^r \to \{0,1\}^n$ fooling any depth-k circuit in $\mathrm{CC}^0[p]$ with error ϵ, where $r = c_{p,k,\epsilon} \cdot \log n$.

Corollary 6 (Pseudorandom generators for $\mathrm{CC}^0[p]$)**.** *Let p be an odd prime number and $\epsilon > 0$ an error parameter. For any $k > 0$ there exists an explicit pseudorandom generator $G : \{0,1\}^r \to \{0,1\}^n$, where $r = c_{p,k,\epsilon} \cdot \log n$, such that for any depth k circuit $C \in \mathrm{CC}^0[p]$, the statistical distance between the two distributions $C(\{0,1\}^n)$ and $C(G(\{0,1\}^r))$ is at most ϵ.*

Our proof of Theorem 5 is based on two new structural results for low degree polynomials, over finite fields, which may be of independent interest:

The first result is on the Fourier spectrum of such polynomials. Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a function. The α-Fourier coefficient of f, for $\alpha \in \mathbb{F}_p^n$, is defined as

$$\widehat{f}(\alpha) = \mathbb{E}_{x \in \mathbb{F}_p^n}\left[\omega^{f(x) - \langle x, \alpha \rangle}\right],$$

where $\omega = e^{2\pi i/p}$ is a primitive p-root of unity, and $\langle x, \alpha \rangle = \sum_{i=1}^n x_i \alpha_i$ is the inner product of x and α. The structural result we prove is that the Fourier coefficients of any low-degree polynomial cannot be spread over many disjoint sets. In other words, we show that one can always find a small set $S \subset [n]$ such that almost all Fourier coefficients intersect S (that is, have some nonzero entry inside S). We note that while Theorem 5 is interesting only for odd p,[3] this structural result is non-trivial also for polynomials over $\mathbb{F}_2$.

Theorem 7 (The Fourier spectrum of low-degree polynomials over finite fields)**.** *For every prime finite field $\mathbb{F}_p$, degree $d \geq 1$ and error $\epsilon > 0$ there exists a constant $C(d,\epsilon) \leq (1/\epsilon)^{O(4^d)}$ such that the following holds. Let $f(x_1, \ldots, x_n)$ be a degree d polynomial over $\mathbb{F}_p$. Then there exists a subset $S \subset [n]$ of size at most*

[3]For $p = 2$ it reduces to the case of pseudorandom distributions.

$|S| \leq C(d,\epsilon)$ *such that*

$$\sum_{\alpha \in \mathbb{F}_p^n : \alpha \neq 0, \alpha_S = 0} |\hat{f}(\alpha)|^2 \leq \epsilon \ ,$$

where α_S is the restriction of α to coordinates in S. In words, there is a constant size subset S such that the total weight of the nonzero Fourier coefficients that do not involve any variable from S is small.

Our second structural result concerns the structure of polynomials with the following property. Denote with $\mathcal{U}_p$ the distribution over $\{0,1\}^n$ where each bit is chosen independently to be 0 with probability $1/p$ and 1 with probability $1-1/p$. We call $\mathcal{U}_p$ the *p-biased distribution*. We show that if the distributions $f(\mathcal{U}_p)$ and $f(\{0,1\}^n)$ are ϵ-far, then f can be approximated, over $\{0,1\}^n$, by a function of a small number of lower degree polynomials. To formally state our theorem we need some definitions.

Definition 8 (Bit-Rank)**.** Let $g : \{0,1\}^n \to \mathbb{F}_p$ be a function. The d-bit-rank of g, denoted bit-rank$_d(g)$, is the minimal number of degree d polynomials over $\mathbb{F}_p$ required to compute g over $\{0,1\}^n$. That is, bit-rank$_d(g) = k$ where k is the minimal number such that there exist k degree d polynomials $f_1, \ldots, f_k : \mathbb{F}_p^n \to \mathbb{F}_p$ and a function $\Gamma : \mathbb{F}_p^k \to \mathbb{F}_p$ such that for all $x \in \{0,1\}^n$

$$g(x) = \Gamma(f_1(x), \ldots, f_k(x)).$$

Example. *Consider the function $g(x) = \sum_{i \neq j} x_i x_j$ over $\mathbb{F}_p$ for $p > 2$. We have that the 1-bit-rank of g is 1, as for all $x \in \{0,1\}^n$*

$$\begin{aligned} g(x) &= (x_1 + \ldots + x_n)^2 - (x_1^2 + \ldots + x_n^2) \\ &= (x_1 + \ldots + x_n)^2 - (x_1 + \ldots + x_n) \ . \end{aligned}$$

Thus, for $x \in \{0,1\}^n$, $g(x)$ is determined by the linear function $\ell(x) = x_1 + \ldots + x_n$. Notice that as a quadratic polynomial over $\mathbb{F}_p$, the rank of g (i.e. the minimal number of linear functions required to compute g on inputs from $\mathbb{F}_p^n$) is either $n-1$ or n, depending on p.

Our second structural result is the following.

Theorem 9 (Structure of bit-biased polynomials)**.** *Let $f(x_1, \ldots, x_n)$ be a degree $d \geq 2$ polynomial over $\mathbb{F}_p$ such that the statistical distance between the distributions $f(\mathcal{U}_p)$ and $f(\{0,1\}^n)$ is at least ϵ. Then, for every $\delta > 0$, there exists a function $g : \{0,1\}^n \to \mathbb{F}_p$ such that $\Pr_{x \in \{0,1\}^n}[g(x) \neq f(x)] \leq \delta$ and bit-rank$_{d-1}(g) \leq p^{O(c)}$ where[4] $c = C((p-1)d, \delta\epsilon^2/p^3)$.*

[4]The function $C(\cdot,\cdot)$ is defined in the statement of Theorem 7.

In fact, for our proof we require such a polynomial g that approximates f with respect to (an affine shift of) $\mathcal{U}_p$, but we find this statement more appealing.

We provide an overview of the proof in the next section. We first give a proof of Theorem 5 assuming Theorems 9 and 7. In Section II-B we sketch the proof of Theorem 9 and in Section II-C we prove Theorem 7. The complete proofs can be found in the full version of the paper.

II. Proof overview

Pseudorandom generators that fool low degree polynomials over $\mathbb{F}_p^n$ were obtained in [8], [9], [10]. In our case we only consider the distribution of the polynomial over $\{0,1\}^n$ (and not over $\mathbb{F}_p^n$ as the aforementioned results), which creates new obstacles, and requires a different approach.

We sketch below the proof of Theorem 5. Our proof is carried by induction on the degree d, and uses Theorem 7 and (a variant of) Theorem 9 as important technical ingredients. We sketch their proofs in the subsequent subsections.

A. Proof overview of Theorem 5

We prove in this section that if $\mathcal{D}$ is a pseudorandom distribution for degree $(p-1)d$ polynomials and K is a k-wise independent distribution, then $\mathcal{D}^{p-1} \oplus K$ is a bit-pseudorandom distribution for degree d polynomials.

Let $f(x) = f(x_1, \ldots, x_n)$ be a polynomial of degree d over $\mathbb{F}_p$. The base case of $d = 1$ was established in [2], [3], hence we assume from now on that $d \geq 2$. We say that a polynomial f is *regular* if it cannot distinguish between the uniform distribution over $\{0,1\}^n$ and the p-biased distribution $\mathcal{U}_p$. We first show that it is simple to construct bit-pseudorandom generators for regular polynomials from pseudorandom generators for somewhat higher degree polynomials. We then proceed to handle the harder case of non-regular polynomials, where the main tool used is a variant of Theorem 9.

1) Regular polynomials: Consider the p-biased distribution $\mathcal{U}_p$. This distribution can be simulated by low-degree polynomials over $\mathbb{F}_p$: let $x \in \mathbb{F}_p^n$ be chosen uniformly at random; then, $x^{p-1} = (x_1^{p-1}, \ldots, x_n^{p-1})$ is distributed according to $\mathcal{U}_p$. Furthermore, it is easy to construct a pseudorandom distribution fooling $f(\mathcal{U}_p)$ as follows. Let $\widetilde{f}(x) = f(x^{p-1})$. Then $\widetilde{f}$ is a polynomial of degree $(p-1)d$, and the distributions $\widetilde{f}(\mathbb{F}_p^n)$ and $f(\mathcal{U}_p)$ are identical. In particular, any distribution fooling degree $(p-1)d$ polynomials over $\mathbb{F}_p$ (such as those guaranteed by Theorem 3) also fools $f(\mathcal{U}_p)$, when raised to the power $p-1$.

Thus, if the polynomial f is regular in the sense that it cannot distinguish between the uniform distribution over $\{0,1\}^n$ and the p-biased distribution $\mathcal{U}_p$, then one can simply use a pseudorandom generator for $\widetilde{f}$ to get a pseudorandom generator for f. Hence, it is not hard to deduce the following lemma.

Lemma 10. *Let $f(x)$ be a degree d polynomial over $\mathbb{F}_p$ such that the distributions $f(\mathcal{U}_p)$ and $f(\{0,1\}^n)$ are ϵ-close. Let $\mathcal{D} \subset \mathbb{F}_p^n$ be a pseudorandom distribution for degree $((p-1)d)$ polynomials over $\mathbb{F}_p$ with error ϵ. Then $f(\mathcal{D}^{p-1})$ and $f(\{0,1\}^n)$ are $O(\epsilon)$-close.*

2) Non-regular polynomials: We now turn to study non-regular polynomials. Namely, polynomials that can distinguish between the uniform distribution over $\{0,1\}^n$ and the p-biased distribution. The main tool in the proof is (a variant of) Theorem 9 that shows that non regular polynomials possess a very special structure. More specifically, that a non-regular polynomial can be well approximated by a function of a small number of lower degree polynomials.

We will start by proving that non-regular polynomials admit a non-uniform distribution when applied to inputs sampled from some shift of the p-biased distribution. For a distribution $\mathcal{D} \subset \{0,1\}^n$ and an element $a \in \{0,1\}^n$ denote by $\mathcal{D} \oplus a$ the distribution generated by bitwise-XORing the element a to all elements of $\mathcal{D}$. It is not hard to obtain the following claim.

Claim 11. *Let $f : \{0,1\}^n \to \mathbb{F}_p$ be a function such that the distributions $f(\mathcal{U}_p)$ and $f(\{0,1\}^n)$ are ϵ-far. Then there exists $a \in \{0,1\}^n$ such that the distribution $f(\mathcal{U}_p \oplus a)$ is $\epsilon/2$-far from the uniform distribution over $\mathbb{F}_p$.*

The following theorem (which is a variant of Theorem 9) shows that non-regular polynomials have a low bit-rank.

Theorem 12. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a polynomial of degree $d \geq 2$. Assume that, for some $a \in \{0,1\}^n$, the distribution $f(\mathcal{U}_p \oplus a)$ is ϵ-far from the uniform distribution over $\mathbb{F}_p$. Then for every $\delta > 0$ there exists a function $g : \{0,1\}^n \to \mathbb{F}_p$ such that*

$$\Pr_{x \in \mathcal{U}_p \oplus a}[g(x) \neq f(x)] \leq \delta$$

and

$$\textit{bit-rank}_{d-1}(g) \leq c + p^c$$

where[5] $c = C((p-1)d, \delta\epsilon^2/p^3)$.

[5]The function $C(\cdot,\cdot)$ is defined in the statement of Theorem 7.

We sketch the proof of Theorem 12 in the next subsection.

We also need the following lemma, which shows that if a degree d polynomial $f(x)$ can be approximated, under some shift of the p-biased distribution, by a function with a low $(d-1)$-bit-rank, then bit-pseudorandom distributions for degree $d-1$ polynomials also fool f.

Lemma 13. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a degree $d \geq 2$ polynomial. Assume that there is a function $g : \{0,1\}^n \to \mathbb{F}_p$ such that bit-rank$_{d-1}(g) = k$ and for some $a \in \{0,1\}^n$ it holds that*

$$\Pr_{x \in \mathcal{U}_p \oplus a}[f(x) \neq g(x)] \leq \delta.$$

Let $\mathcal{D} \subset \{0,1\}^n$ be a bit-pseudorandom distribution for degree $d-1$ polynomials with error ϵ. Then $f(\mathcal{D})$ and $f(\{0,1\}^n)$ are $(c_1^k\epsilon + c_2\delta)$-close, for $c_1 = p^{2^{(p-1)d}}$ and $c_2 = 4p \cdot 2^{(p-1)d}$.

Lemma 13 is proved as follows. The first step in the proof is showing that if f is a degree d polynomial which can be approximated by a function g of low $(d-1)$-bit-rank, then there is a *distribution* on functions H, such that every function in the support of H has a low $(d-1)$-bit-rank and such that for every $x \in \mathbb{F}_p^n$ it holds that $\Pr_{h \in H}[f(x) = h(x)] \geq 1 - \delta$. That is, we move from one function that computes f on most of the space to a distribution that is 'good' for every point x. The main idea behind the proof of this step is to use the self-correction properties of low degree polynomials. This step is the main technical part of the proof. Given the distribution H, the remainder of the proof of Lemma 13 is rather straightforward. We show that if a function has a low $(d-1)$-bit-rank then any bit-pseudorandom distribution for degree $d-1$ polynomials fools it, and then show that if a function can be approximated by a distribution on functions that have low $(d-1)$-bit-rank (as achieved in the 1st step) then this function is also fooled by bit-pseudorandom distributions for degree $d-1$ polynomials.

Theorem 5 now follows immediately from the combination of Claim 11, Theorem 12 and Lemma 13.

B. Proof overview of Theorem 12

We now explain the idea behind the proof of Theorem 12. Bogdanov and Viola proved that if $f(x)$ is a degree d polynomial over $\mathbb{F}_p$ such that $f(\mathbb{F}_p^n)$ is far from the uniform distribution over $\mathbb{F}_p$, then f can be well-approximated by a function of a few polynomials of lower degree [8]. Following this motivating example, we would like to prove that if $f(\mathcal{U}_p)$ is far from uniform then f can be well-approximated over $\mathcal{U}_p$ by a function

of few lower degree polynomials. However, the case of $f(\mathbb{F}_p^n)$ being far from uniform is easy to handle via directional derivatives, as the input space is invariant under shifts (i.e. the mapping $x \to x+y$ for $y \in \mathbb{F}_p^n$ maps the uniform distribution over $\mathbb{F}_p^n$ to itself). In our case, the input distribution $\mathcal{U}_p$ is not invariant under shifts, which creates a major obstacle for using existing techniques.

To overcome this obstacle we first 'complete' f to a polynomial over $\mathbb{F}_p^n$ that carries similar properties: For $a \in \{0,1\}^n$ define $f^{\oplus a}(x) = f(x^{p-1} \oplus a)$. Then $f^{\oplus a}$ is a polynomial of degree $d' = (p-1)d$ and the distributions $f^{\oplus a}(\mathbb{F}_p^n)$ and $f(\mathcal{U}_p \oplus a)$ are identical. We show that as f is non-regular, there exists $a \in \{0,1\}^n$ such that $f^{\oplus a}$ is biased. Similarly to [8] we get that $f^{\oplus a}$ can be approximated by a few of its *directional derivatives*, where the directional derivative of $f^{\oplus a}$ in direction $y \in \mathbb{F}_p^n$ is defined as $f_y^{\oplus a}(x) = f^{\oplus a}(x+y) - f^{\oplus a}(x)$. However, in our case we need a stronger property to hold. Define the *support* of y to be the set of nonzero entries in y, $\mathrm{Supp}(y) = \{i \in [n] : y_i \neq 0\}$. We would like to show that $f^{\oplus a}$ can be approximated by a few directional derivatives having small supports. To obtain this we need Theorem 7 that shows that most of the Fourier weight of $f^{\oplus a}$ is supported on coefficients that intersect a relatively small set S. Using this theorem we get

Claim 14 (informal statement)**.** *Let $\widetilde{f}$ be a polynomial over $\mathbb{F}_p$ of degree d' which is biased. For every $\delta > 0$ there exist a small number of directions $y_1, \ldots, y_k \in \mathbb{F}_p^n$ such that $|\mathrm{Supp}(y_1) \cup \ldots \cup \mathrm{Supp}(y_k)|$ is small, and such that $\widetilde{f}$ can be well-approximated by some function Γ of $\widetilde{f}_{y_1}, \ldots, \widetilde{f}_{y_k}$. Namely,*

$$\Pr_{x \in \mathbb{F}_p^n}[\widetilde{f}(x) \neq \Gamma(\widetilde{f}_{y_1}(x), \ldots, \widetilde{f}_{y_k}(x))] \leq \delta.$$

Claim 14 is proved as follows. Let $\widetilde{f}$ be a biased polynomial, say $\left|\mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{\widetilde{f}(x)}]\right| = \tau > 0$, where $\omega = e^{2\pi i/p}$ is a primitive p-root of unity. Any biased polynomial can be computed by a function of its derivatives,

$$\omega^{\widetilde{f}(x)} = \tau' \cdot \mathbb{E}_{y \in \mathbb{F}_p^n}\left[\omega^{-\widetilde{f}_y(x)}\right],$$

where τ' is appropriately chosen such that $|\tau'| = |\tau|^{-1}$. This follows as for any $x \in \mathbb{F}_p^n$, the value of $\mathbb{E}_{y \in \mathbb{F}_p^n}[\omega^{\widetilde{f}(x+y)}] = \mathbb{E}_{y \in \mathbb{F}_p^n}[\omega^{\widetilde{f}(y)}]$ is independent of x and is not too close to zero. This exact computation can be transformed into an approximation by a few derivatives by sampling independently a few derivatives $y_1, \ldots, y_k \in \mathbb{F}_p^n$, and then applying a standard Chernoff argument to deduce that

$$\omega^{\widetilde{f}(x)} \approx \tau' \cdot \frac{1}{k} \sum_{i=1}^{k} \left[\omega^{-\widetilde{f}_{y_i}(x)}\right].$$

From this is is easy to conclude that $\widetilde{f}$ can be approximated by a function of $\widetilde{f}_{y_1}, \ldots, \widetilde{f}_{y_k}$. The problem with this approach is that the derivatives do not have to be sparse, which is necessary for our proof. In order to overcome this we apply Theorem 7 to the polynomial $\widetilde{f}$. We get that there is a small set of variables $S \subset [n]$ such that most of the Fourier coefficients of $\widetilde{f}$ involve some variable from S. By choosing parameters accordingly we can get that

$$\sum_{\alpha \in \mathbb{F}_p^n : \alpha \neq 0, \alpha_S = 0} |\widehat{\widetilde{f}}(\alpha)|^2 \ll |\tau|.$$

As a corollary, we get that for most x, the value of $\mathbb{E}_{y \in \mathbb{F}_p^S}[\omega^{\widetilde{f}(x+y)}]$ is very close to the global average $\mathbb{E}_{y \in \mathbb{F}_p^n}[\omega^{\widetilde{f}(y)}]$ which does not depend on x. From this we deduce that we can in fact approximate $\widetilde{f}(x)$ very well by averaging only over derivatives in directions supported on S,

$$\omega^{\widetilde{f}(x)} \approx \tau'' \cdot \mathbb{E}_{y \in \mathbb{F}_p^S}\left[\omega^{-\widetilde{f}_y(x)}\right],$$

where $|\tau''| \approx |\tau'|$. Thus, $\widetilde{f}(x)$ can be well approximated by a function of $\{\widetilde{f}_y(x) : y \in \mathbb{F}_p^S\}$.

Thus, applying Claim 14 for $\widetilde{f} = f^{\oplus a}$, we get that $f^{\oplus a}$ can be well approximated by a function of a small number of its sparse derivatives. This is still not enough as the derivatives of $f^{\oplus a}$ have degree $(p-1)d-1$. However, we further show that sparse directional derivatives of $f^{\oplus a}$ can be calculated by directional derivatives of f and a few variables.

Claim 15 (informal statement)**.** *Any directional derivative $f_y^{\oplus a}(x)$, such that $\mathrm{Supp}(y) \subseteq S$, can be computed by some function of $\{f_z(\cdot)\}_{\mathrm{Supp}(z) \subseteq S}$ and $\{x_i : i \in S\}$.*

We prove this claim by showing that any derivatives of $f^{\oplus a}$, with respect to a direction supported on S, satisfies $(f^{\oplus a})_y(x) = f_w(x^{p-1} \oplus a)$ for some w that is supported on S and depends only on y, a and x_S. Theorem 12 follows from Claims 14 and 15.

C. Proof of Theorem 7

In this section we give the proof of Theorem 7. We start by defining the notion of an S-correlated distribution over $\mathbb{F}_p^n$, for a subset $S \subset [n]$. We recall that for $x \in \mathbb{F}_p^n$ we denote by $x_S \in \mathbb{F}_p^S$ the restriction of x to coordinates in S, and we denote the complement of S by $\bar{S} = [n] \setminus S$.

Definition 16. Let $S \subset [n]$. The S-correlated distribution is a joint distribution over pairs $(X, Y) \in \mathbb{F}_p^n \times \mathbb{F}_p^n$ defined as follows. Choose $X_{\bar{S}} = Y_{\bar{S}}$ uniformly in $\mathbb{F}_p^{\bar{S}}$, and choose independently and uniformly $X_S, Y_S \in \mathbb{F}_p^S$. We denote the S-correlated distribution (X, Y) by $\mathcal{D}_S$. For $f, g : \mathbb{F}_p^n \to \mathbb{F}_p$ and $S \subset [n]$, we define the S-correlation of f and g to be

$$\Delta_S(f, g) = \sum_{\alpha \in \mathbb{F}_p^n : \alpha_S = 0, \alpha \neq 0} \widehat{f}(\alpha)\overline{\widehat{g}(\alpha)} \ .$$

Note that an equivalent definition of $\mathcal{D}_S$ is to first sample $X \in \mathbb{F}_p^n$ uniformly, then to set $Y = X$ and finally to resample Y_S. We now restate Theorem 7 in terms of Δ_S.

Theorem 17 (Theorem 7, restated)**.** *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a degree d polynomial. For every $\epsilon > 0$ there exists $S \subset [n]$, of size $|S| \leq C(d, \epsilon) = O(1/\epsilon)^{O(4^d)}$, such that $\Delta_S(f, f) \leq \epsilon$.*

Before giving the formal proof we explain the idea behind it. We will prove the theorem by induction on the degree. The case of linear polynomials will be easy to handle by a direct calculation. For a general degree d we will use the following useful claims.

Claim 18. *Let A be any linear subspace of $\mathbb{F}_p^n$. For every $f : \mathbb{F}_p^n \to \mathbb{F}_p$ and $S \subset [n]$ it holds that $\Delta_S(f, f)^2 \leq \mathbb{E}_{a \in A}[\Delta_S(f_a, f_a)] + \mathbb{E}_{a \in A}[|\widehat{f_a}(0)|^2]$.*

Claim 19. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$. Let A be a random linear subspace of $\mathbb{F}_p^n$ of dimension r (i.e. A is picked at random amongst all r-dimensional subspaces of $\mathbb{F}_p^n$). Then*

$$\mathbb{E}_A \left[\mathbb{E}_{a \in A}[|\widehat{f_a}(0)|^2] \right] \leq \frac{1}{p^r} + \max_{\alpha} |\widehat{f}(\alpha)|^2 \ ,$$

where $\mathbb{E}_A$ means averaging over a random choice of A.

These claims indicate that we have to consider two cases.

Case 1. All the Fourier coefficients of f are small: In this case, the claims above imply that if we set r to a large enough value and pick a random r-dimensional subspace A then setting S be the union of the corresponding sets for f_a, for $a \in A$, we get the required result (using the induction hypothesis).

Case 2. Some Fourier coefficient of f is large: In this case we first approximate f by a function of a small number of (linear shifts of) its partial derivatives. A simple calculation then gives that for some k, δ^* and σ we have

$$\Delta_S(f, f) \leq \frac{1}{k\delta^*} \sum_{i=1}^{k} |\Delta_S(\widetilde{h_{y_i}}, f)| + 2\sigma \ ,$$

where $\{\widetilde{h_{y_i}}\}_{i=1}^k$ is a set of (shifted) derivatives used to approximate f. Observing that for any g and $S \subseteq S'$ it holds that

$$|\Delta_{S'}(f, g)| \leq (\Delta_S(f, f))^{1/2} (\Delta_S(g, g))^{1/2} \ ,$$

we complete the proof for this case as well by picking S' to be the union of the corresponding sets for the polynomials $\widetilde{h_{y_i}}$.

D. Proofs of two useful claims

Following the proof outline above we start by proving Claims 18 and 19. As a first step we prove the following lemma.

Lemma 20. *Let $f, g : \mathbb{F}_p^n \to \mathbb{F}_p$. Then for any $S \subset [n]$ it holds that $\Delta_S(f, g) =$*

$$\mathbb{E}_{(x,y) \in \mathcal{D}_S}[\omega^{f(x) - g(y)}] - \mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{f(x)}]\overline{\mathbb{E}_{y \in \mathbb{F}_p^n}[\omega^{g(y)}]}$$

and for every $S' \supseteq S$ it holds that

$$|\Delta_{S'}(f, g)| \leq (\Delta_S(f, f))^{1/2} (\Delta_S(g, g))^{1/2} \ .$$

Proof: Recall that $\hat{f}(0) = \mathbb{E}[\omega^{f(x)}]$ and similarly for g. Let $\mu = \sum_{\alpha : \alpha_S = 0} \widehat{f}(\alpha)\overline{\widehat{g}(\alpha)}$. Calculating we get,

$$\mu = \sum_{\alpha : \alpha_S = 0} (\mathbb{E}_x \omega^{f(x)} \omega^{-\langle x, \alpha \rangle})(\mathbb{E}_y \omega^{-g(y)} \omega^{\langle y, \alpha \rangle}).$$

Further simplification gives

$$\mu = \frac{1}{p^{2n}} \sum_{x,y} \omega^{f(x) - g(y)} \sum_{\alpha : \alpha_S = 0} \omega^{\langle y - x, \alpha \rangle}.$$

Simple manipulation of the right hand side of the above equation shows,

$$\begin{aligned} \mu &= \frac{1}{p^{2n}} \sum_{x_{\bar{S}} = y_{\bar{S}}} p^{n - |S|} \omega^{f(x) - g(y)} \\ &= \mathbb{E}_{(x,y) \in \mathcal{D}_S}[\omega^{f(x) - g(y)}] \ . \end{aligned}$$

Hence, $\Delta_S(f, g) = \mathbb{E}_{(x,y) \in \mathcal{D}_S}[\omega^{f(x) - g(y)}] - \widehat{f}(0)\overline{\widehat{g}(0)}$. To show the second claim we apply the Cauchy-Schwarz inequality,

$$|\Delta_{S'}(f, g)| = \left| \sum_{\alpha \neq 0, \alpha_{S'} = 0} \widehat{f}(\alpha)\overline{\widehat{g}(\alpha)} \right| \leq$$

$$\left(\sum_{\alpha \neq 0, \alpha_{S'} = 0} |\widehat{f}(\alpha)|^2 \right)^{1/2} \left(\sum_{\alpha \neq 0, \alpha_{S'} = 0} |\widehat{g}(\alpha)|^2 \right)^{1/2} \leq$$

$$\left(\sum_{\alpha\neq 0,\alpha_S=0}|\widehat{f}(\alpha)|^2\right)^{1/2}\left(\sum_{\alpha\neq 0,\alpha_S=0}|\widehat{g}(\alpha)|^2\right)^{1/2}=$$
$$(\Delta_S(f,f))^{1/2}(\Delta_S(g,g))^{1/2}\ .$$

■

We now give the proofs of Claims 18 and 19.

Proof of Claim 18: By Lemma 20 we have

$$\Delta_S(f,f)=\mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{f(x)-f(y)}]-|\widehat{f}(0)|^2\leq$$
$$\mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{f(x)-f(y)}]\ .$$

For any fixed $a\in A$, the distribution $\{(x+a,y+a):(x,y)\in\mathcal{D}_S\}$ is identical to $\mathcal{D}_S$. So we can express $\Delta_S(f,f)$ as follows,

$$\Delta_S(f,f)\leq\mathbb{E}_{a\in A}\mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{f(x+a)-f(y+a)}]\ .$$

Applying the Cauchy-Schwarz inequality (and using the fact that A is a linear subspace) we get

$$\Delta_S(f,f)^2\leq\mathbb{E}_{(x,y)\in\mathcal{D}_S}\left[|\mathbb{E}_{a\in A}[\omega^{f(x+a)-f(y+a)}]|^2\right]$$

Then using the standard trick of handling the squared term we get that the left hand side of the above inequality is

$$\leq\mathbb{E}_{a,a'\in A}\mathbb{E}_{(x,y)\in\mathcal{D}_S}\left[\omega^{f(x+a)-f(x+a')}\omega^{f(y+a')-f(y+a)}\right]$$
$$=\mathbb{E}_{a,a'\in A}\mathbb{E}_{(x',y')\in\mathcal{D}_S}\left[\omega^{f(x'+a-a')-f(x')}\omega^{f(y')-f(y'+a-a')}\right]$$

Simplifying the right hand side further we get,

$$\Delta_S(f,f)^2\leq\mathbb{E}_{a\in A}\mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{f_a(x)-f_a(y)}]$$
$$=\mathbb{E}_{a\in A}[\Delta_S(f_a,f_a)+|\widehat{f_a}(0)|^2]\ .$$

■

Proof of Claim 19: We begin by showing an identity on $\mathbb{E}_{a\in A}[|\widehat{f_a}(0)|^2]$, for any subspace A.

Claim 21. *For any function $f:\mathbb{F}_p^n\to\mathbb{F}_p$ and any subspace $A\subset\mathbb{F}_p^n$*

$$\mathbb{E}_{a\in A}[|\widehat{f_a}(0)|^2]=\sum_{\beta\in\mathbb{F}_p^n,\gamma\in A^\perp}|\widehat{f}(\beta)|^2|\widehat{f}(\beta+\gamma)|^2,$$

where $A^\perp$ is the dual space of A.

Proof: Using the Fourier decomposition formula, the R.H.S of the above expression is

$$\sum_{\beta\in\mathbb{F}_p^n,\gamma\in A^\perp}(\mathbb{E}_{x,x'\in\mathbb{F}_p^n}[\omega^{f(x)-f(x')}\omega^{\langle\beta,x'-x\rangle}])\cdot$$
$$(\mathbb{E}_{y,y'\in\mathbb{F}_p^n}[\omega^{f(y)-f(y')}\omega^{\langle\beta+\gamma,y'-y\rangle}])$$

which is equivalent to

$$\sum_{\gamma\in A^\perp}\mathbb{E}_{x,x',y,y'\in\mathbb{F}_p^n}\left[\omega^{f(x)-f(x')+f(y)-f(y')}\omega^{\langle\gamma,y'-y\rangle}\cdot\right.$$
$$\left.\sum_{\beta\in\mathbb{F}_p^n}\omega^{\langle\beta,x'-x+y'-y\rangle}\right]\ .$$

Considering the inner sum over β, the above expression can be simplified as

$$\frac{1}{p^{3n}}\sum_{x-x'=y'-y}\omega^{f(x)-f(x')+f(y)-f(y')}\sum_{\gamma\in A^\perp}\omega^{\langle\gamma,y'-y\rangle}\ .$$

Now the inner sum over γ is nonzero only when $y'-y\in A$. Denote $a=y'-y\in A$. Recalling that we sum over $x-x'=y'-y=a$, we can further simplify the above expression as

$$\frac{|A^\perp|}{p^{3n}}\sum_{a\in A}\sum_{x',y\in\mathbb{F}_p^n}\omega^{f(x'+a)-f(x')+f(y)-f(y+a)}$$
$$=\mathbb{E}_{a\in A}[|\widehat{f_a}(0)|^2]\ .$$

■

We now have that

$$\mathbb{E}_{a\in A}[|\widehat{f_a}(0)|^2]=\sum_{\beta\in\mathbb{F}_p^n,\gamma\in A^\perp}|\widehat{f}(\beta)|^2|\widehat{f}(\beta+\gamma)|^2=$$
$$\sum_{\beta\in\mathbb{F}_p^n,\alpha\in\mathbb{F}_p^n}|\widehat{f}(\beta)|^2|\widehat{f}(\alpha)|^2\chi_{A^\perp}(\alpha-\beta)\ ,$$

where $\chi_{A^\perp}$ is the characteristic function of $A^\perp$. Let A be a random subspace of dimension r. The probability for $\alpha\neq\beta$ that $(\alpha-\beta)\in A^\perp$ is $1/p^r$. Since $\sum_\alpha|\hat{f}(\alpha)|^2=1$ by Parseval's identity, we obtain that

$$\mathbb{E}_A\left[\mathbb{E}_{a\in A}[|\widehat{f_a}(0)|^2]\right]=$$
$$\sum_{\beta\neq\alpha\in\mathbb{F}_p^n}|\widehat{f}(\beta)|^2|\widehat{f}(\alpha)|^2\mathbb{E}_A[\chi_{A^\perp}(\alpha-\beta)]+\sum_{\alpha\in\mathbb{F}_p^n}|\widehat{f}(\alpha)|^4$$
$$\leq\frac{1}{p^r}+\sum_{\alpha\in\mathbb{F}_p^n}|\widehat{f}(\alpha)|^4\leq\frac{1}{p^r}+\max_\alpha|\widehat{f}(\alpha)|^2\ .$$

■

E. Proof of Theorem 17

The proof is by induction on d. The base case is $d=1$. Let $f(x)=\sum_{i=1}^n a_ix_i$ be any linear polynomial. Consider $S=\{i\}$ such that $a_i\neq 0$. Then for any $\alpha\in\mathbb{F}_p^n$ such that $\alpha_S=0$ we get $\widehat{f}(\alpha)=\mathbb{E}_{x_i\in\mathbb{F}_p}[\omega^{a_ix_i}]\prod_{j\neq i}\mathbb{E}_{x_j\in\mathbb{F}_p}[\omega^{(a_j-\alpha_j)x_j}]=0$. Hence, $\sum_{\alpha:\alpha_S=0}|\widehat{f}(\alpha)|^2=0$ and the claim is proved.

By induction hypothesis, let the result be true for any degree $\leq d-1$ polynomial. As outlined above, the proof proceeds by considering two cases, whether f has some large Fourier coefficient or not.

Case 1: Assume that $|\widehat{f}(\alpha)| \leq \delta^*$, for all $\alpha \in \mathbb{F}_p^n$, for an appropriate choice of δ^* (that we will suitably fix later). Let $\epsilon_d = \epsilon$. By Claim 21 we get that for any $S \subset [n]$ and a subspace $A \subseteq \mathbb{F}_p^n$

$$\Delta_S(f,f)^2 \leq \mathbb{E}_{a\in A}[\Delta_S(f_a,f_a)] + \mathbb{E}_{a\in A}[|\widehat{f_a}(0)|^2] \ .$$

Notice that for each $a \in A$, $\deg f_a \leq d-1$. Hence, by induction hypothesis, for each $a \in A$, there exist S_a of size $C(d-1,\epsilon_{d-1})$ such that $\Delta_{S_a}(f_a,f_a) \leq \epsilon_{d-1}$ (for some ϵ_{d-1} that will be soon determined). Let A be a linear subspace of dimension r that minimizes $\mathbb{E}_{a\in A}[|\widehat{f_a}(0)|^2]$. Let $S = \cup_{a\in A} S_a$. Claim 19 implies that

$$\Delta_S(f,f)^2 \leq \epsilon_{d-1} + \frac{1}{p^r} + \max_\alpha |\widehat{f}(\alpha)|^2 \ .$$

Now it is enough to choose r, ϵ_{d-1} and δ^* such that $\epsilon_{d-1} + \frac{1}{p^r} + (\delta^*)^2 \leq \epsilon_d^2$. Also, notice that $|S| = C(d,\epsilon_d) \leq p^r C(d-1,\epsilon_{d-1})$.

Case 2: Let β be a Fourier coefficient such that $|\widehat{f}(\beta)| \geq \delta^*$. Set $\delta = \overline{\widehat{f}(\beta)}$. Let $h(x) = f(x) - \langle x,\beta\rangle$. Then the bias of $-h(x)$ is $\mathbb{E}_{x\in\mathbb{F}_p^n}[\omega^{-h(x)}] = \delta$. Notice that for every $x \in \mathbb{F}_p^n$ we have $\omega^{h(x)}\mathbb{E}_y[\omega^{-h(x+y)}] = \mathbb{E}_y[\omega^{-h_y(x)}]$. As for every fixed x we have $\mathbb{E}_y[\omega^{-h(x+y)}] = \delta$ it is clear that we can get the following decomposition of $f(x)$

$$\omega^{f(x)} = \omega^{\langle x,\beta\rangle}\cdot\omega^{h(x)} = \omega^{\langle x,\beta\rangle}\cdot\frac{1}{\delta}\mathbb{E}_y[\omega^{-h_y(x)}] = \frac{1}{\delta}\mathbb{E}_y[\omega^{\langle x,\beta\rangle - h_y(x)}] \ .$$

Define $\widetilde{h_y}(x) = \langle x,\beta\rangle - h_y(x)$. Notice that since $h(x)$ has degree $d \geq 2$ then $\deg(\widetilde{h_y}) \leq d-1$. Now, if we sample enough y's uniformly and independently at random, and take the average of the corresponding $\omega^{\widetilde{h_y}(x)}$, then we get a good estimate of $\omega^{f(x)}$. Fix a parameter $\sigma \in (0,1)$ (to be determined later), using Chebyshev's inequality we find k such that the following holds

$$\mathbb{E}_{x,y_1,\ldots,y_k\in\mathbb{F}_p^n}[|\omega^{f(x)} - \frac{1}{\delta k}\sum_{i=1}^k \omega^{\widetilde{h_{y_i}}(x)}|] \leq \sigma.$$

Claim 22. *To get an approximation $\mathbb{E}_{x,y_1,\ldots,y_k\in\mathbb{F}_p^n}[|\omega^{f(x)} - \frac{1}{\delta k}\sum_{i=1}^k \omega^{\widetilde{h_{y_i}}(x)}|] \leq \sigma$, it is enough to take $k = O(|\delta|^{-3}\sigma^{-3})$.*

Proof: It is enough to choose k such that $\mathbb{E}[|\mathrm{Re}(\omega^{f(x)} - \frac{1}{\delta k}\sum_{i=1}^k \omega^{\hat{h}_{y_i}(x)})|] \leq \sigma/2$, and $\mathbb{E}[|\mathrm{Img}(\omega^{f(x)} - \frac{1}{\delta k}\sum_{i=1}^k \omega^{\hat{h}_{y_i}(x)})|] \leq \sigma/2$. Let $Y_i = \mathrm{Re}(\frac{1}{\delta}\omega^{\widetilde{h_{y_i}}(x)})$. Then $\mathbb{E}_{y_i}[Y_i] = \mathrm{Re}(\omega^{f(x)})$. It is clear that $\mathrm{Var}(Y_i) \leq \frac{1}{|\delta|^2}$. Hence, by Chebyshev's inequality we get that

$$\Pr(|\mathrm{Re}(\omega^{f(x)}) - \frac{1}{k}\sum_{i=1}^k Y_i| \geq \frac{\sigma}{4}) \leq \frac{16}{|\delta|^2 k\sigma^2} \ .$$

Therefore, as always $|\mathrm{Re}(\omega^{f(x)}) - \frac{1}{k}\sum_{i=1}^k Y_i| \leq 1 + \delta^{-1} \leq 2\delta^{-1}$ we get that $\mathbb{E}[|\mathrm{Re}(\omega^{f(x)}) - \frac{1}{k}\sum_{i=1}^k Y_i|] \leq \sigma/2$ for $k \geq \frac{128}{|\delta|^3\sigma^3}$. The imaginary part can be approximated similarly. ∎

Fix $\{y_i\}_{i\in[k]}$ in such a way that $\mathbb{E}_{x\in\mathbb{F}_p^n}[|\omega^{f(x)} - \frac{1}{\delta k}\sum_{i=1}^k \omega^{\widetilde{h_{y_i}}(x)})|] \leq \sigma$. Let $F(x) = \frac{1}{k\delta}\sum_{i=1}^k \omega^{\widetilde{h_{y_i}}(x)}$. As $\mathbb{E}_{x\in\mathbb{F}_p^n}[|\omega^{f(x)} - F(x)|] \leq \sigma$ we can upper bound $\Delta_S(f,f)$ as follows

$$\Delta_S(f,f) =$$

$$\mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{f(x)-f(y)}] - \mathbb{E}_{x\in\mathbb{F}_p^n}[\omega^{f(x)}]\cdot\overline{\mathbb{E}_{y\in\mathbb{F}_p^n}[\omega^{f(y)}]} \leq$$

$$\Big|\mathbb{E}_{(x,y)\in\mathcal{D}_S}[(\omega^{f(x)} - F(x))\omega^{-f(y)}] -$$

$$\mathbb{E}_{x\in\mathbb{F}_p^n}[\omega^{f(x)} - F(x)]\cdot\overline{\mathbb{E}_{y\in\mathbb{F}_p^n}[\omega^{f(y)}]}\Big| +$$

$$\Big|\mathbb{E}_{(x,y)\in\mathcal{D}_S}[F(x)\omega^{-f(y)}] - \mathbb{E}_{x\in\mathbb{F}_p^n}[F(x)]\cdot\overline{\mathbb{E}_{y\in\mathbb{F}_p^n}[\omega^{f(y)}]}\Big| \leq$$

$$2\sigma + \Big|\mathbb{E}_{(x,y)\in\mathcal{D}_S}[F(x)\omega^{-f(y)}] - (\mathbb{E}_x[F(x)])(\mathbb{E}_y[\omega^{-f(y)}])\Big| \leq$$

$$2\sigma + \frac{1}{k\delta}\sum_{i=1}^k \Big|\mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{\widetilde{h_{y_i}}(x)-f(y)}] -$$

$$(\mathbb{E}_x[\omega^{\widetilde{h_{y_i}}(x)}])(\mathbb{E}_y[\omega^{-f(y)}])\Big| \leq$$

$$2\sigma + \frac{1}{k\delta^*}\sum_{i=1}^k |\Delta_S(\widetilde{h_{y_i}},f)|.$$

As $\deg(\widetilde{h_{y_i}}) \leq d-1$ we get, by the induction hypothesis, that for each $\widetilde{h_{y_i}}$ there exists a set S_i, of size $C(d-1,\epsilon_{d-1})$, such that $\Delta_{S_i}(\widetilde{h_{y_i}},\widetilde{h_{y_i}}) \leq \epsilon_{d-1}$. Consider $S = \cup_{i=1}^k S_i$. Obviously, $|S| \leq kC(d-1,\epsilon_{d-1})$. Lemma 20 implies that

$$|\Delta_S(\widetilde{h_{y_i}},f)| \leq (\Delta_{S_i}(\widetilde{h_{y_i}},\widetilde{h_{y_i}}))^{1/2}(\Delta_{S_i}(f,f))^{1/2} \leq$$
$$(\Delta_{S_i}(\widetilde{h_{y_i}},\widetilde{h_{y_i}}))^{1/2} \leq \epsilon_{d-1}^{1/2} \ .$$

In order to achieve $\Delta_S(f,f) \leq \epsilon_d$ we need to fix the parameters $\delta^*, \epsilon_{d-1}, k, \sigma$ so that $\frac{1}{\delta^*}\epsilon_{d-1}^{1/2} + 2\sigma \leq \epsilon_d$.

We now show how to pick the parameters adequately. We need to satisfy both $\epsilon_{d-1} + \frac{1}{p^r} + (\delta^*)^2 \leq \epsilon_d^2$ and $\frac{1}{\delta^*}\epsilon_{d-1}^{1/2} + 2\sigma \leq \epsilon_d$. Fix $\sigma = \frac{\epsilon_d}{4}$ and $\delta^* = \frac{\epsilon_d}{2}$. Then it is enough to choose $\epsilon_{d-1} = O(\epsilon_d^4)$ and $r = \log_p(\epsilon_d^2/4)$. We now estimate $|S|$. Recall that $|S| \leq \max(p^r, k)C(d-1, \epsilon_{d-1})$ where $k = O(|\delta^*|^{-3}\sigma^{-3})$. This yields the following bound

$$|S| \leq O(\epsilon_d^{-6})C(d-1, \Omega(\epsilon_d^4))$$

Solving the recurrence for $C(d, \epsilon)$ we get that $C(d, \epsilon) \leq O(\epsilon)^{O(4^d)}$. This completes the proof of Theorem 17.

III. Conclusions and open problems

We construct efficient and explicit bit-pseudorandom generators for constant degree polynomials over finite fields. These yield pseudorandom generators for $\mathrm{CC}^0[p]$ which achieve any small constant error while using only $O(\log n)$ random bits. The proof is based on a new characterization of the Fourier spectrum of low degree polynomials over finite fields.

We state several open problems.

- Construct pseudorandom generators for $\mathrm{AC}^0[p]$. The next step, following this work, is to construct pseudorandom generators for sparse polynomials over $\mathbb{F}_p$ (i.e. polynomials of degree $O(\log n)$ with only a polynomial number of monomials). Any such polynomial can be realized by a depth-2 $\mathrm{AC}^0[p]$ circuit.
- Generalize our results for $\mathrm{CC}^0[m]$ for composite m. As a first step, generalize our results for bit-pseudorandom generators for low degree polynomials over $\mathbb{Z}_m$. The result of [2] constructs a bit-pseudorandom generator for linear forms over $\mathbb{Z}_m$ using $O(\log n)$ random bits.
- Improve the parameters of Theorem 7. For $d = 1$ it is an easy observation that a set S of size $|S| = 1$ suffices. For $d = 2$, it is not difficult to see that all nonzero Fourier coefficients of a quadratic polynomial form an affine space and have the same absolute value. Using this observation one can get a set of size $|S| = O(\log 1/\epsilon)$. We do not have any example of a constant degree polynomial requiring sets of size $\omega(\log 1/\epsilon)$.
- Improve the dependence of the seed length on ϵ in Theorem 5. Currently, the seed length is logarithmic in n but a tower of height $O(d)$ in $1/\epsilon$.

References

[1] M. Luby, B. Velickovic, and A. Wigderson, "Deterministic approximate counting of depth-2 circuits," in *Proceedings of the* 2^{nd} *ISTCS*, 1993, pp. 18–24.

[2] S. Lovett, O. Reingold, L. Trevisan, and S. Vadhan, "Pseudorandom bit generators that fool modular sums," in *Proceedings of the* 13^{th} *RANDOM*, 2009, pp. 615–630.

[3] R. Meka and D. Zuckerman, "Small-bias spaces for group products," in *Proceedings of the* 13^{th} *RANDOM*, 2009, pp. 658–672.

[4] O. Goldreich, *Computational Complexity: A Conceptual Perspective*. Cambridge University Press, 2008.

[5] S. Arora and B. Barak, *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[6] J. Naor and M. Naor, "Small-bias probability spaces: Efficient constructions and applications." *SIAM J. on Computing*, vol. 22, no. 4, pp. 838–856, 1993.

[7] N. Alon, O. Goldreich, J. Håstad, and R. Peralta, "Simple construction of almost k-wise independent random variables." *Random Structures and Algorithms*, vol. 3, no. 3, pp. 289–304, 1992.

[8] A. Bogdanov and E. Viola, "Pseudorandom bits for polynomials," in *Proceedings of the* 48^{th} *FOCS*, 2007.

[9] S. Lovett, "Unconditional pseudorandom generators for low degree polynomials," in *Proceedings of the* 40^{th} *STOC*, 2008, pp. 557–562.

[10] E. Viola, "The sum of d small-bias generators fools polynomials of degree d," *Computational Complexity*, vol. 18, no. 2, pp. 209–217, 2009.

[11] J. Håstad, "Computational limitations for small-depth circuits," Ph.D. dissertation, MIT, 1986.

[12] N. Nisan, "Pseudorandom bits for constant depth circuits," *Combinatorica*, vol. 11, no. 1, pp. 63–70, 1991.

[13] L. M. J. Bazzi, "Polylogarithmic independence can fool DNF formulas," in *Proceedings of the* 48^th *FOCS*, 2007, pp. 63–73.

[14] M. Braverman, "Poly-logarithmic independence fools AC_0 circuits," in *Proceedings of the* 24^{th} *CCC*, 2009.

[15] N. Linial and N. Nisan, "Approximate inclusion-exclusion," *Combinatorica*, vol. 10, pp. 349–365, 1990.

[16] A. A. Razborov, "Lower bounds on the size of bounded depth circuits over a complete basis with logical addition," *Math. Notes*, vol. 41, no. 4, pp. 333–338, 1987.

[17] R. Smolensky, "Algebraic methods in the theory of lower bounds for Boolean circuit complexity," in *Proceedings of the* 19^{th} *STOC*, 1987, pp. 77–82.