Adversarial Low Degree Testing*

Dor Minzer† Kai Zhe Zheng‡

Abstract

In the t-online-erasure model in property testing, an adversary is allowed to erase t values of a queried function for each query the tester makes. This model was recently formulated by Kalemaj, Raskhodnikova and Varma, who showed that the properties of linearity of functions as well as quadraticity can be tested in $O_t(1)$ many queries: $O(\log(t))$ for linearity and $2^{2^{O(t)}}$ for quadraticity. They asked whether the more general property of low-degreeness can be tested in the online erasure model, whether better testers exist for quadraticity, and if similar results hold when "erasures" are replaced with "corruptions".

We show that, in the t-online-erasure model, for a prime power q, given query access to a function $f : \mathbb{F}_q^n \to \mathbb{F}_q$, one can distinguish in $\mathrm{poly}(\log^{d+q}(t)/\delta)$ queries between the case that f is degree at most d, and the case that f is δ-far from any degree d function (with respect to the fractional hamming distance). This answers the aforementioned questions and brings the query complexity to nearly match the query complexity of low-degree testing in the classical property testing model.

Our results are based on the observation that the property of low-degreeness admits a large and versatile family of query efficient testers. Our testers operates by querying a uniformly random, sufficiently large set of points in a large enough affine subspace, and finding a tester for low-degreeness that only utilizes queries from that set of points. We believe that this tester may find other applications to algorithms in the online-erasure model or other related models, and may be of independent interest.

1 Introduction

The main purpose of this paper is to further investigate the t-online-erasure model which was recently initiated by Kalemaj, Raskhodnikova, and Varma [13]. The motivation behind this model stems from scenarios where computations are performed on datasets that may contain erasures which can occur in places that are chosen by an adversary but are limited in their number. For instance, consider an algorithm aiming to detect fraud, while an adversarial party deliberately erases data to conceal the fraudulent activity.

The focus of this paper is the problem of *low degree testing* in the t-online erasure model. In this setting, an algorithm is given query access to a function $f : \mathbb{F}_q^n \to \mathbb{F}_q$, an erasure parameter t, a distance parameter δ, and a degree parameter d. At the start of the computation, all entries in f's truth table are unerased. An algorithm may query the value of $f(x)$ for $x \in \mathbb{F}_q^n$, but after each query, an adversary is allowed to *erase* any t-entries of f's truth table. If the algorithm attempts to query an entry $f(x)$ after it is erased, then $\perp$ is returned. The goal of the algorithm is to output accept with probability 1 if f is degree at most d and to output reject with probability at least $2/3$ if f is δ-far from low degree. In the standard property testing literature, the former stipulation is often referred to as completeness, while the latter is often referred to as soundness. Algorithms accomplishing this task in the t-online erasure model are called *t-online-erasure-resilient*.

Low degree testing is one of the most basic and well studied problems in property testing and can be traced all the way back to Blum, Luby, and Rubinfeld's seminal work on linearity testing [7]. Linearity testing can be thought of as the $d = 1$ case of low degree testing, and after the work of [7], further low degree tests were constructed and analyzed over the next decade - first for arbitrary degrees of univariate polynomials [22], and later for arbitrary degrees of multivariate polynomials [1, 15, 12, 6, 11, 21, 14, 17]. These tests were also relevant within coding theory

*The full version of the paper can be accessed at https://arxiv.org/pdf/2308.15441

†Department of Mathematics, Massachusetts Institute of Technology, Cambridge, USA. Supported by NSF CCF award 2227876 and NSF CAREER award 2239160.

‡Department of Mathematics, Massachusetts Institute of Technology, Cambridge, USA. Supported by the NSF GRFP DGE-2141064.

as polynomials (both univariate and multivariate) over finite fields are the basis of two of the most well known error-correcting codes - the Reed-Solomon and Reed-Muller Codes. Due to their efficient query complexity, low degree tests for multivariate polynomials also found many applications elsewhere in theoretical computer science - perhaps most notably to the construction Probabilistically Checkable Proofs [9, 3, 2, 23, 4]. The extensive work regarding low degree testing also led to the testing of other properties of functions over finite fields, and more generally the study of algebraic property testing [16].

Our work adds to the vast literature on low degree testing by giving the first such algorithms over all degrees and field sizes in the online erasure model. Previously, the only known results were due to [13] who gave algorithms for linearity and quadraticity ($d = 1, 2$) cases over the field $\mathbb{F}_2$.

We find that designing a tester in the online-erasure model requires one to depart significantly from classical testers. To see why, consider what happens when one tries to implement the Blum-Luby-Rubinfeld linearity tester of [7] over $\mathbb{F}_2$. The BLR tester is based on the observation that any linear $f : \mathbb{F}_2^n \to \mathbb{F}_2$ satisfies $f(x+y) = f(x) + f(y)$ for any $x, y \in \mathbb{F}_2^n$ and proceeds by choosing uniformly random $x, y \in \mathbb{F}_2^n$, querying $f(x), f(y), f(x+y)$, and finally checking if $f(x+y) = f(x) + f(y)$ is satisfied. The difficulties that arise in the online erasure model now become clear. After the algorithm queries $f(x)$ and $f(y)$, the adversary can erase the point $f(x+y)$ making it impossible to directly carry out the BLR test.

Faced with this difficulty, the authors of [13] noted that one can attempt to "overwhelm" the adversary's erasure capacity with volume. One can first query a large number, say $m = 100t$, of uniformly random points, $x_1, \ldots, x_m$ so that there are $\binom{m}{2} \geqslant 4000t^2$ possible $x_i + x_j$'s. At this point the adversary could have erased at most $100t^2$ of these sums, making it possible to successfully query a random $x_i + x_j$, and perform the check $f(x_i + x_j) = f(x_i) + f(x_j)$. This idea is adapted and refined into a tester for linearity over $\mathbb{F}_2$ by [13].

Dealing with online erasures becomes significantly more complicated for higher degrees, namely for $d \geqslant 2$. The classical test here proceeds by choosing a $d+1$-dimensional affine space V and then querying $f(x)$ for any $x \in V$, [1]. Given the large number of points needed as well as the large number of dependencies between these points, it is no longer clear how to craft enough strategies to obtain an entire $d+1$-dimensional affine subspace against an adversary. Indeed, the tester of [13] in the online erasure model is much more involved, and it is not clear how to generalize it to larger fields or higher degrees.

Our approach. Our algorithm forgoes the task of mimicking the classical low-degree testers by attempting to query all of the points in a subspace of $\mathbb{F}_q^n$ altogether. Instead, we show that given any suitable number of points inside of a fixed affine subspace of dimension large enough in terms of d and t, one has enough points to design a degree d tester. Thus, our algorithm circumvents the adversary by querying uniformly random points inside of some suitably large affine subspace. As it turns out, the adversary cannot foil our plan if there is no plan.

In other words, the main new observation behind our result is that the collection of local testers for the Reed-Muller code is versatile — so much so that we can fit a tester through a typical, small set of inputs. To the best of our knowledge, this observation is new even in the setting without erasures, and we believe that it may have other applications and be of independent interest.

1.1 Our Results We present our testers for low degree polynomials over prime fields and non-prime fields separately. Although the testers are essentially the same, the analysis in the non-prime field case requires more care and must be handled separately. This also leads to slightly different parameters for our two testers. Formally, our main theorems are the following.

THEOREM 1.1. (PRIME FIELD CASE) *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be the input function over a prime-field vector space, d be the degree parameter, and δ be the distance parameter. Then for $t \leqslant \frac{\delta}{30} p^{n/(20d)}$, there is a t-online-erasure resilient tester with query complexity $O\left(\frac{\left(\log^{3d+3}(t/\delta)\right)}{\delta}\right)$ satisfying:*

- ***Completeness:*** *If f is degree d then the algorithm outputs accept with probability* 1.
- ***Soundness:*** *If f is δ-far from degree d then the algorithm outputs reject with probability at least* $2/3$.

THEOREM 1.2. (NON-PRIME FIELD CASE) *Let $f : \mathbb{F}_q^n \to \mathbb{F}_q$ be the input function, d be the degree parameter,*

and δ be the distance parameter. Then for $t \leqslant \frac{\delta}{100} q^{\frac{n}{100(d+q)}-1}$, there is a t-online-erasure resilient tester with query complexity $\frac{q^{O(1)}}{\delta} O\left(\left(\log^{3d+3q}(t/\delta)\right)\right)$ satisfying:

- ***Completeness:** If f is degree d then the algorithm outputs accept with probability 1.*
- ***Soundness:** If f is δ-far from degree d then the algorithm outputs reject with probability at least $2/3$.*

In the $p = 2$ and $d = 1$ case, our algorithm is essentially the same as that of [13]. Furthermore, we remark that our testers also hold with two sided error in the t-online corruption model with two sided error. In this model, the adversary may change the value of points $f(x)$ instead of erasing them. Our algorithms are still effective in the online-corruption because no matter what points the adversary chooses to erase, the probability that our algorithm queries a corrupted point is bounded by a small constant. Therefore the corruptions only induce a small additive error to the completeness and soundness cases. Our result for corruptions follows from a reduction from [13], showing that online erasure-resilience implies online corruption-resilience if the probability of querying a corrupted point is small (see [13, Lemma 1.8] for a precise statement).

THEOREM 1.3. *The algorithms of Theorems 1.1 and 1.2 are also t-online-corruption resilient and satisfy:*

- ***Completeness:** If f is degree d then the algorithm outputs accept with probability $2/3$.*
- ***Soundness:** If f is δ-far from degree d then the algorithm outputs reject with probability at least $2/3$.*

In the online corruption model, one can never hope for perfect completeness as there is always a nonzero probability that a large enough number of queries made are corrupted, and there is no way for the algorithm to tell if a query has been corrupted or not. In contrast, in the erasure model, the algorithm can always tell if a query is erased as $\perp$ is returned.

1.2 Proof Overview The starting point of our algorithm is the observation that the degree of a function when restricted to a k-dimensional affine subspace can be tested using *any* set of points of size only polynomial in k. This result is obtained by considering inner product testers and reviewing properties of linear affine-invariant functions [16], which are explained in Section 2. Specifically, we rely on the fact that a degree d tester for functions over prime fields is sound if there exists even a single degree $d+1$ monomial that it can reject. Thus our task now becomes designing a tester that accepts all degree at most d monomials while rejecting at least one degree $d+1$ monomial. The view of testing as taking the inner product with functions that are dual to the desired property, as in is done in [21, 17], turns out to be useful for this task. This further reduces our algorithm to the following algebraic observation: for any $S \subset \mathbb{F}_q^k$ of size roughly k^d, there is a function $h : \mathbb{F}_q^k \to \mathbb{F}_q$ whose support is contained within S such that the inner product of h with any monomial of degree at most d is 0, and the inner product of h with some specific degree $d+1$ monomial is nonzero. This fact follows from counting/ dimension arguments, and armed with it the algorithm follows rather naturally.

The case of non-prime fields is a bit harder but is very similar in spirit. Here, we once again show that testing low-degreeness over $\mathbb{F}_q$ admits a versatile and large enough family of inner product testers, enough so that we can fit a test through every large enough set of points.

1.3 Related Work Prior to this work, the t-online erasure model was presented and studied in [13]. In [13], it is shown that the query complexity of linearity testing over $\mathbb{F}_2$ with erasure parameter t is $\Theta(\log(t))$, and that quadraticity testing can be accomplished using a number of queries that is doubly exponential in t. As such, the quadraticity testing result only applies for constant t and furthermore does not apply in the online corruption model. Other properties including the sorted and Lipschitz properties of sequences are also considered in [13] as examples that cannot be tested in the online-erasure model, even for $t = 1$.

The online erasure model we study is also similar to the offline erasure model of Dixit, Raskhodnikova, Thakurta, and Varma [8] where all erasures are made by the adversary at the start of the computation, as well as the tolerant testing model, where there is the additional requirement that functions close to the desired property should be accepted with constant probability [18].

In a concurrent work, Ben-Eliezer, Kelman, Meir and Raskhodnikova [5] investigate the online erasure model in the $q = 2$ case. Of interest to this paper are the following results. For linearity testing ($d = 1$), they obtain a tester with query complexity matching the lower bound in [13]. For general degree, d, they show that any tester requires at least $\Omega\left(\log^d(t)\right)$ queries.

For constant distance parameter δ, in the $q = 2$ case, our tester in Theorem 1.1 achieves query complexity $O(\log^{3d+3}(t))$. Thus, when compared to the lower bound of [5] the polynomial dependence in $\log^d(t)$ is optimal, but improving the exponent to match the lower bound is an intriguing open question. We remark that we did not attempt to optimize the query complexity of our testers and instead focused on obtaining a tester with $\mathrm{poly}(\log^d(t))$ query complexity that also worked for all degrees and field sizes. Optimizing just the analysis in this paper, however, will not close the gap to the lower bound. Instead, what appears to be missing is an "optimal analysis" of the soundness of our basic tester (see Algorithm 1), that makes $O(\log^{d+1}(t))$ queries. In this paper, we rely on a generic result from [16] to obtain soundness roughly $\Omega(\frac{1}{\log^{2d+2}(t)})$, which in turn requires us to repeat the basic test $O(\log^{2d+2}(t))$-times in order to obtain constant soundness. This ultimately results in $O(\log^{3d+3}(t))$ queries. If instead, we had an optimal analysis, similar to what is achieved for the testers in [17], then the soundness of the basic test would be $\Omega(1)$, and only $O(1)$ repetitions would be needed. Unfortunately the analysis of [17] does not apply to our testers.

2 Preliminaries

2.1 Notations For a prime power $q = p^\ell$, let $\mathbb{F}_q$ denote the finite field with q elements and $\mathbb{F}_q^*$ denote the nonzero elements of $\mathbb{F}_q$. When $\ell = 1$, then $q = p$ and we refer to $\mathbb{F}_q = \mathbb{F}_p$ as a prime field. For a fixed field, $\mathbb{F}_q$, let $\mathcal{T}_{n,k}$ denote the set of affine transformations $T : \mathbb{F}_q^k \to \mathbb{F}_q^n$. Though the definition of $\mathcal{T}_{n,k}$ also depends on a field parameter q, we will often drop it from the notation and it will always be clear from the context. For $T \in \mathcal{T}_{n,k}$ let $f \circ T$ denote the function over $\mathbb{F}_q^k$ given by $f \circ T(x) := f(T(x))$, for $x \in \mathbb{F}_q^k$. For a function $h : \mathbb{F}_q^k \to \mathbb{F}_q$, we define its support as the set of points z such that $h(z) \neq 0$.

For $\alpha \in \mathbb{F}_q^n$, and $e \in \{0, \ldots, q-1\}^n$, let $\alpha^e = \prod_{i=1}^n \alpha_i^{e_i}$, where α_i and e_i are the ith coordinates of α and e respectively. Let x^e denote the monomial $\prod_{i=1}^n x_i^{e_i}$ where the x_i's are variables and the number of variables parameter n will be clear from context. The degree of x^e is $\deg(x^e) = |e|_1 = \sum_{i=1}^n e_i$. For a function $f : \mathbb{F}_q^n \to \mathbb{F}_q$, recall that it has a unique polynomial expression, $f(x) = \sum_{e \in \{0,\ldots,q-1\}^n} C_e x^e$, where each coefficient $C_e \in \mathbb{F}_q$. The degree of f is the maximum degree of a monomial appearing in its expansion, or equivalently $\deg(f) = \max_{e:C_e \neq 0} |e|_1$. For each $C_e \neq 0$, we say that the monomial x^e *appears* in f or that f *contains* x^e.

As mentioned, the set of degree d polynomials is often referred to as the degree d Reed-Muller Code. We use the notation $\mathrm{RM}[n, q, d] = \{f : \mathbb{F}_q^n \to \mathbb{F}_q \mid \deg(f) \leqslant d\}$. The notion of distance that we use is fractional hamming distance. That is, for two functions f, g over $\mathbb{F}_q^n$, define $\delta(f, g) = \Pr_{x \in \mathbb{F}_q^n}[f(x) \neq g(x)]$. The distance of f to degree d, is $\delta_d(f) = \min_{g \in \mathrm{RM}[n,q,d]} \delta(f, g)$, and we say that f is δ-far from degree d if $\delta_d(f) \geqslant \delta$.

2.2 Properties of Linear Affine Invariant Families We require a few basic facts about linear, affine invariant families of functions, which we present in this section. These results can all be found or derived from [16]. For a family of polynomials $\mathcal{F}$, let

$$\mathrm{MonomialSupp}(\mathcal{F}) = \{x^e \mid e \in \{0, \ldots, q-1\}^n,\ \exists f \in \mathcal{F} \text{ containing } x^e\}$$

denote the set of monomials that appear in at least one of these polynomials. It is well known that these monomials form a basis for $\mathcal{F}$ if $\mathcal{F}$ is *linear* and *affine-invariant*. These notions are defined as follows.

DEFINITION 1. *A family of functions, $\mathcal{F} \subset \{f : \mathbb{F}_q^n \to \mathbb{F}_q\}$ is linear if*

- $0 \in \mathcal{F}$, *where* 0 *is the constant zero function.*
- *For any $\alpha, \beta \in \mathbb{F}_q$, if $f, g \in \mathcal{F}$, then $\alpha \cdot f + \beta \cdot g \in \mathcal{F}$.*

DEFINITION 2. *A family of functions, $\mathcal{F} \subset \{f : \mathbb{F}_q^n \to \mathbb{F}_q\}$, is called affine invariant if it is closed under compositions with affine transformations. That is, if $f \in \mathcal{F}$ and $T \in \mathcal{T}_{n,n}$, then $f \circ T \in \mathcal{F}$.*

Linear, affine-invariant function families have the following nice properties. The first asserts that affine invariant codes are defined by the set of monomials, $\mathrm{MonomialSupp}(\mathcal{F})$.

LEMMA 2.1. (MONOMIAL EXTRACTION LEMMA [16]) *If $\mathcal{F}$ is a linear, affine-invariant family of polynomials then $\mathcal{F} = \mathrm{span}(\mathrm{MonomialSupp}(\mathcal{F}))$.*

The next property is called p-shadow closed, and requires some additional notions. For two integers $a, b \in \{0, \ldots, q-1\}$, let $a = \sum_{i=0}^{\ell-1} p^i a_i$ and $b = \sum_{i=0}^{\ell-1} p^i b_i$ be their base p representations. We say a is in the p-shadow of b if $a_i \leqslant b_i$ for $i = 0, \ldots, \ell-1$, and denote this by $a \leqslant_p b$. Then for two exponent vectors $e = (e_1, \ldots, e_n)$ and $e' = (e'_1, \ldots, e'_n)$, we say $e \leqslant_p e'$ if $e_i \leqslant_p e'_i$ for every i. Linear, affine-invariant families of polynomials have the following shadow closed property.

LEMMA 2.2. *Let $\mathcal{F} = \mathrm{span}(\mathrm{MonomialSupp}(\mathcal{F}))$ be a linear, affine invariant family of polynomials. If $e \leqslant_p e'$ and $x^{e'} \in \mathrm{MonomialSupp}(\mathcal{F})$, then $x^e \in \mathrm{MonomialSupp}(\mathcal{F})$ as well.*

In the case where $q = p$, the shadow lemma simplifies to the following statement.

LEMMA 2.3. *Let $\mathcal{F} = \mathrm{span}(\mathrm{MonomialSupp}(\mathcal{F}))$ be a linear, affine invariant family of polynomials over a prime field vector space. If $x^e \in \mathrm{MonomialSupp}(\mathcal{F})$ and each entry of $e - e'$ is nonnegative, then $x^{e'} \in \mathrm{MonomialSupp}(\mathcal{F})$.*

Finally, we will need the a lemma which will allow us to go from one monomial in $\mathcal{F}$ to another with the same degree, but with the distribution of the individual degrees shifted. In order to show this lemma, we will need Lucas' Theorem.

THEOREM 2.1. (LUCAS' THEOREM) *For any r_1 and r_2, we have $\binom{r_2}{r_1}$ is nonzero modulo p if and only if $r_1 \leqslant_p r_2$.*

LEMMA 2.4. *Suppose $x^e \in \mathcal{F}$ and suppose $m \leqslant_p e_2$. Then $x^{e'} \in \mathcal{F}$, where $e' = (e_1 + m, e_2 - m, e_3, \ldots, e_n)$.*

Proof. Let T be the affine transformation $T(x) = (x_1, x_1 + x_2, x_3, \ldots, x_n)$. Then,

$$x^e \circ T = x_1^{e_1}(x_1 + x_2)^{e_2} \prod_{j=3}^{n} x_j^{e_j} = \left(\sum_{i=0}^{e_2} \binom{d_2}{i} x_1^{e_1+i} x_2^{e_2-i} \right) \prod_{j=3}^{n} x_j^{e_j}.$$

By Lucas's Theorem and the assumption $m \leqslant_p e_2$, $\binom{d_2}{m} \neq 0$ in $\mathbb{F}_q$, and so $x^e \circ T = x_1^{e_1}(x_1+x_2)^{e_2} \prod_{j=3}^{n} x_j^{e_j}$ contains the monomial $x^{e'}$. The result then follows from Lemma 2.1. ∎

Clearly Lemma 2.4 also holds with the indices 1 and 2 replaced by arbitrary, distinct indices i and j.

2.3 Functions over Finite Fields We now present some additional facts about functions over finite fields. These facts will lead to the inner-product view of testing which was used to construct low-degree tests in [21, 17]. For two functions $f, g : \mathbb{F}_q^n \to \mathbb{F}_q$, define their inner product as

$$\langle f, g \rangle = \sum_{\alpha \in \mathbb{F}_q^n} f(\alpha) g(\alpha).$$

It is clear that this inner product is bi-linear. A key part of our testing algorithm is the observation that the presence of high degree monomials can be "detected" by using inner products with polynomials of relatively sparse support. In order to see this, we will first need a couple of basic facts about finite fields. First, it is a well known fact that $\mathbb{F}_q^*$ has a multiplicative generator. From this, we can deduce the following two lemmas.

LEMMA 2.5. *For any prime power q and integer $i \in \{0, \ldots, q-1\}$,*

$$\sum_{\alpha \in \mathbb{F}_q} \alpha^i = \begin{cases} -1, & \text{if } i = q-1, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. If $i = q-1$, then $\alpha^i = 1$ for all $\alpha \neq 0$, while $0^i = 0$. Therefore, the sum is one summed up $q-1$ times which is -1 in $\mathbb{F}_q$. For $i \in \{1, \ldots, q-2\}$, recall that $\mathbb{F}_q^*$ has a generator γ. That is, $\mathbb{F}_q^* = \{1, \gamma, \ldots, \gamma^{q-2}\}$. Since $\gamma^i \neq 1$, we may write

$$\sum_{\alpha \in \mathbb{F}_q} \alpha^i = \sum_{j=0}^{q-2} (\gamma^i)^j = \frac{(\gamma^i)^{q-1} - 1}{\gamma^i - 1} = \frac{1-1}{\gamma^i - 1} = 0.$$

On the other hand, if $i = 0$ then the sum on the left hand side of the lemma is equal to $\sum_{\alpha \in \mathbb{F}_q} \alpha^0 = \sum_{\alpha \in \mathbb{F}_q} 1 = q = 0$. ∎

LEMMA 2.6. *For $e, e' \in \{0, \ldots, q-1\}^n$, suppose that $x^{e''} = x^e \cdot x^{e'}$ where $x^{e''}$ is reduced so that its individual degrees are in $\{0, \ldots, q-1\}$.*

$$\langle x^e, x^{e'} \rangle = \begin{cases} (-1)^n, & \text{if } e'' = (q-1, \ldots, q-1), \\ 0, & \text{otherwise}, \end{cases}$$

Proof. This is a straightforward application of Lemma 2.5. By definition

$$\langle x^e, x^{e'} \rangle = \sum_{(\alpha_1, \ldots, \alpha_n) \in \mathbb{F}_q^n} \prod_{i=1}^n \alpha_i^{e_i + e_i'} = \prod_{i=1}^n \sum_{\alpha \in \mathbb{F}_q} \alpha^{e_i''}.$$

The result follows from Lemma 2.5. ∎

Lemma 2.6 suggest one method for using inner products to test low degree. Suppose that $q = 2$ and we want to detect if f contains the degree $d+1$ monomial $x_1 \cdots x_{d+1}$. Then by Lemma 2.6 this is equivalent to checking whether or not $\langle f, x_{d+2} \cdots x_n \rangle$ is nonzero.

2.4 Local Characterizations and Testing Using the ideas from the last section, we can first attempt to find functions $h : \mathbb{F}_q^k \to \mathbb{F}_q$ that are *local characterizations* of $\mathrm{RM}[n, q, d]$. These are defined as follows.

DEFINITION 3. *For $h : \mathbb{F}_q^k \to \mathbb{F}_q$, define $\mathcal{F}_n(h) = \{f : \mathbb{F}_q^n \to \mathbb{F}_q \mid \langle f \circ T, h \rangle = 0, \forall T \in \mathcal{T}_{n,k}\}$. We call h a local characterization for $\mathcal{F}_n(h)$.*

Whenever we use the notation $\mathcal{F}_n(h)$, the field size q will be clear from context. Once we are able to construct a local characterization, h, a candidate test becomes clear. Choose a random $T \in \mathcal{T}_{n,k}$, and accept if and only if $\langle f \circ T, h \rangle = 0$. Carrying out this test requires querying $f \circ T(x)$ for every $x \in \mathrm{supp}(h)$. It is clear from the definition of local characterization that this test satisfies completeness, while for soundness we may appeal to a general result from [16]. For our purposes, this lemma states the following.

THEOREM 2.2. (LEMMA 2.9 [16]) *For any local characterization $h : \mathbb{F}_q^k \to \mathbb{F}_q$ and function $f : \mathbb{F}_q^n \to \mathbb{F}_q$, we have*

$$\Pr_{T \in \mathcal{T}_{n,k}} [\langle f \circ T, h \rangle \neq 0] \geqslant \frac{\delta}{4Q^2},$$

where $Q = |\mathrm{supp}(h)|$ and $\delta = \delta(f, \mathcal{F}_n(h))$.

3 The Prime Field Case: Proof of Theorem 1.1

Fix a degree parameter d, a prime field size $q = p$, and an erasure parameter t. In this section we prove Theorem 1.1 by describing and analyzing a t-erasure-resilient degree d tester for functions over $\mathbb{F}_p^n$. For distance parameter δ, our test is t-online-erasure resilient for all $t \leqslant \frac{\delta}{30} p^{n/(20d)}$.

3.1 A Local Characterization for $\mathrm{RM}[n,p,d]$ To start, we give a sufficient condition for a function h to be a local characterization $\mathrm{RM}[n,p,d]$. This requires the following Corollary of Lemma 2.4 applied to the prime field case.

LEMMA 3.1. *Let $\mathcal{F}$ be a linear affine invariant family of functions over $\mathbb{F}_p^n$ and suppose $x^e \in \mathcal{F}$ for some $e \in \{0,\dots,p-1\}^n$. Then $x^{e'} \in \mathcal{F}$ for any $e' \in \{0,\dots,p-1\}^n$ such that $|e'|_1 \leqslant |e|_1$,*

Proof. Note that for $a,b \in \{0,\dots,p-1\}$ the relation $a \leqslant_p b$ is equivalent to $a \leqslant b$. Thus, by repeatedly apply Lemma 2.4, it is clear that $x^v \in \mathcal{F}$ for every $v \in \{0,\dots,p-1\}^n$ such that $|v|_1 = |e|_1$. In particular, there exists a v such that $e' \leqslant_p v$ and $x^v \in \mathcal{F}$. By Lemma 2.2, it follows that $x^{e'} \in \mathcal{F}$. ∎

Broadly, Lemma 3.1 states that in order for h to be a local characterization for $\mathrm{RM}[n,p,d]$, it is sufficient for h to detect any monomial of degree $d+1$. This idea is formalized in the lemma below.

LEMMA 3.2. *For any $k \geqslant \lceil \frac{d+1}{p-1} \rceil$, if $h : \mathbb{F}_p^k \to \mathbb{F}_2$ satisfies $\deg(h) = k(p-1)-(d+1)$ then, h is a local characterization for $\mathrm{RM}[n,p,d]$. Equivalently $\mathrm{RM}[n,p,d] = \mathcal{F}_n(h)$.*

Proof. Let $h : \mathbb{F}_p^k \to \mathbb{F}_p$ have degree $k(p-1)-(d+1)$. We show that $\mathrm{RM}[n,p,d] = \mathcal{F}_n(h)$.

By Lemma 2.6, $\mathrm{RM}[n,p,d] \subseteq \mathcal{F}_n(h)$. Indeed, for any $f \in \mathrm{RM}[n,p,d]$ and any affine transformation $T \in \mathcal{T}_{n,k}$, we have the degree of $f \circ T$ is at most d. Thus if x^e is a monomial appearing in $f \circ T$ and $x^{e'}$ is a monomial appearing in h, then $|e+e'|_1 \leqslant k(p-1)-(d+1)+d \leqslant k(p-1)-1$. It follows that $e+e' \neq (q-1,\dots,q-1)$, and by Lemma 2.6, $\langle x^e, x^{e'} \rangle = 0$. Since this holds for every pair of monomials appearing in $f \circ T$ and h, by the bilinearity of the inner product, $\langle f \circ T, h \rangle = 0$ for every $T \in \mathcal{T}_{n,k}$.

To complete the proof we show that $\mathcal{F}_n(h) \subseteq \mathrm{RM}[n,p,d]$. Suppose $\prod_{i=1}^k x_i^{e_i}$ is one of the degree $k(p-1)-(d+1)$ monomials in h. Let $e_i' = p-1-e_i$ and define

$$e' = (e_1',\dots,e_k',0,\dots,0) \in \{0,\dots,p-1\}^n.$$

Let T be the affine transformation that is identity on the first k-coordinates and sends all other coordinates to 0, i.e. $T(x_1,\dots,x_n) = (x_1,\dots,x_k,0,\dots,0)$. By Lemma 2.6, $\langle x^{e'} \circ T, h \rangle \neq 0$ it follows that $x^{e'} \notin \mathcal{F}_n(h)$. Since $|e'|_1 = d+1$, it follows from Lemma 3.1 that $\mathcal{F}_n(h)$ contains no monomials of degree greater than d. Thus, $\mathcal{F}_n(h) \subseteq \mathrm{RM}[n,p,d]$. ∎

3.2 An Erasure Resilient Low Degree Tester In order to design a low degree tester that works in the presence of online erasures, we would like to have as little dependency between the queried points as possible. To this end, we show that simply choosing $\approx k^d$ uniformly random points in a k-dimensional tester for $k \geqslant d+1$ to be determined later suffices for low degree testing. To make this tester t-online-erasure resilient, we pick k sufficiently large relative to the erasure budget t.

LEMMA 3.3. *For every dimension $k \geqslant d+1$ and every $S \subseteq \mathbb{F}_p^k$ such that $|S| \geqslant \binom{d+k+1}{k}+1$, there exist a function $h : \mathbb{F}_p^k \to \mathbb{F}_p$ such that*

- $\deg(h) = k(p-1)-(d+1)$
- $\mathrm{supp}(H) \subseteq |S|$

Proof. Let $\mathcal{H}$ be the set of functions from $\mathbb{F}_p^k \to \mathbb{F}_p$ whose support is contained in S. Then

$$|\mathcal{H}| \geqslant p^{|S|} \geqslant p^{\binom{d+k+1}{k}+1}.$$

On the other hand, the number of monomials over $\mathbb{F}_p^k$ of degree at least $k(p-1)-(d+1)$ is $\binom{d+k+1}{k}$. To see this, we can bound the number of $e \in \{0,\dots,p-1\}^k$ such that $\sum_{i=1}^k p-1-e_i \leqslant d+1$. The number of linear combinations of such monomials over $\mathbb{F}_p$ is at most $p^{\binom{d+k+1}{k}}$. By the pigeonhole principle there must exist distinct $h_1,h_2 \in \mathcal{H}$ that have the exact same coefficient for each monomial of degree at least $k(p-1)-(d+1)$. Set $h' = h_1 - h_2$,

so that all of these monomials are cancelled out in h'. Then, $\deg(h') \leqslant k(p-1)-(d+1)$. If this inequality is in fact an equality, then we are done. Otherwise, let x^e be a maximum degree monomial appearing in h'. Let x^v be a monomial of minimum degree such that $x^e \cdot x^v$ has degree $k(p-1)-(d+1)$. Such a monomial indeed exists and we can take $h = h' \cdot x^v$ to get the desired function. Indeed, $x^e \cdot x^v$ is a maximum degree monomial of h and $\operatorname{supp}(h) \subseteq \operatorname{supp}(h') \subseteq S$. ∎

Combining Lemmas 3.3 and 3.2, we get that on any set $S \subseteq \mathbb{F}_p^k$ of size $\binom{d+k+1}{k}+1$, there is a local characterization h whose support is contained inside S. This motivates the basic tester described in Algorithm 1. We will analyze Algorithm 1 ignoring erasures. Thus, we give a lower bound on its rejection assuming that no points are erased. To deal with erasures, we note that each query is uniformly random in a k-dimensional affine subspace. Hence by choosing k large enough later, we can give an upper bound on the probability that any query we make is erased. For the remainder of this section, fix

$$Q(k) = \binom{d+k+1}{k} + 1.$$

Algorithm 1 A Basic Low Degree Tester over $\mathbb{F}_p$

1: **procedure** RANDOMPOINTS(f, d, k)
2: Choose $T \in \mathcal{T}_{n,k}$ uniformly at random.
3: Choose $S \subseteq \mathbb{F}_p^k$ of size $Q(k)$ uniformly at random.
4: Query $f \circ T(x)$ for each $x \in S$.
5: $\mathcal{H}(S) \leftarrow$ the set of h with support contained in S that satisfy $\mathrm{RM}[n,p,d] = \mathcal{F}_n(h)$.
6: **if** $\mathcal{H}(S)$ is empty **then**
7: **return** Accept.
8: **end if**
9: Choose $h \in \mathcal{H}(S)$ uniformly at random.
10: **if** $\langle f \circ T, h\rangle = 0$ **then**
11: **return** Accept.
12: **end if**
13: **if** $\langle f \circ T, h\rangle \neq 0$ **then**
14: **return** Reject.
15: **end if**
16: **end procedure**

LEMMA 3.4. *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be a function. If $\deg(f) \leqslant d$, then Algorithm 1 outputs accept with probability 1. If f is δ-far from degree d, Algorithm 1 outputs rejects with probability at least $\frac{\delta}{4Q(k)^2}$.*

Proof. We begin with the completeness of the tester, and towards this end assume that f has degree at most d. Since $\deg(h) \leqslant k(p-1)-(d+1)$ for all $h \in \mathcal{H}$, and $\deg(f \circ T) \leqslant \deg(f)$, it follows from Lemma 2.6 and the bilinearity of inner product that $\langle f \circ T, h\rangle = 0$ for all T.

We move on to the second part of the lemma. For $h \in \mathcal{H}$, let $\Pr(h)$ denote the probability that h is chosen in step 3 and let $\mathrm{rej}_h(f)$ denote the probability over $T \in \mathcal{T}_{n,k}$ that $\langle f \circ T, h\rangle \neq 0$. The random T and h in Algorithm 1 induce a product distribution over $\mathcal{T}_{n,k} \times \mathcal{H}$ where the distribution of $\mathcal{T}_{n,k}$ is uniform. Conditioned on any arbitrary $h \in \mathcal{H}$ being chosen in line 9, T is uniform over $\mathcal{T}_{n,k}$ and the test rejects with probability at least $\mathrm{rej}_h(f)$. By Theorem 2.2, for every $h \in \mathcal{H}$, $\mathrm{rej}_h(f) \geqslant \frac{\delta}{4|\operatorname{supp}(h)|^2} \geqslant \frac{\delta}{4Q(k)^2}$.

Thus, by generating the distribution over T and h by first choosing $h \in \mathcal{H}$ with probability $\Pr(h)$ and then choosing $T \in \mathcal{T}_{n,k}$ uniformly. This makes it clear that Test 1 rejects with probability at least

$$\sum_{h \in \mathcal{H}} \Pr(h)\,\mathrm{rej}_h(f) \geqslant \frac{\delta}{4Q(k)^2}.$$

∎

In order to output reject with probability at least $2/3$ in the far from degree d case, we will repeat Algorithm 1. Henceforth, set $k = 20d\log_p(30t/\delta)$. We will show that the resulting tester, described in Algorithm 2, is t-online-erasure resilient. Before showing completeness and soundness, we give a bound on $Q(k)$ which will be helpful for calculations,

$$Q(k) \leqslant (3k/d)^{d+1} \leqslant (60\log_p(30t/\delta))^{d+1}.$$

By the bound on t stated at the start of the section, we have $k \leqslant n$, so our tester is valid, in that $f \circ T$ is always well defined for $T \in \mathcal{T}_{n,k}$.

Algorithm 2 A t-online-erasure resilient tester over $\mathbb{F}_p$.

procedure ERASURERESILIENT(f, d, δ)
  Set $k = 100d\log\left(\frac{100dt}{\delta}\right) + q$.
  for $i = 1 \to \frac{100Q(k)^2}{\delta}$ **do**
    Run `RandomPointsTest`(f, d, k)
    If Reject is outputted, **return** Reject.
  end for
  return Accept.
end procedure

The following result shows completeness and soundness for Algorithm 1 and hence proves Theorem 1.1

THEOREM 3.1. *Given $f : \mathbb{F}_p^n \to \mathbb{F}_p$, a degree parameter d, a distance parameter δ, and an erasure parameter t, Algorithm 2 accepts f with probability 1 if $\deg(f) \leqslant d$ and rejects f with probability at least $2/3$ if f is δ-far from degree d. The number of queries made is $O\left(\left(\frac{\log(t/\delta)}{\delta}\right)^{3d+3}\right)$.*

Proof. It is clear from Lemma 3.4 that if $\deg(f) \leqslant d$, then f is accepted with probability 1.

Now suppose that f is δ-far from degree d. By Lemma 3.4, with probability at least $3/4$, there is at least one iteration where the basic tester outputs reject, so it remains to bound the probability that any queried point is erased. The total number of erased points is at most $\frac{100Q(k)^3t}{\delta}$ and each query is uniformly random in some k-dimensional affine subspace. Therefore by a union bound, the probability that any individual query is erased is at most,

$$\frac{100Q(k)^3t}{\delta p^k} \cdot \frac{Q(k)^3t}{\delta} \leqslant \frac{100t^2(60\log_p(30t/\delta))^{6d+6}}{\delta^2(30t/\delta)^{20d}} \leqslant \frac{1}{100}.$$

Overall, this shows that if f is δ-far from degree d then Algorithm 2 outputs rejects with probability at least $3/4 - 1/100 \geqslant 2/3$. ∎

4 The Non-Prime Field Case

Our algorithm for low degree testing over non-prime fields follows the same outline as that of the prime field case, but its analysis includes a few more complications. For the remainder of this section, write $d + 1 = s \cdot (q - q/p) + r$ for an integer s and an integer $0 \leqslant r < q - q/p$. For distance parameter δ, our test is t-online-erasure resilient for all $t \leqslant \frac{\delta}{100}q^{n/(100(d+q))-1}$.

4.1 Local Characterizations for $\mathrm{RM}[n, q, d]$

Our first goal is to show the following lemma which gives sufficient conditions for $h : \mathbb{F}_q^k \to \mathbb{F}_q$ to be a local characterization. Its proof has appeared before, first in [21] and later restated closer to the language used here in the appendix of [17]. We include it below for completeness.

LEMMA 4.1. *If $h : \mathbb{F}_q^k \to \mathbb{F}_q$ satisfies*

- $\deg(h) \leqslant (q-1)\cdot k - (d+1)$.
- *h contains the monomial $\left(\prod_{i=1}^{s} x_i^{q/p-1}\right) x_{s+1}^{q-1-r'} \left(\prod_{j=s+2}^{k} x_j^{q-1}\right)$ for all $r \leqslant r' \leqslant q-1$,*

then $\mathrm{RM}[n,q,d] = \mathcal{F}_n(h)$.

We split the proof of Lemma 4.1 into two claims. Fix $h : \mathbb{F}_q^k \to \mathbb{F}_q$ to be a function satisfying the conditions above.

LEMMA 4.2. $\mathrm{RM}[n,q,d] \subseteq \mathcal{F}_n(h)$.

Proof. This is a direct application of Lemma 2.6 and the same as the first part of the proof of Lemma 3.2 ∎

LEMMA 4.3. $\mathcal{F}_n(h) \subseteq \mathrm{RM}[n,q,d]$.

Proof. Suppose for the sake of contradiction that there is $g \in \mathcal{F}_n(h)$ with degree greater than d and let x^e be a maximum degree monomial of g. By Lemma 2.1, $x^e \in \mathcal{F}_n(h)$. Take the smallest index i such that $e_1 + \cdots + e_i > d$, and let $e' = (e_1, \ldots, e_i, 0, \ldots, 0) \in \{0, \ldots, q-1\}^k$. Then, $e' \leqslant_p e$ so $x^{e'} \in \mathcal{F}_n(h)$, and $|e'|_1 \geqslant s(q - q/p) + r'$ for some $r \leqslant r' \leqslant q-1$.

Next we will repeatedly use Lemma 2.4 to obtain a monomial in $\mathcal{F}_n(h)$ where each of the first s individual degrees are at least $q - q/p$. To this end, define

$$c(e) = \sum_{i=1}^{s} \max(0, (q - q/p) - e_i).$$

We will abuse notation and still refer to the monomial as $x^{e'}$ after each application of Lemma 2.4. If e' is not of the desired form, then one of the following must be true:

- There is $j > s$ such that $e_j > 0$, in which case we simply find some $p^m \leqslant_p e_j$ and apply Lemma 2.4 to obtain the monomial $x_i^{e_i + p^m} x_j^{e_j - p^m}$ in place of $x_i^{e_i} x_j^{e_j}$,
- There is $j \leqslant s$ such that $e_j > q - q/p$. In this case we can find p^m such that $e_j - p^m \geqslant q - q/p$, and apply Lemma 2.4 to obtain the monomial $x_i^{e_i + p^m} x_j^{e_j - p^m}$ in place of $x_i^{e_i} x_j^{e_j}$.

In either case, $c(e')$ strictly decreases, while $|e'|_1$ does not change, so we will eventually end up with e' such that $e'_i \geqslant q - q/p$ for $1 \leqslant i \leqslant s$, and $|e'|_1 = s(q - q/p) + r'$. At this point $\sum_{i=1}^{s} e'_i - (q - q/p) + \sum_{i=s+1}^{k} e'_i = r' \leqslant q-1$ so we may apply Lemma 2.4 to shift all degree above $q - q/p$ in the first s coordinates, and all degree in the $s+2$ through kth coordinates, onto e'_i, to obtain the monomial

$$\left(\prod_{i=1}^{s} x_i^{q - q/p} \right) x_{s+1}^{r'},$$

where $r \leqslant r' \leqslant q-1$. As we only applied Lemma 2.4, this monomial must be in $\mathcal{F}_n(h)$. However, by the second condition of Lemma 4.1 and Lemma 2.6, this monomial and h have nonzero inner product, leading to a contradiction. It follows that $\mathcal{F}_n(h)$ does not contain any functions of degree greater than d, completing the proof. ∎

4.2 An Erasure Resilient Low Degree Tester We now describe our t-erasure resilient degree d tester. Similar to the prime field case, we show that a set of random points contains a local characterization, and then we repeat this tester to reject with constant probability in the far from degree d case.

Fix a dimension k to be determined later. Set $e^\star \in \{0, \ldots, q-1\}^k$ such that

- $e^\star_i = q - q/p$ for $1 \leqslant i \leqslant s$,
- $e^\star_{s+1} = q - 1$,
- $e^\star_j = 0$ for $j \geqslant s+2$.

Let $d^\star = |e^\star|_1$, let $\mathcal{G} = \{g : \mathbb{F}_q^k \to \mathbb{F}_q \mid \deg(g) \leqslant d^\star - 1\}$, and let $\mathcal{E} = \{e \in \{0, \ldots, q-1\}^k \mid |e|_1 \leqslant d^\star - 1\}$. Note that $d^\star \leqslant d + q$, $s \leqslant \lceil \frac{d+1}{q - q/p} \rceil$, $|\mathcal{E}| \leqslant \binom{k+d^\star}{d^\star}$, and $|\mathcal{G}| \leqslant q^{\binom{k+d^\star}{d^\star}}$. Now construct the following system of equations over the variables $\{z_\alpha\}_{\alpha \in S}$:

- For each $e' \in \mathcal{E}$, include the equation $\sum_{\alpha \in S} z_\alpha \alpha^{e'} = 0$.
- Include the equation $\sum_{\alpha \in S} z_\alpha \alpha^{e^\star} = 1$.

We first claim that over uniformly random $S \subseteq \mathbb{F}_q^k$ of size $Q(k) = 2q^{s+1}\log(q)\binom{k+d^\star}{d^\star}$, this system of equations is solvable with high probability. To show this, we first state the following simple observation, which can also be seen through the Schwartz-Zippel lemma or combinatorial nullstellensatz.

LEMMA 4.4. *For any $g \in \mathcal{G}$,*

$$\Pr_{\alpha \in \mathbb{F}_q^k}[g(\alpha) \neq x^{e^\star}(\alpha)] \geqslant \frac{1}{q^{s+1}}.$$

Proof. For any $(\alpha_{s+2}, \ldots, \alpha_k) \in \mathbb{F}_q^{k-s-1}$ the $s+1$-variate polynomial obtained by setting $x_i = \alpha_i$ for $s+2 \leqslant i \leqslant k$ into $g(x) - x^{e^\star}$ is nonzero as it contains the monomial $\prod_{i=1}^{s+1} x_i^{e_i^\star}$. Thus for any setting of the last $k-s-1$ variables, there exists $(\alpha_1, \ldots, \alpha_{s+1})$ such that $g(x) - x^{e^\star} \neq 0$ for $x = (\alpha_1, \ldots, \alpha_{s+1}, \alpha_{s+2}, \ldots, \alpha_k)$ and the lemma follows. ∎

LEMMA 4.5. *Choose $S \subseteq \mathbb{F}_q^k$ uniformly at random of size $Q(k)$. Then the probability that the above system of equations over $\{z_\alpha\}_{\alpha \in S}$ has no solution is at most $q^{-\binom{k+d^\star}{d^\star}}$.*

Proof. The system will have a solution as long as the coefficients in the last equation are not in the span of the coefficients of the other equation. In other words, if

$$(\alpha^{e^\star})_{\alpha \in S} \notin \operatorname{span}\{(\alpha^{e'})_{\alpha \in S}\}_{e' \in \mathcal{E}}.$$

However, the span on the right side is simply the set $\{(g(\alpha)_{\alpha \in S}) \mid g \in \mathcal{G}\}$. Therefore the above event will happen if for every $g \in \mathcal{G}$ there exists at least one $\alpha \in S$ such that $g(\alpha) \neq \alpha^{e^\star}$. By Lemma 4.4, for any individual g, we have

$$\Pr_S[g(\alpha) = \alpha^{e^\star}, \forall \alpha \in S] \leqslant (1 - q^{-(s+1)})^{Q(k)} \leqslant e^{-q^{-(s+1)}Q(k)}.$$

Therefore, by a union bound, the probability that there is at least one $g \in \mathcal{G}$ equal to $x^{e^\star}$ on all of S is at most

$$|\mathcal{G}| e^{-q^{-(s+1)}Q(k)} \leqslant e^{-\log(q)\binom{k+d^\star}{d^\star}} \leqslant q^{-\binom{k+d^\star}{d^\star}}.$$

∎

Finally, we note that if the system of equations we constructed is solvable, then there exists $h : \mathbb{F}_q^k \to \mathbb{F}_q$ with support contained in S that satisfies the conditions of Lemma 4.1, and is thus a local characterization of $\mathrm{RM}[n, q, d]$.

LEMMA 4.6. *Choose $S \subseteq \mathbb{F}_q^k$ uniformly at random of size $Q(k)$. Then with probability at least $1 - q^{-\binom{k+d^\star}{d^\star}}$, there exists $h : \mathbb{F}_q^k \to \mathbb{F}_q$ with support contained in S such that $\mathrm{RM}[n, q, d] = \mathcal{F}_n(h)$.*

Proof. By Lemma 4.5 with probability at least $1 - q^{-\binom{k+d^\star}{d^\star}}$ the previously described system of equations is solvable. Suppose this is the case, and abusing notation, let $\{z_\alpha\}_{\alpha \in S}$ denote the solution. Define $h_1 : \mathbb{F}_q^k \to \mathbb{F}_q$ so that $h_1(\alpha) = z_\alpha$ if $\alpha \in S$ and $h_1(\alpha) = 0$ if $\alpha \notin S$. Then each of the previous equations states the value of an inner product involving h_1. In particular we have $\langle h, x^e \rangle = 0$ for all $e \in \{0, \ldots, q-1\}^k$ such that $|e|_1 \leqslant d^\star - 1$ and $\langle h, x^{e^\star} \rangle = 0$. By Lemma 2.6, h_1 contains the monomial $\prod_{i=1}^s x_i^{q/p-1} \prod_{j=s+2}^k x_j^{q-1}$ and all other monomials in h_1 are degree at most $(q-1) \cdot k - (d^\star + 1)$. We may write,

$$h_1(x) = \prod_{i=1}^{s} x_i^{q/p-1} \prod_{j=s+2}^{k} x_j^{q-1} + g(x),$$

where $\deg(g) \leqslant k(q-1) - (d^\star + 1)$.

To complete the proof, take

$$h = h_1 \cdot \left(\sum_{r'=r}^{q-1} x_{s+1}^{q-1-r'} \right).$$

From $\prod_{i=1}^{s} x_i^{q/p-1} \prod_{j=s+2}^{k} x_j^{q-1} \left(\sum_{r'=r}^{q-1} x_{s+1}^{q-1-r'} \right)$ we get each of the monomials in the second condition of Lemma 4.1 and these cannot get cancelled out by $g(x) \left(\sum_{r'=r}^{q-1} x_{s+1}^{q-1-r'} \right)$. Moreover the highest degree monomial of h is

$$\left(\prod_{i=1}^{s} x_i^{q/p-1} \right) x_{s+1}^{q-1-r} \prod_{j=s+2}^{k} x_j^{q-1},$$

and its degree is precisely $k(q-1) - (d+1)$, so the first condition is satisfied as well.

Therefore we show that with probability at least $1 - q^{-\binom{k+d^\star}{d^\star}}$ over S, there exists h with support contained in S such that $\mathrm{RM}[n,q,d] = \mathcal{F}_n(h)$. □

We are now ready to present our basic tester, described in Algorithm 3. Henceforth, we will fix

$$k = 100 d^\star \log_q \left(\frac{100tq}{\delta} \right).$$

For $t \leqslant \frac{\delta}{100} q^{n/(100d^\star)-1}$, we have $k \leqslant n$ and our basic tester is valid. The following bound on $Q(k)$ will be helpful for computations:

$$Q(k) \leqslant 2q^{\frac{d+1}{q-q/p}+1} \log(q) \left(300 \log_q \left(\frac{100tq}{\delta} \right) \right)^{d^\star}.$$

Algorithm 3 A Basic Tester over $\mathbb{F}_q$

procedure GENERALRANDOMPOINTS(f, d, k)
  Choose $T \in \mathcal{T}_{n,k}$ uniformly at random.
  Choose $S \subseteq \mathbb{F}_q^k$ of size $Q(k)$ uniformly at random.
  Query $f \circ T(x)$ for each $x \in S$.
  $\mathcal{H}(S) \leftarrow$ the set of h with support contained in S that satisfy $\mathrm{RM}[n,q,d] = \mathcal{F}_n(h)$.
  if $\mathcal{H}(S)$ is empty **then**
    return Accept.
  end if
  Choose $h \in \mathcal{H}(S)$ uniformly at random.
  if $\langle f \circ T, h \rangle = 0$ **then**
    return Accept.
  end if
  if $\langle f \circ T, h \rangle \neq 0$ **then**
    return Reject.
  end if
end procedure

LEMMA 4.7. *Suppose $f : \mathbb{F}_q^k \to \mathbb{F}_q$ is δ-far from $\mathrm{RM}[n,q,d]$. Then Algorithm 3 on inputs f, d, k rejects with probability at least $\frac{\delta}{5Q(k)^2}$.*

Proof. Suppose that $\mathcal{H}(S)$ is nonempty. Then by Theorem 2.2 combined with an argument identical to Lemma 3.4, the tester rejects with probability at least $\frac{\delta}{4Q(k)^2}$. The probability that $\mathcal{H}(S)$ is empty is at most

$$q^{-\binom{k+d^\star}{d^\star}} \leqslant \frac{\delta}{q^{\binom{k+d^\star}{d^\star}-1}} \leqslant \frac{\delta}{q^{100d+10d\log(k)}} \leqslant \frac{\delta}{100Q(k)^2}.$$

Hence, we may subtract this probability out and get that f is rejected with probability at least $\frac{\delta}{5Q(k)^2}$. ∎

Finally, repeating our basic tester yields a t-online-erasure resilient tester for $\mathrm{RM}[n, q, d]$.

Algorithm 4 A t-online-erasure resilient tester over $\mathbb{F}_q$.

procedure GENERALERASURERESILIENT(f, d, δ)
  Set $k = 100dt \log\left(\frac{100dt}{\delta}\right) + q$.
  for $i = 1 \to \frac{100Q(k)^2}{\delta}$ **do**
    Run `GeneralRandomPointsTest`(f, d, k)
    If Reject is outputted, **return** Reject.
  end for
  return Accept.
end procedure

THEOREM 4.1. *Given $f : \mathbb{F}_q^n \to \mathbb{F}_q$, degree parameter d, and a distance parameter δ in the t-online erasure model, Algorithm 4 outputs accept with probability 1 if $\deg(f) \leqslant d$ and outputs reject with probability $2/3$ if f is δ-far from degree d. The query complexity of Algorithm 4 is $\frac{100Q(k)^3}{\delta}$.*

Proof. The query complexity is clear, and it is also clear from Lemma 4.7 that if $\deg(f) \leqslant d$ then Algorithm 4 outputs accept with probability 1.

To see the soundness, note that if every query is obtained successfully, i.e. no queried point is erased, then by Lemma 4.7, reject is outputted with probability at least $3/4$ when f is δ-far from degree d. It remains to upper bound the probability that any query is erased. The total number of erased points is at most $\frac{100Q(k)^3 t}{\delta}$ and each query is uniformly random in some k-dimensional affine subspace. Therefore by a union bound, the probability that any individual query is erased is at most,

$$\frac{100Q(k)^3 t}{\delta q^k} \cdot \frac{100Q(k)^3 t}{\delta} \leqslant \frac{10^6 q^{6d} \log(q)^6 (300 \log_q(100tq/\delta))^{6d^\star}}{\delta^2 \left(100tq/\delta\right)^{100d^\star}} \leqslant \frac{1}{100}.$$

Overall, this shows that if f is δ-far from degree d then Algorithm 4 outputs rejects with probability at least $3/4 - 1/100 \geqslant 2/3$. ∎

5 Online-Corruption Resilient Testers

As noted, our testers also work with two sided error with the same parameters in the t-online-corruption model, since the probability that a corrupted point is queried at all by our algorithms is at most $1/100$. Recall that in this model, the adversary is allowed to alter entries of the input function's truth table, $f(x)$, and thus one can only hope for testers with two sided error as the probability of a querying a corrupted point must now be subtracted from both the completeness and the soundness.

A formal reduction from online-erasure resilient to online-corruption resilient testing is given in [13, Lemma 1.8]. Applying this lemma to our results yields the following theorems.

THEOREM 5.1. (PRIME FIELD CASE) *Let $f : \mathbb{F}_p^n \to \mathbb{F}_p$ be the input function over a prime-field vector space, d be the degree parameter, and δ be the distance parameter. Then for $t \leqslant \frac{\delta}{30} p^{n/(20d)}$, there is a t-online-corruption resilient tester with query complexity $O\left(\frac{(\log(t/\delta))^{3d+3}}{\delta}\right)$ satisfying:*

- ***Completeness:*** *If f is degree d then the algorithm outputs accept with probability at least $2/3$.*
- ***Soundness:*** *If f is δ-far from degree d then the algorithm outputs reject with probability at least $2/3$.*

THEOREM 5.2. (NON-PRIME FIELD CASE) *Let $f : \mathbb{F}_q^n \to \mathbb{F}_q$ be the input function, d be the degree parameter, and δ be the distance parameter. Then for $t \leqslant \frac{\delta}{100} q^{\frac{n}{100(d+q)}-1}$, there is a t-online-corruption resilient tester with query complexity $\frac{q^{O(1)}}{\delta} O\left((\log(t/\delta))^{3d+3q}\right)$ satisfying:*

- ***Completeness:*** *If f is degree d then the algorithm outputs accept with probability at least $2/3$.*
- ***Soundness:*** *If f is δ-far from degree d then the algorithm outputs reject with probability at least $2/3$.*

6 Further Directions

In this work we take a step towards investigating problems in algebraic property testing within the online erasure model. Regarding further directions, recall the gap in the $q = 2$ case between our upper bound and the lower bound of [5] discussed in Section 1.3. Closing this gap is an intriguing avenue for future work.

Another potential direction is to develop testers for other algebraic properties or to show generally that local characterizations (as defined in Definition 3) yield local tests in the online erasure model. The latter result in the classical model is the main theorem of Kaufman and Sudan's seminal work on algebraic property testing [16]. Their main theorem is restated in this paper as Theorem 2.2.

In our current work, we require there to be many local characterizations so that any large enough set of points in a k-dimensional affine subspace will contain the support of some local characterization with probability close to 1. Naively, one could try to make the same kind of argument work for a *single* characterization and appeal to the Density Hales-Jewett Theorem [10, 19]. This would give that for a fixed local characterization h and any fraction $c > 0$ there is some dimension k_c such that any set of cq^{k_c} points in a k-dimensional affine subspace will contain the support of h (where h is extended to the entire subspace in some manner). However, in order for this approach to work even with a constant erasure parameter, one would need $cq^{k_c} < q^{k_c/2}$, and such a statement is far stronger than what is guaranteed by the Density Hales-Jewett theorem.

Yet another intriguing direction is testing other algebraic properties that are not necessarily affine-invariant, for example, the dual BCH property which is equivalent to being the trace of a low degree polynomial from $\mathbb{F}_{q^n} \to \mathbb{F}_q$. The set of such functions still has a 2-transitive symmetry group but this group is much less rich than the affine-invariant symmetries that we work with.

Finally, it may be possible that our techniques can answer other questions regarding Reed-Muller codes in the online erasure model. For example, it could be interesting to explore analogues of other well studied tasks in coding theory like local decoding or local correction for Reed-Muller codes. We remark that local decoding in an (offline) erasure model appears in [20].

7 Acknowledgements

This work was done in part while the authors were visiting the Simons Institute for the Theory of Computing.

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] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[3] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998.

[4] Sanjeev Arora and Madhu Sudan. Improved low-degree testing and its applications. *Comb.*, 23(3):365–426, 2003.

[5] Omri Ben-Eliezer, Esty Kelman, Uri Meir, and Sofya Raskhodnikova. Personal Communication.

[6] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of Reed-Muller codes. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, Las Vegas, NV, USA*, pages 488–497, 2010.

[7] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, 47(3):549–595, 1993.

[8] Kashyap Dixit, Sofya Raskhodnikova, Abhradeep Thakurta, and Nithin Varma. Erasure-resilient property testing. *SIAM J. Comput.*, 47(2):295–329, 2018.

[9] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996.

[10] Hillel Furstenberg and Yitzhak Katznelson. A density version of the hales-jewett theorem. *J. Anal. Math.*, 57:64–119, 1991.

[11] Elad Haramaty, Amir Shpilka, and Madhu Sudan. Optimal testing of multivariate polynomials over small prime fields. *SIAM J. Comput.*, 42(2):536–562, 2013.

[12] 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.

[13] Iden Kalemaj, Sofya Raskhodnikova, and Nithin Varma. Sublinear-time computation in the presence of online erasures. In Mark Braverman, editor, *13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 - February 3, 2022, Berkeley, CA, USA*, volume 215 of *LIPIcs*, pages 90:1–90:25. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.

[14] Tali Kaufman and Dor Minzer. Improved Optimal Testing Results from Global Hypercontractivity. In *Proceedings of the 63rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2022, October 31 - November 3, Denver, CO, USA*, 2022.

[15] Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM J. Comput.*, 36(3):779–802, 2006.

[16] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In Cynthia Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 403–412. ACM, 2008.

[17] Dor Minzer and Kai Zhe Zheng. Optimal testing of generalized reed-muller codes in fewer queries. In *To Appear in FOCS 2023*, 2023.

[18] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *J. Comput. Syst. Sci.*, 72(6):1012–1042, 2006.

[19] D.H.J. Polymath. A new proof of the density hales-jewett theorem. *Ann. of Math.*, 175(2):1283–1327, 2012.

[20] Sofya Raskhodnikova, Noga Ron-Zewi, and Nithin Varma. Erasures vs. errors in local decoding and property testing. In Avrim Blum, editor, *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, volume 124 of *LIPIcs*, pages 63:1–63:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[21] Noga Ron-Zewi and Madhu Sudan. A New Upper Bound on the Query Complexity of Testing Generalized Reed-Muller Codes. *Theory of Computing*, 9(25):783–807, 2013.

[22] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Computing*, 25(2):252–271, 1996.

[23] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.