

# Improved Optimal Testing Results from Global Hypercontractivity

Tali Kaufman \*

Dor Minzer †

## Abstract

The problem of testing low-degree polynomials has received significant attention over the years due to its importance in theoretical computer science, and in particular in complexity theory. The problem is specified by three parameters: field size  $q$ , degree  $d$  and proximity parameter  $\delta$ , and the goal is to design a tester making as few as possible queries to a given function, which is able to distinguish between the case the given function has degree at most  $d$ , and the case the given function is  $\delta$ -far from any degree  $d$  function.

With respect to these parameters, we say that a tester is *optimal* if it makes  $O(q^d + 1/\delta)$  queries (which are known to be necessary). For the field of size  $q$ , such tester was first given by Bhattacharyya et al. for  $q = 2$ , and later by Haramaty et al. [7] for all prime powers  $q$ . In fact, they showed that the natural  $t$ -flat tester is an optimal tester for the Reed-Muller code, for an appropriate  $t$ . Here, the  $t$ -flat tester is the tester that picks a uniformly random affine subspace  $A$  of dimension  $t$ , and checks that  $\deg(f|_A) \leq d$ . Their analysis proves that the dependency of the  $t$ -flat tester on  $\delta$  and  $d$  is optimal, however the dependency on the field size, i.e. the hidden constant in the  $O$ , is a tower-type function in  $q$ .

We improve the result of Haramaty et al., showing that the dependency on the field size is polynomial. Our technique also applies in the more general setting of lifted affine invariant codes, and gives the same polynomial dependency on the field size. This answers a problem raised in [6].

Our approach significantly deviates from the strategy taken in earlier works [2, 7, 6], and is based on studying the structure of the collection of erroneous subspaces, i.e. subspaces  $A$  such that  $f|_A$  has degree greater than  $d$ . Towards this end, we observe that these sets are poorly expanding in the affine version of the Grassmann graph and use that to establish structural results on them via global hypercontractivity. We then use this structure to perform local correction on  $f$ .

## 1 Introduction

The Reed-Muller code is one of the most basic and useful codes in theoretical computer science. A key aspect of the Reed-Muller code, which plays a significant role in its applications to complexity theory and in particular in the construction of probabilistically checkable proofs, is its local testability. Namely, given a truth table of a function over a field, we wish to be able to distinguish between the case that this truth table represents a Reed-Muller codeword, i.e. a low degree function, and the case it is far from any Reed-Muller codeword.<sup>1</sup>

Usually, the notion of local-testability of the Reed-Muller codes asserts that when the degree  $d$ , the field size  $q$  and proximity parameter  $\delta$  are all thought of as constants, then there is a tester whose query complexity is constant. With regards to this definition, earlier works [1, 9, 8] showed that the Reed-Muller

\*Department of Computer Science, Bar-Ilan University.

†Department of Mathematics, Massachusetts Institute of Technology, Cambridge, USA. Supported by a Sloan Research Fellowship.

<sup>1</sup>Variations of this problems exists, such as when instead of giving the truth table of a function, one is given a table of supposed restrictions of the function to higher dimensional objects such as lines or planes; see for example [14].

code is testable. The current work is mainly concerned with the stronger notion of *optimal testers* for the Reed-Muller codes. Here, we wish to get a tester whose query complexity is tight with respect to  $d$ ,  $q$  and  $\delta$  when they are not thought of as constant. A typical setting to think about is when the proximity parameter is fairly small,  $0 < \delta \leq q^{-d}$ , in which case it is clear that any tester for the corresponding Reed-Muller code must make at least  $\Omega(1/\delta)$  queries.

With respect to this notion, it was shown that the Reed-Muller codes are optimally testable: first in [2] for  $\mathbb{F}_2$ , and then to general fields [7]. These works get an optimal dependency of the query complexity on the degree parameter and the proximity parameter, however they only apply in the case the field size is relatively small. Indeed, the dependency of the rejection probability on the field size is inverse tower-type, which stems from the fact that their proof utilizes the Density Hales-Jewett theorem.

The main result of this paper is an improvement of the above mentioned results, getting an optimal dependency on the degree parameter while simultaneously getting a polynomial dependency of the field size. Our approach significantly deviates from the strategy taken in [2, 7], and is based on studying expansion properties in the associated affine Grassmann graph. As a side contribution, we prove versions of expansion theorems that were previously shown for the Grassmann graph [3, 13] to the affine Grassmann graph, which turns out to include slight additional complications.

We hope the approach presented herein could be useful in proving optimal testing results for other codes. Indeed, while our argument does use specific properties of polynomials, it does so “minimally” and the structure of the underlying structure of the queries plays a more important role.

## 1.1 Local testability of Reed-Muller codes

Throughout this paper,  $p$  denotes a prime number and  $q$  denotes a power of  $p$ .

**Definition 1.1.** For a function  $f: \mathbb{F}_q^n \rightarrow \mathbb{F}_q$ , we denote

$$\delta_d(f) = \min_{\substack{g: \mathbb{F}_q^n \rightarrow \mathbb{F}_q \\ \text{of degree } d}} \Pr_{x \in \mathbb{F}_q^n} [f(x) \neq g(x)].$$

In this paper, we consider the  $t$ -flat tester which is parameterized by a dimension  $t$ . Here and throughout, a  $t$ -flat of a given vector space  $W$  (say  $W = \mathbb{F}_q^n$ ) is a  $t$ -dimensional affine subspace of it. The  $t$ -flat tester works by sampling a random  $t$ -flat  $T \subseteq \mathbb{F}_q^n$ , and checking that  $f|_T$  has degree at most  $d$ . The  $t$  that we pick is the minimal one that makes sense – i.e. the minimal  $t$  such that each  $f: \mathbb{F}_q^n \rightarrow \mathbb{F}_q$  of degree larger than  $d$  fails the test with positive probability, which turns out to be  $t = \lceil \frac{d+1}{q-q/p} \rceil$  [8, 9].

**Definition 1.2.** Given a function  $f: \mathbb{F}_q^n \rightarrow \mathbb{F}_q$ , and  $t, d \in \mathbb{N}$ , the  $t$ -flat test proceeds by picking an affine subspace  $T$  of dimension  $t$ , and testing if  $f|_T$  is a degree  $d$  polynomial. The rejection probability of this test is denoted by  $\varepsilon_{t,d}(f)$ .

Let us focus, for a moment, on the case that  $q = 2$ . In this case the  $t$ -flat test was first analyzed in [1], who proved that  $\varepsilon_{t,d}(f) \geq q^{-t} \delta_d(f)$ . An improved analysis of the tester was given in [2], who showed that the  $t$ -flat tester is in fact an optimal tester, and in particular that  $\varepsilon_{t,d}(f) \geq \min(c, q^t \delta_d(f))$  for some absolute constant  $c > 0$ . The result was later generalized to general fields in [7], which reads:

**Theorem 1.3.** For all primes  $p$  and  $q$  powers of  $p$ , there is  $c(q) > 0$  such that for all  $d \in \mathbb{N}$  and  $f: \mathbb{F}_q^n \rightarrow \mathbb{F}_q$  it holds that for  $t = \lceil \frac{d+1}{q-q/p} \rceil$  we have

$$\varepsilon_{t,d}(f) \geq c(q) \min(1, q^t \delta_d(f)).$$

The analysis of both [2] and [7] follows the same high-level inductive approach on the dimension  $n$ . Assuming the rejection probability of a given function  $f: \mathbb{F}_q^n \rightarrow \mathbb{F}_q$  is small, one considers the restriction of  $f$  to hyperplanes (i.e., to subspaces of  $\mathbb{F}_q^n$  of co-dimension 1), on which the inductive hypothesis gives, for each hyperplane  $W$ , a candidate degree  $d$  function that is close to  $f|_W$ . The main task then (both in [2] and in [7]) is to “sew” together these candidate functions. Towards this end, a careful choice of the collection of hyperplanes that are most convenient for the task must be made. This choice is rather simple for  $\mathbb{F}_2$ , but becomes much more complex in  $\mathbb{F}_q$ , and to do so the authors use Ramsey-type results, more specifically the Density Hales-Jewett theorem. This ultimately leads to an inverse tower-type bound dependency of  $c(q)$  on  $q$ .

Our main result is an improved quantitative version of Theorem 1.3. Namely, we prove:

**Theorem 1.4.** *For all primes  $p$  a prime power  $q$  of  $p$ , there is  $c(q) = q^{-O(1)}$  such that for all  $d \in \mathbb{N}$  and  $f: \mathbb{F}_q^n \rightarrow \mathbb{F}_q$  it holds that for  $t = \lceil \frac{d+1}{q-q/p} \rceil$  we have*

$$\varepsilon_{t,d}(f) \geq c(q) \min(1, q^t \delta_d(f)).$$

Our approach is significantly different from the previously mentioned inductive hypothesis. At a high level, we consider the set of erroneous  $t$ -flats, i.e.  $S = \{T \mid \dim(T) = t, f|_T \text{ is not degree } d\}$  and establish a structural result on it. Interestingly, our starting point is a lemma from [7] (which is a variant of a lemma already appearing in [2]), which in our language upper bounds the measure of the upper shadow of  $S$  as a function of the measure of  $S$ . Here, the upper shadow of  $S$  is

$$S \uparrow = \{B \mid \dim(B) = t + 1, \exists T \in S \text{ such that } T \subseteq B\},$$

and the lemma from [7] asserts that  $\mu(S \uparrow) \leq q \cdot \mu(S)$ . In [2, 7] this lemma is used to relate the rejection probability of the  $t$ -flat tester and the  $(t+k)$ -flat tester; in particular it implies that the rejection probability of the  $t$ -flat tester is at least  $q^{-k}$  times the rejection probability of the  $(t+k)$ -flat tester.

We use this lemma in a different way. The point here is that as  $S$  is a small set, the condition that  $\mu(S \uparrow) \leq q \cdot \mu(S)$  is already itself very restrictive. Examples of  $S$  that exhibit such behaviours can be thought of as the subspace analog of collections of subsets that are nearly tight for the classical Kruskal-Katona theorem. Indeed, a natural type of such small set is

$$H_x = \{T \mid \dim(T) = t, T \ni x\}.$$

We show (simplifying matters somewhat) that indeed, any small  $S$  such that  $\mu(S \uparrow) \leq q\mu(S)$  must almost contain a copy of  $H_x$  for some  $x$ . This suggests that an error occurs at  $x$  and that we should change the value of  $f(x)$ . Indeed, this is the high level strategy we pursue, and we defer a more detailed description to Section 1.4.

## 1.2 Lifted affine invariant linear codes

Our argument in the proof of Theorem 1.4 also applies in the more general setting of lifted affine invariant codes. In this case, an analogous result to Theorem 1.3 was proved in [6] with the same type of inverse-type tower dependency on the field size. Our proof gives a polynomial dependency on the field size making progress along an open problem raised in [6].

To present our result for lifted affine invariant codes, we quickly recall the setting. Let  $q$  be a prime power,  $t \in \mathbb{N}$  and suppose  $\mathcal{B} \subseteq \{g: \mathbb{F}_q^t \rightarrow \mathbb{F}_q\}$  is an affine invariant set of function. By that, we mean

that  $\mathcal{B}$  is closed under composition with affine transformations. Given  $n \geq t$ , the  $n$ -lift of  $\mathcal{B}$  denoted by  $\mathcal{F} = \text{Lift}_n(\mathcal{B})$  is defined as

$$\mathcal{F} = \{ f: \mathbb{F}_q^n \rightarrow \mathbb{F}_q \mid \forall t\text{-flats } A \subseteq \mathbb{F}_q^n, f|_A \in \mathcal{B} \}.$$

For  $k \geq t$ , the  $k$ -flat tester proceeds by taking a  $k$ -flat  $A \subseteq \mathbb{F}_q^n$  randomly, and checking that  $f|_A \in \text{Lift}_k(\mathcal{B})$ , in which case we say the test accepts. We denote by  $\varepsilon_k(f)$  the probability that the  $k$ -flat tester rejects on  $f$ .

**Theorem 1.5.** *For all prime powers  $q$ , there is  $c(q) = q^{-O(1)}$  such that for  $t \in \mathbb{N}$ , if  $\mathcal{B} \subseteq \{g: \mathbb{F}_q^t \rightarrow \mathbb{F}_q\}$  is an affine invariant linear code, and  $f: \mathbb{F}_q^n \rightarrow \mathbb{F}_q$ , then*

$$\varepsilon_t(f) \geq c(q) \min(1, q^t \Delta(f, \mathcal{F})),$$

where  $\Delta(f, \mathcal{F})$  is the relative Hamming distance between  $f$  and  $\mathcal{F}$ .

Using a reduction from [6], one may use Theorem 1.5 in order to get the following slightly more general result. The proof is exactly as in [6, Section 7], and is hence omitted.

**Theorem 1.6.** *For all primes  $p$ , a power  $q$  of  $p$ , and  $Q$  a power of  $q$ , there is  $c(Q) = Q^{-O(1)}$  such that for  $t \in \mathbb{N}$ , if  $\mathcal{B} \subseteq \{g: \mathbb{F}_Q^t \rightarrow \mathbb{F}_q\}$  is an affine invariant linear code, and  $f: \mathbb{F}_Q^n \rightarrow \mathbb{F}_q$ , then*

$$\varepsilon_t(f) \geq c(Q) \min(1, Q^t \Delta(f, \mathcal{F})),$$

where  $\Delta(f, \mathcal{F})$  is the relative Hamming distance between  $f$  and  $\mathcal{F}$ .

### 1.3 The affine Grassmann graph

To execute our approach we consider the affine Grassmann graph along with an appropriate random walk on it. Given an affine space  $W$  of dimension  $k$  over  $\mathbb{F}_q$  and an integer  $\ell < k$ , the affine Grassmann graph  $\text{AffGras}(W, \ell)$  contains as vertices all  $\ell$ -flats of  $W$ , which we denote by  $V(k, \ell)$  when the space  $W$  is clear from the context. The edges of the graph are thought of as weighted according to the following randomized process: starting from a  $\ell$ -flat  $A$ , we take a random  $(\ell + 1)$ -flat satisfying  $A \subseteq B \subseteq W$ , and then take a random  $\ell$ -flat  $A_2 \subseteq B$ ; the weight of the edge  $(A, A_2)$  is the probability it is sampled by this process.

Our first observation is that combining the lemma from [7] with sharp-threshold type result from [11], one concludes that

$$1 - \Phi(S) \geq \frac{1}{q}.$$

Here,  $\Phi(S)$  is the expansion of the set  $S$ , defined as  $\Pr_{A \in S, A' \sim A \text{ neighbour}} [A' \notin S]$ . Thus, we would be able to gain significant insight into the structure of  $S$  provided we could give sufficiently good characterization of sets in the affine Grassmann graph that are poorly expanding. This is exactly the type of question that was studied recently in the context of the 2-to-2 Games Theorem [12, 4, 3, 13], and we leverage insights gained from there in our case of interest.

The affine variant of the Grassmann graph includes further complications, which we explain next. Roughly speaking, the eigenvalues of it are  $q^{-i}$  for  $i = 0, \dots, \ell$ , hence it can be shown that for sets  $S$  of size smaller than  $\varepsilon$ , one always has  $\Phi(S) \geq 1 - 1/q - O(\varepsilon)$ . Thus, in our case we are interested in studying the structure of sets  $S$  that nearly attain this minimum.

The two very natural analogs of small poorly expanding sets are the analogs of zoom-in and zoom-out sets from the non-affine version of the Grassmann graph, and are defined as follows. For a vector  $z \in W$

and an affine hyperplane  $W' \subseteq W$ , and zoom-in with respect to  $z$  and the zoom-out with respect to  $W'$  sets are defined as

$$H_z = \{A \in V(W, \ell) \mid z \in A\}, \quad H_{W'} = \{A \in V(W, \ell) \mid A \subseteq W'\}.$$

It can be shown without much difficulty that  $H_z$  and  $H_{W'}$  have small fractional size, and that  $1 - \Phi(H_z) \geq \frac{1}{q}$ ,  $1 - \Phi(H_{W'}) \geq \frac{1}{q}$ . These are the natural analogs of sets that were shown in [13] to capture, in some sense, the structure of all small non-expanding sets in the Grassmann graph. However, in the affine version of the Grassmann graph there are more examples.

Given  $z \in W \setminus \{0\}$  and a hyperplane  $W' \subseteq W$ , one may consider *the zoom in and zoom-out with respect to the linear part*. To define these, first let us note that given an affine subspace  $A \in V(W, \ell)$ , one may write  $A = x + A'$  where  $A' \subseteq W$  is a linear space, and  $x \in A$  is some vector (we note that  $A'$  is unique but  $x$  is not). Thus, we may define

$$H_{z,\text{lin}} = \left\{ A \in V(W, \ell) \mid A = x + \tilde{A} \text{ for some } x \in W \text{ and a linear space } \tilde{A}, \text{ and } z \in \tilde{A} \right\},$$

$$H_{W',\text{lin}} = \left\{ A \in V(W, \ell) \mid A = x + \tilde{A} \text{ for some } x \in W \text{ and a linear space } \tilde{A}, \text{ and } \tilde{A} \subseteq W' \right\}.$$

It is easy to see that these sets are also small and have expansion roughly  $1 - \frac{1}{q}$ , and furthermore that they are “genuinely” new examples (i.e., they are linear combinations of the basic zoom-in/zoom-out sets). We show that in a sense, these examples entirely capture the structure of sets  $S$  with  $1 - \Phi(S) \geq \frac{1}{q}$ .

To be more precise, we say a set  $S$  is  $\xi$ -pseudo-random with respect to zoom-ins/ zoom-outs/ zoom-ins on the linear part/ zoom outs of the linear part – say zoom-ins for concreteness – if  $\mu(S \cap H_z) \leq \xi \mu(H_z)$  for all  $z \in W$  (see Definition 2.3 for a more formal definition). In this language, our main expansion result, Theorem 2.4, asserts that if  $S$  is  $\xi$ -pseudo-random with respect to zoom-outs (standard and on the linear part), and with respect to zoom-ins on the linear part, and  $1 - \Phi(S) \geq \frac{1}{q}$ , then  $S$  is highly non-pseudo-random with respect to zoom-ins. Namely, there is  $z$  such that  $\mu(S \cap H_z) \geq (1 - o(1))\mu(H_z)$ , or in words  $S$  almost contains  $H_z$ ; see Theorem 2.4 for a precise statement. Here and throughout,  $\mu$  represents the uniform measure over  $V(W, \ell)$ .

With our testing question in mind, this sort of structure appears natural as it suggests that we may want to change the value of  $f$  on  $z$ .

**Remark 1.7.** *A few remarks are in order:*

1. *Our expansion result here is tailored for our application, however our technique can be used to establish weaker structural for a small set  $S$  so long as  $1 - \Phi(S) \geq \frac{1}{q^2} + \delta$ .*
2. *The diligent reader may notice that in the statement above, there is an asymmetric role to each one of the zoom-sets. This is a by-product again of the application we have in mind as we can show that for the set  $S$  of erroneous subspaces, these pseudo-randomness conditions hold. In more generality though, it may be proved that if a set  $S$  is very pseudo-random with respect to 3 of the zoom notions (i.e.  $\xi$ -pseudo-random where  $\xi$  is small), then it is very not pseudo-random with respect to the last notion of zoom (i.e. almost containing a copy of such set).*
3. *It would be interesting to prove expansion theorems in the affine Grassmann graph in greater generality similarly to the way it was done in [13]. Namely, proving that if  $1 - \Phi(S) \geq \frac{1}{q^r} + \delta$ , then  $S$  cannot be  $\varepsilon = \varepsilon(\delta, r) > 0$  pseudo-random with respect to  $r$ -wise intersections of zoom-sets. That is, there must be copies of zoom-sets  $H^1, \dots, H^r$  that intersect non-trivially such that  $\mu(S \cap \bigcap_{i=1}^r H_i) \geq \xi \mu(\bigcap_{i=1}^r H_i)$ .*

## 1.4 Our techniques

Our expansion theorem is proved using Fourier analysis similarly to [13] and is deferred to the appendix. Next, we explain how it is used in order to prove Theorem 1.4.

The proof has two components. We consider the collection of erroneous subspaces, i.e.

$$S = \{ A \in \text{AffGras}(\mathbb{F}_q^n, t) \mid f|_A \text{ is not of degree } d \}.$$

First, as explained earlier we observe that  $\mu(S \uparrow) \leq q\mu(S)$ , and deduce that  $1 - \Phi(S) \geq \frac{1}{q}$ . We then wish to apply our expansion theorem, and towards this end we show that  $S$  is pseudo-random with respect to zoom-outs (both standard and with respect to the linear part), as well as on zoom-ins with respect to the linear part. Proving pseudo-randomness with respect to zoom-outs is fairly easy as these sets enlarge considerably when taking an upper shadow. The proof that  $S$  is pseudo-random with respect to zoom-ins on the linear part is more tricky. In a sense, the idea is that given a  $(t+1)$ -flat  $A = x + \tilde{A}$  such that  $z \in \tilde{A}$  but otherwise  $A$  is random, we may find  $t$ -flats  $B_1, \dots, B_{t+1} \subseteq A$  such that marginally each one of them is distributed uniformly, and together they cover  $A$  entirely. Thus, as errors do not really accumulate on these  $B$ 's, they cannot be concentrated on  $A$ 's of this type. The formal proof proceeds a bit differently and makes use again of our expansion theorem in a lower-order affine Grassmann graph; see Claim 3.4 for details.

We then deduce, using our expansion theorem, that  $S$  nearly contains a copy of  $H_x$  for some  $x \in \mathbb{F}_q^n$ . In words, the test almost always fails if it is being conducted on a subspace  $A$  that contains the point  $x$ . This suggests that  $x$  is a point in which we should change the value of  $f$  in order to get closer to a degree  $d$  polynomial, and we indeed argue this way.

This is the correction step of the argument. The simplest case is  $q = 2$ , which is instructive to consider. Indeed, in this case we have that  $t = d + 1$ , and we argue that if we flip the value of the point  $x$ , the rejection probability of the tests drops additively by  $\Theta(2^{d-n})$ . The point here is that if  $g$  is a polynomial of degree  $d + 1$  on a subspace of dimension  $d + 1$  over  $\mathbb{F}_2$ , then flipping any single value of  $g$  results in a polynomial of degree at most  $d$ . Iterating this argument shows that after we change the values of  $f$  on at most  $O(2^{n-d}\varepsilon)$  points, the rejection probability drops to 0, at which point our function must be a degree  $d$  polynomial.

In the more general case of  $\mathbb{F}_q$ , the correction step is not as simple and requires more work. Here, given such point  $x$ , we consider a random affine subspace  $A$  of dimension  $t + 100$  containing  $x$ . We now focus on affine subspaces  $B \subseteq A$  of dimension  $t$ , and note that expectedly over the random choice of  $A$ :

1. the fraction of such  $B$ 's containing  $x$  on which  $f|_B$  is degree  $d$  is  $O(\varepsilon)$ ;
2. the fraction of such  $B$ 's not containing  $x$ , on which  $f|_B$  is degree  $d$ , is  $1 - O(\varepsilon)$ .

By Markov's inequality, we have that with probability at least 0.99 both of these events hold simultaneously. We fix such  $A$ , and next claim that provided  $\varepsilon$  is sufficiently small (depending only on  $q$ ), we can change the value of  $f(x)$  in some way so that the fraction of  $B$ 's containing  $x$  as in the first item above, would be at least  $1/(2q)$  (thus lowering the rejection probability of the test on such subspaces). We establish that via two steps:

**Bootstrapping errors on  $B \not\ni x$ .** We show that provided that  $\varepsilon$  is small, having chosen  $A$  as above, if  $f|_B$  is degree  $d$  for at least  $1 - O(\varepsilon)$  fraction of the  $t$ -flats  $B \subseteq A$  not containing  $x$ , then the test must pass in fact on all of these  $t$ -flats. Intuitively, the idea here is to consider the random walk on  $B$ 's that moves from  $B$  of dimension  $t$  to  $B'$  of dimension  $t + 1$  that doesn't contain  $x$ , and then back to  $B'' \subseteq B'$  of dimension  $t$  and show that, as before, due to expansion considerations, the errors must be very structured as zoom ins. However, as  $\varepsilon$  is very small, zoom-ins are too large and hence the set of errors must be empty.

The precise execution of this step is done differently, as we do not really wish to study this random walk operation as described above; instead, we look at intermediate  $(t + 50)$ -flats  $C \subseteq A$  that do not contain  $x$ , and perform the standard random walk on  $\text{AffGras}(C, t)$ .

**Correcting errors on  $B \ni x$ .** Having established the previous step, we look at  $A' \subseteq A$  of dimension  $t + 1$  that contains  $x$ , and note that all erroneous affine subspaces  $B \subseteq A'$  must contain  $x$ . As these constitute only  $1/q$  fraction of the subspaces contained in  $A'$ , due to Lemma 2.1 they must all be erroneous. Next, we show it is possible to change  $f(x)$  and make at least one of these  $B$ 's pass the test, which then by the second bullet in Lemma 2.1 guarantees that  $f|_{A'}$  must be degree  $d$ . Indeed,  $f|_{A'}$  has degree at most  $(t + 2)(q - 1)$ , and we can add to it a multiple of  $g(z) = 1_{z=x}$  to eliminate its highest degree monomial (as there is only one such monomial), say it is  $f + g$ . The function  $(f + g)|_{A'}$  then has degree strictly smaller than  $(t + 2)(q - 1)$ , hence by Lemma 2.1 unless  $(f + g)|_{A'}$  is degree at most  $d$ , it must be the case that more than  $1/q$  fraction of the  $B \subseteq A'$  of dimension  $t$  fail the test. However, these can only still be subspaces containing  $x$ , which are at most  $1/q$  fraction. Thus,  $(f + g)|_{A'}$  has degree  $d$ , so that we showed that we may change  $f(x)$  and make  $f|_{A'}$  degree  $d$ .

## 2 Preliminaries

### 2.1 Relating different testers

In this section, we provide several basic facts that will be used throughout the proof. Below, the first bullet is [7][Lemma 4.6], and the second bullet is a slight refinement which elaborates on when a given function  $f$  may be tight for the first bullet. Since the proof of the refinement is a small tweak on the original proof from [7], we fully record it here.

**Lemma 2.1.** *Let  $p$  be prime,  $q \in \mathbb{N}$  be a power of  $p$  and  $d \in \mathbb{N}$  and set  $t = \lceil \frac{d+1}{q-q/p} \rceil$ . Suppose that  $k \geq t$ , and let  $f: \mathbb{F}_q^{k+1} \rightarrow \mathbb{F}_q$ . Then*

1. *If  $\deg(f) > d$ , then  $\varepsilon_{k,d}(f) \geq \frac{1}{q}$ .*
2. *If  $d < \deg(f) < (k + 1)(q - 1)$ , then  $\varepsilon_{k,d}(f) > \frac{1}{q}$ .*

*Proof.* Let  $f(x)$  has degree strictly larger than  $d$ . We shall think about restrictions to  $k$ -flats as taking a non-constant linear  $L: \mathbb{F}_q^{k+1} \rightarrow \mathbb{F}_q$ , and then considering  $f|_{L=0}$ . We shall use the notion of canonical monomials from [7], which in our context reads: a monomial  $M(x) = \prod_{j \leq m} x_j^{e_j}$  is canonical if it appears in  $f$ ,  $q - q/p \leq e_1, \dots, e_{m-1} \leq q - 1$  and  $e_m \leq q - 1$ . From [7][Lemma 4.3] we may compose  $f$  with an invertible affine linear transformation, and get to assume that  $f$  has max-monomial of degree  $\text{ddeg}(f)$  which is canonical; clearly, once we prove the statement for this composition, the lemma immediately follows for the original function. We henceforth assume without loss of generality that this transformation is the identity.

Let  $M$  be a canonical max-monomial of  $f$ , and write  $M(x) = \prod_{j \leq m} x_j^{e_j}$  and  $m \leq k + 1$ . We consider two cases:

- **Case 1:**  $m \leq k$ . In this case, we note that any linear transform  $L$  that does not depend on the variables  $x_1, \dots, x_m$  preserves the degree of  $f$ , i.e.  $\deg(f|_{L=0}) = \deg(f)$ .

For any other linear transformation  $L$ , it must depend on one of the variables  $x_1, \dots, x_m$ , say without loss of generality it depends on  $x_1$ , and say  $L(x) = a_1 x_1 + L'(x_2, \dots, x_k) + a_{k+1} x_{k+1}$  for  $a_1 \neq 0$ .

Denote  $L_z(x) = a_1x_1 + L'(x_2, \dots, x_k) + zx_{k+1}$ , so that  $L(x) = L_{a_{k+1}}(x)$ . In this case, we may think of  $f|_{L_z=0}$  as  $f(-a_1^{-1}(L'(x_2, \dots, x_{k+1}) + zx_{k+1}), x_2, \dots, x_{k+1})$ , and we show that there is  $z \in \mathbb{F}_q$  such that the degree of  $f|_{L_z=0}$  is greater than  $d$ . Thus, we conclude in this case that if  $a_{k+1}$  was already this  $z$  the degree of  $f|_{L=0}$  would have been higher than  $d$ , and so  $L$ 's that depend on the variables  $x_1, \dots, x_m$  we have that  $\deg(f|_{L=0}) > d$  with probability at least  $1/q$ . Together with the previous paragraph this establishes both items of the lemma in this case, and we next show the existence of this  $z$ .

The idea is to look at  $f(-a_1^{-1}(L'(x_2, \dots, x_{k+1}) + zx_{k+1}), x_2, \dots, x_{k+1})$ , and more specifically at the coefficient of the monomial  $M'(x) = \prod_{1 < j \leq m} x_j^{e_j} \cdot x_{k+1}^{e_1}$ . The max-monomial  $M$  from  $f$  would give

us this monomial with coefficient  $-a_1^{-1}z^{e_1}$ , and since  $M$  was max-monomial any other monomials will be able to contribute only  $z$ 's with lower power. Hence, the coefficient of  $M'$  is some non-zero polynomial in  $z$  of degree at most  $e_1 \leq q - 1$ , and hence we may choose  $z$  for which it is non-zero.

• **Case 2:**  $m = k + 1$ . Suppose  $e_1 \geq e_2 \geq \dots \geq e_m$ . Here we consider two subcases.

– First, consider the case that  $e_1 + \dots + e_m < (k + 1)(q - 1)$  (which is the only case we need for the second bullet in the lemma), so that  $e_m < q - 1$ . Choose a non-constant linear transformation

$L(x_1, \dots, x_{k+1}) = \sum_{i=1}^{k+1} a_i x_i + c$  randomly, and note that the probability that  $a_{k+1}$  is 0 is strictly smaller than  $1/q$  (indeed, the distribution of  $(a_1, \dots, a_{k+1})$  is over non-zero vectors). We shall

focus on  $L$ 's such that  $a_{k+1} \neq 0$ . Let  $L_z(x_1, \dots, x_{k+1}) = \sum_{i=1}^{k+1} a_i x_i + z$ , and we argue that for each  $L$ , there are at least 2 values of  $z$  for which  $f|_{L_z=0}$  has degree greater than  $d$ . Indeed,

the argument is exactly the same as before, except that we look at the monomial  $M' = \prod_{i=1}^k x_i$ ,

and note that from  $M$  we have a contribution  $a_{k+1}^{-1}z^{e_{k+1}}$ , and as  $M$  is a max-monomial all other contributions are lower degree in  $z$ . Hence, choosing  $z$  at random the probability this coefficient is non-zero is at least  $\frac{q - e_{k+1}}{q} \geq \frac{2}{q}$ , and in this case the degree of  $f|_{L_z=0}$  is at least  $e_1 + \dots + e_k \geq k(q - q/p) \geq d + 1$ .

Thus, we get that the probability that  $f|_{L=0}$  has degree greater than  $d$  is

$$> \frac{q - 1}{q} \cdot \frac{2}{q} \geq \frac{1}{q},$$

where the first factor comes from the event  $a_1 \neq 0$ , and the second factor comes from the event that  $a_{k+1}$  is one of the two  $z$ 's which keeps the degree of  $f|_{L_z=0}$  high.

– Next, consider the case that  $e_1 + \dots + e_m \geq (k + 1)(q - 1)$ . This case is similar to case 1.

A non-constant affine transformation  $L$  either has the form  $L(x) = \sum_{i=1}^{k+1} a_i x_i + c$ , and letting

$L_z(x) = \sum_{i=1}^{k+1} a_i x_i + z$ , we show that there is  $z \in \mathbb{F}_q$  such that  $\deg(f|_{L_z=0}) \geq d + 1$ . Indeed,

suppose without loss of generality that  $a_{k+1} \neq 0$ , then the coefficient of the monomial  $\sum_{i=1}^k a_i x_i$

in  $f|_{L_z=0}$  is a non-zero polynomial in  $z$  of degree at most  $q - 1$ , hence there is  $z$  for which this coefficient is non-zero, and hence  $f|_{L_z=0}$  has degree  $k(q - 1) \geq d + 1$ .



As  $c = z$ , which happens with probability  $1/q$ , we have that the degree of  $f|_L$  is strictly larger than  $d$ .  $\square$

**Lemma 2.2.** *Let  $p$  be a prime,  $q$  be a power of  $p$ ,  $d \in \mathbb{N}$  and let  $t = \lceil \frac{d+1}{q-q/p} \rceil$ . Suppose that  $k \geq t$ , then for all  $f: \mathbb{F}_q^n \rightarrow \mathbb{F}_q$  we have  $\varepsilon_{k+1,d}(f) \leq q\varepsilon_{k,d}(f)$ .*

*Proof.* Let  $B \subseteq \mathbb{F}_q^n$  be a uniform  $k+1$  dimensional flat, and let  $A \subseteq B$  be a uniform  $k$ -dimensional flat. Then

$$\varepsilon_{k,d}(f) = \Pr_{A,B} [\deg(f|_A) > d] = \Pr_{A,B} [\deg(f|_B) > d] \Pr_{A,B} [\deg(f|_A) > d \mid \deg(f|_B) > d].$$

The first probability on the right hand side is  $\varepsilon_{k+1,d}(f)$ , and the second probability is at least  $1/q$  by Lemma 2.1.  $\square$

## 2.2 Expansion and pseudo-randomness

Denote by  $V_q(k, \ell)$  the set of dimension  $\ell$  affine subspaces in  $\mathbb{F}_q^k$ ; we often omit the subscript  $q$  when it is clear from context. In this section, we discuss expansion in the affine Grassmann graph over  $V_q(k, \ell)$ . Similarly to the works [3, 13], we too consider certain structures of sets that forbid strong expansion properties, but in our case there are additional types of structures (due to the fact we are working in the affine case).

**Definition 2.3.** *Let  $S \subseteq V_q(k, \ell)$ , and let  $\xi \in [0, 1]$ .*

1. *We say  $S$  is  $\xi$ -pseudo-random with respect to hyperplanes, if for each affine hyperplane  $W$  we have that*

$$\mu(S_W) \stackrel{\text{def}}{=} \Pr_{A \in V(W, \ell)} [A \in S] \leq \xi.$$

2. *We say  $S$  is  $\xi$ -pseudo-random with respect to hyperplanes on its linear part, if for each hyperplane  $W$  we have that*

$$\mu(S_{W, \text{lin}}) \stackrel{\text{def}}{=} \Pr_{A \in V(W, \ell), x \notin W} [x + A \in S] \leq \xi.$$

3. *We say  $S$  is  $\xi$ -pseudo-random with respect to points on its linear part if for each point  $y$ , we have that*

$$\mu(S_{x, \text{lin}}) \stackrel{\text{def}}{=} \Pr_{A=x+A' \in \text{AffGrass}(k, \ell)} [A \in S \mid A' \ni y] \leq \xi.$$

4. *We say  $S$  is  $\xi$ -pseudo-random with respect to points, if for each point  $x$  we have that*

$$\mu(S_x) \stackrel{\text{def}}{=} \Pr_{A \in \text{AffGrass}(k, \ell)} [A \in S \mid x \in A] \leq \xi.$$

In [3, 13] it is proved that pseudo-random sets have strong expansion properties in the Grassmann graph. Here, we require a similar statement. Consider  $W$  a  $k$ -dimensional affine space over  $\mathbb{F}_p$ , and consider the random walk on  $\text{AffGras}(W, \ell)$  as described in the introduction.

**Theorem 2.4.** *Let  $\xi > 0$  and let  $q \in \mathbb{N}$  be a prime power. Suppose that  $S \subseteq V_q(W, \ell)$  is a set with:*

1.  $\mu(S) \leq \xi$ ;

2.  $S$  is  $\xi$ -pseudo-random with respect to hyperplanes, with respect to hyperplanes on its linear part, as well as with respect to points on its linear part;
3.  $1 - \Phi(S) \geq \frac{1}{q}$ .

Then there exists a point  $x \in \mathbb{F}_q^n$  such that  $\mu(S_x) \geq 1 - q^2(867\xi^{1/4} + q^{-\ell})$

The proof of Theorem 2.4 proceeds similarly to the proof presented in [3] for the degree 1 case, however we use some of the machinery of [13] to simplify the presentation. The proof is deferred to the appendix.

### 2.3 Expansion and sharp thresholds

**Definition 2.5.** For  $h \geq \ell$  and  $S \subseteq V(k, \ell)$ , we define

$$S \uparrow^h = \{L \mid \dim(L) = h, \exists K \in S, K \subseteq L\}.$$

When  $h = \ell + 1$ , we omit the superscript  $h$ .

The following lemma is very similar to [11, Proposition III.3.4.].

**Lemma 2.6.**  $\mu(S \uparrow) \geq \frac{\mu(S)}{1 - \Phi(S)}$ .

*Proof.* For  $B \in V(k, \ell)$ , denote by  $T \uparrow B$  the uniform distribution over subspaces  $B' \in V(k, \ell + 1)$  containing  $B$ , and for  $B' \in V(k, \ell + 1)$  denote by  $T \downarrow B'$  the uniform distribution over  $B \in V(k, \ell)$  contained in  $B'$ . We consider real-valued functions over  $V(k, \ell)$ ,  $V(k, \ell + 1)$  and view  $T \uparrow, T \downarrow$  as operators,  $T \downarrow: L_2(V(k, \ell)) \rightarrow L_2(V(k, \ell + 1))$ ,  $T \uparrow: L_2(V(k, \ell + 1)) \rightarrow L_2(V(k, \ell))$  defined as

$$T \downarrow f(B') = \mathbb{E}_{B \sim T \downarrow B'} [f(B)], \quad T \uparrow g(B) = \mathbb{E}_{B' \sim T \uparrow B} [g(B')],$$

for  $f: V(k, \ell) \rightarrow \mathbb{R}$ ,  $g: V(k, \ell + 1) \rightarrow \mathbb{R}$ . We note that  $T \downarrow$  is the adjoint of  $T \uparrow$ .

Fix  $S$ , and let  $f = 1_S, g = 1_{S \uparrow}$ . Then

$$\mu(S) = \mathbb{E}_{B \in V(k, \ell)} [f(B)] = \mathbb{E}_{B' \in V(k, \ell + 1)} \left[ \mathbb{E}_{B \sim T \downarrow B'} [f(B)] \right] = \mathbb{E}_{B' \in V(k, \ell + 1)} [g(B') T \downarrow f(B')] = \langle g, T \downarrow f \rangle.$$

Thus, using Cauchy-Schwarz

$$\mu(S)^2 \leq \|g\|_2^2 \|T \downarrow f\|_2^2 = \mu(S \uparrow) \langle T \downarrow f, T \downarrow f \rangle = \mu(S \uparrow) \langle f, T \uparrow T \downarrow f \rangle.$$

Thus,

$$\mu(S \uparrow) \geq \frac{\mu(S)}{\frac{1}{\mu(S)} \langle f, T \uparrow T \downarrow f \rangle}.$$

We note that for  $B$ , the distribution of  $\tilde{B} \sim T \uparrow T \downarrow B$  is distributed according to the random walk of  $\text{AffGras}(\mathbb{F}_q^k, \ell)$ , hence  $\frac{1}{\mu(S)} \langle f, T \uparrow T \downarrow f \rangle = 1 - \Phi(S)$ , finishing the proof.  $\square$

## 3 Testing Reed-Muller codes: proof of Theorem 1.4

In this section, we present the formal proof of Theorem 1.4. For a high level description of our proof strategy we defer the reader to Section 1.4.

### 3.1 Step 1: locating a potential error

Fix  $f$  as in the statement of the theorem, and denote

$$S = \{A \mid \dim(A) = t, \deg(f|_A) > d\}.$$

We note that  $\varepsilon_{t,d}(f) = \mu(S) \stackrel{\text{def}}{=} \varepsilon$ , and that  $\varepsilon_{t+1,d}(f) = \mu(S \uparrow)$ . Throughout, we will assume that  $\varepsilon \leq q^{-M}$  for a sufficiently large (but absolute) constant  $M$ . We will also assume that  $t \geq M$ , otherwise the result follows from [9].

Let  $\varepsilon' = \max(\varepsilon, q^{-d})$ . Our aim in this section is to prove the following proposition.

**Proposition 3.1.** *There exists  $x^* \in \mathbb{F}_q^n$  such that  $\mu(S_{x^*}) \geq 1 - C(q)\varepsilon'^{1/4}$  for  $C(q) = 2000q^2$ .*

We will prove this proposition using Theorem 2.4, and towards this end we first show that the conditions of Theorem 2.4 hold.

**Claim 3.2.**  $1 - \Phi(S) \geq \frac{1}{q}$ .

*Proof.* By Lemma 2.2 we have  $\mu(S \uparrow) \leq q\mu(S)$  and by Lemma 2.6 we have  $\mu(S \uparrow) \geq \frac{\mu(S)}{1 - \Phi(S)}$ . Combining these two inequalities gives the statement of the claim.  $\square$

To apply Theorem 2.4, we first argue that  $S$  is pseudo-random.

**Claim 3.3.** *The set  $S$  is  $2q\mu(S)$  pseudo-random with respect to zoom-outs, also with respect to the linear part.*

*Proof.* Let  $W \subseteq \mathbb{F}_q^n$  be a hyperplane (either affine or not), and sample a  $(t+1)$ -flat uniformly  $A \subseteq \mathbb{F}_q^n$ . We note that the probability that  $W \cap A$  has dimension  $t$  is  $1 - q^{-(t+1)}$ . To see that, we may think of  $W$  as being defined by an equation  $\langle x, h \rangle = c$  for some non-zero vector  $h$  and  $c \in \mathbb{F}_q$ , and  $A$  as being defined by a collection of linearly independent equations  $\langle x, h_i \rangle = c_i$  for  $i = 1, \dots, n - t - 1$ . Whenever  $h \notin \text{span}(h_1, \dots, h_{n-t-1})$ ,  $A \cap W$  has dimension  $t$ . Conditioned on this event,  $A \cap W$  is a uniform  $t$ -flat in  $W$ , and so  $A \cap W \in S$  with probability  $\mu(S_W)$ . Also, if  $A \cap W \in S$  then  $A \in S \uparrow$ , so we get that

$$\mu(S \uparrow) = \Pr[A \in S \uparrow] \geq (1 - q^{-(t+1)})\mu(S_W) \geq \frac{\mu(S_W)}{2}.$$

Thus,  $\mu(S_W) \leq 2\mu(S \uparrow) \leq 2q\mu(S)$ .  $\square$

Next, we prove that  $S$  is pseudo-random with respect to zoom-ins on its linear part. This argument is more involved, and requires a bootstrapping-style argument as described in the proof overview; namely, we show that if there are very little errors on specific type of subspaces, then there must be no errors at all on these type of subspaces.

**Claim 3.4.** *The set  $S$  is  $q^{200-M}$  pseudo-random with respect to zoom-in on its linear part.*

*Proof.* Suppose otherwise, then there is  $z \in \mathbb{F}_q^n$  such that

$$\{x + A \in S \mid A \subseteq \mathbb{F}_q^n \text{ linear subspace of dimension } t, z \in A\},$$

has fractional size  $\alpha > q^{200-M}$  inside  $P_{z,t} = \{x + A \mid A \subseteq \mathbb{F}_q^n \text{ linear subspace of dimension } t, z \in A\}$ . Clearly,

$$S'_z = \{x + A \in S \uparrow^{t+100} \mid A \subseteq \mathbb{F}_q^n \text{ linear subspace of dimension } t + 100, z \in A\}$$

also has at least  $\alpha$  fractional size inside  $P_{z,t+100}$ . Take  $x + A$  a  $(t + 100)$ -flat uniformly, consider the event it is in  $S'_z$ , and take a  $t$ -flat  $B = x' + A' \subseteq x + A$  uniformly; note that

$$\begin{aligned} \Pr_{\substack{x+A \\ B=x'+A' \subseteq x+A}} [\deg(f|_B) > d \mid x + A \in S'_z, z \notin A'] &\leq \frac{\Pr_{x+A, B=x'+A'} [\deg(f|_B) > d \mid z \in A, z \notin A']}{\Pr [x + A \in S'_z \mid z \in A, z \notin A']} \\ &\leq \frac{\Pr_{B=x'+A'} [\deg(f|_B) > d \mid z \notin A']}{\Pr [x + A \in S'_z \mid z \in A] \Pr [z \notin A' \mid x + A \in S'_z]}. \end{aligned}$$

The numerator is at most  $\varepsilon$ , and the denominator is at least  $\alpha/2$ , so we get this probability is at most  $\frac{2}{\alpha}\varepsilon$ . Thus, there exists  $x + A \in S'_z$  such that conditioned on it this probability is at most  $\frac{2}{\alpha}\varepsilon$ , and we fix it henceforth.

We now work over the affine  $t$ -dimensional Grassmann graph over  $x + A$ . Consider  $t$ -flats  $B = x' + A' \subseteq x + A$  conditioned on  $A'$  not containing  $z$ , and let

$$\mathcal{B} = \{B = x' + A' \in \text{AffGras}(x + A, t) \mid z \notin A', \deg(f|_B) > d\}.$$

We argue that  $\mathcal{B}$  must be empty; suppose towards contradiction otherwise. Let  $W = y + \tilde{W} \subseteq x + A$  be randomly chosen where  $\tilde{W}$  is uniformly chosen linear subspace of dimension  $t + 40$  not containing  $z$ , and  $y \in x + A$  is uniformly chosen. Denote

$$\mathcal{B}_W = \{B \in \mathcal{B} \mid B \subseteq W\}.$$

We denote by  $\mu_W$  the uniform measure over  $\text{AffGras}(W, t)$ . We argue that for all  $W$ , if  $\mu_W(\mathcal{B}_W) \leq q^{-100}$ , then  $\mu_W(\mathcal{B}_W) = 0$ . Indeed, if  $\mu_W(\mathcal{B}_W) \leq q^{-100}$  then  $\mathcal{B}_W$  is  $q^{-60}$  pseudo-random with respect to zoom-ins (also with respect to its linear part), as those have measure at least  $q^{-40}$ . Also, by an argument as in Claim 3.3 we have that  $\mathcal{B}_W$  is  $q^{-50}$  pseudo-random with respect to zoom out (also with respect to its linear part). Finally, by an argument as in Claim 3.2 we have

$$1 - \Phi_W(\mathcal{B}_W) \geq \frac{1}{q},$$

where  $\Phi_W$  is expansion with respect to  $\text{AffGras}(W, t)$ . This is now a contradiction to Theorem 2.4. We thus conclude that either  $\mu_W(\mathcal{B}_W) = 0$  or  $\mu_W(\mathcal{B}_W) \geq q^{-100}$ ; as  $\mathcal{B}$  is non-empty (by our assumption), we may find  $W$  such that  $\mu_W(\mathcal{B}_W) \geq q^{-100}$ , and we fix such one.

Next, we take a uniform  $Y = u + \tilde{Y} \subseteq x + A$  of dimension  $t + 99$  conditioned on  $z \notin \tilde{Y}$ , sample a  $(t + 60)$ -flat  $A_2 \subseteq Y$ , and consider  $A_2 \cap W$ . We may think of  $W$  as being defined by a system of 60 independent linear equations  $\langle h_1, x \rangle = c_1, \dots, \langle h_{60}, x \rangle = c_{60}$  over  $x + A$ , and  $A_2$  as being defined by a set of 39 linear equations  $\langle h'_1, x \rangle = c'_1, \dots, \langle h'_{39}, x \rangle = c'_{39}$  where  $h_1, \dots, h'_{39}$  are random linearly independent. Note that the probability that  $\text{span}(h_1, \dots, h_{60}, h'_1, \dots, h'_{39})$  has dimension 99 is at least

$$\prod_{j=0}^{38} \frac{q^{99} - q^{60+j}}{q^{99}} \geq e^{-2 \sum_{j=1}^{\infty} q^{-j}} \geq e^{-4/q},$$

in which case the distribution of  $A_2 \cap W$  is uniform from  $\text{AffGras}(W, t)$ . Hence,  $A_2 \cap W$  is in  $\mathcal{B}_W$  with probability at least  $e^{-4/q} q^{-100}$ . In this case, we have that  $A_2 \in \mathcal{B}_Y \uparrow^{t+60}$  where upper shadow is taken with respect to  $\text{AffGras}(Y, t)$ . Thus, we get that

$$\mathbb{E}_Y [\mu_Y(\mathcal{B}_Y \uparrow^{t+60})] \geq e^{-4/q} \mu_W(\mathcal{B}_W) \geq e^{-4/q} q^{-100}.$$

However, by Lemma 2.2 for each  $Y$  we have that

$$\mu_Y(\mathcal{B}_Y \uparrow^{t+60}) \leq q^{60} \mu_Y(\mathcal{B}_Y),$$

and plugging that in above we get that

$$\mathbb{E}_Y [\mu_Y(\mathcal{B}_Y)] \geq e^{-4/q} q^{-160}.$$

Finally, the left hand side is at most the probability that  $f|_{x'+A'}$  has degree  $> d$  when  $x' + A' \subseteq x + A$  is a random  $t$ -flat conditioned on  $A \not\ni z$ , hence at most  $\frac{2}{\alpha}\varepsilon$  by the choice of  $x + A$ . Overall, we get that

$$\alpha \leq e^{4/q} q^{160} 2\varepsilon \leq q^{200-M},$$

and contradiction. This contradiction implies that  $\mathcal{B}$  is empty, and we quickly finish the argument now.

Let us look at  $x + A$ ; as  $f|_{x+A}$  has degree larger than  $d$ , we may find a  $(t+1)$ -flat  $B = x' + \tilde{B} \subseteq x + A$  such that  $f|_B$  has degree larger than  $d$ . Sample a  $t$ -flat  $x'' + B' \subseteq B$  uniformly. By the above, if  $z \notin B'$ , we have that  $f|_{x''+B'}$  has degree  $d$ . Note that the probability that  $z \in B'$  is at most

$$\frac{q^t - 1}{q^{t+1} - 1} < \frac{1}{q},$$

so we get that for less than  $1/q$  fraction of the  $t$ -flats  $x'' + B' \subseteq B$  we have that  $\deg(f|_{x''+B'}) > d$ . This contradicts Lemma 2.2.  $\square$

We can now prove Proposition 3.1.

*Proof of Proposition 3.1.* From Claims 3.2, 3.3, 3.4 we have that the conditions of Theorem 2.4 hold, and hence we may find  $x^* \in \mathbb{F}_p^n$  such that  $\mu(S_{x^*}) \geq 1 - C(q)\varepsilon^{1/4}$ , for  $C(q) = 2000q^2$ .  $\square$

### 3.2 Step 2: correcting the value on $x^*$

The goal of this section is to prove the following proposition.

**Proposition 3.5.** *There exists  $c \in \mathbb{F}_q$  such that changing the value of  $f(x^*)$  to  $c$ , we have that*

$$\Pr_{A'' \text{ } t\text{-flat}} [\deg(f|_{A''}) \leq d \mid x^* \in A''] \geq \frac{1}{2q}.$$

The rest of this section is devoted to proving Proposition 3.5. Take a uniform  $(t+100)$ -flat  $A$  containing  $x^*$ , and let

$$\mathcal{B}_A = \{B \in \text{AffGras}(A, t) \mid x^* \notin B, \deg(f|_B) > d\},$$

then  $\mathbb{E}_A [\mu_A(\mathcal{B}_A)] \leq O(\varepsilon)$ , so with probability at least  $1/2$  over  $A$  we have that  $\mu_A(\mathcal{B}_A) \leq O(\varepsilon)$ .

Take a  $(t+40)$  flat  $W \subseteq A$  randomly not containing  $x^*$ , and let

$$\mathcal{B}_W = \{B \in \text{AffGras}(W, t) \mid B \in \mathcal{B}_A\}.$$

We argue that for each  $W$ , either  $\mu_W(\mathcal{B}_W) = 0$  or  $\mu_W(\mathcal{B}_W) \geq q^{-100}$ . Otherwise,  $0 < \mu_W(\mathcal{B}_W) < q^{-100}$ . Therefore,  $\mathcal{B}_W$  is  $q^{-60}$  pseudo-random with respect to zoom ins (also with respect to their linear part), and from an argument as in Claim 3.3 we have that  $\mathcal{B}_W$  is  $q^{-98}$  pseudo-random with respect to zoom-outs (as well as their linear parts). Finally, as in the argument in Claim 3.2 we have  $1 - \Phi_W(\mathcal{B}_W) \geq 1/q$ , so we get a contradiction to Theorem 2.4.

**Claim 3.6.**  $\mathcal{B}_A = \emptyset$ .

*Proof.* Otherwise, we may find  $W$  such that  $\mu_W(\mathcal{B}_W) \geq q^{-100}$ . The argument is similar to the end of the argument in Claim 3.4. Take a  $(t + 99)$  flat  $Y \subseteq A$  randomly not containing  $x^*$ , and take a  $(t + 60)$ -flat  $A_2 \subseteq Y$  randomly. Then  $A_2 \cap W$  has dimension  $t$  with probability at least  $e^{-4/q}$ , and then its distribution is uniform in  $\text{AffGras}(W, t)$ . Thus, it is in  $\mathcal{B}_W$  with probability at least  $q^{-100}$ . Therefore, we get that

$$\mathbb{E}_Y [\mu_Y(\mathcal{B}_Y \uparrow^{t+60})] \geq \Pr_{Y, A_2} [A_2 \cap W \in \mathcal{B}_W] \geq e^{-4/q} \cdot q^{-100}.$$

On the other hand, by Lemma 2.2

$$\mathbb{E}_Y [\mu_Y(\mathcal{B}_Y \uparrow^{t+60})] \leq q^{60} \mathbb{E}_Y [\mu_Y(\mathcal{B}_Y)] \leq q^{60} 2\mu_A(\mathcal{B}) \leq 2q^{60} C(q) \varepsilon^{1/4}.$$

Combining the two, we get that

$$q^{-M/4} \geq \varepsilon^{1/4} \geq \frac{1}{C(q)q^{160}},$$

which is a contradiction for large enough  $M$ .  $\square$

We are now ready to prove Proposition 3.5.

*Proof of Proposition 3.5.* Take any  $(t + 1)$ -flat  $A' \subseteq A$  containing  $x^*$ , and define  $g = f|_{A'}$ . Consider the polynomial  $M(x) = 1_{x \neq x^*}$  on  $A'$ . Note that  $M$  has degree  $(t + 2)(q - 1)$ , so we may find a constant  $c \in \mathbb{F}_p$  such that  $g' = g + cM$  has degree strictly smaller than  $(t + 2)(q - 1)$ . We claim that  $\deg(g') \leq d$ . Otherwise, from Lemma 2.1 the fraction of  $t$ -flats  $B \subseteq A'$  such that  $g'|_B$  has degree greater than  $d$  is strictly larger than  $1/q$ . As the fraction of  $B$ 's that contain  $x^*$  is exactly  $1/q$ , it follows that there is  $B \subseteq A'$  not containing  $x^*$  such that  $\deg(g'|_B) > d$ . But for such  $B$ 's we have  $\deg(g'|_B) = \deg(f|_B) \leq d$ , and contradiction. Thus,  $\deg(g') \leq d$ . Stated otherwise, we may change the value of  $f(x^*)$  and make the degree of  $f|_{A'}$  at most  $d$ . In particular, we get that for each  $t$ -flat  $A'' \subseteq A$  we may change  $f(x^*)$  and make the degree of  $f|_{A''}$  at most  $d$ .

Sampling  $A$  a  $(t + 100)$  flat containing  $x^*$  randomly and then a  $t$ -flat  $A'' \subseteq A$  containing  $x^*$ , we get that with probability at least  $1/2$  we may change  $f(x^*)$  and make the degree of  $f|_{A''}$  at most  $d$ . Thus, taking the plurality vote we may choose  $f(x^*)$  that appeases at least  $\frac{1}{2q}$  of the  $t$ -flats containing  $x^*$ .  $\square$

### 3.3 Fixing the error and iterating

**Proposition 3.7.** *We may find  $x \in \mathbb{F}_q^n$  and function  $f'$  which is identical to  $f$  at all points except at  $x$ , such that*

$$\varepsilon_{t,d}(f') \leq \varepsilon_{t,d}(f) - q^{t-d} \frac{1}{4q}$$

*Proof.* Using Proposition 3.1 we find  $x^*$  such that  $\mu(S_{x^*}) \geq 1 - C(q)\varepsilon'$ , and using Proposition 3.5 we find  $c \in \mathbb{F}_q$  such that taking  $f'$  to be identical to  $f$  at all points except at  $x^*$  where it is equal to  $c$ , we have that  $f'$  passes at least  $\frac{1}{2q}$  fraction of the tests containing  $x^*$ . We compare the probability that  $f$  and  $f'$  pass the  $t$ -flat test. Sample a  $t$ -flat  $A$ . Clearly, if  $A$  does not contain  $x^*$  they perform the same; otherwise,  $f$  passes with probability at most  $C(q)\varepsilon'$ , and  $f'$  passes with probability at least  $\frac{1}{2q}$ . As the probability that  $x^* \in A$  is  $q^{t-n}$ , we get that

$$\varepsilon_{t,d}(f') \leq \varepsilon_{t,d}(f) - q^{t-d} \left( (1 - O(\varepsilon')) - \left(1 - \frac{1}{2q}\right) \right) \leq \varepsilon_{t,d}(f) - q^{t-d} \frac{1}{4q}. \quad \square$$

From Proposition 3.7 we get that as long as  $\varepsilon_{t,d}(f) > 0$ , we may find a point  $x$  and change  $f(x)$  so as to decrease  $\varepsilon_{t,d}(f)$  by at least  $q^{t-d}\frac{1}{4q}$ . Thus, after at most  $\frac{\varepsilon_{t,d}(f)}{q^{t-d}/4q}$  invocations of the proposition we will end up with a function that passes the test with probability 1, which by the choice of  $t$  implies we will end up with a degree  $d$  function. We therefore get that

$$\delta_d(f)q^n \leq \frac{\varepsilon_{t,d}(f)}{q^{t-d}/4q},$$

hence  $\delta_d(f) \leq 4q^{1-t}\varepsilon_{t,d}(f)$ . □

## 4 Lifted affine invariant codes: proof of Theorem 1.5

In this section, we argue that the method above used to prove optimal testing for Reed-Muller codes applies to lifted affine invariant codes as well, thereby proving Theorem 1.5. Towards this end, it turns out that the only part that has to be adjusted are Lemmas 2.1 and 2.2. Thus, we begin by proving them for affine invariant codes and then quickly explain how the rest of the proof proceeds.

### 4.1 Facts about affine invariant codes

**Definition 4.1.** *Let  $m, n \in \mathbb{N}$ , let  $p$  be prime and write  $m = \sum_{i=0}^r m_i p^i$ ,  $n = \sum_{i=0}^r n_i p^i$  the base  $p$  expansion of  $m$  and  $n$ . We say  $m$  dominates  $n$  with respect to the  $p$ -base expansion if  $m_i \geq n_i$  for all  $i$ .*

For a polynomial  $f$ , we denote by  $\text{supp}(f)$  the collection of monomials in  $f$  that have a non-zero coefficient. Also, for a set of functions  $\mathcal{B}$ , we denote by  $\text{supp}(\mathcal{B})$  the set of monomials that appear in at least one of these functions. Lastly, we will use the fact that the support of an affine invariant set is affine invariant.

**Lemma 4.2.** *[Monomial spreading [10, Lemma 4.6]] Suppose that  $\mathcal{B}$  is affine invariant, and let  $M = x_1^{d_1+e} x_2^{d_2} x_3^{d_3} \cdots x_t^{d_t}$  and  $M' = x_1^{d_1} x_2^{d_2+e} \cdots x_t^{d_t}$  be monomials such that  $d_1 + e$  dominates  $e$ . If  $M \in \text{supp}(\mathcal{B})$ , then  $M' \in \text{supp}(\mathcal{B})$ .*

Finally, we will use the following characterization of affine invariant codes, saying that they can be characterized by a monomial basis.

**Lemma 4.3.** *[ [10, Lemma 4.2]] If  $\mathcal{B}$  is an affine invariant linear code, then  $\mathcal{B} = \text{span}(\text{supp}(\mathcal{B}))$ .*

Thus, to show that  $f \notin \mathcal{B}$  it suffices to show that the support of  $f$  contains a monomial not in  $\mathcal{B}$ .

### 4.2 The relation lemma

We begin by adapting Lemma 2.1 to our case, following the argument in [6].

**Lemma 4.4.** *Let  $p$  be prime,  $q \in \mathbb{N}$  be a power of  $p$  and let  $t \in \mathbb{N}$ . Let  $\mathcal{B} \subseteq \{g: \mathbb{F}_q^t \rightarrow \mathbb{F}_q\}$  be an affine invariant code, and denote  $\mathcal{F} = \text{Lift}_{k+1}(\mathcal{B})$ . Suppose that  $k \geq t$ , and let  $f: \mathbb{F}_q^{k+1} \rightarrow \mathbb{F}_q$  be such that  $f \notin \mathcal{F}$ . Then*

1.  $\varepsilon_k(f) \geq \frac{1}{q}$ .

2. If  $\varepsilon_k(f) = \frac{1}{q}$ , and the set  $\mathcal{H}$  of hyperplanes  $H$  for which  $f|_H \notin \mathcal{F}$  is of the form

$$\mathcal{H} = \left\{ H \subseteq \mathbb{F}_q^{k+1} \mid x^* \in H \right\}$$

for some  $x^* \in \mathbb{F}_q^{k+1}$ , then there exists  $g \in \mathcal{F}$  that agrees with  $f$  on all points except on  $x^*$ .

*Proof.* Assume without loss of generality that  $x^* = 0$ . We will closely follow the argument in [6, Lemma 5.3] (we note that our assumption about  $x^*$  does not conflict with the assumption therein that  $T$  is the identity), which already establishes the first bullet. Our goal henceforth will be to establish the second bullet.

A hyperplane therein is indexed by  $\vec{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_{k+1})$ , which encodes the hyperplane

$$H = \left\{ x \mid \alpha_0 + \sum_{i=1}^{k+1} \alpha_i x_i = 0 \right\}.$$

For each hyperplane, let  $c_\alpha$  be the smallest  $i \geq 1$  such that  $\alpha_i \neq 0$ . The argument in [6] proceeds as follows:

1. If  $c_\alpha > t$ , the authors show that  $H \in \mathcal{H}$  given that  $\alpha_0 = 0$ . Hence, among the hyperplanes for which  $c_\alpha > t$ , at least  $1/q$  of them lie in  $\mathcal{H}$ .
2. If  $1 \leq c_\alpha \leq t$ , then one may alter  $\alpha_n$  and cause  $H$  to be in  $\mathcal{H}$ . Hence, at least  $1/q$  fraction of these hyperplanes are in  $\mathcal{H}$ . We note that if for some  $\alpha$ , there were at least 2 ways of choosing  $\alpha_n$  so that  $H \in \mathcal{H}$ , then we would get that the fraction of hyperplanes in  $\mathcal{H}$  from this case is strictly greater than  $1/q$ . Thus, since we assume that  $\varepsilon_k(f) = 1/q$ , there is precisely one way of choosing  $\alpha$  so that  $H \in \mathcal{H}$ .

We now look more closely at their analysis in the second case, starting from  $c_\alpha = 1$ . Consider as there

$$B(x_1, \dots, x_n) = \left( x_1 - \sum_{1 < j \leq n} \frac{\alpha_j}{\alpha_1} x_j - \frac{\alpha_0}{\alpha_1}, x_2, \dots, x_n \right),$$

and  $B': \mathbb{F}_q^t \rightarrow \mathbb{F}_q^t$  defined as

$$B'(x_1, \dots, x_t) = \left( x_1 - \sum_{1 < j \leq t} \frac{\alpha_j}{\alpha_1} x_j - \frac{\alpha_0}{\alpha_1}, x_2, \dots, x_t \right).$$

Note that  $(f \circ B)_{x_{t+1}=0, \dots, x_n=0} = f|_{x_{t+1}=0, \dots, x_n=0} \circ B' \notin \mathcal{B}$ , as  $f|_{x_{t+1}=0, \dots, x_n=0} \notin \mathcal{B}$  and  $\mathcal{B}$  is affine invariant. Thus, there is a monomial  $M$  in the support of  $f \circ B$  that is not in  $\text{supp}(\mathcal{B})$ , say

$$M = \prod_{i=1}^t x_i^{d_i}.$$

Let  $\alpha(z) = (\alpha_0, \alpha_1, \dots, \alpha_{n-1}, z)$ , let  $H_z$  be the hyperplane defined by  $\alpha(z)$ , and let

$$f_{\alpha(z)}(x_2, \dots, x_n) = f \left( - \sum_{1 < j \leq n} \frac{\alpha(z)_j}{\alpha_1} x_j - \frac{\alpha_0}{\alpha_1}, x_2, \dots, x_n \right)$$



be the restriction of  $f$  to  $H_z$ . Looking at  $f_{\alpha(z)}$  as a function of  $x_2, \dots, x_n$  and  $z$ , we get that the monomial

$$z^{d_1} x_n^{d_1} \prod_{i=2}^t x_i^{d_i}$$

appears in  $f_{\alpha(z)}$ . This is because  $f_{\alpha(z)}$  is the same as  $f \circ B$  when we replace  $x_1$  with  $\ell(z)x_n$  for some linear function  $\ell(z)$ . Thus, there are at least  $q - d_1$  choices for  $z$  to make that monomial survive in  $H_z$ , in which case we would have that  $H_z \in \mathcal{H}$ . Since by our assumption there is at most 1 such  $z$ , we get that  $d_1 = q - 1$ .

We now observe that the monomial  $M = \prod_{i=1}^t x_i^{d_i}$  must be in the support of  $f$ . Indeed, to have the monomial  $z^{d_1} x_n^{d_1} \prod_{i=2}^t x_i^{d_i}$  in  $f_{\alpha(z)}$ , as  $d_1 = q - 1$ , we must have a monomial whose degree in  $x_1$  is full (i.e.  $q - 1$ ), and expanding  $\left( \sum_{i=2}^n \frac{\alpha(z)_i}{\alpha(z)_1} x_i \right)^{q-1}$  (which would be what that monomial gives on  $x_1$ ), we must have that the contribution from it would have full degree in  $z$ , i.e. it must pick the term  $\alpha(z)_n^{q-1} x_n^{q-1}$ . This says that this part of the monomial does not contribute any  $x_j$  factors for  $j > 1$ , and hence those must be contributed from the original monomial itself.

Thus, we now have that  $d_1 = q - 1$ , and  $M \in \text{supp}(f) \setminus \text{supp}(\mathcal{B})$ . We now move on to the case  $c_\alpha = 2$ , and consider this monomial  $M$  and whether it stays alive in  $f_{\alpha(z)}$ . We look at the corresponding hyperplane as

$$-x_2 = \sum_{j>2} \frac{\alpha_j}{\alpha_2} x_j + \frac{\alpha_0}{\alpha_2},$$

and look at  $f_{\alpha(z)}$  and in particular in the monomial  $z^{d_2} x_n^{d_2} x_1^{q-1} \prod_{i=3}^t x_i^{d_i}$ . There are a few cases that have to be considered.

1. If it exists in  $f_{\alpha(z)}$ , we get that the coefficient of  $x_n^{d_2} x_1^{q-1} \prod_{i=3}^t x_i^{d_i}$  is a non-zero polynomial in  $z$  of degree at most  $d_2$ , and for each  $z$  for it is non-zero we get that  $H_z \in \mathcal{H}$ . Thus there must be a unique choice for  $z$  that would make it alive and necessarily  $d_2 = q - 1$ . We continue to the next  $c$ .
2. Otherwise, it means it has been canceled by some other monomial in  $f$ . We note that any such monomial must be of the form

$$M' = x_1^{q-1} x_2^{d'_2} \cdots x_t^{d'_t},$$

where  $d'_2 > d_2$ .

We argue that  $M' \notin \text{supp}(\mathcal{B})$ . Indeed, assume towards contradiction this is not the case. For this monomial to cancel  $M$ , we look at what happens when we plug in  $x_2$  as in  $\alpha(z)$ :

$$x_2^{d'_2} = \left( \sum_{j>2} \frac{\alpha_j(z)}{\alpha_2} x_j + \frac{\alpha_0}{\alpha_2} \right)^{d'_2} = \left( \frac{z}{\alpha_2} x_n + S \right)^{d'_2} = \sum_{r \leq d'_2} \binom{d'_2}{r} \left( \frac{z}{\alpha_2} x_n \right)^r S^{d'_2 - r},$$

where  $S = \sum_{2 < j < n} \frac{\alpha_j(z)}{\alpha_2} x_j + \frac{\alpha_0}{\alpha_2}$ . The contribution from this that may cancel  $M$  comes from  $r$ , so it is

$$\binom{d'_2}{d_2} \left( \frac{z}{\alpha_2} x_n \right)^{d_2} S^{d'_2 - d_2}.$$

By Lucas's theorem, for  $\binom{d'_2}{d_2}$  to be non-zero mod  $p$  (in which case the last expression is 0 as the characteristic of  $\mathbb{F}_q$  is  $p$ ) we need  $d'_2$  to dominate  $d_2$  in the  $p$ -basis. We then expand  $S^{d'_2-d_2}$ , and should get from it  $\prod_{i=3}^t x_i^{d_i-d'_i}$ . We will do so under the assumption that  $d_i \geq d'_i$  for  $i \geq 3$ ; the argument is similar otherwise. For example, if  $d_3 < d'_3$ , then below every occurrence of the difference  $(d_3 - d'_3)$  is to be replaced by  $(q - 1 + d_3 - d'_3)$ .

Doing the analysis term by term, we should have that  $d'_2 - d_2$  dominates  $d_3 - d'_3$  in the  $p$ -basis, and setting  $e_i = (d'_2 - d_2) - \sum_{j=3}^i (d_j - d'_j)$ , we should have that  $e_i$  dominates  $d_{i+1} - d'_{i+1}$  in the  $p$ -basis. Eventually, we must have that  $e_t = 0$ .

We now use the monomial spreading, i.e. Lemma 4.2. As  $d'_2$  dominated  $d_2$ , we may get that the monomial

$$M'' = x_1^{q-1} x_n^{d_2} x_2^{d'_2-d_2} x_3^{d'_3} x_4^{d'_4} \cdots x_t^{d'_t} = x_1^{q-1} x_n^{d_2} x_2^{e_2} x_3^{d'_3} x_4^{d'_4} \cdots x_t^{d'_t}$$

is in  $\text{supp}(\mathcal{B})$ . As  $e_2$  dominates  $d_3 - d'_3$ , we conclude again using Lemma 4.2 that the monomial

$$M''' = x_1^{q-1} x_n^{d_2} x_2^{e_2-(d_3-d'_3)} x_3^{d_3} x_4^{e_3} \cdots x_t^{d'_t} = x_1^{q-1} x_n^{d_2} x_3^{d_3} x_2^{e_3} x_4^{d_4} \cdots x_t^{d'_t},$$

is in  $\text{supp}(\mathcal{B})$ . Continuing in this way, we eventually conclude that  $M \in \text{supp}(\mathcal{B})$ , and contradiction.

It follows that we had  $M' \notin \text{supp}(\mathcal{B})$ , and we start the iteration for  $c = 2$  again with  $M'$ . Clearly, we will get stuck at  $c = 2$  at most  $q - 1$  as the degree of  $x_2$  increases each time, hence eventually we will hit  $d_2 = q - 1$  and proceed to the next variable.

Hence, we conclude that under the assumption of the lemma and  $x^* = 0$ , we have that the monomial  $\prod_{i=1}^t x_i^{q-1}$  must appear in  $\text{supp}(f)$ . Define  $g = f + s1_{x=x^*}$  for some  $s \in \mathbb{F}_q$  such that  $\prod_{i=1}^t x_i^{q-1} \notin \text{supp}(g)$ ; this is clearly possible, as the support of  $1_{x=x^*}$  is full. We would get that the set of hyperplanes  $H$  for which  $g|_H \notin \mathcal{F}$  is contained in  $\mathcal{H}$ , as we only changed  $f$  in  $x^*$  and any  $H \notin \mathcal{H}$  does not contain it. Hence,  $\varepsilon_k(g) \leq \varepsilon_k(f) = 1/q$ . We claim that  $g \in \mathcal{F}$ . Indeed, otherwise we would run the above argument on  $g$  and conclude that  $g$  must contain the monomial  $\prod_{i=1}^t x_i^{q-1}$  in its support, which is clearly impossible. This is a contradiction, and therefore  $g \in \mathcal{F}$  as desired.  $\square$

The following corollary is immediate:

**Corollary 4.5.** *Let  $p$  be prime,  $q \in \mathbb{N}$  be a power of  $p$  and let  $t \in \mathbb{N}$ . Let  $\mathcal{B} \subseteq \{g: \mathbb{F}_q^t \rightarrow \mathbb{F}_q\}$  be an affine invariant code, and denote  $\mathcal{F} = \text{Lift}_{k+1}(\mathcal{B})$ . Suppose that  $k \geq t$ , and let  $f: \mathbb{F}_q^{k+1} \rightarrow \mathbb{F}_q$  be such that  $f \notin \mathcal{F}$ . Then if  $k \geq k' \geq t$ , then  $\varepsilon_k(f) \leq q^{k-k'}(f)$ .*

### 4.3 Proof of Theorem 1.5

In this section, we explain how to adapt the argument in Section 3 to prove Theorem 1.5. First, the set  $S$  in this context is defined to be

$$S = \{A \mid \dim(A) = t, f|_A \notin \mathcal{B}\}.$$

For the argument there we need  $t$  to be a sufficiently large constant, say larger than  $M$ , and we claim we may indeed assume that. Indeed, otherwise we may look at the  $t + M$  flat tester and get that  $\mu(S \uparrow^{t+M}) \leq$

$q^M \mu(S)$  is the rejection probability (where we used Corollary 4.5). We then look at the problem as trying to understand the lifted code of  $\mathcal{B} = \text{Lift}_{t+M}(\mathcal{B})$ , which is also affine invariant, and we now have that the new  $t$  is large enough. We henceforth assume that  $t$  is large enough to begin with.

Claims 3.2, 3.3 remain unchanged except that we appeal to Lemma 4.4 instead of Lemma 2.2. The proof of Claim 3.4 also remains unchanged, except that in the end we appeal again to Lemma 4.4 instead of Lemma 2.2. This establishes Proposition 3.1 in this case.

The discussion before Claim 3.6 and the claim itself continue to hold as is in this case, and we explain the slight adaptation to the rest of the argument in Section 3.2.

**Proposition 4.6.** *There exists  $c \in \mathbb{F}_q$  such that changing the value of  $f(x^*)$  to  $c$ , we have that*

$$\Pr_{A' \text{ } t\text{-flat}} [\deg(f|_{A'}) | x^* \in A'] \geq \frac{1}{2q}.$$

*Proof.* Take any  $(t+1)$ -flat  $A' \subseteq A$  containing  $x^*$ , and define  $g = f|_{A'}$ . We claim that we may change  $f$  at  $x^*$  and have that  $g \in \mathcal{F}$ . Otherwise, from the second item in Lemma 4.4, the fraction of  $t$ -flats  $B \subseteq A'$  such that  $g|_B \notin \mathcal{B}$  is larger than  $1/q$ . As the fraction of  $B$ 's that contain  $x^*$  is exactly  $1/q$ , it follows that there is  $B \subseteq A'$  not containing  $x^*$  such that  $g|_B \notin \mathcal{B}$ . But for such  $B$ 's we have  $g|_B = f|_B$ , and contradiction.

Sampling  $A$  a  $(t+100)$  flat containing  $x^*$  randomly and then a  $t$ -flat  $A'' \subseteq A$  containing  $x^*$ , we get that with probability at least  $1/2$  we may change  $f(x^*)$  and have  $f|_{A''} \in \mathcal{B}$ . Thus, taking the majority vote we may choose  $f(x^*)$  that appears at least  $\frac{1}{2q}$  of the  $t$ -flats containing  $x^*$ .  $\square$

Given Proposition 4.6, Section 3.3 goes through as well, completing the proof of Theorem 1.5.  $\square$

## 5 Discussion and open questions

Our work explores a potential connection between testing questions in codes and expansion in the underlying test graph, using the idea that the error set exhibits some non sharp-threshold type behaviour. This connection highlights several problems that we think may be of interest.

1. Stability results for Kruskal-Katona type theorems. What can we say about the structure of small sets  $S \subseteq V_q(k, \ell)$  for which  $\mu(S \uparrow) \leq q\mu(S)$ ? Using our techniques, it follows that such sets must be correlated with a zoom-in set or a zoom-in with respect to the linear part (which we are able to eliminate in our case), but it would be interesting to get a more thorough understanding of this problem. Similarly, it would be interesting to understand the structure of large sets with non-perfect shadow, i.e.  $\mu(S \uparrow) \leq 1 - \delta$ .
2. Beyond lifted codes. Can we use expansion type results on structures such as the Grassmann graph (but maybe more) to prove more testing results on other codes? As we have seen, the proof goes through relatively easily for the class of lifted affine invariant codes (improving the dependency on the field size  $q$  over the result of [6]), and we suspect our method should apply in other settings as well.
3. Characterization of near degree 1 functions on the Affine Grassmann graph. As shown in Lemma B.3, small sets  $S$  for which  $1 - \Phi(S) \geq 1/q$  have almost all of their Fourier degree on the first level. In this case, we establish a relatively weak structural result, and it is tempting to ask whether a more detailed structural result holds in this case similarly to the classical FKN theorem from the Boolean cube [5].

4. Beyond the 99% regime. Can the approach suggested herein, or similar ones, be applied to study the testing question for the Reed-Muller code wherein the success probability of the tester is only guaranteed to be at least  $1/q + \delta$ , i.e. the notorious 1% regime?

## References

- [1] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Trans. Inf. Theory*, 51(11):4032–4039, 2005.
- [2] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of Reed-Muller codes. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 488–497, 2010.
- [3] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. On non-optimally expanding sets in Grassmann graphs. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 940–951, 2018.
- [4] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 376–389, 2018.
- [5] Ehud Friedgut, Gil Kalai, and Assaf Naor. Boolean functions whose Fourier transform is concentrated on the first two levels. *Advances in Applied Mathematics*, 29(3):427–437, 2002.
- [6] Elad Haramaty, Noga Ron-Zewi, and Madhu Sudan. Absolutely sound testing of lifted codes. *Theory Comput.*, 11:299–338, 2015.
- [7] Elad Haramaty, Amir Shpilka, and Madhu Sudan. Optimal testing of multivariate polynomials over small prime fields. *SIAM J. Comput.*, 42(2):536–562, 2013.
- [8] Charanjit S. Jutla, Anindya C. Patthak, Atri Rudra, and David Zuckerman. Testing low-degree polynomials over prime fields. *Random Struct. Algorithms*, 35(2):163–193, 2009.
- [9] Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM J. Comput.*, 36(3):779–802, 2006.
- [10] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 403–412, 2008.
- [11] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Global hypercontractivity and its applications. *arXiv preprint arXiv:2103.04604*, 2021.
- [12] Subhash Khot, Dor Minzer, and Muli Safra. On independent sets, 2-to-2 games, and Grassmann graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 576–589, 2017.
- [13] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in Grassmann graph have near-perfect expansion. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 592–601, 2018.

[14] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing, El Paso, Texas, USA, May 4-6, 1997*, pages 475–484, 1997.

## Appendix

This section is devoted to the proof of Theorem 2.4. Our approach closely follows the approach in [13], however as we are only concerned with the special case associated with zoom-ins/ zoom-outs of dimension/ co-dimension 1, our analysis is considerably simpler. Roughly speaking, our proof consists of the following three components:

1. First, we define a Cayley graph that closely resembles the affine Grassmann graph, and show that studying expansion over the two is roughly equivalent (up to some loss in the parameters).
2. Second, we show that for the expansion parameters in question, the problem reduces to studying the structure of functions that have almost all of their Fourier mass on the first level component in the natural degree decomposition.
3. Finally, we perform a 4th-moment vs 2nd-moment type analysis and deduce the structural result.

Throughout this section, we think of  $W$  as a linear space over  $\mathbb{F}_q$  with dimension  $k$ ; without loss of generality  $W = \mathbb{F}_q^k$ . We consider the affine Grassmann graph over  $\ell$ -flats.

## A The Cayley graph construction

Consider the edge-weighted graph  $H = (V, E)$  defined as follows. The set of vertices  $V$  consist of tuples  $(s, x_1, \dots, x_\ell)$  where  $s, x_1, \dots, x_\ell \in \mathbb{F}_q^k$ . The edge weights are described according to the following randomized process; to sample a neighbour of  $(s, x_1, \dots, x_\ell)$ :

1. sample  $y \in \mathbb{F}_q^k$  uniformly;
2. sample  $b_0, b_1, \dots, b_\ell \in \mathbb{F}_q$  uniformly;
3. output  $(s + b_0 y, x_1 + b_1 y, \dots, x_\ell + b_\ell y)$ .

Given a set of vertices in the affine Grassmann graph  $S \subseteq V(\mathbb{F}_q^k, \ell)$ , we associate with it the set  $S^*$  in the Cayley graph defined as

$$S^* = \{(s, x_1, \dots, x_\ell) \mid s + \text{span}(x_1, \dots, x_\ell) \in S\}.$$

We establish some properties of  $S$  and  $S^*$ . First, we show that the non-expansion of  $S^*$  may be lower bounded by the non-expansion of  $S$  (in fact the two are close, but we only need this direction).

**Claim A.1.**  $1 - \Phi(S^*) \geq 1 - \Phi(S) - q^{-\ell}$ .

*Proof.* Recall that  $1 - \Phi(S^*)$  is the probability that starting from a random vertex in  $S^*$  and taking a step, we stay in the set  $S^*$ . Denote by  $v = (s, x_1, \dots, x_\ell)$  the starting point of the walk, by  $y, b_0, \dots, b_\ell$  the parameters that define the step of the walk, and by  $u$  the endpoint of the random walk. There are a few cases:

1.  $b_0 = b_1 = \dots = b_\ell = 0$ , which happens with probability  $q^{-(\ell+1)}$  and corresponds to a self-loop.
2.  $b_0 \neq 0, b_1 = \dots = b_\ell = 0$ , which corresponds to the case the hyperplanes defined by  $v, u$  are parallel. This happens with probability  $\frac{q-1}{q^{\ell+1}}$ .
3.  $\text{span}(x_1 + b_1 y, \dots, x_\ell + b_\ell y)$  has dimension less than  $\ell$ , which happens with probability at most  $q^{\ell-k}$ .
4. Otherwise,  $u$  is a random affine space of dimension  $\ell$  that intersects  $v$  in size  $q^{\ell-1}$ . This happens with probability  $(1 - q^{-\ell} - q^{\ell-k})$ .

We note that in the case of the 3rd item, we always escape the set and hence this doesn't contribute to  $1 - \Phi(S^*)$ . We compare the rest of these probabilities to the corresponding walk on the affine Grassmann graph. Starting at an affine space  $V$  of dimension  $\ell$ , going to  $K \supseteq U$  of dimension  $\ell + 1$  and then to a random  $U \subseteq K$  of dimension  $\ell$ , we have:

1. The probability that  $U = V$  is  $\frac{1}{q^{\ell+1}-1} \frac{q-1}{q}$ .
2. The probability that  $V$  and  $U$  are parallel is  $\frac{q-1}{\frac{q^{\ell+1}-1}{q-1}q} = \frac{(q-1)^2}{q(q^{\ell+1}-1)}$ .
3. Otherwise,  $U$  is random affine space of dimension  $\ell$  that intersects  $V$  in size  $q^{\ell-1}$ . The probability here is  $1 - \frac{q-1}{q^{\ell+1}-1}$ .

Looking at the ratios between the probability of a case in the Cayley graph and the probability of a case in the affine Grassmann graph, the first two are at least 1, whereas the last one is at least  $1 - q^{-\ell}$ . Thus,

$$1 - \Phi(S^*) \geq (1 - q^{-\ell})(1 - \Phi(S)) \geq 1 - \Phi(S) - q^{-\ell}. \quad \square$$

Next, we consider the analogous notions of zoom-ins for sets in the Cayley graph.

**Definition A.2.** *Let  $T$  be a set in the Cayley graph.*

1. For  $z \in \mathbb{F}_p^k$ , the zoom-in of  $T$  with respect to  $z$  is the set

$$\{(s, x_1, \dots, x_\ell) \mid z \in s + \text{span}(x_1, \dots, x_\ell)\}.$$

2. For  $z \in \mathbb{F}_p^k \setminus \{0\}$ , the zoom-in of  $T$  with respect to  $z$  on the linear part is the set

$$\{(s, x_1, \dots, x_\ell) \mid z \in \text{span}(x_1, \dots, x_\ell)\}.$$

3. For an affine hyperplane  $W \subseteq \mathbb{F}_p^k$ , the zoom-out of  $T$  with respect to  $W$  is the set

$$\{(s, x_1, \dots, x_\ell) \mid s + \text{span}(x_1, \dots, x_\ell) \subseteq W\}.$$

4. For a hyperplane  $W \subseteq \mathbb{F}_p^k$ , the zoom-in of  $T$  with respect to  $W$  on the linear part is the set

$$\{(s, x_1, \dots, x_\ell) \mid \text{span}(x_1, \dots, x_\ell) \subseteq W\}.$$

For each one of these cases, say for zoom-ins, we say that  $T$  is  $\xi$ -pseudo-random with respect to it if

$$\mu(T_z) \stackrel{\text{def}}{=} \frac{|\{(s, x_1, \dots, x_\ell) \in T \mid z \in s + \text{span}(x_1, \dots, x_\ell)\}|}{|\{(s, x_1, \dots, x_\ell) \mid z \in s + \text{span}(x_1, \dots, x_\ell)\}|} \leq \xi.$$

We now show that notions of pseudo-randomness of  $S$  transfer to the same notions of pseudo-randomness for  $S^*$ .

**Claim A.3.** *If  $S$  is  $\xi$ -pseudo-random against zoom-ins, then  $S^*$  is  $\xi$  pseudo-random with respect to zoom-ins. Same goes for zoom-outs etc.*

*Proof.* Sampling  $v = (s, x_1, \dots, x_\ell)$  from the Cayley graph conditioned on it representing an affine subspace of dimension  $\ell$  and containing  $z$ , the subspace it represents is distributed uniformly among all subspaces containing  $z$ , hence in  $S_z$  with probability  $\mu(S_z) \leq \xi$ . If  $v$  does not represent an affine subspace of dimension  $\ell$ , we clearly have  $v \notin (S^*)_z$ . Thus,

$$\mu((S^*)_z) = \Pr_v[v \in S^* \wedge v \text{ is dimension } \ell \mid z \in v] \leq \Pr_v[v \in S^* \mid z \in v, v \text{ is dimension } \ell] = \mu(S_z) \leq \xi. \quad \square$$

As a special case of the previous claim, we get that a good zoom-in for  $S^*$  (i.e., one on which the measure of this set is almost 1) is also be good for  $S$ .

**Corollary A.4.** *Suppose that  $\mu((S^*)_z) \geq 1 - \delta$ . Then  $\mu(S_z) \geq 1 - \delta$ .*

## B Decompositions

### B.1 The Fourier decomposition

Let  $F = 1_{S^*}$ . We shall now think of  $F: \mathbb{F}_q^k \rightarrow \{0, 1\}$  as a function, and develop it according to the basis of characters. In this context, writing  $q = p^r$  where  $p$  is prime, we consider the trace map  $\text{Tr}: \mathbb{F}_q \rightarrow \mathbb{F}_p$  defined as  $\text{Tr}(a) = \sum_{i=1}^{r-1} a^{p^i}$ . A character of  $\mathbb{F}_q$  is then defined as  $\chi_a(x) = \omega^{\text{Tr}(ax)}$  for  $a \in \mathbb{F}_q$ , where  $\omega$  is the  $p$ th root of unity. A character of  $\mathbb{F}_q^k$  is indexed by  $\vec{a} \in \mathbb{F}_q^k$  and is defined as  $\chi_{\vec{a}}(x) = \prod \chi_{a_i}(x_i) = \omega^{\sum_{i=1}^k \text{Tr}(a_i x_i)}$ . Finally, a character of  $(\mathbb{F}_q^k)^{\ell+1}$  is indexed by  $\alpha = (\alpha_0, \dots, \alpha_\ell) \in (\mathbb{F}_q^k)^{\ell+1}$  and is defined as

$$\chi_\alpha(s, x_1, \dots, x_\ell) = \chi_{\alpha_0}(s) \prod_{i=1}^{\ell} \chi_{\alpha_i}(x_i).$$

We will use the abbreviation  $x = (x_1, \dots, x_\ell)$ , and then write

$$F(s, x) = \sum_{\alpha} \widehat{F}(\alpha) \chi_\alpha(s, x), \quad \text{where } \widehat{F}(\alpha) = \mathbb{E}_{(s, x)} \left[ F(s, x) \overline{\chi_\alpha(s, x)} \right].$$

**Claim B.1.** *Suppose we have  $\alpha, \beta$  such that  $\alpha_0 = \beta_0$  and  $\text{span}(\alpha_0, \alpha_1, \dots, \alpha_\ell) = \text{span}(\beta_0, \beta_1, \dots, \beta_\ell)$ . Then  $\widehat{F}(\alpha) = \widehat{F}(\beta)$ .*

*Proof.* Follows as  $f$  is invariant under  $(s, x_1, \dots, x_\ell) \rightarrow (s + z_0, z_1, \dots, z_\ell)$  where  $z_1, \dots, z_\ell$  are linearly independent linear combinations of  $x_1, \dots, x_\ell$ , and  $z_0$  is a linear combination of  $x_1, \dots, x_\ell$ .  $\square$

Next, we calculate the eigenvalues of the characters with respect to the random walk on the Cayley graph.

**Claim B.2.** *Let  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_\ell)$  be such that  $\dim(\text{span}(\alpha_0, \dots, \alpha_\ell)) = d$ . Then  $\chi_\alpha$  is an eigenfunction with respect to the random walk on the Cayley graph with eigenvalue  $q^{-d}$ .*

*Proof.* The eigenvalue is easily seen to be equal to  $\mathbb{E}_{b_0, b_1, \dots, b_\ell, y} \left[ \chi_{\sum_{i=0}^{\ell} b_i \alpha_i} (y) \right]$ . Note that if the dimension of  $\text{span}(\alpha_0, \dots, \alpha_\ell)$  is  $d$ , then the probability that  $\sum_{i=0}^{\ell} b_i \alpha_i = 0$  is  $q^{-d}$ . In that case, the expectation is 1, and otherwise it is 0.  $\square$

## B.2 The level decomposition

For  $i = 0, 1, \dots, \ell$ , define

$$F_{\text{lin}, i}(s, x) = \sum_{\substack{\alpha: \alpha_0 \in \text{span}(\alpha_1, \dots, \alpha_\ell) \\ \dim(\text{span}(\alpha_1, \dots, \alpha_\ell)) = i}} \widehat{F}(\alpha) \chi_\alpha(s, x), \quad F_{\text{aff}, i}(s, x) = \sum_{\substack{\alpha: \alpha_0 \notin \text{span}(\alpha_1, \dots, \alpha_\ell) \\ \dim(\text{span}(\alpha_1, \dots, \alpha_\ell)) = i}} \widehat{F}(\alpha) \chi_\alpha(s, x),$$

and for simplicity  $F_i(s, x) = F_{\text{lin}, i}(s, x) + F_{\text{aff}, i-1}(s, x)$ . Clearly

$$F(s, x) = \sum_{i=0}^{\ell} F_i(s, x).$$

Denoting by  $H$  the normalized adjacency operator of the Cayley graph, we have that  $HF(s, x) = \sum_{i=0}^{\ell} q^{-i} F_i$ , and so

$$1 - \Phi(S^*) = \frac{1}{\mu(S^*)} \langle F, HF \rangle = \frac{1}{\mu(S^*)} \left( \|F_0\|_2^2 + \sum_{i=0}^{\ell} q^{-i} \|F_i\|_2^2 + q^{-\ell-1} \|F_{\text{aff}, \ell}\|_2^2 \right).$$

As  $1 - \Phi(S^*) \geq \frac{1}{q} - \frac{1}{q^\ell}$  from Claim A.1,  $\|F_0\|_2 = \mu(S^*)$  and  $\sum_{i \geq 2} \|F_i\|_2^2 = \mu(S^*) - \mu(S^*)^2 - \|F_1\|_2^2$  by Parseval, we get that

$$\frac{1}{q} - \frac{1}{q^\ell} \leq \frac{1}{\mu(S^*)} \left( \mu(S^*)^2 + \frac{1}{q} \|F_1\|_2^2 + \frac{1}{q^2} (\mu(S^*) - \mu(S^*)^2 - \|F_1\|_2^2) \right).$$

Rearranging we get

$$\frac{1}{q} - \frac{1}{q^2} - \frac{1}{q^\ell} \leq \frac{1}{\mu(S^*)} \left( \frac{1}{q} - \frac{1}{q^2} \right) \|F_1\|_2^2 + \mu(S^*),$$

and so

$$\frac{\|F_1\|_2^2}{\mu(S^*)} \geq 1 - q^{2-\ell} - q^2 \xi.$$

We summarize this discussion with the following lemma.

**Lemma B.3.** *Let  $S$  be as in Theorem 2.4, and let  $S^*$  be the corresponding set in the Cayley graph. Then letting  $F = 1_{S^*}$  and looking at the level decomposition above, we have*

$$\frac{\|F_1\|_2^2}{\mu(S^*)} \geq 1 - q^{2-\ell} - q^2 \xi.$$



### B.3 Lower bounding the fourth norm of $F_1$ and stating the upper bound

We now move on to the heart of the argument which handles the fourth norm of  $F_1$ . First, we show an easy lower bound on it:

**Corollary B.4.**  $\frac{\|F_1\|_4^4}{\mu(S^*)} \geq (1 - q^{2-\ell} - q^2\xi)^4$ .

*Proof.* By Hölder's inequality we have

$$\|F_1\|_2^2 = \langle F_1, F_1 \rangle = \langle F_1, F \rangle \leq \|F_1\|_4 \|F\|_{4/3} = \|F_1\|_4 \mu(S^*)^{3/4},$$

using the lower bound on the left hand side from Lemma B.3 establishes the claim.  $\square$

Next, we state the upper bound on it, and then show how the two bounds imply Theorem 2.4. The rest of the appendix is then devoted into proving this upper bound.

**Lemma B.5.** *Suppose  $S$  is*

1.  $\xi$  pseudo-random with respect to zoom-outs (as well as on its linear part),
2.  $\mu(S) \leq \xi$ ,
3.  $\xi$  pseudo-random zoom ins with respect to their linear part,
4. a pseudo-random with respect to zoom-ins.

Then

$$\|F_1\|_4^4 \leq \mu(S^*)a^2 + 863q^2\mu(S^*)\xi^{1/4}.$$

We now show the quick derivation of Theorem 2.4.

*Proof of Theorem 2.4.* Combining Corollary B.4 and Lemma B.5 we get that

$$a^2 + 863q^2\xi^{1/4} \geq 1 - 4q^{2-\ell} - 4q^2\xi,$$

so  $a \geq 1 - q^2(867\xi^{1/4} + q^{-\ell})$  provided  $\xi$  is small enough with respect to  $q$  ( $\xi \leq q^{-10}$  will do).  $\square$

### B.4 An alternative description to $F_1$

To handle  $F_1$ , we shall need a different combinatorial description for  $F_1$ . Define  $f_{1,\text{lin}}, f_{1,\text{aff}}: \mathbb{F}_q^\ell \rightarrow [-1, 1]$  as

$$f_{1,\text{lin}}(x) = \mu((S^*)_{x,\text{lin}}) - \mu(S^*), \quad f_{1,\text{aff}}(x) = \mu((S^*)_{x,\text{aff}}) - \mu(S^*).$$

Let  $\mathcal{M} = \mathbb{F}_q^\ell \setminus \{0\}$ . We define the equivalence relation on  $\mathcal{M}$  which is  $M \sim M'$  if  $M = iM'$  for some  $i \in \mathbb{F}_q$ , and let  $\mathcal{B}$  be the equivalency classes of this relations; we choose a representative element from each equivalency class (arbitrarily).

**Claim B.6.**  $F_1(s, x) = \sum_{M \in \mathcal{B}} f_{1,\text{lin}}(\langle M, x \rangle) + \sum_{M \in \mathbb{F}_q^\ell} f_{1,\text{aff}}(s + \langle M, x \rangle)$ .

*Proof.* By definition, we have

$$F_{1,\text{lin}}(s, x) = \sum_{\beta \in \mathcal{B}} \sum_{M \in \mathcal{M}, v \in \mathbb{F}_q} \widehat{F}(v\beta, M_1\beta, \dots, M_\ell\beta) \chi_{v\beta, M_1\beta, \dots, M_\ell\beta}(s, x).$$

We split this sum according to  $v = 0$  and  $v \neq 0$ .

**Contribution from  $v = 0$ .** For  $v = 0$ , using Claim B.1 we get contribution of

$$\begin{aligned}
& \sum_{\beta \in \mathcal{B}} \sum_{M \in \mathcal{M}} \widehat{F}(0, \beta, 0, \dots, 0) \chi_{0, M_1 \beta, \dots, M_\ell \beta}(s, x) \\
&= \frac{1}{q-1} \sum_{\beta \in \mathbb{F}_q^k \setminus \{0\}} \mathbb{E}_{s', x'} [F(s', x') \chi_\beta(-x'_1)] \sum_{M \in \mathcal{M}} \chi_\beta \left( \sum_{i=1}^{\ell} M_i x_i \right) \\
&= \frac{1}{q-1} \mathbb{E}_{s', x'} \left[ F(s', x') \sum_{M \in \mathcal{M}} \sum_{\beta \in \mathbb{F}_q^k \setminus \{0\}} \chi_\beta \left( -x'_1 + \sum_{i=1}^{\ell} M_i x_i \right) \right].
\end{aligned}$$

Adding  $\frac{|\mathcal{M}|}{q-1} \mu(S^*)$  to this expression amounts to also including  $\beta = 0$ , hence we get that

$$\frac{|\mathcal{M}|}{q-1} F_0 + F_{1, \text{lin}}(s, x) = \frac{1}{q-1} \mathbb{E}_{s', x'} \left[ F(s', x') \sum_{M \in \mathcal{M}} \sum_{\beta \in \mathbb{F}_q^k} \chi_\beta \left( -x'_1 + \sum_{i=1}^{\ell} M_i x_i \right) \right].$$

If  $-x'_1 + \sum_{i=1}^{\ell} M_i x_i \neq 0$ , the sum over  $\beta$  is 0 and otherwise it is  $q^k$ , so we get

$$\begin{aligned}
\frac{q^k}{q-1} \mathbb{E}_{s', x'} \left[ F(s', x') \sum_{M \in \mathcal{M}} 1_{x'_1 = \sum_{i=1}^{\ell} M_i x_i} \right] &= \frac{q^k}{q-1} \sum_{M \in \mathcal{M}} \mathbb{E}_{s', x'} [F(s', x') 1_{x'_1 = \langle M, x \rangle}] \\
&= \frac{1}{q-1} \sum_{M \in \mathcal{M}} \mu((S^*)_{\langle M, x \rangle, \text{lin}}) \\
&= \sum_{M \in \mathcal{B}} \mu((S^*)_{\langle M, x \rangle, \text{lin}}).
\end{aligned}$$

**Contribution from  $v \neq 0$ .** For  $v \neq 0$  we get from similar computations that the contribution is

$$\sum_{\beta \in \mathcal{B}} \sum_{M \in \mathcal{M}, v \neq 0} \widehat{F}(v \beta, 0, \dots, 0) \chi_{v \beta, M_1 \beta, \dots, M_\ell \beta}(s, x) = \sum_{\beta \in \mathbb{F}_q^k \setminus \{0\}} \sum_{M \in \mathcal{M}} \widehat{F}(\beta, 0, \dots, 0) \chi_{\beta, M_1 \beta, \dots, M_\ell \beta}(s, x).$$

We add to that  $F_{1, \text{aff}}$ , which is the term corresponding to  $M = 0$ ; we then add  $q^\ell F_0$ , which corresponds to taking  $\beta = 0$  as well. Hence we get that the contribution from  $v \neq 0$  plus  $F_{1, \text{aff}}(s, x) + q^\ell \mu(S^*)$  is

$$\begin{aligned}
& \sum_{\beta \in \mathbb{F}_q^k} \mathbb{E}_{s', x'} [F(s', x') \chi_\beta(-s')] \sum_{M \in \mathbb{F}_q^\ell} \chi_\beta \left( s + \sum_{i=1}^{\ell} M_i x_i \right) \\
&= \mathbb{E}_{s', x'} \left[ F(s', x') \sum_{M \in \mathbb{F}_q^\ell} \sum_{\beta \in \mathbb{F}_q^k} \chi_\beta \left( s + \sum_{i=1}^{\ell} M_i x_i - s' \right) \right].
\end{aligned}$$

If  $-s' + s + \sum_{i=1}^{\ell} M_i x_i = 0$  we get that the sum over  $\beta$  is  $q^k$  and otherwise it is 0. Hence we get

$$\begin{aligned} q^k \mathbb{E}_{s', x'} \left[ f(s', x') \sum_{M \in \mathbb{F}_q^\ell} 1_{-s' + s + \sum_{i=1}^{\ell} M_i x_i = 0} \right] &= q^k \sum_{M \in \mathbb{F}_q^\ell} \mathbb{E}_{s', x'} [f(s', x') 1_{\langle M, x \rangle + s = s'}] \\ &= \sum_{M \in \mathbb{F}_q^\ell} \mu((S^*)_{s + \langle M, x \rangle}). \end{aligned}$$

Combining all, and moving the multiples of  $\mu(S^*)$  we have added to the other side, we get that

$$F_{1, \text{lin}}(s, x) + F_{1, \text{aff}}(s, x) = \sum_{M \in \mathcal{B}} f_{1, \text{lin}}(\langle M, x \rangle) + \sum_{M \in \mathbb{F}_q^\ell} f_{1, \text{aff}}(s + \langle M, x \rangle). \quad \square$$

## C Properties of $f_{1, \text{lin}}$ and $f_{1, \text{aff}}$

### C.1 Orthogonality and symmetries

**Claim C.1.** *We have*

$$\mathbb{E}_{x \in \mathbb{F}_q^k \setminus 0} [f_{1, \text{lin}}(x)] = 0, \quad \mathbb{E}_{x \in \mathbb{F}_q^k} [f_{1, \text{aff}}(x)] = 0.$$

*Proof.* This is obvious by the definition of these functions. □

### C.2 Second moment

**Claim C.2.** *We have*

$$\mathbb{E}_{x \in \mathbb{F}_q^k \setminus 0} [f_{1, \text{lin}}(x)^2] \leq \frac{\|F_1\|_2^2}{|\mathcal{B}|} \leq \frac{\mu(S^*)}{|\mathcal{B}|}, \quad \mathbb{E}_{x \in \mathbb{F}_q^k} [f_{1, \text{aff}}(x)^2] \leq \frac{\|F_1\|_2^2}{q^\ell} \leq \frac{\mu(S^*)}{q^\ell}.$$

*Proof.* Expanding  $\|F_1\|_2^2$ , it is equal to

$$\mathbb{E}_{s, x} \left[ \sum_{M \in \mathcal{B}} |f_{1, \text{lin}}(\langle M, x \rangle)|^2 + \sum_{M \in \mathcal{B}, M' \in \mathbb{F}_q^\ell} f_{1, \text{lin}}(\langle M, x \rangle) f_{1, \text{aff}}(s + \langle M, x \rangle) + \sum_{M' \in \mathbb{F}_q^\ell} |f_{1, \text{aff}}(s + \langle M, x \rangle)|^2 \right].$$

We note that for each  $x, M, M'$ , the expectation of  $f_{1, \text{lin}}(\langle M, x \rangle) f_{1, \text{aff}}(s + \langle M, x \rangle)$  over  $z$  is 0 by Claim C.1, hence the middle sum vanishes. The other two sums are non-negative so it follows that each one of them is at most  $\|F_1\|_2^2$  in expectation, and the claim follows by translating them into expectations □

### C.3 Fourier coefficients

We shall now think of  $f_{1, \text{lin}}, f_{1, \text{aff}}$  as functions from  $\mathbb{F}_q^k$  to  $\mathbb{R}$  and may therefore discuss their Fourier coefficients.

**Claim C.3.** *Let  $\alpha \in \mathbb{F}_q^k$  index a Fourier coefficient. Then*

$$\widehat{f_{1, \text{lin}}}(\alpha) = \frac{1}{q-1} \widehat{F}(0, \alpha, \dots, \alpha), \quad \widehat{f_{1, \text{aff}}}(\alpha) = \widehat{F}(\alpha, 0, \dots, 0).$$

*Proof.* By definition

$$F_1(s, x) = \sum_{M \in \mathcal{B}} f_{1,\text{lin}}(\langle M, x \rangle) + \sum_{M \in \mathbb{F}_q^\ell} f_{1,\text{aff}}(s + \langle M, x \rangle), \quad (1)$$

and we expand the right hand side, as well as the left hand side, according to Fourier decomposition. The first term on the right hand side is equal to

$$\frac{1}{q-1} \sum_{M \in \mathcal{M}} \sum_{\alpha \in \mathbb{F}_q^k} \widehat{f_{1,\text{lin}}}(\alpha) \chi_\alpha(\langle M, x \rangle).$$

We have

$$\chi_\alpha(\langle M, x \rangle) = \chi_\alpha(M_1 x_1 + \dots + M_\ell x_\ell) = \chi_{M_1 \alpha}(x_1) \cdots \chi_{M_\ell \alpha}(x_\ell) = \chi_{(0, M \alpha)}(s, x),$$

hence

$$\sum_{M \in \mathcal{B}} f_{1,\text{lin}}(\langle M, x \rangle) = \frac{1}{q-1} \sum_{M \in \mathcal{M}} f_{1,\text{lin}}(\langle M, x \rangle) = \frac{1}{q-1} \sum_{M \in \mathcal{M}} \sum_{\alpha \in \mathbb{F}_q^k} \widehat{f_{1,\text{lin}}}(\alpha) \chi_{(0, M \alpha)}(s, x). \quad (2)$$

Similarly, we get that

$$\sum_{M \in \mathbb{F}_q^\ell} f_{1,\text{aff}}(s + \langle M, x \rangle) = \sum_{M \in \mathbb{F}_q^\ell} \sum_{\alpha \in \mathbb{F}_q^k} \widehat{f_{1,\text{aff}}}(\alpha) \chi_\alpha(s + \langle M, x \rangle) = \sum_{M \in \mathbb{F}_q^\ell} \sum_{\alpha \in \mathbb{F}_q^k} \widehat{f_{1,\text{aff}}}(\alpha) \chi_{(\alpha, M \alpha)}(s, x). \quad (3)$$

Finally, we have by definition that

$$F_1(s, x) = \sum_{\substack{\alpha: \alpha_0 \in \text{span}(\alpha_1, \dots, \alpha_\ell) \\ \dim(\text{span}(\alpha_1, \dots, \alpha_\ell))=1}} \widehat{F}(\alpha) \chi_\alpha(s, x) + \sum_{\alpha_0 \in \mathbb{F}_q^k \setminus \{0\}} \widehat{F}(\alpha_0, 0, \dots, 0) \chi_{\alpha_0, 0, \dots, 0}(s, x).$$

Expanding the first sum and using Claim B.1 we get it is equal to

$$\sum_{\alpha_0 \in \mathbb{F}_q^k \setminus \{0\}, M \in \mathcal{M}} \widehat{F}(0, M \alpha_0) \chi_{0, M \alpha_0}(s, x) + \sum_{\alpha_0 \in \mathbb{F}_q^k \setminus \{0\}, M \in \mathcal{M}} \widehat{F}(\alpha_0, M \alpha_0) \chi_{\alpha_0, M \alpha_0}(s, x).$$

Hence

$$F_1(s, x) = \sum_{\substack{\alpha_0 \in \mathbb{F}_q^k \setminus \{0\} \\ M \in \mathcal{M}}} \widehat{F}(0, M \alpha_0) \chi_{0, M \alpha_0}(s, x) + \sum_{\substack{\alpha_0 \in \mathbb{F}_q^k \setminus \{0\} \\ M \in \mathbb{F}_q^\ell}} \widehat{F}(\alpha_0, 0, \dots, 0) \chi_{\alpha_0, 0, \dots, 0}(s, x). \quad (4)$$

We plug in (2), (3), (4) into (1) and equate coefficients to get the statement of the claim.  $\square$

**Corollary C.4.** *Suppose  $S^*$  is  $\xi$ -pseudo-random against zoom-out as well as with respect to the linear part. Then for all  $\alpha$ ,*

$$\left| \widehat{f_{1,\text{lin}}}(\alpha) \right| \leq \frac{1}{(q-1)(q^\ell-1)} \xi, \quad \left| \widehat{f_{1,\text{aff}}}(\alpha) \right| \leq \frac{\xi}{q^{\ell+1}}.$$

*Proof.* We begin with the first inequality. From Claim C.3 we have

$$\widehat{f_{1,\text{lin}}}(\alpha) = \frac{1}{q-1} \mathbb{E}_{s,x} [F(s,x) \chi_{0,\alpha,\dots,\alpha}(s,x)] = \frac{1}{q-1} \mathbb{E}_{s,x, M \in \mathcal{M}} [F(s, Mx) \chi_{0,\alpha,\dots,\alpha}(s, Mx)].$$

Using the symmetries of  $F$  we have that this is equal to

$$\frac{1}{q-1} \mathbb{E}_{s,x} \left[ F(s,x) \mathbb{E}_{M \in \mathcal{M}} [\chi_{0,\alpha,\dots,\alpha}(s, Mx)] \right] = \frac{1}{(q-1)(q^\ell-1)} \mathbb{E}_{s,x} \left[ F(s,x) \left( \sum_{M \in \mathbb{F}_q^\ell} \chi_{0,\alpha,\dots,\alpha}(s, Mx) - 1 \right) \right],$$

where in the last transition we turned expectation into sum and added/subtracted  $M = 0$ . Note that the sum over  $M$  is  $q^\ell$  if  $\langle x_i, \alpha \rangle = 0$  for all  $i$  and 0 otherwise, so the last expression is equal to

$$\frac{1}{(q-1)(q^\ell-1)} \left( \mathbb{E}_{s,x} [F(s,x) q^\ell \mathbf{1}_{\text{span}(x) \subseteq W_\alpha}] - \mu(S^*) \right),$$

where  $W_\alpha$  is the subspace  $\{z \in \mathbb{F}_q^k \mid \langle z, \alpha \rangle = 0\}$ . This is equal to

$$\frac{1}{(q-1)(q^\ell-1)} (\mu((S^*)_{W_\alpha, \text{lin}}) - \mu(S^*)).$$

The result now follows from the pseudo-randomness of  $S^*$  with respect to zoom-outs.

We now move on to the second inequality. For  $\alpha \in \mathbb{F}_q^k \setminus \{0\}$  and  $j \in \mathbb{F}_q$ , denote

$$W_{\alpha,j} = \left\{ z \in \mathbb{F}_q^k \mid \langle z, \alpha \rangle = j \right\}.$$

Then

$$\begin{aligned} \mathbf{1}_{z \in W_{\alpha,j}} &= \sum_{v \in \mathbb{F}_q} \chi_v(\langle z, \alpha \rangle - j) = \sum_{v \in \mathbb{F}_q} \chi_v(-j) \chi_v(\langle z, \alpha \rangle) = \sum_{v \in \mathbb{F}_q} \chi_v(-j) \omega^{\text{Tr}(v(z_1 \alpha_1 + \dots + z_k \alpha_k))} \\ &= \sum_{v \in \mathbb{F}_q} \chi_v(-j) \prod_{i=1}^k \omega^{\text{Tr}(v z_i \alpha_i)} \\ &= \sum_{v \in \mathbb{F}_q} \chi_v(-j) \chi_{v\alpha}(z). \end{aligned}$$

We now invert this formula. We multiply this equality by  $\chi_1(j)$  and average over  $j$  to get that

$$\chi_\alpha(z) = \frac{1}{q} \sum_{j \in \mathbb{F}_q} \chi_1(j) \mathbf{1}_{z \in W_{\alpha,j}}, \tag{5}$$

and we use this equality to establish the second inequality of the lemma.

$$\begin{aligned} \widehat{f_{1,\text{aff}}}(\alpha) &= \widehat{F}(\alpha, 0, \dots, 0) = \mathbb{E}_{s,x} [F(s,x) \chi_\alpha(s)] = \mathbb{E}_{s,x, M} [F(s - Mx, x) \chi_\alpha(s)] \\ &= \mathbb{E}_{s,x} \left[ F(s,x) \mathbb{E}_M [\chi_\alpha(s + Mx)] \right] \\ &= \mathbb{E}_{s,x} \left[ F(s,x) \chi_\alpha(s) \mathbb{E}_M [\chi_\alpha(Mx)] \right]. \end{aligned}$$

As before, the expectation over  $M$  is 1 if  $\langle \alpha, x_i \rangle = 0$  for all  $i$  and 0 otherwise, so

$$\widehat{f_{1,\text{aff}}}(\alpha) = \mathbb{E}_{s,x} [F(s,x)\chi_\alpha(s)\mathbf{1}_{\text{span}(x)\subseteq W_{\alpha,0}}].$$

Plugging in (5) now yields

$$\begin{aligned} \widehat{f_{1,\text{aff}}}(\alpha) &= \mathbb{E}_{j\in\mathbb{F}_q} \left[ \chi_1(j) \mathbb{E}_{s,x} [F(s,x)\mathbf{1}_{s\in W_{\alpha,j}}\mathbf{1}_{\text{span}(x)\subseteq W_{\alpha,0}}] \right] = \mathbb{E}_{j\in\mathbb{F}_q} \left[ \chi_1(j) \mathbb{E}_{s,x} [F(s,x)\mathbf{1}_{s+\text{span}(x)\subseteq W_{\alpha,j}}] \right] \\ &= \mathbb{E}_{j\in\mathbb{F}_q} \left[ \chi_1(j)q^{-(\ell+1)}\mu((S^*)_{W_{\alpha,j}}) \right]. \end{aligned}$$

Taking absolute value, applying the triangle inequality and using the pseudo-randomness of  $S^*$  finishes the proof.  $\square$

## D Proof of Lemma B.5

In this section we prove Lemma B.5. The proof proceeds by opening up the 4-norm and upper bounding different terms in an appropriate way. Write

$$g(s,x) = \sum_{M\in\mathcal{B}} f_{1,\text{lin}}(\langle M,x \rangle), \quad h(s,x) = \sum_{M\in\mathbb{F}_q^\ell} f_{1,\text{aff}}(s + \langle M,x \rangle).$$

Clearly

$$F_1(s,x)^4 = g(s,x)^4 + 4g(s,x)^3h(s,x) + 6g(s,x)^2h(s,x)^2 + 4g(s,x)h(s,x)^3 + h(s,x)^4, \quad (6)$$

and we prove that the expectation of all but the last term is very small. As we will see, it is enough for us to upper bound the expectation of  $g(s,x)^4$  and  $h(s,x)^4$ , but we remark that it is possible to directly analyze each one of these terms separately in order to establish better bounds.

**Claim D.1.**  $\mathbb{E}_{s,x} [g(s,x)^4] \leq \xi^2\mu(S^*) + 4(q-1)^2\xi\mu(S^*) + 24\frac{\xi^2}{(q-1)^2}\mu(S^*)$ . In particular, we have that  $\mathbb{E}_{s,x} [g(s,x)^4] \leq 30q^2\xi\mu(S^*)$ .

*Proof.* We open up according to the definition of  $g(s,x)$ :

$$\begin{aligned} g(s,x)^4 &= \sum_{M_1,M_2,M_3,M_4\in\mathcal{B}} f_{1,\text{lin}}(\langle M_1,x \rangle) \cdots f_{1,\text{lin}}(\langle M_4,x \rangle) \\ &= \frac{1}{(q-1)^4} \sum_{M_1,M_2,M_3,M_4\in\mathcal{M}} f_{1,\text{lin}}(\langle M_1,x \rangle) \cdots f_{1,\text{lin}}(\langle M_4,x \rangle). \end{aligned}$$

We partition the last sum according to  $\dim(\text{span}(M_1, \dots, M_4))$ . Denote by  $H_i$  the collection of  $(M_1, \dots, M_4)$  for which this dimension is  $i$ .

**The contribution from  $H_1$ .** Note that the summands corresponding to  $H_1$  may be written as

$$\frac{1}{(q-1)^4} \sum_{M_1 \in \mathcal{M}, M_2, M_3, M_4 \in \text{span}(M_1) \setminus \{0\}} f_{1,\text{lin}}(\langle M_1, x \rangle) \cdots f_{1,\text{lin}}(\langle M_4, x \rangle) = \frac{1}{q-1} \sum_{M \in \mathcal{M}} f_{1,\text{lin}}(\langle M, x \rangle)^4.$$

Taking expectation over  $x$  we get that the contribution from  $H_1$  is at most

$$\frac{1}{q-1} |\mathcal{M}| \mathbb{E}_z [f_{1,\text{lin}}(\langle M_1, x \rangle)^4] \leq |\mathcal{B}| \|f_{1,\text{lin}}\|_\infty^2 \|f_{1,\text{lin}}\|_2^2.$$

Using Claim C.2 we bound  $\|f_{1,\text{lin}}\|_2^2 \leq \frac{\mu(S^*)}{|\mathcal{B}|}$ , and using the  $\xi$  pseudo-randomness of  $S^*$  with respect to zoom ins on the linear part we have  $\|f_{1,\text{lin}}\|_\infty \leq \xi$ , so the contribution from  $H_1$  is at most  $\xi^2 \mu(S^*)$ .

**The contribution from  $H_2$ .** There are two cases. Either we can partition  $M_1, M_2, M_3, M_4$  into two sets, such that the dimension of the space spanned by each one is 2, or we cannot. The contribution of the first type is at most

$$\begin{aligned} & \frac{1}{(q-1)^4} \sum_{\substack{M_1, M_2 \in \mathcal{M} \text{ linearly ind} \\ M_3, M_4 \in \text{span}(M_1, M_2) \text{ linearly ind}}} |f_{1,\text{lin}}(\langle M_1, x \rangle) \cdots f_{1,\text{lin}}(\langle M_4, x \rangle)| \\ & \leq \frac{2}{(q-1)^4} \sum_{\substack{M_1, M_2 \in \mathcal{M} \text{ linearly ind} \\ M_3, M_4 \in \text{span}(M_1, M_2) \text{ linearly ind}}} |f_{1,\text{lin}}(\langle M_1, x \rangle) f_{1,\text{lin}}(\langle M_2, x \rangle)|^2 + |f_{1,\text{lin}}(\langle M_3, x \rangle) f_{1,\text{lin}}(\langle M_4, x \rangle)|^2. \end{aligned}$$

Taking expectation, the contribution from  $H_2$  is at most

$$4 \sum_{M_1, M_2 \in \mathcal{M} \text{ linearly independent}} \mathbb{E}_{s,x} \left[ |f_{1,\text{lin}}(\langle M_1, x \rangle) f_{1,\text{lin}}(\langle M_2, x \rangle)|^2 \right].$$

As  $\langle M_1, x \rangle$  and  $\langle M_2, x \rangle$  are independently uniformly distributed in  $\mathbb{F}_q^k$ , we get that the last expression is

$$4 |\mathcal{M}| \|f_{1,\text{lin}}\|_2^4 \leq 4 |\mathcal{M}| \left( \frac{\mu(S^*)}{|\mathcal{B}|} \right)^2 \leq 4(q-1)^2 \xi \mu(S^*),$$

where we used Claim C.2.

The contribution of the second type is a multiple of

$$\frac{1}{(q-1)^4} \sum_{M_1, M_2 \in \mathcal{M} \text{ linearly independent}} f_{1,\text{lin}}(\langle M_1, x \rangle) f_{1,\text{lin}}(\langle M_2, x \rangle) f_{1,\text{lin}}(\langle M_1, x \rangle) f_{1,\text{lin}}(\langle M_1, x \rangle),$$

and taking expectation the contribution of this type is proportional to

$$\frac{1}{(q-1)^4} \sum_{M_1, M_2 \in \mathcal{M} \text{ linearly ind}} \mathbb{E}_{s,x} [f_{1,\text{lin}}(\langle M_1, x \rangle)^3 f_{1,\text{lin}}(\langle M_2, x \rangle)],$$

which is equal to 0 as  $\langle M_1, x \rangle$  and  $\langle M_2, x \rangle$  are uniform and independent in  $\mathbb{F}_q^k$ , and the expectation of  $f_{1,\text{lin}}(\langle M_2, x \rangle)$  is 0 by Claim C.1.

**The contribution from  $H_3$ .** The contribution of this case is a constant multiple, not more than  $4!$ , of

$$\frac{1}{(q-1)^4} \sum_{\substack{M_1, M_2, M_3 \in \mathcal{M} \text{ linearly ind} \\ M_4 \in \text{span}(M_1, M_2, M_3)}} \mathbb{E}_{s,x} [f_{1,\text{lin}}(\langle M_1, x \rangle) f_{1,\text{lin}}(\langle M_2, x \rangle) f_{1,\text{lin}}(\langle M_3, x \rangle) f_{1,\text{lin}}(\langle M_4, x \rangle)].$$

If  $M_4 \in \text{span}(M_1, M_2)$ , the contribution is shown to be 0 as in the second type in the analysis of  $H_2$ . Otherwise, we get

$$\frac{1}{(q-1)^4} \sum_{\substack{M_1, M_2, M_3 \in \mathcal{M} \\ \text{linearly ind} \\ j_1, j_2, j_3 \in \mathbb{F}_q \setminus \{0\}}} \mathbb{E}_{s,x} [f_{1,\text{lin}}(\langle M_1, x \rangle) f_{1,\text{lin}}(\langle M_2, x \rangle) f_{1,\text{lin}}(\langle M_3, x \rangle) f_{1,\text{lin}}(\langle j_1 M_1 + j_2 M_2 + j_3 M_3, x \rangle)].$$

Taking expectation, we get that the contribution is proportional to

$$\frac{1}{(q-1)^4} |\{M_1, M_2, M_3 \in \mathcal{M} \text{ linearly ind}\}| \sum_{j_1, j_2, j_3 \in \mathbb{F}_q \setminus \{0\}} \mathbb{E}_{u,v,w} [f_{1,\text{lin}}(u) f_{1,\text{lin}}(v) f_{1,\text{lin}}(w) f_{1,\text{lin}}(j_1 u + j_2 v + j_3 w)].$$

Taking the proportionality constant into consideration, and taking  $j_1, j_2, j_3$  that maximize this expectation, the contribution from  $H_3$  is at most

$$\frac{4!}{q-1} |\mathcal{M}|^3 \left| \mathbb{E}_{u,v,w} [f_{1,\text{lin}}(u) f_{1,\text{lin}}(v) f_{1,\text{lin}}(w) f_{1,\text{lin}}(j_1 u + j_2 v + j_3 w)] \right|, \quad (7)$$

and to upper bound the last expectation we move to the Fourier domain. A straightforward computation shows that

$$\begin{aligned} & \left| \mathbb{E}_{u,v,w} [f_{1,\text{lin}}(u) f_{1,\text{lin}}(v) f_{1,\text{lin}}(w) f_{1,\text{lin}}(j_1 u + j_2 v + j_3 w)] \right| \\ &= \left| \sum_{\alpha} \widehat{f_{1,\text{lin}}}(-j_1 \alpha) \widehat{f_{1,\text{lin}}}(-j_2 \alpha) \widehat{f_{1,\text{lin}}}(-j_3 \alpha) \widehat{f_{1,\text{lin}}}(\alpha) \right|. \end{aligned}$$

Using Claim C.4 we get that this is at most

$$\begin{aligned} \frac{\xi^2}{(q-1)^2 (q^\ell - 1)^2} \sum_{\alpha} \left| \widehat{f_{1,\text{lin}}}(-j_3 \alpha) \widehat{f_{1,\text{lin}}}(\alpha) \right| &\leq \frac{\xi^2}{(q-1)^2 (q^\ell - 1)^2} \sum_{\alpha} \left| \widehat{f_{1,\text{lin}}}(\alpha) \right|^2 \\ &= \frac{\xi^2}{(q-1)^2 (q^\ell - 1)^2} \|f_{1,\text{lin}}\|_2^2 \\ &\leq \frac{\xi^2 \mu(S^*)}{(q-1)(q^\ell - 1)^3}, \end{aligned}$$

where in the last transition we used Claim C.2. Plugging this into (7) yields that the contribution from  $H_3$  is at most

$$24 \frac{\xi^2}{(q-1)^2} \mu(S^*).$$



**The contribution from  $H_4$ .** This is shown to be 0 similarly to the second type in the analysis of  $H_2$ .  $\square$

Next, we upper bound the expectation of  $h(s, x)^4$ .

**Claim D.2.** *We have*

$$\mathbb{E}_{s,x} [h(s, x)^4] \leq \mu(S^*) \|f_{1,\text{aff}}\|_\infty^2 + 32\xi\mu(S^*) + \xi^2 q\mu(S^*).$$

*In particular:*

1.  $\mathbb{E}_{s,x} [h(s, x)^4] \leq \mu(S^*) \|f_{1,\text{aff}}\|_\infty^2 + 33q\xi\mu(S^*);$
2. *and weakening further,*  $\mathbb{E}_{s,x} [h(s, x)^4] \leq 34q\mu(S^*).$

*Proof.* We open up according to the definition of  $h(s, x)$ :

$$\mathbb{E}_{s,x} [h(s, x)^4] = \sum_{M_1, M_2, M_3, M_4 \in \mathbb{F}_q^\ell} \mathbb{E}_{s,x} [f_{1,\text{aff}}(s + \langle M_1, x \rangle) \cdots f_{1,\text{aff}}(s + \langle M_4, x \rangle)].$$

We make the change of variables  $s \leftarrow s + \langle M_1, x \rangle$  and get that

$$\begin{aligned} & \mathbb{E}_{s,x} [h(s, x)^4] \\ &= \sum_{M_1, M_2, M_3, M_4 \in \mathbb{F}_q^\ell} \mathbb{E}_{s,x} [f_{1,\text{aff}}(s) f_{1,\text{aff}}(s + \langle M_2 - M_1, x \rangle) f_{1,\text{aff}}(s + \langle M_3 - M_1, x \rangle) f_{1,\text{aff}}(s + \langle M_4 - M_1, x \rangle)]. \end{aligned}$$

We partition the last sum according to  $\dim(\text{span}(M_2 - M_1, M_3 - M_1, M_4 - M_1))$ . For  $i = 0, \dots, 3$  denote by  $H_i$  the collection of  $(M_1, \dots, M_4)$  for which this dimension is  $i$ .

**The contribution from  $H_0$ .** The contribution here is

$$\sum_{M_1 \in \mathbb{F}_q^\ell} \mathbb{E}_{s,x} [f_{1,\text{aff}}(s)^4] \leq q^\ell \|f_{1,\text{aff}}\|_\infty^2 \|f_{1,\text{aff}}\|_2^2.$$

Using Claim C.2, this is upper bounded by  $\mu(S^*) \|f_{1,\text{aff}}\|_\infty^2$ .

**The contribution from  $H_1$**  There are three subcases we consider. Either there are two differences, say  $M_2 - M_1, M_3 - M_1$  which are 0, in which case the contribution is

$$\sum_{M_1, M_4 \in \mathbb{F}_q^\ell} \mathbb{E}_{s,x} [f_{1,\text{aff}}(s)^3 f_{1,\text{aff}}(s + \langle M_4 - M_1, x \rangle)].$$

The points  $s$  and  $s + \langle M_4 - M_1, x \rangle$  are jointly distributed uniformly on  $\mathbb{F}_q^k$ , so the expectation above may be broken into the product of two expectation, and the expectation of  $f_{1,\text{aff}}(s + \langle M_4 - M_1, x \rangle)$  is 0 by Claim C.1. Hence, the contribution of this sub-case is 0.

In the second subcase,  $M_2 - M_1 = M_3 - M_1 = M_4 - M_1$ , and the contribution here is 0 just like in the previous subcase. In the last subcase, we consider  $M_1, M_2, M_3$  that maximize the absolute value of the expectation and upper bound the contribution as

$$q^{2\ell} q^2 \left| \mathbb{E}_{s,x} [f_{1,\text{aff}}(s) f_{1,\text{aff}}(s + \langle M_2 - M_1, x \rangle) f_{1,\text{aff}}(s + \langle M_3 - M_1, x \rangle) f_{1,\text{aff}}(s + \langle M_4 - M_1, x \rangle)] \right|.$$

1. If one of the differences is 0, say  $M_2 - M_1 = 0$ , then we conclude that  $M_3 - M_1$  and  $M_4 - M_1$  are difference (otherwise we would have been in a previous subcase), and the contribution here is at most

$$\begin{aligned} & q^{2\ell} q^2 \left| \mathbb{E}_{s,x} [f_{1,\text{aff}}(s)^2 f_{1,\text{aff}}(s + \langle M_3 - M_1, x \rangle) f_{1,\text{aff}}(s + \langle M_4 - M_1, x \rangle)] \right| \\ & \leq 2 \cdot q^{2\ell} q^2 \left| \mathbb{E}_{s,x} [f_{1,\text{aff}}(s)^2 f_{1,\text{aff}}(s + \langle M_3 - M_1, x \rangle)^2] + \mathbb{E}_{s,x} [f_{1,\text{aff}}(s)^2 f_{1,\text{aff}}(s + \langle M_4 - M_1, x \rangle)^2] \right|. \end{aligned}$$

Each one of these expectations is equal to  $\|f_{1,\text{aff}}\|_2^4$ , so we get an upper bound of

$$4 \cdot q^{2\ell} q^2 \|f_{1,\text{aff}}\|_2^4 \leq 4 \cdot q^{2\ell} q^2 \left( \frac{\mu(S^*)}{q^\ell} \right)^2 \leq 4q^2 \xi \mu(S^*),$$

where we used Claim C.2.

2. Otherwise, all three differences are non 0 and at least two are different, say  $M_3 - M_1 \neq M_4 - M_1$ . We thus bound the contribution by

$$\begin{aligned} & q^{2\ell} q^2 \left| \mathbb{E}_{s,x} [f_{1,\text{aff}}(s) f_{1,\text{aff}}(s + \langle M_2 - M_1, x \rangle) \right. \\ & \quad \left. \cdot f_{1,\text{aff}}(s + \langle M_3 - M_1, x \rangle) f_{1,\text{aff}}(s + \langle M_4 - M_1, x \rangle)] \right| \\ & \leq 2q^{2\ell} q^2 \left| \mathbb{E}_{s,x} [f_{1,\text{aff}}(s)^2 f_{1,\text{aff}}(s + \langle M_2 - M_1, x \rangle)^2 \right. \\ & \quad \left. + f_{1,\text{aff}}(s + \langle M_3 - M_1, x \rangle)^2 f_{1,\text{aff}}(s + \langle M_4 - M_1, x \rangle)^2] \right|. \end{aligned}$$

The last expectation is equal to  $2\|f_{1,\text{aff}}\|_2^4$ , so we get contribution of  $4q^{2\ell} q^2 \left( \frac{\mu(S^*)}{q^\ell} \right)^2 \leq 4q^2 \xi \mu(S^*)$ .

**The contribution from  $H_2$ .** Let  $M_1, M_2, M_3, M_4$  that maximize this case. Then we need to bound

$$q^{3\ell} q^3 \left| \mathbb{E}_{s,x} [f_{1,\text{aff}}(s) f_{1,\text{aff}}(s + \langle M_2 - M_1, x \rangle) f_{1,\text{aff}}(s + \langle M_3 - M_1, x \rangle) f_{1,\text{aff}}(s + \langle M_4 - M_1, x \rangle)] \right|.$$

Suppose without loss of generality  $M_2 - M_1, M_3 - M_1$  constitute a basis for  $\text{span}(M_2 - M_1, M_3 - M_1, M_4 - M_1)$ . Let  $j_3, j_2$  be such that  $M_4 - M_1 = j_3(M_3 - M_1) + j_2(M_2 - M_1)$ , and make the change of variables  $u = s + \langle M_2 - M_1, x \rangle, w = s + \langle M_3 - M_1, x \rangle$  and note that  $(s, u, w)$  are distributed uniformly on  $(\mathbb{F}_q^k)^3$ . Thus, the above expectation is

$$\mathbb{E}_{s,u,w} [f_{1,\text{aff}}(s) f_{1,\text{aff}}(u) f_{1,\text{aff}}(w) f_{1,\text{aff}}((1 - j_3 - j_2)s + j_2u + j_3w)].$$

If  $j_2 = 0, j_3 = 0$  or  $j_2 + j_3 = 1$ , then this expectation is 0. Indeed, say  $j_2 = 0$ , then  $u$  only appears in the second term and is thus independent of the rest, and by Claim C.1 its expectation is 0. We thus assume otherwise, and move to the Fourier domain. A straightforward computation shows that

$$\sum_{\alpha} \widehat{f_{1,\text{aff}}}((j_2 + j_3 - 1)\alpha) \widehat{f_{1,\text{aff}}}(-j_2\alpha) \widehat{f_{1,\text{aff}}}(-j_3\alpha) \widehat{f_{1,\text{aff}}}(\alpha).$$

Taking absolute value, the absolute value of this sum is at most

$$\|\widehat{f_{1,\text{aff}}}\|_\infty^2 \left| \sum_\alpha \widehat{f_{1,\text{aff}}}(-j_3\alpha) \widehat{f_{1,\text{aff}}}(\alpha) \right| \leq \|\widehat{f_{1,\text{aff}}}\|_\infty^2 \sum_\alpha \widehat{f_{1,\text{aff}}}(\alpha)^2 \leq \|\widehat{f_{1,\text{aff}}}\|_\infty^2 \|\widehat{f_{1,\text{aff}}}\|_2^2.$$

Using Claim C.2 and Corollary C.4 we may bound this by  $\frac{\xi^2}{q^{3\ell+2}}\mu(S^*)$ , and plugging this above we get that the contribution from  $H_2$  is at most

$$q^{3\ell} q^3 \frac{\xi^2}{q^{3\ell+2}} \mu(S^*) = \xi^2 q \mu(S^*).$$

**The contribution from  $H_3$ .** In this case, the joint distribution of  $s, s + \langle M_2 - M_1, x \rangle, s + \langle M_3 - M_1, x \rangle, s + \langle M_4 - M_1, x \rangle$  is uniform over  $(\mathbb{F}_q^k)^4$ , so the contribution is 0 by Claim C.1.  $\square$

**Claim D.3.**  $\mathbb{E}_{s,x} [4g(s,x)^3 h(s,x) + 6g(s,x)^2 h(s,x)^2 + 4g(s,x) h(s,x)^3] \leq 800q^2 \xi^{1/4} \mu(S^*)$ .

*Proof.* Using Holder's inequality, we have

$$\mathbb{E}_{s,x} [4g(s,x)^3 h(s,x) + 6g(s,x)^2 h(s,x)^2 + 4g(s,x) h(s,x)^3] \leq 4\|g\|_4^3 \|h\|_4 + 6\|g\|_4^2 \|h\|_4^2 + 4\|g\|_4 \|h\|_4^3.$$

Use Claim D.1 and the second item of Claim D.2 to bound each term on the right hand side, we get that it is at most

$$4(30q^2 \xi \mu(S^*))^{3/4} (34q\mu(S^*))^{1/4} + 6(30q^2 \xi \mu(S^*))^{1/2} (34q\mu(S^*))^{1/2} + 4(30q^2 \xi \mu(S^*))^{1/4} (34q\mu(S^*))^{3/4}.$$

Further upper bounding this we get it is at most

$$14 \cdot 34q^2 \xi^{1/4} \mu(S^*) \leq 800q^2 \xi^{1/4} \mu(S^*). \quad \square$$

We are now ready to prove Lemma B.5.

*Proof of Lemma B.5.* Take expectation over (6) and use Claims D.1, D.2 (first item) and D.3 to get that

$$\|F_1\|_4^4 \leq 30q^2 \xi \mu(S^*) + 800q^2 \xi^{1/4} \mu(S^*) + \mu(S^*) \|f_{1,\text{aff}}\|_\infty^2 + 33q \xi \mu(S^*),$$

which implies

$$\|F_1\|_4^4 \leq 863q^2 \xi^{1/4} \mu(S^*) + \mu(S^*) \|f_{1,\text{aff}}\|_\infty^2.$$

Finally, note that  $\|f_{1,\text{aff}}\|_\infty \leq a$ , so we conclude that

$$\|F_1\|_4^4 \leq 863q^2 \xi^{1/4} \mu(S^*) + a^2 \mu(S^*). \quad \square$$