Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 65 (2010)

New extension of the Weil bound for character sums with applications to coding

Tali Kaufman
Bar-Ilan University and The Weizmann Institute of Science
kaufmant@mit.edu

Shachar Lovett *
Institute of Advanced Study
slovett@math.ias.edu

November 10, 2010

Abstract

The Weil bound for character sums is a deep result in Algebraic Geometry with many applications both in mathematics and in the theoretical computer science. The Weil bound states that for any polynomial $f(x)$ over a finite field $\mathbb{F}$ and any additive character $\chi : \mathbb{F} \to \mathbb{C}$, either $\chi(f(x))$ is a constant function or it is distributed close to uniform. The Weil bound is quite effective as long as $\deg(f) \ll \sqrt{|\mathbb{F}|}$, but it breaks down when the degree of f exceeds $\sqrt{|\mathbb{F}|}$. As the Weil bound plays a central role in many areas, finding extensions for polynomials of larger degree is an important problem with many possible applications.

In this work we develop such an extension over finite fields $\mathbb{F}_{p^n}$ of small characteristic: we prove that if $f(x) = g(x) + h(x)$ where $\deg(g) \ll \sqrt{|\mathbb{F}|}$ and $h(x)$ is a sparse polynomial of arbitrary degree but bounded weight degree, then the same conclusion of the classical Weil bound still holds: either $\chi(f(x))$ is constant or its distribution is close to uniform. In particular, this shows that the subcode of Reed-Muller codes of degree $\omega(1)$ generated by traces of sparse polynomials is a code with near optimal distance, while Reed-Muller of such a degree has no distance (i.e. $o(1)$ distance) ; this is one of the few examples where one can prove that sparse polynomials behave differently from non-sparse polynomials of the same degree.

As an application we prove new general results for affine invariant codes. We prove that any affine-invariant subspace of quasi-polynomial size is (1) indeed a code (i.e. has good distance) and (2) is locally testable. Previous results for general affine invariant codes were known only for codes of polynomial size, and of length 2^n where n needed to be a prime. Thus, our techniques are the first to extend to general families of such codes of super-polynomial size, where we also remove the requirement from n to be a

*Supported by NSF grant DMS-0835373.

ISSN 1433-8092

prime. The proof is based on two main ingredients: the extension of the Weil bound for character sums, and a new Fourier-analytic approach for estimating the weight distribution of general codes with large dual distance, which may be of independent interest.

Contents

1 Introduction

In this work we provide a new extension to the Weil bound for character sums. Additionally, we develop a new approach for estimating the weight distribution of general codes whose dual has large distance, which greatly extends the method of Krawtchouk polynomials. We combine these results for obtaining better understanding of general families of affine invariant codes of quasi-polynomial size, extending previous results which could only handle such codes of polynomial size.

1.1 Character sums

Let $\mathbb{F}$ be a finite field. An *additive character* is a function $\chi : \mathbb{F} \to \mathbb{C}$ for which $\chi(x+y) = \chi(x)\chi(y)$ (and which is not the identically zero function). For $\mathbb{F} = \mathbb{F}_{p^n}$, the additive characters are given by $\chi_a(x) = e^{\frac{2\pi i}{p}\mathrm{Tr}(ax)}$, where $a \in \mathbb{F}_{p^n}$ and the Trace operator $\mathrm{Tr} : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is defined as $\mathrm{Tr}(x) = \sum_{i=0}^{n-1} x^{p^i}$.

The Weil bound for character sums [23] is a deep result from Algebraic Geometry. The result deals with character sums of low-degree polynomials over a finite field $\mathbb{F}$. Let $f(x) \in \mathbb{F}[x]$ be a univariate polynomial of degree at most $\sqrt{|\mathbb{F}|}$. Let $\chi : \mathbb{F} \to \mathbb{C}$ be any additive character. Weil's bound states that either $\chi(f(x))$ is constant, or is distributed close to uniform when $x \in \mathbb{F}$ is uniformly chosen.

Theorem 1.1 (Weil bound [23])**.** *Let $f(x)$ be a univariate polynomial over $\mathbb{F}$ of degree $\leq |\mathbb{F}|^{1/2-\delta}$. Let $\chi : \mathbb{F} \to \mathbb{C}$ be any additive character. Then either $\chi(f(x))$ is constant for all $x \in \mathbb{F}$, or*

$$|\mathbb{E}_{x\in\mathbb{F}}[\chi(f(x))]| \leq |\mathbb{F}|^{-\delta}.$$

The Weil bound, being a general and powerful result, has found many applications in mathematics and also in theoretical computer science, in particular in the areas of pseudorandomness, explicit constructions and coding theory. For example, it has been used in the study of extractors ([9, 24]) and in the study of locally testable codes ([8, 16]). The Weil bound is very effective for polynomials of degree $\ll \sqrt{|\mathbb{F}|}$, however it fails for polynomials of degree exceeding $\sqrt{|\mathbb{F}|}$. We establish a general result in fields of small characteristics $\mathbb{F}_{p^n}$ which allows to extend polynomials by a small number of monomials of larger degree, as long as they have small *weight degree*. In particular, in some range of the parameters we may add $O(n)$ monomials, while in another range we can add monomials of degree $p^{n-\log n}$. Both of these extend the classic Weil bound significantly.

Definition 1.2 (Weight degree)**.** Let $t \in \{0, \ldots, p^n - 1\}$. The weight degree of t is the hamming weight of the digits of t in base p. That is, let $t = \sum_{i=0}^{n-1} t_i p^i$ be the representation of t in base p, where $0 \leq t_i \leq p-1$. The weight degree of t is

$$\mathrm{wt}(t) = \sum_{i=0}^{n-1} t_i.$$

The weight degree of a monomial x^t is the weight degree of t, and the weight degree of a univariate polynomial $f(x)$ is the maximal weight degree of a monomial in it with a nonzero coefficient.

We prove the following extension of the Weil bound in case $f(x)$ is the sum of a low degree polynomial and a small number of monomials of bounded weight degree (but of arbitrary degree).

Theorem 1.3 (Extension of the Weil bound)**.** *Let $f(x) = g(x) + h(x)$ be a univariate polynomial over $\mathbb{F}_{p^n}$, where $g(x)$ is a polynomial of degree $\leq |\mathbb{F}|^{1/2-\delta}$ and $h(x)$ is the sum of at most $k \geq 1$ monomials, each of weight degree at most d. Let $\chi : \mathbb{F}_{p^n} \to \mathbb{C}$ be an additive character. Then either $\chi(f(x))$ is constant for all $x \in \mathbb{F}_{p^n}$, or*

$$|\mathbb{E}_{x\in\mathbb{F}}[\chi(f(x))]| \leq |\mathbb{F}_{p^n}|^{-\frac{\delta}{2d^2 2^d k}}.$$

Note that in order to get a meaningful bound, we need our parameters to obey $kd^2 2^d \leq O(n)$. Note that for $d \leq (1-\epsilon)\log_2(n)$ we may have $k = n^{O(1)}$. This can be compared to a relatively recent result of Bourgain [2] of a similar flavor. We state it below informally, as the exact formulation is somewhat complex, and we will not require it in the paper.

Theorem 1.4 (Bourgain's extension of Weil bound [2])**.** *Let $f(x) = g(x)+h(x)$ be a univariate polynomial over a prime finite field $\mathbb{F}_q$, where $g(x)$ is a polynomial of degree $\leq |\mathbb{F}_q|^{1/2-\delta}$ and $h(x)$ is the sum of at most $k = O(1)$ monomials, each of degree at most $|\mathbb{F}_q|^{1-\epsilon}$. Let $\chi : \mathbb{F}_q \to \mathbb{C}$ be an additive character. Then either $\chi(f(x))$ is constant for all $x \in \mathbb{F}_q$, or*

$$\left|\mathbb{E}_{x\in\mathbb{F}_q}[\chi(f(x))]\right| \leq |\mathbb{F}_q|^{-\Omega(1)}.$$

Comparing our result with the result of Bourgain, we note several important advantages of our work: first, we can handle non-prime finite fields; second, when $d \leq O(\log n)$ is small enough, we may have $k = poly(n)$ monomials of high degree, while in the result of Bourgain one can take at most $k = O(1)$ such monomials; Third, we can handle additional monomials with degree up to $p^{n-\log n}$, while Bourgain result (even if worked for non prime fields) would allow degree bounded by $p^{n/c}$ for some constant $c < 1$. In contrast, the result of Bourgain does not assume a bound on the weight degree of the monomials. The advantages of our work are crucial for our applications to estimating the weight distributions of codes, and for local testability of codes.

Finally, we view Theorem 1.3 as an important step towards understanding sparse polynomials. Sparse polynomials arise naturally in many areas of theoretical computer science, most notably in circuit complexity and learning theory. To date, our understanding of the behavior of sparse polynomial has been quite limited. An immediate corollary of Theorem 1.3 gives what is, to the best of our knowledge, the first result which separates the behavior of sparse polynomials from general polynomials (of the same degree), in the context of small finite fields.

The Reed-Muller code $\mathrm{RM}_p(n,d)$ is a code generated by all n-variate polynomials over $\mathbb{F}_p$ of total degree at most d. It can equivalently be described as $\mathcal{T}(\{e \in \mathbb{F}_p^n : \mathrm{wt}(e) \leq d\})$, i.e. codewords are traces of univariate polynomials of $\mathbb{F}_{p^n}$ of weight degree at most d. The minimal distance of $\mathrm{RM}_p(n,d)$ is well known; in particular, whenever $d = \omega(1)$ the minimal distance is $o(1)$. To the contrast, let $\mathcal{C}$ be a (nonlinear) code generated by traces of sparse polynomials. Our results show that the code $\mathcal{C} \subset \mathrm{RM}_p(n,d)$ has far better minimal distance; in fact, it has near optimal distance. This argument holds even when $d = O(\log n)$ and the sparsity is $n^{O(1)}$. Previous similar results were only known for constant sparsity.

Corollary 1.5. *Fix $d \le O(\log n)$. Let $t_1, \dots, t_k \in [p^n - 1]$ be chosen of weight degree at most d, where $k = O(\frac{n}{d^2 2^d})$. Consider the code $\mathcal{C} = \{\mathrm{Tr}(\sum_{i=1}^{k} a_i x^{t_i}) : a_i \in \mathbb{F}_{p^n}\}$. Then*

1. *$\mathcal{C}$ is a subcode of $\mathrm{RM}_p(n, d)$;*
2. *The minimal distance of $\mathcal{C}$ is at least $1 - 1/p - p^{-\Omega(n)}$.*

1.2 Weight distribution of linear codes

Using a Fourier-analytic technique we show new estimates of the weight distribution of linear codes with large dual distance. This result combined with our new extension to the Weil bound imply estimation of the weight distribution of every affine-invariant code of super-polynomial size.

A code is a subset $\mathcal{C} \subset \mathbb{F}_p^N$, which can equivalently be viewed as a family of functions $\mathcal{C} = \{f : [N] \to \mathbb{F}_p\}$. All codes we consider in this work are linear[1]. The dimension of a code is $\dim(\mathcal{C}) = \log_p(|\mathcal{C}|)$.

Let $\mathcal{C} \subset \mathbb{F}_p^N$ be any linear code. Let $\mathrm{Const} = \{a^N : a \in \mathbb{F}_p\}$ be the linear code of constant words. Note that as $\mathcal{C}$ is a linear code, then either $\mathrm{Const} \subset \mathcal{C}$ or $\mathrm{Const} \cap \mathcal{C} = \{0^N\}$. We define the distance between $\mathcal{C}$ and the code of constant words as the minimal distance between a nonconstant codeword of $\mathcal{C}$ and a constant word,

$$\mathrm{dist}(\mathcal{C}, \mathrm{Const}) = \min_{f \in \mathcal{C} \setminus \mathrm{Const}} \min_{a \in \mathbb{F}_p} \frac{|\{i \in [N] : f_i \neq a\}|}{N}.$$

The dual of a linear code is defined as

$$\mathcal{C}^\perp = \{g \in \mathbb{F}_p^N : \sum_{i=1}^{N} f_i g_i = 0\}.$$

We prove the following theorem, which gives a tight estimation on the weight distribution of $\mathcal{C}^\perp$ based on the distance between $\mathcal{C}$ and the constant word codes. Previous results on the weight distribution of general codes (e.g. [11]) were based on the use of Krawtchouk polynomials. These results applied only to binary codes whose duals have distance very close to 1/2. I.e. they didn't apply to codes with some arbitrary constant distance as we have here.

Theorem 1.6 (Weight distribution result). *Let $\mathcal{C} \subset \mathbb{F}_p^N$ be a linear code, and assume that $\mathrm{dist}(\mathcal{C}, \mathrm{Const}) = \delta > 0$. For every $\epsilon > 0$ there exist $\ell_{\min} = O(\frac{1}{\delta}\log(|\mathcal{C}|) + \log(1/\epsilon))$ and $\ell_{\max} = O(\sqrt{\epsilon N})$, such that for any $\ell \in [\ell_{\min}, \ell_{\max}]$ the following holds. The number of codewords $g \in \mathcal{C}^\perp$ of weight exactly ℓ is given by $\alpha \cdot \frac{N^\ell}{|\mathcal{C}|}(1 \pm \epsilon)$, where*

- $\alpha = \frac{(p-1)^\ell}{\ell!}$ *if* $\mathcal{C} \cap \mathrm{Const} = \{0^N\}$.
- $\alpha = \frac{C(p,\ell)}{\ell!}$ *if* $\mathrm{Const} \subset \mathcal{C}$, *where* $C_{p,\ell} = |\{v_1, \dots, v_\ell \in \mathbb{F}_p \setminus \{0\} : \sum_{i=1}^{\ell} v_i = 0\}|$.

[1] A code $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$ is linear if for any $f(x), g(x) \in \mathcal{C}$ also $h(x) = \alpha f(x) + \beta g(x) \in \mathcal{C}$ where $\alpha, \beta \in \mathbb{F}_p$.

1.3 Application to Locally testable codes

Let $\mathbb{F}_N = \mathbb{F}_{p^n}$ be a finite field, where we think of p as either constant or small. In this context, a code is a family of functions $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$. A code is *locally testable* if there is a randomized algorithm, which when given as input a function $f : \mathbb{F}_{p^n} \to \mathbb{F}_p$, probes f in a small number of locations and determines (with high probability) whether $f \in \mathcal{C}$ or f is far[2] from all codewords of $\mathcal{C}$. A code is q-locally testable if the number of probes is at most q, where q is sublinear in the code length, i.e. $q = o(N)$.

A recent line of research in property testing focuses on characterization of *general* families of codes that are locally testable [11, 14, 8, 17]. The known results for general codes that are locally testable apply only to *sparse* codes over binary fields $\mathbb{F}_2$, which are codes of size $N^{O(1)}$. This is in contrast to result for specific families of codes (such as Reed-Muller codes) for which much better results are known. It is an important problem to better understand general codes. One reason is that such an understanding might aid in finding specific codes with better parameters; while another is to understand the extremal properties of such codes.

In this work we break the sparsity requirement for local testability of general codes. Namely, we exhibit a general family of codes of size $N^{(\log N)^{O(1)}}$ that are locally testable with $\log N^{O(1)}$ queries. We achieve this by studying *affine invariant codes.* A code $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$ is affine invariant if it is invariant under affine transformation of the coordinates of the input space. That is, if $f(x) \in \mathcal{C}$ then also $g(x) = f(ax + b) \in \mathcal{C}$ for any $a, b \in \mathbb{F}_{p^n}, a \neq 0$. Previous results [8] showed that sparse affine invariant codes over $\mathbb{F}_2$ of length 2^n for *prime* n (i.e., codes of size $N^{O(1)}$) are locally testable. We significantly extend this to codes of size super-polynomial in N, i.e. to codes of size at most $N^{(\log N)^{O(1)}}$. Moreover, we remove the requirement from n to be a prime.

Theorem 1.7 (Testing result (informal))**.** *Let $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$ be a linear subspace which is affine invariant of size $N^{(\log N)^\epsilon}$ for any $\epsilon < \frac{\log p}{\log 2p}$ and n large enough. Then the following holds:*

- *C is a code, namely it has a constant distance.*
- *$\mathcal{C}$ is locally testable with query complexity $q = poly(\dim(\mathcal{C})/n)$.*

In particular, any sparse affine invariant code (i.e. with $\dim(\mathcal{C}) = O(n)$) is locally testable with constant query complexity $q = O(1)$.

Our result generalizes the result of Grigorescu, Kaufman and Sudan [8] in few aspects: First, the result of [8] applies only to sparse codes (i.e codes of size $N^{O(1)}$) while our result applies to codes with super polynomial number of codewords (i.e. to codes of size at most $N^{O(\log N)}$). Second, the result of [8] could work only for fields of size 2^n where n needed to be prime, while we remove the requirement for n to be prime. Third, we provide a *self-contained* proof of a generalization of [8], which used complex machinery (such as Bourgain's extension to the Weil bound, and properties of Krawtchouk polynomials). Moreover, previous results on the testability of sparse codes applied only to binary fields $\mathbb{F}_2$, while our result applies to

[2] If f has distance ϵ from $\mathcal{C}$, i.e. if $\min_{g \in \mathcal{C}} \Pr_{x \in \mathbb{F}_{p^n}}[f(x) \neq g(x)] = \epsilon$, we require the local test to reject f with probability at least $\Omega(\epsilon)$.

any field of small characteristic. The testing result uses our new extension to the Weil bound as well as our new estimation on the weight distribution of codes with large dual distance.

1.4 Proofs overview

1.4.1 New extension to the Weil bound

The proof of our new extension for the Weil bound relies on techniques borrowed from additive combinatorics. This demonstrates yet another connection between additive combinatorics and theoretical computer science. Such connections were used before to establish results regarding pseudorandom generators [5, 18, 22] and list-decoding of codes [12].

We sketch in high level how we achieve the new extension to the Weil bound. Let $f(x) = g(x) + h(x)$ be a univariate polynomial over $\mathbb{F}_{p^n}$, where $\deg(g) \leq |\mathbb{F}_{p^n}|^{1/2-\delta}$ and $h(x)$ is the sum of k monomials, each of weight degree at most d. We need to prove that either $\mathrm{Tr}(f) : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is a constant function, or that it is highly unbiased (note that proving the result for the Trace operator implies it immediately for all additive characters).

The analysis divides into two cases: either g has high weight-degree $\mathrm{wt}(g) \geq d+1$, or g has low weight-degree $\mathrm{wt}(g) \leq d$. The first case is the easier one, and both cases rely on an analysis of directional derivatives of polynomials. The directional derivative of a polynomial $f(x)$ in direction $y \in \mathbb{F}_{p^n}$ is given by $f_y(x) = f(x+y) - f(x)$, and iterated derivatives are defined as $f_{y_1,\ldots,y_k}(x) = (f_{y_1,\ldots,y_{k-1}})_{y_k}(x)$.

The case of high weight g The first case, where $\mathrm{wt}(g) \geq d+1$ is easy to analyze by taking enough derivatives that eliminate $h(x)$, and reducing to a theorem of Deligne [7], which is a multivariate analog of Weil's bound. Specifically, For any $y_1, \ldots, y_{d+1}$ one can verify that since $\mathrm{wt}(h) \leq d$ then

$$h_{y_1,\ldots,y_{d+1}} \equiv 0,$$

hence $f_{y_1,\ldots,y_{d+1}} \equiv g_{y_1,\ldots,y_{d+1}}$. An iterated application of the Cauchy-Schwarz inequality yields that

$$\left|\mathbb{E}_{x\in\mathbb{F}_{p^n}}\left[\omega^{\mathrm{Tr}(f(x))}\right]\right|^{2^{d+1}} \leq \left|\mathbb{E}_{x,y_1,\ldots,y_{d+1}\in\mathbb{F}_{p^n}}\left[\omega^{\mathrm{Tr}(f_{y_1,\ldots,y_{d+1}}(x))}\right]\right|$$

where $\omega = e^{\frac{2\pi i}{p}}$. Hence to prove that $\mathrm{Tr}(f(x))$ in unbiased for uniform x, it is sufficient to prove that $\mathrm{Tr}(f_{y_1,\ldots,y_{d+1}}(x))$ is unbiased for uniform $x, y_1, \ldots, y_{d+1}$. We then verify that as g is of weight degree at least $d+1$, it is not eliminated by taking generic $d+1$ derivatives, and we get that $f_{y_1,\ldots,y_{d+1}}(x)$ is a nonzero polynomial in the variables $x, y_1, \ldots, y_{d+1}$ of total degree at most $\deg(g) \leq |\mathbb{F}_{p^n}|^{1/2-\delta}$. Moreover, we can prove that $\mathrm{Tr}(f_{y_1,\ldots,y_{d+1}}(x))$ is not a constant function; hence by Deligne's theorem we deduce that

$$\left|\mathbb{E}_{x,y_1,\ldots,y_{d+1}\in\mathbb{F}_{p^n}}\left[\omega^{\mathrm{Tr}(f_{y_1,\ldots,y_{d+1}}(x))}\right]\right| \leq |\mathbb{F}|^{-\delta}$$

and the bound on the bias of $\mathrm{Tr}(f(x))$ follows.

The case of low weight g The harder case is handling g of small weight $\mathrm{wt}(g) \leq d$, since h cannot simply be eliminated by taking enough iterated derivatives, without eliminating f altogether. We solve this problem by taking a smaller number of derivatives, such that f is not eliminated, but instead is transformed into a special class of polynomials (p-multilinear polynomials). We then proceed to study this family of polynomials, and are able to bound the bias of such polynomials, given that they came from a polynomial $f = g + h$ where g has low degree and h is the sum of a small number of low weight degree monomials. Most of the technical challenges of the proof are in this part.

1.4.2 Estimations on weight distribution of codes

Following we describe our approach for estimating the weight distribution of general codes whose duals have large distance. A central notion that is useful here is the *bias* of a code. The bias of a codeword $f \in \mathcal{C}$ is defined as

$$\mathrm{bias}(f) = \left|\mathbb{E}_{x\in[N]}[\omega^{f(x)}]\right| = \left|\frac{1}{N}\sum_{x\in[N]}\omega^{f(x)}\right|,$$

where $\omega = e^{2\pi i/p}$. We define the bias of a code as the maximal bias of a nonconstant codeword:

$$\mathrm{bias}(\mathcal{C}) = \max_{f\in\mathcal{C}\setminus\mathrm{Const}} \mathrm{bias}(f).$$

Note that always $\mathrm{bias}(\mathcal{C}) < 1$ and that as the distance of the code gets larger the bias of the code gets smaller.

We relate the codewords in $\mathcal{C}^\perp$ with the following sets. For $v = (v_1, \ldots, v_\ell) \in \{1, \ldots, p-1\}^\ell$ define the sets

$$A_\ell(v) = \{(x_1, \ldots, x_\ell) \in [N]^\ell : \sum_{i=1}^{\ell} v_i f(x_i) = 0 \quad \forall f \in \mathcal{C}\}$$

and

$$B_\ell(v) = \{(x_1, \ldots, x_\ell) \in A_\ell(v) : x_1, \ldots, x_\ell \text{ are all distinct}\}.$$

It follows from the definition that number of codewords in $\mathcal{C}^\perp$ of weight ℓ is $\frac{1}{\ell!}\sum_{v\in\{1,\ldots,p-1\}^\ell}|B_\ell(v)|$. Hence, to obtain our estimation on the weight distribution of $\mathcal{C}^\perp$ we need to show that $|B_\ell(v)| \approx N^\ell/|\mathcal{C}|$. The main step is to show that $|A_\ell(v)| \approx N^\ell/|\mathcal{C}|$. From the last we deduce the estimate for $|B_\ell(v)|$. For estimating $|A_\ell(v)|$, we take $(x_1, \ldots, x_\ell) \in [N]^\ell$, and consider

$$\mathbb{E}_{f\in\mathcal{C}}\left[\omega^{v_1 f(x_1)+\ldots+v_\ell f(x_\ell)}\right],$$

The above expectation is 1 iff $(x_1, \ldots, x_\ell) \in A_\ell(v)$ and otherwise it is 0. This holds since $\mathcal{C}$ is a linear subspace. I.e., either the inner product of v with $(f(x_1), \ldots, f(x_\ell))$ is always zero; or it is uniformly distributed over $\mathbb{F}_p$ when $f \in \mathcal{C}$ is uniformly chosen. Hence we have

$$N^{-\ell}|A_\ell(v)| = \mathbb{E}_{x_1,\ldots,x_\ell\in[N]}\mathbb{E}_{f\in\mathcal{C}}\left[\omega^{v_1 f(x_1)+\ldots+v_\ell f(x_\ell)}\right] = \mathbb{E}_{f\in\mathcal{C}}\prod_{i=1}^{\ell}\mathbb{E}_{x_i\in[N]}\left[\omega^{v_i f(x_i)}\right]$$

We partition the expectation to the cases where $f = 0^N$ and $f \neq 0^N$. When $f = 0^N$ then for all $i = 1, \ldots, \ell$ we have that $\mathbb{E}_{x_i \in [N]}\left[\omega^{v_i f(x_i)}\right] = 1$. If f is non constant $\left|\mathbb{E}_{x_i \in [N]}\left[\omega^{v_i f(x_i)}\right]\right| \leq \text{bias}(\mathcal{C}) \leq \delta$. Hence we deduce that

$$|A_\ell(v)| = \frac{N^\ell}{|\mathcal{C}|}(1+\eta)$$

where $|\eta| \leq |\mathcal{C}|\delta^\ell$. I.e. for codes with small bias (that is of large distance) $|A_\ell(v)| \approx N^\ell/|\mathcal{C}|$. We use our extension to the Weil bound to show that affine invariant subspaces of super-polynomial size have very small bias (and hence are in fact codes), and form this we deduce estimation on the weight distribution of their duals.

1.4.3 The connection between character sums and the testability of affine invariant codes

We sketch in high level how we achieve our improved testability result for affine-invariant codes using the new extension to the Weil bound. Basically, we follow the proof idea of [8]. They use Bourgain's result for character sums, as well as properties of Krawtchouk polynomials. We replace these ingredients with our new expansion to the Weil bound and our new estimation to the weight distribution of linear codes.

Affine invariant codes can be characterized by trace codes. Let $S \subseteq \{0, \ldots, p^n - 1\}$. The S-trace code over $\mathbb{F}_{p^n}$ is defined as the family of functions $f : \mathbb{F}_{p^n} \to \mathbb{F}_p$ given by

$$\mathcal{T}(S) = \left\{ \left(\text{Tr}(\sum_{e \in S} a_e x^e) : F_{p^n} \to \mathbb{F}_p \right) : a_e \in \mathbb{F}_{p^n} \right\}.$$

where we recall that the Trace function $\text{Tr} : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is given by $\text{Tr}(x) = \sum_{i=0}^{n-1} x^{p^i}$. For example, Generalized Reed-Muller codes $\text{RM}_p(n, d)$, which are the family of functions $f : \mathbb{F}_p^n \to \mathbb{F}_p$ where f is an n-variate polynomial of total degree at most d, can be equivalently characterized as

$$\text{RM}_p(n, d) = \mathcal{T}(\{e \in \{0, \ldots, p^n - 1\} : \text{wt}(e) \leq d\}).$$

We define two important properties of trace codes.

Definition 1.8 (Shift closed)**.** Let $S \subseteq \{0, \ldots, p^n - 1\}$. The set S is said to be *shift closed* if, for every $e \in S$, we also have that $ep^\ell \pmod{p^n} \in S$ for all $\ell = 1, \ldots, n$.

The term *shift closed* comes from viewing elements $e \in S$ as vectors in $\mathbb{F}_p^n$, given by the representation of e in base p. In this case, $ep^\ell \pmod{p^n}$ corresponds to a cyclic shift of the vector by ℓ coordinates.

Definition 1.9 (Shadow closed)**.** Let $S \subseteq \{0, \ldots, p^n - 1\}$. The set S is said to be *shadow closed* if the following holds. For any $e \in S$, let $e = \sum_{i=0}^{n-1} e_i p^i$ be the representation of e in base p. Define the *support* of e to be the set of nonzero digits of e,

$$\text{support}(e) = \{0 \leq i \leq n-1 : e_i \neq 0\}.$$

Let e' be obtained from e by changing some of the non-zero digits of e, i.e.

$$e' = \sum_{i \in \text{support}(e)} e'_i p^i.$$

Then we should have that also $e' \in S$. That is, S is shadow closed if

$$\left\{ \sum_{i \in \text{support}(e)} e'_i p^i : e \in S, (e'_i)_{i \in \text{support}(e)} \in \mathbb{F}_p \right\} \subseteq S.$$

A set S is said to be *affine closed* if it is both shift closed and shadow closed. The following general result was established by Kafuman and Sudan [15]. They show that the class of affine invariant linear codes is equivalent to the class of trace codes of affine closed sets.

Theorem 1.10 (Monomial extraction [15]). *Let $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$ be an affine invariant linear code. Then there exists an affine closed set $S \subseteq \{0, \ldots, p^n - 1\}$ such that $\mathcal{C} = \mathcal{T}(S)$. Moreover, for any affine closed set S the code $\mathcal{T}(S)$ is linear and affine invariant.*

Thus, to study affine invariant codes, we need to study trace codes. Recall, the dual of a code $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$ is defined as

$$\mathcal{C}^\perp = \left\{ (g : \mathbb{F}_{p^n} \to \mathbb{F}_p) : \sum_{x \in \mathbb{F}_{p^n}} f(x)g(x) = 0 \quad \forall f \in \mathcal{C} \right\}.$$

The *affine closure* of a function $g : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is the set of functions obtained by applying affine transformations on the coordinates of the input space of f, that is

$$\overline{\text{affine}}(g) = \left\{ (g(ax+b) : \mathbb{F}_{p^n} \to \mathbb{F}_p) : a, b \in \mathbb{F}_{p^n} \right\}.$$

It is easy to verify that if $\mathcal{C}$ is an affine invariant code, and $g \in \mathcal{C}^\perp$, then in fact $\overline{\text{affine}}(g) \subseteq \mathcal{C}^\perp$. An important case is when in fact $\overline{\text{affine}}(g)$ spans the entire code $\mathcal{C}^\perp$.

Definition 1.11 (Single orbit property). Let $g \in \mathcal{C}^\perp$. We say that $\mathcal{C}$ has the *single orbit property* for g if the affine closure of g is a spanning set for $\mathcal{C}^\perp$, that is if

$$\mathcal{C} = \text{Span}(\overline{\text{affine}}(g))^\perp.$$

We will shortly see that the single orbit property is tightly connected to locally testing properties of the code $\mathcal{C}$. First, define the *weight* of $g : \mathbb{F}_{p^n} \to \mathbb{F}_p$ to be the number of coordinates where g evaluates to a nonzero value,

$$\text{wt}(g) = |\{x \in \mathbb{F}_{p^n} : g(x) \neq 0\}|.$$

The following result was established by Kaufman and Sudan [15]. If $\mathcal{C}$ is an affine invariant code which has the single orbit property for a codeword $g \in \mathcal{C}^\perp$ of small weight, then $\mathcal{C}$ can be locally tested[3].

[3] In fact, the local test for $\mathcal{C}$ is performed by computing $\sum f(ax+b)g(x)$ for a small random subset of $a, b \in \mathbb{F}_{p^n}$. Note that to perform each such test, we only need to query $f(x)$ only on $x \in \mathbb{F}_{p^n}$ for which $g(x) \neq 0$.

Theorem 1.12 (Theorem 2.9 in [15]). *Let $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$ be a linear code which is affine invariant. Assume there exists $g \in \mathcal{C}^\perp$ such that $\mathcal{C}$ has the single orbit property for g. Then $\mathcal{C}$ can be locally tested with $O(\mathrm{wt}(g)^2)$ queries.*

Hence, to show that $\mathcal{C}$ can be locally tested, it is sufficient to demonstrate that $\mathcal{C}^\perp$ is spanned by the orbit of a short codeword under the affine group.

Let $\mathcal{C} = \mathcal{T}(S)$ for some affine closed set $S \subseteq \{0, \ldots, p^n - 1\}$. The dual code of $\mathcal{C}$ is a dual-trace code $d\mathcal{T}(S)$, which can be verified (Claim 4.3) to be

$$d\mathcal{T}(S) = \left\{ (f : \mathbb{F}_{p^n} \to \mathbb{F}_p) : \sum_{x \in \mathbb{F}_{p^n}} f(x)x^e = 0 \quad \forall e \in S \right\}.$$

We need to establish that there exists $f \in d\mathcal{T}(S)$ of small weight such that $\mathrm{Span}(\overline{\mathrm{affine}}(f)) = d\mathcal{T}(S)$. Assume that this is false, i.e. that $\mathrm{Span}(\overline{\mathrm{affine}}(f)) \subsetneq d\mathcal{T}(S)$. Using the fact that S is affine invariant, we show (Corollary 4.21) that in fact $f \in d\mathcal{T}(S \cup \{e\})$ where $e \in \{0, \ldots, p^n - 1\} \setminus S$ has small weight.

Hence, in order to conclude the proof, we will show that for a suitably chosen weight ℓ, there exist codewords on weight ℓ in $d\mathcal{T}(S)$ which are not in any of $d\mathcal{T}(S \cup \{e\})$ for any $e \notin S$ which has small weight. The main tool we develop in order to do so, is a tight estimate on the number of codewords of weight ℓ in dual-trace codes. We show the following result.

Lemma (Lemma 4.14, informal statement). *Let $S \subseteq \{0, \ldots, p^n - 1\}$ be affine closed. Define $S' = \{e \in S : (p, e) = 1\}$ to be the set of elements in S which are co-prime to p, and assume that $|S'| \le n^\epsilon$ where $\epsilon < \log p / \log 2p$ and n is large enough. Then there exists $\ell_{\min} = O(|S|)$ and $\ell_{\max} = p^{\Omega(n)}$, such that for any $\ell \in [\ell_{\min}, \ell_{\max}]$ the following holds. The number of codewords in $d\mathcal{T}(S)$ of weight exactly ℓ is given by*

$$\frac{C(p,\ell)}{\ell!} p^{n(\ell - |S'|)}(1 + o(1))$$

where and where $C(p, \ell)$ is given by

$$C(p,\ell) = \left| \left\{ (v_1, \ldots, v_\ell) \in (\mathbb{F}_p \setminus \{0\})^\ell : v_1 + \ldots + v_\ell = 0 \right\} \right|.$$

Similar results were previously obtained over binary fields $\mathbb{F}_2$ using properties of Krawtchouk polynomials [11, 14]. Our technique is different, and relies on methods from additive combinatorics and Fourier analysis. In particular it allows us to extend the result to arbitrary fields and allows to obtain bounds for a wider range of values of ℓ. The proof of this lemma relies on the new extension of the Weil bound we establish, as well as the new estimation of the weight distribution of codes with large dual distance.

Given the lemma, the proof of Theorem 1.7 can be easily concluded. Recall that we showed that in order to prove local testability of an affine invariant code $\mathcal{T}(S)$, we need to show that there is a short codeword whose affine closure linearly spans $d\mathcal{T}(S)$. We showed that any $f \in d\mathcal{T}(S)$ for which this does not occur, is in fact contained in some $d\mathcal{T}(S \cup \{e\})$

for some $e \notin S$ of small weight. Thus, to conclude the proof we need to show that there exist small weight codewords in

$$d\mathcal{T}(S) \setminus \bigcup_{e \notin S: e \text{ has small weight}} d\mathcal{T}(S \cup \{e\}).$$

To this end we apply the tight bounds we obtain for the number of codewords of weight ℓ in dual-trace codes. We first show that if $\mathcal{C}$ is affine invariant of size $|\mathcal{C}| \le p^{n^{1+\epsilon}}$ then in fact $\mathcal{C} = d\mathcal{T}(S)$ where S is affine invariant, and $|S'| \le n^{\epsilon}$, so our estimates for the number of codewords apply for $d\mathcal{T}(S)$. Fix a suitable weight ℓ. The number of codewords of weight ℓ in $d\mathcal{T}(S)$ is given by

$$W_\ell = \frac{C(p,\ell)}{\ell!} p^{n(\ell-|S'|)}(1+o(1)),$$

where we recall that $S' = \{e \in S : (e,p) = 1\}$. On the other hand, as S is affine closed and $e \notin S$, we can bound the number of codewords of weight ℓ in any of the codes $d\mathcal{T}(S \cup \{e\})$ by

$$\le \frac{C(p,\ell)}{\ell!} p^{n(\ell-|S'|-1)}(1+o(1)) \approx p^{-n} W_\ell.$$

Thus to conclude we just need to verify that the number of distinct e of small weight is $\ll p^n$. This then can be verified by a routine calculation.

1.5 Paper organization

We prove the extension to Weil bound for character sums, Theorem 1.3, in Section 2. We prove the new estimation of the weight distribution of codes with large dual distance, i.e. Theorem 1.6 in Section 3. Both sections are written in a self-contained manner, so that readers that are interested in the details of only one of these results can read only the relevant section. Finally, the result about the testability of affine-invariant codes with super polynomial size (i.e. Theorem 1.7) is proved in Section 4. We note that throughout the paper we do not attempt to optimize constants.

2 Extension of the Weil bound

We prove in this section a new extension of the Weil bound for character sums. We recall several definitions and theorems from the introduction, for the sake of self containment. Let $\mathbb{F} = \mathbb{F}_{p^n}$ be a finite field. An additive character $\chi : \mathbb{F} \to \mathbb{C}$ is a mapping such that $\chi(x+y) = \chi(x)\chi(y)$ and χ is not identically zero. The following is a classical result by Weil. It shows that if $f(x)$ is a low-degree polynomial, then if $\chi(f(x))$ is not a constant function, then its distribution is very close to uniform.

Theorem 2.1 (Weil bound [23]). *Let $f(x)$ be a univariate polynomial over $\mathbb{F}_{p^n}$ of degree $\le p^{(1/2-\delta)n}$. Let $\chi : \mathbb{F}_{p^n} \to \mathbb{C}$ be any additive character. Then one of the following must hold:*

1. *$\chi(f(x))$ is constant for all $x \in \mathbb{F}_{p^n}$; or*

2. $\left|\mathbb{E}_{x\in\mathbb{F}_{p^n}}[\chi(f(x))]\right| \le p^{-\delta n}$.

We prove an extension of the Weil bound by allowing a few additional monomials of high degree but of low *weight degree*. The weight degree of a monomial x^t is defined as follows. Let $t = \sum_{i=0}^{n-1} a_i p^i$ be the representation of t in base p, where $0 \le a_i \le p-1$. The weight degree of x^t is defined to be $wt(x^t) = \sum a_i$. The weight degree of a polynomial $f(x)$ is the maximal weight of a monomial in f.

Note 2.2. We note that the weight degree of a polynomial can be equivalently defined also as a *derivative degree*, defined as follows. The directional derivative of $f(x)$ in direction $y \in \mathbb{F}_{p^n}$ is defined as $f_y(x) = f(x+y) - f(x)$. Define iterative derivatives in directions $y_1, \ldots, y_k$ as $f_{y_1,\ldots,y_k} = (f_{y_1,\ldots,y_{k-1}})_{y_k}$. The *derivative degree* of f is the minimal d such that for any $d+1$ derivatives $y_1, \ldots, y_{d+1} \in \mathbb{F}_{p^n}$, $f_{y_1,\ldots,y_{d+1}}(x) \equiv 0$. It can be verified that the derivative degree of a polynomial is exactly its weight degree. We do not prove this here, and will not require this fact in the proof.

We prove an extension of the Weil bound in case f is the sum of a low degree polynomial and a small number of monomials of bounded weight (but of arbitrary degree).

Theorem (Theorem1.3 - Extension of the Weil bound)**.** *Let $f(x) = g(x) + h(x)$ be a univariate polynomial over $\mathbb{F}_{p^n}$, where $g(x)$ is a polynomial of degree $\le p^{(1/2-\delta)n}$ and $h(x)$ is the sum of at most $k \ge 1$ monomials, each of weight degree at most d. Let $\chi : \mathbb{F}_{p^n} \to \mathbb{C}$ be an additive character. Then one of the following must hold:*

1. *$\chi(f(x))$ is constant for all $x \in \mathbb{F}_{p^n}$; or*

2. *$\left|\mathbb{E}_{x\in\mathbb{F}_{p^n}}[\chi(f(x))]\right| \le p^{-\frac{\delta}{2kd^2 2^d}n}$.*

2.1 Technical claims

In this subsection we provide some technical claims that will be needed for the proof of Theorem 1.3.

2.1.1 The trace operator

The trace operator $Tr : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is defined as $\mathrm{Tr}(x) = \sum_{i=0}^{n-1} x^{p^i}$. We give in this subsection some simple properties of the Trace operator.

Claim 2.3 (Characterization of additive characters). Let $\chi : \mathbb{F}_{p^n} \to \mathbb{C}$ be an additive character. Then there exists $a \in \mathbb{F}_{p^n}$ such that $\chi(x) \equiv \omega^{\mathrm{Tr}(ax)}$ where $\omega = e^{2\pi i/p}$.

Proof. We first prove that $\chi(x) = \omega^{\ell(x)}$ where $\ell : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is a linear map. Note that we must have $\chi(0) = 1$ since $\chi(0) = \chi(0+0) = \chi(0)^2$, and we cannot have $\chi(0) = 0$ as this will imply that $\chi \equiv 0$. Thus, we get that the image of χ is a p-th root of unity since $\chi(x)^p = \chi(px) = \chi(0) = 1$. Thus we can write $\chi(x) = \omega^{\ell(x)}$ for some mapping $\ell : \mathbb{F}_{p^n} \to \mathbb{F}_p$. The mapping ℓ is linear since

$$\omega^{\ell(x+y)} = \chi(x+y) = \chi(x)\chi(y) = \omega^{\ell(x)+\ell(y)}.$$

Now we argue that any linear mapping $\ell : \mathbb{F}_{p^n} \to \mathbb{F}_p$ can be represented as $\ell(x) \equiv \mathrm{Tr}(ax)$ for some $a \in \mathbb{F}_{p^n}$. This is proved by a counting argument. Each linear map $\ell : \mathbb{F}_{p^n} \to \mathbb{F}_p$ can be uniquely identified by its image on a basis for $\mathbb{F}_{p^n}$ as a linear space over $\mathbb{F}_p$. Thus, the number of such linear mappings is at most p^n. On the other hand, for each $a \in \mathbb{F}_{p^n}$ the mapping $x \to \mathrm{Tr}(ax)$ is linear (since Trace is a linear mapping), and the total number of theses mappings is the number of distinct $a \in \mathbb{F}_{p^n}$, that is p^n. To conclude we just need to show that for any distinct $a \neq b \in \mathbb{F}_{p^n}$ the mappings $\mathrm{Tr}(ax)$ and $\mathrm{Tr}(bx)$ are distinct. Equivalently, since Trace is a linear mapping, we need to show that $\mathrm{Tr}((a-b)x) \not\equiv 0$. This is clear however because the Trace mapping is not identically zero and $a-b \neq 0$ is invertible. $\square$

Claim 2.4 (Trace of a p-power is unbiased). For every $c \neq 0$ and $0 \leq L \leq n-1$ we have

$$\mathbb{E}_{x \in \mathbb{F}_{p^n}}[\omega^{\mathrm{Tr}(cx^{p^L})}] = 0.$$

Proof. We have $\mathrm{Tr}(cx^{p^L}) = \mathrm{Tr}(c^{p^{n-L}}x)$, so it suffices to prove the claim for $L = 0$. Let $\ell : \mathbb{F}_{p^n} \to \mathbb{F}_p$ defined as $\ell(x) = \mathrm{Tr}(cx)$. The mapping ℓ is linear, and as it is not identically zero, its output is uniform over $\mathbb{F}_p$. Thus we have that $\mathbb{E}_{x \in \mathbb{F}_{p^n}}[\omega^{\ell(x)}] = 0$. $\square$

2.1.2 Reduced forms

We define in this subsection reduced forms of polynomials. We show that for studying character sums it is the sufficient to restrict to reduced polynomials. We start by considering univariate polynomials, and then generalize the definitions and claims to multivariate polynomials.

Definition 2.5 (Reduced form: univariate polynomials)**.** Let $m(x) = ax^t$ be a monomial. We say m is *reduced* if $p \nmid t$. If $t = p^k r$ for $p \nmid r$ we define the *reduced form* of $m(x)$ to be $m(x)^{p^{n-k}} \equiv a^{p^{n-k}} x^r$. A constant term $c \in \mathbb{F}_{p^n}$ is reduced if $c \in \mathbb{F}_p$, otherwise its reduced form is $\mathrm{Tr}(c) \in \mathbb{F}_p$. We say a polynomial is reduced if all its monomials are reduced, and the reduced form of a polynomial is the sum of the reduced forms of its monomials.

Claim 2.6 (Equivalence of reduced form: univariate polynomials). Let $f(x)$ be a univariate polynomial over $\mathbb{F}$. Let $f'(x)$ be its reduced form. Then

1. $\mathrm{Tr}(f(x)) \equiv \mathrm{Tr}(f'(x))$.
2. $deg(f') \leq deg(f)$.
3. $wt(f') \leq wt(f)$.

Proof. For a monomial $m(x) = ax^t$ with $t = p^k r$, $p \nmid r$, let $m'(x) = a^{p^{n-k}} x^r$ be its reduced form. Note that $m'(x) = m(x)^{p^{n-k}}$. Since $\mathrm{Tr}(x) = \mathrm{Tr}(x^p)$ we have that $\mathrm{Tr}(m(x)) = \mathrm{Tr}(m'(x))$ for all $x \in \mathbb{F}$. Note that $wt(m') = wt(m)$ and $\deg(m') = r \leq t = \deg(m)$. For a general polynomial $f(x) = \sum m_i(x)$ we have that $f'(x) = \sum m'_i(x)$. Hence we get that $\mathrm{Tr}(f) \equiv \mathrm{Tr}(f')$, and since cancelations among the m'_i can only reduce the degree and weight degree of f', we get that $deg(f') \leq deg(f)$ and $wt(f') \leq wt(f)$. $\square$

Claim 2.7 (Trace of reduced non-constant polynomial is non-constant: univariate polynomials). Let $f(x)$ be a non-constant reduced univariate polynomial. Then $\mathrm{Tr}(f(x))$ is not constant.

Proof. Assume for contradiction that $\mathrm{Tr}(f(x)) \equiv c$ for some $c \in \mathbb{F}_p$. Let $f(x) = a_0 + \sum_{i \in I} a_i x^i$ where $a_0 \in \mathbb{F}_p$, $a_i \in \mathbb{F}_{p^n}$ for $i \in I$ and $I \subseteq \{0, \ldots, p^n - 1\}$ is nonempty such that $p \nmid i$ for all $i \in I$. Define $g(x) = \mathrm{Tr}(f(x)) - c$. We have that

$$g(x) = -c + \mathrm{Tr}(f(x)) = (a_0 - c) + \sum_{i \in I} \sum_{j=0}^{n-1} a_i^{p^j} x^{ip^j} = (a_0 - c) + \sum_{i \in I} \sum_{j=0}^{n-1} a_i^{p^j} x^{ip^j \pmod{p^n}}.$$

Notice that all the monomials in this representation are distinct, since all $i \in I$ are not divisible by p. Thus this is a non-zero polynomial of degree at most $p^n - 1$, and so it cannot evaluate to zero on all elements of $\mathbb{F}_{p^n}$. $\square$

We now generalize some of the definitions and claims to multivariate polynomials. When we refer to the degree of a multivariate polynomial we always mean is its total degree. The weight degree of a monomial $x_1^{e_1} \ldots x_s^{e_s}$ is the sum of the weight degrees of the variables, that is $wt(x_1^{e_1} \ldots x_s^{e_s}) = wt(x_1^{e_1}) + \ldots + wt(x_s^{e_s})$. The weight degree of a multivariate polynomial is the maximal weight degree of its monomials.

Note 2.8. As in the univariate case, the weight degree of a multivariate degree is equivalent to its derivative degree, which is defined in an analogous way to the univariate case.

Definition 2.9 (Reduced form: multivariate polynomials)**.** Let $m(x_1, \ldots, x_s) = a x_1^{e_1} \ldots x_s^{e_s}$ be a monomial. We say m is *reduced* if $p \nmid gcd(e_1, \ldots, e_s)$ (that is, at least one e_i is co-prime to p). If $e_i = p^k r_i$ where $p \nmid gcd(r_1, \ldots, r_s)$ we define the *reduced form* of $m(x_1, \ldots, x_s)$ to be $a^{p^{n-k}} x_1^{r_1} \ldots x_s^{r_s}$. We say a polynomial is reduced if all its monomials are reduced, and the reduced form of a polynomial is the sum of the reduced forms of its monomial.

Claim 2.10 (Equivalence of reduced form: multivariate polynomials). Let $f(x_1, \ldots, x_s)$ be a multivariate polynomial over $\mathbb{F}$. Let $f'(x_1, \ldots, x_s)$ be its reduced form. Then

1. $\mathrm{Tr}(f(x_1, \ldots, x_s)) \equiv \mathrm{Tr}(f'(x_1, \ldots, x_s))$.
2. $deg(f') \leq deg(f)$.
3. $wt(f') \leq wt(f)$.

Proof. The proof is identical to the proof of Claim 2.10 for the univariate case. $\square$

Claim 2.11 (Trace of reduced non-constant polynomial is non-constant: multivariate polynomials). Let $f(x_1, \ldots, x_s)$ be a non-constant reduced multivariate polynomial. Then $\mathrm{Tr}(f(x_1, \ldots, x_s))$ is not constant.

Proof. The proof is very similar to the proof of Claim 2.7 for the univariate case. If f is not a constant polynomial, that is if I is not empty, then for any $c \in \mathbb{F}_p$ the polynomial $\mathrm{Tr}(f(x_1, \ldots, x_s)) - c$ is a non-zero polynomial of individual degree at most $p^n - 1$ in each variable, and such a polynomial cannot evaluate to zero on all points in $(\mathbb{F}_{p^n})^s$. $\square$

2.1.3 Properties of derivatives

Let $f(x)$ be a univariate polynomial. For every $s \geq 1$ define the s-iterated derivative polynomial of f, $\Delta f(x; y_1, \dots, y_s)$, to be the multivariate polynomial in variables $x, y_1, \dots, y_s \in \mathbb{F}$ defined as

$$\Delta f(x; y_1, \dots, y_s) = f_{y_1,\dots,y_s}(x) = \sum_{I \subseteq [s]} (-1)^{|I|+s} f(x + \sum_{i \in I} y_i).$$

Derivatives play a crucial role in the proof of Theorem 1.3. We study in this subsection some of their properties, and prove some structural results on polynomials of the form $\Delta f(x; y_1, \dots, y_s)$.

Claim 2.12 (Derivation maintains degree). Let $m(x) = x^t$ be a monomial. Then for any k, all the monomials appearing in $\Delta m(x; y_1, \dots, y_k)$ have total degree t (or $\Delta m(x; y_1, \dots, y_k) \equiv 0$).

Proof. The polynomial $\Delta m(x; y_1, \dots, y_k)$ is a linear combination of $(x + \sum_{i \in I} y_i)^t$ for subsets $I \subseteq [k]$, each of which is homogeneous of degree t. $\square$

We show that the character sum of a polynomial can be bounded by a character sum of its iterated derivatives polynomial.

Claim 2.13 (Bias can be bounded by bias of derivatives). For any univariate polynomial $f(x)$ and $s \geq 1$

$$\left|\mathbb{E}_{x \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}]\right| \leq \left(\mathbb{E}_{x, y_1, \dots, y_s \in \mathbb{F}}[\omega^{\mathrm{Tr}(\Delta f(x; y_1, \dots, y_s))}]\right)^{1/2^s}$$

Proof. Consider first the case $s = 1$. We have

$$\begin{aligned}
&\left|\mathbb{E}_{x \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}]\right|^2 = \mathbb{E}_{x, x' \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x))} \overline{\omega^{\mathrm{Tr}(f(x'))}}] = \\
&\mathbb{E}_{x, x' \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x)) - \mathrm{Tr}(f(x'))}] = \mathbb{E}_{x, y \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x+y)) - \mathrm{Tr}(f(x))}] = \\
&\mathbb{E}_{x, y \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x+y) - f(x))}] = \mathbb{E}_{x, y \in \mathbb{F}}[\omega^{\mathrm{Tr}(\Delta f(x; y))}].
\end{aligned}$$

Hence

$$\left|\mathbb{E}_{x \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}]\right| \leq \left(\mathbb{E}_{x, y \in \mathbb{F}}[\omega^{\mathrm{Tr}(\Delta f(x; y))}]\right)^{1/2}.$$

For $s > 1$ we prove the result by induction. By the base case of $s = 1$ and the Cauchy-Schwartz inequality, we have that

$$\left|\mathbb{E}_{x \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}]\right|^{2^s} \leq \left(\mathbb{E}_{x, y_1 \in \mathbb{F}}[\omega^{\mathrm{Tr}(\Delta f(x; y_1))}]\right)^{2^{s-1}} \leq \mathbb{E}_{y_1 \in \mathbb{F}}\left[\left(\mathbb{E}_{x \in \mathbb{F}}[\omega^{\mathrm{Tr}(\Delta f(x; y_1))}]\right)^{2^{s-1}}\right].$$

For every value of $y_1 \in \mathbb{F}$ we have by the $s - 1$ case that

$$\left(\mathbb{E}_{x \in \mathbb{F}}[\omega^{\mathrm{Tr}(\Delta f(x; y_1))}]\right)^{2^{s-1}} \leq \mathbb{E}_{x, y_2, \dots, y_s \in \mathbb{F}}[\omega^{\mathrm{Tr}(\Delta f(x; y_1, \dots, y_s))}],$$

hence we get that

$$\left|\mathbb{E}_{x \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}]\right|^{2^s} \leq \mathbb{E}_{x, y_1, y_2, \dots, y_s \in \mathbb{F}}[\omega^{\mathrm{Tr}(\Delta f(x; y_1, \dots, y_s))}].$$

$\square$

We now define a special family of multivariate polynomials that will play an important role in the proof. Such polynomials arise when taking d-iterated derivatives from a polynomial of weight degree d.

Definition 2.14 (p-multilinear polynomials)**.** A multivariate polynomial $f(x_1, \dots, x_s)$ over $\mathbb{F}_{p^n}$ is p-multilinear if all its monomials are of the form $x_1^{p^{i_1}} \dots x_s^{p^{i_s}}$. In particular, if it is nonzero it has weight degree s.

Claim 2.15 (Structure of derivatives of monomials). Let $m(x) = x^t$ be a monomial of weight degree d. The d-iterated derivatives polynomial $\Delta m(x; y_1, \dots, y_d)$ of m is given as follows. Let $t = \sum_{j=1}^{k} a_{\ell_j} p^{\ell_j}$ where $1 \le a_{\ell_1}, \dots, a_{\ell_k} \le p-1$ and $\sum a_\ell = d$. Let $\mathcal{S}$ be the family of all partitions of $\{1, \dots, d\}$ into k subsets of sizes $a_{\ell_1}, \dots, a_{\ell_s}$, that is

$$\mathcal{S} = \{(S_1, \dots, S_k) : S_1 \cup \dots \cup S_k = \{1, \dots, d\}, |S_1| = a_{\ell_1}, \dots, |S_k| = a_{\ell_k}\}.$$

Then we have

$$\Delta m(x; y_1, \dots, y_d) = c \sum_{(S_1, \dots, S_k) \in \mathcal{S}} \prod_{j=1}^{k} \prod_{i \in S_j} (y_i)^{p^{\ell_j}}.$$

where $c = \prod_{j=1}^{k} a_{\ell_j}! \neq 0$ in $\mathbb{F}$. In particular, Δm is a non-zero p-multilinear polynomial in $y_1, \dots, y_d$ which does not depend on x.

Proof. We have

$$\Delta m(x; y_1, \dots, y_d) = \sum_{I \subseteq [d]} (-1)^{d+|I|} m(x + \sum_{i \in I} y_i) = \sum_{I \subseteq [d]} (-1)^{d+|I|} (x + \sum_{i \in I} y_i)^t.$$

Substituting $t = \sum a_{\ell_j} p^{\ell_j}$, and using the linearity of the Frobenius map $x \to x^{p^{\ell_j}}$ we get that

$$\Delta m(x; y_1, \dots, y_d) = \sum_{I \subseteq [d]} (-1)^{d+|I|} \prod_{j=1}^{k} (x^{p^{\ell_j}} + \sum_{i \in I} (y_i)^{p^{\ell_j}})^{a_{\ell_j}}.$$

Since $\sum a_{\ell_j} = d$ we get that Δm is a degree-d polynomial in the Frobenius images of $x, y_1, \dots, y_d$, i.e. in the monomials $\{x^{p^j}, (y_1)^{p^j}, \dots, (y_d)^{p^j} : 0 \le j \le n-1\}$.

We first claim that Δm does not depend on x, and is p-linear in $y_1, \dots, y_d$. That is, all the monomials of Δm consist of a product $(y_1)^{p^{j_1}} \dots (y_d)^{p^{j_d}}$, where $0 \le j_1, \dots, j_d \le n-1$. Otherwise, there exists some monomial in Δm which does not depend on at least one of $y_1, \dots, y_d$. This is because all monomials of Δm are products of d Frobenius images of $x, y_1, \dots, y_d$, and by the pigeonhole principle, if either a single variable y_i has two images appearing, or an image of x appears in the monomial, then there must exists a variable y_j not participating in the monomial.

Assume w.l.o.g that Δm contains monomials in which y_1 does not participate. Substituting $y_1 = 0$ in the definition of Δm, since $\Delta f(x; 0) = f(x) - f(x) \equiv 0$ for any polynomial f, we get that

$$\Delta m(x; 0, y_2, \dots, y_d) \equiv 0.$$

Hence, if there exist monomials in $\Delta m(x; y_1, \ldots, y_d)$ which do not depend on y_1, they are left intact by the substitution $y_1 = 0$, while all monomials depending on y_1 vanish. Thus since $\Delta m(x; 0, y_2, \ldots, y_d) \equiv 0$ all the monomials in $\Delta m(x; y_1, \ldots, y_d)$ must depend on y_1.

We have thus proved that $\Delta m(x; y_1, \ldots, y_d)$ does not depend on x, and is p-linear in $y_1, \ldots, y_d$. To conclude we need to compute the exact form of $\Delta m(x; y_1, \ldots, y_d)$. Any monomial depending on all $y_1, \ldots, y_d$ must come from the term corresponding for $I = \{1, \ldots, d\}$,

$$(x + \sum_{i \in [d]} y_i)^t = \prod_{j=1}^{k} (x^{p^{\ell_j}} + \sum_{i \in [d]} (y_i)^{p^{\ell_j}})^{a_{\ell_j}}.$$

The individual degree of each y_i is some p^{ℓ_j}, and there are exactly a_{ℓ_j} variables among $y_1, \ldots, y_d$ which has individual degree p^{ℓ_j}. Since the number of variables d is exactly the sum $\sum a_{\ell_j}$, all the monomials depending on all of $y_1, \ldots, y_d$ must be of the form $\prod_{j=1}^{k} \prod_{i \in S_j} (y_i)^{p^{\ell_j}}$, where $(S_1, \ldots, S_k) \in \mathcal{S}$ is a partition of $\{1, \ldots, d\}$ into sets of sizes $a_{\ell_1}, \ldots, a_{\ell_k}$. The coefficient of the monomial $\prod_{j=1}^{k} \prod_{i \in S_j} (y_i)^{p^{\ell_j}}$ is equal to the number of times this monomial appears in the last term, which is exactly $\prod_{j=1}^{k} a_{\ell_j}!$. □

Claim 2.16 (Derivative of reduced monomial is nonzero). Let $m(x)$ be a nonzero reduced monomial of weight degree d. Then $\Delta m(x; y_1, \ldots, y_d)$ is a nonzero reduced polynomial.

Proof. Let $m(x) = x^t$ for $t = \sum a_{\ell_j} p^{\ell_j}$. Since m is reduced we must have $a_0 \neq 0$. By Claim 2.15 we know that

$$\Delta m(x; y_1, \ldots, y_d) = c \sum_{(S_1, \ldots, S_k) \in \mathcal{S}} \prod_{j=1}^{k} \prod_{i \in S_j} (y_i)^{p^{\ell_j}}.$$

Thus any monomial of $\Delta m(x; y_1, \ldots, y_d)$ contains at least one variable of degree 1, thus it is reduced. □

Claim 2.17 (Derivative of distinct reduced monomials is distinct). Let $m'(x), m''(x)$ be two distinct monomials of weight degree d. Then $\Delta m'(x; y_1, \ldots, y_d)$ and $\Delta m''(x; y_1, \ldots, y_d)$ are nonzero polynomials which do not share any common monomial.

Proof. Let $m'(x) = x^{t'}$ and $m''(x) = x^{t''}$ for $t' \neq t''$. By Claim 2.15 we have that $\Delta m'(x; y_1, \ldots, y_d)$ is a nonzero polynomial such that all its monomials have total degree exactly t'. Similarly $\Delta m''(x; y_1, \ldots, y_d)$ is a nonzero polynomial such that all its monomials have total degree exactly t''. Since $t' \neq t''$ the polynomials $\Delta m'(x; y_1, \ldots, y_d)$ and $\Delta m''(x; y_1, \ldots, y_d)$ contain no common monomial. □

Claim 2.18 (High derivative vanishes). Let $f(x)$ be a polynomial of weight degree at most $d-1$. Then $\Delta m(x; y_1, \ldots, y_d) \equiv 0$.

Proof. It is enough to prove the claim for monomials. Let $m(x) = x^t$ be some monomial, and let $d' = wt(m) \leq d-1$ be its weight degree. By Claim 2.15 we have that $\Delta m(x; y_1, \ldots, y_{d'})$ does not depend on x, thus

$$\Delta m(x; y_1, \ldots, y_{d'}, y_{d'+1}) = \Delta m(x + y_{d'+1}; y_1, \ldots, y_{d'}) - \Delta m(x; y_1, \ldots, y_{d'}) \equiv 0.$$

□

Lemma 2.19 (Highest non-vanishing derivative)**.** *Let $f(x)$ be a nonzero reduced polynomial of weight degree d. Then $\Delta f(x; y_1, \dots, y_d)$ is a nonzero reduced polynomial which does not depend on x and is p-linear in $y_1, \dots, y_d$.*

Proof. Let $f(x) = \sum c_t x^t$. Let $m(x) = c_t x^t$ be some monomial of f. If $wt(m) \leq d - 1$ then by Claim 2.18 we have $\Delta m(x; y_1, \dots, y_d) \equiv 0$. Thus it is enough to consider just the monomials of weight degree exactly d. By Claim 2.16 the derivative polynomial of each reduced monomial of weight degree d is a reduced polynomial, and these polynomials for two distinct monomials contain no shared monomials, and so cannot cancel each other. Thus the derivative polynomial $\Delta f(x; y_1, \dots, y_d)$ is a nonzero reduced polynomial. By Claim 2.15 is does not depend on x, and it is p-linear in $y_1, \dots, y_d$. $\square$

Lemma 2.20 (General non-vanishing derivatives)**.** *Let $f(x)$ be a nonzero reduced polynomial of weight degree d. For any $k \leq d$ the polynomial $\Delta f(x; y_1, \dots, y_k)$ is a nonzero reduced polynomial in $x, y_1, \dots, y_k$.*

Proof. Let $f(x) = \sum c_t x^t$. Let $m(x) = c_t x^t$ be some monomial of f. Observe that all monomials in the polynomial $\Delta m(x; y_1, \dots, y_k)$ have the same total degree t. Thus, if $m(x)$ is reduced then so is $\Delta m(x; y_1, \dots, y_k)$, since if $x^{e_0} y_1^{e_1} \dots y_k^{e_k}$ is a monomial of $\Delta m(x; y_1, \dots, y_k)$ which is not reduced, then $p \mid gcd(e_0, \dots, e_k)$. However $t = e_0 + \dots + e_k$ and since $m(x)$ is reduced we have that $p \nmid t$. Contradiction, hence $\Delta m(x; y_1, \dots, y_k)$ must be reduced. Hence, we get that if $f(x)$ is a reduced polynomial, then $\Delta f(x; y_1, \dots, y_k)$ is also reduced. To conclude we need to prove that $\Delta f(x; y_1, \dots, y_k)$ is nonzero. Assume by contradiction it is zero; then so is $\Delta f(x; y_1, \dots, y_d) = \sum_{I \subseteq \{k+1, \dots, d\}} (-1)^{|I|+d-k} \Delta f(x + \sum_{i \in I} y_i; y_1, \dots, y_k)$. However by Lemma 2.19 we know that if f is a nonzero reduced polynomial, then $\Delta f(x; y_1, \dots, y_d)$ is nonzero. Hence also $\Delta f(x; y_1, \dots, y_k)$ must be nonzero. $\square$

2.1.4 Additional claims

We give in this subsection some more claims we will require. The first is the Schwarz-Zippel lemma.

Claim 2.21 (Schwarz-Zippel). Let $f(x_1, \dots, x_s)$ be a polynomial over $\mathbb{F}$ of total degree e. Then

$$\Pr_{x_1, \dots, x_s \in \mathbb{F}}[f(x_1, \dots, x_s) = 0] \leq \frac{e}{|\mathbb{F}|}.$$

The second result we will need is a theorem of Deligne [7] which is a multivariate analog of Weil's bound.

Theorem 2.22 (Deligne theorem [7])**.** *Let $f(x_1, \dots, x_s)$ be a multivariate polynomial over $\mathbb{F}$ of degree $|\mathbb{F}|^{1/2-\delta}$. Let $\chi : \mathbb{F} \to \mathbb{C}$ be an additive character. Then either $\chi(f(x_1, \dots, x_s))$ is constant or*

$$|\mathbb{E}_{x_1, \dots, x_s \in \mathbb{F}}[\chi(f(x))]| \leq |\mathbb{F}|^{-\delta}.$$

2.2 The case of high weight g

In this subsection we prove Theorem 1.3 in the case that g has high weight degree, $\mathrm{wt}(g) \geq d+1$. This is captured by the following lemma, which we prove in this subsection. This is the easier case for Theorem 1.3.

Lemma 2.23 (The case of high weight g)**.** *Let $f(x) = g(x) + h(x)$ be a nonzero reduced univariate polynomial over $\mathbb{F}_{p^n}$, where $g(x)$ is a polynomial of degree $|\mathbb{F}|^{1/2-\delta}$ and weight degree at least $d+1$, and $h(x)$ has weight degree at most d. Then*

$$\left|\mathbb{E}_{x\in\mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}]\right| \leq |\mathbb{F}|^{-\frac{\delta}{2^{d+1}}}.$$

Proof. The polynomial f is nonzero reduced and of weight degree at least $d+1$. By Lemma 2.20 we know that $\Delta f(x; y_1, \ldots, y_{d+1})$ is nonzero and reduced. However, since $\mathrm{wt}(h) \leq d$ we have that $\Delta h(x; y_1, \ldots, y_{d+1}) \equiv 0$ by Claim 2.18, hence we get that $\Delta f(x; y_1, \ldots, y_{d+1}) = \Delta g(x; y_1, \ldots, y_{d+1})$. Also, since derivation cannot increase total degree, we have that $\deg(\Delta f(x; y_1, \ldots, y_{d+1})) \leq \deg(g) \leq |\mathbb{F}|^{1/2-\delta}$.

So, we have that $f'(x, y_1, \ldots, y_{d+1}) = \Delta f(x; y_1, \ldots, y_{d+1})$ is a nonzero reduced polynomial of degree at most $|\mathbb{F}|^{1/2-\delta}$. By Claim 2.11 we have that $\mathrm{Tr}(f')$ is a non-constant function. Thus by Deligne's Theorem (Theorem 2.22) we get that is must be highly unbiased, that is

$$\left|\mathbb{E}_{x,y_1,\ldots,y_{d+1}\in\mathbb{F}}[\omega^{\mathrm{Tr}(f'(x,y_1,\ldots,y_{d+1}))}]\right| \leq |\mathbb{F}|^{-\delta}.$$

To conclude we apply Claim 2.13 to get that

$$\left|\mathbb{E}_{x\in\mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}]\right| \leq \left|\mathbb{E}_{x,y_1,\ldots,y_{d+1}\in\mathbb{F}}[\omega^{\mathrm{Tr}(f'(x,y_1,\ldots,y_{d+1}))}]\right|^{\frac{1}{2^{d+1}}} \leq |\mathbb{F}|^{-\frac{\delta}{2^{d+1}}}.$$

□

2.3 The case of low weight g

In this subsection we prove Theorem 1.3 in the case that g has low weight degree, $\mathrm{wt}(g) \leq d$. This is captured by the following lemma, which we prove in this subsection. This is the harder case for Theorem 1.3.

Lemma 2.24 (The case of low weight g)**.** *Let $f(x) = g(x) + h(x)$ be a nonzero reduced univariate polynomial over $\mathbb{F}_{p^n}$, where $g(x)$ is a polynomial of degree $|\mathbb{F}|^{1/2-\delta}$ and weight degree at most d, and $h(x)$ has weight degree d and is the sum of k monomials. Then*

$$\mathbb{E}_{x\in\mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}] \leq |\mathbb{F}|^{-\frac{\delta}{d^2 2^d k}+O(1/n)}.$$

To prove Lemma 2.24 we require some claims.

Claim 2.25 (Structure of derivative of g). Let $g(x)$ be a polynomial of degree at most $|\mathbb{F}|^{1/2-\delta}$ and weight degree at most d. For $L = \lceil n(1/2-\delta)\rceil$ there exists a p-multilinear polynomial $u(y_2, \ldots, y_d)$ such that

$$\mathrm{Tr}(\Delta g(x; y_1, \ldots, y_d)) \equiv \mathrm{Tr}(y_1^{p^L} \cdot u(y_2, \ldots, y_d)).$$

and such that $\deg(u) \leq p^{2L} \leq |\mathbb{F}|^{1-2\delta+2/n}$.

Proof. By linearity, it suffices to show that for every monomial $m(x)$ appearing in g, there exists a p-multilinear polynomial $u_m(y_2,\ldots,y_d)$ such that $\mathrm{Tr}(\Delta m(x;y_1,\ldots,y_d)) \equiv \mathrm{Tr}(y_1^{p^L} \cdot u_m(y_2,\ldots,y_d))$ and $\deg(u_m) \le p^{2L}$.

Let $m(x) = cx^t$ be such a monomial. If $wt(m) < d$ we have by Claim 2.18 that $\Delta m(x;y_1,\ldots,y_d) \equiv 0$. Otherwise assume that $\mathrm{wt}(m) = d$. By Claim 2.15 we know that $\Delta m(x;y_1,\ldots,y_d)$ does not depend on x and is p-multilinear in $y_1,\ldots,y_d$. Moreover, if $t = \sum_{j=1}^{k} a_{\ell_j} p^{\ell_j}$ where $1 \le a_{\ell_j} \le p-1$ we know that

$$\Delta m(x;y_1,\ldots,y_d) = \sum_{j=1}^{k} y_1^{p^{\ell_j}} w_j(y_2,\ldots,y_d)$$

where $w_j(y_2,\ldots,y_d)$ is a homogeneous p-multilinear polynomial of total degree $t-p^{\ell_j}$. Since $t \le |\mathbb{F}|^{1/2-\delta}$ we have that $\ell_1,\ldots,\ell_k \le n(1/2-\delta) \le L$. Thus, taking $u_m(y_2,\ldots,y_d)$ to be

$$u_m(y_2,\ldots,y_d) = \sum_{j=1}^{k} w_j(y_2,\ldots,y_d)^{p^{L-\ell_j}}$$

we get that

$$\mathrm{Tr}(y_1^{p^L} \cdot u_m(y_2,\ldots,y_d)) \equiv \sum_{j=1}^{k} \mathrm{Tr}(y_1^{p^L} w_j(y_2,\ldots,y_d)^{p^{L-\ell_j}}) \equiv$$
$$\sum_{j=1}^{k} \mathrm{Tr}(y_1^{p^{\ell_j}} w_j(y_2,\ldots,y_d)) = \mathrm{Tr}(\Delta m(x;y_1,\ldots,y_d)).$$

To conclude we need to bound $\deg(u_m)$. Since $\deg(w_j) \le \deg(m) \le p^{n(1/2-\delta)}$ and $L-\ell_j \le L$ we get that $\deg(u_m) \le \deg(m)\cdot p^L \le p^{2L}$. □

Claim 2.26 (Structure of derivative of h)*.* Let $h(x)$ be a polynomial of weight degree d which is the sum of k monomials. For every $0 \le L \le n-1$ there exists a p-multilinear polynomial $v(y_2,\ldots,y_d)$ such that

$$\mathrm{Tr}(\Delta h(x;y_1,\ldots,y_d)) \equiv \mathrm{Tr}(y_1^{p^L} \cdot v(y_2,\ldots,y_d)).$$

and the number of distinct total degrees of monomials appearing in v is at most kd.

Proof. By linearity, it suffices to show that for every monomial $m(x)$ appearing in h, there exists a p-multilinear polynomial $v_m(y_2,\ldots,y_d)$ such that $\mathrm{Tr}(\Delta m(x;y_1,\ldots,y_d)) \equiv \mathrm{Tr}(y_1^{p^L} \cdot v_m(y_2,\ldots,y_d))$ and the monomials appearing in v_m have at most d distinct total degrees.

Let $m(x) = cx^t$ be such a monomial. If $wt(m) < d$ we have by Claim 2.18 that $\Delta m(x;y_1,\ldots,y_d) \equiv 0$. Otherwise assume that $\mathrm{wt}(m) = d$. By Claim 2.15 we know that $\Delta m(x;y_1,\ldots,y_d)$ does not depend on x and is p-multilinear in $y_1,\ldots,y_d$. Moreover, if $t = \sum_{j=1}^{k} a_{\ell_j} p^{\ell_j}$ where $1 \le a_{\ell_j} \le p-1$ we know that

$$\Delta m(x;y_1,\ldots,y_d) = \sum_{j=1}^{k} y_1^{p^{\ell_j}} w_j(y_2,\ldots,y_d)$$

where $w_j(y_2,\ldots,y_d)$ is a homogeneous p-multilinear polynomial of total degree $t-p^{\ell_j}$. Let

$$v_m(y_2,\ldots,y_d)=\sum_{j=1}^{k} w_j(y_2,\ldots,y_d)^{p^{L-\ell_j+n}}$$

where we reduce individual powers of $y_2,\ldots,y_d$ modulo p^n (that is, we replace each y_i^e with $y_i^{e \bmod p^n}$, which are equivalent as functions over the field $\mathbb{F}_{p^n}$). Thus we get that

$$\mathrm{Tr}(y_1^{p^L}\cdot v_m(y_2,\ldots,y_d))\equiv\sum_{j=1}^{k}\mathrm{Tr}(y_1^{p^L}w_j(y_2,\ldots,y_d)^{p^{L-\ell_j+n}})\equiv$$

$$\sum_{j=1}^{k}\mathrm{Tr}(y_1^{p^{\ell_j}}w_j(y_2,\ldots,y_d))=\mathrm{Tr}(\Delta m(x;y_1,\ldots,y_d)).$$

To conclude we need to bound the number of distinct total degrees of monomials appearing in v_m. Each polynomial w_j is homogeneous, and so also $w_j^{p^{L-\ell_j+n}}$ is homogenous, hence contributing a unique total degree to monomials in v_m. As the number of distinct w_j is bounded by $k\le d$ we get the required bound. $\square$

Claim 2.27 (Covering argument for a single element). Let $0\le e\le p^n-1$ such that $\mathrm{wt}(e)=d$. For $0\le s\le n-1$ define $e_s=e\cdot p^s \bmod p^n$, such that also $0\le e_s\le p^n-1$. For $a\le n$ let

$$S=\{0\le s\le n-1: e_s\ge p^{n-a}\}.$$

Then $|S|\le a\cdot d$.

Proof. For every $0\le e\le p^n-1$ let $\vec{e}\in\{0,\ldots,p-1\}^n$ denote the vector corresponding to the base-p representation of e, that is $e=\sum_{i=0}^{n-1}\vec{e}(i)p^i$. Observe that $\vec{e}_s$ is just the cyclic shift of $\vec{e}$ by s coordinates, that is $\vec{e}_s(i)=\vec{e}(i-s \pmod n)$. Note that the weight of e is just the hamming weight of $\vec{e}$, and that $e_s\ge p^{n-a}$ if and only if the vector $\vec{e}_s$ contains some nonzero entry in the indices $n-a\le i\le n-1$. As $\vec{e}$ contains only d nonzero entries, there are at most $a\cdot d$ cyclic shift of $\vec{e}$ such that some of these entries moves to indices $i\in\{n-a,\ldots,n-1\}$. Thus we get that $|S|\le a\cdot d$. $\square$

Claim 2.28 (Covering argument for sum of monomials). Let $h(y_1,\ldots,y_b)$ be a polynomial over $\mathbb{F}_{p^n}$ of weight degree at most d, such that the number of distinct total degrees of its monomial is z. Let $h_s(y_1,\ldots,y_b)=h(y_1,\ldots,y_b)^{p^s}$ reducing each individual degree of $y_1,\ldots,y_b$ modulo p^n. Then for every a there exists $0\le s\le a$ such that

$$\deg(h_s)<p^{n-\lfloor\frac{a}{dz}\rfloor}.$$

Proof. Let $q=\lfloor\frac{a}{dz}\rfloor$. Let $\{e_1,\ldots,e_z\}$ be the set of total degrees occurring in monomials of h. The number of $0\le s\le n-1$ such that $(e_i\cdot p^s \bmod p^n)\ge p^{n-q}$ is bounded by $d\cdot q\le a/z$ by Claim 2.27. Thus, there are at most a values for s such that for some e_i we have $e_i\cdot p^s \bmod p^n\ge p^{n-q}$. Since there are $a+1$ possible values for $0\le s\le a$, by the pigeonhole principle there exists a value for which for all $i=1,\ldots,k$,

$$(e_i\cdot p^s \bmod p^n)<p^{n-q}$$

hence we get that $\deg(h_s)<p^{n-q}$. $\square$

Claim 2.29 (Structure of derivative of f). Let $f(x) = g(x) + h(x)$ be a nonzero reduced univariate polynomial over $\mathbb{F}_{p^n}$, where $g(x)$ is a polynomial of degree $|\mathbb{F}|^{1/2-\delta}$ and weight degree at most d, and $h(x)$ has weight degree d and is the sum of k monomials. Then there exists $M \in \{0, \dots, n-1\}$ and a p-multilinear polynomial $r(y_2, \dots, y_d)$ such that

$$\mathrm{Tr}(\Delta f(x; y_1, \dots, y_d)) \equiv \mathrm{Tr}(y_1^{p^M} \cdot r(y_2, \dots, y_d))$$

and $\deg(r) \le |\mathbb{F}|^{1-\frac{2\delta}{d^2k+1}+3/n}$.

Proof. Let $L = \lceil n(1/2-\delta) \rceil$. By Claim 2.25 there is a p-multilinear polynomial $u(y_2, \dots, y_d)$ such that $\mathrm{Tr}(\Delta g(x; y_2, \dots, y_d)) \equiv \mathrm{Tr}(y_1^{p^L} \cdot u(y_2, \dots, y_d))$ and $\deg(u) \le p^{2L}$. By Claim 2.26 there is a p-multilinear polynomial $v(y_2, \dots, y_d)$ such that $\mathrm{Tr}(\Delta h(x; y_2, \dots, y_d)) \equiv \mathrm{Tr}(y_1^{p^L} \cdot v(y_2, \dots, y_d))$ and the number of distinct total degrees of monomials in v is bounded by kd.

For s define $r_s(y_2, \dots, y_d) = p^s(u(y_2, \dots, y_d) + v(y_2, \dots, y_d))$ where individual degrees of $y_2, \dots, y_d$ are reduced modulo p^n, and set $a = \alpha n$ to be determined later. We will show there exists $0 \le s \le n - 2L - a$ such that $\deg(r_s) \le p^{n-a}$. This will establish the result as for every s,

$$\mathrm{Tr}(\Delta f(x; y_1, \dots, y_d)) \equiv \mathrm{Tr}(y_1^{p^{L+s}} r_s(y_2, \dots, y_d)).$$

First, notice that since $\deg(u) \le p^{2L}$ we have that for any $0 \le s \le n - 2L - a$ we have that

$$\deg(u^{p^s}) \le \deg(u) \cdot p^s \le p^{2L+s} \le p^{n-a}.$$

We now move to consider v. By Claim 2.28 we have that there exists $0 \le s \le n - 2L - a$ such that if we let $v_s(y_2, \dots, y_d) = v(y_2, \dots, y_d)^{p^s}$ reducing individual degrees modulo p^n, we have that

$$\deg(v_s) \le p^{n - \lfloor \frac{n-2L-a}{d^2k} \rfloor}.$$

Combining the two bounds, we get that

$$\deg(r_s) \le \max(p^{n-a}, p^{n - \lfloor \frac{n-2L-a}{d^2k} \rfloor}).$$

Setting $a = \lfloor \frac{n-2L-d^2k}{d^2k+1} \rfloor$ to optimize the bound we get that

$$\deg(r_s) \le p^{n-a} \le p^{n(1-\frac{2\delta}{d^2k+1})+3}.$$

$\square$

We are now ready to prove Lemma 2.24.

Proof of Lemma 2.24. We will bound the bias of $\mathrm{Tr}(f(x))$ by the bias of $\mathrm{Tr}(\Delta f(x; y_1, \dots, y_d))$. By Claim 2.13 we have that

$$\left|\mathbb{E}_{x \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}]\right| \le \left|\mathbb{E}_{x, y_1, \dots, y_d \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x; y_1, \dots, y_d))}]\right|^{1/2^d}.$$

To bound the bias of $\mathrm{Tr}(\Delta f(x; y_1, \dots, y_d))$, we apply Claim 2.29. We have

$$\mathrm{Tr}(\Delta f(x; y_1, \dots, y_d)) \equiv \mathrm{Tr}(y_1^{p^M} \cdot r(y_2, \dots, y_d))$$

where $\deg(r) \le |\mathbb{F}|^{1-\frac{2\delta}{d^2 k+1}+3/n}$. Moreover since f is nonzero and reduced, then by Lemma 2.19 $\Delta f(x; y_1, \ldots, y_d)$ is nonzero, hence $r(y_2, \ldots, y_d)$ must also be nonzero.

Whenever $y_2, \ldots, y_d$ are such that $r(y_2, \ldots, y_d) \neq 0$, we have that $\mathbb{E}_{y_1 \in \mathbb{F}}[\omega^{\mathrm{Tr}(y_1^{p^M} \cdot r(y_2,\ldots,y_d))}] = 0$ by Claim 2.4. The probability that $r(y_2, \ldots, y_d) = 0$ is bounded by Claim 2.21 by

$$\Pr_{y_2,\ldots,y_d \in \mathbb{F}}[r(y_2, \ldots, y_d) = 0] \le \frac{\deg(r)}{|\mathbb{F}|} \le |\mathbb{F}|^{-\frac{2\delta}{d^2 k+1}+3/n}.$$

Combining the results, we get that

$$\left|\mathbb{E}_{x \in \mathbb{F}}[\omega^{\mathrm{Tr}(f(x))}]\right| \le |\mathbb{F}|^{-\frac{2\delta}{(d^2 k+1)2^d}+\frac{3}{2^d n}} \le |\mathbb{F}|^{-\frac{\delta}{d^2 2^d k}+O(1/n)}.$$

□

3 Weight distribution of codes with large dual distance

We begin with some definitions and then state our theorems formally.

3.1 Basic coding definitions

Let $\mathbb{F}_p$ be a finite field. A linear code over $\mathbb{F}_p$ is a linear subspace $\mathcal{C} \subset \mathbb{F}_p^N$. The dimension of a code $\dim(\mathcal{C})$ is the dimension of the linear space. We will view codewords both as elements $f \in \mathbb{F}_p^N$ and as functions $f : [N] \to \mathbb{F}_p$. For a linear code $\mathcal{C}$, its dual $\mathcal{C}^\perp$ is the set functions which are orthogonal to all codewords of $\mathcal{C}$,

$$\mathcal{C}^\perp = \left\{ g \in \mathbb{F}_p^N : \sum_{x \in [N]} f(x)g(x) = 0 \quad \forall f \in \mathcal{C} \right\}.$$

Note that the dual of the dual is the original code, i.e. $(\mathcal{C}^\perp)^\perp = \mathcal{C}$. We next define the weight and support of a codeword. The *support* of a codeword $f \in \mathcal{C}$ is the set of $x \in [N]$ for which $f(x) \neq 0$,

$$\mathrm{support}(f) = \{x \in [N] : f(x) \neq 0\}.$$

The *weight* of a codeword is the size of its support,

$$\mathrm{wt}(f) = |\mathrm{support}(f)| = |\{x \in [N] : f(x) \neq 0\}|.$$

The *distance* of a linear code $\mathcal{C}$ is the minimal hamming distance between two distinct codewords. Equivalently, it is the minimal weight of a nonzero codeword,

$$\mathrm{dist}(\mathcal{C}) = \min_{f \in \mathcal{C} \setminus \{0^N\}} \mathrm{wt}(f).$$

We would be interested in a related notion, which is the distance between $\mathcal{C}$ and constant codewords. Let $\mathrm{Const} = \{a^N : a \in \mathbb{F}_p\}$ be the code of constant codewords. Note that as $\mathcal{C}$

is linear, we either have that Const $\subset \mathcal{C}$ or that Const $\cap\, \mathcal{C} = \{0^N\}$. We define dist$(\mathcal{C}, \text{Const})$ to be the minimal distance between a nonconstant codeword of $\mathcal{C}$ and constant functions.

$$\text{dist}(\mathcal{C}, \text{Const}) = \min_{f \in \mathcal{C} \setminus \text{Const}} \min_{a \in \mathbb{F}_p} \Pr_{x \in [N]}[f(x) \neq a].$$

Note that $0 \leq \text{dist}(\mathcal{C}, \text{Const}) \leq 1 - 1/p$. A related notion, which sometimes is more convenient, is that of *bias*. The bias of a codeword $f \in \mathcal{C}$ is defined as

$$\text{bias}(f) = \left|\mathbb{E}_{x \in [N]}[\omega^{f(x)}]\right| = \left|\frac{1}{N} \sum_{x \in [N]} \omega^{f(x)}\right|,$$

where $\omega = e^{2\pi i/p}$. Note that $0 \leq \text{bias}(f) \leq 1$, where $\text{bias}(f) = 1$ iff $f \in$ Const. We define the bias of a code as the maximal bias of a nonconstant codeword,

$$\text{bias}(\mathcal{C}) = \max_{f \in \mathcal{C} \setminus \text{Const}} \text{bias}(f).$$

Note that always $\text{bias}(\mathcal{C}) < 1$. We now establish a relation between distance in bias both in the case where the distance is small and where it is near maximal.

Claim 3.1. Let $\mathcal{C} \subset \mathbb{F}_p^N$ be a linear code.

(i) If $\text{dist}(\mathcal{C}, \text{Const}) \geq \delta$ then
$$\text{bias}(\mathcal{C}) \leq 1 - \Omega(\delta/p^2).$$

(ii) If $\text{dist}(\mathcal{C}, \text{Const}) \geq 1 - 1/p + \delta$ then
$$\text{bias}(\mathcal{C}) \leq 2p\delta.$$

Proof. Fix $f \in \mathcal{C}$. Let $q(a) = \Pr_{x \in [N]}[f(x) = a]$. Then

$$\text{bias}(f) = \left|\sum_{a \in \mathbb{F}_p} q(a)\omega^a\right|. \tag{1}$$

We first prove (i). Note that by our assumptions on the distance, $q(a) \leq 1 - \delta$ for all $a \in \mathbb{F}_p$. We can assume w.l.o.g that $\delta \leq 1/2$, as otherwise the bound will follow the bound for $\delta = 1/2$. One can verify that for $\delta \leq 1/2$ the RHS of (1) is maximized when $q(0) = 1 - \delta$ and $q(1) = \delta$; hence

$$\text{bias}(f) \leq |(1 - \delta) + \delta\omega| = 1 - \Omega(\delta/p^2).$$

We now prove (ii). Since the distance is at least $1 - 1/p + \delta$, we have $q(a) \leq 1/p + \delta$ for all $a \in \mathbb{F}_p$. Hence $\sum_{a \in \mathbb{F}_p} |q(a) - 1/p| = 2\sum_{a:q(a)>1/p}(q(a) - 1/p) \leq 2p\delta$. Using the fact that $\sum_{a \in \mathbb{F}_p} \omega^a = 0$ we get that

$$\text{bias}(f) = \left|\sum_{a \in \mathbb{F}_p}(q(a) - 1/p)\omega^a\right| \leq \sum_{a \in \mathbb{F}_p} |q(a) - 1/p| \leq 2p\delta.$$

□

Let $\mathcal{C}$ be a code. The next theorem provides a tight estimate on the number of codewords in $\mathcal{C}^\perp$ of weight ℓ for a range of values of ℓ which depends on the bias of $\mathcal{C}$ and the required error of approximation. For simplicity of notation, we denote by $t(1 \pm \epsilon)$ an unspecified quantity in the range $[t - t\epsilon, t + t\epsilon]$.

Theorem (Theorem 1.6 - Weight distribution of codes). *Let $\mathcal{C} \subset \mathbb{F}_p^N$ be a linear code with $\mathrm{bias}(\mathcal{C}) = \delta < 1$. Fix $\epsilon > 0$, and let $\ell_{\min} = \log_{1/\delta} |\mathcal{C}| + \log(1/\epsilon)$ and $\ell_{\max} = \sqrt{\epsilon N}$. Then for any $\ell \in [\ell_{\min}, \ell_{\max}]$, the number of codewords $g \in \mathcal{C}^\perp$ of weight exactly ℓ is given by*

(i) If $\mathcal{C} \cap \mathrm{Const} = \{0\}^n$:

$$\textit{Number of codewords in } \mathcal{C}^\perp \textit{ of weight } \ell = \frac{(p-1)^\ell}{\ell!} \frac{N^\ell}{|\mathcal{C}|}(1 \pm 2\epsilon).$$

(ii) If $\mathrm{Const} \subset \mathcal{C}$:

$$\textit{Number of codewords in } \mathcal{C}^\perp \textit{ of weight } \ell = \frac{C(p,\ell)}{\ell!} \frac{N^\ell}{|\mathcal{C}|}(1 \pm 2\epsilon).$$

where $C(p, \ell)$ is defined as

$$C(p,\ell) = \left| \left\{ (v_1, \ldots, v_\ell) \in (\mathbb{F}_p \setminus \{0\})^\ell : v_1 + \ldots + v_\ell = 0 \right\} \right|.$$

3.2 Proof of Theorem 1.6

We start by proving (i). For any $v = (v_1, \ldots, v_\ell) \in \{1, \ldots, p-1\}^\ell$ define the sets

$$A_\ell(v) = \{(x_1, \ldots, x_\ell) \in [N]^\ell : \sum_{i=1}^{\ell} v_i f(x_i) = 0 \quad \forall f \in \mathcal{C}\}$$

and

$$B_\ell(v) = \{(x_1, \ldots, x_\ell) \in A_\ell(v) : x_1, \ldots, x_\ell \text{ are all distinct}\}.$$

Let $g \in \mathcal{C}^\perp$ be such that g has weight exactly ℓ. Equivalently, there are distinct points $x_1, \ldots, x_\ell \in [N]$ such that $\sum f(x_i) g(x_i) = 0$ for all $f \in \mathcal{C}$. We can identify g uniquely by the list of points $(x_1, \ldots, x_\ell)$ and the evaluation of g on these points $v = (g(x_1), \ldots, g(x_\ell)) \in \{1, \ldots, p-1\}^\ell$. Since the order of $x_1, \ldots, x_\ell$ does not matter, and they are all distinct, there are $\ell!$ elements in $\uplus B_\ell(v)$ which correspond to g, (i.e. these elements correspond to all orderings of $x_1, \ldots, x_\ell$). Thus we obtain the following identity,

$$\text{Number of codewords in } \mathcal{C}^\perp \text{ of weight } \ell = \frac{1}{\ell!} \sum_{v \in \{1,\ldots,p-1\}^\ell} |B_\ell(v)|.$$

Hence, to conclude the proof we will show that $|B_\ell(v)| \approx N^\ell / |\mathcal{C}|$. In fact, we will first show that $|A_\ell(v)| \approx N^\ell / |\mathcal{C}|$ and then deduce the estimate for $|B_\ell(v)|$.

Fix some $v \in \{1, \ldots, p-1\}^\ell$. We will now show an estimate on $|A_\ell(v)|$, where the main tool we use is Fourier analysis. Take any tuple $(x_1, \ldots, x_\ell) \in [N]^\ell$, and consider

$$\mu(x_1, \ldots, x_\ell) = \mathbb{E}_{f \in \mathcal{C}} \left[\omega^{v_1 f(x_1) + \ldots + v_\ell f(x_\ell)}\right],$$

where $\omega = e^{\frac{2\pi i}{p}}$ is a p-root of unity. We claim that if $(x_1, \ldots, x_\ell) \in A_\ell(v)$ then $\mu(x_1, \ldots, x_\ell) = 1$, and if $(x_1, \ldots, x_\ell) \notin A_\ell(v)$ then $\mu(x_1, \ldots, x_\ell) = 0$. This holds since $\mathcal{C}$ is a linear subspace. Hence, either the inner product of v with $(f(x_1), \ldots, f(x_\ell))$ is always zero; or it is uniformly distributed over $\mathbb{F}_p$ when $f \in \mathcal{C}$ is uniformly chosen. Hence we have

$$\begin{aligned} N^{-\ell}|A_\ell(v)| &= \mathbb{E}_{x_1, \ldots, x_\ell \in [N]} [\mu(x_1, \ldots, x_\ell)] \\ &= \mathbb{E}_{x_1, \ldots, x_\ell \in [N]} \mathbb{E}_{f \in \mathcal{C}} \left[\omega^{v_1 f(x_1) + \ldots + v_\ell f(x_\ell)}\right] \\ &= \mathbb{E}_{f \in \mathcal{C}} \prod_{i=1}^{\ell} \mathbb{E}_{x_i \in [N]} \left[\omega^{v_i f(x_i)}\right]. \end{aligned}$$

We partition the expectation to the cases where $f = 0^N$ and $f \neq 0^N$. When $f = 0^N$ then for all $i = 1, \ldots, \ell$ we have that

$$\mathbb{E}_{x_i \in [N]} \left[\omega^{v_i f(x_i)}\right] = 1.$$

Consider now any $f \neq 0^N$ and any $i = 1, \ldots, \ell$. Since we assumed $\mathcal{C} \cap \mathrm{Const} = \{0\}^n$, f is not constant. Let $f_i : [N] \to \mathbb{F}_p$ be defined by $f_i(x) = v_i f(x)$. Note that since $\mathcal{C}$ is linear we have $f_i \in \mathcal{C}$; and since $v_i \in \mathbb{F}_p \setminus \{0\}$ then also f_i is not constant. Hence

$$\left|\mathbb{E}_{x_i \in [N]} \left[\omega^{v_i f(x_i)}\right]\right| \leq \mathrm{bias}(\mathcal{C}) \leq \delta.$$

Hence we deduce that

$$|A_\ell(v)| = \frac{N^\ell}{|\mathcal{C}|}(1 + \eta)$$

where $|\eta| \leq |\mathcal{C}|\delta^\ell$. In particular, if $\ell \geq \log_{1/\delta} |\mathcal{C}| + \log(1/\epsilon)$ we get that $\eta \leq \epsilon$.

To conclude, we need to derive an estimate on $|B_\ell(v)|$. Let $C_\ell(v) = A_\ell(v) \setminus B_\ell(v)$. We will show that $|C_\ell(v)| \ll |B_\ell(v)|$, and hence $|B_\ell(v)| \approx |A_\ell(v)|$. To derive this, note that if $(x_1, \ldots, x_\ell) \in C_\ell(v)$, then $x_1, \ldots, x_\ell$ are not all distinct, that is, $x_i = x_j$ for some distinct $i < j$. Define $v^{(i,j)} \in \{1, \ldots, p-1\}^{\ell-1}$ by "joining" x_i and x_j, i.e. $v^{(i,j)}_a = v_a$ for $1 \leq a < i$ and $i < a < j$, $v^{(i,j)}_i = v_i + v_j$, $v^{(i,j)}_a = v_{a+1}$ for $a > j$. Then we can identify uniquely $(x_1, \ldots, x_\ell) \in C_\ell(v)$ with $x^{(i,j)} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_\ell) \in A_{\ell-1}(v^{(i,j)})$. Hence we get

$$|C_\ell(v)| \leq \sum_{i<j} |A_{\ell-1}(v^{i,j})| \leq \binom{\ell}{2} |A_{\ell-1}(\cdot)| \leq \frac{\ell^2}{N} \frac{N^\ell}{|\mathcal{C}|}$$

Hence we get that as long as $\ell^2 \leq \epsilon N$ we have

$$|B_\ell(v)| = \frac{N^\ell}{|\mathcal{C}|}(1 \pm 2\epsilon).$$

This concludes the proof of (i).

The proof of (ii) is completely analogous. Assume Const $\subset \mathcal{C}$. Define $\mathcal{C}' = \{f \in \mathcal{C} : f(0) = 0\}$ so that $\mathcal{C} = \{f' + f'' : f' \in \mathcal{C}', f'' \in \text{Const}\}$, $\text{bias}(\mathcal{C}') = \text{bias}(\mathcal{C})$ and $\mathcal{C}' \cap \text{Const} = \{0\}^n$. We apply the same argument as in (i) for the code $\mathcal{C}'$. The only additional requirement is that $v_1 + \ldots + v_\ell = 0$. Thus one should not consider $A_\ell(v)$ for all $v \in (\mathbb{F}_p \setminus \{0\})^\ell$, but only those corresponding to $v \in C(p, \ell)$. Thus we have

$$\text{Number of codewords in } \mathcal{C}^\perp \text{ of weight } \ell = \frac{1}{\ell!} \sum_{v \in C(p,\ell)} |B_\ell(v)|.$$

and the proof follows by the estimates we proved on $|B_\ell(v)|$.

4 Testing of affine invariant codes with super-polynomial size

We begin with some definitions and then state our theorems formally.

4.1 Trace codes

We study codes $\mathcal{C} \subset \mathbb{F}_p^{\mathbb{F}_p^n}$ where we view codewords as functions $f : \mathbb{F}_p^n \to \mathbb{F}_p$.

Definition 4.1 (trace codes)**.** Let $S \subseteq \{0, \ldots, p^n - 1\}$. The S-trace code is a code whose codewords are evaluations of functions $f : \mathbb{F}_{p^n} \to \mathbb{F}_p$ given by

$$\mathcal{T}(S) = \left\{ \left(\sum_{e \in S} \text{Tr}(\alpha_e x^e) : \mathbb{F}_{p^n} \to \mathbb{F}_p \right) : \alpha_e \in \mathbb{F}_p \right\},$$

where the Trace function $\text{Tr} : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is given by $\text{Tr}(x) = \sum_{i=0}^{n-1} x^{p^i}$.

For example, dual-BCH codes of weight t correspond to the special case

$$\text{dBCH}(t) = \mathcal{T}(\{1, 2, \ldots, t\}).$$

Generalized Reed-Muller codes over $\mathbb{F}_p^n$ of total degree d are equivalent to

$$\text{RM}_p(n, d) = \mathcal{T}(\{e \in \{0, \ldots, p^n - 1\} : \text{wt}(e) \leq d\}).$$

The following fact gives some simple properties of the Trace operator. For a proof, see any standard Algebra textbook, e.g. [3].

Fact 4.2 (Facts on the trace operator)**.** *Let* $\text{Tr}(x) = \sum_{i=0}^{n-1} x^{p^i}$ *be the trace operator over* $\mathbb{F}_{p^n}$*. Then*

1. *For any* $x \in \mathbb{F}_{p^n}$*,* $\text{Tr}(x) \in \mathbb{F}_p$*. That is,* $\text{Tr} : \mathbb{F}_{p^n} \to \mathbb{F}_p$*.*

2. *The trace operator is linear. That is, for any $x, y \in \mathbb{F}_{p^n}$ and $a, b \in \mathbb{F}_p$ we have*

$$\mathrm{Tr}(ax + by) = a\mathrm{Tr}(x) + b\mathrm{Tr}(y).$$

3. *The trace operator is invariant under the Frobenius map. That is, for any $x \in \mathbb{F}_{p^n}$ and $0 \le i \le n-1$ we have*

$$\mathrm{Tr}(x^{p^i}) = \mathrm{Tr}(x).$$

4. *Let $x \in \mathbb{F}_{p^n}$, and assume that for any $\alpha \in \mathbb{F}_{p^n}$ we have $\mathrm{Tr}(\alpha x) = 0$. Then $x = 0$.*

We denote the dual codeword to $\mathcal{T}(S)$ by $d\mathcal{T}(S) = \mathcal{T}(S)^{\perp}$. The following claim characterizes dual-trace codes.

Claim 4.3 (Characterization of dual-trace codes)*.* Let $S \subseteq \{0, \dots, p^n - 1\}$. Then

$$d\mathcal{T}(S) = \left\{(g : \mathbb{F}_{p^n} \to \mathbb{F}_p) : \sum_{x \in \mathbb{F}_{p^n}} g(x)x^e = 0 \quad \forall e \in S\right\}.$$

Proof. Let $g : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a function such that $\sum g(x)x^e = 0$ for all $e \in S$. We first verify that $g \in d\mathcal{T}(S)$. To do so, we need to show that $\sum_x f(x)g(x) = 0$ for any $f \in \mathcal{T}(S)$. Let $f = \sum_{e \in S} \mathrm{Tr}(\alpha_e x^e) \in \mathcal{T}(S)$. Then we have

$$\begin{aligned}\sum_{x \in \mathbb{F}_{p^n}} f(x)g(x) &= \sum_{x \in \mathbb{F}_{p^n}} \sum_{e \in S} \mathrm{Tr}(\alpha_e x^e)g(x) \\ &= \sum_{e \in S} \mathrm{Tr}(\alpha_e \sum_{x \in \mathbb{F}_{p^n}} x^e g(x)) = 0,\end{aligned}$$

where we used the fact that Trace is a linear operator over $\mathbb{F}_{p^n}$, thus $\mathrm{Tr}(ax+by) = a\mathrm{Tr}(x) + b\mathrm{Tr}(y)$ for any $a, b \in \mathbb{F}_p$ and $x, y \in \mathbb{F}_{p^n}$. Thus, to prove the claim we need to establish that for any $g \in d\mathcal{T}(S)$ and any $e \in S$ we have $\sum g(x)x^e = 0$. Note that for any $\alpha_e \in \mathbb{F}_{p^n}$ we have $f(x) = \alpha_e x^e \in \mathcal{T}(S)$, thus we have

$$\sum_{x \in \mathbb{F}_{p^n}} \mathrm{Tr}(\alpha_e x^e g(x)) = 0.$$

Let $z = \sum_{x \in \mathbb{F}_{p^n}} g(x)x^e$. We obtained that for any $\alpha_e \in \mathbb{F}_{p^n}$ we have

$$\mathrm{Tr}(\alpha_e z) = 0.$$

This can only hold if $z = 0$, thus we conclude that we must have that $\sum_x g(x)x^e = 0$ for all $e \in S$. ∎

The next claim shows that if $S_1 \subseteq S_2$ then $\mathcal{T}(S_1) \subseteq \mathcal{T}(S_2)$ and $d\mathcal{T}(S_1) \supseteq d\mathcal{T}(S_2)$.

Claim 4.4 (Monotonicity of trace codes)*.* Let $S_1 \subseteq S_2 \subseteq \{0, \dots, p^n - 1\}$. Then we have the following inclusions

1. $\mathcal{T}(S_1) \subseteq \mathcal{T}(S_2)$.

2. $d\mathcal{T}(S_1) \supseteq d\mathcal{T}(S_2)$.

Proof. The claim follows immediately from the definition of trace codes and of dual codes. $\square$

We will consider in the following few claims only trace codes for $S \subseteq \{1, \ldots, p^n - 1\}$, i.e. we disallow $0 \in S$. We will later also deal with sets containing 0. We now define irreducible degrees and reduced forms. We will see that it is enough to study trace codes over reduced form sets.

Definition 4.5 (Irreducible degrees and reduced form)**.** We define R as the set of co-prime elements to p,

$$R = \{1 \leq e \leq p^n - 1 : (e, p) = 1\}.$$

For $1 \leq e \leq p^n - 1$ define its *reduced form* $e' \in R$ as follows. Let $e = p^k m$ where $(p, m) = 1$. Then the reduced form of e is $e' = m$. For a subset $S \subseteq \{1, \ldots, p^n - 1\}$ define its reduced form $S' \subseteq R$ as $S' = \{e' : e \in S\}$.

Claim 4.6 (Trace codes are defined over reduce form sets). Let $S \subseteq \{1, \ldots, p^n - 1\}$. Let $S' \subseteq R$ be the reduced form of S. Then $d\mathcal{T}(S) = d\mathcal{T}(S')$ and $\mathcal{T}(S) = \mathcal{T}(S')$.

Proof. By Claim 4.3 we have that $g \in d\mathcal{T}(S)$ iff $\sum g(x)x^e = 0$ for all $e \in S$. For any $0 \leq k \leq n - 1$ we have

$$\left(\sum g(x)x^e\right)^{p^k} = \sum g(x)x^{ep^k} = \sum g(x)x^{ep^k \pmod{p^n}},$$

where we used the facts that $x \to x^{p^k}$ is a linear map over $\mathbb{F}_{p^n}$, and that for any $x \in \mathbb{F}_{p^n}$ we have $x^{p^n} = x$. Hence we get that $\sum g(x)x^e = 0$ iff $\sum g(x)x^{e'} = 0$ for any e' such that $e' = ep^k \pmod{p^n}$. This shows that $d\mathcal{T}(S) = d\mathcal{T}(S')$, since for every element $e \in S$ there is some $e' = ep^k \pmod{p^n} \in S'$ and vice versa. Since $d\mathcal{T}(S) = d\mathcal{T}(S')$ we also get by the uniqueness of dual codes that $\mathcal{T}(S) = d\mathcal{T}(S)^\perp = d\mathcal{T}(S')^\perp = \mathcal{T}(S')$. $\square$

The next claim establishes the size of trace codes defined over reduced form sets $S \subseteq R$.

Claim 4.7 (Size of trace codes). Let $S \subseteq \{1, \ldots, p^n - 1\}$. Let $S' \subseteq R$ be the reduced form of S. Then $|\mathcal{T}(S)| = p^{n|S'|}$.

Proof. By Claim 4.6 we know that $\mathcal{T}(S) = \mathcal{T}(S')$. The codewords of $\mathcal{T}(S')$ are functions of the form

$$f(x) = \sum_{e \in S'} \mathrm{Tr}(\alpha_e x^e),$$

where $\alpha_e \in \mathbb{F}_{p^n}$. The number of combinations of $\{\alpha_e : e \in S'\}$ is $|\mathbb{F}_{p^n}|^{|S'|} = p^{n|S'|}$. Hence to conclude we need to show any two such settings are distinct. Since the code is linear, it is enough to show that if the coefficients α_e are not all zero, then the codeword is not the all zeros codeword, i.e. there is some $x \in \mathbb{F}_{p^n}$ such that

$$\sum_{e \in S'} \mathrm{Tr}(\alpha_e x^e) \neq 0.$$

Let $p(x) = \sum_{e \in S'} \mathrm{Tr}(\alpha_e x^e)$, and note that

$$p(x) = \sum_{e \in S'} \sum_{i=0}^{n-1} \alpha_e^{p^i} x^{ep^i}$$
$$= \sum_{e \in S'} \sum_{i=0}^{n-1} \alpha_e^{p^i} x^{ep^i \pmod{p^n}},$$

where we used the facts that $\mathrm{Tr}(x) = \sum_{i=0}^{n-1} x^{p^i}$ as well as the identity $x^t = x^{t \pmod{p^n}}$ which holds for any t. Since $S' \subseteq R$ is a set of

all the monomials x^{ep^i} for $e \in S'$ are disjoint. Hence $p(x)$ is not the all zeros polynomial. As $\deg(p) \leq p^n - 1$ there must exist some $x \in \mathbb{F}_{p^n}$ such that $p(x) \neq 0$, and the codeword defined by f is not the all zeros codeword. □

4.2 Characterization of affine invariant codes by trace codes

We start by recalling *affine invariant codes*, which are codes that are closed under an affine transformation of the input space coordinates.

Definition 4.8 (Affine closure, and affine invariant codes)**.** Let $f : \mathbb{F}_{p^n} \to \mathbb{F}_p$ be a function. The *affine closure* of f is the set of functions

$$\overline{\mathrm{affine}}(f) = \left\{ (f(ax+b) : \mathbb{F}_{p^n} \to \mathbb{F}_p) : a, b \in \mathbb{F}_{p^n} \right\}.$$

A code $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$ is called *affine invariant* if for any $f \in \mathcal{C}$, we have $\overline{\mathrm{affine}}(f) \subseteq \mathcal{C}$. A codeword $f \in \mathcal{C}$ *affinely generates* $\mathcal{C}$ if

$$\mathcal{C} = \mathrm{Span}(\overline{\mathrm{affine}}(f)).$$

We can characterize linear codes which are affine invariant as a special subfamily of trace codes. To this end we will require some definitions. We first define shift closure of a set, which is tightly related to the reduced form we previously defined.

Definition 4.9 (Shift closed)**.** Let $e \in \{0, \ldots, p^n - 1\}$. The *shift closure* of e is defined as the set

$$\overline{\mathrm{shift}}(e) = \{ep^\ell \pmod{p^n} : \ell = 1, \ldots, n\}.$$

The shift closure of a set $S \subseteq \{0, \ldots, p^n - 1\}$ is defined as the union of the shift closures of its elements,

$$\overline{\mathrm{shift}}(S) = \cup_{e \in S} \overline{\mathrm{shift}}(e).$$

A set $S \subseteq \{0, \ldots, p^n - 1\}$ is said to be *shift closed* if $S = \overline{\mathrm{shift}}(S)$.

The term *shift closed* comes from viewing elements $e \in S$ as vectors in $\mathbb{F}_p^n$, given by the representation of e in base p. In this case, $ep^\ell \pmod{p^n}$ corresponds to a cyclic shift of the vector by ℓ coordinates. The following claim shows that trace codes are invariant under shift closure.

Claim 4.10. Let $S \subseteq \{0, \ldots, p^n - 1\}$. Then

$$d\mathcal{T}(S) = d\mathcal{T}(\overline{\text{shift}}(S)), \quad \mathcal{T}(S) = \mathcal{T}(\overline{\text{shift}}(S)).$$

Proof. The proof is identical to the proof of Claim 4.6. □

We next define the notion of shadow closed sets.

Definition 4.11 (Shadow closed)**.** Let $S \subseteq \{0, \ldots, p^n - 1\}$. The set S is said to be *shadow closed* if the following holds. For any $e \in S$, let $e = \sum_{i=0}^{n-1} e_i p^i$ be the representation of e in base p. Define the *support* of e to be the set of nonzero digits of e,

$$\text{support}(e) = \{0 \leq i \leq n-1 : e_i \neq 0\}.$$

Let e' be obtained from e by changing some of the non-zero digits of e, i.e.

$$e' = \sum_{i \in \text{support}(e)} e'_i p^i.$$

Then we should have that also $e' \in S$. That is, S is shadow closed if

$$\left\{ \sum_{i \in \text{support}(e)} e'_i p^i : e \in S, (e'_i)_{i \in \text{support}(e)} \in \mathbb{F}_p \right\} \subseteq S.$$

Definition 4.12 (Affine closed)**.** A set $S \subseteq \{0, \ldots, p^n - 1\}$ is *affine closed* if it is both shift closed and shadow closed.

We recall the following theorem of Kaufman and Sudan [15] that we presented in the introduction. It shows that affine invariant linear codes are equivalent to trace codes over affine closed sets.

Theorem 4.13 (Equivalence of affine invariant codes and trace codes of affine closed sets [15])**.** *Let $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$ be an affine invariant linear code. Then there exists an affine closed set $S \subseteq \{0, \ldots, p^n - 1\}$ such that $\mathcal{C} = \mathcal{T}(S)$. Moreover, for any affine closed set S the code $\mathcal{T}(S)$ is linear and affine invariant.*

4.3 Weight distribution of affine invariant codes

Theorem 4.13 tells us that in order to study affine invariant codes, it suffices to study trace codes of affine closed sets. In this subsection we establish the following lemma, which gives a tight estimate on the number of codewords in $d\mathcal{T}(S)$ for affine closed sets S. For the statement of the lemma recall that $R = \{1 \leq e \leq p^n - 1 : (e, p) = 1\}$ is the set of elements co-prime to p.

Lemma 4.14 (Weight distribution of dual trace affine closed codes)**.** *Let $S \subseteq \{0, \ldots, p^n - 1\}$ be affine closed of size $|S \cap R| \leq O(n^{\frac{\log p}{\log 2p}} / \log^2 n)$. Then there exists $\ell_{\min} = O(|S \cap R|)$ and*

$\ell_{\max} = \Omega(p^{n/2})$ *such that for any* $\ell \in [\ell_{\min}, \ell_{\max}]$ *the following holds. The number of codewords in* $d\mathcal{T}(S)$ *of weight* ℓ *is given by*

$$\frac{C(p,\ell)}{\ell!} p^{n(\ell - |S \cap R|)} (1 \pm 0.1)$$

where $C(p,\ell)$ *is defined as*

$$C(p,\ell) = \left| \left\{ (v_1, \ldots, v_\ell) \in (\mathbb{F}_p \setminus \{0\})^\ell : v_1 + \ldots + v_\ell = 0 \right\} \right|.$$

We start by showing a general bound on the weight degree of elements of affine closed sets, in terms of the size of the set.

Claim 4.15 (Weight degree bound on affine closed sets). Let $S \subseteq \{0, \ldots, p^n - 1\}$ such that S is affine closed. Then for any $e \in S$,

$$\mathrm{wt}(e) \leq \log_p |S \cap R| + 1.$$

Proof. Let $S' = S \cap R$. Let $e \in S$ be of weight $k \geq 1$. By taking some shift of e we may assume $e \in R$ (that is, $0 \in \mathrm{support}(e)$), hence $e \in S' = S \cap R$. Consider the set

$$E' = \left\{ \sum_{i \in \mathrm{support}(e)} e'_i p^i : e'_i \in \mathbb{F}_p, \; e'_0 \neq 0 \right\}.$$

Note that as S is shadow closed, we have $E' \subseteq S$. Moreover since $e'_0 \neq 0$ we have $E' \subseteq R$, hence $E' \subseteq S' = S \cap R$. Thus $|E'| \leq |S'|$. On the other hand,

$$|E'| = (p-1)p^{\mathrm{wt}(e)-1}.$$

Hence we conclude that $\mathrm{wt}(e) \leq \log_p(\frac{p}{p-1}|S'|) \leq \log_p |S'| + 1$. □

We will need the following simple claim.

Claim 4.16 (Trace is not constant). Let $f(x) = \sum_{e \in R} \alpha_e x^e$ be a nonzero polynomial. Then $\mathrm{Tr}(f(x)) : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is not a constant function.

Proof. Assume for contradiction that $\mathrm{Tr}(f(x)) = a$ for all $x \in \mathbb{F}_{p^n}$. Let $q(x) = \mathrm{Tr}(f(x)) - a$. We have

$$q(x) = -a + \sum_{i=0}^{n-1} (\sum_{e \in R} \alpha_e x^e)^{p^i} = -a + \sum_{i=0}^{n-1} \sum_{e \in R} (\alpha_e)^{p^i} x^{ep^i \pmod{p^n}}.$$

Since $e \in R$ all the degrees $ep^i \pmod{p^n}$ are distinct and different from 0. Thus $q(x)$ is not the zero polynomial. Since $\deg(q) \leq p^n - 1$ we have that there must be x such that $q(x) \neq 0$, hence $\mathrm{Tr}(f(x)) \neq a$. □

The next lemma gives an estimation on the weight distribution of $d\mathcal{T}(S)$ where $S' = S \cap R$ is a relatively small set of elements of small weight degree. We will then show that the lemma can be applied to any affine invariant set S which is not too large.

Lemma 4.17 (Weight distribution of dual trace codes of reduced form sets)**.** *Fix $d \leq O(\log n)$. Let $S' \subseteq R$ be such that for any $e \in S'$ its weight degree is at most $\mathrm{wt}(e) \leq d$ and $|S'| \leq O(\frac{n}{d^2 2^d})$. Then there exist $\ell_{\min} = O(|S'|)$ and $\ell_{\max} = \Omega(p^{n/2})$, such that for any $\ell \in [\ell_{\min}, \ell_{\max}]$, the number of codewords in $d\mathcal{T}(S \cup \{0\})$ of weight ℓ is given by*

$$\frac{C(p,\ell)}{\ell!} p^{n(\ell - |S'|)}(1 \pm 0.1).$$

Proof. The proof follows immediately from Theorem 1.6 using the bias bounds given by Theorem 1.3. □

We can now deduce Lemma 4.14 from Claim 4.15 and Lemma 4.17.

Proof of Lemma 4.14. Let $S \subseteq \{0, \ldots, p^n - 1\}$ be affine closed. Let $S' = S \cap R$. We have that $d\mathcal{T}(S) = d\mathcal{T}(S' \cup \{0\})$. By Claim 4.15 the maximal weight of elements in S is at most $d \leq \log_p |S'| + 1$. Assume that $|S'| \leq O(n^{\frac{\log p}{\log 2p}} / \log^2 n)$. It can be verified that this satisfy the condition in Lemma 4.17 that $|S'| = O(\frac{n}{d^2 2^d})$. The claim thus follows by an application of Lemma 4.17. □

4.4 Trace codes of quasi-polynomial size are generated by a single orbit

We prove in this subsection that any affine invariant linear code of dimension up to $n^{1+\epsilon}$ is generated by a single orbit of a dual codeword for any $\epsilon < \log(p)/\log(2p)$ and large enough n. Combining this with Theorem 1.12 we get that any such code is locally testable, which prove our testing result, Theorem 1.7. We now state the main theorem we prove in this subsection.

Theorem 4.18 (Affine invariant codes are generated by a single orbit)**.** *Let $\mathcal{C} = \{f : \mathbb{F}_{p^n} \to \mathbb{F}_p\}$ be an affine invariant linear code such that $\dim(\mathcal{C}) \leq O(n^{1+\frac{\log p}{\log 2p}} / \log^2 n)$. Then there exists $f \in \mathcal{C}^{\perp}$ such that*

$$\overline{\mathrm{affine}}(f)^{\perp} = \mathcal{C}$$

and of weight

$$\mathrm{wt}(f) \leq O(\dim(\mathcal{C})/n).$$

Let $\mathcal{C} = \mathcal{T}(S)$ be an affine invariant code where $S \subseteq \{0, \ldots, p^n - 1\}$ is affine closed. We start by showing that if some $f \in \mathcal{C}^{\perp} = d\mathcal{T}(S)$ does not generate $d\mathcal{T}(S)$, then in fact $f \in d\mathcal{T}(S \cup \{e\})$ where $e \in \{1, \ldots, p^n - 1\} \backslash S$ has small weight (Corollary 4.21). From this and the exact estimates for the weight distribution for dual trace codes we derive Theorem 4.18. Before proving Corollary 4.21 we will require two technical claims.

Claim 4.19. Let $S \subseteq \{0, \ldots, p^{n-1}\}$ be affine closed. Let $f \in d\mathcal{T}(S)$ be a codeword which does not affinely generate $d\mathcal{T}(S)$, i.e.

$$\overline{\mathrm{affine}}(f) \subsetneq d\mathcal{T}(S).$$

Then

$$\overline{\mathrm{affine}}(f) = d\mathcal{T}(T)$$

for some affine closed $T \supsetneq S$.

Proof. The code $\overline{\text{affine}}(f)$ is an affine invariant code which is a proper subset of $d\mathcal{T}(S)$. By Theorem 4.13 we know that $\overline{\text{affine}}(f) = d\mathcal{T}(T)$ for some affine closed $T \subseteq \{0, \ldots, p^n - 1\}$. Since $d\mathcal{T}(T) \subsetneq d\mathcal{T}(S)$ we must have that $T \supsetneq S$. $\square$

Claim 4.20. Let $S \subsetneq T \subseteq \{0, \ldots, p^n - 1\}$ such that both S and T are affine closed. Then there exist an element $e \in (T \setminus S) \cap R$ such that

$$\text{wt}(e) \leq \log_p |S \cap R| + 2.$$

Proof. Let $S' = S \cap R$ and $T' = T \cap R$. We have $S' \subsetneq T'$ as otherwise, if $S' = T'$, we would have $S = \overline{\text{affine}}(S') = \overline{\text{affine}}(T') = T$.

Let $k = \lfloor \log_p |S'| \rfloor + 2$. We argue there is $e \in T' \setminus S'$ of weight at most k. Otherwise, let $e \in S' \setminus T'$ such that $\text{wt}(e) > k$. Consider the set

$$E = \overline{\text{shadow}}(e) \cap R = \left\{ \sum_{i \in \text{support}(e)} e'_i p^i : e'_i \in \mathbb{F}_p,\ e'_0 \neq 0 \right\},$$

where we use the fact that since $e \in R$ then $0 \in \text{support}(e)$. Note that by definition, $E \subseteq T'$, since T is affine closed hence in particular shadow closed.

Let $e' \in E \subseteq T'$ such that $\text{wt}(e') = k$ (by setting $\text{wt}(e) - k$ digits of e in base p to zero). Consider the set

$$E' = \overline{\text{shadow}}(e') \cap R = \left\{ \sum_{i \in \text{support}(e')} e''_i p^i : e''_i \in \mathbb{F}_p,\ e''_0 \neq 0 \right\}.$$

Note that since $|E'| = (p-1)p^{\text{wt}(e')-1} = (p-1)p^{k-1} > |S'|$ we cannot have that $e' \in S'$. Hence we found an element $e' \in T' \setminus S'$ such that $\text{wt}(e') \leq k$. $\square$

Corollary 4.21. *Let $S \subseteq \{0, \ldots, p^{n-1}\}$ be affine closed. Let $f \in d\mathcal{T}(S)$ be a codeword which does not affinely generate $d\mathcal{T}(S)$, i.e.*

$$\overline{\text{affine}}(f) \subsetneq d\mathcal{T}(S).$$

Then there must exist $e \in R \setminus S$ of weight $\text{wt}(e) \leq \log_p |S \cap R| + 2$ such that

$$f \in d\mathcal{T}(S \cup \{e\}).$$

Proof. By Claim 4.19 we have $\overline{\text{affine}}(f) = d\mathcal{T}(T)$ where $T \supsetneq S$. By Claim 4.20 there is $e \in (T \setminus S) \cap R \subseteq R \setminus S$ such that $\text{wt}(e) \leq \log_p |S \cap R| + 2$. Hence we conclude sicne

$$f \in d\mathcal{T}(T) \subseteq d\mathcal{T}(S \cup \{e\}).$$

$\square$

We are now ready to prove Theorem 4.18.

Proof of Theorem 4.18. Let $\mathcal{C}$ be a linear affine invariant code. By theorem 4.13 we have $\mathcal{C} = \mathcal{T}(S)$ where $S \subseteq \{0, \ldots, p^n - 1\}$ is affine closed. By Claims 4.4, 4.6 and 4.7 we have that

$$|\mathcal{C}| = \mathcal{T}((S \cap R) \cup \{0\}) = p^{n|S \cap R|+1}.$$

Let $\ell = O(|S \cap R|)$. We count the number of codewords in $d\mathcal{T}(S)$ of weight ℓ. To this end we apply Lemma 4.14. The number of codewords in $d\mathcal{T}(S)$ of weight ℓ is given by

$$W_\ell = \frac{C(p, \ell)}{\ell!} p^{n(\ell - |S \cap R|)}(1 \pm 0.1).$$

Let $f \in d\mathcal{T}(S)$ be such that $\overline{\text{affine}}(f) \subsetneq d\mathcal{T}(S)$. By Corollary 4.21 we know that there exists some $e \in R \setminus S$ of weight $\text{wt}(e) \leq k$, where $k \leq \log_p(|S \cap R|) + 2$, such that $f \in d\mathcal{T}(S \cup \{e\})$. Let E be the set of all such possible e,

$$E = \{e \in R \setminus S : \text{wt}(e) \leq k\}.$$

Fix some $e \in E$. Let $S_e = \overline{\text{affine}}(S \cup \{e\})$. Note that as $e \in R \setminus S$ we have $|S_e \cap R| \geq |S \cap R| + 1$. Hence for ℓ in the permissible range for S_e we get that the number of codewords of weight ℓ in $d\mathcal{T}(S_e)$ is given by

$$\frac{C(p, \ell)}{\ell!} p^{n(\ell - |S_e \cap R|)}(1 \pm 0.1) \leq 2p^{-n} W_\ell.$$

So, as long as $|E| \ll p^n$, we can deduce that there must exist some $f \in d\mathcal{T}(S)$ of weight ℓ which is not in any of $d\mathcal{T}(S \cup \{e\})$ for any $e \in E$ (in fact, almost any $f \in d\mathcal{T}(S)$ of weight ℓ will do). Thus, to conclude the theorem we need to bound $|E|$, which is simple since

$$|E| \leq \sum_{i=1}^{k} \binom{n}{i} p^i \leq n^{O(\log n)} \ll p^n.$$

□

References

[1] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn and Dana Ron, *Testing Low Degree Polynomials Over GF(2)*, Proceedings of 7th International Workshop on Randomization and Computation,(RANDOM), Lecture Notes in Computer Science 2764, 188-199, 2003. Also, IEEE Transactions on Information Theory, Vol. 51(11), 4032-4039, 2005.

[2] J. Bourgain, *Mordell's exponential sum estimate revisited*, J. Amer. Math. Soc., 18(2):477-499 (electronic), 2005.

[3] G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*. third edition, MacMillan, New York, 1965.

[4] Blum, M., Luby, M., Rubinfeld, R., *Self-Testing/Correcting with Applications to Numerical Problems*, In J. Comp. Sys. Sci. Vol. 47, No. 3, December 1993.

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

[6] L. Carlitz and S. Uchiyama, *Bounds for exponential sums*, Duke Math. J., 24:37-41, 1957.

[7] P. Deligne, *Aplications de la formule des traces aux sommes trigonometriques*, in SGA $4\frac{1}{2}$ Springer Lecture Notes in Math 569, 1978.

[8] Elena Grigorescu, Tali Kaufman and Madhu Sudan, *Succinct Representation of Codes with Applications to Testing*, manuscript.

[9] Ariel Gabizon, Ran Raz, *Deterministic extractors for affine sources over large fields*, Combinatorica 28(4): 415-440 (2008).

[10] Charanjit S. Jutla, Anindya C. Patthak, Atri Rudra and David Zukcerman , *Testing low-degree polynomials over prime fields*, Proceedings of the 45th Annual Symposium on Foundations of Computer Science (FOCS), pp. 423-432, 2004.

[11] Tali Kaufman and Simon Litsyn, *Almost Orthogonal Linear Codes are Locally Testable*, FOCS 2005: 317-326.

[12] Tali Kaufman and Shachar Lovett, *The List-Decoding Size of Reed-Muller Codes*, ICS 2010.

[13] Tali Kaufman and Dana Ron, *Testing polynomials over general fields*, Proceedings of the 45th Annual Symposium on Foundations of Computer Science (FOCS), pp. 413-422, 2004.

[14] Tali Kaufman and Madhu Sudan, *Sparse random linear codes are locally decodeable and testable*, FOCS 2007, pp. 590–600.

[15] Tali Kaufman and Madhu Sudan, *Algebraic Property Testing: The Role of Invariance*, Proceedings of the 40th ACM Symposium on Theory of Computing (STOC), 2008.

[16] Tali Kaufman, Avi Wigderson, *Symmetric LDPC Codes and Local Testing*, ICS 2010, 406-421.

[17] Swastik Kopparty and Shubhangi Saraf, *Local List-Decoding and Testing of Random Linear Codes from High-Error*, to appear in the Proceedings of STOC 2010.

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

[19] F. J. MacWilliams and N. J. A. Sloan, *The Theory of Error Correcting Codes*, North Holland, Amsterdam, 1977.

[20] Ronitt Rubinfeld and Madhu Sudan, *Robust characterizations of polynomials with applications to program testing*, SIAM Journal on Computing, 25(2):252-271, April 1996.

[21] Madhu Sudan *Invariance in Property Testing* ECCC, TR10-051, 2010.

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

[23] A. Weil, *Sur les courbes algebriques et les varietes qui s'en deduisent*, Actualities Sci. et Ind. no. 1041. Hermann, Paris, 1948.

[24] Avi Wigderson, *Deterministic Extractors - Lecture Notes*, www.math.ias.edu/ avi/TALKS/Pseudorandomness-mini-workshop2009.pdf.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de