Electronic Colloquium on Computational Complexity, Report No. 110 (2017)

On Axis-Parallel Tests for Tensor Product Codes

Alessandro Chiesa
alexch@berkeley.edu
UC Berkeley

Peter Manohar
manohar@berkeley.edu
UC Berkeley

Igor Shinkar
igors@berkeley.edu
UC Berkeley

Abstract

Many low-degree tests examine the input function via its restrictions to random hyperplanes of a certain dimension. Examples include the line-vs-line (Arora, Sudan 2003), plane-vs-plane (Raz, Safra 1997), and cube-vs-cube (Bhangale, Dinur, Livni 2017) tests.

In this paper we study tests that only consider restrictions along *axis-parallel* hyperplanes, which have been studied by Polishchuk and Spielman (1994) and Ben-Sasson and Sudan (2006). While such tests are necessarily "weaker", they work for a more general class of codes, namely tensor product codes. Moreover, axis-parallel tests play a key role in constructing LTCs with inverse polylogarithmic rate and short PCPs (Polishchuk, Spielman 1994; Ben-Sasson, Sudan 2008; Meir 2010). We present two results on axis-parallel tests.

(1) Bivariate low-degree testing with low-agreement. We prove an analogue of the Bivariate Low-Degree Testing Theorem of Polishchuk and Spielman in the low-agreement regime, albeit with much larger field size. Namely, for the 2-wise tensor product of the Reed–Solomon code, we prove that for sufficiently large fields, the 2-query variant of the axis-parallel line test (row-vs-column test) works for *arbitrarily small agreement.* Prior analyses of axis-parallel tests assumed high agreement, and no results for such tests in the low-agreement regime were known.

Our proof technique deviates significantly from that of Polishchuk and Spielman, which relies on algebraic methods such as Bézout's Theorem, and instead leverages a fundamental result in extremal graph theory by Kövári, Sós, and Turán. To our knowledge, this is the first time this result is used in the context of low-degree testing.

(2) Improved robustness for tensor product codes. Robustness is a strengthening of local testability that underlies many applications. We prove that the axis-parallel hyperplane test for the m-wise tensor product of a linear code with block length n and distance d is $\Omega(\frac{d^m}{n^m})$-robust. This improves on a theorem of Viderman (2012) by a factor of $1/\operatorname{poly}(m)$. While the improvement is not large, we believe that our proof is a notable simplification compared to prior work.

Keywords: tensor product codes; locally testable codes; low-degree testing; extremal graph theory

ISSN 1433-8092

Contents

1 Introduction

Locally testable codes (LTCs) are error-correcting codes for which, given an input word, one can verify whether the word belongs to or is far from the code by inspecting the word in a few random locations. LTCs have been studied extensively in different contexts, including program checking, interactive proofs, and probabilistically checkable proofs (PCPs) [BLR93; RS96; BFL91; BFLS91; PS94; GS06]. Goldreich and Sudan [GS06] describe LTCs as "combinatorial counterparts of the complexity theoretic notion of PCPs", motivating the study of these objects separately.

LTC constructions. The first constructions of LTCs were algebraic in nature, and relied on multivariate polynomials. Starting with the seminal work of Blum, Luby, and Rubinfeld [BLR93], there has been much work on such algebraic LTCs by way of results on *linearity testing* and *low-degree testing* in numerous settings [BLR93; BCLR08; BSVW03; BCHKS96; AKKLR05; BKSSZ10]. Many other constructions [Mei09; Vid13; KMRS16] further optimize parameters of these codes, including rate, distance, and the number of queries made by the tester.

Ben-Sasson and Sudan [BS06] suggested a *combinatorial* approach to construct LTCs starting from any linear code by (i) applying the *tensor product* operation [Wol65; WE63] to the code, and (ii) testing the resulting code via the *axis-parallel hyperplane test*. We now discuss both.

The 2-wise tensor product of a linear code $C \subseteq \mathbb{F}^n$, denoted C^2, is the code in $\mathbb{F}^{n^2}$ consisting of all 2-dimensional matrices whose n rows and n columns are codewords in C; similarly, the m-wise tensor product of C, denoted C^m, is the code in $\mathbb{F}^{n^m}$ consisting of all m-dimensional matrices M whose restrictions to any axis-parallel $(m-1)$-dimensional hyperplane is a codeword in C^{m-1}. For example, the code of evaluations of all m-variate polynomials of individual degree at most r is the m-wise tensor product of the code of evaluations of all univariate polynomials of degree at most r.

The axis-parallel hyperplane test for the code C^m works as follows: given a word M, sample a random axis-parallel hyperplane and check if the restriction of M to this hyperplane is a codeword in C^{m-1}. This natural test extends ideas of axis-parallel line tests used in early PCP constructions [BFL91; BFLS91; AS98] to arbitrary tensor product codes.

We study two aspects of the axis-parallel hyperplane test for tensor product codes.

(1) Low-agreement regime. All of the aforementioned works study the axis-parallel hyperplane test in the "high-agreement regime", in which the given codeword is within the unique decoding radius of the tensor product code. What can be said about the "low-agreement regime", in which the given codeword may be as far as the list-decoding radius? This setting is more challenging because one wishes to deduce that a given word has some noticeable global correlation with a codeword, or a short list of codewords, by only assuming that local views of the test have some non-trivial agreement with accepting views (but may not necessarily be very close to such views).

Results in the low-agreement regime are known for *other* tests, such as tests for the Hadamard code [BCHKS96] and the long code [Hås01] as well as random *non*-axis-parallel hyperplane tests in various dimensions [RS97; AS03; MR08]. Moreover, these have applications to PCP constructions and hardness of approximation. However, to our knowledge *prior to our work no results are known for the low-agreement regime of axis-parallel tests.*

(2) Robustness. Ben-Sasson and Sudan [BS06] analyze the axis-parallel hyperplane test via the notion of *robustness*, a stronger notion of local testability borrowed from the PCP literature [BGHSV06; DR04]. Informally, a test for a code is robust if, given any input that is far from the code, the local view of the test is also far from an accepting view on average. For example, the axis-parallel hyperplane test is robust if, given any M that is far from C^m, the restriction of M to a random hyperplane is far from C^{m-1} on average.

Robustness thus relates the global distance to the expected local view distance and, as shown in [BS06], facilitates query reduction via a natural way to compose tests; this notion has also found applications to proof composition in the setting of PCPs [BGHSV06]. These works have motivated the study of the robustness of the axis-parallel hyperplane test for tensor product codes, establishing both positive results [DSW06; BV08; BV15; Vid12] and limitations [Val05; CR05; GM12].

Despite significant progress, robustness results for the axis-parallel hyperplane test seem to be *far from tight*. The best known relation between the global distance and the local distance is due to Viderman [Vid12], but no examples that come anywhere close to his proven bound are known.

2 Main results

We present two main results about tests for tensor product codes. First, we prove an analogue of the Bivariate Low-Degree Testing Theorem of Polishchuk and Spielman [PS94] in the low-agreement regime, albeit with much larger field size. Second, we improve on the robustness of the hyperplane test for testing the tensor product code C^m, for $m \geq 3$. We now discuss our results.

2.1 Bivariate low-degree testing in the low-agreement regime

One of the applications of locally testable codes is constructing PCPs, where it is often desirable to reduce the number of queries made by the test. Typically this is done by increasing the alphabet size so that each "large" symbol bundles together several "small" symbols from different locations of the given word. This bundling now introduces a *consistency* problem, because two large symbols may in principle disagree about the same location in the word.

For example, in [RS97; AS03; MR08; BDN17] the test has access to (alleged) restrictions of a low-degree polynomial to all lines, planes, cubes, or other low-degree manifolds. The test samples several queries that intersect, and checks that their answers are consistent on the intersection. These works establish that if the test accepts with probability above a certain threshold, then the restrictions are close to the restrictions of some low-degree polynomial.

We study this problem in a modified setting, where the test only has access to *axis-parallel* restrictions. Restricting the test in this way makes its task more difficult, but doing so provides other advantages. First, axis-parallel restrictions are sometimes the only natural restrictions, such as when testing the m-wise tensor product of a general linear code C (one may consider restrictions to all $(m-1)$-dimensional hyperplanes). Second, having fewer restrictions enables more efficiency, e.g., it facilitates the construction of short PCPs [PS94; BS08].

Indeed, for this very reason, Polishchuk and Spielman [PS94] study the above problem for bivariate polynomials, where $m = 2$ and C is the degree-r Reed–Solomon code. That is, the test has access to a table of row polynomials and a table of column polynomials, and its goal is to check if these are consistent with restrictions of a bivariate polynomial of individual degree r. The test works as follows: pick a random $(x, y) \in \mathbb{F}^2$, read the row and column polynomials through this point, and accept if and only if the two polynomials are equal on (x, y).

Clearly, if all the row polynomials and column polynomials are restrictions of a bivariate polynomial of individual degree r, then the test always accepts. They prove that, conversely, if the test accepts with probability close to 1, then the given polynomials are "close" to being restrictions (to axis-parallel lines) of some low-degree bivariate polynomial, as written below. In the statement, we say that a bivariate polynomial in variables x and y has degree (a, b) if the degree in x is at most a and that in y is at most b. This means that the table of row polynomials, $\mathcal{R}(x, y)$, has degree (r, n) and the table of column polynomials, $\mathcal{C}(x, y)$, has degree (n, r), where n is the size of the table.

Theorem 2.1 ([PS94])**.** *Let $\mathbb{F}$ be a field and $X, Y \subseteq \mathbb{F}$ subsets of size $n := |X| = |Y|$. Let $\mathcal{R}(x, y)$ be a polynomial of degree (r, n) and $\mathcal{C}(x, y)$ a polynomial of degree (n, r) such that*

$$\Pr_{(x,y)\in X\times Y}[\mathcal{C}(x, y) = \mathcal{R}(x, y)] = 1 - \gamma^2$$

for some $\gamma > 0$. If $n > 2\gamma n + 2r$, then there exists a polynomial $Q(x, y)$ of degree (r, r) such that

$$\Pr_{(x,y)\in X\times Y}[\mathcal{C}(x, y) = \mathcal{R}(x, y) = Q(x, y)] \geq 1 - 2\gamma^2 \ .$$

The theorem above assumes that $n > 2\gamma n + 2r$, which means that $\gamma^2 < (1/2 - r/n)^2 < 1/4$. In other words, it requires the row polynomials and column polynomials to agree on (at least) more than three quarters of the points in $X \times Y$. A slight improvement in the parameters of this theorem is shown in [BCGT13]. However, their result still requires the polynomials to agree on a large fraction of the points in $X \times Y$. But what, if anything, can be said if we only assume that they agree, for example, on more than a 0.1-fraction of those points?

There are several results on low-degree testing that show that, even if we only assume that the test accepts with noticeable probability (for the row-vs-column test this probability equals the agreement between row and column polynomials), one can *still* prove the existence of a *short list* of polynomials that 'explain' most of this probability, and this in turn has applications to constructing PCPs with small errors (see, e.g., [RS97; AS03; MR08]).

Our next result gives a positive answer to the question above, stating that even in the low-agreement regime, we can still deduce some structure about the polynomials $\mathcal{R}$ and $\mathcal{C}$, assuming that the field size is sufficiently large.

Theorem 1. *Let $\mathbb{F}$ be a field of size n, $r \in \mathbb{N}$, and $\delta, \varepsilon \in \mathbb{R}$ be such that $\delta > \varepsilon > 6\sqrt{r/n}$. Let $\mathcal{R}(x, y)$ be a polynomial of degree (r, n) and $\mathcal{C}(x, y)$ a polynomial of degree (n, r) such that*

$$\Pr_{(x,y)\in\mathbb{F}^2}[\mathcal{C}(x, y) = \mathcal{R}(x, y)] = \delta \ .$$

If $n > \exp(\Omega(\frac{r}{\varepsilon}\log(\frac{1}{\varepsilon})))$, then there exist $t = O(\frac{1}{\varepsilon})$ polynomials $Q_1(x, y), \ldots, Q_t(x, y)$ of degree (r, r) such that

$$\Pr_{(x,y)\in\mathbb{F}^2}[\exists i \in [t]\ \mathcal{C}(x, y) = \mathcal{R}(x, y) = Q_i(x, y)] \geq \delta - \varepsilon \ .$$

We remark that Theorem 1 holds in general for the 2-wise tensor of *any* linear code $C \subseteq \mathbb{F}^n$ with minimal distance $\geq n - r$ such that $n > \exp(\Omega(\frac{r}{\varepsilon}\log(\frac{1}{\varepsilon})))$. In particular, this means that the minimal distance of C is at least $n - O(\log n)$. See the paragraph *Beyond polynomials* on page 4 for details.

Note that in the above theorem, δ is the agreement probability, while γ^2 in Theorem 2.1 is the disagreement probability. Also, since $r = o(n)$, both δ and ε can be *sub-constant*. This is the first result that analyzes the row-vs-column test in the low acceptance regime that we are aware of.

The row-vs-column test and its higher-dimensional analogues underly many known PCP constructions [BFL91; BFLS91; PS94; BS08]. However, in all these constructions the low degree tests are only analyzed in the high agreement regime. We believe that analyzing the test in the low-agreement regime may imply short PCP constructions with small (sub-constant) soundness. A weakness of the result stated in Theorem 1 is the requirement that the field size must be very large, which restricts us from getting PCPs with polynomial-size proof length. Nonetheless, we consider Theorem 1 as a promising first step in this direction. More generally, our result suggests that the low-agreement regime for tensor product codes merits further study.

To prove the theorem we leverage a fundamental result in extremal graph theory by Kövári, Sós, and Turán. To our knowledge, this is the first time this result is used in the context of low-degree testing. See Section 4.1 below for a high-level description of our proof.

We observe that in the above theorem, for $\delta > \varepsilon$ sufficiently small, it is *necessary* to have a list of at least $\frac{\delta/\varepsilon}{\mathrm{polylog}(\delta/\varepsilon)}$ polynomials in order to explain all but ε of the agreements of $\mathcal{R}$ and $\mathcal{C}$. In particular, if δ is a small constant and $\varepsilon > 0$ is sufficiently small, then the bound in Theorem 1 is tight up to a $\mathrm{polylog}(1/\varepsilon)$ factor.

Proposition 2.2. *Let $\mathbb{F}$ be a field of size n, $r \in \mathbb{N}$, and $\delta, \varepsilon \in \mathbb{R}$ be such that $\delta > \varepsilon > \frac{r}{n}$. There exists a polynomial $\mathcal{R}(x,y)$ of degree $(0,n)$ and a polynomial $\mathcal{C}(x,y)$ of degree $(n,0)$ such that*

$$\Pr_{(x,y)\in\mathbb{F}^2}[\mathcal{C}(x,y) = \mathcal{R}(x,y)] = \delta$$

but for any $t < \frac{\delta/\varepsilon}{\mathrm{polylog}(\delta/\varepsilon)}$ and polynomials $Q_1(x,y), \ldots, Q_t(x,y)$ of degree (r,r) it holds that

$$\Pr_{(x,y)\in\mathbb{F}^2}[\exists i \in [t] \text{ s.t. } \mathcal{C}(x,y) = \mathcal{R}(x,y) = Q_i(x,y)] < \delta - \varepsilon \ .$$

Thus, unlike other results in this area [RS97; AS03; BDN17] where the list size depends only on δ, the list size *must* grow as a function of ε in our setting. The reason for this difference comes from the restriction of our test, which considers only axis-parallel lines, as opposed to arbitrary lines (or planes or cubes) used in the other works. In particular, in our setting once we choose a point (x,y), the lines going through this point are fixed by the design of the test, while in the aforementioned works above there are many lines (or planes or cubes) going through this point, and the performed queries are chosen at random conditioned on the chosen random point.

Beyond polynomials. While [PS94]'s proof relies on polynomials (a key step is Bézout's Theorem), we rely on combinatorial techniques, so that our Theorem 1 holds in general for the 2-wise tensor of *any* linear code $C \subseteq \mathbb{F}^n$ with minimal distance $\geq n - r$ such that $n > \exp(\Omega(\frac{r}{\varepsilon}\log(\frac{1}{\varepsilon})))$. In particular, this means that the minimal distance of C is at least $n - O(\log n)$. The row-vs-column test is now given two matrices $\mathcal{R}, \mathcal{C} \in \mathbb{F}^{n\times n}$ such that every row of $\mathcal{R}$ is in C, and every column of $\mathcal{C}$ is in C. If $\Pr_{(x,y)\in[n]^2}[\mathcal{R}(x,y) = \mathcal{C}(x,y)] = \delta$, then there exist $t = O(1/\varepsilon)$ codewords $Q_1, \ldots, Q_t \in C^2$ such that $\Pr_{(x,y)\in[n]^2}[\exists i \in [t] \text{ s.t. } \mathcal{R}(x,y) = \mathcal{C}(x,y) = Q_i(x,y)] > \delta - \varepsilon$.

In this context it is worth mentioning that there has been a lot of work on the robustness of the axis-parallel line test for 2-wise tensor products, proving both positive results [BS06; DSW06] and negative ones [Val05; CR05; GM12]. We find it quite remarkable that this result holds for general pairwise tensor codes, albeit with very high distance, as the closely related notion of robustness does not hold for general 2-wise tensor products.

Finally, in the high-agreement regime there is a *correspondence* between the robustness of the axis-parallel line test and the soundness of the row-vs-column test (the matrix is given as a collection of lines rather than explicitly).[1] Yet this correspondence *does not hold* in the low-agreement regime. Consider a matrix M whose rows are random independent codewords: the tensor product test passes with probability at least 0.5 (when reading a row), but M is typically far from a tensor codeword.

Open problems. We raise two questions on the low-agreement regime of axis-parallel line tests.

- *Smaller field size.* Our result (Theorem 1) assumes that the field size n is exponential in the degree r. Can one prove a similar result for smaller fields, such as $n = \mathrm{poly}(r)$?
- *Higher dimensions.* Polishchuk and Spielman [PS94] explain that their result (in the high-acceptance regime) also holds in higher dimensions, where now the test is given a table of low-degree polynomials for each axis-parallel line in $\mathbb{F}^m$ and works as follows: pick a random $p \in \mathbb{F}^m$, read the polynomials along the m axis-parallel lines through p, and check that all polynomials agree on p. Can one prove a high-dimensional analogue of Theorem 1? Namely, is it true that if this test accepts with probability $\delta > 0$, then there is a short list of low-degree polynomials that explain most of the agreements?

2.2 Improved robustness for the axis-parallel hyperplane test

We study the robustness of the axis-parallel hyperplane test for the tensor product code $C^m \subseteq \mathbb{F}^{n^m}$, for an arbitrary linear code C with minimal distance d and block length n over the field $\mathbb{F}$. Let $\mathcal{H}$ be the test that, given a word $M \in \mathbb{F}^{n^m}$, samples a random axis-parallel $(m-1)$-dimensional hyperplane H and checks if $M|_H \in C^{m-1}$. For a word $M \in \mathbb{F}^{n^m}$, we define $\delta(M)$ to be the relative distance of the word M to the code C^m and $\rho(M)$ to be $\mathbb{E}_H[\delta(M|_H, C^{m-1})]$, the expected local distance of M. The test $\mathcal{H}$ is *α-robust* if $\rho(M) \geq \alpha \cdot \delta(M)$ for every word $M \in \mathbb{F}^{n^m}$. The 'strength' of the test increases with α, so the goal is to establish the largest α for which this inequality holds.

What is known. There are two main prior works that study the robustness of the test $\mathcal{H}$ for general m. We state the results of these works, starting with one of Ben-Sasson and Sudan [BS06].

Theorem 2.3 ([BS06])**.** *Let $C \subseteq \mathbb{F}^n$ be a linear code with minimal distance d. For $m \geq 3$ and $\left(\frac{d-1}{n}\right)^m \geq 7/8$, the test $\mathcal{H}$ is α-robust for C^m with $\alpha = 2^{-16}$.*

The above theorem is limited in that the proved robustness is small and, moreover, only provides a guarantee when C has a very large distance. Viderman [Vid12] shows that this condition on the distance is not necessary in order to show *some* robustness guarantee.

Theorem 2.4 ([Vid12])**.** *Let $C \subseteq \mathbb{F}^n$ be a linear code with minimal distance d. For $m \geq 3$, the test $\mathcal{H}$ is α-robust C^m with $\alpha = \frac{1}{2m^2}\left(\frac{d}{n}\right)^m$.*

The above theorem, the state of the art in this setting, improves on the previous one as (i) even if $\left(\frac{d-1}{n}\right)^m \geq 7/8$, the robustness provided by Theorem 2.4 is larger than that provided by Theorem 2.3 for $m \leq 169$; (ii) a robustness guarantee is provided for any choice of m, d, n (as long as $m \geq 3$).

Our result. We present a simpler proof of Theorem 2.4, which also achieves a $\frac{1}{m^2}$ improvement in the robustness by showing that the hyperplane test is $\Omega(\frac{d^m}{n^m})$-robust. This improved value for the robustness appears more "natural", because $\frac{d^m}{n^m}$ is the distance of the code C^m.

[1] Let $M \in \mathbb{F}^{n \times n}$ be such that the average relative distance of a row/column of M to some codeword is $1 - \varepsilon$. One can verify that by considering the closest codewords in each row and in each column, the obtained table of row/column codewords passes the row-vs-column test with probability at least $1 - 2\varepsilon$. Therefore, there exists a tensor codeword that agrees with most of the rows and most of the columns, which in turn implies its agreement with M.

Theorem 2. *Let $C \subseteq \mathbb{F}^n$ be a linear code with minimal distance d. For $m \geq 3$, the test $\mathcal{H}$ is α-robust for C^m with $\alpha = \frac{1}{12}\left(\frac{d}{n}\right)^m$.*

Tight or not? Several works have studied the test $\mathcal{H}$ and all resulting analyses have an exponential dependence on m in the robustness. Yet, there is no evidence indicating that this dependence is necessary. Perhaps a "dream" result of constant robustness, for all codes C and $m \geq 3$, is possible. Like previous results, we too incur the same exponential dependence in the robustness. We present some observations that may suggest that this dependence is not necessary.

- Under certain conditions on M, we can prove that $\rho(M) \geq \max\{\frac{1}{m+c}, c' \frac{d^m}{n^m}\} \cdot \delta(M)$ for constants $c, c' > 0$. These two expressions are *incomparable*, as we can set the parameters m, d, n to make either expression bigger than the other. (See Claim A.1.)
- The guarantees of Theorem 2.3, Theorem 2.4, and Theorem 2 all degrade as $\frac{d^m}{n^m}$ decreases. In particular, the proven value of α in all these cases tends to 0 as $\frac{d}{n}$ tends to 0. However, if C is the Reed–Solomon code (or any other code with a similar interpolation property), then we can prove that $\delta(M) \leq \rho(M) + \frac{d}{n}$ for all M. (See Claim A.4.)

We, thus, think that determining the optimal robustness of $\mathcal{H}$ is an intriguing open problem:

> What is the optimal robustness of the hyperplane test $\mathcal{H}$?
> Can one prove that $\alpha = \Omega\left(\max\left(\frac{1}{m}, \frac{d^m}{n^m}\right)\right)$, or even $\alpha = \Omega(1)$, for all codes?

In [BS06], [Vid12], and our result, the proof shows that when $\rho(M)$ is below some threshold (related to the code's unique decoding radius), then $\delta(M)$ is also small. However, when $\rho(M)$ is not below this threshold, the analysis says nothing about $\delta(M)$, and naively uses $\delta(M) \leq 1$ to prove robustness in this regime. We believe that progress on understanding the optimal robustness of $\mathcal{H}$ hinges on understanding what techniques (if any) can be used to bound $\delta(M)$ in terms of $\rho(M)$ for a larger range of $\rho(M)$.

3 Roadmap

The rest of this paper is organized as follows. Section 4 describes the techniques used to prove our results. Section 5 introduces preliminaries for Theorem 1. Section 6 provides the proof of Theorem 1 and Proposition 2.2 (which justifies the list size in Theorem 1). Section 7 introduces basic notations and definitions for Theorem 2. Section 8 provides the proof of Theorem 2.

4 Techniques

We give an overview of the proof techniques behind Theorem 1 and Theorem 2.

4.1 Theorem 1: bivariate testing in the low agreement regime

Polishchuk and Spielman [PS94] prove their result (Theorem 2.1) using the following approach. Given $\mathcal{R}$ and $\mathcal{C}$ (as in the theorem) such that $\Pr_{x,y}[\mathcal{R}(x,y)=\mathcal{C}(x,y)]>1-\delta$, they define an "error polynomial" E that equals 0 for all (x,y) such that $\mathcal{R}(x,y)\neq\mathcal{C}(x,y)$. Since the fraction of points where $\mathcal{R}(x,y)\neq\mathcal{C}(x,y)$ is small, E is a low-degree polynomial. However, in the low-agreement regime that we consider, the degree of E is rather large, which seems to preclude their approach. In particular, a key step based on Bézout's Theorem in their proof appears to break down.

We take a completely different approach, which relies on a combinatorial statement from extremal graph theory. Given $\mathcal{R}$ and $\mathcal{C}$ such that $\Pr_{x,y}[\mathcal{R}(x,y)=\mathcal{C}(x,y)]=\delta$, we define $A\in\{0,1\}^{n\times n}$ to be the 'agreement matrix': $A(x,y)=1$ if and only if $\mathcal{R}(x,y)=\mathcal{C}(x,y)$. By the assumption it follows that A has at least δn^2 ones. By invoking the Kövári-Sós-and Turán Theorem (which may be thought of as an analogue of Ramsey's Theorem for bipartite graphs) it follows that there are some $S,T\subseteq[n]$ such that $|S|,|T|>\Omega(\log(n))\gg r$ and $A|_{S\times T}\equiv 1$. Since the rows of $\mathcal{R}$ and the columns of $\mathcal{C}$ are polynomials of degree r, we deduce that there exists a unique polynomial Q of degree (r,r) such that for all $(x,y)\in S\times T$ it holds that $\mathcal{R}(x,y)=\mathcal{C}(x,y)=Q(x,y)$.

The argument above may appear to be good progress toward our goal. However, there is a total of $\approx\delta n^2$ ones in A, and the rectangle $S\times T$ is of size $O(\log(n))$, i.e., tiny compared to n. This means that the progress is actually rather small!

Nevertheless, we can now set $A|_{S\times T}$ to be zero, and repeat the same argument again, thus covering all but a small fraction of ones of A with small rectangles. However, this raises a new problem. Each rectangle $S\times T$ found in the previous step can be *very* small, and so there are potentially many different polynomials Q that explain the agreements of $\mathcal{R}$ and $\mathcal{C}$. Our next goal is therefore to "stitch" these rectangles together to show that, in fact, there is only a *small* number of distinct polynomials. We do so by "making the rectangles larger", as we now explain.

Consider a rectangle $S\times T$ from the first step, and let $t'\in\mathbb{F}\setminus T$. Note that if there are $r+1$ points $s'\in S$ such that $A(s',t')=1$, then the row polynomial $\mathcal{R}(\cdot,t')$ is uniquely defined by these $r+1$ points, and hence $A(s,t')=1$ for all $s\in S$. Therefore, we can increase T by adding t' to it. On the other hand, if there are less that $r+1$ such points $s'\in S$, then we may disregard these points as they amount to only a small fraction of the points (since $|S|\gg r$). Thus, on a typical rectangle $S\times T$, we can go from size $O(\log(n))\times O(\log(n))$ to size roughly $O(\log(n))\times\Omega(n)$.

In the last step, we show that if we have many rectangles of size $O(\log(n))\times\Omega(n)$ then it is possible to "stitch" them together using the fact that if we have two rectangles $S_1\times T_1$ and $S_2\times T_2$ with corresponding polynomials Q_1 and Q_2 such that $|T_1\cap T_2|>r$, then $Q_1\equiv Q_2$. Indeed, this follows by the fact that if two univariate polynomials of degree r agree on more than r points, then they are equal. We then use the inclusion-exclusion principle to show that for $\varepsilon>\sqrt{\frac{2r}{n}}$ we cannot have more than $\frac{2}{\varepsilon}$ subsets $T_i\subseteq[n]$ of size at least εn such that $|T_i\cap T_j|\leq r$ for all $i\neq j$.

The full proof of Theorem 1 is provided in Section 6.

4.2 Theorem 2: improved robustness for the hyperplane test

Our goal is to prove that the axis-parallel hyperplane test $\mathcal{H}$ is α-robust for $\alpha = \frac{1}{12}\left(\frac{d}{n}\right)^m$. We prove this statement via a careful combination of the approaches taken by [BS06] and [Vid12]. Specifically, we analyze $\rho(M)$ and $\delta(M)$ by studying the following combinatorial object: the *inconsistency graph* G of the hyperplane test $\mathcal{H}$, which we now informally describe.

The test $\mathcal{H}$ has access to a word $M \in \mathbb{F}^{n^m}$, allegedly in C^m. For any axis-parallel hyperplane H, we denote by g_H the closest codeword to $M|_H$ in C^{m-1} (breaking ties by picking an arbitrary closest codeword). The vertex set of the graph G is the set of $(m-1)$-dimensional hyperplanes, which are the local views of the test. There is an edge between two different hyperplanes H and H' if g_H and $g_{H'}$ disagree on the intersection of the hyperplanes, $H \cap H'$. (See Definition 8.1 for details.) In other words, the graph has an edge between two planes if the local codewords assigned to the planes are inconsistent. The graph G that we study is similar to the inconsistency graph analyzed in [BS06]. The difference is that, for some threshold parameter τ, the graph used in [BS06] adds an edge from H to every other H' in the graph if $\delta(M|_H, g_H) > \tau$.

First, we show that if G has a large independent set I, then there is a codeword f in C^m that agrees with the local codewords g_H on *every* hyperplane H in I. For an independent set I, we define I_b to be the set of $i \in [n]$ such that the hyperplane $\{p \in [n]^m : p_b = i\}$ is in I. A key property of tensor product codes is the unique extension property, which we formally state later on as Claim 7.2. Using the unique extension property of tensor product codes, we show that if there are two axes b_1 and b_2 where I_{b_1} and I_{b_2} both have at least $n-d+1$ planes, then there is a word f in C^m where $f|_H = g_H$ for every H in the independent set. Without loss of generality assume $b_1 = 1$ and $b_2 = 2$. Intuitively, we fill in the restricted hypercube in $\mathbb{F}^{I_1 \times I_2 \times n^{m-2}}$ with the values of the closest codewords to $M|_H$ for each H in the independent set. Since the independent set is large, the restricted hypercube is large enough so that we can extend the partially filled-in hypercube to a unique codeword f in C^m. The uniqueness of the extension implies that $f|_H = g_H$ for every H in I.

Next, we analyze the structure of G to show that every edge is adjacent to a vertex of degree at least $(m-2)d/2$. The key point is that two different C^{m-2} codewords must disagree on at least d^{m-2} points, and these points have a particular structure. For two distinct C^{m-2} codewords, we prove that on each of the $m-2$ remaining axes there must be at least d planes, parallel to that axis, that contain points of disagreement. If not, then using the unique extension property we show that the two codewords must be equal, which is a contradiction. For any edge (H, H'), this gives us a total of $(m-2)d$ planes that disagree with at least one of g_H and $g_{H'}$ on $H \cap H'$, which shows that $\deg(H) + \deg(H')$ is at least $(m-2)d$. Therefore, at least one of H and H' has degree at least $(m-2)d/2$. As an immediate consequence, the set of planes with degree at least $(m-2)d/2$, which we denote by L, is a *vertex cover*, and the set of planes not in L is an *independent set* I.

With some algebraic manipulation, we relate the size of this vertex cover to the expected local distance $\rho(M)$. By expressing $\rho(M)$ as a sum over pairs of intersecting planes, we show that

$$\rho(M) \geq \frac{1}{n^m m(m-1)} \sum_{(H,H'):H\cap H' \neq \emptyset} \Delta|_{H\cap H'}(g_H, g_{H'}) \ .$$

This allows us to express the robustness of the test $\mathcal{H}$ in terms of the size of the vertex cover L.

Similar to the analysis of [Vid12], we break up the proof into two cases. If $|L|$ is somewhat large, then $\rho(M) \geq \frac{1}{12}\left(\frac{d}{n}\right)^m$, and the theorem follows immediately because $\delta(M)$ is anyways at most 1. If

$|L|$ is small, then the corresponding independent set has two axes where $|I_b| \geq n-d+1$. Therefore, there is a global codeword f that is consistent with all the hyperplanes in the independent set. We use this fact to show that $\delta(M)$ must be small when $\rho(M)$ is small, which concludes the proof.

The full proof of Theorem 2 is provided in Section 8.

5 Preliminaries for Theorem 1

5.1 Low-degree polynomials

We will use the following lemmas about low-degree polynomials in the proof of Theorem 1. These are standard interpolation lemmas, and direct proofs can be found in [PS94].

Lemma 5.1. *Let $S, T \subseteq \mathbb{F}$ be two sets each of size at least $r+1$. Suppose that for two polynomials $Q_1(x,y), Q_2(x,y)$ of degree (r,r), if holds that $Q_1(x,y) = Q_1(x,y)$ for all $(x,y) \in (S,T)$. Then $Q_1 \equiv Q_2$.*

Lemma 5.2. *Let $S, T \subseteq \mathbb{F}$ be two sets each of size at least $r+1$. Suppose that there is polynomial $\mathcal{R}(x,y)$ of degree (r,n), and a polynomial $\mathcal{C}(x,y)$ of degree (n,r) such that $\mathcal{R}(x,y) = \mathcal{C}(x,y)$ for all $(x,y) \in (S,T)$. Then, there exists a polynomial $Q(x,y)$ of degree (r,r) such that $Q(x,y) = \mathcal{C}(x,y) = \mathcal{R}(x,y)$ for all $(x,y) \in (S,T)$.*

Corollary 5.3. *Let $S, T \subseteq \mathbb{F}$ be two sets each of sizes $|S| \geq r+2$ and $|T| \geq r+2$, and let $(x_0, y_0) \in (S,T)$. Suppose that there is a polynomial $\mathcal{R}(x,y)$ of degree (r,n), and a polynomial $\mathcal{C}(x,y)$ of degree (n,r) such that $\mathcal{R}(x,y) = \mathcal{C}(x,y)$ for all $(x,y) \in (S,T) \setminus \{(x_0,y_0)\}$. Then $\mathcal{C}(x_0,y_0) = \mathcal{R}(x_0,y_0)$.*

5.2 The Kövári–Sós–Turán theorem

We first define the density of a binary matrix.

Definition 5.4. *Let $A \in \{0,1\}^{k \times \ell}$ be a binary matrix. Define the* density *of A to be $\delta(A) = \frac{\sum_{i \in [k], j \in [\ell]} A_{i,j}}{k \cdot \ell}$. We say that A is τ-*dense *if $\delta(A) \geq \tau$.*

In the proof of Theorem 1 we will use a result due to Kövári, Sós, and Turán [KST54], which states that any sufficiently dense binary matrix contains a large submatrix where every entry is 1.

Theorem 5.5 (Kövári, Sós, Turán)**.** *Let N, M, t, s be natural numbers that satisfy $N \geq s$ and $M \geq t \geq s$, and let $A \in \{0,1\}^{N \times M}$ be a binary matrix. If A is $\left(\sqrt[s]{\frac{t-1}{M}} + \frac{s}{N}\right)$-dense, then there are $S \subseteq [N]$ and $T \subseteq [M]$ of sizes $|S| = s$ and $|T| = t$ such that $A|_{S \times T} \equiv 1$.*

Remark 5.6. *The Kövári–Sós–Turán theorem is usually stated as saying that any sufficiently dense bipartite graph contains a large bipartite clique. It is clear, however, that the matrix formulation above is equivalent by associating a bipartite graph with its adjacency matrix, where the rows correspond to the vertices on the left, and the columns correspond to the vertices on the right.*

6 Proof of Theorem 1

The key step in the proof of Theorem 1 is the following lemma.

Lemma 6.1 (Key lemma)**.** *Suppose that* $|\mathbb{F}| > \exp(\Omega(\frac{r}{\varepsilon}\log(\frac{1}{\varepsilon})))$*. Then, for any* $\varepsilon > \sqrt{\frac{2r}{|\mathbb{F}|}}$ *there are* $t \leq \frac{2}{\varepsilon}$ *polynomials* $Q_1, \ldots, Q_t$ *each of degree* (r, r)*, and subsets* $S_1, \ldots, S_t, B_1, \ldots, B_t \subseteq \mathbb{F}$ *such that*

1. *For all* $i \in [t]$ *and* $(x, y) \in (S_i, B_i)$ *it holds that* $\mathcal{C}(x, y) = \mathcal{R}(x, y) = Q_i(x, y)$*.*
2. *All* S_i*'s are pairwise disjoint.*
3. $\frac{|\cup_{i\in[t]} S_i \times B_i|}{|\mathbb{F}|^2} \geq \delta - 3\varepsilon$*, where* $\delta = \Pr[\mathcal{C}(x, y) = \mathcal{R}(x, y)]$*.*

Before proving Lemma 6.1 let us see how it immediately implies Theorem 1.

Proof of Theorem 1 using Lemma 6.1. Let $\varepsilon > 6\sqrt{\frac{r}{n}}$, and apply Lemma 6.1 with $\varepsilon/3 > \sqrt{\frac{2r}{n}}$. By Lemma 6.1 for some $t \leq \frac{2}{\varepsilon/3} = \frac{6}{\varepsilon}$ there are disjoint subsets $S_1 \times B_1, \ldots, S_t \times B_t \subseteq \mathbb{F}^2$ such that $\frac{|\cup_{i\in[t]} S_i \times B_i|}{|\mathbb{F}|^2} \geq \delta - \varepsilon$, and for all $i \in [t]$ and $(x, y) \in (S_i, B_i)$ it holds that $\mathcal{R}(x, y) = \mathcal{C}(x, y) = Q_i(x, y)$. This implies that

$$\Pr_{(x,y)\in\mathbb{F}^2}[\exists i \in [t] \text{ s.t. } \mathcal{R}(x, y) = \mathcal{C}(x, y) = Q_i(x, y)] \geq \Pr[(x, y) \in \cup_{i\in[t]} S_i \times B_i] \ ,$$

which is at least $\delta - \varepsilon$, as required. □

We devote the rest of this section to proving Lemma 6.1.

6.1 Proof of Lemma 6.1

Let $n = |\mathbb{F}|$, and define the binary matrix $A \in \{0, 1\}^{n\times n}$ where $A(x, y) = 1$ if $\mathcal{C}(x, y) = \mathcal{R}(x, y)$ and $A(x, y) = 0$ otherwise. Note that by the assumption of Theorem 1, we have $\frac{\sum_{x,y\in[n]} A(x,y)}{n^2} = \delta$, i.e., the matrix A is δ-dense.

6.1.1 Step 1

In the first step we apply Theorem 5.5 iteratively to show that there exists a collection of disjoint sets $S_1, \ldots, S_u \subseteq [n]$ with $|S_i| \geq \frac{r}{\varepsilon}$ such that for most points (x, y) it holds that if $A(x, y) = 1$, then $x \in \cup S_i$, and for each $i \in [u]$ there exists $T_i \subseteq [n]$ of size $|T_i| \geq \frac{r}{\varepsilon}$ such that $A_{S_i\times T_i} \equiv 1$.

Claim 6.2. *Let* $n, r \in \mathbb{N}$*,* $\delta > \varepsilon > 0$*, and let* $k = \lceil r/\varepsilon \rceil$*. Let* $A \in \{0, 1\}^{n\times n}$ *be a* δ*-dense matrix as above, and suppose that* $n > 2k^2 \left(\frac{1}{\varepsilon}\right)^{k+1}$*. Then, there exist* $u \in \mathbb{N}$ *and two sequences* $S_i \subseteq [n], T_i \subseteq [n]$ *with* $i = 1, \ldots, u$ *satisfying the following conditions.*

1. *The* S_i*'s are pairwise disjoint.*
2. $|S_i| = |T_i| = k$*.*
3. $A(x, y) = 1$ *for every* $(x, y) \in (S_i, T_i)$ *and* $i \in [u]$*.*

4. $\sum_{(x,y)\in([n]\setminus(\cup S_i),[n])} A(x,y) < \varepsilon n^2$.

Proof. We will use Theorem 5.5 to find a submatrix of A of size $k \times k$ whose entries are all 1s. By the choice of k and the assumption that n is sufficiently large we have that $(\varepsilon - \frac{k}{n})^k = \varepsilon^k(1-\frac{k}{\varepsilon n})^k > \varepsilon^k(1-\frac{k^2}{\varepsilon n}) > \varepsilon^k/2 > \frac{k-1}{\varepsilon n}$, and hence $\varepsilon > \sqrt[k]{\frac{k-1}{\varepsilon n}} + \frac{k}{n}$. Hence, since A is δ-dense, we have $\delta(A) \geq \delta \geq \varepsilon > \sqrt[k]{\frac{k-1}{n}} + \frac{k}{n}$. Therefore, by Theorem 5.5 there exist $S_1 \subseteq [n], T_1 \subseteq [n]$ each of size $|S_1| = |T_1| = k$ such that $A|_{S_1 \times T_1} \equiv 1$.

Next, we remove the rows contained in S_1 from A, and apply the same argument again. Let $M_1 = [n] \setminus S_1$ and define A_1 to be the $(n-k) \times n$ submatrix of A whose rows are indexed by M_1. Note that if $\sum_{x\in M_1, y\in[n]} A_1(x,y) > \varepsilon n^2$ then $\delta(A_1) \geq \frac{\varepsilon n}{|M_1|}$, and thus we have $\delta(A_1) \geq \frac{\varepsilon n}{n-k} > \varepsilon > \sqrt[k]{\frac{k-1}{|M_1|}} + \frac{k}{n}$. Therefore, we can apply Theorem 5.5 again, and find $S_2 \subseteq M_1$ and $T_2 \subseteq [n]$ of size $|S_2| = |T_2| = k$ such that $A|_{S_2 \times T_2} \equiv 1$.

We repeat the same argument again, for each $i \geq 2$ defining the the subset $M_i = M_{i-1} \setminus S_{i-1}$, and letting $A_i = A_{M_i \times [n]}$. Note that if $\sum_{x\in M_i, y\in[n]} A(x,y) \geq \varepsilon n^2$ then $|M_i| \geq \varepsilon n$, and $\delta(A_i) \geq \frac{\varepsilon n}{|M_i|} \geq \varepsilon > \sqrt[k]{\frac{k-1}{|M_i|}} + \frac{k}{n}$. Therefore, by Theorem 5.5 there exist $S_i \subseteq M_i$ and $T_i \subseteq [n]$ of size $|S_i| = |T_i| = k$ such that $A|_{S_i \times T_i} \equiv 1$.

We stop the process after u iterations when $\sum_{x\in M_u, y\in[n]} A(x,y) < \varepsilon n^2$. By definition of the S_i's and T_i's, this gives us the subsets with the desired properties. □

By the assumption $|\mathbb{F}| = n > \exp(\Omega(\frac{r}{\varepsilon}\log(\frac{1}{\varepsilon})))$ in Theorem 1 we have $n > 2k^2\left(\frac{1}{\varepsilon}\right)^{k+1}$. Therefore, we can apply Claim 6.2 on A to get S_i's and T_i's as in the claim.

6.1.2 Step 2

Next, we show that the sets T_i from the previous step can be chosen to be of size at least εn.

Claim 6.3. *Let $\{(S_i,T_i)\}_{i=1}^u$ be the sets from Claim 6.2. For each $i \in [u]$ define $B_i = \{y_0 \in [n] : \sum_{x\in S_i} A(x,y_0) \geq r+1\}$. Then*

1. $\sum_{i\in[u]} \sum_{\substack{x\in S_i \\ y\in[n]\setminus B_i}} A(x,y) \leq \varepsilon n^2$.

2. *For every $i \in [u]$ if $y_0 \in B_i$ then $A(x,y_0) = 1$ for all $x \in S_i$.*

Proof. The first item is by the choice of $k \geq r/\varepsilon$. In each $i \in [u]$ and $y \in [n] \setminus B_i$ it holds that less than ε fraction of the entries are ones, and hence the total number of ones in all $i \in [u]$ and $y \in [n] \setminus B_i$ is less that εn^2. Formally, we have

$$\sum_{i\in[u]} \sum_{\substack{x\in S_i \\ y\in[n]\setminus B_i}} A(x,y) \leq \sum_{i\in[u]} \sum_{y\in[n]\setminus B_i} r \leq u \cdot n \cdot r \leq \varepsilon n^2 \ ,$$

where the last inequality uses the fact that $u \leq n/k$, and $k \geq r/\varepsilon$.

To prove the second item, we use Corollary 5.3. Suppose that $A(x_0,y_0) = 0$ for some $x_0 \in S_i$ and $y_0 \in B_i$. By the assumption on B_i, it holds that $|\{x \in S_i : A(x,y_0) = 1\}| \geq r+1$. Let $S = \{x_0\} \cup \{x \in S_i : A(x,y_0) = 1\}$, and let $T = \{y_0\} \cup T_i$, so that $A(x,y) = 1$ for all $(x,y) \in S \times T \setminus \{(x_0,y_0)\}$. Recall that, by definition of A, $\mathcal{R}(x,y) = \mathcal{C}(x,y)$ for all such (x,y), and hence, by Corollary 5.3 we also have $\mathcal{R}(x_0,y_0) = \mathcal{C}(x_0,y_0)$, and thus $A(x_0,y_0) = 1$. □

Note that the ones not covered by $\cup_i(S_i \times B_i)$ are the $\leq \varepsilon n^2$ ones omitted in Claim 6.2 and the $\leq \varepsilon n^2$ ones disregarded in the proof of Claim 6.3 above. Let us also disregard all S_i's and B_i's such that $|B_i| \leq \varepsilon n$, and consider only the remaining subsets. Note that the set of B_i's with $|B_i| \leq \varepsilon n$ can contain at most εn^2 ones. Redefining u to be the number of remaining sets, we get two collections of subsets $\{S_i \subseteq [n], B_i \subseteq [n]\}_{i=1}^u$ such that

1. the S_i's are pairwise disjoint.
2. $|B_i| > \varepsilon n$ for all $i \in [u]$.
3. $\sum_{(x,y)\in\cup_{i=1}^u S_i\times B_i} \geq (\delta - 3\varepsilon)n^2$.
4. $A|_{S_i\times B_i} \equiv 1$ for all $i \in [u]$.

In particular, by Lemma 5.2 for each $i = 1, \dots, u$ there is a polynomial P_i of degree (r, r) such that $\mathcal{R}(x, y) = \mathcal{C}(x, y) = P_i(x, y)$ for all $(x, y) \in S_i \times B_i$.

6.1.3 Step 3

Next, we observe that if two sets B_i, B_j from the previous step have large intersection, then the corresponding polynomials P_i and P_j are equal.

Claim 6.4. *Suppose that $|B_i \cap B_j| \geq r + 1$ for some $i \neq j \in [u]$. Then $P_i = P_j$ and $B_i = B_j$.*

Proof. Denote $B = B_i \cap B_j$. Note that, for each $y \in B$, $P_i(x, y) = \mathcal{C}(x, y)$ for all $|S_i| = k > r + 1$ values of $x \in S_i$, and hence $P_i(x, y) = \mathcal{C}(x, y)$ for all $x \in [n]$. In particular, $P_i(x, y) = \mathcal{C}(x, y)$ for all $(x, y) \in S_j \times B$. Therefore, $P_i|_{S_j\times B} \equiv P_j|_{S_j\times B}$, and thus $P_i \equiv P_j$ by Lemma 5.1. Applying Corollary 5.3, we conclude that $P_i(x, y) = P_j(x, y) = \mathcal{C}(x, y) = \mathcal{R}(x, y)$ for all $(x, y) \in (S_i \cup S_j) \times (B_i \cup B_j)$. This implies that $B_i = B_j$, as required. ☐

6.1.4 Completing the proof

In the last step we will show that there is a short list of $t \leq \frac{2}{\varepsilon}$ polynomials $Q_1, \dots, Q_t$ such that each of the P_i's is in fact equal to one of the Q_j's. Indeed, denote the number of different B_i's by t. By Claim 6.4, if $B_i \neq B_j$ then $|B_i \cap B_j| \leq r$, and thus by the inclusion-exclusion principle we have

$$n \geq \left|\cup_{i=1}^t B_i\right| \geq \sum_{i=1}^t |B_i| - \sum_{i\neq j} |B_i \cap B_j| \geq t \cdot \varepsilon n - \binom{t}{2} r \ ,$$

where in the last inequality we used the bound $|B_i| > \varepsilon n$ for all i. If $t \geq \frac{2}{\varepsilon}$, then $n \geq t \cdot \varepsilon n - \binom{t}{2} r \ \geq 2n - \frac{2}{\varepsilon^2} r$, and thus $\varepsilon < \sqrt{\frac{2r}{n}}$, which contradicts the assumption on ε. Therefore $t < \frac{2}{\varepsilon}$, as required.

6.2 Proof of Proposition 2.2

We prove that in Theorem 1, for $\delta > \varepsilon$ sufficiently small, it is *necessary* to have a list of $\frac{\delta/\varepsilon}{\text{polylog}(\delta/\varepsilon)}$ polynomials in order to explain all but ε of the agreements of $\mathcal{R}$ and $\mathcal{C}$.

Let $\mathbb{F}$ be a prime field that is sufficiently large, and let $N \in \mathbb{N}$ be an integer that depends on ε, to be chosen later. For all $i \in [N]$ let $a_i = \frac{1}{i \log^2 i}$, and let $c = \sum_{i=1}^{N} a_i^2$. Note that $\sum_{i=1}^{\infty} a_i < \infty$, and $c = \sum_{i=1}^{\infty} a_i^2 < \infty$. Let $A_1, \ldots, A_N \subseteq \mathbb{F}$ be disjoint subsets of size $|A_i| = \left(\sqrt{\frac{\delta}{c}}\right) a_i |\mathbb{F}|$. (We assume for simplicity that $\mathbb{F}$ is sufficiently large and there are no rounding issues.) Note that $\sum_{i=1}^{N} |A_i| = \sqrt{\frac{\delta}{c}} \sum_{i=1}^{N} a_i |\mathbb{F}|$, which is smaller than $|\mathbb{F}|$ for $\delta > 0$ sufficiently small.

For every $y \in A_i$, set the row polynomial $\mathcal{R}(x, y)$ equal to i, and for every $x \in A_i$, set the column polynomial $\mathcal{C}(x, y)$ equal to i. For all $y \notin \cup A_i$ let $\mathcal{R}(x, y) = N + 1$, and for all $x \notin \cup A_i$ let $\mathcal{C}(x, y) = N + 2$. Note that $\mathcal{R}(x, y) = \mathcal{C}(x, y)$ if and only if $x, y \in A_i$ for some i. In particular,

$$\Pr_{x,y}[\mathcal{R}(x, y) = \mathcal{C}(x, y)] = \sum_{i=1}^{N} \left(\frac{|A_i|}{|\mathbb{F}|}\right)^2 = \sum_i \frac{\delta}{c} a_i^2 = \delta.$$

It is clear that if $|A_i| > r + 1$ for all $i \in [N]$, then the only bivariate polynomials that agree with $\mathcal{R}$ and $\mathcal{C}$ on more than $\frac{r}{|\mathbb{F}|}$ fraction of point are the constant polynomials $Q_i \equiv i$ for all $i = 1, \ldots, N$. By construction, for each $i \in [N]$ we have $\Pr_{x,y}[\mathcal{R}(x, y) = \mathcal{C}(x, y) = Q_i(x, y)] = \left(\frac{|A_i|}{|\mathbb{F}|}\right)^2 = \frac{\delta}{c} a_i^2$, and for different Q_i's the domains of the agreements are disjoint. Therefore, in order to choose t polynomials Q_i that maximize the probability $\Pr_{x,y}[\exists i \in [t] \text{ s.t. } \mathcal{R}(x, y) = \mathcal{C}(x, y) = Q_i(x, y)]$ we should choose the first t polynomials $Q_1, \ldots, Q_t$ to get

$$\Pr_{x,y}[\exists i \in [t] \text{ s.t. } \mathcal{R}(x, y) = \mathcal{C}(x, y) = Q_i(x, y)] = \delta - \sum_{i=t+1}^{N} \frac{\delta}{c} a_i^2.$$

Note that $\sum_{i=t+1}^{\infty} a_i^2 > \frac{1}{t \cdot \text{polylog}(t)}$. Therefore, if we choose N to be sufficiently large, as a function of t, we will get that

$$\Pr_{x,y}[\exists i \in [t] \text{ s.t. } \mathcal{R}(x, y) = \mathcal{C}(x, y) = Q_i(x, y)] < \delta - O\left(\delta \cdot \frac{1}{t \cdot \text{polylog}(t)}\right).$$

By letting $\varepsilon = O\left(\delta \cdot \frac{1}{t \cdot \text{polylog}(t)}\right)$ we get that $t > \frac{\delta/\varepsilon}{\text{polylog}(\delta/\varepsilon)}$, as required.

7 Preliminaries for Theorem 2

7.1 Linear codes

A linear code C over a field $\mathbb{F}$ is a linear subspace C of the vector space $\mathbb{F}^n$. Each codeword w in C is a string of length n, which is the block length of the code. The dimension of the code $\dim(C)$ is the dimension of C as a vector space in $\mathbb{F}^n$. For any two words w and v in $\mathbb{F}^n$, the Hamming distance between w and v, denoted by $\Delta(w, v)$, is the number of indices where i where $w_i \neq v_i$. Formally, $\Delta(w, v) = |\{i \in [n] : w_i \neq v_i\}|$. The relative distance between w and v is $\delta(w, v) = \Delta(w, v)/n$, which is the fraction of points where w and v disagree. For any subset S of $[n]$, we will define $\Delta|_S(w, v)$ to be $|\{i \in S : w_i \neq v_i\}|$, which is the Hamming distance between w and v on the subset S. Similarly, $\delta|_S(w, v) = \Delta|_S(w, v)/|S|$. The distance d of a code C is the minimum Hamming distance between any two distinct codewords of C, i.e. $d = d(C) = \min_{w \neq v \in C} \Delta(w, v)$. For any w in $\mathbb{F}^n$, the distance from w to C is defined as $\Delta(w, C) = \min_{v \in C} \Delta(w, v)$, and the relative distance is defined similarly. For any subset $S \subseteq [n]$, the distance from w to C on S is $\Delta|_S(w, C) = \min_{v \in C} \Delta|_S(w, v)$. We will write $\delta(w)$ instead of $\delta(w, C)$ when the code is clear from the context.

Linear codes have a unique extension property.

Claim 7.1 (Unique Extension). *Let I be a subset of $[n]$ of size at least $n - d + 1$. Let C' be the restriction of the code C to the subset I. Then, for every codeword $w \in C'$ there exists a unique $v \in C$ such that $v|_I = w$.*

Proof. By definition, for every w in C' there must exist at least one v in C such that $v|_I = w$. Suppose there exists v_1 and v_2 such that $v_1|_I = v_2|_I = w$. Then v_1 and v_2 agree on S, so $\Delta(v_1, v_2) \leq n - |I| \leq d - 1$. Since v_1 and v_2 are codewords, $\Delta(v_1, v_2) < d$ if and only if $v_1 = v_2$. Therefore, the codeword w has a unique extension to C. □

7.2 Tensor product codes

For any linear code C, the 2-wise tensor product of C, denoted by $C^2 = C \otimes C$ is the linear code in $\mathbb{F}^{n^2}$, where every codeword $M \in \mathbb{F}^{n^2}$ is an $n \times n$ matrix whose each row and column is a codeword of C. The m-wise tensor of C, denoted by C^m, is defined recursively as $C^{m-1} \otimes C$. The code C^m has block length n^m and distance d^m. Furthermore, each $f \in C^m$ can be written as an $n \times n \times \cdots \times n$ (m times) matrix where the entries are values in $\mathbb{F}$, and each axis-parallel line is in C. It is easy to see that f is in C^m if and only if the restriction of f to any $(m-1)$-dimensional axis-parallel hyperplane H is in C^{m-1}. It is also worth noting that the fractional distance of the code C^m is $(d/n)^m$, so the fractional distance of the code decays exponentially in m.

Tensor product codes have a unique extension property that will be used many times in the proof of Theorem 2.

Claim 7.2 (Unique Extension for Tensor Product Codes). *Let $\{C_b\}_{b=1}^m$ be codes with blocklength n_b and distance d_b. Let $I_b \subseteq [n_b]$ be a set of size at least $n_b - d_b + 1$, and let C'_b be the projection of C_b to I_b. Then for every $w \in C' = C'_1 \otimes \cdots \otimes C'_m$, there exists a unique v in $C = C_1 \otimes \cdots \otimes C_m$ such that $v|_{I_1 \times \cdots \times I_m} = w$.*

Proof. By Claim 7.1, for all $b \in [m]$ the projection map $\pi_b \colon C_b \to C'_b$ is bijective. We can extend π_b to be a bijective map from the hybrid code $C'_1 \otimes \cdots \otimes C'_{b-1} \otimes C_b \otimes \cdots \otimes C_m$ to $C'_1 \otimes \cdots \otimes C'_b \otimes C_{b+1} \otimes \cdots \otimes C_m$. For any v in the first hybrid code, define $\pi_b(v) = v|_{I_1 \times \cdots \times I_b \times n_{b+1} \times \cdots \times n_m}$, which is the projection of v to I_b along the bth axis, and the identity map everywhere else. Clearly, π_b is still a bijection, and so the composition of maps $\pi = \pi_m \circ \pi_{m-1} \circ \cdots \circ \pi_1$ is therefore a bijection from C to C', which proves the claim. □

7.3 Locally testable codes and robust tests

A q-query test $\mathcal{T}$ for a code $C \subseteq \mathbb{F}^n$ is a probabilistic algorithm that, given oracle access to a word $w \in \mathbb{F}^n$, makes q (non-adaptive) queries to w and then accepts or rejects. Informally, C is locally testable if there is a test $\mathcal{T}$ that accepts (with probability 1) whenever w is in C, and rejects (say with probability at least 0.5) when w is far from C.

The expected local view distance $\rho^{\mathcal{T}}(w)$ of $\mathcal{T}$ on a word w is the average, over the local views of $\mathcal{T}$, of the distance of w to an accepting view. Instead of analyzing the local testability of C^m, we will instead consider a stronger notion of local testability called robustness, that was introduced in [BS06]. The test $\mathcal{T}$ is *α-robust* if $\rho^{\mathcal{T}}(w) \geq \alpha \cdot \delta(w, C)$ for every word $w \in \mathbb{F}^n$. The 'strength' of the test increases with α, so the goal is to establish the largest α for which this inequality holds.

7.4 The axis-parallel hyperplane test

Definition 7.3. *Let C be a linear code, and let C^m be the m-wise tensor of C. The axis-parallel hyperplane test $\mathcal{H}$ for C^m is the test that given a word $M \in \mathbb{F}^{n^m}$ samples a random axis-parallel $(m-1)$-dimensional hyperplane H and checks if $M|_H \in C^{m-1}$.*

We introduce several observations about the test $\mathcal{H}$ that will be useful in the proof of Theorem 2. Since the hyperplanes sampled by $\mathcal{H}$ are axis-parallel, each hyperplane $H \subseteq [n]^m$ must be a set of the form $H = \{p \in [n]^m : p_b = i\}$, for some $b \in [m]$ and $i \in [n]$. This means that there are nm hyperplanes in total, and each hyperplane can be specified by the pair (b, i). We will use (b, i) to refer to the hyperplane $\{p \in [n]^m : p_b = i\}$.

For $M \in \mathbb{F}^{n^m}$ and an axis-parallel hyperplane H in $[n]^m$, we define g_H to be the closest C^{m-1} codeword to $M|_H$. If this codeword is not unique, then we break ties by picking an arbitrary closest codeword. Using this notation, the expected local view distance $\rho(M)$ can be expressed as

$$\rho(M) = \mathbb{E}_H[\delta|_H(M, C^{m-1})] = \mathbb{E}_H[\delta|_H(M, g_H)] \ ,$$

where the expectation is taken over all axis-parallel hyperplanes H.

Definition 7.4. *The test $\mathcal{H}$ is* α-robust *if $\rho(M) \geq \alpha \cdot \delta(M, C^m)$ for every word $M \in \mathbb{F}^{n^m}$, where $\delta(M, C^m)$ is the relative distance of the word M to the code C^m, and $\rho(M)$ is the expected local distance of M.*

Note that robustness α for the test $\mathcal{H}$ is at most 1.

Lemma 7.5. *The robustness of the axis-parallel hyperplane test $\mathcal{H}$ is $\alpha \leq 1$.*

Proof. Let f be any C^m codeword such that $\delta(M) = \delta(M, f)$. Then,

$$\delta(M) = \delta(M, f) = \frac{1}{nm} \sum_H \delta|_H(M, f) \geq \frac{1}{nm} \sum_H \delta|_H(M, g_H) = \rho(M)$$

since g_H is closer to $M|_H$ than $f|_H$, as $f|_H \in C^{m-1}$. Thus $\alpha \leq \rho(M)/\delta(M) \leq 1$. □

In Appendix B we discuss the difference between robustness and soundness for the test $\mathcal{H}$.

8 Proof of Theorem 2

We prove Theorem 2. We start by recalling some notation. Let C be a linear code with distance d and block length n over $\mathbb{F}$, and let C^m be the m-wise tensor product of C, for some $m \geq 3$. Let M be the input to the test $\mathcal{H}$, which is an evaluation table of a function from $[n]^m \to \mathbb{F}$. Define g_H to be the closest C^{m-1} word to $M|_H$, where ties are broken by picking an arbitrary closest codeword. We will view M as fixed throughout the analysis.

We need to show that $\rho(M) \geq \alpha \cdot \delta(M)$, for $\alpha = \frac{1}{12}\left(\frac{d}{n}\right)^m$. The main idea in the proof is to upper bound $\delta(M)$ by figuring out how to "stitch" together the g_H's to make a global codeword f. We begin by defining the inconsistency graph G. The graph G has each hyperplane as a vertex, and has an edge between two hyperplanes H and H' if they have nonzero intersection and their respective local codewords g_H and $g_{H'}$ are inconsistent, i.e., they disagree on some point p in their intersection $H \cap H'$.

Definition 8.1 (Inconsistency Graph)**.** *The inconsistency graph G of the test $\mathcal{H}$ is a graph where V is the set of hyperplanes, and $E = \{(H, H') : \exists p \in H \cap H' \text{ s.t. } g_H(p) \neq g_{H'}(p)\}$.*

The proof will be divided into several steps. First, we will show that if G contains a large independent set, namely a large set of planes which are all consistent with each other, then there is a global codeword f that stitches together all of the local codewords g_H for every H in the independent set. Then, we will show that every edge in G is adjacent to a vertex of (somewhat) large degree. This will imply that the set of vertices that have large degree is a vertex cover, and its complement is an independent set. We will then show that $\rho(M)$ is lower bounded by some function that is linear in the number of vertices that have large degree. Using these components, we will conclude the proof.

8.1 Step 1: the case of a large independent set

We will show that if G has a large independent set I, then there is an f in C^m that agrees with g_H on H for every H in I. In other words, f is the codeword of C^m that stitches together all of the g_H's in the independent set. The proof relies on Claim 7.2.

Lemma 8.2 (Interpolation)**.** *If G has an independent set I of size $|I| > (m-1)(n-d)+n$, then there exists f in C^m such that $f|_H = g_H$ for every $H \in I$.*

Our proof of this lemma is similar to the proof of a different lemma in [BS06].

Proof. Define I_b to be the set of $i \in n$ such that the plane (b, i) is in I. Since $|I| > (m-1)(n-d)+n$, there must exist $b_1 \neq b_2$ such that $|I_{b_1}|$ and $|I_{b_2}|$ are at least $n-d+1$, as otherwise $|I| = \sum_{b=1}^{m} |I_b| \leq (m-1)(n-d)+n$. Without loss of generality assume $b_1 = 1$ and $b_2 = 2$. Let $S = I_1 \times I_2 \times [n]^{m-2}$ and let $g \colon S \to \mathbb{F}$ be a matrix in $\mathbb{F}^S$. Define $g(p) = g_H(p)$ for every $p \in S$, where H is some plane in $I_1 \cup I_2$ such that $p \in H$. Note that g is well-defined since all the planes in I are consistent with each other, as I is an independent set.

We claim that $g \in C|_{I_1} \otimes C|_{I_2} \otimes C^{m-2}$. This is because for any $H \in I_1$ it holds that $g|_H \in C|_{I_2} \otimes C^{m-2}$, as $g|_H = g_H$ except that the second axis is now restricted to I_2. This means that for every axis $b \neq 1, 2$ and for every line ℓ_b parallel to the b-th axis it holds that $g|_{\ell_b} \in C$. Also, for every line ℓ_2 parallel to the second axis we have that $g|_{\ell_2} \in C|_{I_2}$, because we took a C^{m-1} codeword and restricted it to the subset I_2. However, by symmetry we can repeat the same argument, swapping axis 1 and axis 2, and hence for every line ℓ_1 parallel to the first axis it must hold that $g_{\ell_1} \in C|_{I_1}$. Thus, $g \in C|_{I_1} \otimes C|_{I_2} \otimes C^{m-2}$. Since $|I_1|$ and $|I_2|$ are at least $n-d+1$, we can apply Claim 7.2 to the code $C|_{I_1} \otimes C|_{I_2} \otimes C^{m-2}$ to extend g to a unique codeword $f \in C^m$.

We still need to show that $f|_H = g_H$ for every $H \in I$. By definition of C^m we have $f|_H \in C^{m-1}$. There are three cases. If $H \in I_1$, then f agrees with g_H on a subset of H of size $I_2 \times [n]^{m-2}$, because $g_H|_{I_2 \times [n]^{m-2}} = g|_{I_2 \times [n]^{m-2}} = f|_{I_2 \times [n]^{m-2}}$. Similarly, if $H \in I_2$, then f agrees with g_H on a subset of size $I_1 \times [n]^{m-2}$, and if $H \in I \setminus (I_1 \cup I_2)$, then f agrees with g_H on a subset of size $I_1 \times I_2 \times [n]^{m-3}$. In all 3 of the cases, since $|I_1|$ and $|I_2|$ are at least $n-d+1$, by Claim 7.2 there is a unique codeword $w \in C^{m-1}$ that equals $f|_H$ (or g_H) on that subset of H. But $f|_H$ is in C^{m-1}, so by the uniqueness of the extension it follows that $f|_H = g_H$. □

8.2 Step 2: the structure of G

We will now show that every edge (H, H') in G is adjacent to a vertex of large degree. The proof uses the structure of C^m to show that if two planes disagree on a point, they must disagree on many points, and these points have a certain structure. Using the structure of these points, we find $(m-2)d$ planes that intersect $H \cap H'$ on at least one point that g_H and $g_{H'}$ disagree, and therefore each of these new planes must be adjacent to at least one of H and H'.

Lemma 8.3. *If $(H, H') \in E$, then $\deg(H) + \deg(H') \geq (m-2)d$.*

A similar lemma appears in [BS06], but the graph they consider is different from ours.

Proof. Without loss of generality assume that $H = (1, i)$ and $H' = (2, j)$. Fix $k \in \{3, \ldots, m\}$. Let I_k be the set of l's such that the plane (k, l) is not adjacent to both H and H'. Suppose $|I_k| \geq n - d + 1$. Then $g_H|_{I_k \times [n]^{m-3}} = g_{H'}|_{I_k \times [n]^{m-3}}$. Since $|I_k| \geq n - d + 1$, by Claim 7.2 $g_H|_{I_k \times [n]^{m-3}}$ can be extended to a unique $w \in C^{m-2}$, and so $w = g_H|_{H \cap H'}$. Similarly, $g_{H'}|_{I_k \times [n]^{m-3}}$ can be extended to a unique $v \in C^{m-2}$, and so $v = g_{H'}|_{H \cap H'}$. However, since both $g_H|_{H \cap H'}$ and $g_{H'}|_{H \cap H'}$ agree on $I_k \times [n]^{m-3}$, the uniqueness of the extension implies that they are equal, contradicting the fact that (H, H') is an edge in the graph. Therefore, $|I_k| \leq n - d$ for every k. This means that for a fixed k, there are at least d planes (k, l) such that g_H and $g_{H'}$ disagree on the intersection of all 3 planes. Since g_H and $g_{H'}$ disagree, $g_{(k,l)}$ can agree with at most one of them, so at least one of $(H, (k, l))$ and $(H', (k, l))$ is an edge. This holds for at least d planes for every k, which is a total of $(m-2)d$ planes. Therefore, $\deg(H) + \deg(H') \geq (m-2)d$. □

Thus, for every edge (H, H') one of H and H' has degree $\geq (m-2)d/2$, so we deduce the following corollary.

Corollary 8.4 (Vertex Cover)**.** *The set L of vertices with degree $\geq (m-2)d/2$ is a vertex cover.*

8.3 Step 3: relating the expected local distance to the vertex cover

We now relate the set of vertices of large degree to the expected local view distance of the test $\mathcal{H}$. The main idea is to put the expression for $\rho(M)$ into a particular form, and then apply the triangle inequality to express $\rho(M)$ as a sum over edges in the graph. Using a simple relation between $|L|$ and $|E|$, the lemma follows.

Lemma 8.5. *Let L be the set of vertices with large degree. Then $\rho(M) \geq \frac{m-2}{4(m-1)} \frac{d^{m-1}}{n^{m-1}} \frac{|L|}{nm}$.*

Proof. By definition, $\rho(M) = \frac{1}{n^m m} \sum_H \Delta(M|_H, g_H)$. For any $H = (b, i)$,

$$\Delta(M|_H, g_H) = \frac{1}{m-1} \sum_{c \in [m] \setminus \{b\}} \sum_{j \in [n]} \Delta|_{H \cap (c,j)}(M, g_H) = \frac{1}{m-1} \sum_{H' : H \cap H' \neq \emptyset} \Delta|_{H \cap H'}(M, g_H) \ .$$

This is because for any point $p \in H$ and for any axis $c \neq b$, the point p is in the intersection $H \cap (c, j)$ for exactly one j. Therefore,

$$\rho(M) = \frac{1}{n^m m} \sum_H \Delta(M|_H, g_H) = \frac{1}{n^m m} \sum_H \frac{1}{m-1} \sum_{H' : H \cap H' \neq \emptyset} \Delta|_{H \cap H'}(M, g_H)$$

Every pair (H, H') with $H \cap H' \neq \emptyset$ appears exactly twice in the sum, contributing $\Delta|_{H \cap H'}(M, g_H)$ and $\Delta|_{H \cap H'}(M, g_{H'})$ to the sum. Therefore,

$$\begin{aligned}
\rho(M) &= \frac{1}{n^m m(m-1)} \sum_{(H,H') : H \cap H' \neq \emptyset} \Delta|_{H \cap H'}(M, g_H) + \Delta|_{H \cap H'}(M, g_{H'}) \\
&\geq \frac{1}{n^m m(m-1)} \sum_{(H,H') : H \cap H' \neq \emptyset} \Delta|_{H \cap H'}(g_H, g_{H'}) = \frac{1}{n^m m(m-1)} \sum_{(H,H') \in E} \Delta|_{H \cap H'}(g_H, g_{H'}) \ .
\end{aligned}$$

as $(H, H') \notin E \implies \Delta|_{H \cap H'}(g_H, g_{H'}) = 0$ by definition. Fix $(H, H') \in E$. The local codewords g_H and $g_{H'}$ are both in C^{m-1}, so $g_H|_{H \cap H'}$ and $g_{H'}|_{H \cap H'}$ are both C^{m-2} codewords. In particular, since $\Delta|_{H \cap H'}(g_H, g_{H'}) > 0$, they are *distinct* codewords, and so $\Delta|_{H \cap H'}(g_H, g_{H'}) \geq d^{m-2}$. Therefore,

$$\rho(M) \geq \frac{1}{n^m m(m-1)} \sum_{(H,H') \in E} \Delta|_{H \cap H'}(g_H, g_{H'}) \geq \frac{|E| \, d^{m-2}}{n^m m(m-1)} \ .$$

Since L is the set of vertices of degree $\geq (m-2)d/2$,

$$2\,|E| = \sum_H \deg(H) \geq \sum_{H \in L} \deg(H) \geq |L| \, \frac{(m-2)d}{2} \implies |E| \geq |L| \, \frac{(m-2)d}{4} \ .$$

Thus,

$$\rho(M) \geq \frac{|E| \, d^{m-2}}{n^m m(m-1)} \geq \frac{(m-2)\,|L|\,d^{m-1}}{4n^m m(m-1)} = \frac{(m-2)}{4(m-1)} \frac{d^{m-1}}{n^{m-1}} \frac{|L|}{nm} \ .$$ □

8.4 Putting things together

We are now ready to prove Theorem 2. The result follows from straightforward applications of the previous steps.

Proof of Theorem 2. If $|L| \geq (m-1)d$, then by Lemma 8.5 we have

$$\rho(M) \geq \frac{(m-2)}{4(m-1)} \frac{d^{m-1}}{n^{m-1}} \frac{|L|}{nm} \geq \frac{(m-2)}{4(m-1)} \frac{d^{m-1}}{n^{m-1}} \frac{(m-1)d}{nm} = \frac{m-2}{4m} \frac{d^m}{n^m} \geq \frac{m-2}{4m} \frac{d^m}{n^m} \delta(M) \ ,$$

where the last inequality holds because $\delta(M) \leq 1$. Therefore, assume that $|L| < (m-1)d$. For every f in C^m, using triangle inequality we have

$$\delta(M) \leq \delta(M, f) = \frac{1}{nm} \sum_H \delta|_H(M, f) \leq \frac{1}{nm} \sum_H \delta|_H(M, g_H) + \frac{1}{nm} \sum_H \delta|_H(g_H, f) \ .$$

Recalling that $\rho(M) = \frac{1}{nm} \sum_H \delta|_H(M, g_H)$ we get that

$$\delta(M) \le \rho(M) + \frac{1}{nm} \sum_H \delta|_H(g_H, f) \ .$$

Since L is a vertex cover, the set $\overline{L} = V \setminus L$ is an independent set. Since $|L| < (m-1)d$, $\left|\overline{L}\right| > nm - (m-1)d = (m-1)(n-d) + n$. By Lemma 8.2, $\exists f^* \in C^m$ such that $f^*|_H = g_H$ for every $H \in \overline{L}$. Thus,

$$\delta(M) \le \rho(M) + \frac{1}{nm} \sum_H \delta|_H(g_H, f^*) = \rho(M) + \frac{1}{nm} \sum_{H \in L} \delta|_H(g_H, f^*) \le \rho(M) + \frac{|L|}{nm} \ .$$

By Lemma 8.5, $\rho(M) \ge \frac{(m-2)}{4(m-1)} \frac{d^{m-1}}{n^{m-1}} \frac{|L|}{nm}$. Therefore, $\frac{|L|}{nm} \le \frac{4(m-1)n^{m-1}}{(m-2)d^{m-1}} \rho(M)$ and so

$$\delta(M) \le \rho(M) + \frac{|L|}{nm} \le \rho(M) \left(1 + \frac{4(m-1)n^{m-1}}{(m-2)d^{m-1}}\right) \implies \rho(M) \ge \frac{1}{1 + \frac{4(m-1)n^{m-1}}{(m-2)d^{m-1}}} \delta(M) \ .$$

Thus, $\forall M$, $\rho(M) \ge \alpha\delta(M)$, for $\alpha = \min\left(\frac{1}{1+\frac{4(m-1)n^{m-1}}{(m-2)d^{m-1}}}, \frac{m-2}{4m}\frac{d^m}{n^m}\right)$. Since $m \ge 3$, we have that $\frac{1}{1+\frac{4(m-1)n^{m-1}}{(m-2)d^{m-1}}} \ge \frac{1}{1+8\frac{n^{m-1}}{d^{m-1}}} \ge \frac{d^{m-1}}{9n^{m-1}}$ and $\frac{m-2}{4m}\frac{d^m}{n^m} \ge \frac{1}{12}\frac{d^m}{n^m}$. Therefore, $\alpha \ge \min(\frac{d^{m-1}}{9n^{m-1}}, \frac{1}{12}\frac{d^m}{n^m}) = \frac{1}{12}\frac{d^m}{n^m}$.

$\square$

Acknowledgements

The authors thank Eli Ben-Sasson, Oded Goldreich, and Madhu Sudan for helpful discussions, and the anonymous referees for their very useful comments. This work was supported in part by the Center for Long-Term Cybersecurity at UC Berkeley.

A Other Results

Here we will prove other results that are incomparable to Theorem 2.

We have already shown in Theorem 2 that $\mathcal{H}$ is robust for $\alpha \geq \frac{1}{12}\left(\frac{d}{n}\right)^m$. Most of the proof was dedicated to analyzing the test when the set of large degree vertices, L, was less than $(m-1)d$. In this same regime, we can prove an incomparable value for α. Specifically, we can show that for every M such that $|L| < (m-1)d$ it holds that $\rho(M) \geq \frac{1}{m+c} \cdot \delta(M)$, where c is a constant.

Claim A.1. *If $|L| < (m-1)d$, then $\rho(M) \geq \frac{1}{m+c} \cdot \delta(M)$, for $c = 32/9$. Combining with Theorem 2, this implies that $\rho(M) \geq \max\left(\frac{1}{m+c}, \frac{1}{12}\left(\frac{d}{n}\right)^m\right) \cdot \delta(M)$ when $|L| < (m-1)d$.*

Proof. Let I be the set of planes that are not in L. By the assumption $|L| < (m-1)d$, we have $|I| > (m-1)(n-d)+n$, and thus, by Lemma 8.2 there exists $f \in C^m$ such that $f|_H = g_H$ for all $H \in I$.

Let $K = \{p : \forall H \in I, p \notin H\}$ be the set of points that are not contained in any plane in I. Writing $I = \cup_{b=1}^m I_b$, where I_b is the set of planes (b,i) that are in I, it is clear that we can rewrite K as $K = \{p : p_b \notin I_b \ \forall b \in [m]\}$. Therefore,

$$|K| = \prod_{b=1}^m (n - |I_b|) \leq \left(n - \frac{1}{m}\sum_{b=1}^m |I_b|\right)^m = n^m\left(1 - \frac{1}{nm}\sum_{b=1}^m |I_b|\right)^m = n^m\left(\frac{|L|}{nm}\right)^m .$$

Now, we show that $\delta(M,f) \leq (m+c)\cdot\rho(M)$. We start by writing $\delta(M,f)$ as follows.

$$\delta(M,f) = \frac{1}{n^m}\left|\{p : M(p) \neq f(p)\}\right| = \frac{1}{n^m}\left|\{p \in K : M(p) \neq f(p)\}\right| + \frac{1}{n^m}\left|\{p \notin K : M(p) \neq f(p)\}\right| .$$

The first term is upper bounded by $\frac{|K|}{n^m}$, and so it is at most $\left(\frac{|L|}{nm}\right)^m$. In order to bound the second term, note that for all $p \notin K$ there exists a plane $H_p \in I$ such that $p \in H_p$, and thus, $f(p) = g_{H_p}(p)$. Therefore,

$$\begin{aligned}
\frac{1}{n^m}\left|\{p \notin K : M(p) \neq f(p)\}\right| &= \frac{1}{n^m}\left|\{p \notin K : M(p) \neq g_{H_p}(p)\}\right| \\
&\leq \frac{1}{n^m}\left|\{p \in [n]^m : M(p) \neq g_{H_p}(p)\}\right| \\
&\leq \frac{1}{n^m}\sum_{p\in[n]^m}\left|\{H : p \in H, M(p) \neq g_H(p)\}\right| \\
&= m \cdot \rho(M) .
\end{aligned}$$

This implies that

$$\delta(M,f) \leq \left(\frac{|L|}{nm}\right)^m + m\cdot\rho(M)$$

Next, using the bound $|L| < (m-1)d$ in the assumption of the claim, as well as the bound $\frac{|L|}{nm} \leq \rho(M)\cdot\frac{4(m-1)}{m-2}\cdot\frac{n^{m-1}}{d^{m-1}}$ from Lemma 8.5, we get that

$$\begin{aligned}
\delta(M,f) &\leq \left(\frac{(m-1)d}{nm}\right)^{m-1}\cdot\left(\rho(M)\cdot\frac{4(m-1)}{m-2}\cdot\frac{n^{m-1}}{d^{m-1}}\right) + m\cdot\rho(M) \\
&= \left(\left(1-\frac{1}{m}\right)^m\cdot\frac{4m}{m-2} + m\right)\cdot\rho(M) .
\end{aligned}$$

For $m \geq 3$ we get that $\delta(M) \leq (m + 32/9)\rho(M)$, as required. □

Remark A.2. *In fact, by a slightly modified argument (writing $\rho(M)$ as the sum over the intersections of k planes) we can prove that for $|L| < (m-1)d$ it holds that $\delta(M) \leq \rho(M)\left(k + c_k \frac{n^{m-k}}{d^{m-k}}\right)$, where c_k is a constant for a fixed $k \in [m]$. The proof of Theorem 2 used $k = 1$.*

We can also show that when $|L| < (m-1)d$, we get a robustness of $\alpha = 1$ plus an additive term of d/n. Note that d is the distance of the code, so when $d = O(n)$, the additive term is not small.

Claim A.3. *If $|L| < (m-1)d$, then $\delta(M) \leq \rho(M) + d/n$.*

Proof. In the proof of Theorem 2, we showed that if $|L| < (m-1)d$, then

$$\delta(M) \leq \rho(M) + \frac{|L|}{nm} \leq \rho(M) + \frac{(m-1)d}{nm} \leq \rho(M) + \frac{d}{n} \ .$$ □

Next, we observe that if C is the Reed–Solomon code (or any code with a similar interpolation property), then the above holds without the constraint on $|L|$.

Claim A.4. *If C is the Reed–Solomon code, then $\delta(M) \leq \rho(M) + d/n$ unconditionally.*

Proof. Define $v_b = \sum_{H=(b,i)} \Delta|_H(M, g_H)$, and without loss of generality assume that $v_1 \leq v_2 \leq \cdots \leq v_m$. Observe that

$$\rho(M) = \frac{1}{n^m m} \sum_{b=1}^{m} v_b \ .$$

Let S be any subset of $(1,i)$ planes of size exactly $n-d+1$. By m-variate polynomial interpolation, there exists f in C^m such that $f|_H = g_H$ for every H in S. Therefore,

$$\begin{aligned}
\delta(M) \leq \delta(M,f) &= \frac{1}{n^m} \sum_{H=(1,i)} \Delta|_H(M,f) \leq \frac{1}{n^m} \sum_{H \in S} \Delta|_H(M,f) + \frac{1}{n^m}\left(n - |S|\right) n^{m-1} \\
&= \frac{1}{n^m} \sum_{H \in S} \Delta|_H(M, g_H) + \frac{d-1}{n} \leq \frac{1}{n^m} v_1 + \frac{d-1}{n} = \frac{1}{n^m m}(m v_1) + \frac{d-1}{n} \\
&\leq \frac{1}{n^m m} \sum_{b=1}^{m} v_b + \frac{d-1}{n} \leq \rho(M) + \frac{d}{n} \ .
\end{aligned}$$ □

B Robustness vs. Soundness

In Theorem 2, we study the robustness of the test $\mathcal{H}$. We now compare the robustness of $\mathcal{H}$ to the soundness of $\mathcal{H}$, and show a tight result for the soundness of $\mathcal{H}$.

Robustness is a stronger guarantee than soundness, as $\Pr[\mathcal{H}^M \text{ rejects}] \geq \rho(M)$. Thus, if we show that $\rho(M) \geq \alpha\delta(M)$, then we can deduce that $\Pr[\mathcal{H}^M \text{ rejects}] \geq \alpha\delta(M)$, which upper bounds the soundness error. However, the converse is not true.

In fact, unlike for robustness, we can prove a tight result on the soundness of $\mathcal{H}$, stated below. Note that the query complexity of $\mathcal{H}$ is n^{m-1}, which is $N^{1-\frac{1}{m}}$ where $N := n^m$ is the block length of C^m, the code being tested.

Claim B.1. *Let $\epsilon_0 = ((n-d)(m-1)+n)/nm$. If $\Pr[\mathcal{H}^M \text{ accepts}] = \epsilon > \epsilon_0$, then $\delta(M) \leq (1-\epsilon)^m$.*

Proof. Observe that $\Pr[\mathcal{H}^M \text{ accepts}] = \epsilon = \frac{1}{nm} \cdot |I|$, where I is the set of planes H such that $M|_H \in C^{m-1}$. The condition $\epsilon > \epsilon_0$ implies that $|I| > (n-d)(m-1)+n$, so by Lemma 8.2, there exists an f in C^m such that $f|_H = M|_H$. Let $K = \{p : p \notin H \ \forall H \in I\}$. By the same logic as in the proof of Claim A.1,

$$\delta(M) \leq \delta(M, f) \leq \frac{|K|}{n^m} \leq \left(1 - \frac{|I|}{nm}\right)^m = (1-\epsilon)^m \ . \qquad \square$$

We remark that the above claim is tight, as there are functions where $\Pr[\mathcal{H}^M \text{ accepts}] = \epsilon$ and $\delta(M) = (1-\epsilon)^m$, and furthermore the threshold ϵ_0 is necessary, as we now explain.

- *Necessity of ϵ.* Fix ϵ, and let S be a subset of $[n]$ of size $(1-\epsilon)n$. Let M be a codeword of C^m on the set $[n]^m \setminus S^m$, and a random function on the set S^m. It is clear that $\delta(M) = |S|^m / n^m = (1-\epsilon)^m$, and that the test $\mathcal{H}$ will accept if it reads the plane (b, i) for any $b \in [m]$ and $i \notin S$. Furthermore, the test $\mathcal{H}$ will reject when it reads a plane (b, i) for any $b \in [m]$ and $i \in S$. Therefore, $\Pr[\mathcal{H}^M \text{ accepts}] = (1 - |S|)/n = \epsilon$.

- *Necessity of ϵ_0.* The threshold ϵ_0 is necessary, as we can get an acceptance probability of $1/m$ "for free" by setting $M|_{(1,i)}$ to be a random C^{m-1} codeword. Then, the test $\mathcal{H}$ will accept with probability 1 when it reads a $(1, i)$ plane, and will reject with high probability when it reads a (b, i) plane for $b \neq 1$. However, M will be very far from C^m, so having an acceptance probability of $1/m$ says little about $\delta(M)$.

References

[AKKLR05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. "Testing Reed-Muller codes". In: *IEEE Transactions on Information Theory* 51.11 (2005), pp. 4032–4039.

[AS03] Sanjeev Arora and Madhu Sudan. "Improved Low-Degree Testing and its Applications". In: *Combinatorica* 23.3 (2003). Preliminary version appeared in STOC '97., pp. 365–426.

[AS98] Sanjeev Arora and Shmuel Safra. "Probabilistic checking of proofs: a new characterization of NP". In: *Journal of the ACM* 45.1 (1998). Preliminary version in FOCS '92., pp. 70–122.

[BCGT13] Eli Ben-Sasson, Alessandro Chiesa, Daniel Genkin, and Eran Tromer. "On the Concrete Efficiency of Probabilistically-Checkable Proofs". In: *Proceedings of the 45th ACM Symposium on the Theory of Computing*. STOC '13. 2013, pp. 585–594.

[BCHKS96] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos A. Kiwi, and Madhu Sudan. "Linearity testing in characteristic two". In: *IEEE Transactions on Information Theory* 42.6 (1996), pp. 1781–1795.

[BCLR08] Michael Ben-Or, Don Coppersmith, Mike Luby, and Ronitt Rubinfeld. "Non-abelian homomorphism testing, and distributions close to their self-convolutions". In: *Random Structures and Algorithms* 32.1 (2008), pp. 49–70.

[BDN17] Amey Bhangale, Irit Dinur, and Inbal Livni Navon. "Cube vs. Cube Low Degree Test". In: *Proceedings of the 8th Innovations in Theoretical Computer Science Conference*. ITCS '17. 2017.

[BFL91] László Babai, Lance Fortnow, and Carsten Lund. "Non-Deterministic Exponential Time has Two-Prover Interactive Protocols". In: *Computational Complexity* 1 (1991). Preliminary version appeared in FOCS '90., pp. 3–40.

[BFLS91] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. "Checking computations in polylogarithmic time". In: *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*. STOC '91. 1991, pp. 21–32.

[BGHSV06] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan. "Robust PCPs of Proximity, Shorter PCPs, and Applications to Coding". In: *SIAM Journal on Computing* 36.4 (2006), pp. 889–974.

[BKSSZ10] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. "Optimal Testing of Reed-Muller Codes". In: *Property Testing - Current Research*. 2010, pp. 269–275.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. "Self-Testing/Correcting with Applications to Numerical Problems". In: *Journal of Computer and System Sciences* 47.3 (1993), pp. 549–595.

[BS06] Eli Ben-Sasson and Madhu Sudan. "Robust locally testable codes and products of codes". In: *Random Structures and Algorithms* 28.4 (2006), pp. 387–402.

[BS08] Eli Ben-Sasson and Madhu Sudan. "Short PCPs with Polylog Query Complexity". In: *SIAM Journal on Computing* 38.2 (2008). Preliminary version appeared in STOC '05., pp. 551–607.

[BSVW03] Eli Ben-Sasson, Madhu Sudan, Salil Vadhan, and Avi Wigderson. "Randomness-efficient low degree tests and short PCPs via epsilon-biased sets". In: *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*. STOC '03. 2003, pp. 612–621.

[BV08] Eli Ben-Sasson and Michael Viderman. "Tensor Products of Weakly Smooth Codes Are Robust". In: *Proceedings of the 11th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, and of the 12th International Workshop on Randomization and Computation*. APPROX-RANDOM '08. 2008, pp. 290–302.

[BV15] Eli Ben-Sasson and Michael Viderman. "Composition of semi-LTCs by two-wise tensor products". In: *Computational Complexity* 24.3 (2015), pp. 601–643.

[CR05] Don Coppersmith and Atri Rudra. *On the Robust Testability of Product of Codes*. ECCC TR05-104. 2005.

[DR04] Irit Dinur and Omer Reingold. "Assignment Testers: Towards a Combinatorial Proof of the PCP Theorem". In: *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*. FOCS '04. 2004, pp. 155–164.

[DSW06] Irit Dinur, Madhu Sudan, and Avi Wigderson. "Robust Local Testability of Tensor Products of LDPC Codes". In: *Proceedings of the 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, and of the 10th International Workshop on Randomization and Computation*. APPROX-RANDOM '06. 2006, pp. 304–315.

[GM12] Oded Goldreich and Or Meir. "The tensor product of two good codes is not necessarily robustly testable". In: *Information Processing Letters* 112.8-9 (2012), pp. 351–355.

[GS06] Oded Goldreich and Madhu Sudan. "Locally testable codes and PCPs of almost-linear length". In: *Journal of the ACM* 53 (4 2006). Preliminary version in STOC '02., pp. 558–655.

[Hås01] Johan Håstad. "Some optimal inapproximability results". In: *Journal of the ACM* 48.4 (2001), pp. 798–859.

[KMRS16] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. "High-rate locally-correctable and locally-testable codes with sub-polynomial query complexity". In: *Proceedings of the 48th ACM Symposium on the Theory of Computing*. STOC '16. 2016, pp. 202–215.

[KST54] T. Kővári, V. T. Sós, and P. Turán. "On a Problem of Zarankiewicz". In: *Colloquium Mathematicae* 3 (1954), pp. 50–57.

[MR08] Dana Moshkovitz and Ran Raz. "Sub-Constant Error Low Degree Test of Almost-Linear Size". In: *SIAM Journal on Computing* 38.1 (2008). Preliminary version in STOC '06., pp. 140–180.

[Mei09] Or Meir. "Combinatorial Construction of Locally Testable Codes". In: *SIAM Journal on Computing* 39.2 (2009). Preliminary version appeared in STOC '08., pp. 491–544.

[PS94] Alexander Polishchuk and Daniel A. Spielman. "Nearly-linear size holographic proofs". In: *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*. STOC '94. 1994, pp. 194–203.

[RS96] Ronitt Rubinfeld and Madhu Sudan. "Robust Characterizations of Polynomials with Applications to Program Testing". In: *SIAM Journal on Computing* 25.2 (1996), pp. 252–271.

[RS97] Ran Raz and Shmuel Safra. "A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP". In: *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*. STOC '97. 1997, pp. 475–484.

[Val05] Paul Valiant. "The Tensor Product of Two Codes Is Not Necessarily Robustly Testable". In: *Proceedings of the 8th International Workshop on Approximation, Randomization, and Combinatorial Optimization*. APPROX-RANDOM '05. 2005, pp. 472–481.

[Vid12] Michael Viderman. "A Combination of Testability and Decodability by Tensor Products". In: *Proceedings of the 15th International Workshop on Approximation, Randomization, and Combinatorial Optimization*. APPROX-RANDOM '12. 2012, pp. 651–662.

[Vid13] Michael Viderman. "Strong LTCs with Inverse Poly-Log Rate and Constant Soundness". In: *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science*. FOCS '13. 2013, pp. 330–339.

[WE63] Jack Keil Wolf and Bernard Elspas. "Error-locating codes - A new concept in error control". In: *IEEE Transactions on Information Theory* 9.2 (1963), pp. 113–117.

[Wol65] Jack Keil Wolf. "On codes derivable from the tensor product of check matrices". In: *IEEE Transactions on Information Theory* 11.2 (1965), pp. 281–284.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il