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

Tali Kaufman
Bar-Ilan University and Weizmann
kaufmant@mit.edu

Shachar Lovett
Institute for Advanced Study
slovett@math.ias.edu

***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 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.

I. 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.

A. 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 [14] 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 I.1 (Weil bound [14]). *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 ([5], [15]) and in the study of locally testable codes ([4], [10]). 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 $\mathbb{F}_{p^n}$

Research of the first author supported in part by the Alon Fellowship. Research of the second author support by NSF grant DMS-0835373.

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 I.2 (Weight degree)**.** Let $t \in \{0, \dots, p^n - 1\}$. 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 I.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 [1] 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 I.4 (Bourgain's extension of Weil bound [1])**.** *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 I.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 I.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 I.5. *Fix $d \leq O(\log n)$. 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 $1 - 1/p - p^{-\Omega(n)}$.*

B. 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. [6]) 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 I.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 $C_{p,\ell} = |\{v_1, \dots, v_\ell \in \mathbb{F}_p \setminus \{0\} : \sum_{i=1}^{\ell} v_i = 0\}|$.*

C. 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 [6], [8], [4], [11]. 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 [4] 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 I.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

[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 $\alpha, \beta \in \mathbb{F}_p$.

[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 $\Omega(\epsilon)$.

complexity $q = O(1)$.

Our result generalizes the result of Grigorescu, Kaufman and Sudan [4] in few aspects: First, the result of [4] 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 [4] 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 [4], 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 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.

II. Proofs overview

In this section we provide overviews of the proofs of our main theorems. Due to space limitations, this proceeding version contains only the full proof of the new estimation of the weight distribution of codes with large dual distance, i.e. Theorem I.6 in Section III. Other full proofs can be found in the full version of this paper.

A. 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 [2], [12], [13] and list-decoding of codes [7].

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 [3], 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.

B. 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

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

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

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

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 $\mathbb{E}_{x_i\in[N]}\left[\omega^{v_i f(x_i)}\right] = 1$. If f is non constant $\left|\mathbb{E}_{x_i\in[N]}\left[\omega^{v_i f(x_i)}\right]\right| \le \mathrm{bias}(\mathcal{C}) \le \delta$. Hence we deduce that

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

where $|\eta| \le |\mathcal{C}|\delta^\ell$. 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.

C. 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 [4]. 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

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

where we recall that the Trace function $\mathrm{Tr} : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is given by $\mathrm{Tr}(x) = \sum_{i=0}^{n-1} x^{p^i}$. For example, Generalized Reed-Muller codes $\mathrm{RM}_p(n, d)$, 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

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

We define two important properties of trace codes.

Definition II.1 (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 II.2 (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,

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

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

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

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

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

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 [9]. They show that the class of affine invariant linear codes is equivalent to the class of trace codes of affine closed sets.

Theorem II.3 (Monomial extraction [9])**.** *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

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

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

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

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

Definition II.4 (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

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

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,

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

The following result was established by Kaufman and Sudan [9]. 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].

Theorem II.5 (Theorem 2.9 in [9])**.** *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 to be

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

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 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. *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 $|S'| \leq n^\epsilon$ 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*

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

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

$$C(p,\ell) = \left| \left\{ (v_1, \ldots, v_\ell) \in (\mathbb{F}_p \setminus \{0\})^\ell : \sum_{i=1}^{\ell} v_i = 0 \right\} \right|.$$

[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 $g(x) \neq 0$.

Similar results were previously obtained over binary fields $\mathbb{F}_2$ using properties of Krawtchouk polynomials [6], [8]. 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 I.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 $|\mathcal{C}| \leq p^{n^{1+\epsilon}}$ then in fact $\mathcal{C} = d\mathcal{T}(S)$ where S is affine invariant, and $|S'| \leq n^{\epsilon}$, 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

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

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.

III. Weight distribution of codes with large dual distance

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

A. 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 $f(x) \neq 0$,

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

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

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

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,

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

We would be interested in a related notion, which is the distance between $\mathcal{C}$ and constant codewords. Let $\text{Const} = \{a^N : a \in \mathbb{F}_p\}$ be the code of constant codewords. Note that as $\mathcal{C}$ is linear, we either have that $\text{Const} \subset \mathcal{C}$ or that $\text{Const} \cap \mathcal{C} = \{0^N\}$. We define $\text{dist}(\mathcal{C}, \text{Const})$ 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 $f \in \text{Const}$. 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 III.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 $\mathrm{dist}(\mathcal{C}, \mathrm{Const}) \geq 1 - 1/p + \delta$ then

$$\mathrm{bias}(\mathcal{C}) \leq 2p\delta.$$

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

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

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

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

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

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

■

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 I.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 $\mathcal{C} \cap \mathrm{Const} = \{0\}^n$, Number of codewords in $\mathcal{C}^\perp$ of weight ℓ:

$$\frac{(p-1)^\ell}{\ell!} \frac{N^\ell}{|\mathcal{C}|}(1 \pm 2\epsilon).$$

(ii) If $\mathrm{Const} \subset \mathcal{C}$, Number of codewords in $\mathcal{C}^\perp$ of weight ℓ:

$$\frac{C(p,\ell)}{\ell!} \frac{N^\ell}{|\mathcal{C}|}(1 \pm 2\epsilon).$$

where $C(p,\ell)$ is equal to the following:

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

B. Proof of Theorem I.6

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

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

and

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

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, 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 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

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

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 $v_i \in \mathbb{F}_p \backslash \{0\}$ then also f_i is not constant. Hence

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

Hence we deduce that

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

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

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. $v_a^{(i,j)} = v_a$ for $1 \le a < i$ and $i < a < j$, $v_i^{(i,j)} = v_i + v_j$, $v_a^{(i,j)} = v_{a+1}$ 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

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

Hence we get that as long as $\ell^2 \le \epsilon N$ 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 $\mathrm{Const} \subset \mathcal{C}$. Define $\mathcal{C}' = \{f \in \mathcal{C} : f(0) = 0\}$ so that $\mathcal{C} = \{f' + f'' : f' \in \mathcal{C}', f'' \in \mathrm{Const}\}$, $\mathrm{bias}(\mathcal{C}') = \mathrm{bias}(\mathcal{C})$ and $\mathcal{C}' \cap \mathrm{Const} = \{0\}^n$. 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 that the number of codewords in $\mathcal{C}^\perp$ of weight ℓ is equal to $\frac{1}{\ell!}\sum_{v \in C(p,\ell)} |B_\ell(v)|$, and the proof follows by the estimates we proved on $|B_\ell(v)|$.

References

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

[2] 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.

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

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

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

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

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

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

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

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

[11] 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.

[12] 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.

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

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

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