

# Testing Polynomials over General Fields\*

Tali Kaufman<sup>†</sup>

School of Computer Science

Tel Aviv University

Tel Aviv 69978 Israel

E-mail: kaufmant@post.tau.ac.il

Dana Ron<sup>‡</sup>

Department of Electrical Engineering-Systems

Tel Aviv University

Tel Aviv 69978, Israel

E-mail: danar@eng.tau.ac.il

## Abstract

In this work we fill in the knowledge gap concerning testing polynomials over finite fields. As previous works show, when the cardinality of the field,  $q$ , is sufficiently larger than the degree bound,  $d$ , then the number of queries sufficient for testing is polynomial or even linear in  $d$ . On the other hand, when  $q = 2$  then the number of queries, both sufficient and necessary, grows exponentially with  $d$ .

Here we study the intermediate case where  $2 < q = O(d)$  and show a smooth transition between the two extremes. Specifically, let  $p$  be the characteristic of the field (so that  $p$  is prime and  $q = p^s$  for some integer  $s \geq 1$ ). Then the number of queries performed by the test grows like  $\ell \cdot q^{2\ell+1}$ , where  $\ell = \left\lceil \frac{d+1}{q-q/p} \right\rceil$ . Furthermore,  $q^{\Omega(\ell)}$  queries are necessary when  $q = O(d)$ . The test itself provides a unifying view of the two extremes: it considers random affine subspaces of dimension  $\ell$  and verifies that the function restricted to the selected subspaces is a polynomial of degree at most  $d$ .

Viewed in the context of coding theory, our result shows that Reed-Muller codes over general fields (usually referred to as *Generalized Reed-Muller (GRM) codes*) are locally testable. In the course of our analysis we provide a characterization of small-weight words that span the code. Such a characterization was previously known only when the field size is a prime or is sufficiently large, in which case the minimum weight words span the code.

---

\*An extended abstract of this work will appear in the proceedings of the Forty-Fifth Annual Symposium on Foundations of Computer Science (FOCS) 2004.

<sup>†</sup>This work is part of the author's Ph.D. thesis prepared at Tel Aviv University under the supervision of Prof. Noga Alon, and Prof. Michael Krivelevich. The research was performed in part while visiting the Radcliffe institute of advanced study at Harvard.

<sup>‡</sup>This research was done during a fellowship year at the Radcliffe institute of advanced study at Harvard.

# 1 Introduction

In this paper we consider the problem of testing, for a given finite field  $F$  and degree-bound  $d$ , whether a function  $f : F^n \rightarrow F$  is a multivariate polynomial of total degree at most  $d$  over  $F$ . Specifically, the testing algorithm is given query access to  $f$  and a parameter  $\epsilon > 0$ . If  $f$  is a polynomial of degree at most  $d$  then the testing algorithm must accept. On the other hand, if  $f$  differs from every such polynomial on more than an  $\epsilon$ -fraction of the domain elements, then the test should reject with probability at least  $2/3$ .

The problem of testing multivariate low-degree polynomials over finite fields has been studied extensively, mainly due to its applications to Probabilistically Checkable Proofs systems (PCPs). This is true both for the special case of linear functions (degree-1 polynomials) [BLR93, BFL91, FGL<sup>+</sup>96, BGLR93, BS94, BCH<sup>+</sup>96, SW04] and for the more general case of degree- $d$  polynomials [BFL91, BFLS91, GLR<sup>+</sup>91, FGL<sup>+</sup>96, RS96, FS95, AS97]. However, all these results apply only to testing polynomials over fields that are *larger than the degree-bound*,  $d$ . In particular, when the field size  $|F|$  is at least  $c \cdot d$ , for some sufficiently large constant  $c$ , then a number of queries that is linear in  $d$  is sufficient [PS94, FS95], and when  $d+2 \leq |F| < c \cdot d$  then the dependence on  $d$  is known to be polynomial [FS95, RS96]. In recent work, Alon et al. [AKK<sup>+</sup>03] studied the same property for the case  $|F| = 2$  and for  $d \geq 2$ . Namely, they considered the case in which the degree-bound may be (much) larger than the field size, but their results hold only for  $F = \text{GF}(2)$ . They showed that the number of queries both necessary and sufficient in this case is *exponential* in  $d$ . Hence we encounter a very large gap in terms of the dependence on  $d$  between the query complexity when  $|F| > d$  and the query complexity when  $|F| = 2$ .

**Our Main Result.** In this work we bridge the gap between the two cases mentioned above and show a smooth transition between them. In particular, we describe and analyze a testing algorithm for polynomials of degree at most  $d$  over finite fields of cardinality  $q$  where  $2 \leq q = O(d)$ . The test performs  $O(\ell \cdot q^{2\ell+1} + 1/\epsilon)$  queries, where for prime  $q$ ,  $\ell = \left\lceil \frac{d+1}{q-1} \right\rceil$ , and more generally, when  $q$  is a power of a prime  $p$  then  $\ell = \left\lceil \frac{d+1}{q-q/p} \right\rceil$ . Thus, as we increase the field size  $q$ , the dependence on  $d$  decreases from being exponential, to being polynomial. We note that this query complexity (when  $q = O(d)$ ) is almost tight: for prime fields (and constant  $\epsilon$ )  $\Omega(q^{\ell-1})$  queries are necessary, and for non-prime fields  $\Omega(q^{\lceil \ell/2 \rceil - 1})$  queries are necessary. As we discuss in more detail subsequently, the “gap phenomenon” that we observe, is not unique to testing polynomials: analogous gaps arise in other property testing problems.

**Characterization of Degree- $d$  Polynomials over  $\text{GF}(q)$ .** One of the building blocks of our analysis is a characterization of (total) degree- $d$  multivariate polynomials over finite fields. In particular, we show that for  $F = \text{GF}(q)$  where  $q = p^s$  and  $p$  is prime, a function  $f : F^n \rightarrow F$  is a polynomial of degree at most  $d$ , if and only if its restriction to every affine subspace of dimension  $\ell = \left\lceil \frac{d+1}{q-q/p} \right\rceil$  is a polynomial of degree at most  $d$ . This generalizes the characterization result of Friedl and Sudan [FS95] that refers to the case  $q - q/p \geq d+1$  (that is,  $\ell = 1$ ). We also note that this value,  $\ell$ , of the dimension of the considered subspaces, is tight. Namely, there exist polynomials of degree greater than  $d$  whose restrictions to affine subspaces of dimension less than  $\ell$  are all degree- $d$  polynomials.

**A Unifying Approach to Testing Low-Degree Polynomials.** The testing algorithm presented in this work utilizes the above characterization (which is shown to be *robust* in the sense defined in [RS96]). Specifically, the algorithm selects random affine subspaces (of dimension  $\ell$  as defined above), and checks that the restriction of the input function  $f$  to each of the selected subspaces is indeed a polynomial of degree at most  $d$ . Such a check is implemented by verifying that various linear combinations of the values of  $f$  on the subspace sum to 0. Observe that when the size of the field  $F$  is sufficiently larger than the degree bound  $d$  then  $\ell = 1$ . That is, when the field is sufficiently large, then the algorithm checks whether the univariate polynomials that correspond to restrictions of the function  $f$  to random *lines* in  $F^n$  all have degree at most  $d$ . But this is essentially the original low-degree test of Rubinfeld and Sudan [RS96].<sup>1</sup>

On the other hand, when  $q = 2$  then the test in [AKK<sup>+</sup>03] works by uniformly selecting  $\frac{d+1}{q-1} = d+1$  points in  $\text{GF}(2)^n$  and verifying that the sum of the values of  $f$  taken over all sums of subsets of these points is 0. This too can be shown to amount to checking whether the restriction of  $f$  to the  $(d+1)$ -dimensional subspace spanned by the selected points is a polynomial of degree at most  $d$ . Thus our test suggests a uniform view of all these tests for low-degree polynomials.

**Relation to Coding.** The *Generalized Reed-Muller (GRM)* code of rank  $d$  over  $\text{GF}(q)$ , which we denote by  $\mathcal{GRM}_q(d, n)$ , consists of all words of length  $q^n$  that correspond to the evaluations of degree- $d$  polynomials over  $\text{GF}(q)^n$ . (When  $q = 2$  then the code is simply referred to as *Reed-Muller (RM)*.) Hence, an equivalent view of our main result, from a coding-theory perspective, is that GRM codes are locally testable. Furthermore, our characterization of low-degree polynomials translates into a characterization of a set of small-weight words<sup>2</sup> that span the dual code (which is the GRM code  $\mathcal{GRM}_q(n(q-1) - (d+1), n)$ ). The question concerning when GRM codes are spanned by their *minimum-weight* words has been studied in the coding theory literature.<sup>3</sup> Specifically, Ding and Key [DK00] have shown that if  $q$  is prime or  $q$  is sufficiently larger than  $d$ , then the minimum-weight words of a GRM code indeed span the code, but this is not true in general. In particular, this is not true when  $q$  is not prime and  $q - q/p < d + 1$ .<sup>4</sup> We complement the result of Ding and Key by showing that a GRM code can always be spanned by words of weight that is at most quadratic in the weight of the minimum-weight words. We note that our interest in small weight words that span a GRM code is due to the way our test works. Similarly to other low-degree tests, our test can be viewed as randomly selecting small weight words from the dual code (which is a GRM code itself) and checking that each is indeed orthogonal to the tested word. Thus the weight of the selected words is an important factor in the query complexity of our algorithm.

As we show in Section 7 our analysis implies certain sufficient conditions for local testability of codes. The first, basic, condition, is that the dual code of the tested code can be spanned by a subset  $Q$ , of small weight words. We further require that the words in  $Q$  have a certain *closure* property: for every word  $w \in Q$  if we apply particular transformations on it, then we always get another word in  $Q$ . Finally, we require that there be “short” dependencies between words in  $Q$ .

---

<sup>1</sup>The reason we say “essentially” is that when  $|F| > d$  then it is not necessary to query  $f$  on all points on a selected line, but rather it suffices to interpolate using  $d + 1$  points and check that the resulting degree- $d$  polynomial agrees with a random point on the line.

<sup>2</sup>The weight of a word is simply the number of its non-zero symbols.

<sup>3</sup>The minimum-weight words of  $\mathcal{GRM}_q(d, n)$  have weight  $(q-t)q^{n-r-1}$  where  $r(q-1) + t = d$  and  $0 \leq t < q-1$ , and the points in the support of each such word belong to an affine subspace of dimension  $(n-r)$ .

<sup>4</sup>To be precise, there are two special cases in which the minimum weight words *do* span the code though  $q$  is not prime and  $q - q/p < d + 1$ : the special case of  $n = 1$  (Reed Salomon codes) and the special case that  $d$  is almost the maximum possible degree  $n \cdot (q - 1)$ .

That is, words in  $Q$  (or more precisely, in some subset of  $Q$ ) can be written as a sum of few other words in  $Q$ .

**The paper of Jutla et al. [JPRZ04].** Independently from our work, Jutla, Patthak, Rudra, and Zuckerman [JPRZ04] studied the problem of testing low-degree polynomials and described a testing algorithm that has the same query complexity as our algorithm. However, their algorithm works only for prime fields. We further discuss the relation between our approaches in Section 6.

**A Broader Perspective: The Gap Phenomenon in Property Testing.** Property testing [RS96, GGR98] in general deals with distinguishing between objects that have a particular predetermined property and objects that are far from having the property. Low-degree testing clearly falls under this framework. As discussed above, for this property we encounter a huge gap in different setting of the problem. Interestingly, a similar gap phenomenon is encountered for the property of graph bipartiteness. Specifically, dense graphs can be tested for bipartiteness with complexity  $\Theta(1)$  [GGR98], while the complexity of testing bipartiteness in constant-degree graphs is  $\Theta(\sqrt{n})$  [GR02, GR99], where  $n$  is the number of vertices in the graph. Recently it has been shown [KKR03] that for this property too there is an algorithm (and matching lower bound) that gives a smooth transition between the two extremes. Additional properties for which the gap remains open are  $k$ -colorability and subgraph freeness (among others).

**Other Related Results on Locally Testing (Linear) Codes.** The problem of designing good codes that are locally testable was explicitly defined, e.g., in [FS95, RS96, Aro94]. In *good* we mean that they have high rate and large distance. This question has regained attention recently. Goldreich and Sudan [GS02] showed (by probabilistic arguments) that there exist locally testable (linear) codes over binary alphabets, with almost constant rate, and linear minimum distance. Their construction was derandomized in [BSSVW03], and further improved in [BSGH<sup>+</sup>04]. In [BSHR03] it was shown that local testing of random low-density parity-check codes, which have linear minimum distance and constant rate, requires  $\Omega(n)$  queries. In [BSS03] it was proved that there are no locally testable cyclic codes that have constant rate and linear minimum distance.

## 2 Preliminaries

Let  $F$  be a field of cardinality  $q$  and characteristic  $p$  (that is,  $q = p^s$  where  $p$  is prime). Let  $\omega \in F$  be a generator of the field  $F$ , so that  $F = \{0, \omega^1 = \omega, \omega^2, \dots, \omega^{q-2}, \omega^{q-1} = 1\}$ . We note that this order of the elements in  $F$  in which 1 is represented by  $\omega^{q-1}$  rather than  $\omega^0$  will serve us better than the more standard order  $F = \{0, \omega^0 = 1, \omega^1 = \omega, \omega^2, \dots, \omega^{q-2}\}$ . In particular, using this order, for any  $n \geq 1$  we consider a one-to-one mapping between  $[q-1]^n$  and  $F^n$  (where  $[q-1] \stackrel{\text{def}}{=} \{0, 1, \dots, q-1\}$ ). Specifically, for each  $\beta \in [q-1]^n$ , the point  $x \in F^n$  that *corresponds* to  $\beta$  is defined as follows:  $x_i = 0$  if  $\beta_i = 0$  and otherwise  $x_i = \omega^{\beta_i}$ .

Consider the standard (and unique) representation of the function  $f$  as a polynomial over  $F$  with degree at most  $q-1$  in each variable:

$$f(x) = \sum_{\alpha \in [q-1]^n} C_\alpha^f \cdot x^\alpha \quad \text{where} \quad x^\alpha = \prod_{i=1}^n x_i^{\alpha_i} \tag{1}$$

Here  $\vec{C}^f$  is the *coefficients vector* (indexed by points  $\alpha \in [q-1]^n$ ). If we view the function  $f$  as a  $q^n$ -dimensional vector  $\vec{f}$  (where for  $\beta \in [q-1]^n$ ,  $f_\beta$  is the evaluation of  $f$  on the point  $x \in F^n$  that corresponds to  $\beta$ ), then we can write Equation (1) in the following equivalent form:

$$\vec{f} = \mathcal{H}_n \cdot \vec{C}^f \quad (2)$$

Here  $\mathcal{H}_n$  is the  $q^n \times q^n$  matrix whose entries are indexed by pairs  $\beta, \alpha \in [q-1]^n$  where  $\mathcal{H}_n(\beta, \alpha)$  is the evaluation of the term  $x^\alpha$  at the point  $x \in F^n$  that corresponds to  $\beta$ . By inverting  $\mathcal{H}_n$  (which is non-singular) we can represent the coefficients vector  $\vec{C}^f$  in terms of  $\vec{f}$  as follows:

$$\vec{C}^f = \mathcal{A}_n \cdot \vec{f} \quad (3)$$

Both  $\mathcal{H}_n$  and  $\mathcal{A}_n$  can be defined recursively using tensor products:  $\mathcal{H}_n = \mathcal{H}_1 \otimes \mathcal{H}_{n-1}$  and  $\mathcal{A}_n = \mathcal{A}_1 \otimes \mathcal{A}_{n-1}$  where

$$\mathcal{H}_1 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & \omega & \omega^2 & \dots & 1 \\ 1 & \omega^2 & \omega^4 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{q-2} & \omega^{2(q-2)} & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix} \quad \mathcal{A}_1 = (-1) \cdot \begin{pmatrix} -1 & 0 & 0 & \dots & 0 \\ 0 & \omega^{-1} & \omega^{-2} & \dots & 1 \\ 0 & \omega^{-2} & \omega^{-4} & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \omega^{-(q-2)} & \omega^{-2(q-2)} & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix} \quad (4)$$

For our purposes, the important thing that should be noted is that every coefficient  $C_\alpha^f$  of the polynomial representation of  $f$  is some (easy to compute) linear combination of the values of  $f$  on different domain elements  $x \in F^n$ . In particular, for  $\alpha = \langle q-1, \dots, q-1 \rangle$ ,  $C_\alpha^f = (-1)^n \cdot \sum_{x \in F^n} f(x)$ .

**Definition 1** Let  $\text{POLY}_{n,d}$  denote the class of all functions  $f : F^n \rightarrow F$  that are polynomials of total degree at most  $d$  (where the degree in each variable is at most  $q-1$ ). Namely, if we consider the representation of  $f$  as defined in Equation (1), then  $C_\alpha^f = 0$  for every  $\alpha \in [q-1]^n$  such that  $\sum_{i=1}^n \alpha_i > d$ .

**Definition 2** For  $m \geq 1$  and any choice of  $m+1$  linearly independent points  $y_0, y_1, \dots, y_m \in F^n$ , let  $S(y_0, y_1, \dots, y_m)$  denote the affine subspace of dimension  $m$  that contains all points of the form  $y_0 + \sum_{i=1}^m a_i y_i$ , where  $a_1, \dots, a_m \in F$ . Note that  $y_0$  has a different role from all other  $y_i$ 's.

**Definition 3** For a function  $f : F^n \rightarrow F$  and linearly independent points  $y_0, y_1, \dots, y_m$ , we denote by  $f|_{(y_0, y_1, \dots, y_m)}$  the restriction of  $f$  to the affine subspace  $S(y_0, y_1, \dots, y_m)$ . Namely,  $f|_{(y_0, y_1, \dots, y_m)} : F^m \rightarrow F$  is defined as follows: for every  $v \in F^m$ ,  $f|_{(y_0, y_1, \dots, y_m)}(v) = f(y_0 + \sum_{i=1}^m v_i y_i)$ . With slight abuse of notation (and for sake of succinctness), we shall sometimes use the notation  $f|_S$  instead, where  $S = S(y_0, y_1, \dots, y_m)$  is the subspace spanned by the points. In case the set of points spanning the subspace  $S$  is not explicitly stated, then  $f|_S$  is determined by some canonical choice of a basis.<sup>5</sup>

**Definition 4** For any two functions  $f, g : F^n \rightarrow F$ , let  $\text{dist}(f, g) = \Pr_{y \in F^n} [f(y) \neq g(y)]$  (where  $y$  is selected uniformly).

---

<sup>5</sup>Our interest lies in the *degree* of these functions (represented as polynomials). Since for any given subspace this degree is invariant with respect to the choice of the basis, the particular choice of the basis is only a matter of convenience.

### 3 The Characterization

In this section we prove the following characterization of polynomials of total degree at most  $d$  over finite fields.

**Theorem 1** *Let  $F$  be a field of cardinality  $q$  and characteristic  $p$ , let  $d$  be an integer, and let  $f : F^n \rightarrow F$ . Then  $f \in \text{POLY}_{n,d}$  if and only if for every affine subspace  $S$  of  $F^n$  having dimension  $\ell = \left\lceil \frac{d+1}{q-q/p} \right\rceil$ , we have that  $f|_S \in \text{POLY}_{\ell,d}$ .*

In particular, when  $q$  is prime, that is,  $q = p$ , we get that  $\ell = \left\lceil \frac{d+1}{q-1} \right\rceil$  and we obtain a set of words having weight at most  $q^{\left\lceil \frac{d+1}{q-1} \right\rceil}$  that span the code  $\mathcal{GRM}_q(n(q-1) - (d+1), n)$  (which is dual to  $\mathcal{GRM}_q(d, n)$ ). If we further have that  $d+1$  is divisible by  $q-1$ , then these are exactly the minimum-weight words of  $\mathcal{GRM}_q(n(q-1) - (d+1), n)$  (i.e., the characteristic vectors of affine subspaces of dimension  $\ell$ ). Otherwise, they have weight which is at most a factor of  $q$  larger than the minimum weight words (which correspond to weighted sums of characteristic vectors of affine subspaces of dimension  $\ell-1$ ). As noted in the introduction, in the case of GRM codes over prime fields, it is known that the minimum-weight words span the code [DK00]. However, when  $q$  is not prime then this is not the case (unless  $n = 1$  or  $d$  is very large) [DK00]. Thus we obtain a new result concerning small weight words that span GRM codes.

We also show that the above theorem is tight. Namely,

**Theorem 2** *For any given  $d$  and  $q = p^s$ , let  $\ell$  be as defined in Theorem 1. Then there exists a function  $f : F^n \rightarrow F$  such that for every affine subspace  $S$  of  $F^n$  having dimension less than  $\ell$ , the function  $f$  restricted to  $S$  is a degree- $d$  polynomial, but  $f$  is not a degree- $d$  polynomial.*

**Proof:** Let  $f = x_1^{(p-1)p^{s-1}} \cdot x_2^{(p-1)p^{s-1}} \cdots x_\ell^{(p-1)p^{s-1}}$ , so that the degree of  $f$  is  $\ell \cdot (p-1)p^{s-1}$ , which is at least  $d+1$ . On the other hand, consider any choice of  $\ell$  points  $y_0, y_1, \dots, y_{\ell-1}$ . Then

$$\begin{aligned} f|_{(y_0, y_1, \dots, y_{\ell-1})} &= \prod_{j=1}^{\ell} \left( y_{0,j} + \sum_{i=1}^{\ell-1} z_i \cdot y_{i,j} \right)^{(p-1)p^{s-1}} \\ &= \prod_{j=1}^{\ell} \left( y_{0,j}^{p^{s-1}} + \sum_{i=1}^{\ell-1} \left( y_{i,j}^{p^{s-1}} \cdot z_i^{p^{s-1}} \right) \right)^{p-1} \end{aligned} \quad (5)$$

(where we have used the fact that  $(a+b)^p = a^p + b^p$ ). Since for each  $i$ ,  $z_i^q = z_i^1$ , the total degree of any term in the above expression (which has the form  $z_1^{k_1 \cdot p^{s-1}} \cdot z_2^{k_2 \cdot p^{s-1}} \cdots z_{\ell-1}^{k_{\ell-1} \cdot p^{s-1}}$ ) is at most

$$(\ell-1) \cdot (p-1)p^{s-1} = \left( \left\lceil \frac{d+1}{(p-1)p^{s-1}} \right\rceil - 1 \right) \cdot (p-1)p^{s-1} < d+1 \quad (6)$$

and since  $(\ell-1) \cdot (p-1)p^{s-1}$  is an integer, we are done. ■

As noted in the introduction, our proof of Theorem 1 generalizes the proof of the special case of  $\ell = 1$ , which is presented in [FS95]. We prove Theorem 1 by induction. Namely, we prove that for every  $m \geq \ell$  and every affine subspace  $S$  of  $F^n$  having dimension  $m$ , if for every affine subspace

$S'$  of  $S$  that has dimension  $\ell$ , we have  $f_{|S'} \in \text{POLY}_{\ell,d}$ , then  $f_{|S} \in \text{POLY}_{m,d}$ . The base case,  $m = \ell$  clearly holds, and so we turn to the induction step.

Assume the induction claim holds for every  $m \geq \ell$ , we prove it for  $m + 1$ . Namely, we take any  $(m + 1)$ -dimensional subspace  $T$  of  $F^n$  and consider the function  $h = f_{|T} : F^{m+1} \rightarrow F$ . We then use the induction hypothesis by which for every affine subspaces  $S$  of  $F^{m+1}$  having dimension  $m$  (which is isomorphic to an affine subspace of  $T$  having dimension  $m$ ), the restriction of  $h = f_{|T}$  to  $S$  is a degree- $d$  polynomial.

Let the coefficients of the polynomial representation of  $h$  be denoted by  $\{C_\alpha^h\}_{\alpha \in [q-1]^{m+1}}$ . Our goal is to show that for each  $\alpha \in [q-1]^{m+1}$  such that  $\sum_{i=1}^{m+1} \alpha_i > d$ , we have that  $C_\alpha^h = 0$ . Let us fix any such  $\alpha$ , denote it by  $\alpha^*$  and prove that  $C_{\alpha^*}^h = 0$ . We break the proof into three cases.

**Case 1:** *There exists a subset  $R \subset \{1, \dots, m+1\}$ , where  $|R| = m$ , such that  $\sum_{i \in R} \alpha_i^* > d$ .*

Assume, without loss of generality (since the variables of  $h$ , and the corresponding  $\alpha_i^*$ 's can be reordered), that  $R = \{1, \dots, m\}$ , and assume, contrary to the claim, that  $C_{\alpha^*}^h > 0$ . We will show that this implies that there exist  $y_0, y_1, \dots, y_m \in F^{m+1}$  and  $\gamma = \gamma_1, \dots, \gamma_m$  where  $\gamma_i \in [q-1]$ , and  $\sum_{i=1}^m \gamma_i > d$ , such that for the affine subspace of  $S = S(y_0, y_2, \dots, y_m)$  of dimension  $m$  we have  $C_\gamma^{h|_S} > 0$ , contradicting the induction hypothesis.

Specifically, let  $y_i = e_i$  for  $i = 1, \dots, m$ , where  $e_i$  is the  $i$ 'th unit vector, and for each  $b \in F$ , let  $y_0(b) = b \cdot e_{m+1}$ . Let  $g_b = h_{|(y_0(b), y_1, \dots, y_m)}$  (so that  $g_b : F^m \rightarrow F$ ). Then for each choice of  $b \in F$  we have

$$\begin{aligned} g_b(z_1, \dots, z_m) &= h(z_1, \dots, z_m, b) \\ &= \sum_{\alpha \in [q-1]^{m+1}} C_\alpha^h \cdot \prod_{i=1}^m z_i^{\alpha_i} \cdot b^{\alpha_{m+1}} \end{aligned} \quad (7)$$

Consider the coefficient of the term  $z_1^{\alpha_1^*} \cdot z_2^{\alpha_2^*} \cdots z_m^{\alpha_m^*}$  in  $g_b$ , that is,  $C_{\alpha_1^*, \dots, \alpha_m^*}^{g_b}$  (where recall that  $\alpha^*$  satisfies  $\sum_{i=1}^{m+1} \alpha_i^* > d$ , as well as the premise of this case). This coefficient has the following form:

$$C_{\alpha_1^*, \dots, \alpha_m^*}^{g_b} = \sum_{j=0}^{q-1} C_{\alpha_1^*, \dots, \alpha_m^*, j}^h \cdot b^j \quad (8)$$

Namely, it is the evaluation, at  $b$ , of the univariate polynomial:  $\sum_{j=0}^{q-1} C_{\alpha_1^*, \dots, \alpha_m^*, j}^h \cdot x^j$ . Note that for  $j = \alpha_{m+1}^*$ , the coefficient of  $x^j$  in this polynomial is  $C_{\alpha^*}^h$ , which is non-zero by our counter-assumption. Hence, this polynomial is a non-zero polynomial of degree at most  $q-1$  over  $F$ . This implies that for at least one value of  $b$ , this polynomial attains a non-zero value. But this means that for some choice of  $b$ ,  $C_{\alpha_1^*, \dots, \alpha_m^*}^{g_b} > 0$ . Since  $\sum_{i=1}^m \alpha_i^* > d$ , we have reached a contradiction, and hence completed the proof for this case.

**Case 2:** *There exist a pair of indices  $i, j \in \{1, \dots, m+1\}$  such that  $\alpha_i^*, \alpha_j^* > 0$  and  $\alpha_i^* + \alpha_j^* < q$ .*

Assume, without loss of generality, that  $i = m$  and  $j = m+1$ . Here too we assume, contrary to the claim, that  $C_{\alpha^*}^h > 0$ , and reach a contradiction to the induction hypothesis.

Let  $y_0$  be the all-0 vector, let  $y_i = e_i$  for  $i = 1, \dots, m-1$ , and for each  $b \in F$ , let  $y_m(b) = \langle 0, \dots, 0, 1, b \rangle$  (recall that  $y_0, \dots, y_m \in F^{m+1}$ ). Here too we denote  $g_b = h_{|(y_0, \dots, y_m(b))}$ . Then for each choice of  $b \in F$  we have

$$\begin{aligned}
g_b(z_1, \dots, z_m) &= h(z_1, \dots, z_m, b \cdot z_m) \\
&= \sum_{\alpha \in [q-1]^{m+1}} C_\alpha^h \cdot \prod_{i=1}^{m-1} z_i^{\alpha_i} \cdot z_m^{\alpha_m} \cdot (z_m \cdot b)^{\alpha_{m+1}} \\
&= \sum_{\alpha \in [q-1]^{m+1}} C_\alpha^h \cdot \prod_{i=1}^{m-1} z_i^{\alpha_i} \cdot z_m^{\alpha_m + \alpha_{m+1}} \cdot b^{\alpha_{m+1}}
\end{aligned} \tag{9}$$

Consider the coefficient of the term  $z_1^{\alpha_1^*} \dots z_{m-1}^{\alpha_{m-1}^*} \cdot z_m^{\alpha_m^* + \alpha_{m+1}^*}$  in  $g_b$  (recall that  $\alpha_m^* + \alpha_{m+1}^* < q$ ). This coefficient has the following form:

$$C_{\alpha_1^*, \dots, \alpha_{m-1}^*, \alpha_m^* + \alpha_{m+1}^*}^{g_b} = \sum_{\substack{j, k \in [q-1]^2 \\ j+k = \alpha_m^* + \alpha_{m+1}^*}} C_{\alpha_1^*, \dots, \alpha_{m-1}^*, j, k}^h \cdot b^k \tag{10}$$

That is, it is the evaluation, at  $b$ , of the univariate polynomial:

$$\sum_{k=0}^{q-1} C_{\alpha_1^*, \dots, \alpha_{m-1}^*, \alpha_m^* + \alpha_{m+1}^* - k, k}^h \cdot x^k \tag{11}$$

Note that for  $k = \alpha_{m+1}^*$ , the coefficient of  $x^k$  in this polynomial is  $C_{\alpha^*}^h$ , which is non-zero by our counter-assumption. Hence, this polynomial is a non-zero polynomial of degree at most  $q-1$  over  $F$ , which implies that for at least one value of  $b$  it attains a non-zero value. But this means that for some choice of  $b$ ,  $C_{\alpha_1^*, \dots, \alpha_{m-1}^*, \alpha_m^* + \alpha_{m+1}^*}^{g_b} > 0$  and the proof of this case follows.

We observe that if  $\ell = \left\lceil \frac{2(d+1)}{q} \right\rceil$  then either Case 1 or Case 2 must hold. This implies that Theorem 1 is established for  $q$  that is a power of 2 (since in this case  $q - q/p = q/2$ ). It also follows that for any value of  $q$ , a variant of Theorem 1, which takes  $\ell$  to be at most a factor of 2 larger than that stated in the theorem, is established as well. In order to get the tighter result, which holds for  $\ell = \left\lceil \frac{d+1}{q - (q/p)} \right\rceil$  and any  $q$ , we need to analyze the third and final case.

**Case 3 (neither Case 1 nor Case 2 hold):** For every subset  $R \subset \{1, \dots, m+1\}$ ,  $|R| = m$ , we have that  $\sum_{i \in R} \alpha_i^* \leq d$ , and for every pair of indices  $i, j \in \{1, \dots, m+1\}$  we have that  $\alpha_i^* + \alpha_j^* \geq q$ .

Our proof of this case is similar in its general structure to the proofs of Cases 1 and 2, but is somewhat more involved since we take into account a larger set of  $m$ -dimensional affine subspaces. Specifically, for every choice of  $a_1, \dots, a_m, b$  each in  $F$ , let  $y_0 = b \cdot e_{m+1}$ , and for  $i = 1, \dots, m$ , let  $y_i = a_i \cdot e_i + e_{m+1}$ . Consider the function  $g_{a_1, \dots, a_m, b} = h_{|(y_0(b), y_1(a_1), \dots, y_m(a_m))} : F^m \rightarrow F$ . By

definition:

$$\begin{aligned}
& g_{a_1, \dots, a_m, b}(z_1, \dots, z_m) \\
&= h\left(a_1 \cdot z_1, \dots, a_m \cdot z_m, \sum_{i=1}^m z_i + b\right) \\
&= \sum_{\alpha \in [q-1]^{m+1}} C_\alpha^h \cdot \prod_{i=1}^m (a_i \cdot z_i)^{\alpha_i} \cdot \left(\sum_{i=1}^m z_i + b\right)^{\alpha_{m+1}} \\
&= \sum_{\alpha \in [q-1]^{m+1}} \sum_{\substack{\delta \in [q-1]^m \\ \sum_{i=1}^m \delta_i \leq \alpha_{m+1}}} \binom{\alpha_{m+1}}{\delta_1, \dots, \delta_m} \cdot C_\alpha^h \cdot \prod_{i=1}^m a_i^{\alpha_i} \cdot b^{\alpha_{m+1} - \sum \delta_i} \cdot \prod_{i=1}^m z_i^{\alpha_i + \delta_i} \quad (12)
\end{aligned}$$

Roughly speaking, for each  $\alpha \in [q-1]^{m+1}$ , the exponent  $\alpha_{m+1}$  “gets distributed” among the different  $z_i$ ’s ( $i = 1, \dots, m$ ), and  $b$ . Note that if  $\alpha_i + \delta_i = q$  then  $z_i^{\alpha_i + \delta_i} = z_i$ , and more generally, if  $\alpha_i + \delta_i \geq q$  then  $z_i^{\alpha_i + \delta_i} = z_i^{(\alpha_i + \delta_i) \bmod (q-1)}$ . In what follows we use the shorthand  $(j)_q$  to denote  $(j \bmod (q-1))$ .

For any choice of  $\gamma = \gamma_1, \dots, \gamma_m$ , we consider the coefficient of the term  $\prod_{i=1}^m z_i^{\gamma_i}$  in the representation of  $g_{a_1, \dots, a_m, b}(z_1, \dots, z_m)$  as a polynomial of degree at most  $q-1$  in each variable (that is,  $C_\gamma^{g_{a_1, \dots, a_m, b}}$ ). It follows from Equation (12) that this coefficient is the evaluation of the following *multivariate* polynomial,

$$\begin{aligned}
& H_\gamma(x_1, \dots, x_{m+1}) \\
&= \sum_{\substack{\alpha \in [q-1]^{m+1} \\ \alpha_{m+1} \geq \sum_{i=1}^m (\gamma_i - \alpha_i)_q}} \binom{\alpha_{m+1}}{(\gamma_1 - \alpha_1)_q, \dots, (\gamma_m - \alpha_m)_q} \cdot C_\alpha^h \cdot \prod_{i=1}^m x_i^{\alpha_i} \cdot x_{m+1}^{\alpha_{m+1} - \sum_{i=1}^m (\gamma_i - \alpha_i)_q} \quad (13)
\end{aligned}$$

at the point  $x_1 = a_1, \dots, x_m = a_m, x_{m+1} = b$ .

As in Cases 1 and 2, we would like to show that under the assumption that  $C_{\alpha^*}^h > 0$  for  $\alpha^*$  such that  $\sum_{i=1}^{m+1} \alpha_i^* > d$ , we can get the following. There exist  $a_1, \dots, a_m$  and  $b$  in  $F$  and  $\gamma_1, \dots, \gamma_m \in [q-1]$  such that  $\sum_{i=1}^m \gamma_i > d$  and the coefficient of the term  $\prod_{i=1}^m z_i^{\gamma_i}$  in the representation of  $g_{a_1, \dots, a_m, b}$  as a polynomial of degree at most  $q-1$  (in each variable) is non-zero. Since we want to exploit the existence of  $\alpha^* \in [q-1]^{m+1}$  as stated above, we shall consider  $\gamma_1, \dots, \gamma_m$  of the form  $\gamma_i = \alpha_i^* + \delta_i$  where  $\delta_1, \dots, \delta_m$  ( $\delta_i \in [q-1]$ ) obey the following three conditions:

- C1.  $\sum_{i=1}^m \delta_i \leq \alpha_{m+1}^*$ ;
- C2.  $\alpha_i^* + \delta_i (= \gamma_i) \leq q-1$  for every  $i, 1 \leq i \leq m$ ;
- C3.  $\sum_{i=1}^m (\alpha_i^* + \delta_i) > d$  (that is,  $\sum_{i=1}^m \gamma_i > d$ ).

We would like to show that there exists a choice of  $\delta_1, \dots, \delta_m$  that satisfies conditions C1–C3 such that for  $\gamma = \alpha_1^* + \delta_1, \dots, \alpha_m^* + \delta_m$  we get that  $H_\gamma$  as defined in Equation (13) is a *non-zero* polynomial. This would imply that there exists a choice of  $a_1, \dots, a_m$  and  $b$  on which the value of  $H_\gamma$  is non-zero, meaning that  $C_\gamma^{g_{a_1, \dots, a_m, b}} > 0$ .

Subcase 1:  $q$  is prime. Consider first the case that  $q$  is prime. That is,  $q = p$ . In this case  $m \geq \ell = \left\lfloor \frac{d+1}{q-1} \right\rfloor$ . Let  $\delta_1 = q-1 - \alpha_1^*$ , and recall that for every  $i, j$  we have  $\alpha_i^* + \alpha_j^* \geq q$ , so

that necessarily  $\delta_1 < \alpha_{m+1}^*$ . Next let  $\delta_2 = \min\{q - 1 - \alpha_2^*, \alpha_{m+1}^* - \delta_1\}$ , and in general,  $\delta_i = \min\{q - 1 - \alpha_i^*, \alpha_{m+1}^* - \sum_{j < i} \delta_j\}$ . Then it is not hard to verify that  $\delta_1, \dots, \delta_m$  satisfy conditions C1–C3 (where we use the fact that by definition of  $\ell$ ,  $m \cdot (q - 1) \geq d + 1$  and that  $\sum_{i=1}^{m+1} \alpha_i^* \geq d + 1$ ). We claim that  $H_\gamma$  in this case includes at least one non-zero coefficient. To verify this note that since  $\sum_{i=1}^m \delta_i \leq \alpha_{m+1}^*$ , the sum in Equation (13) includes  $\alpha = \alpha^*$ . Since  $C_{\alpha^*}^h > 0$  by our counter assumption and  $\binom{\alpha_{m+1}^*}{\delta_1, \dots, \delta_m} > 0$ , we get that  $\binom{\alpha_{m+1}^*}{\delta_1, \dots, \delta_m} \cdot C_{\alpha^*}^h$  is a non-zero coefficient of the term  $\prod_{i=1}^m x_i^{\alpha_i^*} \cdot x_{m+1}^{\alpha_{m+1}^* - \sum_{i=1}^m \delta_i}$  in the polynomial  $H_\gamma$ .

Subcase 2:  $q$  is not prime. When  $q$  is not a prime number, so that  $q = p^s$  for  $s > 1$ , then the above argument breaks down due to the multinomial coefficient, which may be 0 modulo  $p$ , causing the term in question to vanish. Hence, in order to complete the proof of this last case we show that there exist a setting of the  $\delta_i$ 's, for which conditions C1–C3 hold and  $\binom{\alpha_{m+1}^*}{\delta_1, \dots, \delta_m}$  is not divisible by  $p$ . Let us denote the latter condition by C4. That is,

C4.  $\binom{\alpha_{m+1}^*}{\delta_1, \dots, \delta_m}$  is not divisible by  $p$ .

Since

$$\binom{\alpha_{m+1}^*}{\delta_1, \dots, \delta_m} = \binom{\alpha_{m+1}^*}{\delta_1} \cdot \binom{\alpha_{m+1}^* - \delta_1}{\delta_2} \dots \binom{\alpha_{m+1}^* - \sum_{j < i} \delta_j}{\delta_i} \dots \binom{\alpha_{m+1}^* - \sum_{j < m} \delta_j}{\delta_m} \quad (14)$$

in order to show that condition C4 holds, it suffices to show that each term in the above product is not divisible by  $p$ .

We shall use the following facts in our setting of the  $\delta_i$ 's:

1. By the premise of this case,  $\alpha_i^* + \alpha_j^* \geq q$  for every pair  $i < j$ .
2. By our choice of  $\ell$  (and since  $m \geq \ell$ ), if  $\gamma_i = \alpha_i^* + \delta_i \geq (p - 1)p^{s-1}$  for  $i = 1, \dots, m$ , then  $\sum_{i=1}^m \gamma_i > d$ .
3. Since, by the premise of this case, the sum of the  $\alpha_i^*$ 's over every subset of size  $m$  is at most  $d$ , we know that for every such subset there exists some  $\alpha_i^*$  such that  $\alpha_i^* < (p - 1)p^{s-1}$ .

We shall also use the following notation: For each  $\alpha_i^*$ , let  $k_{i,j}$  for  $j \in [p - 1]$  be such that  $\alpha_i^* = \sum_{j=0}^{s-1} k_{i,j} p^j$ . Let us assume without loss of generality that  $\alpha_{m+1}^*$  is the smallest  $\alpha_i^*$  and that  $\alpha_1^* \leq \alpha_2^* \leq \dots \leq \alpha_m^*$  (we may re-order the variables to get that). We know that  $\alpha_1^* < (p - 1)p^{s-1}$  (or else  $\sum_{i=1}^m \alpha_i^* \geq (p - 1)p^{s-1} \cdot \ell \geq d + 1$ ). Since  $\alpha_i^* + \alpha_1^* \geq q = p^s$  for every  $i > 1$ , necessarily  $\alpha_i^* > p^{s-1}$  for  $i > 1$ , and similarly  $\alpha_1^* > p^{s-1}$  (since  $\alpha_1^* + \alpha_{m+1}^* \geq q$ ). Hence for each  $\alpha_i^*$  we have that  $1 \leq k_{i,s-1} \leq p - 1$ . For  $1 \leq i \leq m$ , let  $t_i \stackrel{\text{def}}{=} p - 1 - k_{i,s-1}$  (for technical purposes  $t_0 \stackrel{\text{def}}{=} 0$ ). Recall that the  $\alpha_i^*$ 's are in ascending order, thus if  $t_i = 0$  then  $t_j = 0$  for every  $j > i > 0$ .

Finally we shall use the following technical claim that was given in [FS95], and whose proof is provided for the sake of completeness in the appendix.

**Claim 1** *Let  $q = p^s$  for a prime number  $p$  and an integer  $s$ , and let  $r$  and  $t$  be integers that satisfy  $0 < r \leq t \leq q - 1$ . If  $r = kp^{s-1}$  for some integer  $k$  then  $\binom{t}{r}$  is not divisible by  $p$ .*

We now show a setting of the  $\delta_i$ 's for which conditions C1–C4 hold. If  $t_i \geq 1$  and  $t_i \leq k_{m+1,s-1} - \sum_{j<i} t_j$ , then set  $\delta_i \stackrel{\text{def}}{=} t_i p^{s-1}$ . By Claim 1,  $\binom{\alpha_{m+1}^* - \sum_{j<i} \delta_j}{\delta_i}$  is not divisible by  $p$ . If  $t_i \geq 1$  and  $t_i > k_{m+1,s-1} - \sum_{j<i} t_j$ , then set  $\delta_i = \alpha_{m+1}^* - \sum_{j<i} \delta_j < t_i p^{s-1}$ . In this case  $\binom{\alpha_{m+1}^* - \sum_{j<i} \delta_j}{\delta_i} = 1$ , and is hence not divisible by  $p$ . Note, that if there exist  $i_0$  such that  $\delta_{i_0} = \alpha_{m+1}^* - \sum_{j<i_0} \delta_j$ , then  $\delta_j = 0$ , for  $j > i_0$ . We have thus established that condition C4 holds for the above setting of the  $\delta_i$ 's. From the definition of the  $\delta_i$ 's it directly follows that condition C1 holds as well. It remains to verify that conditions C2 and C3 hold.

By the definition of the  $\delta_i$ 's, for every  $1 \leq i \leq m$ :

$$\alpha_i^* + \delta_i \leq (p-1) \cdot p^{s-1} + \sum_{\ell=2}^s k_{i,s-\ell} \cdot p^{s-\ell} < q \quad (15)$$

and so condition C2 holds.

We next verify that  $\sum_{i=1}^m (\alpha_i^* + \delta_i) > d$ . if there exist  $i_0$  s.t.  $\delta_{i_0} = \alpha_{m+1}^* - \sum_{j<i_0} \delta_j$  then

$$\sum_{i=1}^m (\alpha_i^* + \delta_i) = \sum_{i=1}^{m+1} \alpha_i^* > d \quad (16)$$

otherwise  $(\alpha_i^* + \delta_i) \geq (p-1)p^{s-1}$  for every  $1 \leq i \leq m$ . Since  $m \geq \ell = \left\lceil \frac{d+1}{(p-1)p^{s-1}} \right\rceil$  then  $\sum_{i=1}^m (\alpha_i^* + \delta_i) \geq d+1$ . Thus, condition C3 holds and we have completed the proof of Case 3 (and hence Theorem 1).

## 4 The Test

In this section we present and analyze our testing algorithm for degree- $d$  polynomials over fields of cardinality  $q = O(d)$ .

### Algorithm 1 Testing Algorithm for Degree- $d$ Polynomials

1. Let  $\ell = \ell(q, d) = \left\lceil \frac{d+1}{q-d/p} \right\rceil$  and repeat the following  $t = \Theta\left(\ell \cdot q^{\ell+1} + \frac{1}{\epsilon \cdot q^\ell}\right)$  times:
  - (a) Uniformly and independently select  $\ell + 1$  linearly independent points  $y_0, y_1, \dots, y_\ell \in F^n$ .
  - (b) If  $f|_{(y_0, y_1, \dots, y_\ell)} \notin \text{POLY}_{\ell, d}$  then output reject.
2. If no step caused rejection then output accept.

Recall that checking whether  $f|_S \notin \text{POLY}_{\ell, d}$  (where  $S = S(y_0, y_1, \dots, y_\ell)$ ) can be done by querying  $f$  on all points in the subspace  $S$  and verifying that all linear constraints corresponding to the coefficients  $C_\alpha^{f|_S}$  such that  $\sum_{i=1}^\ell \alpha_i > d$ , hold. Hence the total number of queries performed by the algorithm is  $O(t \cdot q^\ell) = O\left(\ell \cdot q^{2\ell+1} + \frac{1}{\epsilon}\right)$  (where the  $q^\ell$  term is due to the number of points in each affine subspace.)

As noted in the introduction, when  $q$  is sufficiently larger than  $d$  so that  $\ell = 1$  (the subspaces are lines), then it is not necessary to query  $f$  on all points on the line, but rather  $d+2$  points suffice.

We note that these checks involving  $d+2$  points on a line can be interpreted as selecting minimum-weight words from the dual GRM code and checking that they are orthogonal to the word defined by  $f$ . Our test can be modified so that instead of checking a set of constraints (several small-weight words in the dual code) in each step, it also selects one random constraint (one small-weight word in the dual code) in each step.<sup>6</sup> However, this will not reduce the query complexity in our case.

**Theorem 3** *If  $f \in \text{POLY}_{n,d}$  then Algorithm 1 accepts with probability 1, and if  $\text{dist}(f, \text{POLY}_{n,d}) > \epsilon$  then Algorithm 1 rejects with probability at least  $2/3$ .*

Theorem 3 shall be proved using the “self-correcting approach”, which has been applied in the analysis of many previous low-degree tests. Namely, given the function  $f$  we define another function  $g$  based on certain “majority votes” of  $f$ . We then show that if  $f$  passes the test with sufficiently high probability, then  $g$  is close to  $f$  and  $g$  is a polynomial of degree at most  $d$ . Bounding the distance between  $f$  and  $g$  follows easily from the definition of  $g$ , and hence the analysis is focused on showing that  $g$  is a polynomial of degree at most  $d$ . The analysis can be viewed as generalizing both the analysis in [RS96] (where the subspaces considered by the test are lines) and the analysis in [AKK<sup>+</sup>03] (where the subspaces are larger but the field is  $GF(2)$ , and the analysis relies on the fact that the field is  $GF(2)$ ).

We start by introducing several notations. In all that follows, whenever we consider the selection of sets of points in  $F^n$ , the selection is restricted to linearly independent points.

**Definition 5** *Let*

$$\eta = \eta(f, d) \stackrel{\text{def}}{=} \Pr_{y_0, y_1, \dots, y_\ell} \left[ f|_{(y_0, y_1, \dots, y_\ell)} \notin \text{POLY}_{\ell, d} \right] \quad (17)$$

*denote the probability that a single step of the algorithm causes  $f$  to be rejected. That is, it is the probability that the restriction of  $f$  to a random affine subspace of dimension  $\ell$  is not a polynomial of degree at most  $d$ .*

**Definition 6** *For each  $\alpha \in [q-1]^\ell$ , let  $C_\alpha^f(y_0, y_1, \dots, y_\ell)$  denote the coefficient  $C_\alpha$  of the polynomial representation of  $f|_{(y_0, y_1, \dots, y_\ell)}$ . We shall use the notation  $B^f(y_0, y_1, \dots, y_\ell)$  as a shorthand for the coefficient  $C_{(q-1, \dots, q-1)}^f(y_0, y_1, \dots, y_\ell)$ . That is,  $C_{(q-1, \dots, q-1)}^f(y_0, y_1, \dots, y_\ell)$  denotes the coefficient of the highest-degree monomial  $x_1^{q-1} \cdot x_2^{q-1} \cdot \dots \cdot x_\ell^{q-1}$  in the polynomial representation of  $f|_{(y_0, y_1, \dots, y_\ell)}$ . We denote by  $V^f(y; y_1, \dots, y_\ell)$  the value that  $f(y)$  “should have” so that  $B^f(y, y_1, \dots, y_\ell) = 0$ . That is,*

$$V^f(y; y_1, \dots, y_\ell) = - \sum_{\substack{b_1, \dots, b_\ell \in F \\ \exists i \text{ s.t. } b_i \neq 0}} f\left(y + \sum_{i=1}^{\ell} b_i \cdot y_i\right). \quad (18)$$

*We refer to  $V^f(y; y_1, \dots, y_\ell)$  as the vote of  $(y_1, \dots, y_\ell)$  on the value assigned to  $y$ .*

*For succinctness of the notation, we shall remove  $f$  from the last two notations (i.e.,  $B(\cdot) = B^f(\cdot)$  and  $V(\cdot) = V^f(\cdot)$ ).*

Note that for  $\eta$  as in Definition 5,

$$\eta \geq \Pr_{y, y_1, \dots, y_\ell} [V(y; y_1, \dots, y_\ell) \neq f(y)] \quad (19)$$

(since the test checks that *all* coefficients  $C_\alpha^f(y, y_1, \dots, y_\ell)$  for which  $\sum_{i=1}^{\ell} \alpha_i > d$  are 0).

---

<sup>6</sup>In case  $q$  is prime then, as shown in [JPRZ04], it suffices to consider a single constraint per affine subspace.

**Definition 7** Let  $g$  be a plurality function that is defined as follows. For each  $y \in F^n$ ,

$$g(y) = \operatorname{argmax}_{a \in F} \left\{ \Pr_{y_1, \dots, y_\ell \in F^n} [V(y; y_1, \dots, y_\ell) = a] \right\} \quad (20)$$

**Lemma 2** For any function  $f$  and for  $\eta$  and  $g$  as defined in Equations (17) and (20) respectively,  $\operatorname{dist}(f, g) \leq 2\eta$ .

**Proof:** First observe that if the test selects points  $y_0, y_1, \dots, y_\ell \in F^n$  such that  $f(y_0) \neq V(y_0; y_1, \dots, y_\ell)$  then this means that  $B(y_0; y_1, \dots, y_\ell) \neq 0$ , which causes the test to reject. Let  $U \subseteq F^n$  consist of all (“bad”) points  $y \in F^n$  such that  $\Pr_{y_1, \dots, y_\ell \in F^n} [f(y) \neq V(y; y_1, \dots, y_\ell)] > 1/2$ . By definition of  $\eta$  we know that  $|U|/q^n < 2\eta$ . But for every  $x \in F^n \setminus U$ , by definition of  $g$  we have that  $f(x) = g(x)$ , and the lemma follows. ■

In the next series of lemmas we prove that if  $\eta$  is sufficiently small then  $g$  is a polynomial of total degree at most  $d$ . In the first, and central lemma, we show that for every  $y$ , the value of  $g(y)$ , which by Definition 7 is the “plurality vote” of  $V(y; y_1, \dots, y_\ell)$ , taken over all  $y_1, \dots, y_\ell$ , equals the vote of a large fraction of the  $\ell$ -tuples  $y_1, \dots, y_\ell$  (assuming  $\eta$  is sufficiently small).

**Lemma 3** For any fixed  $y \in F^n$ , let

$$\gamma(y) \stackrel{\text{def}}{=} \Pr_{y_1, \dots, y_\ell} [V(y; y_1, \dots, y_\ell) = g(y)] \quad (21)$$

Then  $\gamma(y) \geq 1 - 2q\ell\eta$ .

In order to prove Lemma 3 it will actually be more convenient to work with another measure of “correctness” (or “consistency”) of a point  $y$ .

**Lemma 4** For any fixed  $y \in F^n$ , let

$$\delta(y) = \Pr_{y_1, \dots, y_\ell, z_1, \dots, z_\ell} [V(y; y_1, \dots, y_\ell) = V(y; z_1, \dots, z_\ell)] \quad (22)$$

and let  $\gamma(y)$  be as defined in Equation (21). Then  $\gamma(y) \geq \delta(y)$ .

**Proof:** Let  $\beta_a(y) = \Pr_{y_1, \dots, y_\ell} [V(y; y_1, \dots, y_\ell) = a]$  (so that in particular,  $\sum_{a \in F} \beta_a(y) = 1$ ). By definition of  $\gamma(y)$  we have that  $\gamma(y) = \max_a \beta_a(y)$ , and by definition of  $\delta(y)$  we have that,  $\delta(y) = \sum_{a \in F} (\beta_a(y))^2$ . By convexity,  $\max_a \beta_a(y) \geq \sum_{a \in F} (\beta_a(y))^2$ , and the claim follows. ■

**An Auxiliary “Voting Graph”.** In order to show that  $\delta(y)$  is large (and hence  $\gamma(y)$  is large), it will be useful to consider the following auxiliary graph. The definition of this graph was inspired by the way Shpilka and Wigderson used Cayley graphs in their work [SW04] and can also be viewed as formalizing and generalizing part of the analysis in [AKK<sup>+</sup>03]. Each vertex in this graph is labeled by a subset  $\{y_1, \dots, y_\ell\}$ ,  $y_i \in F^n$ . There is an edge between the vertex labeled by  $\{y_1, \dots, y_\ell\}$ , and the vertex labeled by  $\{z_1, \dots, z_\ell\}$  if and only if  $|\{y_1, \dots, y_\ell\} \cup \{z_1, \dots, z_\ell\}| = \ell + 1$ . In other words, neighbors are of the form  $\{y_1, \dots, y_\ell\}$  and  $\{y_2, \dots, y_{\ell+1}\}$ . Each vertex corresponds to an  $\ell$ -tuple that can “vote” on the value of  $f(y)$  for any given  $y$  and hence we refer to it as the *voting graph*.

For a fixed point  $y \in F^n$ , we say that an edge between  $\{y_1, \dots, y_\ell\}$  and  $\{y_2, \dots, y_{\ell+1}\}$  is *good with respect to  $y$*  if  $V(y; y_1, \dots, y_\ell) = V(y; y_2, \dots, y_{\ell+1})$ .

**Lemma 5** For any choice of  $y, y_1, \dots, y_{\ell+1} \in F^n$ ,

$$\begin{aligned} & V(y; y_1, \dots, y_\ell) - V(y; y_2, \dots, y_{\ell+1}) \\ &= \sum_{a \in F, a \neq 0} B(y + a \cdot y_{\ell+1}, y_1, \dots, y_\ell) - \sum_{a \in F, a \neq 0} B(y + a \cdot y_1, y_2, \dots, y_{\ell+1}) \end{aligned}$$

**Proof:** By definition of  $V(y; \cdot)$  we have:

$$\begin{aligned} & V(y; y_1, \dots, y_\ell) - V(y; y_2, \dots, y_{\ell+1}) \\ &= - \sum_{\substack{b_1, \dots, b_\ell \in F \\ b_1 \neq 0}} f\left(y + \sum_{i=1}^{\ell} b_i \cdot y_i\right) + \sum_{\substack{b_2, \dots, b_{\ell+1} \in F \\ b_{\ell+1} \neq 0}} f\left(y + \sum_{i=2}^{\ell+1} b_i \cdot y_i\right) \end{aligned} \quad (23)$$

$$\begin{aligned} &= - \sum_{\substack{b_1, \dots, b_\ell \in F \\ b_1 \neq 0}} f\left(y + \sum_{i=1}^{\ell} b_i \cdot y_i\right) - \sum_{\substack{b_1, \dots, b_{\ell+1} \in F \\ b_1, b_{\ell+1} \neq 0}} f\left(y + \sum_{i=1}^{\ell+1} b_i \cdot y_i\right) \\ &+ \sum_{\substack{b_2, \dots, b_{\ell+1} \in F \\ b_{\ell+1} \neq 0}} f\left(y + \sum_{i=2}^{\ell+1} b_i \cdot y_i\right) + \sum_{\substack{b_1, \dots, b_{\ell+1} \in F \\ b_1, b_{\ell+1} \neq 0}} f\left(y + \sum_{i=1}^{\ell+1} b_i \cdot y_i\right) \end{aligned} \quad (24)$$

$$\begin{aligned} &= - \sum_{\substack{b_1, \dots, b_{\ell+1} \in F \\ b_1 \neq 0}} f\left(y + b_1 \cdot y_1 + \sum_{i=2}^{\ell+1} b_i \cdot y_i\right) + \sum_{\substack{b_1, \dots, b_{\ell+1} \in F \\ b_{\ell+1} \neq 0}} f\left(y + b_{\ell+1} \cdot y_{\ell+1} + \sum_{i=1}^{\ell} b_i \cdot y_i\right) \end{aligned} \quad (25)$$

$$\begin{aligned} &= \sum_{a \in F, a \neq 0} B(y + a \cdot y_{\ell+1}, y_1, \dots, y_\ell) - \sum_{a \in F, a \neq 0} B(y + a \cdot y_1, y_2, \dots, y_{\ell+1}) \end{aligned} \quad (26)$$

■

**Proof of Lemma 3.** Given Lemma 4, it suffices to show that for every  $y \in F^n$ ,  $\delta(y) \geq 1 - 2q\ell\eta$ . For any (random) choice of  $y_1, \dots, y_\ell$  and  $z_1, \dots, z_\ell$ , consider a shortest path between the two corresponding vertices in the voting graph. Such a path is of length at most  $\ell$ , and each edge on the path is of the form  $(\{y_1, \dots, y_i, z_{i+1}, \dots, z_\ell\}, \{y_1, \dots, y_{i-1}, z_i, \dots, z_\ell\})$ . By Lemma 5 such an edge is good if

$$\sum_{a \in F, a \neq 0} B(y + a \cdot y_i, y_1, \dots, y_{i-1}, z_i, \dots, z_\ell) - \sum_{a \in F, a \neq 0} B(y + a \cdot z_i, y_1, \dots, y_i, z_{i+1}, \dots, z_\ell) = 0 \quad (27)$$

Since the  $z_i$ 's and the  $y_i$ 's are selected uniformly at random, each of the  $B(\cdot)$ 's in the above summation is non-zero with probability at most  $\eta$ . Hence, by applying a union bound, the probability, taken over the uniform choice of the  $y_i$ 's and  $z_i$ 's, that an edge is good is at least  $1 - 2q\eta$ . By taking a union bound once more, the probability that all the edges on the path are good is at least  $1 - 2q\ell\eta$ . In such a case,  $V(y; y_1, \dots, y_\ell) = V(y; y_1, \dots, y_{\ell-1}, z_\ell) = \dots = V(y; z_1, \dots, z_\ell)$ , and the lemma follows. ■

We can now establish:

**Lemma 6** If  $\eta < \frac{1}{2(\ell+1)q^{\ell+1}}$  then  $g \in \text{POLY}_{n,d}$ .

**Proof:** Consider any fixed set of points  $y_0, y_1, \dots, y_\ell \in F^n$ . We shall show that  $g_{|(y_0, y_1, \dots, y_\ell)} \in \text{POLY}_{\ell, d}$ . Lemma 6 follows by applying Theorem 1.

Each point in the affine subspace  $S(y_0, y_1, \dots, y_\ell)$  is of the form  $y_0 + \sum_{i=1}^{\ell} b_i y_i$ , where  $b_i \in F$ . Now consider  $(\ell + 1) \cdot \ell$  elements in  $F^n$ , denoted  $\{z_{i,j}\}_{i=0, \dots, \ell}^{j=1, \dots, \ell}$ . Assume first that for every choice of  $b_1, \dots, b_\ell \in F$ ,

$$g\left(y_0 + \sum_{i=1}^{\ell} b_i \cdot y_i\right) = V\left(y_0 + \sum_{i=1}^{\ell} b_i \cdot y_i; z_{0,1} + \sum_{i=1}^{\ell} b_i \cdot z_{i,1}, \dots, z_{0,\ell} + \sum_{i=1}^{\ell} b_i \cdot z_{i,\ell}\right) \quad (28)$$

If we select the elements  $\{z_{i,j}\}$  uniformly and at random, then by Lemma 3, this event occurs with probability at least  $1 - 2q\ell\eta \cdot q^\ell$ . We assume from this point on that Equation (28) holds for every choice of  $b_1, \dots, b_\ell \in F$ .

In order to show that  $g_{|(y_0, y_1, \dots, y_\ell)} \in \text{POLY}_{\ell, d}$  we need to show that for every  $\alpha \in [q-1]^\ell$  such that  $\sum_{i=1}^{\ell} \alpha_i > d$ , we have that  $C_\alpha^g(y_0, y_1, \dots, y_\ell) = 0$ . Let us fix any such  $\alpha$ , and let  $R_\alpha$  denote the row vector  $\mathcal{A}_\ell(\alpha, \cdot)$ . Recall that the coordinates of  $R_\alpha$  are indexed by strings  $\beta \in [q-1]^\ell$  (where we denote the corresponding coordinate by  $R_\alpha^\beta \in F$ ). In what follows we use the notation:  $ex(\beta_i) = \omega^{\beta_i}$  if  $\beta_i \neq 0$ , and  $ex(\beta_i) = 0$  if  $\beta_i = 0$ . Consider the structure of  $\mathcal{A}_\ell$  presented in Section 2. We need to show that

$$\sum_{\beta \in [q-1]^\ell} R_\alpha^\beta \cdot g\left(y_0 + \sum_{i=1}^{\ell} ex(\beta_i) \cdot y_i\right) = 0 \quad (29)$$

For any fixed  $\beta \in [q-1]^\ell$ , by our assumption that Equation (28) holds, and by definition of  $V(\cdot)$  we have:

$$\begin{aligned} & g\left(y_0 + \sum_{i=1}^{\ell} ex(\beta_i) \cdot y_i\right) \\ &= - \sum_{\substack{\gamma \in [q-1]^\ell \\ \gamma \neq (0,0,\dots,0)}} f\left(y_0 + \sum_{i=1}^{\ell} ex(\beta_i) \cdot y_i + \sum_{j=1}^{\ell} ex(\gamma_j) \cdot \left(z_{0,j} + \sum_{i=1}^{\ell} ex(\beta_i) \cdot z_{i,j}\right)\right) \\ &= - \sum_{\substack{\gamma \in [q-1]^\ell \\ \gamma \neq (0,0,\dots,0)}} f\left(y_0 + \sum_{j=1}^{\ell} ex(\gamma_j) \cdot z_{0,j} + \sum_{i=1}^{\ell} ex(\beta_i) \cdot \left(y_i + \sum_{j=1}^{\ell} ex(\gamma_j) \cdot z_{i,j}\right)\right) \end{aligned} \quad (30)$$

This implies (by switching the order of summations) that

$$\begin{aligned} & \sum_{\beta \in [q-1]^\ell} R_\alpha^\beta \cdot g\left(y_0 + \sum_{i=1}^{\ell} ex(\beta_i) \cdot y_i\right) \\ &= - \sum_{\substack{\gamma \in [q-1]^\ell \\ \gamma \neq (0,0,\dots,0)}} \sum_{\beta \in [q-1]^\ell} R_\alpha^\beta \cdot f\left(y_0 + \sum_{j=1}^{\ell} ex(\gamma_j) \cdot z_{0,j} + \sum_{i=1}^{\ell} ex(\beta_i) \cdot \left(y_i + \sum_{j=1}^{\ell} ex(\gamma_j) \cdot z_{i,j}\right)\right) \end{aligned} \quad (31)$$

But for any given choice of  $\gamma = \gamma_1, \dots, \gamma_\ell$ ,

$$\begin{aligned} & \sum_{\beta \in [q-1]^\ell} R_\alpha^\beta \cdot f \left( y_0 + \sum_{j=1}^{\ell} \text{ex}(\gamma_j) \cdot z_{0,j} + \sum_{i=1}^{\ell} \text{ex}(\beta_i) \cdot \left( y_i + \sum_{j=1}^{\ell} \text{ex}(\gamma_j) \cdot z_{i,j} \right) \right) \\ &= C_\alpha^f \left( y_0 + \sum_{j=1}^{\ell} \text{ex}(\gamma_j) \cdot z_{0,j}, y_1 + \sum_{j=1}^{\ell} \text{ex}(\gamma_j) \cdot z_{1,j}, \dots, y_\ell + \sum_{j=1}^{\ell} \text{ex}(\gamma_j) \cdot z_{\ell,j} \right) \end{aligned} \quad (32)$$

Since  $\gamma_1, \dots, \gamma_\ell$  are not all 0, then we know that for each setting of  $\gamma_1, \dots, \gamma_\ell$ , with probability at least  $1 - \eta$  over the choice of the  $z_{i,j}$ 's, for every  $\alpha$  such that  $\sum_{i=1}^{\ell} \alpha_i > d$ ,

$$C_\alpha^f \left( y_0 + \sum_{j=1}^{\ell} \text{ex}(\gamma_j) \cdot z_{0,j}, y_1 + \sum_{j=1}^{\ell} \text{ex}(\gamma_j) \cdot z_{1,j}, \dots, y_\ell + \sum_{j=1}^{\ell} \text{ex}(\gamma_j) \cdot z_{\ell,j} \right) = 0 \quad (33)$$

By taking a union bound over all  $\gamma_1, \dots, \gamma_\ell$  and adding the probability that Equation (28) does not hold for some  $\beta \in [q-1]^\ell$ , we get that with probability greater than 0 there exist  $z_{i,j}$ 's that satisfy all required constraints. ■

By combining Lemmas 2 and 6 we obtain that if  $f$  is  $\Omega\left(\frac{1}{\ell q^\ell}\right)$ -far from  $\text{POLY}_{n,d}$ , then  $\eta = \Omega\left(\frac{1}{\ell q^\ell}\right)$ , and so the algorithm rejects  $f$  with sufficiently high constant probability.

The next lemma, which will help us deal with the case in which  $\eta$  is small, is a variant of a very similar lemma that was proved in [AKK<sup>+</sup>03]. For the sake of completeness we provide its proof in the appendix.

**Lemma 7** *Let  $\zeta \stackrel{\text{def}}{=} \frac{1 - q^\ell \text{dist}(f,g)}{1 + q^\ell \text{dist}(f,g)} \cdot q^\ell \cdot \text{dist}(f,g)$ . If we uniformly and independently select  $y_0, y_1, \dots, y_\ell \in F^n$  then the probability that for exactly one choice of  $b_1, \dots, b_\ell \in F$ , we have that  $f\left(y_0 + \sum_{i=1}^{\ell} b_i \cdot y_i\right) \neq g\left(y_0 + \sum_{i=1}^{\ell} b_i \cdot y_i\right)$ , is at least  $\zeta$ .*

We are now ready to wrap-up the proof of Theorem 3.

**Proof of Theorem 3.** As we have noted previously, if  $f$  is in  $\text{POLY}_{n,d}$ , then by Theorem 1 the tester accepts (with probability 1). We next show that if  $f$  is  $\epsilon$ -far from  $\text{POLY}_{n,d}$ , then the tester rejects with probability at least  $\frac{2}{3}$ .

Suppose that  $\text{dist}(f, \text{POLY}_{n,d}) > \epsilon$ . If  $\eta < \frac{1}{2(\ell+1)q^{\ell+1}}$  then by Lemma 6 we know that  $g \in \text{POLY}_{n,d}$ . We claim that by Lemma 7 and Lemma 2, we also have that  $\eta \geq \frac{1}{3} \cdot q^\ell \cdot \text{dist}(f,g) > \frac{1}{3} \cdot q^\ell \cdot \epsilon$ . To verify this observe that if  $f$  and  $g$  disagree on exactly one point in a subspace  $S$  of dimension  $\ell$ , then  $f|_S \notin \text{POLY}(\ell, d)$ . In other words, by Lemma 7 and the definition of  $\eta$ , if  $0 < \eta < \frac{1}{2(\ell+1)q^{\ell+1}}$  then  $\eta \geq \zeta$  (where  $\zeta$  is as defined in Lemma 7). In particular, since (by Lemma 2)  $\text{dist}(f,g) \leq 2\eta \leq \frac{1}{(\ell+1)q^{\ell+1}}$  and  $\ell \geq 1$ , we get that  $\eta \geq \frac{1}{3} \cdot q^\ell \cdot \text{dist}(f,g)$ . Hence,  $\eta \geq \min\left\{\frac{1}{3}q^\ell\epsilon, \frac{1}{2(\ell+1)q^{\ell+1}}\right\}$ . Since it is enough to perform  $\Theta\left(\frac{1}{\eta}\right)$  rounds of the algorithm in order to detect a violation with probability at least  $\frac{2}{3}$ , the theorem follows. ■

## 5 A Lower Bound

**Theorem 4** *Every algorithm for testing  $\text{POLY}_{n,d}$  with distance parameter  $\epsilon$  must perform  $\Omega\left(\max\left\{\frac{1}{\epsilon}, q^{\ell-1}\right\}\right)$  queries when  $q$  is prime, and  $\Omega\left(\max\left\{\frac{1}{\epsilon}, q^{\lceil \ell/2 \rceil - 1}\right\}\right)$  queries otherwise.*

In order to establish Theorem 4, we consider the relation between polynomials and codes. Specifically, recall that the family  $\text{POLY}_{n,d}$  over a field  $F = \text{GF}(q) = \text{GF}(p^s)$ , corresponds to the Generalized Reed-Muller (GRM) code  $\mathcal{GRM}_q(d, n)$ . Namely, each codeword (having length  $q^n$ ) is determined by the evaluation of a polynomial in  $\text{POLY}_{n,d}$  on all points in the domain  $F^n$ . The minimum distance,  $\Delta(\mathcal{GRM}_q(d, n))$ , of the code is the following (cf. [DK00]): If  $d = r(q - 1) + t$ , where  $0 \leq t < q - 1$ , and  $r$  is an integer, then  $\Delta(\mathcal{GRM}_q(d, n)) = (q - t)q^{n-r-1}$ . The dual code of  $\Delta(\mathcal{GRM}(d, n))$  is the GRM code  $\mathcal{GRM}_q(n(q - 1) - (d + 1), n)$ , so that it has distance  $\Omega\left(q^{\left\lceil \frac{d+1}{q-1} \right\rceil - 1}\right)$ . Let us denote the distance of the dual code by  $\overline{\Delta}(\mathcal{GRM}_q(d, n))$ , and let  $\ell = \left\lceil \frac{d+1}{q-q/p} \right\rceil$  be as in our previous notation. Hence, if  $q$  is prime then  $\overline{\Delta}(\mathcal{GRM}_q(d, n)) = \Omega(q^{\ell-1})$ , and for non-prime  $q$  we can say that  $\overline{\Delta}(\mathcal{GRM}_q(d, n)) = \Omega(q^{\lceil \ell/2 \rceil - 1})$ .

Theorem 4 follows by applying the theorem below, which is a straightforward generalization of a similar theorem proved in [AKK<sup>+</sup>03] for binary codes. For the sake of completeness we include the proof in the appendix.

**Theorem 5** *Let  $\mathcal{F}$  be any family of functions  $f : F^n \rightarrow F$  that corresponds to a linear code  $\mathcal{C}$ . Let  $\Delta(\mathcal{C})$  denote the minimum distance of the code  $\mathcal{C}$  and let  $\overline{\Delta}(\mathcal{C})$  denote the minimum distance of the dual code of  $\mathcal{C}$ .*

*Every testing algorithm for the family  $\mathcal{F}$  must perform  $\Omega(\overline{\Delta}(\mathcal{C}))$  queries, and if the distance parameter  $\epsilon$  is at most  $\Delta(\mathcal{C})/(2q^n)$ , then  $\Omega(1/\epsilon)$  is also a lower bound for the necessary number of queries.*

## 6 The paper of Jutla et al. [JPRZ04]

As noted in the introduction, independently from our work, Jutla et al. [JPRZ04] give a testing algorithm for low-degree polynomials over prime fields. They too provide a characterization of low-degree polynomials (over prime fields) and define their test based on this characterization. As we discuss below, our characterization (in the case of prime fields) is related to the one in [JPRZ04]. However, our approach and hence the proofs for the characterizations are different, and in particular our characterization holds for all fields. Our testing algorithms and their analyses have a similar structure (which follows that of previous low-degree tests), but there are several technical and expositional differences (partly due to our unifying view of low-degree testing). We next discuss how the characterization in [JPRZ04] relates to ours.

Recall that we show that for  $F = \text{GF}(q)$  (where  $q = p^s$  and  $p$  is prime), a function  $f : F^n \rightarrow F$  is a polynomial of degree at most  $d$ , if and only if its restriction to every affine subspace of dimension  $\ell = \left\lceil \frac{d+1}{q-q/p} \right\rceil$  is a polynomial of degree at most  $d$ . In other words, if we consider the unique representation of each such restriction of  $f$  as a polynomial over  $\ell$  variables, then all coefficients that correspond to monomials having degree greater than  $d$  must be 0. Each such coefficient can be shown to equal a certain linear combination of the values of  $f$  in the subspace (see Section 2).

The characterization of Jutla et al. for prime fields ( $q = p$ ) is that one particular linear constraint over each subspace must hold. They do not approach the problem as we do (that is, by characterizing low-degree polynomials as functions whose restrictions to lower dimensional spaces are low-degree polynomials). However, translating their result using our terminology, it can be shown that the linear constraint they consider corresponds to the coefficient of exactly one

monomial of degree  $d + 1$ . This monomial has the following form:  $x_1^{q-1} \cdot x_2^{q-1} \cdots x_{\ell-1}^{q-1} \cdot x_\ell^t$ , where  $1 \leq t \leq q - 1$  and  $d + 1 = (\ell - 1)(q - 1) + t$ .

In retrospect we observe that our analysis can be slightly extended so as to show that in the case that  $q$  is prime then the characterization can be restricted to a single polynomial coefficient. When  $q$  is not prime then it is not clear whether the characterization can be restricted in a similar manner.

## 7 Sufficient Conditions for Local Testability

In this section we extract from our analysis certain sufficient conditions for the local testability of codes over finite fields. Let  $\mathcal{C}$  be a code and  $\bar{\mathcal{C}}$  its dual code. We shall view words in the two codes as functions. We assume that the domain of these functions is  $F^n$ . Our starting point is that the dual code  $\bar{\mathcal{C}}$  can be spanned by a subset  $Q \subset \bar{\mathcal{C}}$  of words, each having weight at most  $L$ . The *support* of a word  $w \in Q$ , denoted  $Sup(w)$ , is the set of all points  $y \in F^n$  such that  $w(y) \neq 0$ .

We say that  $Q$  is *affine-transitive* if the following conditions hold: (1) With each word  $w \in Q$  we can associate a subset of linearly independent points  $y_0, y_1, \dots, y_t \in F^n$  such that  $Sup(w) \subseteq S(y_0, y_1, \dots, y_t)$  and  $w(y_0) \neq 0$ . (2) Consider any other choice of linearly independent points  $z_0, z_1, \dots, z_t \in F^n$ , and any choice of  $b_1, \dots, b_t \in F$ . We require that the word  $w'$ , which is defined by  $w'(z_0 + \sum_{i=1}^t b_i \cdot z_i) = w(y_0 + \sum_{i=1}^t b_i \cdot y_i)$  and  $w'(z) = 0$  otherwise, belongs to  $Q$ .

This notion of transitivity lends itself to defining *classes* of words (where the classes are not necessarily disjoint). Specifically, for each choice of  $\vec{v} \in F^{q^t}$  where  $\vec{v}$  is indexed by all possible settings of  $b_1, \dots, b_t \in F$  and  $v_{0, \dots, 0} \neq 0$ , the class  $Q_{\vec{v}}$  consists of all words  $w$  for which there exists a choice of linearly independent  $y_0, y_1, \dots, y_t \in F^n$  that satisfies:  $w(y_0 + \sum_{i=1}^t b_i \cdot y_i) = v_{b_1, \dots, b_t}$ , for every  $b_1, \dots, b_t \in F$ . Note that if  $Q$  is transitive, then every non-empty class  $Q_{\vec{v}}$  must contain a word as defined above for *every* choice of linearly independent points  $y_0, y_1, \dots, y_t \in F^n$ .

For a non-empty class  $Q_{\vec{v}}$  we say that  $Q_{\vec{v}}$  is *R-randomly-closed under differences* for some integer  $R$ , if the following condition holds. For each point  $y \in F^n$ , let  $Q_{\vec{v}}(y) \subset Q_{\vec{v}}$  consist of all words  $w \in Q_{\vec{v}}$  that satisfy:  $Sup(w) \subset S(y, y_1, \dots, y_t)$  for every choice of  $y_1, \dots, y_t \in F^n$  such that  $y, y_1, \dots, y_t$  are linearly independent. For each  $y$ , consider any two words  $w, w' \in Q_{\vec{v}}(y)$ . Then  $w - w'$  can be represented as a sum of at most  $R$  other words  $u_1(w, w'), \dots, u_R(w, w') \in Q$  for which the following holds. Suppose we fix  $y$ , and select at random  $w, w' \in Q_{\vec{v}}(y)$ . Then for some class  $Q_{\vec{v}'}$  each  $u_i(w, w')$  is uniformly distributed in  $Q_{\vec{v}'}$ .<sup>7</sup>

**Theorem 6** *Let  $\mathcal{C}$  be a code and  $\bar{\mathcal{C}}$  its dual code. Suppose that  $\bar{\mathcal{C}}$  is spanned by a subset of words  $Q$  all having weight at most  $L$ , where  $Q$  is affine-transitive and every class in  $Q$  contains at least  $\theta \cdot |Q|$  words. If there exists a class  $Q_{\vec{v}^*} \subseteq Q$  that is  $R$ -randomly-closed under differences, then  $\mathcal{C}$  is locally testable using  $O\left(\frac{1}{\theta} \cdot (L^2 \cdot R + \frac{1}{\epsilon})\right)$  queries.*

We shall prove Theorem 6 by establishing the correctness of the following algorithm. In what follows we shall sometimes view the word we want to test for membership in  $\mathcal{C}$  as a function  $f : F^n \rightarrow F$ , and sometimes as a string  $f \in F^{q^n}$ .

<sup>7</sup>This last requirement can be generalized to allow other distributions, but for simplicity we state it in this way.

**Algorithm 2** Local Testing Algorithm for  $\mathcal{C}$

1. Repeat the following  $\Theta\left(\frac{1}{\epsilon\theta L} + \frac{L \cdot R}{\theta}\right)$  times:
  - (a) Uniformly and independently select a word  $w$  from  $Q$ .
  - (b) If  $w \cdot f \neq 0$  then output reject.
2. If no step caused rejection then output accept.

We note that instead of selecting a single random word  $w$  from  $Q$  in each iteration, we could instead consider several words, analogously to what is done in Algorithm 1. Namely, we could select  $t + 1$  points  $y_0, \dots, y_t \in F^n$  and check that all words  $w$  such that  $\text{Sup}(w) \subseteq S(y_0, \dots, y_t)$  are orthogonal to  $f$ . In some cases this would actually be more efficient in terms of the query complexity, but we have chosen to present the simpler test described above.

We turn to proving Theorem 6 by adapting the proof of Theorem 3. We start by modifying the definitions and notations presented in that proof. Let

$$\eta \stackrel{\text{def}}{=} \Pr_{w \in Q}[w \cdot f \neq 0] \quad (34)$$

be the probability that the test rejects in any single step.

For  $\vec{v}^*$  as determined in the theorem statement, let

$$U_{\vec{v}^*} \stackrel{\text{def}}{=} \{(b_1, \dots, b_t) \in F^t : v_{b_1, \dots, b_t}^* \neq 0\}, \quad (35)$$

and for any linearly independent  $y, y_1, \dots, y_t$ , let

$$V(y; y_1, \dots, y_t) = -\frac{1}{v_{0, \dots, 0}^*} \cdot \sum_{\substack{(b_1, \dots, b_t) \in U_{\vec{v}^*} \\ (b_1, \dots, b_t) \neq (0, \dots, 0)}} v_{b_1, \dots, b_t}^* \cdot f\left(y + \sum_{i=1}^t b_i \cdot y_i\right) \quad (36)$$

be the *vote* of  $y_1, \dots, y_t$  on the value of  $y$ . That is, we consider the word in  $Q_{\vec{v}^*}(y)$  that is determined by  $y, y_1, \dots, y_t$ , and compute the value that  $f(y)$  “should have” so that this word is orthogonal to  $f$  (given the values that  $f$  gives to all other points in the support of the word). Note that here, by the premise of the theorem concerning the classes in  $Q$ ,

$$\eta \geq \theta \cdot \Pr_{y, y_1, \dots, y_t}[V(y; y_1, \dots, y_t) \neq f(y)]. \quad (37)$$

Finally, we define  $g : F^n \rightarrow F$  as follows:

$$g(y) = \operatorname{argmax}_{a \in F} \left\{ \Pr_{y_1, \dots, y_t \in F^n}[V(y; y_1, \dots, y_t) = a] \right\}. \quad (38)$$

Given the definition of  $g$ , here we can show that

$$\operatorname{dist}(f, g) \leq 2\eta/\theta. \quad (39)$$

Let

$$\gamma(y) \stackrel{\text{def}}{=} \Pr_{y_1, \dots, y_t}[V(y; y_1, \dots, y_t) = g(y)] \quad (40)$$

and

$$\delta(y) \stackrel{\text{def}}{=} \Pr_{y_1, \dots, y_t, z_1, \dots, z_t}[V(y; y_1, \dots, y_t) = V(y; z_1, \dots, z_t)] \quad (41)$$

so that by Lemma 4,  $\gamma(y) \geq \delta(y)$ . Here we shall show that

$$\gamma(y) \geq 1 - R \cdot \eta / \theta \quad (42)$$

by showing that  $\delta(y) \geq 1 - R \cdot \eta / \theta$ . But the latter readily follows from the fact that  $Q_{\bar{v}^*}$  is  $R$ -closed under differences. To see why this is true, for each choice of  $y_1, \dots, y_t$  and  $z_1, \dots, z_t$ , consider the two words  $w = w_{y, y_1, \dots, y_t} \in Q_{\bar{v}^*}$  and  $w' = w_{y, z_1, \dots, z_t} \in Q_{\bar{v}^*}$ , where  $w_{y, y_1, \dots, y_t}$  denotes the word that has value  $v_{b_1, \dots, b_t}^*$  on  $y + \sum_{i=1}^t b_i \cdot y_i$  and is 0 elsewhere, and  $w_{y, z_1, \dots, z_t}$  is defined analogously. Then, by the premise of the theorem concerning  $Q_{\bar{v}^*}$ , the difference between these two words is a sum of  $R$  words,  $u_1, \dots, u_R$ , where when  $y_1, \dots, y_t$  and  $z_1, \dots, z_t$  are chosen uniformly at random, each of these words is uniformly distributed in one of the classes of  $Q$ . But this implies that the probability that  $V(y; y_1, \dots, y_t) - V(y; z_1, \dots, z_t) = 0$  (which is the same as  $(w - w') \cdot f = 0$ ), is lower bounded by the probability that all these words are orthogonal to  $f$ . Since for each  $u_i$ , the probability that  $u_i \cdot f \neq 0$  is at most  $\eta / \theta$ , by a union bound, the probability of this event is at least  $1 - R \cdot \eta / \theta$ .

It remains to prove that if  $\eta$  is sufficiently small, then  $w \cdot g = 0$  for every  $w \in Q$ , from which it follows that  $g \in \mathcal{C}$ . Let us fix any word  $w \in Q$ , and let  $y_0, y_1, \dots, y_t$  be such that  $\text{Sup}(w) \subseteq S(y_0, y_1, \dots, y_t)$ . Let  $U(w) = \left\{ (a_1, \dots, a_t) \in F^n : y_0 + \sum_{i=1}^t a_i \cdot y_i \in \text{Sup}(w) \right\}$  and for each  $(a_1, \dots, a_t) \in U(w)$ , let  $v(w)_{a_1, \dots, a_t} = w(y_0 + \sum_{i=1}^t a_i \cdot y_i)$  (so that  $w \in Q_{\bar{v}(w)}$ ). Here we consider  $(t+1) \cdot t$  elements in  $F^n$ , denoted  $\{z_{i,j}\}_{i=0, \dots, t}^{j=1, \dots, t}$  (which will be selected uniformly and at random). Assume first that for every choice of  $(a_1, \dots, a_t) \in U(w)$

$$g\left(y_0 + \sum_{i=1}^t a_i \cdot y_i\right) = V\left(y_0 + \sum_{i=1}^t a_i \cdot y_i ; z_{0,1} + \sum_{i=1}^t a_i \cdot z_{i,1}, \dots, z_{0,t} + \sum_{i=1}^t a_i \cdot z_{i,t}\right) \quad (43)$$

If we select the elements  $\{z_{i,j}\}$  uniformly and at random, then by Equation (42) and a union bound, this event occurs with probability at least  $1 - L \cdot R \cdot \eta / \theta$ . We assume from this point on that Equation (43) holds for every such choice of  $a_1, \dots, a_t \in U(w)$ .

By definition of  $V(\cdot)$  (and the above assumption), for every choice of  $(a_1, \dots, a_t) \in U(w)$ ,

$$\begin{aligned} & V\left(y_0 + \sum_{i=1}^t a_i \cdot y_i ; z_{0,1} + \sum_{i=1}^t a_i \cdot z_{i,1}, \dots, z_{0,t} + \sum_{i=1}^t a_i \cdot z_{i,t}\right) \\ &= -\frac{1}{v_{0, \dots, 0}^*} \cdot \sum_{\substack{(b_1, \dots, b_t) \in U_{\bar{v}^*} \\ (b_1, \dots, b_t) \neq (0, \dots, 0)}} v_{b_1, \dots, b_t}^* \cdot f\left(y_0 + \sum_{i=1}^t a_i \cdot y_i + \sum_{j=1}^t b_j \cdot \left(z_{0,j} + \sum_{i=1}^t a_i \cdot z_{i,j}\right)\right) \end{aligned} \quad (44)$$

$$= -\frac{1}{v_{0, \dots, 0}^*} \cdot \sum_{\substack{(b_1, \dots, b_t) \in U_{\bar{v}^*} \\ (b_1, \dots, b_t) \neq (0, \dots, 0)}} v_{b_1, \dots, b_t}^* \cdot f\left(y_0 + \sum_{j=1}^t b_j \cdot z_{0,j} + \sum_{i=1}^t a_i \cdot \left(y_i + \sum_{j=1}^t b_j \cdot z_{i,j}\right)\right) \quad (45)$$

Therefore,

$$\begin{aligned}
w \cdot g &= \sum_{(a_1, \dots, a_t) \in U(w)} w \left( y_0 + \sum_{i=1}^t a_i \cdot y_i \right) \cdot g \left( y_0 + \sum_{i=1}^t a_i \cdot y_i \right) \\
&= -\frac{1}{v_{0, \dots, 0}^*} \cdot \sum_{\substack{(b_1, \dots, b_t) \in U_{\vec{v}^*} \\ (b_1, \dots, b_t) \neq (0, \dots, 0)}} v_{b_1, \dots, b_t}^* \\
&\quad \cdot \sum_{(a_1, \dots, a_t) \in U(w)} v(w)_{a_1, \dots, a_t} \cdot f \left( y_0 + \sum_{j=1}^t b_j \cdot z_{0,j} + \sum_{i=1}^t a_i \cdot \left( y_i + \sum_{j=1}^t b_j \cdot z_{i,j} \right) \right)
\end{aligned}$$

That is,  $w \cdot g$  is a sum over at most  $L$  terms, where each term is of the form  $-\frac{v_{b_1, \dots, b_t}^*}{v_{0, \dots, 0}^*}$  times the inner product of a word in  $Q_{\vec{v}(w)}$  with  $f$ . If we select the  $z_{i,j}$ 's uniformly at random, then these words are uniformly distributed in  $Q_{\vec{v}(w)}$ . For each one of them, the probability that its inner product with  $f$  is non-zero is at most  $\eta \cdot |Q|/|Q_{\vec{v}(w)}| \geq \eta/\theta$ . By a union bound, the probability that they are all orthogonal to  $f$  is at least  $1 - L \cdot \eta/\theta$ . We thus obtain that with probability at least  $1 - L \cdot \eta \cdot (R+1)/\theta$  over the choice of the  $z_{i,j}$ 's, all required constraints are satisfied to ensure that  $w \cdot g = 0$ . Hence, for  $\eta < \frac{\theta}{L \cdot (R+1)}$ , there exists a choice of  $z_{i,j}$ 's that satisfies all constraints, implying that  $w \cdot g = 0$  as required.

**Wrapping up the proof of Theorem 6.** By combining Equation (39) with the fact that for  $\eta < \frac{\theta}{L \cdot (R+1)}$  we have that  $g \in \mathcal{C}$ , we obtain that if  $f$  is  $\Omega\left(\frac{\theta}{L \cdot R}\right)$ -far from  $\mathcal{C}$ , then  $\eta = \Omega\left(\frac{\theta}{L \cdot R}\right)$ , and so the algorithm rejects  $f$  with sufficiently high constant probability.

The next lemma is a variant of Lemma 7 (which, as noted before, is a variant of a lemma from [AKK<sup>+</sup>03]). Let  $Q_{\vec{v}} \subseteq Q$  be a class whose words are of length  $L$  (there must be such a class, otherwise  $L$  is not the minimum weight of words in  $Q$ )

**Lemma 8** *Let  $\zeta \stackrel{\text{def}}{=} \frac{1-L \cdot \text{dist}(f,g)}{1+L \cdot \text{dist}(f,g)} \cdot L \cdot \text{dist}(f,g)$ . If we uniformly select a word  $w \in Q_{\vec{v}}$ , then the probability that for exactly one point  $y \in \text{Sup}(w)$  we have that  $f(y) \neq g(y)$  is at least  $\zeta$ .*

**Proof of Theorem 6.** If  $f \in \mathcal{C}$ , then the tester accepts (with probability 1). We next show that if  $f$  is  $\epsilon$ -far from  $\mathcal{C}$ , then the tester rejects with probability at least  $\frac{2}{3}$ .

Suppose that  $\text{dist}(f, \mathcal{C}) > \epsilon$ . If  $\eta < \frac{\theta}{4L \cdot (R+1)}$ , then, as we have shown above,  $g \in \mathcal{C}$ . We claim that by Lemma 8 and by Equation (39), we also have that  $\eta = \Omega(\theta L \cdot \text{dist}(f, g)) = \Omega(\epsilon \theta L)$ . To verify this, consider the class  $Q_{\vec{v}} \subseteq Q$  whose words have weight  $L$ . Observe that if  $f$  and  $g$  disagree on exactly one point in  $\text{Sup}(w)$ , then  $w \cdot f \neq 0$ . Hence, by Lemma 8 and the definition of  $\eta$ , if  $0 < \eta < \frac{\theta}{L \cdot (R+1)}$  then  $\eta/\theta \geq \zeta$ . Next we establish a bound on  $\zeta$  so as to bound  $\eta$ . By Equation (39),  $\text{dist}(f, g) \leq 2\eta/\theta \leq \frac{2}{4L \cdot (R+1)} \leq \frac{1}{2L}$ , we get that  $\zeta > \frac{1}{3}L \cdot \text{dist}(f, g)$  and thus,  $\eta/\theta \geq \frac{1}{3}L \cdot \text{dist}(f, g)$ . Hence,  $\eta \geq \min \left\{ \frac{1}{3}\theta L \epsilon, \frac{\theta}{L \cdot (R+1)} \right\}$ . Since the algorithm performs  $\Theta\left(\frac{1}{\epsilon \theta L} + \frac{L \cdot R}{\theta}\right) = \Omega(1/\eta)$  iterations of selecting a word  $w$  from  $Q$  and checking whether  $w \cdot f \neq 0$ , the theorem follows. ■

## Acknowledgments

We are greatly indebted to Madhu Sudan who suggested the unifying view for testing polynomials over finite fields that we apply in this work. We would also like to thank Simon Litsyn, Alex Samorodnitsky, and Adam Smith for helpful discussions.

## References

- [AKK<sup>+</sup>03] N. Alon, M. Krivelevich, T. Kaufman, S. Litsyn, and D. Ron. Testing low-degree polynomials over  $\text{GF}(2)$ . In *Proceedings of the Seventh Annual Workshop on Randomization and Approximation Techniques in Computer Sciences (RANDOM)*, pages 188–199, 2003.
- [Aro94] S. Arora. *Probabilistic checking of proofs and the hardness of approximation problems*. PhD thesis, UC Berkeley, 1994.
- [AS97] S. Arora and M. Sudan. Improved low-degree testing and its applications. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing*, pages 485–495, 1997.
- [BCH<sup>+</sup>96] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, and M. Sudan. Linearity testing over characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.
- [BFL91] L. Babai, L. Fortnow, and C. Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991.
- [BFLS91] L. Babai, L. Fortnow, L. Levin, and M. Szegedy. Checking computations in polylogarithmic time. In *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*, pages 21–31, 1991.
- [BGLR93] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. Efficient probabilistically checkable proofs and applications to approximation. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on the Theory of Computing*, pages 294–304, 1993.
- [BLR93] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1993.
- [BS94] M. Bellare and M. Sudan. Improved non-approximability results. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on the Theory of Computing*, pages 184–193, 1994.
- [BSGH<sup>+</sup>04] E. Ben-Sasson, O. Goldreich, P. Harsha, M. Sudan, and S. Vadhan. Robust PCPs of proximity, shorter PCPs and applications to coding. In *Proceedings of the Thirty-Sixth Annual ACM Symposium on the Theory of Computing*, pages 1–10, 2004.
- [BSHR03] E. Ben-Sasson, P. Harsha, and S. Raskhodnikova. 3CNF properties are hard to test. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on the Theory of Computing*, pages 345–354, 2003.

- [BSS03] L. Babai, A. Shpilka, and D. Stefankovic. Locally testable cyclic codes. In *Proceedings of the Forty-Fourth Annual Symposium on Foundations of Computer Science*, pages 116–125, 2003.
- [BSSVW03] E. Ben-Sasson, M. Sudan, S. Vadhan, and A. Wigderson. Derandomizing low degree tests via epsilon-biased spaces. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on the Theory of Computing*, pages 612–621, 2003.
- [DK00] P. Ding and J.D. Key. Minimum-weight codewords as generators of generalized Reed-Muller codes. *IEEE Transactions on Information Theory*, 46(6):2152–2157, 2000.
- [FGL<sup>+</sup>96] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Approximating clique is almost NP-complete. *Journal of the Association for Computing Machinery*, pages 268–292, 1996.
- [FS95] K. Friedl and M. Sudan. Some improvements to total degree tests. In *Proceedings of the 3rd Annual Israel Symposium on Theory of Computing and Systems*, pages 190–198, 1995. Corrected version available online at <http://theory.lcs.mit.edu/~madhu/papers/friedl.ps>.
- [GGR98] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the Association for Computing Machinery*, 45(4):653–750, 1998.
- [GLR<sup>+</sup>91] P. Gemmell, R. Lipton, R. Rubinfeld, M. Sudan, and A. Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*, pages 32–42, 1991.
- [GR99] O. Goldreich and D. Ron. A sublinear bipartite tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999.
- [GR02] O. Goldreich and D. Ron. Property testing in bounded degree graphs. *Algorithmica*, pages 302–343, 2002.
- [GS02] O. Goldreich and M. Sudan. Locally testable codes and PCPs of almost-linear length. In *Proceedings of the Forty-Third Annual Symposium on Foundations of Computer Science*, pages 13–22, 2002.
- [JPRZ04] C. S. Jutla, A. C. Patthak, A. Rudra, and D. Zuckerman. Testing low-degree polynomials over prime fields. In *Proceedings of the Forty-Fifth Annual Symposium on Foundations of Computer Science*, 2004.
- [KKR03] T. Kaufman, M. Krivelevich, and D. Ron. Tight bounds for testing bipartiteness in general graphs. In *Proceedings of the Seventh Annual Workshop on Randomization and Approximation Techniques in Computer Sciences (RANDOM)*, pages 341–353, 2003.
- [MS77] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error Correcting Codes*. North Holland, 1977.

- [PS94] A. Polishchuk and D. Spielman. Nearly-linear size holographic proofs. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on the Theory of Computing*, pages 194–203, 1994.
- [RS96] R. Rubinfeld and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.
- [SW04] A. Shpilka and A. Wigderson. Derandomizing homomorphism testing in general groups. In *Proceedings of the Thirty-Sixth Annual ACM Symposium on the Theory of Computing*, pages 427–435, 2004.

## A Proofs of Claim 1, Lemma 7 and Theorem 5

**Claim 1** *Let  $q = p^s$  for a prime number  $p$  and an integer  $s$ , and let  $r$  and  $t$  be integers that satisfy  $0 < r \leq t \leq q - 1$ . If  $r = kp^{s-1}$  for some integer  $k$  then  $\binom{t}{r}$  is not divisible by  $p$ .*

**Proof:** For any positive integer  $j$ , the largest power of  $p$  that divides  $j!$  is

$$\lfloor j/p \rfloor + \lfloor j/p^2 \rfloor + \lfloor j/p^3 \rfloor + \dots .$$

But for  $r = kp^{s-1}$ , the identity  $\lfloor n/p^i \rfloor = \lfloor r/p^i \rfloor + \lfloor (n-r)/p^i \rfloor$  holds. Thus the largest power of  $p$  that divides  $n!$  is

$$\sum_{i=1}^{\infty} \lfloor n/p^i \rfloor = \sum_{i=1}^{\infty} (\lfloor r/p^i \rfloor + \lfloor (n-r)/p^i \rfloor) .$$

Therefore  $n!$  and  $r!(n-r)!$  are divisible by exactly the same power of  $p$ . ■

**Lemma 7** *Let  $\zeta \stackrel{\text{def}}{=} \frac{1-q^\ell \text{dist}(f,g)}{1+q^\ell \text{dist}(f,g)} \cdot q^\ell \text{dist}(f,g)$ . If we uniformly and independently select  $y_0, y_1, \dots, y_\ell \in F^n$ , then the probability that for exactly one choice of  $b_1, \dots, b_\ell \in F$ , we have that  $f(y_0 + \sum_{i=1}^{\ell} b_i \cdot y_i) \neq g(y_0 + \sum_{i=1}^{\ell} b_i \cdot y_i)$ , is at least  $\zeta$ .*

**Proof:** For each  $\beta = \beta_1, \dots, \beta_\ell, \beta_i \in [q-1]$  let  $X_\beta$  be the indicator random variable whose value is 1 if and only if  $f(y_0 + \sum_{i=1}^{\ell} \text{ex}(\beta_i) y_i) \neq g(y_0 + \sum_{i=1}^{\ell} \text{ex}(\beta_i) y_i)$ . Obviously,  $\Pr[X_\beta = 1] = \text{dist}(f, g)$  for every  $\beta$ . It is not difficult to verify that the random variables  $X_\beta$  are pairwise independent. This is true since for any two distinct  $\beta^1, \beta^2$ , the points  $(y_0 + \sum_{i=1}^{\ell} \text{ex}(\beta_i^1) y_i)$  and  $(y_0 + \sum_{i=1}^{\ell} \text{ex}(\beta_i^2) y_i)$  attain each pair of distinct values in  $F^n$  with equal probability when the vectors  $y_i$  are chosen randomly and independently. It follows that the random variable  $X = \sum_{\beta} X_\beta$ , which counts the number of points  $v = (y_0 + \sum_{i=1}^{\ell} \text{ex}(\beta_i) y_i)$  in which  $f(v) \neq g(v)$ , has expectation  $\mathbb{E}[X] = q^\ell \cdot \text{dist}(f, g)$  and variance  $\text{Var}[X] = q^\ell \cdot \text{dist}(f, g) \cdot (1 - \text{dist}(f, g)) \leq \mathbb{E}[X]$ . Our objective is to lower bound the probability that  $X = 1$ . We need the well known fact that for a random variable  $X$  that attains nonnegative, integer values,

$$\Pr[X > 0] \geq \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]} . \quad (46)$$

Indeed, if  $X$  attains the value  $i$  with probability  $\nu_i$  for  $i > 0$ , then, by Cauchy-Schwartz,

$$(\mathbb{E}[X])^2 = \left( \sum_{i>0} i \nu_i \right)^2 = \left( \sum_{i>0} i \sqrt{\nu_i} \sqrt{\nu_i} \right)^2 \leq \left( \sum_{i>0} i^2 \nu_i \right) \cdot \left( \sum_{i>0} \nu_i \right) = \mathbb{E}[X^2] \cdot \Pr[X > 0] . \quad (47)$$

In our case, this implies

$$\Pr[X > 0] \geq \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]} \geq \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X] + (\mathbb{E}[X])^2} = \frac{\mathbb{E}[X]}{1 + \mathbb{E}[X]}. \quad (48)$$

Therefore

$$\mathbb{E}[X] \geq \Pr[X = 1] + \left( \frac{\mathbb{E}[X]}{1 + \mathbb{E}[X]} - \Pr[X = 1] \right) \cdot 2 = \frac{2\mathbb{E}[X]}{1 + \mathbb{E}[X]} - \Pr[X = 1], \quad (49)$$

implying that

$$\Pr[X = 1] \geq \frac{\mathbb{E}[X] - (\mathbb{E}[X])^2}{1 + \mathbb{E}[X]}. \quad (50)$$

Substituting the value of  $\mathbb{E}[X]$ , the desired result follows. ■

**Theorem 5** *Let  $\mathcal{F}$  be any family of functions  $f : F^n \rightarrow F$  that corresponds to a linear code  $\mathcal{C}$ . Let  $\Delta(\mathcal{C})$  denote the minimum distance of the code  $\mathcal{C}$  and let  $\overline{\Delta}(\mathcal{C})$  denote the minimum distance of the dual code of  $\mathcal{C}$ .*

*Every testing algorithm for the family  $\mathcal{F}$  must perform  $\Omega(\overline{\Delta}(\mathcal{C}))$  queries, and if the distance parameter  $\epsilon$  is at most  $\Delta(\mathcal{C})/(2q^n)$ , then  $\Omega(1/\epsilon)$  is also a lower bound for the necessary number of queries.*

**Proof:** We start by showing that  $\Omega(\overline{\Delta}(\mathcal{R}(d, n)))$  queries are necessary. A well known fact from coding theory (see [MS77, Chap. 1, Thm. 10]) states the following: for every linear code  $\mathcal{C}$  whose dual code has distance  $\overline{\Delta}(\mathcal{C})$ , if we examine a sub-word having length  $\Delta'$ , where  $\Delta' < \overline{\Delta}(\mathcal{C})$ , of a uniformly selected codeword in  $\mathcal{C}$ , then the resulting sub-word is uniformly distributed in  $F^{\Delta'}$ . Hence it is not possible to distinguish between a random codeword in  $\mathcal{C}$  and a random word in  $F^n$  (which with high probability is far from any codeword) using less than  $\overline{\Delta}$  queries.

We now turn to the case  $\epsilon < \Delta/2q^n$ . To prove the lower bound here, we apply, as usual, the Yao principle by defining two distributions, one of positive instances, and the other of negative ones, and then showing that in order to distinguish between those distributions any algorithm must perform  $\Omega(1/\epsilon)$  queries. The positive distribution has all its mass at the zero vector  $\vec{0} = (0, \dots, 0)$ . To define the negative distribution, partition the set of all coordinates randomly into  $t = 1/\epsilon$  nearly equal parts  $I_1, \dots, I_t$  and give weight  $1/t$  to each of the characteristic vectors  $w_i$  of  $I_i$ ,  $i = 1, \dots, t$ . (Observe that indeed  $\vec{0} \in \mathcal{C}$  due to linearity, and  $\text{dist}(w_i, \mathcal{C}) = \epsilon$  due to the assumption on the minimum distance of  $\mathcal{C}$ ). Finally, a random instance is generated by first choosing one of the distributions with probability  $1/2$ , and then generating a vector according to the chosen distribution.

Consider the two distributions that were defined. Let  $A$  be a deterministic testing algorithm with query complexity  $s$  (where  $s$  is a function of  $\epsilon$ ). We need to show that if  $A$  gives an incorrect answer with probability at most  $1/3$ , it must be that  $s > 1/(3\epsilon)$ . If  $A$  is incorrect on  $\vec{0}$  (that is, it does not accept it), then it is already incorrect with probability at least  $1/2$ . Otherwise  $A$  should accept the input if all the  $s$  queried bits are 0. Therefore it accepts as well at least  $t - s$  (where  $t = 1/\epsilon$  is as defined above) of the inputs  $w_i$ . This shows that  $A$  gives an incorrect answer with probability at least  $(t - s)/2t$ . for this to be smaller than  $1/3$  it must be the case that  $s > 1/(3\epsilon)$ . ■