Electronic Colloquium on Computational Complexity, Report No. 138 (2017)

Local decoding and testing of polynomials over grids

Srikanth Srinivasan* Madhu Sudan†

September 17, 2017

Abstract

The well-known DeMillo-Lipton-Schwartz-Zippel lemma says that n-variate polynomials of total degree at most d over grids, i.e. sets of the form $A_1 \times A_2 \times \cdots \times A_n$, form error-correcting codes (of distance at least 2^{-d} provided $\min_i\{|A_i|\} \geq 2$). In this work we explore their local decodability and local testability. While these aspects have been studied extensively when $A_1 = \cdots = A_n = \mathbb{F}_q$ are the same finite field, the setting when A_i's are not the full field does not seem to have been explored before.

In this work we focus on the case $A_i = \{0, 1\}$ for every i. We show that for every field (finite or otherwise) there is a test whose query complexity depends only on the degree (and not on the number of variables). In contrast we show that decodability is possible over fields of positive characteristic (with query complexity growing with the degree of the polynomial and the characteristic), but not over the reals, where the query complexity must grow with n. As a consequence we get a natural example of a code (one with a transitive group of symmetries) that is locally testable but not locally decodable.

Classical results on local decoding and testing of polynomials have relied on the 2-transitive symmetries of the space of low-degree polynomials (under affine transformations). Grids do not possess this symmetry: So we introduce some new techniques to overcome this handicap and in particular use the hypercontractivity of the (constant weight) noise operator on the Hamming cube.

*Department of Mathematics, IIT Bombay. srikanth@math.iitb.ac.in.

†Harvard John A. Paulson School of Engineering and Applied Sciences. madhu@cs.harvard.edu.

ISSN 1433-8092

Contents

1 Introduction

Low-degree polynomials have played a central role in computational complexity. (See for instance [26, 8, 5, 20, 22, 18, 27, 3, 2] for some of the early applications.) One of the key properties of low-degree n-variate polynomials underlying many of the applications is the "DeMillo-Lipton-Schwartz-Zippel" distance lemma [10, 25, 28] which upper bounds the number of zeroes that a non-zero low-degree polynomial may have over "grids", i.e., over domains of the form $A_1 \times \cdots \times A_n$. This turns the space of polynomials into an error-correcting code (first observed by Reed [23] and Muller [19]) and many applications are built around this class of codes. These applications have also motivated a rich collection of tools including polynomial time (global) decoding algorithms for these codes, and "local decoding" [4, 17, 9] and "local testing" [24, 1, 14] procedures for these codes.

Somewhat strikingly though, many of these tools associated with these codes don't work (at least not immediately) for all grid-like domains, but work only for the specific case of the domain being the vector space $\mathbb{F}^n$ where $\mathbb{F}$ is the field over which the polynomial is defined and $\mathbb{F}$ is finite. The simplest example of such a gap in knowledge was the case of "global decoding". Here, given a function $f : \prod_{i=1}^{n} A_i \to \mathbb{F}$ as a truth-table, the goal is to find a nearby polynomial (up to half the distance of the underlying code) in time polynomial in $|\prod_i A_i|$. When the domain equals $\mathbb{F}^n$ then such algorithms date back to the 1950s. However the case of general A_i remained open till 2016 when Kim and Kopparty [16] finally solved this problem.

In this paper we initiate the study of local decoding and testing algorithms for polynomials when the domain is not a vector space. For uniformity, we consider the case of polynomials over hypercubes (i.e., when $A_i = \{0, 1\} \subseteq \mathbb{F}$ for every i). We describe the problems formally next and then describe our results.

1.1 Distance, Local Decoding and Local Testing

We start with some brief notation. For finite sets $A_1, \ldots, A_n \subseteq \mathbb{F}$ and functions $f, g : A_1 \times \cdots A_n \to \mathbb{F}$, let the distance between f and g, denoted $\delta(f, g)$ be the quantity $\Pr_a[f(a) \neq g(a)]$ where a is drawn uniformly from $A_1 \times \cdots \times A_n$. We say f is δ-close to g if $\delta(f, g) \leq \delta$, and δ-far otherwise. For a family of functions $\mathcal{F} \subseteq \{h : A_1 \times \cdots \times A_n \to \mathbb{F}\}$, let $\delta(\mathcal{F}) = \min_{f \neq g \in \mathcal{F}}\{\delta(f, g)\}$.

To set the context for some of the results on local decoding and testing, we first recall the distance property of polynomials. If $|A_i| \geq 2$ for every i, the polynomial distance lemma asserts that the distance between any two distinct degree d polynomials is at least 2^{-d}. Of particular interest is the fact that for fixed d this distance is bounded away from 0, independent of n or $|\mathbb{F}|$ or the structure of the sets A_i. In turn this behavior effectively has led to "local decoding" and "local testing" algorithms with complexity depending only on d — we define these notions and elaborate on this sentence next.

Given a family of functions $\mathcal{F}$ from the domain $A_1 \times \cdots \times A_n$ to $\mathbb{F}$, we say $\mathcal{F}$ is (δ, q)-*locally decodable* if there exists a probabilistic algorithm that, given $a \in A_1 \times \cdots \times A_n$ and oracle access to a function $f : A_1 \times \cdots \times A_n \to \mathbb{F}$ that is δ-close to some function $p \in \mathcal{F}$, makes at most q oracle queries to f and outputs $p(a)$ with probability at least $3/4$. (The existence of a (δ, q)-local decoder for $\mathcal{F}$ in particular implies that $\delta(\mathcal{F}) \geq 2\delta$.) We say that $\mathcal{F}$ is (δ, q)-*locally testable* if there exists a probabilistic algorithm that makes q queries to an oracle for $f : A_1 \times \cdots \times A_n \to \mathbb{F}$ and accepts with probability at least $3/4$ if $f \in \mathcal{F}$ and rejects with probability at least $3/4$ if f is δ-far from every function in $\mathcal{F}$.

When $A_1 = \cdots = A_n = \mathbb{F}$ (and so $\mathbb{F}$ is finite) it was shown by Kaufman and Ron [14] (with similar results in Jutla et al. [13]) that the family of n-variate degree d polynomials over $\mathbb{F}$ is (δ, q)-locally decodable and (δ, q)-locally testable for some $\delta = \exp(-d)$ and $q = \exp(d)$. In particular both q and $1/\delta$ are bounded for fixed d, independent of n and $\mathbb{F}$. Indeed in both cases δ is lower bounded by a constant factor of $\delta(\mathcal{F}(n,d))$ and q is upper bounded by a polynomial in the inverse of $\delta(\mathcal{F}(n,d))$ where $\mathcal{F}(n,d)$ denotes the family of degree d n-variate polynomials over $\mathbb{F}$, seemingly suggesting that the testability and decodability may be consequences of the distance. If so does this phenomenon should extend to the case of other sets $A_i \neq \mathbb{F}$ - does it? We explore this question in this paper.

In what follows we say that the family of degree d n-variate polynomials is locally decodable (resp. testable) if there is bounded $q = q(d)$ and positive $\delta = \delta(d)$ such that $\mathcal{F}(n,d)$ is (δ, q)-locally decodable (resp. testable) for every n. The specific question we address below is when are the family of degree d n-variate polynomials locally decodable and testable when the domain is $\{0,1\}^n$. (We stress that the choice of $\{0,1\}^n$ as domain is partly for simplicity and is equivalent to the setting of $|A_i| = 2$ for all i. Working with domains of other (and varying) sizes would lead to quantitative changes and we do not consider that setting in this paper.)

1.2 Main Results

Our first result (Theorem 3.2) shows that even the space of degree 1 polynomials is *not locally decodable* over fields of zero characteristic or over fields of large characteristic. This statement already stresses the main difference between the vector space setting (domain being $\mathbb{F}^n$) and the "grid" setting (domain $= \{0,1\}^n$). One key reason underlying this difference is that the domain $\mathbb{F}^n$ has a rich group of symmetries that preserve the space of degree d polynomials, where the space of symmetries is much smaller when the domain is $\{0,1\}^n$. Specifically the space of degree d polynomials over $\mathbb{F}^n$ is "affine-invariant" (invariant under all affine maps from $\mathbb{F}^n$ to $\mathbb{F}^n$). The richness of this group of symmetries is well-known to lead to local decoding algorithms (see for instance [1]) and this explains the local decodability of $\mathcal{F}(n,d)$ over the domain $\mathbb{F}^n$. Of course the absence of this rich group of symmetries does not rule out local decodability — and so some work has to be done to establish Theorem 3.2. We give an overview of the proof in Section 1.3 and then give the proof in Section 5.

Our second result (Theorem 3.3) shows, in contrast, that the class of *degree d polynomials over fields of small characteristic are locally decodable*. Specifically, we show that there is a $q = q(d,p) < \infty$ and $\delta = \delta(d,p) > 0$ such that $\mathcal{F}(n,d)$ over the domain $\{0,1\}^n$ over a (possibly infinite) field $\mathbb{F}$ of characteristic p is (δ, q)-locally decodable. This is perhaps the first local-decodability result for polynomials over infinite fields. A key technical ingredient that leads to this result, which may be of independent interest, is that when $n = 2p^t$ (twice a power of the characteristic of $\mathbb{F}$) and g is a degree d polynomial for $d < n/2$ then $g(0)$ can be determined from the value of g on the ball on Hamming weight $n/2$ (see Lemma 6.1). Again, we give an overview of the proof in Section 1.3 and then give the actual proof in Section 6.

Our final, and main technical, result (Theorem 3.1) shows somewhat surprisingly that $\mathcal{F}(n,d)$ *is always (i.e., over all fields) locally testable*. This leads to perhaps the simplest natural example of a locally testable code that is not locally decodable. We remark there are of course many examples of such codes (see, for instance, the locally testable codes of Dinur [11]) but these are results of careful constructions and in particular not very symmetric. On the other hand $\mathcal{F}(n,d)$ over $\{0,1\}^n$ does possess moderate symmetry and in particular the automorphism group is transitive. We remark

that for both our positive results (Theorems 3.3 and 3.1), the algorithms themselves are not obvious and the analysis leads to further interesting questions. We elaborate on these in the next section.

1.3 Overview of proofs

Impossibility of local decoding over fields of large characteristic. In Section 5 we show that even the family of affine functions over $\{0,1\}^n$ is not locally decodable. The main idea behind this construction and proof is to show that the value of a affine function $\ell : \{0,1\}^n \to \mathbb{F}$ at 1^n can not be determined from its values on any set S if $|S|$ is small (specifically $|S| = o(\log n / \log\log n)$) and S contains only "balanced" elements (i.e., $x \in S \Rightarrow |\sum_i x_i - (n/2)| = O(\sqrt{n})$. Since the space of affine functions from $\{0,1\}^n$ to $\mathbb{F}$ forms a vector space, this in turn translates to showing that no set of up to $|S|$ balanced vectors contain the vector 1^n in their affine span (over $\mathbb{F}$) and we prove this in Lemma 5.2.

Going from Lemma 5.2 to Theorem 5.3 is relatively standard in the case of finite fields. We show that if one picks a random linear function and simply erase its values on imbalanced inputs, this leads to only a small fraction of error, but its value at 1^n is not decodable with $o(\log n / \log\log n)$ queries. (Indeed many of the ingredients go back to the work of [6], who show that a canonical non-adaptive algorithm is effectively optimal for linear codes, though their results are stated in terms of local testing rather than local decoding.) In the case of infinite fields one has to be careful since one can not simply work with functions that are chosen uniformly at random. Instead we work with random linear functions with bounded coefficients. The bound on the coefficients leads to mild complications due to border effects that need care. In Section 5.2 we show how to overcome these complications using a counting (or encoding) argument.

The technical heart of this part is thus the proof of Lemma 5.2 and we give some idea of this proof next. Suppose $S = \{x^1, \ldots, x^t\}$ contained $x^0 = 1^n$ in its affine span and suppose $|\sum_{j=1}^n x^i_j - (n/2)| \leq n/s$ for all i. Let $a_1, \ldots, a_t \in \mathbb{F}$ be coefficients such that $x^0 = \sum_i a_i x^i$ with $\sum_i a_i = 1$. Our proof involves reasoning about the size of the coefficients $a_1, \ldots, a_t$. To get some intuition why this may help, note that

$$\frac{n}{2} = \left|\sum_{j=1}^n x^0_j - \frac{n}{2}\right| = \left|\sum_{i=1}^t a_i \cdot \left(\sum_{j=1}^n x^i_j - \frac{n}{2}\right)\right| \leq \sum_{i=1}^t |a_i| \cdot \left|\sum_{j=1}^n x^i_j - \frac{n}{2}\right| \leq \frac{n}{s} \cdot \sum_j |a_j|.$$

So in particular if the a_j's are small, specifically if $|a_j| \leq 1$ then we conclude $t = \Omega(s)$. But what happens if large a_j's are used? To understand this, we first show that the coefficients need not be too large (as a function of t) - see Lemma 5.1, and then use this to prove Lemma 5.2. The details are in Section 5.1.

Local decodability over fields of small characteristic. The classical method to obtain a q-query local decoder is to find, given a target point $x^0 \in \mathbb{F}^n$, a distribution on queries $x^1, \ldots, x^q \in \mathbb{F}^n$ such that (1) $P(x^0)$ is determined by $P(x^1), \ldots, P(x^q)$ for every degree d polynomial P, and (2) the query x^i is independent of x^0 (so that an oracle f that usually equals P will satisfy $P(x^i) = f(x^i)$ for all i, with probability at least 3/4). Classical reductions used the "2-transitivity" of the underlying space of automorphisms to guarantee that x^i is independent of x^j for every pair $i \neq j \in \{0, \ldots, q\}$ — a stronger property than required! Unfortunately, our automorphism space is not "2-transitive" but it turns out we can still find a distribution that satisfies the minimal needs.

Specifically, in our reduction we identify a parameter $k = k(p, d)$ and map each variable x_ℓ to either y_j or $1 - y_j$ for some $j = j(\ell) \in [k]$. This reduces the n-variate decoding task with oracle access to $f(x_1, \ldots, x_k)$ to a k-variate decoding task with access to the function $g(y_1, \ldots, y_k)$. Since there are only 2^k distinct inputs to g, decoding can solved with at most 2^k queries (if it can be solved at all). The choice of whether x_ℓ is mapped to y_j or $1 - y_j$ is determined by x^0_j so that $f(x^0) = g(0^k)$. Thus given x^0, the only randomness is in the choice of $j(\ell)$. We choose $j(\ell)$ uniformly and independently from $[k]$ for each ℓ. For $y \in \{0,1\}^k$, x^y denote the corresponding query in $\{0,1\}^n$ (i.e., $g(y) = f(x^y)$). Given our choices, x^y is not independent of x^0 for every choice of y. Indeed if y has Hamming weight 1, then x^y is very likely to have Hamming distance $\approx n/k$ from x^0 which is far from independent. However if $y \in \{0,1\}^k$ is a balanced vector with exactly $k/2$ 1s (so in particular we will need k to be even), then it turns out x^y is indeed independent of x^0. So we query only those x^y for which y is balanced. But this leads to a new challenge: can $P(0^k)$ be determined from the values of $P(y)$ for balanced ys? It turns out that for a careful choice of k (and this is where the small characteristic plays a role) the value of a degree d polynomial at 0 is indeed determined by its values on balanced inputs (see Lemma 6.1) and this turns out to be sufficient to build a decoding algorithm over fields of small characteristic. Details may be found in Section 6.

Local testability over all fields. We now turn to the main technical result of the paper, namely the local testability of polynomials over grids. All previous analyses of local testability of polynomials with query complexity independent of the number of variables have relied on symmetry either implicitly or explicitly. (See for example [15] for further elaboration.) Furthermore many also depend on the local decodability explicitly; and in our setting we seem to have insufficient symmetry and definitely no local decodability. This forces us to choose the test and analysis quite carefully.

It turns out that among existing approaches to analyses of local tests, the one due to Bhattacharyya et al [7] (henceforth BKSSZ) seems to make the least use of local decodability and our hope is to be able to simulate this analysis in our case — but the question remains: "which tester should we use?". This is a non-trivial question since the BKSSZ test is a natural one in a setting with sufficient symmetry; but their analysis relies crucially on the ability to view their test as a sequence of restrictions: Given a function $f : \mathbb{F}^n \to \mathbb{F}$ they produce a sequence of functions $f = f_n, f_{n-1}, \ldots, f_k$, where the function f_r is an r-variate function obtained by restricting f_{r+1} to a codimension one affine subspace. Their test finally checks to see if f_k is a degree d polynomial. To emulate this analysis, we design a somewhat artificial test: We also produce a sequence of functions $f_n, f_{n-1}, \ldots, f_k$ with f_r being an r-variate function. Since we do not have the luxury to restrict to arbitrary subspaces, we instead derive f_r from $f_{r+1}(z_1, \ldots, z_{r+1})$ by setting $z_i = z_j$ or $z_i = 1 - z_j$ for some random pair i, j (since these are the only simple affine restrictions that preserve the domain). We stop when the number of variables k is small enough (and hopefully a number depending on d alone and not on n or $\mathbb{F}$). We then test that the final function has degree d.

The analysis of this test is not straightforward even given previous works, but we are able to adapt the analyses to our setting. Two new ingredients that appear in our analyses are the hypercontractivity of hypercube with the constant weight noise operator (analyzed by Polyanskiy [21]) and the intriguing stochastics of a random set-union problem. We explain our analysis and where the above appear next.

We start with the part which is more immediate from the BKSSZ analysis. This corresponds

to a key step in the BKSSZ analysis where it is shown that if f_{r+1} is far from degree d polynomials then, with high probability, so also is f_r. This step is argued via contradiction. If f_r is close to the space of degree d polynomials for many restrictions, then from the many polynomials that agree with f_r (for many of the restrictions) one can glue together an $r+1$-variate polynomial that is close to f_{r+1}. This step is mostly algebraic and works out in our case also; though the actual algebra is different and involves more cases. (See Lemma 4.6 and its proof in Section 4.2.)

The new part in our analysis is in the case where f_n is moderately close to some low-degree polynomial P. In this case we would still like to show that the test rejects f_n with positive probability. In both BKSSZ and in our analysis this is shown by showing the the 2^k queries into f_n (that given the entire truth table of the function f_k) satisfy the property that exactly f_n is not equal to P on exactly one of the queried points. Note that the value of $f_k(y)$ is obtained by querying f at some point, which we denote x^y. In the BKSSZ analysis x^a and x^b are completely independent given $a \neq b \in \{0,1\}^k$. (Note that the mapping from y to x^y is randomized and depends on the random choices of the tester.) In our setting the behavior of x^a and x^b is more complex and depends on both the set of coordinates j such that where $a_j \neq b_j$ and on the number of indices $i \in [n]$ such that the variable x_i is mapped to variable y_j. Our analysis ends up depending on two new ingredients: (1) The number of variables x_i that map to any particular variable y_j is $\Omega(n/k)$ with probability at least $2^{-O(k)}$ (see Corollary 4.9). This part involves the analysis of a random set-union process elaborated on below. (2) Once the exact number of indices i such that x_i maps to y_j is fixed for every $j \in [k]$ and none of the sets is too small, the distribution of x^a and x^b is sufficiently independent to ensure that the events $f(x^a) = P(x^a)$ and $f(x^b) = P(x^b)$ co-occur with probability much smaller than the individual probabilities of these events. This part uses the hypercontractivity of the hypercube but under an unusual noise operator corresponding to the "constant weight operator", fortunately analyzed by Polyanskiy [21]. Invoking his theorem we are able to conclude the proof of this section.

We now briefly expand on the "random set-union" process alluded to above. Recall that our process starts with n variables, and at each stage a pair of remaining variables is identified and given the same name. (We may ignore the complications due to the complementation of the form $z_i = 1 - z_j$ for this part.) Equivalently we start with n sets $X_1, \ldots, X_n$ with $X_i = \{i\}$ initially. We then pick two random sets and merge them. We stop when there are k sets left and our goal is to understand the likelihood that one of the sets turn out to be too tiny. (The expected size of a set is n/k and too tiny corresponds to being smaller than $n/(4k)$.) It turns out that the distribution of set sizes produced by this process has a particularly clean description as follows: Randomly arrange the elements 1 to n on a cycle and consider the partition into k sets generated by the set of elements that start with a special element and end before the next special element as we go clockwise around the cycle, where the elements in $\{1, \ldots, k\}$ are the special ones. The sizes of these partitions are distributed identically to the sizes of the sets S_j! For example, when $k = 2$ the two sets have sizes distributed uniformly from 1 to $n - 1$. In particular the sets size are not strongly concentrated around n/k - but nevertheless the probability that no set is tiny is not too small and this suffices for our analysis.

Details of this analysis may be found in Section 4.

Organization. In Section 2 we start with some preliminaries including the main definitions and some of the tools we will need later. In Section 3 we give a formal statement of our results. In Section 4 we present and analyze the local tester over all fields. In Section 5 we show that over

fields of large (or zero) characteristic, local decoding is not possible. Finally in Section 6 we give a local decoder and its analysis over fields of small characteristic.

2 Preliminaries

2.1 Basic notation

Fix a field $\mathbb{F}$ and an $n \in \mathbb{N}$. We consider functions $f : \{0,1\}^n \to \mathbb{F}$ that can be written as *multilinear* polynomials of total degree at most d. We denote this space by $\mathcal{F}(n, d; \mathbb{F})$. The space of all functions from $\{0,1\}^n$ to $\mathbb{F}$ will be denoted simply as $\mathcal{F}(n; \mathbb{F})$. (We will simplify these to $\mathcal{F}(n, d)$ and $\mathcal{F}(n)$ respectively, if the field $\mathbb{F}$ is clear from context.)

Given $f, g \in \mathcal{F}(n)$, we use $\delta(f, g)$ to denote the fractional Hamming distance between f and g. I.e.,

$$\delta(f, g) := \Pr_{x \in \{0,1\}^n} [f(x) \neq g(x)]$$

For a family $\mathcal{F}' \subseteq \mathcal{F}(n)$, we use $\delta(f, \mathcal{F}')$ to denote $\min_{g \in \mathcal{F}'}\{\delta(f, g)\}$. Given an $f \in \mathcal{F}(n)$ and $d \geq 0$, we use $\delta_d(f)$ to denote $\delta(f, \mathcal{F}(n, d))$.

2.2 Local Testers and Decoders

Let $\mathbb{F}$ be any field. We define the notion of a local tester and local decoder for subspaces of $\mathcal{F}(n)$.

Definition 2.1 (Local tester)**.** *Fix $q \in \mathbb{N}$ and $\delta \in (0, 1)$. Let $\mathcal{F}'$ be any subspace of $\mathcal{F}(n)$.*

We say that a randomized algorithm T is a (δ, q)-local tester for $\mathcal{F}'$ *if on an input $f \in \mathcal{F}(n)$, the algorithm does the following.*

- *T makes at most q non-adaptive queries to f and either accepts or rejects.*
- *(Completeness) If $f \in \mathcal{F}'$, then T accepts with probability at least $3/4$.*
- *(Soundness) If $\delta(f, \mathcal{F}') \geq \delta$, then T rejects with probability at least $3/4$.*

We say that a tester is adaptive *if the queries it makes to the input f depend on the answers to its earlier queries. Otherwise, we say that the tester is* non-adaptive.

Definition 2.2 (Local decoder)**.** *Fix $q \in \mathbb{N}$ and $\delta \in (0, 1)$. Let $\mathcal{F}'$ be any subspace of $\mathcal{F}(n)$.*

We say that a randomized algorithm T is a (δ, q)-local decoder for $\mathcal{F}'$ *if on an input $f \in \mathcal{F}(n)$ and $x \in \{0,1\}^n$, the algorithm does the following.*

- *T makes at most q queries to f and outputs $b \in \mathbb{F}$.*
- *If $\delta(f, \mathcal{F}') \leq \delta$, then the output $b = f(x)$ with probability at least $3/4$.*

We say that a decoder is adaptive *if the queries it makes to the input f depend on the answers to its earlier queries. Otherwise, we say that the tester is* non-adaptive.

2.3 Some basic facts about binomial coefficients

Fact 2.3. *For integer parameters $0 \le b \le a$, let $\binom{a}{b}$ denote the size of a Hamming ball of radius b in $\{0,1\}^a$; equivalently, $\binom{a}{\le b} = \sum_{j \le b} \binom{a}{j}$. Then, we have*

$$\binom{a}{\le b} \le 2^{aH(b/a)}$$

where $H(\cdot)$ is the binary entropy function.

2.4 Hypercontractivity theorem for spherical averages.

In this section, let $\mathbb{R}$ be the underlying field. Let $\eta \in (0,1)$ be arbitrary. We define a smoothing operator T_η, which maps $\mathcal{F}(r) = \{f : \{0,1\}^r \to \mathbb{R}\}$ to itself. For $F \in \mathcal{F}(r)$, we define $T_\eta F$ as follows

$$T_\eta F(x) = \mathop{\mathbf{E}}_{J \in \binom{[r]}{\eta r}} [F(x \oplus J)]$$

where $x \oplus J$ is the point $y \in \{0,1\}^r$ obtained by flipping x at exactly the coordinates in J.

Recall that for any $F \in \mathcal{F}(r)$ and any $p \ge 1$, $\|F\|_p$ denotes $\mathbf{E}_{x \in \{0,1\}^r} [|F(x)|^p]^{1/p}$.

We will use the following hypercontractivity theorem of Polanskiy [21].

Theorem 2.4 (Follows from Theorem 1 in [21])**.** *Assume that $\eta \in [1/20, 19/20]$ and $\eta_0 = 1/20$. For any $F \in \mathcal{F}(r)$, we have*

$$\|T_\eta F\|_2 \le C \cdot \|F\|_p$$

for $p = 1 + (1-2\eta_0)^2$ and C is an absolute constant.

Corollary 2.5. *Assume that η_0, η are as in the statement of Theorem 2.4 and let $\delta \in (0,1)$ be arbitrary. Say $E \subseteq \{0,1\}^r$ s.t. $|E| \le \delta \cdot 2^r$. Assume that $(x', x'') \in \{0,1\}^r$ are chosen as follows: $x' \in \{0,1\}^r$ and $I' \in \binom{[r]}{\eta r}$ are chosen i.u.a.r., and we set $x'' = x' \oplus I'$. Then we have*

$$\Pr_{x', I'} \left[x' \in E \wedge x'' \in E\right] \le C \cdot \delta^{1+(1/40)}$$

where C is the constant from Theorem 2.4.

Proof. Let $F : \{0,1\}^n \to \{0,1\} \subseteq \mathbb{R}$ be the indicator function of the set E. Note that we have

$$\Pr_{x', I'} \left[x' \in E \wedge x'' \in E\right] = \mathop{\mathbf{E}}_{x', I'} \left[F(x')F(x' \oplus I')\right] = \mathop{\mathbf{E}}_{x'} \left[F(x') T_\eta F(x')\right].$$

By the Cauchy-Schwarz inequality and Theorem 2.4 we get

$$\mathop{\mathbf{E}}_{x'} \left[F(x') T_\eta F(x')\right] \le \|F\|_2 \cdot C \cdot \|F\|_p \tag{1}$$

for $p = 1 + (1-2\eta_0)^2$. Note that we have

$$\begin{aligned} \|F\|_p &\le \delta^{1/p} = \delta^{\frac{1}{1+(1-2\eta_0)^2}} \\ &= \delta^{\frac{1}{2(1-2\eta_0(1-\eta_0))}} \le (\sqrt{\delta})^{1+\min\{\eta_0, 1-\eta_0\}} = \sqrt{\delta}^{1+(1/20)} \end{aligned}$$

where for the last inequality we have used the fact that for $\eta_0 \in [0,1]$ we have

$$\frac{1}{1-2\eta_0(1-\eta_0)} \geq 1 + 2\eta_0(1-\eta_0) \geq 1 + \min\{\eta_0, 1-\eta_0\}.$$

Putting the upper bound on $\|F\|_p$ together with the fact that $\|F\|_2 \leq \sqrt{\delta}$ and (1), we get the claim. □

3 Results

We show upper and lower bounds for testing and decoding polynomial codes over grids. All our upper bounds hold in the non-adaptive setting, while our lower bounds hold in the stronger adaptive setting.

Our first result is that for any choice of the field $\mathbb{F}$ (possibly even infinite), the space of functions $\mathcal{F}(n,d)$ is locally testable. More precisely, we show the following.

Theorem 3.1 ($\mathcal{F}(n,d)$ has a local tester for any field)**.** *Let $\mathbb{F}$ be any field. Fix a positive integer d and any $n \in \mathbb{N}$. Then the space $\mathcal{F}(n,d;\mathbb{F})$ has a non-adaptive (ε, q)-local tester for $q = 2^{O(d)} \cdot \mathrm{poly}(1/\varepsilon)$.*

In contrast, we show that the space $\mathcal{F}(n,d)$ is *not* locally *decodable* over fields of large characteristic, even for $d = 1$.

Theorem 3.2 ($\mathcal{F}(n,d)$ does not have a local decoder for large characteristic)**.** *Let $n \in \mathbb{N}$ be a growing parameter. Let $\mathbb{F}$ be any field such that either $\mathrm{char}(\mathbb{F}) = 0$ or $\mathrm{char}(\mathbb{F}) \geq n^2$. Then any adaptive (ε, q)-local decoder for $\mathcal{F}(n,1;\mathbb{F})$ that corrects an ε fraction of errors must satisfy $q = \Omega_\varepsilon(\log n / \log\log n)$.*

Complementing the above result, we can show that if $\mathrm{char}(\mathbb{F})$ is a constant, then in fact the space $\mathcal{F}(n,d)$ does have a local decoding procedure.

Theorem 3.3 ($\mathcal{F}(n,d)$ has a local decoder for constant characteristic)**.** *Let $\mathrm{char}(\mathbb{F}) = p$ be a positive constant. Fix any $d, n \in \mathbb{N}$. There is a $k \leq pd$ such that the space $\mathcal{F}(n,d;\mathbb{F})$ has a non-adaptive $(1/2^{O(k)}, 4^k)$-local decoder.*

4 A local tester for $\mathcal{F}(n,d)$ over any field

We now present our local tester and its analysis. The reader may find the overview from Section 1.3 helpful while reading the below.

We start by introducing some notation for this section. Throughout, fix any field $\mathbb{F}$. We consider functions $f : \{0,1\}^I \to \mathbb{F}$ where I is a finite set of positive integers and indexes into the set of variables $\{X_i \mid i \in I\}$. We denote this space as $\mathcal{F}(I)$. Similarly, $\mathcal{F}(I,d)$ is defined to be the space of functions of degree at most d over the variables indexed by I.

The following is the test we use to check if a given function $f : \{0,1\}^I \to \mathbb{F}$ is close to $\mathcal{F}(I,d)$.

Test $T_{k,I}(f_I)$

Notation. Given two variables X and Y and $a \in \{0,1\}$, "replacing X by $a \oplus Y$" refers to substituting X by Y if $a = 0$ and by $1 - Y$ if $a = 1$.

- If $|I| > k$, then
  - Choose a random $a \in \{0,1\}$ and distinct $i_0, j_0 \in I$ at random and replace X_{j_0} by $a \oplus X_{i_0}$. Let f'_I denote the resulting restriction of f_I.
  - Run $T_{k,I\setminus\{j_0\}}(f'_I)$ and output what it outputs.
- If $|I| = k$ then
  - Choose a uniformly random bijection $\sigma : I \to [k]$.
  - Choose an $a \in \{0,1\}^k$ uniformly at random.
  - Replace each X_i $(i \in I)$ with $Y_{\sigma(i)} \oplus a_i$.
  - Check if the restricted function $g(Y_1, \ldots, Y_k) \in \mathcal{F}(k,d)$ by querying g on all its inputs. Accept if so and reject otherwise.

Remark 4.1. *It is not strictly necessary to choose a* random *bijection σ in the test $T_{k,I}$ and a fixed bijection $\sigma : I \to [k]$ would do just as well. However, the above leads to a cleaner reformulation of the test in Section 4.1 below.*

Observation 4.2. *Test $T_{k,I}$ has query complexity 2^k.*

Observation 4.3. *If $f_I \in \mathcal{F}(I,d)$, then $T_{k,I}$ accepts with probability 1.*

The following theorem is the main result of this section and implies Theorem 3.1 from Section 3.

Theorem 4.4. *For each positive integer d, there is a $k = O(d)$ and $\varepsilon_0 = 1/2^{O(d)}$ such that for any I of size at least $k+1$ and any $f_I \in \mathcal{F}(I)$,*

$$\Pr\left[\textit{Test } T_{k,I} \textit{ rejects } f_I\right] \geq \frac{1}{2^{O(d)}} \cdot \min\{\delta_d(f_I), \varepsilon_0\}.$$

Theorem 3.1 immediately follows from Theorem 4.4 since to get an $(\varepsilon, 2^{O(d)})$-tester, we repeat the test $T_{k,[n]}$ $t = 2^{O(d)} \cdot \mathrm{poly}(1/\varepsilon)$ many times and accept if and only if each iteration of the test accepts. If the input function $f \in \mathcal{F}(n)$ is of degree at most d, this test accepts with probability 1. Otherwise, this test rejects with probability at least $3/4$ for suitably chosen t as above. The number of queries made by the test is $2^k \cdot t = 2^{O(d)} \cdot \mathrm{poly}(1/\varepsilon)$.

Parameters. For the rest of this section, we use the following parameters. We choose

$$k = M \cdot d \tag{2}$$

for a large absolute constant $M \in \mathbb{N}$ and set

$$\varepsilon_1 = \frac{1}{(4C \cdot 2^{k \cdot H(1/M)})^{40}} \tag{3}$$

where C is the absolute constant from Corollary 2.5. The constant M is chosen so that

$$H(1/M) < \frac{1}{20} \qquad \text{and} \qquad k \geq 100 \log \frac{2}{\varepsilon_1}. \tag{4}$$

Note that the second constraint is satisfied for a large enough absolute constant M since we have

$$\frac{100 \log(2/\varepsilon_1)}{k} \leq \frac{40kH(1/M) + 40 \log C + O(1)}{k} \leq 40H(1/M) + \frac{40 \log C + O(1)}{M}$$

which can be made arbitrary small for large enough constant M. Further, we set

$$\ell = \log \frac{2}{\varepsilon_1} \qquad \text{and} \qquad \varepsilon_0 = \frac{\varepsilon_1}{100\ell}. \tag{5}$$

The following are the two main lemmas used to establish Theorem 4.4.

Lemma 4.5 (Small distance lemma)**.** *Fix any I such that $|I| = r > k+1$ and $f_I : \{0,1\}^I \to \mathbb{F}$ such that $\delta_d(f_I) = \delta \leq \varepsilon_1$.*

$$\Pr\left[T_{k,I} \text{ rejects } f_I\right] \geq \frac{\delta}{2^{O(k)}}.$$

Lemma 4.6 (Large distance lemma)**.** *Fix any I such that $|I| = r$ satisfies $r^2 > 100\ell^2$ and $f_I : \{0,1\}^I \to \mathbb{F}$ such that $\delta_d(f_I) > \varepsilon_1$. Then*

$$\Pr\left[\delta_d(f'_I) < \varepsilon_0\right] < \frac{100\ell^2}{r^2}.$$

With the above lemmas in place, we show how to finish the proof of Theorem 4.4.

Proof of Theorem 4.4. Fix any I and consider the behaviour of the test $T_{k,I}$ on f_I. Assume $|I| = n$.

A single run of $T_{k,I}$ produces a sequence of functions $f_n = f_I, f_{n-1}, \ldots, f_k$, where f_r is a function on r variables. Let $I_n = I, I_{n-1}, \ldots, I_k$ be the sequence of index sets produced. We have $f_r : \{0,1\}^{I_r} \to \mathbb{F}$. Note that $k \geq 100\ell$ by (4) and (5).

Define the following pairwise disjoint events for each $r \in \{k, \ldots, n\}$.

- $\mathfrak{F}_r$ is the event that $\delta_d(f_r) > \varepsilon_1$.
- $\mathfrak{C}_r$ is the event that $\delta_d(f_r) < \varepsilon_0$.
- $\mathcal{E}_r$ is the event that $\delta_d(f_r) \in [\varepsilon_0, \varepsilon_1]$.

For any f_I, one of $\mathfrak{F}_r, \mathfrak{C}_r$, or $\mathcal{E}_r$ occurs with probability 1. If either $\mathcal{E}_n$ or $\mathfrak{C}_n$ occurs, then by Lemma 4.5 we are done. Therefore, we assume that $\mathfrak{F}_n$ holds.

We note that one of the following possibilities must occur: either all the f_r satisfy $\delta_d(f_r) > \varepsilon_1$; or there is some f_r such that $\delta_d(f_r) \in [\varepsilon_0, \varepsilon_1]$; or finally, there is some f_r such that $\delta_d(f_{r+1}) > \varepsilon_1$ but $\delta_d(f_r) < \varepsilon_0$. We handle each of these cases somewhat differently.

Clearly, if $\mathfrak{F}_k$ holds, then $\deg(f_{I_k}) > d$ and hence T_{k,I_k} rejects f_{I_k} with probability 1. On the other hand, by Lemma 4.5, we see that

$$\Pr\left[T_{k,I} \text{ rejects } f_I \mid \bigvee_{r=k}^{n-1} \mathcal{E}_r\right] \geq \frac{\varepsilon_0}{2^{O(k)}}.$$

Thus, we have

$$\Pr\left[T_{k,I} \text{ rejects } f_I\right] \geq \frac{\varepsilon_0}{2^{O(k)}} \cdot \Pr\left[\bigvee_{r=k}^{n-1} \mathcal{E}_r \vee \bigwedge_{r=k}^{n-1} \mathfrak{F}_r\right] \tag{6}$$

Let $\mathcal{E}$ denote the event $\neg(\bigvee_{r=k}^{n-1} \mathcal{E}_r \vee \bigwedge_{r=k}^{n-1} \mathfrak{F}_r)$. Notice that if event $\mathcal{E}$ occurs, there must be an $r \geq k$ such that $\mathfrak{C}_r$ occurs but we also have $\mathfrak{F}_{r+1} \wedge \mathfrak{F}_{r+2} \wedge \cdots \wedge \mathfrak{F}_n$. By Lemma 4.6, the probability of this is upper bounded by $100\ell^2/r^2$ for each $r \geq k$.

By a conditional probability argument, we see that

$$\Pr\left[\neg\mathcal{E}\right] \geq \prod_{r \geq k} \left(1 - \frac{100\ell^2}{r^2}\right) \geq \exp\left(-200\ell^2 \sum_{r \geq k} \frac{1}{r^2}\right) \geq \exp(-O(\ell)) = \frac{1}{2^{O(k)}}$$

where we have used the fact that $k \geq 100\ell$ and for the second inequality we also use $(1 - x) \geq \exp(-2x)$ for $x \in [0, 1/2]$. Plugging the above into (6), we get the theorem. □

It remains to prove Lemmas 4.5 and 4.6 which we do in Sections 4.1 and 4.2 respectively.

4.1 Proof of Small Distance Lemma (Lemma 4.5)

We start with a brief overview of the proof of Lemma 4.5. Suppose f_I is δ-close to some polynomial P for some $\delta \leq \varepsilon_1$. As mentioned in Section 1.3, our aim is to show that the (random) restriction g of f obtained above and the corresponding restriction Q of P differ at only one point. Then we will be done since any two distinct degree-d polynomials on $\{0,1\}^k$ must differ on at least 2 points (if $k > d$) and hence the restricted function g cannot be a degree-d polynomial.

Note that the restriction is effectively given by $a \in \{0,1\}^I$ and $\phi : I \to [k]$ such that $g(y) = f_I(x(y))$ where $x(y) = (x_i(y))_{i \in I}$ is given by $x_i(y_1, \ldots, y_k) = y_{\phi(i)} \oplus a_i$. ($\phi$ is obtained by a sequence of replacements followed by the bijection σ.) Similarly we define $Q(y) = P(x(y))$. To analyze the test, we consider the queries $\{x(y)\}_{y \in \{0,1\}^k}$ made to the oracle for f_I. For every fixed $y \in \{0,1\}^k$ the randomness (in a and ϕ) leads to a random query $x(y) \in \{0,1\}^I$ to f_I and it is not hard to show that for each fixed y, $x(y)$ is uniformly distributed over $\{0,1\}^I$. Hence, the probability that g and Q differ at any fixed $y \in \{0,1\}^k$ is exactly δ.

We would now like to say that for distinct $y', y'' \in \{0,1\}^k$, the probability that g and Q differ at *both* y' and y'' is much smaller than δ. This would be true if, for example, $x(y')$ and $x(y'')$ were independent of each other, but this is unfortunately not the case. For example, consider the case when no X_i ($i \in I$) is identified with the variable Y_k (i.e., for every $i \in I$, $\phi(i) \neq k$).[1] In this case, $x(y') = x(y'')$ for every y' and y'' that differ only at the kth position. More generally, if the number of variables that are identified with Y_k is very small (much smaller than the expected number r/k) then $x(y')$ and $x(y'')$ would be heavily correlated if y' and y'' differed in only the kth coordinate.

So, the first step in our proof is to analyze the above restriction process and show that with reasonable probability, for every Y_j there are many variables (close to the expected number) mapped to it, i.e., $|\phi^{-1}(j)|$ is $\Omega(r/k)$ for every $j \in [k]$. To get to this analysis we first give an alternate (non-iterative) description of the test $T_{k,I}$ and analyze it by exploring the random set-union process mentioned in Section 1.3. We note that this process and its analysis may be independently interesting.

[1] Strictly speaking this case can not occur due to the way ϕ is constructed, but it is is useful to think about this case anyway.

Once we have a decent lower bound on $\min_j |\phi^{-1}(j)|$, we can use the hypercontractivity theorem of Polyanskiy (Theorem 2.4) to argue that for any $y' \neq y''$, the inputs $x(y')$ and $x(y'')$ are somewhat negatively correlated (see Corollary 2.5). We note that since the distribution of the pair $(x(y'), x(y''))$ is not the usual noisy hypercube distribution and so the usual hypercontractivity does not help. But this is where the strength of Polyanskiy's hypercontractivity comes in handy — even after we fix the Hamming distance between $x(y')$ and $x(y'')$ the symmetry of the space leads to enough randomness to apply Theorem 2.4. This application already allows us to show a weak version of Lemma 4.5 and hence a weak version of our final tester.

To prove Lemma 4.5 in full strength as stated, we note that stronger parameters for the lemma are linked to stronger negative correlation between $x(y')$ and $x(y'')$ for various y' and y''. It turns out that this is directly related to the Hamming distance of y' and y'': specifically, we would like their Hamming distance to not be too close to 0 or to k. Hence, we would like to restrict our attention to a subset T of the query points of $\{0,1\}^k$ that form such a "code". At the same time, however, we need to ensure that, as for $\{0,1\}^k$, any two distinct degree-d polynomials cannot differ at exactly one point in T. We construct such a set T in Claim 4.10, and use it to prove Lemma 4.5.

We now begin the formal proof with an alternate but equivalent (non-recursive) description of test $T_{k,I}$ for $|I| = r > k$.

Test $T_{k,I}$ (Alternate description)

- Choose $a \in \{0,1\}^r$ uniformly at random.
- Choose a bijection $\pi : [r] \to I$ uniformly at random.
- Choose $p : \{k+1, \ldots, r\} \to \mathbb{Z}$ so that each $p(i)$ is uniformly distributed over the set $\{1, \ldots, i-1\}$ and the $p(i)$s are mutually independent. (Here $p(i)$ stands for the "parent of i").
- For i in $r, r-1, \ldots, k+1$
  - Substitute $X_{\pi(i)}$ by $a_i \oplus X_{\pi(p(i))}$.
- For $i \in 1, \ldots, k$
  - Replace each $X_{\pi(i)}$ with $a_i \oplus Y_i$ for each $i \in [k]$.
- Check if the restricted function $g(Y_1, \ldots, Y_k)$ is of degree at most d by querying g on all its inputs. Accept if so and reject otherwise.

Proposition 4.7. *The iterative description above is equivalent to test $T_{k,I}$.*

We now begin the analysis of the test $T_{k,I}$. As stated above, the first step is to understand the distribution of the number of X_i ($i \in I$) eventually identified with Y_j (for various $j \in [k]$). We will show (Corollary 4.9) that with reasonable probability, each Y_j has $\Omega(r/k)$ X_is that are identified with it.

Fix any bijection $\pi : [r] \to [r]$. For i, j such that $i \geq j$ and $i \in \{k, \ldots, r\}$, we define $B_{j,i}$ to be the index set of those variables that are identified with $X_{\pi(j)}$ (or its complement) in the first $r - i$ rounds of substitution. Formally,

$$B_{j,i} = \begin{cases} \{\pi(j)\} & \text{if } i = r, \\ B_{j,i+1} & \text{if } i < r \text{ and } p(i+1) \neq j. \\ B_{j,i+1} \cup B_{i+1,i+1} & \text{if } i < r \text{ and } p(i+1) = j. \end{cases}$$

For $j \in [k]$, let $B_j = B_{j,k}$. This is the set of i such that $X_{\pi(i)}$ is "eventually" identified with $X_{\pi(j)}$ (or its complement). For $i \in [r]$, we define $b(i) = j$ if $i \in B_j$.

To analyze the distribution of the "buckets" $B_1, \ldots, B_k$, it will be helpful to look at an equivalent way of generating this distribution. We do this by sampling the buckets in "reverse": i.e., we start with the jth bucket being the singleton set $\{j\}$ and for each $i = k+1, \ldots, r$, we add i to the jth bucket if i falls into the the jth bucket.

Formally, for each $j \in [k]$, define the set $B'_{j,i}$ to be $B_j \cap [i]$. Note that we have

$$B'_{j,i+1} = \begin{cases} \{j\} & \text{if } i = k, \\ B'_{j,i} & \text{if } i > k \text{ and } p(i+1) \notin B'_{j,i}. \\ B'_{j,i} \cup \{i+1\} & \text{if } i > k \text{ and } p(i+1) \in B'_{j,i}. \end{cases}$$

In particular, we see that for any $i \geq k+1$,

$$\Pr_p \left[\pi(i+1) \in B'_{j,i+1} \mid B'_{1,i}, \ldots, B'_{k,i}\right] = \Pr_p \left[p(i+1) \in B'_{j,i} \mid p(k+1), \ldots, p(i)\right] = \frac{|B'_{j,i}|}{i}. \quad (7)$$

This yields the following equivalent way of sampling sets from the above distribution.

Lemma 4.8. *Consider the following sampling algorithm that partitions $[r]$ into k parts. Choose a random permutation σ of the set $\{1, \ldots, r\}$ as follows. First choose a uniform element $i_1 \in \{1, \ldots, k\}$. Now choose a uniformly random permutation σ of $[r]$ such that $\sigma(1) = 1$ and assume that the elements of $[k] \setminus \{i_1\}$ appear in the order $i_2, \ldots, i_k$ in σ ($\sigma(i_2) < \cdots < \sigma(i_k)$). Define $C_1, \ldots, C_k$ as follows:*

- $C_1 = \{j > k \mid \sigma(j) < \sigma(i_2)\} \cup \{1\}$,
- $C_2 = \{j > k \mid \sigma(i_2) < \sigma(j) < \sigma(i_3)\} \cup \{2\}$,
- $\ldots$
- $C_{k-1} = \{j > k \mid \sigma(i_{k-1}) < \sigma(j) < \sigma(i_k)\} \cup \{k-1\}$,
- $C_k = \{j > k \mid \sigma(i_k) < \sigma(j)\} \cup \{k\}$.

Then the distribution of $(C_1, \ldots, C_k)$ is identical to the distribution of $(B_1, \ldots, B_k)$.

Proof. Assume σ is sampled by starting with the element 1 and then inserting the elements $i = 2, \ldots, r$ one by one in a random position *after* 1 (since we are sampling σ such that $\sigma(1) = 1$). Simultaneously, consider the evolution of the jth bucket. Let $C_{j,i}$ denote the jth bucket after elements $2, \ldots, i$ have been inserted.

Note that no matter how the first k elements are ordered in σ, the element $j \in [k]$ goes to the jth bucket at the end of the sampling process. Thus, after having inserted $2, \ldots, k$, we have $C_{j,k} = \{j\}$.

We now insert $(i+1)$ for each i such that $k \leq i < r$. The position of $i+1$ is a uniform random position after the first position. For each i, the probability that $i+1$ ends up in the jth bucket can be seen to be $|C_{j,i}|/i$, exactly as in (7). This shows that $(C_1, \ldots, C_k)$ has the same distribution as $(B_1, \ldots, B_k)$. □

Corollary 4.9. *With probability at least $\frac{1}{2^{O(k)}}$ we have $|B_j| \geq \frac{r}{4k}$ for each $j \in [k]$.*

Proof. We assume that $r > 4k$ since otherwise the statement to be proved is trivial (as each $|B_j| \geq 1$ with probability 1.) By Lemma 4.8 it suffices to prove the above statement for the sets $(C_1, \ldots, C_k)$.

Now, say a permutation σ of $[r]$ fixing 1 is chosen u.a.r. and we set C_j as in Lemma 4.8. We view the process of sampling σ as happening in two stages: we first choose a random linear ordering of $A = \{k+1, \ldots, r\}$, i.e. a random function $\sigma' : A \to [r-k]$, and then inserting the elements $2, \ldots, k$ one by one at random locations in this ordering. (The position of the element 1 is of course determined.)

Condition on any choice of σ'. For $j \in \{2, \ldots, k\}$, let $C'_j = \{i \mid (j-1)r/k \leq \sigma'(i) \leq (j-1)r/k + \lceil r/2k \rceil\}$. Fix any bijection $\tau : \{2, \ldots, k\} \to \{2, \ldots, k\}$.

Consider the probability that on inserting $2, \ldots, k$ into the ordering σ', each $j \in \{2, \ldots, k\}$ is inserted between two elements of $C'_{\tau(j)}$. Call this event $\mathcal{E}_\tau$. Conditioned on this event, it can be seen that for each $j \in \{2, \ldots, k\}$, the jth bucket C_j has size at least

$$\min\{a \mid a \in C'_j\} - \max\{a \mid a \in C'_{j-1}\} \geq \frac{jr}{k} - (\frac{(j-1)r}{k} + \frac{r}{2k} + 1) = \frac{r}{2k} - 1 \geq \frac{r}{4k}$$

where we have defined $C'_1 = \{0\}$. Similarly, conditioned on $\mathcal{E}_\tau$, we have $|C_1| \geq r/k \geq r/(4k)$.

Since this holds for each τ and the events $\mathcal{E}_\tau$ are mutually exclusive, we have

$$\Pr\left[\forall j \in [k],\ |C_j| \geq \frac{r}{4k}\right] \geq \sum_\tau \Pr\left[\mathcal{E}_\tau\right].$$

We now analyze $\Pr[\mathcal{E}_\tau]$ for any fixed τ. Conditioned on the positions of $2, \ldots, j-1$, the probability that $\sigma(j) \in C'_{\tau(j)}$ is at least $(r/(2k)) \cdot (1/r) = 1/(2k)$. Therefore we have

$$\Pr\left[\mathcal{E}_\tau\right] \geq 1/2^{k-1}k^{k-1}.$$

Thus, we get

$$\Pr\left[\forall j \in \{2, \ldots, k\},\ |C_j| \geq \frac{r}{4k}\right] \geq \sum_\tau \Pr\left[\mathcal{E}_\tau\right] \geq \frac{(k-1)!}{2^{k-1} \cdot k^{k-1}} \geq \frac{1}{2^{O(k)}},$$

where we have used the Stirling approximation for the final inequality. This concludes the proof of the corollary. □

Note that the sets B_j are determined by our choice of p. For the rest of the section, we condition on a choice of $p = p_0$ such that Corollary 4.9 holds. We now show how to finish the proof of Lemma 4.5.

Fix a polynomial $P \in \mathcal{F}(I, d)$ such that $\delta(f_I, P) = \delta_d(f_I) = \delta$ as in the lemma statement. Let $E \subseteq \{0,1\}^I$ be the set of points where f and P differ. We have $\frac{|E|}{2^r} = \delta \leq \varepsilon_1$.

For $y', y'' \in \{0,1\}^k$, we use $\Delta(y', y'')$ to denote the Hamming distance between them and $\Delta'(y', y'')$ to denote the quantity $\min\{\Delta(y', y''), k - \Delta(y', y'')\}$.

We prove the following two claims.

Claim 4.10. *There is a non-empty set $T \subseteq \{0,1\}^k$ such that:*

- $|T| \leq \binom{k}{\leq d} + 1$,
- *Given distinct $y', y'' \in T$, $\Delta'(y', y'') \geq k/4$,*

- *No pair of polynomials $P, P' \in \mathcal{F}(I,d)$ can differ at exactly one input from T.*[2]

For each input $y \in \{0,1\}^k$ to the restricted polynomial g, let $x(y) \in \{0,1\}^I$ be the corresponding input to f_I. Let S denote the multiset $\{x(y) \mid y \in T\}$. This is a subset of the set of inputs on which f_I is queried.

Claim 4.11. *Let $p = p_0$ be as chosen above. With probability at least $\delta \cdot (|T|/2)$ over the choice of π and a, we have $|S \cap E| = 1$ (i.e. there is a unique $y \in T$ such that $x(y) \in E$).*

Assuming Claims 4.10 and 4.11, we have proved Lemma 4.5 since with probability at least $\frac{1}{2^{O(k)}} \cdot \delta \cdot (|T|/2)$ (cf. Corollary 4.9 and Claim 4.11), the restricted function $g(Y_1, \ldots, Y_k)$ differs from the restriction $P'(Y_1, \ldots, Y_k)$ of P at exactly 1 point in T. However, by our choice of the set T, any two polynomials from $\mathcal{F}(k,d)$ that differ on T must differ on at least two inputs. Hence, g cannot be a degree d polynomial, and thus the test rejects.

4.1.1 Proof of Claim 4.10

Given functions $f, g \in \mathcal{F}(k)$, we define their inner product $\langle f, g \rangle$ by $\langle f, g \rangle = \sum_{y \in \{0,1\}^k} f(y)g(y)$. Recall that $\mathcal{F}(k,d)^\perp$ is defined to be the set of all $f \in \mathcal{F}(k)$ such that $\langle f, g \rangle = 0$ for each $g \in \mathcal{F}(k,d)$.

We will construct T by finding a suitable non-zero $f \in \mathcal{F}(k,d)^\perp$ and setting $T = \mathrm{Supp}(f)$, where $\mathrm{Supp}(f) = \{y \in \{0,1\}^k \mid f(y) \neq 0\}$. Thus, we need f to satisfy the following properties.

1. $|\mathrm{Supp}(f)| \leq \binom{k}{\leq d} + 1$,
2. Given distinct $y', y'' \in \mathrm{Supp}(f)$, $\Delta'(y', y'') \geq k/4$,
3. No pair of polynomials $P, P' \in \mathcal{F}(I,d)$ can differ at exactly one input from $\mathrm{Supp}(f)$.

We first observe that Property 3 is easily satisfied. To see this, assume that $g_1, g_2 \in \mathcal{F}(k,d)$ differ at exactly one point, say y', from $\mathrm{Supp}(f)$. Then, since $g = g_1 - g_2 \in \mathcal{F}(k,d)$ and $f \in \mathcal{F}(k,d)^\perp$, we must have $\langle f, g \rangle = 0$. On the other hand since $\mathrm{Supp}(g) \cap \mathrm{Supp}(f) = \{y'\}$, we have

$$\langle f, g \rangle = \sum_{y \in \{0,1\}^k} f(y)g(y) = f(y')g(y') \neq 0$$

which yields a contradiction. Hence, we see that g_1 and g_2 cannot differ at exactly one point in $\mathrm{Supp}(f)$.

We thus need to choose a non-zero $f \in \mathcal{F}(k,d)^\perp$ so that Properties 1 and 2 hold. Note that to ensure that $f \in \mathcal{F}(k,d)^\perp$, it suffices to ensure that for each $A \subseteq [k]$ of size at most d we have

$$\sum_{y \in \{0,1\}^k} f(y) \cdot \prod_{i \in A} y_i = 0. \tag{8}$$

The number of such A is $N = \binom{k}{\leq d}$.

To ensure that Properties 1 and 2 hold, it suffices to ensure that $\mathrm{Supp}(f) \subseteq U$ where $U \subseteq \{0,1\}^k$ is a set of size $N+1$ so that any distinct $y', y'' \in U$ satisfy $\Delta(y', y'') \in [k/4, 3k/4]$. (Note that this implies that $\Delta'(y', y'') \geq k/4$.)

[2] Note that it could be that two distinct polynomials in $\mathcal{F}(I,d)$ agree everywhere in T.

To see that such a set U exists, consider the following standard greedy procedure for finding such a set U: starting with an empty set, we repeatedly choose an arbitrary point z to add to U and then remove all points at Hamming distance at most $k/4$ and at least $3k/4$ from z from future consideration. Note that this procedure can produce up to $2^k/(2\binom{k}{\leq k/4})$ many points. By Fact 2.3 and our choice of k (see (2) and (4)) we have

$$\frac{2^k}{2\binom{k}{\leq k/4}} \geq 2^{k(1-H(1/4))-1} \geq 2^{k/20}$$
$$N = \binom{k}{\leq d} \leq 2^{kH(d/k)} < 2^{k/20}.$$

Hence, the above greedy procedure can be used to produce a set U of size $N+1$ as required.

Since we assume that $\mathrm{Supp}(f) \subseteq U$, ensuring (8) reduces to ensuring the following for each $A \subseteq [k]$ of size at most d:

$$\sum_{y \in U} f(y) \cdot \prod_{i \in A} y_i = 0. \tag{9}$$

Choosing $f(y)$ $(y \in U)$ so that the above holds reduces to solving a system of N homogeneous linear equations (one for each A) with $|U| = N+1$ constraints. By standard linear algebra, this system has a non-zero solution. This yields a non-zero $f \in \mathcal{F}(k,d)^{\perp}$ with the required properties.

4.1.2 Proof of Claim 4.11

Let y', y'' be any two distinct points in T. Let Δ denote $\Delta(y', y'')$ and Δ' denote $\Delta'(y', y'')$. We show that

$$\Pr_{\pi,a}\left[x(y') \in E\right] = \delta \tag{10}$$
$$\Pr_{\pi,a}\left[x(y') \in E \wedge x(y'') \in E\right] \leq C \cdot \delta^{1+(1/40)}. \tag{11}$$

where C is the absolute constant from the statement of Corollary 2.5.

Given (10) and(11) we are done since we can argue by inclusion exclusion as follows.

$$\begin{aligned}
\Pr_{\pi,a}\left[|S \cap E| = 1\right] &\geq \sum_{y \in T} \Pr\left[x(y) \in E\right] - \sum_{y' \neq y'' \in T} \Pr\left[x(y') \in E \wedge x(y'') \in E\right] \\
&\geq \delta \cdot |T| - |T|^2 \cdot C \cdot \delta^{1+(1/40)} && \text{(by (10) and (11))} \\
&\geq \delta \cdot |T|(1 - (\binom{k}{\leq d} + 1) \cdot C \cdot \varepsilon_1^{1/40}) && (\because\ \delta \leq \varepsilon_1, |T| \leq \binom{k}{\leq d} + 1)
\end{aligned}$$

Note that by our choice of ε_1 (see (3)) and Fact 2.3 we have

$$\varepsilon_1^{1/40} \leq \frac{1}{4C2^{kH(d/k)}} \leq \frac{1}{2C \cdot (\binom{k}{\leq d} + 1)},$$

which along with our previous computation yields

$$\Pr_{\pi,a}\left[|S \cap E| = 1\right] \geq \delta \cdot \frac{|T|}{2}.$$

This finishes the proof of the Claim using (10) and (11). We now prove (10) and (11).

To prove (10), we consider the distribution of $x(y')$ for any fixed $y' \in \{0,1\}^k$. Condition on any choice of π. For any $i \in [r]$, let $A_i = \{j \mid i \in \bigcup_{i'} B_{j,i'}\}$. Note that $\pi(j) < \pi(i)$ for each $j \in A_i$. We have

$$x(y')_{\pi(i)} = a_i \oplus \bigoplus_{j \in A_i} a_j \oplus y'_{b(i)}. \tag{12}$$

which is a uniform random bit even after conditioning on all a_j for $j < i$. In particular, it follows that for each choice of π, $x(y')$ is a uniformly random element of $\{0,1\}^I$. This immediately implies (10). Also note that since $x(y')$ has the same distribution *for each choice of* π, the random variables $x(y')$ and π are independent from each other.

To prove (11), we will use our corollary to Polyanskiy's Hypercontractivity theorem (Corollary 2.5). Let $D \subseteq [k]$ be the set of coordinates where y' and y'' differ. Condition on any choice of $x(y') \in \{0,1\}^n$. By (12), the point $x(y'')$ satisfies, for each i,

$$x(y'')_{\pi(i)} \oplus x(y')_{\pi(i)} = y''_{b(i)} \oplus y'_{b(i)}.$$

Or equivalently, for any $h \in I$, we have

$$x(y'')_h \oplus x(y')_h = y''_{b(\pi^{-1}(h))} \oplus y'_{b(\pi^{-1}(h))} = \begin{cases} 1 & \text{if } \pi^{-1}(h) \in \bigcup_{j \in D} B_j \\ 0 & \text{otherwise.} \end{cases}$$

Now, we may rewrite the condition $\pi^{-1}(h) \in \bigcup_{j \in D} B_j$ as $h \in \pi(B_D)$ for $B_D := \bigcup_{j \in D} B_j$. Note that $\pi(B_D) \subseteq I$ is a uniformly random subset of I of size $|B_D|$.

Hence, we may equivalently sample the pair $(x(y'), x(y''))$ as follows: Choose $x(y') \in \{0,1\}^I$ uniformly at random, and choose independently a random set $I' \subseteq I$ of size $|B_D|$ and flip $x(y')$ exactly in the coordinates in I' to get $x(y'')$.

Note that $|B_D| = \sum_{j \in D} |B_j| \geq (\Delta \cdot r)/4k$ since $|D| = \Delta$ and $|B_j| \geq r/4k$ for each j by Corollary 4.9. At the same time, we also have $|[r] \setminus B_D| = \sum_{j \in [k] \setminus D} |B_j| \geq (k - \Delta) \cdot r/(4k)$. Thus, $|B_D| = \eta r$ for some $\eta \in [\Delta/(4k), 1 - (k-\Delta)/(4k)] \subseteq [\Delta'/(4k), 1 - (\Delta'/(4k))]$.

By Claim 4.10, we know that $\Delta'(y', y'') \geq k/4$ and hence we have $\eta \in [1/16, 15/16]$. Applying Corollary 2.5, we see that this implies

$$\Pr_{x(y'), I'} \left[x(y') \in E \wedge x(y'') \in E \right] \leq C \cdot \delta^{1+(1/40)}.$$

This proves (11) and hence finishes the proof of the claim.

4.2 Proof of Large Distance Lemma (Lemma 4.6)

We follow the proof of [7, Lemma 12].

Given a triple $(i, j, b) \in I^2 \times \{0,1\}$ with i, j distinct, call (i, j, b) a *bad triple* if the restricted function f'_I obtained when the test chooses $i_0 = i$, $j_0 = j$ and $a = b$ is ε_0-close to $\mathcal{F}(I \setminus j, d)$. To prove Lemma 4.6, it suffices to show that the number of bad triples is at most $100\ell^2$. To do this, we bound instead the number of *bad pairs*, which are defined to be pairs (i, j) for which there exists $b \in \{0,1\}$ such that (i, j, b) is a bad triple. Note that (i, j) is a bad pair iff (j, i) is. Hence, the set of bad pairs (i, j) defines an undirected graph G_{bad}. If there are fewer than $25\ell^2$ edges in G_{bad}, we are done since this implies that there are at most $50\ell^2$ bad pairs and hence at most $100\ell^2$ bad triples. Otherwise, G_{bad} has more than $25\ell^2$ edges and it is easy to see that one of the following two cases must occur:

- G_{bad} has a matching with at least $\ell + 1$ edges, or
- G_{bad} has a star with at least $\ell + 1$ edges.

We show that in each case, we can find a polynomial $P \in \mathcal{F}(I, d)$ is ε_1-close to f_I, which will contradict the assumption that $\delta_d(f_I) > \varepsilon_1$ and hence finish the proof of the lemma.

We first note that in either the matching or the star case, we can replace some variables X with $1 \oplus X$ in f_I (note that this does not change $\delta_d(f_I)$) to ensure that the bad triples that give rise to the bad pairs are all of the form $(X, X', 0)$: i.e., all the bad triples come from identifying variables (and *not* from identifying a variable with the complement of another).

Let $t_1 = (X_{i_1}, X_{j_1}, 0), \ldots, t_{\ell+1} = (X_{i_{\ell+1}}, X_{j_{\ell+1}}, 0)$ denote the bad triples obtained above (in either the matching or the star case). Each triple t_h defines the subset $R_h \subseteq \{0,1\}^I$ where the variables X_{i_h} and X_{j_h} take the same values; let R'_h denote the complement of R_h. Note that each $|R_h| = 2^{r-1}$. Furthermore, it follows from the form of the triples that for each h we have $|S_1 \cap S_2 \cap \cdots S_h| = 2^{r-h}$ for any choice of $S_1 \in \{R_1, R'_1\}, \ldots, S_h \in \{R_h, R'_h\}$.

By assumption, for each triple t_h, there is a polynomial $P^{(h)}$ such that $P^{(h)}$ is ε_0-close to $f^{(h)}$, where the latter function is obtained by identifying the variables X_{i_h} and X_{j_h} in f_I. We will show the following claim.

Claim 4.12. *There is a $P \in \mathcal{F}(I, d)$ such that for each $h \in [\ell + 1]$, $P(x) = P^{(h)}(x)$ for all $x \in R_h$.*

Assuming the above claim, we show that the polynomial P above is actually ε_1-close to f_I, which contradicts our assumption about $\delta_d(f_I)$.

Consider a uniformly random input $x \in \{0,1\}^n$. We have

$$\Pr_x [f_I(x) \neq P(x)] \leq \sum_{h=1}^{\ell} \Pr_x \left[f_I(x) \neq P(x) \mid x \in R_h \setminus \bigcup_{h'<h} R_{h'} \right] \cdot \Pr \left[x \in R_h \setminus \bigcup_{h'<h} R_{h'} \right] + \Pr_x \left[x \notin \bigcup_{h \leq \ell} R_h \right] \tag{13}$$

For each h, we have

$$\Pr_x \left[x \in R_h \setminus \bigcup_{h'<h} R_{h'} \right] = \Pr_x \left[x \in R_h \cap R'_1 \cap \cdots \cap R'_{h-1} \right] = \frac{1}{2^h}$$

$$\Pr_x \left[x \in R_h \setminus \bigcup_{h'<h} R_{h'} \mid x \in R_h \right] = \frac{\Pr_x \left[x \in R_h \cap R'_1 \cap \cdots \cap R'_{h-1} \right]}{\Pr_x [x \in R_h]} = \frac{1}{2^{h-1}}$$

$$\Pr_x \left[x \notin \bigcup_{h \leq \ell} R_h \right] = \Pr_x \left[x \in R'_1 \cap \cdots \cap R'_\ell \right] = \frac{1}{2^\ell}$$

Since $P(x)$ agrees with $P^{(h)}(x)$ for each $x \in R_h$, we have

$$\Pr_x [f_I(x) \neq P(x) \mid x \in R_h] = \Pr_x \left[f_I(x) \neq P^{(h)}(x) \mid x \in R_h \right] \leq \varepsilon_0.$$

Hence, we obtain

$$\Pr_x\left[f_I(x) \neq P(x) \mid x \in R_h \setminus \bigcup_{h'<h} R_{h'}\right] \leq \frac{\Pr_x\left[f_I(x) \neq P(x) \mid x \in R_h\right]}{\Pr_x\left[x \in R_h \setminus \bigcup_{h'<h} R_{h'} \mid x \in R_h\right]}$$
$$= 2^{h-1} \Pr_x\left[f_I(x) \neq P(x) \mid x \in R_h\right] \leq 2^{h-1}\varepsilon_0.$$

Plugging the above into (13), we get

$$\Pr_x\left[f_I(x) \neq P(x)\right] \leq \left(\sum_{h=1}^{\ell} 2^{h-1}\varepsilon_0 \cdot \frac{1}{2^h}\right) + \frac{1}{2^\ell}$$
$$\leq \frac{\varepsilon_0 \ell}{2} + \frac{1}{2^\ell} < \varepsilon_1$$

where the final inequality follows from our choice of ε_0 and ℓ (see (5)). This is a contradiction to our assumption on $\delta_d(f_I)$, which concludes the proof of Lemma 4.6 assuming Claim 4.12.

4.2.1 Proof of Claim 4.12

We now prove Claim 4.12. The proof is a case analysis based on whether G_{bad} has a large matching or a large star. For any $h \in [\ell+1]$ and any polynomial $Q \in \mathcal{F}(I,d)$, we denote $Q|_h$ the polynomial obtained by identifying the variables X_{i_h} and X_{j_h}. We want to define a polynomial P such that for each $h \in [\ell+1]$, we have

$$P|_h = P^{(h)}. \tag{14}$$

As in [7], the crucial observation that will help us find a P as above is the following. Fix any distinct h, h' and consider $P^{(h)}|_{h'}$ and $P^{(h')}|_h$. Note that these polynomials are both naturally defined on the set of inputs $R_{h,h'} := R_h \cap R_{h'}$. However, since f_I is ε_1-close to $P^{(h)}$ and $P^{(h')}$ on R_h and $R_{h'}$ respectively, we see that

$$\Pr_{x \in R_{h,h'}}\left[P^{(h)}|_{h'}(x) \neq P^{(h')}|_h(x)\right] = \Pr_{x \in R_{h,h'}}\left[P^{(h)}(x) \neq P^{(h')}(x)\right]$$
$$\leq \Pr_{x \in R_{h,h'}}\left[P^{(h)}(x) \neq f(x)\right] + \Pr_{x \in R_{h,h'}}\left[P^{(h')}(x) \neq f(x)\right]$$
$$\leq 2\varepsilon_0 + 2\varepsilon_0 \leq 4\varepsilon_0 < \frac{1}{2^d},$$

where for the second inequality we have used the fact that

$$\Pr_{x \in R_{h,h'}}\left[P^{(h)}(x) \neq f(x)\right] \leq \frac{\Pr_{x \in R_h}\left[P^{(h)}(x) \neq f(x)\right]}{\Pr_{x \in R_h}\left[x \in R_{h'}\right]} \leq \frac{\varepsilon_0}{1/2} = 2\varepsilon_0.$$

Since any pair of distinct polynomials of degree d disagree on at least a $(1/2^d)$ fraction of inputs in $R_{h,h'}$, we see that $P^{(h)}|_{h'} = P^{(h')}|_h$ as polynomials. We record this fact below.

Claim 4.13. *For any distinct $h, h' \in [\ell+1]$, $P^{(h)}|_{h'} = P^{(h')}|_h$.*

The Matching case of Claim 4.12. Let $(X_{i_1}, X_{j_1}, 0), \ldots, (X_{i_{\ell+1}}, Y_{i_{\ell+1}}, 0)$ be the set of bad triples that give rise to the distinct edges of the matching in G_{bad}. By renaming variables we assume that $I = [r]$ and the bad triples are all of the form $(X_1, X_2, 0), \ldots, (X_{2\ell+1}, X_{2\ell+2}, 0)$.

Assume that for each $h \in [\ell + 1]$,

$$P^{(h)}(X) = \sum_{S \subseteq I \setminus \{2h\} : |S| \leq d} \alpha_S^{(h)} X^S$$

where $X^S = \prod_{i \in S} X_i$ (note that $P^{(h)} \in \mathcal{F}(I \setminus \{2h\}, d)$ and hence does not involve X_{2h}). For any h, if $|S| > d$ or $S \ni 2h$, we define $\alpha_S^{(h)} = 0$.

Note that we have for any distinct $i, j \in [\ell + 1]$

$$P^{(i)}(X)|_j = \sum_{S \cap \{2j-1, 2j\} = \emptyset} \alpha_S^{(i)} X^S + \sum_{S \cap \{2j-1, 2j\} = \emptyset} (\alpha^{(i)}_{S \cup \{2j-1\}} + \alpha^{(i)}_{S \cup \{2j\}} + \alpha^{(i)}_{S \cup \{2j-1, 2j\}}) X^{S \cup \{2j-1\}}. \quad (15)$$

In particular, Claim 4.13 implies the following for $S \subseteq I$ such that $|S| \leq d$ and i, j distinct such that $S \cap \{2i - 1, 2i, 2j - 1, 2j\} = \emptyset$,

$$\alpha_S^{(i)} = \alpha_S^{(j)} \quad (16)$$

$$\alpha^{(i)}_{S \cup \{2i-1\}} = \alpha^{(j)}_{S \cup \{2i-1\}} + \alpha^{(j)}_{S \cup \{2i\}} + \alpha^{(j)}_{S \cup \{2i-1, 2i\}} \quad (17)$$

Let $\alpha_S^{(i)}|_j$ denote the coefficient of X^S in $P^{(i)}|_j$.

We define the polynomial

$$P(X) = \sum_{S \subseteq I : |S| \leq d} \alpha_S X^S$$

as follows. For each $S \in \binom{I}{\leq d}$, set $\alpha_S = \alpha_S^{(j)}$ for any S such that $S \cap \{2j - 1, 2j\} = \emptyset$: since $|S| \leq d \leq \ell$, there is at least one such $j \in [\ell + 1]$. By (16), we see that any choice of j as above yields the same coefficient α_S.

Note that

$$P|_i = \sum_{S \cap \{2i-1, 2i\} = \emptyset} \alpha_S X^S + \sum_{S \cap \{2i-1, 2i\} = \emptyset} (\alpha_{S \cup \{2i-1\}} + \alpha_{S \cup \{2i\}} + \alpha_{S \cup \{2i-1, 2i\}}) X^{S \cup \{2i-1\}}. \quad (18)$$

Let $\alpha_S|_j$ denote the coefficient of X^S in $P|_j$.

Now we show that $P|_i = P^{(i)}$ for each choice of $i \in [\ell+1]$ by comparing coefficients of monomials and showing that $\alpha_S|_i = \alpha_S^{(i)}$ for each S such that $|S| \leq d$. That will conclude the proof of the matching case of Claim 4.12. Fix any S such that $|S| \leq d$. We consider three cases.

- $S \ni 2i$: In this case, $\alpha_S|_i = \alpha_S^{(i)} = 0$ and hence we are done.
- $S \cap \{2i - 1, 2i\} = \emptyset$: In this case, we have $\alpha_S|_i = \alpha_S^{(i)}$ by definition and hence we are done.

- $S \cap \{2i-1, 2i\} = 2i-1$: In this case, let $T = S \setminus \{2i-1\}$ and fix $j \in [\ell+1]$ such that $j \neq i$ and $T \cap \{2j-1, 2j\} = \emptyset$. We see that

$$\begin{aligned}
\alpha_S|_i &= \alpha_{T\cup\{2i-1\}} + \alpha_{T\cup\{2i\}} + \alpha_{T\cup\{2i-1,2i\}} && \text{(by (18))}\\
&= \alpha^{(j)}_{T\cup\{2i-1\}} + \alpha^{(j)}_{T\cup\{2i\}} + \alpha^{(j)}_{T\cup\{2i-1,2i\}} && \text{(by definition of } P)\\
&= \alpha^{(i)}_{T\cup\{2i-1\}} && \text{(by (17))}\\
&= \alpha^{(i)}_S.
\end{aligned}$$

The Star case of Claim 4.12. We proceed as in the matching case, except that the definition of P will be somewhat more involved. By renaming variables we assume that $I = [r]$ and that the bad triples are all of the form $(X_1, X_r, 0), (X_2, X_r, 0), \ldots, (X_{\ell+1}, X_r, 0)$.

Assume that for each $h \in [\ell+1]$,

$$P^{(h)}(X) = \sum_{S\subseteq[r-1]:|S|\leq d} \alpha^{(h)}_S X^S$$

where $X^S = \prod_{i\in S} X_i$ (note that $P^{(h)} \in \mathcal{F}(I \setminus \{r\}, d)$ and hence does not involve X_r). For any h, if $|S| > d$ or $S \ni r$, we define $\alpha^{(h)}_S = 0$.

For any distinct $i, j \in [\ell+1]$ with $i < j$, we assume that $P^{(i)}|_j$ and $P^{(j)}|_i$ are obtained by replacing X_j with X_i. We thus have

$$P^{(i)}(X)|_j = \sum_{S\cap\{i,j\}=\emptyset} \alpha^{(i)}_S X^S + \sum_{S\cap\{i,j\}=\emptyset} (\alpha^{(i)}_{S\cup\{i\}} + \alpha^{(i)}_{S\cup\{j\}} + \alpha^{(i)}_{S\cup\{i,j\}}) X^{S\cup\{i\}}. \tag{19}$$

$$P^{(j)}(X)|_i = \sum_{S\cap\{i,j\}=\emptyset} \alpha^{(j)}_S X^S + \sum_{S\cap\{i,j\}=\emptyset} (\alpha^{(j)}_{S\cup\{i\}} + \alpha^{(j)}_{S\cup\{j\}} + \alpha^{(j)}_{S\cup\{i,j\}}) X^{S\cup\{i\}}. \tag{20}$$

Using Claim 4.13 and comparing coefficients of $P^{(i)}|_j$ and $P^{(j)}|_i$, we get for $i \neq j$ and S such that $S \cap \{i, j\} = \emptyset$,

$$\alpha^{(i)}_S = \alpha^{(j)}_S \tag{21}$$

$$\alpha^{(i)}_{S\cup\{i\}} + \alpha^{(i)}_{S\cup\{j\}} + \alpha^{(i)}_{S\cup\{i,j\}} = \alpha^{(j)}_{S\cup\{i\}} + \alpha^{(j)}_{S\cup\{j\}} + \alpha^{(j)}_{S\cup\{i,j\}} \tag{22}$$

We now define the polynomial

$$P(X) = \sum_{S\subseteq I\setminus\{r\}:|S|\leq d} \beta_S X^S + \sum_{S\subseteq I: S\ni r, |S|\leq d} \gamma_S X^S$$

as follows.

- For $S \not\ni r$, we define β_S to be $\alpha^{(i)}_S$ for any $i \in [\ell+1]$ such that $i \notin S$. Since $|S| \leq d < \ell+1$ there is such an i. Note that by (21), the choice of i is immaterial.
- For $S \ni r$, we let $T = S \setminus \{r\}$. Note that $|T| < d$. We define $\gamma_{T\cup\{r\}}$ by downward induction on $|T|$ as follows:

  $$\gamma_{T\cup\{r\}} \triangleq \alpha^{(i)}_{T\cup\{i\}} - \beta_{T\cup\{i\}} - \gamma_{T\cup\{i,r\}} \text{ for any fixed } i \in [\ell+1] \setminus T$$

  where we assume that $\gamma_{T\cup\{r\}} = 0$ for $|T| \geq d$.

We will show first by downward induction on $|T|$ that these coefficients are independent of the choice of $i \in [\ell+1] \setminus T$. Fix $i, j \in [\ell+1] \setminus T$. In the base case $|T| = d-1$, we have

$$\begin{aligned}
\alpha^{(i)}_{T\cup\{i\}} - \beta_{T\cup\{i\}} &= \alpha^{(i)}_{T\cup\{i\}} - \alpha^{(j)}_{T\cup\{i\}} && \text{(by definition of } \beta_{T\cup\{i\}}) \\
&= \alpha^{(j)}_{T\cup\{j\}} - \alpha^{(i)}_{T\cup\{j\}} && \text{(by (21) and } \alpha_{T\cup\{i,j\}} = 0) \\
&= \alpha^{(j)}_{T\cup\{j\}} - \beta_{T\cup\{j\}} && \text{(by definition of } \beta_{T\cup\{j\}})
\end{aligned}$$

When $|T| < d-1$, we have

$$\begin{aligned}
&\alpha^{(i)}_{T\cup\{i\}} - \beta_{T\cup\{i\}} - \gamma_{T\cup\{i,r\}} \\
&= \alpha^{(i)}_{T\cup\{i\}} - \alpha^{(j)}_{T\cup\{i\}} - \gamma_{T\cup\{i,r\}} \\
&= \alpha^{(i)}_{T\cup\{i\}} - \alpha^{(j)}_{T\cup\{i\}} - \left(\alpha^{(j)}_{T\cup\{i,j\}} - \beta_{T\cup\{i,j\}} - \gamma_{T\cup\{i,j,r\}}\right) && \text{(Applying definition of } \gamma_{T\cup\{i,r\}} \text{ and induction)} \\
&= \alpha^{(j)}_{T\cup\{j\}} - \alpha^{(i)}_{T\cup\{j\}} - \left(\alpha^{(i)}_{T\cup\{i,j\}} - \beta_{T\cup\{i,j\}} - \gamma_{T\cup\{i,j,r\}}\right) && \text{(From (22) and rearranging terms.)} \\
&= \alpha^{(j)}_{T\cup\{j\}} - \alpha^{(i)}_{T\cup\{j\}} - \gamma_{T\cup\{j,r\}} && \text{(Applying definition of } \gamma_{T\cup\{j,r\}} \text{ and induction)} \\
&= \alpha^{(j)}_{T\cup\{j\}} - \beta_{T\cup\{j\}} - \gamma_{T\cup\{j,r\}}
\end{aligned}$$

We now conclude by showing that the restriction of P obtained by replacing X_r by X_i equals the polynomial $P^{(i)}$. Let $P|_i$ denote the restriction of P obtained by replacing X_r by X_i and let $\alpha_S|_i$ be its coefficients. Note that

$$\begin{aligned}
P|_i &= \sum_{S \not\ni i : |S| \le d} \beta_S X^S + \sum_{T \not\ni i : |T| < d} (\beta_{T\cup\{i\}} + \gamma_{T\cup\{r\}} + \gamma_{T\cup\{i,r\}}) X^{T\cup\{i\}}. \\
&= \sum_{S \not\ni i : |S| \le d} \alpha^{(i)}_S X^S + \sum_{T \not\ni i : |T| < d} \alpha^{(i)}_{T\cup\{i\}} X^{T\cup\{i\}} && \text{(By definition of } \beta_S \text{ and } \gamma_{T\cup\{r\}}) \\
&= P^{(i)}.
\end{aligned}$$

This concludes the proof for the star case.

5 Impossibility of local decoding when $\mathrm{char}(\mathbb{F})$ is large

In this section, we prove Theorem 5.3 which is a more detailed version of Theorem 3.2. Again we remind the reader that an overview may be found in Section 1.3.

Let n be a growing parameter and $\mathbb{F}$ a field of characteristic 0 or positive characteristic greater than n^2. For the results in this section, it will be easier to deal with the domain $\{-1,1\}^n$ rather than $\{0,1\}^n$. Since there a natural invertible linear map that maps $\{0,1\}$ to $\{-1,1\}$ (i.e. $a \mapsto 1-2a$), this change of input space is without loss of generality.

5.1 Local linear spans of balanced vectors

Let $u \in \mathbb{F}^n$ and $U \subseteq \mathbb{F}^n$. For any integer $t \in \mathbb{N}$, we say that u is in the t-span of U if it can be written as a linear combination of at most t elements of U. For $x \in \{-1,1\}^n$, we use $|x|$ to denote

the sum of the entries of x over $\mathbb{Z}$. In this section, we wish to show that if the vector 1^n is in the t-span of balanced vectors, i.e., vectors x with $|x| \leq n/s$ then t is must be growing as a function of s.

As explained earlier we first establish a bound on the size of the solutions of linear equations in systems over $\mathbb{Q}$ with few variables or few constraints. This fact is well-known, but we prove it here for completeness.

Lemma 5.1. *Let $r, s \in \mathbb{N}$ and let $t = \min\{r, s\}$. Let $Mx = u$ be a system of linear equations with $M \in \{-1, 0, 1\}^{r \times s}$ and $u \in \{-1, 0, 1\}^r$.*

- *If $\mathbb{F}$ is a field of characteristic and the system has a solution in $\mathbb{F}^s$, then there exist integers $a_1, \ldots, a_s, b \in \mathbb{Z}$ with $|a_i|, |b| \leq t!$ such that $x_i = a_i/b$ is a solution to $Mx = u$. In particular, there is a solution in $\mathbb{Q}^s$.*
- *If $\mathbb{F}$ is a field of characteristic p and if the system has a solution in $\mathbb{F}^s$, then there exist integers $a_1, \ldots, a_s, b \in \mathbb{Z}$ with $|a_i|, |b| \leq t!$ such that $x_i = a_i/b \pmod p$ is a solution to $Mx = u$. In particular, there is a solution in $\mathbb{F}_p^s$.*

Proof. Note that we can assume that M has full column rank. This is because $Mx = u$ has a solution iff $\tilde{M}\tilde{x} = u$ has a solution where $\tilde{M}$ is the submatrix of M obtained by a keeping a maximal set of linearly independent columns of M. When the columns are linearly independent, we have s is at most r and hence $t = \min\{r, s\} = s$.

We start with the zero characteristic case. Let M' be an invertible $s \times s$ submatrix of M containing the set of s linearly independent rows of M and let $u' \in \mathbb{F}^s$ be the vector corresponding to these rows. Note that the solution x is uniquely determined by $M'x = u'$. We now apply Cramer's rule to see that the solution is given by

$$x_i = \frac{\det(M_i')}{\det(M')}$$

for $i \in [s]$, where M_i' is the $s \times s$ matrix obtained by replacing the ith column of M' bu u'. Since M' and M_i' are matrices with entries in $\{-1, 0, 1\}$, we have $\det(M') \in \mathbb{Z}$ with $|\det(M')| \leq s!$ for each $i \in [s]$ and similarly for $\det(M_i')$. Therefore, we have the claim with $a_i = \det(M_i')$ and $b = \det(M')$.

The characteristic p case is similar with only differnece being the solution now is given by $x_i = a_i/b \pmod p$. □

We now turn to the main technical lemma of this section showing that 1^n is not in linear span of a small number of nearly balanced elements of $\{-1, 1\}^n$.

Lemma 5.2. *Let $n, s = s(n) \in \mathbb{N}$ with $s(n) \leq n$. Let $S = \{x \in \{-1, 1\}^n \mid |x| \leq n/s)\}$. Then $x^0 = 1^n$ is not in the t-span of S unless $t \geq \log s / \log\log s$ provided $\mathbb{F}$ is field of zero characteristic or of characteristic $p \geq 2n^2$.*

Proof. We first consider the case when $\mathbb{F}$ is of zero characteristic. Note that in this case $\mathbb{Q} \subseteq \mathbb{F}$. Suppose $x^0 \in \text{Span}\{x^1, \ldots, x^t\}$ with $x^0 = \sum_{i=1}^n c_i x^i$. Note that the c_i's are expressible as the solution to a linear system whose $Mz = u$ where M and u have entries in $\{-1, 0, 1\}$ and M is a $n \times t$ matrix. By Lemma 5.1 we have that $c_i \in \mathbb{Q}$ with $|c_i| \leq t!$ (more specifically we have $c_i = a_i/b$ with $|a_i| \leq t!$ and this implies $|c_i| \leq t!$). We thus have

$$n = \left|\sum_{j=1}^n x_j^0\right| = \left|\sum_{i=1}^t c_i \sum_{j=1}^n x_j^i\right| \leq \sum_{i=1}^t |c_i| \cdot \left|\sum_{j=1}^n x_j^i\right| \leq \sum_{i=1}^t (t!) \cdot (n/s) \leq (t+1)! \cdot (n/s).$$

We thus conclude that $(t+1)! \geq s$ and thus $t \geq \log s / \log\log s$.

In the case of finite field $\mathbb{F}$, we proceed as above and let $x^0 = \sum_{i=1}^{t} c_i x^i$. By Lemma 5.1 we have that there are integers a_i, b with $|a_i|, |b| \leq t!$ such that $c_i = a_i/b \pmod p$ is a solution to $x^0 = \sum_{i=1}^{t} c_i x^i$. Now consider $b \cdot n$ and we get $b \cdot n = \sum_{i=1}^{t} a_i \sum_{j=1}^{n} x_j^i \pmod p$. We now show that this implies $(t+1)! \geq \min\{p/(2n), s\} = s$ (where the equality follows from $p \geq 2n^2$ and $s \leq n$). Assume $(t+1)! \leq p/(2n)$. Then we have $n \leq |b \cdot n| \leq t! \cdot n < p/2$ over the integers, and $\left|\sum_{i=1}^{t} a_i \sum_{j=1}^{n} x_j^i\right| \leq (t+1)!(n/s) < p/2$ also over the integers. We again conclude that $n \leq (t+1)!(n/s)$ and so $(t+1)! \geq s$ as claimed. The lemma follows. □

5.2 Proof of Theorem 3.2

We now state and prove Theorem 5.3 which immediately implies Theorem 3.2.

Theorem 5.3. *Let $n \in \mathbb{N}$ be a growing parameter and $\varepsilon \in (0,1)$ such that $\varepsilon \geq 2\exp(-n/2s^2)$ for some $s \in \mathbb{N}$ with $100 \leq s \leq \sqrt{n}/100$. Let $\mathbb{F}$ be any field such that either $\mathrm{char}(\mathbb{F}) = 0$ or $\mathrm{char}(\mathbb{F}) \geq n^2$. Then any adaptive (ε, q)-local decoder for $\mathcal{F}(n,1)$ that corrects an ε fraction of errors must satisfy $q = \Omega(\log s / \log\log s)$.*

Proof. The proof of the theorem will use the minimax principle. Specifically, we design a "hard" probability distribution $\mathcal{D}$ over functions that are ε-close to $\mathcal{F}(n,1)$ such that any *deterministic* decoder that decodes the value of a random function (chosen according to $\mathcal{D}$) at the point 1^n while making very few queries will fail to decode the value with probability at least $1/4$.

We start with the case of positive characteristic which is somewhat simpler to describe. Let $\mathrm{char}(\mathbb{F}) = p > n^2$. We define the hard distribution $\mathcal{D}$ as follows. Let

$$E = \{x \in \{-1,1\}^n \mid |\sum_i x_i| \geq 2n/s\},$$

so that, by the Chernoff bound, (see, e.g. [12]) $|E| \leq \varepsilon 2^n$. Let $S = \{-1,1\}^n \setminus E$.

We now sample a random function $f \sim \mathcal{D}$ as follows:

- Choose $a_1, \ldots, a_n \in \mathbb{F}_p \subseteq \mathbb{F}$ uniformly at random independently. Let $\ell(X_1, \ldots, X_n) = \sum_i a_i X_i \in \mathcal{F}(n,1)$.
- Let $f(x) = 0$ if $x \in E$ and $f(x) = \ell(x)$ if $x \in S$.

Since $f(x) = \ell(x)$ for $x \notin E$, we have $\delta(f, \ell) \leq \varepsilon$. In particular $\delta_1(f) \leq \varepsilon$.

Let $\mathcal{A}$ be any deterministic decoding algorithm for decoding $f(1^n)$. Assume that the worst case number of queries t made by $\mathcal{A}$ satisfies $t < \log s / \log\log s$. W.l.o.g. we assume that $\mathcal{A}$ always makes exactly t queries and also that none of these queries are made to inputs $x \in E$ (since at these points $f(x)$ is known to take the value 0). Additionally, we may assume that these queries correspond to linearly independent inputs since if a query point x is a linear combination of previous queries, then $\ell(x) = \langle a, x \rangle$ can be determined from the answers to previous queries.

Let $x^1, \ldots, x^t$ be the (adaptive) queries made by $\mathcal{A}$ on the random function f. After these queries are made, the algorithm has $\ell(x^i) = \langle a, x^i \rangle$ for each $i \in [t]$, where $a = (a_1, \ldots, a_n)$. However, by Lemma 5.2, we know that 1^n is not in the t-span of the inputs in S and hence, given the values $\ell(x^1), \ldots, \ell(x^t)$, $\ell(1^n) = \sum_i a_i$ is still distributed uniformly over $\mathbb{F}_p$. Hence, the probability that the algorithm outputs $\ell(1^n)$ correctly is at most $1/p < 3/4$.

Now consider the case when $\mathrm{char}(\mathbb{F}) = 0$. We define our hard distribution $\mathcal{D}$ exactly as above except that the coefficients $a_1, \ldots, a_n$ are chosen i.u.a.r. from $\{-N, \ldots, N\}$ where $N = n^{\lceil \log s / \log\log s \rceil}$.

Let $\mathcal{A}$ be any deterministic decoding algorithm for decoding $f(1^n)$ as above. Again, we assume that $\mathcal{A}$ always makes $t \leq \log s / \log\log s$ many queries corresponding to linearly independent inputs, and also that none of these queries are made to inputs $x \in E$.

Let $A \subseteq \{-N, \ldots, N\}^n$ be the set of coefficients of linear polynomials ℓ such that $\mathcal{A}$ is able to decode $\ell(1^n) = \sum_i a_i$ correctly.

To bound the size of $|A|$, we use an encoding argument. Considre any $(a_1, \ldots, a_n) \in A$ and let $\ell(X) = \sum_i a_i X_i$. Let $x^1, \ldots, x^t$ be the queries made on input ℓ. Given $\langle a, x^i \rangle$ for $i \in [t]$, the algorithm determines $\sum_i a_i = \langle a, 1^n \rangle$. Hence, at this point the algorithm has $\langle a, x \rangle$ for $x \in I' = \{x^1, \ldots, x^t, 1^n\}$. Note that I' is a set of dimension $t+1$ since by Lemma 5.2, 1^n is not in the t-span of S. We can thus a subset $I'' = \{e^{i_1}, \ldots, e^{i_{n-t-1}}\}$ of the set of standard basis vectors $\{e^1, \ldots, e^n\}$ of size $n - t - 1$ so that $I = I' \cup I''$ is a basis for $\mathbb{F}^n$.

Define an encoding function

$$\mathcal{E} : A \to \{-Nn, \ldots, Nn\}^t \times \{-N, \ldots, N\}^{n-t-1}$$

as follows. For each $x \in A$, we choose I as above and set

$$\mathcal{E}(a) = (\langle a, x^1 \rangle, \ldots, \langle a, x^t \rangle, \langle a, e^1 \rangle, \ldots, \langle a, e^{i_{n-t-1}} \rangle).$$

Note that each $\langle a, x^j \rangle \in \{-Nn, \ldots, Nn\}$ since $a \in \{-N, \ldots, N\}^n$ and $x^j \in \{-1, 1\}^n$.

We claim that $\mathcal{E}$ is 1-1. This is because, on being given $\mathcal{E}(a)$ as above, we can determine $\langle a, x \rangle$ for each $x \in I$ by the following argument: $\mathcal{E}(a)$ directly gives us $\langle a, x \rangle$ for each $x \in I \setminus \{1^n\}$ and by construction of $x^1, \ldots, x^t$, we know that $\langle a, x^1 \rangle, \ldots, \langle a, x^t \rangle$ determines the value of $\langle a, 1^n \rangle$. Thus, we have $\langle a, x \rangle$ for each $x \in I$ and as I is a basis for $\mathbb{F}^n$, we can obtain $a \in \mathbb{F}^n$ as well.

Since $\mathcal{E}$ is 1-1, we see that

$$|A| \leq (2Nn+1)^t \cdot (2N+1)^{n-t-1} \leq (2N+1)^{n-1} \cdot n^t \leq (2N+1)^{n-1} \cdot N \leq (2N+1)^n \cdot \frac{3}{4}.$$

which implies that the relative size of A inside $\{-N, \ldots, N\}^n$ is at most $3/4$. This concludes the proof. □

6 Local decoding when $\mathrm{char}(\mathbb{F})$ is small

In this section, we give a local decoder over fields of small characteristic. An overview of this construction may be found in Section 1.3.

Let p be a prime of constant size and let $\mathbb{F}$ be any (possibly infinite) field of characteristic p. Let d be the degree parameter and k be the smallest power of p that is strictly greater than d. Note that $k \leq pd$. We show that the space $\mathcal{F}(n, d)$ has a $(1/(4 \cdot \binom{2k}{k}), \binom{2k}{k})$-local decoder, hence proving Theorem 3.3.

The main technical tool we use is a suitable linear relation on the space $\mathcal{F}(2k, d)$, which we describe now. We say that a set $S \subseteq \{0,1\}^{2k}$ is *useful* if for every polynomial $G \in \mathcal{F}(2k, d)$, $G(0^{2k})$ is determined by the restriction of the function G to the inputs in S. Let $B \subseteq \{0,1\}^{2k}$ denote the set of all balanced inputs (i.e. inputs of Hamming weight exactly k).

Lemma 6.1. *Fix d, k as above. Then the set $B \subseteq \{0,1\}^{2k}$ of balanced inputs is useful.*

The proof of the above lemma will use Lucas' theorem, which we recall below.

Theorem 6.2 (Lucas' theorem)**.** *Let p be any prime and $a, b \in \mathbb{N}$. Let $a_1, \ldots, a_\ell \in \{0, \ldots, p-1\}$ and $b_1, \ldots, b_\ell \in \{0, \ldots, p-1\}$ be the digits in the p-ary expansion of a and b, i.e., $a = \sum_{j \in [\ell]} a_j p^{j-1}$ and $b = \sum_{j \in [\ell]} b_j p^{j-1}$. Then, we have*

$$\binom{a}{b} \equiv \prod_{i \leq \ell} \binom{a_i}{b_i} \pmod{p}$$

where $\binom{a_i}{b_i}$ is defined to be 0 if $a_i < b_i$.

Corollary 6.3. *For $i \in \{0, \ldots, d\}$, we have $\binom{d+k-i}{k-i} \not\equiv 0 \pmod{p}$ if and only if $i = 0$.*

Proof. Note that by Lucas' theorem (Theorem 6.2), $\binom{a}{b} \equiv 0 \pmod{p}$ if and only if there are digits a_j, b_j in the p-ary expansions of a and b respectively with $a_j < b_j$.

Consider first the case when $i = 0$. Let $a = d + k$ and $b = k$. Let

$$a = \sum_{j=1}^{\ell} a_j \cdot p^{j-1} \qquad b = \sum_{j=1}^{\ell} b_j \cdot p^{j-1} \tag{23}$$

where $a_j, b_j \in \{0, \ldots, p-1\}$ and $k = p^{\ell-1}$. Then, we have $b_j = 0$ for each $j < \ell$ and $b_\ell = a_\ell = 1$. Hence by Lucas' theorem, we have $\binom{a}{b} \neq 0 \pmod{p}$.

Now consider the case when $i \in [d]$. Let $a = d + k - i$ and $b = k - i$. Again write a, b as in (23) with $k = p^{\ell-1}$. In this case, we have $a_\ell = 1$ but $b_\ell = 0$, the latter due to the fact that $b < k$. Hence if we consider $a' = \sum_{j \in [\ell-1]} a_j p^{j-1}$ and $b' = \sum_{j \in [\ell-1]} b_j p^{j-1}$, we get $a' = d - i < b' = k - i$. Therefore, there must exist $j \in [\ell - 1]$ such that $a_j < b_j$. From Lucas' theorem, it now follows that $\binom{a}{b} \equiv 0 \pmod{p}$. □

Proof of Lemma 6.1. Fix any $G \in \mathcal{F}(2k, d)$. Assume that

$$G(Y_1, \ldots, Y_{2k}) = \sum_{I \subseteq [2k] : |I| \leq d} \alpha_I Y^I$$

where Y^I denotes $\prod_{i \in I} Y_i$.

Let B' denote all those inputs in B where the last $k - d$ bits are set to 0. We will compute the sum of G on inputs from B'. But let us first consider a monomial Y^I and see what its sum over $y \in B'$ looks like. The monomial evaluates to 1 on $y \in B'$ if $y_i = 1$ for every $i \in I$, and evaluates to 0 otherwise. There are exactly $\binom{d+k-|I|}{k-|I|}$ choices of $y \in B'$ that satisfy $y_i = 1$ for every $i \in I$. Thus summing over $y \in B'$ we get $\sum_{y \in B'} y^I = \binom{d+k-|I|}{k-|I|}$. Summing over all monomials we get:

$$\begin{aligned} \sum_{y \in B'} G(y) &= \sum_{I \subseteq [2k] : |I| \leq d} \alpha_I \cdot \sum_{y \in B'} Y^I \\ &= \sum_{I \subseteq [2k] : |I| \leq d} \alpha_I \cdot \binom{d + k - |I|}{k - |I|} \end{aligned} \tag{24}$$

By Corollary 6.3, it follows that for $i \in \{0, \ldots, d\}$, we have

$$\binom{d+k-i}{k-i} \not\equiv 0 \pmod{p}$$

if and only if $i = 0$ and so $\sum_{y \in B'} G(y) = \binom{d+k}{k} \cdot \alpha_\emptyset$. Let $c = \binom{d+k}{k} \pmod{p}$. We have $c \in \mathbb{F}_p^* \subseteq \mathbb{F}^*$ and in particular c is invertible in $\mathbb{F}$, and $\sum_{y \in B'} G(y) = c \cdot \alpha_\emptyset = c \cdot G(0^{2k})$. Hence, we get $G(0^{2k}) = c^{-1} \cdot \sum_{y \in B'} G(y)$. Therefore, $G(0^{2k})$ is determined by the restriction of G to B' and hence also by its restriction to B. □

We now show that $\mathcal{F}(n, d)$ has a $(1/(4 \cdot \binom{2k}{k})), \binom{2k}{k})$-local decoder.

The decoder. We now give the formal description of the decoder. Let the decoder be given oracle access to f with the promise that f is $1/(4 \cdot \binom{2k}{k}))$-close to some $F \in \mathcal{F}(n, d)$. Let the input to the decoder be $x \in \{0,1\}^n$. The problem is to find $F(x)$.

We describe the decoder below:

Decoder $D_k^f(x)$.

- Partition $[n]$ into $2k$ parts by choosing a *uniformly* random map $h : [n] \to [2k]$. I.e. each $h(j)$ is chosen i.u.a.r. from $[2k]$.
- For $i \in [2k]$ and $j \in [n]$ such that $h(j) = i$, identify X_j with $Y_i \oplus x_j$.
- Let $g(Y_1, \ldots, Y_{2k})$ and $G(Y_1, \ldots, Y_{2k})$ be the restrictions of f and F respectively. Assuming $g|_B = G|_B$, query g at all inputs in B and decode $G(0^{2k})$ from $G|_B$. Output the value decoded.

The main theorem of this section is the following. Note that this implies Theorem 3.3.

Theorem 6.4. *Let $\mathbb{F}$ be a field of characteristic p. For integer $d \geq 0$, let k be the smallest power of p greater than d. Then the decoder D_k is a $(1/(4 \cdot \binom{2k}{k})), \binom{2k}{k})$-local decoder for $\mathcal{F}(n, d; \mathbb{F})$.*

Proof. The bound on the query complexity of the decoder is clear from the description of D_k. So we only need to argue that the decoder outputs the value of $F(x)$ correctly with probability at least $3/4$.

The crucial observation is that for each fixed $y \in B$, querying $g(y)$ amounts to querying f at a *uniformly* random point $z \in \{0,1\}^n$, where the randomness comes from the choice of h. This is because for each $j \in [n]$, we have

$$z_j = y_{h(i)} \oplus x_j$$

where $h : [n] \to [2k]$ is a uniformly random function. Since y is *balanced*, each $y_{h(i)}$ is a uniformly random bit. Hence we see that $z \in \{0,1\}^n$ is distributed uniformly over $\{0,1\}^n$.

Thus, if $\delta(f, F) \leq 1/(4 \cdot \binom{2k}{k}))$, with probability at least $3/4$, all the random queries made lie outside the error set $E = \{z \in \{0,1\}^n \mid f(z) \neq F(z)\}$ and in this case, the decoder is able to access the function $G|_B$ at each input $y \in B$. By Lemma 6.1, this allows the decoder to determine $G(0^{2k})$. Noting that the image of 0^{2k} in $\{0,1\}^n$ is exactly x, we thus see that the decoder outputs $F(x)$ correctly. □

References

[1] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Trans. Information Theory*, 51(11):4032–4039, 2005. 1, 2

[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. 1

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

[4] Donald Beaver and Joan Feigenbaum. Hiding instances in multioracle queries. In C. Choffrut and T. Lengauer, editors, *Proceedings of the 7th Annual Symposium on Theoretical Aspects of Computer Science*, pages 37–48, Rouen, France, 22–24 February 1990. Springer. Lecture Notes in Computer Science, Volume 415. 1

[5] Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract). In Janos Simon, editor, *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 1–10. ACM, 1988. 1

[6] Eli Ben-Sasson, Prahladh Harsha, and Sofya Raskhodnikova. Some 3CNF properties are hard to test. *SIAM J. Comput.*, 35(1):1–21, 2005. 3

[7] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of Reed-Muller codes. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 488–497. IEEE Computer Society, 2010. 4, 17, 19

[8] Manuel Blum, Ashok K. Chandra, and Mark N. Wegman. Equivalence of free Boolean graphs can be decided probabilistically in polynomial time. *Inf. Process. Lett.*, 10(2):80–82, 1980. 1

[9] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993. 1

[10] Richard A. DeMillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Inf. Process. Lett.*, 7(4):193–195, 1978. 1

[11] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007. 2

[12] Devdatt Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, New York, NY, USA, 1st edition, 2009. 24

[13] 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. 2

[14] Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM J. Comput.*, 36(3):779–802, 2006. 1, 2

[15] 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. 4

[16] John Y. Kim and Swastik Kopparty. Decoding Reed-Muller codes over product sets. In Ran Raz, editor, *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, volume 50 of *LIPIcs*, pages 11:1–11:28. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016. 1

[17] Richard Lipton. New directions in testing. In *Distributed Computing and Cryptography*, volume 2 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 191–202. AMS, 1991. 1

[18] Carsten Lund, Lance Fortnow, Howard J. Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. *J. ACM*, 39(4):859–868, 1992. 1

[19] D. E. Muller. Application of Boolean algebra to switching circuit design and to error detection. *IEEE Transactions on Computers*, 3:6–12, 1954. 1

[20] Ketan Mulmuley, Umesh V. Vazirani, and Vijay V. Vazirani. Matching is as easy as matrix inversion. *Combinatorica*, 7(1):105–113, 1987. 1

[21] Yury Polyanskiy. Hypercontractivity of spherical averages in Hamming space. *CoRR*, abs/1309.3014, 2013. 4, 5, 7

[22] Alexander A. Razborov. On the method of approximations. In David S. Johnson, editor, *Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washigton, USA*, pages 167–176. ACM, 1989. 1

[23] Irving S. Reed. A class of multiple-error-correcting codes and the decoding scheme. *IEEE Transactions on Information Theory*, 4:38–49, 1954. 1

[24] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, April 1996. 1

[25] Jacob T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, 1980. 1

[26] Adi Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979. 1

[27] Adi Shamir. IP=PSPACE. In *31st Annual Symposium on Foundations of Computer Science, St. Louis, Missouri, USA, October 22-24, 1990, Volume I*, pages 11–15. IEEE Computer Society, 1990. 1

[28] Richard Zippel. Probabilistic algorithms for sparse polynomials. In Edward W. Ng, editor, *Symbolic and Algebraic Computation, EUROSAM '79, An International Symposiumon Symbolic and Algebraic Computation, Marseille, France, June 1979, Proceedings*, volume 72 of *Lecture Notes in Computer Science*, pages 216–226. Springer, 1979. 1

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il