Electronic Colloquium on Computational Complexity, Report No. 33 (2010)

Pseudorandom generators for $\mathrm{CC}_0[p]$ and the Fourier spectrum of low-degree polynomials over finite fields

Shachar Lovett [*] Partha Mukhopadhyay[†] Amir Shpilka[‡]

March 5, 2010

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 [LVW93] or that could only fool the distribution generated by linear functions over $\mathbb{F}_p$, when evaluated on the Boolean cube [LRTV09, MZ09].

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, up to error δ, by a function of a small number (depending only on ϵ, δ and d) of lower degree polynomials.

[*] The Weizmann Institute of Science, shachar.lovett@weizmann.ac.il. Research supported by the Israel Science Foundation (grant 1300/05).

[†] Technion - Israel Institute of Technology, partha@cs.technion.ac.il. Research supported in part at the Technion by an Aly Kaufman Fellowship.

[‡] Technion - Israel Institute of Technology, shpilka@cs.technion.ac.il. Research supported by the Israel Science Foundation (grant 439/06).

ISSN 1433-8092

Contents

1 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 [Gol08, AB09]).

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 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 [NN93] who gave PRGs with $O(\log n)$ seed length. The distributions constructed in [NN93] are also known as ϵ-biased distributions. Alon et al. extended this construction to work over arbitrary finite fields [AGHP92]. In [LVW93] a pseudorandom generator for the class of bounded degree polynomials over finite fields was given.[2] The seed length of [LVW93] was not optimal and was later improved in a sequence of works [BV07, Lov08, Vio09]. Note that all these generators take as input vectors from $\mathbb{F}_p^r$ and output vectors in $\mathbb{F}_p^n$. In [LRTV09, MZ09] 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 $\{0,1\}^n$. 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

[1] One should think of D as either the Boolean cube $\{0,1\}^n$ or as $\mathbb{F}_p^n$.

[2] This is not explicitly stated in [LVW93], but it follows from their result for depth 2 circuits with a symmetric function at the top.

gates. AC_0 is probably the most intensively studied amongst classes of small-depth circuits. Håstad [Hås86] 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 [Nis91] for constructing efficient pseudorandom generators for AC_0 (these pseudorandom generators use $r = \text{polylog(n)}$ bits). Recently, following a breakthrough by Bazzi [Baz07], Braverman [Bra09] showed that any polylog-wise independent distribution is pseudorandom for AC_0 circuits, thus settling a conjecture of Linial and Nisan [LN90]. $\mathrm{AC}_0[p]$ 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 [Raz87, Smo87] 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, [Raz87, Smo87] showed that the MOD_q function has polynomially small correlation with $\mathrm{AC}_0[p]$ when q and q are co-prime. 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!

1.1 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 [Smo87], no pseudorandom generator better than the one constructed in [LVW93] (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 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 $\mathrm{CC}_0[p]$ we have to fool the distribution 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 [LRTV09, MZ09] 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 [LVW93, BV07, Lov08, Vio09]. The best result to date is by Viola.

Theorem 3 (Theorem 1 in [Vio09]). *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 construction bit-pseudorandom generators for linear polynomials (i.e. the case of $d = 1$) was first studied by [LRTV09, MZ09] in the context of small-space computations. Before describing their generator we need a few notations. For $a = (a_1, \dots, a_n) \in \mathbb{F}_p^n$ define $a^{p-1} = (a_1^{p-1}, \dots, a_n^{p-1}) \in \{0,1\}^n$ 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. [LRTV09, MZ09] 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 [LRTV09, MZ09]). *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 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 $|S| \leq C(d, \epsilon)$ 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, almost all nonzero Fourier coefficients of f intersect S.

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 $\text{bit-rank}_d(g)$, is the minimal number of degree d polynomials over $\mathbb{F}_p$ required to compute g over $\{0,1\}^n$. That is, $\text{rank}_d(g) = k$ 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 $x \in \{0,1\}^n$*

$$g(x) = (x_1 + \ldots + x_n)^2 - (x_1^2 + \ldots + x_n^2) = (x_1 + \ldots + x_n)^2 - (x_1 + \ldots + x_n) .$$

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.

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

Our second structural result is the following.

Theorem 9 (Structure of bit-biased polynomials)**.** *Let $f(x_1,\dots,x_n)$ be a degree d 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$_d(g) \leq p^{O(c)}$ where*[4] *$c = C((p-1)(d+1), \delta\epsilon^2/p^3)$.*

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.

1.2 Proof overview

Pseudorandom generators that fool low degree polynomials over $\mathbb{F}_p^n$ were obtained in [BV07, Lov08, Vio09]. 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. Let $f(x) = f(x_1,\dots,x_n)$ be a polynomial of degree d over $\mathbb{F}_p$. The base case of $d = 1$ was established in [LRTV09, MZ09], hence we assume from now on $d \geq 2$.

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, $x^{p-1} = (x_1^{p-1},\dots,x_n^{p-1})$ 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 $f(\mathcal{U}_p)$.

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 (Lemma 11, informal statement)**.** *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.*

Non-regular polynomials We now have to deal with non-regular polynomials, i.e polynomials that distinguish between uniform bits and the p-biased distribution. This is the main challenge we tackle in the paper. In fact, we will show that this property is so strong that bit-pseudorandom generators for degree $d-1$ polynomials with small enough error suffice to fool any such degree d polynomial. The proof consists of two steps. First we prove Theorem 13 (which is close in spirit to Theorem 9) that shows that f can be well-approximated, with respect to (an affine shift of) $\mathcal{U}_p$, by a few polynomials of degree $d-1$. We then prove that any distribution that fools degree $d-1$ polynomials (over $\{0,1\}^n$) also fools f (Lemma 14).

We now explain the idea behind the proof of Theorem 13. 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

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

$\mathbb{F}_p$, then f can be well-approximated by a function of a few polynomials of lower degree [BV07]. Following this motivating example, we would like to prove that if $f(\mathcal{U}_p)$ is far from uniform (a similar property can be easily obtained from the fact that f is not regular, see Claim 12) then f can be well-approximated over $\mathcal{U}_p$ by 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 $x \to x + a$ for $a \in \mathbb{F}_p^n$ 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: Define $f^{\oplus a} = f(x^{p-1} \oplus a)$, for some $a \in \{0,1\}^n$. 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 (Corollary 23). Similarly to [BV07] 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 (Claim 25, informal statement)**.** *Let $\widetilde{f}$ be a polynomial over $\mathbb{F}_p$ of degree d'. 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,*

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

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 (Claim 27, informal statement)**.** *Any directional derivative $f_y^{\oplus a}(x)$ can be computed by some function of $f_y(x)$ and $\{x_i : i \in \mathrm{Supp}(y)\}$.*

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 depends only on y and a, and is supported on S. Combining Claims 25 and 27 yields the required approximation of f.

To complete the picture we shortly remark on the proof of Theorem 7. The proof is by induction on the degree using Fourier analysis. The basic idea is that for every linear subspace $A \subseteq \mathbb{F}_p^n$ we have that

$$\sum_{\alpha \in \mathbb{F}_p^n : \alpha \neq 0, \alpha_S = 0} |\hat{f}(\alpha)|^2 \leq \mathbb{E}_{a \in A}\left[\sum_{\alpha \in \mathbb{F}_p^n : \alpha \neq 0, \alpha_S = 0} |\widehat{f_a}(\alpha)|^2\right] + \mathbb{E}_{a \in A}\left[|\widehat{f_a}(0)|^2\right] .$$

Using this useful inequality we break the analysis to two cases depending on whether f has a high Fourier coefficient or not. If all of f's Fourier coefficients are small, then we construct S in the following way: we pick a constant dimensional subspace A at random. For each derivative f_a, where

$a \in A$, we find a set S_a as guaranteed by the induction hypothesis (for some ϵ' depending on ϵ and d). Finally, we set S to be the union of all the S_a-s. When f has a high Fourier coefficient, we approximate f using a small number of lower degree polynomials and set S to be the union of their corresponding sets.

1.3 Paper organization

In Section 2 we fix some notations. We prove our main theorem, Theorem 5, in Section 3. The proof is based on Theorem 13 whose proof is given in Section 5, where we also prove Theorem 9. The proof of Theorem 13 relies on Theorem 7 that we prove in Section 6. We conclude and give some open problems in Section 7. For completeness, we sketch the proof for the linear case of Theorem 5 (i.e. $d = 1$) in Appendix A.

2 Preliminaries

We will be working over a fixed prime finite field $\mathbb{F}_p$. Let $f(x) = f(x_1, \dots, x_n)$ be a degree d polynomial over $\mathbb{F}_p$. Let $\mathcal{D}$ be a distribution. The support of $\mathcal{D}$ is the set of elements which have positive probability under $\mathcal{D}$. If the support of $\mathcal{D}$ is contained in a set S, we denote this by $\mathcal{D} \subseteq S$. For a distribution $\mathcal{D} \subset \mathbb{F}_p^n$, we denote by $f(\mathcal{D})$ the output distribution of f given inputs samples according to $\mathcal{D}$. For a subset $S \subset \mathbb{F}_p^n$ we denote by $f(S)$ the distribution of f over inputs chosen uniformly from S. In particular, $f(\mathbb{F}_p^n)$ denotes the distribution of f over uniform field elements, and $f(\{0,1\}^n)$ denotes the distribution of f over uniform bits.

The statistical distance between two distributions $\mathcal{D}', \mathcal{D}''$ is given by $\mathrm{sd}(\mathcal{D}', \mathcal{D}'') = \frac{1}{2}\sum_x |\Pr[\mathcal{D}' = x] - \Pr[\mathcal{D}'' = x]|$. If the statistical distance is at most ϵ, the distributions are said to be *ϵ-close.* If the statistical distance is at least ϵ, the distributions are said to be *ϵ-far.* It is easy to verify that statistical distance satisfies the triangle inequality.

Denote $[n] = \{1, 2, \dots, n\}$. A distribution $K \subset \{0,1\}^n$ is said to be *k-wise independent* if for any k distinct indices $i_1, \dots, i_k \in [n]$, the distribution K restricted to these indices is uniform over $\{0,1\}^k$.

For a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ we denote by $\widehat{f} : \mathbb{F}_p^n \to \mathbb{C}$ its Fourier transform, defined as $\widehat{f}(\alpha) = \mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{f(x) - \langle x, \alpha \rangle}]$, where $\omega = e^{2\pi i/p}$ and $\langle x, \alpha \rangle = \sum_{i=1}^n \alpha_i x_i$ is the inner product of x and α. The Fourier representation of $f(x)$ is given by $f(x) = \sum_{\alpha \in \mathbb{F}_p^n} \widehat{f}(\alpha) \omega^{\langle x, \alpha \rangle}$. Parseval's identity gives that $\sum_{\alpha \in \mathbb{F}_p^n} |\widehat{f}(\alpha)|^2 = 1$.

3 Bit pseudorandom generator for low degree polynomials

In this section we prove Theorem 5. As sketched in Section 1.2 we first prove the theorem for the (easy) case of *regular* polynomials (a notion that we shall soon define) and then for non-regular polynomials.

3.1 Regular polynomials

Definition 10. The *p-biased distribution* $\mathcal{U}_p \subset \{0,1\}^n$ is the distribution in which we choose each bit independently to be 0 with probability $\frac{1}{p}$ and to be 1 with probability $1 - \frac{1}{p}$.

We call a polynomial $f : \mathbb{F}_p^n \to \mathbb{F}_p$ ϵ-regular if $\mathrm{sd}(f(\mathcal{U}_p), f(\{0,1\}^n)) \leq \epsilon$. The following lemma shows that if f is a regular polynomial then it is fooled by the $p-1$ power of a pseudorandom distribution for degree $(p-1)d$ polynomials.

Lemma 11. *Let $f(x_1, \ldots, x_n)$ be an ϵ-regular polynomial of degree d over $\mathbb{F}_p$. 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 $\mathrm{sd}(f(\mathcal{D}^{p-1}), f(\{0,1\}^n)) \leq 2\epsilon$.*

Proof. Let $\tilde{f} : \mathbb{F}_p^n \to \mathbb{F}_p$ be defined as $\tilde{f}(x_1, \ldots, x_n) = f(x_1^{p-1}, \ldots, x_n^{p-1})$. As f is a degree d polynomial, $\tilde{f}$ is a polynomial of degree $(p-1)d$. Since $\mathcal{D}$ is pseudorandom against polynomials of degree $(p-1)d$, we have that $\tilde{f}(\mathcal{D})$ and $\tilde{f}(\mathbb{F}_p^n)$ are ϵ-close. By the definition of $\tilde{f}$ it follows that $f(\mathcal{D}^{p-1})$ and $f(\mathcal{U}_p)$ are ϵ-close. Hence, $\mathrm{sd}(f(\{0,1\}^n), f(\mathcal{D}^{p-1})) \leq \mathrm{sd}(f(\{0,1\}^n), f(\mathcal{U}_p)) + \mathrm{sd}(f(\mathcal{U}_p), f(\mathcal{D}^{p-1})) \leq 2\epsilon$. □

3.2 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. Namely, 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 a 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 $\mathcal{D}$.

Claim 12. *Let $f(x_1, \ldots, x_n)$ be a degree d polynomial over $\mathbb{F}_p$ 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 $\mathbb{F}_p$.*

Proof. If $f(\mathcal{U}_p)$ and $f(\{0,1\}^n)$ are ϵ-far, at least one of them is $\epsilon/2$-far from the uniform distribution over $\mathbb{F}_p$. If it is $f(\mathcal{U}_p)$, then we are done with $a = 0$. Otherwise assume that $f(\{0,1\}^n)$ is $\epsilon/2$-far from the uniform distribution over $\mathbb{F}_p$. We can generate the uniform distribution over $\{0,1\}^n$ by first choosing $a \in \{0,1\}^n$ uniformly at random, and then bitwise-XORing it to the distribution $\mathcal{U}_p$. In other words, the uniform distribution over $\{0,1\}^n$ is a convex combination of the distributions $\{\mathcal{U}_p \oplus a : a \in \{0,1\}^n\}$. Thus, the distribution $f(\{0,1\}^n)$ is a convex combination of the distributions $\{f(\mathcal{U}_p \oplus a) : a \in \{0,1\}^n\}$. In particular, there must exist some $a \in \{0,1\}^n$ such that the distribution $f(\mathcal{U}_p \oplus a)$ is $\epsilon/2$-far from uniform. □

We recall the definition of bit-rank given in Subsection 1.1.

Definition (Bit-Rank)**.** Let $g : \{0,1\}^n \to \mathbb{F}_p$ be a function. The d-bit-rank of g, denoted $\text{bit-rank}_d(g)$, is the minimal number of degree d polynomials over $\mathbb{F}_p$ required to compute g over $\{0,1\}^n$. That is, $\mathrm{rank}_d(g) = k$ 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 $x \in \{0,1\}^n$

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

The following theorem, shows that non-regular polynomials have a low bit-rank. We shall later deduce Theorem 9 from it. We defer the proof of the theorem to Section 5.

Theorem 13. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a polynomial of degree $d+1$ for some $d \geq 1$. Assume that, for some $a \in \{0,1\}^n$, the distribution of $f(\mathcal{U}_p \oplus a)$ is ϵ-far from uniform. 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 bit-rank$_d(g) \leq c + p^c$ where*[5] *$c = C((p-1)(d+1), \delta\epsilon^2/p^3)$.*

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

Lemma 14. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a degree $d+1$ polynomial. Assume that there is a function $g : \{0,1\}^n \to \mathbb{F}_p$ such that bit-rank$_d(g) = k$ and for some $a \in \{0,1\}^n$ it holds that*

$$\Pr_{x \in \mathcal{U}_p \oplus a}[f(x) = g(x)] \geq 1 - \delta.$$

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

To ease the reading we first show how to obtain Theorem 5 using Theorem 13 and Lemma 14. The proof of Lemma 14 is given in Section 4 and the proof of Theorem 13 is given in Section 5.

3.3 Proof of Theorem 5

For convenience we repeat the statement of the theorem.

Theorem. *Let $\mathbb{F}_p$ be an odd prime finite field, $d \geq 1$ be a degree and $\epsilon > 0$ be 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 bit-pseudorandom 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 on $\mathbb{F}_p$ and d.

Proof. The proof is by induction on the degree d. The case $d = 1$ was established in [LRTV09, MZ09] (Theorem 4). We restate their result here. For completeness we give a sketch of the proof in Appendix A.

Theorem (Bit-pseudorandom distribution for linear polynomials)**.** *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 against linear polynomials over $\mathbb{F}_p$ with error $O(\epsilon)$.*

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

We now proceed with the induction. Let $d > 1$, and let $f(x)$ be a polynomial of degree $d+1$. We divide the analysis into two cases. Assume first that $f(\mathcal{U}_p)$ is $\epsilon/2$-close to $f(\{0,1\}^n)$. Lemma 11 implies that if $\mathcal{D}_1 \subset \mathbb{F}_p^n$ is a pseudorandom distribution for degree $(p-1)(d+1)$ polynomials, with error $\epsilon/2$, then $f(\mathcal{D}_1^{p-1})$ and $f(\{0,1\}^n)$ are ϵ-close.

We now handle the case that $f(\mathcal{U}_p)$ is $\epsilon/2$-far from $f(\{0,1\}^n)$. By Claim 12 there exists some $a \in \{0,1\}^n$ such that $f(\mathcal{U}_p \oplus a)$ is $(\epsilon/4)$-far from uniform. Let $\delta > 0$ be determined later. Applying Theorem 13 there exists a function $g : \{0,1\}^n \to \mathbb{F}_p$ such that $\Pr_{x \in \mathcal{U}_p \oplus a}[f(x) \neq g(x)] \leq \delta$ and $\text{bit-rank}_d(g) \leq p^c + c$ for $c = C((p-1)(d+1), \delta(\epsilon/4)^2/p^3) = O(p\delta^{-1}\epsilon^{-1})^{O(4^{(p-1)(d+1)})}$. Lemma 14 implies that if $\mathcal{D}' \subset \{0,1\}^n$ is a bit-pseudorandom distribution for degree d polynomials with error ξ (that will be determined soon) then $f(\mathcal{D}')$ and $f(\{0,1\}^n)$ are τ-close for

$$\tau = c_1^{p^c+c}\xi + c_2\delta$$

where $c_1 = p^{2^{(p-1)(d+1)}}$ and $c_2 = 4p \cdot 2^{(p-1)(d+1)}$. In order to get $\tau \leq \epsilon$ we set $\delta = \epsilon/2c_2$ and $\xi = \epsilon/2c_1^{p^c+c}$. Substituting the parameters yields the bound

$$1/\xi \leq \exp(\exp((1/\epsilon)^{O(4^{(p-1)(d+1)})})\ .$$

We now put things together. Let $\mathcal{D}_2 \subset \mathbb{F}_p^n$ to be a pseudorandom distribution for degree $(p-1)d$ polynomials with error $\delta = \delta(p,d,\xi)$. Let $K \subset \{0,1\}^n$ be a k-wise independent distribution for $k = k(p,d,\xi)$. By the induction hypothesis, $\mathcal{D}' = \mathcal{D}_2^{p-1} \oplus K$ is a bit-pseudorandom distribution for degree d polynomials with error ξ. Thus, if f is not $\epsilon/2$-regular then $\mathcal{D}'$ fools f with error ϵ. To combine the two case in our analysis we note that if $\mathcal{D} \subset \mathbb{F}_p^n$ is a pseudorandom distribution against degree $(p-1)(d+1)$ polynomials with error ξ then $\mathcal{D}$ satisfies the requirements of both $\mathcal{D}_1$ and $\mathcal{D}_2$ (recall that $\xi \ll \epsilon$). Hence $\mathcal{D}^{p-1} \oplus K$ is a bit-pseudorandom distribution against any polynomial of degree $d+1$ with error ϵ. To conclude the proof we note that as

$$\delta(p, d+1, \epsilon) = \delta\left(p, d, 1/\exp(\exp((1/\epsilon)^{O(4^{(p-1)(d+1)})}))\right)$$

and

$$k(p, d+1, \epsilon) = k\left(p, d, 1/\exp(\exp((1/\epsilon)^{O(4^{(p-1)(d+1)})}))\right)$$

then there is a constants $c_{p,d} > 0$ such that

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

as claimed.

□

4 Approximately low bit-rank polynomials

In this section we give the proof of Lemma 14. We first give an overview of the proof.

Step 1. The first step in the proof is showing that if f is a degree $d+1$ polynomial which can be approximated by a function g of low d-bit-rank, then there is a *distribution* on functions G, such that every function in the support of G has a low d-bit-rank and such that for every $x \in \mathbb{F}_p^n$

it holds that $\Pr_{h\in G}[f(x) = h(x)] \geq 1 - \delta$ (Lemma 18). That is, we move from one function that compute 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 (Claim 15). This step is the main technical part of the proof

Step 2. In the second step we show that if a function has a low d-bit-rank then any bit-pseudorandom distribution for degree d polynomials fools it. The argument here is quite straightforward (Claim 19).

Step 3. Finally, we show that if a function can be computed using a distribution on functions that have low d-bit-rank (as we achieved in **Step 1** above) then it is fooled by bit-pseudorandom distributions for degree d polynomials (Claim 21).

4.1 Step 1: from average case to worst case approximation

As in the overview above we start by showing that there exists a distribution on low d-bit-rank functions that correctly computes f everywhere (w.h.p.). To construct such a distribution we shall refer to the self correction properties of polynomials over $\mathbb{F}_p$. Using these properties we will show that we can construct G by (roughly) considering many shifts of g (the polynomial that computes f on a $1-\delta$ fraction of $\mathbb{F}_p^n$). We start with the following well known fact.

Claim 15. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a degree $d+1$ polynomial. For every $x, y_1, \ldots, y_{d+2} \in \mathbb{F}_p^n$ the following holds*

$$f(x) = \sum_{I\subseteq[d+2],|I|\geq 1} (-1)^{|I|+1} f(x + \sum_{i\in I} y_I) \ .$$

Proof. Taking $d+2$ partial derivatives of $f(x)$, in directions[6] $y_1, \ldots, y_{d+2}$, iteratively, we obtain the constant zero function. That is $f_{y_1,\ldots,y_{d+2}}(x) \equiv 0$. The claim follows as, by definition, $f_{y_1,\ldots,y_{d+2}}(x) = (-1)^{d+2} \sum_{I\subseteq[d+2]} (-1)^{|I|} f(x + \sum_{i\in I} y_I)$. □

The following is an easy corollary.

Corollary 16. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a degree $d+1$ polynomial. Let $t = (p-1)(d+1)+1$. For every $x, y_1, \ldots, y_t \in \mathbb{F}_p^n$ and $a \in \{0,1\}^n$ the following holds*

$$f(x^{p-1} \oplus a) = \sum_{I\subseteq[t],|I|\geq 1} (-1)^{|I|+1} f((x + \sum_{i\in I} y_I)^{p-1} \oplus a) \ .$$

Proof. Define $g^{\oplus a}(x) = f(x^{p-1} \oplus a)$. Note that $g^{\oplus a}$ is a polynomial of degree $(p-1)(d+1)$ since

$$g^{\oplus a}(x_1, \ldots, x_n) = f(\alpha_1 x_1^{p-1} + \beta_1, \ldots, \alpha_n x_n^{p-1} + \beta_n)$$

where α_i, β_i are defined as follows. If $a_i = 0$ then $\alpha_i = 1, \beta_i = 0$ and if $a_i = 1$ then $\alpha_i = -1, \beta_i = 1$. The claim is proved by applying Claim 15 to the polynomial $g^{\oplus a}$. □

[6] Recall that the derivative of f in direction y is defined as $f_y(x) = f(x+y) - f(x)$. It is easy to verify that if f has degree $d+1$ then f_y has degree at most d.

We now show that a shift of a low bit-rank polynomial also has low bit-rank.

Claim 17. *Let $g : \{0,1\}^n \to \mathbb{F}_p$ be a function. For $a, c \in \{0,1\}^n, b \in \mathbb{F}_p^n$ define $g^{\oplus a,+b,\oplus c} : \{0,1\}^n \to \mathbb{F}_p$ by $g^{\oplus a,+b,\oplus c}(x) = g(((x \oplus c) + b)^{p-1} \oplus a)$. Then bit-rank$_d(g^{\oplus a,+b,\oplus c}) \leq$ bit-rank$_d(g)$.*

Proof. Assume bit-rank$_d(g) = k$. Consequently, there are k degree d polynomials $f_1, \ldots, f_k$ and a mapping $\Gamma : \mathbb{F}_p^k \to \mathbb{F}_p$ such that $g(x) = \Gamma(f_1(x), \ldots, f_k(x))$. Thus

$$g^{\oplus a,+b,\oplus c}(x) = g(((x \oplus c) + b)^{p-1} \oplus a) = \Gamma(f_1(((x \oplus c) + b)^{p-1} \oplus a), \ldots, f_k(((x \oplus c) + b)^{p-1} \oplus a)) \ .$$

We will conclude the proof by showing that each $f_j(((x \oplus c) + b)^{p-1} \oplus a)$ is a polynomial of degree at most d (in x). Define $f_j'(x_1, \ldots, x_n) = f_j(\alpha_1 x_1 + \beta_1, \ldots, \alpha_n x_n + \beta_n)$ where α_i, β_i are defined such that $\alpha_i x_i + \beta_i = ((x_i \oplus c_i) + b_i)^{p-1} \oplus a_i$ for $x_i \in \{0,1\}$ (that is, $\beta_i = (c_i + b_i)^{p-1} \oplus a_i$ and $\alpha_i = -\beta_i + (((1 \oplus c_i) + b_i)^{p-1} \oplus a_i))$. As we applied an affine linear transformation to the inputs $x_1, \ldots, x_n$, we have $\deg(f_j') \leq \deg(f_j) \leq d$. We conclude that for any $x \in \{0,1\}^n$

$$g^{\oplus a,+b,\oplus c}(x) = \Gamma(f_1'(x), \ldots, f_k'(x)),$$

hence bit-rank$_d(g^{\oplus a,+b,\oplus c}) \leq k$. □

Let G be a distribution over functions $g : \{0,1\}^n \to \mathbb{F}_p$. The d-bit-rank of G is defined to be the maximal d-bit-rank of a function in the support of G. The following lemma concludes the idea sketched above and shows that if f is close to a function with a low bit-rank then there is a distribution on low bit-rank functions that pointwise computes f.

Lemma 18. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a degree $d+1$ polynomial. Assume that there is a function $g : \{0,1\}^n \to \mathbb{F}_p$ such that bit-rank$_d(g) = k$ and such that for some $a \in \{0,1\}^n$ it holds that*

$$\Pr_{x \in \mathcal{U}_p \oplus a}[f(x) = g(x)] \geq 1 - \delta \ .$$

Then, there is a distribution G on functions such that bit-rank$_d(G) \leq (2^{(p-1)(d+1)+1} - 1)k$ and $\Pr_{h \in G}[f(x) = h(x)] \geq 1 - (2^{(p-1)(d+1)+1} - 1)\delta$.

Proof. We start by noting that the distribution $\mathcal{U}_p \oplus a$ is equivalent to the distribution of $x^{p-1} \oplus a$ for uniform $x \in \mathbb{F}_p^n$. By our assumption we have that

$$\Pr_{x \in \mathbb{F}_p^n}[f(x^{p-1} \oplus a) \neq g(x^{p-1} \oplus a)] \leq \delta.$$

Applying Corollary16 to f, which is a degree $d+1$ polynomial, we obtain

$$f(x^{p-1} \oplus a) = \sum_{I \subseteq [t], |I| \geq 1} (-1)^{|I|+1} f((x + \sum_{i \in I} y_I)^{p-1} \oplus a) \tag{1}$$

for $t = (p-1)(d+1)+1$ and any $y_1, \ldots, y_t \in \mathbb{F}_p^n$. Fix some $x \in \{0,1\}^n$. Let $y_1, \ldots, y_t \in \mathbb{F}_p^n$ be chosen uniformly at random, and note that for any non-empty $I \subseteq [t]$, the distribution of $x + \sum_{i \in I} y_i$ is uniform over $\mathbb{F}_p^n$. Therefore, for every $I \neq \emptyset$ it holds that

$$\Pr_{y_1, \ldots, y_t \in \mathbb{F}_p^n}\left[f((x + \sum_{i \in I} y_i)^{p-1} \oplus a) \neq g((x + \sum_{i \in I} y_i)^{p-1} \oplus a)\right] \leq \delta \ .$$

As $x^{p-1} = x$ for $x \in \{0,1\}^n$ we have that for such x-s $f(x^{p-1} \oplus a) = f(x \oplus a)$. Therefore, by Equation (1) and the union bound we get

$$\Pr_{y_1,\ldots,y_t \in \mathbb{F}_p^n}\left[f(x \oplus a) \neq \sum_{I \subseteq [t], |I| \geq 1} (-1)^{|I|+1} g((x + \sum_{i \in I} y_i)^{p-1} \oplus a)\right] \leq (2^t - 1)\delta \ .$$

Hence, for every $x \in \{0,1\}^n$ we have

$$\Pr_{y_1,\ldots,y_t \in \mathbb{F}_p^n}\left[f(x) \neq \sum_{I \subseteq [t], |I| \geq 1} (-1)^{|I|+1} g(((x \oplus a) + \sum_{i \in I} y_i)^{p-1} \oplus a)\right] \leq (2^t - 1)\delta \ .$$

For any setting of $y_1, \ldots, y_t \in \mathbb{F}_p^n$, define

$$h^{(y_1,\ldots,y_t)}(x) = \sum_{I \subseteq [t], |I| \geq 1} (-1)^{|I|+1} g(((x \oplus a) + \sum_{i \in I} y_i)^{p-1} \oplus a) \ .$$

Let G denote the distribution over the functions $h^{(y_1,\ldots,y_t)}$ obtained by sampling $y_1, \ldots, y_t \in \mathbb{F}_p^n$ uniformly at random. We conclude that for every $x \in \{0,1\}^n$ it holds that

$$\Pr_{h \in G}[f(x) = h(x)] \geq 1 - (2^t - 1)\delta \ .$$

To complete the proof we bound the d-bit-rank of G. Each function $h \in G$ is a linear combination of $g(((x \oplus a) + \sum_{i \in I} y_i)^{p-1} \oplus a) = g^{\oplus a, +\sum_{i \in I} y_i, \oplus a}(x)$, and by Claim 17 we know that $\text{bit-rank}_d(g^{\oplus a, +\sum_{i \in I} y_i, \oplus a}) \leq \text{bit-rank}_d(g) = k$. Therefore, $\text{bit-rank}_d(h) \leq (2^t - 1)k$. Consequently, $\text{bit-rank}_d(G) \leq (2^t - 1)k$. □

4.2 Steps 2 and 3: fooling approximately low bit-rank polynomials

We start by arguing that bit-pseudorandom distributions for degree d polynomials also fool functions with low d-bit-rank.

Claim 19. *Let $g : \{0,1\}^n \to \mathbb{F}_p$ be a function with $\text{bit-rank}_d(g) = k$. Let $\mathcal{D} \subset \{0,1\}^n$ be a bit-pseudorandom distribution for degree d polynomials with error ϵ. Then $g(\mathcal{D})$ and $g(\{0,1\}^n)$ are $(p^{k/2}\epsilon)$-close.*

Proof. Let $g = \Gamma(f_1(x), \ldots, f_k(x))$ be a representation of g as a function of k polynomials of degree $\leq d$. Denote with $\mathcal{D}_1 \subset \mathbb{F}_p^k$ the joint distribution of $(f_1(x), \ldots, f_k(x))$ when $x \in \{0,1\}^n$ is chosen uniformly at random. Similarly, denote with $\mathcal{D}_2 \subset \mathbb{F}_p^k$ the joint distribution of $(f_1(x), \ldots, f_k(x))$ when $x \in \mathcal{D}$. We will prove that $\mathcal{D}_1$ and $\mathcal{D}_2$ are $(p^{k/2}\epsilon)$-close and hence $g(\mathcal{D})$ and $g(\{0,1\}^n)$ are $(p^{k/2}\epsilon)$-close.

For $\alpha \in \mathbb{F}_p^k$ define $\langle \mathcal{D}_i, \alpha \rangle \subset \mathbb{F}_p$ to be the distribution of the inner product $\langle y, \alpha \rangle$ where $y \in \mathbb{F}_p^k$ is sampled according to $\mathcal{D}_i$. In other words, for $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{F}_p^k$ we have that $\langle \mathcal{D}_1, \alpha \rangle$ is the distribution of $f_\alpha(x) = \sum \alpha_i f_i(x)$ over uniform $x \in \{0,1\}^n$, and that $\langle \mathcal{D}_2, \alpha \rangle$ is the distribution of $f_\alpha(x)$ for $x \in \mathcal{D}$. Since $\mathcal{D}$ is a bit-pseudorandom distribution for degree d polynomials with error ϵ and as each f_α is a degree d polynomial (it is a linear combination of polynomials of degree d), we get that the distributions $\langle \mathcal{D}_1, \alpha \rangle$ and $\langle \mathcal{D}_2, \alpha \rangle$ are ϵ-close. The following well-known fact shows that two distributions with similar Fourier coefficients must be close. For completeness we give the proof in Appendix B.

Fact 20. *Let $\mathcal{D}_1, \mathcal{D}_2 \subset \mathbb{F}_p^k$ be two distributions. Assume that for every $\alpha \in \mathbb{F}_p^k$ the distributions $\langle \mathcal{D}_1, \alpha \rangle$ and $\langle \mathcal{D}_2, \alpha \rangle$ are ϵ-close. Then $\mathcal{D}_1$ and $\mathcal{D}_2$ are $(p^{k/2}\epsilon)$-close.*

It follows that $\mathcal{D}_1$ and $\mathcal{D}_2$ are $(p^{k/2}\epsilon)$-close which concludes the proof. □

We next prove that if a degree $(d+1)$ polynomial f can be pointwise approximated by a distribution with a low d-bit-rank, then f is in fact fooled by bit-pseudorandom distributions for degree d polynomials.

Claim 21. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a degree $d+1$ polynomial. Let G be a distribution over functions $h : \{0,1\}^n \to \mathbb{F}_p$ such that bit-rank$_d(G) = k$, and such that for every $x \in \{0,1\}^n$*

$$\Pr_{h \in G}[f(x) = h(x)] \geq 1 - \delta.$$

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

Proof. We need to bound

$$\mathrm{sd}(f(\mathcal{D}), f(\{0,1\}^n)) = \tfrac{1}{2} \sum_{t \in \mathbb{F}_p} \left| \Pr_{x \in \mathcal{D}}[f(x) = t] - \Pr_{x \in \{0,1\}^n}[f(x) = t] \right| .$$

Let $E \subset \{0,1\}^n$ be some distribution. We now prove that for every $t \in \mathbb{F}_p$ it holds that

$$\left| \Pr_{x \in E}[f(x) = t] - \Pr_{x \in E, h \in G}[h(x) = t] \right| \leq \delta.$$

First, note that

$$\begin{aligned}
\Pr_{x \in E}[f(x) = t] &= \Pr_{x \in E, h \in G}[f(x) = t \ \wedge \ f(x) = h(x)] + \Pr_{x \in E, h \in G}[f(x) = t \ \wedge \ f(x) \neq h(x)] \\
&= \Pr_{x \in E, h \in G}[h(x) = t \ \wedge \ f(x) = h(x)] + \Pr_{x \in E, h \in G}[f(x) = t \ \wedge \ f(x) \neq h(x)]
\end{aligned}$$

and

$$\Pr_{x \in E, h \in G}[h(x) = t] = \Pr_{x \in E, h \in G}[h(x) = t \ \wedge \ f(x) = h(x)] + \Pr_{x \in E, h \in G}[h(x) = t \ \wedge \ f(x) \neq h(x)] .$$

Therefore, we get that

$$\begin{aligned}
&\left| \Pr_{x \in E}[f(x) = t] - \Pr_{x \in E, h \in G}[h(x) = t] \right| = \\
&\left| \Pr_{x \in E, h \in G}[f(x) = t \ \wedge \ f(x) \neq h(x)] - \Pr_{x \in E, h \in G}[h(x) = t \ \wedge \ f(x) \neq h(x)] \right| \leq \\
&\Pr_{x \in E, h \in G}[f(x) \neq h(x)] = \mathbb{E}_{x \in E} \Pr_{h \in G}[f(x) \neq h(x)] \leq \delta.
\end{aligned}$$

The claim now follows as

$$\begin{aligned}
2 \cdot \mathrm{sd}(f(\mathcal{D}), f(\{0,1\}^n)) &= \sum_{t \in \mathbb{F}_p} |\Pr_{x \in \mathcal{D}}[f(x)=t] - \Pr_{x \in \{0,1\}^n}[f(x)=t]| \\
&\leq \sum_{t \in \mathbb{F}_p} |\Pr_{x \in \mathcal{D}, h \in G}[h(x)=t] - \Pr_{x \in \{0,1\}^n, h \in G}[h(x)=t]| \\
&+ \sum_{t \in \mathbb{F}_p} |\Pr_{x \in \mathcal{D}}[f(x)=t] - \Pr_{x \in \mathcal{D}, h \in G}[h(x)=t]| \\
&+ \sum_{t \in \mathbb{F}_p} |\Pr_{x \in \{0,1\}^n}[f(x)=t] - \Pr_{x \in \{0,1\}^n, h \in G}[h(x)=t]| \\
&\leq E_{h \in G}[\sum_{t \in \mathbb{F}_p} |\Pr_{x \in \mathcal{D}}[h(x)=t] - \Pr_{x \in \{0,1\}^n}[h(x)=t]|] + 2p\delta \\
&\leq E_{h \in G}[2 \cdot \mathrm{sd}(h(\mathcal{D}), h(\{0,1\}^n))] + 2p\delta \\
&\leq 2p^{k/2}\epsilon + 2p\delta.
\end{aligned}$$

□

The proof of Lemma 14 now follows easily.

Proof of Lemma 14. By Lemma 18 there is a distribution G on functions such that $\text{bit-rank}_d(G) \leq (2^{(p-1)(d+1)+1} - 1)k$ and $\Pr_{h \in G}[f(x) = h(x)] \geq 1 - (2^{(p-1)(d+1)+1} - 1)\delta$. Applying Claim 21 we get that the distance between $f(\mathcal{D})$ and $f(\{0,1\}^n)$ is bounded by

$$\mathrm{sd}(f(\mathcal{D}), f(\{0,1\}^n)) \leq p^{2^{(p-1)(d+1)}k}\epsilon + p2^{(p-1)(d+1)+2}\delta\ .$$

□

5 The structure of non-regular polynomials

In this section prove of Theorem 13. To ease the reading we repeat it here.

Theorem (Theorem 13)**.** *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a polynomial of degree $d+1$ for some $d \geq 1$. Assume that, for some $a \in \{0,1\}^n$, the distribution of $f(\mathcal{U}_p \oplus a)$ is ϵ-far from uniform. 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 bit-rank$_d(g) \leq c + p^c$ where*[7] *$c = C((p-1)(d+1), \delta\epsilon^2/p^3) = (p^3/\delta\epsilon^2)^{O(4^{(p-1)(d+1)})}$.*

Before giving the actual proof we first give an overview of the main steps.

Step 1: We start by showing that if f is non-regular then it must have a somewhat large 'Fourier coefficient' with respect to a shifted p-biased distribution (Corollary 23).

Step 2: Defining $f^{\oplus a}(x) = f(x^{p-1} \oplus a)$ it follows that $f^{\oplus a}$ is a degree $(p-1)(d+1)$ polynomial that has a (relatively) high bias. In addition, Theorem 7 implies that there is a relatively small set

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

of variables S such that the weight of the Fourier mass of $f^{\oplus a}$, that is supported on $\bar{S}$, is small.

Step 3: Next we show that if a polynomial has the two properties found in **Step 2** then it can be well approximated by (a function of) a small number of its derivatives and the variables in S. This is the main technical part of the proof (Claim 25 and Corollary 26). Intuitively, the idea is the following: Note that when $\widehat{f}(0) \neq 0$ we have that $\omega^{-f(x)} = \widehat{f}(0)^{-1}\mathbb{E}_{y\in\mathbb{F}_p^n}\omega^{f_y(x)}$. In our case we show that we can actually get $\omega^{-f(x)} \approx \widehat{f}(0)^{-1}\mathbb{E}_{y\in\mathbb{F}_p^S}\omega^{f_y(x)}$ for most x's, where $\mathbb{F}_p^S$ is the set of n-tuples in $\mathbb{F}_p$ that are supported on S. The reason being that w^f takes discrete values that are 'far' from each other and the average contribution of the derivatives in directions that are not supported on S is small (this follows, after some manipulations, from the structure guaranteed by Theorem 7).

Step 4: Using the fact that $f^{\oplus a}(x) = f(x^{p-1} \oplus a)$ we show that each derivative of $f^{\oplus a}$ is actually a function of a small number of the partial derivatives of f and the variables in S (Claim 27).

Step 5: From the above steps we get that $f(x^{p-1} \oplus a)$ can be well approximated by the variables in S and a small number of derivatives $f_z(x^{p-1} \oplus a)$. In the last step of the proof we show that we can actually replace the variables in $\{x_i : i \in S\}$ with $\{x_i^{p-1} \oplus a : i \in S\}$. From this we shall conclude that f can be well approximated, with respect to the shifted p-biased distribution, by (a function of) a small number of its derivatives and the variables in S.

We now go to the formal proof according to the steps sketched above.

5.1 Steps 1 and 2: finding structure in the Fourier spectrum

We start with an easy claim regarding Fourier coefficients of distributions. Abusing notations, given a distribution $\mathcal{D} \subseteq \mathbb{F}_p$ we identify it with the function $\mathcal{D} : \mathbb{F}_p \to [0,1]$ in the following way $\mathcal{D}(y) = \Pr_{x\in\mathcal{D}}[x = y]$. Notice that under this definition $\hat{\mathcal{D}}(0) = 1/p$ and in general,[8] $\hat{\mathcal{D}}(t) = \mathbb{E}_{x\in\mathbb{F}_p}[\mathcal{D}(x)\cdot\omega^{-t\cdot x}] = \frac{1}{p}\mathbb{E}_{x\in\mathcal{D}}[\omega^{-t\cdot x}]$.

Claim 22. *Let $\mathcal{D} \subset \mathbb{F}_p$ be a distribution which is ϵ-far from uniform. Then there exists some $t \in \mathbb{F}_p \setminus 0$ such that*

$$p \cdot \hat{\mathcal{D}}(t) = \mathbb{E}_{x\in\mathcal{D}}[\omega^{-t\cdot x}] \geq \epsilon \cdot p^{-1/2} \ .$$

Proof. Let $\mathcal{U}$ denote the uniform distribution on $\mathbb{F}_p$. We have

$$\begin{aligned} 4\epsilon^2 \leq 4\cdot \mathrm{sd}(\mathcal{D},\mathcal{U})^2 &= \left(\sum_{t\in\mathbb{F}_p} |Pr[\mathcal{D} = t] - \Pr[\mathcal{U} = t]|\right)^2 \\ &\leq p^2\mathbb{E}_{t\in\mathbb{F}_p}[|Pr[\mathcal{D} = t] - \Pr[\mathcal{U} = t]|^2] = p^2\sum_{t\in\mathbb{F}_p}|\hat{\mathcal{D}}(t) - \hat{\mathcal{U}}(t)|^2, \end{aligned}$$

where the last equality follows from the Parseval identity. As $\hat{\mathcal{U}}(t) = 0$ for $t \neq 0$, and $\hat{\mathcal{D}}(0) = \hat{\mathcal{U}}(0) = 1/p$ we get

$$\sum_{t\in\mathbb{F}_p\setminus 0} |\hat{\mathcal{D}}(t)|^2 \geq 4\epsilon^2/p^2,$$

[8] When speaking of distributions we do not consider the function $\omega^{\mathcal{D}}$ as we do with polynomials.

hence there is some $t \in \mathbb{F}_p \setminus 0$ such that $|\hat{\mathcal{D}}(t)| \geq \sqrt{\frac{4}{p-1}}\epsilon/p \geq \epsilon \cdot p^{-3/2}$. □

We obtain the following corollary.

Corollary 23. *If the distribution $f(\mathcal{U}_p \oplus a)$ is ϵ-far from uniform then there is some $0 \neq t \in \mathbb{F}_p$ such that*

$$\mathbb{E}_{x \in \mathcal{U}_p \oplus a}[\omega^{t \cdot f(x)}] \geq \epsilon \cdot p^{-1/2} .$$

Note that we can assume w.l.o.g that $t = 1$. Indeed, let $f' = t \cdot f(x)$. We shall prove that there is 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 $\text{bit-rank}_d(g') \leq c + p^c$. Setting $g(x) = t^{-1} \cdot g'(x)$ we get the required polynomial for f. Thus from now on we assume that $t = 1$, i.e. that

$$\mathbb{E}_{x \in \mathcal{U}_p \oplus a}[\omega^{f(x)}] \geq \epsilon \cdot p^{-1/2} .$$

Let $f^{\oplus a} : \mathbb{F}_p^n \to \mathbb{F}_p$ be defined as $f^{\oplus a}(x) = f(x^{p-1} \oplus a)$. Since the distribution of $x^{p-1} \oplus a$ for uniform $x \in \mathbb{F}_p^n$ is exactly $\mathcal{U}_p \oplus a$ we get

$$\mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{f^{\oplus a}(x)}] \geq \epsilon \cdot p^{-1/2} .$$

Let $\gamma = \delta\epsilon^2/p^2$, for some $\delta > 0$. As $f^{\oplus a}(x)$ is a polynomial of degree at most $(p-1)(d+1)$, Theorem 7 implies that there exists a subset $S \subset [n]$ of size $|S| \leq C((p-1)(d+1), \gamma)$ such that

$$\sum_{\alpha \in \mathbb{F}_p^{\overline{S}} \setminus 0} |\widehat{f^{\oplus a}}(\alpha)|^2 \leq \gamma . \tag{2}$$

5.2 Step 3: approximating $f^{\oplus a}$ by a few derivatives

Next, we show that the function $f^{\oplus a}(x)$ can be well approximated by a small set of its derivatives. We start with some definitions and a simple yet useful equality. For a subset $S \subset [n]$ let $\mathbb{F}_p^S$ denote the set of vectors $v \in \mathbb{F}_p^n$ which are supported on S, that is,

$$\mathbb{F}_p^S = \{v \in \mathbb{F}_p^n : v_i = 0 \ \forall i \notin S\} .$$

Similarly let $F_p^{\overline{S}}$ denote the set of vectors supported on $\overline{S} = [n] \setminus S$. For $x \in \mathbb{F}_p^n$ let $x_S \in \mathbb{F}_p^S$ denote the part of x which is supported on S, and $x_{\overline{S}}$ the part of x supported on $[n] \setminus S$.

Claim 24. *For any function $h : \mathbb{F}_p^n \to \mathbb{F}_p$ and $S \subseteq [n]$ we have*

$$\mathbb{E}_{x \in \mathbb{F}_p^n, y \in \mathbb{F}_p^S}[\omega^{h_y(x)}] = \sum_{\alpha \in \mathbb{F}_p^{\bar{S}}} |\hat{h}(\alpha)|^2 .$$

Proof. Using the Fourier decomposition $\omega^{h(x)} = \sum_{\alpha \in \mathbb{F}_p^n} \hat{h}(\alpha)\omega^{\langle \alpha, x \rangle}$ we get

$$\begin{aligned}
\mathbb{E}_{x \in \mathbb{F}_p^n, y \in \mathbb{F}_p^S}[\omega^{h_y(x)}] &= \mathbb{E}_{x \in \mathbb{F}_p^n, y \in \mathbb{F}_p^S}[\omega^{h(x+y)-h(x)}] \\
&= \mathbb{E}_{x \in \mathbb{F}_p^n, y \in \mathbb{F}_p^S} \sum_{\alpha, \beta \in \mathbb{F}_p^n} \hat{h}(\alpha)\overline{\hat{h}(\beta)}\omega^{\langle \alpha, x+y \rangle - \langle \beta, x \rangle} \\
&= \sum_{\alpha, \beta \in \mathbb{F}_p^n} \hat{h}(\alpha)\overline{\hat{h}(\beta)} \left(\mathbb{E}_{x \in \mathbb{F}_p^n}\omega^{\langle \alpha - \beta, x \rangle}\right) \left(\mathbb{E}_{y \in \mathbb{F}_p^S}\omega^{\langle \alpha, y \rangle}\right) \\
&= \sum_{\alpha \in \mathbb{F}_p^n : \alpha_S = 0} |\hat{h}(\alpha)|^2 = \sum_{\alpha \in \mathbb{F}_p^{\overline{S}}} |\hat{h}(\alpha)|^2 .
\end{aligned}$$

□

We next show that a function h that has a high bias and that satisfy that $\sum_{\alpha \in \mathbb{F}_p^{\overline{S}} \setminus 0} |\hat{h}(\alpha)|^2$ is small can be well approximated by its derivatives in directions from $\mathbb{F}_p^S$. The proof is based on the idea described in **Step 3** above.

Claim 25. *Let $h : \mathbb{F}_p^n \to \mathbb{F}_p$ be a function such that $|\mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{h(x)}]| \geq \epsilon$. Let $\delta > 0$ be an error parameter, and assume there is a subset $S \subset [n]$ such that $\sum_{\alpha \in \mathbb{F}_p^{\overline{S}} \setminus 0} |\hat{h}(\alpha)|^2 \leq \gamma$ for $\gamma = \delta\epsilon^2/p^2$. Then h can be approximated, on a $1-\delta$ fraction of $\mathbb{F}_p^n$, by a function of its derivatives in directions supported on S. That is, there exists a function $\Gamma : \mathbb{F}_p^{p^{|S|}} \to \mathbb{F}_p$ such that*

$$\Pr_{x \in \mathbb{F}_p^n}[h(x) \neq \Gamma(\{h_y(x) : y \in \mathbb{F}_p^S\})] \leq \delta.$$

Proof. Let $\hat{h}(0) = \mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{h(x)}]$. By the assumption in the claim, $|\hat{h}(0)| \geq \epsilon$. Define $\phi(x) = \hat{h}(0)^{-1}\mathbb{E}_{y \in \mathbb{F}_p^S}\omega^{h_y(x)}$. We will show that $\phi(x)$ can be used to compute $h(x)$ on most of $\mathbb{F}_p^n$. Define $\Delta(x) = |\omega^{-h(x)} - \phi(x)|$. Note that the minimal distance between different p-th roots of unity, i.e. distinct elements in $\{\omega^t : t \in \mathbb{F}_p\}$, is given by $|1-\omega| = 2\sin(\frac{\pi}{p}) \geq 2/p$. We will show that for most $x \in \mathbb{F}_p^n$ we have $\Delta(x) < 1/p$, and hence we can deduce $\omega^{-h(x)}$ uniquely (and therefore $h(x)$) given $\phi(x)$. To achieve this we shall bound $\mathbb{E}_{x \in \mathbb{F}_p^n}[\Delta(x)^2]$ and then use Markov's inequality.

$$\begin{aligned}
&\mathbb{E}_{x \in \mathbb{F}_p^n}[\Delta(x)^2] = \mathbb{E}_{x \in \mathbb{F}_p^n}[(\omega^{-h(x)} - \phi(x))\overline{(\omega^{-h(x)} - \phi(x))}] = \\
&\mathbb{E}_{x \in \mathbb{F}_p^n, y,z \in \mathbb{F}_p^S}[(\omega^{-h(x)} - \hat{h}(0)^{-1}\omega^{h_y(x)})(\omega^{h(x)} - \overline{\hat{h}(0)^{-1}}\omega^{-h_z(x)})] = \\
&\mathbb{E}_{x \in \mathbb{F}_p^n, y,z \in \mathbb{F}_p^S}[\omega^{-h(x)+h(x)} - \hat{h}(0)^{-1}\omega^{h_y(x)+h(x)} - \overline{\hat{h}(0)^{-1}}\omega^{-h_z(x)-h(x)} + |\hat{h}(0)|^{-2}\omega^{h_y(x)-h_z(x)}] = \\
&\mathbb{E}_{x \in \mathbb{F}_p^n, y,z \in \mathbb{F}_p^S}[1 - \hat{h}(0)^{-1}\omega^{h(x+y)} - \overline{\hat{h}(0)^{-1}}\omega^{-h(x+z)} + |\hat{h}(0)|^{-2}\omega^{h(x+y)-h(x+z)}] = \\
&\mathbb{E}_{x \in \mathbb{F}_p^n, y \in \mathbb{F}_p^S}[1 - 1 - 1 + |\hat{h}(0)|^{-2}\omega^{h(x+y)-h(x)}] = \\
&\mathbb{E}_{x \in \mathbb{F}_p^n, y \in \mathbb{F}_p^S}[-1 + |\hat{h}(0)|^{-2}\sum_{\alpha \in \mathbb{F}_p^{\overline{S}}} |\hat{h}(\alpha)|^2]
\end{aligned}$$

where the last equality hollows from Claim 24. We thus have

$$\mathbb{E}_{x \in \mathbb{F}_p^n}[\Delta(x)^2] = |\hat{h}(0)|^{-2} \sum_{\alpha \in \mathbb{F}_p^{\overline{S}} \setminus 0} |\hat{h}(\alpha)|^2 \leq \frac{\gamma}{\epsilon^2} .$$

By Markov's inequality we obtain that

$$\Pr_{x \in \mathbb{F}_p^n}[\Delta(x) \geq 1/p] = \Pr_{x \in \mathbb{F}_p^n}[\Delta(x)^2 \geq 1/p^2] \leq \frac{\mathbb{E}_{x \in \mathbb{F}_p^n}[\Delta(x)^2]}{1/p^2} \leq \frac{p^2\gamma}{\epsilon^2} = \delta.$$

We now define $\Gamma : \mathbb{F}_p^{p^{|S|}} \to \mathbb{F}_p$ as the value of $h(x)$ for which $|\phi(x) - \omega^{-h(x)}|$ is minimized (breaking ties arbitrarily). Since $\phi(x)$ depends only on $\{h_y(x) : y \in \mathbb{F}_p^S\}$ so does Γ and, by the argument above, as long as $\Delta(x) < 1/p$ we know that $\Gamma(x) = h(x)$. Since $\Pr[\Delta(x) \geq 1/p] \leq \delta$, we conclude that

$$\Pr_{x \in \mathbb{F}_p^n}[h(x) \neq \Gamma(\{h_y(x) : y \in \mathbb{F}_p^S\})] \leq \delta .$$

□

From Equation (2) and Claim 25 we obtain the following corollary.

Corollary 26. *Let f, a, ϵ be as in the statement of Theorem 13. Then, for every $\delta > 0$ there is a function $\Gamma_1 : \mathbb{F}_p^{p^{|S|}} \to \mathbb{F}_p$ and a set $S \subset [n]$, of size $|S| \leq C((p-1)(d+1), \delta\epsilon^2/p^2)$, such that*

$$\Pr_{x \in \mathbb{F}_p^n}[f^{\oplus a}(x) \neq \Gamma_1(\{f_y^{\oplus a}(x) : y \in \mathbb{F}_p^S\})] \leq \delta \ .$$

5.3 Step 4: 'fixing' the derivatives

We now show that we can replace the derivatives of $f^{\oplus a}(x)$ in Corollary 26 by derivatives of f itself.

Claim 27. *For any fixed $y \in \mathbb{F}_p^S$ we have that $(f^{\oplus a})_y(x) = f^{\oplus a}(x+y) - f^{\oplus a}(x)$ is determined by the variables $\{x_i : i \in S\}$ and the derivatives of f supported on S when evaluated on $x^{p-1} \oplus a$. Namely, for every $y \in \mathbb{F}_p^S$ there is a function $\Psi^{(y)} : \mathbb{F}_p^{|S|+p^{|S|}} \to \mathbb{F}_p$ such that for every $x \in \mathbb{F}_p^n$*

$$(f^{\oplus a})_y(x) = \Psi^{(y)}(\{x_i : i \in S\}, \{f_z(x^{p-1} \oplus a) : z \in \mathbb{F}_p^S\}) \ .$$

Proof. Fix $x \in \mathbb{F}_p^n$. Let $w \in \mathbb{F}_p^n$ be defined by the equation $w = ((x+y)^{p-1} \oplus a) - (x^{p-1} \oplus a)$ (interpreted as a pointwise equality). Note that as $y \in \mathbb{F}_p^S$ we also have that $w \in \mathbb{F}_p^S$, i.e. $w_i = 0$ for all $i \notin S$. Moreover, note that we can compute w_S (hence also w) as a fixed function (depending on y, a) of $\{x_i : i \in S\}$. Hence we get

$$(f^{\oplus a})_y(x) = f((x+y)^{p-1} \oplus a) - f(x^{p-1} \oplus a) = f((x^{p-1} \oplus a) + w) - f(x^{p-1} \oplus a) = f_w(x^{p-1} \oplus a) \ .$$

Consequently, $(f^{\oplus a})_y(x)$ is a function of $\{x_i : i \in S\}$ and $\{f_z(x^{p-1} \oplus a) : z \in \mathbb{F}_p^S\}$. □

Combining Corollary 26 and Claim 27 we obtain the following corollary.

Corollary 28. *Let f, a, ϵ be as in the statement of Theorem 13. Then, for every $\delta > 0$ there is a set $S \subset [n]$, of size $|S| \leq C((p-1)(d+1), \delta\epsilon^2/p^2)$, and a function $\Gamma_2 : \mathbb{F}_p^{p^{|S|}+|S|} \to \mathbb{F}_p$ such that*

$$\Pr_{x \in \mathbb{F}_p^n}[f(x^{p-1} \oplus a) \neq \Gamma_2(\{x_i : i \in S\}, \{f_z(x^{p-1} \oplus a) : z \in \mathbb{F}_p^S\})] \leq \delta \ .$$

5.4 Step 5: putting it all together

We now prove Theorem 13. Given Corollary 28 we basically have to complete **Step 5** in order to conclude the proof. That is, we have to show that f can be well approximated by a function of a small number of its derivatives and the variables in S.

Proof of Theorem 13. By Corollary 28 there is a function $\Gamma_2 : \mathbb{F}_p^{p^{|S|}+|S|} \to \mathbb{F}_p$ such that

$$\Pr_{x \in \mathbb{F}_p^n}\left[f(x^{p-1} \oplus a) \neq \Gamma_2(\{x_i : i \in S\}, \{f_z(x^{p-1} \oplus a) : z \in \mathbb{F}_p^S\})\right] \leq \delta \ .$$

Thus, we have a function approximating f on Boolean inputs (under the distribution $x^{p-1} \oplus a$) which depends on $\{x_i : i \in S\}$. We will next show how these variables can be replaced by variables of the form $x_i^{p-1} \oplus a_i$, for $i \in S$. Let $u \in (\mathbb{F}_p \setminus 0)^n$ be chosen uniformly at random. Observe that the joint distribution of (x^{p-1}, x) over uniform $x \in \mathbb{F}_p^n$ is identical to the joint distribution of $(x^{p-1}, x^{p-1} \cdot u)$, where the product $x^{p-1} \cdot u$ is taken element-wise. It follows that

$$\Pr_{x\in\mathbb{F}_p^n, u\in(\mathbb{F}_p\setminus 0)^n}\left[f(x^{p-1}\oplus a) \neq \Gamma_2(\{(x_i)^{p-1}\cdot u_i : i \in S\}, \{f_z(x^{p-1}\oplus a) : z \in \mathbb{F}_p^S\})\right] \leq \delta \ .$$

By averaging, there exist a value $u^* \in (\mathbb{F}_p \setminus 0)^n$ such that

$$\Pr_{x\in\mathbb{F}_p^n}\left[f(x^{p-1}\oplus a) \neq \Gamma_2(\{(x_i)^{p-1}\cdot u_i^* : i \in S\}, \{f_z(x^{p-1}\oplus a) : z \in \mathbb{F}_p^S\})\right] \leq \delta \ .$$

Notice that given a, u^* we can compute $x_i^{p-1} \cdot u_i^*$ as a function of $x_i^{p-1} \oplus a_i$. Hence, we can define a function $\Gamma : \mathbb{F}_p^{p^{|S|}+|S|} \to \mathbb{F}_p$ such that

$$\Gamma(\{(x_i)^{p-1}\oplus a_i : i \in S\}, \{f_z(x^{p-1}\oplus a) : z \in \mathbb{F}_p^S\}) =$$
$$\Gamma_2(\{(x_i)^{p-1}\cdot u_i^* : i \in S\}, \{f_z(x^{p-1}\oplus a) : z \in \mathbb{F}_p^S\}) \ .$$

Therefore,

$$\Pr_{x\in\mathbb{F}_p^n}[f(x^{p-1}\oplus a) \neq \Gamma(\{x_i^{p-1}\oplus a_i : i \in S\}, \{f_z(x^{p-1}\oplus a) : z \in \mathbb{F}_p^S\})] \leq \delta \ .$$

As the distribution of $x^{p-1} \oplus a$, for uniform $x \in \mathbb{F}_p^n$, is the same as $\mathcal{U}_p \oplus a$, we conclude that

$$\Pr_{x\in\mathcal{U}_p\oplus a}[f(x) \neq \Gamma(\{x_i : i \in S\}, \{f_z(x) : z \in \mathbb{F}_p^S\})] \leq \delta \ .$$

Set $g(x) = \Gamma(\{x_i : i \in S\}, \{f_z(x) : z \in \mathbb{F}_p^S\})$. Clearly, $\text{bit-rank}_d(g) \leq |S| + p^{|S|}$. This completes the proof of the theorem. □

We now give the proof of Theorem 9.

Proof of Theorem 9. Combining Claim 12, Theorem 13 and Lemma 18 we get that there is a distribution G on functions such that for $c = C((p-1)(d+1), \delta\epsilon^2/4p^3)$ it holds that $\text{bit-rank}_d(G) \leq O(2^{(p-1)(d+1)}p^c)$ and $\Pr_{h\in G}[f(x) = h(x)] \geq 1 - (2^{(p-1)(d+1)+1} - 1)\delta$. A simple averaging argument implies that there is some $h \in G$ such that $\Pr_{x\in\{0,1\}^n}[f(x) = h(x)] \geq 1 - (2^{(p-1)(d+1)+1} - 1)\delta$. □

6 The Fourier spectrum of low degree polynomials

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 29. 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 30 (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 $|S| \le C(d,\epsilon) = O(1/\epsilon)^{O(4^d)}$, such that $\Delta_S(f,f) \le \epsilon$.*

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

Claim 32. *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*

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

where $\mathbb{E}_A$ means that we are 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

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

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

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

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

6.1 Proofs of two useful claims

Following the proof outline above we start by proving Claims 31 and 32. As a first step we prove the following lemma.

Lemma 33. *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

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

Proof. Recall that $\hat{f}(0) = \mathbb{E}[\omega^{f(x)}]$ and similarly for g. Calculating we get,

$$\begin{aligned}
\sum_{\alpha:\alpha_S=0} \widehat{f}(\alpha)\overline{\widehat{g}(\alpha)} &= \sum_{\alpha:\alpha_S=0} (\mathbb{E}_x \omega^{f(x)}\omega^{-\langle x,\alpha\rangle})(\mathbb{E}_y \omega^{-g(y)}\omega^{\langle y,\alpha\rangle}) \\
&= \frac{1}{p^{2n}} \sum_{x,y} \omega^{f(x)-g(y)} \sum_{\alpha:\alpha_S=0} \omega^{\langle y-x,\alpha\rangle} \\
&= \frac{1}{p^{2n}} \sum_{x_{\bar{S}}=y_{\bar{S}}} p^{n-|S|}\omega^{f(x)-g(y)} \\
&= \mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{f(x)-g(y)}] .
\end{aligned}$$

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,

$$\begin{aligned}
|\Delta_{S'}(f,g)| = \left| \sum_{\alpha\neq 0,\alpha_{S'}=0} \widehat{f}(\alpha)\overline{\widehat{g}(\alpha)} \right| &\le \left(\sum_{\alpha\neq 0,\alpha_{S'}=0} |\widehat{f}(\alpha)|^2 \right)^{1/2} \left(\sum_{\alpha\neq 0,\alpha_{S'}=0} |\widehat{g}(\alpha)|^2 \right)^{1/2} \\
&\le \left(\sum_{\alpha\neq 0,\alpha_{S}=0} |\widehat{f}(\alpha)|^2 \right)^{1/2} \left(\sum_{\alpha\neq 0,\alpha_{S}=0} |\widehat{g}(\alpha)|^2 \right)^{1/2} = (\Delta_S(f,f))^{1/2}(\Delta_S(g,g))^{1/2} .
\end{aligned}$$

□

We now give the proofs of Claims 31 and 32.

Proof of Claim 31. By Lemma 33 we have

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

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,

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

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

$$\begin{aligned}
\Delta_S(f,f)^2 &\le \mathbb{E}_{(x,y)\in\mathcal{D}_S}\left[|\mathbb{E}_{a\in A}[\omega^{f(x+a)-f(y+a)}]|^2 \right] \\
&= \mathbb{E}_{(x,y)\in\mathcal{D}_S}\left[(\mathbb{E}_{a\in A}[\omega^{f(x+a)-f(y+a)}])(\mathbb{E}_{a'\in A}[\omega^{-f(x+a')+f(y+a')}]) \right] \\
&= \mathbb{E}_{a,a'\in A}\mathbb{E}_{(x,y)\in\mathcal{D}_S}\left[\omega^{f(x+a)-f(x+a')}\omega^{f(y+a')-f(y+a)} \right] \\
&= \mathbb{E}_{a,a'\in A}\mathbb{E}_{(x',y')\in\mathcal{D}_S}\left[\omega^{f(x'+a-a')-f(x')}\omega^{f(y')-f(y'+a-a')} \right] \\
&= \mathbb{E}_{a\in A}\mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{f_a(x)-f_a(y)}] \\
&= \mathbb{E}_{a\in A}[\Delta_S(f_a,f_a) + |\widehat{f_a}(0)|^2] .
\end{aligned}$$

□

Proof of Claim 32. We begin by showing an identity on $\mathbb{E}_{a\in A}[|\widehat{f_a}(0)|^2]$ for any subspace A.

Claim 34. *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

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

which is equivalent to

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

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

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

□

6.2 Concluding the proof

We now have all the required ingredients to prove Theorem 30.

Proof of Theorem 30. The proof is by induction on d. The base case is $d = 1$. Let $f(x) = \sum_{i=1}^{n} a_i x_i$ 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 $\widehat{f}(\alpha) = \mathbb{E}_{x_i \in \mathbb{F}_p}[\omega^{a_i x_i}] \prod_{j \neq i} \mathbb{E}_{x_j \in \mathbb{F}_p}[\omega^{(a_j - \alpha_j) x_j}] = 0$. 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 34 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 $\mathbb{E}_{a \in A}[|\widehat{f_a}(0)|^2]$. Consider $S = \cup_{a \in A} S_a$. Claim 32 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 $d \geq 2$ we have $\deg(\widetilde{h_y}) \leq d - 1$. Now we can expect that 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 can get a good estimate of $\omega^{f(x)}$. In particular for a parameter $\sigma \in (0, 1)$ to be determined later, we find k such that the following holds

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

By simple application of Chebyshev's inequality, we estimate the parameter k.

Claim 35. *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

$$\begin{aligned}
\Delta_S(f,f) &= \mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{f(x)-f(y)}] - \mathbb{E}_{x\in\mathbb{F}_p^n}[\omega^{f(x)}]\cdot\overline{\mathbb{E}_{y\in\mathbb{F}_p^n}[\omega^{f(y)}]} \\
&\leq |\mathbb{E}_{(x,y)\in\mathcal{D}_S}[(\omega^{f(x)} - F(x))\omega^{-f(y)}] - \mathbb{E}_{x\in\mathbb{F}_p^n}[\omega^{f(x)} - F(x)]\cdot\overline{\mathbb{E}_{y\in\mathbb{F}_p^n}[\omega^{f(y)}]}| \\
&\quad +|\mathbb{E}_{(x,y)\in\mathcal{D}_S}[F(x)\omega^{-f(y)}] - \mathbb{E}_{x\in\mathbb{F}_p^n}[F(x)]\cdot\overline{\mathbb{E}_{y\in\mathbb{F}_p^n}[\omega^{f(y)}]}| \\
&\leq 2\sigma + |\mathbb{E}_{(x,y)\in\mathcal{D}_S}[F(x)\omega^{-f(y)}] - (\mathbb{E}_x[F(x)])(\mathbb{E}_y[\omega^{-f(y)}])| \\
&\leq 2\sigma + \frac{1}{k\delta}\sum_{i=1}^{k}|\mathbb{E}_{(x,y)\in\mathcal{D}_S}[\omega^{\widetilde{h_{y_i}}(x)-f(y)}] - (\mathbb{E}_x[\omega^{\widetilde{h_{y_i}}(x)}])(\mathbb{E}_y[\omega^{-f(y)}])| \\
&\leq 2\sigma + \frac{1}{k\delta^*}\sum_{i=1}^{k}|\Delta_S(\widetilde{h_{y_i}},f)|.
\end{aligned}$$

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 33 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 the theorem. □

7 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 $\mathbb{Z}_m$.

- 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

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach.* Cambridge University Press, 2009.

[AGHP92] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple construction of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992.

[Baz07] Louay M. J. Bazzi. Polylogarithmic independence can fool dnf formulas. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS '07)*, pages 63–73, Washington, DC, USA, 2007. IEEE Computer Society.

[Bra09] Mark Braverman. Poly-logarithmic independence fools AC_0 circuits. In *Proceedings of the* 24^{th} *Conference on Computational Complexity (CCC '09)*, 2009.

[BV07] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. In *Proceedings of the* 48^{th} *Annual IEEE Symposium on Foundations of Computer Science (FOCS '07)*, pages 41–51, Washington, DC, USA, 2007. IEEE Computer Society.

[Gol08] Oded Goldreich. *Computational Complexity: A Conceptual Perspective.* Cambridge University Press, 2008.

[Hås86] Johan Håstad. *Computational limitations for small-depth circuits.* PhD thesis, MIT, 1986.

[LN90] Nati Linial and Noam Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10:349–365, 1990.

[Lov08] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. In *Proceedings of the* 40^{th} *annual ACM symposium on Theory of computing (STOC '08)*, pages 557–562, New York, NY, USA, 2008. ACM.

[LRTV09] Shachar Lovett, Omer Reingold, Luca Trevisan, and Salil Vadhan. Pseudorandom bit generators that fool modular sums. In *Proceedings of the* 12^{th} *International Workshop and* 13^{th} *International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX '09 / RANDOM '09)*, pages 615–630, Berlin, Heidelberg, 2009. Springer-Verlag.

[LVW93] Michael Luby, Boban Velickovic, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proceedings of the* 2^{nd} *ISTCS*, pages 18–24, 1993.

[MZ09] Raghu Meka and David Zuckerman. Small-bias spaces for group products. In Irit Dinur, Klaus Jansen, Joseph Naor, and José D. P. Rolim, editors, *APPROX-RANDOM*, volume 5687 of *Lecture Notes in Computer Science*, pages 658–672. Springer, 2009.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993.

[Raz87] Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Math. Notes*, 41(4):333–338, 1987.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the* 19^{th} *annual ACM symposium on Theory of computing (STOC '87)*, pages 77–82, New York, NY, USA, 1987. ACM.

[Vio09] Emanuele Viola. The sum of small-bias generators fools polynomials of degree . *Computational Complexity*, 18(2):209–217, 2009.

A Proof for linear polynomials

In this section we give the proof Theorem 4. For convenience we repeat it here.

Theorem (Bit-pseudorandom distribution for linear polynomials)**.** *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 $k = O(p^2 \log 1/\epsilon)$. Then $\mathcal{D}^{p-1} \oplus K$ is bit-pseudorandom distribution against linear polynomials over $\mathbb{F}_p$ with error $O(\epsilon)$.*

Proof. Let $f(x) = \sum_{i=1}^n a_i x_i$ be some linear polynomial. Define the *weight* of f, $\mathrm{wt}(f)$, to be the number of nonzero coefficients in f. We consider two cases. Consider first the case that $\mathrm{wt}(f) \leq k$. In such a case any k-wise independent distribution fools f completely. The distribution $\mathcal{D}^{p-1} \oplus K$ is k-wise independent, hence it fools f with error 0.

We now move to the second case, where $\mathrm{wt}(f) > k$. We will prove that in this case the distribution of $f(x)$ is $O(\epsilon)$-close to the uniform distribution over $\mathbb{F}_p$, both when $x \in \{0,1\}^n$ is chosen uniformly at random and when we choose $x \in \mathcal{D}^{p-1} \oplus K$. Hence these two distributions are $O(\epsilon)$-close to each other. In fact, we shall prove a stronger claim: for any fixed $v \in \{0,1\}^n$, the distribution of $f(x)$ where $x \in \mathcal{D}^{p-1} \oplus v$ is $O(\epsilon)$-close to uniform.

We first note that if $X \in \mathbb{F}_p$ is a distribution, then Fact 20 shows that in order to prove that X is ϵ-close to uniform it suffices to prove that for any $c \in \mathbb{F}_p \setminus 0$ it holds that $\mathbb{E}[\omega^{cX}] \le \epsilon/\sqrt{p}$. Since multiplying by $c \neq 0$ does not change the weight of f, it is enough to prove that if $\mathrm{wt}(f) > k$ then $|\mathbb{E}_{x \in \{0,1\}^n}[\omega^{f(x)}]| \le O(\epsilon)$ and $|\mathbb{E}_{x \in \mathcal{D}^{p-1} \oplus a}[\omega^{f(x)}]| \le O(\epsilon)$.

We first prove the claim for uniform inputs. Note that if $z \in \{0,1\}$ is uniform and $a \neq 0$, then

$$\left|\mathbb{E}_{z \in \{0,1\}}[\omega^{az}]\right| \le 1 - \Omega(1/p) \ .$$

Therefore, as $\mathrm{wt}(f) > k$ we get

$$\left|\mathbb{E}_{x \in \{0,1\}^n}[\omega^{f(x)}]\right| = \prod_{i=1}^{n} \left|\mathbb{E}_{x_i \in \{0,1\}}[\omega^{a_i x_i}]\right| \le (1 - \Omega(1/p))^k = O(\epsilon) \ .$$

We now move to proving the claim for $x \in \mathcal{D}^{p-1} \oplus v$. That is, we wish to prove that

$$|\mathbb{E}_{x \in \mathcal{D}}[\omega^{\sum a_i (x_i^{p-1} \oplus v_i)}]| = O(\epsilon) \ .$$

Define $g : \mathbb{F}_p^n \to \mathbb{F}_p$ as $g(x) = \sum a_i (x_i^{p-1} \oplus v_i)$. Note that g is a polynomial of degree $p-1$, as $x_i^{p-1} \oplus v_i$ is equal to x_i^{p-1} when $v_i = 0$ and is equal to $1 - x_i^{p-1}$ when $v_i = 1$. Since $\mathcal{D}$ is a pseudorandom distribution for degree $p-1$ polynomials with error ϵ, we get that

$$\left|\mathbb{E}_{x \in \mathcal{D}}[\omega^{g(x)}] - \mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{g(x)}]\right| \le \epsilon.$$

Hence it is enough to prove that $\left|\mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{g(x)}]\right| = O(\epsilon)$. For that end, let $y \in \{0,1\}^n$ be distributed as follows: $y_1, \ldots, y_n$ are chosen independently such that $\Pr[y_i = v_i] = 1/p$. Then, for $x \in \mathbb{F}_p^n$ chosen uniformly at random, the distributions of $x^{p-1} \oplus v$ and of y are identical. Moreover it is straightforward to verify that for any $a_i \neq 0$ we have

$$|\mathbb{E}_{y_i}[\omega^{a_i y_i}]| \le 1 - \Omega(1/p^2) \ .$$

The claim now follows as

$$\begin{aligned}\left|\mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{g(x)}]\right| &= \left|\mathbb{E}_{x \in \mathbb{F}_p^n}[\omega^{\sum a_i (x_i^{p-1} \oplus v_i)}]\right| = \left|\mathbb{E}_y[\omega^{\sum a_i y_i}]\right| \\ &= \prod_{i=1}^{n} |\mathbb{E}_{y_i}[\omega^{a_i y_i}]| \le (1 - \Omega(1/p^2))^k = O(\epsilon) \ .\end{aligned}$$

□

B Proof of Fact 20

For completeness we give the proof of the following well known fact.

Fact (Fact 20)**.** *Let $\mathcal{D}_1, \mathcal{D}_2 \subset \mathbb{F}_p^k$ be two distributions. Assume that for every $\alpha \in \mathbb{F}_p^k$ the distributions $\langle \mathcal{D}_1, \alpha \rangle$ and $\langle \mathcal{D}_2, \alpha \rangle$ are ϵ-close. Then $\mathcal{D}_1$ and $\mathcal{D}_2$ are $(p^{k/2}\epsilon)$-close.*

Proof of Fact 20. We need to bound

$$\mathrm{sd}(\mathcal{D}_1, \mathcal{D}_2) = \tfrac{1}{2} \sum_{x \in \mathbb{F}_p^k} |\Pr[\mathcal{D}_1 = x] - \Pr[\mathcal{D}_2 = x]| \; .$$

By the Cauchy-Schwarz inequality we get

$$4 \cdot \mathrm{sd}(\mathcal{D}_1, \mathcal{D}_2)^2 \leq p^k \sum_{x \in \mathbb{F}_p^k} |\Pr[\mathcal{D}_1 = x] - \Pr[\mathcal{D}_2 = x]|^2 \; .$$

Let $\widehat{\mathcal{D}_i} : \mathbb{F}_p^k \to \mathbb{C}$ be the Fourier transform of $\mathcal{D}_i$, when we think of $\mathcal{D}_i$ as the function $\mathcal{D}_i(y) = \Pr_{x \in \mathcal{D}_i}[x = y]$. In other words, $\widehat{\mathcal{D}_i}(\alpha) = \mathbb{E}_{x \in \mathbb{F}_p^k}[\Pr[\mathcal{D}_i = x]\omega^{-\langle x, \alpha \rangle}]$. By the Parseval identity we get that

$$4 \cdot \mathrm{sd}(\mathcal{D}_1, \mathcal{D}_2)^2 \leq p^{2k} \cdot \sum_{\alpha \in \mathbb{F}_p^k} |\widehat{\mathcal{D}_1}(\alpha) - \widehat{\mathcal{D}_2}(\alpha)|^2 \; .$$

From the assumption that for every $\alpha \in \mathbb{F}_p^k$ the distributions $\langle \mathcal{D}_1, \alpha \rangle$ and $\langle \mathcal{D}_2, \alpha \rangle$ are ϵ-close we obtain

$$\begin{aligned} |\widehat{\mathcal{D}_1}(\alpha) - \widehat{\mathcal{D}_2}(\alpha)| &= \left| E_{t \in \mathbb{F}_p^k}[(Pr[\langle \mathcal{D}_1, \alpha \rangle = t] - Pr[\langle \mathcal{D}_2, \alpha \rangle = t]) \cdot \omega^{-t}] \right| \\ &\leq \mathbb{E}_{t \in \mathbb{F}_p^k} |[Pr[\langle \mathcal{D}_1, \alpha \rangle = t] - Pr[\langle \mathcal{D}_2, \alpha \rangle = t]]| \\ &\leq 2\epsilon / p^k \; . \end{aligned}$$

Thus we conclude that

$$4 \cdot \mathrm{sd}(\mathcal{D}_1, \mathcal{D}_2)^2 \leq p^{2k} \sum_{\alpha \in \mathbb{F}_p^k} (2\epsilon/p^k)^2 = 4p^k\epsilon^2 \; .$$

The bound $\mathrm{sd}(\mathcal{D}_1, \mathcal{D}_2) \leq p^{k/2}\epsilon$ now follows. □